Cluster Structures, Orientifolds of Brane Tilings and Higher Laminations

Valdo Tatitscheff

Thèse

INSTITUT DE RECHERCHE MATHÉMATIQUE AVANCÉE

UMR 7501

Strasbourg

présentée pour obtenir le grade de docteur de
l'Université de Strasbourg
Spécialité MATHÉMATIQUES

Valdo Tatitscheff

Cluster Structures, Orientifolds of Brane Tilings and Higher Laminations

Soutenue le 16 Septembre 2022
devant la commission d'examen

Vladimir V. Fock, directeur de thèse
Alexander B. Goncharov, rapporteur
Iñaki García-Etxebarria, rapporteur
Cyril Closset, examinateur
Olivier Guichard, examinateur
Yang-Hui He, examinateur
Andrew Neitzke, examinateur
Nikita Nekrasov, examinateur

irma.math.unistra.fr

IRMA
Institut de Recherche
Mathématique Avancée

cnrs

Université
de Strasbourg

Contents

Introduction française **9**
Fondations . . . 9
Avant 1905 . . . 9
Relativité restreinte . . . 13
Physique pré-quantique . . . 15
Relativité générale et cosmologie . . . 17
Théorie quantique . . . 20
Electrodynamique quantique . . . 23
Interactions nucléaires faibles et fortes . . . 26
Modèle Standard de la physique des particules . . . 32
Mathématique physique contemporaine . . . 35
Théories de jauge et géométrie . . . 35
Supersymétrie . . . 36
Théorie des cordes . . . 39
Information quantique, holographie, et correspondance AdS–CFT . . . 43
Espaces de Teichmüller et généralisations . . . 45

Sommaire narratif **49**

Remerciements – Acknowledgments **51**

Introduction **53**

I Cluster higher Teichmüller theory **57**

1 Cluster algebras **61**
1.1 Total positivity . . . 62
1.1.1 The totally positive Grassmanian $\mathrm{Gr}^+_{2,m}$ . . . 62
1.1.2 Positivity in $\mathrm{GL}_n(\mathbb{C})$ and $\mathrm{SL}_n(\mathbb{C})$ . . . 64
1.2 Mutations of quivers and matrices . . . 65
1.2.1 The mutation class of a quiver . . . 67
1.2.2 Skew-symmetric and skew-symmetrizable matrices . . . 67
1.2.3 Quivers with potential and bipartite graphs on oriented surfaces . . . 69
1.3 Cluster algebras . . . 72
1.3.1 Rank-1 cluster algebras . . . 73
1.3.2 Rank-2 cluster algebras without frozen variables . . . 74
1.3.3 A rank-2 cluster algebra with frozen variables . . . 75
1.4 The Laurent phenomenon . . . 77
1.5 Three amusing consequences of the Laurent phenomenon . . . 78
1.5.1 The Somos-4 sequence . . . 78
1.5.2 Euler's counterexample of the primality of Fermat numbers . . . 78
1.5.3 Markov numbers . . . 79
1.6 Y-patterns . . . 80

2 Teichmüller theory **83**
2.1 Generalities on Teichmüller spaces . . . 83
2.1.1 Hyperbolic geometry . . . 83
2.1.2 Generalities on Teichmüller spaces . . . 85
2.2 Ciliated surfaces and triangulations . . . 88
2.2.1 Definitions . . . 88
2.2.2 Combinatorial type of a triangulation . . . 90
2.2.3 Mapping class groups of ciliated surfaces . . . 92
2.3 Teichmüller $\mathcal{X}$-space (Teichmüller space with holes) . . . 93
2.3.1 Parametrization . . . 94
2.3.2 Reconstruction . . . 96
2.3.3 Flips, hole lengths and Poisson structure . . . 98
2.4 Teichmüller $\mathcal{A}$-space (decorated Teichmüller space) . . . 99
2.4.1 Parametrization . . . 100
2.4.2 Reconstruction . . . 102
2.4.3 Flips and the closed 2-form . . . 105

3 Laminations on surfaces **107**
3.1 Thurston's measured laminations . . . 107
3.1.1 Laminations from three-dimensional hyperbolic spaces . . . 107
3.1.2 Laminations from curve systems . . . 109
3.1.3 Train tracks . . . 111
3.1.4 Measured laminations from train tracks . . . 112
3.1.5 Laminations and Thurston's compactification of Teichmüller spaces . . . 113
3.2 Rational laminations on ciliated surfaces . . . 114
3.2.1 Singular and Borel–Moore homologies . . . 114
3.2.2 Rational $\mathcal{A}$-laminations . . . 115
3.2.3 Rational $\mathcal{X}$-laminations . . . 117
3.3 Pairings . . . 120
3.3.1 Additive and intersection pairings . . . 120
3.3.2 Multiplicative pairings . . . 122
3.3.3 Duality conjectures (I) . . . 123

4 Higher Teichmüller theory **125**
4.1 Positive algebraic geometry . . . 126
4.1.1 Generalities . . . 126
4.1.2 Points over a semifield and tropical compactifications . . . 127
4.2 Cluster ensembles . . . 128
4.2.1 Cluster $\mathcal{X}$- and $\mathcal{A}$-tori . . . 128
4.2.2 Cluster transformations . . . 130
4.2.3 Properties of cluster ensembles . . . 131
4.2.4 Duality conjectures (II) . . . 132
4.3 Hitchin components and cluster higher Teichmüller spaces . . . 133
4.3.1 Hitchin components . . . 133
4.3.2 Total positivity and positive configurations of flags . . . 135
4.3.3 Framed local systems and Teichmüller $\mathcal{X}$-spaces . . . 136
4.3.4 Decorated local systems and Teichmüller $\mathcal{A}$-spaces . . . 140
4.4 Other higher Teichmüller spaces . . . 143
4.4.1 Spaces of maximal representations . . . 143
4.4.2 Θ-positivity . . . 144

II Supersymmetry, string theory and holography **149**

5 Supersymmetry in four dimensions **151**
5.1 General definitions and properties . . . 153
5.2 Renormalization and anomalies. . . . 156

5.3 Perturbative spontaneous supersymmetry breaking . . . 159
5.4 $\mathcal{N} = 1$ quantum dynamics: SQCD and Seiberg duality. . . . 161
5.5 Dynamical supersymmetry breaking . . . 164
5.5.1 Supersymmetry-breaking criteria . . . 166
5.5.2 Two DSB Models . . . 167
5.6 $\mathcal{N} = 2$ quantum dynamics: Seiberg–Witten theory . . . 168
5.7 $\mathcal{N} = 4$ super Yang–Mills . . . 171

6 String theory 173
6.1 Perturbative type II string theories . . . 174
6.2 Branes . . . 177
6.3 String dualities and M-theory . . . 179
6.4 Supersymmetric quantum field theories from brane configurations . . . 182
6.4.1 Four-dimensional theories with eight supercharges . . . 184
6.4.2 Five-dimensional theories with eight supercharges . . . 186
6.4.3 Four dimensional theories with four supercharges . . . 188

7 String geometry 191
7.1 Compactifications without fluxes . . . 191
7.2 Flux compactifications . . . 193
7.3 Branes at abelian Calabi–Yau orbifolds of $\mathbb{C}^3$ . . . 195
7.3.1 Regular D3-branes . . . 195
7.3.2 Fractional D3-branes . . . 197
7.4 Toric Calabi–Yau singularities . . . 198
7.5 Brane tilings . . . 208
7.5.1 Generalities . . . 208
7.5.2 Forward and inverse algorithms . . . 211
7.5.3 Some properties of brane tilings. . . . 213
7.5.4 Cluster integrable systems . . . 217
7.5.5 Fractional branes revisited . . . 219
7.6 Orientifolds . . . 220
7.6.1 Generalities . . . 220
7.6.2 Orientifolds of brane tilings . . . 221

8 Holography and gauge–gravity dualities 225
8.1 $\mathrm{AdS}_5 \times S^5$ / $\mathcal{N} = 4$ $d = 4$ super Yang–Mills correspondence. . . . 226
8.2 Conical singularities and toric geometry . . . 229
8.3 More general gauge–gravity correspondences . . . 230
8.3.1 Branes at the conifold and the duality cascade . . . 231
8.3.2 Fractional branes of different kinds . . . 233

III Dynamical supersymmetry breaking in string theory : the Octagon 237

9 DSB on branes at toric affine CY3 241
9.1 Introduction . . . 242
9.1.1 Review of previous implementations of DSB models in brane tilings . . . 242
9.1.2 $\mathcal{N} = 2$ Fractional Branes Decay . . . 243
9.1.3 Holes in brane tilings and zig-zag paths . . . 246
9.2 SU(5) Models . . . 247
9.2.1 Fixed Point Orientifolds . . . 248
9.2.2 Fixed Line Orientifolds . . . 252
9.3 $3 - 2$ Models . . . 255
9.3.1 General Features . . . 255
9.3.2 Fixed Point Orientifolds . . . 258
9.3.3 Fixed Line Orientifolds . . . 259
9.3.4 Twin $3 - 2$ models? . . . 261

10 Inverse algorithm and triple diagrams **263**
10.1 Triple diagrams and fast-inverse algorithm . . . 265
10.1.1 Generalities . . . 265
10.1.2 Consistent dimer models from triple diagrams . . . 266
10.2 Implementation of symmetries and substructures . . . 268
10.2.1 The local structure of the hexagonal cluster . . . 268
10.2.2 An example . . . 269
10.2.3 Another example . . . 272
10.3 Reflection symmetries in dimer models . . . 272

11 Dimers orientifolds and anomalies **275**
11.1 Gauge anomalies and dimer models . . . 275
11.1.1 Anomalies and Zig-Zag Paths . . . 276
11.1.2 Examples . . . 277
11.2 Anomaly Cancellation Conditions in Orientifolds . . . 278
11.2.1 Dimers and Orientifolds . . . 279
11.2.2 The Adjacency Matrix of Orientifolded Theories . . . 280
11.2.3 The Homogeneous Problem . . . 282
11.2.4 The Non-Homogeneous Problem . . . 283
11.3 A Zig-Zag Algorithm for Orientifolds . . . 284
11.3.1 Fixed Line Orientifolds . . . 284
11.3.2 Fixed Point Orientifolds . . . 290
11.4 General Criteria for Anomaly-Free Orientifolds . . . 293
11.4.1 Diagonal Line Orientifolds . . . 293
11.4.2 Horizontal/Vertical Line Orientifolds . . . 298
11.4.3 Fixed Point Orientifolds . . . 299
11.5 Conclusions . . . 303

12 The Octagon **305**
12.1 The rise of the Octagon . . . 305
12.2 Analysis of the model . . . 308

IV New orientifolds of brane tilings **311**

13 Dimers in a bottle **315**
13.1 Torus involutions and Orientifolds . . . 315
13.1.1 Orientifold projections and Dimers . . . 315
13.1.2 Torus involutions . . . 317
13.2 Glide Orientifolds . . . 318
13.2.1 Orbifold $\mathbb{C}^2/\mathbb{Z}_2$ . . . 318
13.2.2 More orbifold examples . . . 320
13.2.3 Conifold-like singularities . . . 321
13.2.4 General properties . . . 325
13.3 T-duals of the Glide Orientifold . . . 327
13.3.1 Type IIA picture and the brane tiling . . . 327
13.3.2 The mirror picture . . . 328
13.4 Involutions and Zig-Zag Paths . . . 331
13.4.1 Glide Orientifold from the Toric Diagram . . . 331
13.4.2 Fractional branes . . . 333
13.4.3 Shift Orientifolds . . . 334
13.5 Conclusion . . . 339

V Higher laminations and Hecke TQFTs 341

14 Towards higher laminations 343
14.1 $\mathcal{A}$-laminations as a basis of regular functions . . . 344
14.1.1 Regular functions on moduli stacks of local systems . . . 344
14.1.2 Classical $\mathcal{A}$-laminations for G of type A_1 . . . 345
14.1.3 G-higher $\mathcal{A}$-laminations . . . 348
14.2 Littlewood–Richardson coefficients from hives . . . 349
14.2.1 The saturation conjecture . . . 349
14.2.2 Hives . . . 350
14.2.3 Honeycombs and Horn's conjecture . . . 353
14.3 The Satake correspondence and Hecke algebras . . . 354
14.3.1 Hecke algebras . . . 354
14.3.2 The Satake isomorphism . . . 357
14.4 Wiring diagrams and higher laminations . . . 358
14.4.1 Wiring diagrams on the cylinder . . . 358
14.4.2 The Satake correspondence revisited . . . 359
14.4.3 Higher laminations in triangles from spherical Hecke algebras . . . 360
14.4.4 Classical laminations from the spherical Hecke algebra of SL_2 . . . 362
14.5 Spectral networks . . . 362
14.5.1 Theories of class S . . . 362
14.5.2 BPS states . . . 365
14.5.3 Wilson–'t Hooft operators and laminations . . . 367
14.6 Higher laminations and ramified covers . . . 367
14.6.1 Higher laminations as Lagrangian surfaces in T^*S . . . 367
14.6.2 From integral Lagrangian surfaces to combinatorial higher laminations . . . 370

15 Topological Quantum Field Theories from Hecke algebras 373
15.1 Introduction . . . 373
15.2 Preliminaries . . . 375
15.2.1 Hecke algebras and their standard basis . . . 375
15.2.2 The Kazhdan–Lusztig basis . . . 377
15.2.3 Ciliated surfaces . . . 377
15.3 Definition of the polynomial and first properties . . . 378
15.3.1 On the standard structure constants in Hecke algebras . . . 378
15.3.2 An invariant for surfaces with punctures . . . 379
15.3.3 Gluing surfaces . . . 381
15.3.4 Invariants of punctured surfaces at $q = 1$ . . . 382
15.4 Graphical calculus . . . 383
15.4.1 Structure constants and flag counting . . . 383
15.4.2 Finite higher laminations for triangles . . . 384
15.4.3 Higher laminations for ciliated surfaces . . . 390
15.5 Hecke topological quantum field theory . . . 393
15.5.1 Triangle invariants and gluing revisited . . . 393
15.5.2 Hecke topological quantum field theories . . . 394
15.5.3 Invariants of n-gons . . . 394
15.6 Explicit expression and positivity . . . 395
15.6.1 Schur elements and Wedderburn decomposition . . . 395
15.6.2 Central elements . . . 397
15.6.3 Explicit expression . . . 399
15.6.4 Positivity properties . . . 401
15.7 Computation using Sage and CHEVIE . . . 403
15.7.1 Sage . . . 403
15.7.2 CHEVIE . . . 404
15.8 Positivity for Schur elements . . . 404
15.9 Perspectives . . . 406

16 Symmetric rings and TQFTs **407**
16.1 Introduction 407
16.1.1 Hecke algebras 407
16.1.2 Symmetric rings 407
16.1.3 Topological Quantum Field Theories 409
16.2 TQFTs and symmetric rings 409
16.2.1 From a TQFT to a symmetric ring 409
16.2.2 From a symmetric ring to a TQFT 410
16.2.3 Application to the Hecke ring 411

Index **413**

References **417**

Introduction française

Cette thèse de physique mathématique présente les travaux de recherche que j'ai effectués lors de mon doctorat à l'UFR de mathématiques et d'informatique de l'Université de Strasbourg et l'Institut de Recherche Mathématique Avancée, de septembre 2018 à décembre 2021. Elle s'articule en deux axes principaux : le premier à motivations plus physiques et le second à motivations davantage mathématiques, bien que les domaines concernés soient à l'interface de la théorie des cordes et des mathématiques fondamentales, et si tant est qu'une telle distinction ait un sens. Cette interface m'intéresse tout particulièrement depuis plusieurs années et a motivé ces travaux.

La visée de cette introduction en français est de placer mes travaux de thèse dans leur contexte de recherche. Certains points, bien qu'importants, ne sont mentionnés qu'hâtivement par économie de place, et ceux que je développe le plus me semblent à la fois fondamentaux et propres à être présentés de manière pédagogique : les choix faits ci-dessous sont imprégnés de mes biais. Des recherches supplémentaires et l'accès à de véritables travaux d'histoire des sciences seront assurément indispensables à quiconque souhaite compléter cette ébauche pour soi-même. J'évoque simultanément des découvertes et résultats appartenant aujourd'hui à des domaines scientifiques différents (mathématique, physique, chimie), car il me semble qu'ils ne peuvent être décorrélés dans le cadre de notre récit historique.

Fondations

La communauté scientifique s'accorde aujourd'hui sur l'existence de quatre forces fondamentales qui gouvernent notre univers : la *gravité*, l'*électromagnétisme*, la *force nucléaire faible* et la *force nucléaire forte*. Dans la première partie de cette introduction, nous allons présenter succinctement l'histoire de deux théories physiques parmi les plus prédictives : la *relativité générale* d'Einstein, qui décrit la gravité à l'échelle des corps célestes, et le *modèle standard de la physique des particules*, qui décrit les trois autres interactions fondamentales sus-mentionnées à l'échelle subatomique.

Avant 1905 (l'*Annus Mirabilis* d'Einstein)

- 1685 – *Philosophiae Naturalis Principia Mathematica*, Newton. Cet ouvrage constitue le fondement de la mécanique classique, expose les trois lois dites de Newton ainsi que la *loi universelle de la gravitation* : deux corps ponctuels de masse m_1 et m_2 sont soumis à une force attractive de norme

$$F = -\mathcal{G}\frac{m_1 m_2}{d^2} \ , \tag{1}$$

  où $\mathcal{G}$ est appelée constante gravitationnelle, de valeur mesurée

$$\mathcal{G} = 6.674\ 30(15) \times 10^{-11}\ \mathrm{m^3 kg^{-1} s^{-2}} \ .$$

  Newton en déduit les *lois de Kepler*, auparavant empiriques. Le *Principia* pose les bases du calcul infinitésimal.

- 1690 – *Traité de la Lumière* de Huygens, dans lequel est développée une théorie ondulatoire de la lumière, confirmée et approfondie par Young et Fresnel – entre autres – au 19ème siècle.

- 1733 – du Fay constate que deux objets frottés avec de l'ambre se repoussent, de même que deux objets frottés avec une baguette en verre, alors que les premiers attirent les seconds. Cette découverte sera ensuite interprétée comme l'existence de charges électriques positives et négatives.

- 1799 - 1805 – *Traité de mécanique céleste* de Laplace, qui reformule notamment les lois de Newton dans le formalisme de calcul différentiel qu'il a lui-même introduit. C'est la formulation "moderne" des lois de Newton. Ce traité est également la pierre angulaire du calcul différentiel, qui permet par exemple le développement ultérieur de la géométrie différentielle.

- 1801 – Expérience des fentes de Young. En plaçant une source lumineuse derrière une feuille opaque percée de deux fentes, on observe sur un écran non pas deux taches lumineuses dans le prolongement des fentes, mais une tache centrale entourée de bandes lumineuses de plus en plus fines, caractéristiques des interférences ondulatoires, comme montré en Figure 1. Cette expérience historique prouve donc le comportement ondulatoire de la lumière.

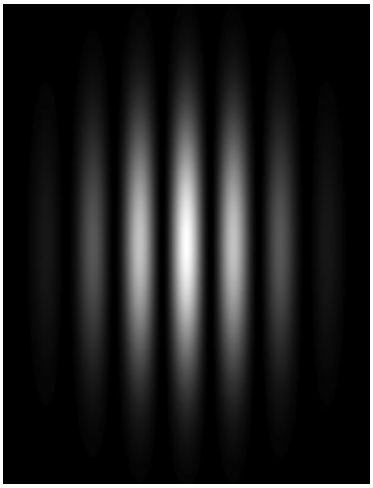


Figure 1: Franges d'interférences de l'expérience des fentes de Young (source : Wikipédia).

- 1824 – *Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance*, Carnot. L'ouvrage contient une formulation préliminaire de la seconde loi de la thermodynamique : "il est impossible de produire de la puissance motrice à moins qu'on ne dispose d'un corps froid et d'un corps chaud".

- 1829 - 1832 – Lobatchevski et Bolyai introduisent la géométrie hyperbolique, une des premières instances de géométrie non-Euclidienne – objet central en mathématique et physique actuelles.

- 1832 – Galois introduit la notion de groupes de permutations et de sous-groupes normaux. La notion de groupe devient par la suite une notion incontournable en mathématique et en physique.

- 1834 – *Loi des gaz parfaits* $PV = nRT$ de Clapeyron, fondement de la thermodynamique des gaz.

- 1838 – Faraday découvre les rayons cathodiques en faisant passer un courant électrique dans un tube de verre sous vide. L'étude de ces mystérieux rayons a été centrale pour les sciences de la deuxième moitié du 19ème siècle.

- 1849 – Expérience de Fizeau entre Montmartre et Suresnes, qui montre que la lumière se propage à vitesse finie : $c \sim 3 \times 10^8$ m/s.

- 1854 – *Über die Hypothesen, welche der Geometrie zu Grunde liegen*, Riemann. Ces travaux constituent les fondements de la géométrie Riemannienne.

- 1865 – *A Dynamical Theory of the Electromagnetic Field* de Maxwell est le point culminant de plusieurs décennies de travaux en électromagnétisme, conduits notamment par Franklin, Volta, Ørsted, Ampère, Poisson, Gauss, Ohm, Faraday, Kirchhoff et Thompson/Kelvin. Les interactions *électromagnétiques* sont décrites de manière unifiée dans la théorie de Maxwell, sous la forme de 20 équations. Ces dernières sont reformulées en 4 équations par Heaviside en 1884 : les champs

électrique $\mathbf{E}$ et magnétique $\mathbf{B}$ satisfont, dans le vide et en présence d'une densité de charge ρ et d'une densité de courant $\mathbf{j}$, les équations couplées suivantes:

$$\begin{cases} \nabla\mathbf{E} = \rho/\epsilon_0 \\ \nabla\mathbf{B} = 0 \\ \nabla \wedge \mathbf{E} = -\partial\mathbf{B}/\partial t \\ \nabla \wedge \mathbf{B} = \mu_0\mathbf{j} + c^{-2}\partial\mathbf{E}/\partial t \end{cases} . \tag{2}$$

Dans ces équations, c est une vitesse, ϵ_0 la permittivité diélectrique du vide et μ_0 la perméabilité magnétique du vide. Ces constantes satisfont $\epsilon_0\mu_0 = c^{-2}$, avec $\mu_0 \sim 10^{-7}$ T$\cdot$m/A. Dans le vide, les équations de Maxwell décrivent la propagation d'ondes à la vitesse c, calculée comme très proche de la vitesse de la lumière. Ceci conduit Maxwell à faire l'hypothèse que la lumière est une onde électromagnétique, en accord avec l'expérience de Young.

- 1865 – Clausius introduit le terme d'*entropie*, grâce auquel le second principe de la thermodynamique peut être reformulé comme : l'entropie d'un système fermé ne peut qu'augmenter.

- 1869 – Mendeleïev propose son *tableau périodique des éléments*. Les atomes peuvent être classés par masse atomique croissante ; les propriétés chimiques de ces atomes satisfont alors une périodicité qui permet de les ranger en un tableau, dans lequel les éléments d'une même colonne ont des caractéristiques chimiques semblables.

- 1872 – *Programme d'Erlangen* de Klein qui a une influence considérable sur les développements futurs de la géométrie.

- 1873 - 1874 – Lie pose les bases de l'étude des groupes continus, c'est-à-dire des groupes et algèbres de Lie. Les algèbres de Lie de dimension finie ont une riche théorie des représentations, et interviennent notamment de façon centrale en physique des particules.

- 1879 – Hall découvre *l'effet Hall* : une plaque métallique parcourue par un courant et soumise à un champ magnétique transverse exhibe une différence de potentiel à ses bords dans la troisième direction. L'effet Hall met notamment en évidence que le courant électrique est transporté par des charges négatives (plus tard identifiées comme étant des *électrons*) dans la majorité des conducteurs.

- 1881 – Poincaré décrit le plan hyperbolique à l'aide du modèle de l'hyperboloïde. Les constructions de Poincaré en géométrie hyperbolique influencent fortement la géométrie discutée dans cette thèse.

- 1885 – *Série de Balmer*. Les quatre premières raies spectrales de l'hydrogène ont pour longueur d'onde 652 nm, 486 nm, 434 nm et 410 nm. Balmer montre que ces longueurs d'ondes satisfont la formule
$$\lambda = B \times \frac{m^2}{m^2 - n^2} ,$$
avec $n = 2$, $m = 3, 4, 5, 6$ et $B = 364$ nm. C'est un premier pas dans l'étude de la structure électronique de l'atome d'hydrogène.

- 1887 – Hertz démontre expérimentalement l'existence des ondes électromagnétiques prédites par Maxwell, en produisant des *ondes Hertziennes*. Les ondes électromagnétiques décrites par la théorie de Maxwell englobent une variété de phénomènes physiques en fonction de leur longueur d'onde λ : des ondes Hertziennes ($\lambda \gtrsim 1$ m) aux rayons γ ($\lambda \lesssim 10^{-12}$ m) (découverts par Villard en 1904, qui prouve également en 1914 que ces rayons γ sont des ondes électromagnétiques) en passant par la lumière visible (400×10^{-9} m $\lesssim \lambda \lesssim 800 \times 10^{-9}$ m). Cela constitue le spectre électromagnétique, présenté en Figure 2. Dans la suite de cette thèse et notamment de cette introduction, le terme *lumière* sera utilisé de manière équivalente à "rayonnement électromagnétique" : la lumière visible, les ondes radio, les rayonnements infrarouges, ultraviolets, X et γ sont de la *lumière*.

- 1887 – Les expériences de Michelson et Morley démontrent que la vitesse de la lumière c est indépendante du référentiel d'observation.

- 1888 – *Formule de Rydberg* pour les raies spectrales de l'hydrogène et d'autres éléments :
$$\frac{1}{\lambda} = R_H \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) ,$$

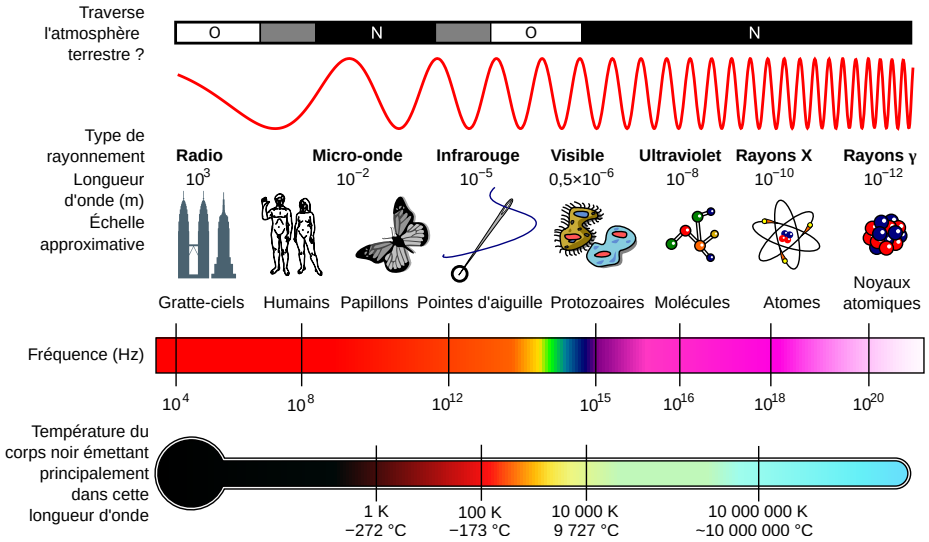


Figure 2: Le spectre électromagnétique (source : image de la NASA retravaillée par Inductiveload et traduite par Berru, sous licence CC BY-SA 3.0).

où $n_1, n_2 \in \mathbb{Z}_{>0}$ et où R_H est la constante de Rydberg. Cette formule généralise la série de Balmer évoquée plus haut.

- 1894 – Cartan défend sa thèse *Sur la structure des groupes de transformations finis et continus*, pierre angulaire de la théorie des algèbres de Lie.

- 1895 – Röntgen découvre les rayons X.

- 1895 – Lorentz introduit les *transformations de Lorentz*, qui sont des changements de coordonnées de l'espace-temps sous lesquelles les équations de Maxwell gardent la même forme. On dit que ces transformations *préservent* les équations de Maxwell.

- 1896 – Becquerel découvre la radioactivité en étudiant des sels d'uranium. Il en distingue deux types qu'il appelle α et β : les rayons α sont chargés positivement et pénètrent peu la matière (quelques centimètres dans l'air) tandis que les rayons β sont chargés négativement, et beaucoup plus pénétrants. De plus, il est trouvé dans les années qui suivent que le spectre d'émission des rayons α est un spectre de raies, tandis que celui des rayons β est continu. Ses travaux sont approfondis notamment par Marie et Pierre Curie qui découvrent le *polonium* et le *radium* (et sont à l'origine du terme "radioactivité").

- 1897 – Thomson, Townsend et Wilson démontrent que les rayons cathodiques sont constitués d'*électrons*. Ce sont des particules de charge électrique négative et de masse $m_e \sim 9 \times 10^{-31}$ kg. Thompson propose le modèle de *plum pudding*, selon lequel la matière est constituée d'un milieu homogène chargé positivement dans lequel des électrons sont répartis et peuvent se déplacer.

- 1900 – Hilbert propose sa liste de 23 problèmes lors du 2ème congrès international des mathématiques à Paris. Ces problèmes ont eu un impact très important sur l'évolution ultérieure des mathématiques.

- 1900 – Villard découvre la radioactivité γ, dont les rayons ne sont pas déviés par un champ magnétique (les rayons γ sont donc électriquement neutres) et plus de 100 fois plus pénétrants que les rayons β.

- 1902 – Lebesgue soutient sa thèse de doctorat *Intégrale, longueur, aire*, dans laquelle est introduite sa théorie de l'intégration, socle des mathématiques modernes et de la théorie des probabilités axiomatisée par Kolmogorov en 1933.

- 1904 – Poincaré énonce sa célèbre conjecture, prouvée en 2002 par Perelman : toute variété différentielle fermée de dimension 3 simplement connexe est homéomorphe à une 3-sphère.

- 1904 – Abegg énonce une loi précurseuse de la *règle de l'octet* pour la valence des atomes, qui explique comment les atomes s'assemblent pour former des molécules.
- 1905 – Théorème de Wedderburn : tout corps fini est commutatif.

Théorie de la relativité restreinte

Les travaux sur la lumière de Young et Fresnel ont été à l'origine du consensus scientifique du 19ème siècle que la lumière est une onde transversale, qui se propage dans un milieu appelé *éther*. Différentes théories de l'éther sont proposées, mais elles sont systématiquement en contradiction avec certaines expériences comme celle de Michelson et Morley de 1887. Les théories de la relativité d'Einstein tiennent leur nom du *principe de relativité*, énoncé ci-dessous. On distingue deux telles théories : la *relativité restreinte* (1905), dans laquelle l'espace-temps est un support rigide où on décrit la cinématique d'objets ponctuels, et la *relativité générale* (1915), dans laquelle l'espace-temps est dynamique et satisfait *l'équation d'Einstein* (et où l'on peut aussi étudier la cinématique d'objets ponctuels).

- 1895 – Électrodynamique de Maxwell–Lorentz, qui repose sur le fait que la vitesse de la lumière est constante dans toutes les référentiels.
- 1898 – *La mesure du temps*, Poincaré, où l'auteur souligne que puisque la lumière a une vitesse finie, la synchronisation de deux horloges par signaux lumineux a des conséquences sur la notion même de simultanéité.
- 1904 - 1905 – Travaux de Lorentz et Poincaré sur une théorie dynamique de l'électron. Wien montre qu'il est impossible d'aller plus vite que la vitesse de la lumière.
- 1905 – *Zur Elektrodynamik bewegter Körper*, Einstein, qui pose les bases de la relativité restreinte. La théorie d'Einstein repose sur deux hypothèses :
  1. Le principe de relativité : Les lois de la physique sont les mêmes pour des observatrices qui se déplacent à vitesse constante les unes par rapport aux autres.
  2. La constance de la vitesse de la lumière pour les observatrices dans tous les systèmes de référence en mouvement uniforme.

Einstein en déduit naturellement les transformations de Lorentz. Il montre que cela entraîne la contraction des longueurs et la dilation du temps. Les transformations relativistes font souvent intervenir le facteur :

$$\gamma = \frac{1}{\sqrt{1-(v/c)^2}}$$

où v est la vitesse d'un référentiel mobile par rapport à un référentiel fixe, et c la vitesse de la lumière. Puisque nécessairement $v < c$, on a $\gamma > 1$.

Considérons une observatrice et son référentiel d'observation, ainsi qu'une fusée qui passe devant elle à une vitesse v et qui émet deux flash lumineux séparés par un intervalle de temps $\Delta\tau$ mesuré dans le référentiel de la fusée. Pour l'observatrice immobile, l'intervalle de temps entre les deux flash lumineux est de $\Delta t = \gamma\Delta\tau$. C'est la *dilatation du temps*. De plus, si la fusée a une longueur L dans son référentiel alors, pour l'observatrice immobile, elle est de taille $\Delta x = \gamma^{-1}L$. C'est la *contraction des longueurs*.

Enfin, Einstein démontre que l'énergie cinétique relativiste d'une particule de masse m s'écrit

$$E = mc^2(\gamma - 1) = \sqrt{m^2c^4 + p^2c^2}\ ,$$

ce qui conduit (dans le référentiel de repos de la particule) à la célèbre formule :

$$E = mc^2\ . \tag{3}$$

- 1907 - 1908 – Minkowski introduit l'espace-temps comme une variété non-euclidienne à quatre dimensions. *L'espace-temps de Minkowski* $\mathbb{R}^{1,3}$ est la variété affine réelle $\mathbb{R}^4$ munie de la métrique pseudo-Riemannienne de signature $(-,+,+,+)$. On distingue trois types de vecteurs selon le signe de leur norme :
  - Un vecteur $\overrightarrow{AB}$ est de genre *temps* si $||\overrightarrow{AB}||^2 < 0$.

– Un vecteur $\overrightarrow{AB}$ est de genre *lumière* si $||\overrightarrow{AB}||^2 = 0$.

– Un vecteur $\overrightarrow{AB}$ est de genre *espace* si $||\overrightarrow{AB}||^2 > 0$.

Les transformations de Lorentz sont naturellement définies comme les isométries de l'espace-temps de Minkowski, et par conséquent elle préservent les normes.

Un évènement A (c'est-à-dire, un point de l'espace-temps de Minkowski) ne peut influencer un évènement B que si une information émise à A peut atteindre B. La vitesse maximale à laquelle cette information transite est c, et par conséquent il faut que le vecteur $\overrightarrow{AB}$ soit de genre lumière et que B soit dans le futur de A. L'ensemble des points B qui peuvent être influencés par A forme le cône futur $\mathcal{C}^+$ de A. De même, on définit le cône passé $\mathcal{C}^-$ de A comme l'ensemble des évènements B qui peuvent influencer A. Les cônes futurs et passés d'un point de l'espace-temps constituent une partie de l'étude de la *structure causale* de l'espace-temps.

L'espace-temps de Minkowski peut être défini en toute dimension $d \geq 2$ comme l'espace affine $\mathbb{R}^d$ muni de la métrique de signature $(-, +, \ldots, +)$. L'espace-temps de Minkowski de dimension 2 est représenté à gauche dans la Figure 3.

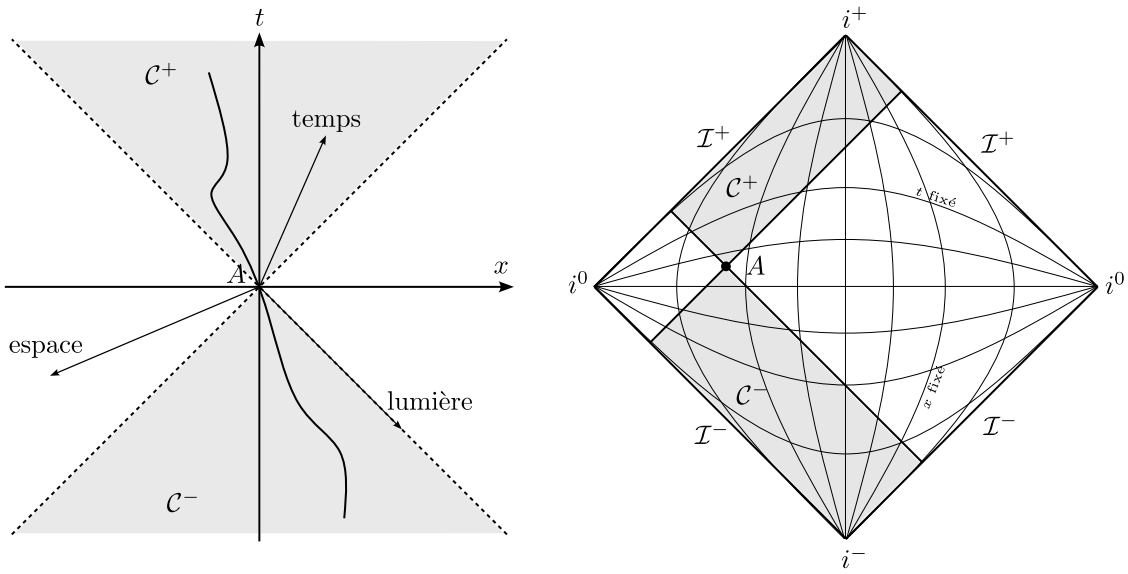


Figure 3: L'espace-temps de Minkowski en 2 dimensions.

Le diagramme de Minkowski présenté en Figure 3 constitue un exemple de *diagramme d'espace-temps*, où comme son nom l'indique à la fois le temps et l'espace sont représentés. Sur un tel digramme, une courbe dont la tangente est (à chaque instant) contenue dans le cône de lumière futur décrit l'ensemble des positions successives occupées par une particule massive : au lieu de représenter un objet en mouvement de manière dynamique dans l'espace, on le voit comme cette courbe statique dans l'espace-temps. On parle de la *ligne d'univers* de cet objet. Un exemple de ligne d'univers est représenté dans le diagramme de Minkowski en Figure 3.

Le diagramme à droite de la Figure 3 est le *diagramme de Penrose* de l'espace-temps de Minkowski en 2 dimensions. C'est une représentation "déformée" de ce dernier, de taille finie. Plus l'on s'approche du bord du diagramme, et plus les distances représentées sont grandes. La déformation est judicieusement choisie pour que les lignes de lumière correspondent toujours à des lignes orientées à 45° (techniquement, on passe du diagramme de Minkowski au diagramme de Penrose par une transformation conforme). Ainsi, le diagramme de Penrose décrit efficacement la structure causale de l'espace-temps de Minkowski. Au bord du diagramme de Penrose, la déformation est telle que certains infinis sont contractés en points : par exemple, le point i^+ (respectivement, i^-, i^0) est l'infini futur de genre temps (respectivement, passé de genre temps, de genre espace). Les infinis de genre lumière futur $\mathcal{I}^+$ et passé $\mathcal{I}^-$ sont, quant à eux, contractés en des segments.

Physique pré-quantique

Dans cette section, nous allons prendre le temps de détailler davantage les deux expériences fondatrices de la mécanique quantique, puisqu'elles permettent de saisir l'essence des idées de la physique quantique développée ultérieurement.

Rayonnement des corps noirs et loi de Planck (1900). Un corps noir est un objet idéal (c'est-à-dire un concept, une idéalisation, plutôt qu'un objet physique) qui absorbe toute l'énergie électromagnétique qu'il reçoit sans en réfléchir ou en transmettre aucune fraction. À température non-nulle, un corps noir émet un rayonnement électromagnétique sous l'effet de l'agitation thermique (de la même manière qu'un morceau de métal chauffé au rouge, ou du rayonnement infrarouge émis par un objet à température ambiante). À l'équilibre thermique, l'énergie émise et absorbée s'équilibrent parfaitement si bien que le spectre du rayonnement émis ne dépend que de la température.

Une cavité hermétique fournit un bon modèle de corps noir. Dans le cas d'une cavité unidimensionnelle de longueur L, l'énergie de la cavité se répartit en ondes stationnaires dont la longueur d'onde λ satisfait

$$\lambda = \frac{2L}{m} \ ,$$

avec m un nombre entier. Le principe d'équipartition de l'énergie de physique statistique implique que chaque degré de liberté du système (c'est-à-dire, chaque onde stationnaire) a une énergie moyenne de $k_B T$, où k_B est la constante de Boltzmann et T la température. Cela est possible classiquement puisque l'amplitude d'une onde peut être arbitrairement petite. Si l'on considère à présent un cube de côtés de longueur L, on obtient la loi de Rayleigh–Jeans :

$$I_{\mathrm{RJ}}(\lambda, T) = \frac{2\pi c k_B T}{\lambda^4} \ , \tag{4}$$

où $I(\lambda, T)$ est la puissance émise à la température T par les ondes stationnaires de longueur d'onde λ (c'est une limite continue). Cette formule varie comme λ^{-4} : la puissance émise diverge pour les petites longueurs d'ondes, ce qui n'est pas possible physiquement car la cavité reçoit une énergie finie. On parle de *catastrophe ultraviolette*.

Dans la Figure 4, la courbe de radiance spectrale (la puissance émise par unité d'angle solide) en fonction de la longueur d'onde est donnée pour un corps noir à 3000 K, 4000 K et 5000 K, tandis que la formule de Rayleigh–Jeans donne la courbe la plus à droite, qui ne coïncide qu'asymptotiquement, pour les grandes longueurs d'ondes.

En 1900, Planck propose une résolution de ce problème dans un article intitulé *Zur Theorie des Gesetzes der Energieverteilung im Normalspectrum*, basée sur l'hypothèse que l'énergie des ondes stationnaires dans la cavité est quantifiée. Plus précisément, il impose que l'énergie d'une onde de longueur d'onde λ doit être un multiple de hc/λ, où h est une constante. Par conséquent, l'amplitude d'une onde ne peut être arbitrairement petite. Planck estime la valeur de cette constante à $h \sim 6,55 \times 10^{-34}$ J·s ; elle prend ensuite le nom de *constante de Planck*. À haute fréquence (i.e. à petite longueur d'onde), les modes ne sont pas excités car la plus petite excitation a une énergie supérieure à $k_B T$, et donc la puissance émise ne diverge plus lorsque λ tend vers 0. La loi de Planck ainsi calculée donne :

$$I_{\mathrm{P}}(\lambda, T) = \frac{2\pi h c^2}{\lambda^5} \frac{1}{\exp\left(\dfrac{hc}{\lambda k_B T}\right) - 1} \ ,$$

et contrairement à la loi de Rayleigh–Jeans énoncée plus haut, elle décrit de manière satisfaisante le spectre d'émission des corps noirs.

Effet photoélectrique (1905). Il s'agit du phénomène selon lequel un matériau métallique soumis à de la lumière (i.e. un rayonnement électromagnétique) émet des électrons. Observé pour la première fois en 1839 par Becquerel et Edmond, l'effet photoélectrique a les propriétés intéressantes suivantes. Les électrons ne sont émis que si la fréquence des ondes lumineuses incidentes est suffisante. La quantité d'électrons émis est alors proportionnelle à l'intensité de la source lumineuse, mais la vitesse de ces électrons ne dépend, encore une fois, que de la fréquence de l'onde lumineuse. Ce phénomène n'est pas explicable par la théorie de Maxwell, dans laquelle l'énergie déposée par une onde électromagnétique dans

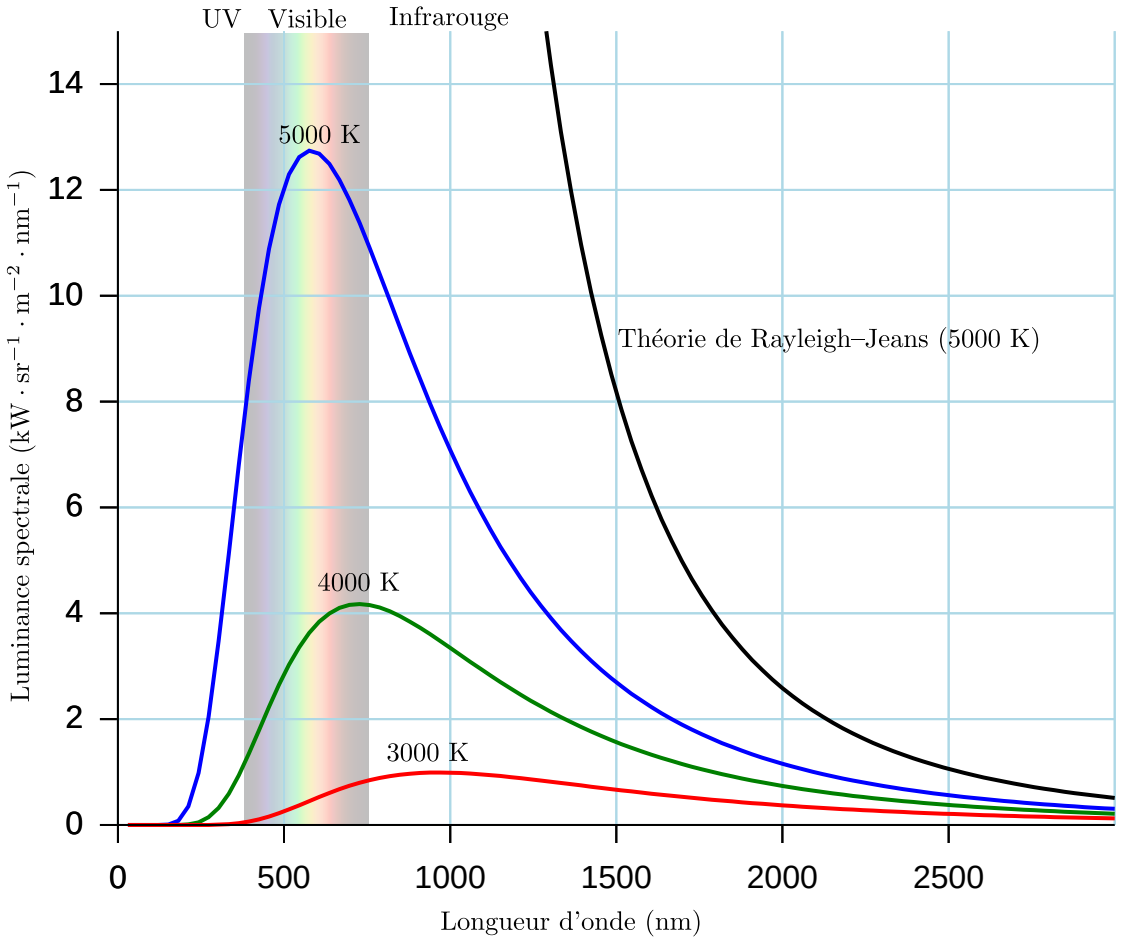


Figure 4: Courbes spectrales de rayonnement de corps noirs et prédiction de Rayleigh–Jeans (source : travail personnel à partir d'un graphique de wikipédia).

un matériau dépend de la fréquence de l'onde et de son intensité, sans que ces deux grandeurs puissent être décorrélées.

En 1905, Einstein publie *Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt*, dans lequel il prolonge l'idée des *quantas* de lumière de Planck : l'énergie d'une onde électromagnétique de fréquence ν est un multiple de $h\nu$, où h est la constante de Planck. Einstein interprète ces niveaux d'énergies de manière corpusculaire : de la lumière monochromatique de fréquence ν est constituée de photons de fréquence ν, et chacun de ces photons a une énergie $h\nu$. Ainsi lorsque la fréquence ν est trop faible, $h\nu$ est inférieure à l'énergie E nécessaire pour déloger un électron du métal. Inversement, lorsque la fréquence ν est suffisamment élevée, chaque photon peut déloger un électron du métal et l'énergie de l'électron émis est de $E - h\nu$, c'est-à-dire indépendante de l'intensité de la source lumineuse. La quantité d'électrons émis par effet photoélectrique est quant à elle proportionnelle à la densité de photons du faisceau lumineux, c'est-à-dire à l'intensité de la source.

Cette explication de l'effet photoélectrique confirme le bien-fondé physique de la quantification des ondes électromagnétiques. Ainsi, la lumière a un comportement corpusculaire en plus de ses caractéristiques ondulatoires mises en exergue par Young et Fresnel.

- 1904 - 1910 – Hilbert pose les bases de la théorie spectrale des opérateurs linéaires, notamment auto-adjoints.
- 1907 – Rutherford et Royds prouvent que les rayons α sont constitués de particules, et plus précisément d'ions d'hélium.
- 1909 – Taylor réalise l'expérience des fentes d'Young avec de la lumière d'intensité si faible que les photons sont émis individuellement, et observe la persistance des interférence. Les photons

introduits par Planck et Einstein étant définis comme des particules, cette expérience prouve qu'ils ont également des propriétés ondulatoires.

- 1909 – En bombardant une feuille d'or avec des particules α, Geiger, Marsden et Rutherford démontrent que la fraction de la matière chargée positivement est contenue dans un tout petit volume : le *noyau atomique*. Cela réfute le modèle du plum pudding de Thompson qui laisse place au *modèle planétaire de l'atome*, dans lequel les électrons, chargés négativement, orbitent autour d'un noyau atomique, chargé positivement.
- 1909 – L'expérience de la goutte d'huile de Millikan démontre la quantification de la charge électrique, et permet la mesure de la charge élémentaire $e \sim 1.6 \times 10^{-19}$ C. Il s'agit en fait de la valeur de la charge de l'électron.
- 1913 – *On the constitution of atoms and molecules*, Bohr, dans lequel celui-ci propose le *modèle atomique de Bohr*. Les électrons sont répartis en orbites circulaires autour du noyau atomique, et les seules orbites autorisées sont celles telles que le moment angulaire $m_e vr$ de l'électron (m_e est la masse de l'électron, r est la rayon de l'orbite et v et la vitesse de l'électron qui doit compenser exactement l'attraction électrique) est quantifié. Plus précisément, Bohr demande que $m_e vr = n\hbar$, où n est un entier et $\hbar = (2\pi)^{-1}h$ est la *constante de Planck réduite*. Le modèle de Bohr permet de retrouver la formule de Rydberg pour la longueur d'onde des raies spectrales de l'hydrogène, et fournit ainsi une explication attrayante de cette série heuristique.
- 1916 – *The Atom of the Molecule*, Lewis, qui pose les bases de la liaison chimique.
- 1917 – Rutherford prouve l'existence des *protons* dans les noyaux atomiques.
- 1919 – *The Arrangement of Electrons in Atoms and Molecules*, Langmuir, dans lequel il propose le terme de *liaison covalente*.
- 1920 – Rutherford propose l'existence de *neutrons* au sein des noyaux atomiques.
- 1922 – L'expérience de Stern et Gerlach démontre la quantification du moment angulaire de systèmes de taille atomique, conduisant à la découverte du *spin* (le un moment angulaire intrinsèque) de l'électron.

Théorie de la relativité générale et cosmologie

La théorie de la relativité générale prolonge naturellement les idées d'Einstein développées dans la théorie de relativité restreinte, et fournit un paradigme nouveau quant à la gravité, par rapport à la théorie de Newton établie plus de 300 ans auparavant. Elle prédit notamment l'existence de trous noirs, qui jouent un rôle particulier dans la motivation des travaux présentés dans cette thèse. La cosmologique moderne (et notamment la théorie du Big Bang) est également basée sur la relativité générale.

- 1908 – *Relativitätsprinzip und die aus demselben gezogenen Folgerungen*, Einstein, dans lequel est exposé le *principe d'équivalence* : une observatrice en chute libre doit percevoir la même physique que si elle était dans un référentiel inertiel.
- 1911 – *Einfluss der Schwerkraft auf die Ausbreitung des Lichtes*, Einstein. Sont proposés d'une part que l'écoulement du temps dépend du champ gravitationnel, et d'autre part que la lumière est défléchie par les corps massifs.
- 1915 – Einstein publie sa théorie de la *relativité générale*, qui explique que la métrique de l'espace-temps $g_{\mu\nu}$, de tenseur de Ricci $R_{\mu\nu}$ et de courbure scalaire R, dépend de la distribution de masse et d'énergie décrite par le tenseur énergie-impulsion $T_{\mu\nu}$, selon *l'équation d'Einstein*:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi\mathcal{G}}{c^4}T_{\mu\nu} \ , \tag{5}$$

  où Λ est la *constante cosmologique* et $\mathcal{G}$ la constante de gravitation universelle. Si on fait abstraction de la constante cosmologique, l'équation d'Einstein indique que la courbure de l'espace-temps (le terme de gauche) est proportionnelle à la densité d'énergie-impulsion (le terme de droite). La linéarisation de cette équation décrit la propagation *d'ondes gravitationnelles*, qui sont des perturbations locales des intervalles de temps et des longueurs.

- 1915 – Hilbert publie également un article sur cette même théorie de la relativité générale. L'équation d'Einstein peut être obtenue par principe variationnel sur l'action de Einstein–Hilbert :

$$S_{\mathrm{EH}} = \frac{c^4}{16\pi\mathcal{G}} \int \left[(R - 2\Lambda) + \mathcal{L}_M\right] \sqrt{-g}\mathrm{d}^4x \ , \tag{6}$$

où $\mathcal{L}_M$ est la densité lagrangienne des champs de matière dans la théorie.

- 1916 – Schwarzschild calcule la première solution exacte aux équations d'Einstein dans le vide (c'est-à-dire pour $T_{\mu\nu} = 0$), dite *métrique de Schwarzschild*. Dans les *coordonnées de Schwarzschild*, elle s'écrit:

$$g = -\left(1 - \frac{r_S}{r}\right) c^2\mathrm{d}t^2 + \left(1 - \frac{r_S}{r}\right)^{-1} \mathrm{d}r^2 + r^2\mathrm{d}\Omega_2^2 \ , \quad r_S = \frac{2\mathcal{G}M}{c^2} \ , \tag{7}$$

où r_S est appelé rayon de Schwarzschild, $\mathrm{d}\Omega_2^2$ est l'élément différentiel de surface sur une 2-sphère et M est une masse. La métrique de Schwarzschild n'est pas complète dans ce jeu de coordonnées ; sa complétion peut-être décrite en utilisant notamment les coordonnées de Kruskal–Szekeres qui datent de 1960. Le diagramme de Penrose correspondant à cette complétion est donné en Figure 5 (dans lequel chaque point correspond à une 2-sphère).

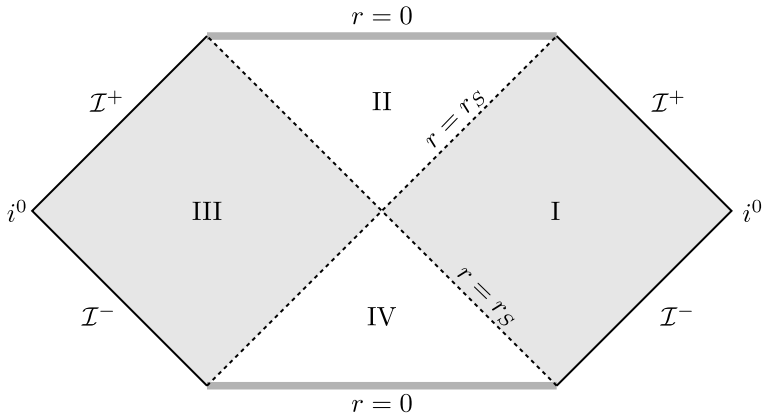


Figure 5: Le diagramme de Penrose de la complétion de la métrique de Schwarzschild.

On y distingue quatre régions, judicieusement dénotées I, II, III et IV. Comme avant, le diagramme de Penrose est un outil pratique pour étudier la causalité dans la solution de Schwarzschild. On y voit notamment que le cône futur de tout point dans la région II termine sur la ligne supérieure $r = 0$. De manière similaire, le cône passé de tout point dans la zone IV commence sur la ligne inférieure $r = 0$. Cela détermine la propriété *d'horizon* des lignes pointillées $r = r_S$. La métrique de Schwarzschild est interprétée comme une métrique de trou noir où $r = r_S$ détermine l'horizon : toute particule, massive ou non, qui traverse cet horizon ne peut en ressortir. La région de la solution de Schwarzschild à l'extérieur de l'horizon est I, l'intérieur du trou noir est II, et la singularité est à $r = 0$. La région IV est interprétée comme un trou blanc, et la région III, comme un univers parallèle décrit par la complétion de la métrique de Schwarzschild.

La solution de Schwarzschild fournit un modèle théorique pour étudier certaines propriétés des trous noirs non chargés et statiques, avec certaines limites, notamment quant à la formation du trou noir lui-même.

- 1916 - 1918 – Reissner et Nordström écrivent une solution aux équations d'Einstein qui décrit un trou noir chargé électriquement.

- 1922 - 1927 – Friedmann et Lemaître proposent une forme de métrique solution de l'équation d'Einstein qui décrit un univers homogène, isotrope en expansion ou contraction, et qui constitue le fondement de la cosmologie moderne.

- 1929 – Hubble découvre que l'univers est en expansion en observant que le *décalage vers le rouge* de la lumière émise par les galaxies lointaines est proportionnel à leur distance : plus les galaxies sont distantes, plus elles s'éloignent vite de la Terre. En extrapolant cette expansion dans le passé, on

arrive à la conclusion que l'univers devait être incroyablement contracté il y a quelques 14 milliards d'années : c'est le début de la *théorie du Big Bang*.

- 1933 – Zwicky observe (par un calcul de gravité de Newton) que la vitesse de certaines galaxies dans l'Amas de la Chevelure de Bérénice ne correspond pas à ce qu'on attend compte tenu de la distribution de masse lumineuse de l'amas, c'est-à-dire de la masse estimée à partir de ses émissions lumineuses. C'est l'un des premiers indices qui mènent ultimement à l'hypothèse de l'existence d'une forme de matière massive qui n'interagit que par gravitation et n'émet pas de lumière : la *matière noire*.

- 1935 – Robertson et Walker prouvent l'unicité de la métrique dite de Friedmann–Lemaître–Robertson–Walker pour décrire un univers homogène et isotrope.

- 1963 – Kerr propose une solution aux équations d'Einstein qui décrit des trous noirs en rotation.

- 1964 – Découverte de Cygnus X-1, dans la constellation du Cygne. Il s'agit du premier trou noir jamais observé, qui fournit une preuve que les trous noirs existent dans notre univers et ne sont pas seulement une solution exotique aux équations d'Einstein.

- 1965 – Newman écrit une solution dite de Kerr–Newman qui décrit des trous noirs chargés électriquement, en rotation.

- 1965 – Penzias et Wilson découvrent le *fond diffus cosmologique* (CMB), qui est une certification importante de la théorie du Big Bang. Le CMB est un rayonnement thermique qui suit presque exactement la distribution des corps noirs à une température de 2,7 K, et émis quasi-uniformément de toutes les directions. Il est interprété comme une relique du rayonnement thermique au moment du découplage lumière-matière, 379 000 ans après le Big Bang, lors duquel l'Univers était à une température de 3000 K.

- 1967 – Théorème dit *de calvitie* : un trou noir classique est entièrement décrit par sa masse, sa charge électrique et son moment angulaire. Autrement dit, un trou noir ne garde aucun information de comment il s'est formé, à partir de quel(s) astre(s), et dans quelles conditions.

- 1970 – Les observations de Rubin, Ford et Freeman sur les courbes de vitesse radiale dans de nombreuses galaxies fournissent des données supplémentaires très convaincantes quant à l'existence de la matière noire : la vitesse de rotation des étoiles en fonction de la distance au centre galactique est incompatible avec les prédictions gravitationnelles basées sur la distribution de masse "lumineuse" de la galaxie.

- 1979 – Guth propose la théorie de l'inflation qui complète de manière élégante la théorie du Big Bang et rend cette dernière plus accommodable avec les observations expérimentales. L'inflation serait une période d'expansion exponentielle de l'univers primordial, entre 10^{-36} et 10^{-32} secondes après le Big Bang.

- 1998 – Premières indications directes de l'expansion accélérée de l'univers. L'explication parcimonieuse de cette accélération requiert une constante cosmologique Λ faiblement positive : c'est *l'énergie noire*. Les observations indiquent que la densité d'énergie dans notre univers se répartit en quelques 70% d'énergie noire, 25% de matière noire et 5% de matière ordinaire. Cette distribution fait désormais partie intégrante du *modèle standard de la cosmologie* ou modèle Λ-CMB, dont les grandes lignes sont présentées en Figure 6.

- 2015 – Première détection d'ondes gravitationnelles dans les interféromètres gravitationnels LIGO et VIRGO. C'est la première preuve directe de l'existence de celles-ci, bien que prédites par Einstein cent ans auparavant. Plusieurs ondes gravitationnelles ont depuis été observées, révolutionnant de nombreux domaines d'astrophysique.

- 2016 – Des observations tendent à montrer qu'une galaxie dénommée Dragonfly 44 est constituée presque entièrement de matière noire.

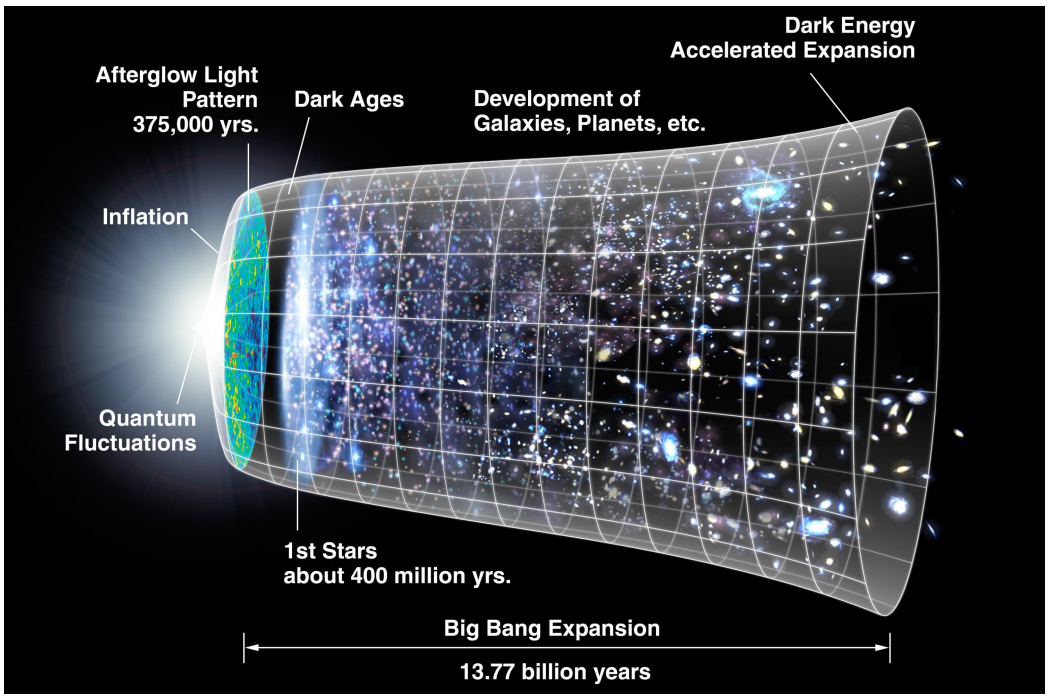


Figure 6: Évolution de l'univers selon le modèle standard de la cosmologie (source : NASA).

- 2019 – Première image directe d'un trou noir supermassif par le "Event Horizon Telescope" (EHT) (M87* dans la galaxie de Messier M87). Les trous noirs observés auparavant ne l'avaient été que de manière indirecte, en mesurant les vitesses des étoiles orbitant autour d'eux. L'image de la Figure 7 cependant, qui est la photographie de l'EHT, montre le disque d'accrétion formé de gaz ionisé en train de chuter dans M87*.

Figure 7: Le trou noir M87* (photo du EHT arrangée par BevinKacon, sous licence CC BY 4.0).

Théorie quantique

- 1923 – *Ondes et quantas*, de Broglie. La relation de Planck–Einstein $E = h\nu$ peut se réécrire en $p = h/\lambda$, où $p = E/c$ est la quantité de mouvement d'un photon de longueur d'onde λ. De

Broglie fait l'hypothèse que cette relation est en fait valable pour toute particule, y compris celles massives, et donc que tout particule a des caractéristiques ondulatoires. Cette hypothèse est d'abord confirmée pour les électrons en 1928 par les expériences de Davisson et Germer. Ultérieurement, l'expérience des fentes d'Young a été réalisée avec des atomes et même des molécules, et montre la validité de l'hypothèse de de Broglie.

- 1925 – Mécanique matricielle de Heisenberg, Born et Jordan, qui est une première formulation de la mécanique quantique, obtenue en prenant en considération l'hypothèse de de Broglie.
- 1926 – Schrödinger publie sa célèbre équation :

$$i\hbar\frac{\partial}{\partial t}\Psi(t) = \hat{H}\Psi(t) \ , \tag{8}$$

où $\hat{H}$ est l'opérateur Hamiltonien (l'analogue quantique de l'énergie) et $\hbar = h/(2\pi)$ est la constante de Planck réduite. Par exemple, un particule de masse m sur une droite paramétrée par x dans un potentiel $V(x)$ est décrite par une fonction d'onde $\Psi : \mathbb{R}_x \times \mathbb{R}_t \to \mathbb{C}$, et le Hamiltonien s'écrit

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x) \ .$$

Dans le formalisme moderne de la physique quantique, un système est décrit par un espace de Hilbert $\mathcal{H}$ appelé espace des états (par exemple, dans le cas de la particule massive mentionnée ci-dessus, $\mathcal{H} = \mathbb{L}^2(\mathbb{R})$). La fonction d'onde indique la probabilité de présence de la particule : dans un état $\Psi(x,t)$, la probabilité de mesurer la particule entre $x = x_-$ et $x = x_+$ est :

$$\mathbb{P} = \int_{x_-}^{x_+} |\Psi(x,t)|^2 \ \mathrm{d}x \ ,$$

où la fonction d'onde est normalisée de sorte que $\int_{\mathbb{R}} |\Psi(x,t)|^2 \ \mathrm{d}x = 1$.

La description d'un système de deux particules identiques dans l'espace $\mathbb{R}^3$ donne lieu à la distinction fondamentale entre les particules qualifiées de bosons et celles nommées fermions.

Si l'on échange les deux particules, la fonction d'onde d'état initial $\Psi(\mathbf{x})$ ne peut qu'être multipliée par une phase $\theta \in \mathbb{R}/2\pi\mathbb{R}$, et devient $\theta\Psi(\mathbf{x})$. Si l'on ré-échange les deux particules, on obtient la fonction d'onde $\theta^2\Psi(\mathbf{x})$ qui doit coïncider avec $\Psi(\mathbf{x})$, si bien que $\theta = \pm 1$. On peut argumenter que cette propriété ne dépend que du type de particule. Dans le cas où $\theta = -1$, on dit que la particule est de type *fermionique*. Elle satisfait alors le *principe d'exclusion de Pauli* (1925) : deux particules fermioniques identiques ne peuvent occuper simultanément le même état quantique. Un système de plusieurs particules fermioniques identiques suit la *statistique de Fermi–Dirac* (1926). Les électrons par exemple sont des fermions, ce qui explique l'organisation des électrons d'un atome en orbitales électroniques autour du noyau, et ainsi les propriétés chimiques de cet atome. Dans le cas où $\theta = +1$, on dit que la particule est de type *bosonique*. Un système de plusieurs bosons identiques suit la *statistique de Bose-Einstein* (1925) : deux bosons peuvent être dans le même état quantique. Les photons sont des exemples de bosons, et la statistique de Bose–Einstein permet notamment l'existence de faisceaux laser.

Une spécificité de la physique quantique est qu'elle permet l'existence d'états intriqués, où l'on a une connaissance maximale de l'état global du système en même temps qu'une ignorance maximale des composantes de cet état. Par exemple, on peut considérer un état de deux électrons de spin total zéro, sans que l'on puisse distinguer quel électron à un spin "up" et quel électron a un spin "down".

Indépendamment du formalisme mathématique de la mécanique quantique présenté ci-dessus, on peut expliquer la différence de nature entre la mécanique quantique et la mécanique classique comme suit. Considérons un objet ponctuel de masse m, qu'on libère sans vitesse initiale au temps $t = t_0$ en point A. Soumis aux forces extérieures, la position de l'objet en fonction du temps est déterminée par la seconde loi de Newton. Admettons qu'à $t = t_1$, l'objet se trouve en un point B. En sachant que l'objet est en A à t_0 et en B à t_1, la trajectoire qu'il suit pour aller de A à B est celle déterminée par la seconde loi de Newton. On peut reformuler cette condition sur les trajectoires en introduisant une quantité appelée action (pour une telle particule massive, l'action est l'intégrale, le long de la trajectoire, de la différence entre l'énergie cinétique et l'énergie potentielle) et dénotée S. L'action est donc une

quantité naturellement associée à toute trajectoire. Celle effectivement prise par la particule massive est déterminée par un principe variationnel : c'est celle, parmi toutes les trajectoires de A à B qui extrémise l'action S. C'est une formulation possible de la physique classique, dite *Lagrangienne*.

Notons qu'une action est homogène à $M \cdot L^2 \cdot T^{-1}$, tout comme la constante de Planck $\hbar$ qui pour cette raison est parfois appelée *quanta d'action*. En physique quantique, la notion de trajectoire n'est pas bien définie puisque chaque particule a une longueur d'onde d'après le principe de de Broglie, et donc une extension spatiale (on peut aussi invoquer les inégalités d'Heisenberg – voir ci-dessous). On peut néanmoins calculer la *probabilité* que la particule, partant de A à t_0, soit en B à t_1, par une *intégrale de chemin* : on fait la somme pondérée de *toutes* les trajectoires, aussi non-physiques soient elles, où à chaque trajectoire est assigné le poids $\exp(iS/\hbar)$. L'intégrale de chemin en question s'écrirait par exemple :

$$\int [\mathcal{D}\gamma] \exp\left(i\frac{S[\gamma]}{\hbar}\right) , \tag{9}$$

où γ est un chemin continu de A à B, $S[\gamma]$ est l'action de γ et $[\mathcal{D}\gamma]$ est l'élément variationnel sur l'espace des chemins, qui est très difficile à définir rigoureusement en général, sauf en dimension 0 et en dimension 1, où cela peut notamment être fait grâce au mouvement Brownien. Ainsi, en physique quantique, il n'y a pas de trajectoire préférée ; toutes les trajectoires sont pondérées par l'exponentielle de l'action multipliée par le nombre imaginaire i, mesurée en unités de $\hbar$. Le lemme de Riemann–Lebesgue assure que les trajectoires qui contribuent le plus à l'intégrale de chemin sont celles qui sont "proches" des trajectoires classiques.

- 1926 – Klein, Gordon et Fock proposent l'équation dite de *Klein–Gordon* qui est une généralisation relativiste de l'équation de Schrödinger :

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2 + \frac{m^2c^2}{\hbar^2}\right)\phi(t,\mathbf{x}) = 0 .$$

- 1927 – *Principe d'incertitude* ou d'indétermination d'Heisenberg : l'incertitude sur la quantité de mouvement Δp et celle sur la position Δx d'une particule massive satisfont l'inégalité

$$\Delta x \Delta p \geq \frac{\hbar}{2} .$$

  C'est une conséquence directe du caractère ondulatoire des objets en mécanique quantique. On interprète souvent l'inégalité ci-dessus en disant qu'il est impossible de connaître très précisément à la fois la position et la vitesse d'une particule quantique.

- 1927 – *Wechselwirkung neutraler Atome und Homöopolare Bindung nach der Quantenmechanik*, Heitler et London. Cet article étudie la liaison covalente de la molécule de dihydrogène H_2 avec une approche quantique. La notion de liaison chimique est ensuite largement développée par Pauling.

- 1928 – Dirac publie son *équation de Dirac* :

$$(i\hbar\gamma^\mu\partial_\mu - mc)\psi(t,\mathbf{x}) = 0 ,$$

  où $\gamma_0, \gamma_1, \gamma_2$ et γ_3 sont des matrices qui satisfont $\{\gamma_\mu, \gamma_\nu\} = 2h_{\mu\nu}$ avec $h_{\mu\nu}$ la métrique de Minkowski, et avec la convention de sommation d'Einstein : $\gamma^\mu\partial_\mu = \gamma^0\partial_0 + \gamma^1\partial_1 + \gamma^2\partial_2 + \gamma^3\partial_3$ qu'on utilisera également dans la suite de cette introduction.

Les solutions aux équations de Klein–Gordon et Dirac ne peuvent pas être interprétées comme des amplitudes de probabilité de présence, à la différence des solutions à l'équation de Schrödinger. En particulier, ces équations prédisent l'existence d'états d'énergie négative, ce qui mène au modèle de la *mer de Dirac* (1930). Un champ d'électrons par exemple est une solution de l'équation de Dirac ; les excitations d'énergie positive sont les électrons tandis que celles d'énergie négative sont interprétées comme les anti-particules de l'électron : les *positrons*.

- 1928 – Gamow explique la radioactivité α comme un effet tunnel, ce qui explique notamment le spectre en raies de la radioactivité α : un noyau atomique ${}^A_Z X$ devient ${}^{A-2}_{Z-2}Y$ (ou l'un de ses états excités), et l'énergie cinétique de la particule alpha émise est la différence d'énergie de liaison entre le noyau initial et les produits.

- 1927 - 1929 – von Neumann développe le concept d'espace de Hilbert abstrait et commence l'étude des opérateurs auto-adjoints non bornés.
- 1930 – Pauli propose l'existence des *neutrinos électroniques* pour expliquer le spectre d'énergie continu observé pour les électrons dans les désintégrations β^- des noyaux atomiques.
- 1932 – Découverte expérimentale des positrons e^+ par Anderson, qui sont des particules chargées positivement et dont toutes les autres propriétés sont similaires à celle des électrons. Un positron et un électron peuvent s'annihiler en des photons.
- 1932 – Découverte expérimentale des neutrons par Chadwick.
- 1924 – *Versuch einer Theorie der Strahlen*, Fermi, qui développe une première théorie des désintégrations β. Cette théorie intègre l'existence des neutrinos et permet ainsi d'expliquer le spectre en énergie des émissions β : un neutron peut spontanément devenir un proton en émettant un électron ainsi qu'un anti-neutrino électronique.

Unités naturelles. La description de la physique fondamentale qui émerge des premières décennies du 20ème siècle dégage trois constantes dimensionnées apparemment universelles : la constante de gravitation $\mathcal{G}$ avec $[\mathcal{G}] = M^{-1} \cdot L^3 \cdot T^{-2}$, la vitesse de la lumière c avec $[c] = L \cdot T^{-1}$ et la constante de Planck réduite $\hbar$ où $[\hbar] = M \cdot L^2 \cdot T^{-1}$. On peut combiner ces trois constantes pour obtenir une masse, une longueur et une unité de temps naturelles, dites de Planck:

$$m_P := \sqrt{\frac{c\hbar}{\mathcal{G}}} \sim 2,2 \times 10^{-8} \text{ kg}, \quad l_P := \sqrt{\frac{\mathcal{G}\hbar}{c^3}} \sim 1,6 \times 10^{-35} \text{ m}, \quad t_P := \sqrt{\frac{\mathcal{G}\hbar}{c^5}} \sim 5,4 \times 10^{-44} \text{ s}. \tag{10}$$

De manière équivalente, la masse, la longueur et le temps de Planck forment un système d'unités dans lequel $\mathcal{G}$, c et $\hbar$ valent 1. Travailler dans ce système d'unités permet notamment de s'affranchir des facteurs de $\mathcal{G}$, c et $\hbar$ dans les calculs, et une analyse dimensionnelle du résultat permet de les restaurer. Ainsi, l'équation de Klein–Gordon se réécrit simplement $(\partial_\mu \partial^\mu - m^2)\phi = 0$, et celle de Dirac : $(i\partial\!\!\!/ - m)\psi := (i\gamma^\mu \partial_\mu - m)\psi = 0$. À partir de ces trois grandeurs, on peut également construire d'autres quantités comme l'énergie de Planck : $E_P = m_P c^2 \sim 2,0 \times 10^9$ J.

Électrodynamique quantique

L'électrodynamique quantique (QED) est une théorie qui décrit les interactions lumière-matière de façon quantique et relativiste, développée notamment par Tomonoga, Schwinger, Feynman et Dyson. La QED est la théorie dans laquelle cela a du sens de dire qu'un photon est émis par un électron, ou que la répulsion électrique de deux électrons correspond à l'échange de photons. Plus techniquement, c'est une théorie de jauge abélienne de groupe de jauge U(1), définie sur l'espace de Minkowski 4-dimensionnel. Le potentiel vecteur A_μ de l'électromagnétisme est interprété comme le champ de photons, et interagit avec un champ de spin-1/2 ψ chargé sous A_μ et de masse m_B qui décrit, par exemple, les électrons et positrons. La dynamique de la théorie peut être exprimée sous la forme d'une action S_{QED} qui s'écrit :

$$S_{\text{QED}} = \int \mathrm{d}^4x \left[-\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \overline{\psi}(i D\!\!\!/ - m_B)\psi \right] ,$$

où $D_\mu = \partial_\mu + ie_B A_\mu$ et $\overline{\psi} = \psi^\dagger \gamma^0$. On peut écrire $S_{\text{QED}} = S_{\text{free}} + S_{\text{int}}$, où

$$S_{\text{free}} = \int \mathrm{d}^4x \left[-\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \overline{\psi}(i\partial\!\!\!/ - m_B)\psi \right] , \qquad S_{\text{int}} = -e_B \int \mathrm{d}^4x \overline{\psi} \gamma^\mu A_\mu \psi \ .$$

L'action S_{free} définit une théorie libre qui décrit un champ de spin-1 A_μ et un champ de spin-1/2 ψ n'interagissant pas entre eux. La constante e_B devant l'intégrale dans S_{int} est une *constante de couplage* dont la valeur indique l'intensité de l'interaction entre les particules correspondantes, ici les électrons/positrons et les photons. Elle est interprétée comme la charge du champ ψ, tandis que m_B est sa masse.

La valeur de e_B est petite devant 1, par conséquent on peut résoudre cette théorie de façon perturbative, en exprimant les quantités recherchées comme des perturbations aux résultats de la théorie libre,

qui peut, elle, être résolue de manière exacte. À cet effet, Feynman introduit en 1948 des diagrammes désormais appelés *diagrammes* ou *graphes de Feynman*. Dans ces diagrammes, les lignes d'univers des particules de spin-1/2 (c'est-à-dire, les excitations élémentaires du champ ψ) sont représentées par des lignes pleines orientées (dans le même sens que le temps pour des électrons et dans le sens contraire pour des positrons) et les photons (c'est-à-dire, les excitations élémentaires du champ A_μ), par des lignes ondulées non orientées. Le terme d'interaction $\overline{\psi}\gamma^\mu A_\mu \psi$ détermine le type de sommet autorisé : une ligne décrivant un photon et deux lignes décrivant les fermions (une orientée vers le sommet et l'autre s'en éloignant) se rencontrent. Ces graphes de Feynman représentent différents phénomènes physiques lorsqu'ils sont dessinés dans des diagrammes d'espace-temps, comme en Figure 8. De gauche à droite : un électron émet un photon, une paire électron-positron s'annihile en un photon, deux électrons se repoussent en échangeant deux photons, un photon devient une paire électron-positron, qui se ré-annihile en un photon.

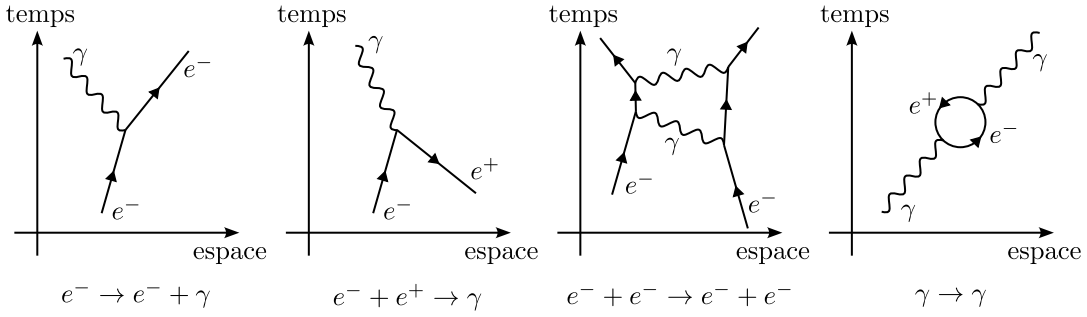


Figure 8: Des graphes de Feynman pour la QED.

A chaque graphe de Feynman on associe un nombre complexe appelé *amplitude*, calculé à partir des *règles de Feynman*. De manière similaire à la discussion ci-dessus quant à la différence entre la mécanique classique et quantique, pour calculer la probabilité d'un processus physique on doit considérer tous les diagrammes de Feynman qui décrivent ce processus, et sommer les amplitudes correspondantes. La probabilité est le carré du module de cette somme. En général, il y a une infinité de diagrammes de Feynman correspondant à un phénomène donné. Cependant, chaque sommet dans un graphe contribue notamment d'un facteur e_B, et ainsi l'amplitude d'un processus (par exemple, la diffusion électron-électron $e^-e^- \to e^-e^-$) s'écrit :

$$\mathcal{A} = \sum_n e_B^n \mathcal{A}_n \ ,$$

où $\mathcal{A}_n$ est l'amplitude totale des diagrammes de Feynman à n sommets.

Puisque $e_B \ll 1$, on peut tronquer la somme en laquelle $\mathcal{A}$ est décomposée à un certain ordre, ce qui donne une approximation du résultat escompté. L'électrodynamique quantique est une des théories les plus fructueuses de la physique, à ce jour, puisqu'elle permet de calculer des grandeurs de manière très précise ; les prédictions théoriques sont en excellent accord avec les observations expérimentales. Par exemple, le calcul théorique du *moment anomal de l'électron* donne :

$$a_{\text{th}} = 0,001\ 159\ 652\ 153\ 5(24\ 0) \ ,$$

tandis que la valeur expérimentale est :

$$a_{\text{exp}} = 0,001\ 159\ 652\ 180\ 85\ (76) \ .$$

L'électrodynamique quantique prédit également des corrections quantiques à la masse m_B et la charge e_B, à cause des interactions entre les particules de la théorie et la mer de particules virtuelles environnantes que cette dernière prédit. Les valeurs mesurables de la charge e et de la masse m des électrons, par exemple, diffèrent des paramètres e_B et m_B qui apparaissent dans l'action. Ces corrections quantiques dépendent de l'énergie des particules, et c'est donc aussi le cas des valeurs de e et m qui dépendent de l'énergie de l'électron. Cette dépendance des paramètres de la théorie en l'énergie s'appelle la *renormalisation*. On introduit souvent la *constante de structure fine*, qui est adimensionnée :

$$\alpha(\mu) = \frac{e(\mu)^2}{4\pi\epsilon_0 \hbar c} \ ,$$

où μ est l'énergie. Expérimentalement, on trouve par exemple $\alpha^{-1}(0\text{ GeV}) \sim 137$, et $\alpha^{-1}(90\text{ GeV}) \sim 127$.

Les énergies sont exprimées ici en puissances d'électronvolts, comme il est courant de le faire en physique des particules : 1 eV $\sim 1,6 \times 10^{-19}$ J. L'énergie de Planck vaut par exemple $E_P \sim 10^{19}$ GeV. Il est également usuel d'exprimer les masses en eV/c^2, ou simplement en puissances d'électronvolts (en unités naturelles). Par exemple, les particules que nous avons rencontrées jusqu'à maintenant ont les masses suivantes :

- Photon : 0 eV,
- Électron : 511 keV,
- Proton : 938 MeV,
- Neutron : 940 MeV,

Porté par les succès de l'électrodynamique quantique, le formalisme mathématique de cette théorie – manifestement adapté à la description des interactions fondamentales – est développé en tant que tel. On parle de *théorie quantique des champs*, ou de *théorie des champs quantiques*. Comme dans le cas de la QED, on définit souvent une théorie quantique des champs en spécifiant le type de champs qu'elle contient, ainsi qu'une fonctionnelle des champs encore appelée *action*, qui détermine la dynamique et dépend de paramètres. Lorsque la théorie est presque libre, on peut utiliser des *techniques perturbatives* – notamment les graphes de Feynman – pour la résoudre de façon approchée. On dit que la théorie est dans son *régime perturbatif*, ou *à couplage faible*. Inversement, lorsque la théorie est loin d'être libre, on parle de *régime non-perturbatif* où la théorie est *à couplage fort*. La dynamique est plus difficile à résoudre car il n'existe pas de méthode générale pour cela.

Les paramètres d'une théorie quantique des champs doivent en général être renormalisés, comme la masse et la charge de l'électron en QED. La dépendance d'un paramètre g en l'échelle d'énergie μ est encodée par la *fonction β associée* :

$$\beta(g) = \frac{\partial g}{\partial \ln(\mu)} \ .$$

En électrodynamique quantique, la fonction β correspondant à la constante de structure fine α, *calculée par des techniques perturbatives*, vaut :

$$\beta(\alpha) = \frac{2\alpha^2}{3\pi} \ ,$$

Lorsque l'énergie est très faible ($\mu \sim 0$), on a vu que $\alpha^{-1} \sim 137,036$. La théorie est donc dans un régime perturbatif, ce qui justifie d'ailleurs l'emploi de techniques perturbatives. La fonction β indique que la valeur de α croit avec μ ; il existe une énergie μ_0 telle que $\alpha(\mu_0) \sim 1$ au delà de laquelle les techniques perturbatives deviennent caduques, car la QED est fortement couplée. La dynamique de la QED est alors en principe difficile à étudier. En pratique, ce *pôle de Landau* Λ_{Landau} n'est pas problématique en tant que tel, en particulier car $\Lambda_{\text{Landau}} \sim 10^{177}$ GeV $\gg \Lambda_P$, échelle à laquelle une théorie unifiant les quatre interactions fondamentales doit exister : individuellement, la QED n'a alors plus vraiment de sens. De plus, nous verrons plus tard que les interactions électromagnétiques et faibles fusionnent en l'interaction *électrofaible* autour de 246 GeV, et donc que la seule chose qui importe est de pouvoir parler de la QED entre 0 eV et quelques centaines de GeV.

Pour sonder un objet de taille L avec une onde, il faut que la longueur d'onde de cette dernière soit comparable avec L. D'après la relation de de Broglie $E = hc/\lambda$, l'énergie est inversement proportionnelle à la distance sondée comme illustré en Figure 9. Pour comparaison, les collisions actuelles au Large Hadron Collider, au CERN, se font à une énergie de 14 TeV et permettent ainsi de sonder jusqu'à 10^{-21} m. Par analogie avec la lumière visible, la limite des grandes énergies (c'est-à-dire des petites distances) est appelée *ultraviolet* (UV), tandis que la limite des petites énergies (c'est-à-dire des grandes distances) est appelée *infrarouge* (IR). Ainsi, on dit que l'électrodynamique quantique est libre dans l'IR et devient de plus en plus fortement couplée dans l'UV.

La renormalisation des constantes des théories quantiques des champs est liée à la *régularisation* de certains infinis qui apparaissent lors de calculs de graphes de Feynman à boucles. Dans les cas favorables, ces infinis peuvent être remplacés par des *contre-termes* qui implémentent la renormalisation dans l'action

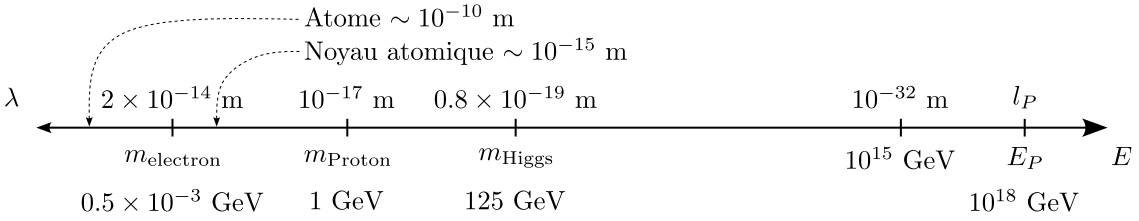


Figure 9: Longueurs d'ondes de de Broglie et énergies à l'échelle subatomique.

de la théorie. Lorsqu'une théorie quantique des champs est telle qu'on peut absorber tous ses infinis dans un nombre fini de tels contre-termes, on dit que cette théorie est *renormalisable*.

En 1939, Fierz énonce une première version du théorème de spin-statistique, reformulé par Pauli en 1940 puis Schwinger et Feynman dans les années 1950. Chaque champs de particule en théorie quantique des champs est caractérisé partiellement par son *spin*, c'est-à-dire son moment angulaire intrinsèque, qui prend des valeurs positives demi-entières : 0, 1/2, 1, 3/2, ... Le théorème de spin-statistique indique que les particules de spin entier sont des bosons (la fonction d'onde est alors paire sous l'échange de deux particules) tandis que les particules de spin strictement demi-entier sont des fermions (la fonction d'onde est impaire sous l'échange de deux particules).

Interactions nucléaires faibles et fortes

Dans les années 30, il est communément admis que les noyaux atomiques sont constitués de protons et de neutrons, et que ces derniers sont des particules élémentaires. Les protons étant électriquement chargés (positivement) et les neutrons étant neutres, les noyaux seraient instables si rien ne compensait la répulsion électrostatique. Ce raisonnement conduit à la conjecture de l'existence d'une nouvelle interaction, appelée *force nucléaire forte*. En 1935, Yukawa propose l'hypothèse que la force nucléaire entre protons et neutrons est transmise par une ou des particules massives. Leur masse implique notamment que ces vecteurs ont un temps de vie court, donc que la portée de cette force nucléaire est limitée, et enfin que les gros noyaux atomiques sont instables car leur rayon est supérieur à la portée de l'interaction forte. Yukawa appelle ces particules des *mésons* (de la racine grecque 'mesos' qui signifie intermédiaire) car, estimant leur masse par des raisonnements heuristiques, il trouve qu'elle doit se situer entre celle d'un électron ($\sim$ 500 keV) et celle d'un proton ($\sim$ 1 GeV).

Par ailleurs, la théorie de Fermi de la désintégration β ne peut vraisemblablement pas être décrite en QED, ni par ces interactions fortes supposées : une quatrième force fondamentale est donc proposée, la *force nucléaire faible*, qui tient son nom de la lenteur des désintégrations β comparativement aux interactions électromagnétiques ou fortes, ce qui indique que l'interaction responsable de ces désintégrations est plus faible que ces dernières.

Le développement des chambres à brouillard (inventées en 1911 par Wilson et améliorées en 1936 par Langsdorf) et des chambres à bulles (inventées en 1952 par Glaser), ainsi que des premiers accélérateurs de particules dans les années 1940, ont permis de nombreuses découvertes et ainsi joué un rôle déterminant, entre les années 1930 et 1980, dans l'accroissement de la compréhension des interactions nucléaires faibles et fortes. La majorité des particules mentionnées ci-dessous sont instables, et se désintègrent en d'autres particules en un temps plus ou moins court. C'est en étudiant la désintégration et ses produits qu'on peut remonter aux propriétés de la particule initiale. Les particules sensibles à l'interaction forte sont appelées *hadrons* (de la racine grecque 'hadros' : fort), et les autres, *leptons* (de la racine grecque 'leptos' : faible).

- 1936 – Découverte du *muon* μ par Anderson à partir de l'étude de rayons cosmiques. Cette particule a une masse d'environ 106 MeV, proche de celle prédite par Yukawa. Cela fait qu'elle est appelée *méson* μ durant les premières années de son histoire, cependant rapidement il devient clair qu'elle ne participe en fait pas aux interactions fortes : il s'agit donc d'un *lepton*. Le muon est une sorte d'électron, 200 fois plus massif que ce dernier.

- 1947 – Découverte des mésons $\pi^{\pm}$, ou *pions*, de masse 140 MeV par Lattes, Occhialini et Powell. Les pions $\pi^{\pm}$ sont instables, et ont un temps de demi-vie de $2,6 \times 10^{-8}$ s. Les pions sont des hadrons.

- 1947 – Découverte des mésons $K^{\pm}$ (de masse 494 MeV), ou *kaons* et K^0 (de masse 498 MeV) par Rochester and Butler. Les kaons sont des hadrons.

- 1950 – Découverte du méson π^0, de masse 135 MeV. Les pions π^0 sont instables, et ont un temps de demi-vie de $8,5 \times 10^{-17}$ s. Les π^0 sont également des hadrons.

- 1950 – Découverte du *baryon* Λ^0 par Hopper et Biswas, qui est un hadron de masse $m(\Lambda^0) \sim$ 1115 MeV. Il se distingue des mésons par le fait qu'il peut se désintégrer en un proton et d'autres particules (ce qui définit les baryons par opposition aux mésons). Son temps de demi-vie est de 10^{-10} s, contrairement aux 10^{-23} s attendues.

- 1953 – Découverte du baryon Ξ^- par Armenteros et al., de masse 1314 MeV.

- 1953 – Gell-Mann, Pais, Nakano et Nishijima introduisent *l'étrangeté* S, un *nombre quantique* associé aux hadrons, qui serait conservée par les interactions fortes et pas par les interactions faibles, afin d'expliquer les temps de demi-vie des kaons et du baryon Λ. Par exemple, on a $S(\pi^{\pm}) = S(\pi^0) = 0$, $S(K^+) = S(K^0) = 1$, $S(K^-) = S(\overline{K^0}) = -1$, $S(\Lambda^0) = 1$, $S(\Xi^-) = -2$.

- 1954 – Yang et Mills étendent le concept de symétrie de jauge auparavant défini pour des groupes abéliens (comme c'est le cas en électrodynamique quantique) à des groupes non-abéliens. Ces théories sont désormais appelées *théories de Yang–Mills*.

- 1956 – Cowan, Reines, Harrison, Kruse and McGuire détectent le neutrino électronique ν_e, qui est un lepton de masse apparemment nulle.

- 1956 – *Question of Parity Conservation in Weak Interactions* de Lee and Yang, qui soulignent que bien que la symétrie de parité ait été vérifiée expérimentalement pour des interactions électromagnétiques et fortes, elle n'a jamais fait l'objet d'une étude en ce qui concerne les interactions faibles. La symétrie de parité $\mathbf{P}$ agit sur les coordonnées dans un repère de l'espace-temps de Minkowski de la manière suivante :

$$(t, x, y, z) \longrightarrow (t, -x, -y, -z) \ .$$

  Qu'une interaction ne respecte pas $\mathbf{P}$ peut-être reformulé comme le fait que cette interaction ne se passe pas de la même manière dans notre univers, et dans une copie miroir de ce dernier.

  Si l'on s'autorise à considérer des particules qui ne sont pas identifiées sous $\mathbf{P}$ (c'est-à-dire qui diffèrent de leur image miroir), les fermions de Dirac (comme l'électron) ne sont pas les plus élémentaires possibles. On peut en effet distinguer deux fermions dits *de Weyl* dans chaque fermion de Dirac : un de *chiralité droite* et l'autre de *chiralité gauche*. Sous la symétrie $\mathbf{P}$, les chiralités sont échangées. Se poser la question de savoir si les interactions faibles respectent la parité revient donc à se demander si les fermions de chiralité gauche subissent les mêmes interactions faibles que les fermions de chiralité droite.

- 1957 – Wu conduit une expérience de désintégration β du Cobalt 60 :

$$^{60}_{27}\mathrm{Co} \longrightarrow {}^{60}_{28}\mathrm{Ni}^* + e^- + \overline{\nu_e} \longrightarrow {}^{60}_{28}\mathrm{Ni} + e^- + \overline{\nu_e} + 2\gamma$$

  qui montre que les interactions faibles ne respectent pas la parité. Cela résout le *mystère* $\tau - \Theta$: les mésons τ^+ et Θ^+ étaient identifiés comme deux particules pouvant notamment se désintégrer en

$$\tau^+ \longrightarrow \pi^+ + \pi^0 \ ,$$
$$\Theta^+ \longrightarrow \pi^+ + \pi^+ + \pi^- \ .$$

  D'une part, $\mathbf{P}(\pi^+ + \pi^0) = 1$ et d'autre part, $\mathbf{P}(\pi^+ + \pi^+ + \pi^-) = -1$, ce qui, en supposant que la parité est conservée dans ces désintégrations, semble indiquer que τ^+ et Θ^+ sont deux particules différentes. Cependant, les données expérimentales ne permettaient pas de distinguer ces deux particules, qui ont notamment exactement la même masse. Il s'agit en fait de la même particule, le kaon K^+, et sa désintégration en deux états de différentes parités est possible puisque les interactions faibles ne préservent pas $\mathbf{P}$.

L'expérience de Wu implique aussi que les interactions faibles violent la conjugaison de charge **C**. Cette symétrie discrète échange les particules avec leurs anti-particules, c'est-à-dire, inverse le signe de toutes les charges.

La dernière symétrie Lorentzienne discrète est le renversement du temps **T**, qui dans un repère agit comme $(t, x, y, z) \to (-t, x, y, z)$. Le célèbre *théorème CPT*, prouvé dans les années 1950 par Schwinger, Lüders et Pauli, affirme que toute théorie quantique des champs relativiste dans l'espace de Minkowski $\mathbb{R}^{1,3}$ est symétrique sous la combinaison **CPT** des trois symétries Lorentziennes discrètes.

- 1957 – Landau propose que la combinaison **CP** est la bonne symétrie à considérer entre la matière et l'antimatière, et que **CP** est conservée dans les interactions faibles.

- 1959 – Découverte du baryon Ξ^0 par Alvarez et al., de masse 1314 MeV et d'étrangeté $S(\Xi^0) = -2$.

- 1961 – *The renormalizability of vector meson interactions*, Glashow, où l'auteur propose un premier modèle pour unifier les interactions faibles et l'électromagnétisme, en supposant que l'interaction faible est *transmise* par un échange de bosons vecteurs, plutôt qu'une interaction de quatre fermions comme dans la théorie de Fermi.

- 1961 – Découverte des mésons η et η' par Pevsner et al., qui ont pour masse $m(\eta) \sim 548$ MeV et $m(\eta') \sim 958$ MeV.

- 1961 – Gell-Mann introduit la *voie octuple* ("Eightfold way", en anglais) pour classifier les mésons et baryons. Il s'agit de classifier ces derniers en représentations du groupe SU(3), ce qu'on obtient en plaçant les particules dans un plan paramétré par la charge électrique et l'étrangeté. Par exemple, le méson η' est un singulet de SU(3) tandis que le méson η, les pions, et les kaons forment un octet, comme montré en Figure 10.

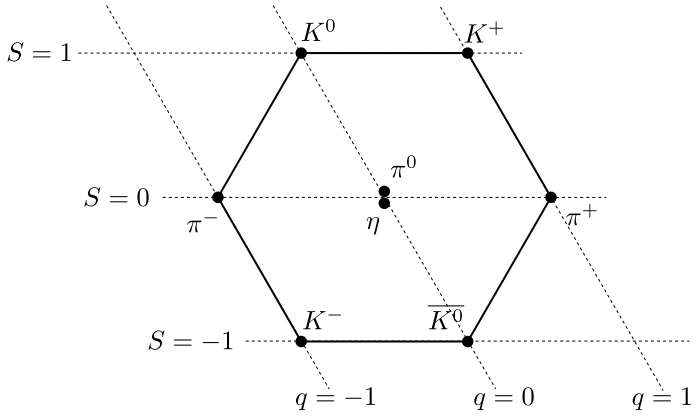


Figure 10: L'octet des pions, kaons et η.

- 1961 – Nambu et Jona-Lasinio étudient les conséquences de l'existence de symétries chirales en théorie quantique des champs et de leur brisure spontannée. Ce formalisme suggère notamment que la masse des nucléons est principalement due à l'auto-interaction d'un champ fermionique élémentaire, et prédit l'existence de particules légères indentifiées aux pions.

- 1962 – Lederman, Schwartz et Steinberger font la première observation d'une interaction impliquant un neutrino muonique ν_μ, un lepton de masse apparemment nulle similaire au neutrino électronique.

- 1964 – *Electromagnetic and weak interactions*, Salam et Ward. Ce deuxième modèle prédit un photon ainsi que trois bosons de jauge massifs, avec la symétrie de jauge correspondante brisée explicitement.

- 1964 – La brisure spontanée de symétrie de jauge en théorie quantique de champs relativiste est développée par Brout et Englert, Higgs, ainsi que Guralnik, Hagen et Kibble.

- 1964 – Cronin et Fitch montrent que la symétrie **CP** peut être brisée lors de la désintégration des kaons, ce qui remet en cause l'hypothèse de Landau de 1957.

- 1964 – Gell-Mann et Zweig introduisent le modèle des *quarks* qui prolonge la classification la voie octuple de Gell-Mann. L'idée centrale est que les hadrons ne seraient pas des particules élémentaires, mais des combinaisons de quarks et d'antiquarks, qui sont des fermions. Ce premier modèle comprend trois quarks, dénommés *up* (u), *down*(d) et *strange* (s) ainsi que leurs antiparticules $\overline{u}$, $\overline{d}$ et $\overline{s}$. Leurs charges électriques q valent $q(u) = -q(\overline{u}) = 2/3$, $q(d) = q(s) = -q(\overline{d}) = -q(\overline{s}) = -1/3$ et leurs étrangetés : $S(u) = S(d) = S(\overline{u}) = S(\overline{d}) = 0$, $S(s) = -1$ et $S(\overline{s}) = 1$. Les compositions conjecturées en quarks de quelques uns des mésons et baryons évoqués ci-dessus sont :

$$\begin{aligned}
\text{proton } p &: uud, \\
\text{neutron } n &: udd, \\
\pi^+ &: u\overline{d}, \\
\pi^- &: d\overline{u}, \\
\pi^0 &: \frac{1}{\sqrt{2}}\left(u\overline{u} - d\overline{d}\right), \\
K^+ &: u\overline{s}, \\
\eta &: \frac{1}{\sqrt{6}}\left(u\overline{u} + d\overline{d} - 2s\overline{s}\right), \\
\eta' &: \frac{1}{\sqrt{6}}\left(u\overline{u} + d\overline{d} + s\overline{s}\right), \\
\Lambda^0 &: uds.
\end{aligned}$$

  La mesure ultérieure de la masse de ces quarks donne $m(u) \sim 2.3$ MeV, $m(d) \sim 4,8$ MeV et $m(s) \sim 95$ MeV.

- 1964 – Glashow et Bjorken prédisent l'existence d'un quatrième quark, le *charm* (c), qui permet une meilleure description des interactions faibles.

- 1964 – Découverte du baryon Ω^- (sss) d'étrangeté -3, de masse 1672 MeV et de spin 3/2, ce qui pose un problème puisque les quarks sont des fermions et satisfont le principe d'exclusion de Pauli. Le même problème existe en ce qui concerne le baryon Σ^{++} (uuu), de spin 3/2.

- 1964 - 1965 – Greenberg, Han et Nambu proposent que les quarks possèdent une symétrie de jauge SU(3) additionnelle.

- 1965 – Struminsky propose que les quarks possèdent des nombres quantiques supplémentaires encore non découverts afin de résoudre le problème de l'existence du baryon Ω^- (et également du Σ^{++}) : ce seront les *couleurs* bleue, rouge et verte qui seront assignées aux quarks en chromodynamique quantique.

- 1967 – *A model of leptons*, Weinberg. Dans ce troisième modèle électrofaible est introduite l'idée de brisure spontanée de symétrie (c'est-à-dire, le mécanisme de Higgs) dans une théorie de jauge de groupe $\mathrm{SU}(2)_I \times \mathrm{U}(1)_Y$ (où I désigne l'isospin faible, et Y, l'hypercharge faible). Le facteur $\mathrm{SU}(2)_I$ est chiral, et n'interagit qu'avec les fermions (électrons, neutrinos et leurs équivalents) de chiralité gauche, tandis que $\mathrm{U}(1)_Y$ n'est pas chiral et interagit avec les fermions indépendamment de leur chiralité. Un doublet de Higgs est introduit, qui brise spontanément $\mathrm{SU}(2)_I \times \mathrm{U}(1)_Y$ en U(1), où ce dernier facteur est le groupe de jauge de l'électromagnétisme. La théorie prévoit notamment l'existence de trois bosons de jauge massifs vecteurs de l'interaction nucléaire faible, dénotés W^-, W^+ et Z^0 et tels que $m_{Z^0} > m_{W^\pm}$, $m_{Z^0} > 80$ GeV et $m_{W^\pm} > 40$ GeV.

- 1968 – Des expériences de *diffusion profondément inélastique* conduites au "Stanford Linear Accelerator Center" (SLAC) montrent que le proton est constitué d'objets ponctuels qui sont ensuite identifiés aux *quarks de valence* du proton : uud. C'est la première observation expérimentale des quarks up et down.

- 1968 - 1973 – Le *modèle dual de résonance* est développé pour expliquer les interactions fortes. C'est une théorie de cordes ouvertes microscopiques, dont les extrémités sont les quarks.

- 1969 – Adler, Bell et Jackiw calculent que le courant chiral axial $\overline{\psi}\gamma^5\gamma^\mu\psi$ en QED est anomal : bien que l'action de la QED (avec des quarks up et down) satisfasse la symétrie chirale correspondant à la conservation de ce courant, cette symétrie est brisée par des effets quantiques à une boucle. Cela explique notamment pourquoi le temps de demi-vie du pion π^0, qui est de $8,5 \times 10^{-18}$ s, est tant plus faible que celui des pions $\pi^\pm$ ($2,6 \times 10^{-8}$ s) : le pion π^0 peut se désintégrer en deux photons via une boucle triangulaire de quarks qui a une amplitude non-nulle (présentée en Figure 11 où les lignes continues sont des (anti)quarks up ou down), tandis que les pions chargés $\pi^\pm$ ne peuvent pas se désintégrer de cette manière. Le temps de demi-vie de ces derniers est donc plus long.

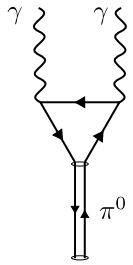


Figure 11: Le graphe de Feynman d'amplitude non-nulle qui correspond à l'anomalie ABJ.

Plus généralement, on dit qu'une symétrie des équations du mouvement classique d'une théorie des champs est *anomale* si cette symétrie est brisée par les corrections quantiques. Dans les théories quantiques des champs *chirales* où les fermions gauches et droits ne couplent pas de la même manière aux autres champs, certains courants *chiraux* souffrent d'anomalies. Lorsque ces derniers correspondent à des symétries globales (comme c'est la cas en QCD – voir plus bas), ces anomalies ne sont en rien problématiques et fournissent des informations précieuses sur la théorie. Inversement, si les courants correspondent à des symétries de jauge, une anomalie implique une *violation de l'unitarité* de la théorie, qui est alors mal définie en tant que théorie quantique. L'absence d'anomalies de jauge impose une condition sur la quantité de fermions chiraux gauches et droits qui couplent à chaque facteur simple du groupe de jauge. Une application importante de l'étude de ces anomalies de jauge est la *condition d'appariement des anomalies* de 't Hooft (1980), qui permet notamment d'obtenir des informations sur la dynamique à couplage fort d'une théorie quantique des champs, à partir de la connaissance de cette théorie à couplage faible.

- 1970 – Glashow, Iliopoulos et Maiani proposent leur *mécanisme GIM* pour expliquer l'absence expérimentale de "Flavor changing neutral currents".

- 1973 – Kobayashi et Maskawa étendent les travaux de Cabibbo, en montrant qu'on peut expliquer la violation de symétrie **CP** dans les interactions faibles si l'on suppose l'existence de deux quarks supplémentaires, les quarks *top* (t) et *bottom* (b). La *matrice CKM* est une matrice unitaire de passage entre la base des états de quarks libres et la base naturelle des états de quarks au regard des interactions faibles. Le carré de la valeur absolue des entrées donne la *probabilité* qu'un quark de type up donné se transforme en un certain quark de type down lors d'une interaction avec les bosons faibles $W^\pm$.

- 1973 – Fritzsch, Gell-Mann et Leutwyler développent la théorie de la chromodynamique quantique (QCD) pour expliquer les interactions fortes. La QCD est une théorie de Yang–Mills de groupe de jauge $\mathrm{SU}(3)_c$ appelé le groupe de *couleur*, en interaction avec les quarks qui sont des fermions de Dirac dans la représentation fondamentale de SU(3). Autrement dit, on assigne aux quarks une couleur (rouge, vert ou bleu) qui est l'équivalent de la charge électrique pour l'interaction électromagnétique. L'équivalent du photon de la QED consiste en huit *gluons* g, qui sont des bosons de masse nulle eux-même chargés sous le groupe de couleur (c'est une propriété des théories de Yang–Mills non-abéliennes). Plus précisément, les gluons sont dans la représentation adjointe de $\mathrm{SU}(3)_c$ et on peut donc leur assigner une couleur et une "anti-couleur".

- 1973 – Gross, Wilczek et Politzer découvrent la *liberté asymptotique* de certaines théories de Yang–Mills non-abéliennes, et en particulier de la QCD. La renormalisation de la constante de couplage g de la QCD est déterminée par la fonction β (à une boucle) :

$$\beta(g_s) = -\left(11 - \frac{2n_f}{3}\right)\frac{g_s^2}{16\pi^2} ,$$

où n_f est le nombre de *saveurs* de quarks, c'est-à-dire six (up, down, charm, strange, top, bottom). Ainsi le coefficient de la fonction β de la QCD est négatif, et donc plus l'énergie est grande plus la constante de couplage est petite. Cette dernière tend vers 0 (et la QCD tend vers une théorie libre) dans l'extrême UV, et c'est pour cela qu'on parle de *liberté asymptotique*. Inversement, la constante de couplage devient de plus en plus grande dans l'infrarouge, et d'ordre 1 autour de $\Lambda_S \sim 332$ MeV. Cette échelle d'énergie ne provient pas d'un terme dans l'action de la QCD, mais de la dynamique de la théorie elle-même. On parle d'*échelle d'énergie dynamique*. Le comportement et la description de la QCD à une échelle d'énergie μ dépendent crucialement de comment μ se compare à Λ_S.

Lorsque $\mu \gg \Lambda_S$, la QCD est dans son régime perturbatif, et est bien décrite par des quarks interagissant faiblement via le groupe de couleur $\mathrm{SU}(3)_c$. Inversement, lorsque $\mu \ll \Lambda_S$, la QCD est dans un régime de couplage fort, et la description perturbative en termes de quarks et de gluons n'est plus adaptée. Il y a même *confinement* de la couleur : les quarks et les gluons forment des états liés, qui sont les baryons et mésons observés à basse énergie. Une partie importante de la masse de ces états liés provient de l'énergie de liaison de ces états composites ; par exemple la somme des masses des quarks *uud* constitutifs du proton est de 9.4 MeV, tandis que $m(p) \sim 938$ MeV. Puisque la couleur est confinée, la chromodynamique quantique donne lieu à une *force nucléaire résiduelle* entre les protons et les neutrons, principalement transmise par l'échange de pions qui correspondent donc bien aux particules dont l'existence a été prédite par Yukawa.

Les six saveurs de quarks de la QCD sont naturellement divisées en deux groupes : les quarks up, down et strange ont une masse de moins de 100 MeV tandis que les quarks charm, top et bottom ont une masse de plus de 1 GeV. Dans la limite où l'on suppose que la masse de ces trois premiers quarks est nulle, la théorie jouit d'une symétrie globale chirale $\mathrm{SU}(3)_L \times \mathrm{SU}(3)_R$ qui agit indépendamment sur les parties de chiralité gauche et droite de ces trois quarks. Cette symétrie chirale est brisée dynamiquement dans le vide de la QCD en le sous-groupe diagonal $\mathrm{SU}(3)_D$, ce qui doit donner lieu à des bosons de Goldstone de masse nulle. En restaurant la masse des quarks u, d et s, on obtient des pseudo bosons de Goldstone dont la masse, bien que non-nulle, est faible. Ce sont les pions, kaons et η constitutifs de l'octet donné en Figure 10.

- 1974 – Observation expérimentale du quark *charm* au SLAC et au Laboratoire National de Brookheaven, à travers la découverte du méson J/ψ ($c\bar{c}$) de masse 3 GeV. Le quark charm a quant à lui une masse de 1,28 GeV.

- 1974 – Wilson introduit les opérateurs maintenant appelés *boucles de Wilson*, qui sont des *paramètres d'ordre* pour le confinement : dans une phase confinée, ces opérateurs suivent une loi d'aire, tandis que dans une phase de Higgs, ils suivent une loi de périmètre.

- 1975 – Découverte du tau (τ) au "Stanford Linear Accelerator Center", qui est un lepton de type électron et de masse $m(\tau) \sim 1,8$ GeV.

- 1977 – Observation expérimentale du quark *bottom* (b), de masse $m(b) \sim 4,18$ GeV.

- 1978 – 't Hooft introduit les *boucles de 't Hooft* qui sont les analogues magnétiques de boucles de Wilson.

- 1983 – Découverte des bosons W^+, W^- et Z^0 par les collaborations UA1 et UA2 au CERN. Leur masse est $m(W^\pm) \sim 80$ GeV et $m(Z^0) \sim 91$ GeV, en accord avec les prédictions du modèle électrofaible de Weinberg de 1967.

- 1995 – Observation expérimentale du quark *top* (t) par les collaborations CDF et DØ au Tévatron. La masse de ce quark est $m(t) \sim 173$ GeV.

- 2000 – Observation du neutrino tauique ν_τ au CERN, un lepton de masse apparemment nulle qui est à la particule τ ce que le neutrino électronique (respectivement, muonique) est à l'électron (respectivement, muon).
- 2012 – Découverte du boson de Higgs H^0 par les collaborations ATLAS et CMS au CERN. Sa masse est de $m(H^0) \sim 125$ GeV.

Le Modèle Standard de la physique des particules

La théorie qui décrit, de manière unifiée, les interactions électromagnétiques, faibles et fortes, est le fruit des décennies de recherche rapidement décrites dans la section précédente, et porte le nom de Modèle Standard de la physique des particules. Il s'agit d'une théorie quantique des champs relativiste définie dans l'espace-temps de Minkowski $\mathbb{R}^{1,3}$. Les particules élémentaires qu'elle contient peuvent être classifiées de la manière suivante :

- Les **bosons**, divisés en deux sous-groupes.
  - Les **bosons de jauge** sont les particules vectrices des interactions elles-mêmes. Ce sont des particules de spin 1 :
    - Le *photon* γ, de masse nulle, vecteur de l'interaction électromagnétique.
    - Les *bosons faibles* W^+, W^- *et* Z^0, vecteurs de l'interaction faible, qui acquièrent spontanément une masse par la brisure de la symétrie de jauge dans le modèle électrofaible déclenchée par le doublet de Higgs. La masse de ces bosons explique la portée finie des interactions faibles, à la différence de l'électromagnétisme. Le boson W^+ (resp. W^-, Z^0) est électriquement chargé positivement (resp. chargé négativement, neutre), et n'a aucune charge de couleur.
    - Les *huit gluons g* qui sont les vecteurs des interactions de couleur en QCD, de masse nulle. Les gluons sont neutres pour l'interaction électrofaible, et chargés pour les interactions de couleur. En dessous de l'échelle d'énergie dynamique Λ_S des interactions fortes, la couleur est confinée, ce qui explique la portée finie des interactions fortes d'une manière très différente de celle des interactions faibles : à des distances de l'ordre du noyau atomique, ce sont les pions qui transmettent l'interaction forte et non plus les gluons.
  - Le **boson de Higgs** H^0 est l'excitation massive qui résulte de la brisure spontanée de la symétrie de jauge électrofaible par le doublet de Higgs. Il s'agit d'une particule de spin 0.
- Les **fermions** du modèle standard sont des particules de spin 1/2 qui constituent la "matière". Ils sont également divisés en deux sous-groupes :
  - Les **leptons** sont, par définition, ceux qui ne sont pas chargés sous l'interaction forte. Il y a trois *générations* rassemblées en doublets :

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}, \quad \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix}, \quad \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix},$$

    respectivement : l'*électron*, le *neutrino électronique*, le *muon*, le *neutrino muonique*, le *tau*, le *neutrino tauique*. Les neutrinos sont électriquement neutres, tandis qu'électrons, muons et taus sont chargés négativement (et leur antiparticule, positivement). Par émission ou absorption d'un boson faible $W^\pm$, un composant d'un doublet est transformé en l'autre.
  - Les **quarks** sont chargés pour l'interaction forte. Il y a six *saveurs* de quarks dans le modèle standard : *up*, *down*, *charm*, *strange*, *top*, *bottom*. Les quarks up, charm et top sont dits de *type up* tandis que les quarks down, strange et botton sont dits de *type down*. De plus, chaque quark a une couleur (rouge, vert ou bleu) qui peut changer par interaction avec les gluons. Par exemple, un quark up rouge peut absorber un gluon vert-antirouge et devient ainsi un quark up vert. Les quarks sont chargés électriquement : les quarks de type up ont une charge $2/3$ et ceux de type down, une charge $-1/3$. Leurs anti-particules ont des charges opposées. Les quarks sont également chargés pour l'interaction faible qui ne préserve pas les saveurs des quarks. La matrice CKM (1973), dont les entrées sont mesurées expérimentalement, décrit ces changement de saveur. Par exemple, la probabilité qu'un quark up devienne un quark down (respectivement, strange, bottom) par émission d'un boson W^+ est de 94,8% (respectivement, 5,0%, 0,001%).

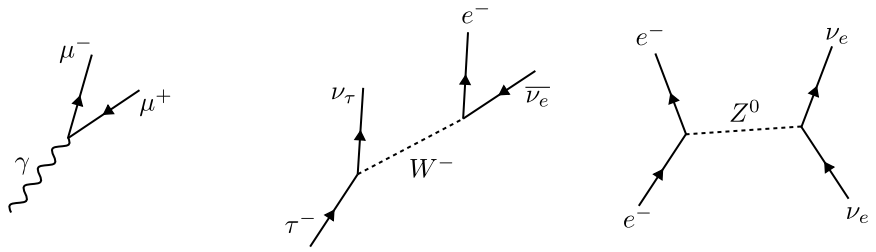


Figure 12: Quelques interactions leptons-leptons décrites par le Modèle Standard.

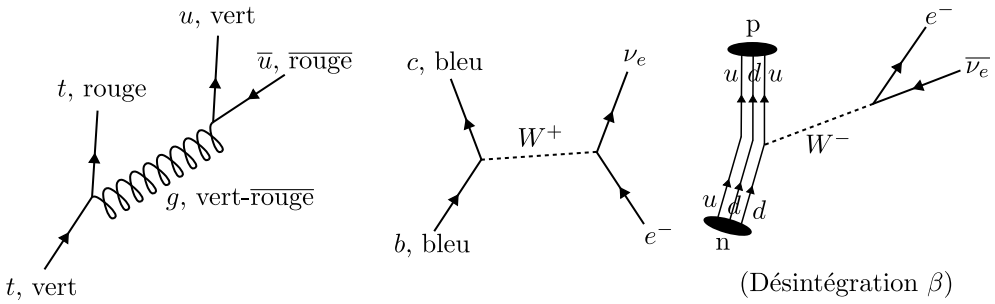


Figure 13: Quelques interactions de quarks décrites par le Modèle Standard.

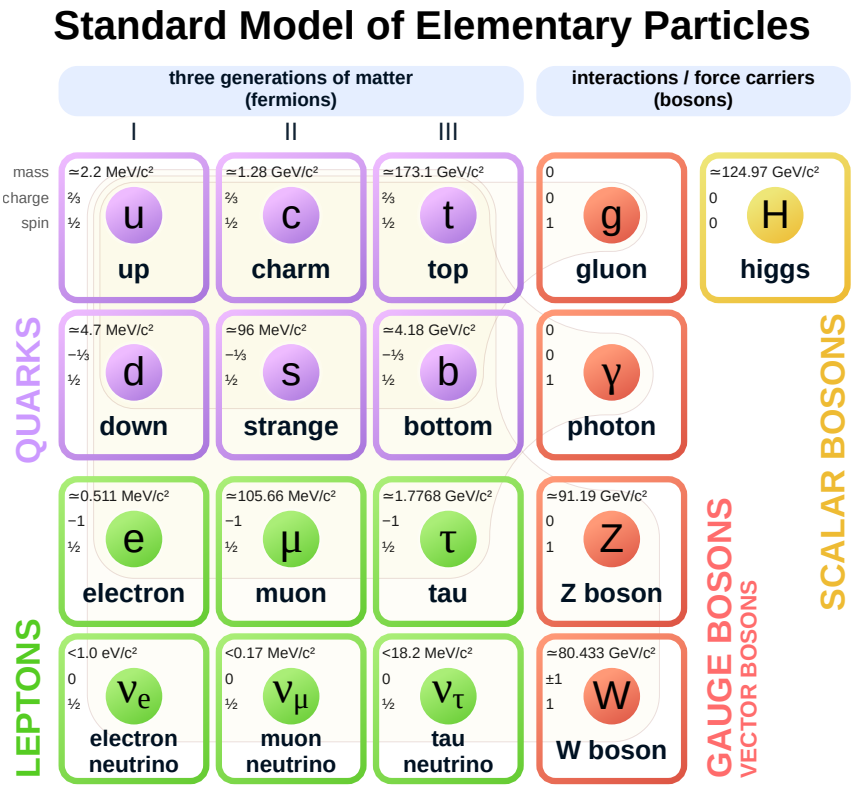


Figure 14: Particules du Modèle Standard (tableau de Cush obtenu à partir d'une version de MissMJ).

L'accord entre les prédictions théoriques du Modèle Standard et les mesures expérimentales est remarquable. Cependant, cette théorie ne peut pas être complète, et ce pour plusieurs raisons, dont les suivantes :

1. Le Modèle Standard de la physique des particules ne décrit pas la gravité. À des échelles d'énergies suffisamment faibles devant l'énergie de Planck $\Lambda_P \sim 10^{18}$ GeV, la gravité a en effet un impact négligeable sur les particules, devant les autres interactions que le Modèle Standard décrit. Par

exemple, la répulsion électrique entre deux électrons statiques séparés de 1 m est de quelques 10^{-28} N, tandis que l'attraction gravitationnelle (calculée selon la loi de Newton) est de 10^{-71} N. Cependant, la gravité ne peux plus être négligée à des énergies de l'ordre de l'échelle de Planck. Ainsi, le Modèle Standard ne peut décrire convenablement les interactions électromagnétiques, faibles et fortes qu'à des énergies suffisamment faibles, inférieures à une certaine échelle Λ_{SM}. Les expériences conduites au LHC montrent que le Modèle Standard décrit correctement la nature à des énergies inférieures à 14 TeV, c'est-à-dire à des distances supérieures à 10^{-21} m. Autrement dit, on sait que $\Lambda_{\mathrm{SM}} \geq 14$ TeV.

Un traitement quantique perturbatif des équations d'Einstein (c'est-à-dire de la gravité) est possible, et par quantification on obtient une particule sans masse de spin 2 appelée *graviton* (qui n'a pas encore été mise en évidence expérimentalement). Cependant, la théorie obtenue n'est pas renormalisable et ne peut donc pas être considérée comme une théorie complète de la gravité quantique : il s'agit tout au plus d'une *théorie effective*, valable seulement jusqu'à une certaine énergie plus faible que l'énergie de Planck. Les recherches dans cette direction sont limitées par les difficultés expérimentales quant aux tests des lois de gravitation pour des petites masses : actuellement, la loi de Newton n'a été vérifiée que pour des masses supérieures à quelques fractions de grammes. Les mécanismes de la gravité à l'échelle de particules très énergétiques (la masse de Planck est de l'ordre de 20 μg) sont encore à élucider.

2. Il a été montré que les particules décrites par le Modèle Standard ne peuvent pas expliquer entièrement la nature de la matière noire.

3. L'observation des neutrinos émis par le soleil par Davis et Bahcall dans les années 1960, confirmée par des expériences au "Sudbury Neutrino Observatory" et au "Super Kamiokande" a permis de montrer le phénomène dit *d'oscillation des neutrinos*, qui ne peut s'expliquer que si ces derniers ont une masse. Le Modèle Standard décrit quant à lui des neutrinos non massifs.

4. La renormalisation de la masse du champ de Higgs du modèle Standard souffre de divergences quadratiques. À cause de cela, et sauf si des suppressions extraordinaires ont lieu entre les différentes contributions à la masse du Higgs H^0 (ce qui est appelé *fine tuning*), on s'attend à ce que $m(H^0)$ soit typiquement de l'ordre de Λ_{SM}, ce qui est incompatible avec la masse mesurée $m(H^0) \sim 125$ GeV et la borne inférieure actuelle sur Λ_{SM}. On parle d'un problème de *naturalité*, ou de hiérarchie : comment expliquer que $m(H^0) \ll \Lambda_{\mathrm{SM}}$?

5. L'énergie du vide du Modèle Standard est estimée à quelques 10^{113} J $\cdot$ m^3, alors que la mesure de la constante cosmologique donne plutôt quelques 10^{-9} J $\cdot$ m^3. Cet écart est appelé *problème de la constante cosmologique*.

6. La violation de la symétrie **CP** dans les interactions faibles et les interactions fortes a été observée expérimentalement. Cependant, un terme dit *terme thêta* dans l'action de la chromodynamique quantique prévoit, encore une fois sauf s'il y a un *fine tuning* extraordinaire, une brisure de la symétrie **CP** beaucoup plus importante que celle observée. C'est un autre problème de naturalité appelé *problème de CP fort*. Peccei et Quinn ont proposé un mécanisme basé sur une brisure spontanée de symétrie permettant de résoudre ce paradoxe en 1977, qui est aujourd'hui très consensuel. Cependant, la théorie de Peccei–Quinn prévoit une nouvelle particule dénommée *axion*, et qui n'a toujours pas été observée. Il n'y a donc pas encore de confirmation directe de cette résolution du problème de CP fort.

Ainsi, malgré ses succès, le Modèle Standard appelle a être étendu en une théorie qui le contient, et qui résout certains des problèmes sus-mentionnés. Toutefois, le fait que les collisionneurs de particules les plus puissants construits aujourd'hui ne puissent pas contredire les prédictions du Modèle Standard donne une saveur nouvelle à cette tâche : il faut pouvoir approfondir la connaissance que nous avons de notre univers avec très peu d'indications expérimentales, voire aucunes. Ce constat explique le nouveau paradigme apparu en physique théorique dans les années 1980, et l'apparition d'une nouvelle branche dans cette discipline parfois appelée *mathématique physique*, d'essence et de contenus plus mathématiques, notamment incarnée par l'étude des théories quantiques des champs *supersymétriques* et par la *théorie des cordes*.

Mathématique physique contemporaine

Théories de jauge et géométrie

L'application des idées des théories de jauge non-abéliennes introduites par Yang et Mills en 1954 a été incroyablement fructueuse, en permettant notamment la construction de nouveaux invariants de variétés lisses, l'étude de structures géométriques comme les variétés hyperkählériennes ou encore l'étude de nombreux *espaces des modules* en géométrie algébrique. En mathématiques, on entend par théorie de jauge l'étude des connections sur des fibrés vectoriels et principaux en géométrie différentielle ou algébrique. L'introduction des fibrés de Higgs par Hitchin a notamment donné une généralisation très intéressante des espaces de Teichmüller dont nous aurons l'occasion de reparler. Quelques articles fondateurs de cette branche de la géométrie sont listés ci-dessous.

- 1977 - 1978 – *Deformation of instantons* et *Self-duality in four-dimensional Riemannian geometry*, Atihay, Hitchin, Singer. Les objets d'étude de ces articles sont les connections de Yang–Mills auto-duales, aussi appelées *instantons*, sur les variétés Riemanniennes en dimension quatre. Ces pseudo-particules ont été introduites en 1975 par Belavin, Polyakov, Schwarz et Tyupkin comme interpolant deux vides de la théorie de Yang–Mills euclidienne de groupe de jauge SU(2).

- 1977 – *Instantons and algebraic geometry*, Atiyah, Ward, où sont étudiées des connections entre les instantons et la géométrie algébrique.

- 1978 – *Construction of instantons*, Atiyah, Drinfeld, Hitchin, Manin, qui décrit l'espace des modules des instantons en termes de données purement linéaires.

- 1983 – *The Yang–Mills equations over Riemann surfaces*, Atiyah, Bott. Cet article est la pierre angulaire de l'utilisation des théories de jauge en géométrie des surfaces de Riemann.

- 1983 – *An application of gauge theory to four-dimensional topology*, Donaldson, dans lequel l'auteur développe la théorie des instantons afin d'étudier la topologie des variétés lisses de dimension quatre. Les invariants de Donaldson permettent notamment de prouver l'existence de variétés topologiques sans structure lisse, et celle de plusieurs structures lisses non-équivalentes sur $\mathbb{R}^4$.

- 1985 – *Anti-self-dual Yang–Mills connections over complex algebraic surfaces and stable vector bundles*, Donaldson et *On the existence of Hermitian Yang–Mills connections in stable vector bundles*, Uhlenbeck, Yau, qui prouvent la correspondance entre les connections de Yang–Mills anti-auto-duales et les fibrés vectoriels holomorphes stables conjecturée par Kobayashi et Hitchin. Cette correspondance prolonge celle de Narashiman et Seshadri (1965) entre les fibrés vectoriels holomorphes stables sur une surface de Riemann Σ et les représentations projectives unitaires du groupe fondamental de Σ.

- 1987 – *The self-duality equations on a Riemann surface*, Hitchin, où sont introduits les *fibrés de Higgs* dont les espaces de modules ont des propriétés très intéressantes, et qui fournissent des généralisations des espaces de Teichmüller. Ces *composantes de Hitchin* sont discutées en Section 4.3.1.

- 1988 – *Flat G-bundles with canonical metric*, Corlette, où est fait le lien entre les solutions des équations de Hitchin et les représentations du groupe fondamental de la surface considérée.

- 1989 – *Quantum Field Theory and the Jones polynomial*, Witten, dans lequel l'auteur développe une approche du polynôme de Jones, un invariant de nœuds, en termes de théories de Chern–Simons qui sont des théories de jauge topologiques en dimension trois.

- 1991 - 1992 – *Nonabelian Hodge theory* et *Higgs bundles and local systems*, Simpson, qui prouve que les fibrés des Higgs polystables correspondent aux solutions des équations de Hitchin. La correspondance de Hodge non-abélienne généralise la décomposition de Hodge pour des coefficients dans le groupe $\mathrm{GL}_n(\mathbb{C})$ au lieu de $\mathbb{C}$.

Supersymétrie

Une *supersymétrie* en théorie quantique des champs est une symétrie de l'espace-temps sous laquelle les champs bosoniques (qui ont nécessairement un spin entier d'après le théorème de spin statistique) sont transformés en champs fermioniques (qui ont nécessairement un spin strictement demi-entier). Mathématiquement, cela revient à définir les particules élémentaires non comme des représentations de l'algèbre de Poincaré de symétries infinitésimales relativistes, mais comme des représentations d'une extension $\mathbb{Z}/2\mathbb{Z}$-graduée de l'algèbre de Poincaré. L'idée de supersymétrie sous la forme qu'on lui attribue aujourd'hui date de 1971.

Certaines théories quantiques des champs supersymétriques peuvent être définies de manière perturbative par une action, comme le modèle de Wess–Zumino (1974), qui est l'un des modèles supersymétriques les plus simples :

$$S = \int \mathrm{d}^4x \left[\partial_\mu \phi^\dagger \partial^\mu \phi - i\chi\sigma^\mu \partial_\mu \overline{\chi} - \left| \frac{\partial W}{\partial \phi} \right|^2 - \frac{1}{2}\frac{\partial^2 W}{\partial \phi^2}\chi\chi - \frac{1}{2}\frac{\partial^2 W^\dagger}{\partial \phi^{\dagger 2}}\overline{\chi\chi} \right] ,$$

où ϕ est un champ scalaire complexe, χ^α (respectivement $\overline{\chi}_{\dot{\alpha}}$) est un fermion de Weyl chiral gauche (respectivement droit) et W est le polynôme

$$W(\phi) = \frac{1}{2}m\phi^2 + \frac{1}{3}\lambda\phi^3 ,$$

avec m et λ des paramètres interprétés respectivement comme la masse du champ ϕ et le couplage de Yukawa entre ϕ et les fermions χ et $\overline{\chi}$.

La supersymétrie de cette théorie est la transformation δ_ϵ définie de la manière suivante, où ϵ est un spineur de Weyl infinitésimal :

$$\begin{aligned} \delta_\epsilon \phi &= \sqrt{2}\epsilon\chi, \\ \delta_\epsilon \chi &= \sqrt{2}i\sigma^\mu \overline{\epsilon} \partial_\mu \phi - \sqrt{2}\epsilon \frac{\partial W^\dagger}{\partial \phi^\dagger} , \end{aligned}$$

avec σ^μ, $\mu = 0, \ldots, 3$ les matrices de Pauli. On peut vérifier que les équations du mouvement sont invariantes sous l'action de cette supersymétrie.

Lorsque la supersymétrie est une symétrie globale d'une théorie quantique des champs (comme c'est le cas dans la théorie de Wess–Zumino) on dit que cette dernière est *globalement supersymétrique*. Dès l'article fondateur de Wess et Zumino, les théories quantiques des champs globalement supersymétriques apparaissent comme remarquables car leur renormalisation est sous bien meilleur contrôle qu'en général. Par conséquent, ces théories supersymétriques sont des bonnes candidates pour des extensions du Modèle Standard pouvant prétendre à la résolution du point 4 ci-dessus.

De plus, les théorèmes de Coleman–Mandula (1967) et Haag–Lopuszanski–Sohnius (1975) fournissent un argument conceptuel solide pour l'étude des théories supersymétriques : en un sens, ce sont les théories quantiques des champs les plus symétriques que l'on peut considérer si l'on accepte certains principes physiques fondamentaux s'appliquant aux diffusions de particules élémentaires. Ces théories sont souvent plus compliquées que des théories génériques car la supersymétrie impose des contraintes sur les particules impliquées, et dans un même temps plus facilement exploitables de part leur supersymétrie. En particulier, il est possible de comprendre certains aspects de la dynamique non-perturbative des théories quantiques des champs supersymétriques, et ainsi d'étudier des phénomènes physiques très intéressants comme le confinement, l'existence de monopoles magnétiques, les instantons et diverses dualités, pour ne citer qu'eux.

Phénoménologie de la supersymétrie. Une première extension supersymétrique du Modèle Standard est proposée par Fayet en 1977, et complétée notamment par Dimopoulos et Georgi en 1981, conduisant à *l'extension supersymétrique minimale du Modèle Standard* (MSSM). Cette théorie améliore la renormalisation du champ de Higgs (point 4 ci-dessus) en *doublant le nombre de particules* du Modèle Standard. En effet, la définition même de la supersymétrie implique un appariement des degrés de liberté bosoniques et fermioniques de la théorie quantique des champs en considération. Les bosons et fermions appariés doivent avoir les mêmes charges locales et globales, ce qui est impossible sans ajout de nouvelles particules, appelées *superpartenaires* de celles du Modèle Standard.

Si la supersymétrie n'est pas brisée spontanément, les fermions et bosons dans une même paire doivent également avoir la même masse, ce qui est contraire aux observations expérimentales. Une brisure spontanée est donc nécessaire, et implique une séparation entre la masse des bosons et des fermions appariés. Le superpartenaire le plus léger est une particule stable selon ce modèle, et constitue un candidat intéressant pour la particule constitutive de la matière noire (point 2 ci-dessus).

Enfin, la théorie MSSM modifie la renormalisation des constantes de couplage des interactions faible, électromagnétique et forte, de sorte qu'elles coïncident à une énergie de quelques $\Lambda_{\mathrm{GUT}} \sim 10^{15}$ GeV (appelée échelle d'énergie de *grande unification*). Cela laisse espérer une unification de ces trois forces fondamentales au delà de cette énergie.

L'implémentation pratique de la brisure spontanée de la supersymétrie dans le modèle MSSM est étonnamment compliquée : soit on accepte une dose importante d'arbitraire, ce qui diminue le pouvoir prédictif de la théorie et son attrait conceptuel, soit on est forcé de considérer une théorie auxiliaire appelée *secteur caché*, qui décrit des interactions et des particules inobservables si ce n'est par les effets ténus et indirects qu'elles ont sur les particules du MSSM.

Sous l'hypothèse d'une résolution complète du problème de hiérarchie 4, une version exploitable du modèle MSSM appelée pMSSM (qui dépend de 19 paramètres au lieu d'une centaine) prédit l'existence de particules supersymétriques qui devraient vraisemblablement déjà avoir été observées au LHC. Ainsi :

- soit pour une raison ou une autre, ces particules existent mais n'ont toujours pas été découvertes,
- soit ces particules n'existent pas, le modèle pMSSM est correct mais on doit accepter un certain réglage fin et une résolution seulement partielle du problème de hiérarchie,
- soit ces particules n'existent pas et les modèles pMSSM ou MSSM ne fournissent pas une description valable de la nature.

Dans tous les cas (même le dernier) ce n'est que le modèle particulier qui est remis en cause, et non pas l'idée que la supersymétrie doive être invoquée pour décrire la nature. Par exemple, un autre modèle dénommé *supersymétrie fractionnée* a été proposé en 2004 par Arkani–Hamed et Dimopoulos ; il ne résout le problème de hiérarchie que partiellement, mais continue à fournir des candidats constitutifs de la matière noire, et une unification des interactions microscopiques à des énergies supérieures à Λ_{GUT}.

Dimension de l'espace-temps et supersymétrie. Un aspect plus théorique très intéressant de l'étude des théories quantiques des champs supersymétriques est le fait que l'implémentation de la supersymétrie est fortement dépendante de la dimension de l'espace-temps. Même si jusqu'à présent nous avons considéré des théories quantiques des champs dans l'espace de Minkowski $\mathbb{R}^{1,3}$, puisque l'espace-temps dans lequel nous vivons est manifestement de dimension quatre macroscopiquement, il est tout à fait possible d'étudier des théories quantiques des champs dans des espaces-temps de dimension n, par exemple les espaces de Minkowski $\mathbb{R}^{1,d-1}$ où $d \in \mathbb{Z}_{>0}$.

Alors que les champs bosoniques ne dépendent que peu de d, les champs fermioniques, eux, en dépendent fortement. C'est une conséquence de la théorie des représentations des algèbres de Lie orthogonales $\mathfrak{so}(d, \mathbb{C})$, qui est la complexification de l'algèbre de Lorentz en dimension d. On distingue plusieurs types de fermions irréductibles : de *Weyl* W, de *Majorana* M, de *Majorana–Weyl* MW, *symplectiques* S ou *sympletiques–Weyl* SW. L'existence de ces fermions en fonction de la dimension d est donnée dans la table suivante :

Dimension d	Fermion	Nombre de composantes
2	MW	1
3	M	2
4	M, W	4
5	S	8
6	SW	8
7	S	16
8	M, W	16
9	M	16
10	MW	16
11	M	32

La suite des natures de fermions irréductibles en fonction de la dimension est 8-périodique ; c'est une manifestation de la *périodicité de Bott* (1957-1959).

Les générateurs des supersymétries échangeant des bosons et des fermions, ils ne préservent pas le spin. Plus précisément, l'action d'un générateur de supersymétrie sur un champ fait varier le spin de 1/2. Plus il y a de supersymétries, et plus la théorie doit contenir des champs fondamentaux de spins différents. Dans les théories quantiques des champs telles que celles présentées plus haut, on se restreint en général à des champs de spin 0, 1/2 ou 1. Cela limite le nombre $\mathcal{N}$ de supersymétries possibles en chaque dimension : il faut que le nombre total des composantes des supersymétries (appelées *supercharges*) soit inférieur à 16. Chaque supersymétrie contribue d'un nombre de supercharges indiqué dans la colonne de droite du tableau, si bien que cette contrainte sur le spin donne par exemple $\mathcal{N} \leq 4$ pour $d = 4$, $\mathcal{N} \leq 2$ pour $d = 6$ et $\mathcal{N} \leq 1$ pour $d = 10$.

Plus la valeur de $\mathcal{N}$ est grande, plus la théorie est rigide et analysable de façon exacte, et moins elle est phénoménologique. Par exemple, en dimension $d = 4$, les théories de Yang–Mills avec $\mathcal{N} = 4$ supersymétries sont entièrement déterminées par leur groupe de jauge et sont *superconformes*, ce qui implique en particulier qu'aucune quantité n'est renormalisée : la dynamique de la théorie est indépendante de l'échelle d'énergie. Ces théories de Yang–Mills exhibent une *dualité électrique-magnétique* étudiée par Montonen et Olive en 1977, qui est une version rigoureuse de la symétrie manifeste entre les champs électriques et magnétiques des équations de Maxwell dans le vide, et a des liens avec la *dualité de Langlands*. Avec $\mathcal{N} = 2$ supersymétries, on peut calculer exactement la dynamique à basse énergie des théories de Yang–Mills asymptotiquement libres ou superconformes à l'aide de la *théorie de Seiberg–Witten*, qui est déjà extrêmement riche. Cette dernière a notamment permis la découverte de théories exotiques, dites *d'Argyres–Douglas*, où l'on observe par exemple une coexistence d'excitations électriques et magnétiques légères. Les théories avec $\mathcal{N} = 1$ supersymétrie sont probablement encore plus intéressantes, mais moins tractables mathématiquement ; en particulier, il est souvent impossible de calculer exactement la dynamique à basse énergie lorsqu'elle est à couplage fort. Néanmoins, la supersymétrie présente a permis la découverte et l'étude de phénomènes profonds comme la *dualité de Seiberg* (1994), qui implémente dans un cadre physique des *transformations amassées tropicales*.

En dimension quatre, les théories avec $\mathcal{N} = 3$ supersymétries définies perturbativement ont toujours $\mathcal{N} = 4$ supersymétries, pour une raison analogue au théorème de Frobenius (1877) qui caractérise les algèbres associatives à division de dimension finie sur $\mathbb{R}$ comme étant $\mathbb{R}$, $\mathbb{C}$ et $\mathbb{H}$. Cependant, l'existence de théories non-définies perturbativement en dimension 4 a été montrée par Garcia-Etxebarria et Regalado en 2015 grâce à une construction de théorie des cordes, les *S-folds*.

Le simple fait qu'il soit possible d'étudier une telle variété de phénomènes physiques grâce à la supersymétrie illustre parfaitement la puissance de cette hypothèse mathématique, ainsi que le changement de paradigme évoqué à la fin de la présentation du Modèle Standard. Il est probable que les théories supersymétriques étudiées n'aient, en grande majorité, aucun rôle direct à jouer dans la description de notre univers ; cependant la profondeur d'étude rendue possible par l'hypothèse de la supersymétrie permet d'étudier le *langage* des théories quantiques des champs en tant que tel et d'accéder à une diversité auparavant inaccessible. L'espoir absolu d'un point de vue phénoménologique est de pouvoir ensuite transposer les connaissances ainsi acquises à la description de la nature ; d'un point de vue mathématique, l'accès à une telle richesse théorique est un trésor bien suffisant.

Si on s'autorise des spins 3/2, la cohérence de la théorie implique également l'existence de particules de spin 2. La contrainte de ne pas avoir de spin plus grand que 5/2 donne alors $\mathcal{N} \leq 8$ pour $d = 4$, $\mathcal{N} \leq 2$ pour $d = 10$ et $\mathcal{N} \leq 1$ pour $d = 11$. Aucune théorie cohérente de particules de spin plus grand que 5/2 n'est connue, ce qui fait donc de 11 la dimension maximale possible pour l'implémentation de la supersymétrie. Une particule non massive de spin 2 a toutes les propriétés d'un *graviton*, le quanta hypothétique des ondes gravitationnelles d'Einstein. On peut montrer que dans les théories supersymétriques qui contiennent des particules de spin 2, la supersymétrie est *locale* et non globale, ce qui implique l'équation d'Einstein. Ces théories où la supersymétrie est une symétrie de jauge sont nommées *théories de supergravité*. Elles fournissent une description semi-classique non-renormalisable de la relativité générale avec d'autres champs de spin $s \leq 3/2$.

Supersymétrie et géométrie. La supersymétrie tisse des liens forts entre les théories quantiques des champs et certaines structures géométriques. Ces liens s'observent naturellement dans les *espaces des*

modules des théories supersymétriques.

Dans une théorie quantique des champs, on définit un *état de vide* comme une configuration de champs invariante de Lorentz qui minimise une fonction appelée *potentiel scalaire*. En général, les effets quantiques impliquent que les vides d'une théorie sont tous isolés dans l'espace des configurations de champs. Cependant, lorsque la théorie est supersymétrique, la situation est radicalement différente. D'une part, les configurations de champs ont toutes une énergie positive. D'autre part, les vides se rangent en deux catégories : ceux qui ont une énergie strictement positive, dans lesquels la supersymétrie est brisée spontanément, et ceux qui ont une énergie nulle, qui préservent la supersymétrie. Cette dernière dompte les corrections quantiques et permet l'existence de *variétés* continues de vides supersymétriques pour la théorie. L'union de ces variétés forme *l'espace des modules* des vides supersymétriques de la théorie.

En fonction de la dimension d'espace-temps d et de la quantité $\mathcal{N}$ de supersymétries de la théorie, des sous-variétés naturelles de l'espace des modules sont dotées de structures mathématiques riches. On trouve par exemple des variétés Kähleriennes, des variétés Kähleriennes spéciales, ou encore des variétés hyperkähleriennes.

Dans le cas spécial où $d = 1$, un article célèbre de Witten de 1982 fait le lien entre la supersymétrie et la théorie de Morse, qui permet d'analyser la topologie d'une variété différentielle en étudiant les fonctions différentiables qui y sont définies, initiée par le *lemme de Morse* (1931) et développée durant la deuxième moitié du 20ème siècle. En dimension $d = 4$, l'étude des théories de Yang–Mills avec $\mathcal{N} = 2$ supersymétries a conduit à la théorie de Seiberg–Witten qui permet notamment de calculer des nombres d'instantons. En 1994, Witten introduit les invariants de Seiberg–Witten issue de la théorie éponyme, qui sont très similaires aux invariants de Donaldson, mais plus faciles à construire.

L'étude des théories des champs supersymétriques en dimension 4 fait l'objet du Chapitre 5.

Théorie des cordes

Définition perturbative. La théorie des cordes a d'abord été développée en tant que modèle des interactions fortes sous le nom de *modèle dual de résonance*.

- 1968 – Veneziano propose son amplitude de diffusion qui s'exprime en termes de la fonction bêta d'Euler, et qui satisfait de nombreuses propriétés nécessaires à la description de mésons à couplage fort.
- 1969 – Les règles de Chan–Paton permettent d'inclure la symétrie de saveur (alors connue sous le nom d'isospin) dans l'amplitude de Veneziano.
- 1969 - 1970 – Nambu, Nielsen et Susskind interprètent l'amplitude de Veneziano en termes de cordes.
- 1971 – Ramond, Neveu et Schwarz proposent une méthode pour inclure une description des fermions dans le modèle dual de résonnance. Ce sont ces recherches qui sont à l'origine de l'étude de la supersymétrie, discutée plus haut.

Les prédictions de la théorie des cordes ainsi construite sont en contradiction avec les observations expérimentales, et le modèle dual de résonance laisse la place à la chromodynamique quantique en 1973, beaucoup plus fructueuse pour décrire les interactions fortes. Cependant, en 1974, Schwarz, Scherk et Yoneya montrent que les théories des cordes décrivent une particule dont les propriétés sont celles d'un graviton. Ainsi renaît l'intérêt pour la théorie des cordes, en tant que candidate d'une description de la gravité quantique. Le postulat fondamental de la théorie des cordes est que les particules élémentaires ne sont pas ponctuelles comme cela est supposé en théorie classique des champs, mais plutôt des *cordes* qui se déplacent dans l'espace en vibrant. L'analogue des lignes d'univers associées aux trajectoires des particules ponctuelles évoquées en dessous de la Figure 3 sont, pour les cordes, des *surfaces d'univers*. Les graphes de Feynman sont adaptés de la même manière.

L'hypothèse que les particules élémentaires sont des cordes et non des points n'est pas en contradiction avec les données expérimentales, à condition que la taille de ces cordes l_s soit bien plus faible que les distances les plus petites sondées à ce jour, c'est-à-dire que $l_s \ll 10^{-21}$ cm ; si c'est le cas, ces cordes paraissent être ponctuelles. En plus de se déplacer dans l'espace, les cordes (contrairement aux points) peuvent vibrer, ces vibrations se décomposant en modes de Fourier. Chaque mode de vibration a une

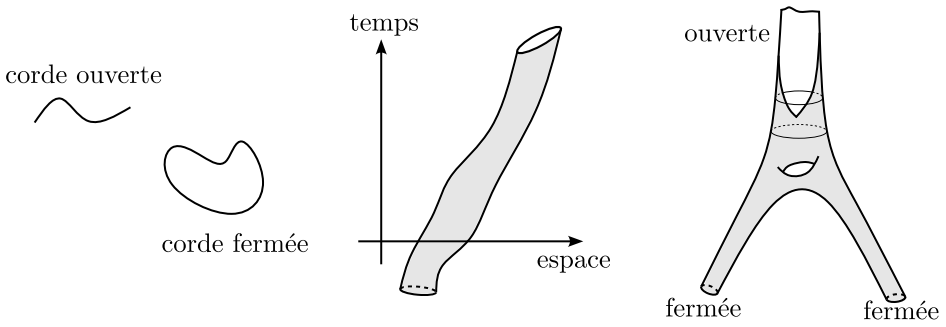


Figure 15: Cordes ouvertes, fermées, surfaces d'univers et "graphes" de Feynman.

énergie associée, qui est d'autant plus grande que la longueur d'onde est petite. S'il est impossible de distinguer l'extension spatiale des cordes alors cette énergie apparaît seulement comme une contribution à la masse de ce qui apparaît comme étant effectivement ponctuel. Ainsi, la théorie des cordes prévoit, de manière *effective*, l'existence d'une "tour" de particules de masses de plus en plus grandes. L'un des espoirs (phénoménologique) de la théorie des cordes est de pouvoir expliquer toutes les particules connues de cette manière. Ainsi on peut espérer donner une explication naturelle à la question de pourquoi les particules élémentaires du Modèle Standard sont ce qu'elles sont, et ont les propriétés qui les caractérisent.

La construction perturbative d'une théorie des cordes dans un espace-temps plat de dimension d (c'est-à-dire l'espace-temps de Minkowski $\mathbb{M}^{1,d-1}$) commence par la construction d'une théorie classique relativiste. Cette dernière est décrite par l'*action de Nambu–Goto*, qui est l'intégrale de l'aire relativiste de la surface d'univers de la corde. Une reformulation intéressante est l'*action de Polyakov*, à laquelle on ajoute des champs fermioniques qui servent (à terme) à décrire des fermions dans l'espace-temps, selon les préceptes de Ramond, Neveu et Schwarz (RNS). Cette action dépend du choix d'une métrique auxiliaire sur la surface d'univers de la corde, mais est crucialement *invariante par transformation conforme* de cette métrique.

La quantification de cette théorie des cordes relativiste fixe la dimension d de l'espace-temps afin que la symétrie conforme de la théorie ne soit pas anomale : $d = 10$. La projection GSO, introduite par Gliozzi, Scherk et Olive en 1979, permet alors d'obtenir une théorie *sans tachyon*, relativiste et quantique, dans $\mathbb{R}^{1,9}$. En 1984, Green et Schwarz proposent un mécanisme d'annulation des anomalies de jauge et gravitationnelles, prouvant ainsi l'existence d'une théorie des cordes quantiques et relativiste cohérente.

En réalité, il y a des choix à faire quant aux types de cordes (fermées, ou ouvertes et fermées) et de surfaces d'univers (orientables, ou orientables et non-orientables) que l'on considère, et quant à la manière selon laquelle les fermions RNS sont introduits. Il n'y a donc pas une, mais plusieurs théories des cordes cohérentes dans $\mathbb{M}^{1,9}$. Cinq d'entre elles sont supersymétriques : les théories de type I (1982), type IIA (1982), type IIB (1982), hétérotique SO(32) (1985) et hétérotique $E_8 \times E_8$ (1985). À basse énergie, les théories des cordes sont décrites de manière effective par des théories de supergravité, et contiennent des champs tensoriels, ainsi que différents champs de jauge généralisés qui sont des formes différentielles de différents degrés. Il existe également une théorie des cordes hétérotique $\mathrm{SO}(16) \times \mathrm{SO}(16)$ non-supersymétrique (1986), dans laquelle l'absence de supersymétrie implique un moins bon contrôle des corrections quantiques ; elle est donc plus difficile à étudier que ses analogues supersymétriques.

Les théories des cordes sont par construction bien définies à toute énergie : les graphes de Feynman sont en effet des surfaces, sans sommets d'interaction. Cette propriété empêche l'existence des *divergences ultraviolettes* qui gâtent certaines théories quantiques des champs. Ainsi, les théories des cordes sont considérées comme des *complétions ultraviolettes* des théories de supergravité correspondantes : ces dernières sont satisfaisantes en dessous d'une certaine énergie Λ_{string}, mais doivent être complétées en des théories des cordes pour avoir un sens à des énergies supérieures à Λ_{string}.

Compactifications. Au lieu de considérer les théories des cordes dans $\mathbb{R}^{1,9}$, on peut les étudier dans des espaces-temps de dimension 10 et de topologies plus compliquées. Une famille de cas parcimonieux est constituée des *compactifications toroïdales* $\mathbb{R}^{1,d-1} \times (S^1)^{10-d}$. Ces espaces-temps sont particulièrement intéressants d'un point de vue phénoménologique : de la même manière qu'on ne peut "résoudre" l'extension spatiale des cordes si elles sont suffisamment petites, si les dimensions circulaires de ces

espaces-temps ont une taille R, elles sont invisibles à des énergies inférieures à R^{-1} (en unités naturelles). On obtient donc des théories effectives en dimension d. Les théories résultantes ont le même nombre de composantes de supersymétries. Par exemple, la théorie de type IIB dans $\mathbb{R}^{1,3} \times (S^1)^6$ est décrite, à basse énergie, par la théorie de supergravité avec $\mathcal{N} = 8$ supersymétries en dimension quatre.

Ce mécanisme de *compactification* permet d'expliquer comment une théorie des supercordes, qui pour être définie perturbativement doit l'être dans un espace-temps de dimension 10, peut décrire une physique macroscopiquement de dimension inférieure ($d = 4$ si l'on s'intéresse à des questions phénoménologiques). En 1985, Candelas, Horowitz, Strominger et Witten montrent que la condition pour obtenir des théories effectives en dimension $d \leq 10$ est de considérer des compactifications de la forme $\mathbb{R}^{1,d-1} \times X_{10-d}$, où X_{10-d} est une variété de dimension $10 - d$ d'*holonomie spéciale*. Dans le cas $d = 4$, on obtient par exemple diverses théories de supergravité avec $\mathcal{N} = 1$ en faisant l'hypothèse que X_6 est une *variété de Calabi–Yau*, c'est-à-dire une variété complexe Kählerienne compacte de dimension 3 dont l'holonomie est contenue dans SU(3). Comme souligné dans la section précédente, plus le nombre de supersymétries est petit et plus la théorie en question est réaliste. Ainsi, il est intéressant de considérer des variétés de Calabi–Yau dont l'holonomie est contenue dans SU(3) mais pas dans SU(2), afin que la théorie résultante en dimension 4 ait exactement $\mathcal{N} = 1$. En 1985, peu de telles variétés algébriques sont connues, mais l'intérêt que leur portent les théoriciens des cordes crée un engouement pour ces objets, et rapidement de nouvelles constructions apparaissent, notamment celle de Batyrev (1994), qui prouve l'existence d'environ un demi-milliard de Calabi–Yau compacts différents de dimension (réelle) $d = 6$.

Afin d'obtenir des vides intéressants en théorie des cordes, il est nécessaire des considérer des *compactifications avec flux*, dans lesquelles en plus de choisir la géométrie de l'espace de compactification X il faut fixer les flux (quantifiés) de certaines formes différentielles de jauge le long des cycles homologiques de X. Cela mène à une quantité considérable de choix de vides, l'estimation la plus basse étant de quelques 10^{500} vides "raisonnables". L'ensemble de ceux-ci est appelé le *paysage de la théorie des cordes*, par analogie avec la notion de "paysage adaptatif" en génétique.

La théorie des cordes (ou les théories des cordes) est un langage, au même titre que la théorie quantique des champs : il s'agit d'un ensemble d'axiomes et de règles qui définissent un cadre mathématique d'étude. Alors qu'en théorie quantique des champs on définit une théorie particulière en spécifiant son contenu en champs et ses couplages à une certaine énergie, ces choix sont essentiellement *géométriques* en théorie des cordes. De plus, le fait qu'il n'existe que quelques théories fondamentales dans $\mathbb{R}^{1,9}$ est particulièrement attrayant conceptuellement.

Dualités. Une dualité est une symétrie entre théories, c'est-à-dire une correspondance entre deux descriptions différentes de la même physique. Une première telle *dualité* entre théories des cordes est la *T-dualité*, découverte par Sathiapalan en 1987. Elle relie notamment la théorie IIA dans $\mathcal{M}^{1,8} \times S^1_R$ de constante de couplage g_s et la théorie IIB dans $\mathcal{M}^{1,8} \times S^1_{\alpha'/R}$ de constante de couplage $\alpha' g_s / R$, où S^1_R signifie que la taille du cercle S^1 est R, et où α' est la *pente de Regge*, qui vaut $\alpha' = l_s^2$ avec l_s la longueur des cordes. De même, les deux théories hétérotiques supersymétriques sont T-duales.

La *S-dualité* est introduite par Sen en 1994. C'est une dualité qui lie une description à couplage fort et une description à couplage faible, ce qui la rend très intéressante. En effet, on peut grâce à elle étudier le régime de couplage fort d'une théorie des cordes à l'aide de méthodes perturbatives dans la description duale. Les théories de type I et hétérotique SO(32) sont par exemple S-duales, et la théorie IIB est S-autoduale.

L'étude des dualités S et T entre théorie des supercordes montre que ces cinq théories sont cinq descriptions différentes d'une même physique. En 1995, Witten argumente en faveur de l'existence d'une théorie appelée *théorie M*, définie en dimension 11 et à partir de laquelle on peut obtenir les cinq supercordes dans certaines limites. La théorie M ne décrit pas de cordes mais des membranes de dimensions 3 et 6. A basse énergie, elle est décrite de manière effective par l'unique théorie de supergravité en dimension maximale 11. Les dualités entre ces théories sont souvent représentées comme en Figure 16.

Un produit dérivé de l'étude de ces dualités a eu un impact considérable en mathématiques : il s'agit de la *symétrie miroir*. L'homologie des Calabi–Yau compacts de dimension 3 est entièrement décrite par la donnée de deux nombres de Hodge : $h^{1,1}$ (qui est la dimension de l'espace des modules de Kähler) et $h^{2,1}$ (qui est la dimension de l'espace des structures complexes). Dans sa formulation la plus faible, la symétrie miroir postule que pour chaque Calabi–Yau compact X de dimension 3, il en existe un autre Y tel que $h^{1,1}(X) = h^{2,1}(Y)$ et $h^{1,1}(Y) = h^{2,1}(X)$. Physiquement, la théorie IIA compactifiée sur X est duale à la théorie IIB compactifée sur Y, et réciproquement.

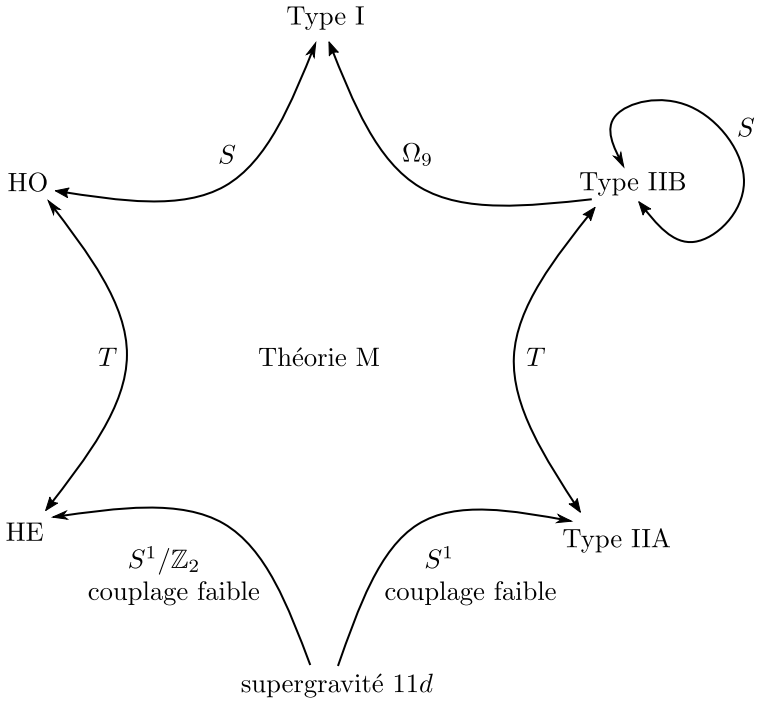


Figure 16: Dualités des théories des cordes.

Branes. Introduites par d'une part par Dai, Leigh et Polchinski et d'autre part par Horava en 1989, les D-branes (où 'brane' est une analogie au terme 'membrane') sont des objets étendus non-perturbatifs décrits par les théories des cordes. Mathématiquement, ce sont des sous-variétés de l'espace-temps munies de fibrés vectoriels. Lorsque le volume d'une D-brane est décrit par p directions d'espace et une direction de temps, on parle de Dp-brane. Ainsi, une Dp-brane est une sous-variété Lorentzienne de dimension $p+1$ de l'espace-temps. Il existe deux points de vue duaux sur les D-branes, ce qui en fait des objets particulièrement riches.

D'une part, les D-branes supportent les extrémités de cordes ouvertes. Lors de la quantification des cordes perturbatives en présence d'une D-brane, les modes de vibration des cordes ouvertes donnent lieu à une théorie de Yang–Mills supersymétrique dans le volume de la D-brane, de manière effective, à basse énergie. Par exemple, le volume d'une unique D3-brane dans $\mathbb{R}^{1,9}$ étendue le long des directions $t = x^0, x^1, x^2, x^3$ et localisée en $x^4 = \cdots = x^9 = 0$ est doté de la théorie de super Yang–Mills en dimension 4 avec $\mathcal{N} = 4$ supersymétries et U(1) comme groupe de jauge. On peut empiler des D-branes les unes sur les autres, ce qui augmente la symétrie de jauge : la superposition de N D3-branes disposées comme avant donne lieu à la théorie de super Yang–Mills en dimension 4 avec $\mathcal{N} = 4$ supersymétries et de groupe de jauge U(N).

D'autre part, à basse énergie, les branes sont décrites par des solutions des équations de supergravité, et correspondent à des trous noirs étendus. Ces solutions sont appelées *branes noires*. En théorie des cordes, cette description des D-branes les définit comme des *états non-perturbatifs* de cordes fermées.

Les D-branes que l'on peut considérer dépendent de la théorie des cordes dans laquelle on travaille. Par exemple, la théorie IIA contient des D0s, D2s, D4s, D6s et D8s tandis que la théorie IIB contient des D(-1)s, D1s, D3s, D5s, D7s et D9s. De plus, les théories des cordes décrivent d'autres types de branes ou d'objets étendus, comme les cordes (parfois appelées NS1-branes), les NS5-branes et les orientifolds. Sous les dualités S et T, les branes sont transformées en d'autres branes, ou éventuellement en géométrie.

En considérant des *configurations de branes*, on peut construire des théories quantiques des champs (effectives, à basse énergie) dont la structure est encodée dans la géométrie de la configuration. L'étude de la dynamique des branes en théorie des cordes fournit alors des informations précieuses sur la dynamique perturbative et non-perturbative de ces théories des champs. Deux exemples de telles configurations de branes sont représentées en Figure 17. Il s'agit de (p, q)-toiles, introduites en 1997 par Aharony, Hanany et Kol, et qui encodent des théories en dimension $d = 5$ avec $\mathcal{N} = 1$ supersymétrie. Celle de gauche

décrit une théorie de super Yang–Mills de groupe de jauge SU(2) avec 3 saveurs, tandis que celle de droite décrit une théorie de super Yang–Mills pure (i.e. sans saveur), de groupe de jauge SU(3). Ces (p, q)-toiles sont la tropicalisation de courbes algébriques tropicales qui apparaissent lorsque ces configurations sont relevées en théorie M.

De manière similaire, on peut considérer des branes qui ne sont pas placées dans l'espace plat $\mathbb{R}^{1,9}$ mais dans un espace de géométrie plus compliquée, par exemple au sommet d'une *singularité Calabi–Yau affine* X. Dans cette thèse, le cas suivant est considéré à de nombreuses reprises : des D3-branes sont placées dans un espace-temps de la forme $\mathbb{R}^{1,3} \times X_6$, où X_6 est une singularité Calabi–Yau affine isolée de dimension complexe 3, et où les branes s'étendent le long des directions de $\mathbb{R}^{1,3}$ et sont localisées au point singulier de X_6. La théorie de jauge dans le volume des D3-branes est alors déterminée par la géométrie de X_6. Ces deux approches sont essentiellement duales par des combinaisons de dualités S et T ; les branes au point singulier de variétés Calabi–Yau peuvent notamment être représentées par des configurations particulières nommées *pavage par branes* ("brane tiling" en anglais) qui sont obtenues par deux T-dualités successives sur la configuration de départ. Cette manière de construire des théories de jauge et d'étudier leur dynamique s'appelle *l'ingénierie de jauge*.

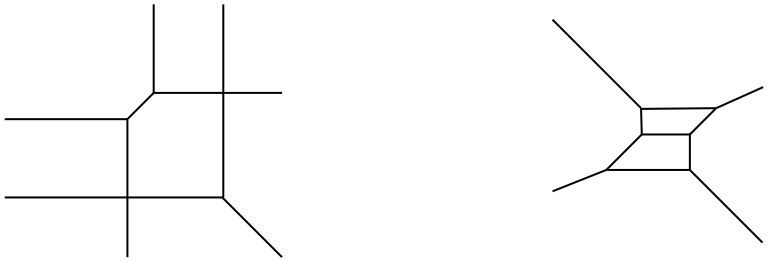


Figure 17: Deux (p, q)-toiles.

L'ingénierie de jauge permet de construire des familles infinies de théories quantiques des champs supersymétriques qui, par construction, admettent une complétion ultraviolette en une théorie de gravité quantique (la théorie des cordes). C'est une donnée intéressante quant à ces théories car si l'on considère une théorie quantique des champs quelconque, supersymétrique ou non, le fait qu'elle puisse être ainsi complétée dans l'UV n'est jamais une évidence. Le *programme du marécage*, initié par Vafa en 2005, tente de dégager les propriétés que les théories quantiques des champs doivent satisfaire pour être compatibles avec la gravité quantique. La théorie des cordes fournit un cadre d'étude idéal pour cela. Un exemple de *conjecture marécageuse*, parmi les plus consensuelles, est la *conjecture de gravité faible* ("weak gravity conjecture" en anglais) introduite par Arkani-Hamed, Motl, Nicolis et Vafa en 2006. Dans l'une de ses versions, elle statue qu'une théorie quantique de jauge en dimension quatre de groupe de jauge U(1) et de constante de couplage g ne peut être compatible avec la gravité quantique que si elle contient une particule chargée suffisamment légère, dont la masse satisfait $m \leq g\Lambda_P$. Un exemple plus récent est la *conjecture des cobordismes* énoncée par McNamara et Vafa en 2019, qui stipule que les classes de cobordismes dans l'espace des configurations d'une théorie quantique des champs doivent toutes être triviales afin que cette théorie puisse être compatible avec la gravité quantique.

Le formalisme de la théorie des cordes est introduit et étudié dans les Chapitres 6 et 7.

Information quantique, holographie, et correspondance AdS–CFT

L'entropie thermodynamique d'un système mesure la mésinformation que l'on a sur l'état microscopique de ce dernier si on ne connaît que son état macroscopique. Par exemple, la donnée d'un gaz parfait constitué de n moles de particules dans un volume V et à une température $T > 0$ ne détermine pas une unique configuration microscopique. Si Ω est le nombre de configurations microscopiques compatibles avec l'état macroscopique du système, l'entropie est $S = k_B \ln(\Omega)$, où k_B est la constante de Boltzmann. La seconde loi de la thermodynamique affirme que l'entropie d'un système fermé ne peut qu'augmenter.

Cependant, si l'on considère un trou noir, il semble que l'entropie de tout ce qui y tombe est irrémédiablement perdue, puisque le théorème de calvitie des trous noirs (qui est un théorème de relativité générale, c'est-à-dire un résultat non quantique) stipule justement que ces derniers ne gardent aucune trace de ce qui traverse l'horizon, si ce n'est dans l'évolution de leur masse, charge et moment angulaire.

Motivé par l'observation de Christodoulou et Hawking que l'aire de l'horizon d'un trou noir ne décroît jamais, Bekenstein propose en 1972 qu'on peut assigner une entropie à tout trou noir, proportionnelle à son aire $\mathcal{A}$ mesurée en unités de Planck : $S_{\mathrm{BH}} \propto k_B l_P^{-2} \mathcal{A}$. Un trou noir ayant une masse, il a a également une une énergie (relativiste), et s'il a une entropie alors il doit également avoir une température. Cette intuition est confirmée par Hawking en 1974, qui montre par un calcul de théorie quantique des champs au voisinage d'un trou noir de masse M que ce dernier rayonne de l'énergie selon la loi de Planck, à une température dite de *Hawking* :

$$T_{\mathrm{H}}(M) = \frac{\hbar c^3}{8\pi k_B \mathcal{G} M} \ .$$

Cette température varie inversement proportionnellement à la masse du trou noir ; pour un trou noir de masse solaire elle est très faible, de l'ordre de 10^{-7} K. L'expression de cette température permet de calculer l'*entropie de Bekenstein–Hawking* :

$$S_{\mathrm{BH}}(\mathcal{A}) = \frac{k_B \mathcal{A}}{l_P^2} \ ,$$

qui fixe l'entropie d'un trou noir à un quart de son aire, mesurée en unités de Planck (multipliée par k_B). Bekenstein argumente également que l'entropie maximale que peut contenir une région d'espace-temps bornée par une surface d'aire $\mathcal{A}$ est $S_{\mathrm{BH}}(\mathcal{A})$. C'est la *borne entropique de Bekenstein*, qui dépend de manière surprenante de l'*aire* bornant la région et non pas du volume de cette dernière, comme on pourrait s'y attendre en extrapolant les exemples habituels de systèmes thermodynamiques. Cette borne permet de définir des trous noirs d'une manière nouvelle : un trou noir d'horizon H est, parmi tous les corps bornés par H, celui qui a une entropie maximale.

Une conséquence importante du calcul d'Hawking est qu'un trou noir isolé dans le vide (quantique) perd de l'énergie en rayonnant : on dit que les *trous noirs s'évaporent*. En pratique, les trous noirs astrophysiques de masse comparable à celle du Soleil ou plus importante sont plus froids que le rayonnement de fond diffus cosmologique, dont la température est de $2,7$ K. Par conséquent en se thermalisant avec leur environnement ils ne s'évaporent pas, au contraire. La situation théorique des trous noirs isolés fournit cependant une expérience de pensée fort intéressante. Dans ce cas, un trou noir de masse solaire s'évapore en 10^{64} ans.

La radiation émise par le trou noir est intriquée avec l'intérieur de ce dernier. L'intrication quantique peut être mesurée par l'*entropie d'intrication* de l'extérieur du trou noir relativement à l'intérieur. Les lois de la physique quantique, et en particulier le fait que l'évolution temporelle d'un système quantique est unitaire, impliquent que l'entropie totale d'intrication est conservée. Cette dernière est nulle pour un état quantique pur (par distinction avec un état mixte), qui le reste sous une évolution unitaire. Considérons donc un trou noir formé par l'effondrement d'une sphère de photons dans un état quantique pur. Le calcul de Hawking suggère que l'entropie d'intrication du rayonnement émis (par rapport à l'intérieur du trou noir) augmente de façon monotone avec le temps. Ainsi, une fois que le trou noir s'est évaporé, l'ensemble de la radiation émise possède une grande entropie d'intrication. Alors :

1. Soit l'évaporation du trou noir laisse un reliquat de taille Planckienne qui possède lui aussi une grande entropie d'intrication afin que l'entropie d'intrication totale de l'univers soit nulle,
2. Soit il faut abandonner l'idée d'évolution unitaire en mécanique quantique couplée à la gravité,
3. Soit l'entropie d'intrication de la radiation émise n'augmente pas de façon monotone, en contradiction avec le raisonnement d'Hawking.

C'est le *paradoxe de l'information*. La première éventualité n'est pas satisfaisante d'un point de vue thermodynamique, ce qui laisse donc une alternative entre les deux autres points.

't Hooft (*Dimensional reduction in Quantum Gravity*, 1993) et Susskind (*The world as a hologram*, 1994) soulignent la conséquence suivante de la borne entropique de Bekenstein : pour toute surface S dans l'espace, il est possible de décrire fidèlement tout ce qu'il se passe dans la région bornée par S en termes de degrés de liberté sur la surface S elle-même. C'est le *principe holographique*, en analogie avec les plaques holographiques qui contiennent l'information de l'hologramme, une image tri-dimensionnelle, sur une plaque photographique bi-dimensionnelle. En particulier, la théorie gravitationnelle décrivant le voisinage d'un trou noir doit pouvoir être encodée de manière équivalente en une théorie non-gravitationnelle

sur une surface bornant ce voisinage (en une dimension de moins). Cela permet potentiellement de décrire l'évaporation d'un trou noir en termes d'une théorie non-gravitationnelle qui, si elle est quantique, a sûrement une évolution unitaire.

En 1998, Maldacena publie *The Large N Limit of Superconformal field theories and supergravity*, dans lequel il utilise les deux points de vue duaux sur les D-branes de Polchinski pour montrer notamment la dualité entre les deux théories suivantes.

- La théorie de supergravité qui décrit le voisinage de branes noires (dont la géométrie est $\mathrm{AdS}_5 \times S^5$), obtenues comme limite à basse énergie d'une superposition de N D3-branes en théorie des supercordes de type IIB, avec $gN \gg 1$ où g est la constante de couplage de la théorie IIB.Ici AdS_5 désigne l'espace-temps anti de Sitter en dimension 5, qui est la version Lorentzienne de l'espace hyperbolique de dimension 5, et S^5 est la 5-sphere.
- La théorie de jauge $\mathcal{N} = 4$ super Yang–Mills de groupe de jauge $\mathrm{SU}(N)$ et de constante de couplage g_{YM} tel que $g^2_{\mathrm{YM}}N \gg 1$, obtenue à basse énergie dans le volume des D3-branes.

C'est le premier exemple explicite de dualité holographique : la théorie de jauge, qui est une théorie quantique des champs en dimension 4 dont l'évolution est unitaire, encode fidèlement la théorie de supergravité IIB dans $\mathrm{AdS}_5 \times S^5$ qui est, de manière effective, une théorie en dimension 5. Il existe des solutions de trous noirs en supergravité IIB dans $\mathrm{AdS}_5 \times S^5$, dont l'évolution est décrite par une théorie unitaire, ce qui montre que le raisonnement d'Hawking n'est pas correct, au moins dans ce cas.

Lorsque $g^2_{\mathrm{YM}}N$ n'est pas grand devant 1, la théorie de jauge doit être duale non plus à la supergravité IIB mais à la théorie des supercordes de type IIB ; la correspondance de Maldacena fournit donc en principe une définition non-perturbative de la théorie IIB dans $\mathrm{AdS}_5 \times S^5$. Maldecena conjecture que cette correspondance doit s'étendre à de nombreuses compactifications de la théorie IIB de la forme $\mathrm{AdS}_5 \times X_5$, où X_5 est une variété de Sasaki–Einstein, duales à diverses *théories conformes des champs* (CFT) généralisant la théorie $\mathcal{N} = 4$ super Yang–Mills. C'est la *correspondance AdS–CFT*.

Cette correspondance a le bon goût de lier une théorie de supergravité, qui peut être étudiée par des méthodes de géométrie classique, et une théorie quantique des champs à couplage fort. C'est donc une excellente opportunité pour étudier ces dernières, et les applications de cette correspondance holographique allant dans ce sens ont été nombreuses, aussi bien pour l'étude de théories supersymétriques que dans d'autres domaines de la physique, comme en chromodynamique quantique (on parle de correspondance AdS–QCD) ou en théorie de la matière condensée (AdS–CMT).

Les correspondances holographiques ont également été généralisées à d'autres théories de gravité, dans l'espoir de pouvoir étudier de plus en plus finement les propriétés quantiques des trous noirs. En particulier, l'étude d'une théorie de gravité dite de *Jackiw–Teitelboim* (JT) en dimension $d = 2$ a suscité beaucoup d'intérêt depuis 2015, année où Kitaev a argumenté que le modèle de *Sachdev–Ye–Kitaev* de matière condensée est le dual holographique de la gravité de JT, dans un certain régime. Les théories de gravité en deux dimensions sont intimement liées aux espaces de Teichmüller des surfaces lisses, que nous allons maintenant brièvement présenter.

Les correspondances AdS–CFT et jauge–gravité sont présentées dans le Chapitre 8.

Espaces de Teichmüller et généralisations

Espaces de Teichmüller classiques. Une surface lisse S connexe est un espace topologique connexe muni d'un *atlas* lisse, qui est la donnée de *cartes* $(U_i, f_i)_{i\in I}$, où I est un ensemble tel que $S \subset \bigcup_{i\in I} U_i$, et $\forall i \in I$, U_i est un ouvert de S et $f_i : U_i \to \mathbb{R}^2$ est un homéomorphisme de son domaine de définition sur son image, et tel que pour tous $i, j \in I$, $f_i \circ f_j^{-1}$ est un difféomorphisme sur son image. Prosaïquement, une surface lisse correspond à l'intuition qu'on se fait d'une surface S infiniment fine possiblement courbe mais sans plis, et à déformations près où l'on s'autorise toutes les déformations qui ne plient pas S.

Une structure complexe sur S est la donnée d'un atlas holomorphe $(V_k, g_k)_{k\in K}$, défini de la même manière qu'auparavant mais avec $g_k : V_k \to \mathbb{C}$ des homéomorphismes tels que pour tous $k, l \in K$, $g_k \circ g_l^{-1}$ est un biholomorphisme de son domaine de définition sur son image. Prosaïquement, un biholomorphisme est un isomorphisme continu du plan complexe qui préserve les angles. Une structure complexe sur S est

la donnée d'un atlas où les cartes sont des ouverts de $\mathbb{C}$, et où les fonctions de transitions d'une carte à une autre, quelconques, préservent les angles. De plus, la structure complexe est également définie à déformations près, mais où cette fois-ci les seules déformations qu'on s'autorise sont celles qui préservent les angles.

Puisqu'un tel isomorphisme du plan qui préserve les angles est un cas (très) particulier d'isomorphisme lisse du plan, la définition d'une structure complexe est plus contraignante que celle d'une structure lisse. On peut donc se demander : étant donnée une surface lisse connexe S, combien de structures complexes différentes compatibles avec la structure lisse existe-t-il sur S ?

On peut donner deux sens distincts au mot "différentes" de la question précédente, car il existe aussi deux sens distincts de la "reparamétrisation" de S. En effet, on ne veut pas compter plusieurs fois les mêmes structures complexes seulement parce que l'on change la paramétrisation de S. On peut définir deux groupes de reparamétrisations qui préservent la structure lisse : le groupe de difféomorphismes $\mathrm{Diff}(S)$ de S, et son sous-groupe des difféomorphismes qui peuvent être continûment déformés en l'identité, noté $\mathrm{Diff}^0(S)$. On définit alors l'espace de Teichmüller $\mathcal{T}(S)$ de S et l'espace des modules $\mathcal{M}(S)$ de S comme les quotients :

$$\mathcal{T}(S) = \mathrm{StrC}(S)/\mathrm{Diff}^0(S) \ , \qquad \mathcal{M}(S) = \mathrm{StrC}(S)/\mathrm{Diff}(S) \ ,$$

où $\mathrm{StrC}(S)$ désigne l'espace des structures complexes sur S.

Une surface lisse connexe est dite de type fini si elle est obtenue à partir de la surface fermée de genre g en lui retirant un nombre fini k de disques fermés disjoints. On note cette surface $S_{g,k}$, et $\chi(S_{g,k}) = 2 - 2g - k$ sa caractéristique d'Euler.

Le théorème d'uniformisation de Riemann affirme que si S est munie d'une structure complexe, le revêtement universel $\widehat{S}$ de S ne dépend que du signe de $\chi(S_{g,k})$: s'il est négatif, $\widehat{S}$ est le plan hyperbolique $\mathbb{H}$. Il s'agit du cas général, puisqu'il n'y a qu'un nombre fini de couples (g,k) qui donnent $\chi(S_{g,k}) \geq 0$, et un nombre infini dans le cas contraire. On peut alors montrer que $\mathcal{T}(S)$ est homéomorphe à $\mathbb{R}^{6g-6+3s}$. L'espace des modules $\mathcal{M}(S)$ est obtenu en quotientant $\mathcal{T}(S)$ par le *groupe modulaire*, défini comme $\mathrm{Diff}(S)/\mathrm{Diff}^0(S)$. Puisque ce dernier n'agit pas librement sur l'espace de Teichmüller, l'espace des modules est en général un orbifold.

Toujours par le théorème d'uniformisation de Riemann, on peut définir l'espace de Teichmüller d'une surface de type fini $S_{g,k}$ avec $2g + k > 2$ comme le quotient

$$\mathcal{T}(S) = \mathrm{StrHyp}(S)/\mathrm{Diff}^0(S) \ ,$$

où $\mathrm{StrHyp}(S)$ désigne l'espace des métriques hyperboliques sur S, c'est-à-dire les métriques de courbure constante négative égale à -1. Cela identifie l'espace des modules de S à l'espace des configurations de la gravité topologique en dimension 2, qui intervient aussi dans les théories de gravité avec dilaton dont la gravité JT est un exemple.

Il existe une troisième définition de l'espace de Teichmüller $S_{g,k}$ avec $2g + k > 2$ en termes du groupe fondamental $\pi_1(S_{g,k})$. Étant donnée une structure complexe sur $S_{g,k}$, l'application de revêtement $\mathbb{H} \to S_{g,k}$ est une application holomorphe, et le groupe fondamental $\pi_1(S_{g,k})$ agit sur $\mathbb{H}$ par automorphismes du revêtement. Les automorphismes de $\mathbb{H}$ forment le groupe $\mathrm{PSL}_2(\mathbb{R})$, si bien que la donnée d'une structure complexe sur $S_{g,k}$ détermine un morphisme discret et fidèle $\pi_1(S_{g,k}) \to \mathrm{PSL}_2(\mathbb{R})$. Le changement de point-base pour le groupe fondamental conjugue ce morphisme par un élément de $\mathrm{PSL}_2(\mathbb{R})$, et on peut vérifier que

$$\mathcal{T}(S_{g,h}) = \mathrm{Hom}^{\mathrm{d,f}}(\pi_1(S_{g,k}), \mathrm{PSL}_2(\mathbb{R}))/\mathrm{PSL}_2(\mathbb{R}) \ ,$$

où $\mathrm{Hom}^{\mathrm{d,f}}(\pi_1(S_{g,k}), \mathrm{PSL}_2(\mathbb{R}))$ est l'ensemble des morphismes $\pi_1(S_{g,k}) \to \mathrm{PSL}_2(\mathbb{R})$ discrets et fidèles, sur lequel $\mathrm{PSL}_2(\mathbb{R})$ agit par conjugaison.

Les espaces de Teichmüller peuvent être munis de structures intéressantes : métriques, structure complexe, divers plongements, crochet de Poisson, ainsi que des systèmes de coordonnées remarquables. Des objets combinatoires et géométriques comme les *laminations*, introduites par Thurston, ont des liens naturels avec les espaces de Teichmüller. L'espace des laminations $\mathsf{T}(S_{g,k})$ sur une surface $S_{g,k}$ de type fini supporte également des structures intéressantes, et notamment des systèmes de coordonnées. Lorsque la surface $S_{g,k}$ est telle que $2g + k > 2$ et $k > 0$, il existe plusieurs versions d'espaces de Teichmüller et

de laminations. Dans cette thèse nous nous intéresserons en particulier à quatre de ces espaces, dénotés $\mathcal{T}^x(S_{g,k})$, $\mathcal{T}^a(S_{g,k})$, $\mathsf{T}^x(S_{g,k})$, $\mathsf{T}^a(S_{g,k})$, qui admettent des systèmes de coordonnées dans lesquels les formules de changement de cartes sont des *transformations amassées*. Ces espaces de Teichmüller sont introduits dans le Chapitre 2, et les espaces de laminations correspondants, dans le Chapitre 3.

Algèbres et variétés amassées. Une algèbre amassée de rang n est une algèbre engendrée par certaines fractions rationnelles en n indéterminées. L'un des exemples les plus simples d'algèbre amassée complexe est l'algèbre dite de type A_2 :

$$\mathcal{A}_{A_2} = \mathbb{C}\left[a_1, a_2, \frac{1+a_1}{a_2}, \frac{1+a_2}{a_1}, \frac{1+a_1+a_2}{a_1 a_2}\right] , \tag{11}$$

qui est de rang 2. Les algèbres amassées on été introduites par Fomin et Zelevinski entre 2001 et 2006.

La structure d'une algèbre amassée est encodée dans un *carquois*, c'est-à-dire un graphe orienté fini décrit par ses sommets et l'ensemble de ses flèches (d'où son nom). Numérotons $1, \ldots, n$ les n sommets d'un carquois $\mathcal{Q}$ et supposons qu'il n'existe aucun 2-cycle $i \to j \to i$. Il existe une transformation élémentaire de carquois appelée *mutation à un sommet*, et qui est définie de la manière suivante. Soit k un sommet de $\mathcal{Q}$.

1. Pour toute séquence de flèches $i \to k \to j$, on ajoute une flèche $i \to j$.
2. On inverse la direction de toutes les flèches incidentes à k.
3. On ôte tous les 2-cycles $i \to j \to i$ obtenus durant les deux premières étapes.

Le carquois résultant est appelé le résultat de la mutation en k de $\mathcal{Q}$, et dénoté $\mu_k(\mathcal{Q})$. On peut noter qu'une mutation n'agit que sur les flèches d'un carquois, et par conséquent la numérotation des sommets de $\mathcal{Q}$ définit naturellement une numérotation des sommets de $\mu_k(\mathcal{Q})$. Un exemple de mutation est présenté en Figure 18.

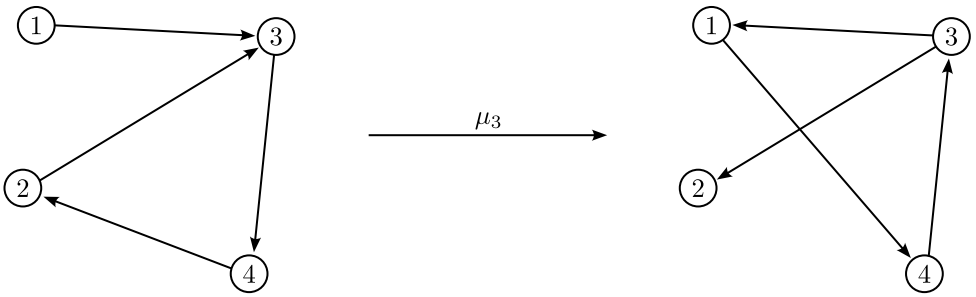


Figure 18: La mutation en 3 d'un carquois.

Étant donné un carquois $\mathcal{Q}$, on peut associer n indéterminées $a_1, \ldots, a_n$ à ses sommets. Le cœur de la théorie des algèbres et variétés amassées réside dans les *formules de mutation* qui définissent des fractions rationnelles en $a_1, \ldots, a_n$ naturellement associées aux sommets de $\mu_k(\mathcal{Q})$. Pour tout $l = 1, \ldots, n$, soit $\mu_k(a_l)$ la fraction rationnelle associée au l-ième sommet de $\mu_k(\mathcal{Q})$. Pour $l \neq k$, on a $\mu_k(a_l) = a_l$. Lorsque $l = k$, on a :

$$\mu_k(a_k) = a_k^{-1}\left(\prod_{i \to k} a_i + \prod_{k \to i} a_i\right) . \tag{12}$$

Ce sont les formules de mutation dites *de type* $\mathcal{A}$. Elles apparaissent dans de nombreux domaines à la fois en mathématique et en physique. Il existe également des formules de mutation *de type* $\mathcal{X}$. Une transformation amassée est une composition de mutations et d'automorphismes du carquois.

L'algèbre amassée définie par un carquois est obtenue en calculant toutes les expressions rationnelles résultats de mutations à partir d'un jeu d'indéterminées initiales, et est l'algèbre polynomiale engendrée par l'ensemble de ces expressions rationnelles. Ces algèbres sont présentées dans le Chapitre 1. Un résultat central de la théorie des algèbres amassées est *le théorème de Laurent*, qui stipule que toutes ces expressions rationnelles sont en fait des polynômes de Laurent en les indéterminées initiales, c'est-à-dire que leur dénominateur est toujours un monôme en ces indéterminées. Par conséquent, les *variétés*

amassées, qui sont (moralement) les duales algébro-géométriques des algèbres amassées, sont obtenues en recollant des tores algébriques déployés le long d'isomorphismes birationnels. Ces variétés algébriques sont introduites dans le Chapitre 4. Pour chaque carquois on construit deux variétés amassées dénotées $\mathcal{A}$ et $\mathcal{X}$. La théorie des variétés amassées a été développée par Fock et Goncharov à partir de 2003 et leur article fondateur : *Moduli space of local systems and higher Teichmüller theory.*

Les formules de mutations sont intéressantes en ce qu'elles ne contiennent pas de soustraction. Elles sont donc définies sur tout semi-corps $(F, +, \times)$, dans lesquels les éléments n'ont pas forcément d'inverse pour la loi $+$. Des exemples de semi-corps de notre intérêt sont $(\mathbb{R}_{>0}, +, \times)$, et, pour $\mathbb{A} = \mathbb{Z}, \mathbb{Q}, \mathbb{R}$, le semi-corps *tropical* $\mathbb{A}^t = (\mathbb{A}, \max, +)$. Puisque les formules de mutation sont bien définies sur le semi-corps F, on peut considérer les F-points $\mathcal{X}(F)$ et $\mathcal{A}(F)$ des variétés $\mathcal{X}$ et $\mathcal{A}$.

Pour chaque surface $S = S_{g,k}$ telle que $2g + k > 2$ et $k > 0$, on peut construire naturellement une classe unique de carquois à mutations près, et donc deux variétés amassées $\mathcal{A}_S$ et $\mathcal{X}_S$. On a alors les résultats suivant :

$$\mathcal{T}^x(S) = \mathcal{X}_S(\mathbb{R}_{>0}) \ , \quad \mathcal{T}^a(S) = \mathcal{A}_S(\mathbb{R}_{>0}) \ , \quad \mathsf{T}^x(S, \mathbb{Q}) = \mathcal{X}_S(\mathbb{Q}^t) \ , \quad \mathsf{T}^a(S, \mathbb{Q}) = \mathcal{A}_S(\mathbb{Q}^t) \ .$$

Composantes de Hitchin. L'étude des représentations du groupe fondamental d'une surface de Riemann S dans des groupes de Lie à l'aide de fibrés de Higgs a été initiée par Hitchin dans *Lie groups and higher Teichmüller spaces*, 1992. Il montre en particulier que si G est un groupe complexe simple adjoint et si G^r est sa forme réelle déployée, deux des composantes connexes de la variété $\mathrm{Hom}(\pi_1(S), G^r)/G^r$ contiennent l'espace de Teichmüller classique $\mathcal{T}(S)$. Elles sont appelées *composantes de Hitchin*. En 2004, Labourrie, et indépendamment, Fock et Goncharov, prouvent que les représentations dans les composantes de Hitchin sont toutes fidèles et discrètes. Ces composantes fournissent donc une généralisation des espaces de Teichmüller présentés ci-dessus, lorsque le groupe $\mathrm{PSL}_2(\mathbb{R})$ est remplacé par G^r.

Dans leur article de 2003, Fock et Goncharov développent une construction combinatoire de ces espaces, et fournissent des systèmes de coordonnées complètement explicites lorsque $G = \mathrm{PGL}_n(\mathbb{R})$. Il s'agit de construire des variétés amassées $\mathcal{X}_{G,S}$ et $\mathcal{A}_{G^L,S}$ qui sont des espaces des modules de G-connections (ou G^L-connections) plates sur S, avec G^L le dual de Langlands de G. Les composantes de Hitchin sont alors identifiées à $\mathcal{X}_{G,S}(\mathbb{R}_{>0})$ et $\mathcal{A}_{G^L,S}(\mathbb{R}_{>0})$. La présentation de la construction de Fock et Goncharov forme le cœur du Chapitre 4.

Dualités amassées. Fock et Goncharov proposent également une généralisation d'une observation sur les laminations dans le cas des espaces de Teichmüller classiques, qui est que les laminations entières de type $\mathcal{X}$ (resp. $\mathcal{A}$) paramétrisent une base des fonctions *universellement positives de Laurent* sur la variété amassée $\mathcal{A}$ (resp. $\mathcal{X}$).

Réseaux spectraux. Les constructions de Fock et Goncharov ont été centrales dans le développement de la théorie des *réseaux spectraux* de Gaiotto, Moore et Neitzke entre 2008 et 2014. Les réseaux spectraux sont l'aboutissement de l'étude des états BPS dans les théories quantiques des champs dites de *classe S*, obtenus en plaçant des M5-branes en théorie M sur une sous-variété de l'espace-temps de la forme $\mathbb{R}^{1,3} \times S$, où S est une surface de Riemann. C'est une vaste généralisation de la théorie de Seiberg et Witten de 1994. Les coordonnées de Fock et Goncharov interviennent dans le calcul des propriétés de *wall-crossing* des états BPS de ces théories.

Sommaire narratif

Ce manuscrit est divisé en cinq parties. Les deux premières ne contiennent pas de résultats nouveaux ; ce sont plutôt des introductions pédagogiques aux deux grandes théories qui interviennent dans la suite.

La première partie est une introduction à la théorie de Teichmüller de Fock et Goncharov (2003-2006). Le premier chapitre porte sur les algèbres amassées ; on introduit les mutations $\mathcal{A}$ et $\mathcal{X}$. Le deuxième chapitre discute des espaces de Teichmüller classiques et de leur paramétrisation amassée. Le troisième chapitre débute par une introduction générale à la théorie des laminations sur les surfaces, et finit par la description de la paramétrisation amassée des espaces de laminations classiques. Enfin, le quatrième chapitre aborde les espaces de Teichmüller supérieurs amassés. On introduit tout d'abord les variétés amassées, puis les composantes de Hitchin, l'approche de Fock et Goncharov, et enfin on discute brièvement des G-espaces de Teichmüller supérieurs où G n'est pas un groupe de Lie réel déployé : il s'agit des espaces de représentations maximales, et enfin des espaces de représentations Θ-positives.

La deuxième partie comporte d'abord un cinquième chapitre sur la supersymétrie, qui tient plus du memento que d'une véritable introduction. Le sixième chapitre porte sur les fondements des théories de supercordes de type II et de la théorie M : définition perturbative des théories IIA et IIB, D-branes, NS-branes et M-branes, dualités S et T. On y discute en particulier des configurations de branes de type Hanany–Witten en dimensions 4 et 5, avec des exemples de l'étude des théories quantiques des champs qu'elles permettent. Le cœur du septième chapitre est l'étude des théories de jauge dans le volume de D3-branes placées au sommet de singularités Calabi–Yau affines toriques de dimension 3. Après une discussion générale des compactifications en théorie des cordes, la théorie sur des D3-branes au point singulier de l'orbifold $\mathbb{C}^3/\mathbb{Z}_5$ est calculée de façon perturbative. Cet exemple motive la géométrie torique, présentée ensuite. On introduit alors les pavages par branes, ou dimères, en discutant notamment de l'implémentation des branes fractionnaires et des orientifolds. Le huitième chapitre porte sur les correspondances "jauge–gravité". Après une présentation de la dualité de Maldacena, celle-ci est généralisée au cas des singularités toriques, puis aux correspondances non-conformes en présence de branes fractionnaires. On classifie ces dernières en fonction de la dynamique de la théorie correspondante à basse énergie.

Les trois parties restantes contiennent les résultats originaux de cette thèse.

La troisième partie discute de l'implémentation de théories quantiques des champs brisant la supersymétrie dynamiquement, dans des systèmes de D3-branes transverses à des orientifolds de singularités (affines toriques Calabi–Yau de dimension 3) décrites par des dimères. Les théories en question sont les modèles de brisure dynamique de supersymétrie (DSB) SU(5) et $3-2$. La motivation originale de cette étude a été la découverte (en 2018 par Buratti, García-Valdecasas et Uranga) d'un canal d'instabilité, dit $\mathcal{N}=2$, dans les dimères alors connus complétant ces deux modèles. Cette instabilité $\mathcal{N}=2$ existe si et seulement si la singularité en question, avant orientifold, n'est pas isolée, ce qui est équivalent à l'existence dans le dimère de branes fractionnaires $\mathcal{N}=2$. Lorsque ces théories DSB sont plongées dans des dimères, l'existence de branes fractionnaires $\mathcal{N}=2$ déstabilise le vide stable prédit par la brisure dynamique de supersymétrie. D'où la problématique : est-il possible de plonger soit le modèle SU(5) soit le modèle $3-2$ dans un dimère définissant une théorie correcte, et sans l'instabilité $\mathcal{N}=2$? L'hypothèse de ne pas utiliser de D7-branes de saveur, par parcimonie, est posée.

Le neuvième chapitre présente des résultats généraux sur les orientifolds de pavages par branes, qui montrent en particulier que presque toutes les implémentations possibles des modèles SU(5) et $3-2$ impliquent immédiatement la présence de l'instabilité $\mathcal{N}=2$. Seule une, dans laquelle sont décrites deux copies de la théorie SU(5), semble ne pas l'engendrer. Le dixième chapitre introduit des méthodes de

construction de pavages par branes avec contraintes, afin d'explorer si ce modèle dit twin-SU(5) peut réellement être plongé dans un dimère correct. Des exemples sont trouvés mais souffrent d'anomalies de jauge. Le onzième chapitre présente des techniques combinatoires pour l'étude des anomalies de jauges dans les orientifolds fixant une droite, deux droites ou quatre points dans le dimère (sans D7-brane de saveur). Dans le cas des orientifolds fixant une ou deux droites, ces techniques montrent que l'existence d'anomalies ne dépend que du diagramme torique, et pas du pavage par branes particulier en question. Armé de ces nouveaux résultats, un dimère satisfaisant, *l'octogone*, est finalement trouvé. Il implémente le modèle twin-SU(5) sans aucune des instabilités connues à ce jour, et est présenté en détails dans le douzième chapitre.

La quatrième partie coïncide avec le treizième chapitre, qui présente la construction de nouveaux orientifolds de dimères sans points fixes. L'involution correspondante est une translation glissée, et donne donc une bouteille de Klein par quotient du tore. Cette construction répond donc à la question de Franco et Vegh (2006) du sens des pavages par branes sur une bouteille de Klein. Ces théories ont les propriétés nécessaires pour définir des théories superconformes. La translation d'une moitié de cellule fondamentale est également étudiée, mais on montre qu'elle conduit toujours à des orientifolds non-supersymétriques. Ces résultats achèvent la classification des types topologiques d'orientifolds de dimères.

La cinquième partie traite de la généralisation des laminations de Thurston aux espaces de Teichmüller supérieurs. Dans le quatorzième chapitre est présenté un tableau conjectural des objets qui, en toute vraisemblance, devraient jouer un rôle dans la définition des laminations supérieures. Le quinzième chapitre est dédié à la construction d'une famille de théories topologiques des champs dérivée des idées du chapitre précédent. Il y a une telle théorie topologique pour chaque algèbre de Hecke associée à un groupe de Coxeter fini, qui associe en particulier à chaque surface topologique épointée un polynôme de Laurent. Ce dernier peut être exprimé en termes de la théorie des représentations de l'algèbre de Hecke correspondante. Le seizième chapitre est une preuve courte et élémentaire de la dualité entre les théories topologiques des champs dites ouvertes-fermées et les anneaux symétriques.

Remerciements – Acknowledgments

Merci Vladimir d'avoir accepté de superviser ma thèse et de m'avoir donné l'occasion de découvrir les mathématiques passionnantes des algèbres amassées et des espaces de Teichmüller supérieurs. I am honoured that Alexander Goncharov, Iñaki Garcia-Etxebarria, Andrew Neitzke, Nikita Nekrasov, Olivier Guichard, Cyril Closset and Yang-Hui He accepted to be part of my jury. I am thankful to Riccardo Argurio, Matteo Bertolini, Maria Chlouveraki, Vladimir Fock, Sebastián Franco, Iñaki García-Etxebarria, Malte Lackmanm, Sebastian Manecke, Ander Retolazza, Daniel Tubbenhauer and Angel Uranga for inputs, discussions and comments at various stages of the projects in which I have been involved.

Cette thèse est une longue aventure qui aura duré bien plus de quatre ans : il y a presque dix ans, je découvrais avec curiosité et excitation la puissance explicative de la chimie organique, éclairé par la passion contagieuse de Clément, Aurélien, Eva, Hakim, Julien, Baptiste et tant d'autres. J'ai des souvenirs vifs et émus des différentes phases de la préparation aux Olympiades Internationales de Chimie, et de l'expérience inoubliable des dix jours passés à Moscou en compagnie de Dorian, Jean, Clément, Ludivine, Sylvie et Aurélien, ainsi que les participants de 72 autres pays. Le monde est grand ! Je n'aurais pu cultiver cette curiosité sans la passion transmise par mes professeur.e.s à l'aube de mon parcours scientifique : Jonathan Etrog, Armel Martin, Marine Pontuer, Henri Lemberg, Yves Dupont et Sophie Martin en classes préparatoires, ainsi que beaucoup d'autres avant le baccalauréat. Le TIPE a été une bonne occasion d'explorer toujours plus de laboratoires de chimie, de jouer avec des machines RMN et des lyophilisateurs, tout ça grâce à la générosité sans bornes de Jean-Marie.

Je garde, de ces années et de celles qui ont suivi durant lesquelles j'ai poursuivi ma scolarité à l'École Normale Supérieure, beaucoup de mes plus chères et chers ami.e.s. Merci Marie, pour la richesse et la diversité de notre amitié, Raphaël, pour nos multiples interactions scientifiques, musicales et marines, Samuel, pour le partage de tes passions et de l'intégrité si forte avec laquelle tu appréhendes la vie, François, pour tout le temps que nous avons passé à refaire le monde à l'ENS, dans notre colloc improvisée au boulevard Davout ou sur des camps. Merci à vous Gauthier, Charles, Hortense, Quitterie, Constantin, Meriem, Andrei, Thomas, Clément, François, Jérôme, Gaspard, pour tout ces moments précieux partagés pendant et depuis ces années. Votre amitié compte beaucoup pour moi. Merci Luc, pour ton accompagnement fidèle et toujours joyeux dans ma pratique de la trompette. Merci également Olivier Debarre, Thomas Duquesnes, Jérémie Bouttier, Ilia Itenberg, Omid Amini, Adel Bilal, Jan Troost, François Loeser, Julien Marché, Costas Bachas, Antoine van Proeyen, Axel Kleinschmidt, Dionysos Anninos, Gui Pimentel, Julian Sonner et Marcel Vonk pour le partage enthousiaste de votre sagesse.

J'ai également eu la chance de travailler un été sur la phénoménologie du couplage trilinéaire du champ de Higgs "au-delà du Modèle Standard", au laboratoire Leprince–Ringuet grâce à Yves Sirois et Roberto Salerno, et ainsi de rencontrer le bon Luca Mastrolorenzo en compagnie de qui le campus de l'X en plein été n'a plus rien de désert. Un an plus tard, j'ai eu la chance de visiter le CERN, et les expériences CMS et LHCb du LHC, avec l'excellent guide Alexandre Zabi.

Je suis très reconnaissant envers mes ami.e.s du *local* rue Madame et des EEUdF: Paul, Isaure, Mathilde, Klervi, Lucas, Raphaëlle, Juliette, Constance, Margaux, Agathe, Hortense et tou.te.s les autres que je n'ai pas la place de nommer, avec qui nous avons tant partagé et de qui j'ai tant appris. Merci Rémy pour la constance de ton amitié et la complicité de nos échanges, et Mathilde pour les multiples découvertes que nous avons faites à deux ainsi que l'enthousiasme de nos envies d'aventure.

Cette thèse serait bien différente si je n'avais pas été accompagné par le guidage bienveillant d'Amir-Kian Kashani-Poor dans ma découverte de la physique mathématique. Yang, I am eternally grateful for your enthusiastic welcome at City, University of London, the mathematical physics through which you guided me, as well as you unfailing support on a variety of matters. I had a wonderful time in London and the chance to meet awesome friends : Abeer, Patrick, Cecilia, Johann, Sonia and her joyful squad.

J'ai pris, un soir de novembre 2017, une décision apparemment anodine : me rendre le lendemain matin à une conférence mystérieuse, "Solvay Doctoral School in Theoretical Physics", à l'Institut Poincaré à Paris. Son impact s'est en fait révélé considérable : j'ai suivi assidûment les trois semaines de cours parisiens sur la supersymétrie, puis suis parti à Amsterdam pour trois semaines supplémentaires et holographiques, pleines de souvenirs avec la fratrie Schlum et une joyeuse bande. Meeting you guys, Antoine, Pierluigi, Pablo, Adrien, Marine and all the others whom I did not get the chance to see again as much in Brussels, was one of the absolute best event in my life, as amazing as unlikely.

Another huge consequence of this tiny unlikely event is the meeting with my dear now-collaborators and friends, Riccardo, Matteo, Seba, Eduardo, Shani and Antoine. I am enormously indebted to you all for the fascinating things you taught me, as well as the benevolent forbearance you have always had for my crackpoteries. Merci mille fois Antoine pour avoir permis cela, toutes les discussions que nous avons eues, scientifiques ou non, et nos groupes de travail improvisés, sur le paradoxe de l'information entre autres. Thank you Eduardo for organizing the Journal Club, which has been a great source of knowledge and thrilling discussions over the years. Thank you Shani for your keen interest in mathematical questions and the chats that ensued. Riccardo, I won't ever be able to thank you enough for kindly allowing me to join the team and making it possible at all, as well as for your gratifying encouragements. I admire the sense of pedagogy that you have demonstrated in lots of Skypes, Riccardo and Matteo. And thanks a million, Riccardo, Matteo and Seba, for your unconditionnal support professionally-wise. Merci aux habitants de la tanière pour votre accueil et notre amitié si précieuse ; merci Andres pour nos discussions.

Mes années à Strasbourg ont été intenses. Merci Francisco, Martin et Pierre-Alexandre pour tous les moments géniaux que nous avons partagés ; merci Yohann et Djibril pour l'ambiance t·r·o·p·i·c·a·l·e dans le burø 114 ; c'était top de donner un cours avec toi Yohann ! La bonne atmosphère d'un laboratoire est une alchimie subtile entre ses membres : merci pour tout ce que vous avez apporté et apportez encore, Nicolas, Guillaume, Amaury (pareil, ce fut top d'enseigner avec toi), Alexander et Alexander, Laura, Cuong, Hao, Archia, Marie, Philippe, Luca, Antoine, Thomas, Clarence, Clément, Basile, Raoul, Adam, et tou.te.s les autres que je n'ai pas la place de citer ici.

Merci de ces moments que l'on a partagés en rando, Martin, Yohann et Xiaolin, de la grimpe au Kronthal, à Block Out et à Roc en Stock, des knöpfle et des escape jeux cidrés.

Pierre, merci pour nos Skypes conditionnellement stables au début de la pandémie, et pour le soutien bienveillant que tu m'as apporté ces dernières années, en faisant don de ton temps et de ton énergie. Ta présence et ton implication ont été plus que rassurantes ; ton aide m'a rendu la thèse plus fluide.

At the conference at CIRM in the Fall of 2020 I had some thrilling discussions with you, Alexander and Georgios. They led to setting up a very fruitful working group with Eugen, Clarence and the three of us. Many thanks guys for the good athmosphere and your enthusiasm; I have learned a lot, and most of all I am very happy I got the chance to know you more. Alexander, merci infiniment pour nos discussion variées, le travail sur les algèbres de Hecke sur lequel nous avons avancé ensemble, et tous les autres moments que nous avons partagés : déjeuners chez toi, balades, jeux et tomates endiablées.

Merci Coline, pour nos beaux échanges ; ton amitié est un présent qui me grandit. Merci Marie, Paul, Pauline, Juliette, Éponine, Eve, Amédée, Jo, Léa, Nina pour les merveilleux moments que nous avons partagés. Merci Marc pour nos discussions toujours passionnantes et la chaleur de ta présence.

En arrivant à Strasbourg il y a quatre ans, je m'y suis immédiatement senti chez moi, et cela grâce à l'accueil sans égal que vous m'avez réservé, Nathalie, Christophe, Jeanne, Lucie et Paul. Lors des semaines que j'ai passées à Hoenheim, nous parlions de la thèse que *j'allais* faire ; je suis heureux de pouvoir vous montrer aujourd'hui l'objet de nos spéculations de jadis.

Merci du fond du cœur à toute ma famille pour son soutien que je sais indéfectible : Hakima, Sophia, Marc, Francis, Aurore, Jacques, Joëlle, Paul-Émile, Pierre-Jean, Michelle, Jean, Vincent, Céphise, Timothée, Odilon, Nicolas, Guillaume, Benjamin et tous leurs compagnes, compagnons et enfants.

Adèle, je suis très fier d'achever la rédaction de ce manuscrit, qui n'aurait tout simplement pas existé sans tes encouragements et ton aide récurrente. J'ai une chance incroyable de t'avoir dans ma vie ; ta présence orangée illumine mon quotidien et l'absolu de ton honnêteté relationnelle me guide invariablement vers le meilleur de moi-même. Un million de mercis pour toutes tes croquetteries et ton amour !

Antoine, Raphaël, Adèle, Papa et Maman, merci d'avoir passé un temps fou à relire de nombreuses parties de cet ouvrage ; il est en partie le vôtre.

Papaul, Mam, Madame Aouici, Monsieur Aouici, Papi, Mamie, Papa, Maman, Tycho et Bethsabée, c'est grâce à vous, avant tout, que je suis qui je suis. Je vous dédie ce manuscrit, avec beaucoup d'amour.

Introduction

A dramatic turn of events occurred in the 1980's in fundamental physics, and particularly in elementary particle physics. After decades of intense experimental discoveries guiding the building of relentlessly improving phenomenological models – culminating with the Standard Model of particle physics – began an era of theoretical questions of a different nature. Many of these new developments were hopelessly out of reach of experiments; that led to a complete new paradigm of a more mathematical essence. Meanwhile, undoubtedly stirred by Michael Atiyah's impulse [Hit20], the study of interactions between gauge theories, supersymmetry, topology and geometry became an important research field. This contributed to popularizing previously unfamiliar mathematical objects in the theoretical physics community, in particular thanks to Edward Witten's catalysis. All this ultimately led to the emergence of a new branch of mathematical physics, sometimes referred to as physical mathematics, experimental mathematics or *theoretical mathematics*, in contrast with the old field of *rigorous mathematics* [JQ93]. Among theoretical mathematics, string theory without doubt holds a very special place. Very abstract mathematical structures naturally come alive in it, such as complex algebraic geometry, manifolds with special holonomy or category theory. The rich mathematical content of string theory, together with the fact that it is pervaded by a deep physical intuition, has many times proved to be very fruitful in order to conjecture mathematical results and correspondences. The best example of this is perhaps the discovery of mirror symmetry.

The common thread in this dissertation is another example of mathematical objects which lie at an interface between pure mathematics and string theory: cluster structures. Cluster algebras have been introduced by Sergei Fomin and Andrei Zelevinsky in [FZ02], as the abstraction of algebraic relations appearing in George Lusztig's description of total positivity in reductive Lie groups and dual canonical bases of quantized universal enveloping algebras. Shortly after, Vladimir Fock and Alexander Goncharov defined cluster varieties in the context of higher Teichmüller theory [FG06]. Seiberg duality of four-dimensional $\mathcal{N} = 1$ quiver gauge theories, proposed by Nathan Seiberg in [Sei95], was subsequently reinterpreted as a tropical mutation. Cluster structures also appear in the study of BPS states of four-dimensional $\mathcal{N} = 2$ gauge theories [GMN13c], and were even more recently shown to play a role in planar amplitudes of four-dimensional $\mathcal{N} = 4$ super Yang-Mills theories [AHBC$^+$16]. The work I have produced as part of my PhD articulates in two main directions. Let us discuss them in turn.

The Teichmüller space $\mathcal{T}(S)$ of a compact oriented topological surface S is a connected component of the $\mathrm{PSL}_2(\mathbb{R})$-character variety $\mathrm{Hom}(\pi_1(S), \mathrm{PSL}_2(\mathbb{R}))/\mathrm{PSL}_2(\mathbb{R})$ of S, consisting solely of discrete and faithful representations. Hitchin has shown that when G is the split real form of an adjoint reductive group, there is a connected component of the Γ-character variety $\mathrm{Hom}(\pi_1(S), G)/G$ of S in which the Teichmüller space $\mathcal{T}(S)$ embeds [Hit92]. It was later proved that Hitchin's components consist also solely of discrete and faithful representations. For any adjoint reductive group G with Langlands dual G^L, and for S possibly non-closed, Fock and Goncharov have constructed a pair of cluster varieties $(\mathcal{X}_{G,S}, \mathcal{A}_{G^L,S})$ whose spaces of real positive points $\mathcal{X}_{G,S}(\mathbb{R}_{>0})$ and $\mathcal{A}_{G^L,S}(\mathbb{R}_{>0})$ are variants of Hitchin's components [FG06]. The space $\mathcal{T}(S)$ can be defined in terms of different structures on S, and is related to other interesting objects on S, such as Thurston's laminations. Rational laminations are systems of weighted curves on S, and the projectivisation of their space's completion provides a spherical compactification of $\mathcal{T}(S)$ to which the action of the mapping class group extends. Fock and Goncharov have also proposed a generalization of two variants of rational laminations spaces when G is a general reductive adjoint group: $\mathcal{X}_{G,S}(\mathbb{Q}^t)$ and $\mathcal{A}_{G^L,S}(\mathbb{Q}^t)$. The projectivisation of these spaces of laminations also provides a spherical compactification of the corresponding varieties $\mathcal{X}_{G,S}(\mathbb{R}_{>0})$ and $\mathcal{A}_{G^L,S}(\mathbb{R}_{>0})$. However, the generalization of rational laminations as systems of weighted curves on S remained elusive.

The first part of my work (in collaboration with Vladimir Fock and Alexander Thomas) is directed

towards this generalization. We conjecture the following:

Conjecture 0.1. *A G-higher lamination is a union of curves on S colored with the positive simple roots of the affine group $\widehat{G^L}$ corresponding to G^L, with vertices of some type determined by the braid and quadratic relations in the (spherical) Hecke algebra of G^L. Moreover, G-higher laminations are in one-to-one correspondence with ramified coverings $\tilde{\Sigma}$ of S in T^*S which are Lagrangian subspaces and such that the restriction of the Liouville 1-form λ to $\tilde{\Sigma}$ is integral, i.e.*

$$\lambda_{|\tilde{\Sigma}} \in H^1(\tilde{\Sigma}, \partial\tilde{\Sigma}, \mathbb{Z}) \ , \tag{13}$$

modulo Hamiltonian diffeomorphisms preserving the zero section.

However, we do not yet have a satisfying description of these higher laminations. The framework we developed nevertheless led us to the interesting construction of non-commutative open-closed 2-dimensional topological quantum field theories, which associate a Laurent polynomial with integer coefficients to every punctured surface. Each such topological quantum field theory corresponds to the Iwahori–Hecke algebra of a finite Coxeter system. Last, we developed a short proof of the correspondence between algebras endowed with a symmetrizing trace and open-closed topological quantum field theories. Besides pursuing towards a definition of higher laminations, the topological quantum field theories we have introduced open interesting new research directions. For example, it is natural to wonder whether a categorification in terms of Soergel bimodules exist, or even motivic generalizations of these Hecke topological quantum field theories.

The AdS–CFT correspondence is one of the greatest theoretical discoveries of the last decades; in particular, it has far-reaching implications in the study of quantum black holes and the dynamics of gauge theories. By considering D3-branes at affine toric Calabi–Yau singularities, the original duality between type IIB superstrings on $\mathrm{AdS}_5 \times S^5$ and four-dimensional $\mathcal{N} = 4$ super Yang–Mills has been generalized to correspondences between type IIB superstrings on $\mathrm{AdS}_5 \times Y_5$ (with Y_5 is a five-dimensional Sasaki–Einstein manifold) and $\mathcal{N} = 1$ quiver CFTs. The latter are nicely described in terms of brane tilings, i.e. dimer models on a torus. Seiberg dualities on brane tilings are encoded as urban renewal, or spider moves. The addition of fractional branes and/or orientifolds breaks the conformal invariance, yielding gauge–gravity correspondences between locally AdS warped-throats and $\mathcal{N} = 1$ quiver gauge theories. In this context, the question of whether it is possible to realize Dynamical Supersymmetry Breaking (DSB) models on deformation fractional branes at singularities is of great relevance. Some instances of orientifold singularities with realizations of the SU(5) and $3 - 2$ DSB models were proposed in [FHK$^+$07]. However, it was shown in [BGVU19] that the vacuum of these DSB models was destabilized into runaway, eventually because the corresponding singularities were not isolated, as proved shortly after in [ABMP19]. A singularity is non-isolated if and only if it hosts $\mathcal{N} = 2$ fractional branes, which are pushed to infinity by the DSB sector along $\mathcal{N} = 2$ flat directions in the moduli space.

As a second part of my work, Ricardo Argurio, Matteo Bertolini, Sebastián Franco, Eduardo García-Valdecasas, Shani Meynet, Antoine Pasternak and myself, tackled the following refinement of the question of above: is it possible to realize Dynamical Supersymmetry Breaking (DSB) models on deformation fractional branes at isolated singularities? A study of the dimer model substructures needed to encode the SU(5) and $3 - 2$ DSB models showed that all of them imply the presence of $\mathcal{N} = 2$ fractional branes, except one. However, the fact that the latter could be embedded in a consistent brane tiling in a satisfactory way was not obvious. In particular, after orientifold the model can suffer gauge anomalies. As a consequence, we performed a general analysis of the anomaly cancellation conditions in orientifolded brane tilings, which constrained the type of orientifolds than could possibly work. Armed with these knowledge, we developed an inverse algorithm with constraints, which led us to a dimer model hosting a twin version of the SU(5) DSB model. It is very likely that it is the simplest of such models, though it is a very complicated brane tiling: it encodes a theory with fourteen simple gauge factors. The mere fact that we were able to compute such a dimer proves the effectiveness of our methods. That such a model exists proves that DSB models exist in string theory, or at least that the previously discovered instabilities do not rule them out. An exciting continuation of this research would be to generalize the model for $M \gg 1$ deformation fractional branes; if the corresponding gauge theory is still DSB, one could hope to obtain a gravitational dual description of dynamical supersymmetry breaking.

Together with Eduardo García-Valdecasas, Shani Meynet and Antoine Pasternak, we explained the physical meaning of dimer models on a Klein Bottle (that they can possibly encode superconformal field

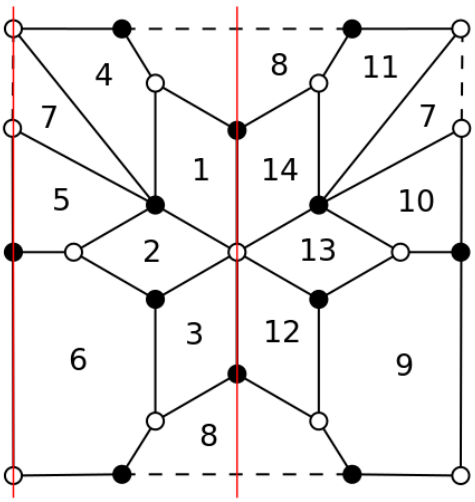


Figure 19: The brane tiling hosting the twin SU(5) model on a fractional brane after orientifold.

theories had been noted in [FV06], hence raising the question of their interpretation). This completed the description of orientifolds of brane tilings as involutions of the underlying bipartite maps.

In [GK11], a class of cluster integrable systems build from dimer models on a torus was introduced. It would be interesting to understand what part they play in the physics of D3-branes at toric affine CY3 singularities, and whether the cluster dualities between them and brane tilings have any physical meaning. Moreover, the orientifolded brane tilings presumably correspond to cluster integrable systems on double Bruhat cells of affine groups not of type $\widehat{A}_N$, hence connecting with [FM16a].

I chose to include two pedestrian introductory parts in this dissertation, one on cluster higher Teichmüller theory, and the other on the physics background needed to discuss DSB models in brane tilings. They are for sure less rigorous than research articles or standard textbooks, however since comprehensive introductions to these topics are rare, I think that they might be of some interest at least for those who would like to have a first overview of either subject, if any.

Since cluster algebras and varieties are the common thread of this manuscript, the introduction to cluster higher Teichmüller theory comes first, and constitutes Part I. The physics introduction follows in Part II. Then, Part III consist of the articles:

1. *Dimers, orientifolds and stability of supersymmetry breaking vacua* [ABF$^+$21b],
   with R. Argurio, M. Bertolini, S. Franco, E. García-Valdecasas, S. Meynet and A. Pasternak, hep-th/2007.13762,
   Journal of High Energy Physics, 2021(1), 61,

2. *Inverse algorithm and triple point diagrams* [Tat21],
   hep-th/2111.02195,

3. *Dimers, orientifolds and anomalies* [ABF$^+$21a],
   with R. Argurio, M. Bertolini, S. Franco, E. García-Valdecasas, S. Meynet and A. Pasternak, hep-th/2009.11291,
   Journal of High Energy Physics, 2021(2), 153,

4. *The Octagon and the non-supersymmetric string landscape* [ABF$^+$21c],
   with R. Argurio, M. Bertolini, S. Franco, E. García-Valdecasas, S. Meynet and A. Pasternak, hep-th/2005.09671,
   Physics Letters B, 815, 136153,

in order of appearance on ArXiv, while Part IV is based on:

1. *Dimers in a bottle* [GVMPT21],
   with E. García-Valdecasas, S. Meynet and A. Pasternak, hep-th/2101.02670
   Journal of High Energy Physics, 2021(4), 274.

The core of Part V is the article on Hecke topological quantum field theories (TQFTs):

1. *Topological quantum field theories from Hecke algebras* [FTT21],
   with V. Fock and A. Thomas, math.QA/2105.09622,

preceded by a chapter casting the characters that presumably play a part in the combinatorial definition of higher laminations, and followed by another containing the short proof (that should appear as a preprint one day) of the equivalence between symmetric algebras and open-closed TQFTs.

Part I

Cluster higher Teichmüller theory

Introduction

The main goal of this part of the dissertation is to introduce, in a pedestrian way, Fock and Goncharov's cluster higher Teichmüller theory [FG06]. Rather than focusing on details and rigor, we aim to follow a pedagogical line of thoughts which hopefully will pinpoint the main ideas in a clear way. Our hope is that this part may be useful as a first introduction to the topic, after which anyone interested would be able to go through the original research papers – and in particular [FG06] – more easily.

This topic is of course not uncorrelated with the sequel of the dissertation; Part V in particular strongly relies on the ideas introduced below. Since the work presented in Part V is mainly motivated by the quest towards the generalization of Thurston's laminations, we will put a special emphasis on classical laminations in the following four introductory chapters.

Cluster algebras – introduced in Chapter 1 – are the common thread throughout the manuscript. On the one hand, cluster algebras (and varieties) play a central part in cluster higher Teichmüller theory, as the name suggest. On the other, they also appear in the context of brane tilings introduced in Section 7.5: under Seiberg duality, the ranks of the unitary gauge groups in a brane tiling transform as tropical cluster $\mathcal{A}$ variables. Brane tilings are central characters in Parts III and IV. An interesting example in geometry where such structures appear is the Teichmüller theory of non-closed surfaces of finite type, or more generally, ciliated surfaces. Chapter 2 is devoted to the presentation of the so-called Teichmüller space with holes and decorated Teichmüller space of a ciliated surface. Then, Chapter 3 discusses laminations, and more specifically rational laminations on ciliated surfaces. Cluster transformations also naturally appear in their description. These Teichmüller and laminations spaces motivate the introduction of cluster ensembles and varieties in Chapter 4. With these objects at hand, we turn to higher Teichmüller theory and more specifically to G-higher Teichmüller theory, where G is the split real form of a reductive group, either with finite center or simply-connected. To every ciliated surface S and adjoint reductive group G, one can associate a pair $(\mathcal{X}^+_{G,S}, \mathcal{A}^+_{G^L,S})$ of higher Teichmüller spaces, where G^L is the *Langlands dual* of G. Prosaically, G^L is the simply-connected group whose root (resp. weight) lattice is the weight (resp. root) lattice of G [1]. Both Teichmüller spaces $\mathcal{X}^+_{G,S}$ and $\mathcal{A}^+_{G^L,S}$ project to the moduli space $\mathcal{L}^+_{G,S}$ of positive representations $\pi_1(S) \to G(\mathbb{R})$ modulo $G(\mathbb{R})$-conjugation, which is a connected component of the $G(\mathbb{R})$-character variety of S:

$$\mathrm{Hom}(\pi_1(S), G(\mathbb{R}))/G(\mathbb{R}) \ , \tag{14}$$

and generalizes Hitchin's components. Here $G(\mathbb{R})$ is the split real form of G. The space $\mathcal{L}^+_{G,S}$ consists solely of discrete and faithful representations, and there is a canonical embedding of the classical Teichmüller space in it.

Let now G be any connected semi-simple real Lie group with finite center. There are fundamental differences in the G-character varieties for G split real and for G complex or compact, which allow in particular the existence of connected components such as $\mathcal{L}^+_{G,S}$ in the first case. A connected component of the character variety consisting solely of discrete and faithful representations is said to be a G-higher Teichmüller space. Apart from split real forms of reductive groups, Hermitian Lie groups of tube type are also known to yield G-higher Teichmüller spaces. Recently, the notion of Θ-positive representations was proposed by Guichard and Wienhard in [GW18]. It defines in turn moduli spaces of Θ-positive representations of $\pi_1(S)$ into G which have been shown to be a union of G-higher Teichmüller spaces [GLW21]. Θ-positivity provides a unified framework which includes both cases of real split groups and Hermitian groups of tube type, as well as two new families of real Lie groups for which higher Teichmüller spaces therefore exist. We will discuss Θ-positivity briefly at the end of Chapter 4.

[1] Similarly, if G is simply-connected, its Langlands dual is the adjoint group G^L with root and weight lattices exchanged.

Chapter 1

Cluster algebras

Cluster algebras are commutative algebras introduced in a series of papers by Fomin and Zelevinsky [FZ02, FZ03, FZ05, FZ07]. To give a first example, a complex cluster algebra $\mathcal{A}$ of rank n satisfies

$$\mathbb{C}[a_1, \ldots, a_n] \subset \mathcal{A} \subset \mathbb{C}(a_1, \ldots, a_n) \ , \tag{1.1}$$

i.e. it is a sub $\mathbb{C}[a_1, \ldots, a_n]$-module of $\mathbb{C}(a_1, \ldots, a_n)$. It is defined as the polynomial algebra $\mathbb{C}[\alpha_i]$ generated by a finite or countable family $(\alpha_i)_{i \in I}$, where $\alpha_i \in \mathbb{C}(a_1, \ldots, a_n)$. Every such α_i is obtained by successive *mutations*, whose exact expression depends on an auxiliary combinatorial object: a quiver. One of the simplest examples of a cluster algebra over $\mathbb{C}$ is the rank-2 cluster algebra denoted $\mathcal{A}_{A_2}(\mathbb{C})$, defined as:

$$\mathcal{A}_{A_2}(\mathbb{C}) = \mathbb{C}\left[a_1, a_2, \frac{1+a_1}{a_2}, \frac{1+a_2}{a_1}, \frac{1+a_1+a_2}{a_1 a_2}\right] \ . \tag{1.2}$$

Cluster algebras generalize the algebraic structure that appears in the study of the (dual) *canonical basis* of the ring $\mathbb{C}[\mathrm{SL}_3/U]$ [GZ86] where U is the subgroup of upper-triangular unipotent matrices, and more generally in the dual canonical bases in $\mathbb{C}[G/U]$ [Lus90], where G is a connected simply-connected semi-simple algebraic group, and U a maximal unipotent subgroup of G. The theory of dual canonical bases is related to total positivity in reductive Lie groups [Lus10, Lus94, FZ99a]; the exchange relations appearing in the study of positivity in double Bruhat cells have also been a motivation for the definition and the study of cluster algebras. We refer to the introduction of [FZ02] for a more detailed discussion of the motivations and the historical context, and well as references.

Cluster algebras also appear in high-energy physics. Some instances are Seiberg duality of four-dimensional $\mathcal{N} = 1$ supersymmetric quantum field theories [Sei95, BD02] (discussed in Chapter 5), the study of BPS states in theories of class S [GMN13c] (evoked in Chapter 14), scattering amplitudes of planar four-dimensional $\mathcal{N} = 4$ super Yang–Mills theory [AHBC$^+$16] and some classes of algebraic integrable systems [GK11, FM16b] (briefly introduced in Chapter 7).

In this chapter we introduce cluster algebras in a pedestrian way. First, we present some aspects of total positivity in Section 1.1. More precisely, we define totally positive Grassmannians, show how efficient total-positivity tests can be designed for elements of $\mathrm{Gr}_{2,m}(\mathbb{C})$ and how these are related to triangulations of an m-gon. We also touch upon total-positivity in $\mathrm{GL}_n(\mathbb{C})$ and $\mathrm{SL}_n(\mathbb{C})$, again showing how efficient total-positivity tests can be designed and how these correspond to combinatorial objects known as pseudo-line arrangements. In Section 1.2 we define (generalized) quivers, which generalize both triangulations and pseudo-line arrangements. Elementary moves in the latter correspond to mutations of the corresponding quiver. We also discuss quivers with potentials, and show how such structures appear as one considers bipartite graphs on oriented surfaces. The latter will play a prominent role in Parts III and IV. The definition of cluster algebras is given in Section 1.3; we also provide classic examples of cluster algebras of rank 1 and 2. The Laurent phenomenon – a cornerstone of the cluster theory, is discussed in Section 1.4. We present some applications in Section 1.5. Lastly, we introduce Y-patterns in Section 1.6. These are also rational expressions in the variables of a cluster algebra, enjoy interesting properties and will be of importance in the sequel of the manuscript.

1.1 Total positivity

We introduce total positivity, first by considering the totally positive grassmannian $\mathrm{Gr}^+_{2,m}$, following [FWZ06], and then total positivity of matrices in GL_n and SL_n, following [FZ99b, Fom10]. These examples are also discussed in the introduction of [FZ02].

1.1.1 The totally positive Grassmanian $\mathrm{Gr}^+_{2,m}$

Let $k \leq m \in \mathbb{Z}_{>0}$. The space of k-dimensional subspaces of an m-dimensional complex vector space V is a complex manifold called the (complex) *Grassmannian* manifold of k planes in $\mathbb{C}^m$ and denoted $\mathrm{Gr}_{k,m}(\mathbb{C})$. Fixing a basis $(e_1, ..., e_m)$ of V allows to identify the points of $\mathrm{Gr}_{k,m}(\mathbb{C})$ with equivalence classes of rank k matrices in $\mathcal{M}_{k,m}(\mathbb{C})$ in the following way: the point in $\mathrm{Gr}_{k,m}(\mathbb{C})$ corresponding to such a matrix is the span of its rows in $\mathbb{C}^m$, and conversely a point $z \in \mathrm{Gr}_{k,m}(\mathbb{C})$ corresponds to the equivalence class of matrices in $\mathcal{M}_{k,m}(\mathbb{C})$ whose rows form a linear basis of the subspace z. The equivalence relation is left-multiplication by $\mathrm{GL}_k(\mathbb{C})$ on $\mathcal{M}_{k,m}(\mathbb{C})$.

Definition 1.1. *Let $l \leq k$, $I \subset \{1, ..., k\}$, $J \subset \{1, ..., m\}$ subsets with l elements and $M \in \mathcal{M}_{k,m}$. The determinant of the $l \times l$ matrix obtained from M by taking only the rows in I and columns in J is the (I, J)-minor of M, denoted $\Delta_{I,J}(M)$. If $I = [|1, k|]$ and if J is a subset of $[|1, m|]$ of cardinal k, the minor $\Delta_{[|1,k|],J} =: \Delta_J(M)$ is the Plücker coordinate corresponding to J.*

The Plücker coordinates $\{\Delta_J(z)\}_{|J|=k}$ are functions

$$\Delta_J(z) : \mathcal{M}_{k,m}(\mathbb{C}) \longrightarrow \mathbb{C} \tag{1.3}$$

which are invariant under left-multiplication by $\mathrm{GL}_k(\mathbb{C})$ except for the fact that they rescale with a common factor $\det(g)$. They can nonetheless be considered as homogeneous coordinates on $\mathrm{Gr}_{k,m}(\mathbb{C})$;

$$\mathrm{Gr}_{k,m}(\mathbb{C}) \longrightarrow \mathbb{P}^N(\mathbb{C}) \tag{1.4}$$

is an embedding called the *Plücker embedding*, where $N = \binom{m}{k} - 1$.

Let us now restrict to the case $k = 2$, and as before let us identify $\mathrm{Gr}_{2,m}(\mathbb{C})$ with the space of equivalence classes in $\mathcal{M}_{2,m}(\mathbb{C})$ under left $\mathrm{GL}_2(\mathbb{C})$ multiplication. The minors of a matrix $M \in \mathcal{M}_{2,m}(\mathbb{C})$ are either 1×1 – these are the entries of the M, or 2×2 – these are the Plücker coordinates on $\mathrm{Gr}_{2,m}(\mathbb{C})$.

Definition 1.2 (Def. 3.1 of [Pos06])**.** *The totally positive Grassmannian $\mathrm{Gr}^+_{2,m}$ is the subset of $\mathrm{Gr}_{2,m}(\mathbb{C})$ consisting of the elements represented by the equivalence classes of matrices in $\mathcal{M}_{2,m}(\mathbb{C})$ whose Plücker coordinates are either all in $\mathbb{R}_{>0}$, or all in $\mathbb{R}_{<0}$.*

The Plücker coordinates on $\mathrm{Gr}_{2,m}(\mathbb{C})$ being labeled by a pair of integers $i < j \in [|1, m|]$, from now on we will denote them P_{ij} instead of $\Delta_{\{i,j\}}$. One can wonder how one could test whether all the maximal minors of a matrix $M \in \mathcal{M}_{2,m}(\mathbb{C})$ are in $\mathbb{R}_{>0}$. The brute-force computation of all the $\binom{m}{2}$ minors surely works, however it is not necessary to do that many tests, since there are relations between the minors.

Proposition 1.3. *For $1 \leq i < j < k < l \leq m$, one has:*

$$P_{ik}P_{jl} = P_{ij}P_{kl} + P_{il}P_{jk} \ . \tag{1.5}$$

The relation of Equation (1.5) is reminiscent of the Ptolemy theorem[1]. Moreover, Plücker relations exhaust the relations between Plücker coordinates: the homogeneous coordinate ring of $\mathrm{Gr}_{2,m}(\mathbb{C})$ is

$$\mathcal{O}_{\mathrm{Gr}_{2,m}(\mathbb{C})} = \mathbb{C}\left[(P_{ij})_{i<j\in[|1,n|]}\right]/\mathrm{PR} \ , \tag{1.6}$$

wher PR is the set of all Plücker relations.

Equation (1.5) implies that if $P_{jl}, P_{ij}, P_{kl}, P_{il}, P_{jk} \in \mathbb{R}_{>0}$, then it also holds that $P_{ik} \in \mathbb{R}_{>0}$. Hence, an interesting refinement of the question of above is: are there efficient total positivity tests for the elements of $\mathrm{Gr}_{2,m}(\mathbb{C})$? It turns out to be the case, as we are going to explain now. In order to ensure total positivity, it suffices to check the positivity of only $2m - 3$ well-chosen minors - this is a drastic reduction of complexity compared to the brute-force algorithm.

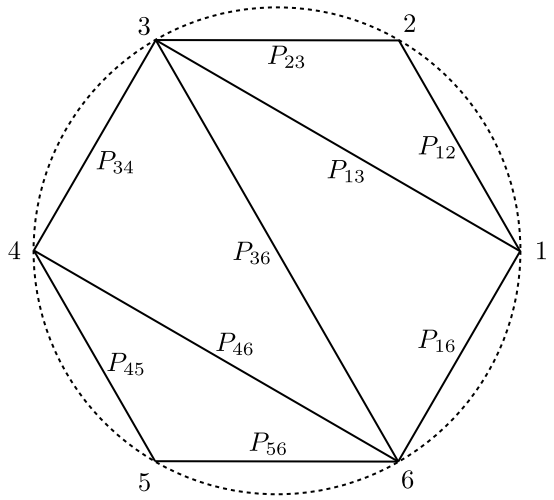


Figure 1.1: A triangulation Γ of a regular hexagon.

Let us consider a convex m-gon P_m with vertices labeled $1, \ldots, m$ clockwise and assign the Plücker coordinate P_{ij} to the chord with endpoints i and j. Let Γ be a triangulation of P_m by pairwise non-crossing diagonals but possibly common endpoints: it consists of m sides and $m-3$ diagonals and hence there is a distinguished collection $\tilde{x}(\Gamma)$ of $2m-3$ Plücker coordinates corresponding to the chords of Γ.

The Plücker coordinates corresponding to the sides of the m-gon P_m are said to be *frozen variables* or coefficients; they belong to every triangulation. The $m-3$ Plücker coordinates which correspond to chords in the interior of P_m are called *cluster variables*; together frozen and cluster variables form the *cluster* $x(\Gamma)$. The union of the cluster and the frozen variables is called *extended cluster* $\tilde{x}(\Gamma)$.

One can check that the $2m-3$ Plücker coordinates associated with the chords of any triangulation T are algebraically independent as polynomials on the entries of the matrices in $\mathcal{M}_{2,k}(\mathbb{C})$. Thus, the following result implies that the positivity of the Plücker coordinates in the extended cluster corresponding to Γ is enough to ensure the total positivity of the corresponding matrix.

Theorem 1.4. *Let $i < j \in [|1, m|]$. The Plücker coordinate P_{ij} can be expressed as a subtraction-free rational expression in the elements of any given extended cluster $\tilde{x}(\Gamma)$.*

The theorem is implied by the following three points:

1. Any P_{ij} appears as an element of an extended cluster $\tilde{x}(\Gamma)$ for some triangulation Γ of P_m.
2. If Γ and Γ' are any two triangulations of P_m, Γ can always be transformed into Γ' by a finite sequence of *flips*.

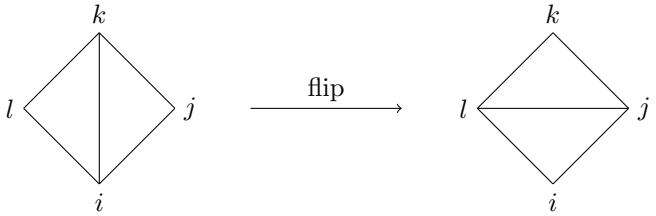


Figure 1.2: The flip of the diagonal in the quadrilateral $ijkl$.

3. Upon the flip of the diagonal of a quadrilateral $ijkl$ appearing in the triangulation Γ as in Figure 1.2, P_{ki} is replaced by

$$P_{il} = \frac{P_{jk}P_{li} + P_{kl}P_{ij}}{P_{kl}} \ , \tag{1.7}$$

[1] In any quadrilateral inscribed in a circle on a plane, the products of the lengths of opposite sides add up to the product of the lengths of the two diagonals.

while all other Plücker coordinates are left unchanged. This transformation is a rational subtraction-free expression of the elements of the extended cluster associated with Γ.

1.1.2 Positivity in $\mathrm{GL}_n(\mathbb{C})$ and $\mathrm{SL}_n(\mathbb{C})$

Definition 1.5. *Let $G = \mathrm{GL}_n$ or $G = \mathrm{SL}_n$. An invertible matrix $m \in G$ is totally non-negative (resp. totally positive) if for all $I, J \subset [|1, n|]$ such that $|I| = |J|$, one has $\Delta_{I,J}(m) \in \mathbb{R}_{\geq 0}$ (resp. $\Delta_{I,J}(m) \in \mathbb{R}_{>0}$).*

The systematic study of such matrices was initiated in [GK02]. In particular, Gantmacher and Krein proved that such matrices have distinct real positive eigenvalues. For a general survey on this notion of total positivity, see [And87]. The following proposition is a direct consequence of the Cauchy–Binet formula:

Proposition 1.6. *Totally non-negative (resp. positive) matrices form a multiplicative sub-semigroup of G denoted $G_{\geq 0}$ (resp. $G_{>0}$).*

Now, the next proposition implies that the study of the semi-group $G_{\geq 0}$ reduces to the study of its sub-semigroup $B_{\geq 0}$, where B is the subgroup of upper-triangular matrices in G.

Proposition 1.7 (Cryer's splitting lemma [Cry73, Cry76])**.** *A matrix $m \in G$ is totally non-negative if and only if it has a Gaussian decomposition $m = u_- du$, where u_- is non-negative lower-triangular unipotent (i.e. with a diagonal of 1's), d is non-negative diagonal and u is non-negative upper-triangular unipotent.*

Furthermore, the Loewer–Whitney theorem [Loe55, Whi52] implies that every non-negative unipotent upper-triangular matrix $u \in U_{\geq 0}$ can be written as a product of totally non-negative matrices $x_i(t)$ where $t \in \mathbb{R}_{\geq 0}$ and where $x_i(t)$ is the matrix in $\mathrm{GL}_n(\mathbb{C})$ with 1's on the diagonal, t at the entry $(i, i+1)$ and 0 elsewhere:

$$x_i(t) = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & t & \cdots & 0 \\ 0 & \dots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{1.8}$$

Definition 1.8. *Let U be the subgroup of $\mathrm{SL}_n(\mathbb{C})$ consisting of unipotent upper-triangular matrices. An element $x \in \mathrm{SL}_n(\mathbb{C})/U$ is totally positive if all its flags minors $\Delta_I(x)$ are in $\mathbb{R}_{>0}$ where the flag minor $\Delta_I(x)$ corresponding to a non-empty proper subset $I \subset [|1, n|]$ is defined as*

$$\Delta_I : x \longmapsto \Delta_I(x) = \det\{x_{ij} | i \in I,\ j \leq |I|\}\ . \tag{1.9}$$

Proposition 1.7 implies that a matrix $x \in \mathrm{SL}_n(\mathbb{C})$ is totally positive if and only if x and its transpose x^T represent totally positive elements in $\mathrm{SL}_n(\mathbb{C})/U$. Flag minors satisfy generalized Plücker relations. As in the case of the totally positive Grassmanian $\mathrm{Gr}^+_{2,m}$, it is enough to check that only a subset of the flag minors are positive in order to ensure that an element $x \in \mathrm{SL}_n(\mathbb{C})/U$ is totally positive. In total, there are $2^n - 2$ flag minors but total positivity is implied by the positivity of only

$$\dim(\mathrm{SL}_n(\mathbb{C})/U) = \frac{(n-1)(n+2)}{2} \tag{1.10}$$

well chosen flag minors.

Efficient positivity tests are provided by *pseudo-line arrangements* which play the role of the triangulations of the m-gon in the case of $\mathrm{Gr}^+_{2,m}$. Two examples of pseudo-line arrangements for $n = 4$ are presented in Figure 1.3, which reproduces Fig. 2 of [Fom10]. A pseudo-line is the graph of a continuous function on $[0, 1]$ and a pseudo-line arrangement is a set of pseudo-lines such that each pair has exactly one crossing point in common, and considered up to homotopy.

Total positivity in $\mathrm{SL}_n(\mathbb{C})/U$ is encoded in pseudo-line arrangements with n pseudo-lines, that we will consider as labeled $1, \dots, n$ from bottom to top. To each region of a pseudo-line arrangement but the bottom and the top ones, one associates a flag minor $\Delta_{I(R)}$ where $I(R)$ is the set of pseudo-lines passing below the region. An example is shown in Figure 1.3.

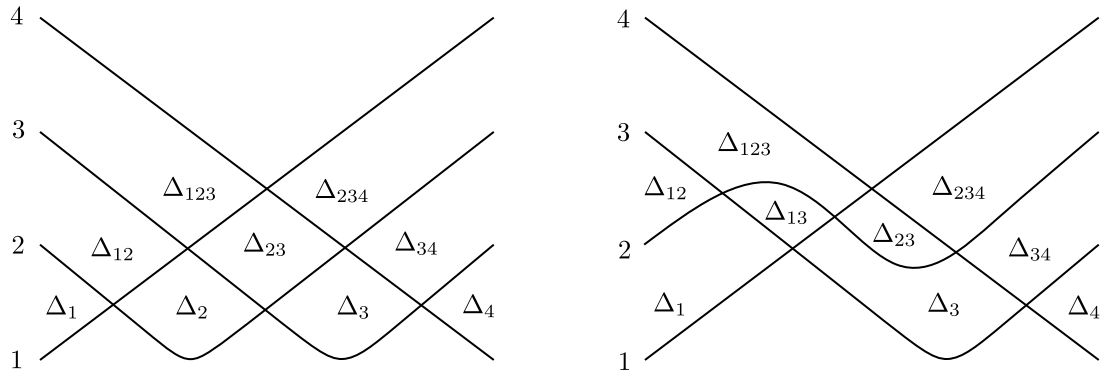


Figure 1.3: Pseudo-line arrangements.

As in the case of $\mathrm{Gr}^+_{2,m}$, the set of $\frac{(n-1)(n+2)}{2}$ flag minors appearing in a pseudo-line arrangement form an extended cluster. The flag minors corresponding to open regions of the pseudo-line arrangement are frozen variables, while the ones corresponding to regions entirely bounded by pseudo-lines are cluster variables. In Figure 1.3 the flag minors $\Delta_1, \Delta_4, \Delta_{12}, \Delta_{34}, \Delta_{123}, \Delta_{234}$ are frozen variables, while the others are cluster variables. Each cluster contains $\binom{n-1}{2}$ flag minors.

Flips of the triangulation of the m-gon that appeared while studying $\mathrm{Gr}^+_{2,m}$ also have an equivalent here, as local moves of the pseudo-line arrangements. They consist of dragging one of the pseudo-lines through an intersection of two others, as shown in Figure 1.4. One can show that for under such a local move of pseudo-line arrangements, the flag minors appearing in the clusters satisfy

$$ef = ac + bd \ . \tag{1.11}$$

These are the generalized Plücker relations. As in the case of the triangulations of m-gons, given any two pseudo-lines arrangements one can always transform the first into the second through a finite sequence of these local moves. For example, in the example of Figure 1.3 one has:

$$\Delta_{13} = \frac{\Delta_1\Delta_{23} + \Delta_{12}\Delta_3}{\Delta_2} \ , \tag{1.12}$$

which is a subtraction-free expression of the flag minors in the extended cluster corresponding to the pseudo-line arrangement on the left of Figure 1.3.

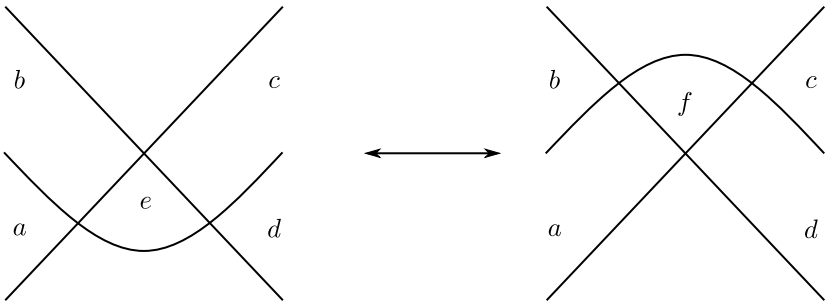


Figure 1.4: Local move in pseudo-line arrangements.

Therefore, a reasoning similar to the one in the proof of Theorem 1.4 shows that any pseudo-line arrangement provides an efficient total positivity test for the elements in $\mathrm{SL}_n(\mathbb{C})/U$. One can prove that these tests, just as the ones given by the triangulation of an m-gon in the case of $\mathrm{Gr}^+_{2,m}$ are the most efficient ones.

1.2 Mutations of quivers and matrices

Quivers generalize both the triangulations of the m-gon in the case of $\mathrm{Gr}^+_{2,m}$ and the pseudo-line arrangements in the case of total positivity in $\mathrm{SL}_n(\mathbb{C})/U$. Flips in triangulations and local moves in pseudo-lines arrangements are both special instances of quiver *mutations*.

Quivers and mutations

Definition 1.9. *A* quiver *is a finite oriented graph possibly with multiple arrows but neither loops nor oriented* 2*-cycles. Two examples are shown in Figure 1.5. We do not require such quivers to be connected. Consistently with the above, one can declare some of the vertices to be frozen. Non-frozen vertices will be referred to as mutable. Edges between frozen vertices are superfluous.*

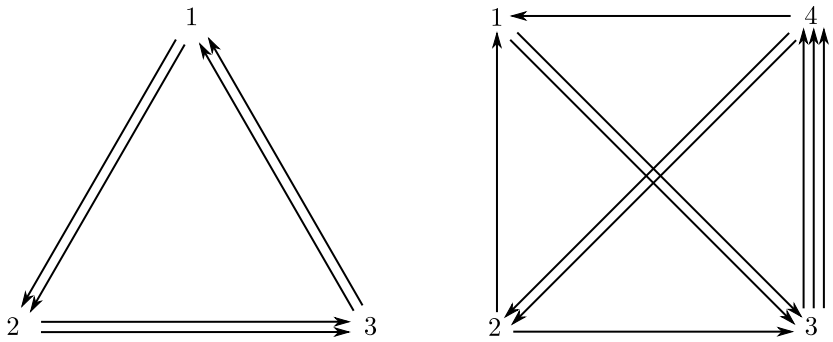


Figure 1.5: The Markov (left) and the Somos-4 (right) quivers.

Let Γ be a triangulation of an oriented m-gon P_m. In order construct a quiver $\mathcal{Q}_\Gamma$ from Γ, one assigns a vertex of $\mathcal{Q}_\Gamma$ to each chord of Γ and declare that the vertices corresponding to the chords on the boundary of Γ are frozen while the others are mutable. Then, inside each triangle of Γ one draws three arrows, linking each edge of the triangle to the next one according to the counterclockwise orientation. Last, one removes all arrows connecting frozen vertices. An example is shown on the left of Figure 1.6. Numbers label the mutable vertices, and letters, the frozen ones.

One can also assign a quiver to a pseudo-line arrangement, in such a way that the local moves correspond to the mutations that will be defined soon. The procedure is more involved than the one for triangulations, and hence we refer to [FWZ06, Section 2.3] for its precise description. Nevertheless, an example is depicted on the right of Figure 1.6.

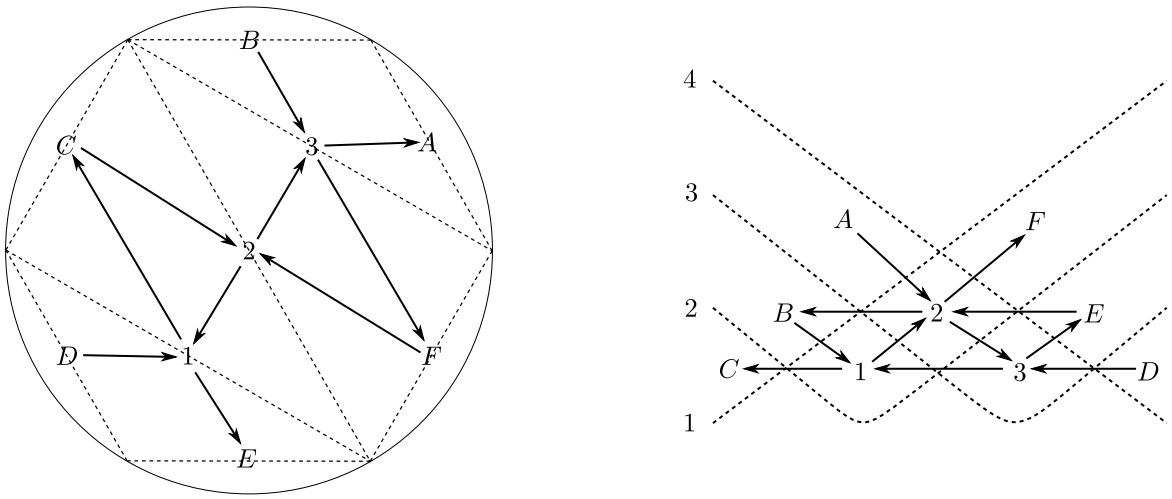


Figure 1.6: Frozen (resp. mutable) vertices are labeled by letters (resp. numbers).

Definition 1.10. *Let k be a mutable vertex of a quiver $\mathcal{Q}$. The* mutation at k *is another quiver $\mu_k(\mathcal{Q})$, obtained as follows:*

1. *First, for every path $i \xrightarrow{e} k \xrightarrow{f} j$ one adds an arrow $i \xrightarrow{[ef]} j$ (unless i and j are frozen).*
2. *Then, one reverses the directions of all arrows incident to k.*
3. *Last, one repeatedly removes oriented* 2*-cycles that appeared during the first two steps.*

One can check easily that the Markov quiver on the left of Figure 1.5 is invariant under the mutation at any of its vertices, while the mutation at the vertex 1 (resp. 2) of the Somos-4 quiver shown on the right of Figure 1.5 yields a copy of the same quiver rotated by $-\pi/2$ (resp. $+\pi/2$). However, the mutation at a vertex of a quiver does not generally yield an isomorphic quiver.

Mutations at the mutable vertices of a quiver corresponding to a triangulation of an m-gon are exactly flips, as advertised. It is also the case that mutations of a quiver obtained from a pseudo-line arrangement with the rules of [FWZ06, Section 2.3] encode local moves. The mutation corresponding to the local move at face 1 on a pseudo-line arrangement is shown in Figure 1.7.

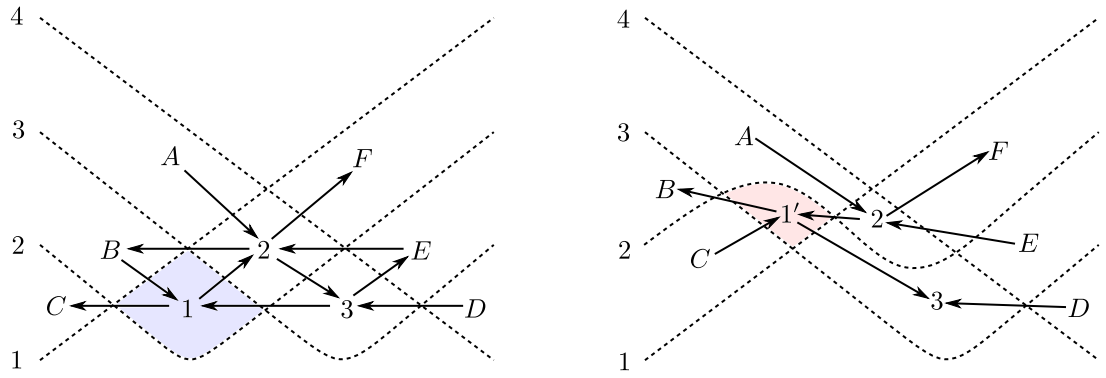


Figure 1.7: The mutation at face 1 on the quiver on the left.

These mutation rules seem to be very ad-hoc at first sight, however they amazingly appear in many different contexts in mathematics and in physics, as underlined in the introduction of this chapter.

1.2.1 The mutation class of a quiver

Let $\mathcal{Q}$ be a quiver, and let $M_{\mathcal{Q}}$ be the set of all quivers that can be obtained from $\mathcal{Q}$ by repeated mutations at mutable vertices. The relation $\mathcal{Q} \sim \mathcal{Q}'$ if $\mathcal{Q}$ and $\mathcal{Q}'$ are related through a finite sequence of mutations is an equivalence relation. The set $M_{\mathcal{Q}}$ is the equivalence class of $\mathcal{Q}$ for $\sim$. When the set $M_{\mathcal{Q}}$ has finite cardinal one says that $\mathcal{Q}$ is of *finite mutation type*, however in general $M_{\mathcal{Q}}$ is infinite. The Markov quiver of Figure 1.5 is an example of a finite mutation type quiver, however the Somos-4 quiver in the same figure is not of finite mutation type.

Quivers of finite mutation type without frozen vertices type have been completely classified in [FST12], and fall into three classes:

1. quivers with two vertices,

2. quivers associated with triangulations of punctured surfaces, as in Figure 1.6,

3. 11 "exceptional" cases denoted $E_6, E_7, E_8, \widetilde{E_6}, \widetilde{E_7}, \widetilde{E_8}, E_6^{(1,1)}, E_7^{(1,1)}, E_8^{(1,1)}, X_6$ and X_7 (see Fig. 6.1 in [FST12]) together with their mutations. The exceptional quivers X_6 and X_7 are shown in Figure 1.8.

1.2.2 Skew-symmetric and skew-symmetrizable matrices

Definition 1.11. *Let $\mathcal{Q}$ be a quiver with mutable vertices labeled $1, ..., n$, and frozen ones, $n+1, ..., m$. The extended exchange matrix associated with $\mathcal{Q}$ is the skew-symmetric $m \times n$ matrix $\tilde{B}_{\mathcal{Q}}$ such that $(\tilde{B}_{\mathcal{Q}})_{ij} = n$ if there are n arrows going from i to j in $\mathcal{Q}$, $(\tilde{B}_{\mathcal{Q}})_{ij} = -n$ if there are n arrows going from j to i in $\mathcal{Q}$, and $(\tilde{B}_{\mathcal{Q}})_{ij} = 0$ otherwise. The corresponding exchange matrix $B_{\mathcal{Q}}$ is the upper $n \times n$ block of $\tilde{B}_{\mathcal{Q}}$.*

The following three matrices are respectively the extended exchange matrices of the Markov quiver, the Somos-4 quiver in Figure 1.5, and the quiver associated with the triangulation Γ of the 6-gon shown

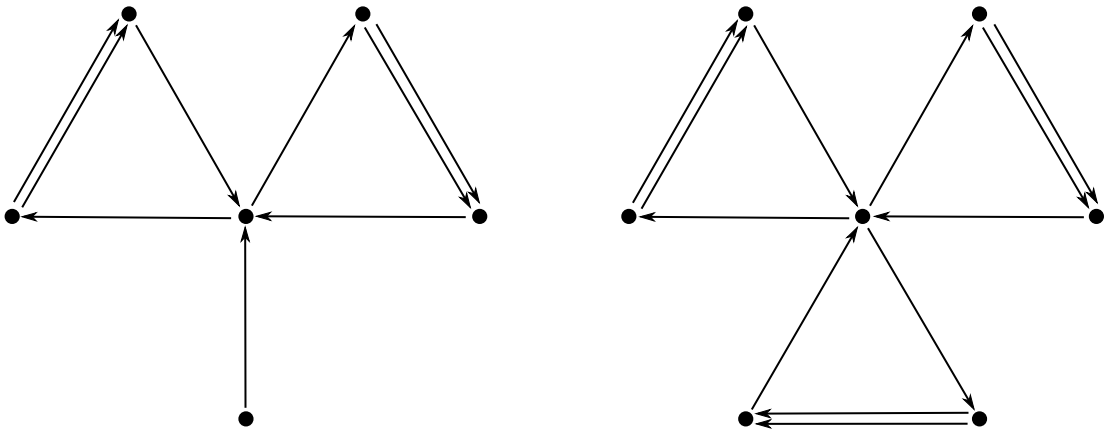


Figure 1.8: Exceptional quivers of finite mutation type X_6 (left) and X_7 (right).

in Figure 1.6.

$$\tilde{B}_{\text{Markov}} = \begin{bmatrix} 0 & 2 & -2 \\ -2 & 0 & 2 \\ 2 & -2 & 0 \end{bmatrix} \quad \tilde{B}_{\text{Somos}-4} = \begin{bmatrix} 0 & -1 & 2 & -1 \\ 1 & 0 & 1 & -2 \\ -2 & -1 & 0 & 3 \\ 1 & 2 & -3 & 0 \end{bmatrix} \quad \tilde{B}_{\Gamma} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \tag{1.13}$$

Conversely, given a matrix of size $m \times n$ with $m \geq n$ and such that the upper $n \times n$ block is skew-symmetric, one can define a quiver with that extended exchange matrix by considering a mutable vertex for each of the first n rows, a frozen vertex for each of the remaining $m - n$ rows, and with the arrows between any two vertices i and j determined by the entries of the matrix.

Proposition 1.12. *Let k be a mutable vertex of a quiver $\mathcal{Q}$ with extended exchange matrix $\tilde{B}_{\mathcal{Q}} = (b_{ij})$. The extended exchange matrix $\tilde{B}' = \tilde{B}_{\mu_k(\mathcal{Q})} = (b'_{ij})$ of the quiver $\mu_k(\mathcal{Q})$ is given by*

$$b'_{ij} = \begin{cases} -b_{ij} & \text{if } i = k \text{ or } j = k \\ b_{ij} + \text{sgn}(b_{ik})[b_{ik}b_{kj}]_+ & \text{otherwise} \end{cases}, \tag{1.14}$$

where $[b_{ik}b_{kj}]_+ = \max(b_{ik}b_{kj}, 0)$.

One can generalize this mutation formula to matrices that are non skew-symmetric but skew-symmetrizable.

Definition 1.13. *An $n \times n$ matrix B with integer entries is skew-symmetrizable if there exists $d_1, ..., d_n \in \mathbb{Z}$ called the multipliers, such that $d_i b_{ij} = -d_j b_{ji}$. An extended skew-symmetrizable matrix is an $m \times n$ matrix with $m \geq n$ such that its top $n \times n$ sub-matrix is skew-symmetrizable.*

Definition 1.14. *Let $\tilde{B} = (b_{ij})$ be an extended skew-symmetrizable $m \times n$ matrix (with $m \geq n$), and let $k \in \{1, ..., n\}$. The mutation of $\tilde{B}$ at k is the extended skew-symmetrizable matrix $\mu_k(\tilde{B})$ with entries satisfying Eq.* 1.14.

It is straightforward to prove the following facts:

Proposition 1.15. *1. $\mu_k(\tilde{B})$ is skew-symmetrizable with the same multipliers as $\tilde{B}$.*

2. $\mu_k \circ \mu_k(\tilde{B}) = \tilde{B}$.

3. If $b_{ij} = b_{ji} = 0$ then $\mu_i \circ \mu_j(\tilde{B}) = \mu_j \circ \mu_i(\tilde{B})$.

Any skew-symmetrizable matrix can be encoded as a *generalized quiver*, where to each mutable vertex i is assigned its multiplier d_i, and with $d_i b_{ij}$ arrows from i to j if $d_i b_{ij} \geq 0$, or $-d_i b_{ij} = d_j b_{ji}$ arrows from j to i otherwise. We will present an explicit example involving those generalized quivers in Section 1.5.2

1.2.3 Quivers with potential and bipartite graphs on oriented surfaces

Quivers with potential appear in string theory [DM96, DGM97] as the combinatorial objects describing the worldvolume theory of D-branes at toric Calabi-Yau singularities (see Chapter 7). More mathematically, topological B-branes and open strings between them on the orbifold $\mathbb{C}^d/G$, where G is a finite subgroup of $\mathrm{SU}(d)$ and $d \leq 3$, are described by the derived category of McKay quiver representations (with relations) [Asp04]. This follows from a result by Bridgeland, King and Reid [BKR99].

Let us sketch here some aspects of the theory of quivers with potential and their mutations, following [DWZ08], as well as how such quivers with potentials arise from bipartite graphs on oriented surfaces. As a first step, let us rewrite the definition of quivers of above somewhat more formally.

Definition 1.16. *A quiver is the data of a quadruple $\mathcal{Q} = (Q_0, Q_1, s, t)$ where Q_0 is the finite set of vertices, Q_1 is the finite set of arrows and $s, t : Q_1 \to Q_0$ are the source and target maps, respectively. As before, we assume that our quivers do not have oriented 2-loops. Let K be a field, $R = K^{Q_0}$ the vector space of K-valued functions on Q_0 and $A = K^{Q_1}$ the vector space of K-valued functions on Q_1.*

The vector A is naturally endowed with an R-bimodule structure [DWZ08, Section 2]. Let A^d be the tensor product of d copies of A with itself over R. One also sets $A^0 = R$.

The set of arrows Q_1 of $\mathcal{Q}$ is naturally identified with a basis of the R-bimodule A. For every $d \geq 1$, the elements $a_1 \dots a_d \in A^d$, such that for all $k = 1, \dots, d-1$ one has $t(a_k) = s(a_{k+1})$, form a K-basis of A^d called the path basis. Each A^d module A^d can be decomposed as:

$$A^d = \bigoplus_{i,j \in Q_0} A^d_{i,j} \ , \tag{1.15}$$

where $A^d_{i,j}$ is spanned by the paths $a_1 \dots a_d$ such that $s(a_1) = i$ and $t(a_d) = j$. Now, for each $d \geq 1$ the cyclic part A^d_{cyc} of A^d is defined as:

$$A^d_{\mathrm{cyc}} = \bigoplus_{i \in Q_0} A^d_{i,i} \ . \tag{1.16}$$

The complete path algebra of $\mathcal{Q}$ is

$$R\langle\langle A \rangle\rangle = \prod_{d=0}^{\infty} A^d \ ; \tag{1.17}$$

its elements are (possibly infinite) K-linear combinations of the elements of a path basis in $R\langle A \rangle$. Let

$$R\langle\langle A \rangle\rangle_{\mathrm{cyc}} = \prod_{d=1}^{\infty} A^d_{\mathrm{cyc}} \ . \tag{1.18}$$

Denote $A^* = \mathrm{Hom}(A, R)$ and let $\xi \in A^*$. The cyclic derivative $\partial_\xi : R\langle\langle A \rangle\rangle_{\mathrm{cyc}} \longrightarrow R\langle\langle A \rangle\rangle$ with respect to ξ is defined as:

$$\partial_\xi(a_1 \dots a_d) = \sum_{k=1}^{d} \xi(a_k) a_{k+1} \dots a_d a_1 \dots a_{k-1} \ . \tag{1.19}$$

Definition 1.17 (3.1 of [DWZ08])**.** *A potential for $\mathcal{Q}$ is an element of $R\langle\langle A \rangle\rangle_{\mathrm{cyc}}$. The Jacobian ideal corresponding to a potential S is the closure in $R\langle\langle A \rangle\rangle$ of the two-sided ideal generated by the elements $\partial_\xi(S)$ for all $\xi \in A^*$; it is also a two-sided ideal. The Jacobian algebra corresponding to S is*

$$R\langle\langle A \rangle\rangle / J(S) \ . \tag{1.20}$$

In physics terminology, a potential is rather called a *superpotential*, the Jacobian ideal is the space of F-term equations and the Jacobian algebra is the space of solutions to the F-term equations, also called *master space* of the quiver gauge theory – we refer to Chapter 5 for more details.

Let now $\mathcal{B} = (V_W \amalg V_B, E)$ be a finite bipartite graph embedded in an oriented surface S, where V_W (resp. V_B) is the set of white (resp. black) vertices and where $E \subset V_W \times V_B$. The faces of $\mathcal{B}$ as the connected components of $S - \mathcal{B}$, and one assumes that they all have the topology of a disk. The embedding of $\mathcal{B}$ in S is equivalent to a fat structure on $\mathcal{B}$, which is the data, at each vertex, of a cyclic orientation of the edges incident to it. Let us assume that $\mathcal{B}$ has no two-valent vertex. One assigns a quiver $\mathcal{Q}$ with potential S to any such $\mathcal{B}$ as follows:

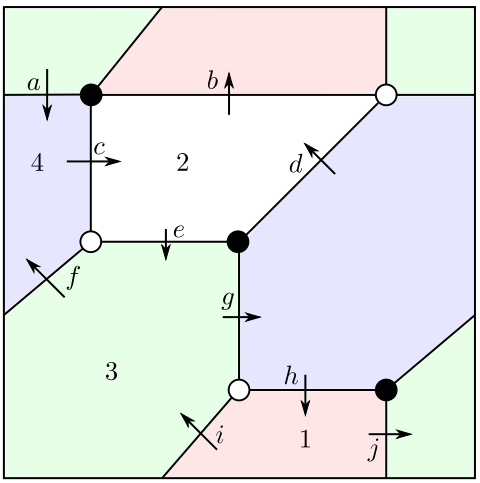
1. The vertices of $\mathcal{Q}$ are in one-to-one correspondence with the faces of $\mathcal{B}$.

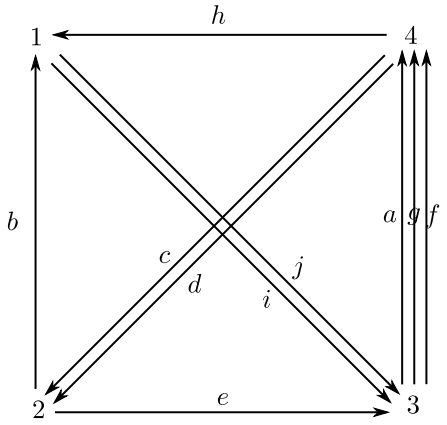
2. The arrows of $\mathcal{Q}$ connect faces of $\mathcal{B}$ to adjacent ones in such a way that each arrow crosses an edge of $\mathcal{B}$ with the black end of the edge to its left.

3. Each white (resp. black) vertex in $\mathcal{B}$ defines an element of $R\langle\langle A\rangle\rangle$ for the quiver we defined, by following the arrows of $\mathcal{Q}$ crossing the edges incident to the vertex clockwise (resp. counterclockwise).

An example is shown in Figure 1.9. The corresponding quiver (displayed on the right) encodes the gauge theory on the worldvolume of D3-branes at the tip of the affine cone over the toric del Pezzo surface dP_1[2]; again we refer to Chapter 7 for more details about this. The potential is shown below the quiver. The corresponding Jacobian ideal is generated by:

$$J(S) = \langle dbj - cbi, jad - iac, ef - bia, bja - eg, fc - gd, ce - hj, hi - de, ig - jf, gh - acb, adb - fh\rangle \ . \quad (1.21)$$





$$S = efc + hig + dbja - acbi - egd - hjf$$

Figure 1.9: A quiver with potential from a bipartite graph on the torus T^2.

Quivers with potentials can be mutated: the mutation of a quiver with potential $(\mathcal{Q}, S)$ at a (mutable) vertex k is another quiver with potential $(\mu_k(\mathcal{Q}), \mu_k(S))$, where $\mu_k(\mathcal{Q})$ is the quiver obtained from $\mathcal{Q}$ by mutating at k with the same rules as for quivers without potential. Assume that vertices of $\mathcal{Q}$ are denoted $i, j, k, \ldots$ while $e, f, \ldots$ stand for arrows. Then:

1. For every path $i \xrightarrow{e} k \xrightarrow{f} j$ one adds an arrow $i \xrightarrow{[ef]} j$.

2. Then one inverts all arrows incident to k: an arrow $i \xrightarrow{e} k$ is replaced by $i \xleftarrow{e'} k$ and an arrow $k \xrightarrow{f} j$ is replaced by $k \xleftarrow{f'} j$.

3. Remove all 2-cycles created by the two first steps.

The mutated potential $\mu_k(S)$ is $[S] + \Delta_k$, where $[S]$ is obtained from S by substituting $[ef]$ for each factor ef with $t(e) = k = s(f)$ of any cyclic path occurring in the expression of S, and where

$$\Delta_k = \sum_{e,f \in Q_1 | t(e)=k=s(f)} [ef]f'e' \ . \quad (1.22)$$

If the mutated potential contains a cyclic path of length two, it has to be removed, accordingly to the fact that the 2-cycles suffer the same fate at step 3 above.

There exists an elementary transformation of bipartite fat graphs called *spider move* and depicted in Figure 1.10, which corresponds to some of the mutations of quivers with potential. Spider moves can

[2]it is the blow-up of $P^2(\mathbb{C})$ at a point

only be applied to any square face of the bipartite graph, for otherwise the mutated quiver with potential cannot be associated with a bipartite graph on an oriented surface anymore. However, a spider move at a square face encodes exactly the mutation of the equivalent quiver with potential, at the corresponding vertex. Similar spider moves where black and white vertices are exchanged are of course also allowed.

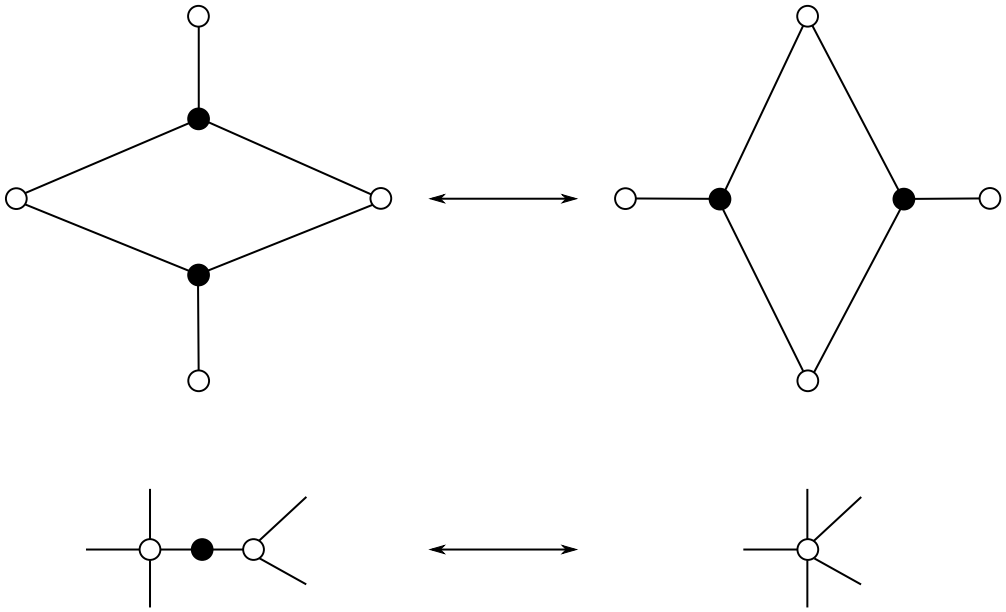
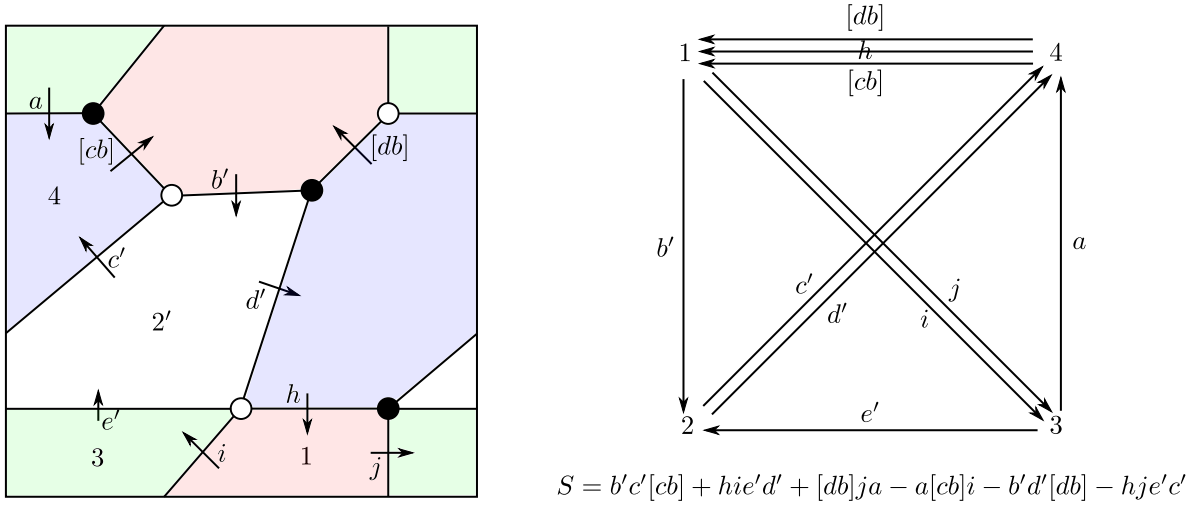


Figure 1.10: A spider move (top). Adding and removing two-valent nodes (bottom).

In order to make a spider move at some face of a bipartite fat graph, one might first need to split vertices of valency four or greater into more vertices of lower valency. In addition, spider move at square faces might create 2-valent vertices, in which case one contracts them afterwards so that their neighbor vertices merge. This splitting-contraction operation is shown at the bottom of Figure 1.10 for black 2 valent vertices; the same operation can also be done for 2-valent vertices.

Let us consider an explicit example of these spider moves. In the bipartite graph in Figure 1.9 one can split the top black node of the face 2 into two 3-valent black vertices connected by a 2-valent white vertex, do a spider move, and then contract the white 2-valent vertex that has appeared. The result is shown in Figure 1.11, and the corresponding quiver with potential is shown on the right. One can check easily that the latter is the result of mutating the quiver with potential in Figure 1.9 at the node 2.



$$S = b'c'[cb] + hie'd' + [db]ja - a[cb]i - b'd'[db] - hje'c'$$

Figure 1.11: The bipartite graph, quiver and superpotential after the mutation at 2 in Figure 1.9.

1.3 Cluster algebras

In this section, we mostly follow [Mar14]. We restrict our presentation to complex cluster algebras of geometric type (see [FZ02] for a precise definition). Since we have introduced a lot of terminology in the previous subsections, let us refine what is written in the introduction: a cluster algebra A of rank $n \in \mathbb{Z}_{>0}$ over $\mathbb{C}$ is an integral domain with distinguished subsets of size n called clusters, the union of which generates A. It satisfies

$$\mathbb{C}[a_1, \dots, a_n] \subset A \subset \mathbb{C}(a_1, \dots, a_n) \ , \tag{1.23}$$

as a ring, and A is generated by rational fractions in $a_1, \dots, a_n$ determined by some combinatorial data encoded in a quiver. Let $m \geq n$ be another integer and let $\mathcal{F} = \mathbb{C}(a_1, ..., a_m)$ be the field of rational fractions in m variables with complex coefficients.

Definition 1.18. *A seed in $\mathcal{F}$ is a pair $(\tilde{a}, \tilde{B})$, where $\tilde{a} = \{a_1, ..., a_m\}$ is a free generating set of $\mathcal{F}$ and where $\tilde{B}$ is an extended skew-symmetrizable matrix of size $m \times n$. Each a_i for $i = 1, ..., n$ is a cluster variable and the set $\{a_1, \dots, a_n\}$ is the cluster of the seed. The frozen variables are $a_{n+1}, \dots, a_m$. The set $\tilde{a}$ is the extended cluster of the seed.*

Definition 1.19. *Let $(\tilde{a}, \tilde{B})$ be a seed, and let $k \in \{1, ..., n\}$. The mutation of $\left(\tilde{a}, \tilde{B}\right)$ at k is the seed $\mu_k\left(\tilde{a}, \tilde{B}\right) = \left(\tilde{a}', \tilde{B}'\right)$ where $\tilde{B}' = \mu_k\left(\tilde{B}\right)$ as defined in the previous section, and where $\tilde{a}' = \{a'_1, ..., a'_m\}$ is such that:*

$$a'_i = \begin{cases} a_i & \text{if } i \neq k \\ a_k^{-1}\left(\prod\limits_{b_{ik}>0} a_i^{b_{ik}} + \prod\limits_{b_{ik}<0} a_i^{-b_{ik}}\right) & \text{if } i = k \end{cases} . \tag{1.24}$$

Let T_n be the n-valent regular tree with edges labeled $1, ..., n$, such that for each vertex the n edges incident to it all have distinct labels.

Definition 1.20. *A seed pattern on T_n is an assignment $\left(\tilde{a}(t), \tilde{B}(t)\right)_{t \in \mathsf{T}_n}$ of seeds to the vertices of T_n, such that two seeds at neighbor vertices linked by an edge labeled k are obtained one from the other by mutation at k.*

Let χ be the set of all cluster variables appearing in a seed pattern $\left(\tilde{a}(t), \tilde{B}(t)\right)_{t \in \mathsf{T}_m}$, and let

$$R = \mathbb{C}[a_{n+1}, ..., a_m] \tag{1.25}$$

be the polynomial ring in the frozen variables $a_{n+1}, ..., a_m$.

Definition 1.21. *The cluster algebra corresponding to the seed pattern $\left(\tilde{a}(t), \tilde{B}(t)\right)_{t \in \mathsf{T}_m}$ is the R-subalgebra of $\mathcal{F}$ generated by the cluster variables, i.e. $A = R[\chi]$.*

The mutation formula of Equation (1.24) is subtraction-free; hence it also makes sense if the indeterminates are evaluated in any multiplicative semifield in place of the field of complex numbers.

Definition 1.22. *A (commutative) semi-field $(\mathbb{S}, \oplus, \otimes)$ is a set $\mathbb{S}$ endowed with two inner laws $\oplus$ and $\otimes$ such that $(\mathbb{S}, \otimes)$ is an abelian group, and such that $\oplus$ is associative, commutative, and $\otimes$ is distributive with respect to $\oplus$. By a slight abuse of notation, the set $(\mathbb{S} \cup \{0\})$, where 0 is the additive unit and a zero element for $\otimes$, is still called semifield when endowed with the inner laws $\oplus$, $\otimes$.*

Fields are clearly semifields, and hence $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are semifields. Examples of semifields which are not fields are $(\mathbb{Q}_{>0}, +, \times)$ and $(\mathbb{R}_{>0}, +, \times)$, to which one can add the zero unit, yielding $(\mathbb{Q}_{\geq 0}, +, \times)$ and $(\mathbb{R}_{\geq 0}, +, \times)$. More exotic examples are the so-called *tropical semifields*. Let A be $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{R}$; the tropical semifield A^t is defined as $A^t = (A, \max, +)$. The additive unit in that case is $\{-\infty\}$, and we will sometimes consider $A^t = (A \cup \{-\infty\}, \max, +)$. Tropical semifields arise naturally in the context of valued fields, in which they encode how valuations behave under addition and multiplication in the field. For example, tropical semifields appear naturally in the study of functional asymptotics, under the name of max-plus algebras.

The tropical semifield $\mathbb{R}^t$ is the limit of a family of isomorphic semifield structures on $\mathbb{R}$: let $\hbar \geq 0$ and let $R_\hbar = (\mathbb{R}, +_\hbar, \times_\hbar)$ where for $a, b \in \mathbb{R}$:

$$a +_\hbar b = \hbar \ln \left(\exp\left(\frac{a}{\hbar}\right) + \exp\left(\frac{b}{\hbar}\right) \right) \ , \tag{1.26}$$

$$a \times_\hbar b = \hbar \ln \left(\exp\left(\frac{a}{\hbar}\right) \times \exp\left(\frac{b}{\hbar}\right) \right) = a + b \ . \tag{1.27}$$

When $\hbar > 0$, $\mathbb{R}_\hbar$ is isomorphic to $(\mathbb{R}_{>0}, +, \times)$ whereas $\mathbb{R}_0 = \mathbb{R}^t$. This is called *Maslov dequantization*.

We will elaborate on this when discussing cluster varieties, in Chapter 4.

1.3.1 Rank-1 cluster algebras

The tree T_1 has exactly two vertices, and thus there are only two seeds and two clusters $\{a_1\}$ and $\{a_1'\}$. The extended exchange matrix $\tilde{B}$ is a column with m entries, and with the first entry necessarily 0 by antisymmetry. The relation of Equation (1.24) in this case reads

$$a_1 a_1' = M_1 + M_2 \ , \tag{1.28}$$

where M_1 and M_2 are monomials in the frozen variables $a_2, \ldots, a_m$ whose precise expressions depends on $\tilde{B}$. The corresponding rank-1 cluster algebra $\mathcal{A}$ is the R-subalgebra of $\mathbb{F}$ generated by a_1 and a_1' subject to Equation (1.28):

$$\mathbb{C}[a_1, a_2, \ldots, a_m] \subset \mathcal{A} = \mathbb{C}\left[a_1, \frac{M_1 + M_2}{a_1}, a_2, \ldots, a_m\right] \subset \mathbb{C}(a_1, \ldots, a_m) \ . \tag{1.29}$$

Consider for example the quiver on the left of Figure 1.12, where the only mutable vertex is the node 1 and the frozen vertices are the boxes 2, 3 and 4.

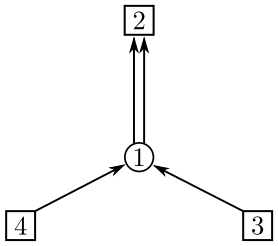


Figure 1.12: A quiver defining a rank-1 cluster algebra.

The mutation at 1 yields the opposite quiver (where all arrows are reversed). The mutation rule in that case is:

$$a_1' = \frac{a_3 a_4 + a_2^2}{a_1} \ , \tag{1.30}$$

and mutating again at $1'$ yields:

$$a_1'' = \frac{a_2^2 + a_3 a_4}{a_1'} = a_1 \ , \tag{1.31}$$

This is a general fact: the mutation rules imply straightforwardly that if k is a mutable vertex in a quiver and if k' is the corresponding vertex in the mutated quiver, then $\mu_{k'} \circ \mu_k$ is the identity transformation on the cluster. Hence in that case the cluster algebra corresponding to this quiver is

$$\mathcal{A} = \mathbb{C}\left[a_1, \frac{a_3 a_4 + a_2^2}{a_1}, a_2, a_3, a_4\right] \ . \tag{1.32}$$

1.3.2 Rank-2 cluster algebras without frozen variables

We start by studying rank-2 cluster algebras without frozen variables. Any 2×2 skew-symmetrizable matrix with integer entries writes:

$$\pm B(b,c) = \pm \begin{bmatrix} 0 & b \\ -c & 0 \end{bmatrix} , \tag{1.33}$$

where $b, c \in \mathbb{Z}$ are either strictly positive or zero. The mutation of such matrices at 1 or 2 is merely multiplication by -1.

Let $\mathcal{A}(b,c)$ be the cluster algebra corresponding to the seed pattern defined by the matrix of Equation (1.33). The case $b = c = 0$ is not particularly interesting, and hence we assume that $b, c > 0$. Let us start with a seed $((a_1, a_2), B(b,c))$. A mutation at 1 yields the seed $((a_1' =: a_3, a_2), -B(b,c))$, where

$$a_3 = \frac{a_2^c + 1}{a_1} . \tag{1.34}$$

Mutating at 2 this new seed yields $((a_3, a_2' := x_4), B(b,c))$, where

$$a_4 = \frac{a_3^b + 1}{a_2} . \tag{1.35}$$

Doing this over and over again in both directions (one can also mutate the original seed at 2 to obtain a_0, and so on), one obtains a sequence $(..., a_0, a_1, a_2, a_3, a_4, ...)$ where for all $i \in \mathbb{Z}$, the cluster variable a_i is a rational fraction in a_1 and a_2. Part of the seed pattern is shown in Figure 1.13.

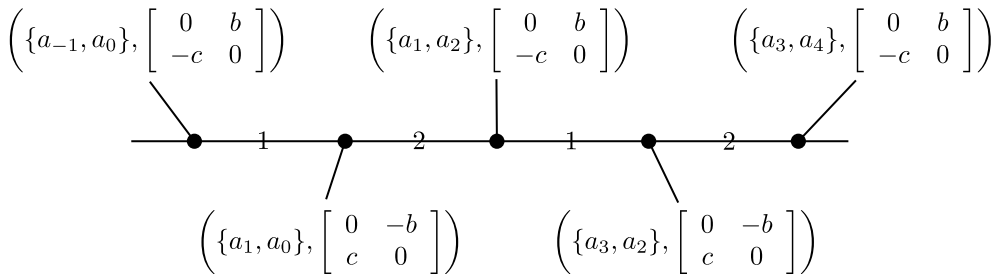


Figure 1.13: The general form of the seed pattern for a rank-2 cluster algebra.

The exchange relation is:

$$a_{k-1}a_{k+1} = \begin{cases} a_k^c + 1, & (k \text{ even}) \\ a_k^b + 1, & (k \text{ odd}) \end{cases} . \tag{1.36}$$

1. The cluster algebra $\mathcal{A}(1,1)$ is said to be of type A_2. The exchange relation yields the following 5-periodic sequence of cluster variables:

$$..., a_1, a_2, \frac{a_2+1}{a_1}, \frac{a_1+a_2+1}{a_1a_2}, \frac{a_1+1}{a_2}, a_1, a_2, ... \tag{1.37}$$

   Hence:

$$\mathbb{C}[a_1, a_2] \subset \mathcal{A}(1,1) = \mathbb{C}\left[a_1, a_2, \frac{a_2+1}{a_1}, \frac{a_1+a_2+1}{a_1a_2}, \frac{a_1+1}{a_2}\right] \subset \mathbb{C}(a_1, a_2) . \tag{1.38}$$

2. The cluster algebra $\mathcal{A}(1,2)$ is said to be of type B_2. The exchange relation yields:

$$\mathcal{A}(1,2) = \mathbb{C}\left[a_1, a_2, \frac{a_2^2+1}{a_1}, \frac{1+a_1+a_2^2}{a_1a_2}, \frac{2a_1+a_1^2+a_2^2+1}{a_1a_2^2}, \frac{a_1+1}{a_2}\right] . \tag{1.39}$$

3. The cluster algebra $\mathcal{A}(1,3)$ is said to be of type G_2. The exchange relation yields:

$$\begin{aligned} \mathcal{A}(1,3) = \mathbb{C}\Bigg[a_1, a_2, \frac{1+a_2^3}{a_1}, \frac{1+a_1+a_2^3}{a_1a_2}, \frac{1+2a_2^3+a_2^6+3a_1+3a_1a_2^3+3a_1^2+a_1^3}{a_1^2a_2^3}, \\ \frac{1+2a_1+a_1^2+a_2^3}{a_1a_2^2}, \frac{1+3a_1+3a_1^2+a_1^3+a_2^3}{a_1a_2^3}, \frac{1+a_1}{a_2}\Bigg] . \end{aligned} \tag{1.40}$$

4. When $b = 1$ and $c = 4$, applying repeatedly the exchange relation of Equation (1.36) gives rise to a non-periodic sequence of cluster variables, hence the corresponding cluster algebra $\mathcal{A}(1, 4)$ is a polynomial algebra over countably many rational fractions in a_1 and a_2. In general, if the sequence of cluster variables is periodic, one says that the corresponding cluster algebra is of finite type. For example, $A(1, 1)$, $\mathcal{A}(1, 2)$ and $\mathcal{A}(1, 3)$ are of finite type, but not $\mathcal{A}(1, 4)$. Cluster algebras of finite type have been classified in [FZ03], and are in one-to-one correspondence with the Dynkin diagrams of finite dimensional simple Lie algebras.

1.3.3 A rank-2 cluster algebra with frozen variables

Now let us consider a rank-2 cluster algebra with frozen variables. The quiver, the sequence of mutations generating all cluster variables as well as the clusters generated via these mutations are shown in Figure 1.14.

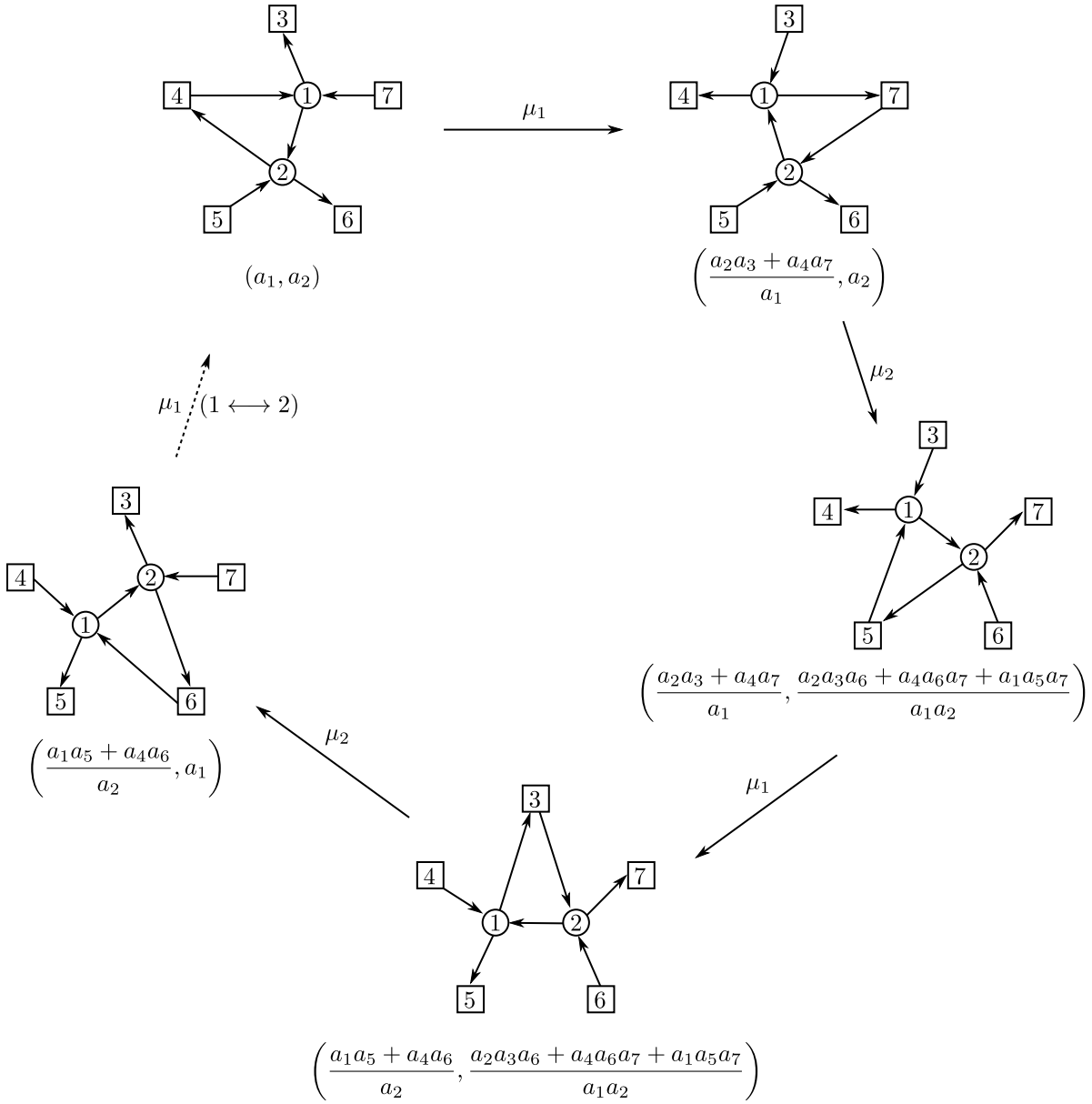


Figure 1.14: A quiver with mutable and frozen vertices generating a cluster algebra.

As before, mutable vertices are represented as circles and labeled 1 and 2, while frozen vertices are represented as squares and labeled 3, 4, 5, 6, 7. The cluster and frozen variables corresponding to the

upper-left quiver in Figure 1.14 are denoted $a_1, a_2, a_3, a_4, a_5, a_6, a_7$. Note that after doing the sequence of mutations $\mu_1 \circ \mu_2 \circ \mu_1 \circ \mu_2 \circ \mu_1$ on the upper-left seed one obtains a seed which is exactly the one we started with, up to the exchange of the labels on the nodes 1 and 2. This is the simplest example of the pentagon relation, which holds in general: given two vertices i and j of a quiver connected by a simple arrow, the sequence of mutations $\mu_i \circ \mu_j \circ \mu_i \circ \mu_j \circ \mu_i$ is an automorphism of the seed.

The cluster algebra corresponding to the upper-left seed in Figure 1.14 is:

$$\mathcal{A} = \mathbb{C}\left[a_1, a_2, a_3, a_4, a_5, a_6, a_7, \frac{a_2a_3 + a_4a_7}{a_1}, \frac{a_2a_3a_6 + a_4a_6a_7 + a_1a_5a_7}{a_1a_2}, \frac{a_4a_6 + a_1a_5}{a_2}\right] . \tag{1.41}$$

Comparing this cluster algebra to $\mathcal{A}(1,1)$ one sees that the variables assigned to frozen vertices appear as coefficients. One can specialize the values of $a_3, \ldots, a_7$ to 1, in which case one is exactly left with $\mathcal{A}(1,1)$.

Each quiver appearing in Figure 1.14 is the quiver corresponding to a triangulation of some pentagon, hence it is natural to expect some relationship between the cluster algebra of Equation (1.41) and the Grassmannian $\mathrm{Gr}_{2,5}(\mathbb{C})$. Indeed, this cluster algebra is the ring $\mathcal{O}(\mathrm{Gr}_{2,5}(\mathbb{C}))$ of regular functions over $\mathrm{Gr}_{2,5}(\mathbb{C})$, as explained in [Mar14]. One can identify the cluster and frozen variables with Plücker variables in the following way: $a = P_{12}$, $b = P_{23}$, $c = P_{34}$, $d = P_{45}$, $e = P_{15}$, $x = P_{25}$, $y = P_{35}$. One can then read directly on Figure 1.14 all the relations between Plücker coordinates. For example:

$$P_{14} = \frac{P_{12}P_{45}P_{35} + P_{23}P_{45}P_{15} + P_{34}P_{15}P_{25}}{P_{25}P_{25}} . \tag{1.42}$$

We could consider any seed as the initial one - for example, the one corresponding to the extended cluster:

$$\left(\frac{a_2a_3 + a_4a_7}{a_1}, \frac{a_2a_3a_6 + a_4a_6a_7 + a_1a_5a_7}{a_1a_2}, a_3, a_4, a_5, a_6, a_7\right) . \tag{1.43}$$

Let us rename these indeterminates $(a'_1, a'_2, a_3, a_4, a_5, a_6, a_7)$. Then the cluster algebra in Equation (1.41) can be equivalently be described as:

$$\mathcal{A} = \mathbb{C}\left[a'_1, a'_2, a_3, a_4, a_5, a_6, a_7, \frac{a'_2a_4 + a_3a_5}{a'_1}, \frac{a'_2a_4a_7 + a_3a_5a_7 + a'_1a_3a_6}{a'_1a'_2}, \frac{a_5a_7 + a'_1a_6}{a'_2}\right] . \tag{1.44}$$

The totally positive Grassmannian $\mathrm{Gr}^+_{2,5}$ can be described as the points in $\mathrm{Gr}_{2,5}(\mathbb{C})$ such that $a_1, a_2, a_3, \ldots, a_7$ are all either positive or negative. Equivalently, it is also the elements in $\mathrm{Gr}_{2,5}(\mathbb{C})$ such that $a'_1, a'_2, a_3, \ldots, a_7$ are either all positive or all negative. The structure of the cluster algebra encodes conveniently the positive part of $\mathrm{Gr}_{2,5}$: a point is in $\mathrm{Gr}^+_{2,5}$ if and only if the elements in one (equivalently, any) extended cluster of $\mathcal{O}(\mathrm{Gr}_{2,5}(\mathbb{C}))$ are all positive at it (or all negative). Each extended cluster can be understood as defining an $(\mathbb{R}_{>0})^7$-chart on $\mathrm{Gr}_{2,5}(\mathbb{C})$, with the elements of the extended cluster as coordinate functions. Transition functions between *cluster charts* are obtained as sequences of mutations. Hence, a point in the Grassmanian is in $\mathrm{Gr}^+_{2,5}$ if and only if its coordinates in one (equivalently, any) cluster chart are all positive or all negative.

One can define the totally non-negative Grassmanian $(\mathrm{Gr}_{2,5})_{\geq 0}$ as the subset of $\mathrm{Gr}_{2,5}(\mathbb{C})$ consisting of those elements such that their Plücker coordinates are all either in $\mathbb{R}_{\geq 0}$ or in $\mathbb{R}_{\leq 0}$. Again, extended clusters form $(\mathbb{R}_{\geq 0})^7$-charts on $(\mathrm{Gr}_{2,5})_{\geq 0}$. However, some of the points in $(\mathrm{Gr}_{2,5})_{\geq 0}$ which are not totally positive do not belong to all charts: for example, the point defined by the coordinates $(a_1, a_2, a_3, a_4, a_5, a_6, a_7) = (0, 3, 1, 2, 1, 2, 1)$ in a first cluster chart corresponds to the points with coordinates

$$\left(\frac{a_1a_5 + a_4a_6}{a_2}, a_1, a_3, a_4, a_5, a_6, a_7\right) = (4/3, 0, 1, 2, 1, 2, 1) \tag{1.45}$$

in this second chart, but it does not correspond to any point in the third chart parameterized by

$$\left(\frac{a_2a_3 + a_4a_7}{a_1}, a_2, a_3, a_4, a_5, a_6, a_7\right) , \tag{1.46}$$

i.e. it is 'at infinity'. Formally, the totally non-negative Grassmannian can be described by charts to $(\mathbb{R}_{\geq 0})^7$ with transition functions which are birational isomorphisms. We will make these ideas precise in Chapter 4.

Looking at all the examples we have seen so far, one fact is striking: the cluster variables obtained through repeated mutations on an initial seed all seem to belong to $\mathbb{C}[a_1^{\pm 1}, \ldots, a_n^{\pm 1}, a_{n+1}, \ldots, a_m]$, where $(a_1, \ldots, a_n, \ldots, a_m)$ is the initial extended cluster. This is surprising: looking at the mutation formula of Equation (1.24), one merely expects the cluster variables to belong to $\mathbb{C}(a_1, \ldots, a_n, \ldots, a_m)$.

1.4 The Laurent phenomenon

Theorem 1.23. *Let $\mathcal{A}$ be a cluster algebra of rank $n > 0$, let T_n be the n-regular tree underlying the seed pattern of $\mathcal{A}$ and let $t_0 \in T_n$ with $(\tilde{a}_0, \tilde{B}_0)$ the extended cluster at t_0, where $\tilde{a}_0 = (a_1, ..., a_m)$ and $\tilde{B}_0 = (b_{ij})_{m\times n}$, as before. If $t \in T_n$ is any other vertex of the seed pattern of $\mathcal{A}$ and if a is a cluster variable at t, then:*

$$a \in \mathbb{C}[a_1^{\pm 1}, \ldots, a_n^{\pm 1}, a_{n+1}, \ldots, a_m] \tag{1.47}$$

In particular, frozen variables do not appear in the denominators of the Laurent polynomials.

Partial proof. It is clear from the mutation formula of Equation (1.24) that

$$a \in \mathbb{C}(a_1, \ldots, a_n, a_{n+1}, \ldots, a_m) \ . \tag{1.48}$$

Let us consider the unique path in T_n which connects t_0 to t and let $d \in \mathbb{Z}_{>0}$ be its length. Let $j, k \in [|1, n|]$ be the labels assigned to the first two edges of this path as one goes from t_0 to t:

$$t_0 \overset{j}{\text{———}} t_1 \overset{k}{\text{———}} t_2 \text{———} \cdots \text{———} t \ . \tag{1.49}$$

By definition of the seed pattern, one has $j \neq k$. The proof goes by induction on d.

If $d = 1$ and $d = 2$, the result of the theorem follows directly from the exchange formula of Equation (1.24). Let us now assume that $d \geq 3$. The proof of [FZ02, FWZ06] distinguishes the cases $b_{jk} = b_{kj} = 0$ and $b_{jk}b_{kj} < 0$. In the latter case the proof is more subtle and longer than the in the former, even if the strategy is similar. For conciseness, we will only discuss the case $b_{jk} = b_{kj} = 0$ for which the strategy is already very instructive. The other case is lengthier but the strategy is similar. For the complete proof of the theorem we refer to [FWZ06].

Let us assume that the result of the theorem holds for paths of length $d-1$, and let t_3 be the vertex of T_n connected to t_0 by a k-edge:

$$t_3 \overset{k}{\text{———}} t_0 \overset{j}{\text{———}} t_1 \overset{k}{\text{———}} t_2 \ . \tag{1.50}$$

Since $b_{jk}b_{kj} = 0$ one has $\mu_j \circ \mu_k(\tilde{a}_0, \tilde{B}_0) = \mu_k \circ \mu_j(\tilde{a}_0, \tilde{B}_0)$. This follows from Proposition 1.15 for the skew-symmetrizable matrix part, and from the exchange formula for the extended cluster part. Hence $\mu_k(\tilde{a}_0, \tilde{B}_0) = \mu_j \circ \mu_k \circ \mu_j(\tilde{a}_0, \tilde{B}_0)$, and the seed at t_3 is the same as the seed at the vertex of T_n connected to t_2 by an edge labeled j. Both seeds at t_1 and t_3 are at a distance $d-1$ away from a seed containing a and hence the induction hypothesis implies that:

$$a \in \mathbb{C}\left[a_1^{\pm 1}, \ldots, a_{j-1}^{\pm 1}, (a_j')^{\pm 1}, a_{j+1}^{\pm 1}, \ldots, a_n^{\pm 1}, a_{n+1}, \ldots, a_m\right] \ , \tag{1.51}$$

where

$$a_j' = \frac{M_1 + M_2}{a_j} \tag{1.52}$$

with $M_1, M_2 \in \mathbb{C}[a_1, ..., a_m]$, and also as an element

$$a \in \mathbb{C}\left[a_1^{\pm 1}, \ldots, a_{k-1}^{\pm 1}, (a_k')^{\pm 1}, a_{k+1}^{\pm 1}, \ldots, a_n^{\pm 1}, a_{n+1}, \ldots, a_m\right] \ , \tag{1.53}$$

where

$$a_k' = \frac{M_3 + M_4}{a_k} \tag{1.54}$$

with $M_3, M_4 \in \mathbb{C}[a_1, ..., a_m]$. Hence we can write

$$a = \frac{P(a_1, \ldots, a_m)}{Q(a_1, \ldots, a_n)(M_1 + M_2)^u} = \frac{R(a_1, \ldots, a_m)}{S(a_1, \ldots, a_n)(M_3 + M_4)^v} \ , \tag{1.55}$$

where $P, R \in \mathbb{C}[a_1, \ldots, a_m]$; $Q, S \in \mathbb{C}[a_1, \ldots, a_n]$ are monomials and $u, v \in \mathbb{Z}_{\geq 0}$. Note that if $(M_1 + M_2)$ and $(M_3 + M_4)$ were coprime then one would necessarily have $u = v = 0$, but this does not hold in general.

Let us now add two additional frozen variables a_{m+1} and a_{m+2} at t_0, in such a way that $b_{j,m+1} = 1$ and $b_{i,m+1} = 0$ for $i \neq j$, as well as $b_{k,m+2} = 1$ and $b_{i,m+2} = 0$ for $i \neq k$. Modify the seed pattern and the cluster algebra correspondingly. What is above is left unchanged upon the addition of a_{m+1} and a_{m+2}, and hence:

$$a = \frac{P'(a_1, \ldots, a_m, a_{m+1}, a_{m+2})}{Q'(a_1, \ldots, a_m)(M_1 + M_2)^{u'}} = \frac{R'(a_1, \ldots, a_m, a_{m+1}, a_{m+2})}{S'(a_1, \ldots, a_m)(M_3 + M_4)^{v'}} \ , \tag{1.56}$$

where $P', R' \in \mathbb{C}[a_1, \ldots, a_m, a_{m+1}, a_{m+2}]$; $Q', S' \in \mathbb{C}[a_1, \ldots, a_n]$ are monomials, and $u', v' \in \mathbb{Z}_{\geq 0}$.

From the mutation formula of Equation (1.24), one knows that $M_1 + M_2 \in \mathbb{C}[a_1, ..., a_m]$ is a binomial of degree 1 in a_{m+1} (resp. degree 0 in a_{m+2}), while $M_3 + M_4 \in \mathbb{C}[a_1, ..., a_m]$ is a binomial of degree 0 in a_{m+1} (resp. degree 1 in a_{m+2}). Both are thus irreducible and coprime; subsequently, one has $u' = v' = 1$. Specializing a_{m+1} and a_{m+2} to 1 now implies that:

$$a \in \mathbb{C}\left[a_1^{\pm 1}, \ldots, a_n^{\pm 1}, a_{n+1}, \ldots, a_m\right] \ . \tag{1.57}$$

This concludes the induction step of the proof, and thus proves the theorem in the case $b_{jk}b_{kj} = 0$. □

1.5 Three amusing consequences of the Laurent phenomenon

In this section we mostly follow [Mar14].

1.5.1 The Somos-4 sequence

We have already emphasized that the Somos-4 quiver (shown in Figure 1.5) has the special property that the mutation at 1 yields the same quiver rotated by $-\pi/2$. Let us start with a seed $((a_1, a_2, a_3, a_4), B)$, where B corresponds to this quiver, and where the cluster variables a_1, a_2, a_3, a_4 correspond to the vertices with labels as in Figure 1.5. Mutation in the direction 1 yields the seed $((a_5 := a_1', a_2, a_3, a_4), B')$, where:

$$a_5 a_1 = a_2 a_4 + a_3^2 \ , \tag{1.58}$$

and where B' is the matrix obtained from B by permuting the lines as well as the rows, as $(1, 2, 3, 4) \to (2, 3, 4, 1)$. One can then mutate at 2, which yields the new cluster variable

$$a_6 a_2 = a_3 a_5 + a_4^2 \ . \tag{1.59}$$

Repeating this procedure yields a sequence of cluster variables $(a_n)_{n \geq 1}$ such that

$$a_{n+4} a_n = a_{n+1} a_{n+3} + a_{n+2}^2 \ . \tag{1.60}$$

The Somos-4 sequence is defined by the recurrence relation of Equation (1.60) and $a_1 = a_2 = a_3 = a_4 = 1$. It consists *a priori* of rational numbers, however Theorem 1.23 implies that each cluster variable a_n is a Laurent polynomial in a_1, a_2, a_3, a_4, and hence the Somos-4 sequence consists only of integers.

1.5.2 Euler's counterexample of the primality of Fermat numbers

Euler disproved Fermat's conjecture that all numbers of the form $F_n = 2^{2^n} + 1$ are prime numbers, by pointing out that

$$F_5 = 2^{32} + 1 = 641 \cdot 6700417 \ . \tag{1.61}$$

This equality can be obtained in the context of cluster algebras. In this example we need to consider quivers with multipliers, encoding skew-symmetrizable matrices as at the end of Section 1.2. Let us start with the extended seed

$$\tilde{\sigma} = \left(\tilde{a} = (a_1, a_2, a_A), \tilde{B} = \begin{bmatrix} 0 & 4 \\ -1 & 0 \\ 1 & -3 \end{bmatrix} \right) \ , \tag{1.62}$$

encoded in the leftmost quiver on the top row of Figure 1.15. Multipliers are depicted as numbers in parentheses near the nodes of the quivers; the arrow multiplicities are shown as numbers on the edges. A sequence of mutations is presented on the same figure, with the generalized quivers and extended clusters shown at each step.

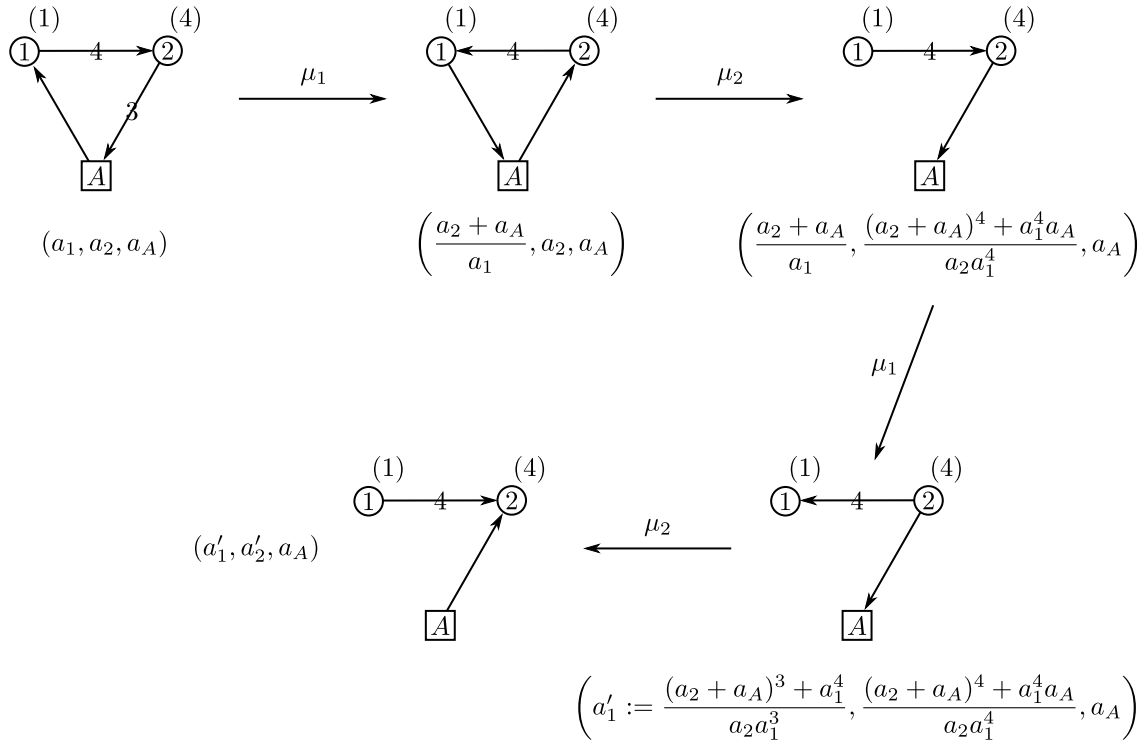


Figure 1.15: A sequence of mutations on an extended seed with generalized quiver.

Let us now specialize the variables in the initial extended cluster to (a_1, a_2, a_A) to $(3, -1, 16)$. Theorem 1.23 implies that every cluster variable appearing somewhere on the seed pattern is of the form $M/3^k$, where $M \in \mathbb{Z}$ and $k \in \mathbb{Z}_{\geq 0}$. Under the mutation of the initial seed at 1, the extended cluster becomes

$$\left(\frac{a_2+a_A}{a_1}, a_2, a_A\right) \ , \tag{1.63}$$

which is $(5, -1, 16)$ when $(a_1, a_2, a_A) = (3, -1, 16)$. Similarly, Theorem 1.23 implies that every cluster variable that appear somewhere on the seed pattern is of the form $N/5^l$, where $N \in \mathbb{Z}$ and $l \in \mathbb{Z}_{\geq 0}$. Hence we conclude that when $(a_1, a_2, a_A) = (3, -1, 16)$ every cluster variable is an integer.

The extended cluster $\mu_1\mu_2\mu_1(a_1, a_2, a_A)$ when $(a_1, a_2, a_A) = (3, -1, 16)$ becomes:

$$\left(\frac{(a_2+a_A)^3+a_1^4}{a_2 a_1^3}, \frac{(a_2+a_A)^4+a_1^4 a_A}{a_2 a_1^4}, a_A\right) = (-128, -641, 16) \ , \tag{1.64}$$

hence applying μ_2 once more, we obtain a new cluster variable a_2' such that $-641 \cdot a_2' = ((a_1')^4 \cdot a_A + 1) = (128^4 \cdot 16 + 1) = (2^{32} + 1) = F_5$.

1.5.3 Markov numbers

The Markov quiver shown on the left of Fig. 1.5 is invariant under mutations, therefore exchange relations will always be of the form $a_i' a_i = a_j^2 + a_k^2$, for $\{i, j, k\} = \{1, 2, 3\}$. If a specialization of a cluster (a_1, a_2, a_3) satisfies the Diophantine equation

$$3a_1 a_2 a_3 = a_1^2 + a_2^2 + a_3^2 \ , \tag{1.65}$$

then the image of this cluster under any mutation is also a solution. For example:

$$\begin{aligned} 3a_1 a_2 a_3' = 3a_1 a_2 \frac{a_1^2 + a_2^2}{a_3} &= \left(a_1^2 + a_2^2 + a_3^2\right) \frac{a_1^2 + a_2^2}{a_3^2} \\ &= \frac{\left(a_1^2 + a_2^2\right)^2}{a_3^2} + a_1^2 + a_2^2 \\ &= (a_3')^2 + a_1^2 + a_2^2 \ . \end{aligned} \tag{1.66}$$

If one specializes (a_1, a_2, a_3) to $(1, 1, 1)$ one obtains a sequence of triples of integers (this is a consequence of Theorem 1.23) that are all solutions to Equation (1.65).

Equation (1.65) is called the *Markov equation.* A triple of integers which is a solution of this equation is said to be a Markov triple, and an element of a Markov triple is a *Markov number.* It can be shown that every Markov triple appears as the image of $(1, 1, 1)$ under multiple mutations.

1.6 Y-patterns

Let $\Big((a_1, \ldots, a_m), \tilde{B} = (b_{ij})\Big)$ be a seed, and for all $j \in [|1, n|]$ let

$$x_j = \prod_{i=1}^{m} a_i^{-b_{ij}} \ , \tag{1.67}$$

so that x_j is the ratio of the two monomials that appear in the numerator of the right-hand side of the exchange relation Equation (1.24) for the mutation at j and the cluster variable a_j.

Proposition 1.24. *For all $j \in [|1, n|]$ let also $\mu_k(x_j)$ be the monomial obtained as in Equation (1.67), but substituting the a_i's for the mutated cluster variables $\mu_k(a_i)$ and the matrix (b_{ij}) for $(b')_{ij}$. Then:*

$$\mu_k(x_j) = \begin{cases} x_k^{-1} & \text{if } k = j, \\ x_j(1 + x_k)^{b_{kj}} & \text{if } j \neq k \text{ and } b_{kj} \leq 0 \ , \\ x_j(1 + x_k^{-1})^{b_{kj}} & \text{if } j \neq k \text{ and } b_{kj} \geq 0 \ . \end{cases} \tag{1.68}$$

Proof. Consider first the case $j = k$. Then

$$\mu_k(x_k) = \prod_i \mu_k(a_i)^{-b'_{ik}} = \prod_{i \neq k} a_i^{-b'_{ik}} = \prod_{i \neq k} a_i^{b_{ik}} = x_k^{-1} \ . \tag{1.69}$$

Now let us assume that $j \neq k$. Then

$$\begin{aligned} \mu_k(x_j) = \prod_i \mu_k(a_i)^{-b'_{ij}} &= \mu_k(a_k)^{b_{kj}} \left(\prod_{i \neq k} a_i^{-b_{ij} - \operatorname{sgn}(b_{ik})[b_{ik} b_{kj}]_+} \right) \\ &= \left(\frac{\displaystyle\prod_{i|b_{ik}>0} a_i^{b_{ik}} + \prod_{i|b_{ik}<0} a_i^{-b_{ik}}}{x_k} \right)^{b_{kj}} \left(\prod_{i \neq k} a_i^{-b_{ij} - \operatorname{sgn}(b_{ik})[b_{ik} b_{kj}]_+} \right) \\ &= \left(\prod_{i|b_{ik}>0} a_i^{b_{ik}} + \prod_{i|b_{ik}<0} a_i^{-b_{ik}} \right)^{b_{kj}} a_k^{-b_{ij}} \left(\prod_{i \neq k} a_i^{-b_{ij} - \operatorname{sgn}(b_{ik})[b_{ik} b_{kj}]_+} \right) \\ &= \left(\prod_{i|b_{ik}>0} a_i^{b_{ik}} + \prod_{i|b_{ik}<0} a_i^{-b_{ik}} \right)^{b_{kj}} \left(\prod_i a_i^{-b_{ij}} \right) \left(\prod_i a_i^{-\operatorname{sgn}(b_{kj})[b_{ik} b_{kj}]_+} \right) \\ &= x_j \left(\prod_{i|b_{ik}>0} a_i^{b_{ik}} + \prod_{i|b_{ik}<0} a_i^{-b_{ik}} \right)^{b_{kj}} \left(\prod_i a_i^{-\operatorname{sgn}(b_{kj})[b_{ik} b_{kj}]_+} \right) \ , \end{aligned} \tag{1.70}$$

and

$$\prod_i a_i^{-\operatorname{sgn}(b_{kj})[b_{ik} b_{kj}]_+} = \begin{cases} \left(\displaystyle\prod_{i|b_{ik}<0} a_i^{-b_{ik}} \right)^{-b_{kj}} & \text{if } b_{kj} < 0, \\ \left(\displaystyle\prod_{i|b_{ik}>0} a_i^{b_{ik}} \right)^{-b_{kj}} & \text{if } b_{kj} > 0, \\ 1 & \text{if } b_{kj} = 0, \end{cases} \tag{1.71}$$

so that

$$\mu_k(x_j) = \begin{cases} x_j(x_k+1)^{b_{kj}} & \text{if } b_{kj} < 0, \\ x_j(1+x_k^{-1})^{b_{kj}} & \text{if } b_{kj} > 0, \\ x_j & \text{if } b_{kj} = 0. \end{cases} \tag{1.72}$$

□

One can define Y-seeds and Y-patterns analogously to the above definitions, as in [FZ07].

Definition 1.25. *A Y-seed of rank n is a pair (x, B) with $x = (x_1, \ldots, x_n)$ and with B a skew-symmetrizable $n \times n$ matrix (with integer entries). The mutation of a Y-seed (x, B) at $k \in [|1, n|]$ is the Y-seed (x', B') such that $B' = \mu_k(B)$ and x' is obtained from x via Equation* (1.68).

Definition 1.26. *A Y-pattern of rank n is a collection of Y-seeds $(x(t), B(t))_{t \in T_n}$ at the vertices of the n-regular tree T_n with edges labeled in $[|1, n|]$, such that no two edges incident to the same vertex have the same label, and such that the Y-seeds on two neighboring vertices linked by an edge carrying the label k are related through the mutation at k.*

As in Equation (1.24), the equations of Equation (1.68) are subtraction-free, and hence they make sense in any semifield. More details on Y-patterns can be found in [FZ07, FWZ06], for example.

* * * * * * *

Cluster algebras are commutative algebras generated by a family of Laurent polynomials in some indeterminates, generated from the latter by mutations. They will play a central role in the following chapters.

In Chapter 2, we will present aspects of the Teichmüller theory of so-called ciliated surfaces, in which the mutation formulas of Equation (1.24) and Equation (1.68) appear naturally. In Chapter 3, we will discuss laminations on ciliated surfaces, in which the tropical versions of the mutation formulae appear naturally. Building on this, cluster varieties will be introduced in Chapter 4. They are the geometric objects dual to cluster algebras, in an algebro-geometric sense. Fock and Goncharov's higher Teichmüller spaces provide examples of cluster varieties; they constitute the core of Chapter 4.

Even if it will be less prominent, cluster algebras and varieties will also be present in Parts II to IV of this dissertation: under Seiberg dualities on toric phases of worldvolume theories on D3-branes at toric Calabi–Yau singularities, the rank of gauge groups transform accordingly to the tropical mutation formula of Equation (1.24).

Chapter 2

Teichmüller theory

We first introduce the (classical) Teichmüller spaces of smooth oriented surfaces of finite type. A slight generalization of this class of surfaces, dubbed the class of *ciliated surfaces*, is then presented. Such oriented smooth surfaces form a good framework to introduce two variants of the classical Teichmüller space, namely the Teichmüller space with holes and the decorated Teichmüller space. Both spaces have interesting coordinates that are naturally associated to the edges of a triangulation of the surface. Under flips of the triangulations, these coordinates change accordingly to the mutation formulae for Y-patterns and cluster variables, respectively. Moreover, they are endowed with additional structure: the Teichmüller space with holes admits a canonical Poisson bracket, while there exist a canonical closed 2-form on the decorated Teichmüller space.

2.1 Generalities on Teichmüller spaces

This first section is mostly based on [FM11, IT12, TM79].

2.1.1 Hyperbolic geometry

The hyperbolic plane $\mathbb{H}$ is the unique complete, simply connected two-dimensional Riemannian manifold with Riemannian metric $d_{\mathbb{H}}$ of constant sectional curvature -1. One model for the hyperbolic plane is the *Poincaré upper half-plane* $\big\{(x,y) \in \mathbb{R}^2 |\ y > 0\big\}$, endowed with the metric:

$$\mathrm{d}s^2 = \frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2} \ . \tag{2.1}$$

The geodesics in the upper half-plane model are vertical lines and half-circles perpendicular to the real axis, and hence any two distinct points in $\mathbb{R} \cup \{\infty\}$ are the endpoints of a unique geodesic. Another model is the Poindaré disk $\big\{(x,y) \in \mathbb{R}^2 |\ r^2 = x^2 + y^2 < 1\big\}$, with the metric

$$\mathrm{d}s^2 = 4\frac{\mathrm{d}x^2 + \mathrm{d}y^2}{(1-r^2)^2} \ , \tag{2.2}$$

in which geodesics are arcs and lines in the disk perpendicular to the unit circle in $\mathbb{R}^2$, and hence any two distinct points on the unit circle S^1 are the endpoints of a unique geodesic. The map

$$z \longmapsto \frac{z-i}{z+i} \tag{2.3}$$

is an isometry from the half-plane to the disk model of $\mathbb{H}$.

Let us define an equivalence relation on the space of unit-speed half geodesics $\gamma : [0,\infty[\to \mathbb{H}$ by imposing that γ and γ' are equivalent if there exists $D \in \mathbb{R}_{>0}$ such that for all $t \in [0,\infty[$:

$$d_{\mathbb{H}}(\gamma(t), \gamma'(t)) \leq D \ . \tag{2.4}$$

Definition 2.1. *The boundary at infinity $\partial_\infty \mathbb{H}$ of $\mathbb{H}$ is the set of equivalence classes of half unit-speed geodesics in $\mathbb{H}$ for the above-defined equivalence relation. One denotes $\overline{\mathbb{H}} = \mathbb{H} \cup \partial_\infty \mathbb{H}$.*

The boundary $\partial_\infty \mathbb{H}$ is naturally identified with $\mathbb{R} \cup \{\infty\} = \mathbb{P}^1(\mathbb{R})$ in the half-plane model, and with the unit circle in the disk model. Moreover, the map of Equation (2.3) extends to the boundary: $\mathbb{P}^1(\mathbb{R}) \to S^1$.

Isometries of $\mathbb{H}$ are naturally represented by elements of $\mathrm{PSL}_2(\mathbb{R})$ acting on $\mathbb{H}$ by homographies:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot z = \frac{az+b}{cz+d} , \tag{2.5}$$

where $z \in \mathbb{C}$ is such that $\mathrm{Im}(z) > 0$, and where $ad - bc = 1$. This action extends to $\partial_\infty \mathbb{H}$, and hence every isometry of $\mathbb{H}$ is a homeomorphism from the closed unit disk in $\mathbb{C}$ to itself. Brouwer fixed point theorem then implies that it has a least one fixed point in $\overline{\mathbb{H}}$. At such a fixed point z_0 Equation (2.5) becomes

$$P(z) := cz_0^2 + (d-a)z_0 - b = 0 , \tag{2.6}$$

and three cases can be distinguished.

- P has two conjugated roots in $\mathbb{C}$, with exactly one in $\mathbb{H}$. Equivalently, $(d-a)^2 + 4bc < 0$ from which one deduces that $(d+a)^2 < 4$ since $ad - bc = 1$, and hence $|d+a| < 2$. The corresponding elements of $\mathrm{PSL}_2(\mathbb{R})$ are said to be *elliptic*. They are rotations about their fixed point.
- P has a double real root. This is equivalent to having the absolute value of the trace equal to 2; the corresponding elements in $\mathrm{PSL}_2(\mathbb{R})$ are said to be *parabolic*.
- P has two distinct real roots. This is equivalent to having the absolute value of the trace strictly bigger than 2; the corresponding elements in $\mathrm{PSL}_2(\mathbb{R})$ are said to be *hyperbolic*. Hyperbolic isometries of $\mathbb{H}$ have a unique geodesic orbit, shown in as dotted red lines in Figure 2.1.

It follows from Equation (2.6) that the only element in $\mathrm{PSL}_2(\mathbb{R})$ which fixes three distinct points or more in $\overline{\mathbb{H}}$ is the identity. The orbits of typical elliptic, parabolic and hyperbolic elements in $\mathrm{PSL}_2(\mathbb{R})$ acting on the half-plane or disk models of $\mathbb{H}$ are shown in Figure 2.1, where the boundary $\partial_\infty \mathbb{H}$ is shown as a dashed line.

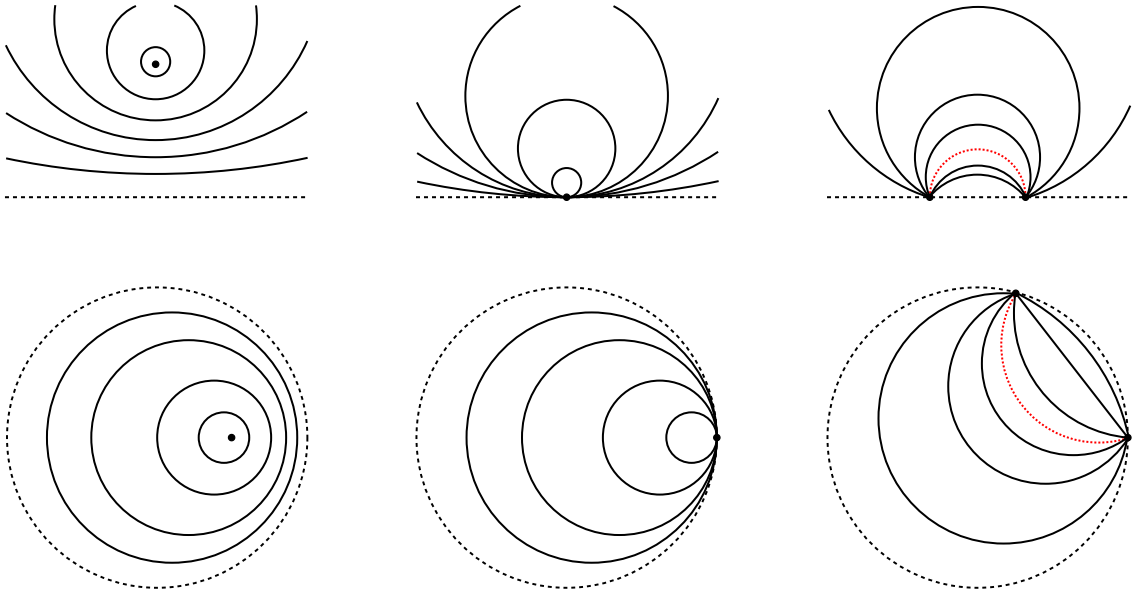


Figure 2.1: Orbits of an elliptic (left), parabolic (middle) and hyperbolic (right) element in $\mathrm{PSL}_2(\mathbb{R})$.

Let now S be a hyperbolic oriented compact smooth surface, i.e. a two dimensional oriented compact Riemannian smooth manifold with metric of constant sectional curvature -1. Its universal cover is $\mathbb{H}$.

Proposition 2.2 (Prop 5.3.1 in [TM79])**.** *Let $\alpha \in \pi_1(S)$ be a free homotopy class of closed curves in S. There exists a unique geodesic in α.*

Let us assume that S has a non-empty boundary, and let $\beta \in \pi_1(S, \{p_0, p_1\})$ be a homotopy class of curves in S from $p_0 \in \partial S$ to $p_1 \in \partial S$ such that the preimages of p_0 and p_1 under the covering map $\mathbb{H} \to S$ lie on $\partial_\infty \mathbb{H}$, and considered relatively to $\{p_0, p_1\}$. There exists a unique geodesic in β.

Proof. Let $M_\alpha \in \mathrm{PSL}_2(\mathbb{R})$ be the covering automorphism of $\mathbb{H}$ corresponding to α. It cannot be elliptic since S is smooth. It cannot be parabolic, for otherwise there would be simple closed curves in α of arbitrarily small length, which goes against the fact that S is compact. Hence M_α is a hyperbolic element of $\mathrm{PSL}_2(\mathbb{R})$, and it preserves a unique geodesic which projects to a closed geodesic on S in the homotopy class α. The geodesics γ and γ' corresponding to two hyperbolic elements M_α and $M_{\alpha'}$ in $\mathrm{PSL}_2(\mathbb{R})$ project to the same geodesic in S if and only if there is a covering transformation which maps the first to the second, i.e. $\alpha' = g\alpha g^{-1}$ for $g \in \pi_1(S)$, which is equivalent to α' and α being in the same free homotopy class.

Let $\tilde{p}_0$ (respectively $\tilde{p}_1$) be a preimage of p_0 (respectively p_1) under the covering map $\mathbb{H} \to S$. The points $\tilde{p}_0$ and $\tilde{p}_1$ are distinct points on $\partial_\infty\mathbb{H}$. There exists a one-parameter family of hyperbolic elements in $\mathrm{PSL}_2(\mathbb{R})$ which preserves $\tilde{p}_0$ and $\tilde{p}_1$ and fixes a unique geodesic from $\tilde{p}_0$ to $\tilde{p}_1$ (again shown as dotted red lines on the right of Figure 2.1). This geodesic projects to a geodesic on S in the class β. If $\tilde{p}'_0$ and $\tilde{p}'_1$ are two other lifts of p_0 and p_1 in $\partial_\infty\mathbb{H}$, the unique geodesic from $\tilde{p}'_0$ to $\tilde{p}'_1$ projects to a geodesic in β if and only if it is related to the unique one from $\tilde{p}_0$ to $\tilde{p}_1$ by a covering transformation, which implies that both geodesics in $\mathbb{H}$ project to the same one in S. □

Let us now turn to the definition of *horocycles*, which are central objects in the definition decorated Teichmüller spaces.

Definition 2.3. *The horocycle in $\mathbb{H}$ at $A \in \mathbb{P}^1(\mathbb{R}) = \partial_\infty\mathbb{H}$ and going through $M \in \mathbb{H}$ is the limit of the family of hyperbolic circles in $\mathbb{H}$ going through M as their hyperbolic center tends to A. Equivalently, it is an orbit of a parabolic element in $\mathrm{PSL}_2(\mathbb{R})$ stabilizing A (see Figure 2.1).*

In the half-plane model and the disk model of $\mathbb{H}$, hyperbolic circles are euclidean circles tangent at a point of the boundary $\partial_\infty\mathbb{H}$: a horocycle at $A \in \mathbb{R} \subset \mathbb{P}^1(\mathbb{R})$ is a circle (hyperbolic, or equivalently, euclidean) in the upper half-plane $\mathbb{H}$ and tangent to the real axis at A. A horocycle at $\infty \in \mathbb{P}^1(\mathbb{R})$ is a horizontal line in the upper half-plane. Any horocycle at A is orthogonal to every geodesic in $\mathbb{H}$ that has A as one of its endpoint. Moreover, the following holds.

Proposition 2.4. *Let H_1 and H_2 be two horocycles based at the same point $A \in \mathbb{P}^1(\mathbb{R})$, and let p and p' be two points on H_1. Then $d_{\mathbb{H}}(p, H_2) = d_{\mathbb{H}}(p', H_2)$.*

There is a convenient description of the set of horocycles [FG07]:

Proposition 2.5. *The map*

$$\left(\mathbb{R}^2 - \{(0,0)\}\right) / \pm 1 \longrightarrow \{\textit{Horocycles in } \mathbb{H}\} \tag{2.7}$$

which associates to $(x, y) \in \left(\mathbb{R}^2 - \{(0,0)\}\right)$ the horocycle at $x/y \in \mathbb{P}^1(\mathbb{R})$ and of euclidean diameter y^{-2} (when $y = 0$ it is x^{-2} which plays the role of y^{-2}) is a canonical isomorphism equivariant with respect to the action of $\mathrm{PSL}_2(\mathbb{R})$.

2.1.2 Generalities on Teichmüller spaces

Three definitions of the Teichmüller space of a smooth oriented surface. Let S be an oriented smooth surface. A *marked complex structure* on S is a pair (X, f_X) where X is a Riemann surface and $f_X : S \to X$ is an orientation preserving diffeomorphism. Two marked complex structures (X, f_X) and (Y, f_Y) on S are said to be equivalent if $f_Y \circ f_X^{-1} : X \to Y$ is isotopic to a biholomorphic map.

Definition 2.6. *The Teichmüller space $\mathcal{T}(S)$ of S is the space of equivalence classes of marked complex structures on S.*

For example, $\mathcal{T}(S^2)$ is a point, $\mathcal{T}(\mathbb{R}^2)$ consists of two points representing the complex plane and the Poincaré disk, and $\mathcal{T}(T^2) = \mathbb{H}$.

Let now $\mathrm{Diff}(S)$ be the group of orientation-preserving diffeomorphisms of S, and let $\mathrm{Diff}_0(S)$ be the normal subgroup of $\mathrm{Diff}(S)$ consisting of the diffeomorphisms of S isotopic to Id. The *mapping class group* $\mathrm{MCG}(S)$ of S is the quotient group $\mathrm{Diff}(S)/\mathrm{Diff}_0(S)$. The mapping class group $\mathrm{MCG}(S)$ acts on $\mathcal{T}(S)$ by changing the marking, and the quotient $\mathcal{M}(S) = \mathcal{T}(S)/MCG(S)$ is naturally identified with the moduli space $\mathcal{M}(S)$ of complex structures on S. In the case of the torus $\mathbb{T}^2$, the mapping class group is

$\mathrm{PSL}_2(\mathbb{Z})$ and $\mathcal{M}(\mathbb{T}^2) = \mathbb{H}/\mathrm{PSL}_2(\mathbb{Z})$.

Let us now recall the following fundamental theorem, for the proof of which we refer to the references listed in [IT12, Section 2.1].

Theorem 2.7 (Uniformization theorem)**.** *Every simply connected Riemann surface is biholomorphically equivalent to the complex sphere $\mathbb{P}^1(\mathbb{C})$, the complex plane $\mathbb{C}$ or the upper half-plane $\mathbb{H}$.*

The data of a Riemannian metric $\mathrm{d}s^2$ on an oriented 2-manifold S induces a complex structure on S through *isothermal coordinates* [IT12, Section 1.5]. Depending on whether the universal cover of the Riemann surface constructed from $(S, \mathrm{d}s^2)$ is $\mathbb{P}^1(\mathbb{C})$, $\mathbb{C}$ or $\mathbb{H}$ respectively, one can show that there is a unique metric on S of constant curvature 1, 0 or -1 respectively, in the same conformal class as $\mathrm{d}s^2$. Which of these three simply-connected Riemann surface is the universal cover of a given compact Riemann surface depends only on the Euler characteristic $\chi(S)$ of S: if $\chi(S) > 0$ the universal cover is $\mathbb{P}^1(\mathbb{C})$, if $\chi(S) = 0$ it is $\mathbb{C}$ and if $\chi(S) < 0$ it is $\mathbb{H}$. If S is closed then $\chi(S) = 2 - 2g$ where g is the genus of S.

Let R' and S' be two Riemann surfaces whose complex structures are induced by the Riemannian metric of two oriented Riemann 2-manifolds $(R, \mathrm{d}s^2_R)$ and $(S, \mathrm{d}s^2_S)$. Then, a map $f : R \to S$ is a conformal isomorphism if and only if $f : R' \to S'$ is a biholomorphism [IT12, Theorem 1.7]. Hence the Teichmüller space of a smooth surface S can be defined as follows.

Definition 2.8. *If $\chi(S) < 0$ (respectively $\chi(S) = 0$, $\chi(S) > 0$), the Teichmüller space of S is the space of metrics of constant curvature -1 (respectively 0, 1) with which S can be endowed, up to $\mathrm{Diff}_0(S)$: two metrics $\mathrm{d}s^2$ and $(\mathrm{d}s')^2$ on S are equivalent is there exists $f \in \mathrm{Diff}_0(S)$ such that $f : (S, \mathrm{d}s^2) \longrightarrow (S, (\mathrm{d}s')^2)$ is a conformal isomorphism.*

Regarding this definition and having already described the Teichmüller spaces of the sphere and the torus, we will assume in the sequel that $\chi(S) < 0$ [1]. A Riemannian metric of constant curvature -1 on a smooth oriented surface S is said to be *hyperbolic*. By extension, a smooth oriented surface S such that $\chi(S) < 0$ will also be said *hyperbolic*. When S is closed and oriented, $\chi(S) < 0$ is equivalent to $g \geq 2$. If S is obtained by removing k disjoint open disks to the closed oriented surface of genus g, in which case S is said to be *of finite type*, one has:

$$\chi(S) = 2 - 2g - k \ . \tag{2.8}$$

Except when $g = 0$ and $k \leq 2$ or $g = 1$ and $k = 0$, the Euler characteristic of S is negative; in other words, being hyperbolic is the norm.

The universal covering transformation group Γ of a hyperbolic surface of finite type S endowed with a hyperbolic metric is a subgroup of $\mathrm{Aut}(\mathbb{H})$ and acts properly discontinuously on $\mathbb{H}$.

Theorem 2.9 (2.5 in [IT12])**.** *For all $p \in S$, the group Γ is isomorphic to the fundamental group $\pi_1(R, p)$.*

There is a natural topology on $\mathrm{PSL}_2(\mathbb{R})$ induced by the one on $\mathrm{SL}_2(\mathbb{R})$: a sequence $(A_n)_{n\geq 1} \in \mathrm{SL}_2(\mathbb{R})$ converges to $A \in \mathrm{SL}_2(\mathbb{R})$ if and only if the entries of (A_n) converge to the entries of A. This topology coincide with the compact-open topology on $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{Aut}(\mathbb{H})$. If a subgroup Γ of $\mathrm{PSL}_2(\mathbb{R})$ is discrete with respect to it, it is called a *Fuchsian group*.

Theorem 2.10 (2.17 in [IT12])**.** *Let Γ be a subgroup of $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{Aut}(\mathbb{H})$. The following are equivalent:*

1. *Γ is Fuchsian.*
2. *Γ acts properly discontinuously on $\mathbb{H}$.*

The Fuchsian group Γ defined as the universal covering group of $\mathbb{H} \to S$ is called *Fuchsian model* of S. We can now state a third definition of Teichmüller spaces.

Definition 2.11. *Let S be a smooth oriented surface with $\chi(S) < 0$, $p \in S$ and $\pi_1(S, p)$ the fundamental group of S based at p. The Teichmüller space of S is the space of faithful group morphisms*

$$\pi_1(S, p) \longrightarrow \mathrm{PSL}_2(\mathbb{R}) \tag{2.9}$$

with discrete image, modulo overall conjugation by $\mathrm{PSL}_2(\mathbb{R})$ (accounting for base point changes)[2]. The class of loop surrounding a puncture needs to be mapped to a parabolic element of $\mathrm{PSL}_2(\mathbb{R})$.

[1] This case is the most general one: every Riemann surface whose universal cover is not $\mathbb{H}$ is either $\mathbb{P}^1(\mathbb{C})$, $\mathbb{C}$, $\mathbb{C} - \{0\}$ or a torus [IT12, Theorem 2.15].

[2] More precisely: the space of discrete and faithful morphisms $\pi_1(S, p) \to \mathrm{PSL}_2(\mathbb{R})$ modulo $\mathrm{PSL}_2(\mathbb{R})$-conjugation consists of two connected components – one of which is the Teichmüller space. This will be discussed more extensively in Chapter 4.

The Fricke embedding of $\mathcal{T}(S)$. One can choose a normalization of the image of the map in Equation (2.9) which fixes a representative of the conjugacy class under the global action of $\mathrm{PSL}_2(\mathbb{R})$. The free entries of the image of a set of generators of $\pi_1(S,p)$ under the map of Equation (2.9) then form a set of coordinates for the Teichmüller space $\mathcal{T}(S)$.

Let $S_{g,k}$ be the surface obtained from the closed oriented smooth surface of genus g by removing k points. The fundamental group of $S_{g,k}$ can be presented as

$$\pi_1(S_{g,k},p) = \left\langle \alpha_1,\dots,\alpha_g,\beta_1,\dots,\beta_g,\gamma_1,\dots,\gamma_k \,\middle|\, \prod_{i=1}^{g}\alpha_i\beta_i\alpha_i^{-1}\beta_i^{-1} = \prod_{i=1}^{k}\gamma_i \right\rangle , \tag{2.10}$$

and any choice of normalization defines an embedding:

$$\mathcal{T}(S_{g,k}) \to \mathbb{R}^{6g-6+3k} , \tag{2.11}$$

which in turns induces a topology on $\mathcal{T}(S_{g,k})$ from the one on $\mathbb{R}^{6g-6+3k}$, with respect to which $\mathcal{T}(S_{g,k})$ is homeomorphic to $\mathbb{R}^{6g-6+3k}$: **topologically, the Teichmüller space is an open ball**.

The topology on $\mathcal{T}(S)$ can be defined in a more intrinsic way. Consider a point in $\mathcal{T}(S_{g,k})$, so that $S_{g,k}$ is endowed with a hyperbolic structure. Proposition 2.2 tells than in each free homotopy class of closed curves on $S_{g,k}$ there exists a unique geodesic for the hyperbolic metric at hand. The hyperbolic length of this geodesic can be considered as a function

$$l : \mathcal{T}(S) \longrightarrow \mathbb{R}_{>0} , \tag{2.12}$$

and denoting SCC the set of all simple closed curves on S, the map

$$\mathcal{T}(S) \longrightarrow \mathbb{R}^{\mathrm{SCC}} \tag{2.13}$$

is an embedding. The product topology on $\mathbb{R}^{\mathrm{SCC}}$ induces a topology on $\mathcal{T}(S)$ for which it is homeomorphic to $\mathbb{R}^{6g-6+3k}$.

Metrics on Teichmüller spaces. There is yet another definition of Teichmüller spaces in terms of quasiconformal mappings, on which we will not extend (see [IT12, Chapter 4,5,6] for details about this). Let us only point out that when S is a closed oriented smooth surface there is a natural open embedding (called the *Bers embedding*) of $\mathcal{T}(S)$ into the space of quadratic differentials on S, which is a complex space of complex dimension $3g-3$. This endows $\mathcal{T}(S)$ with a natural complex structure.

The definition of Teichmüller spaces in terms of quasi-conformal mappings allow the definition of metrics on $\mathcal{T}(S)$. On the first hand, the *Teichmüller metric* induces the topology on $\mathcal{T}(S)$ as defined above, and $\mathcal{T}(S)$ is complete with respect to it. On the other hand, the *Weil–Peterson metric* is defined as follows. The tangent space of $\mathcal{T}(S)$ at some point (R,f) is identified with the space of quadratic differentials on S, which is endowed with an $\mathbb{L}^2$-hermitian product called the Petersson inner product. The latter induces the Weil–Peterson metric on $\mathcal{T}(S)$; it is a Kähler metric with negative Ricci, scalar and holomorphic sectional curvatures [IT12, Chapter 8].

Fenchel–Nielsen coordinates. Any decomposition of a hyperbolic surface S of finite type in pairs of pants yields a set of coordinates on $\mathcal{T}(S)$.

A pair of pants is a smooth oriented surface diffeomorphic to a sphere with three holes. Given a pair of pants P and $a,b,c \in \mathbb{R}_{>0}$ there is a unique hyperbolic structure on P such that the boundary components of P are geodesics of respective hyperbolic length a,b and c. Given two hyperbolic pairs of pants P and P' such that both have a boundary component (denoted C and C' respectively) of hyperbolic length $a \in \mathbb{R}_{>0}$, one can glue P and P' together by identifying C and C'. This can be done in $\mathbb{R}$ inequivalent ways ensuring that the hyperbolic metrics on the two original surfaces glue to a hyperbolic metric on the resulting one. This real parameter is called the twist corresponding to the simple closed curve $C = C'$ in the resulting surface. One can also glue together different boundary components of the same hyperbolic pair of pants.

In the case of the finite type surface $S_{g,k}$ obtained by removing $k > 0$ disks to a genus $g \geq 0$ oriented closed surface where g and k are such that $\chi(S_{g,k}) = 2-2g-k < 0$, one needs $3g-3+k$ simple closed curves to define a pants decomposition, and the latter contains $2g-2+k$ pairs of pants. The boundary components of S are hyperbolic geodesics entirely described by their lengths, whereas the

simple closed curves in the interior of S are assigned a hyperbolic length and a twist parameter; these are the *Fenchel–Nielsen coordinates*. They form a chart

$$\mathbb{R}_{>0}^{3g-3+2k} \times \mathbb{R}^{3g-3+k} \longrightarrow \mathcal{T}(S_g) \ . \tag{2.14}$$

An example of a pants decomposition of the closed oriented smooth surface of genus $g = 3$ is shown on the left of Figure 2.2, and another of the surface of finite type $S_{g,k}$ with $g = 2$ and $k = 1$ is shown on the right of Figure 2.2.

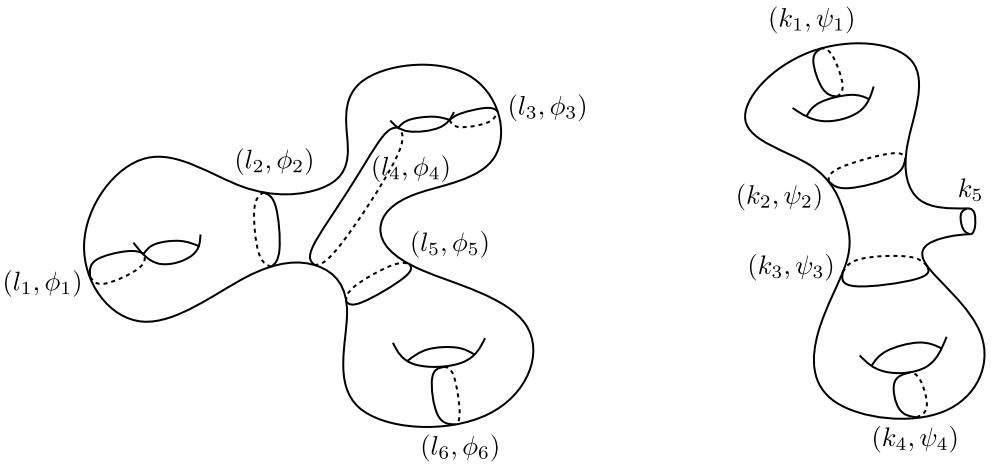


Figure 2.2: Hyperbolic lengths are denoted l_i and k_i while the twists are the ϕ_i and ψ_i.

There are in general many different pants decompositions of the same surface, and each one yields a different set of Fenchel–Nielsen coordinates on the Teichmüller space $\mathcal{T}(S)$ of the surface S. These sets are different global charts on $\mathcal{T}(S)$. One can compute the transitions functions from one chart to another.

When S is closed, the mapping class group $\mathrm{MCG}(S)$ of S is generated by the *Dehn twists* about simple closed curves in S. A Dehn twist about a simple closed curve c on S can be described as follows: let $\mathcal{N} \simeq S^1 \times [0,1] \subset S$ the a tubular neighborhood of c in S. The Dehn twist about c is the map $f : S \to S$ which is the identity in $S - \mathcal{N}$ and defined in $\mathcal{N}$ by:

$$f : (s,t) \in S^1 \times [0,1] \longmapsto (e^{2i\pi t}s, t) \ . \tag{2.15}$$

In the remainder of this chapter we are going to introduce two slightly different notions of Teichmüller spaces, namely *Teichmüller space with holes* $\mathcal{T}^x(S)$ and *decorated Teichmüller spaces* $\mathcal{T}^a(S)$. They are defined on *ciliated surfaces*, i.e. finite type smooth oriented surfaces with marked points on the boundary. Each version of Teichmüller space is endowed with coordinate systems, which as in the Fenchel-Nielsen case are associated to decompositions of the surface into elementary pieces – triangles in these cases. The set of coordinates associated to a given triangulation of a ciliated surface can be seen as a global chart on the Teichmüller space, and the transition functions between any two charts are sequences of Y-patterns mutations as in Equation (1.68) and cluster mutations as in Equation (1.24). Hence, both Teichmüller spaces hint for the notion of *cluster varieties*, i.e. algebraic varieties whose regular functions are related to cluster algebras. Teichmüller spaces with holes are endowed with a canonical Poisson bracket (with Casimirs), whereas on the decorated ones there is a canonical closed 2-form.

We will discuss these two types of cluster Teichmüller spaces in turn, after having defined carefully ciliated surfaces and studied the notion of triangulation we are interested in.

2.2 Ciliated surfaces and triangulations

2.2.1 Definitions

Definition 2.12. *A ciliated surface is a compact oriented smooth surface obtained from a closed oriented smooth surface by removing a finite number of disjoint open disks and by marking a finite set of points on the boundary. The latter are called cilia. A boundary component without cilium is said to be a hole.*

Let $k \in \mathbb{Z}_{\geq 0}$ be the number of disjoint open disks one removes from the closed oriented smooth surface S_g of genus g, and let us label the boundary components of the resulting surface $1, \dots, k$ in such a way that $1, \dots, h$ are holes and with $p_i \in \mathbb{Z}_{>0}$ the number of marked points on the i-th boundary circle, for $i = h+1, \dots, k$. The resulting ciliated surface is completely determined by g, h and the set $\{p_{h+1}, \dots, p_k\}$ as a smooth (equivalently, topological) surface, and hence we will denote it $S_{g,h,\{p_{h+1},\dots,p_k\}}$ (and $S_{g,h}$ for short when there is no cilium). Three examples of ciliated surfaces are shown in Figure 2.3.

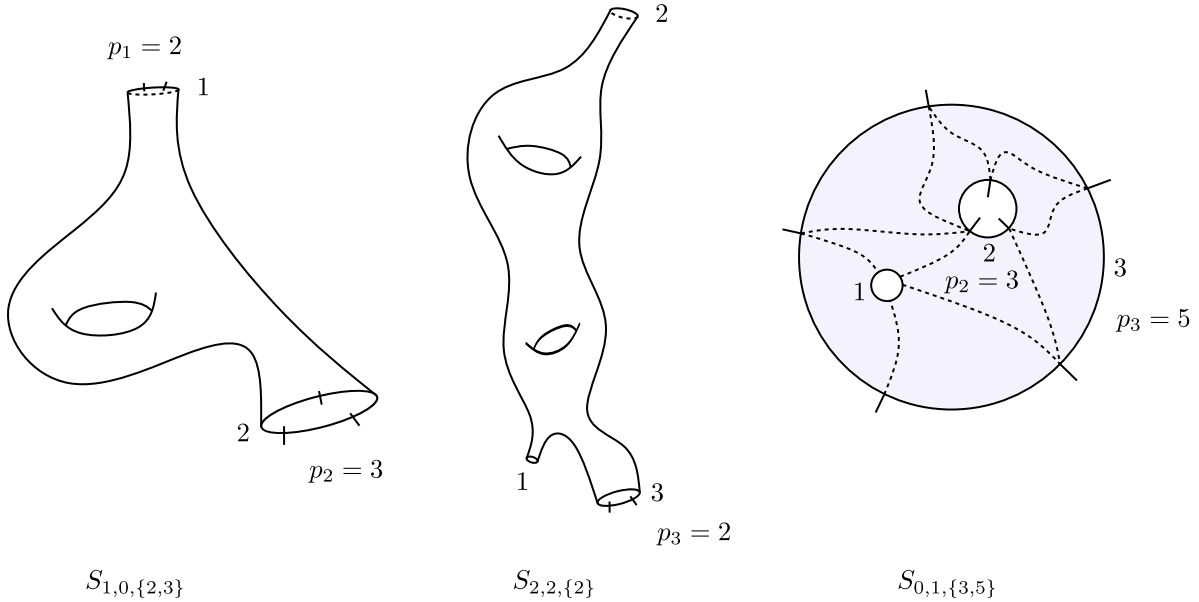


Figure 2.3: Three ciliated surfaces.

In the sequel, unless explicitly specified, we will assume every ciliated surface to be non-closed: $k \geq 1$, and hyperbolic: $\chi(S) < 0$. These conditions are equivalent to either $g = 0$ and $k \geq 3$, or $g \geq 1$ and $k \geq 1$.

Definition 2.13. *The boundary of a ciliated surface is the disjoint union of the boundary segments connecting two adjacent cilia.*

Definition 2.14. *A triangulation Γ of a ciliated surface S such that $k \geq 1$ and $\chi(S) < 0$ is a maximal isotopy class of non self-intersecting, pairwise non-intersecting and non-isotopic curves on S whose endpoints are at holes or cilia. A triangulation of S with holes shrunk to punctures decomposes it into topological triangles such that each vertex is either a cilium or a shrunk hole.*

A triangulation of $S_{0,1,\{3,5\}}$ is shown on the right of Figure 2.3. If Γ is a triangulation of a ciliated surface S, the curves in Γ are the edges of Γ, the endpoints of these curves are its vertices, while the connected components of $S - \Gamma$ are its faces. One can distinguish the *external edges*, which belong to the boundary of S, and the other edges, dubbed *internal edges*.

Let $V(\Gamma)$, $E(\Gamma)$, $F(\Gamma)$, $E_e(\Gamma)$, $E_i(\Gamma)$ be respectively the set of vertices, edges, faces, external edges and internal edges of Γ, and for each such finite set A let $\#A$ denote the cardinal of A. Let also $c = \sum p_i$ be the total number of cilia in S. One has:

$$\#E_e(\Gamma) = c \ , \tag{2.16}$$

$$\#V(\Gamma) = h + c \tag{2.17}$$

by Definition 2.14. The Euler characteristic of S with shrunk holes is

$$\#F(\Gamma) - \#E(\Gamma) + \#V(\Gamma) = 2 - 2g + h - k \ , \tag{2.18}$$

and since Γ decomposes S in triangles:

$$3\#F(\Gamma) = 2\#E(\Gamma) - \#E_e(\Gamma) = 2\#E(\Gamma) - c \ . \tag{2.19}$$

From these two last equations one deduces that:

$$\#E(\Gamma) = 6g - 6 + 2c + 3k \ , \tag{2.20}$$

$$\#E_i(\Gamma) = 6g - 6 + c + 3k \ , \tag{2.21}$$

$$\#F(\Gamma) = 4g - 4 + c + 2k \ . \tag{2.22}$$

Hence $\#E(\Gamma)$, $\#E_e(\Gamma)$, $\#E_i(\Gamma)$, $\#V(\Gamma)$, $\#F(\Gamma)$ are entirely determined by the topology of S.

The number of distinct triangulations of a ciliated surface is in general infinite, except for $S_{0,0,\{p\}}$ and $S_{0,1,\{p\}}$ with $p \geq 1$. Two triangulations, respectively of $S_{0,0,\{5\}}$ and $S_{0,1,\{4\}}$ are shown in Figure 2.4, together with their corresponding quiver constructed as in Section 1.2. The mutable part of the quiver on the left of Figure 2.4 is a Dynkin quiver of type A_2, and the one on the right of Figure 2.4 is a Dynkin quiver of type D_4; the corresponding cluster algebras are of finite type as proved in [FZ03].

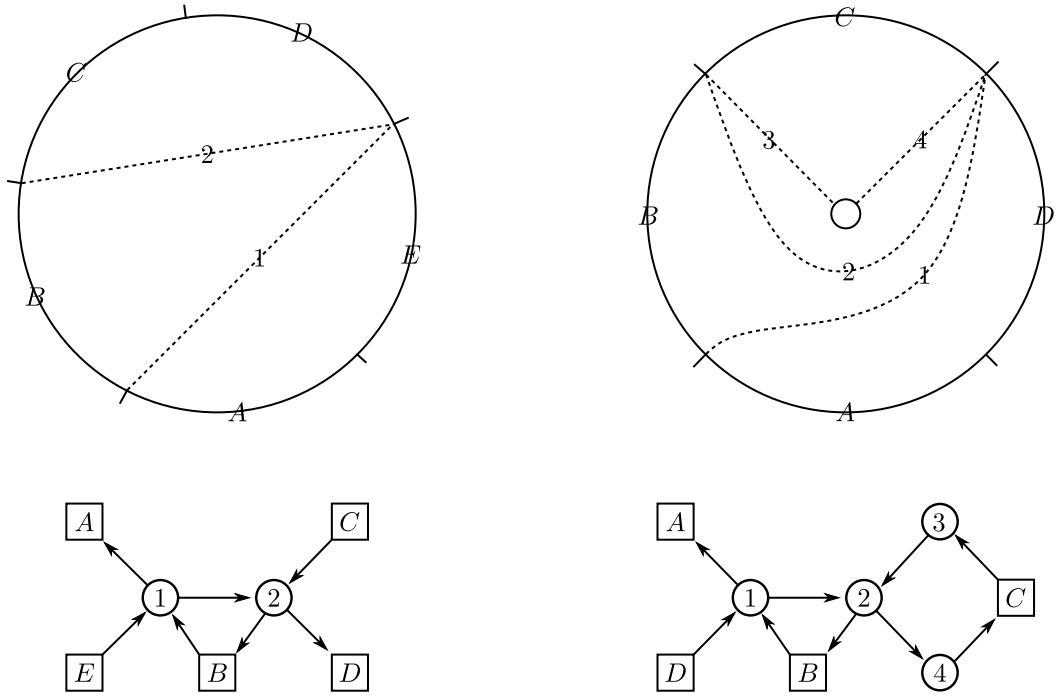


Figure 2.4: A triangulation of $S_{0,0,\{5\}}$ (left) and of $S_{0,1,\{4\}}$ (right).

The following classical result will be of prime importance to us, since the coordinates to be constructed on $\mathcal{T}^x(S)$ and $\mathcal{T}^a(S)$ are associated to the edges of a triangulation Γ of S. It implies that in order to compute the transition function between any two charts, i.e. any two triangulations of S, one only needs to know what the transition function is when the two triangulations are related by a mere flip of an internal edge.

Theorem 2.15 ([Mos88, Hat91, Bur99])**.** *Any two triangulations of ciliated surface can be related through a finite sequence of flips (or Whitehead moves) displayed in Figure 2.5.*

2.2.2 Combinatorial type of a triangulation

While there are, in general, infinitely many inequivalent triangulations of a fixed ciliated surface S, the number of different *combinatorial types* of triangulations of S is always finite. The combinatorial type of the triangulation corresponds exactly to what is encoded in the quiver dual to it.

As in Section 1.2 the quiver corresponding to a triangulation can be equivalently described as a skew-symmetric extended exchange matrix. In the sequel we will denote the edges of a triangulation $\alpha, \beta, \gamma \in E(\Gamma)$, and the extended exchange matrix will be written $\epsilon^{\alpha\beta}$. It is straightforward to check that:

$$\epsilon^{\alpha\beta} = \sum_{i \in F(\Gamma)} \langle \alpha | i | \beta \rangle \ , \tag{2.23}$$

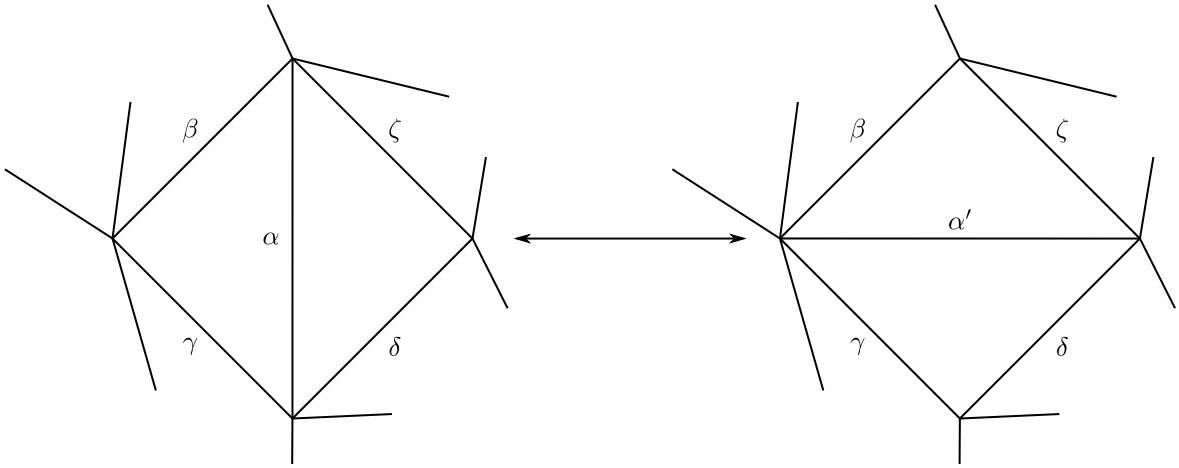


Figure 2.5: A flip of a triangulation at the inner edge e.

where $\langle\alpha|i|\beta\rangle = +1$ if α and β are two sides of the face i and if β is the edge directly after α with respect to their common vertex, in the clockwise direction. Likewise, $\langle\alpha|i|\beta\rangle = -1$ if α and β are two sides of the face i and if β is the edge directly after α with respect to their common vertex, in the counterclockwise direction. The extended exchange matrix ϵ encoding the topology of the triangulation is rectangular $E(\Gamma) \times E_i(\Gamma)$ (since external edges of Γ cannot be flipped), and its entries belong to $\{0, \pm 1, \pm 2\}$. For example, the exchange matrices corresponding to the triangulations of $S_{0,0,\{5\}}$ and $S_{0,1,\{4\}}$ shown in Figure 2.4 are respectively:

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \\ -1 & 0 \\ 1 & -1 \\ 0 & 1 \\ 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & -1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix} . \tag{2.24}$$

Triangulations of ciliated surfaces can be flipped at internal edges α as shown in Figure 2.5. There is a natural identification between the edges of the original triangulation and the edges of the resulting one, and hence there is a natural choice of edge labels after a flip, in terms of the ones before the flip. The extended exchange matrix varies in the following way under the flip of some edge $\alpha \in E_i(\Gamma)$:

$$(\epsilon')^{\beta\gamma} = \begin{cases} -\epsilon^{\beta\gamma} & \text{if } \alpha = \beta \text{ or } \alpha = \gamma \\ \epsilon^{\beta\gamma} + \mathrm{sgn}(\epsilon^{\beta\alpha})[\epsilon^{\beta\alpha}\epsilon^{\alpha\gamma}]_+ & \text{otherwise} \end{cases} , \tag{2.25}$$

i.e. just as in Equation (1.14). If $\alpha \in E_i(\Gamma)$ we will denote μ_α the flip at α so that we can write:

$$\mu_\alpha : E_i(\Gamma) \longrightarrow E_i(\mu_\alpha(\Gamma)) . \tag{2.26}$$

Proposition 2.16. *Let $\alpha, \beta \in E_i(\Gamma)$. Under the identification of the edges of a triangulation Γ with the edges of the triangulation obtained by the flip of an internal edge, one has:*

1. $\mu_\alpha \circ \mu_\alpha = \mathrm{Id}$,
2. *If $\epsilon^{\alpha\beta} = 0$, then $\mu_\alpha \circ \mu_\beta = \mu_\beta \circ \mu_\alpha$: flips at disjoint edges commute.*

Let us emphasize once more that distinct triangulations can have the same combinatorial type. For example the two triangulations of the once-punctured torus shown in Figure 2.6 cannot be deformed one into the other, however they have the same topological type, which corresponds to the Markov quiver. We have seen in Section 1.2 that this quiver is invariant under mutations, and therefore the topological type of any triangulation of $S_{1,1}$ is described by the Markov quiver.

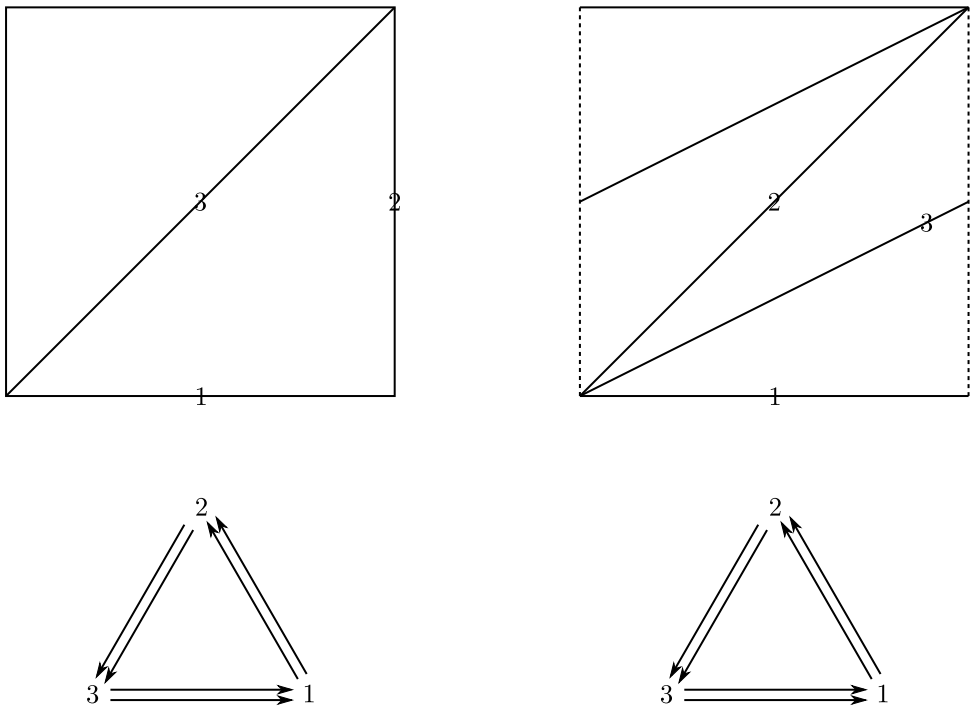


Figure 2.6: Two different triangulations of $S_{1,1}$.

2.2.3 Mapping class groups of ciliated surfaces

Definition 2.17. *The mapping class group* $\mathrm{MCG}(S)$ *of a ciliated surface* S *is the group of connected components of the diffeomorphisms of* S *preserving the orientation and the set of cilia.*

Let us quickly and without proof present the description of the modular groupoid of S given in [FG07, Appendix A]. A *groupoid* is a category $\mathcal{G}$ where all morphisms are invertible and where any two objects are isomorphic. To each such category one can associate a group called the *fundamental group of the groupoid* since for all $A, B \in \mathcal{O}b(\mathcal{G})$ the automorphism groups $\mathrm{Hom}(A, A)$ and $\mathrm{Hom}(B, B)$ are isomorphic. For example, the *fundamental groupoid* of a topological space X has the points of X as objects, and for all $x, y \in X$, $\mathrm{Hom}(x, y)$ is the set of homotopy classes of paths from x to y. From the definition it is clear that $\mathrm{Hom}(x, x) = \pi_1(X, x)$ and $\mathrm{Hom}(y, y) = \pi_1(X, y)$. Any $\phi \in \mathrm{Hom}(x, y)$ induces an isomorphism $\mathrm{Hom}(y, y) \to \mathrm{Hom}(x, x)$.

Let $\mathsf{Tr}(S)$ be the set of pairs (Γ, f) where Γ is a triangulation of S and f is a marking of the edges of Γ, i.e. an isomorphism:

$$f : E(\Gamma) \longrightarrow \{1, \ldots, 6g - 6 + 2c + 3k\} \ . \tag{2.27}$$

A pair (Γ, f) is said to be a marked triangulation of S. Let also $|\mathsf{Tr}|(S)$ be the set of combinatorial classes of triangulations of S. The mapping class group of S acts freely on the set of marked triangulations (the marking deals with possible degeneracies of the action on triangulations with non-trivial automorphism groups), and:

$$|\mathsf{Tr}|(S) = \mathsf{Tr}(S)/MCG(S) \ . \tag{2.28}$$

Definition 2.18. *The modular groupoid* $\mathcal{MCG}(S)$ *is the category whose set of objects is* $|\mathsf{Tr}|(S)$ *and where for* $C_1, C_2 \in |\mathsf{Tr}|(S)$ *the set* $\mathrm{Hom}(C_1, C_2)$ *consists of the classes*

$$[\Gamma_1, \Gamma_2] = \mathrm{MCG}(S) \cdot (\Gamma_1, \Gamma_2) \ , \tag{2.29}$$

where Γ_1 *is of type* C_1 *and* Γ_2 *of type* C_2*, and where* $\mathrm{MCG}(S)$ *acts diagonally on the pairs. If* Γ_1, Γ_2 *and* Γ_3 *are triangulations of* S *then* $[\Gamma_1, \Gamma_2] \circ [\Gamma_2, \Gamma_3] = [\Gamma_1, \Gamma_3]$.

One can check that if Γ_1, Γ_2 are any two triangulations of S then $[\Gamma_1, \Gamma_1]$ is the identity and $[\Gamma_1, \Gamma_2] = [\Gamma_2, \Gamma_1]^{-1}$. Moreover for any $C \in |\mathsf{Tr}|(S)$ the group $\mathrm{Hom}(C, C)$ is isomorphic to $\mathrm{MCG}(S)$. There are morphisms in $\mathcal{MCG}(S)$ corresponding to the flips: if Γ is a triangulation of S and if $\alpha \in E_i(\Gamma)$, then $[\Gamma, \mu_\alpha \Gamma]$ is a morphism from the combinatorial type of Γ to the one of $\mu_\alpha \Gamma$. The main proposition is:

Proposition 2.19. *The modular groupoid $\mathcal{MCG}(S)$ of S is generated by the flips and subjects to the following three types of relations only:*

1. *If $\alpha \in E_i(\Gamma)$ and if $[\Gamma, \mu_\alpha\Gamma]$ is a flip then $[\mu_\alpha\Gamma, \Gamma]$ is also a flip and*
$$[\Gamma, \mu_\alpha\Gamma] \circ [\mu_\alpha\Gamma, \Gamma] = \mathrm{Id}_{|\Gamma|} \ , \tag{2.30}$$
2. *If α, β are disjoint edges in $E_i(\Gamma)$ then*
$$[\mu_\alpha\Gamma, \mu_\beta\mu_\alpha\Gamma] \circ [\Gamma, \mu_\alpha\Gamma] = [\mu_\beta\Gamma, \mu_\alpha\mu_\beta\Gamma] \circ [\Gamma, \mu_\beta\Gamma] \tag{2.31}$$
where we used implicitly the natural identification $E(\Gamma) \to E(\mu_\alpha\Gamma)$,
3. *For any $\alpha, \beta \in E_i(\Gamma)$ such that $\epsilon^{\alpha\beta} = \pm 1$, the triangulations $\mu_\alpha\mu_\beta\Gamma$ and $\mu_\beta\mu_\alpha$ are related by a flip and:*
$$[\Gamma, \mu_\beta\Gamma] \circ [\mu_\alpha\mu_\beta\Gamma, \mu_\beta\Gamma] \circ [\mu_\beta\mu_\alpha\Gamma, \mu_\alpha\mu_\beta\Gamma] \circ [\mu_\alpha\Gamma, \mu_\beta\mu_\alpha\Gamma] \circ [\Gamma, \mu_\alpha\Gamma] = \mathrm{Id}_{|\Gamma|} \ . \tag{2.32}$$
This is the pentagon identity that we already encountered in Figure 1.14.

We refer to [FG07, Appendix B] for the proof of this proposition as well as more details on the combinatorial description of $\mathcal{MCG}(S)$.

2.3 Teichmüller $\mathcal{X}$-space (Teichmüller space with holes)

Let S be a ciliated surface (hyperbolic and with $k \geq 1$). The Teichmüller space $\mathcal{T}(S)$ of S parametrizes equivalence classes of hyperbolic metrics on S such that the preimages of every cilium under the universal covering map $\mathbb{H} \longrightarrow S$ lie on $\partial_\infty\mathbb{H}$, and such that circles bounding holes are geodesics of finite length while segments of the boundary between adjacent cilia are geodesics of infinite length. The length of the geodesic corresponding to a hole can be zero, in which case the hole is actually a puncture. The Teichmüller space with holes $\mathcal{T}^x(S)$ is very similar to $\mathcal{T}(S)$: it is a ramified cover of it of degree 2^h, where h is the number of holes in S. The only difference between $\mathcal{T}(S)$ and $\mathcal{T}^x(S)$ is that in the latter, a point also encodes an orientation for each hole which is not a puncture: we will denote it $+$ or $-$ depending on whether it coincides with the orientation induced by the one of the surface, or not. Studying $\mathcal{T}^x(S)$ instead of $\mathcal{T}(S)$ allows for the construction of a very nice set of coordinates which behave as Y-patterns, as already advertised.

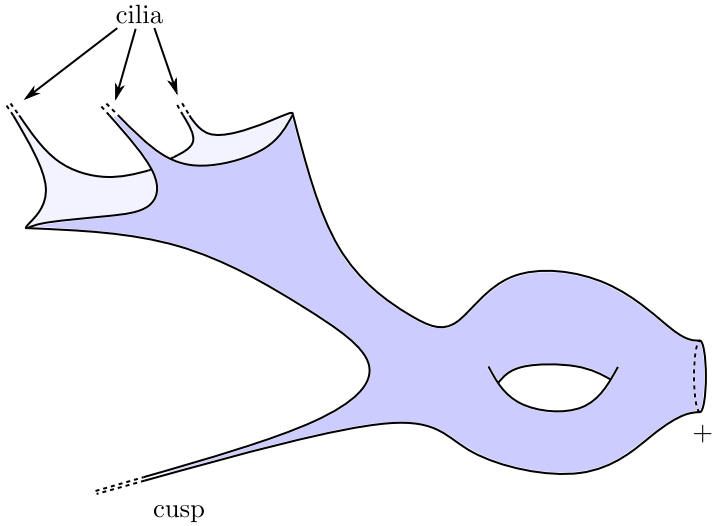


Figure 2.7: What a point in $\mathcal{T}^x(S_{1,2,\{3\}})$ corresponds to.

At a generic point of $\mathcal{T}^x(S)$, the geodesic bounding any hole is a finite non-zero length and hence if $S = S_{g,h,\{p_{h+1},\dots,p_k\}}$, there are 2^h possible choices of orientation for the holes. All of these project to the

same hyperbolic metric on S. The covering ramifies over the equivalence classes of hyperbolic metrics on S for which some holes are cusps. Hence $\mathcal{T}^x(S)$ is indeed a $2^h:1$ ramified cover over the space we are interested in. An illustration of how one may represent oneself what a point of $\mathcal{T}^x(S_{1,2,\{3\}})$ looks like is proposed in Figure 2.7. The three cilia are 'at infinity' and the segments on the boundary between cilia are geodesics of infinite length. There is one cusp and one hole which is not a cusp, and which is assigned the orientation $+$.

The coordinates on $\mathcal{T}^x(S)$ that we are going to describe are often called Thurston's *shear coordinates*, and are thourougly reviewed in [Pen87, Foc97, Thu98, FG07, BBFS09]. We follow the presentation of [FG07].

Let Γ be a triangulation of S. Shear coordinates, to which we will refer as $\mathcal{X}$-coordinates, are assigned to the internal edge of Γ. They parametrize $\mathcal{T}^x(S)$.

2.3.1 Parametrization

Coordinates on $\mathcal{T}^x(S)$ can be constructed as follows. At a point of $\mathcal{T}^x(S)$, the surface S is endowed with a hyperbolic structure with respect to which the edges of the triangulation Γ can be taken to be geodesics, as follows from Proposition 2.2. Note that with that hyperbolic structure, some holes of S are punctures while the others correspond to a unique geodesic of finite size.

Actually, one can easily assume an edge of Γ to be geodesic only if it connects two cilia, two punctures, or a puncture to a cilium – since the preimages of punctures and cilia for $\mathbb{H}\to S$ lie on $\partial_\infty\mathbb{H}$. As for the edges of Γ connecting a puncture or a cilium to a hole or two holes together, one proceeds as follows. For each hole one considers the unique geodesic surrounding it, and after having chosen an arbitrary point on it one deforms the edge under consideration to a geodesic there. Then, one rotates this point along the geodesic in the direction prescribed by the orientation of the hole, infinitely many times. This yields an limiting geodesic of infinite length, corresponding to the edge of Γ.

After having done this at every hole of S, each face of Γ has become an ideal hyperbolic triangle, i.e. a hyperbolic triangle with geodesic sides of infinite length homeomorphic to $\mathbb{R}$.

This yields a triangulation of the *convex core* of S. Starting from any ciliated surface S endowed with a hyperbolic metric, one can consider the completion $\tilde{S}$ of S from which the convex core of S is obtained as follows: each hole defines a free homotopy class, in which there is a unique geodesic from Proposition 2.2. Cut $\tilde{S}$ along this geodesic for each hole, and remove the semi-infinite throats created by the cuts. What remains is the convex core of S. Therefore, $\mathcal{T}^x(S)$ describes equivalence classes of hyperbolic metrics on S such that the boundary components corresponding to holes are geodesics of finite length (and zero length for the cusps, in which case the boundary component is at infinity). One may however equivalently think of $\mathcal{T}^x(S)$ as describing complete hyperbolic metrics on S.

Two ideal hyperbolic triangles can be smoothly glued along an edge in $\mathbb{R}$ different ways: one can shear one triangle with respect to the other along their common edge. Let us lift the triangulation Γ of S to the universal cover $\mathbb{H}$, and consider two adjacent triangles of this lift. The vertices of this triangulated quadrilateral are four distinct points on the boundary at infinity $x_1,x_2,x_3,x_4\in\partial_\infty\mathbb{H}=\mathbb{P}^1(\mathbb{R})=\mathbb{R}\cup\{\infty\}$ – say that the two triangles have vertices at $x_1<x_2<x_3$ and $x_1<x_3<x_4$, respectively. There exists a unique element in $\mathrm{PSL}_2(\mathbb{R})$ which maps (x_1,x_2,x_3) to $(\infty,-1,0)$ since the action of $\mathrm{PSL}_2(\mathbb{R})$ on $\partial_\infty\mathbb{H}$ is 3-transitive, as follows from the discussion in Section 2.1.1. Let $z\in\mathbb{R}$ be the image of x_4 under it, as shown in Figure 2.8: it is the shear parameter associated with this gluing. Since $\mathrm{PSL}_2(\mathbb{R})$ preserves the cyclic ordering of triples of points on $\mathbb{P}^1(\mathbb{R})$ one has $z\in\mathbb{R}_{>0}$.

Since the element of $\mathrm{PSL}_2(\mathbb{R})$ mapping (x_1,x_2,x_3) to $(\infty,-1,0)$ is unique, z is uniquely defined by the ordered quadruple (x_1,x_2,x_3,x_4) and the choice of a first element x_1. One may as well take x_3 as first element and consider the quadruple (x_3,x_4,x_1,x_2). Then, it is the triangle with vertices x_3,x_4,x_1 which is mapped to $\infty,-1,0$ while the second one has vertices at x_1,x_2,x_3. One can check easily that the shear parameters in both cases actually coincide. Hence the shear parameter we have defined is $\mathrm{PSL}_2(\mathbb{R})$-invariant and depends only on the vertices x_1,x_2,x_3,x_4 of the quadrilateral, and on the way it is triangulated.

Explicitly, one has:

$$z=-\frac{(x_1-x_2)(x_3-x_4)}{(x_2-x_3)(x_4-x_1)}\ , \tag{2.33}$$

and indeed z can be intrinsically defined as (minus) the cross-ratio of the four lines $x_1,x_2,x_3,x_4\in\mathbb{P}^1(\mathbb{R})$. If x_1,x_2,x_3 and x_4 are pairwise distinct, $z\in\mathbb{R}_{>0}$.

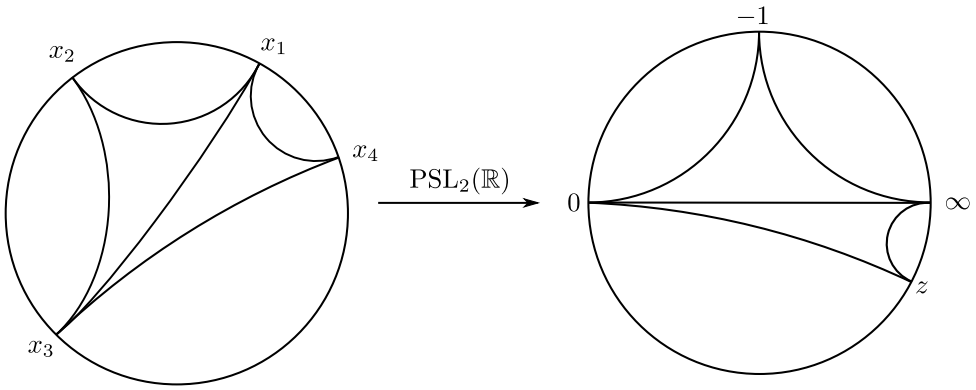


Figure 2.8: A normal form for a triangulated quadrilateral in $\mathbb{H}$.

Since $\mathbb{H} \to S$ is a local isometry, for any internal edge $e \in E_i(\Gamma)$ one can assign a shear parameter which is the one corresponding to a lift (equivalently, all of them) of e to $\mathbb{H}$. Shear coordinates corresponding to the triangulation Γ can be seen as a map

$$\mathcal{X}_\Gamma : (\mathbb{R}_{>0})^{E_i(\Gamma)} \longrightarrow \mathcal{T}^x(S) \ . \tag{2.34}$$

The coordinates on the internal edges of Γ can be expressed as the monodromy of an abelian connection on an auxiliary bipartite graph Λ_B on S, defined as follows. There is one white vertex for each vertex of Γ, and one black vertex for each face of Γ. There is an edge between a white node and a black one in Λ_B if the vertex of Γ corresponding to the white node belongs to the face of Γ corresponding to the black one. An example for a triangulation of the pentagon $S_{0,0,5}$ is shown in Figure 2.9.

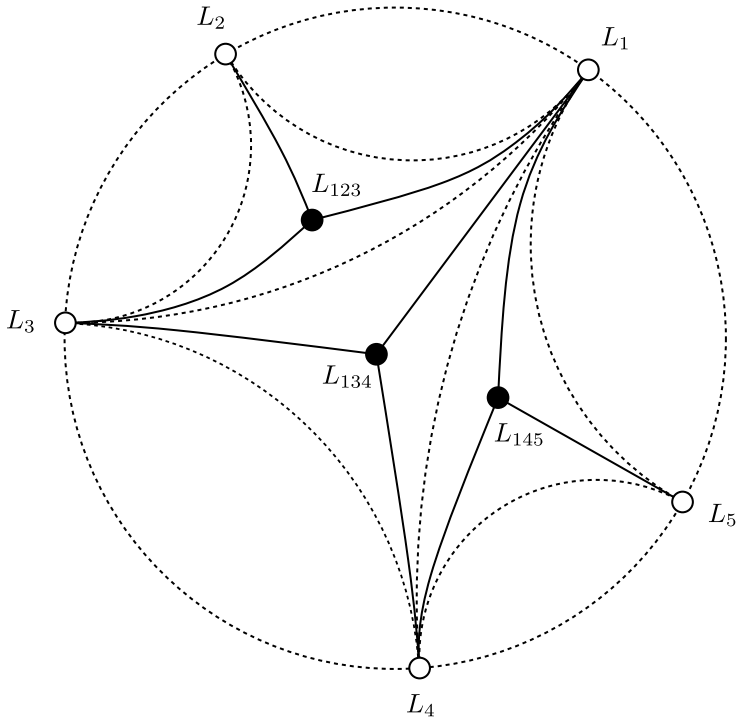


Figure 2.9: A bipartite graph (plain lines) from a triangulation of $S_{0,0,5}$ (dashed lines).

Likewise, one can construct a bipartite graph from the lift of Γ to the universal cover $\mathbb{H}$ of S: it is the lift of Λ_B, that we will still denote in the same way. In the universal cover, each white vertex of Λ_B is on the boundary of $\mathbb{H}$ and hence it corresponds naturally to a point in $\mathbb{P}^1(\mathbb{R})$, i.e. to a line in $\mathbb{R}^2$. These are denoted $L_1, \ldots, L_5$ in Figure 2.9. Each black node of Λ_B, or rather, its lift, can also be associated to a line in $\mathbb{R}^2$. Let L_i, L_j and L_k be the lines corresponding to the white nodes connected to it, with

$i < j < k$. These three lines are in $\mathbb{R}^2$, and hence generically the map

$$L_i \oplus L_j \oplus L_k \longrightarrow \mathbb{R}^2 \tag{2.35}$$

has a one-dimensional kernel L_{ijk}, which we associate to the corresponding black vertex, as in Figure 2.9. The collection of real lines at white and black vertices of Λ_B can be considered as a line bundle on Λ_B. Choosing a trivialization of it amounts to pick a non-zero vector in the line at each vertex of Λ_B. Upon such a choice, the natural maps $L_{ijk} \to L_i$, $L_{ijk} \to L_j$ and $L_{ijk} \to L_k$ are represented as non-zero real numbers that can be assigned to the edges of Λ_B with the convention that every edge is oriented from black to white. Hence we obtain an element of $(\mathbb{R}^\times)^{E(\Lambda_B)}$.

Changing the trivialization amounts to an action of $(\mathbb{R}^\times)^{V(\Lambda_B)}$ on $(\mathbb{C}^\times)^{E(\Lambda_B)}$. This is a (discrete) gauge action, and the space of configurations modulo gauge is naturally identified with

$$H^1(\Lambda_B, \mathbb{R}^\times) \ , \tag{2.36}$$

which in turns corresponds to assigning a non-zero real number to each edge of Γ since faces of Λ_B correspond to quadruples of white nodes forming a quadrilateral of Γ, and one can show that the monodromy is minus the cross-ratio of the corresponding four lines. As before, if $L_i < L_j < L_k < L_l \in \mathbb{P}^1(\mathbb{R})$, the monodromy corresponding to this quadruple is in $\mathbb{R}_{>0}$.

This way of computing $\mathcal{X}$-coordinates will be generalized in Section 4.3.3.

2.3.2 Reconstruction

Let us construct the reciprocal of the map of Equation (2.34). More precisely, we are going to explain how one constructs a Fuchsian model $G < \mathrm{PSL}_2(\mathbb{R})$ of S from an element $(x^\alpha)_{\alpha \in E_i(\Gamma)} \in (\mathbb{R}_{>0})^{E_i(\Gamma)}$ such that this is indeed the reciprocal of Equation (2.34).

One wishes to associate an element of $\mathrm{PSL}_2(\mathbb{R})$ to every closed path on S, using the fact that G acts as deck transformations of the covering $\mathbb{H} \to S$. Every such element will be defined as a product of the matrices I, I^{-1} and $B(x^\alpha)$ for $x^\alpha \in \mathbb{R}_{>0}$, where:

$$I = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}, \quad B(x^\alpha) = \begin{pmatrix} 0 & (x^\alpha)^{1/2} \\ -(x^\alpha)^{-1/2} & 0 \end{pmatrix} \ . \tag{2.37}$$

Note that both are matrices in $\mathrm{PSL}_2(\mathbb{R})$. One the one hand, $I \in \mathrm{PSL}_2(\mathbb{R})$ maps $(\infty, -1, 0)$ to $(-1, 0, \infty)$ when acting on $\partial_\infty H$. On the other hand, $B(x^\alpha)$ maps $(\infty, -1, 0)$ to $(0, x^\alpha, \infty)$. In terms of ideal hyperbolic triangles, I is a counterclockwise rotation of angle $2\pi/3$ of the triangle with vertices $(\infty, -1, 0)$, while $B(x^\alpha)$ maps the latter to the triangle with vertices $(0, x^\alpha, \infty)$.

Let Δ and Δ' be two neighbor triangles of the lift of Γ to $\mathbb{H}$ sharing a common edge. Suppose that the latter carries the coordinate $x^\alpha \in \mathbb{R}_+^*$. There exists a unique $M_\Delta \in \mathrm{PSL}_2(\mathbb{R})$ such that $M_\Delta \Delta$ is the triangle with vertices $(\infty, -1, 0)$ and $M_\Delta \Delta'$ is the triangle with vertices $(0, x^\alpha, \infty)$. Thus, the unique element of $\mathrm{PSL}_2(\mathbb{R})$ which maps Δ to Δ' is

$$M_\Delta^{-1} B(x^\alpha) M_\Delta \ , \tag{2.38}$$

and the unique one which rotates cyclically the vertices of Δ counterclockwise is:

$$M_\Delta^{-1} I M_\Delta \ . \tag{2.39}$$

Let γ be a closed loop on S based at a point in the interior of a face of the triangulation Γ, and lift it to a path $\tilde{\gamma}$ in $\mathbb{H}$. This path starts and ends in the interior of faces of the lift $\tilde{\Gamma}$ of Γ to $\mathbb{H}$. Possibly after an isotopy of $\tilde{\gamma}$ in $\mathbb{H}$, one can assume that $\tilde{\gamma}$ intersects the edges of $\tilde{\Gamma}$ transversely. Let $\Delta_1, \dots, \Delta_k, \dots, \Delta_N$ be the faces of $\tilde{\Gamma}$ met along $\tilde{\gamma}$. Since γ is closed, Δ_1 and Δ_N are two lifts of the same face of Γ. We are going to compute a sequence of elements $\mathrm{Id} = M_1, M_2', M_2, \dots, M_N', M_N \in \mathrm{PSL}_2(\mathbb{R})$ such that for $k \in [|2, N|]$ both M_k and M_k' map Δ_1 to Δ_k, though not in the same way.

Up to $\mathrm{PSL}_2(\mathbb{R})$ one can assume that Δ_1 has vertices $(\infty, -1, 0)$ and that the edge through which $\tilde{\gamma}$ exits this face is the one opposite to -1. This is illustrated in Figure 2.10. One starts with $M_1 = \mathrm{Id}$, which trivially maps $(\Delta_1, \clubsuit)$ to $(\Delta_1, \clubsuit)$ in a trivial way ($\clubsuit$ marks a vertex, as in the figure).

As $\tilde{\gamma}$ goes through the edge shared by Δ_1 and Δ_2 one considers the matrix $M_2' = B(x^\alpha)$. It maps $(\Delta_1, \clubsuit)$ to $(\Delta_2, \clubsuit)$ in the only orientation-preserving way.

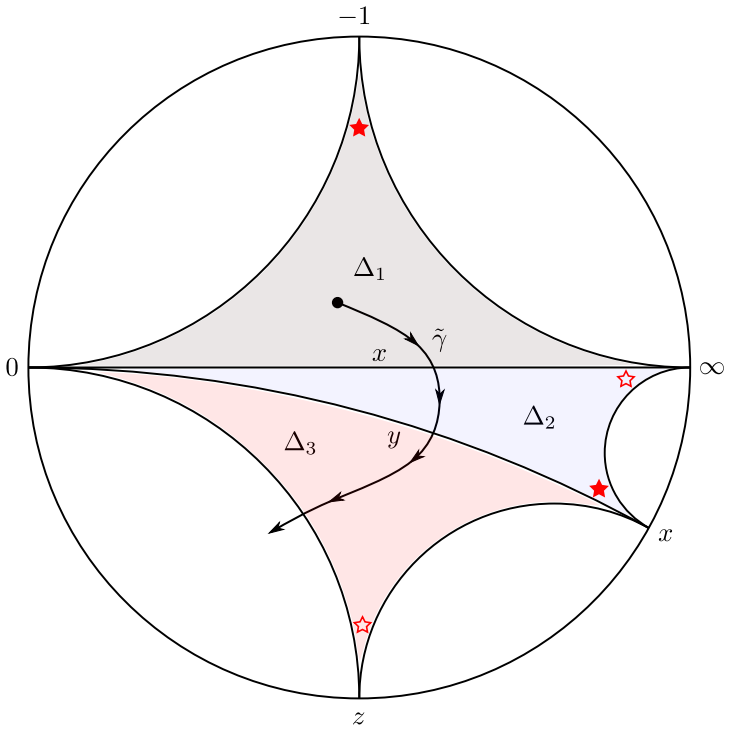


Figure 2.10: From shear coordinates to monodromies.

At this point, one would like to rotate Δ_2 in such a way that $\tilde{\gamma}$ exits it through the edge opposite to the starred vertex. From Equation (2.39) one obtains:

$$M_2 = M_2' I^b \left(M_2'\right)^{-1} M_2' = B(x^\alpha) I^b \ , \tag{2.40}$$

where $b = 0, 1$ or 2 depending on through which edge $\tilde{\gamma}$ exits. In Figure 2.10 one has $b = 1$, and the matrix $B(x^\alpha)I$ maps $(\Delta_1, \clubsuit)$ to $(\Delta_2, \spadesuit)$.

As $\tilde{\gamma}$ crosses the edge β shared by Δ_2 and Δ_3 and corresponding to the shear parameter x^β, Equation (2.38) implies that

$$M_3' = M_2 B(x^\beta) M_2^{-1} = M_3' = B(x^\alpha) I^b B(x^\beta) \in \mathrm{PSL}_2(\mathbb{R}) \ . \tag{2.41}$$

In Figure 2.10, the matrix M_3' maps $(\Delta_1, \clubsuit)$ to $(\Delta_3, \spadesuit)$. One continues likewise until Δ_N is reached.

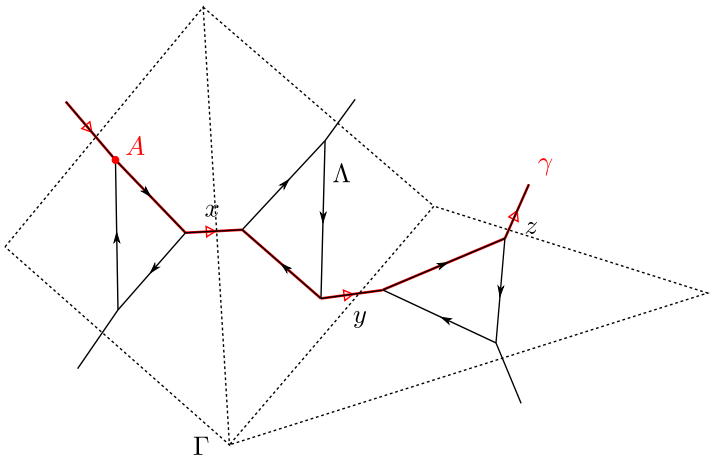


Figure 2.11: The graph Λ constructed from the triangulation Γ of S.

Building on this, there is a convenient way to compute the monodromy $M_\gamma \in \mathrm{PSL}_2(\mathbb{R})$ of a closed loop $\gamma \in \pi_1(S)$ corresponding to a point in $\mathcal{T}^x(S) \simeq (\mathbb{R}_{>0})^{E_i(\Gamma)}$. Let us first define again another graph Λ on S from the triangulation Γ: one draws a small edge for Λ transverse to each edge of Γ, and connects the endpoints of these small edges by clockwise oriented edges in each face of Γ. This construction is displayed in Figure 2.11.

Let γ be a closed loop on S representing $[\gamma] \in \pi_1(S)$. One can isotope it to an oriented loop on Λ, which can moreover be chosen with the minimal possible edge-length. Let us take an arbitrary vertex of this path on Λ as starting point, and construct inductively a matrix in $\mathrm{PSL}_2(\mathbb{R})$ from the identity element, by following the oriented path on Λ isotopic to γ.

1. Multiply on the right by $B(x^\alpha)$ for each edge of Λ transverse to an edge of Γ labeled by a shear parameter $x^\alpha \in \mathbb{R}_+^*$. Note that $B(x^\alpha)$ is its own inverse in $\mathrm{PSL}_2(\mathbb{R})$ and thus the edges of Λ transverse to those of Γ do not need to be oriented.
2. Multiply on the right by I (respectively, I^{-1}) for each edge belonging to a small triangle in Λ if the orientation of the path agrees (respectively, disagrees) with the orientation of the small edge.

For example, an oriented path on Λ is shown on Figure 2.11 as the thick red edges, with the arbitrary starting point denoted A. The first steps of the construction of the monodromy matrix corresponding to this path yield

$$M = IB(x)I^{-1}B(y)IB(z)\dots \tag{2.42}$$

2.3.3 Flips, hole lengths and Poisson structure

Coordinate transformation under a flip. Given two triangulations Γ_1 and Γ_2 of Σ, one would like to relate the set of shear coordinates corresponding to Γ_1 and the one corresponding to Γ_2. From Theorem 2.15 we know that there exists a finite sequence of flips through with Γ_1 is transformed to Γ_2, and hence we wish to understand how the shear coordinates vary under a flip. In the case of two triangles glued along a single edge the coordinate change under the flip of that edge is given in Figure 2.12. All other shear coordinates not displayed in the figure remain the same.

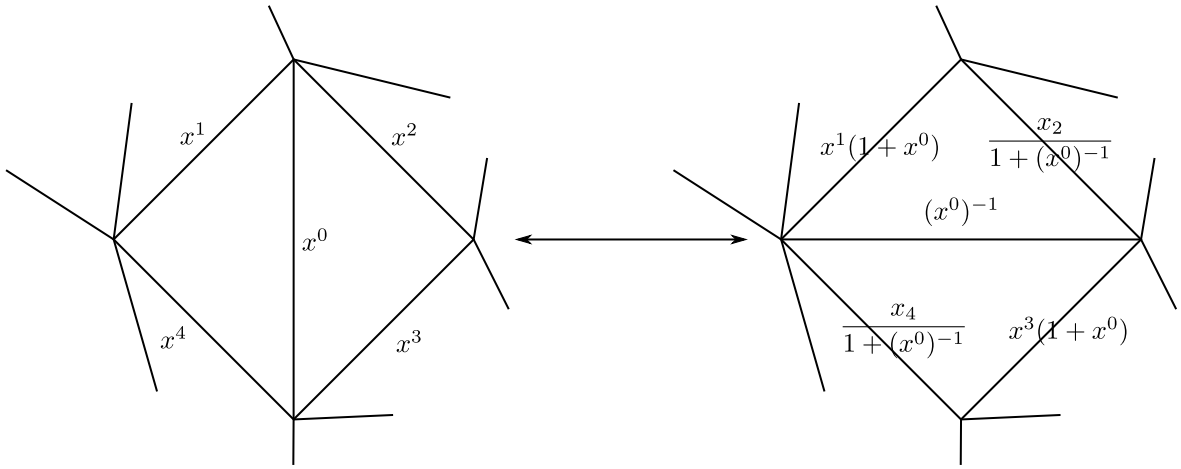


Figure 2.12: The $\mathcal{X}$-coordinates change rule under a flip.

In general, if $\left(\epsilon^{\alpha\beta}\right)_{\alpha,\beta\in E(\Gamma)}$ is the extended exchange matrix of the triangulation Γ, the mutation formula under the flip of the edge α is exactly the Y-pattern mutation formula, from now on called the *cluster $\mathcal{X}$-mutation formula* of Equation (1.68):

$$\mu_\alpha(x^\beta) = \begin{cases} (x^\alpha)^{-1} & \text{if } \beta = \alpha \\ (x^\beta)(1+x^\alpha)^{\epsilon^{\alpha\beta}} & \text{if } \epsilon^{\alpha\beta} \geq 0 \\ (x^\beta)(1+(x^\alpha)^{-1})^{\epsilon^{\alpha\beta}} & \text{if } \epsilon^{\alpha\beta} \leq 0 \end{cases} . \tag{2.43}$$

As already emphasized, these formulae are subtraction-free, hence if Γ and Γ' are two triangulations of the same ciliated surface related by a flip at $e \in E_i(\Gamma)$ these mutation rules define a map

$$\mu_e : \mathbb{R}_+^{*\,E_i(\Gamma)} \to \mathbb{R}_+^{*\,E_i(\Gamma')} . \tag{2.44}$$

Moreover, if e' is the flipped edge in Γ' one can easily check that $\mu_{e'} : {\mathbb{R}_+^*}^{E_i(\Gamma')} \longrightarrow {\mathbb{R}_+^*}^{E_i(\Gamma)}$ is the reciprocal of μ_e. This mutation rule is the transition functions from the chart on $\mathcal{T}^x(S)$ corresponding to Γ to the one corresponding to Γ':

$$\begin{array}{ccc} {\mathbb{R}_+^*}^{E_i(\Gamma)} & & \\ \uparrow \mu_{e'} \quad \downarrow \mu_e & \searrow^{\mathcal{X}_\Gamma} & \\ & & \mathcal{T}^x(S) \\ & \nearrow_{\mathcal{X}_{\Gamma'}} & \\ {\mathbb{R}_+^*}^{E_i(\Gamma')} & & \end{array} \tag{2.45}$$

Hole lengths. For any hole ρ in S, let

$$r^\rho : \mathcal{T}^x(S) \longrightarrow \mathbb{R}_{>0} \tag{2.46}$$

be the map given by the exponent of the length (resp. minus the exponent of the length) of the geodesic surrounding ρ if the orientation of ρ coincides (resp. disagrees) with the orientation of Σ. In terms of shear coordinates, a direct computation shows that the map r^ρ is:

$$r^\rho = \prod_\alpha x^\alpha \ , \tag{2.47}$$

where the product runs over all edges incident to the hole ρ, counted with multiplicity. This follows from the reconstruction described in the previous subsection: the monodromy of a simple loop going around a hole ρ can be written as

$$M_\rho = \prod_\alpha I^b B(x^\alpha) \ , \tag{2.48}$$

where b is either 1 or -1, and does not depend on i. The terms $(I^b B(x^\alpha))$ appearing in the product are either all upper-triangular or all lower-triangular, depending on the value of b, and hence one writes easily M_ρ. For example if $b = 1$:

$$M_\rho = \begin{bmatrix} \prod_\alpha (x^\alpha)^{-1/2} & -\prod_\alpha (x^\alpha)^{-1/2} \\ 0 & \prod_\alpha (x^\alpha)^{1/2} \end{bmatrix} . \tag{2.49}$$

The length of the geodesic preserved by M_ρ, which is the one bounding the hole ρ, is given by the absolute value of the logarithm of the ratio of the two eigenvalues of M_ρ, which implies Equation (2.47).

A Poisson structure. The *log-canonical Poisson structure* on $(\mathbb{R}_{>0})^{E_i(\Gamma)}$ is defined by

$$\{f, g\} = \sum_{\alpha,\beta \in E_i(\Gamma)} \epsilon^{\alpha\beta} \ x^\alpha x^\beta \frac{\partial f}{\partial x^\alpha} \frac{\partial g}{\partial x^\beta} \ , \tag{2.50}$$

where f, g are functions on $(\mathbb{R}_{>0})^{E_i(\Gamma)}$. The main property of this Poisson bracket is that it is independent of the triangulation Γ, as follows from a direct calculation using Equation (2.43) and Equation (1.14). Hence it defines an intrinsic Poisson bracket on $\mathcal{T}^x(S)$. This Poisson bracket is a prominent feature of cluster $\mathcal{X}$-varieties defined in Chapter 4.

2.4 Teichmüller $\mathcal{A}$-space (decorated Teichmüller space)

Let S be a ciliated surface (hyperbolic and with $k \geq 1$). The decorated Teichmüller space $\mathcal{T}^a(S)$ of S describes hyperbolic structures on S such that holes are all punctures, together with horocycles at the cilia and the punctures. It contains more information than the usual Teichmüller space $\mathcal{T}(S)$, however it

projects naturally to it. As in the last section, one of the nicest properties of this version of the Teichmüller space is that it admits a nice parametrization. This construction is due to Penner [Pen87, Pen92].

Horocycles, as introduced in Section 2.1.1, can also be defined on hyperbolic ciliated surfaces. Let $\mathbb{H}/\Delta$ be a Fuchsian model of S. A horocycle at a puncture or a cilium A in S is a Δ-invariant set consisting of one horocycle at each preimage of A in $\mathbb{H}$. If the horocycles are small enough in $\mathbb{H}$, their image in S is a small curve close to A, to which all the geodesics emanating at A are orthogonal. However, if the horocycles in $\mathbb{H}$ are too big, their image is a curve on S winding non-trivially. Two horocycles on S are said to be *tangent* along a path γ connecting their base points if the corresponding horocycles at the ends of a (equivalently, any) lift of γ to $\mathbb{H}$ are tangent.

Definition 2.20. *A decorated ciliated hyperbolic surface is a ciliated surface endowed with a hyperbolic metric such that every hole is a puncture, and with a horocycle chosen at each puncture and cilium. The Teichmüller space of decorated ciliated surfaces $\mathcal{T}^a(S)$ is called the decorated Teichmüller space.*

An illustration of how one may represent oneself what a point on $\mathcal{T}^a(S_{1,2,\{3\}})$ looks like is proposed in Figure 2.13. Cilia are labeled $1, 2, 3$ and their corresponding horocycles, H_1, H_2, H_3; the cusps are labeled A, B and their corresponding horocycles, H_A, H_B.

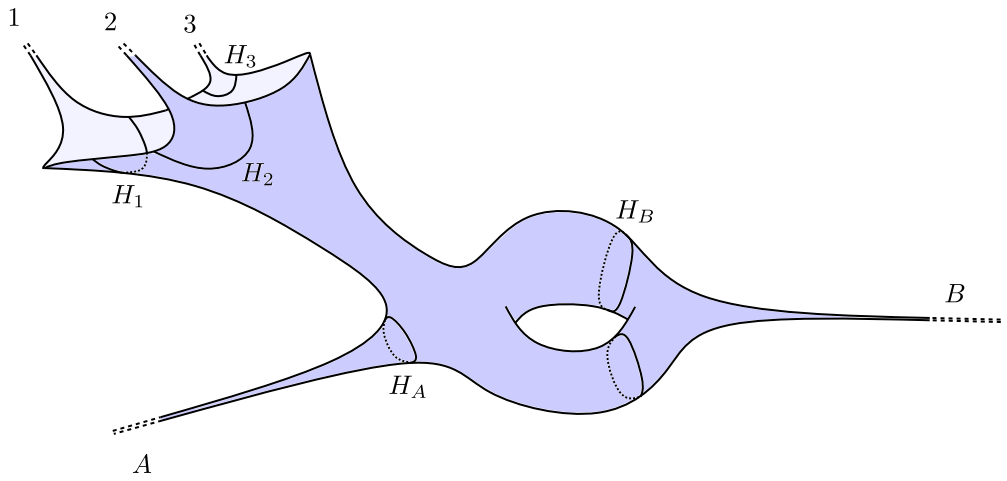


Figure 2.13: What a point in $\mathcal{T}^a(S_{1,2,\{3\}})$ corresponds to.

Clearly, if one forgets the horocycles one is left with a hyperbolic structure on S, which corresponds to a special point in $\mathcal{T}^x(S)$ at which the holes all have zero length. In other words, there is a natural map:

$$p : \mathcal{T}^a(S) \longrightarrow \mathcal{T}^x(S) \ . \tag{2.51}$$

Let us emphasize that it is in general neither surjective (since the only points of $\mathcal{T}^x(S)$ in its image are those at which the holes are all punctures) nor injective (since two points in $\mathcal{T}^a(S)$ which differ only by the choice of horocycles are mapped to the same point in $\mathcal{T}^x(S)$).

The space $\mathcal{T}^a(S)$ has the canonical subspace $\mathcal{T}^a_0(S)$, which consists of all decorated ciliated Riemann surfaces such that for each path connecting to adjacent cilia (hence retractable to the boundary) the horocycles at its two ends are tangent. For every cilium and for every hole ρ there is an action r_ρ of $(\mathbb{R}_{>0}, \times)$ on $\mathcal{T}^a(S)$:

$$r_\rho : \mathbb{R}_{>0} \times \mathcal{T}^a(S) \longrightarrow \mathcal{T}^a(S) \ . \tag{2.52}$$

which corresponds to changing the size of the corresponding horocycle.

Let Γ be a triangulation of S. Assigning a strictly positive real number to every edge of Γ (and not only the internal ones) provides the interesting parametrization of $\mathcal{T}^a(S)$ we are after.

2.4.1 Parametrization

Let us fix a point of $\mathcal{T}^a(S)$. Let α be any edge of Γ that we can assume to be geodesic, as follows from Proposition 2.2, and let $\tilde{\alpha}$ be any lift of α to the universal cover $\mathbb{H}$. There is a horocycle in $\mathbb{H}$ based at

each of the endpoints of $\tilde{\alpha}$. Let l be the length of the segment in $\tilde{\alpha}$ between the intersection points with each of these horocycles (any other lift of α yields the same l). The coordinate one assigns to α is

$$a_\alpha = e^{\pm l/2} \ , \tag{2.53}$$

with a plus (respectively, minus) sign when the two horocycles do not (respectively, do) intersect, as shown in Figure 2.14.

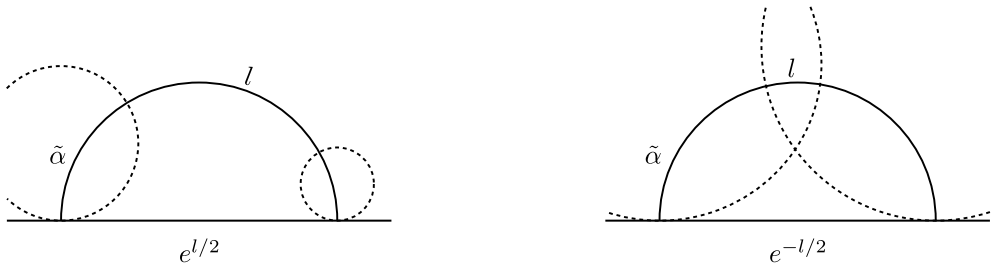


Figure 2.14: The definition of $\mathcal{A}$-coordinates.

The coordinate a_α can be computed easily when the horocycles at the two ends of $\tilde{\alpha}$ are represented as elements of $(\mathbb{R}^2 - \{(0,0)\}/\pm 1)$ as in Proposition 2.5: if the two horocycles are represented by (x_1, y_1) and (x_2, y_2), then

$$a_\alpha = \left| \det \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} \right| \ . \tag{2.54}$$

The coordinates $(a_\alpha)_{\alpha \in E(\Gamma)}$ corresponding to the triangulation Γ can be seen as a map

$$\mathcal{A}_\Gamma : (\mathbb{R}_{>0})^{E(\Gamma)} \longrightarrow \mathcal{T}^a(S) \ . \tag{2.55}$$

The points lying in the subspace $\mathcal{T}_0^a(S)$ clearly correspond to those whose $\mathcal{A}$-coordinates on external edges are 1's.

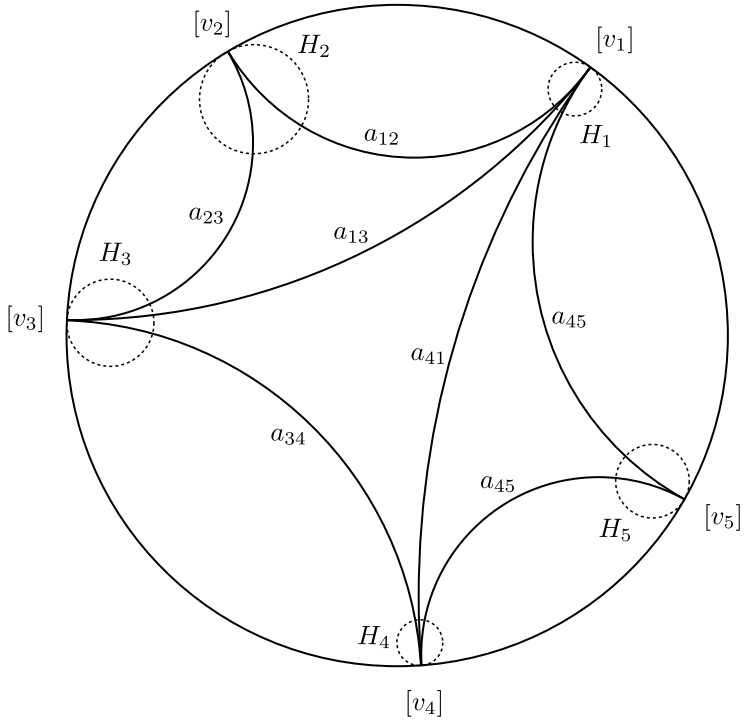


Figure 2.15: Constructing coordinates on the decorated Teichmüller space.

Similarly to the computation of $\mathcal{X}$-coordinates as the monodromy of an abelian connection on a graph, let us reformulate the definition of $\mathcal{A}$-coordinates in a way similar to the construction of $\mathcal{A}$-coordinates on higher $\mathcal{A}$-Teichmüller spaces presented in Section 4.3.4. We follow [FG06, Section 11].

Let us consider the lift of Γ with geodesic edges to $\mathbb{H}$. The endpoints of these edges lie on the boundary $\mathrm{P}^1(\mathbb{R})$, and there is a horocycle in $\mathbb{H}$ at each such endpoint, to which one can associate a vector (up to sign) in $\mathbb{R}^2 - \{(0,0)\}/\pm 1$, as follows from Proposition 2.5. The line in $\mathbb{R}^2$ generated by one such vector is the point in $\mathbb{P}^1(\mathbb{R})$ at which the horocycle is. The example of the pentagon $S_{0,0,\{5\}}$ is shown in Section 2.4.1, with $[v_1], \dots, [v_5] \in (\mathbb{R}^2 - \{(0,0)\}/\pm 1)$ the vectors (up to sign) corresponding to the horocycles.

The $\mathcal{A}$-coordinates can be computed as before; for example, in the case of the pentagon one has

$$a_{ij} = |[v_i] \wedge [v_j]| \ . \tag{2.56}$$

There is also another way to compute the $\mathcal{A}$-coordinates which trades the configuration of the sign-ambiguous $[v_i]$'s for a *twisted configuration* of ordinary vectors. For definiteness we will stick to the case of the pentagon. Let $L_1, \dots, L_5$ be the lines in $\mathbb{R}^2$ generated by $[v_1], \dots, [v_5]$. Pick any line $L \subset \mathbb{R}^2$ which does not contain any v_i for $i = 1, \dots, 5$, and fix one of the complements of $\mathbb{R}^2 - L$. For each $i = 1, \dots, 5$ it contains either v_i or $-v_i$. Then, one associates the line L_i together with the vector $w_i = \pm v_i \in L_i$, depending on the choice of L, to the corresponding point on $\mathbb{P}^1(\mathbb{R})$. One example is shown in Figure 2.16, which determines the choice $w_5 = -v_5, w_1 = v_1, w_2 = v_2, w_3 = v_3, w_4 = v_4$.

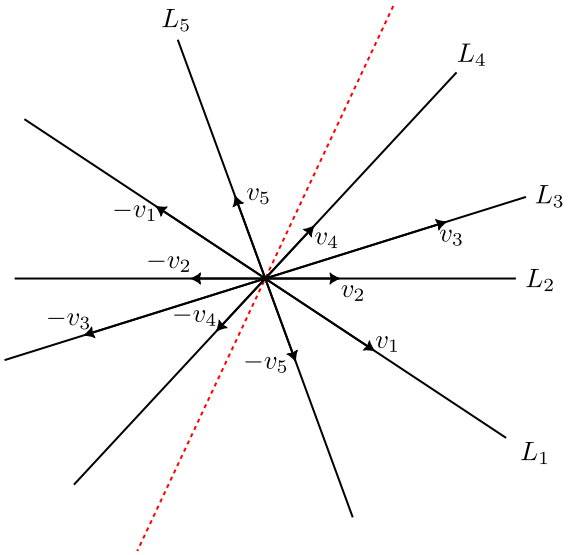


Figure 2.16: Constructing a twisted configuration of vectors.

The computation of $\mathcal{A}$-coordinates goes as follow. Let $i, j \in [|1, 5|]$ such that there is an edge of the lift of Γ connecting the two corresponding points on the boundary of $\mathbb{H}$. Then:

$$a_{ij} = w_i \wedge \epsilon(w_i, j) v_j \ , \tag{2.57}$$

where $\epsilon(w_i, j)$ has value 1 (resp. -1) if the representative of $[v_j]$ directly after w_j as one goes counter-clockwise in Figure 2.16 starting at w_j is v_j (resp. $-v_j$). This definition of the coordinates is independent of the choice of i among i, j. For example:

$$a_{45} = v_4 \wedge v_5 = (-v_5) \wedge v_4 \ . \tag{2.58}$$

It also does not depend on the choice of the line L.

2.4.2 Reconstruction

Proposition 2.21. *In coordinates, the canonical map of Equation* (2.51) *is:*

$$x^\alpha = \prod_\beta (a_\beta)^{-\epsilon^{\alpha\beta}} \ , \tag{2.59}$$

where the sum runs over all edges of Γ*, and where* $\epsilon^{\alpha\beta}$ *is the exchange matrix of* Γ*.*

Proof. Consider an ideal triangulated quadrilateral in $\mathbb{H}$ with edges $\alpha, \beta, \gamma, \delta$, diagonal α and vertices A, B, C and D a shown in Figure 2.17. Assume that a horocycle H_A (respectively H_B, H_C and H_D) has been fixed at A (respectively B, C and D).

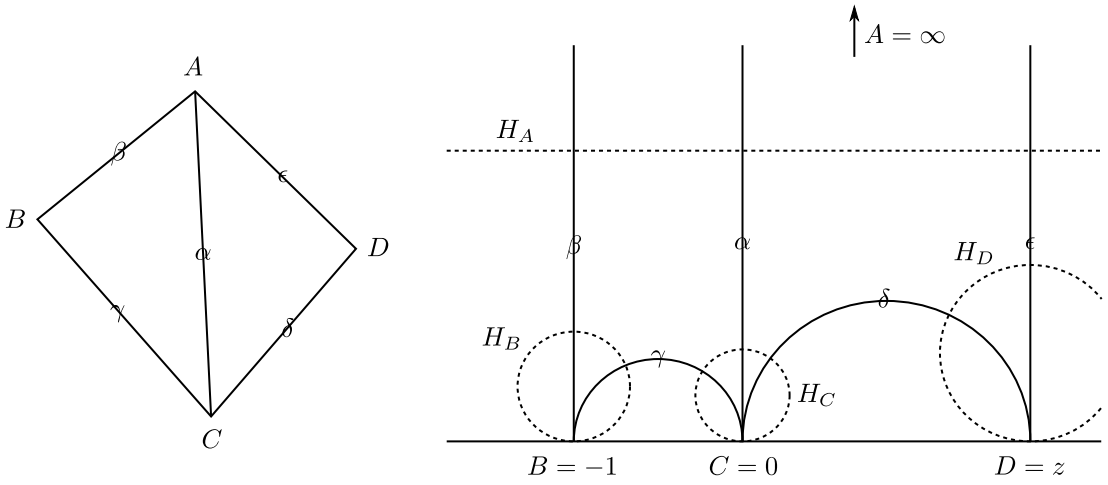


Figure 2.17: Proof of the explicit expression of the canonical map $\mathcal{T}^a(S) \longrightarrow \mathcal{T}^x(S)$.

Up to $\mathrm{PSL}_2(\mathbb{R})$ one can assume that $A = \infty$, $B = -1$ and $C = 0$. Let $z \in \mathbb{R}_{>0}$ be the coordinate of $D \in \partial_\infty \mathbb{H}$, as in Figure 2.17.

Let (x_A, y_A) (respectively $(x_B, y_B), (x_C, y_C)$ and (x_D, y_D)) be the vectors corresponding to H_A (respectively H_B, H_C and H_D) as in Proposition 2.5. Clearly one has $y_A = 0$, $x_B = -y_B$, $x_C = 0$, and moreover one can assume that $x_A, x_B, y_C, x_D, y_D > 0$. It follows from Equation (2.54) that

$$a_\alpha = x_A y_C \ , \quad a_\beta = -x_A y_B \ , \quad a_\gamma = x_B y_C \ , \quad a_\delta = y_C x_D \ , \quad a_\epsilon = x_A y_D \ , \tag{2.60}$$

and since $z = x_D/y_D$ it implies that $z = (a_\delta a_\beta)(a_\epsilon a_\gamma)^{-1}$, which is what was expected. All other cases are proved in the same way. □

Given $(a_\alpha)_{\alpha \in E(\Gamma)} \in (\mathbb{R}_{>0})^{E(\Gamma)}$ one can thus reconstruct the hyperbolic structure on S as a point of $\mathcal{T}^x(S)$, using the reconstruction for $\mathcal{T}^x(S)$ presented in the last section.

The horocycles still need to be reconstructed. Let us consider a face of Γ with vertices $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ enumerated clockwise, and corresponding sides α, β and γ as in Figure 2.18.

Lift the triangulation Γ to the universal cover $\mathbb{H}$ and consider one of the lifts of the triangle $\tilde{A}\tilde{B}\tilde{C}$. Let A, B and C be its vertices, projecting to $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ respectively. One can consider a first coordinate z on $\mathbb{H}$ which is such that $A = \infty$, $B = 0$ and the horocycle at A is the horizontal line $\mathrm{Im} z = 1$. Let $x_C \in \mathbb{R}_{>0}$ be the coordinate of C in this frame. This is depicted in Figure 2.18.

Note that the affix of the intersection point E between the geodesic from B to A and the horocycle at B is ia_α^{-2}, since the hyperbolic distance between E and D (which has affix i) is $l_\alpha = 2 \log a_\alpha$. The element of $\mathrm{PSL}_2(\mathbb{R})$ which exchanges A and B and maps the horocycle at B to $\mathrm{Im} z' = 1$ in terms of the new coordinate z' is the matrix:

$$D(a_\alpha) = \begin{pmatrix} 0 & a_\alpha^{-1} \\ -a_\alpha & 0 \end{pmatrix} \ . \tag{2.61}$$

The horocycle at B in Figure 2.18 corresponds to the pair $(0, a_\alpha)$ while the one at C is $(a_\gamma x_C, a_\gamma)$, so that Equation (2.54) implies:

$$|x_C| = \frac{a_\beta}{a_\alpha a_\gamma} \ , \tag{2.62}$$

which, since $x_C \in \mathbb{R}_{<0}$, leads to:

$$x_C = -\frac{a_\beta}{a_\alpha a_\gamma} \ . \tag{2.63}$$

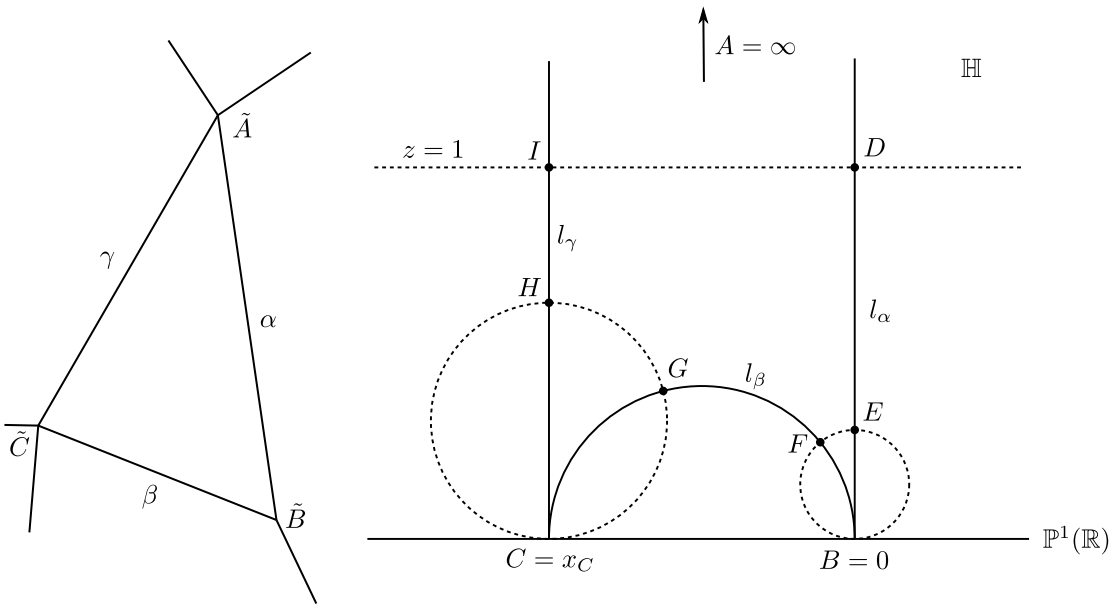


Figure 2.18: The reconstruction of horocycles from the data of Penner's coordinates.

The element in $\mathrm{PSL}_2(\mathbb{R})$ which fixes $B = 0$ and maps C to ∞ and the horocycle at C to the horizontal line $\mathrm{Im} z'' = 1$ in terms of the new coordinate z'' is

$$F\left(\frac{a_\alpha}{a_\beta}, a_\gamma\right) = \begin{pmatrix} \frac{a_\alpha}{a_\beta} & 0 \\ a_\gamma & \frac{a_\beta}{a_\alpha} \end{pmatrix} . \tag{2.64}$$

It maps 0 to itself, C to ∞ and ∞ to $a_\alpha(a_\gamma a_\beta)^{-1}$. Moreover, the affix of the point H in Figure 2.18 is $x_C + i a_\gamma^{-2}$, hence under the transformation of Equation (2.64) it is mapped to $i + a_\alpha(a_\gamma a_\beta)^{-1}$, which is the point on the geodesic linking ∞ to $a_\alpha(a_\gamma a_\beta)^{-1}$ with imaginary part 1. Hence the horocycle at C going through H is mapped to the horocycle at ∞ given by $\mathrm{Im} z'' = 1$, as wanted.

Hence, the data of $\mathcal{A}$-coordinates $(a_\alpha)_{\alpha \in E(\Gamma)}$ unambiguously determines the horocycles at the vertices. As in the (shear) $\mathcal{X}$-coordinates case, we can express the monodromies corresponding to a point in $\mathcal{T}^a(S)$ using an auxiliary graph Λ on S build from the triangulation Γ. The procedure is illustrated in Figure 2.19: the triangulation Γ is shown as dashed lines and Λ as plain lines.

The graph Λ is constructed from the triangulation Γ in the following way. On each edge α of Γ one assigns two vertices of Λ (one near each end of α). These are all the vertices of Λ, connected in the following way: each edge of Γ gives rise to an edge of Λ which connects the two vertices of Λ on it. Moreover, inside each triangle of Γ there are three additional clockwise oriented 'small' edges for Λ, in such a way that each vertex of Γ is connected to its neighbor as in Figure 2.19.

To the edge of Λ corresponding to an edge α of Γ with $\mathcal{A}$-coordinate a_α, one assigns the element $D(\alpha) \in \mathrm{PSL}_2(\mathbb{R})$. To each clockwise oriented edge in Λ one assigns $F(a_\alpha/a_\beta, a_\gamma)$. This is shown in Figure 2.19.

Let γ be an oriented closed loop γ on S, and let $a = (a_\alpha)_{\alpha \in E(\Gamma)} \in \mathcal{T}^a(S)$. Deform γ to a loop on Λ in such a way that its orientation coincides with the one on the small edges of Λ. Multiplying together the matrices one meets on the way (from left to right) yields the monodromy matrix along γ at a.

Using Λ one can also easily compute the distance between the horocycles at any two holes or cilia A and B. Let H_A and H_B be the horocycles at A and B corresponding to a point $a \in \mathcal{T}^a(S)$. Let γ be the unique geodesic with endpoints A and B, and let i_A (resp. i_B) be the intersection point $\gamma \cap H_A$ (resp. $\gamma \cap H_B$). Deform γ to a path on Λ from one of the neighboring vertices of A in Λ to one of the neighboring vertices of B, in such a way that its orientation coincides with the one on the small edges of Λ. The hyperbolic distance between i_A and i_B is the upper-right element of the product of the matrices corresponding to the elementary segments on the path from A to B on Λ. Since the F-matrices

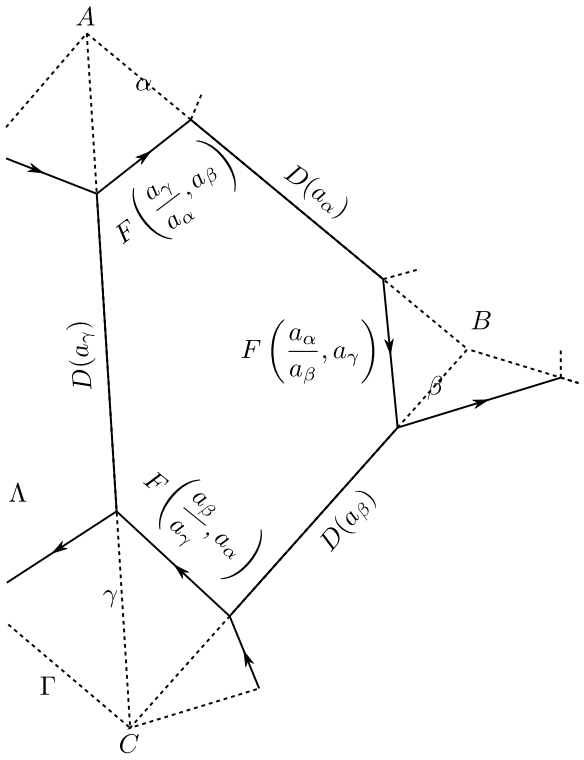


Figure 2.19: The auxiliary graph to compute the monodromies from Penner's coordinates.

of Equation (2.64) are lower-triangular, this element does not depend on the choice of endpoints for the path on Λ.

2.4.3 Flips and the closed 2-form

Mutation rule. The transformation of $\mathcal{A}$-coordinates under the flip of an edge $\alpha \in E_i(\Gamma)$ in given in Figure 2.20 in the special case where α is the unique edge shared by the two triangles it bounds. The $\mathcal{A}$-coordinates on the edges of the triangulation not represented in the picture do not vary.

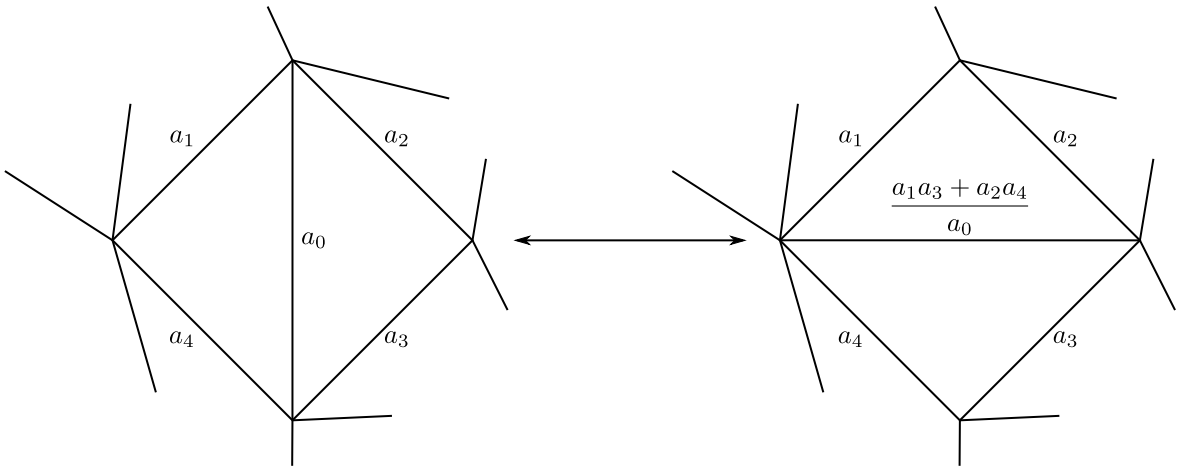


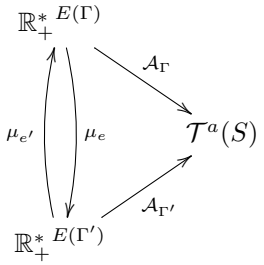
Figure 2.20: The $\mathcal{A}$ coordinates change rule under a flip.

In general if $\left(\epsilon^{\alpha\beta}\right)_{\alpha,\beta\in E(\Gamma)}$ is the extended exchange matrix of the triangulation Γ, the variation of

$\mathcal{A}$-coordinates μ_α corresponding to the flip of some edge α can be written for any $\gamma \in E(\Gamma)$ as:

$$\begin{aligned} \mu_\alpha(a_\gamma) &= a_\gamma && \text{if } \gamma \neq \alpha \ , \\ \mu_\alpha(a_\alpha) &= a_\alpha^{-1} \left(\prod_{\beta \in E(\Gamma),\ \epsilon^{\alpha\beta}>0} a_\beta + \prod_{\beta \in E(\Gamma),\ \epsilon^{\alpha\beta}<0} a_\beta \right) && \text{if } \gamma = \alpha \ . \end{aligned}$$

This is exactly the mutation formula for cluster variables of Equation (1.24). From now on it will be called the *cluster $\mathcal{A}$-mutation formula* These rules are subtraction-free, and hence if Γ and Γ' are two triangulations of the same ciliated surface related by a flip at some edge α of Γ, these mutation rules μ_α indeed map $\mathbb{R}_{>0}^{E(\Gamma)}$ to $\mathbb{R}_{>0}^{E(\Gamma')}$. Moreover, if α' is the flipped edge in Γ' one sees easily that $\mu_{\alpha'} : \mathbb{R}_{>0}^{E(\Gamma')} \longrightarrow \mathbb{R}_{>0}^{E(\Gamma)}$ is the reciprocal map of μ_α. These coordinate changes are transition functions:



(2.65)

A closed 2-form on $\mathcal{T}^a(S)$. In the cluster chart corresponding to Γ, one can check directly that the 2-form

$$\omega = \sum_{\alpha,\beta \in E(\Gamma)} \epsilon^{\alpha\beta} \frac{\mathrm{d}a_\alpha}{a_\alpha} \wedge \frac{\mathrm{d}a_\beta}{a_\beta} \tag{2.66}$$

is closed. In general, it is degenerate. Moreover, the mutation formulae of Equation (1.14) and Equation (2.65) imply that it is independent of the triangulation. As such, it defines a degenerate symplectic structure on $\mathcal{T}^a(S)$.

* * * * * * *

The Teichmüller space with holes $\mathcal{T}^x(S)$ and the decorated Teichmüller space $\mathcal{T}^a(S)$ of a ciliated surface S are interesting variants of Teichmüller spaces in which cluster structures naturally appear. They are charted by copies of $(\mathbb{R}_{>0})^N$ which describe the whole space, where $N = \#E_i(\Gamma)$ in the first case and $N = \#E(\Gamma)$ in the second. Transition functions between charts are sequences of $\mathcal{X}$-mutations and $\mathcal{A}$-mutations, respectively. The space $\mathcal{T}^x(S)$ is endowed with a Poisson bracket, and $\mathcal{T}^a(S)$, with a closed 2-form. In fact, these two Teichmüller spaces are the $\mathbb{R}_{>0}$-points of cluster varieties, which will be defined in Chapter 4.

Before that however, we will discuss laminations on ciliated surfaces. There are also two variants of lamination spaces denoted $\mathcal{A}$ and $\mathcal{X}$ and they correspond to tropical points of cluster varieties.

In Chapter 4 we will introduce the G-higher and Teichmüller spaces of type $\mathcal{X}$ and $\mathcal{A}$, where G is the split real form of any reductive algebraic group, either adjoint or simply connected. These spaces are the generalization of what has been done in this chapter, for general reductive algebraic groups.

Chapter 3

Laminations on surfaces

The Teichmüller space of a surface S is related to various structures on S, as the three different definitions presented at the beginning of last chapter indicate: $\mathcal{T}(S)$ is at the same time the space of hyperbolic structures on S modulo $\mathrm{Diff}_0(S)$, the space of complex structures on S modulo $\mathrm{Diff}_0(S)$ and a connected component of $\mathrm{Hom}(\pi_1(S), \mathrm{PSL}_2(\mathbb{R}))/\mathrm{PSL}_2(\mathbb{R})$ consisting of discrete and faithful morphisms. Laminations form another class of objects on surfaces which connect with Teichmüller theory.

We present measured laminations, as well as the close notions of curve systems and train tracks in Section 3.1. An important property of the space of measured laminations on a surface S is that it provides a compactification of the Teichmüller space $\mathcal{T}(S)$ to which the action of the mapping class group $\mathrm{MCG}(S)$ extends. Curve systems provide a definition of integral and rational laminations, which differ from the real (measured) laminations. The latter are in general more difficult to describe explicitly than the former: real laminations correspond to a singular foliations of S together with a transverse real measure, whereas integral and rational laminations are mere systems of curves on S with integral or rational weights.

The definition of integral and rational laminations can be extended to ciliated surfaces. They come in two flavors denoted $\mathcal{A}$ and $\mathcal{X}$, as in the case of Teichmüller spaces. The notation comes from the fact that the space of laminations of type $\mathcal{A}$ (resp. $\mathcal{X}$) admits coordinate systems corresponding to triangulations of S, and with the transition functions being sequences of (tropical) mutations of type $\mathcal{A}$ (resp. $\mathcal{X}$). This will be presented in Section 3.2. Just as the Teichmüller spaces of the last chapter can be interpreted as the $\mathbb{R}_{>0}$-points of cluster varieties, integral (resp. rational) laminations spaces naturally appear as the $\mathbb{Z}^t$-points (resp. $\mathbb{Q}^t$-points) of cluster varieties, as we will see in Chapter 4.

Laminations play an important role in the cluster Teichmüller theory of ciliated surfaces, because they encode functions on the Teichmüller space. More precisely, an $\mathcal{A}$-lamination corresponds to a function on $\mathcal{T}^x(S)$ while an $\mathcal{X}$-lamination can be interpreted as a function on $\mathcal{T}^a(S)$. These properties are nicely described in terms of *p*airings, as we will do in Section 3.3. The *duality conjectures* discussed in Section 3.3.3 state that integral laminations of type $\mathcal{A}$ (resp. of type $\mathcal{X}$) provide a basis of a special subset of functions on the Teichmüller space of type $\mathcal{X}$ (resp. of type $\mathcal{A}$).

3.1 Thurston's measured laminations

The goal of this section is to provide general background for Thurston's laminations. We will (briefly) discuss three-dimensional hyperbolic manifolds with boundary, curve systems, train tracks, measured laminations and Thurston's compactification of the Teichmüller space.

3.1.1 Laminations from three-dimensional hyperbolic spaces

In this section we follow [TM79, Section 8].

Generalities. Let $n \geq 2$ be an integer, $\mathbb{H}^n$ the n-dimensional real hyperbolic space, S^{n-1}_∞ its boundary at infinity defined as in Definition 2.1, and Γ a discrete group of orientation preserving isometries of $\mathbb{H}^n$. Let x be any point in $\mathbb{H}^n$, and define the *limit set* $\mathcal{L}_\Gamma \subset S^{n-1}_\infty$ *of* Γ as the set of accumulation points of the orbit $\{\Gamma \cdot x\}$.

Proposition 3.1. *$\mathcal{L}_\Gamma$ does not depend on the choice of $x \in \mathbb{H}^n$.*

Proof. Let $y \in \mathbb{H}^n$ be any other point, and let $\{\gamma_i\}_{i\in\mathbb{N}}$ be a sequence in Γ such that $\gamma_i x \to p \in \mathcal{L}_\Gamma$ as $i \to \infty$. Since Γ is a group of isometries of $\mathbb{H}^n$, if $d(\cdot,\cdot)$ is the hyperbolic metric in $\mathbb{H}^n$:

$$d(\gamma_i x, \gamma_i y) = d(x, y) \tag{3.1}$$

is independent of i. Hence, if $d_E(\cdot,\cdot)$ is the euclidean metric on the boundary S^{n-1}_∞, $d_E(\gamma_i x, \gamma_i y) \to 0$ as $i \to \infty$, and hence $\lim_{i\to\infty} \gamma_i x = \lim_{i\to\infty} \gamma_i y$. □

Let us define the *domain of discontinuity* of Γ as $D_\Gamma := S^{n-1}_\infty - \mathcal{L}_\Gamma$, and set:

$$M_\Gamma = \text{ConvHull}(\mathcal{L}_\Gamma)/\Gamma \ , \tag{3.2}$$

which is a hyperbolic manifold if Γ acts freely on $\text{ConvHull}(\mathcal{L}_\Gamma)$. From now on we restrict to $n = 3$.

Definition 3.2. *A Kleinian group is a discrete subgroup of* $\text{Isom}(\mathbb{H}^3) = \text{PSL}_2(\mathbb{C})$ *whose domain of discontinuity is non-empty.*

Let K be any closed set on the boundary S^2_∞ of $\mathbb{H}^3$, and let $H(K)$ be its convex hull in $\mathbb{H}^3$. Any point in $\partial H(K) - K$ belongs to some line segment in $\partial H(K)$, and hence $\partial H(K) - K$ can be developed in the hyperbolic plane. Usually $\partial H(K) - K$ is not smooth, however if Γ is a torsion-free Kleinian group, ∂M_Γ is a smooth hyperbolic surface bent along a **lamination**. Let us give our first definition of a lamination in a fairly general setting.

Definition 3.3. *A lamination L on an n-dimensional manifold M is a closed subset $A \subset M$ called the support of L, with a local product structure for A: there is a covering of a neighborhood of A in M by open sets U_i with charts $\phi_i : U_i \to \mathbb{R}^{n-k} \times \mathbb{R}^k$ such that*

$$\phi(U_i \cap A) = \mathbb{R}^{n-k} \times B_i \ , \tag{3.3}$$

where $B \subset \mathbb{R}^k$. The transition functions $\phi_{ij} = \phi_i \circ \phi_j^{-1}$ must be of the form

$$\phi_{ij}(x, y) = (f_{ij}(x, y), g_{ij}(y)) \ , \tag{3.4}$$

where $y \in B_i$, and $g_{ij}(y) \in B_j \in \mathbb{R}^k$.

Equivalently, a lamination is a regular foliation of a closed subset of M.

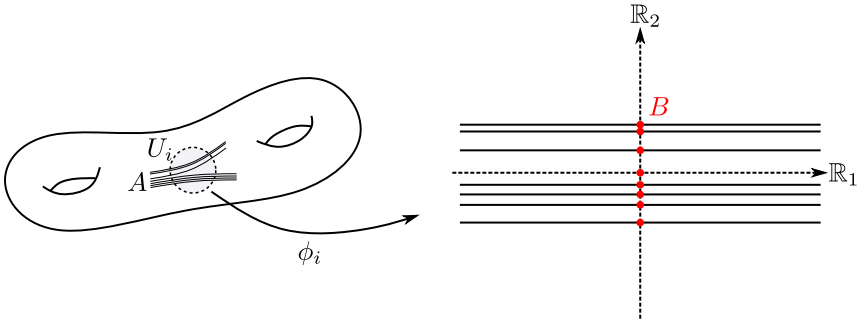


Figure 3.1: A chart of lamination.

Geodesic laminations and transverse measures. Let us assume that $M = S$ is a hyperbolic surface. A lamination is said to be *geodesic* if the leaves of the corresponding foliation are geodesics.

Let γ be a geodesic foliation of a hyperbolic surface S and let us consider the complement of the lamination $S - \gamma$. The Euler characteristic of the support of γ is zero since it admits a regular foliation, and since $\chi(S) < 0$ the complement of γ in S cannot be empty. It consists of regions bounded by geodesics which are either closed or open. For each connected component of $S - \gamma$ we can consider its double, obtained by identifying the boundaries of two copies of this connected component, as show in Figure 3.2. Each of these doubles D is a complete hyperbolic surface with finite area $-2\pi\chi(D)$, hence the number of connected components of $S - \gamma$ is bounded from above by $2|\chi(S)|$, and in particular it is finite.

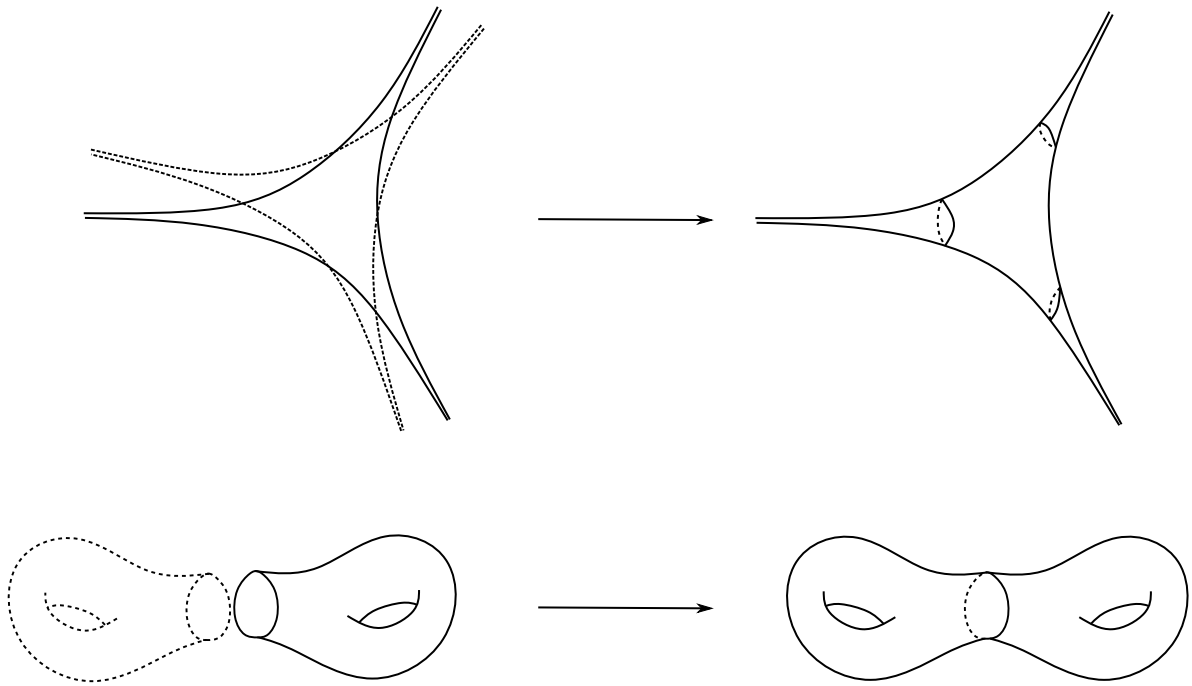


Figure 3.2: Doubles of an ideal hyperbolic triangle and a torus with one hole.

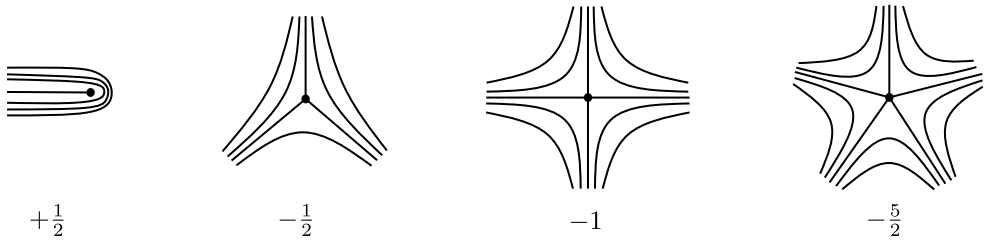


Figure 3.3: Some values for the index.

Every geodesic lamination γ on S can be extended to a foliation with isolated singularities on $S - \gamma$. There is an index formula for the Euler characteristic of S in terms of the index of these singularities. Some values of the index are shown in Figure 3.3. The index formula implies that $\chi(S)$ is half the Euler characteristic of the double of $S - \gamma$, and hence $\text{Area}(S) = \text{Area}(S - \gamma)$, thus γ has measure 0.

In the case of the boundary ∂M_Γ of the three-dimensional convex hyperbolic manifold M_Γ corresponding to a Kleinian group Γ, one can add assign a transverse measure in order to account quantitatively for the bending along a curve of the corresponding lamination. More precisely:

Definition 3.4. *Let S be a hyperbolic surface and let γ be a lamination on S. A transverse measure μ for γ is a measure defined on each local leaf space B_i and such that the coordinate changes are measure-preserving. Equivalently, a transverse measure for γ can be thought of as a measure defined on every 1-dimensional submanifold T of S transverse to γ, supported on $T \cap \gamma$, and invariant under local projections along the leaves of γ. A geodesic lamination together with such a transverse measure is called a measured geodesic lamination.*

3.1.2 Laminations from curve systems

In this subsection we follow [Hat88]. Let S be a compact surface, with or without boundary.

Definition 3.5. *A curve system on S is a finite collection of non self- and pairwise-intersecting curves which are either:*

- *a circle not bounding a disk in S and non-isotopic to a connected component of ∂S,*

- *an arc with endpoints in ∂S non-isotopic (relatively to its endpoints) to arcs in ∂S.*

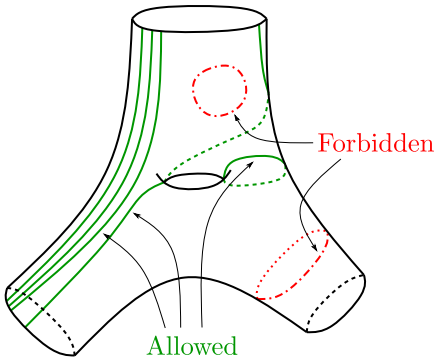


Figure 3.4: Examples and non-examples of the curves that can belong to a curve system.

Let $\mathcal{CS}(S)$ be the set of isotopy classes of such curve systems in S. Let $\mathcal{PS}(S)$ be the set of projective isotopy classes of curve systems in S, i.e. the set obtained from $\mathcal{CS}(S)$ by identifying a non-empty curve systems with any positive number of parallel copies of itself.

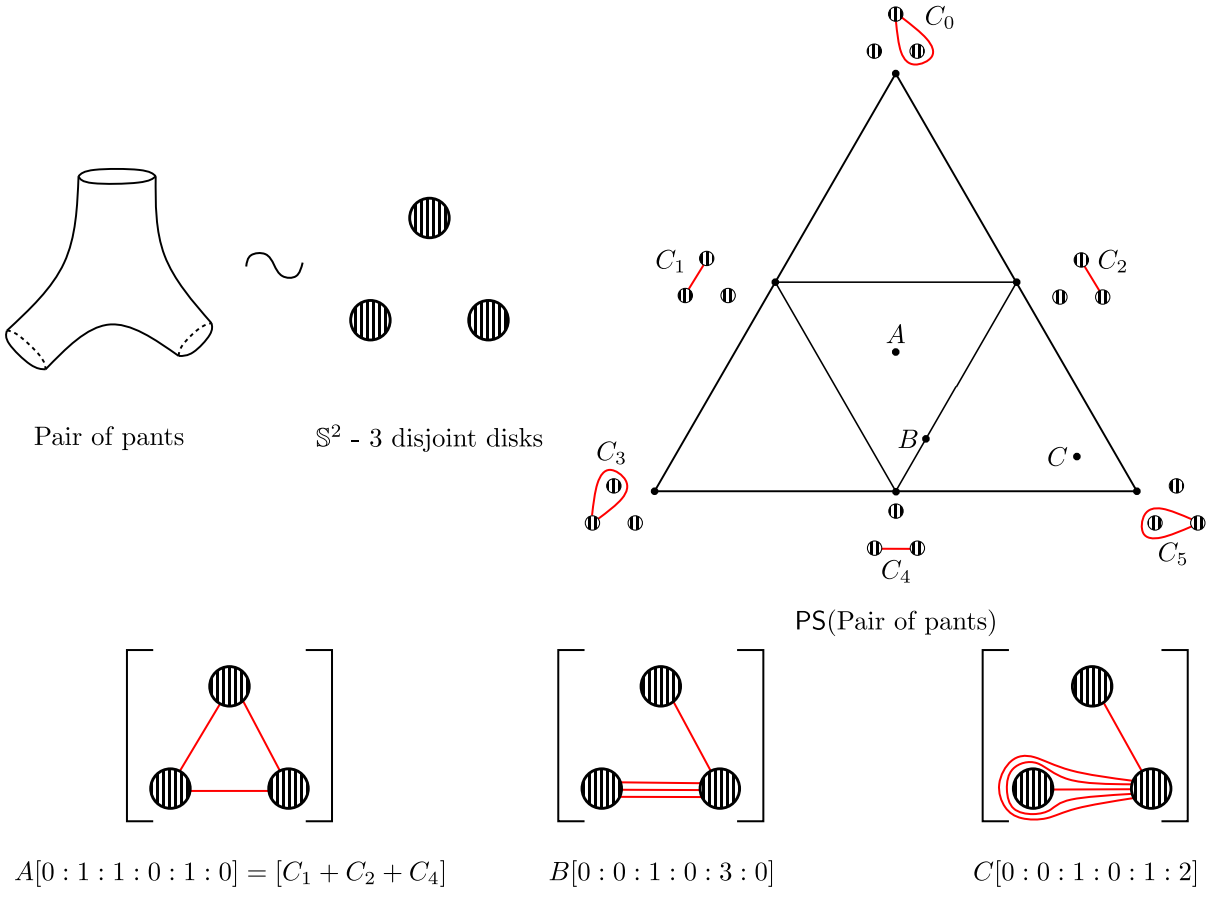


Figure 3.5: The simplicial complex $\mathsf{PS}(S)$ when S is a pair of pants.

Remark 3.6. *Any point in $\mathcal{CS}(S)$ can be written as*

$$n_0 C_0 + \cdots + n_k C_k \tag{3.5}$$

for $k \in \mathbb{N}$, where the C_i's are curves of one of the two types described in Definition 3.5 and the n_i's are strictly positive integers, so that $n_i C_i$ denotes n_i parallel copies of C_i.

Let us define a simplicial complex $\mathsf{PS}(S)$ whose k-simplices are in bijection with isotopy classes of such $(k+1)$-tuples $(C_0, \dots, C_k)$ forming a curve system, and endowed with homogeneous coordinates $[x_0 : \cdots : x_k]$, where $x_0, \dots, x_k \in \mathbb{R}$. Let $i \in [|0, k|]$. The i-th facet of the k-simplex corresponding to $(C_0, \dots, C_k)$ is defined by the condition $x_i = 0$.

Any point $n_0 C_0 + \cdots + n_k C_k \in \mathcal{CS}(M)$ projects to a point with homogeneous coordinates $[n_0 : \cdots : n_k]$ in $\mathcal{PS}(S)$. Hence one can identify $\mathcal{PS}(S)$ with the points in $\mathsf{PS}(S)$ whose homogeneous coordinates are integral, or equivalently, rational. We will see below that the non-rational points of $\mathsf{PS}(S)$ can be identified with measured projective laminations on S. The example of $\mathsf{PS}(S)$ when S is a pair of pants is depicted in Figure 3.5.

3.1.3 Train tracks

Definition 3.7. *A train track τ on a hyperbolic surface S is a closed subset of S meeting the boundary ∂S transversely, and locally diffeomorphic to the model shown on the left of Figure 3.6. Equivalently it a compact submanifold of S meeting ∂M transversely, except at finitely many branch points in the interior of S where two arcs merge into one with the same tangential direction. We will assume that our train tracks are good, meaning that they have none of the complementary regions listed in Figure 3.7. An example of a good train track is displayed on the right of Figure 3.6.*

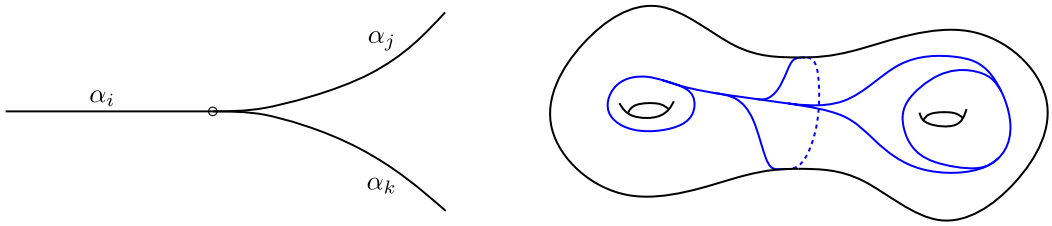


Figure 3.6: The only allowed junction in train tracks.

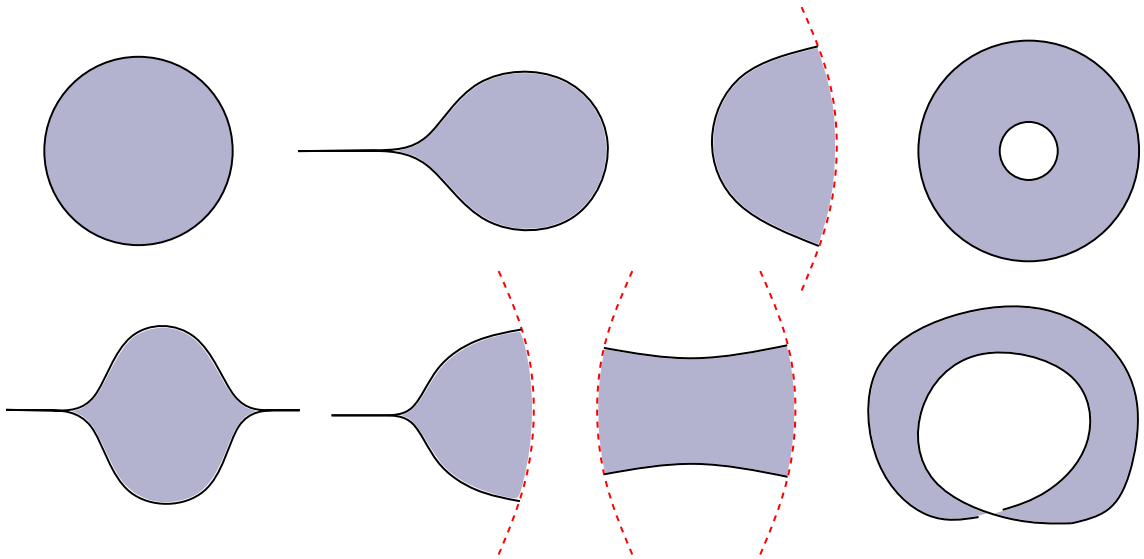


Figure 3.7: Bad train tracks. The dashed red lines represent the segments of the boundary of the surface.

Definition 3.8. *A measure on a track τ is an assignment of real positive weights $\alpha_i \geq 0$ to the connected components of the non-branching locus of τ satisfying the branch equation $\alpha_i = \alpha_j + \alpha_k$ at each branch point, with the notation of Figure 3.6.*

The set of measures on a given train track τ forms a cone $C(\tau)$ in $\mathbb{R}^n$, which is the intersection of the hyperplanes defined by the branching equation with the quadrant $[0, \infty[^n$.

Given a train track τ and an n-tuple $(\alpha_1, \cdots, \alpha_n) \in C(\tau) \cap \mathbb{Z}^n$, one can take α_i parallel copies to the i-th non-singular arc of τ and match these arcs together at the branch points according to the branch equations, as in Figure 3.8. It yields a curve system $S_\alpha \subset M$ said to be carried by τ.

Figure 3.8: From a train track to a curve system.

The main point is the following lemma 1.1 of [Hat88], to which we refer for the proof:

Lemma 3.9. *If τ is a good train track and if S_α is carried by τ, then $S_\alpha \in \mathcal{CS}(M)$.*

One can define a simplicial complex $\mathsf{MS}(S)$ from a surface S, by considering the so-called *standard train tracks* on S determined by a decomposition of S into pairs of pants. We refer to [Hat88] for a precise definition. Each such standard train track τ defines a cone $C(\tau)$ such that if τ' is a subtrack of τ, then $C(\tau')$ is a face of $C(\tau)$. Doing the identifications as prescribed by combinatorics, one obtains the simplicial complex $\mathsf{ML}(S)$ (where ML stands for measured laminations) that can be projectivised into a projective simplicial complex $\mathsf{PL}(M)$.

Proposition 3.10 (1.2 of [Hat88])**.**

$$\mathsf{ML}(S)_{\mathbb{Z}} \simeq \mathcal{CS}(S) \tag{3.6}$$

$$\mathsf{PL}(S)_{\mathbb{Q}} \simeq \mathcal{PS}(S) \tag{3.7}$$

Proposition 3.11 (1.5 of [Hat88])**.** $\mathsf{ML}(S)$ *is piecewise linearly homeomorphic to* $\mathbb{R}^{-3\chi(S)-b} \times [0, \infty[^b$.

More on train tracks can be found in [TM79, PH92].

3.1.4 Measured laminations from train tracks

Given a train track τ on a surface S and $\alpha \in C(\tau)$ a (real) measure on τ one can construct a foliation $\mathcal{N}_\alpha$ of the fibered neighborhood $\mathcal{N}_\tau$ of τ. For each non-singular arc i of τ one considers the fibered neighborhood of i of size α_i and connect these neighborhoods at each branch point as displayed in Figure 3.9. Then, one removes all the singular leaves. An example of such a singular leaf is shown as the dashed line in Figure 3.9.

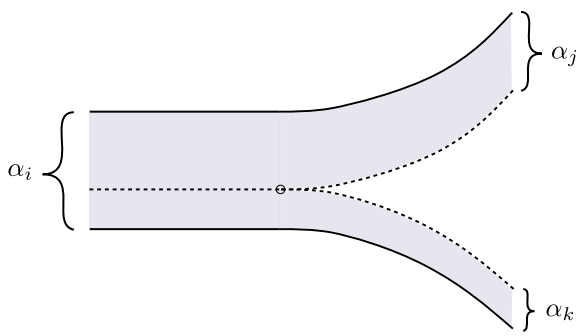


Figure 3.9: The construction of laminations from measured train tracks.

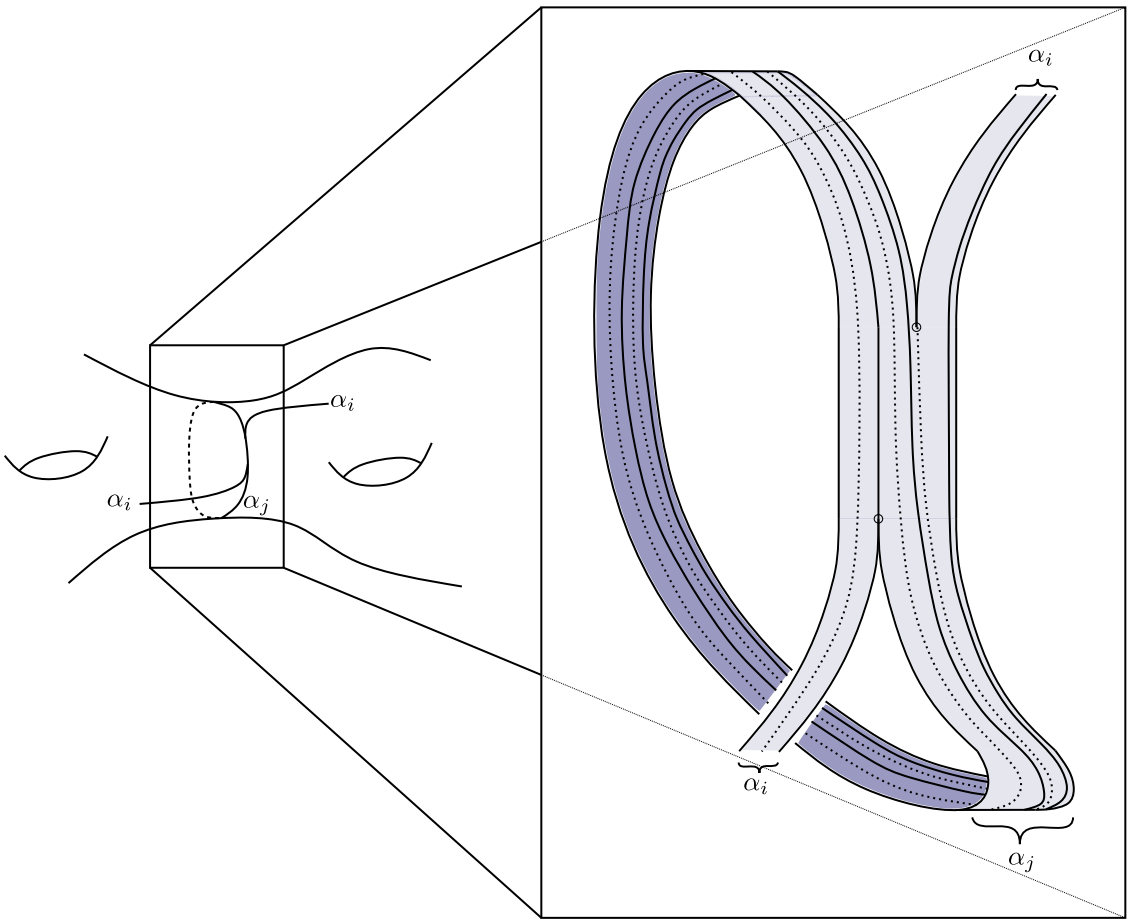


Figure 3.10: How Cantor sets arise in generic laminations.

Figure 3.10 shows how this construction works near a closed curve on S. Generically, the resulting lamination can not be interpreted as a projective curve system anymore. This gives a geometric interpretation of train tracks on a surface, with generic real measure.

When the measure assigned to the train track τ is rational, removing the singular leaves during the construction of the lamination from the measured train track as in Figure 3.9 yields a collection of strips on the surface[1] of finite rational width, which is a fattening of a projective curve system.

Proposition 3.12 (2.3 of [Hat88])**.** $\mathsf{ML}(S)$ *is isomorphic to the space of measured laminations on* S*.*

Hence every point of $\mathsf{PL}(S)$ can be naturally interpreted as a projective measured lamination on S, which generalizes the interpretation of the rational points of $\mathsf{PL}(S)$ as projective curve systems on S.

3.1.5 Laminations and Thurston's compactification of Teichmüller spaces

The space of projective measured laminations on a surface S forms a boundary for the Teichmüller space of S, to which the action of the mapping class group extends [Thu88]. Heuristically, this can be understood as follows. A point in the Teichmüller space of S consists of the data of a hyperbolic metric on S, which is a smooth rank 2 section of the symmetric product $\mathrm{Sym}^2(T^*M)$ of the cotangent space of S with itself: at each point, the metric can be written in coordinates as

$$g = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \qquad \mu,\nu \in \{0,1\}\ , \tag{3.8}$$

where $g_{\mu\nu}$ is a rank-2 matrix:

$$\begin{pmatrix} g_{00} & g_{01} \\ g_{10} & g_{11} \end{pmatrix} \tag{3.9}$$

[1]Note that there is no contradiction with the fact that laminations have zero measure, since everything we are doing here is purely topological.

A boundary for the Teichmüller space of S can be defined as all the smooth sections of $\mathrm{Sym}^2(T^*M)$ which degenerate at every point of S into a rank-1 tensor. Locally, each such section can be written as $(df)^2$, where f is a smooth function on S. Just as usual differential 1-forms correspond to oriented level lines, rank-1 sections of $\mathrm{Sym}^2(T^*M)$ correspond to unoriented level lines and carry some kind of measure determined by f: they are measured laminations. Actually, the set of level lines does not change when f is rescaled uniformly, and this explains why it is the space of projective measured laminations which appear.

After having developed the theory of cluster varieties, we will have an interesting perspective which will allow us to understand conceptually why projective measured laminations naturally provide a spherical compactification of the Teichmüller space, and how this can be extended to general positive varieties. This will be conducted in Section 4.1.2.

3.2 Rational laminations on ciliated surfaces

Throughout this section, we will say that a curve on a ciliated surface S (without self-intersection and either closed or connecting two points on the boundary disjoint from the cilia) is *special* if it is retractable to a hole or to an interval on the boundary containing exactly one cilium. A curve is *contractible* if it can be retracted to a point within this class of curves. This section is based on [Foc97, FG06, FG07]. We will stick to the notation for ciliated surfaces introduced in Section 2.2.

Before introducing $\mathcal{A}$-laminations and $\mathcal{X}$-laminations on ciliated surfaces, we will discuss briefly singular (a.k.a. bounded) and Borel–Moore (a.k.a. unbounded) homologies of topological spaces, for the way they are defined is very similar to laminations. By analogy, $\mathcal{A}$-laminations (resp. $\mathcal{X}$-laminations) are also called bounded (resp. unbounded) laminations in [FG07].

3.2.1 Singular and Borel–Moore homologies

Let X be a locally compact space, and let $i \in \mathbb{Z}_{\geq 0}$. A singular i-simplex in X is a continuous map

$$\Delta^i \to X \ , \tag{3.10}$$

where Δ^i is the standard i-simplex. In bounded homology, also called homology with compact support, for all $i \in \mathbb{Z}_{\geq 0}$ one defines the space of i-chains $C_i(X)$ as the free abelian group generated by all possible singular i-simplices in X. Then one considers the chain complex

$$\ldots \xrightarrow{\phi_3} C_2(X) \xrightarrow{\phi_2} C_1(X) \xrightarrow{\phi_1} C_0(X) \xrightarrow{0} 0 \tag{3.11}$$

and from it one defines the singular homology groups as:

$$H_i = \ker(\phi_i)/\mathrm{im}(\phi_{i+1}) \ . \tag{3.12}$$

In Borel-Moore homology [BM60], also called homology with closed support or unbounded homology, one defines the space $C_i^{\mathrm{BM}}(X)$ as the space of formal (infinite) sums of i-simplices in X with integer coefficients:

$$u = \sum_\sigma a_\sigma \sigma \ , \tag{3.13}$$

where σ runs over all i-simplices in X and where $a_\sigma \in \mathbb{Z}$, and such that for any compact subspace $S \subset X$ only finitely many simplices σ whose image meets S have a non-zero coefficient a_σ. Then one considers the chain complex

$$\ldots \xrightarrow{\phi_3^{\mathrm{BM}}} C_2^{\mathrm{BM}}(X) \xrightarrow{\phi_2^{\mathrm{BM}}} C_1^{\mathrm{BM}}(X) \xrightarrow{\phi_1^{\mathrm{BM}}} C_0^{\mathrm{BM}}(X) \xrightarrow{0} 0 \tag{3.14}$$

and defines from it the singular homology groups as:

$$H_i^{\mathrm{BM}} = \ker(\phi_i^{\mathrm{BM}})/\mathrm{im}(\phi_{i+1}^{\mathrm{BM}}) \ . \tag{3.15}$$

Consider the cylinder $C =]-1,1[\times S^1$ shown in Figure 3.11, and the dashed red circle $\{0\} \times S^1 \subset C$. In bounded homology it is a 1-chain and not a boundary, hence it represents a non-trivial element in

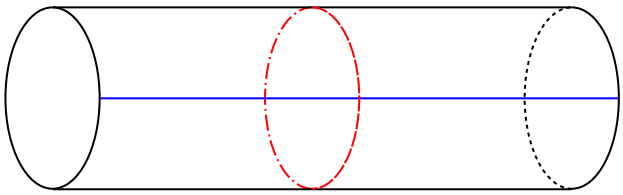


Figure 3.11: The cylinder $C =]-1,1[\times S^1$.

$H_1(C)$. In unbounded homology it is also a 1-chain, however it is a boundary and hence it is trivial in $H_1^{\mathrm{BM}}(C)$. Conversely, the blue line $]-1,1[\times\{0\} \subset C$ is not even a 1-chain in homology with compact support, however in homology with closed support it is a 1-chain and it is not a boundary, hence it represents a non-trivial element in $H_1^{\mathrm{BM}}(C)$.

Borel-Moore homology extends Poincaré duality for non-compact spaces X. More precisely:

Proposition 3.13. *Let X be a locally compact topological space with a closed embedding into an oriented manifold M of dimension m. Then for all $i \in [|0,m|]$:*

$$H_i^{\mathrm{BM}}(X) = H^{m-i}(M, M\backslash X)\ . \tag{3.16}$$

where $H^{m-i}(M, M\backslash X)$ is the $(m-i)$-th cohomology group of M relatively to $M\backslash X$.

The proof can be found in [Ive12, Theomem IX.4.7]. When $X = M$ this result indeed extends Poincaré duality to a duality between homology with closed support and ordinary cohomology. Subsequently:

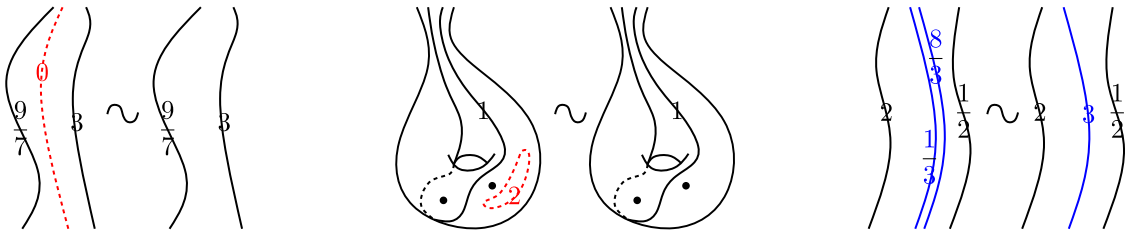
Proposition 3.14. *If X be a compact topological space, homologies with compact and closed support on X coincide:*

$$H_*^{\mathrm{BM}}(X) = H_*(X)\ . \tag{3.17}$$

3.2.2 Rational $\mathcal{A}$-laminations

Definition 3.15. *A rational $\mathcal{A}$-lamination on S, or rational bounded measured lamination, is a homotopy class of a finite number of self and mutually non-intersecting unoriented curves, either closed or connecting two points of the boundary disjoint from cilia and the holes, with rational weight and such that the weight of a curve is non-negative, unless it is special. Moreover, rational $\mathcal{A}$-laminations are subject to the following equivalence relation:*

1. *A lamination containing a curve of weight zero is equivalent to the lamination with this curve removed.*
2. *A lamination containing a contractible curve is equivalent to the lamination with that curve removed.*
3. *A lamination containing two homotopy equivalent curves with weights u and v is equivalent to the lamination with one of these curves removed and with the weight $u+v$ on the other.*



1. 2. 3.

Figure 3.12: Equivalence relations for rational $\mathcal{A}$-laminations.

Let $\mathsf{T}^a(S,\mathbb{Q})$ be the space of rational bounded measured laminations. It has the space $\mathsf{T}^a(S,\mathbb{Z})$ consisting of bounded measured laminations with integral weights as a natural subset. Moreover, given a

rational bounded measured laminations one can multiply all the weights by any $\lambda \in \mathbb{Q}_{>0}$; in other words there is an action:

$$\mathbb{Q}_{>0} \times \mathsf{T}^a(S, \mathbb{Q}) \longrightarrow \mathsf{T}^a(S, \mathbb{Q}) \ . \tag{3.18}$$

The space $\mathsf{T}^a(S)$ has another natural subspace $\mathsf{T}^a_0(S)$; which consists of the rational bounded measured laminations such that for any segment of the boundary of S between two cilia the total weight of the curves ending at it vanishes. Let us emphasize that curves of an $\mathcal{A}$-lamination cannot end at holes.

Let us now describe how the set of rational bounded measured laminations can be parameterized by rational numbers on the edges of a triangulation Γ of S.

Parametrization. One first deforms the curves of the lamination in such a way that they intersect the edges of the triangulation at the minimal possible number of points. Then, to each edge $\alpha \in \Gamma$ one assigns half the sum of the weights of the curves of the lamination intersecting α.

This parametrization is a chart:

$$\mathcal{A}^t_\Gamma : \mathbb{Q}^{E(\Gamma)} \longrightarrow \mathsf{T}^a(S, \mathbb{Q}) \ , \tag{3.19}$$

where as before $E(\Gamma)$ is the set of (internal and external) edges of Γ. Note that the image of $\mathsf{T}^a(S, \mathbb{Z})$ is contained in $(\frac{1}{2}\mathbb{Z})^{E(\Gamma)}$ and the one of $\mathsf{T}^a_0(S)$ consists of those edge assignments such that the rationals assigned to the exterior edges of Γ vanish.

Reconstruction. Let $\{a_\alpha\}_{\alpha \in E(\Gamma)} \in (\mathbb{Q})^{E(\Gamma)}$. For all $u, v \in \mathbb{Q}$ and $\alpha \in E(\Gamma)$ one sets:

$$\tilde{a}_\alpha = u a_\alpha + v \ . \tag{3.20}$$

Choosing wisely u and v one can always ensure that the $\tilde{a}_\alpha$ are positive integers and satisfy the triangle inequality for every face Γ: if α, β and γ are the sides of a triangle f we ask that

$$|\tilde{a}_\alpha - \tilde{a}_\beta| \leq \tilde{a}_\gamma \leq \tilde{a}_\alpha + \tilde{a}_\beta \ . \tag{3.21}$$

On each edge $\alpha \in E(\Gamma)$ one marks $2\tilde{a}_\alpha$ points, and joins them pairwise inside each triangle with non-intersecting and non-self-intersecting curves, in the up-to-homotopy unique way allowed by the triangular inequality of Equation (3.21). An example is shown in Figure 3.13.

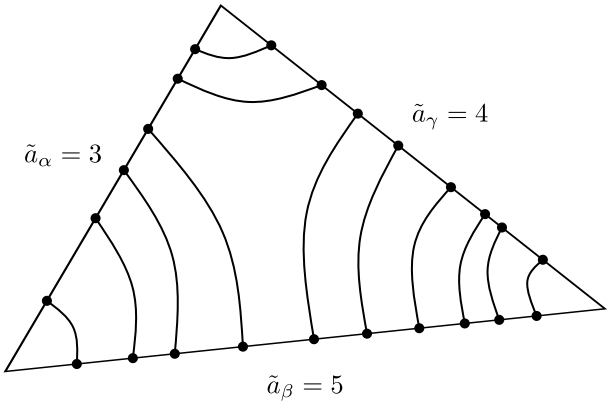


Figure 3.13: The construction of the curves of the element of $\mathsf{T}^a(S)$ in each triangle of Γ.

Last, one adds a special curve around each hole and each cilium with weight $-v$, and one divides the weights of all other curves by u. This yields the lamination corresponding to the coordinates $\{a_\alpha\}_{\alpha \in E(\Gamma)}$.

Mutation rules Under the flip of an internal edge $\alpha \in E_i(\Gamma)$ the coordinates constructed above on the space of rational bounded measured laminations transform in the following way:

$$\mu_\alpha(a_\beta) = \begin{cases} \max\left(\sum\limits_{\gamma,\ \epsilon^{\alpha\gamma}>0} a_\gamma, \sum\limits_{\gamma,\ \epsilon^{\alpha\gamma}<0} a_\gamma \right) - a_\alpha & \text{if } \beta = \alpha \\ a_\beta & \text{otherwise} \end{cases} \ , \tag{3.22}$$

where $\beta, \gamma \in E(\Gamma)$. This formula is the usual $\mathcal{A}$-mutation formula of Equation (2.65), in the tropical semi-field $\mathbb{Q}^t$. In the case of a flip of the single common edge of two adjacent triangles, the transformation rule is presented explicitly in Figure 3.19. Restricting to the laminations whose coordinates in the chart $\mathcal{A}^t_\Gamma$ are integers (resp. half-integers), one replaces $\mathbb{Q}^t$ by the sub-semifield $\mathbb{Z}^t$ (resp. $\frac{1}{2}\mathbb{Z}^t$).

Borrowing notation from the next chapter, let $\mathcal{A}(S, \frac{1}{2}\mathbb{Z}^t)$ (resp. $\mathcal{A}(S, \mathbb{Z}^t)$) denote the space of laminations on S whose coordinates in one (equivalently, every) cluster chart are half-integers (resp. integers).

Integral $\mathcal{A}$-laminations as defined in the previous section correspond to systems of curves on S with integral weights. They form the space $\mathsf{T}^a(S, \mathbb{Z})$ which, by the above, satisfies:

$$\mathcal{A}(S, \mathbb{Z}) \subset \mathsf{T}^a(S, \mathbb{Z}) \subset \mathcal{A}(S, \frac{1}{2}\mathbb{Z}^t) \ . \tag{3.23}$$

In fact, one can show that the elements of $\mathsf{T}^a(S, \mathbb{Z})$ are exactly the points in $\mathcal{A}(S, \frac{1}{2}\mathbb{Z}^t)$ such that at each vertex v of Γ, the sum of the coordinates on the edges incident to v is an integer.

As in the case of Teichmüller spaces, the $\mathcal{A}^t_\Gamma$ are cluster charts on the lamination space $\mathsf{T}^a(S, \mathbb{Q})$ or $\mathsf{T}^a(S, \mathbb{Z})$. In the next chapter we will see that these lamination spaces relate to the tropical points of cluster varieties.

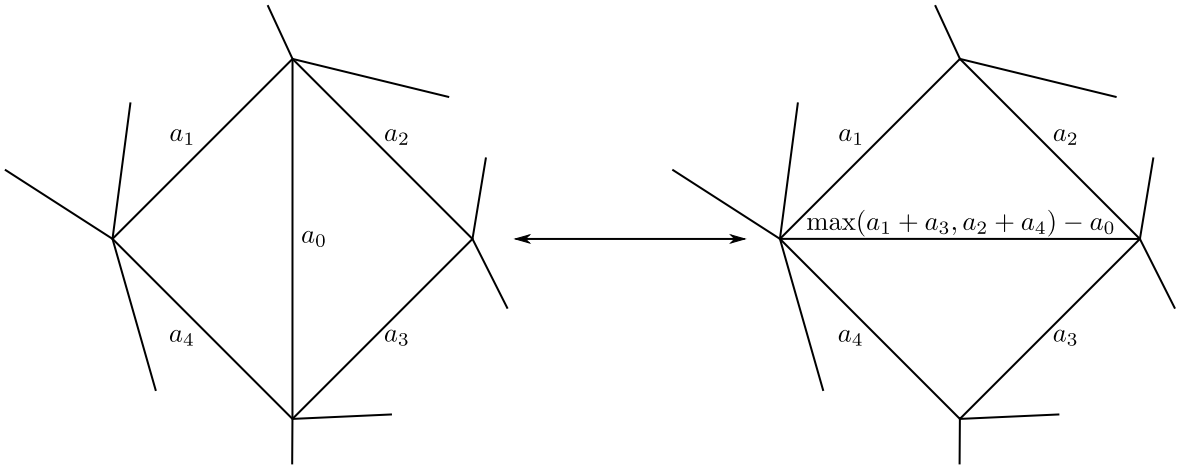


Figure 3.14: The coordinate changes under a flip for tropical $\mathcal{A}$ coordinates.

3.2.3 Rational $\mathcal{X}$-laminations

Definition 3.16. *A rational $\mathcal{X}$-lamination γ on a ciliated surface S, or rational unbounded measured lamination, is an isotopy class of a finite collection of non-self-intersecting and pairwise non-intersecting curves on S with positive rational weights, either closed or ending at the boundary (including holes) and disjoint from cilia, and subject to the following relations:*

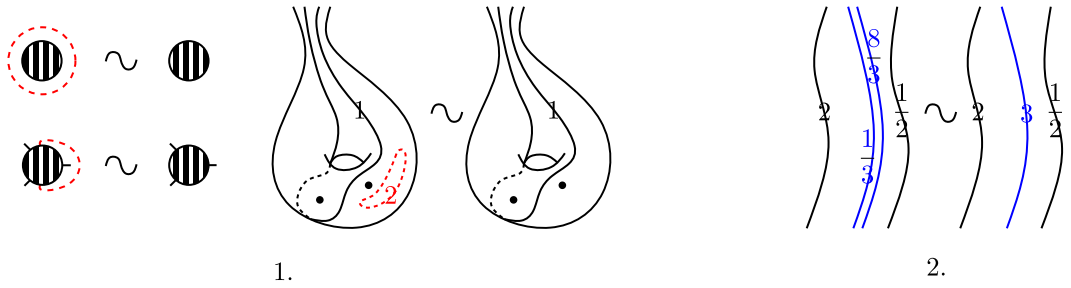
1. *A lamination containing a contractible or special curve is equivalent to the lamination with this curve removed,*
2. *A lamination containing two homotopy equivalent curves of weights u and v is equivalent to the lamination with one of these curves removed and with the weight $u + v$ on the other,*

Moreover, a rational $\mathcal{X}$-lamination encodes a choice of orientation of the holes of S, but the ones which are disjoint from the curves.

Let $\mathsf{T}^x(S, \mathbb{Q})$ be the space of rational unbounded measured laminations. The space $\mathsf{T}^x(S, \mathbb{Z})$ of integral unbounded measured laminations is a natural subset of $\mathsf{T}^x(S, \mathbb{Q})$. Moreover, given a rational unbounded measured laminations one can multiply all the weights by any $\lambda \in \mathbb{Q}_{>0}$, which amounts to the action

$$\mathbb{Q}_{>0} \times \mathsf{T}^x(S, \mathbb{Q}) \longrightarrow \mathsf{T}^x(S, \mathbb{Q}) \ . \tag{3.24}$$

Let us now describe how the set of rational unbounded measured laminations can be parameterized by rational numbers on the internal edges of a triangulation Γ of S.



1. 2.

Figure 3.15: Equivalence relations for rational $\mathcal{X}$-laminations.

Parametrization First, one rotates each hole infinitely many times in the direction prescribed by the orientation, as shown in Figure 3.16.

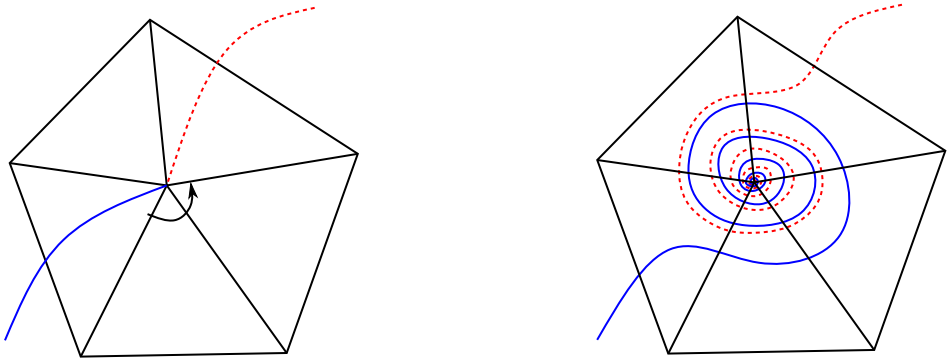


Figure 3.16: Rotating the holes.

Then, one deforms the curves in Γ in such a way that they do not cross any edge of Γ consecutively in opposite directions. Given a quadrilateral formed by two adjacent triangles in Γ sharing an edge $\alpha \in E_i(\Gamma)$, one can distinguish two types of curves which enter and exit the quadrilateral through opposite sides, dubbed positive and negative, as in Figure 3.17.

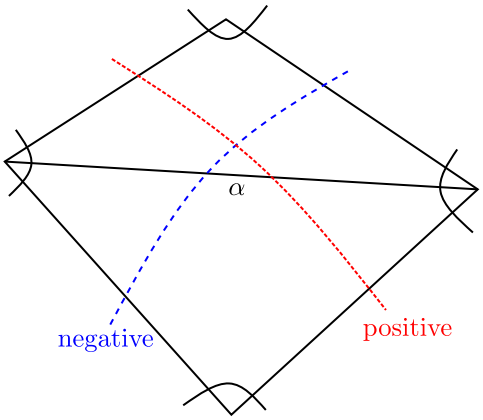


Figure 3.17: The definition of positive and negative curves with respect to the edge $\alpha \in E(\Gamma)$.

Note that since the curves of a lamination do not intersect, with respect to an edge $\alpha \in E_i(\Gamma)$, a lamination contains only positive curves or only negative ones, but not both. Note also that the curves of a lamination not entering and exiting a quadrilateral through opposite edges are neither positive not negative with respect to the diagonal.

To any internal edge of the triangulation $\alpha \in E_i(\Gamma)$, one assigns the sum of the weights of positive curves minus the sum of the weights of negative curves intersecting it.

This parametrization assigns a rational number to each internal edge of Γ, which is seen as a chart:

$$\mathcal{X}^t_\Gamma : \mathbb{Q}^{E_i(\Gamma)} \longrightarrow \mathsf{T}^x(S, \mathbb{Q}) \ . \tag{3.25}$$

The image of the integral $\mathcal{X}$-lamination under this map are the points $\mathbb{Z}^{E_i(\Gamma)}$.

Reconstruction Thanks to the action of Equation (3.24), it is enough to understand how to reconstruct a rational unbounded measured lamination corresponding to a integral coordinates $\{x^\alpha\}_{\alpha \in E_i(\Gamma)} \in \mathbb{Z}^{E_i(\Gamma)}$.

One starts by considering the faces of Γ, whose sides are homeomorphic to the real line $\mathbb{R}$. For all $k \in \mathbb{Z}_{\geq 0}$, one connects the point corresponding to $k + 1/2 \in \mathbb{R}$ one one of the sides of a triangle to the point corresponding to $-k - 1/2 \in \mathbb{R}$ one the side directly after the first one, with respect to the counterclockwise orientation. One does so in such a way that all these segments are non-self-intersecting and mutually non-intersecting. This step of the reconstruction is shown in Figure 3.18.

Now, given an internal edge α of Γ carrying the coordinate $x^\alpha \in \mathbb{Z}$, one glue the triangles sharing the edge α by identifying the point corresponding to $k + 1/2 \in \mathbb{R}$ on α in the first triangle with the point corresponding to $x^\alpha - k - 1/2 \in \mathbb{R}$ on α in the second triangle. In Figure 3.18 the gluing of the two triangles corresponds to $x^\alpha = -2$.

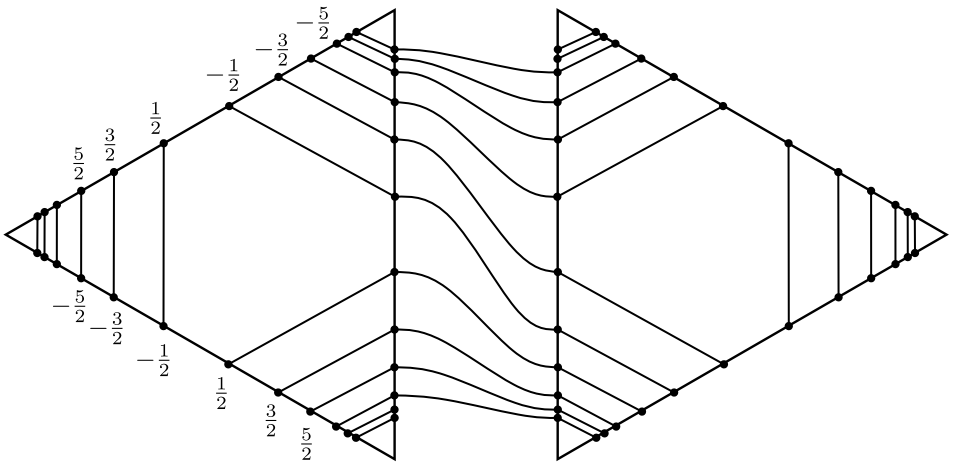


Figure 3.18: The reconstruction of rational unbounded measured laminations.

As one reconstructs S by gluing the faces of Γ together, the infinite set of curves in each triangle becomes a collection of open or closed curves on the surface. Either such a curve intersects a least one internal edge of Γ positively or negatively, or it is closed. The total number of positive and negative intersections between internal edges of Γ and the curves is

$$\sum_{\alpha \in E_i(\Gamma)} |x^\alpha| \ , \tag{3.26}$$

and in particular it is finite. Hence, the resulting lamination contains no more that this number of connected components, together with possibly infinitely many special curves which are to be considered as removed.

There is a natural map:

$$\mathsf{p} : \mathsf{T}^a(S, \mathbb{Q}) \longrightarrow \mathsf{T}^x(S, \mathbb{Q}) \ , \tag{3.27}$$

obtained by removing all special curves in a rational bounded measured lamination. One can check that in coordinates, it reads:

$$x_i = \sum_j \epsilon_{ij} a_j \ . \tag{3.28}$$

Mutation rules Under the flip of an internal edge α the coordinates constructed above on the space of rational unbounded measured laminations transform in the following way:

$$\mu_\alpha(x^\beta) = \begin{cases} -x^\alpha & \text{if } \beta = \alpha \\ x^\beta + \epsilon^{\alpha\beta}\max(0, x^\alpha) & \text{if } \epsilon^{\alpha\beta} \geq 0 \\ x^\beta + \epsilon^{\alpha\beta}\max(0, -x^\alpha) & \text{if } \epsilon^{\alpha\beta} \leq 0 \end{cases} \tag{3.29}$$

where $\beta, \gamma \in E(\Gamma)$ and where the coordinates on external edges of Γ are always zero. This formula is the usual $\mathcal{X}$-mutation formula of Equation (2.43) in the tropical semi-field $\mathbb{Q}^t$. In the case of a flip of the single common edge of two adjacent triangles, the transformation rule is presented explicitly in Figure 3.19. Integral $\mathcal{X}$-laminations have integral $\mathcal{X}$-coordinates, and the way the latter change under a flip is also given by the above mutation formula, except that it is now interpreted in $\mathbb{Z}^t$ instead of $\mathbb{Q}^t$.

As in the case of Teichmüller spaces and $\mathcal{A}$-laminations, the $\mathcal{X}_\Gamma^t$ are cluster charts. We will see in the next chapter that $\mathsf{T}^x(S, \mathbb{Q})$ is the space of tropical rational points of a cluster variety.

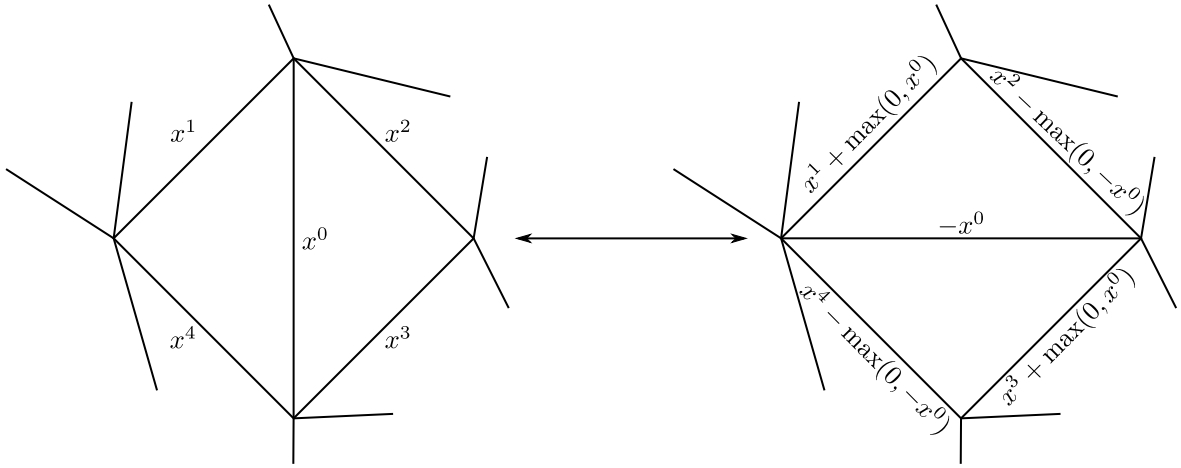


Figure 3.19: The coordinate changes under a flip for tropical $\mathcal{X}$ coordinates.

3.3 Pairings

Teichmüller and laminations spaces relate through pairings that we are going to introduce in this section, following [FG06, Sections 12.2, 12.3, 12.4] as well as [FG07].

3.3.1 Additive and intersection pairings

The choice of a point in the Teichmüller space of a surface S allows to associate a positive real number to any element of $\pi_1(S)$, by taking the hyperbolic length of the unique geodesic representing this homotopy class. This leads to the definition of additive pairings between Teichmüller and lamination spaces.

Additive pairing between $\mathcal{T}^x(S)$ and $\mathsf{T}_0^a(S, \mathbb{Q})$. The pairing

$$I : \mathcal{T}^x(S) \times \mathsf{T}_0^a(S, \mathbb{Q}) \longrightarrow \mathbb{R} \tag{3.30}$$

is defined recursively as follows.

If γ is a simple closed curve on S, interpreted as an element of $\mathsf{T}_0^a(S, \mathbb{Q})$, the pairing $I(\cdot, \gamma)$ is the function which to each point of $\mathcal{T}^x(S)$ associates $\pm$ half of the length of γ with respect to the hyperbolic metric defined by the point of $\mathcal{T}^x(S)$. The sign is $+$ unless γ surrounds a hole which is negatively oriented at that point of $\mathcal{T}^x(S)$.

If $l_1, l_2 \in \mathsf{T}_0^a(\mathbb{Q})$ and if l_1 and l_2 are non-intersecting, for all $\alpha, \beta \in \mathbb{Q}$ one sets:

$$I(\cdot, \alpha l_1 + \beta l_2) = \alpha I(\cdot, l_1) + \beta I(\cdot, l_2)\ . \tag{3.31}$$

Part of [FG06, Proposition 12.1] is that the pairing I is continuous. This follows from the fact that $I(m,\cdot)$ is a convex function of the coordinates $\mathcal{A}^t_\Gamma : \mathsf{T}^a_0(\mathbb{Q}) \to \mathbb{Q}^{E(\Gamma)}$, where $m \in \mathcal{T}^x(S,\mathbb{Q})$:

$$I(m,\mathcal{A}^t_\Gamma(l_1)) + I(m,\mathcal{A}^t_\Gamma(l_2)) \geq I(m,\mathcal{A}^t_\Gamma(l_1)+\mathcal{A}^t_\Gamma(l_2)) \ . \tag{3.32}$$

Since the mutation rules of Equation (3.22) are continuous, the cluster charts define a topology on $\mathsf{T}^a(S,\mathbb{Q})$. Its completion defines the space of real $\mathcal{A}$-laminations on S, denoted $\mathsf{T}^a(S,\mathbb{R})$. The cluster charts extend naturally to:

$$\mathcal{A}^t_\Gamma : \mathbb{R}^{E(\Gamma)} \longrightarrow \mathsf{T}^a(S,\mathbb{R}) \ . \tag{3.33}$$

Rational curve systems on S are dense in the space of transverse measured laminations, with respect to the topology of the length function which is exactly the pairing $I(\cdot,m)$ for $m \in \mathcal{T}^x(S)$. Hence we conclude that:

Corollary 3.17 (12.1 in [FG06])**.** *Elements of* $\mathsf{T}^a_0(S,\mathbb{R})$ *are Thurston's measured laminations, generalized to ciliated surfaces.*

Additive pairing between $\mathsf{T}^x(S,\mathbb{Q})$ and $\mathcal{T}^a_0(S)$. As in the previous case, the pairing

$$I : \mathsf{T}^x(S,\mathbb{Q}) \times \mathcal{T}^a_0(S) \longrightarrow \mathbb{R} \tag{3.34}$$

is defined recursively. If $l \in \mathsf{T}^x(S,\mathbb{Q})$ is a weight-one simple closed curve or a curve ending on the boundary of S but not on a hole, the pairing $I(l,\cdot)$ is the function which to each point of $\mathcal{T}^a_0(S)$ associates the length of l with respect to the hyperbolic structure corresponding to that point. If l has one of its ends or both ending on a hole, the pairing $I(l,\cdot)$ is the function which to each point of $\mathcal{T}^a_0(S)$ associates the signed length of the segment of the geodesic obtained after rotating infinitely many times in the direction prescribed by the orientation of the hole(s), and intersected with the horocycle(s) at that (those) holes, as in Figure 2.14.

If $l_1, l_2 \in \mathsf{T}^x(\mathbb{Q})$ and if l_1 and l_2 are non-intersecting, for all $\alpha,\beta \in \mathbb{Q}$ one sets:

$$I(\alpha l_1 + \beta l_2,\cdot) = \alpha I(l_1,\cdot) + \beta I(l_2,\cdot) \ . \tag{3.35}$$

Let us list some of the properties stated and proved in [FG06] and [FG07]. This pairing is also continuous. Moreover, let p and p denote the maps of Equation (2.51) and Equation (3.27), respectively. Then for $m \in \mathcal{T}^a_0(S)$ and $l \in \mathsf{T}^a(S,\mathbb{Q})$ one has:

$$I(p(m),l) = I(\mathsf{p}(l),m) \ . \tag{3.36}$$

In coordinates, if an $\mathcal{X}$-lamination l has positive coordinates $(x^\alpha)_{\alpha\in E_i(\Gamma)}$ in the chart $\mathcal{X}^t_\Gamma$ and if a point $m \in \mathcal{T}^a_0(S)$ has coordinates $(a_\alpha)_{\alpha\in E_i(\Gamma)}$ in the chart $\mathcal{A}_\Gamma$ corresponding to the same triangulation, then

$$I(l,m) = \sum_\alpha a_\alpha x_\alpha \ . \tag{3.37}$$

As in the previous paragraph, let us note the the space $\mathsf{T}^x(S,\mathbb{Q})$ can be completed into the space of real $\mathcal{X}$-laminations $\mathsf{T}^s(S,\mathbb{R})$, and the additive pairing of Equation (3.34) extends naturally to

$$I : \mathsf{T}^x(S,\mathbb{R}) \times \mathcal{T}^a_0(S) \longrightarrow \mathbb{R} \ . \tag{3.38}$$

Intersection pairing between $\mathsf{T}^x(S)$ and $\mathsf{T}^a_0(S)$. Last, the intersection pairing

$$\mathcal{I} : \mathsf{T}^x(S,\mathbb{Q}) \times \mathsf{T}^a_0(S,\mathbb{Q}) \longrightarrow \mathbb{R} \tag{3.39}$$

is defined as follows. Let $\gamma_1 \in \mathsf{T}^x(S,\mathbb{Q})$ and $\gamma_2 \in \mathsf{T}^a_0(S,\mathbb{Q})$ are two simple closed curves with weight one. If one or both ends of γ_1 lie on a segment of the boundary of S between two cilia, one moves it or them counterclockwise until the nearest cilium. Then one defines $\mathcal{I}(\gamma_1,\gamma_2)$ to be the minimal number of intersections between γ_1 and γ_2. If $l_1, l_2 \in \mathsf{T}^x(\mathbb{Q})$ and if l_1 and l_2 are non-intersecting, for all $\alpha,\beta \in \mathbb{Q}$ one sets:

$$I(\alpha l_1 + \beta l_2,\cdot) = \alpha I(l_1,\cdot) + \beta I(l_2,\cdot) \ . \tag{3.40}$$

If $l_1, l_2 \in \mathsf{T}^a_0(\mathbb{Q})$ and if l_1 and l_2 are non-intersecting, for all $\alpha, \beta \in \mathbb{Q}$ one sets:

$$I(\cdot, \alpha l_1 + \beta l_2) = \alpha I(\cdot, l_1) + \beta I(\cdot, l_2) \ . \tag{3.41}$$

The intersection pairing also extend to the real lamination spaces. It is a degeneration of the additive pairings defined above, in the following sense. Let Γ be a triangulation of S, and let $m \in \mathcal{T}^x(S)$ with coordinates $(x^\alpha)_{\alpha \in E_i(\Gamma)}$ in the chart $\mathcal{X}_\Gamma$. For any positive real number C, denote m^C the point in $\mathcal{T}^x(S)$ with coordinates $((x^\alpha)^C)_{\alpha \in E_i(\Gamma)}$. Let also $l \in \mathsf{T}^x(S, \mathbb{R})$ with coordinates $(\log x^\alpha)_{\alpha \in E_i(\Gamma)}$. Then

$$\lim_{C \to \infty} \frac{1}{C} I(m^C, \cdot) = \mathcal{I}(l, \cdot) \ . \tag{3.42}$$

The same holds with the roles of the $\mathcal{X}$- and $\mathcal{A}_0$-spaces exchanged.

3.3.2 Multiplicative pairings

Another way to construct functions on Teichmüller spaces from a closed curve γ on S is to consider the absolute value of the monodromy along γ with respect to the hyperbolic structure determined by the point in the Teichmüller space. This extends to laminations, and yields multiplicative pairings, which are algebraic functions on the Teichmüller space.

Multiplicative pairing between $\mathsf{T}^x(S, \mathbb{Z})$ and $\mathcal{T}^a_0(S)$. Following [FG06] and [FG07], the pairing

$$\mathbf{I} : \mathsf{T}^x(S, \mathbb{Z}) \times \mathcal{T}^a_0(S) \longrightarrow \mathbb{R}_{>0} \tag{3.43}$$

is defined as follows. If $\gamma \in \mathsf{T}^x(S, \mathbb{Z})$ is a single closed curve with weight k and if $m \in \mathcal{T}^a(S)$, then $\mathbf{I}(\gamma, m)$ is the absolute value of the trace of the monodromy along γ^k. If $\gamma \in \mathsf{T}^x(S, \mathbb{Z})$ is a single curve of weight k connecting two points of the boundary of S and if $m \in \mathcal{T}^a(S)$, then $\mathbf{I}(\gamma, m) = \exp I(\gamma, m)$.

Last, if $u, v \in \mathbb{Z}_{\geq 0}$, if $l_1, l_2 \in \mathsf{T}^x(S, \mathbb{Z})$ such that no curve in l_1 intersects or coincide with a curve in l_2, and if $m \in \mathcal{T}^a(S)$, then

$$\mathbf{I}(ul_1 + vl_2, m) = \mathbf{I}(l_1, m)^u \mathbf{I}(l_2, m)^v \ . \tag{3.44}$$

This pairing can be considered as a family of functions on $\mathcal{T}^a_0(S)$ indexed by $\mathcal{X}$-laminations. An important property is the following.

Proposition 3.18. *Let $(x^\alpha)_{\alpha \in E_i(\Gamma)}$ be the coordinates of an integral $\mathcal{X}$-lamination l in a chart $\mathcal{X}_\Gamma$. The function $\mathbf{I}(l, \cdot)$ on $\mathcal{T}^a_0(S)$ is a Laurent polynomial with positive integral coefficients in the coordinates $(a_\alpha)_{\alpha \in E_i(\Gamma)}$ in the chart $\mathcal{A}_\Gamma$ on $\mathcal{T}^a_0(S)$.*

In other words, the function $\mathbf{I}(l, \cdot)$ is a *universally positive Laurent polynomial* on $\mathcal{T}^a_0(S)$, in the sense that it a Laurent polynomial with positive coefficients in every cluster chart.

Multiplicative pairing between $\mathcal{T}^x(S)$ and $\mathsf{T}^a_0(S, \mathbb{Z})$. Following [FG06] and [FG07], the pairing

$$\mathbf{I} : \mathcal{T}^x(S) \times \mathsf{T}^a_0(S, \mathbb{Z}) \to \mathbb{R}_{>0} \tag{3.45}$$

is defined as follows. If $m \in \mathcal{T}^x(S)$ and if $\gamma \in \mathsf{T}^a(S, \mathbb{Z})$ is a single closed curve with weight k which is not retractable to a hole, then $\mathbf{I}(m, \gamma)$ is the absolute value of the trace of the monodromy along γ^k. If $m \in \mathcal{T}^x(S)$ and if $\gamma \in \mathsf{T}^a(S, \mathbb{Z})$ is a single closed curve with weight k which is retractable to a positively (resp. negatively) oriented hole, then $\mathbf{I}(m, \gamma)$ is the absolute value of the trace of the largest (resp. smallest) eigenvalue of the monodromy of γ^k.

Last, if $m \in \mathcal{T}^a(S)$, if $u, v \in \mathbb{Z}_{\geq 0}$ and if $l_1, l_2 \in \mathsf{T}^a(S, \mathbb{Z})$ such that no curve in l_1 intersects or coincide with a curve in $l_2 a$, then

$$\mathbf{I}(m, ul_1 + vl_2) = \mathbf{I}(m, l_1)^u \mathbf{I}(m, l_2)^v \ . \tag{3.46}$$

This pairing can be considered as a family of functions on $\mathcal{T}_x(S)$ indexed by $\mathcal{A}_0$-laminations. An important property is the following.

Proposition 3.19. *Let $(a_\alpha)_{\alpha \in E_i(\Gamma)} \in (\frac{1}{2}\mathbb{Z})^{E_i(\Gamma)}$ be the coordinates of an integral $\mathcal{A}_0$-lamination l in a chart $\mathcal{A}_\Gamma$. The function $\mathbf{I}(\cdot, l)$ on $\mathcal{T}^x(S)$ is a Laurent polynomial with positive integral coefficients in $((x^\alpha)^{1/2})_{\alpha \in E_i(\Gamma)}$, where $(x^\alpha)_{\alpha \in E_i(\Gamma)}$ are the coordinates in the chart $\mathcal{X}_\Gamma$ on $\mathcal{T}^a_0(S)$. It has highest term is $\prod_\alpha (x^\alpha)^{a_\alpha}$, and lowest one $\prod_\alpha (x^\alpha)^{-a_\alpha}$.*

In other words, the function $\mathbf{I}(\cdot, l)$ is a *universally positive Laurent polynomial* on $\mathcal{T}^x(S)$, in the sense that it a Laurent polynomial with positive coefficients in every cluster chart.

3.3.3 Duality conjectures (I)

In order to motivate the Conjecture 3.20, let us study a toy model consisting of the ciliated surface $S = S_{0,0,5}$ in which we have artificially frozen one of the two cluster variables. Since we are interested in $\mathcal{X}$-spaces and $\mathcal{A}_0$-spaces only, we completely forget about the frozen variables on the sides of the pentagon. There are only two charts on the spaces $\mathcal{T}^x(S)$, $\mathcal{T}_0^a(S)$, $\mathsf{T}^x(S)$ and $\mathsf{T}^a(S)$, related through the mutation at the unique mutable edge.

We will restrict to the pair of spaces $\mathcal{T}^x(S)$ and $\mathsf{T}_0^a(S,\mathbb{Z})$. Let x be the cluster variable and y the frozen variable in one chart on $\mathcal{T}^x(S)$. Likewise, let a be the cluster variable and b the frozen variable in one chart on $\mathcal{T}_0^a(S)$. Under the flip of the non-frozen edge, these variables transform according to the mutation formulae of type $\mathcal{X}$ and $\mathcal{A}^t$:

$$\left\{\begin{array}{l} x \\ y \end{array}\right. \longrightarrow \left\{\begin{array}{l} x' = x^{-1} \\ y' = y(1+x) \end{array}\right. , \qquad \left\{\begin{array}{l} a \\ b \end{array}\right. \longrightarrow \left\{\begin{array}{l} a' = \max(b,0) - a \\ b' = b \end{array}\right. . \tag{3.47}$$

Note that $y = y'(1+(x')^{-1})^{-1}$ and $a = \max(b',0) - a$. These mutations are depicted in Figure 3.20.

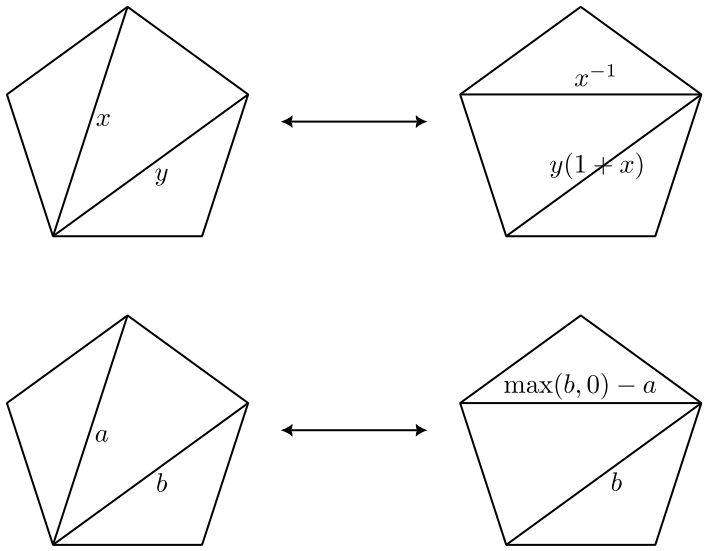


Figure 3.20: Coordinate changes under the only possible flip.

We are interested in studying the universally positive Laurent polynomials on $\mathcal{T}^x(S)$. Let us consider the monomial $x^k y^l$, which is clearly universally positive Laurent in the first chart. In terms of the variables in the second chart however, it writes $(x')^{-k}(y')^l(1+(x')^{-1})^{-l}$ and it is Laurent only if $l \leq 0$ (however it is still positive). If the latter condition holds, the highest order term reads $(x')^{-k}(y')^l = (x')^{\max(l,0)-k}(y')^l$.

Conversely, a monomial $(x')^m(y')^n$ with $m,n \in \mathbb{Z}$ can be rewritten as $x^{-m}y^n(1+x)^n$, and it is universally positive Laurent in both charts if and only if $n \geq 0$. If the latter condition holds, the highest order term is $x^{n-m}y^n = x^{\max(n,0)-m}y^n$.

Hence in this example the only universally positive Laurent polynomials in the cluster coordinates on $\mathcal{T}^x(S)$ are those such that the exponents of the highest order term (as well as the lowest order one) transform according to the tropical $\mathcal{A}$-mutation formulae, i.e. as the coordinates of laminations in $\mathcal{A}_0(S,\mathbb{Z})$. This motivates the following conjecture, that we will discuss more in the next chapter.

Conjecture 3.20 (12.2 in [FG06])**.** *Here $\mathcal{T}(S)$ stands for either version of Teichmüller spaces of S. Let $\mathbb{L}(\mathcal{T}(S))$ be the ring of algebraic functions on $\mathcal{T}(S)$ which are Laurent polynomials in every cluster chart, and let $\mathbb{L}_+(\mathcal{T}(S))$ be the sub-semiring of those polynomials which moreover have positive integer coefficients only in every cluster chart. It is a cone, of which we denote $\mathbb{E}(\mathcal{T}(S))$ the set of extremal elements in $\mathbb{L}_+(\mathcal{T}(S))$. Then:*

$$\mathcal{A}_0(S,\mathbb{Z}) \simeq \mathbb{E}(\mathcal{T}^x(S)) , \qquad \mathsf{T}^x(S,\mathbb{Z}) \simeq \mathbb{E}(\mathcal{T}_0^a(S)) , \tag{3.48}$$

where $\mathcal{A}_0(S,\mathbb{Z})$ is, as before, the space of $\mathcal{A}$-laminations in $\mathsf{T}_0^a(S,\mathbb{Z})$ whose coordinates in one (equivalently, every) cluster chart are integers.

* * * * * * *

We have defined two different versions of laminations on ciliated surfaces, namely $\mathcal{A}$-laminations and $\mathcal{X}$-laminations. Both spaces admit coordinate systems associated with triangulations of the underlying surfaces, and coordinates are naturally associated to the edges of the triangulation. Under the flip of an internal edge, they undergo cluster mutations of type $\mathcal{A}$ and $\mathcal{X}$, respectively.

Teichmüller and laminations spaces are related through additive and multiplicative pairings: an $\mathcal{A}_0$-lamination (resp. $\mathcal{X}$-lamination) provides naturally a function on the $\mathcal{X}$- (resp. $\mathcal{A}$-) Teichmüller space. Laminations whose coordinates are integers play a special role. The functions on Teichmüller spaces they define are extremal universally positive Laurent polynomials. Conjecturally, the converse also holds: every extremal universally positive Laurent polynomial on either Teichmüller space corresponds to a point in the other lamination space.

In the next chapter we will introduce cluster varieties, which allow to consider uniformly the Teichmüller and lamination space of one kind, as the space of points over different semifields of a single cluster variety. Many of the properties introduced above extend naturally to this general setup.

Chapter 4

Higher Teichmüller theory

In Chapter 1, we introduced the totally positive Grassmanian $\mathrm{Gr}^+_{2,m}$ and the semigroup $(\mathrm{SL}_n)_{>0}$ of totally positive matrices in (SL_n). Both are subspaces in an algebraic variety over $\mathbb{C}$ – respectively $\mathrm{Gr}_{2,m}(\mathbb{C})$ and $\mathrm{SL}_n(\mathbb{C})$, on which the cluster variables take real positive values. The Teichmüller spaces $\mathcal{T}^x(S)$ and $\mathcal{T}^a(S)$ of a ciliated surface S were introduced in Chapter 2, and are parameterized by positive real numbers on the edges of a triangulation. Importantly, the fact the elements of a single cluster are real positive at a point ensures that the same holds for every cluster variable. This points towards the definition of *totally positive varieties*.

More generally, one can wonder what the algebraic variety corresponding to a cluster algebra is, and how the positivity properties of the latter are encoded geometrically. In the first part of this chapter we will tackle this question and introduce *cluster ensembles*, which consist of a pair $(\mathcal{X}, \mathcal{A})$ of *cluster varieties* associated with a mutation class of a seed, as introduced in [FG06, FG09]. Both cluster varieties are *positive varieties*, which implies that one can consider their points $(\mathcal{X}(\mathbb{S}), \mathcal{A}(\mathbb{S}))$ over any semifield $\mathbb{S}$ – as expected. In this framework, both Teichmüller spaces $\mathcal{T}^x(S)$ and $\mathcal{T}^a(S)$ are naturally identified with the $\mathbb{R}_{>0}$-points of the elements of a cluster ensemble. The $\mathcal{X}$-space is endowed with a canonical Poisson structure, and the $\mathcal{A}$ space, with a canonical closed 2-form. In Section 4.1 we discuss positive algebraic geometry. Then, we define the cluster ensemble associated to a seed in Section 4.2, and present some of its properties in Section 4.2.3. We follow [FG09] closely.

Using Lusztig's generalization of total positivity to an arbitrary reductive Lie group G, one can consider the spaces $(\mathcal{X}^+_{G,S}, \mathcal{A}^+_{G,S})$ obtained as the $(\mathbb{R}_{>0})$-points of a cluster ensemble. They are analogous to the Teichmüller spaces introduced in Chapter 2 except that now G plays the role of $\mathrm{PSL}_2(\mathbb{R})$. Just as the Teichmüller space $\mathcal{T}(S)$ of a ciliated surface S is a connected component of $\mathrm{Hom}(\pi_1(S), \mathrm{PSL}_2(\mathbb{R}))/\mathrm{PSL}_2(\mathbb{R})$ consisting of discrete and faithful representations, both $\mathcal{X}^+_{G,S}$ and $\mathcal{A}^+_{G,S}$ project to the space of positive representations $\mathcal{L}^+_{G,S}$, which is a connected component of the *G-character variety* $\mathrm{Hom}(\pi_1(S), G)/G$ of S, consisting solely of discrete and faithful representations. Defining these spaces, which is the main goal of this part of the dissertation, will be done in Section 4.3, following [FG06].

When G is a complex reductive Lie group or a compact real Lie group, the connected components of the G-character variety of a surface S are in one-to-one correspondence with the elements of $\pi_1(G)$ [AB83, Li93]. When G is split real, the situation is drastically different: there are additional connected components in the G-character variety, such as the Hitchin components. A connected component of $\mathrm{Hom}(\pi_1(S), G)/G$ consisting of discrete and faithful representations is called a *G-higher Teichmüller space*. For example, the spaces $\mathcal{L}^+_{G,S}$ are higher Teichmüller spaces. One can wonder whether higher Teichmüller spaces exist for real Lie groups G which are not split real.

It turns out to be the case when G is a real Lie group of Hermitian type [BIW10]: the spaces of *maximal representations* $\pi_1(S) \to G$ are unions of higher Teichmüller spaces. Recently, positive and maximal representations were both described as special cases of Θ-positive representations [GW18]. The spaces of Θ-positive representations are (unions of) higher Teichmüller spaces, and exist for real Lie groups G which are neither split nor of Hermitian type, thereby leading to new families of higher Teichmüller spaces [GLW21]. This story will be reviewed in Section 4.4.

4.1 Positive algebraic geometry

4.1.1 Generalities

Let $\Gamma_m = \mathrm{Spec}\ \mathbb{Z}[X, X^{-1}]$ denote the *multiplicative group scheme*: over any field $\mathbb{F}$ one has $\Gamma_m(\mathbb{F}) = \mathbb{F}^\times$. A *split algebraic torus* H is a scheme Γ_m^n, where $n \in \mathbb{Z}_{>0}$. Over $\mathbb{C}$ for example one has $\Gamma_m^n = (\mathbb{C}^\times)^n$, with ring of regular functions $\mathcal{O}_{\Gamma_m^n} = \mathbb{C}[X_1^{-1}, \dots, X_n^{\pm 1}]$, i.e. the ring of Laurent polynomials, and function field $K_H = \mathbb{C}(X_1, \dots, X_n)$, i.e. the field of rational fractions in n indeterminates.

To every split algebraic torus $H = \Gamma_m^n$ correspond two rank-n lattices (equivalently, abelian groups of finite rank). The *lattice of characters* is $X^\bullet(H) = \mathrm{Hom}(H, \Gamma_m)$ and the *lattice of cocharacters* is $X_\bullet(H) = \mathrm{Hom}(\Gamma_m, H)$. Characters are regular functions on H. Over $\mathbb{C}$, they are:

$$H(\mathbb{C}) \ni (t_1, \dots, t_n) \longmapsto \prod_i t_i^{\alpha_i} \in \mathbb{C}^\times \ , \tag{4.1}$$

for $(\alpha_1, \dots, \alpha_n)$ and n-tuple of integers. Likewise, cocharacters of $H(\mathbb{C})$ correspond to n-tuples of integers $(\beta_1, \dots, \beta_n)$:

$$\mathbb{C}^\times \ni t \longmapsto (t^{\beta_1}, \dots, t^{\beta_n}) \in H(\mathbb{C}) \ . \tag{4.2}$$

A rational function $f \in K_H$ on a split algebraic torus $H = \Gamma_m^n$ is said to be *positive* if it belongs to the sub-semifield generated by the characters, i.e. if it can be written as the ratio of two elements of $\mathbb{Z}_{>0}[X^\bullet(H)]$. A rational map $H_1 \to H_2$ between two algebraic tori is a *positive rational map* if f^* induces a morphism of the semifields of positive rational functions. For example, the map

$$\begin{array}{ccc} \Gamma_m^2 & \longrightarrow & \Gamma_m^2 \\ (x, y) & \longmapsto & (xy + 1, xy^{-1}) \end{array} \tag{4.3}$$

is a positive regular map, while $(x, y) \mapsto (x - y, y)$ is not. Importantly, mutations such as:

$$(a_1, a_2, a_3) \longmapsto \left(\frac{a_2 a_3 + 1}{a_1}, a_2, a_3 \right) \ , \qquad (x_1, x_2, x_3) \longmapsto \left(\frac{x_1}{1 + x_2^{-1}}, \frac{1}{x_2}, x_3(1 + x_2) \right) \tag{4.4}$$

are positive rational maps. Note that the inverse of a positive map is not necessarily positive: $(x, y) \mapsto (x + y, y)$ is positive but its inverse $(x', y') \mapsto (x' - y', y')$ is not. A *positive divisor* on a split algebraic torus H is a divisor $f = 0$ where $f \in K_H$ is a positive rational function. Let **Pos** be the category whose objects are split algebraic tori and whose morphisms are positive rational functions.

Let now X be an irreducible scheme or stack over $\mathbb{Q}$. A *positive atlas* on X is a family of birational isomorphisms

$$\psi_\alpha : H_\alpha \to X \ , \qquad \alpha \in C_X \ , \tag{4.5}$$

where C_X is a non-empty set, and where the H_α are split algebraic tori, such that each ψ_α is regular on the complement of a positive divisor in H_α, and such that for all $\alpha, \beta \in C_X$, the map

$$\psi_\beta^{-1} \circ \psi_\alpha : H_\alpha \to H_\beta \tag{4.6}$$

is a positive rational map. The maps $\psi_\alpha : H_\alpha \to X$ are said to be *positive coordinate systems* or *positive charts* on X, and a scheme X endowed with a positive atlas is said to be a *positive scheme*. If Γ is a group of automorphism on X, one says that a positive atlas is Γ-equivariant if Γ acts on C_X and if each $\gamma \in \Gamma$ induces an isomorphism $i_\gamma : H_\alpha \to H_{\gamma(\alpha)}$ such that $\psi_{\gamma(\alpha)} \circ i_\gamma = \gamma \circ \psi_\alpha$.

Now we can state the definition of a positive space as in [FG09, Definition 1.3]. Recall from Section 2.2.3 that a *groupoid* is a category where morphisms are isomorphisms, and where any two objects are isomorphic. The *fundamental group of a groupoid* is the automorphism group of any object in the category. Let $\mathcal{G}_X$ be a groupoid. A *positive space* is a functor

$$\psi_X : \mathcal{G}_X \to \mathbf{Pos} \ . \tag{4.7}$$

If X is a positive scheme with the notation of Equation (4.5), the category $\mathcal{G}_X$ has the positive charts as objects hence $\mathrm{Ob}(\mathcal{G}_X) = C_X$, and the morphisms are given by the elements in $C_X \times C_X$ consisting of pairs of charts with non-zero intersection. The functor ψ_X satisfies $\psi_X(\alpha) = H_\alpha$ and $\psi_x(\alpha \to \beta) = \psi_\beta^{-1} \circ \psi_\alpha$. Let Π_X be the fundamental group of $\mathcal{G}_X$. A positive space $\mathcal{G}_X \to \mathbf{Pos}$ gives rise to a (non-necessarily

separated) scheme X^* endowed with a Π_X-equivariant positive atlas, obtained by gluing together the algebraic split tori H_α along the birational maps $\psi_\beta^{-1} \circ \psi_\alpha$.

The ring of regular functions $\mathbb{L}(X)$ on X^* consists of all rational functions which are regular on every coordinate torus H_α; it is called the ring of *universally Laurent polynomials* on X. Since X^* admits a positive structure, there is a well-defined semifield $\mathbb{Q}_+(X)$ of positive rational functions over X. Intersecting it with $\mathbb{L}(X)$ yields the semiring denoted $\widetilde{\mathbb{L}}_+(X)$.

A regular rational function might be positive without the coefficients of the corresponding Laurent polynomials being necessarily positive as well. For example:

$$\frac{1+x^3}{1+x} = 1 - x + x^2 \ . \tag{4.8}$$

Let $\mathbb{L}_+(X)$ be the sub-semiring of $\widetilde{\mathbb{L}}_+(X)$ consisting of the regular positive rational functions on X whose restriction to one (equivalent, any) positive chart is a Laurent polynomial with positive integral coefficients. It is called the semiring of *positive universally Laurent polynomials* on X. Let $\mathbb{E}(X)$ be the set of extremal elements in $\mathbb{L}_+(X)$, i.e. those elements that cannot be written as the sum of two non-zero elements of $\mathbb{L}_+(X)$.

4.1.2 Points over a semifield and tropical compactifications

An important property of positive spaces is that one can consider their points valued in any semifield. If H is an algebraic split torus and if $(\mathbb{S}, \oplus, \odot)$ is an arbitrary semifield defined as in Section 1.3, the set of $\mathbb{S}$-valued points of H is defined as

$$H(\mathbb{S}) = X_\bullet(H) \otimes_{\mathbb{Z}} \mathbb{S} \ , \tag{4.9}$$

where the tensor product is with respect to the abelian group $(\mathbb{S}, \odot)$. Any positive birational isomorphism $\psi : H_1 \to H_2$ induces a map $\psi_* : H_1(\mathbb{S}) \to H_2(\mathbb{S})$. If the reciprocal ψ^{-1} of ψ is positive as well, then ψ_* is an isomorphism. For example, the map $\Gamma_m^3 \to \Gamma_m^3$ defined by $a' = a^{-1}(bc+1)$, $b' = b$, $c' = c$ becomes the piecewise linear map $\alpha' = \max(\beta+\gamma, 0) - \alpha$, $\beta' = \beta$, $\gamma' = \gamma$ between the set of $\mathbb{R}^t$-points $\Gamma_m^3(\mathbb{R}^t) \to \Gamma_m^3(\mathbb{R}^t)$, where $\alpha, \beta, \gamma, \alpha', \beta', \gamma' \in \mathbb{R}$.

If X is a positive space, one defines the set of $\mathbb{S}$-points of X as:

$$X(\mathbb{S}) = \coprod_\alpha H_\alpha(\mathbb{S}) / \sim \ , \tag{4.10}$$

where $\sim$ denotes the identifications induced by the isomorphisms $(\psi_\beta^{-1} \circ \psi_\alpha)_*$.

Let X be a positive variety. As advertised in Section 3.1.5, we will now explain, following the lines of [FG06, Section 4.5], why the space $X(\mathbb{R}^t)$ provides a spherical compactification of $X(\mathbb{R}_{>0})$. Let $A = \mathbb{Q}$ or $\mathbb{R}$, and let $H = \Gamma_m^n$ be a split algebraic torus. The set of A^t-points of H is the abelian group $(A^n, +)$, with 0 as neutral element. The multiplicative group $(A_{>0}, \times)$ acts naturally on it. The spherization $\mathbf{Sph}H(A^t)$ is defined as

$$\mathbf{Sph}H(A^t) = \big(H(A^t) - \{0\}\big) / A_{>0} \ . \tag{4.11}$$

It is a sphere, as the name suggests.

Let $\phi : H \to H'$ be a positive rational map, and let $\phi_* : H(A^t) \to H'(A^t)$ be the corresponding piecewise linear map. It satisfies $\phi_*(0) = 0$, and the action of $A_{>0}$ commutes with ϕ_*. Hence ϕ_* induces a map

$$\overline{\phi}_* : \mathbf{Sph}H(A^t) \to \mathbf{Sph}H'(A^t) \ . \tag{4.12}$$

Hence one can defined the spherization of $X(\mathcal{A}^t)$ as

$$\mathbf{Sph}X(A^t) = \{\{p_\alpha \in \mathbf{Sph}H_\alpha(A^t)\} \mid \overline{\psi}_{\alpha,\beta}(p_\alpha) = p_\beta \ \ \forall \alpha, \beta \in \mathcal{G}_X\} \ . \tag{4.13}$$

Let now $\mathbb{R}((\epsilon^{-1}))_{>0}$ be the semifield of formal Laurent series $\sum_N^\infty a_s \epsilon^{-s}$ with $N \in \mathbb{Z}$ and $a_N \in \mathbb{R}_{>0}$. Each element of $\mathbb{R}((\epsilon^{-1}))_{>0}$ can be interpreted as a function to $\mathbb{R}_{>0}$ defined when $\epsilon \in \mathbb{R}_{>0}$ is real and large enough. Likewise, the elements of $H(\mathbb{R}((\epsilon^{-1}))_{>0})$ modulo reparametrizations $\epsilon^{-1} \to f(\epsilon^{-1})$ for $f \in \mathbb{R}((\epsilon^{-1}))_{>0}$ can be interpreted as lines in $H(\mathbb{R}_{>0})$ defined only in a neighborhood of infinity. The asymptotics of these lines is encoded in the degree $-\deg : \sum_N^\infty a_s \epsilon^{-s} \to -N$, which induces a map

$$-\deg : X(\mathbb{R}((\epsilon^{-1}))_{>0}) \longrightarrow X(\mathbb{Z}^t) \ . \tag{4.14}$$

The image of any element of $X(\mathbb{R}((\epsilon^{-1}))_{>0})$ under it projects to $\mathbf{Sph}X(\mathbb{Q}^t)$ if its degree is not zero. Hence the completion $\mathbf{Sph}X(\mathbb{R}^t)$ of $\mathbf{Sph}X(\mathbb{Q}^t)$ provides a spherical boundary for $\mathcal{X}(\mathbb{R}_{>0})$, to which the action of the fundamental group Π_X of the groupoid $\mathcal{G}_X$ on $\mathcal{X}(\mathbb{R}_{>0})$ extends, in a natural way.

4.2 Cluster ensembles

In the last section we have explained how to obtain positive schemes from positive spaces by gluing algebraic split tori along birational isomorphisms. Before defining cluster ensembles formally, let us briefly recollect some facts from the previous chapters.

Cluster algebras are defined starting with an initial seed consisting of a (generalized) quiver together with a set of variables, either mutable or frozen, associated to the vertices of that quiver. By mutating the latter in all possible ways, one generates all cluster variables. These, together with the frozen variables, generate the cluster algebra as a polynomial ring. The Laurent phenomenon implies that the cluster algebra is a subring of the Laurent algebra in the cluster variables of any seed, tensored with the polynomial algebra in the frozen variables.

We have introduced two versions of Teichmüller spaces of ciliated surfaces, both endowed with coordinate systems associated with triangulations of the ciliated surface, in which the transition functions are positive rational maps. More precisely, the Teichmüller spaces are charted by copies of $(\mathbb{R}_{>0})^{E_i(\Gamma)}$ or $(\mathbb{R}_{>0})^{E(\Gamma)}$ depending on the version, where Γ is a triangulation. Denoting by N either $E_i(\Gamma)$ or $E(\Gamma)$, the space $(\mathbb{R}_{>0})^N$ is the set of $\mathbb{R}_{>0}$-points of the algebraic split torus Γ_m^N. Regarding rational laminations spaces, they are charted by copies of $\mathbb{Q}^N$ which is the set of $\mathbb{Q}^t$-points of the algebraic split torus Γ_m^N, and likewise for laminations with integral or real coordinates.

This suggests that there exists a pair of positive varieties $(\mathcal{X}_S, \mathcal{A}_S)$ for any ciliated surface S, both endowed with a positive atlas in which the charts correspond to the triangulations of S, i.e. $C_{\mathcal{X}} = C_{\mathcal{A}}$ is the set of all triangulations of S. This would yield the identifications

$$\mathcal{X}_S(\mathbb{R}_{>0}) = \mathcal{T}^x(S)\ , \quad \mathcal{X}_S(\mathbb{Q}^t) = \mathsf{T}^x(S, \mathbb{Q})\ , \quad \mathcal{A}_S(\mathbb{R}_{>0}) = \mathcal{T}^a(S)\ , \quad \mathcal{A}_S(\mathbb{Q}^t) = \mathsf{T}^a(S, \mathbb{Q})\ . \tag{4.15}$$

It turns out to be the case, as we are going to discuss now. In fact for every seed s, as in the case of cluster algebras, one can construct a pair of positive varieties $(\mathcal{X}_{[\mathsf{s}]}, \mathcal{A}_{[\mathsf{s}]})$ which depends only on the mutation class $[\mathsf{s}]$ of s. Both positive spaces are called *cluster varieties*, and the pair is said to be the *cluster ensemble* corresponding to $[\mathsf{s}]$.

4.2.1 Cluster $\mathcal{X}$- and $\mathcal{A}$-tori

Since we need algebraic split tori in order to define positive varieties, we first need to understand how to obtain such objects from a seed. This will come naturally upon a slight change of perspective on the definition of seeds from Chapter 1. Let us rewrite here, for convenience, the Definition 1.13 of skew-symmetrizable matrices with integer entries:

Definition 4.1. *An $n \times n$ matrix ϵ with integer entries is skew-symmetrizable if there exists $d_1, ..., d_n \in \mathbb{Z}$ called the multipliers, such that $d_i \epsilon_{ij} = -d_j \epsilon_{ji}$. Equivalently one could write $\epsilon_{ij} d_j^{-1} = -\epsilon_{ji} d_i^{-1}$. An extended skew-symmetrizable matrix is an $m \times n$ matrix with $m \geq n$ such that its top $n \times n$ sub-matrix is skew-symmetrizable.*

Definition 4.2. *A seed is a quadruple $(I, I_0, \epsilon, \{d_i\})$ such that*

i) I is a finite set and $I_0 \subset I$,

ii) $\epsilon = (\epsilon_{ij})$ is a $\mathbb{Q}$-valued function on $I \times I$ such that $\epsilon_{ij} \in \mathbb{Z}$ when $i, j \in I - I_0$,

iii) $\{d_i\}_{i \in I}$ is a set of positive rational numbers such that $(\epsilon_{ij} d_j^{-1})$ is skew-symmetric.

Since we wish to have an algebraic split torus associated with each seed, we consider the lattice (i.e. the free abelian group) generated by elements in one-to-one correspondence with the rows of the extended exchange matrix, since the categories of finite rank lattices and of split algebraic tori are mutually dual. The definition of a seed is naturally rewritten as:

Definition 4.3. *A seed is a quadruple $\mathsf{s} = (\Lambda, (*, *), \{e_i\}, \{d_i\})$, where:*

i) Λ *is a lattice,*

ii) $(*,*)$ *is a skew-symmetric* $\mathbb{Q}$*-valued bilinear form on* Λ*,*

iii) $\{e_i\}_{i\in I}$ *is a basis of the lattice* Λ*, and there is a subset* I_0 *of* I *whose elements are said to be frozen,*

iv) $\{d_i\}_{i\in I}$ *are positive integers such that for all* $(i,j)\in I^2\backslash I_0^2$ *one has* $\epsilon_{ij}:=(e_i,e_j)d_j\in\mathbb{Z}$*.*

Cluster $\mathcal{X}$-torus of a seed. The lattice Λ is dual to the split algebraic torus $\mathcal{X}_\Lambda=\mathrm{Hom}(\Lambda,\Gamma_m)$. Any $v\in\Lambda$ defines a character $X_v\in\mathrm{Hom}(\mathcal{X}_\Lambda,\Gamma_m)$, whose value at $x\in\mathcal{X}_\Lambda$ is $x(v)$. Conversely, any $a\in\Lambda^*=\mathrm{Hom}(\Lambda,\mathbb{Z})$ gives rise to a cocharacter $\phi_a\in\mathrm{Hom}(\Gamma_m,\mathcal{X}_\Lambda)$. For any field $\mathbb{F}$, at the level of $\mathbb{F}$-points if $f\in\Gamma_m(\mathbb{F})=\mathbb{F}^\times$ then $\phi_a(f)$ is the homomorphism $v\to f^{a(v)}$.

Definition 4.4. *The split algebraic torus* $\mathcal{X}_\Lambda$ *is called the* $\mathcal{X}$*-torus corresponding to the seed* s*. It carries a natural log-canonical Poisson structure induced by the form* $(*,*)$*:*

$$\{X_v,X_w\}=(v,w)X_vX_w\ . \tag{4.16}$$

The basis $\{e_i\}$ *yields special characters* $\{X_i\}$ *dubbed cluster* $\mathcal{X}$*-coordinates. They form a basis in the abelian group of characters of* $\mathcal{X}_\Lambda$*. Note that accordingly to [FG09] but contrarily to what we have done above, this defines cluster* $\mathcal{X}$*-coordinates associated with the frozen vectors. In general we will forget about those.*

Cluster $\mathcal{A}$-torus of a seed. Let $\{e_i^*\}_{i\in I}$ be the basis of Λ^* dual to $\{e_i\}$. For all $i\in I$ let $f_i=d_i^{-1}e_i$, and let $\Lambda^\circ\subset\Lambda^*\otimes_{\mathbb{Z}}\mathbb{Q}$ be the lattice generated by the f_i. Let $\mathcal{A}_\Lambda=\mathrm{Hom}(\Lambda^\circ,\Gamma_m)$. As before, every $v\in\Lambda^\circ$ defines a character on $\mathcal{A}_\Lambda$ and every $a\in(\Lambda^\circ)^*$ defines a cocharacter on $\mathcal{A}_\Lambda$. The skew-symmetric $\mathbb{Q}$-valued bilinear form on Λ can be naturally interpreted as an element of the antisymmetric tensor space $\bigwedge^2\Lambda^\circ$, and therefore it defines an element $W\in\mathcal{O}(\mathcal{A}_\Lambda)^*\wedge\mathcal{O}(\mathcal{A}_\Lambda)^*$. The image of W under the map

$$\mathrm{d}\log\wedge\mathrm{d}\log : O(\mathcal{A}_\Lambda)^*\wedge\mathcal{O}(\mathcal{A}_\Lambda)^*\longrightarrow\Omega^2(\mathcal{A}_\Lambda) \tag{4.17}$$

is a closed two-form Ω on $\mathcal{A}_\Lambda$.

Definition 4.5. *The split algebraic torus* $\mathcal{A}_\Lambda$ *is called the* $\mathcal{A}$*-torus corresponding to the seed* s*. It carries a closed 2-form* Ω *induced by the form* $(*,*)$*. The basis* $\{f_i\}$ *provides the cluster* $\mathcal{A}$*-coordinates* $\{A_i\}$*.*

Moreover, the natural map $p^*:\Lambda\to\Lambda^\circ$, defined by:

$$v\longmapsto\sum_j(v,e_j)e_j^*=\sum_j[(v,e_j)d_j]f_j\ . \tag{4.18}$$

yields a homomorphism $p:A_\Lambda\to\mathcal{X}_\Lambda$.

Proposition 4.6 (Lemma 1.5 in [FG09])**.** *The fibers of* p *are the leaves of the null-foliation of the closed 2-form* Ω*. The image* $p(\mathcal{A}_\Lambda)$ *is a symplectic leaf of the Poisson structure on* $\mathcal{X}_\Lambda$*, and the symplectic structure on* $p(\mathcal{A}_\Lambda)$ *induced by* Ω *coincides with the one induced by restriction of the Poisson structure.*

Let $\epsilon_{ij}=(e_i,e_j)d_j$ be the exchange matrix corresponding to the seed s. In coordinates, the Poisson structure on the $\mathcal{X}$-torus reads:

$$\{X_i,X_j\}=\frac{\epsilon_{ij}}{d_j}X_iX_j\ , \tag{4.19}$$

the 2-form Ω on $\mathcal{A}_\Lambda$ writes:

$$\Omega=\sum_{i,j\in I}d_i\epsilon_{ij}d\log A_i\wedge d\log A_j\ , \tag{4.20}$$

and the homomorphism from $\mathcal{A}_\Lambda$ to $\mathcal{X}_\Lambda$ is:

$$p^*X_i=\prod_{j\in I}A_j^{\epsilon_{ij}}\ . \tag{4.21}$$

These last equations are the analogues of Equations (2.50), (2.51), (2.66) and (3.27).

4.2.2 Cluster transformations

Among all positive varieties, cluster varieties are singled out by the choice of birational isomorphisms $\psi_\beta^{-1} \circ \psi_\alpha$ as in Equation (4.6): they are compositions of cluster mutations.

Definition 4.7. *Let $k \in [|1, n|]$. The mutation of a seed $\mathsf{s} = (\Lambda, (*, *), \{e_i\}, \{d_i\})$ in the direction k is a seed $\mu_k(\mathsf{s}) = (\Lambda, (*, *), \{e'_i\}, \{d_i\})$, where*

$$e'_i = \begin{cases} e_i + [\epsilon_{ik}]_+ e_k & \text{if} \quad i \neq k \\ -e_k & \text{if} \quad i = k \end{cases} . \tag{4.22}$$

Denoting $\epsilon'_{ij} = (e'_i, e'_j)d_j$ the mutated exchange matrix, one has

$$(\epsilon')^{\beta\gamma} = \begin{cases} -\epsilon^{\beta\gamma} & \text{if } \alpha = \beta \text{ or } \alpha = \gamma \\ \epsilon^{\beta\gamma} + \mathrm{sgn}(\epsilon^{\beta\alpha})[\epsilon^{\beta\alpha}\epsilon^{\alpha\gamma}]_+ & \text{otherwise} \end{cases} , \tag{4.23}$$

as in Equation (1.14). As we have already emphasized, the mutation is involutive on the exchange matrix (but not necessarily on the basis, see [FG09, remark below Eq. 9]).

Any seed mutation in the direction k induces a positive rational map between the corresponding seed $\mathcal{X}$- and $\mathcal{A}$- tori. Let us denote X'_i (resp. A'_i) the cluster coordinates on the seed torus $\mu_k(\mathcal{X}_\Lambda)$ (resp. $\mu_k(\mathcal{A}_\Lambda)$), and set

$$\mu_k^* X'_i := \begin{cases} X_k^{-1} & \text{if } i = k \\ X_i(1 + X_k^{\mathrm{sgn}(\epsilon_{ik})})^{-\epsilon_{ik}} & \text{if } i \neq k \end{cases} , \tag{4.24}$$

as in Equation (1.68), and

$$A_k \cdot \mu_k^* A'_k = \prod_{j:\epsilon_{kj}>0} A_j^{\epsilon_{kj}} + \prod_{j:\epsilon_{kj}<0} A_j^{-\epsilon_{kj}} , \qquad \mu_k^* A'_i = A_i \text{ for } i \neq k \tag{4.25}$$

as in Equation (1.24). An isomorphism of seeds induces isomorphisms between the corresponding $\mathcal{X}$- and $\mathcal{A}$-tori. A *seed cluster transformation* is a composition of seed isomorphisms and mutations. It gives rise to a *cluster transformation* of the corresponding $\mathcal{X}$- and $\mathcal{A}$- tori. A seed cluster transformation is trivial if the corresponding cluster transformations of $\mathcal{X}$- and $\mathcal{A}$-tori are trivial.

Let s be a seed. The *cluster modular groupoid* $\mathcal{G}_{[\mathsf{s}]}$ is the category whose objects are all the seeds mutation equivalent to s, and where morphisms are the cluster transformations modulo the trivial ones. The fundamental group $\Pi_{[\mathsf{s}]}$ of $\mathcal{G}_{[\mathsf{s}]}$ is the *cluster modular group* of the class.

The functor from $\mathcal{G}_{[\mathsf{s}]}$ to the category whose objects are $\mathcal{X}$-tori and morphisms cluster transformations is a positive space denoted $\mathcal{X}_{[\mathsf{s}]}$. Similarly, the functor from $\mathcal{G}_{[\mathsf{s}]}$ to the category whose objects are $\mathcal{A}$-tori and morphisms cluster transformations is a positive space denoted $\mathcal{A}_{[\mathsf{s}]}$. The pair of positive spaces $(\mathcal{X}_{[\mathsf{s}]}, \mathcal{A}_{[\mathsf{s}]})$ is the cluster ensemble associated to the mutation class of the seed s. They give rise to positive varieties called cluster varieties, endowed with a $\Pi_{[\mathsf{s}]}$-equivariant positive atlas.

Let S be a ciliated surface and Γ a triangulation of S. Let s be the seed obtained from Γ as in Section 1.2 and Section 2.2.2. For reasons that will become clear in the next chapter, the corresponding cluster varieties are denoted $\mathcal{X}_{S,\mathrm{SL}_2}$ and $\mathcal{A}_{S,\mathrm{SL}_2}$. One deduces from the previous chapters that:

$$\mathcal{X}_{\mathrm{SL}_2,S}(\mathbb{R}_{>0}) = \mathcal{T}^x(S) , \quad \mathcal{X}_{\mathrm{SL}_2,S}(\mathbb{Q}^t) = \mathsf{T}^x(S, \mathbb{Q}) , \quad \mathcal{X}_{\mathrm{SL}_2,S}(\mathbb{Z}^t) = \mathsf{T}^x(S, \mathbb{Z}) , \tag{4.26}$$

as well as

$$\mathcal{A}_{\mathrm{SL}_2,S}(\mathbb{R}_{>0}) = \mathcal{T}^a(S) , \quad \mathcal{A}_{\mathrm{SL}_2,S}(\mathbb{Q}^t) = \mathsf{T}^a(S, \mathbb{Q}) . \tag{4.27}$$

The space $\mathcal{A}_{\mathrm{SL}_2,S}(\mathbb{Z}^t)$ (resp. $\mathcal{A}_{\mathrm{SL}_2,S}(\frac{1}{2}\mathbb{Z}^t)$) consists of the rational $\mathcal{A}$-laminations on S whose coordinates in one (equivalently, every) cluster chart are in $\mathbb{Z}$ (resp. $\frac{1}{2}\mathbb{Z}^t$). We have already seen that:

$$\mathcal{A}_{\mathrm{SL}_2,S}(\frac{1}{2}\mathbb{Z}^t) \subset \mathsf{T}^a(S, \mathbb{Z}) \subset \mathcal{A}_{\mathrm{SL}_2,S}(\mathbb{Z}^t) . \tag{4.28}$$

The reasoning of Section 4.1.2 implies that $\mathbf{Sph}\mathcal{X}_{\mathrm{SL}_2,S}(\mathbb{R}^t)$ (resp. $\mathbf{Sph}\mathcal{A}_{\mathrm{SL}_2,S}(\mathbb{R}^t)$) provides a spherical boundary for $\mathcal{X}_{\mathrm{SL}_2,S}(\mathbb{R}_{>0})$ (resp. $\mathcal{A}_{\mathrm{SL}_2,S}(\mathbb{R}_{>0})$) to which the action of the cluster modular group on $\mathcal{X}_{\mathrm{SL}_2,S}(\mathbb{R}_{>0})$ (resp. $\mathcal{A}_{\mathrm{SL}_2,S}(\mathbb{R}_{>0})$) extends, as in Thurston's compactification of the Teichmüller space.

4.2.3 Properties of cluster ensembles

Cluster $\mathcal{X}$-varieties

Let $\mathcal{X}_\Lambda$ be the cluster $\mathcal{X}$-torus corresponding to a seed $\mathsf{s} = (\Lambda, (*,*), \{e_i\}, d_i)$ and let us consider the bilinear form corresponding to the exchange matrix ϵ_{ij}:

$$[e_i, e_j] = (e_i, e_j) d_j \ . \tag{4.29}$$

Its left-kernel is the sublattice of Λ defined by

$$\ker_L[*,*] = \left\{ \{\alpha_i\} \in \Lambda \ \middle| \ \sum_{i \in I} \alpha_i \epsilon_{ij} = 0 \ \forall j \in I \right\} \ . \tag{4.30}$$

Let $H_\mathcal{X}$ be the split algebraic with group of characters $\ker_L[*,*]$. The inclusion $\ker_L[*,*] \hookrightarrow \Lambda$ provides a surjection

$$\mathcal{X}_\Lambda \longrightarrow H_\mathcal{X} \ , \tag{4.31}$$

which glues well under mutations, i.e. *the $\mathcal{X}$-space is fibered over the torus $H_\mathcal{X}$* [FG09, Section 2.1].

Proposition 4.8 (Lemma 2.6 in [FG09])**.** *A cluster $\mathcal{X}$-variety admits a Poisson structure preserved under the action of the cluster modular group.*

This follows from the fact that the Poisson structure defined by *Equation* (4.19) at the level of a seed $\mathcal{X}$-torus, is preserved under mutations. Remarkably, this Poisson structure admits a canonical deformation quantization.

Let Λ be a lattice, and let $(*,*)$ be a skew-symmetric bilinear form. The Heisenberg group $\mathcal{H}_\Lambda$ associated to $(\Lambda, (*,*))$ is the central extension $0 \to \mathbb{Z} \to \mathcal{H}_\Lambda \to \Lambda \to 0$ with group law defined for all $v_1, v_2 \in \Lambda$ and $n_1, n_2 \in \mathbb{Z}$ by:

$$\{v_1, n_1\} \circ \{v_2, n_2\} = \{v_1 + v_2, n_1 + n_2 + (v_1, v_2)\} \ . \tag{4.32}$$

Let us denote by q the element $(0,1) \in \mathcal{H}_\Lambda$, and by X_v the element $(v,0) \in \mathcal{H}_\Lambda$, for all $v \in \Lambda$. Equation (4.32) can be rewritten as:

$$q^{-(v_1, v_2)} X_{v_1} X_{v_2} = X_{v_1 + v_2} \tag{4.33}$$

for all $v_1, v_2 \in \Lambda$. There is an involutive antiautomorphism $* : \mathbf{T}_\Lambda \to \mathbf{T}_\Lambda$, defined by:

$$*(X_v) = X_v \ , \qquad *(q) = q^{-1} \ . \tag{4.34}$$

The *quantum torus algebra* $\mathbf{T}_\Lambda$ corresponding to $(\Lambda, (*,*))$ is the group ring of $\mathcal{H}_\Lambda$. Given a basis $\{e_i\}_{i \in I}$ of Λ, and letting $X_i := X_{e_i}$, the algebra $\mathbf{T}_\Lambda$ is the algebra of non-commutative polynomials in the X_i's over $\mathbb{Z}[q, q^{-1}]$ with relations

$$q^{-(e_i, e_j)} X_i X_j = q^{-(e_j, e_i)} X_j X_i \qquad \forall i, j \in I \ . \tag{4.35}$$

The quantum dilogarithm $\Psi_q(x)$ is the q-analogue of the gamma function. It satisfies the difference equation:

$$\Psi_q(q^2 x) = (1 + qx) \Psi_q(x) \ . \tag{4.36}$$

as well as $\Psi_{q^{-1}}(x) = \Psi_q(x)^{-1}$. Let s be a seed and let $\mu_k(\mathsf{s})$ be the result of the mutation in the direction k. Let $\mathbb{T}_\mathsf{s}$ (resp. $\mathbb{T}_{\mu_k(\mathsf{s})}$) be the non-commutative fraction field of the quantum torus algebra $\mathbf{T}_\mathsf{s}$ (resp. $\mathbf{T}_{\mu_k(\mathsf{s})}$). The quantum mutation can be described as the homomorphism $\mu_k^q : \mathbb{T}_{\mu_k(\mathsf{s})} \to \mathbb{T}_\mathsf{s}$ defined by:

$$\mathbb{T}_{\mu_k(\mathsf{s})} \xrightarrow{\ \sim\ } \mathbb{T}_\Lambda \xrightarrow{\mathrm{Ad}_{\Psi_{q_k}(X_k)}} \mathbb{T}_\Lambda \xrightarrow{\ \sim\ } \mathbb{T}_\mathsf{s} \ . \tag{4.37}$$

Specializing q to 1 yields the classical $\mathcal{X}$-mutation formula [FG09, Corollary 3.6].

Cluster $\mathcal{A}$-varieties

Let $\mathcal{A}$ be the cluster $\mathcal{A}$-variety defined by the mutation class of a seed $\mathsf{s} = (\Lambda, (*,*), \{e_i\}, d_i)$. The right-kernel of the bilinear form of Equation (4.29) is:

$$\ker_R[*,*] = \left\{ \{\beta_j\} \in \Lambda \;\middle|\; \sum_{j\in I} \epsilon_{ij}\beta_j = 0 \;\forall i \in I \right\} . \tag{4.38}$$

The map $\delta : \Lambda \to (\Lambda^\circ)^*$ defined on the basis $\{e_i\}$ by $\delta(e_i) = d_i e_i$ is an isomorphism of lattices. Let $K_{\mathcal{A}}$ be the image of $\ker_R[*,*]$ under it; it acts on $(\Lambda^\circ)^*$ in a natural way. Let $\mathcal{H}_{\mathcal{A}}$ be the split algebraic torus with group of cocharacters K_A. The action of $K_{\mathcal{A}}$ on $(\Lambda^\circ)^*$ transposes into a morphism of tori:

$$\mathcal{H}_{\mathcal{A}} \times \mathcal{A}_\Lambda \longrightarrow \mathcal{A}_\Lambda , \tag{4.39}$$

defined for all $\chi_\beta(t) \in \mathcal{H}_{\mathcal{A}} = \mathrm{Hom}(\Gamma_m, K_A)$ and $(a_1, \ldots, a_n)$, by:

$$\chi_\beta(t) \times (a_1, \ldots, a_n) = (t^{\beta_1} a_1, \ldots, t^{\beta_n} a_n) . \tag{4.40}$$

This map glues well under mutation and defines an action of $\mathcal{H}_A$ on the cluster $\mathcal{A}$-variety defined by s [FG09, Lemma 2.10].

Proposition 4.9 (Corollary 6.9 in [FG09])**.** *A cluster $\mathcal{A}$-variety admits a closed 2-form preserved under the action of the cluster modular group.*

This follows from the fact that the closed 2-form defined by *Equation* (4.20) at the level of seed $\mathcal{A}$-tori is preserved under mutations. It is shown in [FG09, Section 6] that this 2-form has a motivic avatar.

4.2.4 Duality conjectures (II)

Let $\mathsf{s} = (\Lambda, (*,*), \{e_i\}, d_i)$ be a seed. The Langlands dual seed of s is the seed $\mathsf{s}^\vee = (\Lambda^\vee, (*,*)_{\Lambda^\vee}, \{e_i^\vee\}, d_i^\vee)$, where $\Lambda^\vee = (\Lambda^\circ)^*$ with basis elements $e_i^\vee = d_i e_i$, multipliers $d_i^\vee = d_i^{-1} D$ with D the lowest common multiple of the d_i's, and where $(e_i^\vee, e_j^\vee)_{\Lambda^\vee} = -(e_j, e_i) d_i d_j D^{-1}$. Langlands duality on seeds amounts to replacing the exchange matrix ϵ_{ij} by $\epsilon_{ij}^\vee = -\epsilon_{ji}$ and the multipliers d_i by $d_i^\vee = d_i^{-1} D$.

Let $(\mathcal{X}, \mathcal{A})$ be the cluster ensemble corresponding to the seed s and let $(\mathcal{X}^\vee, \mathcal{A}^\vee)$ be the one corresponding to the Langlands dual seed $\mathsf{s}^\vee$.

Conjecture 4.10 (4.1 in [FG09])**.** *There exist isomorphisms*

$$\mathcal{A}(\mathbb{Z}^t) = \mathbb{E}(\mathcal{X}^\vee) \qquad \textit{and} \qquad \mathcal{X}(\mathbb{Z}^t) = \mathbb{E}(\mathcal{A}^\vee) \tag{4.41}$$

equivariant with respect to the cluster modular group. They imply in turn the isomorphisms:

$$\mathbb{I}_{\mathcal{A}} : \mathbb{Z}_+\{\mathcal{A}(\mathbb{Z}^t)\} \to \mathbb{L}_+(\mathcal{X}^\vee) , \qquad \mathbb{I}_{\mathcal{X}} : \mathbb{Z}_+\{\mathcal{X}(\mathbb{Z}^t)\} \to \mathbb{L}_+(\mathcal{A}^\vee) , \tag{4.42}$$

which are in particular expected to have the following properties.

1. *If $(a_1, \ldots, a_n)$ are the coordinates of a point $l \in \mathcal{A}(\mathbb{Z}^t)$ in some cluster chart, one has:*

$$\mathbb{I}_{\mathcal{A}}(l) = \prod_i X_i^{a_i} + \ldots , \tag{4.43}$$

   where "$\ldots$" stands for lower order terms, and where the X_i's are the cluster coordinates on $\mathcal{X}^\vee$ in the same chart.

2. *If the coordinates $(a_1, \ldots, a_n)$ of a point $l \in \mathcal{X}(\mathbb{Z}^t)$ in some cluster chart are non-negative, one has:*

$$\mathbb{I}_{\mathcal{X}}(l) = \prod_i A_i^{x_i}. \tag{4.44}$$

   where $\ldots$ stands for lower order terms, and where the A_i's are the cluster coordinates on $\mathcal{A}^\vee$ in the same chart.

3. Let $l \in \mathcal{A}(\mathbb{Z}^t)$ and let $m \in \mathcal{A}^\vee$. Then

$$\mathbb{I}_{\mathcal{A}}(l)(p(m)) = \mathbb{I}_{\mathcal{X}}(p(l))(m) \ . \tag{4.45}$$

It is argued in [FG09] that this conjecture implies the existence of pairings

$$\mathcal{A}(\mathbb{R}^t) \times \mathcal{X}^\vee(\mathbb{R}_{>0}) \to \mathbb{R} \ , \qquad \mathcal{X}(\mathbb{R}^t) \times \mathcal{A}^\vee(\mathbb{R}_{>0}) \to \mathbb{R} \ , \tag{4.46}$$

as well as an intersection pairing

$$\mathcal{A}(\mathbb{R}^t) \times \mathcal{X}^\vee(\mathbb{R}^t) \to \mathbb{R} \ , \tag{4.47}$$

all three equivariant with respect to the cluster modular group. These generalize the pairings discussed at the end of the last chapter, in Section 3.3.

4.3 Hitchin components and cluster higher Teichmüller spaces

4.3.1 Hitchin components

In [Hit92], Hitchin introduced a family of higher Teichmüller spaces for G the adjoint (i.e., with trivial center) group of a split real form of a complex simple Lie group $G^{\mathbb{C}}$. The construction is based on the theory of Higgs bundles on Riemann surfaces developed in [AB83, Hit87, Sim90, Sim92, Cor93]. For other, more detailed and nicer introductions we refer to [BGPG07, GRR15, Wen16, Hua20].

First, let us recall the classical Riemann-Hilbert isomorphism. Let Σ be a Riemann surface, $E \to \Sigma$ a complex rank-n vector bundle over Σ where $n \in \mathbb{Z}_{>0}$, and ∇ a flat connection on E. Since ∇ is flat, the holonomy along a closed path γ on Σ depends only on the homotopy class of γ, and hence it defines a representation

$$\pi_1(\Sigma) \longrightarrow \mathrm{GL}_n(\mathbb{C}) \ . \tag{4.48}$$

This morphism is independent of the gauge action of $\mathrm{GL}_n(\mathbb{C})$ on E. Changing the base point in Σ amounts to conjugating by $\mathrm{GL}_n(\mathbb{C})$.

Conversely, a representation $\rho : \pi_1(\Sigma) \longrightarrow \mathrm{GL}_n(\mathbb{C})$ defines an associated bundle over Σ. The universal cover $\widehat{\Sigma} \to \Sigma$ is a principal $\pi_1(\Sigma)$-bundle over Σ, and the associated bundle of interest is

$$\widehat{\Sigma} \times_\rho \mathbb{C}^n \ . \tag{4.49}$$

It admits locally constant transition functions, and hence a flat connection.

Theorem 4.11 (Riemann–Hilbert correspondence)**.** *Let Σ be a Riemann surface. The space of gauge equivalence classes of flat rank-n vector bundles E over Σ is isomorphic to the space of equivalence classes of representations $\pi_1(\Sigma) \to \mathrm{GL}_n(\mathbb{C})$ modulo conjugation by $\mathrm{GL}_n(\mathbb{C})$.*

The Riemann–Hilbert correspondence can also be stated in algebro-geometric terms [Del06]. A *G-local system* on an algebraic curve S is a sheaf $\mathcal{L}$ on S such that locally it is the constant sheaf with value G, i.e. for all $x \in S$ there exists an open neighborhood U of x such that $\mathcal{L}_{|U}$ is the constant G-sheaf. On nice curves S, G-local systems are equivalent to principal G-bundles on S with a flat connection. The Riemann–Hilbert correspondence is between isomorphism classes of G-local systems on S and algebraic G-connections on S with regular singularities.

There exist other correspondences involving representations of $\pi_1(\Sigma)$, such as the so-called non-abelian Hodge correspondence [Hit87, Sim90], which states the homeomorphisms

$$\mathcal{M}_{\mathrm{Dol}} \simeq \mathcal{M}_{\mathrm{dR}} \simeq \mathcal{M}_{\mathrm{B}} \ , \tag{4.50}$$

where $\mathcal{M}_{\mathrm{Dol}}$ is the moduli space of polystable Higgs bundles on Σ (to be defined in a few lines) modulo complex gauge transformations, $\mathcal{M}_{\mathrm{dR}}$ is the moduli space of flat connections on E modulo gauge, and $\mathcal{M}_{\mathrm{B}}$ is the moduli space of representations $\mathrm{Hom}(\pi_1(\Sigma), \mathrm{GL}_n(\mathbb{C}))$ modulo $\mathrm{GL}_n(\mathbb{C})$-conjugation.

Definition 4.12. *A Higgs bundle over Σ is a pair (E, Φ), where $E \to \Sigma$ is a holomorphic vector bundle and Φ is a holomorphic section of $\mathrm{End}(E) \otimes \Omega^1(\Sigma)$ called the Higgs field.*

A Higgs bundle is said to be *stable* if for each subbundle $F \subset E$ such that $\Phi(F) \subset F \otimes \Omega^1(\Sigma)$ one has

$$\mu(F) := \frac{\deg(F)}{\mathrm{rk}(F)} < \frac{\deg(E)}{\mathrm{rk}(E)} =: \mu(E) \ . \tag{4.51}$$

The rational $\mu(E)$ is called the *slope* of E. A Higgs bundle is *polystable* if it is a direct sum of stable Higgs bundles of the same slope.

Theorem 4.13. *If (E, Φ) is stable and if $\deg E = 0$, then there is a unique unitary connection A on E compatible with the holomorphic structure and satisfying*

$$F_A + [\Phi, \Phi^*] = 0 \ , \tag{4.52}$$

where F_A is the curvature of A. Equation (4.52) is called the Hitchin equation.

Equation (4.52) also makes sense when A is a unitary connection on a principal G-bundle P over Σ, where G is the compact real form of $G^{\mathbb{C}}$, with Cartan involution $A \mapsto -A^*$. Correspondingly, there is also a principle bundle version of Higgs bundles.

Definition 4.14. *Let G be a compact real Lie group. A G-Higgs bundle is a pair (P, Φ) where $P \to \Sigma$ is a principal G-bundle and where Φ is a holomorphic section of $\mathrm{Ad}P \otimes_{\mathbb{C}} \Omega^{1,0}(\Sigma)$, with $\mathrm{Ad}P$ the associated bundle corresponding to the adjoint representation $G \to \mathrm{End}(\mathfrak{g})$.*

Let $\mathcal{M}_G$ be the moduli space of G-Higgs bundles modulo complex gauge transformations, where $g \in G^{\mathbb{C}}$ acts as $(E, \Phi) \to (g \cdot E, g\Phi g^{-1})$. The space $\mathcal{M}_G$ is referred to as the *Hitchin moduli space* of G-Higgs bundles. Let us assume that G is simple, and that $p_1, \ldots, p_l$ are the elements of a basis of the invariant polynomials on $\mathfrak{g} \otimes \mathbb{C}$, of degree $n_1, \ldots, n_l$. The *Hitchin fibration* is the holomorphic map:

$$\begin{array}{rcl} \mathcal{M}_G & \longrightarrow & \displaystyle\bigoplus_{i=1}^{l} H^0(\Sigma, K^{\otimes n_i}) \\ (A, \Phi) & \longmapsto & (p_1(\Phi), \ldots, p_l(\Phi)) \end{array} \ , \tag{4.53}$$

where K is the canonical bundle on Σ. This map is proper. From a unitary connection A on the principal G-bundle P which satisfies Equation (4.52) one can canonically construct a flat $G^{\mathbb{C}}$-connection:

$$\nabla = \nabla_A + \Phi + \Phi^* \ . \tag{4.54}$$

The corresponding representation of $\pi_1(\Sigma)$ provided by the Riemann–Hilbert correspondence is completely reducible. A theorem by Corlette [Cor88] asserts the converse: if E is a vector bundle over Σ with a completely reducible flat connection ∇, there exists a metric on E such that ∇ can be written as in Equation (4.54) with (A, Φ) a solution of Equation (4.52). Therefore $\mathcal{M}_G$ is identified with the moduli space of flat, completely reducible, $G^{\mathbb{C}}$ connections on Σ, and hence with the moduli space of completely reducible morphisms $\pi_1(\Sigma) \to G^{\mathbb{C}}$, modulo $G^{\mathbb{C}}$-conjugation.

The goal now is to identify a connected component consisting of representations $\pi_1(\Sigma) \to G^r$ within this space, where G^r is the real split fort of $G^{\mathbb{C}}$. This is done explicitly in [Hit92] by constructing the so-called *Hitchin section*:

$$\mathcal{M}_G \longleftarrow \bigoplus_{i=1}^{l} H^0(\Sigma, K^{\otimes n_i}) \ , \tag{4.55}$$

which determines a subset of the moduli space $\mathcal{M}_G$ as its image.

Theorem 4.15 (7.5 in [Hit92])**.** *Let $G^{\mathbb{C}}$ be an adjoint complex simple Lie group G, and let G^r be its split real form. The image of the Hitchin section is a connected component of the moduli space of flat completely reducible G^r-connections on Σ, called the Hitchin component.*

The Teichmüller space of Σ embeds into the Hitchin component. Moreover, Labourie proved, using *Anosov representations*, that:

Theorem 4.16 (1.5 in [Lab04])**.** *Every representation in Hitchin component is discrete and faithful.*

In other words, the Hitchin components are G^r-higher Teichmüller spaces. They provide a class of such spaces for all adjoint split real simple Lie groups: $\mathrm{PSL}_n(R)$, $\mathrm{SO}_{n,n+1}(\mathbb{R})$, $\mathrm{PSp}_{2n}(\mathbb{R})$, $\mathrm{PSO}_{n,n}(\mathbb{R})$ together with the five exceptional adjoint split real cases.

4.3.2 Total positivity and positive configurations of flags

Total positivity in $\mathrm{SL}_n(\mathbb{C})$ introduced in Section 1.1.2 was generalized for arbitrary split complex reductive groups by Lusztig in [Lus94]. Let G be a split reductive group over $\mathbb{Q}$, let B^+ be a Borel subgroup of G containing a Cartan subgroup H and let $W = \mathrm{Norm}(H)/H$ be the Weyl group of G. Eventually, let B^- be the Borel subgroup opposite to B^+ and let U^+ (resp. U^-) be the unipotent radical of B^+ (resp B^-). Set $H_{>0} := H(\mathbb{R}_{>0})$.

Let $\{s_i\}_{i\in I}$ be the set of simple reflections in W, and let $x_i(t)$, $y_i(t)$, $h_i(t)$ be the corresponding one-parameter subgroups of G. The *non-negative unipotent subgroup* $U^+_{\geq 0} = U^+(\mathbb{R}_{\geq 0})$ of G is defined as the semigroup generated by the $x_i(t)$ where $t \in \mathbb{R}_{>0}$. Let now w_0 be the longest element in W.

Proposition 4.17 (2.7 in [Lus94], 5.1 in [FG06])**.** *Every reduced expression $s_{i_1} \dots s_{i_k}$ of w_0 gives rise to an open regular embedding $\Gamma^k_m \hookrightarrow U^+$:*

$$(t_1, \dots, t_k) \longrightarrow x_{i_1}(t_1) \dots x_{i_k}(t_k) \ . \tag{4.56}$$

The image of $(\mathbb{R}_{>0})^k$ under this map defines the totally positive unipotent semigroup $U^+_{>0}$ in G. A different choice of reduced expression for w_0 gives a different parametrization of $U^+_{>0}$. The transition function from any such parametrization to another is a positive birational map. In other words, the set of these maps for all possible reduced expression of w_0 forms a positive atlas on U^+.

Replacing the x_i's in Equation (4.56) by y_i's, one obtains a positive atlas on U^- and a set of parametrizations of $U^-_{>0}$. As proved in [FZ99a, FG06], a positive structure can also be defined on double Bruhat cells in G:

$$G^{u,v} = BuB \cap B^- vB^- \ . \tag{4.57}$$

For any set C, group G and set X on which G acts, one denotes

$$\mathrm{Conf}_C(X) = X^C/G \tag{4.58}$$

the *configuration space of points in X parameterized by C*, where the action of G on the set of maps X^C is the one induced by the action of G on X. When $C = [|1, n|]$ we write $\mathrm{Conf}_C(X) = \mathrm{Conf}_n(X)$, for short. Assume now that there is another group π acting on C. In that case one can consider the set of π-equivariant maps:

$$\mathrm{Conf}_{C,\pi}(X) = \mathrm{Conf}_C(X)^\pi \ . \tag{4.59}$$

A set C is said to be a *cyclic set* if the choice of any element in C defines a total order on C, with the orders corresponding to different elements related in the obvious way. As a defining example, a set of points on a circle is cyclically ordered by the choice of an orientation of the circle. An example that will be of great importance is the set $\mathcal{F}_\infty(S) \subset \partial\mathbb{H}$ of the preimages of punctures and cilia of a ciliated surface S endowed with a hyperbolic structure. There is a natural action of $\pi_1(S)$ on $\mathcal{F}_\infty(S)$. As a cyclic set endowed with a $\pi_1(S)$-action, $\mathcal{F}_\infty(S)$ is independent of the hyperbolic structure on S.

Let $\mathcal{B} = G/B^+$ be the flag variety of G. It parametrizes Borel subgroups in G, and the opposite Borel subgroups B^+ and B^- project to two flags in generic position, still denoted B^+ and B^-. The group G acts on $\mathcal{B}$ on the left by translation. For example, the flag variety $\mathcal{B}$ when $G = \mathrm{GL}_m$ parametrizes the complete flags in an m-dimensional vector space V:

$$0 = V_0 \subset V_1 \subset \dots \subset V_m = V \ , \quad \text{where } \dim V_i = i \text{ for all } i = 1, \dots, m \ . \tag{4.60}$$

Definition 4.18. *An element $(B_1, \dots, B_n) \in \mathrm{Conf}_n(\mathcal{B})$ is positive if:*

$$(B_1, \dots, B_n) \sim (B^+, B^-, u_1B^-, u_1u_2B^-, \dots, u_1 \dots u_{n-2}B^-) \ , \tag{4.61}$$

where $u_1, \dots, u_{n-2} \in U^+(\mathbb{R}_{>0})$. Let $\mathrm{Conf}^+_n(\mathcal{B})$ be the space of positive configurations of n flags in $\mathcal{B}$. More intrinsically, $\mathrm{Conf}_n(\mathcal{B})$ is endowed with a positive atlas, under which the image of $(\mathbb{R}_{>0})^N$ is $\mathrm{Conf}^+_n(\mathcal{B})$.

Theorem 4.19 (1.2 in [FG06])**.** *Let $(B_1, \dots, B_n) \in \mathrm{Conf}^+_n(\mathcal{B})$. Then*

$$(B_2, \dots, B_n, B_1) \in \mathrm{Conf}^+_n(\mathcal{B}) \quad \textit{and} \quad (B_n, B_{n-1}, \dots, B_1) \in \mathrm{Conf}^+_n(\mathcal{B}) \ . \tag{4.62}$$

Definition 4.20 (1.5 in [FG06]). *Let C be a cyclic set. A map $\beta : C \to \mathcal{B}(\mathbb{R})$ is said to be positive if for any finite cyclic subset $x_1, \dots, x_n$ in C one has:*

$$(\beta(x_1), \dots, \beta(x_n)) \in \mathrm{Conf}_n^+(\mathcal{B}) \ . \tag{4.63}$$

Let now $\mathcal{A} = G/U^+$ be the principal affine variety of G, whose elements are called *affine flags*, and on which G acts on the left by translation. For example when $G = \mathrm{GL}_m$ the affine flag variety $\mathcal{A}$ parametrizes the complete decorated flags in an m-dimensional vector space V:

$$0 \subset \mathrm{Vect}(v_1) \subset \mathrm{Vect}(v_1, v_2) \subset \cdots \subset \mathrm{Vect}(v_1, \dots, v_m) = V \ , \tag{4.64}$$

where $v_1, \dots, v_m \in V$ and for all $i = 2, \dots, m$, one has $v_i \notin \langle v_1, \dots, v_{i-1} \rangle$.

The identification $W \simeq \mathrm{Norm}(H)/H$ provides a way to lift the elements of W to G (more technically, one needs a *pinning* [Lus94]). Let $s_G \in G$ be the lift of w_0^2 to G. It is a central element of G such that $s_G^2 = e$, where $e \in G$ is the identity [FG06, Corollary 2.1]. For example, when $G = \mathrm{SL}_m$ one has $s_G = (-1)^{m-1} e$.

Definition 4.21 (2.5 in [FG06]). *A twisted cyclic configuration of affine flags in $\mathcal{A}$ is an orbit in $\mathrm{Conf}_n(\mathcal{A})$ for the twisted cyclic shift map:*

$$(A_1, \dots, A_n) \longrightarrow (A_2, \dots, A_n, s_G A_1) \ . \tag{4.65}$$

Let $\widetilde{Conf}_n(\mathcal{A})$ be the moduli space of twisted cyclic configurations of affine flags in $\mathcal{A}$.

4.3.3 Framed local systems and Teichmüller $\mathcal{X}$-spaces

Let S be a ciliated surface and let $\mathcal{L}_{G,S}$ be the moduli space of G-local systems on S with holes shrunk to punctures. Let $\mathcal{L}$ be a G-local system on S, and let $\mathcal{L}_{\mathcal{B}} = \mathcal{L} \times_G \mathcal{B}$ be the *associated flag bundle*.

Definition 4.22 (2.1 in [FG06]). *Let G be a split reductive group defined over $\mathbb{Q}$. A framed G-local system on S is a pair $(\mathcal{L}, \beta)$ where $\mathcal{L}$ is a G-local system on S and β is a flat section of the restriction of $\mathcal{L}_{\mathcal{B}}$ to the set $\widetilde{\partial S}$ containing the circles bounding a hole in S, and the segments obtained as follows. For each boundary component of S carrying cilia, one removes a point in the interior of each segment between adjacent cilia. This decomposes the boundary component into a disjoint union of segments, each of which containing exactly one cilium. Let $\mathcal{X}_{G,S}$ be the moduli space of framed local systems on S.*

This definition is a slight modification of the one in [FG06]. The restriction of $\mathcal{L}_{\mathcal{B}}$ to a segment in $\widetilde{\partial S}$ is equivalent to the choice of a flag in $\mathcal{B}$ at each of these segments, i.e. at each cilium of S. The restriction of $\mathcal{L}_{\mathcal{B}}$ to the circle components of the boundary bounding holes is equivalent to the data of a flag invariant under the monodromy around the hole at each of these circles. For example when $S = S_{0,0,\{n\}} = D_n$ is the disk with n cilia, one has $\mathcal{X}_{G,D_n} = G\backslash\mathcal{B}^n = \mathrm{Conf}_n(\mathcal{B})$. When S is an annulus (which is *not* a hyperbolic ciliated surface), the moduli space $\mathcal{X}_{G,S}$ is the space of triples (g, B_1, B_2) modulo G, where g is the monodromy around the topologically non-trivial class of $\pi_1(S)$, and B_1, B_2 are two Borel subgroups of G such that $g \in B_1 \cap B_2$. The group G acts on such triples by conjugation on g and left-multiplication on B_1, B_2. In that case, $\mathcal{X}_{G,S}$ coincides with the so-called Steinberg variety [CG97].

Proposition 4.23 (Lemma 1.1 in [FG06]). *There exists a natural bijection*

$$\mathcal{X}_{G,S}(\mathbb{C}) \simeq \mathrm{Conf}_{\mathcal{F}_\infty(S),\pi_1(S)}(\mathcal{B}(\mathbb{C})) \ . \tag{4.66}$$

Theorem 4.24 (Decomposition theorem). *Let us assume that G is a split semi-simple adjoint algebraic group over $\mathbb{Q}$. Then, the moduli space $\mathcal{X}_{G,S}$ is birationnally equivalent to*

$$\prod_{f \in F(\Gamma)} \mathcal{X}_{G,f} \times \prod_{e \in E(\Gamma)} H \ , \tag{4.67}$$

where Γ is a triangulation of S.

A consequence is that the set of these birational isomorphisms, when Γ runs over the set of all triangulations of S, provide a positive atlas on $\mathcal{X}_{G,S}$ equivariant with respect to the cluster modular groupoid [FG06, Theorem 6.1]. It is obtained from the positive atlas on each $\mathcal{X}_{G,f}$ and the trivial one on $H \simeq \Gamma_m^{\mathrm{rk}(G)}$. In fact, $\mathcal{X}_{G,S}$ is a cluster $\mathcal{X}$-variety as defined in Section 4.2. We will see this explicitly below for $G = \mathrm{PGL}_m$.

Definition 4.25. *Let S be a ciliated surface and G a split semi-simple adjoint algebraic group over $\mathbb{Q}$. The G-higher framed Teichmüller space of S, or Teichmüller $\mathcal{X}$-space of S, is defined by:*

$$\mathcal{X}^+_{G,S} = \mathrm{Conf}^+_{\mathcal{F}_\infty(S),\pi_1(S)}(\mathcal{B}) \ . \tag{4.68}$$

This definition coincides with the one coming from the positive structure on $\mathcal{X}_{G,S}$.

$$\mathcal{X}^+_{G,S} = \mathcal{X}_{G,S}(\mathbb{R}_{>0}) \ . \tag{4.69}$$

The elements of $\mathcal{X}^+_{G,S}$ are called positive framed G-local systems on S. There is a natural map $\mathcal{X}_{G,S} \to \mathcal{L}_{G,S}$ obtained by forgetting the framing, and the image of $\mathcal{X}^+_{G,S}$ under it is called the space of positive G-local systems on S – or, Riemann–Hilbert-equivalently, positive representations $\pi_1(S) \to G$. The moduli space of such positive representations is denoted $\mathcal{L}^+_{G,S}$. It consists of classes of representations $\pi_1(S) \to G(\mathbb{R})$ modulo $G(\mathbb{R})$.

Subsequently, after taking logarithms of the coordinates one sees that the G-higher Teichmüller spaces $\mathcal{X}^+_{G,S}$ are homeomorphic to open balls $\mathbb{R}^N$ for some $N \in \mathbb{R}$. Moreover there is a natural embedding $\mathcal{X}^+_{\mathrm{PGL}_2,S} \hookrightarrow \mathcal{X}^+_{G,S}$. The name higher Teichmüller space is justified by the following proposition.

Proposition 4.26 (Theorems 1.9 and 1.10 in [FG06])**.** *Any positive representation $\rho \in \mathcal{X}^+_{G,S}$ is discrete and faithful.*

Two generalizations of the above are proposed in [FG06], namely the universal G-higher Teichmüller space $\mathcal{X}^+_G$ as well as G-higher Teichmüller spaces $\mathcal{X}^+_{G,S}$ for closed surfaces. The latter coincides with Hitchin's construction, which is proved by showing that when S has no boundary, the moduli space of positive representations $\mathcal{L}^+_{G,S}$ coincides with Labourie's moduli space of Anosov representations [Lab04], which was itself proved to correspond to Hitchin's components when $G = \mathrm{PGL}_m(\mathbb{R})$ by Labourie, using a result from [Gui08]. Interestingly, despite the correspondence between the $\mathcal{L}^+_{G,S}$ and Hitchin's components, the way they are constructed is very different: in the first case the construction is combinatorial and algebraic, and does not require the choice of an analytic structure on S, whereas in the second case the construction is mostly analytic and requires S to be a Riemann surface.

As emphasized in the introduction of [FG06], the discreteness condition that appears in the definition of (higher) Teichmüller spaces is of a non-algebraic nature. However, positivity provides a good algebraic framework in which one can possibly accounts for such a constraint.

When $G = \mathrm{PGL}_2$, the space $\mathcal{X}^+_{\mathrm{PGL}_2,S}$ is shown to coincide with the Teichmüller space with holes of S [FG06, Section 11]:

$$\mathcal{X}^+_{\mathrm{PGL}_2,S} = \mathcal{T}^x(S) \ . \tag{4.70}$$

Moreover, $\mathcal{L}^+_{\mathrm{PGL}_2,S}$ is the classical Teichmüller space $\mathcal{T}(S)$ of S.

Special coordinates when $G = \mathrm{PGL}_m$. One can construct coordinates on $\mathcal{X}^+_{\mathrm{PGL}_m,S}$ as follows. This generalizes the construction of coordinates on $\mathcal{T}^x(S)$ presented in Chapter 2. Let Γ be a triangulation of a ciliated surface S as defined in Section 2.2, and consider the m-subtriangulation obtained by subdividing each face of Γ into m^2 triangles, $m(m+1)/2$ pointing up and $m(m-1)/2$ pointing down, as shown on the left of Figure 4.1 for $m = 4$. Edges of Γ are the plain lines, while the edges of the subtriangulation are dashed.

The vertices of the m-triangulation of a triangle are in one-to-one correspondence with triples of non-negative integers (i, j, k) such that $i + j + k = m$, as on the left of Figure 4.1. The moduli space $\mathcal{X}_{\mathrm{PGL}_m,S}$ is parameterized by complex numbers at each vertex of the subtriangulation, but the ones of Γ. At each point of $\mathcal{X}_{\mathrm{PGL}_m,S}$ there is a flag at each vertex of Γ, from which these coordinates are constructed. Here we follow the presentation of [Pal13].

Let us first consider a face of Γ, at the vertices of which are the flags $A = (0 = A_0 \subset A_1 \subset \cdots \subset A_m = \mathbb{C}^m)$, $B = (0 = B_0 \subset B_1 \subset \cdots \subset B_m = \mathbb{C}^m)$ and $C = (0 = C_0 \subset C_1 \subset \cdots \subset C_m = \mathbb{C}^m)$ of $\mathcal{B}(\mathbb{C})$ as on the right of Figure 4.1. To the vertices of the subtriangulation inside this face are associated triple-ratios $X_{i,j,k}(A, B, C)$ as follows, where $i, j, k \in [|1, m-1|]$ and $i + j + k = m$. Generically A, B and C are transverse, and hence $A_{i-1} \oplus B_{i-1} \oplus C_{i-1}$ is a subspace of $\mathbb{C}^m$ of codimension 3. Let

$$V_{i,j,k} = \mathbb{C}^m / A_{i-1} \oplus B_{i-1} \oplus C_{i-1} \ . \tag{4.71}$$

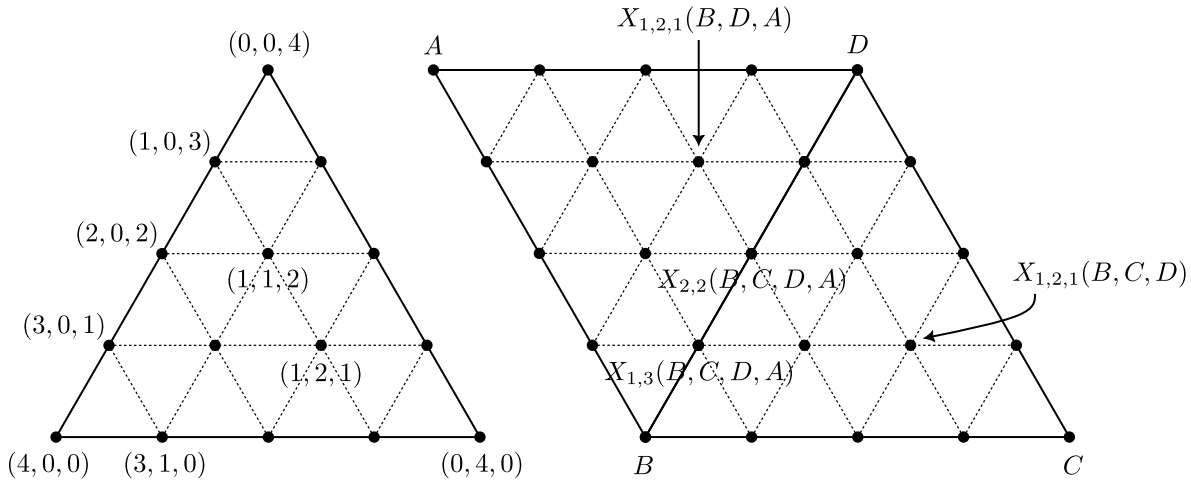


Figure 4.1: The construction of coordinates on $\mathcal{X}^+_{\mathrm{PGL}_4,S}$.

Let $(\overline{A}_i, \overline{A_{i+1}})$, $(\overline{B}_i, \overline{B_{i+1}})$ and $(\overline{C}_i, \overline{C_{i+1}})$ be the projections of (A_i, A_{i+1}), (B_i, B_{i+1}) and (C_i, C_{i+1}) to $V_{i,j,k}$. They form a triple of (generically transverse) flags in $V_{i,j,k}$. Let v_A (respectively, v_B, v_C) be a non-zero vector in $\overline{A}_i$ (respectively $\overline{B}_i$, $\overline{C}_i$), and let f_A (respectively, f_B, f_C) be a linear form whose kernel is $\overline{A}_{i+1}$ (respectively $\overline{B}_{i+1}$, $\overline{C}_{i+1}$). Then one sets:

$$X_{i,j,k}(A, B, C) = \frac{f_A(v_B) f_B(v_C) f_C(v_A)}{f_A(v_C) f_B(v_A) f_C(v_B)} \ . \tag{4.72}$$

Note that $X_{i,j,k}(A, B, C) = X_{j,k,i}(B, C, A)$.

Consider now an internal edge of Γ, with flags B and D assigned to its ends, and C, D at the other vertices, as in Figure 4.1. To the vertices of the subtriangulation in the interior of this edge are assigned the coordinates $X_{i,j}(B, C, D, A)$ where $i, j \in [|1, m-1|]$ and $i + j = k$, constructed as follows. The direct sum $B_{i-1} \oplus D_{i-1}$ is generically a subspace of $\mathbb{C}^m$ of codimension 2. Let

$$V_{i,j} = \mathbb{C}^m / B_{i-1} \oplus C_{i-1} \ . \tag{4.73}$$

The projection of B_i (resp. C_1, D_j, A_1) in $V_{i,j}$ is a line $\overline{B}_i$ (resp. $\overline{C}_1$, $\overline{D}_i$ and $\overline{A}_1$). Then one sets:

$$X_{i,j}(B, C, D, A) = \chi(\overline{B}_i, \overline{C}_1, \overline{D}_j, \overline{A}_1) \ , \tag{4.74}$$

where χ is (minus) the cross-ratio. Note that $X_{i,j}(B, C, D, A) = X_{j,i}(D, A, B, D)$.

The set of coordinates $X_{i,j,k}$ for each face of Γ and $X_{i,j}$ for each internal edge of Γ parametrizes $\mathcal{X}_{\mathrm{PGL}_m,S}$. More precisely:

Proposition 4.27. *The map*

$$\Gamma^N_m \to \mathcal{X}_{\mathrm{PGL}_m,S} \ , \tag{4.75}$$

where $N = |F(\Gamma)| \frac{(m-1)(m-2)}{2} + |E_i(\Gamma)|(m-1)$ *and where* Γ^N_m *accounts for the aforementioned coordinates, is a birational isomorphism. It is a positive chart on* $\mathcal{X}_{\mathrm{PGL}_m,S}$. *The image of* $\mathbb{R}_{>0}$ *under it parametrizes the* PGL_m*-higher Teichmüller space* $\mathcal{X}^+_{\mathrm{PGL}_m,S}$.

The subtriangulation can be flipped at any internal edge, yielding another positive chart on $\mathcal{X}_{\mathrm{PGL}_m,S}$ related to the original one by a cluster $\mathcal{X}$-mutation. The quiver underlying the cluster variety $\mathcal{X}_{\mathrm{PGL}_m,S}$ is the one obtained from the m-subtriangulation of Γ, as in Section 1.2. Note however that in most cases, after a sequence of flips one does not obtain a subtriangulation of a triangulation of S, and the coordinates cannot be expressed as triple ratios and cross-ratios of the flags at the punctures and cilia of S. However these positive charts do exist: they are defined by sequences of mutations starting at a subtriangulation of a triangulation of S. For more details on this construction of the coordinates we refer to [FG06, Pal13]. Since $\mathcal{X}_{\mathrm{PGL}_m,S}$ is a cluster $\mathcal{X}$-variety, it is endowed with a canonical Poisson structure.

As in Section 2.3.1 one can construct the coordinates corresponding to an m-subtriangulation of a triangulation Γ of S as the monodromy of an auxiliary bipartite graph Λ_B. Consider a triangle of Γ together with its m-subtriangulation, oriented as on the left of Figure 4.1. Define a white vertex of Λ_B in each of the $m(m+1)/2$ triangles pointing up, and a black vertex in each of the $m(m-1)/2$ triangles pointing down. The edges of Γ are determined by the adjacency of these small triangles. Note that the small triangles pointing up are in bijection with the triples of positive integers (ι, κ, λ) such that $\iota + \kappa + \lambda = 2m + 1$, whereas the ones pointing down are in bijection with the triples of positive integers (ι, κ, λ) such that $\iota + \kappa + \lambda = 2m + 2$. For example, to the bottom-left (resp. bottom-right) small triangle pointing up is assigned the triple $(1, m, m)$ (resp. $(m, 1, m)$), and to the pointing-down-one next to it is assigned $(2, m, m)$ (resp. $(m, 2, m)$).

Let A, B, C be the flags in $\mathcal{B}(\mathbb{C})$ assigned to the vertices of this triangle of Γ, enumerated counterclockwise and starting at the bottom left vertex. Let us assume that they are in general position. For every triple of positive integers (ι, κ, λ) such that $\iota + \kappa + \lambda = 2m + 1$, one assigns the line $A_\iota \cap B_\kappa \cap C_\lambda \subset \mathbb{C}^m$ to the white vertex of Λ_B corresponding to (ι, κ, λ). The dimension of $A_\iota \cap B_\kappa \cap C_\lambda$ is $\iota - (\kappa - 1) - (\lambda - 1) = 1$ under the genericity assumption, i.e. it is indeed a line in $\mathbb{C}^m$.

Let us now consider a triple of positive integers (ι, κ, λ) such that $\iota + \kappa + \lambda = 2m + 2$, i.e. a black vertex of Λ_B. The white vertices of Λ_B adjacent to it correspond to the triples $(\iota - 1, \kappa, \lambda)$, $(\iota, \kappa - 1, \lambda)$ and $(\iota, \kappa, \lambda - 1)$, and the lines $A_{\iota-1} \cap B_\kappa \cap C_\lambda$, $A_\iota \cap B_{\kappa-1} \cap C_\lambda$ and $A_\iota \cap B_\kappa \cap C_{\lambda-1}$ are all contained in the plane $A_\iota \cap B_\kappa \cap C_\lambda$. We assign the kernel of the map

$$(A_{\iota-1} \cap B_\kappa \cap C_\lambda) \oplus (A_\iota \cap B_{\kappa-1} \cap C_\lambda) \oplus (A_\iota \cap B_\kappa \cap C_{\lambda-1}) \to A_\iota \cap B_\kappa \cap C_\lambda \tag{4.76}$$

to this black vertex. We are thus in the same situation as in Section 2.3.1: choosing a trivialization of the lines over every vertex of Λ_B yields an assignment of numbers in $(\mathbb{C})^\times$ to the edges of Λ_B, and changing the trivialization amounts to a natural $(\mathbb{C}^\times)^{V(\Lambda_B)}$ action on $(\mathbb{C}^\times)^{E(\Lambda_B)}$. The example of this construction for $G = \mathrm{PGL}_4$ is depicted in Figure 4.2.

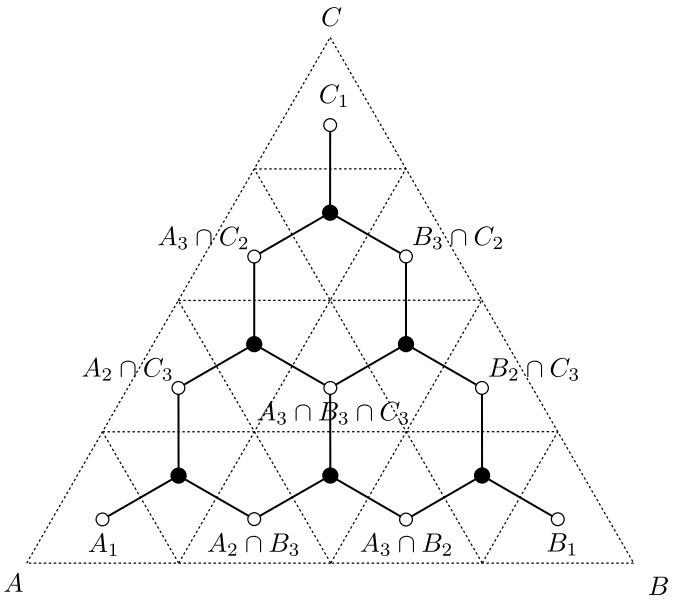


Figure 4.2: Coordinates as the monodromy of an abelian connection on a graph.

This generalizes straightforwardly to a construction of a bipartite graph Λ_B on a general ciliated surface S from a triangulation Γ of S: one constructs Λ_B in the faces of Γ first, and then one glues the latter together. After this, one merges the white vertices in adjacent small triangles; this is compatible with the line assignment on them. The invariants $X_{i,j}$ and $X_{i,j,k}$ of above are then obtained as the monodromy of the abelian $(\mathbb{C}^\times)$-connection on Λ_B.

G-higher $\mathcal{X}$-laminations. The space $\mathcal{X}_{G,S}$ being a positive space, one can consider its points valued in a tropical semi-field $\mathbb{Z}^t$, $\mathbb{Q}^t$ or $\mathbb{R}^t$. When $G = \mathrm{PGL}_m$ we have the positive charts defined in the last paragraph (one for each triangulation Γ of S), and hence the image of $\mathbb{Z}^N$, $\mathbb{Q}^N$ or $\mathbb{R}^N$ under the map

of Equation (4.75) parametrizes the corresponding space in each case. As before the internal edges of the subtriangulation can be flipped, which yields new coordinate systems that in general do not have a simple geometric interpretation. Under a flip, the coordinates change accordingly to the tropical cluster $\mathcal{X}$-mutations.

The space $\mathcal{X}_{\mathrm{PGL}_2,S}(\mathbb{Z}^t)$ (resp. $\mathcal{X}_{\mathrm{PGL}_2,S}(\mathbb{Q}^t)$, $\mathcal{X}_{\mathrm{PGL}_2,S}(\mathbb{R}^t)$) corresponds to the space of integral (resp. rational, real) $\mathcal{X}$-laminations on S. That follows from the analysis of Chapter 3. By analogy, in the general G-case one speaks of G-higher $\mathcal{X}$-lamination spaces. The cluster duality conjectures reviewed in Section 4.2.4 interpret the points of $\mathcal{X}_{G,S}(\mathbb{Z}^t)$ as parametrizing the elements of $\mathbb{E}((\mathcal{A}_{G^L,S})_0)$, where $(\mathcal{A}_{G,S})_0$ is the subset with trivial frozen coordinates of the cluster variety $\mathcal{A}_{G^L,S}$ defined in the next section, and where G^L is the Langlands dual of G. The reasoning of Section 4.1.2 implies that $\mathbf{Sp}\mathcal{X}_{G,S}(\mathbb{R}^t)$ is a spherical compactification of $\mathcal{X}_{G,S}(\mathbb{R}_{>0})$.

4.3.4 Decorated local systems and Teichmüller $\mathcal{A}$-spaces

The Cartan subgroup H of G is canonically isomorphic to the quotient B^+/U^+. Let

$$i_{B^+} : H \to B^+/U^+ \tag{4.77}$$

be the canonical isomorphism. The Cartan subgroup H acts on the principle affine variety $\mathcal{A} = G/U^+$ on the right as follows. Let $g \in G$ and $h \in H$. Then

$$gU^+ \cdot h = hi_B(h)U \ . \tag{4.78}$$

The quotient $\mathcal{A}/H$ is identified with $\mathcal{B} = G/B$. Let $\mathcal{L}$ be a G-local system on S considered with the natural action of G on the right. The *principal affine bundle* corresponding to $\mathcal{L}$ is $\mathcal{L}_{\mathcal{A}} = \mathcal{L}/U$.

Let $T'_S = T_S - S$ be the tangent space to S with the zero section removed. The fundamental group $\pi_1(T'S, x)$ for $x \in T'S_y$ with $y \in S$ is a central extension:

$$1 \to \mathbb{Z} \to \pi_1(T'S, x) \to \pi_1(S, y) \to 1 \ ; \tag{4.79}$$

the inclusion $T'S_y \subset T'S$ induces an isomorphism between the $\mathbb{Z}$ appearing in the previous equation and $\pi_1(T'S_y, x)$. Let σ_S be a generator of this central subgroup of $\pi_1(T'S, x)$, defined up to a sign.

Definition 4.28. *A twisted G-local system on S is a G-local system on $T'S$ with monodromy s_G along σ_S, modulo gauge. Since $s_G^2 = e$, this definition is well-posed.*

Consider the quotient $\overline{\pi}_1(T'S, x)$ of $\pi_1(T'S, x)$ by the central subgroup $2\mathbb{Z} \subset \mathbb{Z}$, so that $\overline{\pi}_1(T'S, x)$ is a central extension of $\pi_1(S, y)$ by $\mathbb{Z}/2\mathbb{Z}$, and let $\overline{\sigma}_S$ be the image of σ_S in this quotient. Then the twisted local systems on S correspond to the representations $\rho : \overline{\pi}_1(T'S, x) \to G$ such that $\rho(\overline{\sigma}_S) = s_G$, modulo G-conjugation.

Let $\mathbf{C}$ be a small neighborhood of a hole in S with boundary curve C. One has a canonical isomorphism $\pi_1(T'\mathbf{C}) \simeq \mathbb{Z} \times \mathbb{Z}$ where one factor is generated by the lift of the homotopy class of C and the other is $\pi_1(T'\mathbf{C}_x)$, so that $\overline{\pi}_1(T'\mathbf{C}) \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}$. Let now D be a boundary component in S with p, and let $x_1, \ldots, x_p \in S$, one in each segment between two adjacent cilia. Let $\mathbf{D}$ be a small neighborhood of D in S homeomorphic to $D\times]0, 1]$, and let $\mathbf{D}' = (D - \{x_1, \ldots, x_p\})\times]0, 1] \subset \mathbf{D}$.

Definition 4.29. *Let us from now on assume that G is simply connected (and still split reductive over $\mathbb{Q}$). Let $\mathcal{L}$ be the representative of a twisted local system on S. A decoration on $\mathcal{L}$ is a locally constant section α of the restriction of $\mathcal{L}_{\mathcal{A}}$ to $\cup_i \mathbf{C}_i \cup_j \mathbf{D}'_j$. A decorated twisted G-local system is a pair $(\mathcal{L}, \alpha)$. Let $\mathcal{A}_{G,S}$ be the moduli space of decorated twisted G-local systems on S.*

Note that when $s_G = e$, a decorated twisted G-local system is simply a pair $(\mathcal{L}, \alpha)$ where $\mathcal{L}$ is a G-local system on S and where α is a locally constant section of the restriction of $\mathcal{L}_{\mathcal{A}}$ to $\cup_i \mathbf{C}_i \cup_j \mathbf{D}'_j$.

Lemma 4.30 (2.3 in [FG09])**.** *Let $S = S_{0,0,\{k\}}$ be the disk with k cilia. Then $\mathcal{A}_{G,S} \simeq \widetilde{\mathrm{Conf}}_k(\mathcal{A})$.*

Proof. The tangent space T'_S in this case can be retracted to a circle, hence there is a unique isomorphism class of twisted G-local system on S. Its monodromy along a curve in $T'S$ obtained by considering a non-zero tangent vector field near the boundary of S is s_G. Let $c_1, \ldots, c_k$ be the cilia on the boundary of S, enumerated clockwise. For all $i = 1, \ldots, k$ let v_i be a non-zero tangent vector to S at x_i pointing

inside D, and consider the restriction of the flat section α of $\mathcal{L}_{\mathcal{A}}$ to $v_1, \dots, v_k \in T'_S$. Consider now for each $i = 2, \dots, k$ the counterclockwise flat transport of each v_i to x_1 along the boundary of S. This yields a configuration $(A_1, \dots, A_k) \in \mathrm{Conf}_k(\mathcal{A})$ at x_1. Basing the same reasoning at x_2 instead of x_1 yields $(A_2, \dots, A_k, s_G A_1) \in \mathrm{Conf}_k(\mathcal{A})$, etc... □

Let $G' = G/Z(G)$ be the adjoint group corresponding to G. Since s_G is in the center, under the canonical projection $G' \to G$ a twisted G-local system on S defines a G'-local system on S. Moreover, given a twisted G-local system on S, a decoration α naturally defines a framing β since $\mathcal{A} \to \mathcal{B}$ is a principal H-bundle. Hence there is a natural map

$$p : \mathcal{A}_{G,S} \to \mathcal{X}_{G',S} \tag{4.80}$$

which generalizes Equation (2.51). The image of $\mathcal{A}_{G,S}$ under it is the set of framed unipotent G-local systems on S, i.e. a framed G-local system on S such that its monodromy along any curve bounding a hole is unipotent.

There is an analogue of Proposition 4.23 for the moduli space $\mathcal{A}_{G,S}$ [FG09, Section 8.6]. Let $\widetilde{\mathcal{F}}_\infty(S)$ be the non-trivial twofold cover of $\mathcal{F}_\infty(S)$ such that it is again a cyclic set with a cyclic structure compatible with the one on $\mathcal{F}_\infty(S)$, and let σ be its non-trivial automorphism. The $\mathbb{Z}/2\mathbb{Z}$-extension $\overline{\pi}_1(S)$ acts by automorphisms on $\widetilde{\mathcal{F}}_\infty(S)$, with $\overline{\sigma}_S$ acting as σ.

Proposition 4.31. *There exists a natural bijection*

$$\mathcal{A}_{G,S}(\mathbb{C}) \simeq \mathrm{Conf}_{\widetilde{\mathcal{F}}_\infty(S), \overline{\pi}_1(S)}(\mathcal{A}(\mathbb{C})) \ . \tag{4.81}$$

Theorem 4.32 (Decomposition theorem)**.** *Let Γ be a triangulation of S. There exists a natural map*

$$\mathcal{A}_{G,S} \to \prod_{f \in F(\Gamma)} \mathcal{A}_{G,f} \times \prod_{e \in E(\Gamma)} \mathcal{A}_{G,e} \tag{4.82}$$

and the image of $\mathcal{A}_{G,S}$ under it is birationnally identified with the subvariety $\mathcal{A}_{G,\Gamma}$ of the right-hand side defined by the equations $q_{f,e} \circ q_f = q_e$ for e and edge of f, and where for any $e \in E(\Gamma)$ and $f \in F(\Gamma)$ the maps $q_e : \mathcal{A}_{G,S} \to \mathcal{A}_{G,e}$, $q_f : \mathcal{A}_{G,S} \to \mathcal{A}_{G,f}$ and $q_{f,e} : \mathcal{A}_{G,f} \to \mathcal{A}_{G,e}$ when e is an edge of f, are the natural projections.

The set of birational isomorphisms

$$\mathcal{A}_{G,\Gamma} \to \mathcal{A}_{G,S} \ , \tag{4.83}$$

where Γ runs over the set of all triangulations of S, provide a positive atlas on $\mathcal{A}_{G,S}$ equivariant with respect to the cluster modular groupoid [FG06, Theorem 8.2]. In fact, $\mathcal{A}_{G,S}$ is a cluster $\mathcal{A}$-variety as defined in Section 4.2. We will see this explicitly below for $G = \mathrm{SL}_m$.

Definition 4.33. *Let S be a ciliated surface and G a split semi-simple simply-connected algebraic group over $\mathbb{Q}$. The G-higher decorated Teichmüller space of S, or Teichmüller $\mathcal{A}$-space of S, is defined by:*

$$\mathcal{A}^+_{G,S} = \mathcal{A}_{G,S}(\mathbb{R}_{>0}) \ . \tag{4.84}$$

There is a notion of positive twisted configuration of affine flags, and hence a definition of the Teichmüller $\mathcal{A}$-space in terms of the boundary map of Equation (4.81), and both definitions coincide. The space $\mathcal{A}^+_{G,S}$ consists of classes of representations $\pi_1(S) \to G(\mathbb{R})$ modulo $G(\mathbb{R})$.

Subsequently, after taking logarithms of the coordinates one sees that the G-higher Teichmüller spaces $\mathcal{A}^+_{G,S}$ are homeomorphic to open balls $\mathbb{R}^N$ for some $N \in \mathbb{R}$. Moreover there is a natural embedding $\mathcal{A}^+_{\mathrm{SL}_2,S} \hookrightarrow \mathcal{A}^+_{G,S}$. That these spaces are higher Teichmüller spaces is justified again by the Proposition 4.26 and the natural map $p : \mathcal{A}_{G,S} \to \mathcal{X}_{G',S}$, where G' is as before the adjoint group $G/Z(G)$.

When $G = \mathrm{SL}_2$, the space $\mathcal{A}^+_{\mathrm{SL}_2,S}$ is shown to coincide with the decorated Teichmüller space of S [FG06, Section 11]:

$$\mathcal{A}^+_{\mathrm{SL}_2,S} = \mathcal{T}^a(S) \ . \tag{4.85}$$

Special coordinates when $G = \mathrm{SL}_m$. Starting from a triangulation Γ of S, we consider as in the previous subsection the m-subtriangulation of Γ (see Figure 4.1). Let us first consider a triangle of Γ. The twisted SL_m-local system on Γ determines a configuration of three affine flags $A = (a_1, \dots, a_m)$, $B = (b_1, \dots, b_n)$ and $C = (c_1, \dots, c_n)$ in $\mathbb{C}^m$ at a fixed vertex of the triangle (going from the latter to the one directly after it with respect to the clockwise orientation yields the triple $(B, C, s_G A)$, and so forth, and so on). To each internal point of the triangle corresponding to the triple of positive integers (i, j, k) such that $i + j + k = m$ one associates the coordinate:

$$\Delta_{i,j,k}(A, B, C) = \int_{\mathbb{C}^m} \mathrm{d}\omega \; a_1 \wedge \cdots a_i \wedge b_1 \wedge \cdots \wedge b_j \wedge c_1 \wedge \cdots \wedge c_k \; , \tag{4.86}$$

where ω is the volume form on $\mathbb{C}^m$. Note that this definition implies that

$$\Delta_{j,k,i}(B, C, A) = (-1)^{i(j+k)} X_{i,j,k}(A, B, C) \; . \tag{4.87}$$

If $m = i+j+k$ is odd then $i(j+k)$ is even. Conversely, if $m = i+j+k$ is odd then $(-1)^{i(j+k)} = (-1)^i$, and thus $\Delta_{j,k,i}(B, C, -A) = (-1)^i \Delta_{j,k,i}(B, C, A) = \Delta_{i,j,k}(A, B, C)$. Hence in both cases $\Delta_{j,k,i}(B, C, (-1)^{m-1}A) = \Delta_{i,j,k}(A, B, C)$. Since $s_G = (-1)^{m-1}$ for SL_m, these invariants are well-defined on twisted cyclic configurations of affine flags.

Let us now consider an edge of Γ with the twisted local system defining a configuration of two affine flags $A = (a_1, \dots, a_m)$ and $B = (b_1, \dots, b_n)$ at one of its ends. For any pair of positive integers i, j such that $i + j = m$ (hence defining a vertex of the subtriangulation in the interior of the edge) one sets

$$\Delta_{i,j}(A, B) = \int_{\mathbb{C}^m} \mathrm{d}\omega \; a_1 \wedge \cdots a_i \wedge b_1 \wedge \cdots \wedge b_j \; . \tag{4.88}$$

As before these invariants are well-defined on twisted configurations:

$$\Delta_{i,j}(A, B) = \Delta_{j,i}(B, (-1)^{m-1}A) \; . \tag{4.89}$$

Proposition 4.34. *The map*

$$\Gamma_m^N \to \mathcal{X}_{\mathrm{SL}_m,S} \; , \tag{4.90}$$

where $N = |F(\Gamma)|\frac{(m-1)(m-2)}{2} + |E(\Gamma)|(m-1)$ and where Γ_m^N accounts for the above-mentioned coordinates, is a birational isomorphism. It is a positive chart on $\mathcal{A}_{\mathrm{SL}_m,S}$. The image of $\mathbb{R}_{>0}$ under it parametrizes the SL_m-higher Teichmüller space $\mathcal{A}^+_{\mathrm{SL}_m,S}$.

The subtriangulation can be flipped at any internal edge, yielding another positive chart on $\mathcal{A}_{\mathrm{PGL}_m,S}$ related to the original one by a cluster $\mathcal{A}$-mutation. The quiver underlying the cluster variety $\mathcal{A}_{G,S}$ is the one obtained from the m-subtriangulation of Γ, as in Section 1.2. As before, after a sequence of flips one does not obtain a subtriangulation of a triangulation of S in general, and therefore the coordinates cannot be expressed easily as Δ-invariants: they are defined by sequences of mutations starting at a subtriangulation of a triangulation of S. The cluster variety $\mathcal{A}_{G,S}$ admits a natural subvariety $(\mathcal{A}_{G,S})_0$ defined by the fact that external edges are assigned trivial coordinates. Again, we refer to [FG06] for a more detailed discussion of these coordinates.

When $G = \mathrm{SL}_2$, these coordinates coincide with those on $\mathcal{T}^s(S)$ introduced in Chapter 2, as sketched in Section 2.4.1 (for a full proof of this point we refer to [FG06, Section 11]).

***G*-higher $\mathcal{A}$-laminations.** The space $\mathcal{A}_{G,S}$ being a positive space, one can consider its points valued in a tropical semi-field $\mathbb{Z}^t$, $\mathbb{Q}^t$ or $\mathbb{R}^t$. When $G = \mathrm{SL}_m$ we have the positive charts defined in the last paragraph (one for each triangulation Γ of S), and hence the image of $\mathbb{Z}^N$, $\mathbb{Q}^N$ or $\mathbb{R}^N$ under the map of Equation (4.75) parametrizes the corresponding space in each case. As before the internal edges of the subtriangulation can be flipped, and the coordinates change accordingly to the tropical cluster $\mathcal{A}$-mutations.

The space $\mathcal{A}_{\mathrm{SL}_2,S}(\mathbb{Z}^t)$ (resp. $\mathcal{A}_{\mathrm{SL}_2,S}(\mathbb{Q}^t)$, $\mathcal{A}_{\mathrm{SL}_2,S}(\mathbb{R}^t)$) corresponds to the space of rational with integer coordinates (resp. rational, real) $\mathcal{A}$-laminations on S. That follows from the analysis of Chapter 3. By analogy, in the general G-case one speaks of G-higher $\mathcal{A}$-lamination spaces. The cluster duality conjectures reviewed in Section 4.2.4 interpret the points of $(\mathcal{A}_{G,S})_0(\mathbb{Z}^t)$ as parametrizing the elements of $\mathbb{E}(\mathcal{X}_{G,S})$. The reasoning of Section 4.1.2 imply that $\mathbf{Sp}\mathcal{A}_{G,S}(\mathbb{R}^t)$ is a spherical compactification of $\mathcal{A}_{G,S}(\mathbb{R}_{>0})$.

4.4 Other higher Teichmüller spaces

The study of another class of higher Teichmüller spaces was pioneered in [BIW10] with the introduction of *spaces of maximal representations*. We discuss it briefly in Section 4.4.1. Maximal representations can be described in terms of positive boundary maps, in the spirit of Proposition 4.23 and Proposition 4.31. However, the notion of positivity in this case differs from the one developed in the previous section. In [GW18], Guichard and Wienhard introduced a broad notion of positivity generalizing both the positivity of configurations of flags introduced in the previous section and the positivity associated with maximal representations, under the name Θ-*positivity*. One defines Θ-positive representations in terms of Θ-positive boundary maps, and the spaces of Θ-positive representations are higher Teichmüller spaces generalizing both Hitchin's components and spaces of maximal representations. Interestingly, this led to the discovery of new Teichmüller spaces. We discuss Θ-positivity in Section 4.4.2.

4.4.1 Spaces of maximal representations

In this section we follow [BIW10] and [GW18].

A real Lie group is said to be of *Hermitian type* if it is connected, semi-simple, with finite center, without compact factor, and if the associated symmetric space is Hermitian. Let G be a Lie group of Hermitian type and let $\mathcal{H} = G/K$ be the associated symmetric space. Let S be a connected oriented compact surface of finite type. Let $\rho : \pi_1(S) \to G$ be a representation. It induces a map $\rho^* : S \to \mathcal{H}$. The *Toledo invariant* $T(S, \rho)$ associated to ρ is defined as the evaluation of the pullback of the class of the Kähler form on $\mathcal{H}$, on the fundamental class of S relative to its boundary. For any representation ρ one has:

$$|T(S, \rho)| \leq |\chi(S)| \cdot \mathrm{rk}(\mathcal{H}) \ , \tag{4.91}$$

where $\mathrm{rk}(\mathcal{H})$ is the rank of $\mathcal{H}$.

Definition 4.35. *A representation $\rho : \pi_1(S) \to G$ is said to be maximal if*

$$T(S, \rho) = |\chi(S)| \cdot \mathrm{rk}(\mathcal{H}) \ . \tag{4.92}$$

Let $\mathrm{Hom}_{\max}(\pi_1(S), G)$ be the subspace of maximal representations in the G-character variety of S.

From now on we assume that $\chi(S) \leq -1$. A Hermitian symmetric space if said to be *of tube type* if it is biholomorphic to a domain of the form $T_\Omega = V + i\Omega$ where V is a real vector space and $\Omega \subset V$ is a sharp convex cone. A Lie group of Hermitian type whose associated Hermitian symmetric space is of tube type is said to be a *Hermitian Lie group of tube type*. For example, the group $\mathrm{SL}_2(\mathbb{R}) \simeq \mathrm{SU}(1,1) \simeq \mathrm{Sp}_2(\mathbb{R})$ is Hermitian of tube type, for its associated Hermitian symmetric space is $\mathbb{H} = \mathbb{R} + i\mathbb{R}_+$. For $n \geq 2$, the group $\mathrm{Sp}_{2n}(\mathbb{R})$ is also Hermitian of tube type, with associated Hermitian symmetric space the Siegel upper-half space $\mathrm{Sym}_n(\mathbb{R}) + i\mathrm{Pos}_n(\mathbb{R})$. Here $\mathrm{Pos}_n(\mathbb{R}) \subset \mathrm{Sym}_n(\mathbb{R})$ is the subset of positive definite symmetric matrices.

Theorem 4.36 (1, 4.1 in [BIW10])**.** *Any maximal representation $\rho : \pi_1(S) \to G$ is faithful and discrete. Moreover, the image of ρ is Zariski-dense if and only if G is Hermitian of tube type.*

Hence we can restrict to the study of maximal G-representations when G is Hermitian of tube type. From now on we assume it to be case. Spaces of maximal G-representations of S generalize the classical Teichmüller space of S, in the following sense.

Theorem 4.37 (3 in [BIW10])**.** *A representation $\pi_1(S) \to\simeq \mathrm{PSL}_2(\mathbb{R})$ is maximal if and only if it is the holonomy representation of a complete hyperbolic metric in the interior of S.*

Maximal representations can be described in terms of boundary maps, as advertised in the introduction of this section. To this end, let us define the Shilov boundary of $\mathcal{H}$. The symmetric space $\mathcal{H}$ can be realized as a bounded domain $\mathcal{D}$ in a complex space $\mathcal{C}$. The *Shilov boundary* $\mathrm{Sh}(\mathcal{H})$ of $\mathcal{H}$ is the smallest closed subset of the boundary $\partial\mathcal{D}$ of $\mathcal{D}$ for which the maximum modulus principle holds, i.e.

$$\max_{z \in \mathrm{Sh}(\mathcal{H})} |f(x)| = \max_{z \in \partial\mathcal{D}} |f(x)| \tag{4.93}$$

for every holomorphic function f defined in a neighborhood of $\mathcal{D}$ in $\mathcal{C}$. For example, when $G = \mathrm{Sp}_{2n}(\mathbb{R})$ the Siegel half-plane is biholomorphic to the so-called Siegel disc, whose Shilov boundary can be identified with the space of Lagrangian subspaces in $\mathbb{R}^{2n}$. There is a notion of transversality for pairs of points in $\mathrm{Sh}(\mathcal{H})$. Any triple of points $(x, y, z) \in \mathrm{Sh}(\mathcal{H})^3$ can be assigned a *generalized Maslov index* $\beta_{\mathrm{Sh}(\mathcal{H})}(x, y, z) \in \mathbb{R}$ (see [Cle07] and references therein) which, under our assumption that G in of tube-type, takes only finitely many values. A triple $(x, y, z) \in \mathrm{Sh}(\mathcal{H})^3$ is *maximal* if its generalized Maslov index is the greatest possible.

Theorem 4.38 (8 in [BIW10])**.** *Let $\hat{S}$ be a hyperbolic surface homeomorphic to the interior of S and of finite area, so that one can consider its universal cover $\mathbb{H}$, and let $\rho : \pi_1(S) \to G$ be a representation. Then ρ is maximal if and only if there exists a $\pi_1(S)$-equivariant left-continuous map $\partial\mathbb{H} \to \mathrm{Sh}(G)$ which maps positively oriented triples in $\partial\mathbb{H}$ to maximal triples in $\mathrm{Sh}(G)$.*

Here the boundary map is stated in a language closer to the one of [Lab04] than to the one of [FG06], however both definitions are equivalent.

Proposition 4.39. *If S is closed, $\mathrm{Hom}_{\max}(\pi_1(S), G)$ is a union of connected components of $\mathrm{Hom}(\pi_1(S), G)$. If S has a non-empty boundary, then*

$$\mathrm{Hom}_{\max}(\pi_1(S), G) \subset \mathrm{Hom}^{\mathrm{Sh}(\mathcal{H})}(\pi_1(S), G) \ , \tag{4.94}$$

where the space on the right-hand side is defined as the space of representations in $\mathrm{Hom}(\pi_1(S), G)$ such that the image of the class of any simple closed curve on S bounding a hole or a puncture has at least one fixed point in $\mathrm{Sh}(\mathcal{H})$. Moreover $\mathrm{Hom}_{\max}(\pi_1(S), G)$ is a union of connected components of $\mathrm{Hom}^{\mathrm{Sh}(\mathcal{H})}(\pi_1(S), G)$.

In other words, spaces of maximal representations into a Hermitian Lie group of tube type provide new examples of higher Teichmüller spaces.

The condition for a triple of points on the Shilov boundary of $\mathcal{H}$ to be maximal can be recast in a form closer to what has been presented in the previous section. Following [GW18], we will consider the specific case of $\mathrm{Sp}_{2n}(\mathbb{R})$. Let us first define a sub-semigroup $\mathrm{Sp}_{2n}^{\succ 0}$ of $\mathrm{Sp}_{2n}(\mathbb{R})$, in the spirit of Cryer's splitting lemma (Proposition 1.7). Here the positivity comes from the sharp convex cone appearing in the tube type presentation of the Hermitian symmetric space corresponding to G rather than conditions on minors. Let:

$$V = \left\{ g \in \mathrm{Sp}_{2n,\mathbb{R}} \mid g = \begin{pmatrix} \mathrm{Id}_n & 0 \\ M & \mathrm{Id}_n \end{pmatrix},\ M \in \mathrm{Sym}_n(\mathbb{R}) \right\} \ , \tag{4.95}$$

$$W = \left\{ g \in \mathrm{Sp}_{2n,\mathbb{R}} \mid g = \begin{pmatrix} \mathrm{Id}_n & N \\ 0 & \mathrm{Id}_n \end{pmatrix},\ N \in \mathrm{Sym}_n(\mathbb{R}) \right\} \ , \tag{4.96}$$

$$H = \left\{ g \in \mathrm{Sp}_{2n,\mathbb{R}} \mid g = \begin{pmatrix} A & 0 \\ 0 & (A^t)^{-1} \end{pmatrix} \right\} \ . \tag{4.97}$$

Then one sets:

$$\mathrm{Sp}_{2n}^{\succ 0} = V^{\succ 0} H^{\circ} W^{\succ 0} \ , \tag{4.98}$$

where $V^{\succ 0}$ (resp. $W^{\succ 0}$) is defined as V (resp. W) except that the condition $M \in \mathrm{Sym}_n(\mathbb{R})$ (resp. $M \in \mathrm{Sym}_n(\mathbb{R})$) is replaced by $M \in \mathrm{Pos}_n(\mathbb{R})$ (resp. $N \in \mathrm{Pos}_n(\mathbb{R})$), and where H° stands for the connected component of the identity in H. Let now $(e_1, \dots, e_n, f_1, \dots, f_n)$ be a symplectic basis of $\mathbb{R}^{2n}$ and let $L_e = \mathrm{Vect}(e_1, \dots, e_n)$ and $L_f = \mathrm{Vect}(f_1, \dots, f_n)$. Any Lagrangian subspace L of $\mathbb{R}^{2n}$ transverse to L_f can be written as $g \cdot L_e$ for some $g \in V$. One says that the triple (L_e, L, L_f) is positive if $g \in V^{\succ 0} \subset V$ [GW18]. This coincides with the notion of maximal triples of points in the Shilov boundary of the Siegel upper half-space, acted on by $\mathrm{Sp}_{2n}(\mathbb{R})$.

4.4.2 Θ-positivity

Let us define Θ-positivity, following [GW18]. It generalizes both notions of positivity evoked above.

Let G be a real semisimple Lie group with finite center and let $\mathfrak{g} = \mathrm{Lie}(G)$ with Cartan decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$, where $\mathfrak{k}$ is the Lie algebra of the maximal compact subgroup of G. Let $\mathfrak{a} \subset \mathfrak{p}$ be a maximal abelian algebra in $\mathfrak{p}$, let $\Sigma = \Sigma(\mathfrak{g}, \mathfrak{a})$ be the restricted root system of $\mathfrak{g}$ with respect to $\mathfrak{a}$ [Gor94] and

$$\mathfrak{g} = \mathfrak{g}_0 \oplus \left(\bigoplus_{\alpha \in \Sigma} \mathfrak{g}_\alpha \right) \tag{4.99}$$

the corresponding root decomposition. Let Σ^+ and Σ^- be the spaces of positive and negative roots with respect to a fixed ordering, and let Δ be the corresponding set of positive simple roots. For any subset $\Theta \subset \Delta$ one sets

$$\mathfrak{u}_\Theta = \sum_{\alpha \in \Sigma^+_\Theta} \mathfrak{g}_\alpha \quad \text{and} \quad \mathfrak{u}^{\mathrm{opp}}_\Theta = \sum_{\alpha \in \Sigma^+_\Theta} \mathfrak{g}_{-\alpha} \ , \tag{4.100}$$

with $\Sigma^+_\Theta = \Sigma^+ \backslash \mathrm{Span}(\Delta - \Theta)$. One also defines

$$l_\Theta = \mathfrak{g}_0 \oplus \bigoplus_{\alpha \in \mathrm{Span}(\Delta - \Theta) \cap \Sigma^+} \mathfrak{g}_\alpha \oplus \mathfrak{g}_{-\alpha} \ . \tag{4.101}$$

Let $P_\Theta = \mathrm{Norm}_G(\mathfrak{u}_\Theta)$ and $P^{\mathrm{opp}}_\Theta = \mathrm{Norm}_G(\mathfrak{u}^{\mathrm{opp}}_\Theta)$, as well as $L_\Theta = P_\Theta \cap P^{\mathrm{opp}}_\Theta$. One has $\mathrm{Lie}(L_\Theta) = l_\Theta$. The group L_Θ acts on $\mathfrak{u}_\Theta$ via the adjoint action $\mathrm{ad} : L_\Theta \to \mathrm{End}(\mathfrak{u}_\Theta)$. Let $\mathfrak{z}_\Theta$ be the center of l_Θ, and for all $\beta \in \mathfrak{z}^*_\Theta$ set

$$\mathfrak{u}_\beta = \{ N \in \mathfrak{u}_\Theta \mid \mathrm{ad}(Z)N = \beta(Z)N, \ \forall Z \in \mathfrak{z}_\Theta \} \ , \tag{4.102}$$

so that

$$\mathfrak{u}_\Theta = \bigoplus_{\beta \in \mathfrak{z}^*_\Theta} \mathfrak{u}_\beta \ . \tag{4.103}$$

Any $\beta \in \mathfrak{z}^*_\Theta$ can be uniquely written as the restriction of a root in $\mathfrak{a}^* \cap \mathrm{Span}(\Theta)$ to $\mathfrak{z}_\Theta$, and hence we consider the β's as elements of $\mathfrak{a}^* \cap \mathrm{Span}(\Theta)$. Moreover, one has $[\mathfrak{u}_\beta, \mathfrak{u}_{\beta'}] = \mathfrak{u}_{\beta+\beta'}$, and hence as a Lie algebra $\mathfrak{u}_\Theta$ is generated by the $\mathfrak{u}_\beta$ for $\beta \in \Theta$.

Definition 4.40. *The group G admits a Θ-positive structure if for all $\beta \in \Theta$ there exists an L°_Θ-invariant sharp convex cone c_β in $\mathfrak{u}_\beta$.*

Theorem 4.41 (4.8 in [GW18])**.** *The group G has a Θ-positive structure if and only if there are two transverse points E_Θ, F_Θ in the partial flag variety G/P_Θ, such that there exists a connected component of $\Omega_{E_\Theta} \cap \Omega_{F_\Theta}$ with a semigroup structure. Here Ω_{E_Θ} denotes the space of flags in G/P_Θ transverse to E_Θ, and similarly for Ω_{F_Θ}.*

In the case of Lusztig's positivity, G is split, $\Theta = \Delta$, $P_\Theta = B$ and $L_\Theta = H$. When G is Hermitian of tube type, Σ is of type C_r so that $\Delta = \{\alpha_1, \dots, \alpha_r\}$. The choice $\Theta = \{\alpha_r\}$, such that P_Θ stabilizes a point on the Shilov boundary of the corresponding Hermitian symmetric space, defines a Θ-positive structure on G which corresponds to the one of the previous section.

In fact, one can understand exactly for which cases the condition of Definition 4.40 is satisfied, which leads to the following classification.

Theorem 4.42 (4.3 in [GW18])**.** *A semi-simple real Lie group G with finite center admits a Θ-positive structure if and only if (G, Θ) belongs to the following four cases:*

1. *G is a split real form and $\Theta = \Delta$,*
2. *G is of Hermitian type and $\Theta = \{\alpha_r\}$,*
3. *G is locally isomorphic to $\mathrm{SO}(p, q)$ for $p \neq q$ and $\Theta = \{\alpha_1, \dots, \alpha_{p-1}\}$,*
4. *G is a real form of F_4, E_6, E_7 or E_8 whose restricted root system is of type F_4, and $\Theta = \{\alpha_1, \alpha_2\}$.*

Let G be a group and $\Theta \subset \Delta$ such that G admits a Θ-positive structure. For every $\beta \in \Theta$ let $C_\beta = \exp(c_\beta) \subset U_\beta = \exp(u_\beta) \subset U_\Theta = \exp(u_\Theta)$. For all $\beta \in \Theta$ let $x_\beta : \mathfrak{u}_\beta \to U_\beta$ be the map

$$v \longmapsto \exp(v) \ . \tag{4.104}$$

Let $U^{\geq 0}_\Theta$ be the sub-semigroup of U_Θ generated by the C_β's, for all $\beta \in \Theta$. One defines $(U^{\mathrm{opp}}_\Theta)^{\geq 0}$ similarly.

Definition 4.43. *The Θ-nonnegative semigroup $G_\Theta^{\geq 0}$ is defined as $G_\Theta^{\geq 0} = U_\Theta^{\geq 0} L_\Theta^\circ (U_\Theta^{\mathrm{opp}})^{\geq 0}$.*

Let $W = \langle s_\alpha \mid \alpha \in \Delta \rangle$ be the Weyl group of G. There is at most one node $\beta_\Theta \in \Theta$ in the Dynkin diagram corresponding to Σ, such that it is connected to $\Delta - \Theta$. For all $\beta \in \Theta - \{\beta_\Theta\}$ let $\sigma_\beta = s_\beta$, and let σ_{β_Θ} be the longest element in the subgroup of the Weyl group generated by the s_α for $\alpha \in \{\beta_\Theta\} \cup (\Delta - \Theta)$. Let now

$$W(\Theta) = \langle \sigma_\beta \mid \beta \in \Theta \rangle \ . \tag{4.105}$$

In every case appearing in Theorem 4.42, $W(\Theta)$ is isomorphic to the Weyl group of a simple root system, in which the σ_β are the simple reflections. For example, when G is split real one has $W(\Theta) = W$, and when G is Hermitian of tube type, $W(\Theta) \simeq W_{A_1}$.

The group $W(\Theta)$ acts on the weight spaces u_β for $\beta \in \mathfrak{a}^* \cap \mathrm{Span}(\Theta)$. Let $w_\Theta^0 \in W(\Theta)$ be the longest element and let $\sigma_{i_1} \cdot \sigma_{i_l}$ be a reduced expression for w_Θ^0. For each $\beta \in \Theta$ let c_β° denotes the interior of the sharp convex cone c_β, and let $F_{\sigma_{i_1} \cdot \sigma_{i_l}} : c^\circ_{\beta_{i_1}} \times \cdots \times c^\circ_{\beta_{i_l}} \to U_\Theta$ be the map:

$$(v_{i_1}, \dots, v_{i_l}) \longmapsto x_{\beta_{i_1}}(v_{i_1}) \dots x_{\beta_{i_l}}(v_{i_l}) \tag{4.106}$$

In the spirit of Lusztig's parametrization of the positive sub-semigroup $G_{>0}$ of Proposition 4.17, one can show that the image $U_\Theta^{>0}$ of the map $F_{\sigma_{i_1} \cdot \sigma_{i_l}}$ is independent of the reduced expression for w_Θ^0.

Definition 4.44. *The sub-semigroup $U_\Theta^{>0}$ of G is called the Θ-positive semigroup of U_Θ. One defines $(U_\Theta^{>0})^{\mathrm{opp}}$ in a similar way. The Θ-positive semigroup $G_\Theta^{>0}$ is defined as the sub-semigroup in G generated by $U_\Theta^{>0}$, $(U_\Theta^{>0})^{\mathrm{opp}}$ and L_Θ^0.*

This notion of positivity allows to define a broad class of higher Teichmüller spaces.

Definition 4.45. *Let G be a semisimple Lie group with a Θ-positive structure, and let S_g be the oriented closed surface of genus g. A representation $\pi_1(S_g) \to G$ is said to be Θ-positive if there exists a ρ-equivariant positive map $\partial \mathbb{H} \to G/P_\Theta$ which maps positively oriented triples in ∂H to positive triples in G/P_Θ.*

Conjectures about the properties of the spaces of Θ-positive representations we proposed in [GW18] and subsequently proved in [GLW21] after having shown that Θ-positive representations are Θ-Anosov. Let us state two main results of [GLW21].

Theorem 4.46 (A in [GLW21])**.** *If G admits a Θ-positive structure, then there exists a connected component of the G-character of variety of a closed oriented surface which consists solely of discrete and faithful representations.*

Theorem 4.47 (D in [GLW21])**.** *Let S be a closed oriented surface and G a semi-simple Lie group admitting a Θ-positive structure. The set of Θ-positive representations $\mathrm{Hom}_{\Theta-\mathrm{pos}}(\pi_1(S), G)$ is a union of connected components of $\mathrm{Hom}^*(\pi_1(S), G)$, where $\mathrm{Hom}^*(\pi_1(S), G)$ is the space of homomorphisms which do not factor through a parabolic subgroup of G, even when restricted to a finite index subgroup of $\pi_1(S)$.*

Therefore Θ-positivity yields higher Teichmüller spaces. In the classification Theorem 4.42, the cases when G is split real and when G is Hermitian of tube type correspond to the already-known Hitchin components and spaces of maximal representations, however the two remaining families ($G = \mathrm{SO}(p, q)$ for $p \neq q$, and the exceptional types) define new higher Teichmüller spaces.

Higher Teichmüller spaces when $G = \mathrm{SO}(p, q)$ for $p \neq q$ have been studied in [BP21, AABC$^+$18, AABC$^+$19, BCGP$^+$21] using various methods. The notion of *magical $\mathfrak{sl}_2$-triple* was introduced recently in [BCGP$^+$21] in the context of G-Higgs bundles. The groups G admitting such magical $\mathfrak{sl}_2$-triple are exactly the ones admitting a Θ-positive structure.

* * * * * * *

The algebro-geometric duals to cluster algebras are cluster varieties, which are special instances of positive varieties. To the mutation class of a seed one can associate a cluster ensemble consisting of a pair $(\mathcal{X}, \mathcal{A})$ of cluster varieties and enjoying various interesting features.

Moduli spaces of framed or decorated G-local systems on a ciliated surface S form examples of cluster ensembles of geometric interest. Their underlying positive structure allows to consider their points over various semi-fields. In particular, the $(\mathbb{R}_{>0})$-points of these cluster varieties generalize the Teichmüller space with holes and the decorated Teichmüller space of a ciliated surface. They form connected components of the G-character variety of S which consist solely of discrete and faithful representations, and are identified with Hitchin's components. It is interesting that the positive structure on these varieties allows to account for a discreteness condition, which is non-algebraic in nature, within algebraic geometry.

By analogy with the classical laminations spaces of S, the spaces of rational G-higher laminations on S are defined as the $\mathbb{Q}^t$-points of the cluster varieties $\mathcal{X}_{G,S}$ and $\mathcal{A}_{G,S}$. In Chapter 14, we will explore directions towards the definition of combinatorial objects on S generalizing the rational laminations on S as systems of curves, as presented in Chapter 3.

When G is of type A, the spaces $\mathcal{X}_{\mathrm{PGL}_m,S}$ and $\mathcal{A}_{\mathrm{SL}_m,S}$ admit special coordinate systems for each triangulation Γ of S. These coordinates are expressed in terms of flags.

Lastly, we have discussed another class of higher Teichmüller spaces defined for Hermitian Lie groups of tube type, as well as the recent notion of Θ-positivity which leads to the definition of a broad class of higher Teichmüller spaces containing both Hitchin's components and spaces of maximal representations, as well as completely new instances.

Part II

Supersymmetry, string theory and holography

Chapter 5

Supersymmetry in four dimensions

Our current understanding of fundamental particle physics at energies below a few TeV is gathered in the Standard Model of particle physics, which is a gauge theory with gauge group $\mathrm{U}(1) \times \mathrm{SU}(2) \times \mathrm{SU}(3)$ (possibly quotiented by a subgroup of $\mathbb{Z}/6\mathbb{Z}$). It describes the electromagnetic, weak and strong nuclear forces in a unified framework. The Standard Model is currently tested at many particle colliders around the world, and the agreement between theoretical predictions and experiments is astonishing [Z+20].

However the Standard Model has some short-comings that are motivations to look for an extension of it. Our Universe is believed to be governed by four fundamental forces at low energies: the aforementioned three and gravity. At energies far below the Planck scale $\Lambda_P = 10^{19}$ GeV, gravity is negligible, and that makes it possible for the Standard Model to account faithfully for most of fundamental physics phenomena – but this cannot be true anymore as soon as gravity has to be taken into account, e.g. at energies comparable to Λ_P. Moreover, there are some features of the Standard Model that still lack a deep understanding, as well as physics besides gravity which cannot be described within the Standard Model. Unraveling such mysteries would provide hints on how to extend the Standard Model into a more complete theory, and hence to an even better description of physics. Let us discuss briefly three such conundrums.

- The first is known as the *hierarchy problem*. The mass of the Higgs boson $M_H \sim 125$ GeV is related to the scale of electroweak symmetry breaking $\Lambda_{\mathrm{EW}} \sim 246$ GeV, for both of which there is now a good amount of experimental evidence. The Higgs boson being a scalar particle, one expects its mass to receive quadratic corrections from quantum effects, which would in principle give it an observable mass comparable to the energy scale up to which the Standard Model is defined, except if some particularly fine-tuned cancellations between these quantum corrections happen. This fine-tuning would seemingly go against a principle known as *naturalness*. The hierarchy problem is often boldly expressed as: why is the Higgs mass so small compared to the Planck scale $\Lambda_P \sim 10^{18}$ GeV?

- The study of galaxy rotation curves, gravitational lensing, anisotropies of the Cosmic Microwave Background, and direct observations such as the Bullet cluster suggest the existence of *dark matter*, presumably constituted of massive particles interacting only very weakly with electromagnetism. The Standard Model does not describe particles with the good properties to be the main constituent of dark matter; this points towards searching for consistent extensions containing such particles.

- Under renormalization, the observable strengths of the gauge couplings in the Standard Model depend on the energy scale, and seem to converge approximately around the so-called Grand Unification scale $\Lambda_{\mathrm{GUT}} \sim 10^{18}$ GeV. This suggest that the three microscopic forces, i.e. electromagnetism as well as the weak and strong nuclear forces, might unify above this energy scale. This would be very interesting theoretically and philosophically, however for the convergence to be exact one needs to consider a more complete theory than the Standard Model.

Extensions of the Standard Model must be constructed with care, for the resulting theory might otherwise be inconsistent phenomenologically. No-go theorems constraining the possible infinitesimal symmetries of reasonable quantum field theories were proved in the 1960's, culminating with the Coleman–Mandula theorem [CM67]: the most general Lie algebra of infinitesimal symmetries of a phenomenologically viable quantum field theory must be a direct product of the Poincaré algebra and an internal algebra [Wei13]. This theorem severely constraints quantum field theories of phenomenological interest.

A way out the Coleman–Mandula theorem is to weaken the required assumptions on the algebraic structure of the set of infinitesimal symmetries of quantum field theories. One may for example allow it to be a Lie superalgebra (i.e. a $\mathbb{Z}_2$-graded Lie algebra) instead of a conventional Lie algebra [HLS75]. In that case some of the infinitesimal symmetries are odd, and they link bosonic states of the theory to fermionic states, and vice–versa. The quantum field theories which have such infinitesimal symmetries are said to be *supersymmetric*.

Extensions of the Standard Model that are supersymmetric solve some of the flaws mentioned above: there are natural candidates for the particle constituting dark matter, the unification of gauge couplings is better or even exact in many models, and quantum corrections are under much more control than in non-supersymmetric gauge theories, which lets one hope for a resolution of the hierarchy problem. For example, some models such as the Minimal Supersymmetric extension of the Standard Model (MSSM) and its variants offer a natural explanation of the weak hierarchy problem [Z+20], even if the latest data from the Large Hadron Collider suggests that there is some fine-tuning in the Higgs mass, after all. Other models such as split supersymmetry [AHD05] embrace this scenario and show that supersymmetric extensions of the Standard Model can be implemented in a rich variety of ways. Even if today's available experimental data might induce pessimistic feelings regarding models such as the MSSM, the non-observation of a particular supersymmetric extension of the Standard Model at LHC[1] does not challenge the very idea of supersymmetry at all, but rather the specific implementation under study.

All the more so as supersymmetry may somehow be necessary to describe quantum gravity; it is in particular the case in string theory, which is one of the more developed and sprawling approach to it: supersymmetry seems to always be necessary in order to obtain phenomenologically interesting string theories. Should this fact be true, the question would not be whether supersymmetry exists at all but rather how it has to be implemented in the description of our universe.

Furthermore, whether supersymmetry must play a role in the description of fundamental physics or not, it can be seen as a mere mathematical constraint which grants much control on quantum field theories. It allows one to wonder about deep questions which are less (if at all) accessible in general quantum field theories: supersymmetry is a very interesting theoretical laboratory. Quantum field theories with the minimal possible amount of supersymmetry in four dimensions satisfy powerful non-renormalization theorems reviewed in Section 5.2, which tame radiative corrections a lot. Holomorphy and holomorphic decoupling allow moreover to peep into the strong coupling behaviour of supersymmetric gauge theories: we will review in Section 5.4 the quantum dynamics of SQCD (the analogue of quantum chromodynamics with $\mathcal{N} = 1$ supersymmetry) which is worked out by extensive use of holomorphy.

The more supersymmetries, the more control: we discuss the quantum dynamics of gauge theories in four dimensions with $\mathcal{N} = 2$ in Section 5.6 and $\mathcal{N} = 4$ in Section 5.7, which is the maximal possible amount of supersymmetries in non-gravitational theories in four dimensions. In these cases one can do much more than a mere peep into the strong coupling dynamics.

Section 5.3 and Section 5.5 tackle questions of a more phenomenological origin: namely, how supersymmetry can be broken, and how it can be implemented in interesting ways to the description of the physics of our world.

A last important feature of supersymmetric quantum field theories is that they often allow for many different choices of vacua. The set of all possibles vacua of such a theory is its *moduli space*. It is always a complex algebraic variety, and is often endowed with additional mathematical structure, therefore becoming Kähler manifolds, special Kähler manifolds or hyperkähler manifolds, for example. When looking at supersymmetric gauge theories with gauge group G, one often uncovers bridges between the representation theory of G, the geometry of the moduli space, and the low-energy dynamics of the quantum field theory.

What follows in this chapter is mostly based on [Ber15]. Other lectures I used include [Arg01] and [IS07], as well as [PT98, SS00] as far as dynamical supersymmetry breaking is concerned.

[1] or rather, its testable avatars [Z+20]

5.1 General definitions and properties

Super-Poincaré algebras.

Let $\mathcal{N} \in \mathbb{Z}_{\geq 0}$. A 4-dimensional super-Poincaré algebra with $\mathcal{N}$ supersymmetries is a super-Lie algebra

$$\mathfrak{g} = \mathfrak{g}_0 \oplus \mathfrak{g}_1 \tag{5.1}$$

satisfying some requirements derived in [HLS75] and listed below. By definition [Kac77] the super Lie bracket

$$[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g} \tag{5.2}$$

is bilinear and for $i, j, k \in \mathbb{Z}_2$ and $g_i \in \mathfrak{g}_i$ it satisfies

$$[\mathfrak{g}_i, \mathfrak{g}_j] \subset \mathfrak{g}_{i+j} \ , \tag{5.3}$$

$$[g_i, g_j] = -(-1)^{ij}[g_j, g_i] \ , \tag{5.4}$$

$$(-1)^{ik}[g_i, [g_j, g_k]] + (-1)^{ji}[g_j, [g_k, g_i]] + (-1)^{kj}[g_k, [g_i, g_j]] = 0 \ . \tag{5.5}$$

The even homogeneous part $\mathfrak{g}_0$ of degree 0 is the direct sum $\mathfrak{g}_0 = \mathfrak{p} \oplus \mathfrak{b}$ of the 4-dimensional Poincaré algebra $\mathfrak{p} = \mathfrak{so}(1,3) \oplus \mathbb{R}^4$ with usual generators $M_{\mu\nu}, P_\mu$ and a reductive Lie algebra $\mathfrak{b}$ which commutes with $\mathfrak{p}$.

The odd part $\mathfrak{g}_1$ of $\mathfrak{g}$ is of the form

$$\left[\left(\frac{1}{2}, 0\right) \otimes V\right] \oplus \left[\left(0, \frac{1}{2}\right) \otimes V^*\right] \ , \tag{5.6}$$

where $\left(\frac{1}{2}, 0\right)$ and $\left(0, \frac{1}{2}\right)$ stand for the two non-isomorphic spin representations of $\mathfrak{so}(1,3)$ and where V is an N-dimensional representation of $\mathfrak{b}$. Let Q^I_α, $\alpha = 1, 2$; $I = 1, ..., N$, denote generators of $\left[\left(\frac{1}{2}, 0\right) \otimes V\right]$ and let $\overline{Q}^I_{\dot\alpha}$, $\dot\alpha = 1, 2$; $I = 1, ..., N$, generators of $\left[\left(0, \frac{1}{2}\right) \otimes V^*\right]$. In our conventions $(Q^I_\alpha)^\dagger = \overline{Q}^I_{\dot\alpha}$. These generators of $\mathfrak{g}_1$ are called *supercharges*; they commute with infinitesimal space-time translations, and:

$$\left[M_{\mu\nu}, Q^I_\alpha\right] = i(\sigma_{\mu\nu})^\beta_\alpha Q^I_\beta \ , \tag{5.7}$$

$$\left[M_{\mu\nu}, \overline{Q}^{I\dot\alpha}\right] = i(\overline{\sigma}_{\mu\nu})^{\dot\alpha}_{\dot\beta} \overline{Q}_{I\dot\alpha} \ , \tag{5.8}$$

$$\left[Q_{I\alpha}, \overline{Q}^J_{\dot\beta}\right] = 2\sigma^\mu_{\alpha\dot\beta} P_\mu \delta^{IJ} \ , \tag{5.9}$$

$$[Q_{I\alpha}, Q_{J\beta}] = \epsilon_{\alpha\beta} Z^{IJ} \ , \tag{5.10}$$

$$\left[\overline{Q}^I_{\dot\alpha}, \overline{Q}^J_{\dot\beta}\right] = \epsilon_{\dot\alpha\dot\beta} \left(Z^{IJ}\right)^* \ . \tag{5.11}$$

In Equation (5.7) and Equation (5.8) $M_{\mu\nu}$ stands for the usual generators of the Lorentz algebra $\mathfrak{so}(1,3)$ and $\sigma_{\mu\nu} = (\sigma^\mu \overline{\sigma}^\nu - \eta^{\mu\nu})$ are the 2-index Pauli matrices: the Q^I_α and $\overline{Q}^{I\dot\alpha}$ transform as left and right Weyl spinors under the Lorentz group. Moreover, one sees that Q^I_1 and $\overline{Q}^I_{\dot 2}$ raise the z component of the spin (J_3) by half a unit, while Q^I_2 and $\overline{Q}^I_{\dot 1}$ lower the z component of the spin by half a unit. Hence the supersymmetry generators Q^I_α and $\overline{Q}^{I\dot\alpha}$ map fermionic states to bosonic ones, and vice versa.

In Equation (5.9) the P_μ are the generators of space-time translations, and the σ^μ are the usual Pauli matrices. This equation indicates that an infinitesimal supersymmetry translation is somehow the square root of an infinitesimal space-time translation. It also implies that in a supersymmetric theory defined as a representation of $\mathfrak{g}$ on a Hilbert space the energy of any state is positive:

$$0 \leq \sum_{\alpha,\dot\alpha=1}^{2} ||(Q_{1\alpha})^\dagger |\phi\rangle||^2 + ||Q_{1\alpha} |\phi\rangle||^2 = \sum_{\alpha,\dot\alpha=1}^{2} 2\sigma^\mu_{\alpha\dot\alpha} \langle\phi| P_\mu |\phi\rangle = 4\delta^{\mu 0} \langle\phi| P_\mu |\phi\rangle = 4 \langle\phi| P_0 |\phi\rangle \ . \tag{5.12}$$

Lastly in Equation (5.10) and Equation (5.11) $\epsilon_{\alpha\beta}$ and $\epsilon_{\dot\alpha\dot\beta}$ are the totally antisymmetric Levi-Civita symbols and hence $Z^{IJ} = -Z^{JI}$. The latter matrix is the central charge of the algebra – it is in the center of $\mathfrak{g}$. Note that both equations are trivial when $\mathcal{N} = 1$.

Moreover, there are generators in $\mathfrak{b}$ which do not commute with the Q^I_α and $\overline{Q}^I_{\dot\alpha}$: these are called *R-symmetry* generators of the algebra.

The Haag–Lopuszanski–Sohnius theorem [HLS75] states that the most general graded Lie algebra of infinitesimal symmetries of the S-matrix in a nice massive quantum field theory in four dimensions is the direct sum of a super-Poincaré algebra as described above and a finitely generated internal part commute with the super-Poincaré algebra.

$\mathcal{N} = 1$ superspace and superfields.

From now on and until the last section of this chapter let us assume parsimoniously that we are considering a $d = 4$ $\mathcal{N} = 1$ quantum field theory, i.e. a representation of the four dimensional $\mathcal{N} = 1$ super Poincaré algebra on a Hilbert space of states $\mathcal{H}$.

In order to describe the fields corresponding to irreducible representations of the super Poincaré algebra is it convenient to introduce the superspace $M^{4|1} = \overline{\mathrm{OSp}(4|1)}/\mathrm{SO}(1,3)$ in which points are described by triples $(x^\mu, \theta_\alpha, \overline{\theta}_{\dot\alpha})$ where the x^μ are the usual space-time coordinates in four-dimensional Minkowski space and where θ_α and $\overline{\theta}_{\dot\alpha}$ for $\alpha, \dot\alpha = 1, 2$ are anti-commuting Grassmannian coordinates. The supersymmetry generators act as translations in superspace:

$$Q_\alpha = -i\partial_\alpha - \sigma^\mu_{\alpha\dot\beta}\overline{\theta}^{\dot\beta}\partial_\mu, \qquad \overline{Q}_{\dot\alpha} = i\partial_{\dot\alpha} - \theta^\beta\sigma^\mu_{\beta\dot\alpha}\partial_\mu \ . \tag{5.13}$$

It is convenient to define the following two covariant derivatives in superspace:

$$D_\alpha = \partial_\alpha + i\sigma^\mu_{\alpha\dot\beta}\overline{\theta}^{\dot\beta}\partial_\mu, \qquad \overline{D}_{\dot\alpha} = \partial_{\dot\alpha} + i\theta^\beta\sigma^\mu_{\beta\dot\alpha}\partial_\mu \ ; \tag{5.14}$$

they anticommute with the supersymmetry generators.

Because of the fact that the Grassmannian coordinates are anti-commuting any (smooth) function on $M^{4|1}$ admits a Taylor-like expansion:

$$Y(x,\theta,\overline{\theta}) = f(x)+\theta\psi(x)+\overline{\theta}\overline{\chi}(x)+\theta\theta m(x)+\overline{\theta}\overline{\theta}n(x) = \theta\sigma^\mu\overline{\theta}v_\mu(x)+\theta\theta\overline{\theta}\overline{\lambda}(x)+\overline{\theta}\overline{\theta}\theta\rho(x)+\theta\theta\overline{\theta}\overline{\theta}d(x) \ . \tag{5.15}$$

A superfield is a function on $M^{4|1}$ such that

$$Y(x+\delta x, \theta+\delta\theta, \overline{\theta}+\delta\overline{\theta}) = e^{-i\left(\epsilon Q+\overline{\epsilon}\overline{Q}\right)}Y(x,\theta,\overline{\theta})e^{i\left(\epsilon Q+\overline{\epsilon}\overline{Q}\right)} \ . \tag{5.16}$$

A general recipe to construct supersymmetric actions is to consider expressions such as:

$$\mathcal{S} = \int \mathrm{d}x^4\mathrm{d}^2\theta\mathrm{d}^2\overline{\theta}Y(x,\theta,\overline{\theta}) = \int \mathrm{d}x^4\mathcal{L}(\phi(x),\psi(x),A_\mu(x),\dots) \ . \tag{5.17}$$

Instead of considering general superfields one can impose supersymmetric constraints to obtain more elementary building blocks for our supersymmetric Lagrangian densities, in fact those corresponding to irreducible representations of the supersymmetry algebra. In what follows since we will most of the time consider quantum field theories in which the spin of particles is smaller than 1, we will be working with two kinds of supermultiplets.

- *Chiral superfields* Φ, which satisfy
$$\overline{D}_{\dot\alpha}\Phi = 0 \ . \tag{5.18}$$
The physical degrees of freedom they describe consist of a complex scalar field $\phi(x)$ and a Weyl fermion $\psi(x)$. Off-shell there is also an auxiliary field F. In fact:
$$\Phi = \phi(x) + \sqrt{2}\theta\psi(x) + i\theta\sigma^\mu\overline{\theta}\partial_\mu\phi(x) - \theta\theta F(x) - \frac{i}{\sqrt{2}}\theta\theta\partial_\mu\psi(x)\sigma^\mu\overline{\theta} - \frac{1}{4}\theta\theta\overline{\theta}\overline{\theta}\Box\phi(x) \ . \tag{5.19}$$
Chiral superfields describe matter, and hence they are sometimes refer to as matter superfields and supermultiplets.

- *Vector superfields* V, also called *real superfields*since they satisfy
$$\overline{V} = V \ . \tag{5.20}$$
They are the supersymmetric generalization of gauge fields. They take value in the Lie algebra of a gauge group G: $V = V_a T^a$, and the supersymmetric generalization of a gauge transformation is
$$e^V \to e^{i\overline{\Lambda}}e^V e^{-i\Lambda} \ . \tag{5.21}$$

In the so-called Wess–Zumino gauge, a vector superfield reads:

$$V_{WZ} = \theta\sigma^\mu\overline{\theta}v_\mu(x) + i\theta\theta\overline{\theta}\overline{\lambda}(x) + i\overline{\theta}\overline{\theta}\theta\lambda(x) + \frac{1}{2}\theta\theta\overline{\theta}\overline{\theta}D(x) \ , \tag{5.22}$$

from which one sees that the degrees of freedom of a vector superfield consists of a gauge boson $v_\mu(x)$, a Weyl fermion $\lambda(x)$ called a gaugino and an auxiliary field $D(x)$.

The supersymmetric generalization of the field strength of a vector field is

$$W_\alpha = -\frac{1}{4}\overline{DD}(e^{-V}D_\alpha e^V) = -i\lambda_\alpha(y) + \theta_\alpha D(y) + i(\sigma^{\mu\nu}\theta)_\alpha F_{\mu\nu}(y) + \theta\theta(\sigma^\mu D_\mu\overline{\lambda})_\alpha(y) \ , \tag{5.23}$$

where $y^\mu = x^\mu + i\theta\sigma^\mu\overline{\theta}$ and where $F_{\mu\nu}$ is the usual field strength of the vector field $v_\mu(x)$.

The most general gauge-matter action (with canonical Kähler potential – see below) writes:

$$\begin{aligned}\mathcal{L} =& \frac{1}{32\pi}\mathrm{Im}\left(\tau\int \mathrm{d}^2\theta\,\mathrm{Tr}\,W^\alpha W_\alpha\right) + \sum_A \xi_A\int \mathrm{d}^2\theta\mathrm{d}^2\overline{\theta}V^A \\ &+ \int \mathrm{d}^2\theta\mathrm{d}^2\overline{\theta}\ \overline{\Phi}e^V\Phi + \int \mathrm{d}\theta^2 W(\Phi) + \int \mathrm{d}\overline{\theta}^2\overline{W}(\overline{\Phi}) \ , \ .\end{aligned} \tag{5.24}$$

In order to obtain a Lagrangian with canonical normalization of the fields after integration over the Grassmanian coordinates one must rescale the vector superfield as $V \to 2gV$, where g is the Yang–Mills coupling. One then obtains:

$$\begin{aligned}\mathcal{L} =& \,\mathrm{Tr}\left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - i\lambda\sigma^\mu D_\mu\overline{\lambda} + D^2\right] + \frac{\Theta}{32\pi^2}g^2\,\mathrm{Tr}\,F_{\mu\nu}\tilde{F}^{\mu\nu} + g\sum_A \xi_A D^A \\ &+ (\overline{D_\mu\Phi})D^\mu\Phi - i\psi\sigma^\mu D_\mu\overline{\psi} + \overline{F}F + i\sqrt{2}g\overline{\phi}\lambda\psi - i\sqrt{2}g\overline{\psi\lambda}\phi + g\overline{\phi}D\phi \\ &- \frac{\partial W}{\partial\phi^i}F^i - \frac{\partial\overline{W}}{\partial\overline{\phi}_i}\overline{F}_i - \frac{1}{2}\frac{\partial^2 W}{\partial\phi^i\partial\phi^j}\psi^i\psi^j - \frac{1}{2}\frac{\partial^2\overline{W}}{\partial\overline{\phi}_i\partial\overline{\phi}_j}\overline{\psi}_i\overline{\psi}_j \ .\end{aligned} \tag{5.25}$$

In Equation (5.24) we have denoted τ the complexified gauge coupling:

$$\tau = i\frac{4\pi}{g^2} + \frac{\Theta}{2\pi} \ , \tag{5.26}$$

where Θ is the usual theta angle of Yang–Mills theories. The first and third terms in Equation (5.24) are the kinetic terms for respectively the vector superfields collectively denoted V and the chiral superfields collectively denoted Φ; g is the charge of Φ under the gauge group G with respect to which V transforms in the adjoint. The second term in Equation (5.24) is called a Fayet–Illiopoulos term, and exists only for abelian factors U(1) of the gauge group G, over which the sum runs. The W which appears in the last two terms of Equation (5.24) is called the superpotential and it must be a holomorphic (analytic) function of the chiral superfield. The superpotential terms contain Yukawa interaction terms for the chiral superfields.

In any integral over superspace one can distinguish between terms that are integrals over half-superspace only (this is the case for the superpotential terms only in Equation (5.24)) and which are called F-terms, and the others which are called D-terms.

The kinetic term for chiral multiplets in Equation (5.24) is actually not the most general one:

$$\int \mathrm{d}^2\theta\mathrm{d}\overline{\theta}^2 K(\overline{\Phi}e^{-2gV},\Phi) \qquad \text{with } K(a,b) = \sum_{m,n=1}^{\infty} c_{mn}b^m a^n, \quad c_{mn} = c_{nm}^* \ , \tag{5.27}$$

is a perfectly fine kinetic term for the chiral supermultiplets. The function K is called the Kähler potential, and the choice in Equation (5.24) is said to be the canonical Kähler potential.

Integrating out the auxiliary fields D^a and F^i in Equation (5.24) using their equation of motion:

$$\overline{F}_i = \frac{\partial W}{\partial\phi^i}, \qquad D^a = -g\overline{\phi}T^a\phi - g\xi^a \ , \tag{5.28}$$

where $\xi^a = \xi_A$ if T^a is the generator of a U(1) factor in G and 0 otherwise, yields

$$\mathcal{L} = \mathrm{Tr}\left[-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - i\lambda\sigma^\mu D_\mu\overline{\lambda}\right] + \frac{\theta_{\mathrm{YM}}}{32\pi^2}g^2\,\mathrm{Tr}\,F_{\mu\nu}\tilde{F}^{\mu\nu} + (\overline{D_\mu\Phi})D^\mu\Phi - i\psi\sigma^\mu D_\mu\overline{\psi} \tag{5.29}$$

$$+ i\sqrt{2}g\overline{\phi}\lambda\psi - i\sqrt{2}g\overline{\psi\lambda}\phi - \frac{1}{2}\frac{\partial^2 W}{\partial\phi^i\partial\phi^j}\psi^i\psi^j - \frac{1}{2}\frac{\partial^2\overline{W}}{\partial\overline{\phi}_i\partial\overline{\phi}_j}\overline{\psi}_i\overline{\psi}_j - V(\phi,\overline{\phi})\ , \tag{5.30}$$

where

$$V(\phi,\overline{\phi}) = \frac{\partial W}{\partial\phi^i}\frac{\partial\overline{W}}{\partial\overline{\phi}_i} + \frac{g^2}{2}\sum_a\left|\overline{\phi}_i(T^a)^i_j\phi^j + \xi^a\right|^2 = \overline{F}F + \frac{1}{2}D^2 \geq 0 \tag{5.31}$$

is the *scalar potential* of the theory (defined as before as the non-derivative scalar part of the Hamiltonian). The equation of motion for the auxiliary fields $\overline{F}_i$ has to be modified if the Kähler potential is not the canonical one. Now comes the following important result:

Proposition 5.1. *The (classical) supersymmetric vacua of the theory defined by Equation* (5.29) *are exactly the zeros of the scalar potential $V(\phi,\overline{\phi})$. The classical moduli space (of vacua) of the theory is the (analytic) algebraic subvariety defined by $V(\phi,\overline{\phi}) = 0$ in the field space, parameterized by the ϕ_i. The equations $F_i = 0$ are called F-term equations, and the ones $D^a = 0$ are called D-term equations.*

Proof. A vacuum is a Lorentz-invariant field configuration of (locally) minimal energy. The result follows from the fact that in a supersymmetric theory the energy of any state is positive and that a state has zero energy if and only if it is supersymmetric. The fact that the classical moduli space of any supersymmetric quantum field theory is an algebraic variety is proved in [LT96]. The idea is that the classical moduli space can be defined by imposing the D-term equations and identifying gauge equivalent configurations on the space of solutions to the F-terms. The latter is clearly is a subvariety of the field space, and the solutions to the D-terms modulo gauge is exactly the G-symplectic quotient of $\mathcal{F}$, which is homeomorphic to the (GIT) quotient of $\mathcal{F}$ by the complexified gauge group $G^{\mathbb{C}}$: this is a version of the Kempf–Ness theorem [KN79]. □

R-symmetry

Remember that the R-symmetry is generated by those elements in the reductive part $\mathfrak{b}$ of the super-Poincaré algebra which do not commute with the supersymmetry generators (but remember that $\mathfrak{b}$ commutes with $\mathfrak{p}$). In four-dimensional theories with $\mathcal{N} = 1$, the group of R-symmetries is at most U(1). It is convenient to define its action on superspace by defining the R-charges of the odd coordinates: $R[\theta_\alpha] = 1$ and $R[\overline{\theta}_{\dot\alpha}] = -1$. This and the rules of Grassmanian calculus imply in turn that $R[\mathrm{d}\theta_\alpha] = -1$ and $R[\mathrm{d}\overline{\theta}_{\dot\alpha}] = +1$. Since the Lagrangian density is uncharged under global symmetries one sees that the superpotential must have R-charge 2 in order for the R-symmetry to be unbroken, and that vector superfields necessarily have R-charge 0. Moreover from the expansion in components of a chiral $\Phi = (\phi,\psi)$ and a vector $V = (\lambda, v_\mu)$ superfield it is necessarily the case that

$$R[\psi] = R[\phi] - 1 = R[\Phi] - 1, \qquad R[\lambda] = R[v_\mu] + 1 = R[V] + 1 = 1\ . \tag{5.32}$$

5.2 Renormalization and anomalies.

Non-renormalization theorems.

One of the most interesting properties of supersymmetric quantum field theories is that quantum corrections are under much better control than in a general quantum field theory. Using Feynman supergraphs techniques [GSR79] (see also [Sei93]), in a supersymmetric quantum field theory containing chiral and vector superfields only one can prove that all contributions to the effective action that can be generated by loop diagrams can be written as an integral over $\mathrm{d}^2\theta\mathrm{d}^2\overline{\theta}$, hence by definition it is a D-term. This contribution cannot affect superpotential terms since they are F-terms.

Proposition 5.2. *The superpotential in a four-dimensional supersymmetric quantum field theory is tree-level exact in perturbation theory.*

The superpotential may still receive corrections at the non-perturbative level and we will see many examples of that later on. D-terms on the contrary in general do receive contributions at any order in perturbation theory.

The perturbative tree-exactness of the superpotential imply that the renormalization of the couplings appearing in the superpotential can be expressed in terms of the renormalization of the chiral fields appearing in it. In general if a chiral field Φ renormalizes as $\overline{\Phi}\Phi \to Z_\Phi \overline{\Phi}\Phi$ one defines the *anomalous dimension* γ_Φ of Φ as

$$\gamma_\Phi = -\frac{\partial \ln Z_\Phi}{\partial \ln \mu} \ . \tag{5.33}$$

The renormalized scaling dimension of Φ reads

$$\dim \Phi = 1 + \frac{1}{2}\gamma_\Phi \ . \tag{5.34}$$

Concerning the gauge part of the theory since the Θ coupling does not enter the equations of motion it does not renormalize. The complexified gauge coupling τ however does renormalize but at one-loop only in perturbation theory, because of the supersymmetric non-renormalization theorems. Let G be the gauge group of the theory, and assume that there are matter superfields Φ_i in representations R_i of G. Then:

$$\beta_\tau = 3T(\mathrm{Adj}) - \sum_i T(R_i) =: b \ , \tag{5.35}$$

where $T(R_i)$ is the Dynkin label of the representation R_i defined as:

$$\mathrm{Tr}\big(T^A_{R_i} T^B_{R_i}\big) = T(R_i)\delta^{AB} \ . \tag{5.36}$$

For example if $G = \mathrm{SU}(N)$ one has

$$T(\mathrm{Adj}) = N, \quad T_{\square} = T_{\overline{\square}} = \frac{1}{2}, \quad T_{\substack{\square \\ \square}} = \frac{N-2}{2}, \quad T_{\square\square} = \frac{N+2}{2} \ . \tag{5.37}$$

It is common to absorb the choice of normalization for the generators of G in the defining representation by dividing all the Dynkin labels by the one of the defining representation.

Equation Equation (5.35) implies that the complexified gauge coupling τ runs with the energy scale as

$$\tau(\mu) = \tau_{\mathrm{UV}} - \frac{b}{2\pi i}\log\frac{\mu}{\Lambda_{\mathrm{UV}}} + \cdots = -\frac{b}{2\pi i}\ln\frac{\mu}{\Lambda} + \dots \ , \tag{5.38}$$

where τ_{UV} is the coupling at the energy scale Λ_{UV}, where the dots stand for non-perturbative contributions, and where $\Lambda = \Lambda_{\mathrm{UV}}^b e^{2\pi i \tau_{\mathrm{UV}}}$ is the (holomorphic) dynamical scale of the theory, which does not renormalize.

The way in which the complexified gauge coupling enters the supersymmetric gauge Lagrangians densities does not correspond to the usual normalization of gauge theories. Instead of considering the normalization of Equation (5.24) one should consider

$$\left(\frac{1}{4g_c^2} - i\frac{\Theta}{32\pi^2}\right)\int \mathrm{d}^2\theta\, \mathrm{Tr}\, W^\alpha(g_c V) W_\alpha(g_c V) + \mathrm{c.c.} \ , \tag{5.39}$$

where the subscript c is here to emphasize that this corresponds to the canonical normalization. The innocent looking rescaling $V \to g_c V$ that seems to be necessary to go from the so-called holomorphic normalization of Equation (5.24) to the one of Equation (5.39) is actually not that innocent, because of a rescaling anomaly. Hence one must distinguish between the *holomorphic gauge coupling* which renormalizes at one-loop only in perturbation theory as in Equation (5.35) and the *physical gauge coupling*, which renormalizes at all orders in perturbation theory: the running with the energy scale of the physical gauge coupling is given by the *NSVZ beta function* [NSVZ86]:

$$\beta_{\frac{8\pi^2}{g_c^2}} = \frac{3T(\mathrm{Adj}) - \sum_i T(R_i)(1-\gamma_i)}{1 - g_c^2 T(\mathrm{Adj})/8\pi^2} \ . \tag{5.40}$$

Fixed points of the renormalization group flow are scale-invariant theories. In most cases it implies in fact that the super Poincaré algebra of infinitesimal symmetries of the theory extends to the superconformal algebra, in which case the theory is said to be *superconformal*. The bosonic part of the

superconformal algebra is $\mathfrak{so}(2,4) \times \mathfrak{u}(1)_R$. Contrarily to the super Poincaré case the $\mathrm{U}(1)_R$ symmetry now fully enters the superconformal algebra, and this implies in particular that the dimension of any pure chiral primary operator $\mathcal{O}$ depends on its R-charge as:

$$\dim \mathcal{O} = \frac{3}{2} R[\mathcal{O}] = 1 + \frac{1}{2}\gamma_{\mathcal{O}} \ . \tag{5.41}$$

The relations in the algebra involving R-symmetry imply that it commutes with bosonic global symmetries and hence operators related by such a bosonic global symmetry have the same anomalous dimension.

Anomalies and 't Hooft anomaly matching.

A global symmetry $\tilde{G}$ is said to be *anomalous* if it is broken by quantum effects. Let us assume that the classical conserved current j_A^μ associated with $\tilde{G}$ is chiral, so that $\tilde{G}$ is a classical chiral global symmetry. A chiral global symmetry in a classical field theory is anomalous when the field integration measure in the path integral description of the quantum theory is not invariant under the symmetry [Fuj84]. For a general discussion on anomalies we refer to [Har05].

Only U(1) factors in $\tilde{G}$ can be anomalous, and they get contributions at one loop only from triangle (in four dimensions) Feynman diagrams. Let us consider a theory with gauge group G with generators t_A, a classical global symmetry group $\tilde{G}$ with generators $\tilde{t}_A$ and a set of Weyl fermions ψ_i transforming in the representations $(R_i, \tilde{R}_i)$ of $G \times \tilde{G}$. The Adler–Bell–Jackiw (ABJ) [Adl69, BJ69] calculation yields:

$$\partial_\mu j_A^\mu = \frac{A}{16\pi^2} F_{B\mu\nu} \tilde{F}^{B\mu\nu} \ , \tag{5.42}$$

where $A = \sum_i \tilde{q}_i T(R_i)$ is the anomaly coefficient, $\tilde{q}_i$ is the charge of ψ_i under the global U(1) factors, $T(R_i)$ is the Dynkin index of the representation R_i, $F_B^{\mu\nu}$ is the B-component of the G field strength and $\tilde{F}_{\mu\nu}$ is the magnetic (Hodge) dual of $F_{\mu\nu}$.

Instantons are classical solutions of the Euclidean Yang–Mills action which approach pure gauge at infinity. They satisfy

$$\frac{\theta}{32\pi^2} \int \mathrm{d}^4x \, \mathrm{Tr}\, F_{\mu\nu} \tilde{F}^{\mu\nu} = n\theta \tag{5.43}$$

where $n \in \mathbb{Z}$ is the instanton number. They are intrinsically strong-coupling effects since they vanish for $\Lambda \to 0$, where Λ is the strong coupling scale of the theory. The global U(1) anomaly of Equation (5.42) can be re-expressed as:

$$\Delta Q = 2An \tag{5.44}$$

after integration of Equation (5.42) in space-time and with ΔQ the amount of charge violation due to the anomaly: anomalous symmetries are violated by a specific amount $2An$ is a given instanton background with instanton number n.

Global U(1) symmetry groups can be anomalous without harming the consistence of the gauge theory, contrarily to the gauge symmetries for which an anomaly violates the unitarity of the theory. ABJ anomalies can thus be interpreted as obstructions to gauge a given U(1) global symmetry. The condition from the ABJ formula that an R-symmetry is free of ABJ anomalies amounts precisely to the vanishing of the NSVZ beta function of Equation (5.40).

By definition, 't Hooft anomalies [tH80] are computed from triangle Feynman diagrams with only global currents at the vertices (while ABJ anomalies have one global and two local current insertions). By weakly gauging a global symmetry and adding spectator fermions so as to cancel the gauge anomaly, one proves easily that 't Hooft anomalies are renormalization group independent quantities: this is the *'t Hooft anomaly matching condition*. It imposes stringent constraints on the low-energy dynamics of supersymmetric gauge theories, as we will see.

Four dimensional superconformal algebras have central charges a and c, and they can be computed as 't Hooft anomalies:

$$a = \frac{3}{32}\left(3\,\mathrm{Tr}\, R^3 - \mathrm{Tr}\, R\right), \qquad c = \frac{1}{32}(9\,\mathrm{Tr}\, R^3 - 5\,\mathrm{Tr}\, R) \ . \tag{5.45}$$

Note that if $\mathrm{U}(1)_R$ is a valid R-symmetry in the supersymmetric (not necessarily superconformal) sense and if U(1) is any other bosonic global non-R symmetry of the theory then $\mathrm{U}(1)_R \times \mathrm{U}(1)$ is also valid as an R-symmetry. Among all the possible R-symmetries obtained in this way, in a superconformal

theory only one of them is the superconformal R-symmetry which enters the infinitesimal symmetry algebra. It was proved in [IW03] that the superconformal R-symmetry is singled out by the fact that it maximizes the central charge a.

5.3 Perturbative spontaneous supersymmetry breaking

Supersymmetry is not a manifest symmetry in our universe, at least at energies lower that a few TeV, hence if it exists at all it must be a hidden symmetry. In order to implement the breaking of a supersymmetric UV theory, either the UV lagrangian of the theory contains relevant supersymmetry breaking terms (these are called soft supersymmetry breaking terms) or supersymmetry is spontaneously broken by the choice of vacuum. The low-energy effective Lagrangians of theories at non-supersymmetric vacua typically contain soft-supersymmetry breaking terms. Let us quickly analyze how supersymmetry may break spontaneously.

In a $d = 4$, $\mathcal{N} = 1$ gauge theory with canonical Kähler potential and Lagrangian given by Equation (5.25), the scalar potential reads:

$$V(\phi, \overline{\phi}) = \overline{F}F + \frac{1}{2}D^2 \ . \tag{5.46}$$

By definition a vacuum of the theory satisfies:

$$\forall i \ , \qquad \frac{\partial V}{\partial \phi^i} = 0 \quad \text{and} \quad \frac{\partial V}{\partial \overline{\phi}_i} = 0 \ , \tag{5.47}$$

and it is supersymmetric if and only if it has zero energy, that is:

$$F^i(\overline{\phi}) = 0 \quad \text{and} \quad D^a(\overline{\phi}, \phi) = 0 \ . \tag{5.48}$$

One can show with supersymmetric Ward identities that at a supersymmetry-breaking vacuum there must be a massless fermion, called the goldstino. This is the analogue of the Goldstone theorem for ordinary bosonic symmetries. In a Lagrangian theory defined by Equation (5.25) it can be expressed as

$$\psi^G_\alpha \sim \langle F^i \rangle \psi^i_\alpha + \langle D^a \rangle \lambda^a_\alpha \ . \tag{5.49}$$

F-term breaking

A first way supersymmetry may break spontaneously is if there does not exist any local minimum of the scalar potential such that all F-terms vanish. Since the superpotential terms in a supersymmetric Lagrangian do not renormalize, if this happens in perturbation theory then it must happen at tree-level.

One of the simplest examples of a theory which exhibits F-term supersymmetry breaking is the so-called *Polonyi model*. It is a theory of a single chiral superfield Φ with Lagrangian density:

$$\mathcal{L} = \int \mathrm{d}\theta^2 \mathrm{d}\overline{\theta}^2 \overline{\Phi}\Phi + \int \mathrm{d}\theta^2 \lambda \Phi + \int \mathrm{d}\overline{\theta}^2 \overline{\lambda\Phi} \ . \tag{5.50}$$

The scalar potential in that case is constant equal to $|\lambda|^2$ and that sets the supersymmetry breaking scale to $|\lambda| = M_s^2$. This model admits an R-symmetry with $R(\Phi) = 2$ which is broken at any choice of vacuum such that $\langle \Phi \rangle \neq 0$. Allowing the Kähler potential to be non-canonical:

$$\mathcal{L} = \int \mathrm{d}\theta^2 \mathrm{d}\overline{\theta}^2 K(\overline{\Phi}, \Phi) + \int \mathrm{d}\theta^2 \lambda \Phi + \int \mathrm{d}\overline{\theta}^2 \overline{\lambda\Phi} \ , \tag{5.51}$$

allows a variety of different behaviors. The scalar potential is

$$V(\phi, \overline{\phi}) = \left(\frac{\partial^2 K}{\partial \phi \partial \overline{\phi}} \right)^{-1} |\lambda|^2 \ , \tag{5.52}$$

and depending on the analytical properties of K one can obtain very different situations, some of which are displayed in Figure 5.1: there can be a stable non-supersymmetric vacuum (i.e. a global minimum of the scalar potential at a strictly positive height), a metastable non-supersymmetric vacuum (i.e. a local

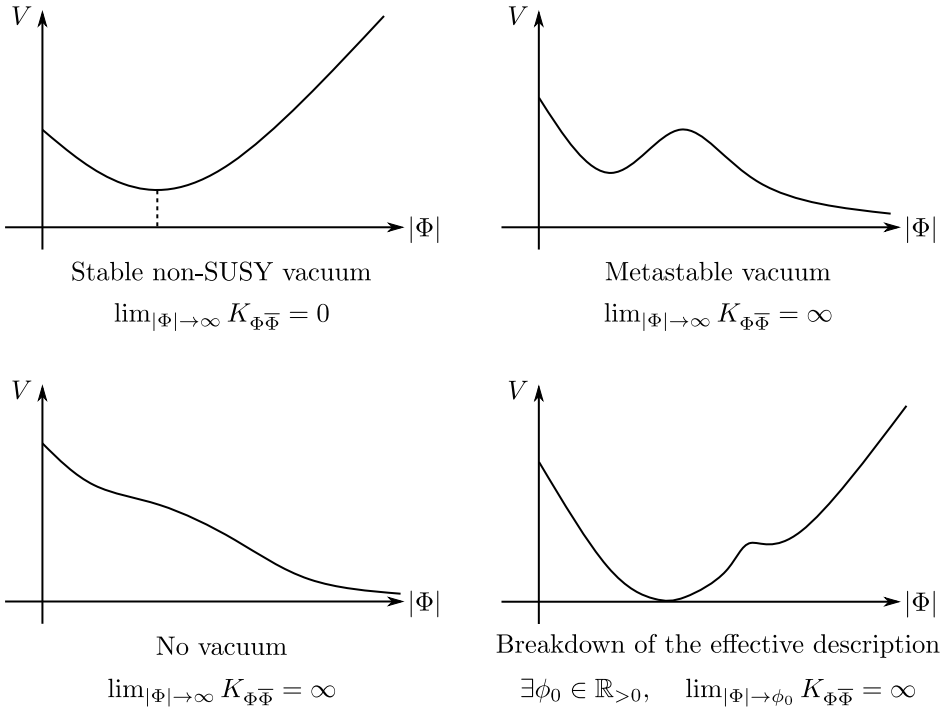


Figure 5.1: Different qualitative behaviors in the Polonyi model with non-canonical Kähler potential.

minimum of the scalar potential at a strictly positive height), no vacuum at all when the scalar potential slopes to zero at infinity (one speaks of a runaway behaviour). The fact that Kähler potential explodes at finite distance in the configuration space hints for the existence of new massless degrees of freedom there, i.e. a breakdown of the effective description.

Another classic family of models exhibiting F-term supersymmetry breaking consists of the so-called *O'Raifeartaigh models*. The simplest such theory describes three chiral superfields X, Φ_1 and Φ_2 with canonical Kähler potential and superpotential

$$W = \frac{1}{2} h X \Phi_1^2 + m \Phi_1 \Phi_2 - \mu^2 X \ . \tag{5.53}$$

This model has an R-symmetry with R-charges $R(X) = 2$, $R(\Phi_1) = 0$ and $R(\Phi_2) = 2$. The F-term equations are

$$\begin{aligned} \overline{F}_X &= \tfrac{1}{2} h \phi_1^2 - \mu^2, \\ \overline{F}_1 &= h x \phi_1 + m \phi_2, \\ \overline{F}_2 &= m \phi_1, \end{aligned} \tag{5.54}$$

where x is the bosonic field in the chiral multiplet X. The equations $\overline{F}_X = 0$ and $\overline{F}_2 = 0$ cannot be solved simultaneously, and hence supersymmetry is broken in any vacuum. If $|\mu| < |m|$ there is a family of vacua defined by

$$\phi_1 = \phi_2 = 0 \tag{5.55}$$

and with any value of x allowed. The energy of these vacua are $|\mu^2|^2$.

D-term breaking.

In a theory with chiral and vector superfields and no Fayet–Illiopoulos terms supersymmetry breaking is governed by the F-terms in the sense that if all F-terms are zero at some point in field space then all the D-terms can also be put to zero with gauge transformations. Hence the most interesting case is when there are Fayet–Illiopoulos terms. The *Fayet–Illiopoulos model* describes two massive chiral superfields Φ_+ and Φ_- with opposite charge with respect to a single U(1) gauge factor, and with Lagrangian density:

$$\frac{1}{32\pi} \mathrm{Im}\left(\tau \int \mathrm{d}^2\theta \, \mathrm{Tr}\, W^\alpha W_\alpha\right) + \int \mathrm{d}^2\theta \mathrm{d}^2\overline{\theta} \left(\xi V + \overline{\Phi}_+ e^{2eV} \Phi_+ + \overline{\Phi}_- e^{-2eV} \Phi_-\right) + m \int \mathrm{d}^2\theta \Phi_+ \Phi_- + c.c. \ . \tag{5.56}$$

The F-term and D-term equations are

$$\overline{F}_{\pm} = m\phi_{\mp} = 0, \tag{5.57}$$

$$D = -\frac{1}{2}\left[\xi + 2e(|\phi_+|^2 - |\phi_-|^2)\right] = 0. \tag{5.58}$$

Both equations cannot be satisfied if $\xi \neq 0$, and hence supersymmetry is broken.

Pseudo-moduli spaces and Coleman–Weinberg potential.

Both the Polonyi model with canonical Kähler potential and the O'Raifeartaigh model have a continuum of vacua at which the scalar potential has a strictly positive constant value. These are called *pseudo-moduli spaces of vacua*, and there is (at least) one massless scalar mode at each of these vacua which implement motion along these flat directions. This mode is called *pseudo-modulus*.

These classical non-supersymmetric flat directions in field space are in general lifted by quantum corrections, i.e. the pseudo-modulus gets a mass, while this is not the case for supersymmetric flat directions because of the non-renormalization theorems: since the superpotential in a supersymmetric theory is tree-level exact the F-term equations are not modified by loop corrections. As far as D-terms are concerned, Fayet–Illiopoulos terms do not renormalize if the theory does not suffer gravitational anomalies. When the gauge group of the theory does not have any U(1) factor one can shown from general grounds [Wit82, FNP$^+$81] that D-terms cannot lift a zero-energy state.

5.4 $\mathcal{N}=1$ quantum dynamics: SQCD and Seiberg duality.

Super Quantum ChromoDynamics (SQCD) will be our prototypical example of a one dimensional gauge theory with matter: it is an SU(N) gauge theory ($N \geq 3$) with F flavors, that is, F pairs of chiral fields $(Q_i, \tilde{Q}_i)$ which transform in $(\square, \overline{\square})$ of SU(N), and Lagrangian density:

$$\mathcal{L} = \frac{1}{32\pi}\mathrm{Im}\left(\tau \int \mathrm{d}^2\theta\, \mathrm{Tr}\, W^\alpha W_\alpha\right) + \int \mathrm{d}^2\theta \mathrm{d}^2\overline{\theta}\left(\overline{Q}_i e^{2V} Q^i + \overline{\tilde{Q}^i} e^{-2V} \tilde{Q}_i\right) , \tag{5.59}$$

The classical global symmetry of SQCD when $F \geq 1$ is $\mathrm{SU}(F)_L \times \mathrm{SU}(F)_R \times \mathrm{U}(1)_A \times \mathrm{U}(1)_B \times \mathrm{U}(1)_R$, where the charges of the chiral fields are:

	$\mathrm{SU}(F)_L$	$\mathrm{SU}(F)_R$	$\mathrm{U}(1)_A$	$\mathrm{U}(1)_B$	$\mathrm{U}(1)_R$
Q	$\overline{\square}$	$\mathbf{1}$	1	1	$\frac{F-N}{F}$
$\tilde{Q}$	$\mathbf{1}$	$\square$	1	-1	$\frac{F-N}{F}$

(5.60)

The ABJ anomaly coefficient corresponding to $\mathrm{U}(1)_A$ is $2F$, hence the $\mathrm{U}(1)_A$ factor is anomalous. The $\mathrm{U}(1)_B$ and $\mathrm{U}(1)_R$ factors however, are not. For example:

$$A_{\mathrm{U}(1)_R} = F\left(\frac{F-N}{F} - 1\right)\frac{1}{2} + F\left(\frac{F-N}{F} - 1\right)\frac{1}{2} + N = 0\ . \tag{5.61}$$

Note that when $F = 0$ the $\mathrm{U}(1)_R$ symmetry in fact is anomalous and only $\mathbb{Z}_{2N}$ survives at the quantum level, while $\mathrm{SU}(F)_L$, $\mathrm{SU}(F)_R$, $\mathrm{U}(1)_A$ and $\mathrm{U}(1)_B$ simply do not exist. It seems legitimate to study the case $F = 0$ on its own. In analogy with usual QCD the gauge fields will be often referred to as gluons and their fermionic partners as gluinos, while the chiral fields with be referred to as quarks.

Super Yang–Mills theory.

SQCD with $F = 0$ is simply SU(N) super Yang–Mills (SYM) theory. The only quantum global symmetry of the theory is a discrete $\mathbb{Z}_{2N}$ R-symmetry. The running of the holomorphic gauge coupling is given by

$$\tau = \frac{3N}{2\pi i}\log\frac{\Lambda}{\mu}, \tag{5.62}$$

where $\Lambda = \mu e^{2\pi i\tau/3N}$ is the holomorphic dynamical scale. SYM theory confines as the non-supersymmetric non-abelian pure Yang–Mills theories, and assuming that SYM has a mass gap the effective Lagrangian should only depend on τ, which implies that the effective superpotential must be of the form:

$$W_{\mathrm{eff}} = N\Lambda^3 \tag{5.63}$$

One can show that the gaugini condensate:

$$\langle\lambda\lambda\rangle = \Lambda^3 e^{i\frac{2k\pi}{N}} \tag{5.64}$$

with $k = 0, \ldots, N-1$, which breaks the R-symmetry spontaneously to $\mathbb{Z}_2$ and leaves N distinct isolated vacua labeled by the phase of the gaugino condensate.

Adding flavors.

Let us now assume that $F \geq 1$, for which SQCD has a quantum global symmetry group $\mathrm{SU}(F)_L \times \mathrm{SU}(F)_R \times \mathrm{U}(1)_B \times \mathrm{U}(1)_R$, where the charges of the chiral fields are:

$$\begin{array}{cccccc} & \mathrm{SU}(F)_L & \mathrm{SU}(F)_R & \mathrm{U}(1)_B & \mathrm{U}(1)_R \\ Q & \overline{\square} & \mathbf{1} & 1 & \frac{F-N}{F} \\ \tilde{Q} & \mathbf{1} & \square & -1 & \frac{F-N}{F} \end{array} \tag{5.65}$$

The chiral superfields Q^i_a and $\tilde{Q}^a_j$, where $a = 1, \ldots, N$ is a color index and $i, j = 1, \ldots, F$ are flavor indices, can be seen as $F \times N$ and $N \times F$ matrices, respectively. On the moduli space the D-term equations

$$D^A = \overline{Q}^b_i (T^A)^c_b Q^i_c - \tilde{Q}^b_i (T^A)^c_b \overline{\tilde{Q}}^i_c = 0 \ , \tag{5.66}$$

must be satisfied, with $A = 1, \ldots, N^2 - 1$ and $(T^A)^c_b := (T^A_{\mathbf{N}})^c_b = -(T^A_{\overline{\mathbf{N}}})^c_b$.

The classical moduli space is parameterized by single-trace gauge invariant operators. Many are obtained as the elements of the $F \times F$ meson matrix:

$$M^i_j = Q^i_a \tilde{Q}^a_j \ . \tag{5.67}$$

We need to distinguish two cases.

- Either $F < N$, in which case the meson matrix M is generically of maximal rank F, hence the moduli space has complex dimension F^2. At a generic point the gauge group is broken to $\mathrm{SU}(N-F)$, which is consistent with the counting of degrees of freedom:

$$2FN - \{N^2 - 1 - [(N-F)^2 - 1]\} = F^2 \tag{5.68}$$

  where FN is the total degrees of freedom of UV chiral fields, $N^2 - 1 - [(N-F)^2 - 1]$ is the number of chiral fields 'eaten' by massless vector superfields in the super Higgs mechanism, and the difference F^2 is the number of left-over degrees of freedom of massless chiral superfields which parametrize the moduli space,

- Or $F \geq N$, in which case the meson matrix is generically of non-maximal rank N. Moreover there are additional single trace gauge invariant operators: baryons. Writing $k = F - N$:

$$B_{i_1 \ldots i_k} = \epsilon_{i_1 \ldots i_k j_1 \ldots j_N} \epsilon^{a_1 \ldots a_N} Q^{j_1}_{a_1} \ldots Q^{j_N}_{a_N} \ , \tag{5.69}$$

$$\tilde{B}^{i_1 \ldots i_k} = \epsilon^{i_1 \ldots i_k j_1 \ldots j_N} \epsilon_{a_1 \ldots a_N} \tilde{Q}^{a_1}_{j_1} \ldots \tilde{Q}^{a_N}_{j_N} \ . \tag{5.70}$$

  There are classical relations between the mesons and the baryons. Since $F \geq N$ at a generic point of the moduli space the gauge group G is fully broken, and hence one expects the dimension of the moduli space to be $2NF - (N^2 - 1)$.

Now we wish to understand the quantum dynamics of SCQD and, in particular, whether an effective superpotential is generated by dynamical effects. This was done in [Sei94, Sei95].

The case F < N: the ADS superpotential.

An effective superpotential called the Affleck–Dine–Seiberg (ADS) superpotential is generated:

$$W_{\mathrm{ADS}} = (N-F) \left(\frac{\Lambda^{3N-F}}{\det M} \right)^{\frac{1}{N-F}} \ . \tag{5.71}$$

It induces a runaway behaviour: the scalar potential has minimum energy at infinity in field space, and the theory does not admit a vacuum a finite vacuum expectation values for the chiral fields.

The case F = N.

The classical moduli space is parameterized by the elements of the meson matrix as well as two baryons B and $\tilde{B}$. There is a classical relation between the gauge invariant operators:

$$\det M - B\tilde{B} = 0 \ . \tag{5.72}$$

One can show that this relation is deformed by quantum effects to

$$\det M - B\tilde{B} = \Lambda^{2N} \ , \tag{5.73}$$

and the quantum moduli space defined by this last equation is smooth. The gauge group is fully broken everywhere contrarily to what one naively expects from the classical analysis. The global chiral symmetry is also broken, and how much as well as how it breaks depends on where one sits on the quantum moduli space. On the mesonic branch defined by $B\tilde{B} = 0$ and $\det M = \Lambda^{2N}$ the global $\mathrm{SU}(F)_L \times \mathrm{SU}(F)_R$ is broken to its diagonal subgroup, whereas on the baryonic branch defined by $M = 0$ and $B\tilde{B} = -\Lambda^{2N}$ it is instead the global $\mathrm{U}(1)_B$ which breaks spontaneously.

The case F = N + 1.

When $F = N+1$ the moduli space is classically exact: there is no quantum correction to the Kähler metric whatsoever. In particular, the moduli space is still singular at its origin. However the interpretation of the singularity differs: classically it translates the presence of massless gluons and quarks since the gauge group is unbroken, however quantum mechanically this description cannot hold since the gauge theory is strongly coupled near the origin of the moduli space and confines. The new massless degrees of freedom which appear at the origin of the moduli space are more likely new massless mesons and baryons which are there because of the fact that the classical relations between mesons and baryons are trivially satisfied at the origin.

Moreover, the chiral symmetry is unbroken at the origin of the moduli space and hence SQCD with $F = N + 1$ exhibits s-confinement there, i.e. confinement without chiral symmetry breaking.

The case F > N + 1. Seiberg duality.

When $F \geq N + 2$ the moduli space of SQCD is also classically exact and hence contains singularities. However, as in the $F = N + 1$ case the interpretation of the singularities is different than in the quantum theory. Away from the origin the gauge group is higgsed: at a point where the meson matrix has rank k and where the baryons have zero expectations values, $\mathrm{SU}(N)$ with F flavors breaks spontaneously to $\mathrm{SU}(N - k)$ with $F - k$ flavors. Hence everything boils down to understanding what happens near the origin of the moduli space: unlike when $F = N+1$ the singularity at the origin cannot correspond to new massless mesons and baryons anymore – for example, 't Hooft anomalies conditions cannot be satisfied with such an assumption.

Seiberg's proposal in [Sei95] is that $\mathrm{SU}(N)$ SCQD with $F > N+1$ is in a non-abelian Coulomb phase near the origin of the moduli space, so that there are massless quarks and gluon fields in the low energy effective field theory. This is clearly the case when $F \geq 3N$ since the effective low-energy theory in that case is a free theory of massless gluons and quarks (for SQCD in that regime is IR free).

When $3N/2 < F < 3N$ one can argue that the low-energy effective field theory at the origin of the moduli space is an interacting SCFT which has two dual descriptions: the original one as an $\mathrm{SU}(N)$ gauge theory with F flavors, and another as an $\mathrm{SU}(F - N)$ gauge theory with F flavors and additional gauge invariant massless fields. We will use mSQCD (m stands for magnetic) to denote for the latter theory. More precisely, it is an $\mathrm{SU}(F - N)$ gauge theory with F flavors q_i and $\tilde{q}^j$ (where $i, j = 1, \dots F$) transforming in the fundamental and anti-fundamental representation of $\mathrm{SU}(F - N)$ respectively, as well as gauge singlets Φ^i_j where again $i, j = 1, \dots F$. There is also a superpotential:

$$W = h q_i \Phi^i_j \tilde{q}^j \ , \qquad h \in \mathbb{C}. \tag{5.74}$$

The fact that 't Hooft anomaly matching conditions hold is a strong argument in favor of this duality besides the fact that the two theories has the same global symmetry group and the same number of infrared degrees of freedom. Seiberg duality is involutive: mmSQCD = SQCD. Moreover if Λ_{el} is the dynamical scale of SQCD and if Λ_{m} is the dynamical scale of mSQCD then

$$\Lambda_{\mathrm{el}}^{3N-F} \Lambda_{\mathrm{m}}^{3(F-N)-F} = (-1)^{F-N} \mu^F \ , \tag{5.75}$$

where μ is an arbitrary matching scale, needed to identify the IR SQCD and mSQCD degrees of freedom. This latter relation imply that the more strongly coupled SQCD, the more weakly coupled mSQCD: Seiberg duality resembles electric-magnetic duality (hence the notation). The quarks and gluons in one description can be thought of as solitons of the quarks and gluons in the other theory. This description of the physics near the origin of the moduli space of SQCD in terms of the dual mSQCD theory actually holds for every $F \geq N + 2$.

- When $N + 1 < F < 3N/2$, the low-energy effective field theory of SQCD near the origin of the moduli space is better understood in mSQCD, since it is IR free in this regime: the massless low-energy fields are the dual gluons of $\mathrm{SU}(F - N)$ as well as the massless fields q_i, $\tilde{q}^j$ and Φ^i_j: this is the magnetic free phase. In that case one can choose $\mu = \Lambda_{\mathrm{el}} = \Lambda_{\mathrm{m}}$, mSQCD is naturally though of as an effective theory defined up to the UV scale Λ_{m} and SCQD as a UV completion of mSQCD.
- The range $3N/2 < F < 3N$ is called the conformal window of SQCD for the reason given above. The map $N \to N' = F - N$ preserves the conformal window.
- When $F \geq 3N$, SCQD is infrared-free: this is the electric free phase. Correspondingly, mSQCD is in a confining phase. In that case one can again choose $\mu = \Lambda_{\mathrm{el}} = \Lambda_{\mathrm{m}}$, SQCD is naturally thought of as an effective theory defined up to the UV scale Λ_{el} and mSCQD as a UV completion of SQCD.

The analysis of the quantum dynamics of SQCD with N fixed as a function of F is schematically depicted in Figure 5.2.

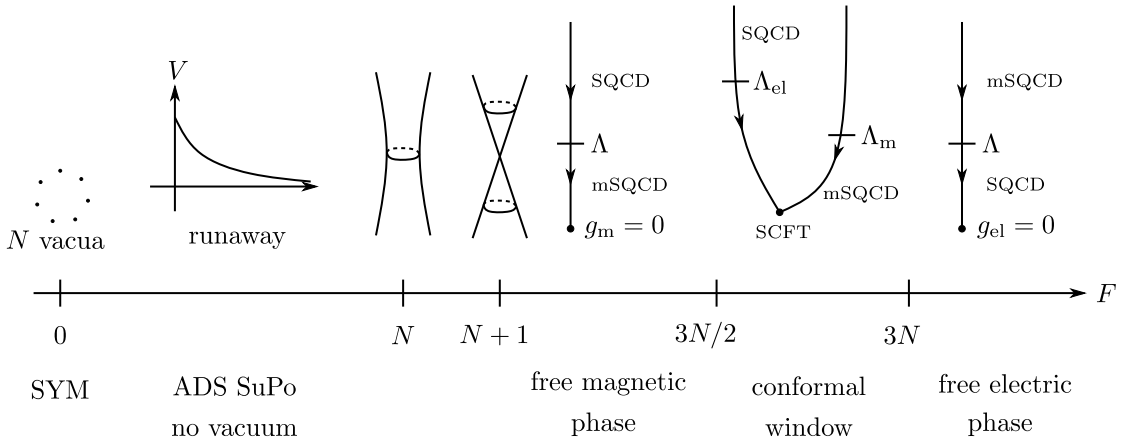


Figure 5.2: The phases of SQCD as a function of F.

5.5 Dynamical supersymmetry breaking

The supersymmetric non-renormalization theorems imply that if supersymmetry is unbroken at tree level it can only be broken by non-perturbative effects. Tree-level supersymmetry breaking occurs at scale M_S of the same magnitude as the natural mass scale of the theory. If M_S scale is much lower than that in our universe we need a theory in which supersymmetry is broken not at tree level but by very small quantum effects, as argued in [Wit81]. The dynamics of gauge theories provide a good way to obtain tiny quantum corrections thanks to dimensional transmutation: the renormalization group flow induces an exponential hierarchy of scales:

$$\Lambda \sim e^{-\frac{\#}{g^2(M_X)}} \ll M_X \ , \tag{5.76}$$

where M_X is a scale at which the theory is weakly coupled.

This analysis calls for a gauge sector which breaks supersymmetry dynamically, dubbed the *hidden sector*. Supersymmetry breaking would be transmitted to any of the supersymmetric extension of the Standard Model via gravitational interactions, gauge interactions or gravitational anomalies.

Much information concerning the phenomenological implementation of supersymmetry breaking in supersymmetric extensions of the Standard Model can be found in [Lut05].

Consider a theory defined by an action S_{μ_0} at some scale μ_0 and let $0 < \mu < \mu_0$ be another energy scale. The Wilsonian action S_μ at μ is defined by integrating out all fluctuations whose momentum is greater than μ:

$$e^{iS} = \int_{\phi(p),|p|>\mu} [\mathrm{D}\phi] e^{iS_0} \ . \tag{5.77}$$

It is the action which describes the physics at the scale μ by its classical couplings.

In a supersymmetric theory any parameter in the Lagrangian can be though of as the VEV of a so-called spurious superfield, and in particular the couplings appearing in F-terms. Hence F-terms are not only holomorphic in the chiral superfields but also in the couplings. Subsequently the (F-terms of the) Wilsonian action at the scale μ is also holomorphic in the coupling constants: the couplings appearing in S_μ are holomorphic quantities of the coupling constants at μ_0. This provides severe constraints on effective actions of supersymmetric quantum field theories.

We will focus on dynamical supersymmetry breaking in stable vacua i.e. absolute minima of the scalar potential. Methods to study dynamical supersymmetry breaking as well as many examples are given in [ADS85].

It is a good idea to consider theories without flat directions because the latter lead in general to runaways, through a dynamically generated superpotential which slopes to zero at infinity. The interplay between the perturbative and non-perturbative contributions to the scalar potential tends to produce non-supersymmetric stable vacua, i.e. dynamical supersymmetry breaking at finite distance in configuration space. This is displayed schematically in Figure 5.3.

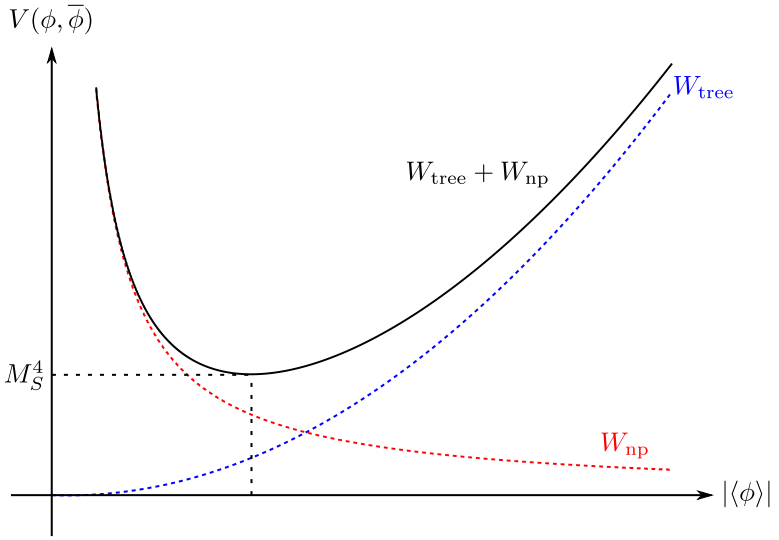


Figure 5.3: Interplay between perturbative and non-perturbative contributions to the scalar potential.

The energy of the scalar potential at the stable supersymmetry breaking vacuum is related to the supersymmetry breaking scale M_S as in Figure 5.3.

After having proven the existence of a stable supersymmetry breaking vacuum one would like to study the low-energy effective field theory around it. This might be difficult depending on the location of this vacuum in configuration space and on how the minimal value of the scalar potential compares with the dynamical scale of the theory (assuming there is a single one). The further away from the origin of configuration space the DSB vacuum, the more weakly coupled the effective theory is. Moreover, if the energy of the scalar potential at the vacuum is smaller than the dynamical scale of the original theory gauge degrees of freedom are integrated out in the effective theory and the latter is a theory of chiral superfields only, which resembles very much O' Raifeartaigh-like models with non canonical Kähler potentials. One can distinguish three different scenarios:

- it may be impossible to compute both the effective superpotential and the Kähler potential, in which case one cannot say much more than the mere fact that supersymmetry is dynamically broken,
- it may be that one can compute the effective superpotential but not the Kähler potential, in which case it is possible to determine the low-energy degrees of freedom but not their dynamics; one

speaks of non-calculable models,

- or one can compute both the effective superpotential as well as the Kähler potential, in which case one has access to lots of properties of the low energy effective field theory. One speaks of calculable models.

5.5.1 Supersymmetry-breaking criteria

It is of interest to determine criteria that show in some situations that supersymmetry is broken without a direct analysis of D-term and F-term equations.

Proposition 5.3. *In a four-dimensional $\mathcal{N}=1$ supersymmetric quantum field theory with a spontaneously broken global symmetry and without any non-compact classical flat direction, supersymmetry is generally broken.*

Proof. The Goldstone boson associated with the breaking of the global symmetry is a massless real boson without potential. Were supersymmetry unbroken it would have a partner real boson such that the two together would form the complex scalar of a chiral supermultiplet. Being in the same multiplet it would also be massless and without potential. In general the partner of a Goldstone boson corresponds to a non-compact classical flat direction and that contradicts the fact that there is no non-compact classical flat direction. □

Proposition 5.4. *In a four-dimensional $\mathcal{N}=1$ supersymmetric quantum field theory containing chiral superfields only and a generic superpotential W there are supersymmetric vacua.*

Proof. There are as many chiral superfields as conditions defining supersymmetric vacua coming from F-terms, and hence if W is generic there exist solutions to the F-term equations. □

In general if the superpotential preserves a global symmetry one can show that it does not change the conclusion of the previous proposition, since the number of independent variables is diminished by the same amount as the number of independent F-term equations.

Proposition 5.5 (Nelson–Seiberg [NS94])**.** *In a 4d $\mathcal{N}=1$ supersymmetric quantum field theory with generic superpotential W and such that the low-energy effective theory is a supersymmetric theory of chiral superfields only, the existence of an R-symmetry is a necessary condition for supersymmetry breaking. The fact that the R-symmetry is spontaneously broken in any vacuum is a sufficient condition for supersymmetry breaking.*

Proof. Let $\Phi^1,\dots,\Phi^F$ denote the low-energy chiral superfields and for all $i=1,\dots,F$ let $R(\Phi^i)=r_i$. There is at least one field – say Φ_1 – charged under the R-symmetry: $r_1\neq 0$. One can write:

$$W=(\Phi^1)^{\frac{2}{r_1}}f(X^i)\qquad \text{with } X^i=\Phi^i(\Phi^1)^{\frac{-r_i}{r_1}}\quad \forall\, i=2,\dots,F \tag{5.78}$$

The F-term equations read:

$$\frac{\partial W}{\partial\phi^j}=(\phi^1)^{\frac{2-r_j}{r_1}}\frac{\partial f}{\partial X^j}=0\qquad \text{when } j\neq 1 \tag{5.79}$$

$$\frac{\partial W}{\partial\phi^1}=\frac{2}{r_1}(\phi^1)^{\frac{2-r_1}{r_1}}f(X^i)+(\phi^1)^{\frac{2}{r_1}}\frac{\partial f}{\partial X^j}\frac{\partial X^j}{\partial\phi^1}=0\ . \tag{5.80}$$

If ϕ_1 is finite and non-zero the first set of equations imply $\partial f/\partial X^j=0$ and hence the $\overline{F}_1$ term becomes

$$\frac{\partial W}{\partial\phi^1}=\frac{2}{r_1}(\phi^1)^{\frac{2-r_1}{r_1}}f(X^i)=0\ . \tag{5.81}$$

In terms of the new function f there are now only $F-1$ independent variables $X^2,\dots,X^F$ but still F equations which are independent generically. Hence this system does not admit solutions and supersymmetry is broken. Because of the remark above the proposition, the existence of an R-symmetry in this class of theories is indeed a necessary condition for supersymmetry breaking. Moreover R-symmetry is spontaneously broken if and only if ϕ_1 is finite and non-zero, which implies supersymmetry breaking. □

Let $\Phi = (\phi, \psi, F)$ be a chiral superfield in a supersymmetric gauge theory such that Φ does not appear in the superpotential but where there is however not any non-compact direction in the moduli space. The so-called *Konishi anomaly* implies in that case that

$$\{Q, \psi\phi\} \sim \lambda_A \lambda^A \ , \tag{5.82}$$

where λ is the gaugino in the vector supermultiplet of the gauge group under which Φ is charged. Hence the vacuum energy is proportional to $\langle \mathrm{Tr}\, \lambda\lambda \rangle$ and:

Proposition 5.6. *If the gaugino condensate $\langle \mathrm{Tr}\, \lambda\lambda \rangle$ forms then supersymmetry is broken.*

This criterion is a special case of the Nelson–Seiberg one when theories which have an R-charge, since the gaugino condensate necessarily breaks the latter.

The *Witten index* I_W is a topological index which counts the difference between the number of bosonic modes and the one of fermionic modes of zero energy in a supersymmetric theory. The Witten index does not change as one varies the parameters of the theory and therefore in can be computed in a convenient corner of the parameter space, for example at weak coupling. More precisely, the Witten index does not change under a variation of the parameters which do not modify the asymptotic behaviour of the potential: if a non-zero parameter is varied to another non-zero value one does not expect the Witten index to change. It might however if a formerly zero parameter is turned on, or if a non-zero parameter is set to vanish.

Proposition 5.7. *When the Witten index is non zero then there exists a zero-energy state in the theory, and hence supersymmetry is unbroken.*

5.5.2 Two DSB Models

The SU(5) model.

The SU(5) model is the supersymmetric gauge theory with gauge group $G = \mathrm{SU}(5)$, a chiral superfield T in the antisymmetric **10** representation of G and a chiral superfield $\tilde{Q}$ in the antifundamental $\overline{\mathbf{5}}$ representation of G [ADS84]. It is asymptotically free (the one-loop beta function coefficient is $b_1 = 13$).

One cannot construct gauge invariants out of T and $\tilde{Q}$ and hence the theory does not have any flat direction. For the same reason there cannot be any superpotential. Hence the classical moduli space of the theory is merely the origin of the configuration space which is a supersymmetric vacuum, where G is unbroken.

The theory has a non-anomalous global symmetry $\mathrm{U}(1) \times \mathrm{U}(1)_R$, where the chiral superfields have charges

$$\begin{array}{ccc} & \mathrm{U}(1) & \mathrm{U}(1)_R \\ T & -1 & 1 \\ \tilde{Q} & 3 & -9 \end{array} \tag{5.83}$$

fixed by the cancellation of the ABJ anomalies. Assuming that the theory confines at the origin of field space, the massless gauge invariant fields X_i with charges that appear as low-energy degrees of freedom must reproduce the 't Hooft anomalies for $\mathrm{U}(1)^3$, $\mathrm{U}(1)^2\mathrm{U}(1)_R$, $\mathrm{U}(1)\mathrm{U}(1)_R^2$ and $\mathrm{U}(1)_R^3$ in the UV theory. It was shown in [ADS84] that one needs at least five low energy fields X_i to satisfy 't Hooft anomaly matching conditions with complicated charges, and hence the breaking of $\mathrm{U}(1) \times \mathrm{U}(1)_R$ seems very plausible. If the global symmetry group indeed breaks spontaneously in the vacuum then supersymmetry is also broken as follows from Proposition 5.3.

There are other compelling arguments for the fact that this model breaks supersymmetry dynamically. One may prove that the gaugino condensate has a non-zero vacuum expectation value [MV84], which allows to conclude thanks to Proposition 5.6. One can also add fields charged under the SU(5) gauge group so that the new theory has flat directions on which more classical arguments can be used to prove that it breaks supersymmetry if these new fields have a mass [Mur95]. Sending this mass to infinity gives back the original theory and one can show that this does not change the Witten index, for example.

The SU(5) model admits a generalization introduced in [ADS85] as an $\mathrm{SU}(N)$ gauge theory with an antisymmetric chiral superfield A_{ij} and $(N-4)$ antifundamental chiral superfields F_i^f where $i = 1, \ldots, N$ and $f = 1, \ldots, N-4$. This model most likely breaks supersymmetry if $N \geq 7$ is odd.

The $3-2$ model.

The $3-2$ model is a theory with gauge group $\mathrm{SU}(3)\times\mathrm{SU}(2)$, chiral superfields $Q_{\alpha i}$ in $(\square_3,\square_2)$, $\overline{U}^i$ and $\overline{D}^i$ in $\overline{\square}_3$ and L_α in $\square_2$ where i (resp. α) is an index for SU(3) (respectively SU(2)), and tree-level superpotential

$$W_{\text{tree}} = \lambda Q\overline{D}L = \lambda\epsilon^{\alpha\beta}Q_{\alpha i}\overline{D}^i L_\beta \ . \tag{5.84}$$

Up to overall normalization there is a single global non-anomalous non-R charge Y, and a non-anomalous R-charge. Local and global charges for the chiral fields in the model can be summarized as follows:

$$\begin{array}{ccccc} & \mathrm{SU}(3) & \mathrm{SU}(2) & \mathrm{U}(1)_Y & \mathrm{U}(1)_R \\ Q_{\alpha i} & \mathbf{3} & \mathbf{2} & 1/3 & 1 \\ \overline{U}^i & \overline{\mathbf{3}} & - & -4/3 & -8 \\ \overline{D}^i & \overline{\mathbf{3}} & - & 2/3 & 4 \\ L & - & \mathbf{2} & -1 & -3 \end{array} \tag{5.85}$$

There are six single trace gauge invariant operators that can built from the four chiral superfields and the superpotential lifts all the corresponding D-flat equations.

In a regime where $\lambda \ll 1$ and $\Lambda_2 \ll \Lambda_3$ and at energy scales $\Lambda_2 \ll \mu \ll \Lambda_3$ the group SU(2) is weakly coupled while SU(3) is strongly coupled and we can consider SU(2) as a global symmetry. Then the SU(3) gauge theory is an SU(3) super quantum chromodynamics (SQCD) with $2 = F-1$ flavors. Hence a dynamical ADS superpotential is generated:

$$W_{\text{np}} = \frac{\Lambda_3^7}{Y} \ , \tag{5.86}$$

where $Y = \epsilon^{\alpha\beta}\left(Q_{\alpha i}\overline{U}^i Q_{\beta j}\overline{D}^j - Q_{\alpha i}\overline{D}^i Q_{\beta j}\overline{U}^j\right)$. The full quantum superpotential is

$$W_{\text{tree}} + W_{\text{np}} \tag{5.87}$$

and the minimum of the scalar potential must be at a non-zero vacuum expectation value for Y. Since Y has R-charge -2, R-symmetry is broken, which implies that supersymmetry is broken as follows from Proposition 5.3. In this regime the minimum is moreover in a weakly-coupled region, and one can compute the Kähler potential in the stable vacuum as well as the following estimate:

$$M_S^4 \sim \Lambda_3^4 \lambda^{\frac{10}{7}} \ . \tag{5.88}$$

In the regimes where $\Lambda_2 \gg \Lambda_3$ and $\Lambda_2 \sim \Lambda_3$ one can show that supersymmetry is still broken by dynamical effects, however the model is now uncalculable.

5.6 $\mathcal{N}=2$ quantum dynamics: Seiberg–Witten theory

There are two supermultiplets of the $\mathcal{N}=2$ super-Poincaré algebra in dimension 4 which describe excitations of spin smaller than one only:

- the $\mathcal{N}=2$ *vector multiplet*, which consists of an $\mathcal{N}=1$ vector multiplet and an $\mathcal{N}=1$ chiral multiplet with the same internal charges $V=(\lambda_\alpha,a_\mu,D)\oplus(\phi,\psi_\alpha,F)$. In particular (ϕ,ψ_α,F) is in the adjoint of the gauge group corresponding to (λ_α,a_μ,D).
- the $\mathcal{N}=2$ *hypermultiplet*, which consists of an $\mathcal{N}=1$ chiral multiplet and an $\mathcal{N}=1$ antichiral multiplet $H=Q\oplus\tilde{Q}=(q,\psi,F)\oplus(\overline{\tilde{q}},\overline{\tilde{\psi}},\overline{\tilde{F}})$, so that (q,ψ,F) and $(\overline{\tilde{q}},\overline{\tilde{\psi}},\overline{\tilde{F}})$ have of opposite charges under internal symmetries. When a hypermultiplet is an a real or pseudo-real representation of a gauge group one can impose a constraint linking (q,ψ,F) to $(\overline{\tilde{q}},\overline{\tilde{\psi}},\overline{\tilde{F}})$ which respects $\mathcal{N}=2$ supersymmetry. The result is called *half-hypermultiplet*.

We will assume that the R-symmetry group of our $\mathcal{N}=2$ supersymmetry algebras is always $\mathrm{SU}(2)_R$. The most general Lagrangian density respecting $\mathcal{N}=2$ supersymmetry and expressed in $\mathcal{N}=1$ language

is of the form:

$$\frac{1}{32\pi}\mathrm{Im}\left(\tau\int \mathrm{d}^2\theta\,\mathrm{Tr}\,W^\alpha W_\alpha\right)+\int \mathrm{d}^2\theta\mathrm{d}^2\overline{\theta}\,\mathrm{Tr}\,\overline{\Phi}e^{2gV}\Phi+\int \mathrm{d}^2\theta\mathrm{d}^2\overline{\theta}\left(\overline{Q}^i e^{2gV}Q_i+\tilde{Q}^i e^{-2gV}\overline{\tilde{Q}}_i\right)$$
$$+\int \mathrm{d}^2\theta\left(\sqrt{2}g\tilde{Q}^i\Phi Q+\mu_i\tilde{Q}^iQ_i\right)+\text{c.c.} \qquad (5.89)$$

The moduli space of vacua $\mathcal{M}$ of an $\mathcal{N}=2$ theory is the space of solutions to D-terms and F-terms, exactly as in the case of $\mathcal{N}=1$ theories. At a generic point in the moduli space both the scalars in the vector multiplets and the ones in the hypermultiplets can have non-zero vacuum expectation values. Interesting subvarieties of the moduli space can be defined:

- The *Coulomb branch* $\mathcal{M}_V$, which is a special Kähler manifold on which the scalar in the hypermultiplets are required to have a zero vacuum expectation value. It is described by

$$[\overline{\Phi},\Phi]=0\ . \qquad (5.90)$$

  The low-energy effective theory on the Coulomb branch is an abelian gauge theory generally, since Φ is in the adjoint of G and hence the latter can at most be higgsed to $\mathrm{U}(1)^{\mathrm{rg}(G)}$. The metric on the Coulomb branch is modified by quantum effects.

- The *Higgs branch* $\mathcal{M}_H$, which is a hyper-Kähler manifold defined only when the mass terms for the hypermultiplets are zero and on which the scalar in the vector multiplets are required to have a zero vacuum expectation value. It is described by

$$(Q_i\overline{Q}^i+\overline{\tilde{Q}}_i\tilde{Q}^i)_{\text{traceless}}=0\ , \qquad (5.91)$$
$$(Q_i\tilde{Q}^i)_{\text{traceless}}=0. \qquad (5.92)$$

  This Higgs branch is always classically exact.

Locally the full moduli space is a direct product $\mathcal{M}_V\times\mathcal{M}_H$. At a generic point the low energy effective action is completely determined by the knowledge of the Kähler potential on the Higgs branch (which can be determined classically) and by the data of the Kähler potential and the matrix of complexified gauge couplings for $\mathrm{U}(1)^{\mathrm{rg}(G)}$ on the Coulomb branch. Both the Kähler potential and this matrix of complexified couplings on the Coulomb branch are determined by a single holomorphic function $\mathcal{F}$ of the chiral superfields in the vector multiplet, dubbed prepotential. Hence solving the quantum dynamics of $\mathcal{N}=2$ theories amounts to computing the prepotential. More precisely, if a_i are the bosonic fields in the low energy $\mathcal{N}=2$ U(1) vector multiplets, the matrix of complexified gauge couplings can be expressed as:

$$\tau^{ij}=\frac{\partial^2\mathcal{F}}{\partial a_i\partial a_j} \qquad (5.93)$$

while for:

$$a_D^i=\frac{\partial\mathcal{F}}{\partial a_i}\ , \qquad (5.94)$$

the Kähler potential is:

$$K=i(\overline{a}_D^i a_i-\overline{a}_i a_D^i)\ . \qquad (5.95)$$

BPS states.

In a $4d$ $\mathcal{N}=2$ super-Poincaré algebra the odd generators satisfy the relations of Equations (5.9) to (5.11) that we reproduce here for convenience. Here $I,J=1,2$, and Z is a complex central charge:

$$\left[Q_{I\alpha},\overline{Q}^J_{\dot\beta}\right]=2\sigma^\mu_{\alpha\dot\beta}P_\mu\delta^{IJ}\ ,$$
$$[Q_{1\alpha},Q_{2\beta}]=-[Q_{2\alpha},Q_{1\beta}]=\epsilon_{\alpha\beta}Z\ ,$$
$$\left[\overline{Q}^1_{\dot\alpha},\overline{Q}^2_{\dot\beta}\right]=-\left[\overline{Q}^2_{\dot\alpha},\overline{Q}^1_{\dot\beta}\right]=\epsilon_{\dot\alpha\dot\beta}Z^*\ .$$

Any massive state $|\psi\rangle$ in the theory always satisfy the bound

$$M_{|\psi\rangle}\geq|Z_{|\psi\rangle}| \qquad (5.96)$$

called the *BPS bound* [2], and the states which saturate it are said to be *BPS states*. These are more rigid representation of the $\mathcal{N} = 2$ super Poincaré algebra, and hence there are robust under generic perturbations of the theory. When the low-energy theory is a weakly coupled $\mathrm{U}(1)^{\mathrm{rg}(G)}$ gauge theory the central charge Z is a linear combination of the electric charges n^i, the magnetic charges m_i and the flavor charges f^j where $i = 1, \dots, \mathrm{rg}(G)$ and $j = 1, \dots, F$ for some $F \geq 0$:

$$Z = a_i n^i + a_D^i m_i + \sum_{j=1}^{F} \mu_j f^j \ . \tag{5.97}$$

When the theory is weakly coupled the coefficients a_i, a_D^i and μ_j admit a natural interpretation in terms of what enters the Lagrangian in Equation (5.89), but at strong coupling the latter relation should instead be thought of as a definition of these coefficients.

Seiberg–Witten theory.

Seiberg and Witten have studied the quantum dynamics of $\mathcal{N} = 2$ SU(2) gauge theory with $F = 0, 1, 2, 3$ and 4 hypermultiplets in two seminal papers [SW94a, SW94b]. Let us study is some details the $F = 0$ case for concreteness – we follow the conventions of [Tac13].

In $\mathcal{N} = 2$ SU(2) gauge theories the Coulomb branch $\mathcal{M}_C$ of the moduli space is parameterized by vacuum expectation values for the adjoint chiral field Φ satisfying Equation (5.90), the solutions of which being $\phi = \mathrm{Diag}(a, -a)$ for $a \in \mathbb{C}$. At any $a \neq 0$ on the Coulomb branch the gauge group breaks spontaneously to $\mathrm{U}(1) \times \mathbb{Z}_2$ where $\mathbb{Z}_2$ exchanges a and $-a$, so that the Coulomb branch is parameterized by $u := a^2$. Classically, there is a singularity at $a = 0$ since the gauge group is unbroken at that point.

When $|u|$ is large the SU(2) gauge theory is still weakly coupled when the breaking to U(1) occurs, and hence the coupling of the low-energy theory (which does not run) can be obtained from the perturbative running of the SU(2) theory and depends on the choice of $a \in \mathcal{M}_C$ as:

$$\tau(a) = -\frac{8}{2\pi i} \log \frac{a}{\Lambda} + \dots \tag{5.98}$$

where Λ is the strong coupling scale of the theory. From Equations (5.93) and (5.94) and integrating this last equation once one obtains

$$a_D = -\frac{8a}{2\pi i} \log \frac{a}{\Lambda} + \dots \tag{5.99}$$

and hence the pair (a, a_D) has a monodromy $M_\infty : (a, a_D) \to (-a, -a_D + 4a)$ at infinity in $\mathcal{M}_C$. In this weakly coupled region of $\mathcal{M}_C$ defined by $|a| >> |\Lambda|$ the semi-classical analysis is enough to account for the quantum corrections to the classical moduli space.

In the strongly coupled region of $\mathcal{M}_C$ however, when $|a| \lesssim |\Lambda|$ one expects more drastic quantum corrections. One can argue that one needs in fact two singularities at $\pm\Lambda^2 \in \mathcal{M}_C$ instead of the unique classical one, with monodromies $M_{-\Lambda^2}$ and M_{Λ^2} for a and a_D prescribed (in particular) by the consistency condition $M_{-\Lambda^2} M_{\Lambda^2} = M_\infty$. A solution is $M_{-\Lambda^2} : (a, a_D) \to (-a - a_D, 4a - 3a_D)$ and $M_{\Lambda^2} : (a, a_D) \to (a - a_D, a_D)$.

One way to solve this Riemann-Hilbert problem is by considering a and a_D as period integrals on a surface Σ, defined by the equation

$$(\Sigma): \qquad \Lambda^2 z + \frac{\Lambda^2}{z} = x^2 - u \ . \tag{5.100}$$

It is a two-folds ramified cover over the sphere $\mathbb{P}^1(\mathbb{C})$ parameterized by z with four simple branch points, and it is endowed with a differential $\lambda_{\mathrm{SW}} = x \mathrm{d}z/z$. One chooses of symplectic basis (A, B) of $H^1(\Sigma, \mathbb{Z})$ and declares that

$$a = \frac{1}{2\pi i} \int_A \lambda_{\mathrm{SW}}, \qquad a_D = \frac{1}{2\pi i} \int_B \lambda_{\mathrm{SW}}. \tag{5.101}$$

When $u = -2\Lambda^2$ (resp. $u = 2\Lambda^2$) a cycle in Σ shrinks. This defines a singularity in the Coulomb branch, and the monodromy around it can be computed from the Picard–Lefschetz monodromy formula. The singularity is interpreted as a BPS excitation of charges given by the coefficients of the shrinking cycle in the basis (A, B) becoming massless.

[2] BPS stands for Bogomol'nyi–Prasad–Sommerfeld.

More generally any BPS state of (electric, magnetic) charges (n_e, n_m) in the low-energy effective field theory correspond to a cycle γ in Σ whose coefficients in the basis (A, B) are (n_e, n_m). The central charge of such a BPS state is easily computed as

$$Z_\gamma = \frac{1}{2\pi i} \int_\gamma \lambda_{\mathrm{SW}} \ , \tag{5.102}$$

from which one deduces its mass as follows from the BPS formula of Equation (5.96).

From the knowledge of the Seiberg–Witten curve one can even compute the instantonic non-perturbative corrections to the coupling constant, i.e. compute the coefficients c_k in

$$\tau(a) = -\frac{8}{2\pi i} \log \frac{a}{\Lambda} + \sum_{k=0}^{\infty} c_k \left(\frac{\Lambda}{a}\right)^{4k} \ . \tag{5.103}$$

These coefficients agree with the ones that are computed directly from an instanton calculation [Nek03].

In general in an $\mathcal{N} = 2$ gauge theory with gauge group G, the quantum a_i, a_D^i and μ_j appearing in the BPS formula of Equation (5.97) are completely encoded in the data of a Riemann surface Σ called the *Seiberg–Witten curve* which is a ramified covering of another Riemann surface C dubbed the *UV curve*: in an asymptotically free $\mathcal{N} = 2$ gauge theory with gauge group G and F flavors, the Seiberg–Witten curve Σ has genus $\mathrm{rk}(G)$ and F punctures and it is endowed with a meromorphic one-form λ_{SW}, the periods of which in a symplectic basis $\mathcal{B} = (A_1, \ldots, A_g, B_1, \ldots, B_g, C_1, \ldots, C_F)$ of $H_1(\Sigma, \mathbb{Z})$ are

$$a_i = \frac{1}{2\pi i} \int_{A_i} \lambda_{\mathrm{SW}}, \qquad a_D^i = \frac{1}{2\pi i} \int_{B_i} \lambda_{\mathrm{SW}}, \qquad \mu_j = \frac{1}{2\pi i} \int_{C_j} \lambda_{\mathrm{SW}} \ . \tag{5.104}$$

The equation defining Σ depends on a choice of point on $\mathcal{M}_C$. Generically Σ is smooth but there are special points on $\mathcal{M}_C$ where at least one cycle in Σ shrinks. These points are singularities of the quantum Kähler metric: some excitations (in fact, BPS states) in the low-energy effective theory become massless. In general, a homology class $\gamma \in H^1(\Sigma, \mathbb{Z})$ corresponds to a BPS state whose electric, magnetic and flavor charges are the coefficients of γ in $\mathcal{B}$, whose central charge is given by the integral of λ_{SW} along γ, and hence whose mass is:

$$M = \left| \frac{1}{2\pi i} \int_\gamma \lambda_{\mathrm{SW}} \right| \ . \tag{5.105}$$

5.7 $\mathcal{N} = 4$ super Yang–Mills

There is only one supermultiplet of the $\mathcal{N} = 4$ super Poincaré algebra (of which the R-charge is assumed to be $\mathrm{SU}(4)_R$) in four dimensions which contains excitations of spin smaller than one only, called the $\mathcal{N} = 4$ vector multiplet. In $\mathcal{N} = 1$ language it is:

$$V = (\lambda_\alpha, v_\mu, D) \oplus \bigoplus_{A=1}^{3} (\phi^A, \psi_\alpha^A, F^A) \ . \tag{5.106}$$

The most general $\mathcal{N} = 4$ invariant Lagrangian expressed in $\mathcal{N} = 1$ superspace is:

$$\frac{1}{32\pi} \mathrm{Im} \left(\tau \int \mathrm{d}^2\theta \, \mathrm{Tr} \, W^\alpha W_\alpha \right) + \int \mathrm{d}^2\theta \mathrm{d}^2\overline{\theta} \, \mathrm{Tr} \sum_A \overline{\Phi}_A e^{2gV} \Phi^A - \int \mathrm{d}^2\theta \sqrt{2} g \, \mathrm{Tr} \, \Phi_1 [\Phi_2, \Phi_3] + \mathrm{c.c.} \tag{5.107}$$

The supersymmetric non-renormalization theorems imply that the gauge coupling $\mathcal{N} = 4$ super Yang–Mills theory does not renormalize at all orders in perturbation theory, as well as when non-perturbative effects are taken into account. This hints to a superconformal invariance, which is indeed the case for $\mathcal{N} = 4$ super Yang–Mills: the superconformal group $\mathrm{SU}(2, 2|4)$ is an exact symmetry of the theory at the origin of the moduli space.

There is a remarkable duality of $\mathcal{N} = 4$ super Yang–Mills theories known as *Montonen–Olive duality* [MO77] under which the complexified gauge coupling τ is acted on by $\mathrm{SL}_2(\mathbb{Z})$ as:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \tau = \frac{a\tau + b}{c\tau + d} \ , \tag{5.108}$$

while dyonic particles in the low-energy effective theory at some point on the Coulomb branch of (electric, magnetic) charges $(n_e^i, n_i^m)_{i=1,\ldots,\mathrm{rk}(G)}$ are mapped to particles of charges $(an_e^i - bn_i^m, -cn_e^i + dn_i^m)_{i=1,\ldots,\mathrm{rk}(G)}$. More intrinsically, $\mathcal{N} = 4$ super Yang–Mills with gauge group G and complexified coupling τ is dual to $\mathcal{N} = 4$ super Yang–Mills with gauge group $\tilde{G}$ and complexified coupling on the right-hand side of Equation (5.108). Here $\tilde{G}$ is either G and its Langlands dual G^L, depending on the decomposition the corresponding element in $\mathrm{SL}_2(\mathbb{Z})$ in the generators S and T of $\mathrm{SL}_2(\mathbb{Z})$, where

$$S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}. \tag{5.109}$$

$\mathcal{N} = 4$ super Yang–Mills is the prototypical example of conformal field theories which appear in the AdS/CFT correspondence that we will review in Chapter 8.

* * * * * * *

We have studied various aspects of supersymmetry in four dimensions. Quantum field theories that enjoy supersymmetries are better-behaved than generic theories, as far as renormalization is concerned. This fact makes such theories ideal candidates to study strong-coupling dynamics as well as their space of vacua.

If supersymmetry breaking occurs perturbatively, then it does at tree level either through D-terms or F-terms. Dynamical supersymmetry breaking relies on non-perturbative effects and allows for an exponential separation between the supersymmetry breaking energy scale and the natural scale of the theory.

We have presented the vacuum structure of $\mathcal{N} = 1$ super Quantum Chromodynamics, and Seiberg duality, which links SQCD with N colors and F flavors (and $F > N + 1$) to 'SQCD' with $F - N$ colors, F flavors and a superpotential term. The dynamics of these SQCD theories will be our fundamental example of $\mathcal{N} = 1$ dynamics in four dimensions, and we will strongly rely on it when studying more complicated $\mathcal{N} = 1$ theories obtained in the worldvolume of branes at singularities in Chapters 7 and 8 and in Parts III and IV.

As soon as $\mathcal{N} > 1$, quantum field theories are necessarily non-chiral. One can study $\mathcal{N} = 2$ super Yang–Mills theories with the techniques developed by Seiberg and Witten which allow for a description of the low-energy dynamics on the Coulomb branch as well as of the BPS states. Besides, $\mathcal{N} = 4$ super Yang–Mills theories enjoy Montonen–Olive duality, which puts on firm grounds the mysterious electric-magnetic symmetry of Maxwell's equations in the vacuum.

Chapter 6

String theory

String theory is a set of theories and techniques aiming to describe gravity together with the three microscopic forces we are aware of at the quantum level and in a unifying framework. Its name comes from the fact that the elementary microscopic objects the theory describes are strings rather than point-like particles. In the perturbative approach to string theory, the latter contains only two fundamental objects: an open string, and a closed string (it is also possible to consider theories of a closed string only). If their typical size is tiny, one would need an astonishing energy to resolve the fact that they are not point-like objects, or equivalently, at energies lower than that these strings would just appear as point-like particles. However, it is crucial that strings can vibrate. The Fourier modes of the vibrations form a tower of states with constant energy gap. The energy of a mode appears macroscopically as a contribution to the mass of what seems to be a point particle, and hence a quantum theory of strings naturally yields a tower of states, which opens-up the possibility to explain the zoology of elementary particles we observe in our universe, in a natural way. Actually, string theory was originally studied as a candidate theory to describe the strong nuclear interaction, before being demoted in profit of quantum chromodynamics. However, it was realized in 1974 that every quantum relativistic string theory contains spin-2 excitations with the correct Ward identities to be gravitons, hence allowing string theories to be plausible descriptions of quantum gravity [Yon74]. String theory as formulated on the *world-sheet* is a two-dimensional quantum field theory whose fields are the coordinates of a map from a surface to a metric *target space* X.

One generally speaks of two *revolutions* that occurred in the history of string theory. The first was triggered in 1984 by a famous paper by Green and Schwarz [GS84] showing anomaly cancellation in some string theories with gauge groups SO(32) and $E_8 \times E_8$ in 10 dimensions, hence proving that there exist consistent, anomaly-free string theories. Supersymmetry is needed in these consistent string theories, which are therefore dubbed superstring theories. In order to cancel the so-called conformal anomalies of the worldsheet, the fields in the $2d$ quantum field theory must satisfy some conditions. The simplest way to satisfy them is to fix the dimension of the target space to ten, hence every superstring theory induces a 10-dimensional quantum field theory in the target space. The target space is usually interpreted as space-time, and we will use these two terms interchangeably. Five superstring theories were discovered and studied afterwards.

- Type I superstring theory has $\mathcal{N} = 1$ supersymmetry in 10 dimensions and describes oriented and unoriented open and closed strings, perturbatively. The endpoints of open strings are associated with an SO(32) gauge symmetry.
- Type IIA superstring theory has $\mathcal{N} = (1, 1)$ supersymmetry in 10 dimensions and describes oriented closed strings only, perturbatively. It is chiral on the world-sheet and non-chiral in target space.
- Type IIB superstring theory has $\mathcal{N} = (2, 0)$ supersymmetry in 10 dimensions and describes oriented closed strings only, perturbatively. It is non-chiral on the world-sheet and chiral in target space.
- Heterotic SO(32) theory (HO) has $\mathcal{N} = (1, 0)$ supersymmetry in 10 dimensions and describes oriented closed strings only, perturbatively. It is chiral and has an SO(32) gauge symmetry.
- Heterotic $E_8 \times E_8$ theory (HE) has $\mathcal{N} = (1, 0)$ supersymmetry in 10 dimensions and describes oriented closed strings only, perturbatively. It is chiral and has an $E_8 \times E_8$ gauge symmetry.

The second revolution happened in 1995 after years of working out many dualities between these five superstring theories. Witten suggested that the strong coupling limit of type IIA superstrings is an 11−dimensional theory which in fact does not describe any strings at all, bur membranes [Wit95], dubbed M-theory[1]. Because of various dualities, M-theory in fact appears as a unifying framework for all the string theories. At low energies M-theory is well approximated by 11-dimensional supergravity, yielding the famous picture in Figure 6.1: S stands for S-duality, T for T-duality. One can obtain type I string theory as an orientifold of type IIB as we will review later, and 11-dimensional supergravity compactified on S^1 (resp. $S^1/\mathbb{Z}_2$) yields the low-energy limit of type IIA (resp. heterotic $E_8 \times E_8$).

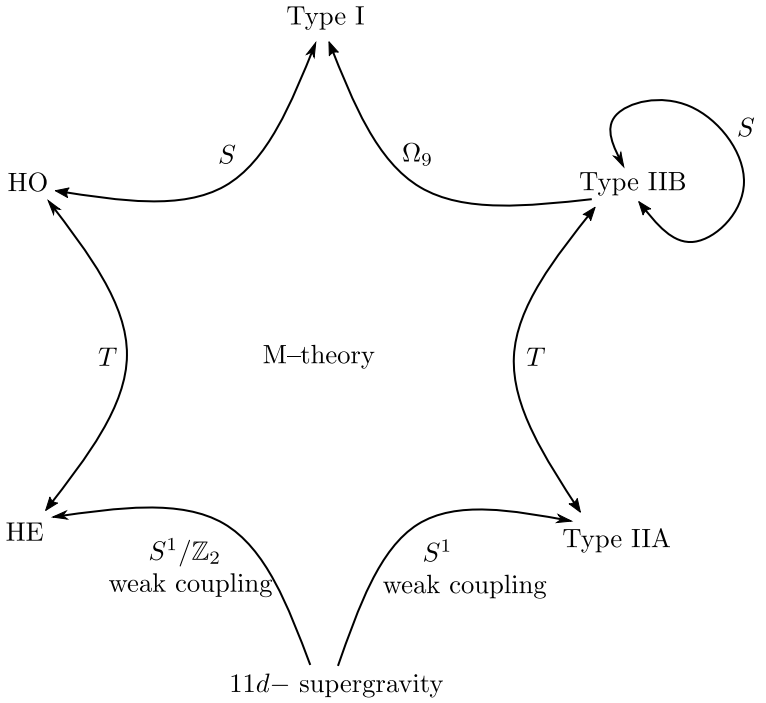


Figure 6.1: A web of theories in $10d$ and $11d$.

One can argue that string theory is the more successful attempt to study quantum gravity and the unification of fundamental physics, as of today. Quantum gravity set aside, string theory also provided a lot of insight into quantum field theories and especially at strong coupling, as it gradually became clear that quantum field theories are ubiquitous in string theories. This will be one of our main interests: we will present tools and techniques allowing for the construction of complicated gauge theories with the help of string theories or M-theory (this is usually referred to as *gauge engineering*) and the study of their dynamics.

In what follows we will mostly be interested in type II string theories and M-theory. We will first present the perturbative definition of type II string theories in Section 6.1. The Section 6.2 presents extended objects known as branes, which arise non-perturbatively in type II string theories. Branes will be among our main characters in what follows. Then we introduce string dualities and M-theory in Section 6.3, together with its extended objects known as M2-branes and M5-branes. Lastly, Section 6.4 sketches the construction of supersymmetric gauge theories in 3, 4, and 5 dimensions from branes and the study of their dynamics.

6.1 Perturbative type II string theories

In the Raymond–Neveu–Schwarz formalism one starts with the world-sheet action

$$S = \frac{1}{4\pi\alpha'} \int_{\mathrm{WS}} \mathrm{d}^2\sigma \left(\partial_\alpha X^\mu \overline{\partial}^\alpha X_\mu + \overline{\psi}^\mu \not{\partial} \psi_\mu \right) , \tag{6.1}$$

[1]There is not universal consensus on what M stands for. Common hypotheses are Mother, Membrane, Matrix, Mystery or Magic.

where the world-sheet WS is a two-dimensional surface. Let assume for now that it is a cylinder (σ_0, σ_1) with $\sigma_1 \sim \sigma_1 + 2\pi$ and Lorentzian metric $\mathrm{d}s^2 = -\mathrm{d}\sigma_0^2 + \mathrm{d}\sigma_1^2$. The two-dimension gamma matrices are $\rho_0 = i\sigma_2$ and $\rho_1 = \sigma_1$ where σ_i are the usual Pauli matrices (and as usual $\partial\!\!\!/ = \partial_\alpha \gamma^\alpha$). The fields X_μ (resp. ψ_μ) are bosonic (resp. Majorana-Weyl fermionic) fields on the worldsheet but both are vectors in target space X of dimension D. The constant α' is called the *Regge slope* and $\alpha' = l_s^2$ where l_s is the string length.

The periodicity condition for the bosonic fields is $X^\mu(\sigma_0, \sigma_1 + 2\pi) = X^\mu(\sigma_0, \sigma_1)$ while for the fermionic fields there are two possibilities:

$$\text{Neveu–Schwarz (NS)}: \quad \psi^\mu(\sigma_0, \sigma_1 + 2\pi) = -\psi^\mu(\sigma_0, \sigma_1), \tag{6.2}$$

$$\text{Ramond (R)}: \quad \psi^\mu(\sigma_0, \sigma_1 + 2\pi) = \psi^\mu(\sigma_0, \sigma_1), \tag{6.3}$$

hence there are four different sectors for closed superstrings: NS–NS, NS–R, R–NS and R–R, since the periodicity conditions are independently set on the left and right movers. In order to study the spectrum one decomposes each given left-moving or right-moving fermion into Fourier modes. For example, in the right moving sector one obtains:

$$\psi^\mu_{\mathrm{NS}} = i^{-1/2} \sum_{r \in \mathbb{Z} + \frac{1}{2}} \psi^\mu_r e^{ir(\sigma_0 - \sigma_1)}, \qquad \psi^\mu_{\mathrm{R}} = i^{-1/2} \sum_{r \in \mathbb{Z}} \psi^\mu_r e^{ir(\sigma_0 - \sigma_1)}. \tag{6.4}$$

In order to quantize the closed superstring in a consistent way one needs to cancel the central charge of the super Virasoro algebra. The easiest way to do so is to fix the dimension of the target space to be $d = 10$: this is the *critical dimension*.

Let us assume for now that the space-time X is the flat Minkowski space $\mathbb{R}^{1,9}$, so that physical states transform in representations of the little group SO(8). The number of left and right movers is constrained by the so-called mass-shell condition.

One defines the ground state in each sector to be a state annihilated by all the $r > 0$ modes. In an NS sector there is not much more to say, while in an R sector the ground state is degenerate due to the ψ^μ_0's. The latter satisfy the Dirac gamma matrix algebra in $9 + 1$ dimensions:

$$\Gamma^\mu = \sqrt{2}\psi^\mu_0, \tag{6.5}$$

and hence the ground states in an R sector form the Dirac representation **32** of this algebra. This representation is reducible to two chiral Weyl representations $\mathbf{16} + \mathbf{16}'$ distinguished by their eigenvalue under $\Gamma = \Gamma^0 \cdots \Gamma^9$ and this extends to the whole string spectrum. Under $\mathrm{SO}(9,1) \to \mathrm{SO}(1,1) \times \mathrm{SO}(8)$:

$$\mathbf{16} \to (\frac{1}{2}, \mathbf{8}) + (-\frac{1}{2}, \mathbf{8}'), \tag{6.6}$$

$$\mathbf{16}' \to (\frac{1}{2}, \mathbf{8}') + (-\frac{1}{2}, \mathbf{8}). \tag{6.7}$$

Let $\exp(\pi i F)$ (resp. $\exp\left(\pi i \tilde{F}\right)$) be the world-sheet fermion number mod 2 in the right (resp. left) moving sector. The Dirac equation leaves an **8** with $\exp(\pi i F)$ and an $\mathbf{8}'$ with $\exp(-\pi i F)$, and likewise in the left moving sector. In the left moving or right moving NS sector physical states transform in the SO(8) vector representation $\mathbf{8}_v$ when $\exp(\pi i F) = 0$ and in the trivial representation **1** when $\exp(\pi i F) = 1$. One still needs to impose the mass-shell level condition, and this leaves 16 different sectors: $(\mathrm{NS}_+, \mathrm{NS}_+)$, $(\mathrm{NS}_-, \mathrm{R}_+), \ldots$. The *Gliozzi–Scherk–Olive projection* [GSO77] ensures that the resulting closed string theory is consistent, i.e. that operator product expansions close, that one-loop amplitudes are modular invariant and that all pairs of vertex operators are mutually local. The idea is to keep only some of the above-mentioned sectors. One finds two independent closed superstring theories which are supersymmetric in space-time $\mathbb{R}^{1,9}$ (and hence which do not have any tachyonic states), that can for example be obtained as:

$$\text{IIA}: \quad (\mathrm{NS}_+, \mathrm{NS}_+) \quad (\mathrm{R}_+, \mathrm{NS}_+) \quad (\mathrm{NS}_+, \mathrm{R}_-) \quad (\mathrm{R}_+, \mathrm{R}_-), \tag{6.8}$$

$$\text{IIB}: \quad (\mathrm{NS}_+, \mathrm{NS}_+) \quad (\mathrm{R}_+, \mathrm{NS}_+) \quad (\mathrm{NS}_+, \mathrm{R}_+) \quad (\mathrm{R}_+, \mathrm{R}_+). \tag{6.9}$$

The lowest energy states have zero energy. They are bosonic in the NS–NS and R–R sectors, and fermionic in the NS–R and R–NS sectors. In type IIA they are:

$$\begin{aligned} &[\mathbf{8}_v \oplus \mathbf{8}] \otimes [\mathbf{8}_v \oplus \mathbf{8}'] \\ &= [\mathbf{1} \oplus \mathbf{28} \oplus \mathbf{35}]_{NS-NS} \oplus [\mathbf{8} \oplus \mathbf{56}']_{NS-R} \oplus [\mathbf{8}' \oplus \mathbf{56}]_{R-NS} \oplus [\mathbf{8}_v \oplus \mathbf{56}_t]_{R-R}\ , \end{aligned} \tag{6.10}$$

and in type IIB:

$$\begin{aligned}&[\mathbf{8}_v \oplus \mathbf{8}] \otimes [\mathbf{8}_v \oplus \mathbf{8}] \\ &= [\mathbf{1} \oplus \mathbf{28} \oplus \mathbf{35}]_{NS-NS} \oplus [\mathbf{8} \oplus \mathbf{56}]_{NS-R} \oplus [\mathbf{8} \oplus \mathbf{56}]_{R-NS} \oplus [\mathbf{1} \oplus \mathbf{28} \oplus \mathbf{35}_+]_{R-R} \ . \end{aligned} \tag{6.11}$$

The target space interpretation of these massless degrees of freedom is as follows.

- The NS–NS sectors in type IIA and type IIB are identical and contain three fields: a dilaton ϕ transforming in $\mathbf{1}$, a 2-form $(B_2)_{\mu\nu}$ which transforms in the $\mathbf{28}$ and a graviton $G_{\mu\nu}$ i.e. a traceless symmetric rank–2 tensor field which transforms in the $\mathbf{35}$.
- The NS–R and R–NS sectors describe target space fermions: two dilatinos which are in the same $\mathcal{N} = 2$ supermultiplet as the dilaton and two gravitinos which are in the same $\mathcal{N} = 2$ supermultiplet as the graviton. Type IIA (resp. type IIB) has $\mathcal{N} = (1, 1)$ (resp. $\mathcal{N} = (2, 0)$) supersymmetry in 10 dimensions and hence dilatinos and gravitinos have opposite (resp. the same) chirality: the dilatinos λ^1 and λ^2 transform in the $\mathbf{8}$ and $\mathbf{8}'$ (resp. in the $\mathbf{8}$) and the gravitinos ψ^1_μ and ψ^2_μ transform in the $\mathbf{56}'$ and $\mathbf{56}$ (resp. in the $\mathbf{56}$).
- The R–R sector describe p-forms in target space. In type IIA there a one-form $(C_1)_\mu$ transforming in the $\mathbf{8}_v$ and a three-form $(C_3)_{\mu\nu\rho}$ transforming in the $\mathbf{56}_t$. In type IIB there is a zero-form (C_0) transforming in the $\mathbf{1}$, a two-form $(C_2)_{\mu\nu}$ transforming in the $\mathbf{28}$ and a self-dual four-form $(C_4)^{(4)}$ transforming in the $\mathbf{35}_+$.

The spectrum of string theories splits, and the difference in energy between consecutive levels is of order α'^{-1}. The low-energy limit of superstring theories is $\alpha' \to 0$, in which the dynamics is given in terms of a low-energy effective action for the zero energy states: supergravity theories.

In 10 dimensions it is known that there are only two supergravity theories with two supersymmetries, called type IIA and type IIB supergravities. They are respectively the low-energy limit of type IIA and type IIB string theories.

The bosonic part of the action in type IIA supergravity (in the *string frame*) is:

$$\begin{aligned}S_{\text{IIA}} = \frac{1}{2\kappa_0^2} \int \mathrm{d}^{10}x\sqrt{-G} \Big\{ e^{-2\phi} \left[R + 4\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}|H_3|^2 \right] \\ - \frac{1}{2}|F_2|^2 - \frac{1}{2}|\tilde{F}_4|^2 \Big\} - \frac{1}{4\kappa_0^2} \int B_2 \wedge F_4 \wedge F_4 \ , \end{aligned} \tag{6.12}$$

where $G_{\mu\nu}$ is the metric, ϕ is the dilaton, $H_3 = \mathrm{d}B_2$ is the NS–NS three-form field strength, $F_2 = \mathrm{d}C_1$ is the field strength of the R–R 1-form, $F_4 = \mathrm{d}C_3$ is the one of the R–R 3-form, and $\tilde{F}_4 = \mathrm{d}C_3 + C_1 \wedge H_3$.

The bosonic part of the type IIB supergravity action in the string frame is:

$$\begin{aligned}S_{\text{IIB}} = \frac{1}{2\kappa_0^2} \int \mathrm{d}^{10}x\sqrt{-G} \Big\{ e^{-2\phi} \left[R + 4\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}|H_3|^2 \right] \\ - \frac{1}{2} \left(|F_1|^2 + |\tilde{F}_3|^2 + \frac{1}{2}|\tilde{F}_5|^2 \right) \Big\} - \frac{1}{4\kappa_0^2} \int C_4 \wedge H_3 \wedge F_3 \ , \end{aligned} \tag{6.13}$$

where now $F_1 = \mathrm{d}C_0$, $F_3 = \mathrm{d}C_2$, $F_3 = \mathrm{d}C_5$, $\tilde{F}_3 = F_3 - C_0 H_3$ and

$$\tilde{F}_5 = F_5 - \frac{1}{2} C_2 \wedge H_3 + \frac{1}{2} B_2 \wedge F_3 \ . \tag{6.14}$$

The fields equation are compatible with the self-duality of the four-form: $\tilde{F}_5 = \star F_5$ but they do not imply it. This constraint has to be imposed as an added constraint on the solutions of the action.

Both actions can be re-expressed in the *Einstein frame* under a redefinition of the metric $G \to G^E$ and the dilaton ϕ, in which the Einstein–Hilbert part of the action has its standard form. One can then read the expression of the 10-dimensional Newton constant G_N:

$$16\pi G_N = 2\kappa^2 = 2\kappa_0^2 g_s^2 = (2\pi)^7 \alpha'^4 g_s^2 \ , \tag{6.15}$$

where the last equality can be derived by computing a graviton scattering.

The asymptotic value ϕ_0 of the dilaton field at infinity sets the string coupling constant:

$$g_s = e^{\phi_0} \ . \tag{6.16}$$

Both theories are invariant under (higher) gauge transformations. The first is

$$B_2 \to B_2 + \mathrm{d}\Lambda_1 \tag{6.17}$$

where Λ_1 is a one-form. Fundamental strings are the electric charges for such a gauge symmetry, since B_2 couples to the worldsheet as

$$\frac{1}{2\pi\alpha'}\int_{\mathrm{WS}} B_2 \ . \tag{6.18}$$

The second class of gauge transformations is

$$C_p \to C_p + \mathrm{d}\Lambda_{p-1} \tag{6.19}$$

where Λ_{p-1} is a $(p-1)$-form, and where $p = 1, 3$ in type IIA and $p = 0, 2, 4$ in type IIB. These higher gauge symmetries have D-branes as non-perturbative electric and magnetic sources.

6.2 Branes

D-branes were introduced in [DLP89] as submanifolds of the target space $\mathbb{R}^{1,9}$ on which open strings can end. Perturbatively, type II superstrings do not describe open strings however D-branes are intrinsically non-perturbative objects: type II superstrings contain open strings in their non-perturbative spectrum.

The simplest occurrence of a D-brane is as one considers an open superstring (i.e. a map from a strip $\mathbb{R} \times [0, \pi]$ into $\mathbb{R}^{1,9}$) such that the first p coordinates satisfy the Neumann boundary conditions and the remaining $10 - p$ satisfy the Dirichlet boundary conditions. Quantizing the open string with such boundary conditions yields (at low energies $\alpha' \to 0$) a $(p+1)$-dimensional U(1) super Yang–Mills theory on the world volume of the D-brane, obtained by dimensional reduction of $\mathcal{N} = 1$ U(1) super Yang–Mills theory in 10 dimensions. In general, Dp-branes for $p = 0, \ldots, 9$ are $(p+1)$-dimensional Lorentzian submanifolds in space-time, and the gauge theory massless degrees of freedom which arise in the worldvolume of such a D-brane correspond to the endpoints of open strings attached to it. The Lorentzian submanifold on which a D-brane lies is called the *worldvolume of that D-brane*.

The GSO projection on this open string sector constrains the spatial dimension of D-branes to be even (resp. odd) in type IIA (resp. type IIB) superstring theory: type IIA contains D0, D2, D4, D6 and D8-branes while type IIB contains D1, D3, D5, D7 and D9-branes. It might also be the case that such a D-brane is not a Lorentzian submanifold of the target space but a Euclidean submanifold. In that case one speaks of euclidean D-branes. Type IIA contains ED1, ED3, ... euclidean D-branes and type IIB ED2, ED4, ...

The dynamics of the U(1) gauge theory on the worldvolume of a D-brane and how it couples to background NS–NS fields is encoded in the *Dirac–Born–Infeld action* (DBI action):

$$S_{\mathrm{DBI}} = -T_{\mathrm{D}p}\int_{WV} \mathrm{d}^{p+1}\xi\sqrt{-\det(G + B + 2\pi\alpha' F)} \ , \tag{6.20}$$

where F is the field strength of the abelian gauge field on the worldvolume WV, and where:

$$T_{\mathrm{D}p} = \frac{1}{(2\pi)^p \alpha'^{\frac{p+1}{2}} g_s} \tag{6.21}$$

is the Dp-brane tension, computed from the exchange of a closed string between two D-branes.

One can label the ends of open strings with non-dynamical *Chan–Paton factors* $1, \ldots, N$. These can be thought of as arising from N coincident D-branes located somewhere in target space, they keep track of to which D-brane of the stack an endpoint of an open string belongs. Quantizing open strings with Chan–Paton factors shows that a stack of N coincident D-branes supports a U(N) gauge theory on their worldvolume. The Dirac–Born–Infeld action can be generalized to such a non-abelian gauge theory.

At leading order in α' the term containing F in the expansion of the (non-abelian) DBI action at low energies is:

$$S_{\mathrm{DBI}}^{\alpha'\to 0} = -\frac{T_{\mathrm{D}p}(2\pi\alpha')^2}{4}\int \mathrm{d}^{p+1}\sigma\sqrt{-G}\,\mathrm{Tr}\,F^{\mu\nu}F_{\mu\nu} \tag{6.22}$$

from which the gauge coupling $g_{\rm YM}$ of the low-energy super Yang–Mills theory on a stack of branes can be expressed as:

$$g^2_{\rm YM} = T^{-1}_{{\rm D}p}(2\pi\alpha')^{-2} = (2\pi)^{p-2}{\alpha'}^{\frac{p-3}{2}} g_s \ . \tag{6.23}$$

D-branes preserve one-half of the supersymmetry: they are BPS states of the 10 dimensional supersymmetric string theory in $\mathbb{R}^{1,9}$. If Q_L and Q_R are the two Majorana-Weyl supercharges, a flat Dp-brane extending along $x^0, \ldots, x^p$ preserves the linear combination

$$\epsilon^L Q_L + \epsilon^R Q_R \tag{6.24}$$

with $\epsilon_L = \Gamma_0 \cdots \Gamma_p \epsilon_R$ and hence they carry conserved charges, which are nothing else than the antisymmetric R–R charges of the closed string theory:

- In type IIA theory D0-branes (resp. D2-branes) are electrically charged under C_1 (resp. C_3) while D6-branes (resp. D4-branes) are magnetically charged under C_1 (resp. C_3). The D8-branes are electrically charged under a non-dynamical 9-form.
- In type IIB theory D1-branes (resp. D3-branes) are electrically charged under C_2 (resp. C_4) while D5-branes (resp. D3-branes) are magnetically charged under C_2 (resp. C_4). The D7-branes are magnetically charged under C_0 for which there also exist electrically charged objects known as D-instantons (they are D(-1)-branes). The D9-branes couple to an R–R 10-form.

Electric–magnetic duality is Hodge duality, as in Maxwell theory. For example, in type IIB D1-branes couple electrically to C_2 whose field strength $F_2 = {\rm d}C_2$ is a three-form. The magnetic dual of the latter is $\star {\rm d}C_2$ which is a seven-form and the field strength of a six-form potential, to which D5-branes couple naturally. Hence D1-branes couple electrically to C_2 while D5-branes couple magnetically to it.

The coupling between a Dp-brane and an R–R $(p+1)$-form potential C_{p+1} is described by

$$S = T_{{\rm D}p} \int_{WV} C_{p+1} \tag{6.25}$$

at leading order in α' as $\alpha' \to 0$. Hence the charge of the Dp brane is nothing else than (g_s times) its tension. This translates the fact that the attraction between two Dp-branes from the exchange of NS–NS fields compensate exactly the repulsion from the exchange of R–R fields, consistently with the fact that D-branes are stable, BPS objects in the theories.

Because D-branes have a tension they are gravitational sources and backreact by deforming the ambient metric. This backreaction is controlled by the product $T_{{\rm D}p}G_N$. At weak string coupling however something interesting happens: the tension of a Dp-brane grows as $\sim g_s^{-1}$ whereas the 10-dimensional Newton constant vanishes as g_s^2, hence the product $T_{{\rm D}p}G_N \sim g_s$ is small when g_s is small. This is very useful in that D-branes can be used as probes of the geometry: at weak string coupling $g_s \to 0$ and low energies $\alpha' \to 0$ the quantum geometry of type II string theory can be explored from the perspective of the world-volume theory of a D-brane, setting aside the gravitational back-reaction of the D-brane.

More precisely, the gravitational backreaction of a N Dp-brane at low energy is well approximated by the corresponding *extremal black Dp-brane* solution of type II supergravity theories. In the string frame:

$${\rm d}s^2 = f_p^{-1/2}(r){\rm d}x\cdot{\rm d}x + f_p^{1/2}(r){\rm d}y\cdot{\rm d}y, \quad e^{\phi(r)} = g_s f_p^{(3-p)/4}(r), \quad {\rm d}C_{p+1} = {\rm d}f_p^{-1}(r)\wedge{\rm d}x^0\wedge\cdots\wedge{\rm d}x^p, \tag{6.26}$$

where x stands for the $p+1$ coordinates in the worldvolume of the Dp-brane and y for the $9-p$ transverse coordinates, and where

$$f_p(r) = 1 + \left(\frac{r_p}{r}\right)^{7-p} \tag{6.27}$$

with

$$\left(\frac{r_p}{l_s}\right)^{7-p} = (2\sqrt{\pi})^{5-p}\Gamma\left(\frac{7-p}{2}\right) g_s N \ . \tag{6.28}$$

The extremal black Dp-brane solution of above is a solution of type IIA supergravity if p is even and type IIB supergravity is p is odd. The dilaton approaches a constant as $r \to \infty$, which defines the string coupling constant g_s. For $p = 3$, the dilaton is actually constant everywhere in space-time.

One sees from Equation (6.28) that when $g_s \to 0$ one also has $r_p \to 0$, and hence that $H_p \to 1$: this is the probe limit.

NS5-branes.

This analysis calls for another: both in type IIA and type IIB superstring theories there is the massless NS–NS B_2-field, which is a two-form and to which the worldsheet couples naturally. The magnetic potential dual to B_2 is a six-form in target space. Hence there is again another type of extended solitonic object in type II theories known as the NS5-brane, for it is a $(5+1)$-brane and couples magnetically to the NS–NS two-form. Its tension is

$$T_{\mathrm{NS5}} = \frac{1}{(2\pi)^5 \alpha'^3 g_s^2} \ , \tag{6.29}$$

and the fact that it scales as g_s^{-2} in terms of g_s is much more familiar than the scaling of the D-brane tensions, since solitons in gauge theories usually scale as g^{-2}, where g is the coupling constant.

Even if both type IIA and type IIB string theories have NS5-branes which have the same tension, it must be emphasized that type IIA NS5-branes are not the same as type IIB NS5-branes. In particular, the worldvolume theories on NS5-branes differ whether one is in type IIA or type IIB string theory: in type IIB the worldvolume theory of an NS5-brane is a usual U(1) super Yang–Mills theory with $\mathcal{N} = (1,1)$ supersymmetry, while in type IIA the worldvolume theory of an NS5-brane has $\mathcal{N} = (0,2)$ supersymmetries and describes an anti self-dual tensor multiplet.

The NS5-brane also corresponds to an extended object in type II supergravities, at low energies: an extremal black NS5-brane. In the string frame in type IIA supergravity for example, it is given by:

$$\mathrm{d}s^2 = \mathrm{d}x \cdot \mathrm{d}x + f(r) \mathrm{d}y \cdot \mathrm{d}y, \quad e^{\phi(r)} = f^{1/2}(r), \quad (C_3)_{\mu\nu\rho} = \epsilon_{\mu\nu\rho\sigma} \partial_q f(r), \tag{6.30}$$

where

$$f(r) = 1 + \frac{N\alpha'}{r^2} \ . \tag{6.31}$$

An NS5-brane in type II string theories is also an half-BPS object. If it is extended along $x^0, \ldots, x^5$ it preserves the combination

$$\epsilon^L Q_L + \epsilon^R Q_R \tag{6.32}$$

with $\epsilon_L = \Gamma_0 \cdots \Gamma_5 \epsilon_R$ and $\epsilon_R = -\Gamma_0 \cdots \Gamma_5 \epsilon_L$.

6.3 String dualities and M-theory

T-duality is a perturbative duality of type II superstring theories which maps type IIA theory to type IIB theory. Let us assume for concreteness that the 10-dimensional space-time is $\mathbb{R}^{1,8} \times S^1_R$ with the flat Lorentzian metric and where S^1 has length R. T-duality along the compact direction asserts that type IIA theory in $\mathbb{R}^{1,8} \times S^1_R$ is dual to type IIB theory in $\mathbb{R}^{1,8} \times S^1_{\alpha'/R}$, where:

- Kaluza–Klein modes and winding modes of closed strings along the compact direction have been exchanged,
- the boundary conditions for open strings (Neumann and Dirichlet) along the compact direction have been exchanged,
- the new string coupling in terms of the old one g_s is

$$g_s \to \frac{\sqrt{\alpha'}}{R} g_s, \tag{6.33}$$

- In a trivial NS–NS background, the R–R sector mixes as:

Type IIA	Type IIB
$(C_1)_a$	$(C_2)_{9a}$
$(C_1)_9$	C_0
$(C_3)_{abc}$	$(C_4)_{9abc}$
$(C_3)_{9bc}$	$(C_2)_{bc}$

(6.34)

- In the NS–NS sector the metric G and the B field mix according to Buscher's rules [Bus88, Bus87].

Under a T-duality in a direction transverse to the worldvolume of a Dp-brane the latter is mapped to a D$(p+1)$-brane whose worldvolume is the one of the Dp supplemented by the direction of the T-duality. Under a T-duality in a direction parallel to the worldvolume of a Dp-brane the latter is mapped to a D$(p-1)$-brane whose worldvolume is the one of the Dp but the direction of the T-duality.

As far as NS5-branes are concerned, T-dualizing in a direction parallel to the worldvolume of an NS5 yields another NS5 (however not of the same type and hence whose worldvolume theory is different), whereas T-dualizing in a direction perpendicular to the NS5 exchanges the brane for pure geometry (this comes from the exchange of the NS–NS fields G and B under T-duality) [EJL98]. For example a stack of NS5-branes under a perpendicular T-duality is mapped to a multi-centered Taub–Nut space.

Let us now restrict to type IIB theory in order to discuss S-duality, which is a non-perturbative $\mathrm{SL}_2(\mathbb{Z})$ duality symmetry. The type IIB supergravity action in the Einstein frame is invariant under $\mathrm{SL}_2(\mathbb{R})$ which acts as

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} B_2 \\ C_2 \end{bmatrix} = \begin{bmatrix} aB_2 + bC_2 \\ cB_2 + dC_2 \end{bmatrix} \tag{6.35}$$

on the NS–NS and R–R 2-forms, as

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \tau = \frac{a\tau + b}{c\tau + d} \,, \tag{6.36}$$

where $\tau = C_0 + ie^{\phi}$ is the axion-dilaton field, whereas $\mathrm{SL}_2(\mathbb{R})$ acts trivially on the Einstein frame metric G^E and on C_4. In the quantum theory where Dirac quantization condition has to be imposed on charged states, the whole group $\mathrm{SL}_2(\mathbb{R})$ is not a symmetry anymore, but only its subgroup $\mathrm{SL}_2(\mathbb{Z})$ which is an exact duality symmetry of type IIB superstring theory.

S-duality of type IIB theory is another name for the action of

$$\mathrm{S} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \,. \tag{6.37}$$

Under it: $B_2 \to C_2$, $C_2 \to -B$ and $\tau \to -1/\tau$ (when $C_0 = 0$, this amounts to $g_s \to g_s^{-1}$) i.e. S-duality is a strong-weak duality.

Because of how S-duality acts on the fields one sees that a fundamental string F1 which couples electrically to B is S-dual to a D1-brane which couples electrically to C_2, and vice-versa; an NS5-brane which couples magnetically to B is S-dual to a D5-brane brane which couples magnetically to C_2.

There are in fact two utterly important consequences of this $\mathrm{SL}_2(\mathbb{Z})$ duality symmetry of type IIB superstring theory. First, under the action of

$$S = \begin{bmatrix} p & q \\ a & b \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \,, \tag{6.38}$$

a fundamental string is mapped to a string-like object coupling to B_2 like p fundamental strings and to C_2 like q D1-strings. Because this action of $\mathrm{SL}_2(\mathbb{Z})$ is an exact duality symmetry of type IIB superstring theory, this object which behaves as a bound state of p F1s and q D1s is as elementary and stable as a fundamental string. It is called (p,q)-string. Likewise, the fact that $\mathrm{SL}_2(\mathbb{Z})$ is an exact duality symmetry implies the existence of elementary and stable (p,q) 5-branes (or fivebranes) which are bound states of p D5's and q NS5's (note the change of convention with respect to (p,q)-strings). The tension of a (p,q)-fivebrane is given by [BBS06]:

$$T_{(p,q)} = |p + \tau q| T_{\mathrm{D5}} \,. \tag{6.39}$$

Independently, D3-branes are preserved by S-duality however from Equation (6.23) one sees that the Yang–Mills coupling of the theory on the worldvolume of D3-branes transforms as:

$$g_{\mathrm{YM}} \to g_{\mathrm{YM}}^{-1} \,. \tag{6.40}$$

The worldvolume theory on a stack of N D3-branes in four-dimensional $\mathcal{N} = 4$ U(N) super Yang–Mills theory, and hence the S-duality symmetry of type IIB superstrings descend to the Montonen–Olive duality of $\mathcal{N} = 4$ SYM in four dimension presented in Section 5.7.

Second, fundamental strings can end on D-branes, by definition. Consider for definiteness a fundamental string streching between two D1's. Under an S-duality this amounts to a D1 streching between

two F1's, hence D1-branes can end on fundamental strings! By electric-magnetic duality this implies in turn that D5's can end on NS5's. Repeated applications of T-duality and S-duality on configurations of branes and fundamental strings tell us that open branes exist [Str96]. For example, consider a D5 streching between two NS5's:

	0	1	2	3	4	5	6	7	8	9
D5	–	–	–	–	–	–	0	0	0	0
$NS5_1$	–	x	–	–	–	–	–	0	0	0
$NS5_2$	–	x'	–	–	–	–	–	0	0	0

(6.41)

After T-dualities in the directions 4 and 5 which are in the worldvolume of the NS5's and of the D5 one obtains a D3-brane stretching between two NS5's:

	0	1	2	3	4	5	6	7	8	9
D3	–	–	–	–	0	0	0	0	0	0
$NS5_1$	–	x	–	–	–	–	–	0	0	0
$NS5_2$	–	x'	–	–	–	–	–	0	0	0

(6.42)

and this is S-dual to a D3 stretching between two D5s: we just learned that a D3 brane can end on D5s (actually, on any (p, q) 5-brane)! These successive dualities are represented in Figure 6.2.

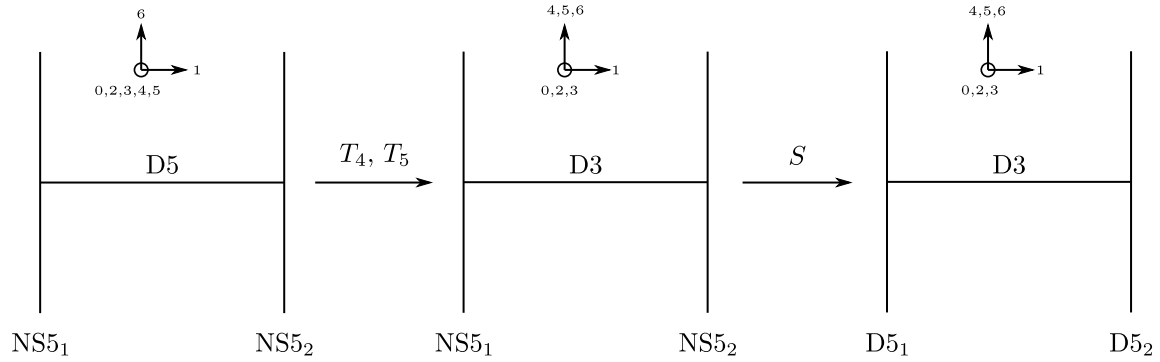


Figure 6.2: A chain of exact string dualities and their action on branes.

M-theory.

Consider D0-branes in type IIA theory. Their tension is given in Equation (6.21):

$$T_{\mathrm{D0}} = (l_s g_s)^{-1} \ . \tag{6.43}$$

It is tempting to interpret this as the first Kaluza–Klein excitation of a massless supergravity multiplet in 11 dimensions (as we will see later there is a unique supergravity theory in 11 dimensions) compactified on a circle of radius

$$R_{11} = l_s g_s \ . \tag{6.44}$$

Likewise, a bound state of N D0-branes can be interpreted as the N-th Kaluza-Klein mode of the same supermultiplet. The radius of the compactification circle goes as the string coupling constant, and one expects that the strong coupling limit of type IIA superstring theory is described by an 11-dimensional theory.

There is another hint that type IIA might be described as the compactification of an 11-dimensional theory. The worldvolume theory of a type IIA NS5-brane describes an anti-self dual tensor multiplet, as well as five scalars. Four of the latter can be interpreted as describing the position of the NS5 in the transverse directions through their vacuum expectation values, and the fifth again hints for an additional hidden direction.

All this is supported by the existence of a unique 11-dimensional $\mathcal{N} = 1$ supergravity theory, whose bosonic action is:

$$S = \frac{1}{2\kappa_{11}^2} \int \mathrm{d}^{11}x \sqrt{-G} \left(R - \frac{1}{2}|F_4|^2 \right) - \frac{1}{6} \int A_3 \wedge F_4 \wedge F_4 \ , \tag{6.45}$$

where R as usual is the scalar curvature, $F_4 = \mathrm{d}A_3$ is the field strength of a three-form and where κ_{11} is the eleven-dimensional gravitational coupling constant. The eleven-dimensional Newton constant is

$$16\pi G_{11} = 2\kappa_{11}^2 = \frac{1}{2\pi}(2\pi l_p)^9 \ , \tag{6.46}$$

where l_p is the eleven-dimensional Planck length.

The theory which describes the strong coupling limit of type IIA superstring theory and which has eleven-dimensional supergravity as its low-energy limit is called M-theory. It does not contain any strings, and it cannot be defined perturbatively which makes it harder to handle than the type II theories. Since there is a 3-form one expects that there are 2-branes and 5-branes in the theory, dubbed M2- and M5-branes, respectively. Fundamental strings in type IIA string theory must arise from M2-branes wrapping the circular additional dimension of M-theory. They have tension

$$T_{\mathrm{F1}} = \frac{1}{2\pi l_s^2}, \tag{6.47}$$

and if it is a wrapped M2-brane it must be that $T_{\mathrm{F1}} = 2\pi R_{11} T_{\mathrm{M2}}$. Hence one deduces the tension of an M2-brane from Equation (6.44) and $l_p = (g_s)^{1/3} l_s$:

$$T_{\mathrm{M2}} = \frac{2\pi}{(2\pi l_p)^3} \ . \tag{6.48}$$

Likewise one expects D4-branes in type IIA theory to be M5-branes wrapped on the circular 11-th direction, from which one deduces that:

$$T_{\mathrm{M5}} = \frac{2\pi}{(2\pi l_p)^6} \ . \tag{6.49}$$

D2-branes are identified with M2 branes not wrapping the circular direction and type IIA NS5 branes are identified with unwrapped M5-branes. D6-branes however cannot be related to M2 or M5 branes in any simple way; it corresponds instead to a Kaluza-Klein monopole in M-theory. Similarly to D-branes and NS5-branes in type II superstrings, at low energies M2 and M5 branes correspond to black brane solutions of 11-dimensional supergravity.

6.4 Supersymmetric quantum field theories from brane configurations

In [HW97a] Hanany and Witten considered configurations of NS5-branes, D5-branes and D3-branes in type IIB string theory in $\mathbb{R}^{1,9}$, where the D5-branes are extended along x^0, x^1, x^2, x^7, x^8 and x^9, the NS5-branes along x^0, x^1, x^2, x^3, x^4 and x^5 and the D3-branes along x^0, x^1, x^2 and x^6:

	0	1	2	3	4	5	6	7	8	9
D5	–	–	–	x_j^3	x_j^4	x_j^5	z_j	–	–	–
NS5	–	–	–	–	–	–	t_i	x_i^7	x_i^8	x_i^9
D3	–	–	–	×	×	×	–	×	×	×

(6.50)

The D3s are finite (or semi-infinite) and end on the NS5s if some conditions on the coordinates of the branes are satisfied. An example of such a brane setup is shown in Figure 6.3. The horizontal direction in the Figure stands for the direction 6 whereas the vertical one stands for 3,4 and 5. The directions 7,8,9 of space-time are depicted transversely to the plane of the figure, where the directions 0, 1 and 2 are swept by all the branes in this setup and hence they are not represented. With these conventions, NS5-branes are vertical lines, D3-branes are horizontal segments of half-line ending on NS5's, and they can stack-up: for example, there is a stack of three D3-branes streching between the two left-most NS5's in Figure 6.3. D5-branes extend along the directions 0,1,2 and the ones transverse to the plane of the figure only, and hence they appear as dots. D3-branes can also end on D5's, provided their coordinate along 3,4 and 5 coincide. This happens between the two right-most NS5's in Figure 6.3. A D3-brane streching between two NS5's can also slide in the directions 3,4 and 5 since the NS5's are extended along

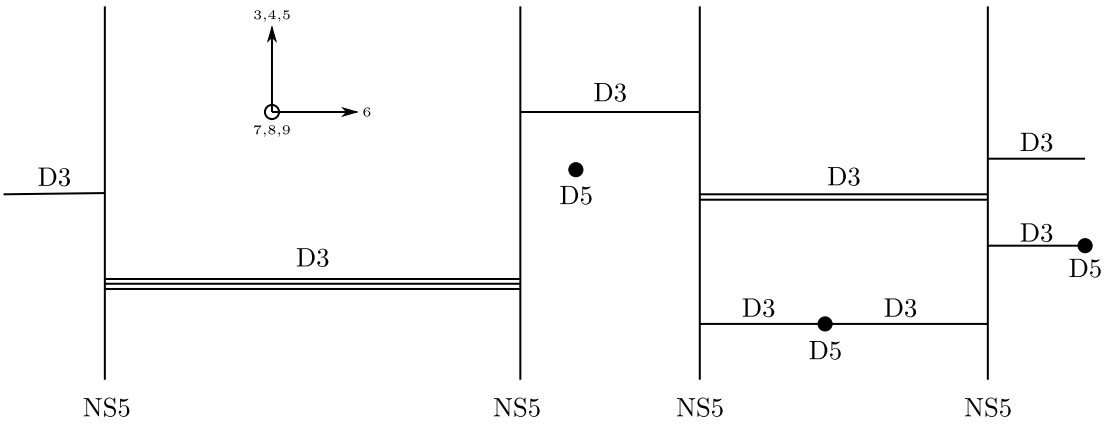


Figure 6.3: A Hanany-Witten brane setup.

them, and hence a stack of D3-branes streching between two NS5's can split, as the one between the two right-most NS5-branes.

Such brane system preserves one fourth of the 32 supercharges of type IIB theory, i.e. 8 real supercharges. Since the fivebranes extend along two directions not shared by the D3-branes they are much heavier than the latter, and one can think the parameters describing the position of the fivebranes as being fixed, and study the quantum dynamics of the worldvolume theory on the D3s only. At low energies this yields a $\mathcal{N} = 4$ quantum field theory in $2 + 1$ dimensions. By dimensional reduction, each stack of N D3-branes extending between two NS5s located at x_i^6 and $x_{i'}^6$ gives rise to an U(N) gauge group with gauge coupling

$$\frac{1}{g^2} \sim |x_i^6 - x_{i'}^6| \ , \tag{6.51}$$

where the proportionality factor is universally determined by the coupling g_s of type IIB superstring theory. Semi-infinite D3-branes are infinitely more massive than the finite ones and they give rise to flavor groups in the low energy theory. Open strings attached to the stack of D3s on the left of an NS5 and to the stack of D3s on the right of the same NS5 gives rise to a bifundamental field in the worldvolume theory of the D3s, which is massless if the x^3, x^4 and x^5 coordinates of the two stack of D3s coincide. Open strings streching between D3s and D5s give rise to flavor fields in the low-energy effective theory in the fundamental representation of the gauge group on the D3s which are massless if the x^3, x^4 and x^5 coordinates of the D3 and the D5 coincide.

From the point of view of the NS5-branes, the endpoint of a D3 is a magnetic monopole. Such an insertion on the worldvolume of an NS5 bends it, hence taking that into account the position of the NS5 along x^6 now becomes a function of the transverse direction to the D3-endpoint within the NS5. Away from the endpoints, it satisfies

$$\nabla^2 x^6(x^3, x^4, x^5) = 0 \tag{6.52}$$

while the endpoint itself acts like a Dirac source, but it is difficult to describe quantitatively such a singularity. The Green function of the Laplacian in three dimensions being $\nabla^2 \delta(\vec{0}) = |\vec{x}|^{-1}$, the x^6 value of the NS5 tends to a constant at infinity in the three-dimensional (x^3, x^4, x^5) space. In Equation (6.50) if one uses x^6 values at infinity of NS5 branes on the right-hand side, the left hand side corresponds to the bare value of the coupling constant; in general $\mu = |(x^3, x^4, x^5)|$ is interpreted as an energy scale at which the coupling satisfies $g^2(\mu)^{-2} = |x_i^6(\mu) - x_{i'}^6(\mu)|$.

Hanany–Witten transitions.

When an NS5 passes through a D5 a three-brane connecting them is created, as explained in [HW97a]: otherwise one finds a discrepancy between the analysis of the brane configuration and known results about the moduli space of the low-energy field theory. This is depicted in Figure 6.4, where we have taken into account the backreaction of the D3-endpoint on the worldvolume of the NS5, so that the NS5 has the asymptotic shape $x^6(\vec{w}) = |\vec{w} - \vec{w}_i|^{-1}$.

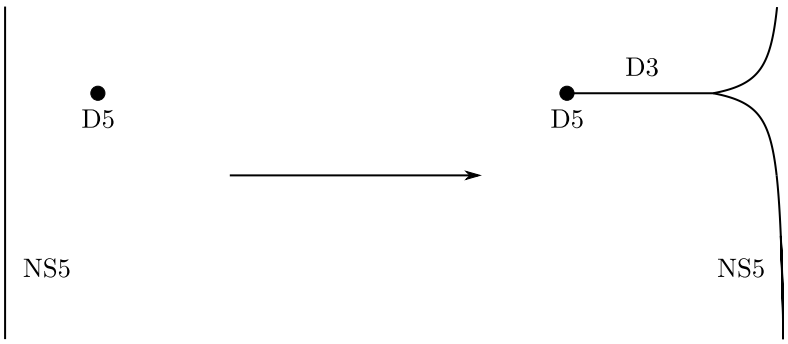


Figure 6.4: The Hanany–Witten effect

An NS5-brane is a magnetic source for the B_2 field in type IIB superstring theory, i.e.:

$$\int_S \frac{H_3}{2\pi} = 1 \ , \tag{6.53}$$

with $H_3 = \mathrm{d}B_2$ as usual, and where S is a three-sphere wrapping once around the NS5. Likewise, a D5 is a magnetic source for the C_2 field:

$$\int_{S'} \frac{F_3}{2\pi} = 1 \ , \tag{6.54}$$

with $F_3 = \mathrm{d}C_2$ as usual and where S' is a three-sphere wrapping once around the D5. The linking number of the worldvolumes of the NS5 and the D5 inside $T = \mathbb{R}^7_{3,4,5,6,7,8,9}$ is

$$L = \int_{Y_D} \frac{H_3}{2\pi} = \int_{Y_{\mathrm{NS}}} \frac{F_3}{2\pi} \ , \tag{6.55}$$

evaluated at any point in $\mathbb{R}^3_{0,1,2}$. In the last equation Y_D (resp. Y_{NS}) is the projection of the worldvolume of the D5 (resp. NS5) in T. As a D5 passes through an NS5 this linking number changes and that must be compensated by the addition of a D3 brane stretching between the D5 and the NS5. It is indeed argued in [HW97a] that the total linking number (which also receives a contribution from the D3 brane) should not change as a brane is passed through another.

6.4.1 Four-dimensional theories with eight supercharges

Under a T-duality along x^3 the setups of the previous subsubsection becomes brane configurations in type IIA superstring theory studied in [Wit97]. Their properties matches the known quantum behaviour of four dimensional $\mathcal{N} = 2$ field theories, and the lift to M-theory even allows a direct calculation of Seiberg-Witten curves (which we introduced in Section 5.6). For simplicity and conciseness we will only consider configurations without D6 branes. The classical setup is:

	0	1	2	3	4	5	6	7	8	9
NS5	–	–	–	–	–	–	×	0	0	0
D4	–	–	–	–	×	×	–	0	0	0

, (6.56)

where the ×s mean that each corresponding brane has a fixed position along that direction.

Let $v = x^4 + ix^5$. The low-energy effective theory on the fourbranes is an $\mathcal{N} = 2$ supersymmetric gauge theory in four dimensions: for example, the low-energy theory on the D4s corresponding to the classical brane configuration shown in Figure 6.5 can be described as the quiver shown below on the same figure, where circles are gauge groups, squares are flavor groups and edges are bifundamental hypermultiplets.

At low energies any such brane setup describes a four dimensional theory along x^0, x^1, x^2 and x^3 whose gauge group is

$$\prod_\alpha SU(k_\alpha) \ , \tag{6.57}$$

where α labels the NS5s from left to right, and where k_α is the number of open D4s between $\mathrm{NS5}_\alpha$ and the $\mathrm{NS5}_{\alpha+1}$. The gauge factors are SU(N)s rather than U(N)s because there is a constraint on the position

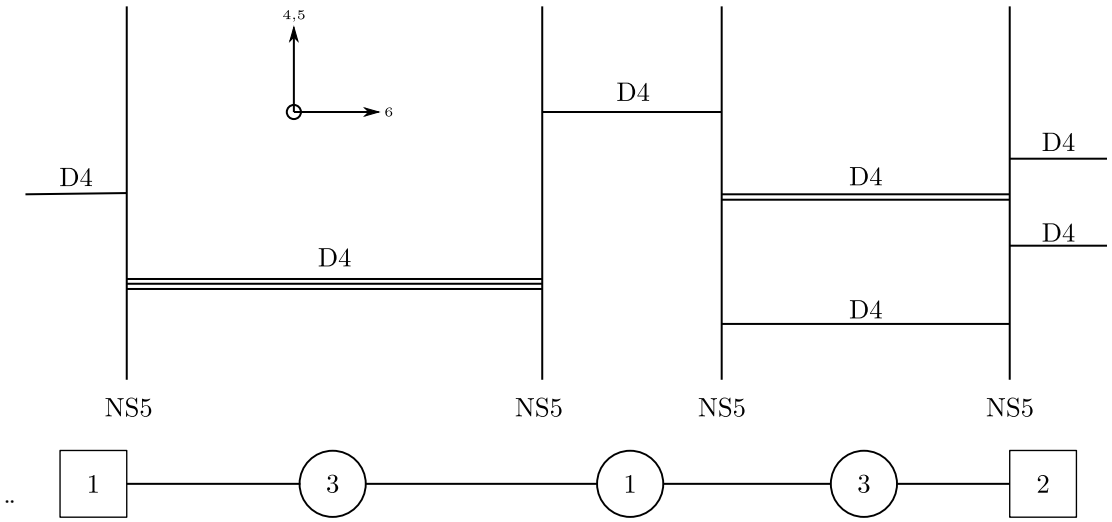


Figure 6.5: A four-dimensional $\mathcal{N} = 2$ gauge theory as the low energy theory on a brane setup in type IIA.

of the D4 endpoints in the worldvolume of each NS5 brane. The coupling constant of a factor $\mathrm{SU}(k_\alpha)$ should be given by an expression of the form

$$\frac{1}{g_\alpha^2} = \frac{|x^6_{\alpha+1} - x^6_\alpha|}{g_s} \tag{6.58}$$

however g_α^2 is a function of v because of the backreaction of the D4s endpoints on the NS5s. Crucially, in this case x^6 is the solution of a two dimensional Laplace equation (away from the D4 endpoints):

$$\nabla^2 x^6(v) = 0 \ . \tag{6.59}$$

Since the Green function of the Laplacian in two dimensions is logarithmic in $|v|$, the x^6 coordinate of an NS5 does not tend to a definite x^6 value at infinity in v in general. This translates the logarithmic one-loop running of the coupling constant in four-dimensional non-abelian gauge theories, with the absolute value $|v|$ interpreted as a mass scale.

At each $\mathrm{NS5}_\alpha$ there is a hypermultiplet in the bifundamental representation of the gauge groups describing the D4 brane stacks on each side, whose mass can be expressed easily in terms of the position of the D4 endpoints.

In the low-energy four dimensional theory x^6 is the real part of a complex field in a vector multiplet. Its imaginary part is a scalar field which propagates on the five-brane (the fifth scalar on the worldvolume of type IIA NS5 branes), better interpreted in M theory instead of type IIA. Let $x^{10} \sim x^{10} + 2\pi R$ be the coordinate on the M-theory circle, with $R = l_s g_s$. With it Equation (6.58) can be rewritten as an equation for the complexified gauge coupling of the factor $\mathrm{SU}(k_\alpha)$ at the energy scale $|v|$:

$$-i\tau_\alpha(v) = s_\alpha(v) - s_{\alpha-1}(v) \ , \tag{6.60}$$

where $Rs_\alpha(v) = x^6_\alpha(v) + ix^{10}_\alpha(v)$ is the coordinates of $\mathrm{NS5}_\alpha$ along x^6 and x^{10} in M-theory units.

The D4s and NS5s in type IIA both descend from M5 branes, and the type IIA setups that we are considering can be reinterpreted as a configuration of a single smooth M5 brane in $\mathbb{R}^{1,9} \times S^1$:

	0	1	2	3	7	8	9	4	5	6	10
M5	−	−	−	−	0	0	0	·····	Σ	·····	

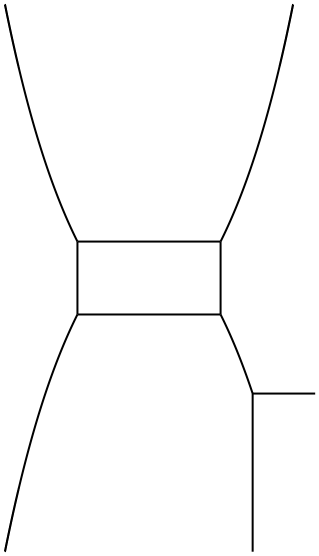
(6.61)

where Σ is a surface in $Q = \mathbb{R}^3_{4,5,6} \times S^1_{10}$. This has the nice consequence of (generically) smoothing out the singularities of the type IIA setups, which are difficult to understand quantitatively, as is usual with singularities.

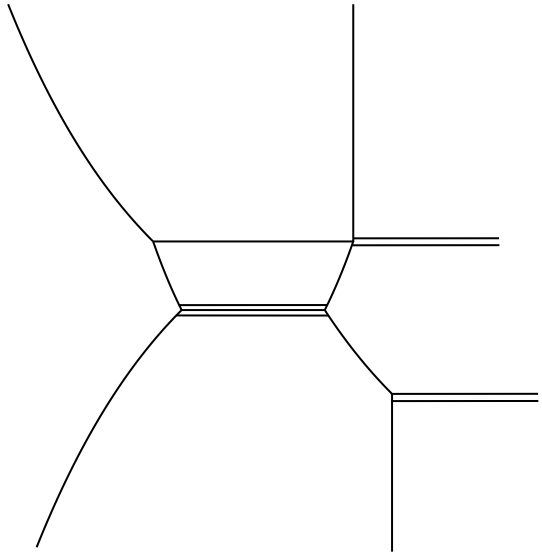
The M5 brane corresponding to a given configuration of D4s and NS5s in type IIA should give the latter back in the limit where R becomes small, and if one forgets about the circle coordinate x^{10}.

Moreover, in order for the M5 brane to preserve one-fourth of the 32 supersymmetries in M-theory so as to obtain a $4d$ $\mathcal{N}=2$ theory at low energies, Σ has to be a Riemann surface in Q in the complex structure in which x^4+ix^5 and x^6+ix^{10} are holomorphic.

The surface Σ is hence given by an equation $F(t,v)=0$ in Q, where $t=\exp\bigl(x^6+ix^{10}\bigr)$. The degree of F in t and v are related to the number of fourbranes and fivebranes in the type IIA configuration. The very nice outcome of all this construction is that these Riemann surfaces are nothing but the Seiberg–Witten curves introduced in Section 5.6, which contain much information about the low-energy effective field theory on the Coulomb branch of the D4 worldvolume theory. Some examples from [Wit97] are given in Figure 6.6, where e and f are constants (that can be removed by rescaling t and v) and where μ and the μ_is are the v-positions of the semi-infinite D4 branes, i.e. the masses of the flavor hypermultiplets in the language of field theory.



SU(2), $N_f=1$



SU(4), $N_f=4$

$$(\Sigma)\ :\ \left(t+\frac{B(v)}{2}\right)^2=\frac{B(v)^2}{4}-f(v-\mu) \qquad B(v)=e(v^2+u)$$

$$(\Sigma)\ :\ \left(t+\frac{B(v)}{2}\right)^2=\frac{B(v)^2}{4}-f\prod_{i=1}^{4}(v-\mu_i) \qquad B(v)=e(v^4+u_2v^2+u_3v+u_4)$$

Figure 6.6: Type IIA brane setups and equations of the corresponding M-theory curve.

6.4.2 Five-dimensional theories with eight supercharges

Let us now consider configurations of fivebranes in type IIB theory studied in [AH97, AHK98]:

	0	1	2	3	4	5	6	7	8	9
D5	–	–	–	–	–	–	×	0	0	0
NS5	–	–	–	–	–	×	–	0	0	0

(6.62)

It preserves one-fourth of the original 32 supersymmetries of type IIB theory and everything that matters about the setup is described in the 56-plane. One can also add (p,q)-fivebranes to such a configuration (which are bound states of p D5s and q NS5s) without breaking further the supersymmetry, provided that a (p,q)-fivebrane has slope $[p:q]$ in the 56-plane.

D5 branes can end on NS5s however if they do their endpoints bend the NS5s. The relevant equation describing the backreaction is now a one-dimensional Laplace equation and hence the brane configuration remains piece-wise linear. An example of such a backreaction induced by the endpoint of a $(1,0)$ brane inside a $(0,1)$ brane is shown in Figure 6.7: there is a half-infinite $(1,1)$ fivebrane emanating out of the vertex. Actually, any junction of arbitrary (p_i,q_i)-fivebranes is allowed if it preserves the fivebrane

charges:

$$\sum_i p_i + \sum_i q_i = 0 \ ; \tag{6.63}$$

they do not break the supersymmetry further.

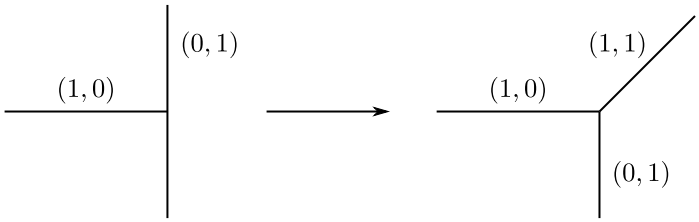


Figure 6.7: The back reaction of a D5 endpoint inside an NS5 (as seen in the 56 plane).

The projection of such a fivebrane configuration in the 56-plane is called a (p, q)-*web*. At low energies the worldvolume theory on the D5s is a five-dimensional gauge theory with $\mathcal{N} = 1$ supersymmetry. One can study the Coulomb branch of the moduli space and the low-energy BPS states on it using (p, q)-webs [AHK98], as well as the Higgs branches (to a lesser extend). More recent work has focused on Higgs branches of such theories and generalizations, starting with [CHS19, BCG+20]. Three examples of (p, q)-webs are displayed in Figure 6.8. The D5 worldvolume theory on the leftmost (resp. the one in the middle, the rightmost) diagram describes a point on the Coulomb branch of a $5d$ $\mathcal{N} = 1$ pure SU(2) (resp. SU(2) with 3 flavors, SU(2) $\times$ SU(3) gauge theory with 3 flavors) gauge theory.

The geometry of these diagrams can be linked to the parameters in the Lagrangian of the corresponding field theory. For example, in the left-most diagram in Figure 6.8 m_W is the mass of the W-boson while $m_0 = g_0^{-2}$ with g_0 the bare gauge coupling.

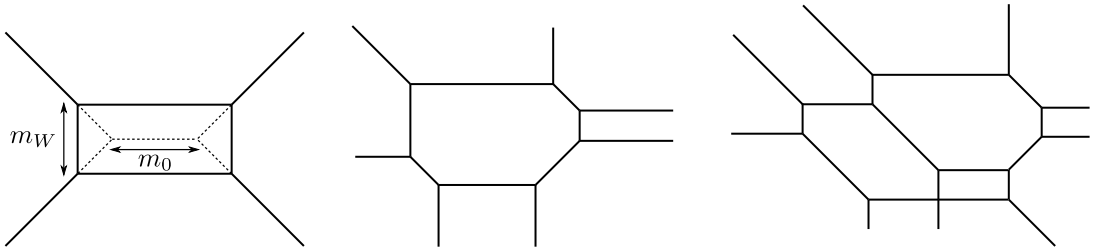


Figure 6.8: Some (p, q)-webs (as seen in the 56 plane).

With the hope of smoothing the singularities at the vertices one can look for an M-theoretic lift of (p, q)-webs. By combining a T-duality with the M-theory limit of type IIA theory one obtains a duality between type IIB theory compactified on a circle of length L_B and M-theory compactified on a torus with a base of length $2\pi L_t$ and modular parameter $\tau = \tau_{\text{IIB}}$ the axion-dilaton field of type IIB theory. Then:

$$\frac{1}{2\pi l_s^2} = 2\pi L_t T_{\text{M2}} \ , \qquad L_B = \frac{1}{2\pi L_t^2 \text{Im}(\tau) T_{\text{M2}}} \ , \tag{6.64}$$

with T_{M2} the M2-brane tension given in Equation (6.48). Let x_t and y_t be the coordinates on the M-theory torus. The (p, q)-web lifts in M-theory to a single M5-brane wrapping a Riemann surface Σ in the complex split torus $(\mathbb{C}^\times)^2$ parameterized by

$$s = \exp\left(\frac{x^5 + ix_t}{L_t}\right), \qquad t = \exp\left(\frac{y^6 + iy_t}{L_t}\right) \ . \tag{6.65}$$

As before, the fact that Σ is a complex curve ensures that 8 supercharges are preserved. Moreover, the analytic expression of Σ can be read easily from the (p, q)-web. The latter can indeed be considered as the dual graph of a triangulated lattice polygon, which is the Newton polygon of the analytic expression of Σ: this is shown on the left of Figure 6.9 for the web in the middle of Figure 6.8. Doing the same for the web on the left of Figure 6.8 yields the following expression for the M-theory curve:

$$(\Sigma): \quad P(s, t) = As + t + ABst + Ast^2 + s^2 t = 0 \ , \tag{6.66}$$

where the coefficients are linked to the physical parameters of the theory as:

$$A \sim 2\exp\left(\frac{L_B}{2g_0^2}\right), \qquad B \sim 2\exp\left(\frac{m_W L_B}{2}\right). \tag{6.67}$$

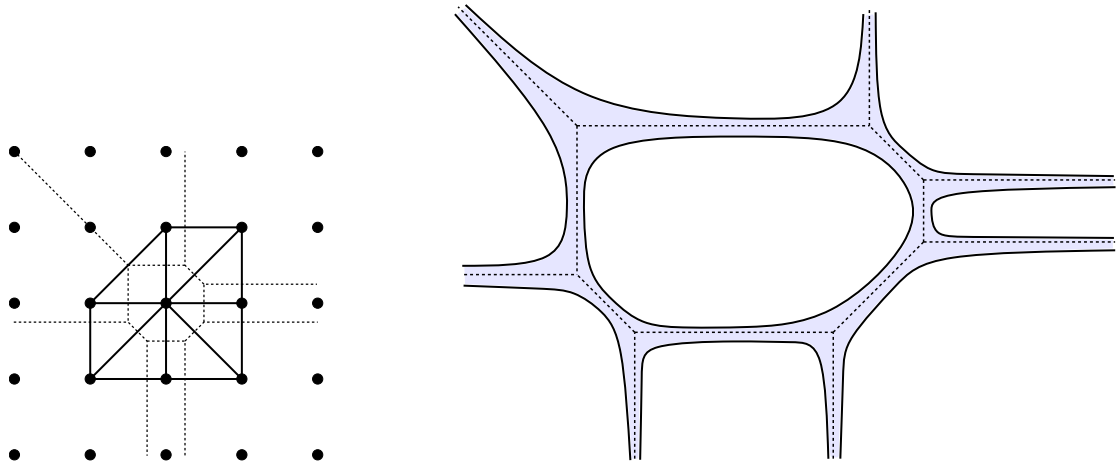


Figure 6.9: The lattice polygon dual to a (p,q)-web and the amoebae of a complex curve in $(\mathbb{C}^\times)^2$.

The link between the M-theory curve and its corresponding web diagram can be formalized as follows. Given any affine complex curve defined by a polynomial $P(s,t)=0$ in $(\mathbb{C}^\times)^2_{s,t}$, its *amoeba projection* is its image under the map

$$\begin{array}{ccc}(\mathbb{C}^\times)^2 & \longrightarrow & \mathbb{R}^2 \\ (s,t) & \longmapsto & (\ln|s|, \ln|t|)\ .\end{array} \tag{6.68}$$

These projections of complex curves onto a real plane have a characteristic shape explaining the name amoeba: the amoebae projection of the curve of corresponding to the web in the middle of Figure 6.8 is shown in blue on the right of Figure 6.9 (for some wise choice of parameters).

Since $\ln|s| = L_t^{-1}x^5$ and $\ln|t| = L_t^{-1}x^6$, the amoeba can be considered as a subspace of the 56-plane in type IIB theory. Drawing the original (p,q)-web on top of the amoeba makes it clear that the latter is a fattened version of the former: the M-theory curve indeed smooths the (p,q)-web.

As $L_t \to 0$ (which corresponds to $L_B \to \infty$) the amoeba shrinks to the (p,q)-web, which is to be identified with the tropicalization of Σ, i.e. the non-smooth locus of:

$$(\Sigma^t): \quad \max\left(\frac{1}{2g_0^2}+x^5, x^6, \frac{1}{2g_0^2}+\frac{m_W}{2}+x^5+x^6, \frac{1}{2g_0^2}+x^5+2x^6, 2x^5+x^6\right). \tag{6.69}$$

The lift of the (p,q)-web to M-theory is an incarnation of *Maslov dequantization*, discussed at the end of Section 1.3.

6.4.3 Four dimensional theories with four supercharges

As a last example of how brane configurations can be used to study the quantum behaviour of supersymmetric field theories, let us consider the following setup in type IIA theory as in [EGK97b, EGK+97a]:

	0	1	2	3	4	5	6	7	8	9
NS5	–	–	–	–	–	–	x_1^6	0	0	0
N D4	–	–	–	–	0	0	–	0	0	0
F D6	–	–	–	–	x_j^4	x_j^5	x_j^6	–	–	–
NS5′	–	–	–	–	0	0	x_2^6	0	–	–

(6.70)

An example is shown in Figure 6.10.

This configuration preserves one eighth of the 32 supersymmetries of type IIA, and hence the world-volume theory on the D4-branes at low energies is an $\mathcal{N}=1$ four-dimensional quantum field theory.

In Figure 6.10 the N D4s stretching between the NS5 and the NS5' give rise to an SU(N) gauge group with gauge coupling g. The separation between the two NS5-branes along the x^6 direction is proportional

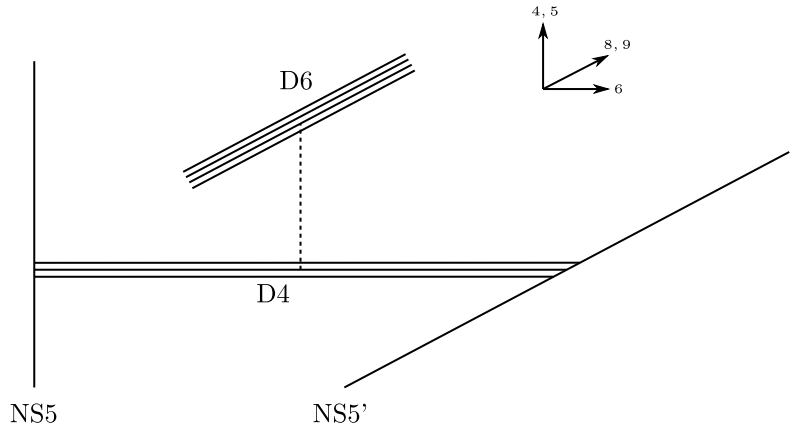


Figure 6.10: A configuration of branes in type IIA string theory.

to g^{-2}. Each D6 brane gives rise to a flavor of quarks, i.e. a pair of chiral multiplets in the fundamental and antifundamental representations of SU(N). The distance between a D6 and the stack of D4-branes is proportional to the bare mass of the corresponding flavor, and hence when the F D6-branes are located at $x^4 = x^5 = 0$ the low-energy theory on the N D4s is (massless) SQCD with N colors and F flavors.

If brane setups are a good tool to study the quantum dynamics of supersymmetric field theories, we should be able to understand Seiberg duality in terms of branes, and it is the case. To this end, let us emphasize that when a stack of D6 branes is at $x^4 = x^5 = 0$ between the NS5 and the NS5', it is possible to split the stack of D4 branes on the D6s and move the part between the D6s and the NS5' along x^8 and x^9. This corresponds to giving a non-zero vacuum expectation value to one or more quarks in the low-energy effective field theory.

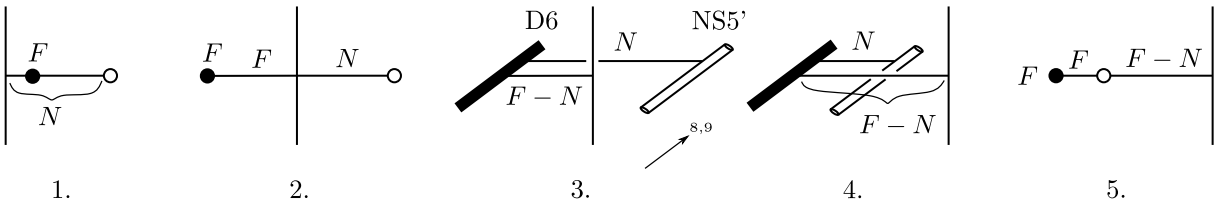


Figure 6.11: Seiberg duality from branes.

Seiberg duality is described in five successive steps displayed in Figure 6.11. For $F > N$:

1. In the starting configuration, there are N D4s stretching between the NS5 and the NS5', as well as F D6s at $x^4 = 0$, $x^5 = 0$ and some x^6 such that $x_1^6 < x^6 < x_2^6$.

2. One moves the NS5-brane along x^6 and through the F D6s. This creates F D4s stretching between the D6s and the NS5 (this is the Hanany-Witten effect).

3. One splits the stack of *D4s* along x^8 and x^9. There are now N D4s at some non-zero (x^8, x^9) stretching between NS5' and the D6s, and $F - N$ D4s at $(x^8, x^9) = (0, 0)$ stretching between the D6s and the NS5.

4. One moves the latter set of $F - N$ D4s and NS5 along x^7, keeping the D4s attached to the $F - N$ D6s. Now that NS5 and NS5' are separated in x^7, one can move NS5 along x^6 to the right of NS5' in a smooth way since the branes do not intersect at any point, and then move back the stack of $F - N$ D4s and NS5 back to their initial x^7 position.

5. Lastly, one moves back the stack of N branes between the N D6s and NS5' to their initial x^6 position. The $F - N$ D4s between NS5 and the D6s split on NS5', and one is left with $F - N$ D4s stretching between NS5' and NS5, and F D4s streching between the F D6s and NS5'. This latter stack of D4s can move along x^8 and x^9, and this gives rise to F^2 gauge neutral fields Φ_j^i in the low-energy theory. World-sheet instantons induce the superpotential of the magnetic theory.

* * * * * * *

Type II string theories can be defined perturbatively in flat, ten-dimensional space-time $\mathbb{R}^{1,9}$. Type IIA (resp. IIB) theory is non-chiral (resp. chiral) in space-time, and its massless modes describe a dilaton, a graviton and a 2-form in the NS–NS sector, two gravitinos in the NS–R and R–NS sectors, as well as differential forms of odd degree (resp. even degree) in the R–R sector. At low-energies, type II superstring theories are well-described by type II supergravity theories.

Type II string theories describe branes, which are of a non-perturbative nature. D-branes naturally couple electrically and magnetically to the R–R fields of the theory. Therefore type IIA theory describes D0, D2, D4, D6 and D8-branes, whereas type IIB contains D(-1), D1, D3, D5, D7 and D9-branes. Both type II theories contain NS5-branes which couple magnetically to the NS–NS 2-form. Moreover, both theories are related through T-duality when compactified on a circle. Type IIB theory is auto S-dual, which notably implies the existence of bound states of 1-branes and 5-branes known as (p, q)-strings and (p, q)-fivebranes. Analyzing type IIA theory suggests that at strong coupling it is better described as an eleven-dimension theory known as M-theory. At low energies, the latter is well described by eleven-dimensional supergravity, whose bosonic degrees of freedom are a graviton and a three-form. This points towards the existence of M2-branes and M5-branes in M-theory.

Branes are of prime interest since string theory implies the existence of massless fields in their world-volume. In particular, D-branes hosts supersymmetric Yang–Mills theories with 16 supercharges. One can consider brane configurations which preserve supersymmetries; this allows the engineering of arbitrarily complicated supersymmetric quantum field theories in various dimensions. The quantum dynamics of the latter can be studied through the geometry of the corresponding brane configurations.

Chapter 7

String geometry

We are now going to study type II string theories and M-theory in non-trivial space-times. There are two main reasons why one would want to do this.

- In order to connect string theories and M-theory to the real world which appears to be four-dimensional at low energies, one can assume that the critical dimensions of superstring theories or M-theory are shaped as $\mathbb{R}^{1,3} \times X_6$ for string theories or $\mathbb{R}^{1,3} \times X_7$ for M-theory, where X_6 and X_7 are compact metric spaces of typical size l_X. The physics at energies smaller than l_X^{-1} is then effectively four-dimensional: one speaks of *compactification*. In order to keep some control on the low-energy theory it is of interest to preserve some of the supersymmetries. This imposes stringent constraints on X_6 or X_7: for example X_6 needs to be a *Calabi–Yau threefold*. Since superstring theories and M-theory are well approximated at low energies by supergravity theories which have differential forms as dynamical fields, one might either consider compactifications in which these forms are trivial (Section 7.1), or not (Section 7.2).

- The massless spectrum of open strings on D-branes at singularities can be very rich. By choosing the singularity appropriately one can construct complicated gauge theories and study their dynamics using string/M theory, which is usually referred to a *geometric engineering*. Again, it is of interest to preserve some supersymmetries, in which case the gauge theories one obtains are supersymmetric. In Section 7.3 we first study how the worldvolume theory on D3 branes changes under the mild replacement of the transverse flat space $\mathbb{C}^3$ that they have in flat space with *Calabi–Yau orbifolds* of $\mathbb{C}^3$. Section 7.4 presents some aspects of *toric geometry* and the construction of (complex) three-dimensional affine toric Calabi–Yau singularities. The worldvolume theory of D3-branes whose transverse space is an affine toric CY3 singularity is derived in Section 7.5 with the help of *brane tilings*. Lastly, we introduce *orientifolds* in Section 7.6.

7.1 Compactifications without fluxes

Let us assume that the ten-dimensional target space M_{10} of a superstring theory is the direct product:

$$M_{10} = M_4 \times M_6 \ . \tag{7.1}$$

We require M_4 to be a maximally symmetric i.e. homogeneous and isotropic four dimensional Lorentzian space-time (which implies in particular that the curvature tensor is proportional to the scalar curvature) and M_6 to be a six-dimensional Riemannian spin manifold. Let us also assume that the NS–NS three-form field as well as the dilaton field vanish identically, both in M_4 and M_{10}. The metric also splits as:

$$\mathrm{d}s^2_{M_{10}} = \mathrm{d}s^2_{M_4} + \mathrm{d}s^2_{M_6} \tag{7.2}$$

If one believes that physics is supersymmetric above some energy scale $\Lambda \ll M_P$ where M_P is the Planck scale, it is of interest to investigate the conditions that M_4 and M_6 must satisfy in order to preserve some of the supersymmetries of string theories in flat space-time. However, as we saw in Chapter 5: the more supersymmetric a model is, the less phenomenological.

A vacuum state is supersymmetric if and only if the SUSY variations of the fermionic fields vanish. The low-energy supergravity approximation always contain a gravitino, whose SUSY variation when the NS–NS three-form is zero can vanish only if there exists a *covariantly constant spinor* ϵ in M_{10}:

$$\nabla_{M_{10}}\epsilon = 0 \ . \tag{7.3}$$

Since we have assumed that M_{10} is a direct product, a covariantly constant spinor ϵ can be written as a tensor product $\zeta \otimes \eta$ with ζ on M_4 and η on M_6.

The fact that the maximally symmetric space M_4 has a covariantly constant spinor implies that it is flat, and hence that $M_4 = \mathbb{R}^{1,3}$. As far as M_6 is concerned, the existence of a covariantly constant spinor implies that:

1. The holonomy of M_6 must be in an SU(3) subgroup of $\mathrm{Spin}(6) \simeq \mathrm{SU}(4)$,

2. M_6 must be a complex manifold, with complex structure

$$J^n_n = i\eta_+^\dagger \gamma^n_m \eta_+ \ , \tag{7.4}$$

   where (η_+, η_-) are the two chiral spinors constituting η and where the γ_m are the gamma matrices on M_6, which implies that M_6 is in fact a Kähler manifold,

3. The three-form

$$\Omega = \frac{1}{6}\Omega_{abc}\mathrm{d}z^a \wedge \mathrm{d}z^b \wedge \mathrm{d}z^c \tag{7.5}$$

   is a nowhere vanishing holomorphic three-form on M_6.

One says that M_6 is a *Calabi–Yau threefold* (CY3). When M_6 is compact, it follows from the Calabi conjecture proved by Yau in 1977–1978[1] that if ω is a Kähler form on M_6 there exists a unique Ricci-flat Kähler metric whose Kähler form is in the same cohomology class as ω.

In general one says that a compact Kähler manifold X of complex dimension n is Calabi–Yau is the following equivalent assumptions hold:

- X has a trivial canonical bundle,
- X has a nowhere vanishing holomorphic n-form,
- X has a Kähler metric with global holonomy in $\mathrm{SU}(n)$.

This implies that X has a Kähler metric with vanishing Ricci curvature.

Calabi–Yau manifolds have been studied extensively since it became clear that they play a central role in string theory compactifications. The CY3 studied in string theory are usually assumed to be simply connected, and this together with Serre duality, complex conjugation and Poincaré duality implies that the *Hodge diamond* of a compact CY3 is of the form:

$$\begin{array}{ccccccc} & & & 1 & & & \\ & & 0 & & 0 & & \\ & 0 & & h^{1,1} & & 0 & \\ 1 & & h^{2,1} & & h^{2,1} & & 1 \\ & 0 & & h^{1,1} & & 0 & \\ & & 0 & & 0 & & \\ & & & 1 & & & \end{array} \ , \tag{7.6}$$

i.e. only depends only on $h^{1,1}$ and $h^{2,1}$. One way to construct many compact Calabi–Yau threefolds is by using Batyrev's construction [Bat94] in toric geometry. It implies in particular that one can construct a compact CY3 for each four-dimensional lattice reflexive polyhedron. These have been classified modulo $\mathrm{SL}_4(\mathbb{Z})$ [KS00b]: there are $473,800,776$ such polyhedra. This gives a lower bound on the number of topologically distinct compact CY3, however examples of non-toric compact CY3 are also known. We refer to [BHHP20] for a tour in the beautiful landscape of Calabi–Yau threefolds.

[1] He was awarded the Fields medal in 1982 in part for this proof.

Any compact CY3 X has a moduli space which is locally a product:

$$\mathcal{M} = \mathcal{M}^{2,1} \times \mathcal{M}^{1,1} \ , \tag{7.7}$$

where $\mathcal{M}^{2,1}$ is the moduli space of complex structures on X, and $\mathcal{M}^{1,1}$ the moduli space of Kähler structures on X. Compactifying either type II string theory on such a compact CY3 yields an effectively four dimensional $\mathcal{N} = 2$ supersymmetric theory with moduli fields which parametrize flat directions in the moduli space. Compactifying type IIA theory on X yields $h^{1,1}$ abelian $\mathcal{N} = 2$ vector multiplets moduli and $(h^{2,1} + 1)$ hypermultiplets moduli. Conversely, compactifying type IIB theory on X yields $h^{2,1}$ abelian vector multiplets moduli and $h^{1,1} + 1$ hypermultiplets moduli. This lead to the *mirror symmetry conjecture*: for each compact CY3 X with Hodge numbers $h^{1,1}_X$ and $h^{2,1}_X$ there exists a compact CY3 Y such that $h^{1,1}_Y = h^{2,1}_X$ and $h^{2,1}_Y = h^{1,1}_X$, and such that type IIA string theory compactified on X is equivalent to type IIB string theory compactified on Y. A famous piece of evidence for the conjecture is the manifest reflection symmetry of Figure 1 in [CLS90] showing the plot of $h^{1,1}+h^{2,1}$ versus $2(h^{1,1}-h^{2,1})$ for a large family of Calabi-Yau threefolds.

A deep generalization of this original mirror symmetry conjecture was proposed by Kontsevich in [Kon95]: for any $2n$ symplectic variety (V, ω) and its mirror W (an n-dimensional complex manifold), the derived Fukaya category $F(V)$ on V should be equivalent to the derived category of coherent sheaves on W. This conjecture can be expressed in physical terms (when $V = X$ and $W = Y$ are Calabi–Yau manifolds) as the correspondence between euclidean BPS branes on X in type IIA and euclidean BPS branes on Y in type IIB. These BPS branes are well described by the topological string models B and A. B-branes on X are described by the derived category on coherent sheaves on X while A-branes on Y are described by the Fukaya category on Y [Dou00, Dou01, AL01, Laz01, Dia01, Asp04].

Another conjecture (or philosophy) has been proposed by Strominger, Yau and Zaslow in [SYZ96]. It explains mirror symmetry between compact Calabi–Yau n-folds as a sequence of n T-dualities.

Riemannian manifolds of dimension greater than three with restricted holonomy are also potentially useful to construct string compactifications. Seven dimensional manifolds with G_2 holonomy break one eight of the 32 supersymmetries of M-theory in flat space-time, so that M-theory compactified on such a manifold yields a four dimensional $\mathcal{N} = 1$ quantum field theory. Eight dimensional manifolds with Spin(7) holonomy break one sixteenth of the 32 supersymmetries of M-theory in flat space-time, so that M-theory compactified on such a manifold yields a three dimensional $\mathcal{N} = 1$ quantum field theory.

7.2 Flux compactifications

Compactifications of string theories or M-theory on manifolds with restricted holonomy always yield moduli, i.e. massless scalar fields without potential in the low-energy effective field theory. This is problematic from a phenomenological point of view, in particular because physical parameters of the low-energy effective field theory such as the couplings depend on the value of these moduli.

One possible approach towards the resolution of such an issue is to consider warped compactifications of string theories or M-theory. For definiteness we will discuss type IIB superstrings on backgrounds which are locally of the form:

$$M_{10} = M_4 \times M_6 \ , \tag{7.8}$$

with a compact M_6 and a warped Einstein frame metric:

$$\mathrm{d}s^2_{M_{10}} = e^{2A(y)}\eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu + e^{-2A(y)}\tilde{g}_{mn}\mathrm{d}y^m \mathrm{d}y^n \ , \tag{7.9}$$

where x^μ are the coordinates on M_4 and y^m the ones on M_6, and where $\eta_{\mu\nu}$ is the four-dimensional Minkowski metric. The warp factor $A(y)$ only depends on the coordinates on the internal manifold; this is a consequence of requiring $4d$ Poincaré invariance. A no-go theorem [MN01] shows that without sources as branes or singularities in the internal geometry, any flat space solution has its warp factor constant and vanishing fluxes. Moreover, de Sitter solution are even completely excluded under these hypotheses. This comes from considering the external part of the ten dimensional Einstein equation, which in the warped metric ansatz of Equation (7.9) leads to an equality between the Lagrangian of e^{4A} and a sum of positive-definite terms, which thus have to vanish.

More precisely, let us write the type IIB supergravity action in the Einstein frame as in [GKP02]:

$$S_{\mathrm{IIB}} = \frac{1}{2\kappa^2}\int \mathrm{d}^{10}x\sqrt{-G}\left[R - \frac{|\partial\tau|^2}{2(\mathrm{Im}\tau)^2} - \frac{|G_3|^2}{2\mathrm{Im}\tau} - \frac{|\tilde{F}_5|^2}{4}\right] + \frac{1}{8i\kappa^2}\int \frac{C_4 \wedge G_3 \wedge \overline{G_3}}{\mathrm{Im}\tau} + S_{\mathrm{loc}} \ , \tag{7.10}$$

where $G_3 = F_3 - \tau H_3$ is the combined three-flux, $\tau = C_0 + ie^{-\phi}$ is the axion-dilaton field,

$$\tilde{F}_5 = F_5 - \frac{1}{2} C_2 \wedge H_3 + \frac{1}{2} B_2 \wedge F_3 \ , \tag{7.11}$$

and where $S_{\rm loc}$ is the action of localized objects. Poincaré invariance and the Bianchi identity impose

$$\tilde{F}_5 = (1 + \star_{10}) \left[\mathrm{d}\alpha \wedge \mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3 \right] \ , \tag{7.12}$$

with α a function on M_6. From the ten dimensional Einstein equation one obtains:

$$\Delta \left(e^{4A} - \alpha \right) = \frac{1}{6 \mathrm{Im} \tau} e^{8A} \left| iG_3 - \star_6 G_3 \right|^2 + e^{-4A} \left| \partial (e^{4A} - \alpha) \right|^2 + 2\kappa^2 e^{2A} (\mathcal{J}_{\rm loc} - T_3 \rho_3^{\rm loc}) \ , \tag{7.13}$$

where

$$\mathcal{J}_{\rm loc} = \frac{1}{4} \left(\sum_m \left(T^{\rm loc} \right)^m_m - \sum_\mu \left(T^{\rm loc} \right)^\mu_\mu \right) \tag{7.14}$$

with $T^{\rm loc}$ the energy-momentum tensor associated with the localized sources and ρ_3 the D3 charge density from localized sources. Moreover fluxes are quantized, because of the generalized Dirac quantization conditions. Working out the conditions for supersymmetric solutions yields the following constraints:

- G_3 is an imaginary-self dual $\star_6 G_3 = iG_3$ primitive $(2,1)$-form,
- The warped factor and the four-form potential are related through $e^{4A} = \alpha$,
- The localized sources saturate the bound $\mathcal{J}^{\rm loc} \geq T_3 \rho_3^{\rm loc}$.

The authors of [GKP02] provide explicit examples of this construction and show that one can obtain hierarchies of scales in these warped compactifications, generalizing the brane-world models of Randall and Sundrum [RS99a, RS99b].

One of the nice outcomes of these compactifications with non-trivial fluxes is that the latter generate a superpotential for some number of moduli fields, which leads to their stabilization. This is a step towards building phenomenologically interesting compactifications of string theories. Moreover this construction can even lead to four dimensional de Sitter space-times as shown by the *KKLT construction* [KKLT03], and with a cosmological constant small enough so that it is compatible with Weinberg's anthropic bound [Wei87]:

$$-10^{120} M_P^4 < \Lambda < 10^{-118} M_P^4 \ , \tag{7.15}$$

as shown in [BP00]. This leaves however a huge number of possible compactifications with fluxes – at least 10^{500} for type IIB superstrings on some Calabi–Yau manifolds, which form the so-called *landscape of string vacua*. One speaks of a discretuum of vacua. The statistical study of the landscape [BP00, Dou03] has shed much light on deep questions such as the fate of supersymmetry (see [DOS05] for example) and also the status of intelligent design and anthropic principles (with popularization books such as [Sus05]).

The swampland program.

The string theory landscape contains quantum field theories which by definition can be UV-completed to consistent theories of quantum gravity, since they are effective field theories obtained as flux compactifications of string theories. In contrast, the *swampland program* [Vaf05] aims to derive criteria that quantum field theories must satisfy in order to be compatible with quantum gravity. More prosaically, the question is: given some quantum field theory considered as an effective theory, is it compatible at all with the precepts of quantum gravity?

There are many swampland conjectures, most of which were derived from string-independent lines of thought based on developments on quantum black holes and black holes evaporation, but which were tested in many non-trivial ways in the framework of string theory, which is the better developed theory of quantum gravity at hand, up to today. Some early examples of the swampland conjectures are:

- There is no global symmetry in quantum gravity [BD88, BS11],
- A gauge theory coupled to gravity contains all the states allowed by Dirac quantization [Pol04],

- In a theory with gauge group U(1) and gauge coupling g coupled to gravity, there must exist a particle in the theory with mass m and charge q such that $m \leq \sqrt{2}gqM_P$, where M_P is the Planck mass. Moreover, the cutoff scale of the theory satisfies $\Lambda \leq gM_P$ [AHMNV07].

Other conjectures were proposed more recently such as the distance conjecture [OV07], the non-supersymmetric AdS conjecture [OV17], the dS conjecture [OOSV18] or the cobordism conjecture [MV19].

The status of string theory in such conjectures being central was erected as the string lamppost principle [MV21], and it is a question of interest to understand to what extent string theory is universal, i.e. is able to describe every consistent quantum gravity theory.

Let us note before closing this section that the KKLT construction of de Sitter vacua in string theory has led to many debates concerning its validity, and two decades after [KKLT03] there is still no global consensus nowadays on whether it is possible at all to construct de Sitter vacua in string theory [DVR18].

7.3 Branes at abelian Calabi–Yau orbifolds of $\mathbb{C}^3$

Now that we are into studying superstring theories in topologically non-trivial backgrounds, let us discuss D-branes at non-compact Calabi–Yau singularities. These backgrounds and non-compact and hence cannot be used as the internal manifolds of the previous section however they play a tremendous role in brane-world models and in holography, as we will see in the next chapter. Here we mainly follow [DM96, DGM97] as well as [Ber03, Chapter 1]. Since we are primarily interested in four-dimensional gauge theories we will focus on D3-branes at singularities in type IIB superstring theory.

7.3.1 Regular D3-branes

We will begin with arguably the simplest orbifold of $\mathbb{C}^3$, namely $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$, which is the quotient of $\mathbb{C}^3$ with its flat metric under the following action of $\mathbb{Z}_2 = \{1, g\}$:

$$g \cdot (z_1, z_2, z_3) = (-z_1, -z_2, z_3) \ , \tag{7.16}$$

where $(z_1, z_2, z_3) = (x^4 + ix^5, x^6 + ix^7, x^8 + ix^9)$. The points $(0, 0, z_3)$ with $z_3 \in \mathbb{C}$ are singular, and so would be a ten dimensional quantum field theory in $\mathbb{R}^{1,3} \times \mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$. String theory however is different in that its extended objects (strings) can wind around singular points. We will be particularly interested in the spectrum of branes at singularities, hence let us consider a single ordinary D3-brane extended along $\mathbb{R}^{1,3}$ and placed at $(0, 0, 0) \in \mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$.

While the Chan–Paton factors for open string states ending on single D3 brane in flat space $\mathbb{C}^3$ would be trivial, when the D3 brane is placed at $(0, 0, 0) \in \mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$, an open string state can begin and end on the D3 in different ways: either it winds once around $(0, 0) \in \mathbb{C}^2/\mathbb{Z}_2$ or it does not. In order to study this is is convenient to move the D3 a tiny bit away from the singular points and to view the resulting configuration not in $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$ but in the universal cover $\mathbb{C}^3$, in which there are now two D3-branes D and D' identified under the action of $\mathbb{Z}_2$ [DM96]. Open strings can then be assigned Chan–Paton factors which are 2×2 matrices and which describe open-string states in the following way:

$$\lambda = \begin{pmatrix} D - D & D - D' \\ D' - D & D' - D' \end{pmatrix} \ . \tag{7.17}$$

The non-trivial element g of $\mathbb{Z}_2$ naturally acts exchanges the branes D and D', and hence:

$$\rho_{\text{reg}}(g) \cdot \begin{pmatrix} D + D' \\ D - D' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} D + D' \\ D - D' \end{pmatrix} \ , \tag{7.18}$$

so that the vector space spanned by D and D' is the regular representation of $\mathbb{Z}_2$. The Chan–Paton factors transform as $\lambda \to \rho(g)\lambda\rho(g)^{-1}$ [GP96, DM96]. In order to preserve supersymmetry, $\mathbb{Z}_2$ must act on the world-sheet fermions exactly as on the bosons: g maps ψ^μ to ψ^μ when $\mu = 0, 1, 2, 3, 8, 9$ and to $-\psi^\mu$ when $\mu = 4, 5, 6, 7$.

An open string state $\lambda\psi^\mu |k\rangle$ is preserved by the orbifold projection if the combined action of $\mathbb{Z}_2$ on λ and on ψ^μ is the identity. In the case at $\lambda\psi^\mu |k\rangle$ is preserved with diagonal Chan-Paton factors when

$\mu = 0, 1, 2, 3, 8, 9$ and anti-diagonal ones when $\mu = 4, 5, 6, 7$. Placing back the D3 brane at $(z_1, z_2, z_3) = (0, 0, 0)$, the massless Neveu–Schwarz open string states thus write:

$$A^\mu = \begin{pmatrix} A_1^\mu & 0 \\ 0 & A_2^\mu \end{pmatrix} , \quad \Phi^3 = \begin{pmatrix} Z_{11} & 0 \\ 0 & Z_{22} \end{pmatrix} , \quad \Phi^1 = \begin{pmatrix} 0 & X_{12} \\ X_{21} & 0 \end{pmatrix} , \quad \Phi^2 = \begin{pmatrix} 0 & Y_{12} \\ Y_{21} & 0 \end{pmatrix} , \tag{7.19}$$

and the Ramond sector furnishes the corresponding space-time fermions so as to have supersymmetry.

The worldvolume theory of stack of N D3 branes at $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$ is a four dimensional $\mathcal{N} = 2$ gauge theory with gauge group $\mathrm{U}(N)_1 \times \mathrm{U}(N)_2$ and two hypermultiplets $(X_{12} \oplus Y_{21})$ and $(Y_{12} \oplus X_{21})$ in the representations $(\overline{\square}_1, \square_2)$ and $(\overline{\square}_2, \square_1)$. More precisely, the low-energy effective theory has gauge group $\mathrm{SU}(N)_1 \times \mathrm{SU}(N)_2$: the diagonal U(1) factor corresponds to the center-of-mass mode of the stack of D-branes and is completely decoupled in the infrared, while the anti-diagonal one is infrared-free and becomes a global U(1)-symmetry of the low-energy theory. There is a low-energy superpotential which descends from the one of the world-volume theory of D3-branes in flat space:

$$W_{\text{flat}} = \Phi^3[\Phi^1, \Phi^2] . \tag{7.20}$$

Plugging in the expression the fields of Equation (7.19) one obtains:

$$W = Z_{11}X_{12}Y_{21} - Z_{11}Y_{12}X_{21} + Z_{22}X_{21}Y_{12} - Z_{22}Y_{21}X_{12} . \tag{7.21}$$

The (mesonic) moduli space of this gauge theory when $N = 1$ and without Fayet–Illiopoulos parameters is exactly the singularity $\mathbb{C}^2/Z_2 \times \mathbb{C}$. The vacuum expectation values of the scalar fields Z_{11} and Z_{22} in the vector multiplet are associated with displacements along the flat directions x^8 and x^9, whereas the vacuum expectation values of the single trace mesonic operators formed out of the X_{ij} and Y_{ij} are associated with displacements along the orbifold directions $x^4, \ldots, x^7$: the F-terms impose $X_{12}Y_{21} = Y_{12}X_{21}$ and one can choose $a = X_{12}X_{21}$, $b = Y_{12}Y_{21}$ and $c = X_{12}Y_{21}$ as a set of independent mesonic operators. One sees that $ab = c^2$, which is the equation defining $\mathbb{C}^2/\mathbb{Z}_2$ in $\mathbb{C}^3$.

This generalizes straightforwardly to any Calabi–Yau abelian orbifold of $\mathbb{C}^3$: let us consider the example of $\mathbb{C}^3/Z_5$ with the following action of a generator g_5 of $\mathbb{Z}_5$:

$$g_5 \cdot (z_1, z_2, z_3) = (g_5 z_1, g_5^2 z_2, g_5^2 z_3), \qquad g_5 = e^{\frac{2\pi i}{5}} . \tag{7.22}$$

Under the action of g_5 one has $\psi^{0,1,2,3} \to \psi^{0,1,2,3}$, $\psi^{4,5} \to g_5\psi^{4,5}$ and $\psi^{6,7,8,9} \to g_5^2\psi^{6,7,8,9}$. The Chan-Paton factors now form a 5×5 matrix on which the diagonalized $\rho_{\text{reg}}(g_5)$ acts as:

$$\lambda \longrightarrow \begin{pmatrix} 1 & & & & \\ & g_5 & & & \\ & & g_5^2 & & \\ & & & g_5^3 & \\ & & & & g_5^4 \end{pmatrix} \lambda \begin{pmatrix} 1 & & & & \\ & g_5 & & & \\ & & g_5^2 & & \\ & & & g_5^3 & \\ & & & & g_5^4 \end{pmatrix}^{-1} . \tag{7.23}$$

The massless Neveu–Schwarz open string states which survive the orbifold projection writes:

$$A^\mu = \begin{pmatrix} A_1^\mu & & & & \\ & A_2^\mu & & & \\ & & A_3^\mu & & \\ & & & A_4^\mu & \\ & & & & A_5^\mu \end{pmatrix} , \quad \Phi^1 = \begin{pmatrix} & X_{12} & & & \\ & & X_{23} & & \\ & & & X_{34} & \\ & & & & X_{45} \\ X_{51} & & & & \end{pmatrix} , \tag{7.24}$$

$$\Phi^2 = \begin{pmatrix} & & Y_{13} & & \\ & & & Y_{24} & \\ & & & & Y_{35} \\ Y_{41} & & & & \\ & Y_{52} & & & \end{pmatrix} , \quad \Phi^3 = \begin{pmatrix} & & Z_{13} & & \\ & & & Z_{24} & \\ & & & & Z_{35} \\ Z_{41} & & & & \\ & Z_{52} & & & \end{pmatrix} . \tag{7.25}$$

The low-energy worldvolume theory on a stack of N D3-branes at a $\mathbb{C}^3/\mathbb{Z}_5$ singularity is a four dimensional $\mathcal{N} = 1$ $\mathrm{SU}(N)_1 \times \mathrm{SU}(N)_2 \times \mathrm{SU}(N)_3 \times \mathrm{SU}(N)_4 \times \mathrm{SU}(N)_5$ gauge theory with chiral multiplets X_{ij}, Y_{ij} and Z_{ij} in the bifundamental representation $(\overline{\square}_i, \square_j)$ for some values of i and j, and superpotential:

$$\begin{aligned} W =& Z_{13}X_{34}Y_{41} - Z_{13}Y_{35}X_{51} + Z_{24}X_{45}Y_{52} - Z_{24}Y_{41}X_{12} + Z_{35}X_{51}Y_{13} - Z_{35}Y_{52}X_{23} \\ &+ Z_{41}X_{12}Y_{24} - Z_{41}Y_{13}X_{34} + Z_{52}X_{23}Y_{35} - Z_{52}Y_{24}X_{45} . \end{aligned} \tag{7.26}$$

The moduli space of this theory is parameterized by gauge invariant mesonic operators built out of the X_{ij}, Y_{ij} and Z_{ij}, and one can check that modulo F-terms it is a copy of $\mathbb{C}^3/\mathbb{Z}_5$.

The matter content of worldvolume theories of D3-branes at such singularities is conveniently encoded in a *quiver*, i.e. an oriented graph: nodes represent gauge groups, and an arrow from node i to the j encoded a bifundamental chiral superfield in the representation $(\overline{\square}_i, \square_j)$. The quivers corresponding to the examples we have presented are given in Figure 7.1.

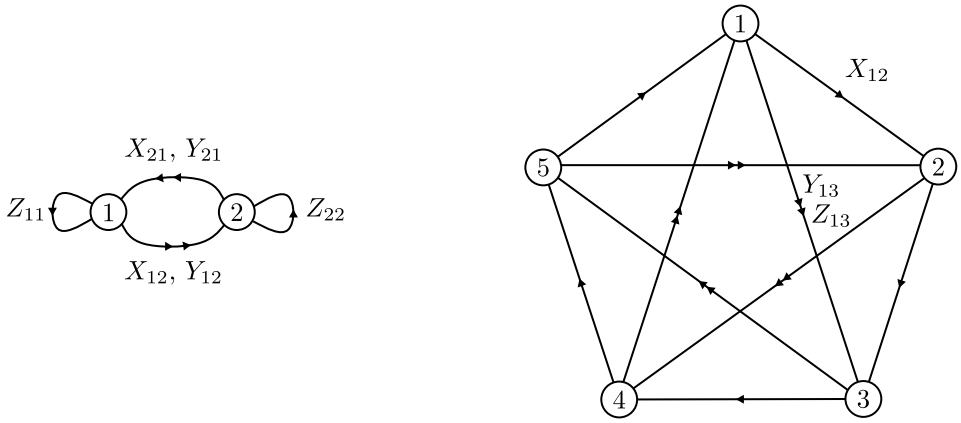


Figure 7.1: The quivers corresponding to D3 branes at $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$ (left) and $\mathbb{C}^3/Z_5$ (right).

An orbifold of $\mathbb{C}^3$ by a discrete group G is Calabi–Yau if and only if its holonomy is in SU(3). Since the holonomy of $\mathbb{C}^3$ is trivial, an orbifold $\mathbb{C}^3/G$ is Calabi–Yau if and only if G is a discrete subgroup of SU(3): G can for example be the abelian group $\mathbb{Z}_n$ generated by g_n acting as $(z_1, z_2, z_3) \to (g_n^a z_1, g_n^b z_2, g_n^c z_3)$ at the condition that $a + b + c = 0 \mod n$. When $\mathbb{C}^3/G$ is Calabi–Yau, the low-energy worldvolume theory on a stack of D3-branes at the singularity is a four dimensional $\mathcal{N} = 1$ gauge theory – actually, it is even a four dimensional $\mathcal{N} = 1$ superconformal theory, as one can check by computing the beta functions of the gauge couplings. When the holonomy is not only a discrete subgroup of SU(3) but of SU(2), the low-energy effective theory has $\mathcal{N} = 2$ supersymmetry. Moving a D3-brane out of the singularity i.e. from $(z_1, z_2, z_3) = (0, 0, 0)$ to a non-zero (z_1, z_2, z_3) amounts to giving a vacuum expectation value to a meson, which breaks the gauge group to the diagonal SU(N) factor, in general: the low-energy theory is then the same as the one of D3-branes in flat space $\mathbb{C}^3$.

When the theory has only $\mathcal{N} = 1$ supersymmetry, the moduli space of a single D3-brane is parameterized by gauge invariant mesonic operators and can be naturally identified with $\mathbb{C}^3/G$ when the Fayet–Illiopoulos parameters vanish. The latter parameterize resolutions of the orbifold [DGM97].

Translating the string theory orbifolding procedure into the gauge theory language, general rules describing directly any orbifold of four-dimensional $\mathcal{N} = 4$ U(N) super Yang–Mills were derived in [KS98, LNV98, BKV98, Kak98].

7.3.2 Fractional D3-branes

In contrast to regular D3-branes at singularities $\mathbb{C}^3/G$ defined as the endpoint locus of open strings whose Chan–Paton factors transform in the regular representation of G, fractional branes are the endpoint locus of open strings whose Chan–Paton factors transform in a representation of G smaller than the regular one (hence the name).

The low-energy theory on a fractional D3-brane at a Calabi–Yau orbifold $\mathbb{C}^3/G$ is a gauge theory which is easily described on the quiver corresponding to a regular D3 brane at the singular point. Let us describe this quiver as $\mathcal{Q} = (V_0, V_1)$ where V_0 is the set of nodes and $V_1 \subset V_0^2$ the set of arrows: if $v, v' \in V_0$ the pair $(v, v') \in V_0^2$ describes an arrow from v to v'. The low-energy worldvolume theory on a fractional brane is described as a rank assignment $(N_v)_{v \in V_0} \in (\mathbb{Z}_{\geq 0})^{V_0}$ on the nodes of the quiver such that:

1. it is not constant, i.e. there exist $v, v' \in V_0$ such that $N_v \neq N_{v'}$,

2. it is simple, i.e. it there is no $k \geq 2$ such that $N_v/k \in \mathbb{Z}_{\geq 0}$ for all $v \in V_0$,

3. at least one node has rank 0.

If the rank assignment is constant one is describing a stack of regular branes, and if it is not simple it corresponds to some bound state of fractional branes (and possibly regular ones).

All possible rank assignment are however not allowed, since the gauge theory described in this way may suffer gauge anomalies and hence be inconsistent quantum mechanically. The gauge anomaly cancellation is easily written in quiver language:

$$\forall v \in V_0 \ , \qquad \sum_{v' \to v} N_{v'} - \sum_{v \to v'} N_{v'} = 0 \ . \tag{7.27}$$

Any rank assignment on the quiver satisfying Equation (7.27) is a bound state of regular and fractional D3 branes at the singularity at hand.

For example, any rank assignment on the nodes of the quiver on the left of Figure 7.1 satisfies Equation (7.27) (and hence there are two fractional branes corresponding to $(N_1, N_2) = (1, 0)$ and $(N_1, N_2) = (0, 1)$), whereas the only anomaly-free rank assignments in the quiver corresponding to $\mathbb{C}^3/\mathbb{Z}_5$ are constant, i.e. there are no fractional branes on $\mathbb{C}^3/\mathbb{Z}_5$.

In general the low-energy effective theories on fractional D3-branes are non conformal $\mathcal{N} = 1$ gauge theories, which makes fractional branes very interesting in order to study non-conformal gauge–gravity correspondences, as we will review in the next chapter.

On an orbifold $\mathbb{C}^3/G$ closed strings excitations can also wind non-trivially around the singular point(s), and hence one decomposes them into twisted and untwisted sectors. Modular invariance imposes the introduction of the twisted sectors in order to obtain a consistent anomaly-free string theory [DHVW85, DHVW86]. Regular branes couple only to the untwisted sector, while fractional branes couple to both. In the twisted sector one finds in particular R–R p-forms which are the zeroth Kaluza–Klein modes of R–R $(p + 2)$-forms compactified on vanishing 2-cycles of the orbifold, if any. Hence one can wonder whether fractional D3-branes may be higher D-branes in disguise, and this turns out to be exactly the case: fractional D3-branes are D5-branes wrapped on vanishing 2-cycles (so that they are effectively threebranes). However not every 2-cycle is allowed: a D5-brane wrapped on a vanishing 2-cycle C sources the R–R four-form and hence the flux of the latter is non zero on the 4-cycle dual to C. In order for the theory to be consistent, R–R tadpoles must vanish and it turns out that they do only when this 4-cycle is non-compact. This condition is actually exactly equivalent to Equation (7.27).

7.4 Toric Calabi–Yau singularities

Generalities.

A complex toric variety X of dimension n is a complex algebraic variety of dimension n which contains the algebraic split torus $(\mathbb{C}^\times)^n$ as an open dense subset and such that the natural action of $(\mathbb{C}^\times)^n$ on itself extends to X. For a first quick example, note that $\mathbb{P}^2(\mathbb{C})$ is a toric variety of dimension 2: if it is parameterized by homogeneous coordinates $[x : y : z]$, the torus $(\mathbb{C}^\times)^2$ is the open dense subset $\{[x : y : z] \mid x, y, z \neq 0\}$ with $(\alpha, \beta) \in (\mathbb{C}^\times)^2$ corresponding to $[\alpha : \beta : 1]$, and the action of $(\mathbb{C}^\times)^2$ on itself extends on $\mathbb{P}^2(\mathbb{C})$ in such a way that $(\alpha, \beta) \in (\mathbb{C}^\times)^2$ acts as:

$$[x : y : z] \to [\beta^{-1}x : \alpha^{-1}y : \alpha^{-1}\beta^{-1}z] \ . \tag{7.28}$$

Any toric variety can be defined from the combinatorial data of a fan, to be defined very soon. Conversely the action of the torus on any toric variety defines a fan. We will briefly review the construction of toric varieties from fans and give some of their properties without proof. We refer to [Ful16, CLS11, Clo09] for more details.

Affine complex toric varieties of dimension n are conveniently described in terms of cones in $L = \mathbb{Z}_n$. A *strongly convex rational polyhedral cone* in $L_\mathbb{R} = L \otimes_\mathbb{Z} \mathbb{R}$ is a subset

$$\sigma = \{r_1 v_1 + \ldots r_s v_s \in L_\mathbb{R} \mid r_1, \ldots, r_s \in \mathbb{R}_{\geq 0}\} \subset L_\mathbb{R} \ , \tag{7.29}$$

where $\{v_1, \ldots, v_s\}$ is a finite set of vectors in L, and such that σ does not contain any line through the origin of $L_\mathbb{R}$. In the following, strongly convex rational polyhedral cones will be referred to as cones, for

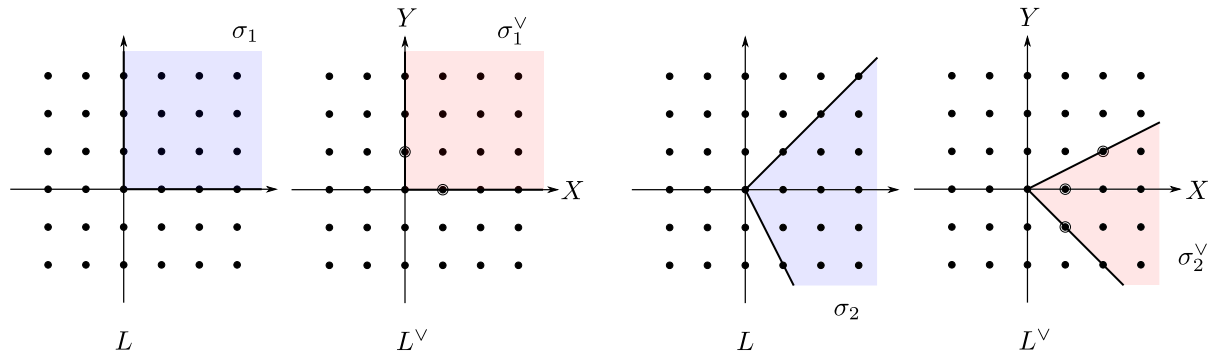
short. Let now $L^\vee = \mathrm{Hom}(L, \mathbb{Z})$ be the dual module of L and let $L^\vee_{\mathbb{R}} = L^\vee \otimes_{\mathbb{Z}} \mathbb{R}$. The dual $\sigma^\vee$ of a cone σ is the set

$$\sigma^\vee = \{w \in L^\vee_{\mathbb{R}} \mid \langle w, v\rangle \geq 0 \ \forall v \in L_{\mathbb{R}}\} \subset L^\vee_{\mathbb{R}} \ . \tag{7.30}$$

By Farkas' theorem, $\sigma^\vee$ is a convex polyhedral cone in $L^\vee_{\mathbb{R}}$. Moreover, Gordon's lemma asserts that $S_\sigma = \sigma^\vee \cap L^\vee$ is a finitely generated semi-group; let $A_\sigma = \mathbb{C}[S_\sigma]$ be the complex commutative algebra determined by S_σ. The *affine toric variety* corresponding to a cone σ is

$$X_\sigma = \mathrm{Spec}\ A_\sigma \ . \tag{7.31}$$

Two examples of cones in $L = \mathbb{Z}^2$, their dual in $L^\vee$, the corresponding complex commutative algebras and their affine toric varieties are shown in Figure 7.2. A choice of generators in $\sigma_1^\vee$ and $\sigma_2^\vee$ is shown as double circles.



$S_{\sigma_1} = \mathbb{C}[X, Y]$
$X_{\sigma_1} = \mathbb{C}^2$

$S_{\sigma_2} = \mathbb{C}[X, XY^{-1}, X^2Y]$
$X_{\sigma_2} = \mathbb{C}^2/\mathbb{Z}_3$

Figure 7.2: Two examples of cones, their dual and the corresponding affine toric variety.

A face of a cone σ is the intersection of σ with any hyperplane in $L_{\mathbb{R}}$. The faces of a cone are also cones. A *fan* Δ in L is a finite set of cones in $L_{\mathbb{R}}$ such that for each $\sigma \in \Delta$ the faces of σ are in Δ, and such that for each $\sigma, \sigma' \in \Delta$ the intersection $\sigma \cap \sigma'$ is a face both of σ and of σ'. The toric variety X_Δ corresponding to a fan Δ in L is obtained by gluing the affine toric varieties corresponding to each higher-dimensional cone in Δ together, along the open dense subsets defined by the lower dimensional cones. The toric variety corresponding to the 0-dimensional cone consisting of the origin of L is always $(\mathbb{C}^\times)^n$. Since every cone in L contains it, every toric variety constructed from a fan in $L = \mathbb{Z}^n$ has $(\mathbb{C}^\times)^n$ as an open dense subset.

The example of $\mathbb{P}^2(\mathbb{C})$ is given in Figure 7.3. The fan consists of three 2-dimensional cones σ_1, σ_2 and σ_3, three 1-dimensional cones on which the 2-dimensional cones intersect and one 0-dimensional cone, which is the origin in L. The diagram on the right shows how each commutative semi-group algebra corresponding to a 2-dimensional cone injects into the semi-group algebras corresponding to the faces along which it intersects other 2-dimensional cones. These monomorphisms are dual to open dense embeddings of the spectrum of the latter into the spectrum of the former.

The affine varieties X_{σ_1}, X_{σ_2} and X_{σ_3} are the standard affine charts on $\mathbb{P}^2(\mathbb{C})$, the toric varieties corresponding to one-dimensional cones in Δ are the open dense subsets of $\mathbb{P}^2(\mathbb{C})$ on which two of the three homogeneous coordinates are non-zero, and the toric variety corresponding to the unique 0-dimensional cone in Δ is the open dense subset of $\mathbb{P}^2(\mathbb{C})$ on which neither of the homogeneous coordinate is zero, i.e. it is $(\mathbb{C}^\times)^2$.

A few important geometric properties of toric complex varieties are the following [Ful16].

Proposition 7.1. *Let X_Δ be the complex toric variety defined by a fan Δ in L. Then:*

1. *X_Δ is normal,*
2. *X_Δ is Cohen-Macaulay,*
3. *X_Δ is non-singular if and only if every cone in Δ is generated by part of a basis of L,*

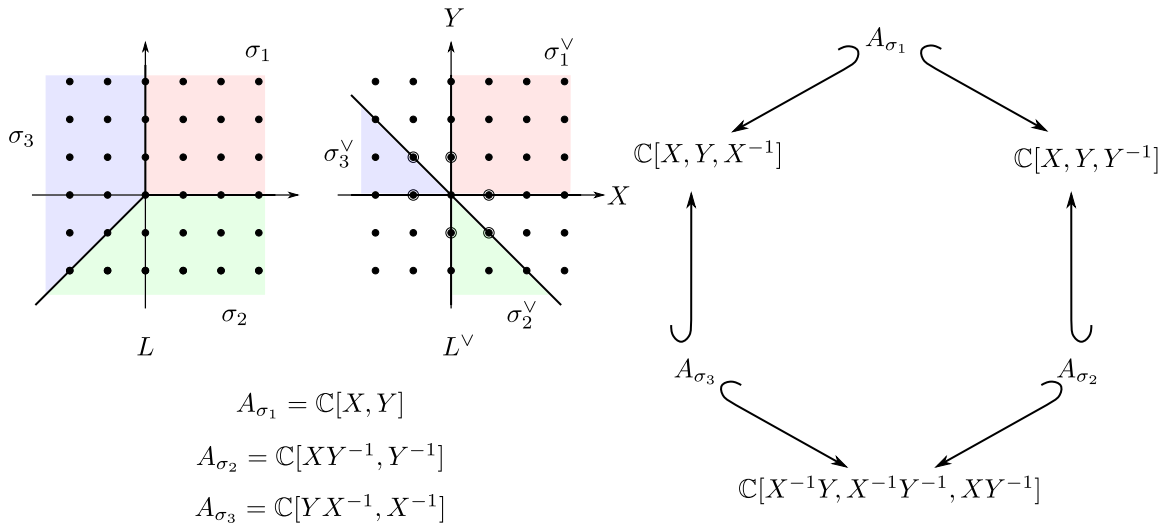


Figure 7.3: Two examples of cones, their dual and the corresponding affine toric variety.

4. *X_Δ is compact if and only if the support $|\Delta|$ of Δ is the whole space L (with the obvious definition of the support of a fan).*

Because of the 4. in Proposition 7.1, the fan Δ corresponding to any compact two-dimensional toric variety (i.e. toric surface) corresponds to a lattice polygon P. The latter is defined by its vertices, which are the generators of the semi-groups defined by each 1-dimensional cone in Δ. Four examples are shown in Figure 7.4.

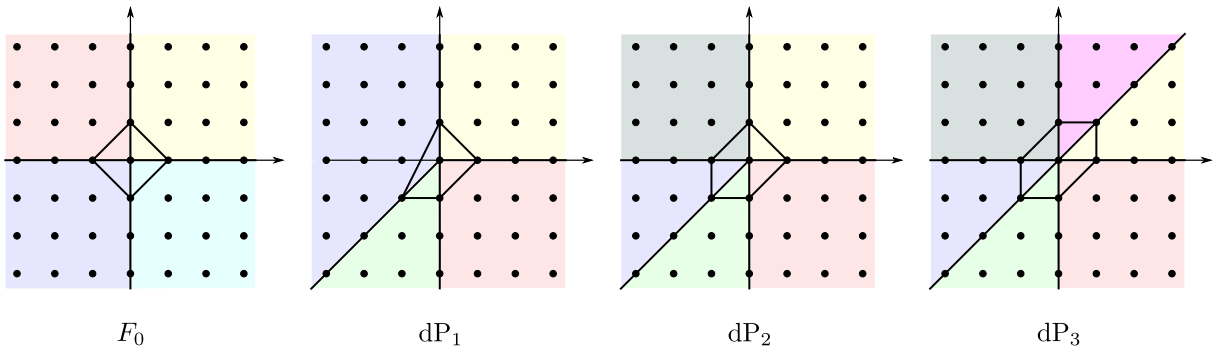


Figure 7.4: Fans from polygons.

The toric surface defined by the left-most fan is called the zeroth *Hirzebruch surface* and denoted F_0; it is $\mathbb{P}^1(\mathbb{C}) \times \mathbb{P}^1(\mathbb{C})$. The one to its right defines the first *del Pezzo surface* dP_1 (which is also the first Hirzebruch surface). It is a non-trivial $\mathbb{P}^1$ bundle over $\mathbb{P}^1$, and can be equivalently described as $\mathbb{P}^2(\mathbb{C})$ blown-up at one generic point. The second and third del Pezzo surfaces (respectively dP_2 and dP_3) on the right of the same figure can similarly be defined as $\mathbb{P}^2(\mathbb{C})$ blown-up at respectively 2 and 3 generic points. Consistently, $\mathbb{P}^2(\mathbb{C})$ is sometimes denoted dP_0.

The homogeneous coordinate ring of a toric variety.

There is another equivalent construction of toric varieties from fans introduced in [Cox93] which is very convenient in practice. Let Δ be a fan in $L = \mathbb{Z}^n$ and let $\lambda_1, \ldots, \lambda_m$ be the one-dimensional cones in Δ. For each $i = 1, \ldots, m$ let l_i be the generator of the semi-group $\lambda_i \cap L$. If $m \geq n$ there are $m - n$ independent linear relations among the l_i that we can write as an $m \times (m - n)$ matrix:

$$\sum_{i=1}^{m} Q_k^i l_i = 0, \qquad k = 1, \ldots, m - n \ . \tag{7.32}$$

Without loss of generality one can assume that the Q_k^i are integers. Let us define an action of $(\mathbb{C}^\times)^{m-n}$ on $\mathbb{C}^m$ such that $(\alpha_1, \ldots, \alpha_{m-n})$ acts as:

$$(z_1, \ldots, z_m) \longrightarrow \left(\prod_{k=1}^{m-n} \alpha_k^{Q_k^1} z_1, \ldots, \prod_{k=1}^{m-n} \alpha_k^{Q_k^m} z_m \right) . \tag{7.33}$$

Now, for every set $S = \{l_{i_1}, \ldots, l_{i_j}\}$ which does not generate a cone in Δ let $V_S = \{z_{i_1} = \cdots = z_{i_j} = 0\} \subset \mathbb{C}^m$, and let

$$Z_\Delta = \bigcup V_S , \tag{7.34}$$

where the sum runs over all such sets. Then:

$$X_\Delta = \frac{\mathbb{C}^m - Z_\Delta}{(\mathbb{C}^\times)^{m-n} \times \Gamma}, \tag{7.35}$$

where Γ is the discrete group L/L' with L' the lattice generated by the l^i.

For example, the fan on the left of Figure 7.3 is such that $Z_\Delta = \{(0,0,0)\}$, and there is only one relation among the 1-dimensional cone which defines the action $(z_1, z_2, z_3) \to (\alpha z_1, \alpha z_2, \alpha z_3)$: in that case Equation (7.35) yields exactly the definition of $\mathbb{P}^2(\mathbb{C})$ as the set of complex lines in $\mathbb{C}^3$.

Consider now the fan consisting of a 3-dimensional cone in $L = \mathbb{Z}^3$ generated by $(0,0,1)$, $(0,1,1)$, $(1,1,1)$ and $(1,0,1)$ and its faces. One can check in that case that $Z_\Delta = \{z_1 = z_3 = 0\} \cup \{z_2 = z_4 = 0\}$, and one needs to consider the action of $(\mathbb{C}^\times)$ on (z_1, z_2, z_3, z_4) given by

$$(z_1, z_2, z_3, z_4) \to (\alpha z_1, \alpha^{-1} z_2, \alpha z_3, \alpha^{-1} z_4) . \tag{7.36}$$

The corresponding toric variety defined as

$$\mathcal{C}_0 = \frac{\mathbb{C}^4 - Z_\Delta}{\mathbb{C}^\times} \tag{7.37}$$

is known as the *conifold* [CdlO90]. The ring of regular functions over $\mathcal{C}_0$ is generated by monomials $z_1^{n_1} z_2^{n_2} z_3^{n_3} z_4^{n_4}$ with $n_1, n_2, n_3, n_4 \geq 0$, that are invariant under the action of $\mathbb{C}^\times$, i.e. those such that $n_1 - n_2 + n_3 - n_4 = 0$. A set of monomial generators is $u = z_1 z_2$, $v = z_3 z_4$, $w = z_1 z_4$ and $x = z_2 z_3$, and there is one relation among them $uv = wx$, so that the ring of regular function over $\mathcal{C}_0$ is

$$\mathcal{O}_{\mathcal{C}_0} = \frac{\mathbb{C}[u, v, w, x]}{uv - wx} . \tag{7.38}$$

This ring can also be interpreted as the one defining the projective surface in $\mathbb{P}^1(\mathbb{C}^3)$ which describes a ordinary double point at $(0,0,0)$, and hence $\mathcal{C}_0$ is the affine cone over it.

This holds in general: let σ be a top-dimensional cone in L viewed together with its faces as a cone Δ, and let X_Δ be the corresponding affine toric variety defined as in Equation (7.35). The ring of regular functions $\mathcal{O}_{X_\Delta}$ is generated by the monomials $z_1^{n_1} \ldots z_m^{n_m}$ which are invariant under the action of $(\mathbb{C}^\times)^{m-n}$, i.e. those such that $n_1, \ldots, n_m \geq 0$ and $(n_1, \ldots, n_m) \cdot Q = 0$. These conditions identify the monomials in $\mathcal{O}_{X_\Delta}$ with the integral elements of the cone $\sigma^\vee$ dual to σ defined in Equation (7.30). For every $x \in \sigma^\vee \cap L^\vee$:

$$n_i = \langle x, v_i \rangle . \tag{7.39}$$

Hence the ring of regular functions $\mathcal{O}_{X_\Delta}$ can equivalently be expressed as

$$\mathcal{O}_{X_\Delta} = \mathbb{C}[\sigma^\vee \cap L^\vee] . \tag{7.40}$$

If Δ is any fan in L each top-dimensional cone in it corresponds to an affine chart on X_Δ whose ring of regular function is given by Equation (7.40).

Let us now again consider a top-dimensional cone σ in L, and let $\{w_1, \ldots, w_p\}$ be a set of generators of the dual cone $\sigma^\vee \cap L^\vee$. For each $i = 1, \ldots, p$ let $(w_i^1, \ldots, w_i^n)$ be the coordinates of w_i in $L^\vee$ and assign the weight w_i^n to w_i. The ring of regular functions on X_Δ (where again Δ consists of σ and its faces) can be written as

$$\mathcal{O}_{X_\Delta} = \frac{\mathbb{C}[x_1, \ldots, x_p]}{\mathcal{I}} , \tag{7.41}$$

where for each $i = 1, \ldots, p$ the variable x_i corresponds to w_i and the additive semi-group relations between the w_i transpose into multiplicative relations between the x_i which generate the ideal $\mathcal{I}$. The latter is homogeneous for the weights we have assigned to the w_i: the same polynomial ring defines a complete variety in the weighted projective space $\mathbb{P}_{(w_1^n,\ldots,w_p^n)}(\mathbb{C})$ and X_Δ is the affine cone over it.

This definition of a toric variety is nearly identical to the *Gauged Linear Sigma Model* (GLSM) introduced in [Wit93]: the relationship between the two approaches is given by the Kempf-Ness theorem [KN79]. The GLSM approach defines the toric variety as the supersymmetric moduli space of an $\mathcal{N} = 2$, $\mathrm{U}(1)^{m-n}$ gauge theory in two dimensions. One can even describe the resolutions of a given toric variety by turning on Fayet-Illiopoulos terms for the U(1) factors in the gauge group.

The Calabi–Yau condition and toric diagrams

The toric divisors in a toric variety are easily expressed in terms of the homogeneous coordinates $z_1, \ldots, z_m$. There is one toric divisor for each $i = 1, \ldots, m$:

$$D_i : \quad \{z_i = 0\} \cap X_\Delta \ . \tag{7.42}$$

The canonical bundle on X_Δ is given by

$$\mathcal{K}_{X_\Delta} = \mathcal{O}\left(-\sum_{i=1}^{m} D_i\right) \ , \tag{7.43}$$

for a proof of which we refer to [Clo09]. We have also the following result:

Proposition 7.2. *The canonical bundle $\mathcal{K}_{X_\Delta}$ is trivial (i.e. X_Δ is Calabi–Yau) if and only if*

$$\sum_{i=1}^{m} Q_k^i = 0, \qquad \forall k = 1, \ldots m - n \ , \tag{7.44}$$

if and only if all the vectors l_i defined as before all lie on the same hyperplane in L.

We are especially interested in three-dimensional Calabi–Yau varieties with the aim to construct non-trivial backgrounds for type IIB superstrings and study the worldvolume theory on D3-branes at Calabi–Yau singularities. In that case the condition of Proposition 7.2 translates into the fact that Δ needs to be generated by vectors all lying in a single 2-plane in $L = \mathbb{Z}^3$. Up to an $\mathrm{SL}_3(\mathbb{Z})$ change of basis in $L = \mathbb{Z}^3$ one can assume that this plane is defined by the equation $z = 1$: a toric CY3 variety is thus entirely defined by some planar data. Note that this implies that a toric CY3 cannot be compact, because of 4. in Proposition 7.1. When the fan Δ defining a CY3 contains a single 3-dimensional cone, it is equivalently given by the data of a planar lattice polygon dubbed the toric diagram of the toric CY3. The subgroup of $\mathrm{SL}_3(\mathbb{Z})$ fixing the hyperplane $z = 1$ is isomorphic to $\mathrm{SL}_2(\mathbb{Z})$ and acts on the hyperplane $z = 1$ as changes of basis. Hence, a toric diagram is only defined up to the action of this $\mathrm{SL}_2(\mathbb{Z})$ group.

Examples and computations.

Let us consider the toric diagram on the left of Figure 7.4, dubbed F_0. It defines a fan in $L = \mathbb{Z}^3$ which contains a single 3-dimensional cone σ generated by $(1, 0, 1)$, $(0, 1, 1)$, $(-1, 0, 1)$ and $(0, -1, 1)$. A set of generators for the dual cone $\sigma^\vee$ in $L^\vee$ is

$$\left\{ \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \right\} \ . \tag{7.45}$$

For each $i = 1, \ldots, 9$ let as before x_i be the coordinate corresponding to w_i. Let I be the ideal of relations among the x_i; one can show that a minimal set of generators for such relations contain 20 polynomials. Thus:

$$\mathcal{O}_{X_{F_0}} = \frac{\mathbb{C}[x_1, \ldots, x_9]}{\mathcal{I}} \ . \tag{7.46}$$

The same polynomial ring defines a projective surface in $\mathbb{P}^1(\mathbb{C}^9)$ which is the zeroth Hirzebruch surface F_0. The toric CY3 defined by Equation (7.46) is the affine cone over F_0.

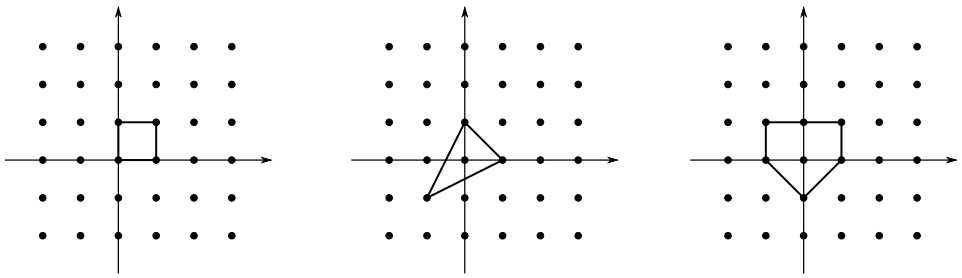


Figure 7.5: Toric diagrams for the conifold $\mathcal{C}_0$, the orbifold $\mathbb{C}^3/\mathbb{Z}_3$ and PdP_3.

Let us now consider the toric diagram on the left of Figure 7.5. It defines a 3-dimensional cone σ in $L = \mathbb{Z}^3$ generated by $(0,0,1)$, $(1,0,1)$, $(1,1,1)$ and $(0,1,1)$. The dual cone $\sigma^\vee$ is generated by $(0,1,0)$, $(-1,0,1)$, $(-1,0,1)$ and $(1,0,0)$. The ring of regular functions over the corresponding toric CY3 is

$$\frac{\mathbb{C}[x_1,x_2,x_3,x_4]}{\langle x_1x_3 - x_2x_4\rangle}\ , \tag{7.47}$$

i.e. the corresponding toric CY3 is the conifold $\mathcal{C}_0$.

The toric diagram in the middle of Figure 7.5 defines a three-dimensional cone σ in $L = \mathbb{Z}^3$ generated by $(1,0,1)$, $(0,1,1)$ and $(-1,-1,1)$. The dual cone $\sigma^\vee$ is generated by:

$$\left\{\begin{pmatrix}-1\\2\\1\end{pmatrix},\begin{pmatrix}0\\1\\1\end{pmatrix},\begin{pmatrix}-1\\-1\\1\end{pmatrix},\begin{pmatrix}1\\0\\1\end{pmatrix},\begin{pmatrix}0\\-1\\1\end{pmatrix},\begin{pmatrix}-1\\0\\1\end{pmatrix},\begin{pmatrix}1\\-1\\1\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix},\begin{pmatrix}-1\\1\\1\end{pmatrix},\begin{pmatrix}2\\-1\\1\end{pmatrix}\right\}. \tag{7.48}$$

Let $x_1,\ldots,x_{10}$ be the corresponding variables. The ideal of relations I among them is minimally generated by 25 polynomials. The variety defined by its ring of regular functions $\mathbb{C}[x_1,\ldots,x_{10}]/\mathcal{I}$ embeds into the affine space $\mathbb{C}^{10}$ and is the affine cone over a projective curve in $\mathbb{P}^1(\mathbb{C}^{10})$. Using now the definition of Equation (7.35) yields directly that this toric variety is $\mathbb{C}^3/\mathbb{Z}_3$.

In general if σ is a 3-dimensional cone in $L = \mathbb{Z}_3$ one may wonder 1) how one can compute the dual cone $\sigma^\vee$, 2) how one can compute a minimal set of generators of the latter, 3) how one can compute a minimal set of generating relations among the generators of $\sigma^\vee$ found in 2). These three quests are in general difficult to handle by hand, even if if is reasonably doable for small toric diagrams. However there exist general algorithms to do this, and using a software like Macaulay2 [GS] is in any case quicker and safer than computing by hand. In order to exemplify the procedure we will show explicitly the computation for the toric diagram on the right of Figure 7.5.

1) Regarding the computation of the dual cone, this is easily done using a Package of Macaulay2 called Polyhedra:

```
loadPackage "Polyhedra"
M = matrix{{-1,0,1},{0,-1,1},{1,0,1},{1,1,1},{0,1,1},{-1,1,1}}
Mt = transpose M
C = posHull Mt
Cv = dualCone C
```

2) Let C be a polyhedral cone in $\mathbb{R}^d$, i.e. the positive hull of a finite set of points in $\mathbb{R}^d$. There exists a set $\mathcal{H}_{\mathbb{Z}^d}(C) \subset C \cap \mathbb{Z}^d$ such that:

1. each $z \in C \cap \mathbb{Z}^d$ is a positive integer combination of elements of $\mathcal{H}_{\mathbb{Z}^d}(C)$, i.e. for each $h \in \mathcal{H}_{\mathbb{Z}^d}(C)$ there exists $z_h \in \mathbb{N}$ such that $z = \sum_{h\in\mathcal{H}_{\mathbb{Z}^d}(C)} z_h h$.

2. $\mathcal{H}_{\mathbb{Z}^d}(C)$ has minimum cardinal among all subsets of $C \cap \mathbb{Z}^d$ for which 1. holds.

Such a set is said to be a *Hilbert basis* of C [Hil90, HW97b, Sch98]. That is exactly what we have in mind when speaking of a minimal set of generators for a cone. There exist algorithms to compute Hilbert bases of polyhedral cones [Hem02]. Typing:

```
hilbertBasis Cv
```

in Macaulay2 after the previous lines yield the set of column vectors:

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \right\} , \tag{7.49}$$

which form indeed a Hilbert basis for the dual cone $\sigma^\vee$. The command

```
rays Cv
```

yields the *extremal generators* of $\sigma^\vee$ instead, which are in one-to-one correspondence with the sides of the toric diagram. The extremal generators are contained in any Hilbert basis of $\sigma^\vee$.

3) The relations between the generators of the ring of regular functions on the toric variety can be obtained as follows:

```
S = QQ[x_1,x_2,x_3,x_4,x_5,x_6,x_7]
Q = QQ[a,a',b,b',c,c']
A = ideal(a*a'-1,b*b'-1,c*c'-1)
R = Q/A
g = map(R,S,{a*c,b*c,a'*c,b'*c,c,a*b*c,a'*b*c})
```

This last line encodes how the generators $x_1, \ldots, x_7$ of the ring of regular functions are related to the generators of the dual cone $\sigma^\vee$: the additive semi-group structure on $\sigma^\vee$ becomes a multiplicative semi-group structure on the monomial in the ring of regular functions, and hence:

$$\mathcal{O}_X = \frac{\mathbb{C}[x_1, \ldots, x_7]}{\ker g} . \tag{7.50}$$

```
I = kernel g
mingens I
```

yields nine polynomial relations:

$$\left| \begin{array}{l} x_3x_6 - x_1x_7 \\ x_3x_5 - x_4x_7 \\ x_2x_5 - x_1x_7 \\ x_1x_5 - x_4x_6 \\ x_2x_4 - x_5^2 \\ x_2x_3 - x_5x_7 \\ x_1x_3 - x_5^2 \\ x_2^2 - x_6x_7 \\ x_1x_2 - x_5x_6 \end{array} \right. , \tag{7.51}$$

which indeed form a minimal generating set for the ideal of relations in the case at hand. One can also compute the relation between relations in a similar way, the relations between relations between relations ... In fact there is a built-in function in Macaulay2 which computes the free resolution of a ring or an ideal. For example,

```
res I
```

yields the chain complex

$$R^1 \longleftarrow R^9 \longleftarrow R^{16} \longleftarrow R^9 \longleftarrow R^1 \longleftarrow 0 , \tag{7.52}$$

and one can obtain the explicit expression of the maps in this complex via the commands `Res.dd_1`, `Res.dd_2`, ...

Triangulations and resolutions

Let Δ be a fan in $\mathbb{Z}^3$ defined by a toric diagram and consisting of a single 3-dimensional cone σ generated by the vertices of the toric diagram together with the faces of σ. The 3. of Proposition 7.1 implies that X_Δ is in singular except if the toric diagram is an elementary triangle, and hence one speaks of affine toric CY3 in these cases. A natural way to resolve the singularity is to fix a triangulation of the toric diagram: the new fan Δ_T now contains as many 3-dimensional cones as the number of triangles in the triangulation of the toric diagram, and each such 3-dimensional cone defines a non-singular affine chart on X_{Δ_T}, so that X_{Δ_T} itself is non-singular.

Let us consider the case of the conifold first. The two possible triangulations of the toric diagram are shown on the left and in the middle of Figure 7.6. Let σ_1 and σ_2 (resp. σ_3 and σ_4) be the two 3-dimensional cones appearing in Δ_T for the leftmost (resp. rightmost) triangulation of the toric diagram of the conifold, read from left to right. Then in the case of the leftmost triangulation:

$$\sigma_1^\vee = \left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\rangle, \quad \sigma_2^\vee = \left\langle \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\rangle, \tag{7.53}$$

so that the first affine chart is Spec $(\mathbb{C}[x_1, y_1, z_1])$ and the second Spec $(\mathbb{C}[x_2, y_2, z_2])$, with the transition function

$$z_1 = z_2^{-1}, \qquad \frac{x_1}{y_1} = \frac{x_2}{y_2}, \qquad x_1 z_1 = x_2, \qquad y_1 z_1 = y_2 \tag{7.54}$$

on the overlap of the two charts, as follows from the integral relations between the vectors in $L^\vee$.

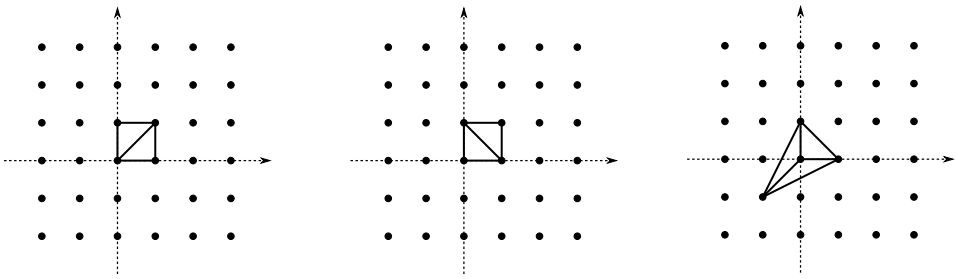


Figure 7.6: Triangulations of toric diagrams.

In the case of the other triangulation:

$$\sigma_3^\vee = \left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \right\rangle, \quad \sigma_4^\vee = \left\langle \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\rangle, \tag{7.55}$$

so that the first affine chart is Spec $(\mathbb{C}[x_1, x_2, w_1])$ and the second Spec $(\mathbb{C}[y_1, y_2, w_2])$, with the transition function

$$w_1 = w_2^{-1}, \qquad \frac{x_1}{x_2} = \frac{y_1}{y_2}, \qquad x_1 w_1 = y_2, \qquad x_2 w_1 = y_2 \tag{7.56}$$

on the overlap of the two charts, as follows from the integral relations between the vectors in $L^\vee$.

In both case when x_1, x_2, y_1 and y_2 are not zero the z_i or the w_i are determined by the x_i and the y_i however it is not the case when $x_1 = x_2 = y_1 = y_2 = 0$, in which case the z_i or the w_i parameterize a copy of $\mathbb{P}^1(\mathbb{C})$ that resolves the conifold singularity. The so-called resolved conifold is the total space of the plane bundle $\mathcal{O}(-1) \oplus \mathcal{O}(-1) \to \mathbb{P}^1(\mathbb{C})$. As one can see from two rightmost relations in Equation (7.54) and Equation (7.56) these two resolutions are not equivalent: they are related by a flop.

Regarding the rightmost diagram in Figure 7.6, the consequence of the triangulation is most easily seen if one writes the variety as in Equation (7.35):

$$\frac{\mathbb{C}^4 - \{z_1 = z_2 = z_3 = 0\}}{\mathbb{C}^\times} \tag{7.57}$$

where $\mathbb{C}^\times \ni \alpha : (z_1, z_2, z_3, z_4) \to (\alpha z_1, \alpha z_2, \alpha z_3, \alpha^{-3} z_4)$. The first three coordinates parameterize $\mathbb{P}^2(\mathbb{C})$ and the last one parametrizes a line bundle over it which is $\mathcal{O}(-3) \to \mathbb{P}^2(\mathbb{C})$. Hence the triangulation in

this case corresponds to the blow-up

$$\left\{ \begin{array}{c} \mathcal{O}(-3) \\ \downarrow \\ \mathbb{P}^2(\mathbb{C}) \end{array} \right\} \longrightarrow \mathbb{C}^3/\mathbb{Z}_3 \ . \tag{7.58}$$

The torus fibration, (p,q)-webs and the topology of affine toric CY3.

By definition any toric variety X of dimension n contains the complex split torus $(\mathbb{C}^\times)^n$ as an open dense subset, and the action of $(\mathbb{C}^\times)^n$ on itself extends to X. Therefore there is an action of $\mathrm{U}(1)^n$ on X, which degenerates at some locus if X is not exactly $(\mathbb{C}^\times)^n$. It is interesting to study how this happens [LV98].

In order to built ourselves some intuition let us study two-dimensional examples first. The toric variety $\mathbb{C}^2$ is parameterized by $(z_1, z_2) = (e^{i\theta}|z_1|, e^{i\phi}|z_2|)$ and hence it is a torus fibration over the upper-right quadrant in the real plane parameterized by $|z_1|$ and $|z_2|$. Over the horizontal axis the θ circle S^1_θ vanishes, while S^1_ϕ shrinks over the vertical axis. This is depicted on the left of Figure 7.7. The toric variety $\mathbb{P}^2(\mathbb{C})$ can be described as the set of $(z_1, z_2, z_3) \in \mathbb{C}^3$ such that $|z_1|^2 + |z_2|^2 + |z_3|^2 = 1$ and modulo an overall $\mathrm{U}(1)$ action: $(z_1, z_2, z_3) \to e^{i\alpha}(z_1, z_2, z_3)$. Thus $\mathbb{P}^2(\mathbb{C})$ is parameterized by $(|z_1|^2, |z_2|^2, \phi, \theta)$ where $(z_1, z_2) = (e^{i\theta}|z_1|, e^{i\phi}|z_2|)$ and it is a torus fibration over the triangle $|z_1|^2 + |z_2|^2 \leq 1$ in the real plane parameterized by $|z_1|^2$ and $|z_2|^2$. Over the segment $|z_2|^2 = 0$ the circle S^1_ϕ shrinks, over the $|z_1|^2 = 0$ the circle S^1_θ shrinks and over $|z_1|^2 + |z_2|^2 = 1$ the circle $S^1_{\theta+\phi}$ vanishes. This is depicted on the right of Figure 7.7.

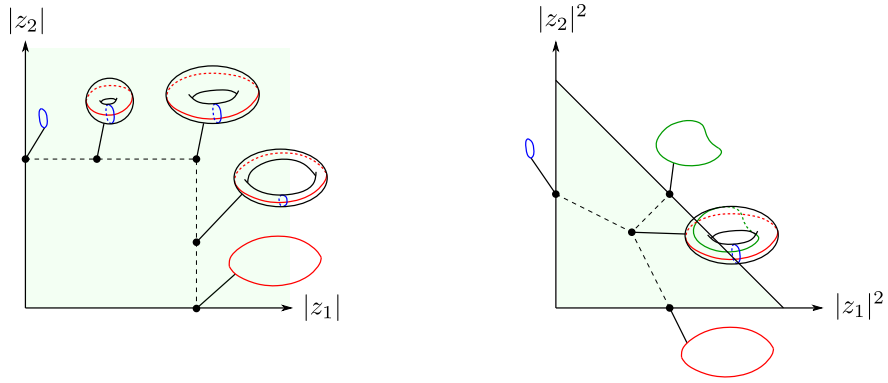


Figure 7.7: Two-dimensional toric varieties are torus fibrations over real surfaces.

The same holds for three-dimensional toric varieties. Let us consider the projections studied in [AKMV05] and reviewed in [Mar05] in two simple examples, namely flat space $\mathbb{C}^3$ and the resolved conifold[2]. Let z_1, z_2 and z_3 be the coordinates on $\mathbb{C}^3$, and let us consider the functions

$$r_1(z) = |z_1|^2 - |z_2|^2, \tag{7.59}$$
$$r_2(z) = |z_1|^2 - |z_3|^2, \tag{7.60}$$
$$r_3(z) = \mathrm{Im}(z_1 z_2 z_3) \ . \tag{7.61}$$

The functions r_1, r_2 and r_3 generate Hamiltonian flows on $\mathbb{C}^3$ endowed with its canonical symplectic form $\omega = \sum \mathrm{d}z_i \wedge \mathrm{d}\overline{z_i}$. These give us the fibration we are looking for: the base is parameterized by the values of the Hamiltonians r_1, r_2 and r_3, the flows corresponding to r_1 and r_2 correspond to

$$(z_1, z_2, z_3) \to \left(e^{i(\alpha+\beta)} z_1, e^{-i\alpha} z_2, e^{-i\beta} z_3\right) \tag{7.62}$$

and generate circles, and the one corresponding to r_3 generates a real line. One finds that the circle actions parameterized by α and β degenerate along the graph shown on the left of Figure 7.8, where the cycle denote (p, q) corresponds to the action of $p\alpha - \beta q$.

[2]In this case the toric varieties are presented as $T^2 \times \mathbb{R}_1$ fibrations over an $\mathbb{R}^2 \times \mathbb{R}_2$ base instead of T^3 fibrations over an $\mathbb{R}^3$ base, but the T^2 fiber appearing in $T^2 \times \mathbb{R}$ is exactly the one we are interested in and which leads to the (p, q)-web.

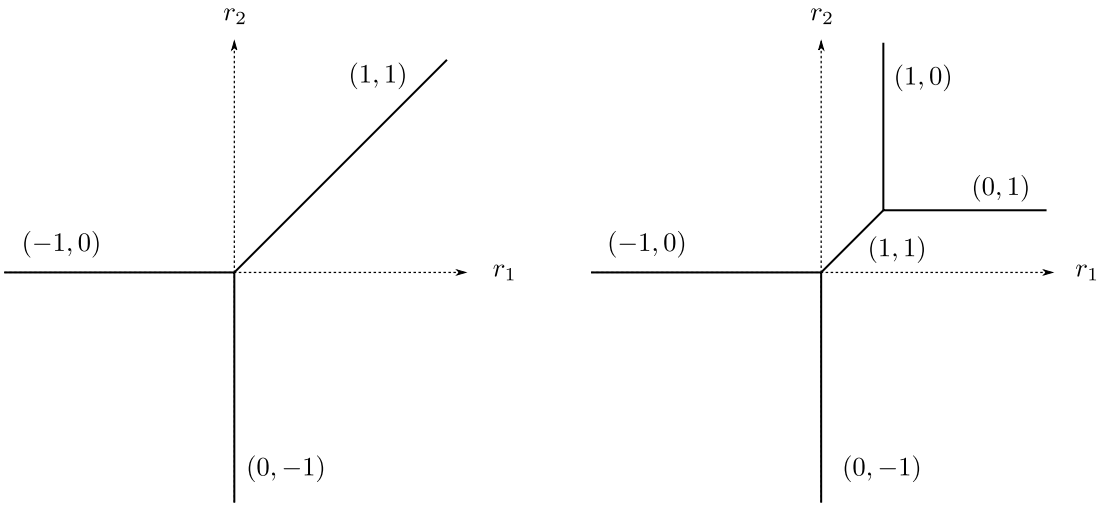


Figure 7.8: Degeneration of the torus fibrations in the case of $\mathbb{C}^3$ and the resolved conifold.

The case of the resolved conifold $\tilde{\mathcal{C}}_0$ works in a similar way. Let z_1, z_2, z_3, z_4 be complex coordinates such that $|z_1|^2 + |z_4|^2 - |z_2|^2 - |z_3|^2 = t$ and up to the U(1) action

$$(z_1, z_2, z_3, z_4) \to (e^{i\alpha} z_1, e^{-i\alpha} z_2, e^{-i\alpha} z_3, e^{i\alpha} z_4) \ . \tag{7.63}$$

The variables z_1 and z_4 can be taken as homogeneous coordinates on $\mathbb{P}^1(\mathbb{C})$ of which t parameterize the size, whereas the variables z_2 and z_3 parameterize the fibers of the bundle $\mathcal{O}(-1) \oplus \mathcal{O}(-1) \to \mathbb{P}^1(\mathbb{C})$. One can consider a first affine $\mathbb{C}^3$ chart on $\tilde{\mathcal{C}}_0$ defined by $z_4 \neq 0$ and parameterized by z_1, z_2 and z_3. Then the Hamiltonians

$$r_1(z) = |z_1|^2 - |z_2|^2 \tag{7.64}$$
$$r_2(z) = |z_1|^2 - |z_3|^2 \tag{7.65}$$

yield just as before the bottom left part of the figure on the right of Figure 7.8. On the other $\mathbb{C}^3$ patch on $\tilde{\mathcal{C}}_0$ defined by $z_1 \neq 0$ and parameterized by z_2, z_3 and z_4, the Hamiltonians rewrite as

$$r_1(z) = |z_3|^2 - |z_4|^2 + t \tag{7.66}$$
$$r_2(z) = |z_2|^2 - |z_4|^2 + t \tag{7.67}$$

and they give rise to the top right part of the figure on the right of Figure 7.8.

This method can be applied to any resolved toric CY3: starting with a triangulated toric diagram the piecewise linear graph in the (r_1, r_2) plane is the graph dual to the triangulated toric diagram. It is called (p, q)-web because of the analogy with the diagrams we have presented in Section 6.4.2. The examples of (p, q)-webs corresponding to resolutions of the affine toric CY3 F_0 and PdP_3 are shown in Figure 7.9. The corresponding toric diagrams are shown on the left of Figure 7.4 and on the right of Figure 7.5. Over an affine segment of the (p, q)-web of slope (p, q) in the (r_1, r_2)-plane the cycle (p, q) of the T^2 fibers degenerates.

The (p, q)-webs of fully resolved toric CY3 teach us a lot of information about the topology of the corresponding affine toric CY3. Consider the (p, q)-webs corresponding to a resolution of PdP_3 shown on the right of Figure 7.9. Bounded faces correspond to compact 4-cycles (which are the exceptional divisors of the resolution), unbounded faces correspond to non-compact 4-cycles whereas finite segments of the (p, q)-web correspond to compact 2-cycles. The singular limit of the resolution corresponds to shrinking all the segments and bounded faces appearing in the (p, q)-web, and hence the compact 2-cycles, compact 4-cycles and non-compact 4-cycles in the resolution vanish in the singular limit. Hence the knowledge of any (p, q)-web teaches us about the vanishing cycles at the singularity [Ber].

The fact that these diagrams resemble (p, q)-webs is not a mere coincidence [LV98, AKMV05]. M-theory compactified on a torus is dual to type IIB superstring theory compactified on a circle. Moreover

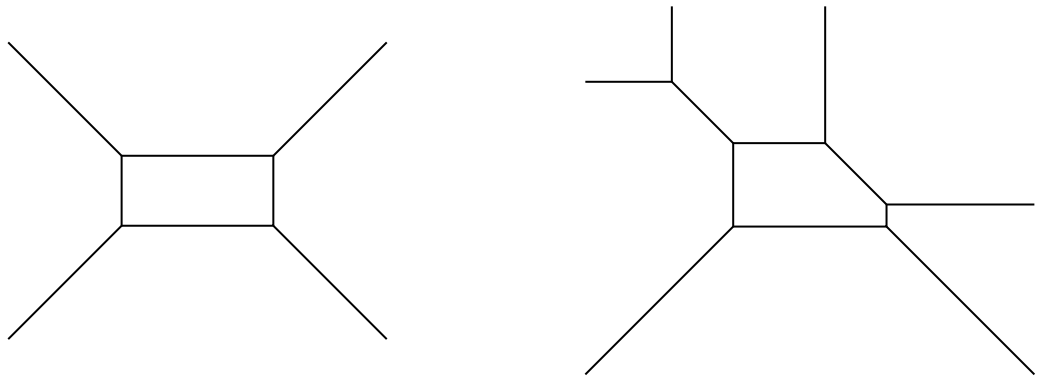


Figure 7.9: The (p, q)-webs of resolutions of F_0 and PdP_3.

when compactifying M-theory on an affine toric CY3 singularity the low-energy effective field theory is a $5d$ $\mathcal{N} = 1$ theory since the CY3 breaks one fourth of the 32 supersymmetries of M-theory. As we have reviewed an affine toric CY3 singularity can be seen as a T^2 fibration over $\mathbb{R}^4$, so that M-theory compactified on it is dual to type IIB theory in flat space compactified on a circle. Under the T-duality between type IIA theory and type IIB theory needed on the way from the M-theory description to the type II theory description, the degeneration of the T^2 fibration over $\mathbb{R}^4$ in the CY3 yields NS5 branes in the IIB picture.

7.5 Brane tilings

7.5.1 Generalities

We have seen in Section 7.3 that when a stack of D3 branes is placed at the singular point of an abelian Calabi–Yau orbifold $\mathbb{C}^3/G$ of flat space, its worldvolume theory is a four-dimensional $\mathcal{N} = 1$ quiver gauge theory with a superpotential. We have also seen in Section 7.4 that abelian Calabi–Yau orbifolds of $\mathbb{C}^3$ are special cases of affine toric CY3 singularities, hence the question: what is the worldvolume theory on a stack of D3 branes extending along x^0, x^1, x^2, x^3 and placed at the singular point of an affine toric CY3 singularity?

In order to answer this question, it is very convenient to introduce combinatorial objects known as *brane tilings*, which encode both the quiver describing the low-energy worldvolume theory as well as the superpotential. Brane tilings describe worldvolume theories which satisfy an additional assumption dubbed toricity: a four-dimensional $\mathcal{N} = 1$ quiver gauge theory is said to be toric if every chiral field appears exactly twice in the superpotential: once in a monomial with a positive coefficient and the other in a monomial with a negative coefficient.

Definition and a first example.

A *bipartite graph* is a graph (V, E) with V the finite set of vertices and E the finite set of edges, such that each vertex is either black or white and such that each edge links a black vertex to a white one.

Definition 7.3. *A brane tiling (or dimer model) is a bipartite graph Γ embedded in a topological surface S, and whose faces (i.e. the connected components of the complement of Γ in T^2) are topological disks.*

A brane tiling encodes a four-dimensional toric $\mathcal{N} = 1$ gauge theory in the following way: the faces of Γ correspond to the simple factors of the gauge group of the theory, edges are in one-to-one correspondence with chiral multiplets and nodes correspond to the monomials in the superpotential: those with a positive (resp. negative) coefficient give rise to white (resp. black) nodes.

As a first example, let us consider the low-energy worldvolume theory on D3 branes at the Calabi–Yau $\mathbb{C}^3/Z_5$ orbifold that we have derived in Section 7.3. It is toric: the quiver is shown on the right of Figure 7.1 and the superpotential is:

$$\begin{aligned} W =& Z_{13}X_{34}Y_{41} - Z_{13}Y_{35}X_{51} + Z_{24}X_{45}Y_{52} - Z_{24}Y_{41}X_{12} + Z_{35}X_{51}Y_{13} - Z_{35}Y_{52}X_{23} \\ &+ Z_{41}X_{12}Y_{24} - Z_{41}Y_{13}X_{34} + Z_{52}X_{23}Y_{35} - Z_{52}Y_{24}X_{45} \ . \end{aligned} \tag{7.68}$$

The brane tiling corresponding to this theory is presented in Figure 7.10. It is drawn on the fundamental cell of a torus: the red and blue edges on the boundary are identified with the orientation prescribed by the arrows on it. There are five faces on the brane tiling labeled $1, 2, 3, 4, 5$ consistently with the fact that the worldvolume theory has gauge group $\mathrm{SU}(N)_1 \times \mathrm{SU}(N)_2 \times \mathrm{SU}(N)_3 \times \mathrm{SU}(N)_4 \times \mathrm{SU}(N)_5$. The matter content of the theory consists of the chiral multiplets X_{ij}, Y_{ij} and Z_{ij} where $j = i + 1 \pmod 5$ for the X's and $j = i + 2 \pmod 5$ for the Y's and the Z's. These are in one-to-one correspondences with the edges of the brane tiling as follows. Since the brane tiling is a bipartite graph, its edges are oriented: one might for example agree on the fact that an arrow transverse to an edge in Γ always has the black end of this edge to its left. The diagram on the right of Figure 7.10 shows how such arrows crossing the edges of the brane tiling perpendicularly are associated to the chiral multiplets in the theory. For example the dashed arrow going from the face 1 to the face 2 on the left of Figure 7.10 corresponds to X_{12}. It is easy to check that the chiral multiplets are indeed in one-to-one correspondence with the edges of this brane tiling. The quiver dual to the brane tilings is hence exactly the quiver we have derived for a stack of D3-branes on $\mathbb{C}^3/\mathbb{Z}_5$ shown on the right of Figure 7.1.

As already emphasized, the brane tiling also encodes the superpotential of the theory: each black (resp. white) node corresponds to a term coming with a negative (resp. positive) coefficient in the superpotential. One considers the elementary sequence of arrows transverse to the edges of Γ which circles around a node. The orientation of this loop is counterclockwise at black nodes and clockwise at white nodes. The trace of the product of chiral multiplets corresponding to the arrows in the loop taken in the same order as the latter is the corresponding monomial in the superpotential (with a $\pm$ sign depending on the color of the node). For example, the two loops of dashed arrows shown on the brane tilings on the left of Figure 7.10 yields

$$W = -Z_{24}Y_{41}X_{12} + Z_{41}X_{12}Y_{24} \ , \tag{7.69}$$

where the trace implied.

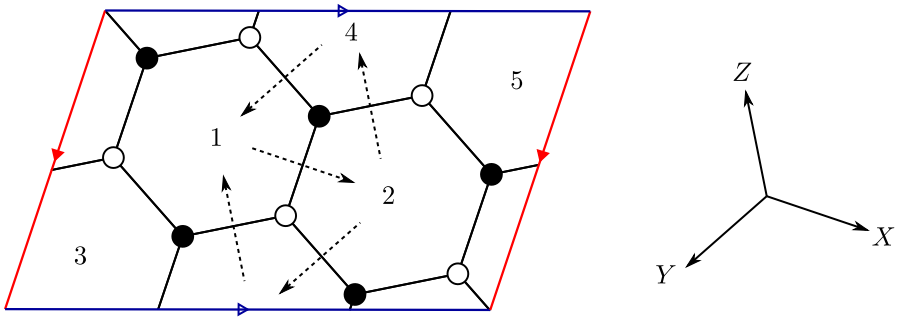


Figure 7.10: The brane tiling encoding the worldvolume theory on D3 branes at $\mathbb{C}^3/\mathbb{Z}_5$.

It is clear that any toric quiver gauge theory can be encoded in such a brane tiling. However it is not obvious whether any such bipartite graph on a surface S encodes the low-energy worldvolume theory on a stack of D3 branes at a singularity, and indeed there are some constraints elucidated in [HV07].

Consistency conditions for brane tilings and combinatorial tools.

The AdS–CFT correspondences that we will review in Chapter 8 implies that one expects the low-energy worldvolume theory on a stack of D3 branes at a general toric singularity to flow to a four-dimensional $\mathcal{N} = 1$ superconformal field theory (SCFT) [Mal99, Wit98a, AGM$^+$00]. The global symmetry of such an SCFT contains a $\mathrm{U}(1)_R$-symmetry which can be determined via the so-called *a-maximization* [IW03], where a is the SCFT central charge. It can be expressed in terms of the $\mathrm{U}(1)_R$ 't Hooft anomalies as in Equation (5.45). Let R_i be the R-charge of any UV field i in the theory i.e. any edge i in the brane tiling. One must have:

$$\sum_i R_i = 2 \ \text{ at each node}, \qquad \sum_i (1 - R_i) = 2 \ \text{ at each face}, \tag{7.70}$$

where the first sum runs over all edges of the brane tiling incident to a given node, and the second over all edges on the boundary of a given face. The first condition comes from the requirement to the

superpotential has R-charge 2, and the second translates the vanishing of the NSVZ beta function 5.40 at the superconformal fixed point of the renormalization group flow.

Let V, E and F be the set of nodes, edges and faces of a brane tiling on a closed surface S. Summing the first equation in Equation (7.70) over all nodes and the second over all faces yields:

$$2\sum_{i\in E} R_i = 2\#V \ , \qquad 2\#E - 2\sum_{i\in E} R_i = 2\#F \ . \tag{7.71}$$

Thus $\#F - \#E + \#V = 0$, and since the brane tiling is a cell decomposition of S:

Proposition 7.4. *A brane tiling on a closed surface S whose Euler characteristic is non-zero cannot describe a theory which flows to a superconformal fixed point at low energies.*

Hence S, if closed, must be either a torus T^2 or a Klein bottle. Let us assume for now that $S = T^2$, so that it is oriented.

Let us now introduce again another combinatorial object which will be of fundamental interest to us.

Definition 7.5. *A zig-zag path (ZZP) on a brane tiling Γ is an oriented path of edges of Γ which turns maximally right at each black vertex, and maximally left at each white one. One can equivalently represent zig-zag paths as strands crossing edges at their middle.*

The three ZZPs in the brane tiling of $\mathbb{C}^3/Z_5$ are shown in Figure 7.11.

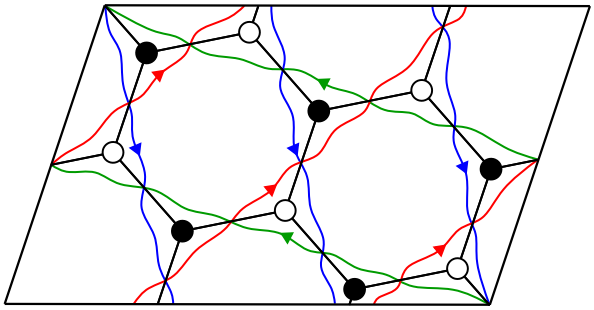


Figure 7.11: The ZZPs in the brane tiling of $\mathbb{C}^3/\mathbb{Z}_5$.

After reinterpreting (π times) the R-charges of chiral fields in a brane tiling Γ as angles in an isoradial embedding [Duf68, Mer01, Ken03], the authors of [HV07] show that some conditions must be imposed on the ZZPs of a brane tiling in order for it to be consistent. Other conditions were worked out later [Bro12, IU10] and can be summarized as follows.

Proposition 7.6. *A brane tiling is consistent if there is no homologically trivial ZZP, if no ZZP has a self-intersection in the universal cover and if no pair of ZZP intersect each other on the universal cover in the same direction more than once. In what follows, we will always assume brane tilings to be consistent.*

Bipartite graphs form an interesting class of combinatorial models on which one can study the statistical mechanics of so-called dimer models. In fact, dimer models were first introduced in chemistry in order to model the adsorption statistics of diatomic molecules on the surface of crystals. Let us briefly introduce the main characters of dimer models as well as the so-called *Kasteleyn matrix*.

Definition 7.7. *Let $\Gamma = (V, E)$ be a bipartite graph. A perfect matching PM on Γ is a subset of E such that each vertex in V belongs to exactly one edge in PM.*

All the perfect matchings on a (consistent) bipartite graph on T^2 are shown in Figure 7.12. In each case, the edges in the perfect matching are plain whereas the ones not in the perfect matching are dashed.

When Γ is embedded on an oriented closed surface S one can distinguish the perfect matchings according to their height. One first needs to choose a fundamental cell (this is implicitly already done in Figure 7.12). Let us assume for concreteness that $S = T^2$ and that the fundamental cell is a square with up, bottom, left and right sides. Since Γ is bipartite, every edge is oriented, for example from black to white. Then, one declares that each edge exiting the fundamental cell through the top (resp. bottom, left, right) is assigned the weight $\pm y$ (resp. $\pm y^{-1}$, $\pm x^{-1}$, $\pm x$), where the sign is given by the Pfaffian

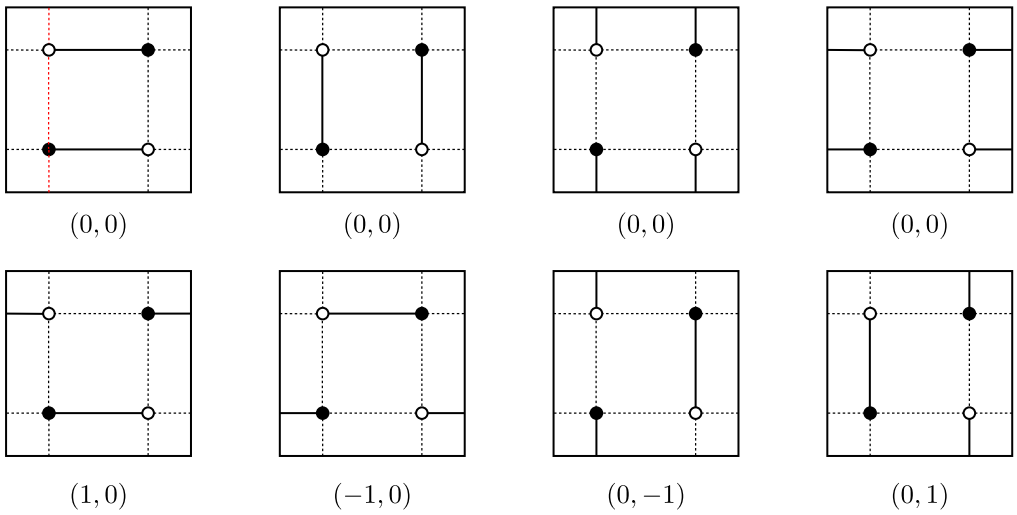


Figure 7.12: All the perfect matchings in a bipartite graph on T^2.

orientation of this edge. All other edges are assigned a weight ± 1, depending on their Pfaffian sign. The *height* of a perfect matching if the exponent in x and y of the product of these weights on all edges appearing in the perfect matchings. The height of each perfect matching is shown in Figure 7.12.

The Kasteleyn matrix is a convenient tool to count all the matchings on a bipartite graph Γ. In order to so, one first needs to choose a *Pfaffian orientation* of Γ, i.e. a choice of sign on each edge such that the number of $-$ signs as one go along a face of Γ is odd (resp. even) if the number of edges on the boundary of this face is 0 mod 4 (resp. 2 mod 4). In the example of Figure 7.12, one possible choice is to assign a $-$ sign to the two edges shown in red in the top left figure, and a $+$ sign to all the others. Let V_W (resp. V_B) be the set of white (resp. black) vertices in Γ. The Kasteleyn matrix K is the $\#V_W \times \#V_B$ weighted adjacency matrix of Γ, supplemented with the weights of before. The determinant of the Kasteleyn matrix counts perfect matchings with heights: the absolute value of the coefficient in front of the monomial $x^a y^b$ in it is the number of perfect matchings of height (a, b).

For example, the Kasteleyn matrix of the bipartite graph of Figure 7.12 is

$$K = \begin{bmatrix} -1 - y^{-1} & 1 + x \\ 1 + x^{-1} & 1 + y \end{bmatrix} , \tag{7.72}$$

where we have numbered black and white vertices from left to right. We obtain

$$\det K = -x - y - 4 - x^{-1} - y^{-1} , \tag{7.73}$$

which is consistent with our listing in Figure 7.12.

7.5.2 Forward and inverse algorithms

Motivated by groundbreaking articles on the AdS/CFT correspondence at the dawn of the 21st century – see Chapter 8, a lot of effort was put into the calculation of low-energy worldvolume theories on D3-branes at the tip of toric singularities, as a generalization of the simpler examples of the orbifolds of flat space [KS98, HU98] and of the conifold [KW98].

A method was proposed in [FHH01a] to derive the low-energy worldvolume theories on a D3 brane probing 'small' toric singularities, using *partial resolution*. One starts with the toric diagram describing the toric singularity of interest and one embeds it into the toric diagram of an abelian orbifold of $\mathbb{C}^3$, for which the low-energy worldvolume theory on a probe D3-brane is known from a perturbative analysis as in Section 7.3. By considering non-zero Fayet–Illiopoulos terms for some of the gauge groups in the latter theory, one partially resolves the abelian orbifold, and hence the worldvolume theory on the probe D3-brane is higgsed into a theory with less gauge groups and different matter content, which corresponds to the worldvolume theory on a probe D3-brane at a toric singularity whose toric diagram is strictly smaller than the one of the abelian orbifold of $\mathbb{C}^3$. It is possible to control how the toric diagram changes during this process, and hence one can obtain the worldvolume theories on probe D3-branes at any toric

singularity – at least in principle, since the algorithm is computationally very hard. In practice, it works well for affine toric CY3 singularities whose toric diagrams are small enough: in [FHH01a] the authors exhibit for example the quiver and the superpotential of the theories corresponding to the affine cones over F_0, dP_1, dP_2 and dP_3 whose toric diagrams are the lattice polygons displayed in Figure 7.4.

An interesting observation in [FHH01a] is that this procedure to determine the worldvolume theories on a D3-brane at a general affine toric CY3 singularity depends in general on choices which are such that the results are not unique: there are in general different theories corresponding to the same singularity. This phenomenon was dubbed *toric duality*. Shortly after it was proved that in fact toric duality is actually Seiberg duality in disguise [BP01, FHHU01].

The authors of [HK05] emphasized the connection between perfect matchings in dimer models and fields in the GLSM description of affine toric CY3 singularities (a mathematical proof of the correspondence was later given in [FV06]), and this motivated in turn the introduction of brane tilings [FHK+06, FHM+06]. Many examples were given in this article, such as the ones displayed in Figure 7.13.

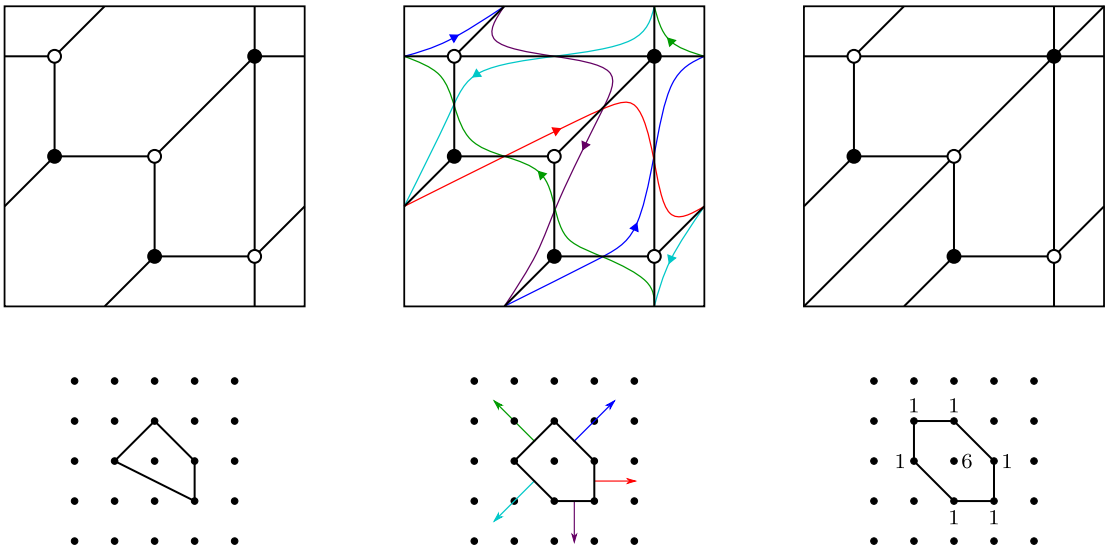


Figure 7.13: Brane tilings corresponding to the affine cones over dP_1 (left), dP_2 (middle) and dP_3 (right).

An important observation of [HV07] is that in all the examples of brane tilings known at that time, the zig-zag paths in the tilings are in one-to-one correspondence with the external legs of the (p, q)-web associated with the toric diagram, i.e. the outward pointing integral vectors normal to the sides of the toric diagram. The homology of a ZZP expressed in the symplectic basis of $H_1(T^2, \mathbb{Z})$ determined by our choice of fundamental cell is a pair of integers (a, b), which is the coordinates of the corresponding integral vector in the lattice of the toric diagram. An example of this is shown in the middle of Figure 7.13. This holds in general; it provides a very efficient way to compute the toric diagram of the singularity corresponding to a given brane tiling. Finding the affine toric CY3 singularity corresponding to a quiver gauge theory with superpotential is usually referred to as the *forward algorithm*.

Former forward algorithms consisted of computing the moduli space of the theory (which is very lengthy) or computing the determinant of the Kasteleyn matrix corresponding to the brane tiling which is a Laurent polynomial in x and y as in Equation (7.73), as introduced in [FHK+06]. The Newton polynomial of this determinant is exactly the toric diagram of the singularity. Let us note here that computing the determinant of the Kasteleyn matrix provides a quick consistency check for a brane tiling, for the multiplicities corresponding to the lattice points on the boundary of the toric diagram are severely constrained by the consistency conditions: the vertices of the toric diagram must have multiplicity one, while the lattice points in the sides of the toric diagram must have multiplicities satisfying Pascal's triangle. For instance, the multiplicities of the perfect matchings on the brane tiling corresponding to dP_3 on the right of Figure 7.13 are shown directly on the toric diagram below; they satisfy these conditions and hence the brane tiling passes this consistency check.

The link between ZZPs in brane tilings and the external legs of the corresponding (p, q)-web also provides a way to construct brane tilings from a given affine toric CY3 singularity. This is usually called the *inverse algorithm*, and the specific inverse algorithm using ZZPs and proposed in [HV07] is dubbed

the *fast inverse algorithm*, in contrast to the more complicated and computationally expensive method of partial resolutions on abelian orbifolds of $\mathbb{C}^3$. The fast inverse algorithm is based on the observation that white vertices (resp. black vertices) of the brane tiling are in one-to-one correspondence with connected components of the complement of the zig-zag paths in the torus T^2 whose sides are oriented clockwise (resp. counterclockwise). Faces of the brane tiling correspond in turn to connected components such that the orientation of their boundary alternates. See again the brane tiling in the middle of Figure 7.13 for an example. Hence starting with a set of simple closed curves on T^2 with homologies the coordinates of the outward pointing integral vectors normal to the sides of the toric diagram, one can look for an arrangement of these curves up to homotopy such that the connected components of the complement have their boundary which is either oriented clockwise, or counterclockwise, or such that its orientation alternate. Such a configuration is then dual to a brane tiling corresponding to the singularity we started with. This shows that the lattice to which the vertices of the polygon belong can be naturally identified with the integral first homology group of the torus which supports a brane tiling. The action of $\mathrm{SL}_2(\mathbb{Z})$ on this lattice translates into choices of fundamental cells for the torus.

More on brane tilings can be found in general introductions such as [Ken07, Yam08, FHSX17]. Dimer models also appear in other fields of mathematics and physics – for instance crystal melting and topological strings [IVNO08].

7.5.3 Some properties of brane tilings.

Massive fields and Seiberg duality.

There are two transformations of brane tilings that are of interest. The first is the *contraction of 2-valent nodes* as shown on the left of Figure 7.14. Physically, a 2-valent node represents a mass term $m^2\Phi_{ij}\Phi_{ji}$ in the lagrangian of the theory. Since we are mostly interested in the low-energy effective theory in the far IR, the corresponding chiral multiplets can be integrated out of the lagrangian, as happens effectively at energies smaller than m. The second is a combinatorial transformation known as *urban renewal*, displayed on the right of Figure 7.14. It encodes Seiberg duality at the corresponding face as explained in [FHK$^+$06], following the ideas developed in [Sch99]: in terms of the quiver, Seiberg duality is a mutation exactly as in Section 1.2. The superpotential varies in a way which has been formalized in [DWZ08] and which we have presented in Section 1.2.3. If the quiver with potential is represented as a brane tiling, Seiberg duality takes the form shown on the right of Figure 7.14. In terms of the spider moves ubiquitous in Part I, a urban renewal is a sequence of two spider moves possibly combined with additions or contractions of 2-valent vertices. Note that in order to preserve toricity so that the result of doing a Seiberg duality on a brane tiling is again a brane tiling, Seiberg duality can only be done on square faces.

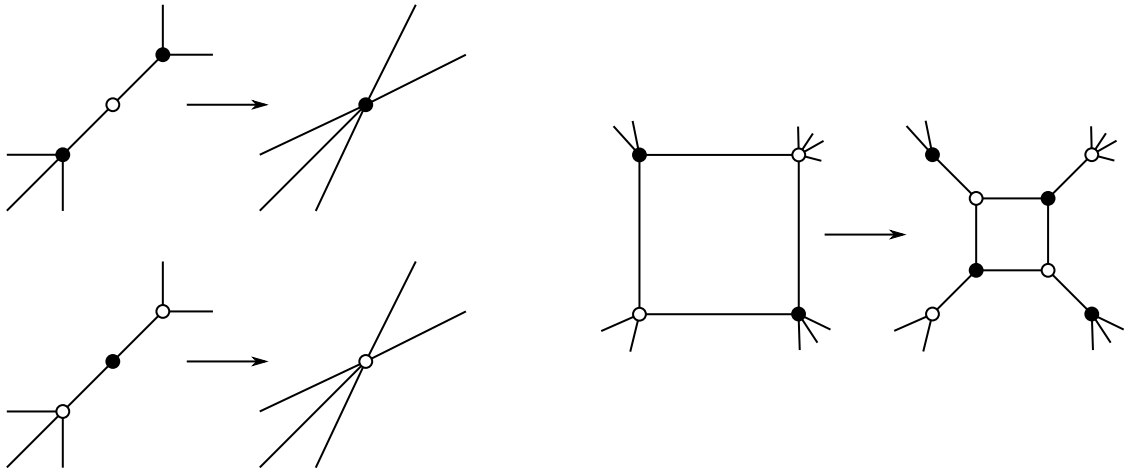


Figure 7.14: Contracting 2-valent vertices (left) and urban renewal (right).

Drawing the ZZPs which appear on the right of Figure 7.14 makes it clear that the set of ZZPs in a dimer model together with their homology is preserved under Seiberg duality. Hence two Seiberg dual brane tilings correspond to the same singularity, as expected.

Brane tilings as brane setups.

A important fact emphasized in [FHK+06] and developed in [FHKV08a] is that brane tilings are not merely combinatorial tools: they are in fact physical brane setups similar to the ones presented in Section 6.4. One starts with type IIB string theory in space-time $\mathbb{R}^{1,3} \times \mathbb{R}^8 \times T^2$, with x^4 and x^6 the coordinates along the two circle directions and where all other coordinates $x^0, \ldots, x^3, x^5, x^7, x^8, x^9$ parameterize a copy of $\mathbb{R}$. Let $z = \exp\big(x^5 + ix^4\big)$ and $w = \exp\big(x^7 + ix^6\big)$ be $\mathbb{C}^\times$ coordinates. The brane configuration is made of one NS5 brane extending along 0123 and wrapping a complex curve Σ defined by $P(z, w) = 0$ in the complex plane parameterized by z and w, and of a stack of N D5s extending along 012346. The brane tiling is the projection of this setup on the torus parameterized by x^4 and x^6. Faces correspond to D5 branes suspended between the NS5, edges correspond to chiral multiplets in the bifundamental representation of their adjacent faces while nodes are exactly the superpotential terms coming from the worldsheet instantons.

This brane configuration is actually dual under to T-dualities to a stack of D3 branes at an affine toric CY3 singularity. We have seen in the last section that such singularities are T^2 fibrations over an $\mathbb{R}^4$ base, and under T dualities along both directions of this T^2 the D3 branes (which are transverse to the singular toric CY3) are mapped to D5 branes wrapping the dual T^2 while the singularities in the torus fibration of the toric singularity are mapped to NS5 branes. More precisely, in the situation we are currently interested in, the affine toric CY3 singularity extends along the transverse space to $x^0, \ldots, x^3$. We take the directions corresponding to the T^2 to be parameterized by $x^{4'}$ and $x^{6'}$, and the $\mathbb{R}^4$ base of the T^2 fibration, by x^5, x^7, x^8, x^9. T-dualizing along $S^1_{4'}$ and $S^1_{6'}$ yields new circles S^1_4 and S^1_6 parameterized by x^4 and x^6. The brane tiling is drawn on the torus $S^1_4 \times S^1_6$. This is depicted schematically on the left of Figure 7.15.

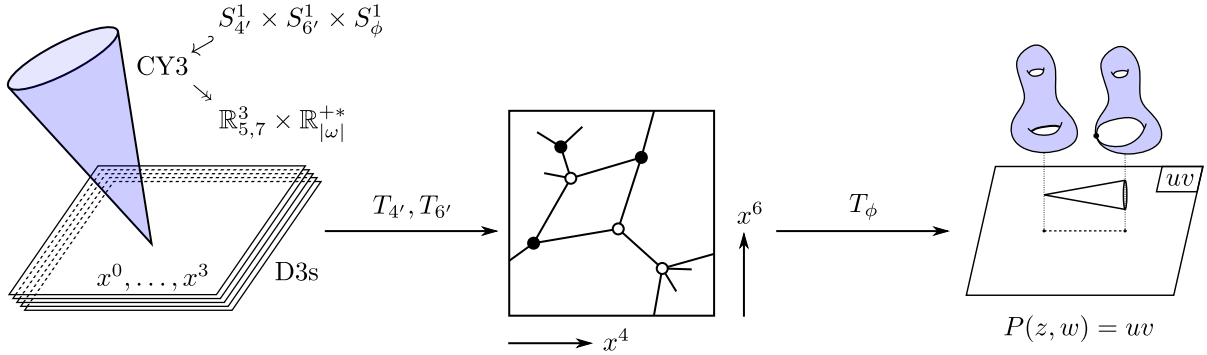


Figure 7.15: From D3 branes at a singularity to the mirror through brane tilings.

Let now $\omega = x^8 + ix^9 = e^{i\phi}|\omega|$. The circle S^1_ϕ is the last independent circle in the toric affine CY3 singularity. T-dualizing along it yields a space-time with again another topology, which is the mirror of the one we started with as follows from the SYZ approach to mirror symmetry [SYZ96]. Mirror symmetry in the context of affine toric CY3 singularities is known as *local mirror symmetry* [HV00, HIV00], and works as follows: the mirror to an affine toric CY3 singularity X described by a toric diagram Δ is the 3-dimensional complex subspace of $\mathbb{C}^\times_z \times \mathbb{C}^\times_w \times \mathbb{C}_u \times \mathbb{C}_v$ defined by

$$P(z, w) = uv \ , \tag{7.74}$$

where P is the Laurent polynomial whose Newton polygon is Δ. The surface defined by $P(z, w) = 0$ in the zw-plane is known as the *mirror curve* of X. The mirror geometry can be seen as a double fibration over a complex plane, which is why it is depicted as it is on the right of Figure 7.15. The T-duality along S^1_ϕ preserves the surface Σ wrapped by the NS5 in the brane tiling setup [FHKV08a], however since the direction of the T-duality is not in the worldvolume of the NS5, the latter becomes pure geometry. Thus, the surface Σ wrapped by the NS5 in the brane tiling is identified with the mirror curve of the original affine toric CY3 singularity. The D5 branes in the brane tiling picture become D6 branes wrapping singular 3-cycles of the mirror geometry.

The surface Σ can be derived directly from the brane tiling via the *untwisting map* [FHKV08a]. The bipartite graph Γ of the brane tiling also embeds naturally in Σ, however the cyclic orientation of the set of edges incident to each nodes (induced by the orientation of the underlying surface) differs for one type

of nodes, e.g. the black ones. Under this untwisting map ZZPs in the brane tiling are mapped to faces of the bipartite graph embedded in Σ, and vice-versa.

Brane tilings as a generalization of the McKay correspondence.

Let G be a discrete group and let V_i, $i = 1, \ldots, r$ be the set of irreducible representations of G. Let also V be the regular representation of G, and for all $i = 1, \ldots, r$ let:

$$V \otimes V_i = \bigoplus_{j=1}^{r} a_{ji} V_j \ . \tag{7.75}$$

The McKay quiver of the pair (G, V) is the quiver with r nodes, one for each irreducible representation V_i and a_{ij} arrow from the i-th node the the j-th one. McKay's observation of [McK80] when G is a discrete subgroup of SU(2) is that the McKay quiver of (G, V) is an affine Dynkin diagram of type ADE. The latter famously also appear in the resolutions of the simple surface singularities $\mathbb{C}^2/G$ where $G < \mathrm{SU}(2)$: in that case nodes correspond to the exceptional divisors and the edges, and edges are given by the intersection numbers between them. The results of [BKR99] imply that the derived category of coherent sheaves on any smooth resolution of $\mathbb{C}^2/G$ (which is exactly the category of B-branes on $\mathbb{C}^2/G$) is equivalent to the derived category of representations of the McKay quiver (with relations) [Asp04].

The McKay graphs which appear in resolutions of simple singularities $\mathbb{C}^2/G$ has a physical implementation in Hanany–Witten brane setups: the quiver encoding the low-energy four-dimensional $\mathcal{N} = 2$ theory in the worldvolume of D4 branes stretching between N NS5 branes on a circle is the affine Dynkin quiver of type $\hat{A}_{N_1}$, and the brane configuration is T-dual to D3 branes at the A_{N-1} singularity $\mathbb{C}^2/\mathbb{Z}_N$. Under another T-duality in the worldvolume of the NS5 branes the configuration is mapped to a brane tiling, whose dual graph is again the affine Dynkin quiver of type $\hat{A}_{N_1}$. The Hanany–Witten setup, the brane tiling as well as the corresponding quiver are displayed in Figure 7.16

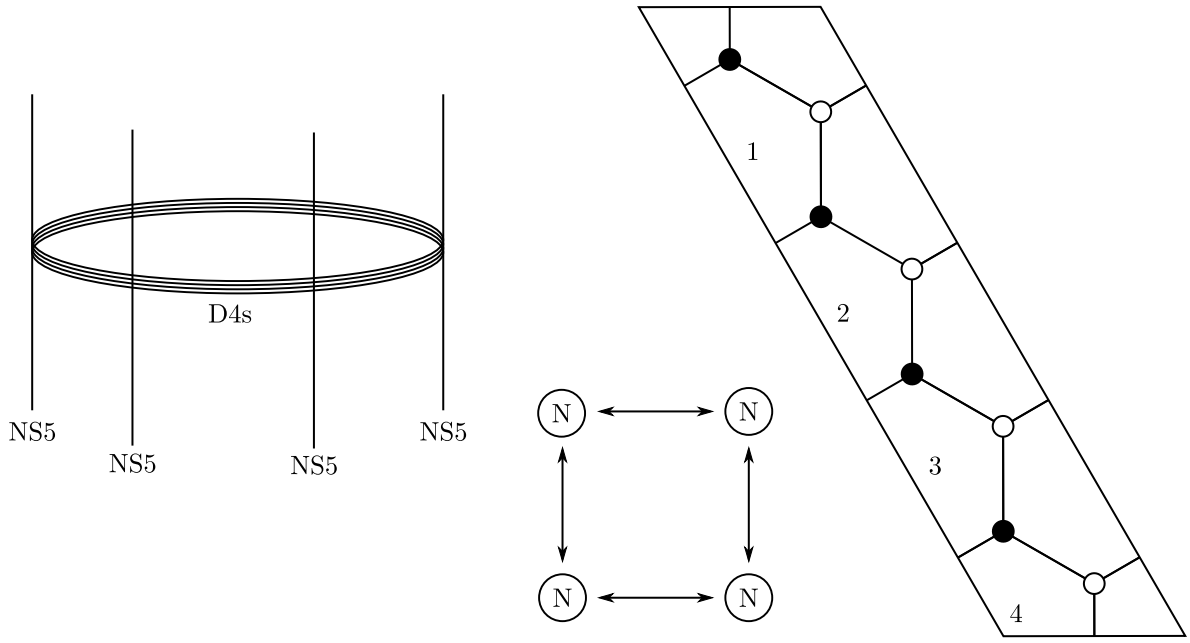


Figure 7.16: D4 branes streching between NS5s on a circle and the corresponding quiver.

General brane tilings can thus be seen as generalizations of the original McKay correspondence for discrete subgroups G of SU(3) when the brane tiling corresponds to a toric Calabi–Yau $\mathbb{C}^3/G$ orbifold, as well as for general affine toric CY3 singularities [FHK+06].

Brane tilings and geometry

The gauge invariant BPS mesonic operators in a four-dimensional $\mathcal{N} = 1$ quiver gauge theory correspond to closed loops in the quiver. When the theory is encoded in a brane tiling, these operators are naturally

defined on the latter as oriented strands crossing edges transversely and in a way compatible with the orientation of the edges of the dimer model. The geometry transverse to a D-brane is identified to the mesonic moduli space of its worldvolume theory [DM96, DGM97], and the gauge invariant BPS mesonic operators of a theory (modulo F-terms) are by definition well-defined algebraic functions on the mesonic moduli space.

When the geometry transverse to a D-brane is an affine toric CY3 singularity X defined by a cone $\sigma \subset \mathbb{Z}^3$, the ring of regular functions on X is a complex polynomial vector space generated by monomials in one-to-one correspondence with the lattice points in the dual cone $\sigma^\vee$. Hence the gauge invariant BPS mesonic operators in a theory described by a brane tiling corresponding to a singularity X are in one-to-one correspondence with the monomial generators of $\mathcal{O}_X$, i.e. with the lattice points in $\sigma^\vee$.

A Hilbert basis of $\sigma^\vee$ (or equivalently, a set of monomials generating $\mathcal{O}_X$ as a ring) hence admits an equivalent in terms of gauge invariant BPS mesonic operators: the mesonic operators modulo F-terms corresponding to the elements of a Hilbert basis of $\sigma^\vee$ are dubbed *fundamental mesonic operators* in [FHK+07]. Since any Hilbert basis of $\sigma^\vee$ contains the extremal generators which correspond to the sides of the toric diagram of the singularity X, i.e. to the ZZP in the brane tilings, the ZZPs are always fundamental mesons. However, there might be more: the example of the fundamental mesonic operators (modulo F-terms) on a brane tiling corresponding to the affine cone over dP_1 is presented in Figure 7.17. The plain paths are ZZPs are they correspond to extremal generators of $\sigma^\vee$ while the dashed one are non-extremal generators.

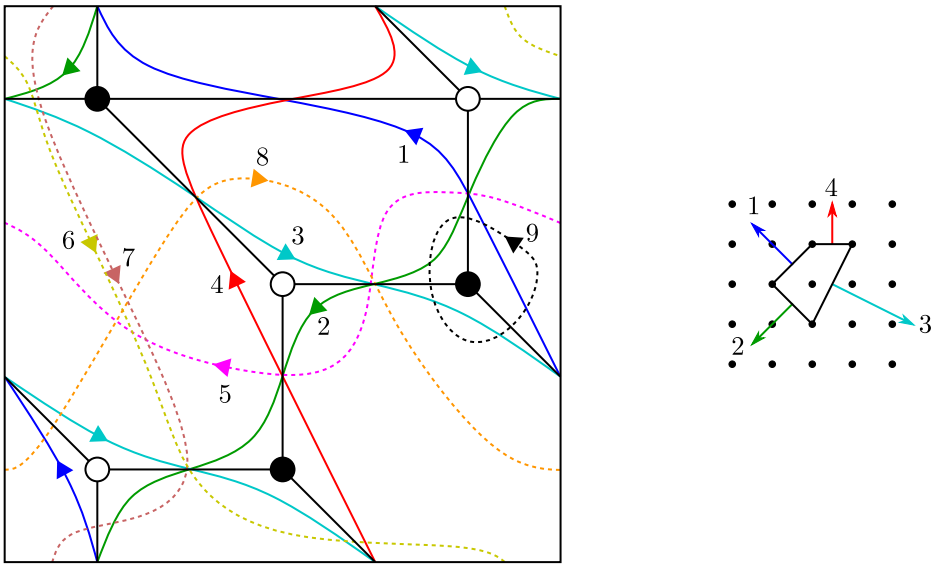


Figure 7.17: The fundamental mesons in a brane tiling corresponding to dP_1.

The paths corresponding to gauge invariant BPS mesonic operators modulo F-terms are entirely described in terms of their homology numbers since the F-term relations acts on the paths at each edge of the tiling as shown in Figure 7.18.

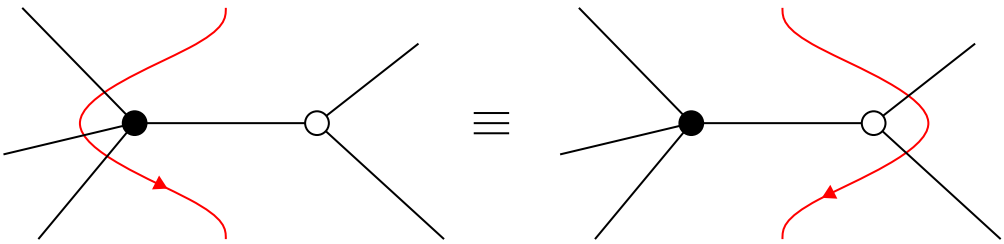


Figure 7.18: F-term relations seen as an equivalence relation between paths on the tiling.

The homologies of the paths corresponding to the 9 fundamental mesons shown in Figure 7.17 are:

$$h_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad h_2 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad h_3 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \quad h_4 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad h_5 = \begin{pmatrix} -1 \\ 0 \end{pmatrix},$$
$$h_6 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad h_7 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad h_8 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad h_9 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{7.76}$$

and one can show (with Macaulay2 for example) that

$$h_1 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \quad h_2 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \quad h_3 = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}, \quad h_4 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad h_5 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix},$$
$$h_6 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \quad h_7 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \quad h_8 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad h_9 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{7.77}$$

is a Hilbert basis of the dual cone corresponding to dP_1.

Isometries of the geometry transverse to a D-brane translate in general into global symmetries of its worldvolume theory. In the cases that we are interested in, the geometry X is an affine toric CY3 singularity and hence its Calabi–Yau metric admits a $\mathrm{U}(1)^3$ isometry.

Therefore the four-dimensional $\mathcal{N} = 1$ gauge theories encoded in brane tilings have a distinguished $\mathrm{U}(1)^3$ global symmetry. It decomposes into $\mathrm{U}(1)_F^2 \times \mathrm{U}(1)_R$, where each factor in $\mathrm{U}(1)_F^2$ is dubbed flavor mesonic symmetry. The charges of any gauge invariant mesonic operator in the brane tiling are easily read as the coordinates of the corresponding lattice point in the dual cone defining X [FHK+07].

7.5.4 Cluster integrable systems

A class of interesting algebraic integrable systems constructed out of bipartite graphs on T^2 satisfying exactly the same conditions as brane tilings (Proposition 7.6) has been constructed in [GK11].

The phase space admits $(\mathbb{C}^\times)^n$-charts glued along birational isomorphisms. Let us describe one such chart before explaining how they are glued together. Let $\Gamma = (V, E)$ be a consistent bipartite graph on T^2 and let $w : E \to \mathbb{C}^\times$ be a weight function on its edges. Each perfect matching D on Γ is assigned is energy:

$$\mathcal{E}_w(D) = \prod_{e \in D} w(e) \ . \tag{7.78}$$

Having chosen a Pfaffian orientation for Γ one writes the Kasteleyn matrix $K(\Gamma, w)$ as before but taking the weight of each edge supplemented by the weight function w. Let D_0 be a perfect matching corresponding to one corner of the Newton polygon corresponding to Γ. The normalized partition function of the dimer model is

$$Z'_{D_0} = \frac{\det K(\Gamma, w)}{\mathcal{E}_w(D_0)} \ . \tag{7.79}$$

The *normalized partition function* is (as the determinant of the Kasteleyn matrix) a Laurent polynomial in two variables x and y with Newton polygon Δ the one corresponding to Γ. Diving the determinant of the Kasteleyn matrix by $\mathcal{E}_w(D_0)$ has the nice consequence that the normalized partition function is independent of the choice of fundamental cell for T^2. One can rewrite Z'_{D_0} as:

$$Z'_{D_0} = \sum_{a,b \in \mathbb{Z}} \sum_{(a+a_0, b+b_0)} \frac{\mathcal{E}_w(D)}{\mathcal{E}_w(D_0)} x^a y^b = \sum_{a,b \in \mathbb{Z}} c_{ab}(w) x^a y^b \ , \tag{7.80}$$

where the second sum runs over all dimer configurations D whose height is $(a + a_0, b + b_0)$. Each term $\mathcal{E}_w(D)\mathcal{E}_w(D_0)^{-1}$ in the sum can be rewritten as $\mathcal{E}(D - D_0)$ where $D - D_0$ is the union of paths on Γ obtained by taking all edges in D with their standard orientation (from black to white) and all edges in D_0 with the opposite one. These paths form naturally oriented loops on Γ. For example, taking the reference perfect matching to be the left-most one in the bottom row in Figure 7.12 yields what is displayed in Figure 7.19. The pair of integers below each case is now the homology of the corresponding oriented loops on Γ.

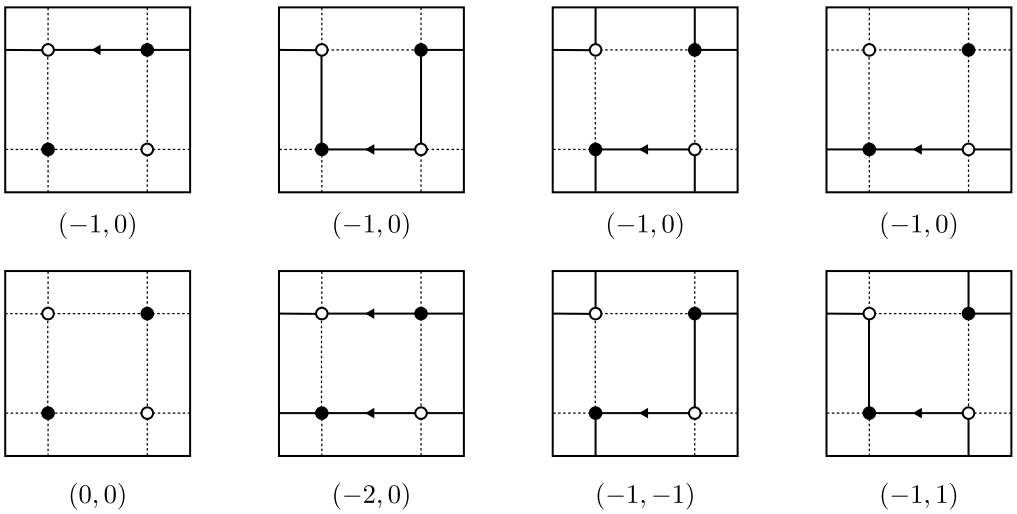


Figure 7.19: Union of closed loops from perfect matchings.

The weight function w is a 1-cochain on Γ. It is easy to check that the normalized partition function does not depend on w itself but merely on its cohomology class $[w] \in H^1(\Gamma, \mathbb{C}^\times)$. The space $H^1(\Gamma, \mathbb{C}^\times)$ is the chart $\mathcal{L}_\Gamma$ on the integrable system, corresponding to Γ. If one chooses two oriented loops γ_x and γ_y on Γ with homology $(1, 0)$ and $(0, 1)$ on T^2, then $H^1(\Gamma, \mathbb{C}^\times)$ is exactly given by the data of the $\mathbb{C}^\times$-monodromy along γ_x, γ_y and all the faces of Γ but one: $\mathcal{L}_\Gamma \simeq \mathbb{C}_x^\times \times \mathbb{C}_y^\times \times (\mathbb{C}^\times)^{\#F-1}$. One can show that the number of faces $\#F$ of Γ on T^2 is always twice the area of the lattice polygon Δ.

The Poisson structure on $\mathcal{L}_\Gamma$ is induced from the intersection pairing on the mirror curve: every closed loop on Γ can either be considered as a closed loop on T^2 or as a closed loop on Σ which is obtained from Γ via the untwisting map: starting from the bipartite fat graph Γ with the fat structure induced by the orientation of T^2 one inverts the cyclic orientation of the edges incident to each black vertex and reconstruct the corresponding surface with cusps Σ. The space of $\mathbb{C}^\times$-valued (algebraic) functions on $H_1(\Gamma)$ is naturally identified with $H^1(\Gamma, \mathbb{C}^\times)$, and hence the intersection pairing on $H_1(\Sigma)$ induces a Poisson bracket $\{\cdot, \cdot\}$ on $\mathcal{L}_\Gamma$.

The coefficients $c_{ab}(w)$ in the normalized partition function of Equation (7.80) are also naturally functions on $\mathcal{L}_\Gamma$. It is shown in [GK11] that the ones corresponding to lattice points on the boundary of Δ are Casimirs of the Poisson bracket, while the ones corresponding to lattice points in the interior of Δ are not Casimirs but mutually commute for the Poisson bracket: they form a family of independent Hamiltonians on $(\mathcal{L}_\Gamma, \{\cdot, \cdot\})$.

The *Pick theorem* [GS93] asserts that

$$2\mathcal{A}(\Delta) = 2i(\Delta) + e(\Delta) - 2 \ , \tag{7.81}$$

where $\mathcal{A}(\Delta)$ is the area of Δ, $i(\Delta)$ the number of lattice points in the interior of Δ and $e(\Delta)$ the number of lattice points on the boundary of Δ. The number $e(\Delta) - 2$ is exactly the dimension of the center of the Poisson algebra, and hence Pick theorem implies that the family of Hamiltonians of before is maximal: $(\mathcal{L}_\Gamma, \{\cdot, \cdot\})$ is an algebraic integrable system.

Theorem 4.7 in [GK11] shows that the rational transformations of the face weights appearing in $\mathcal{L}_\Gamma = \simeq \mathbb{C}_x^\times \times \mathbb{C}_y^\times \times (\mathbb{C}^\times)^{\#F-1}$ which preserve the modified partition function are uniquely determined to be sequences of the maps in Figure 7.20.

The underlying transformation of Γ is a spider move as introduced in Section 1.2.3. Under such elementary transformations of Γ one can keep track of the paths γ_x and γ_y, and one can glue the $(\mathbb{C}^\times)^{\#F-1}$ parts of $\mathcal{L}_\Gamma$ together along the birational transformation of Figure 7.20. The Poisson bracket as well as the Hamiltonians glue well under such identification and hence one can declare the total space of the integrable system corresponding to Γ to be the union of all $\mathcal{L}_{\Gamma'}$ where Γ' can be obtained from Γ via a sequence of spider moves, and with the identification provided by the cluster $\mathcal{X}$ mutations.

Attempts have been made to link these dimer integrable systems to brane tilings [Fra11, EFS12, FGH12, AFM12]. Dimer integrable systems have been generalized to integrable systems on double Bruhat cells of affine Lie groups in [FM16b]: dimer models encode double Bruhat cells in the affine Poisson-Lie

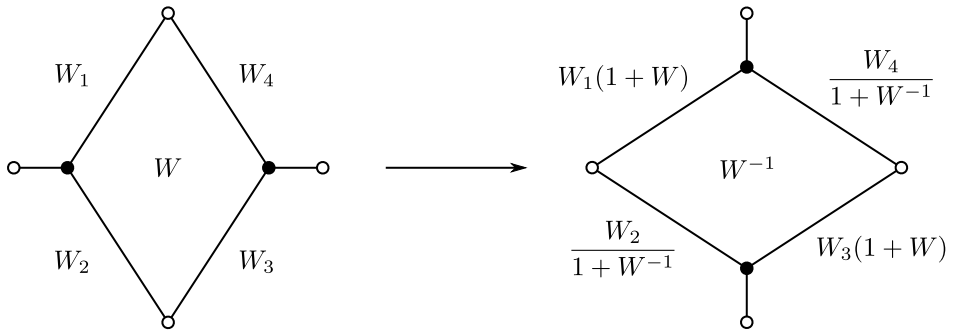


Figure 7.20: Elementary rational transformations of the face weights which preserve the partition function.

groups $\widehat{\mathrm{PGL}}(N)$ [FM16a].

Brane tilings are a very efficient way to encode the geometry of an affine toric CY3 singularity X: they are the result of two T-dualities in the T^2 fiber of X above $\mathbb{R}^4$ as in the previous section. They can also be used to study the worldvolume theory of M2 branes at the singular point of X, as initiated in [HZ08, HVZ09, DHMT09]. These M2 worldvolume theories are interesting generalizations of the ABJM theory introduced in [ABJM08]. Developments in the study of such three dimensional supersymmetric Chern–Simons theories have led to a conjecture for the exact quantization conditions of the integrable systems of Goncharov and Kenyon [HM16, FHMn16].

7.5.5 Fractional branes revisited

Fractional branes in dimer models admit a convenient combinatorial description in terms of zig-zag paths as first described in [But06]. Here we will refer to as fractional brane any bound state of regular and fractional branes, i.e. a rank assignment on the faces of Γ such that there is no gauge anomaly. For all face f of Γ let $N_f \in \mathbb{Z}_{\geq 0}$ the rank assigned to it. A rank assignment is a fractional brane if

$$\forall f \in F, \quad \sum_{f' \to f} N_{f'} - \sum_{f \to f'} N_{f'} = 0 \ , \tag{7.82}$$

where the arrows refer to the quiver dual to the brane tiling. This last equation is a generalization of Equation (7.27). Examples of fractional branes ot the singular point of the affine cone over dP_2 are shown on the two left-most brane tilings in Figure 7.21.

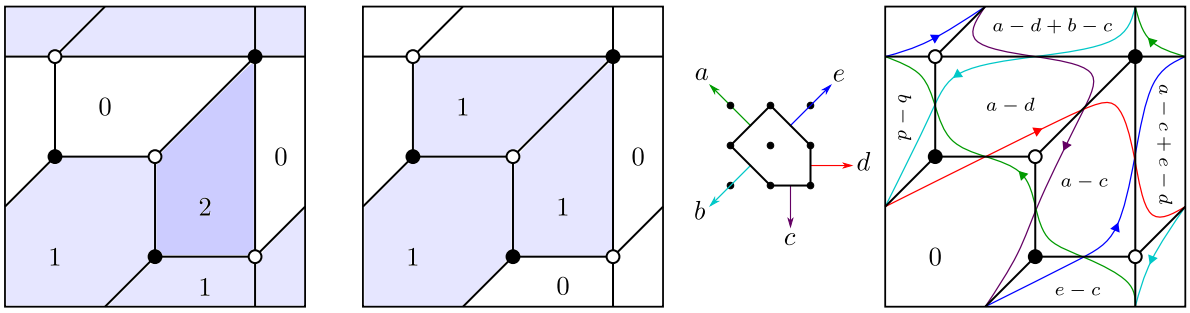


Figure 7.21: Fractional branes at the singular point of the affine cone over dP_2.

One can obtain all fractional branes as follows: let us consider an affine toric CY3 singularity described by its toric diagram, and label all outgoing integral normal vectors to the sides of the toric diagram corresponding to a singularity by $a, b, c, \ldots$. Recall that the zig-zag paths in any brane tiling corresponding to this singularity are in one-to-one correspondence with these outgoing integral normal vectors. Let us now fix a brane tiling associated with the singularity at hand, and assign the rank 0 to one of its faces. As one goes from this face to any of the adjacent ones, one has to cross two ZZPs going in opposite directions: the intersection between the arrow from the original face to the target and one of the ZZP is positive while the other is negative. Hence one can assign the formal difference of two ZZPs to the

target face. One can repeat this procedure again and again until all faces of the brane tiling are assigned a linear combination of ZZPs with integer coefficients, and such that the sum of all coefficients is zero. This is displayed on the right of Figure 7.21 for a brane tiling corresponding to the affine cone over dP_2.

Since any ZZP crossing an edge of the brane tiling in the boundary of a given face actually crosses two edges in the boundary of this face with opposite orientation, assigning integer weights $N_a, N_b, N_c, \dots$ to the outgoing integral normal vectors in the toric diagram always yields a rank assignment on the faces of the brane tiling such that the anomaly cancellation condition of Equation (7.82) holds. There are however constraints coming from the fact that the rank assignment on the faces of the brane tiling must be well defined: there must be no monodromy as one circles around the torus. Note that since ZZPs are closed loops on T^2, as one goes around the torus along a loop of homology (n, m) in some basis of $H_1(T^2, \mathbb{Z})$, one crosses a ZZP of homology (n', m') exactly $nm' - mn'$ times, where intersections are counted with sign. Hence the conditions in order to have well defined rank assignments can be read directly from the toric diagram since the coordinates $(a_1, a_2), (b_1, b_2), (c_1, c_2), \dots$ of the outgoing integral normal vectors are the homology of the corresponding ZZPs. One must have:

$$\sum_{k=a,b,c,\dots} k_1 N_k = 0 \ , \qquad \sum_{k=a,b,c,\dots} k_2 N_k = 0 \ . \tag{7.83}$$

For example, in the case of the affine cone over dP_2 displayed on the right of Figure 7.21 these conditions become:

$$N_a + N_e - N_b - N_c = 0 \ , \qquad N_e + N_d - N_a - N_b = 0 \ . \tag{7.84}$$

Considering integer assignments to a, b, c, d, e satisfying these conditions ensures that

$$N_a - N_c + N_e - N_d = N_b - N_d \ , \qquad N_a - N_d + N_b - N_c = N_e - N_c \ , \tag{7.85}$$

so that the rank assignment is indeed well-defined.

One result of [But06] is that assignments of integers to the outgoing integral normal vectors to the sides of a toric diagram satisfying Equation (7.83) correspond exactly to anomaly-free rank assignments on the faces of any brane tiling associated with this toric diagram.

Note however that this method 1) may yield faces with negative rank, which is unphysical 2) is blind to a global shift of the ranks on the faces, i.e. to regular branes. One can use the second point to shift the rank on all faces of the brane tiling to positive values. If one does that in a minimal way so that at least one face has rank zero after the shift, one obtains a purely fractional brane. On the contrary, if all the faces of the brane tiling have strictly positive rank, the configuration describes a bound state of purely fractional branes with regular ones.

7.6 Orientifolds

7.6.1 Generalities

Orientifolds of type II theories are perturbative quotients of some type II superstring theory in Calabi–Yau backgrounds under a symmetry group which contains the world-sheet parity reversal operator ω (hence the name). More precisely [IU12], let us consider one of the two type II superstring theories on $\mathbb{R}^{1,3} \times X_6$, where X_6 is a Calabi–Yau threefold (either compact or non-compact) with complexified Kähler form J and nowhere vanishing holomorphic three-form Ω_3.

A type IIA orientifold is defined by an action of the form

$$\omega \mathcal{R} (-1)^{F_L} \ , \tag{7.86}$$

where $R(-1)^{F_L}$ is needed in order for this action to be an involution, and where since type IIA is chiral on the worldsheet one requires $\mathcal{R}$ to be an antiholomorphic involution of X_6, i.e.

$$\mathcal{R}^* J = -J \ , \qquad \text{and } \mathcal{R}^* \Omega_3 = \overline{\Omega}_3 \ . \tag{7.87}$$

A type IIB orientifold is defined by an action of the form

$$\omega \mathcal{R} (-1)^{F_L} \ , \tag{7.88}$$

where $R(-1)^{F_L}$ is needed in order for this action to be an involution, and where since type IIB is non-chiral on the worldsheet one requires $\mathcal{R}$ to be an holomorphic involution of X_6.

The fixed locus of $\mathcal{R}$ in $\mathbb{R}^{1,3} \times X_6$ is, by definition, an *orientifold plane*. Similarly to D-branes, we will say an orientifold is an Op-plane when its world-volume is a $(p+1)$-dimensional Lorentzian manifold. In type IIA orientifolds one typically obtains O6 planes, whereas type IIB orientifolds might yield O3, O5, O7 and O9 planes. In general an Op-plane couples to the R–R $(p+1)$-form potential C_{p+1} and its charge can be computed through the calculation of the amplitude of the emission of an R–R closed string, in the zero momentum limit. This charge might yield to non-zero R–R tadpole diagrams, which must cancel for the consistency of the theory. This is achieved by adding some number of Dp-branes on top of the Op-plane.

The typical example is the construction of type I superstring theory as an orientifold of type IIB in $\mathbb{R}^{1,3} \times \mathbb{C}^3$ [Dab97]. The orientifold action is given by $\omega\mathcal{R}(-1)^{F_L}$ where $\mathcal{R}$ acts as the identity on $\mathbb{C}^3$, so that one is left with an O9-plane. The closed string sector of the orientifold theory contains unoriented closed strings which are invariant under ω, and the massless bosonic spectrum consists of the graviton field $G_{\mu\nu}$, the dilaton ϕ, the R–R 2-form C_2 and the symmetric product of the two gravitini. Type I superstring theory has indeed $\mathcal{N} = 1$ in ten dimensions.

The O9-plane has -32 units of charge with respect to the (non-dynamical) 10-form potential of type IIB, and hence the R–R tadpoles cancel provided one introduces 32 D9-branes on top of the O9. These additional branes induce an U(32) gauge symmetry in space-time, of which only the SO(32) subgroup is invariant under the action of ω.

In general in orientifolds of type IIB superstring theory on $\mathbb{R}^{1,3} \times X_6$, depending on the dimension of the fixed locus of $\mathcal{R}$ in X_6 one may end up with O3/D3, O5/D5, O7/D7 or O9/D9 brane systems. In O3/D3 or O7/D7 systems the geometric involution $\mathcal{R}$ is such that $\mathcal{R}^* J = J$ and $\mathcal{R}^* \Omega_3 = -\Omega_3$ whereas in O5/D5 and O9/D9 systems one has rather $\mathcal{R}^* J = -J$ and $\mathcal{R}^* \Omega_3 = \Omega_3$ [IU12]. Moreover, O3 and O7 branes preserve the same supersymmetries as D3 branes and hence one interesting way to generalize even more the gauge engineering techniques described so far is to consider configurations of D3 branes at the singular point of an affine toric CY3 singularity together with O3/O7 orientifolds and their companion D3/D7 branes needed to cancel the R–R tadpoles.

7.6.2 Orientifolds of brane tilings

Since the low-energy worldvolume theory on D3 branes at toric CY3 singularities are nicely encoded in brane tilings, it is of interest to understand how orientifolds can be implemented in brane tilings. This has been first studied in the seminal article [FHK$^+$07].

Starting with a brane tiling describing the low-energy worldvolume theory on a D3 brane at an affine toric CY3 singularity, the orientifolded theory is obtained by a $\mathbb{Z}_2$ identification of gauge groups, chiral multiplets and superpotential terms. This identification must therefore correspond to an involutive automorphism of the brane tiling, seen as a bipartite map on T^2. This automorphism may have a non-empty fixed locus, and each connected component of it can be assigned a sign $\pm$. Faces of the tiling might either be identified pairwise or self-identified under the $\mathbb{Z}_2$ automorphism. In the first case, let us fix one face of each pair and denote them $a, b, \ldots$, and their image: $a', b', \ldots$. Let also the self-identified faces be denoted $k, l, \ldots$. The theory after orientifold can be described as in [FHK$^+$07] and in the following way.

- Each selected face of the first type $a, b, \ldots$ gives rise to a gauge factor $\mathrm{U}(N_a)$,
- Each self-identified face $k, l, \ldots$ gives rise to a gauge factor $\mathrm{SO}(N_a)$ (resp. $\mathrm{Sp}(N_a/2)$) if it lies on a connected component of the fixed locus carrying a + (resp. $-$) sign,
- A pair of bifundamental chiral multiplet $(\square_a, \overline{\square}_b)$ and $(\overline{\square}_{a'}, \square_{b'})$ with $a \neq b'$ gives rise to a bifundamental chiral multiplet $(\square_a, \overline{\square}_b)$,
- A bifundamental chiral multiplet $(\square_a, \overline{\square}_{b'})$ together with its image give rise to a bifundamental $(\square_a, \square_b)$. Similarly, $(\square_{a'}, \overline{\square}_b)$ and its image give rise to $(\overline{\square}_a, \overline{\square}_b)$,
- A bifundamental chiral multiplet $(\square_a, \overline{\square}_{a'})$ gives rise to a chiral multiplet in the symmetric $\square\!\square_a$ (resp. antisymmetric $\begin{smallmatrix}\square\\\square\end{smallmatrix}_a$) representation of $\mathrm{SU}(N_a)$ if it lies on a connected component of the fixed locus with a + (resp. $-$) sign.

These rules make it clear that it is the fixed locus of the $\mathbb{Z}_2$ symmetry of a brane tiling which gives rise to non-unitary SO or Sp gauge groups as well as matter in rank-2 tensor representations. Being able to construct models with non-unitary gauge groups and matter in tensor representations is very interesting from a gauge engineering perspective, and hence the authors of [FHK+07] focused on three types of torus involutions which have a non-empty fixed locus, listed below.

- The π-rotation of the fundamental cell about a fixed point. It is always possible to choose the unit cell of the dimer in such a way that its corners coincide with a fixed point. Additionally, due to the dimer's toroidal periodicity, there will also be fixed points at the center of the boundaries of the unit cell, and in the center of the unit cell itself, see Figure 7.22.

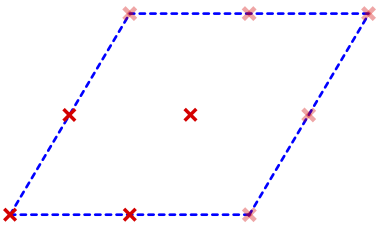


Figure 7.22: A schematic representation of a dimer unit cell with orientifold fixed points. The shaded points are the periodic images of the four basic ones.

- The reflection through a vertical or horizontal line: the unit cell of the dimer can be taken to be rectangular, and the dimer is invariant under a reflection leaving fixed the lines going along one of the boundaries of the unit cell. By the periodicity of the dimer, there must be a second fixed line parallel to the first one, and going through the middle of the unit cell. Vertical and horizontal fixed lines will be considered on the same footing here. This is depicted on the left of Figure 7.23.
- The reflection through a single diagonal line: the unit cell can be taken to have the shape of a rhombus, and the dimer is invariant under reflections about a fixed line which goes along one of the diagonals of the rhombus. The periodicity of the dimer does not imply the presence of other fixed lines in the unit cell. Again, we will not make the distinction between the two diagonals. This is depicted on the right of Figure 7.23.

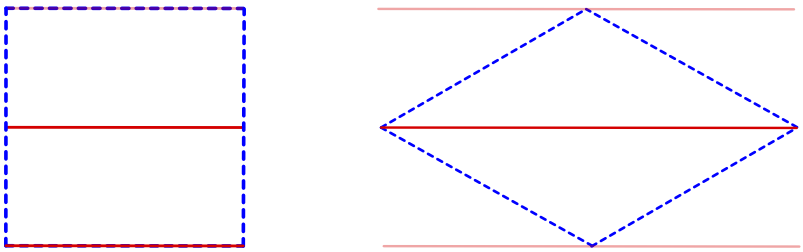


Figure 7.23: A schematic representation of orientifold fixed lines going through the dimer unit cell: two fixed lines on the left, a single fixed line on the right.

These involutions respectively have four fixed points, one fixed line or two fixed lines. In the following, we will use the two nomenclatures "double and single" or "horizontal/vertical and diagonal fixed lines" interchangeably. One example of each such involutions is displayed in Figure 7.24, with the toric diagram corresponding to the unorientifolded brane tiling and the quiver of the orientifold theory.

Some rules constraining possible involutions or the parity of mesonic operators under the orientifold have been worked out in [FHK+07]. For example, involutions with fixed points must map nodes of the tilings to nodes of the opposite color, while involutions with fixed lines must map nodes of the tiling to nodes of the same color (this comes from the perturbative definition of the orientifold only). Moreover, mesonic operators appearing in the superpotential must be odd under the orientifold action, in order for the orientifold to preserve $\mathcal{N} = 1$ supersymmetry. In the case of orientifold projections with four fixed points, the signs assigned to the fixed points must satisfy the sign rule: their product must be $(-1)^{n_W/2}$ where n_W is the number of superpotential terms [FHK+07].

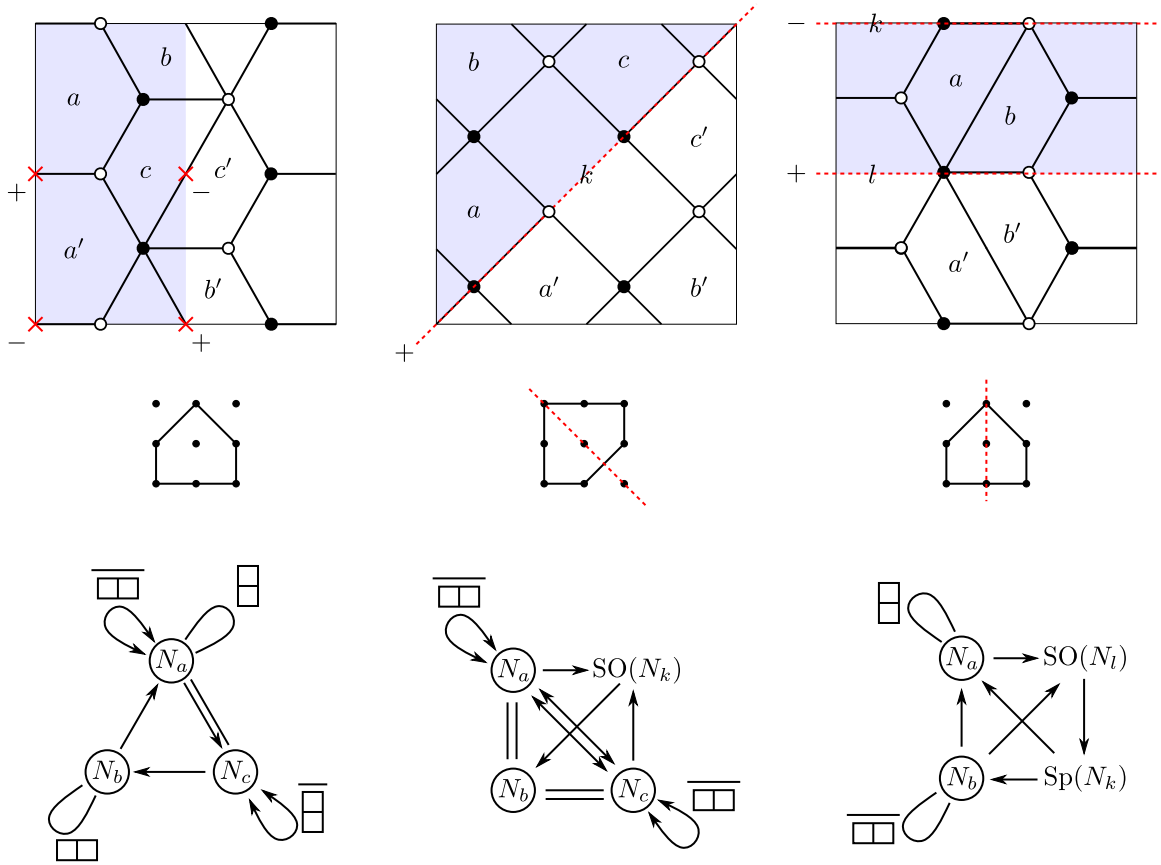


Figure 7.24: Orientifolds as $\mathbb{Z}_2$ involutions of dimers.

The signs carried by the connected components of the fixed locus of the torus involutions can be related to the charge of the corresponding O-planes [IKY08].

Constraints on the toric diagram

The existence of brane tilings admitting a given involution of one of the types of above puts constraints on the toric diagram of the corresponding singularity, as studied in [RU16b]. Under an involution of the torus underlying a brane tiling, the following holds:

- if the involution is a π-rotation, a ZZP of winding (p, q) is mapped to a ZZP of winding (p, q).
- if the involution is the reflection through a fixed diagonal line, a ZZP of winding (p, q) is mapped to a ZZP of winding $(-q, -p)$ or $(q, -p)$, depending on the orientation of the fixed line.
- if the involution has two fixed horizontal (resp. vertical) lines, a ZZP of winding (p, q) is mapped to a ZZP of winding $(-p, q)$ (resp. $(p, -q)$).

Hence, in order to possibly admit a brane tiling symmetric with respect to a diagonal (resp. horizontal, vertical) line, an affine toric CY3 singularities must admit a toric diagram symmetric with respect to a diagonal (resp. vertical, horizontal) line. This is not an equivalence: there exist toric diagrams with the good properties to possibly admit a brane tiling with a prescribed symmetry, but for which one can in fact prove that the latter cannot exist. We will present the general reasoning leading to this result as well as an explicit example in Section 10.3.

* * * * * * *

In order to make contact with real-world macroscopically-four-dimensional physics, one needs to come up with ways to go from the ten dimensions of string theory to four dimensions.

One way to do so is to compactify string theory on six-dimensional Riemannian manifolds. The condition to preserve some amount of supersymmetry below the compactification scale, which is an appealing assumption both phenomenologically and technically, translates into the Riemannian manifold being a Calabi–Yau manifold. One can turn on quantized fluxes as one compactifies, which provides ways to obtain fully-fledged stable string vacua with few or none massless moduli.

Another way to obtain four-dimensional physics from type IIB string theory is to consider D3-brane worldvolume theories. When the branes are placed at the singular points of Calabi–Yau varieties, their worldvolume hosts a four-dimensional (non-gravitational) quantum field theory with $\mathcal{N} = 1$ supersymmetry. Among the simplest examples of such setups are the Calabi–Yau orbifolds of $\mathbb{C}^3$, for which the worldvolume theory of branes at their tip can be computed perturbatively. Another interesting singularity is the conifold, which is defined by the polynomial equation $xy - zw$ in $\mathbb{C}^4$.

These examples are special cases of the more general class of affine toric Calabi–Yau singularities, which can be conveniently described with the tools of toric geometry. The worldvolume theory of D3-branes at such a singular point is a four-dimensional $\mathcal{N} = 1$ quiver gauge theory, with the quiver encoding the simple gauge factors and the chiral matter fields of the theory. The latter is actually entirely described by its quiver and its superpotential, which is a polynomial in the matter fields. When the theory satisfies the toricity assumption, both the quiver and the superpotential can be encoded as a bipartite graph on the torus T^2. This graph is the dimer model – or brane tiling – of the theory.

There are, in general, many brane tilings corresponding to a given affine toric Calabi–Yau threefold. They describe the different toric phases of the theory on D3-branes transverse to this singularity. Any two such toric phases can be related through a finite sequence of Seiberg dualities, which in terms of brane tilings correspond to urban renewals or spider moves, with the ranks on the faces transforming according to the tropical $\mathcal{A}$-mutation formula.

Lastly, one can describe orientifolds of brane tilings as involutions of the bipartite maps. Those with a non-empty fixed locus display either a fixed line, two fixed lines or four fixed points in the brane tiling. This in turn opens up the possibility of describing orthogonal or symplectic gauge factors, as well as chiral matter fields in rank-2 tensor representations of the gauge factors. Therefore, it enlarges a great deal the class of four-dimensional $\mathcal{N} = 1$ theories which can be given such a stringy ascendancy.

Chapter 8

Holography and gauge–gravity dualities

The idea of holography arose after it was understood that the area of classical black holes – e.g. Schwarzschild ones – behaves in a very similar way to the entropy of thermodynamic systems since it can only increase, which happens when matter falls inside the black hole. Bekenstein viewed this fact as something deeper than a mere analogy: the entropy of a black hole should – just as in classical thermodynamics – count the possible microstates of some sort corresponding to a given macroscopic black hole [Bek73, Bek74]. The (classical) general relativity "no-hair" theorem states that black holes are entirely described by their mass, their electric charge and their angular momentum. Therefore in order to speak of black hole microstates it is necessary to use a theory of quantum gravity. Even without a full knowledge of the latter, investigating the thermodynamics of black holes led to compelling yet highly paradoxical pictures.

What has been learned from the deep insights of Bekenstein and Hawking [Haw75, Haw76] is that black holes have an *entropy* and a *temperature*. For example the entropy of a Schwarzschild black hole of mass M is

$$S_{\mathrm{BH}}(M_0) = \frac{k_B A c^3}{4\mathcal{G}\hbar} \ , \tag{8.1}$$

which is one fourth of the area $A = 4\pi r_S^2$ of the horizon (where $r_S = 2\mathcal{G}M$) measured in Planck units, and its temperature is

$$T_{\mathrm{H}} = \frac{\hbar c^3}{8\pi k_B \mathcal{G} M} \ . \tag{8.2}$$

One of the main point of Bekenstein and Hawking is that the entropy of Equation (8.1) is the maximal amount of entropy that a region of space-time of radius r_S can contain. Interestingly, this entropy scales as r_S^2 and not as r_S^3 as one would ingenuously expect.

Since black holes have a temperature, they radiate energy away (if their surroundings are colder): this process is called *black hole evaporation*. For a solar mass black hole, the entropy is huge (around 10^{60} erg $\cdot K^{-1}$, to be compared with the entropy of the sun which is of 10^{42} erg $\cdot K^{-1}$ [Bek73]) and the temperature tiny ($\sim 10^{-7}$ K, whereas the temperature of the cosmic microwave background is ~ 2.73 K). Were a solar black hole in empty space (without cosmic microwave background), it would take around 10^{66} years to evaporate completely.

Black hole evaporation leads to the *black hole information paradox*. Hawking's calculation seems to show that the entanglement entropy of the radiation emitted away by the black holes increases monotonically with time since the radiation seems to be emitted in a thermal state. If this indeed holds, after the full evaporation of a black hole one is left with a final state which has a huge non-zero entanglement entropy. If the black hole has been created by the collapsing of a pure quantum state (whose entanglement entropy is zero), this seems to go against the fact that quantum evolution is unitary. Three options have been considered to explain this phenomenon:

1. Either the black hole does not completely disappear after the evaporation process and one is left with a Planck-sized *remnant* which has also a huge non-zero entropy, so that the total entanglement entropy of the system is zero. This is not satisfactory since it violates the Bekenstein–Hawking entropy bound.

2. Or the final state after the black hole creation from a pure quantum state and its evaporation is not a pure state, i.e. it has non-zero entanglement entropy. This is not satisfactory since it violates the unitarity of quantum evolution.

3. Or Hawking's formula for the entanglement entropy of the outgoing radiation when the black hole evaporates is valid in a coarse grained sense, i.e. the radiation does not come out as a mixed state exactly. This is not satisfactory since it is in contradiction with Hawking's calculation, which seemingly relies on very consensual principles.

't Hooft [tH93] and Susskind [Sus95] impulsed the scrutiny of option 3. and proposed that the degrees of freedom of the four-dimensional black hole are equivalently encoded on the three-dimensional (stretched) horizon, motivated by the fact that bits of hidden information making the black hole's entropy seem to correspond to elementary (Planckian) units of surface on the black hole's horizon. Since an optic hologram stores the data of a three-dimensional image on a two-dimensional holographic film, this was dubbed the *holographic principle*. The idea is that the quantum-gravitational theory describing the black hole should be equivalent to a non-gravitational quantum theory in one-dimension less, and that the evolution of the latter theory should describe the black hole's evaporation in a manifestly unitary way. The holographic principle found its first explicit incarnation in the AdS–CFT correspondence proposed by Maldacena in [Mal99], as we will review in the next section.

The developments that led to the holographic principle are reviewed in a popularized way in [Sus08]. More on black hole information with a view towards the AdS–CFT correspondence can be found in [Har16, Har18].

Another idea that was realized through AdS–CFT correspondences, or more generally through string–gauge correspondences, was the expectation that large N expansions of gauge theories were somehow related to string theories [tH74, Col80, Pol98]. Consider for example the generalization of QCD to the gauge theory with gauge group $\mathrm{SU}(N)$ and N flavors of quarks in the (anti)fundamental representation of the gauge group. Denoting $\lambda = g_{\mathrm{YM}}^2 N$ the 't Hooft coupling of the theory, the smaller λ, the better the perturbative description of the theory via Feynman diagrams. The large N expansion is the limit where $N \to \infty$ while λ is kept to a fixed value.

Assuming that λ is small, each Feynman diagram can be represented as a *fat graph*, i.e. a graph with a cyclic ordering of the set of edges incident to each vertex, provided each gluon propagator is represented as a double line. Fat graphs have vertices V, edges E and faces F, i.e. they can be naturally embedded on surfaces. This is the so-called 't Hooft double-line representation. Each diagram is associated with a power of N, and in the large N expansion the higher this power the more a diagram contributes. 't Hooft noticed that the power of N of a graph is $F - E + V$, which is the Euler characteristic of the surface corresponding to the fat graph under consideration. Therefore, the amplitude of a process described by the theory can be written as a genus expansion which is very reminiscent of the one of string theories.

We present the original AdS–CFT correspondence in Section 8.1, its generalizations to more general geometries in Section 8.2 and to non-conformal gauge–gravity correspondence in Section 8.3.

8.1 $\mathrm{AdS}_5 \times S^5$ / $\mathcal{N} = 4$ $d = 4$ super Yang–Mills correspondence.

The idea in [Mal99] is to consider a stack of branes in type II string theory of M-theory whose worldvolume theory is a conformal field theory (CFT). The three original examples consist of stacks of D3-branes in type IIB string theory, M2-branes or M5-branes in M-theory. We will be mostly interested in the type IIB case, and hence we will restrict to it from now on. We follow the presentation of [AGM$^+$00] and [BBS06].

Let us consider a stack of N coincident D3-branes in type IIB string theory in $\mathbb{R}^{1,9}$, extending along x^0, x^1, x^2, x^3 and localized along $x^4, \dots, x^9$. At low energies $\alpha' \to 0$, the stack of branes is well approximated by the extremal black 3-brane solution of type IIB supergravity given in Equation (6.26), with N units of charge. The metric is:

$$\mathrm{d}s^2 = f_3^{-1/2} \mathrm{d}x \cdot \mathrm{d}x + f_3^{1/2} \mathrm{d}y \cdot \mathrm{d}y \ , \tag{8.3}$$

where x stands for the coordinates along the branes and y for those along the transverse space $\mathbb{R}^6$. Let us also introduce spherical coordinates in transverse space: $\mathrm{d}y \cdot \mathrm{d}y = \mathrm{d}r^2 + r^2 \mathrm{d}\Omega_5^2$. The function f_3 only

depends on r, and:

$$f_3(r) = 1 + \left(\frac{R}{r}\right)^4 , \qquad R^4 = 4\pi\alpha'^2 g_s N . \tag{8.4}$$

The dilaton is constant and hence there is no ambiguity in the definition of the string coupling constant. The R–R field strength is

$$F_5 = N(\omega_5 + \star\omega_5) , \tag{8.5}$$

where ω is the volume form on the 5-sphere at constant radius in transverse space. The metric of Equation (8.3) has a horizon at $r = 0$. In the near-horizon limit i.e. when r is small, the metric takes the form

$$\mathrm{d}s^2 = \left(\frac{r}{R}\right)^2 \mathrm{d}x \cdot \mathrm{d}x + \left(\frac{r}{R}\right)^2 \mathrm{d}r^2 + R^2 \mathrm{d}\Omega_5^2 , \tag{8.6}$$

which can be rewritten with the new variable $z = R^2/r$ as:

$$\mathrm{d}s^2 = R^2 \frac{\mathrm{d}x \cdot \mathrm{d}x + \mathrm{d}z^2}{z^2} + R^2 \mathrm{d}\Omega_5^2 . \tag{8.7}$$

Hence near the horizon the geometry is $\mathrm{AdS}_5 \times S^5$, where both AdS_5 and S^5 have radius R. The supergravity approximation is justified when stringy effects can be neglected: $R \gg l_s$, which is equivalent to $g_s N \gg 1$ because of Equation (8.4).

As for usual four-dimensional black holes there is a potential barrier separating the near-horizon region and the exterior one. It has been computed in [GKT97, Kle97] that the absorption cross-section at low energies goes like $\sim \omega^3 R^8$, where ω is the energy of a mode send from infinity towards the black brane. Hence at low energies the system of massless modes outside the potential barrier (i.e. in flat space) and the one of massless modes close to the horizon (i.e. in $\mathrm{AdS}_5 \times S^5$) decouple, and the physics at low energies is schematically:

$$\text{Free gravitons in flat space} \oplus \text{type IIB supergravity on } \mathrm{AdS}_5 \times S^5. \tag{8.8}$$

The limit we have loosely referred to as *low-energy limit* is more precisely described as the *decoupling limit* [Mal99]. If one considers a stack of D3-branes separated by a distance r, the decoupling limit is the one in which $\alpha' \to 0$ while r is modified accordingly, so that the masses of strings streching from one brane to one of its neighbors are fixed, i.e. r/α' is constant.

The stack of D3-branes in flat space can also be considered from the perspective of string theory, in which open strings end on the D3's and induce a low-energy four-dimensional $\mathcal{N} = 4$ super Yang–Mills theory on their worldvolume. Closed string excitations couple to the D3-branes but are free to move away from them i.e. in the *bulk*. There are interaction terms between the open string modes and the closed string modes, but they vanish in the low-energy limit. Hence the physics at low energies is schematically:

$$\text{Free gravitons in flat space} \oplus d = 4, \mathcal{N} = 4 \text{ U}(N) \text{ super Yang–Mills.} \tag{8.9}$$

Comparing Equation (8.8) and Equation (8.9) leads to the conjectural AdS–CFT correspondence:

Type IIB supergravity on $\mathrm{AdS}_5 \times S^5 \longleftrightarrow \mathcal{N} = 4$ super Yang–Mills in four dimensions.

This is the weakest formulation of the AdS–CFT correspondence. In fact, as argued in [Mal99], the duality should hold beyond the supergravity approximation:

Type IIB superstrings on $\mathrm{AdS}_5 \times S^5 \longleftrightarrow \mathcal{N} = 4$ super Yang–Mills in four dimensions.

It turns out that as one considers a stack of N D3-branes the U(1) subgroup of U(N) does not participate in the duality, which is therefore refined to a duality between type IIB supergravity or superstrings on $\mathrm{AdS}_5 \times S^5$ and $\mathcal{N} = 4$ SU(N) super Yang–Mills in four dimensions. On the gravity side there are N units of F_5 flux through the sphere S^5.

It is known that $\mathcal{N} = 4$ SU(N) super Yang–Mills in four dimensions is a UV-finite superconformal field theory. The gauge coupling g_{YM} does not run, and perturbation theory is justified when the *'t Hooft coupling*

$$\lambda = g_{\mathrm{YM}}^2 N \tag{8.10}$$

is small. From Equation (6.23) one knows that $g_{\mathrm{YM}}^2 = 2\pi g_s$ and hence

$$\lambda = \frac{R^4}{2l_s^2}. \tag{8.11}$$

The supergravity approximation on the gravity side of the correspondence is justified when the 't Hooft coupling of the super Yang–Mills theory is large, i.e. when the gauge theory is in a non-perturbative regime. Conversely, when the gauge theory is in a perturbative regime the dual gravitational theory is strongly coupled and is in a stringy regime. It is believed that $\mathcal{N} = 4$ SU(N) super Yang–Mills in four dimensions provides a non-perturbative definition of type IIB superstrings in $\mathrm{AdS}_5 \times S^5$.

Sending N to infinity while keeping λ fixed is known as the large N expansion of the gauge theory and can be studied through the so-called $1/N$ expansion [Col85]. On the gravity side since

$$\frac{\lambda}{2\pi N} = g_s \ , \tag{8.12}$$

this limit corresponds the perturbative limit of type IIB superstrings, i.e. the large N expansion of $\mathcal{N} = 4$ SU(N) super Yang–Mills in four dimensions at fixed 't Hooft coupling corresponds to the perturbative genus expansion in string theory, echoing some deep old ideas [tH74].

Both sides of the duality enjoy a PSU$(2,2|4)$ symmetry. On the string theory side the bosonic part of this supergroup is the universal cover SU$(2,2) \times$ SU(4) of the isometry group SO$(4,2) \times$ SO(6) of $\mathrm{AdS}_5 \times S^5$, and the fermionic part comes from the conserved supercharges which transform in the representation $(\mathbf{4},\mathbf{4}) + (\overline{\mathbf{4}},\overline{\mathbf{4}})$ of SU$(2,2) \times$ SU(4). The generators of PSU$(2,2|4)$ are the Killing vectors and Killing spinors of the geometry. On the gauge theory side the bosonic part of PSU$(2,2|4)$ comes from the conformal group SO$(4,2)$ in four dimensions together with the R-symmetry SU(4), and PSU$(2,2|4)$ is the complete superconformal group of symmetries of the theory.

The holographic correspondence.

The metric on AdS_5 in the Poincaré patch is given by

$$\mathrm{d}s^2 = R^2 \frac{\mathrm{d}x \cdot \mathrm{d}x + \mathrm{d}z^2}{z^2} \ , \tag{8.13}$$

however the AdS–CFT correspondence is better understood in the universal cover of AdS_5, sometimes denoted CAdS_5 [BBS06]. The space CAdS_5 has the topology of a solid cylinder $B_4 \times \mathbb{R}_t$ with boundary $S^3 \times \mathbb{R}_t$. It is sometimes cleaner to work in the Euclidean version EAdS_5 of AdS_5, which can be described as the hypersurface in $\mathbb{R}^{d+1,1}$ defined by

$$y_1^2 + \cdots + y_d^2 - t_1^2 + t_2^2 = -R^2 \ . \tag{8.14}$$

The metric on EAdS_5 space can be described as

$$\mathrm{d}s^2 = 4\frac{\sum \mathrm{d}u_i^2}{(1 - \sum u_i^2)^2} \ , \qquad \sum_{i=1}^{5} u_i^2 \leq 1 \ , \tag{8.15}$$

i.e. EAdS_5 is the hyperbolic 5-space.

The AdS_5 or EAdS_5 space is called the *bulk*, whereas the cylinder $S^4 \times \mathbb{R}_t$ at infinity is the *boundary*. A point of coordinates (x^μ, z) in the bulk (in the notation of Equation (8.13)) corresponds to a point of coordinates x^μ on the boundary. The radial coordinate $z = R^2/r$ in the bulk is identified with the energy scale on the gauge theory side: four-dimensional $\mathcal{N} = 4$ SU(N) super Yang–Mills theory at the energy scale E (in the Wilsonian sense) 'lives' on the cylinder defined by $z \sim E^{-1}$ inside the bulk. The gauge theory in the extreme UV without any degrees of freedom integrated out is considered to be located on the boundary at infinity $r \to \infty$.

A precise correspondence between the observables of the gauge theory and those of the gravity theory has been proposed in [GKP98, Wit98a]. Building on this one can establish an *AdS–CFT dictionary* which links physical quantities on both sides of the correspondence: for example, masses of excitations on the gravity side correspond to dimensions of operators on the gauge side.

8.2 Conical singularities and toric geometry

The ideas of [Mal99, Wit98a] suggest that Maldacena's original AdS–CFT correspondence can be extended to other backgrounds of type IIB string theory of the form $\mathrm{AdS}_5 \times X_5$ with X_5 a five-dimensional *Sasaki–Einstein manifold* bearing F_5-flux. A Sasaki–Einstein manifold is a Riemannian manifold (S, g) such that the metric cone over S (with metric $\mathrm{d}s^2 = dr^2 + r^2 d_S^2$) is Kähler (this is what is meant by Sasakian) and such that the Ricci curvature on S is a constant times the metric (this is what is meant by Einstein) [Spa11]. Type IIB superstrings on $\mathrm{AdS}_5 \times X_5$ should be dual to conformal field theories generalizing four-dimensional $\mathcal{N} = 4$ super Yang–Mills.

Type IIB superstrings on the background $\mathrm{AdS}_5 \times X_5$ with $X_5 = T^{1,1}$ has been studied in [KW98]. The manifold $T^{1,1}$ is a homogeneous space

$$T^{1,1} = (\mathrm{SU}(2) \times \mathrm{SU}(2))/\mathrm{U}(1) \ , \tag{8.16}$$

with U(1) a diagonal subgroup of $\mathrm{SU}(2) \times \mathrm{SU}(2)$. Equivalently, $T^{1,1}$ is a U(1) bundle over $S^2 \times S^2$ with Sasaki–Einstein metric

$$\mathrm{d}s^2_{T^{1,1}} = \frac{1}{9}(\mathrm{d}\psi + \cos\theta_1 \mathrm{d}\phi_1 + \cos\theta_2 \mathrm{d}\phi_2)^2 + \frac{1}{6}\sum_{i=1}^{2}\left[\mathrm{d}\theta_i^2 + \sin^2\theta_i \mathrm{d}\phi_i^2\right] \ , \tag{8.17}$$

where (θ_1, ϕ_1) and (θ_2, ϕ_2) parameterize each S^2 while $\psi \in [0, 4\pi[$ parametrizes the U(1) fiber.

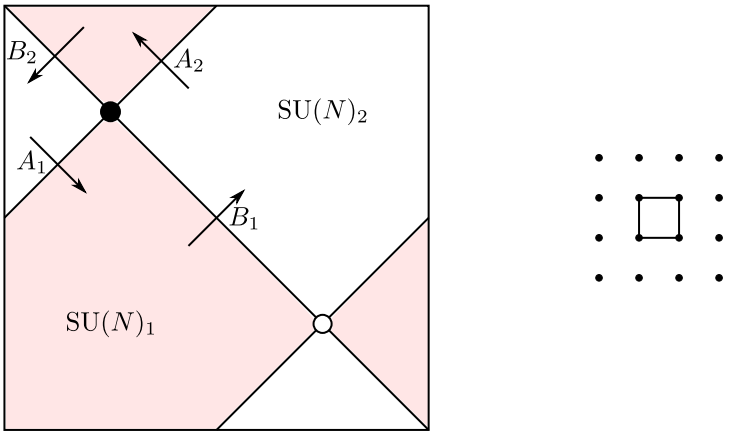


Figure 8.1: The brane tiling encoding the theory dual to type IIB on $\mathrm{AdS}_5 \times T^{1,1}$.

It is argued in [KW98] that the gauge theory dual to type IIB superstrings on $\mathrm{AdS}_5 \times T^{1,1}$ is a four-dimensional $\mathcal{N} = 1$ gauge theory with gauge group $\mathrm{SU}(N)_1 \times \mathrm{SU}(N)_2$, two fields A_1 and A_2 in $(\overline{\square}_1, \square_2)$, two fields B_1 and B_2 in $(\overline{\square}_2, \square_1)$, and a superpotential

$$W = \frac{\lambda}{2}\epsilon^{ij}\epsilon^{kl}\,\mathrm{Tr}\, A_i B_k A_j B_l \ . \tag{8.18}$$

This gauge theory is conveniently encoded in the brane tiling shown on the left of Figure 8.1. Even if it is anachronous to do so, one can use the forward algorithm to see easily that this brane tiling describes the worldvolume theory on D3 branes probing the conifold singularity, whose toric diagram is shown on the right of Figure 8.1.

The Klebanov–Witten generalization of the AdS–CFT correspondence can be obtained through the same reasoning as Maldacena's, except that one needs to consider D3-branes at the tip of the conifold singularity instead of in flat space.

This construction can be generalized to any arbitrary conical singularity Y_6 with metric

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2 g_{ij}\mathrm{d}x^i \mathrm{d}x^j \ , \tag{8.19}$$

where g_{ij} is the metric on a five-dimensional Sasaki–Einstein manifold X_5, so that Y_6 is Calabi–Yau and hence the worldvolume theory on D3-branes at the singularity preserves a certain amount of supersymmetries, which is crucial to keep some control on the AdS–CFT correspondence. It is however of interest

to lower the number of preserved supersymmetries as much as possible, since many interesting physical phenomena cannot exist in four-dimensional gauge theories with $\mathcal{N} > 1$.

Let us consider type IIB superstring theory on $\mathbb{R}^{1,3} \times Y_6$ with a stack of N D3-branes extending along $\mathbb{R}^{1,3}$ and placed at the singular point of Y_6. At low energies the gravitational backreaction of the D3-branes is encoded in the metric:

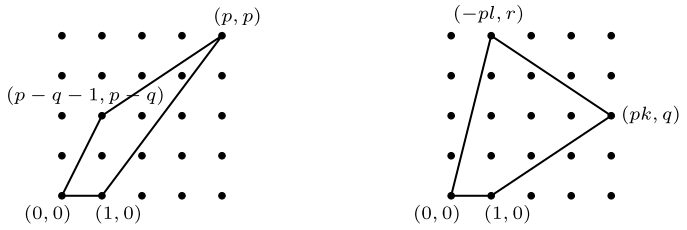
$$\mathrm{d}s^2 = f_3^{-1/2}(r)[-\mathrm{d}t^2 + \mathrm{d}\vec{x}^2] + f_3^{1/2}(r)(\mathrm{d}r^2 + r^2 g_{ij}\mathrm{d}x^i\mathrm{d}x^j) \ , \tag{8.20}$$

with

$$f_3(r) = 1 + \left(\frac{R}{r}\right)^4 \ , \qquad R^4 = 4\pi\alpha'^2 g_s N \ . \tag{8.21}$$

The geometry has a horizon at $r \to 0$, and in the near horizon limit the geometry is $\mathrm{AdS}_5 \times X_5$. These setups have also been studied in [MP99].

A rich playground to study such extensions of the AdS–CFT correspondence is the set of affine toric CY3 singularities. All of them are metric cones over five-dimensional Sasaki-Einstein manifolds, and the worldvolume theory on D3-branes at the singular point of the CY3 singularity is easily computed using brane tiling techniques. Even if in general it is very hard to derive the Sasaki-Einstein metric on X_5 (or equivalently, the Calabi–Yau metric on the singularity) explicitly, infinite countable families were found in [GMSW04b, GMSW04a, GMSW04c, CLPP05, CLPP09]. One family contains Sasaki–Einstein manifolds which are usually denoted $Y^{p,q}$ with $q < p$ positive integers, and the second contains Sasaki–Einstein manifolds denoted $L^{p,q,r}$ with $p \leq q$ and $r < p + q$ with p, q, r positive integers. All these manifolds have topology $S^2 \times S^3$. The corresponding Calabi–Yau cones are toric [MS06a, MS05], and the toric diagrams of $Y^{p,q}$ and $L^{p,q,r}$ are shown in Figure 8.2, following [FHM$^+$06]. The $Y^{p,q}$ gauge theories were derived in [MS06a, BFH$^+$05], and brane tilings encoding the worldvolume theory on D3-branes at $Y^{p,q}$ and $L^{p,q,r}$ are given in [FHM$^+$06].



metric cone over $Y^{p,q}$ metric cone over $L^{p,q,r}$

Figure 8.2: The toric diagrams of the metric cones over $Y^{p,q}$ and $L^{p,q,r}$ (with $rk + ql = 1$).

These infinite families of holographic dual pairs have been used to test and extend the AdS–CFT dictionary. An important point is the volume minimization of [MSY06, BZ05] which is dual to the a-maximization of [IW03]. The holographic dual to the central charge a of the CFT living on the boundary of $\mathrm{AdS}_5 \times X_5$ is the volume of X_5:

$$a = \frac{\pi^3}{4\mathrm{Vol}(X_5)} \ , \tag{8.22}$$

and hence the gravity dual of a-maximization is the minimization of $\mathrm{Vol}(X_5)$.

8.3 More general gauge–gravity correspondences

We have seen in the previous section that Maldacena's original AdS–CFT correspondence could be generalized to less supersymmetric setups by studying stacks of D3-branes at general affine toric CY3 singularities, in which case one obtains a duality between type IIB superstring theory on $\mathrm{AdS}_5 \times X_5$ with X_5 some Sasaki–Einstein 5-manifold and a four-dimensional $\mathcal{N} = 1$ conformal field theory. The AdS–CFT dictionary related the geometry of X_5 to field theoretic quantities of the CFT.

It is also of great interest to study generalizations of these AdS–CFT correspondences to non-conformal setups. Then, the (non-conformal) gauge theory is dual to type IIB string theory in some background which is *not* a direct product $\mathrm{AdS}_5 \times X_5$ but a more general warped geometry. Instead of referring

to such dualities as AdS–CFT correspondences, one rather speaks of *gauge–gravity dualities*. A great introduction to some techniques that can be used to study such correspondences is [Ber03].

Starting from a stack of N D3-branes at the singular point of a toric affine CY3 singularity, there are different ways to break the conformal invariance of the worldvolume theory on the D3-branes. One can for example add *fractional branes*, and since these are easily implemented in brane tilings we will concentrate on this approach. Besides, most of these seemingly different ways to break conformal invariance are mutually dual [Ber03].

8.3.1 Branes at the conifold and the duality cascade

We have reviewed the AdS–CFT correspondence between type IIB superstrings on $\mathrm{AdS}_5 \times T^{1,1}$ and a four-dimensional $\mathcal{N} = 1$ gauge theory with gauge group $\mathrm{SU}(N)_1 \times \mathrm{SU}(N)_2$, two fields A_1 and A_2 in $(\overline{\square}_1, \square_2)$, two fields B_1 and B_2 in $(\overline{\square}_2, \square_1)$, and a superpotential given in Equation (8.18). The correspondence is obtained by considering a stack of N D3-branes at the conifold in two different ways. The conifold is the metric cone over $T^{1,1} \simeq S^2 \times S^3$, and hence there is one vanishing 2-cycle and one vanishing 3-cycle at the singularity, on which branes can be wrapped. In particular one can wrap a D5-brane on the vanishing 2-cycle, equivalently described as a fractional D3-brane at the singularity. From the brane tiling of Figure 8.1 one sees that the ranks of the two gauge groups can be chosen arbitrarily, i.e. there is one independent fractional brane on the conifold, consistently with the analysis of Section 7.5.5.

Since fractional D3-branes are wrapped D5-branes, the 3-cycle dual to the 2-cycle in $T^{1,1}$ on which the D5s are wrapped carries M units of R–R 3-form flux. If one considers N regular branes and M fractional branes at the conifold in the limit $N \to \infty$ with M fixed, one expects the geometry on the gravity side of the correspondence to be the one corresponding to N regular branes at the conifold with sub-leading corrections in M/N [KN00, KT00]. It is shown in [KN00] that the R–R 3-form flux through the 3-cycle of $T^{1,1}$ induces a radial variation of the NS–NS 2-form potential:

$$B_2 = e^{\phi} f(r) \omega_2 \ , \tag{8.23}$$

where ω_2 is the volume form on the 2-cycle of $T^{1,1}$.

The gauge theory on the worldvolume of N regular and M fractional D3-branes at the singular point of the conifold is the four-dimensional $\mathcal{N} = 1$ gauge theory with gauge group $\mathrm{SU}(N+M)_1 \times \mathrm{SU}(N)_2$, two fields A_1 and A_2 in $(\overline{\square}_1, \square_2)$, two fields B_1 and B_2 in $(\overline{\square}_2, \square_1)$, and the superpotential of Equation (8.18). This quantum field theory does not flow to an interacting conformal fixed point anymore, and:

$$\frac{1}{g_1^2} - \frac{1}{g_2^2} \sim e^{-\phi} \left(\int_{S^2} B_2 - \frac{1}{2} \right) \ , \tag{8.24}$$

as emphasized in [KW98, MP99]. Hence:

> The radial dependence of the NS–NS two-form B_2 is (holographically) dual to the running of the difference of the (inverse) gauge couplings along the renormalization group flow.

Near the conformal fixed point of the renormalization group flow one expects in the limit $M/N \to 0$, one has:

$$\frac{1}{g_1^2} - \frac{1}{g_2^2} = \frac{3M}{\pi} \ln\Big(\frac{\mu}{\Lambda}\Big) + \text{const.} \ , \tag{8.25}$$

and moreover:

$$\frac{1}{g_1^2} + \frac{1}{g_2^2} = \frac{1}{4\pi g_s} \ . \tag{8.26}$$

On the gravity side, the dilaton of type IIB supergravity is constant at leading order in $1/N$. Taking the R–R 0-form to vanish and the R–R 3-form field strength as induces by the presence of the fractional D3-branes, one finds:

$$f(r) \sim M \log(r/r_0) \tag{8.27}$$

at leading order in M/N [KN00] for the function f appearing in Equation (8.23). This reproduces the logarithmic running of Equation (8.25).

Again another step was made in [KT00], where the supergravity solution corresponding to N regular and M fractional D3-branes (with $M \ll N$) at the conifold was computed to all order in M/N, taking into account the backreaction of the fluxes H_3 and F_3 on other fields. The main features of the Klebanov–Tseytlin solution are that:

- The dilaton field $e^{\phi} = g_s$ is exactly constant,
- The field strengths are $F_3 = M\omega_3$ on the one hand, $B_2 = 3g_sM\omega_2 \ln(r/r_0)$ and $H_3 = \mathrm{d}B_2 = 3g_sMr^{-1}\mathrm{d}r \wedge \omega_2$ [KS00a],
- The ten-dimensional metric is:

$$\mathrm{d}s_{10}^2 = h^{-1/2}(r)\mathrm{d}x \cdot \mathrm{d}x + h^{1/2}(r)(\mathrm{d}r^2 + r^2\mathrm{d}s_{T^{1,1}}^2) \ , \tag{8.28}$$

  with warp factor

$$h(r) = b_0 + 4\pi\frac{g_sN + a(g_sM)^2\ln(r/r_0) + a(g_sM)^2/4}{r^4} \tag{8.29}$$

  where a and b_0 are constants,
- The $F_5 = \mathrm{d}C_4 + B_2 \wedge F_3 = \mathcal{F}_5 + \star\mathcal{F}_5$ flux of the supergravity solution acquires a radial dependence:

$$\mathcal{F}_5 = \big(N + ag_sM^2\ln(r/r_0)\big)\,\mathrm{Vol}(T^{1,1}) \ , \tag{8.30}$$

  i.e. the flux decreases as r becomes smaller, until some $r = \tilde{r}$ where the F_5 flux vanishes. This was dubbed *RG cascade* in [KS00a] since it implies that the ranks of the gauge groups in the dual gauge theory decrease monotonically along the renormalization group flow.

As argued in [KS00a] one needs $b_0 = 0$ in order to have the good behaviour in the UV, i.e. at large r. This implies however that there is a naked singularity at some $r = r_s$ where $h(r_s) = 0$. As r decreases, the cascade must stop at $r = \tilde{r}$ since the rank of gauge groups cannot be strictly negative. The gauge dual to the RG cascade is called *duality cascade* [KS00a, Str05], and it is well described in brane tilings language, as displayed in Figure 8.3.

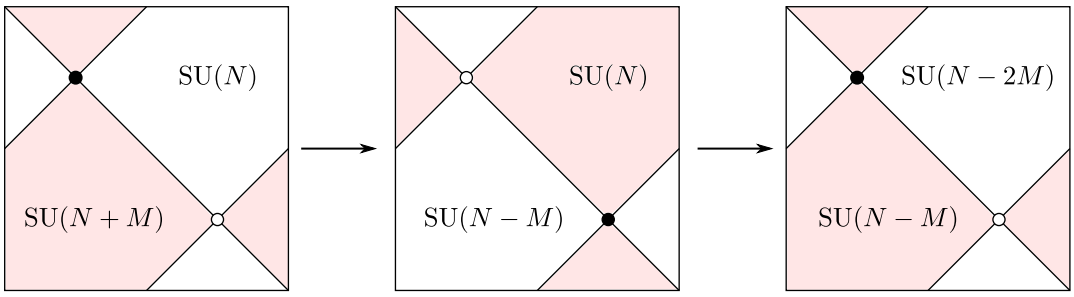


Figure 8.3: Two steps of the duality cascade.

Starting with the four-dimensional $\mathcal{N} = 1$ gauge theory described by the brane tiling on the left of this figure, the holomorphic beta functions for both gauge groups are

$$\beta_1 \propto N + 3M \ , \qquad \beta_2 \propto N - 2M \ , \tag{8.31}$$

so that the first gauge group of rank $N + M$ flows to strong coupling faster than the second, of rank N. As the dynamical scale of the first gauge group one can use Seiberg duality to extend the description of the RG flow in terms of another dual theory, which is presented in the middle of Figure 8.3. Then the story repeats: one dualizes the gauge group $\mathrm{SU}(N)$ which flows to strong coupling faster than the one of rank $\mathrm{SU}(N - M)$; this yields the brane tiling on the right of Figure 8.3, and so forth, and so on. If one started with M fractional and $N = kM$ regular D3-branes, the cascade brings the gauge theory to a $\mathrm{SU}(2M) \times \mathrm{SU}(M)$ theory after some number of steps, and this theory is in turn dual to four-dimensional $\mathcal{N} = 1$ $\mathrm{SU}(M)$ super Yang–Mills theory with no chiral flavor, which is known to confine in the IR as we have reviewed in Section 5.4.

The duality cascade is also well described in the type IIA picture of Seiberg duality, in the spirit of the brane setups of Section 6.4. The worldvolume theory of N regular and M fractional D3-branes at the conifold is described by N D4-branes on a circle with an NS5 brane and an NS5' brane in perpendicular directions, together with M open D4-branes streching between the NS5 and the NS5'.

	0	1	2	3	4	5	6	7	8	9
D4	–	–	–	–	0	0	–	0	0	0
NS5	–	–	–	–	–	–	×	0	0	0
NS5′	–	–	–	–	0	0	×	–	–	0

(8.32)

The backreaction of the D4-endpoints in the NS5 worldvolume bends the latter, because there are M open D4-branes which create an unbalance on each side of the NS5 and the NS5'. Since both NS5 branes extend in perpendicular directions the bending does not force them to intersect, however their positions along the circle direction coincide at some point. This is interpreted as the fact that the gauge coupling of the SU$(N+M)$ gauge group diverges, since $g_1^{-2} \propto |x^6_{\text{NS5}} - x^6_{\text{NS5}'}|$.

In order to remove this divergence one may move one of the NS5 brane around the circle once, in the direction prescribed by the bending. Because of the *Hanany–Witten effect* this leaves $N-M$ D4's wrapping the circle and M open D4's streching between the two NS5 branes. Doing this over and over again decreases the rank of each gauge group by N every two turns. This is depicted in Figure 8.4.

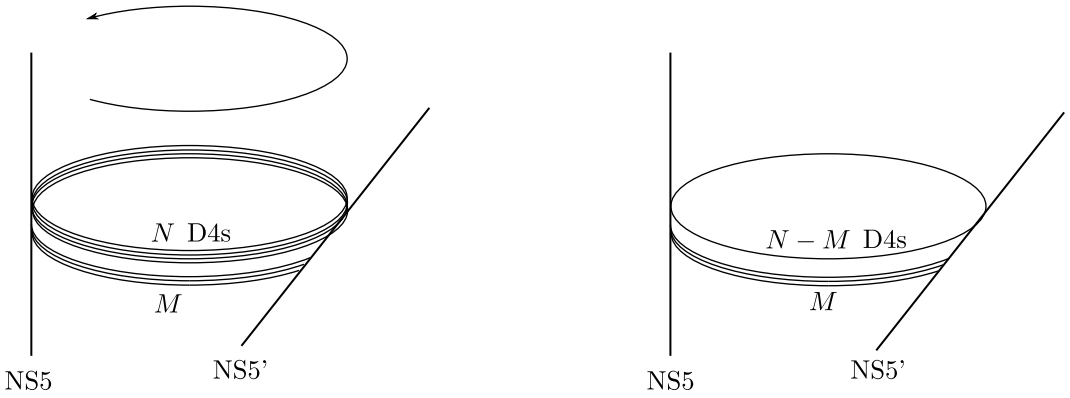


Figure 8.4: The duality cascade as seen from a IIA brane setup.

Something drastic might happen when the ranks of both gauge groups become of order M, and it does [KS00a]: the space-time geometry is modified by the strong coupling dynamics of the gauge theory. Studying the moduli space of a probe D3-brane one learns that the M fractional D3-branes deform the geometry: the conifold is to be replaced by the deformed conifold, defined by:

$$z_1^2 + z_2^2 + z_3^2 + z_4^2 = \epsilon^2 \ , \tag{8.33}$$

with $\epsilon = r_s^{3/2}$: the naked singularity at $r = r_s$ is resolved.

The presence of the fractional branes on top of the regular ones induces an anomaly for the R-symmetry of the gauge theory and only a $\mathbb{Z}_{2M}$ non-anomalous R-symmetry is preserved. The gaugino condensate in the super Yang–Mills theory determines the choice of vacuum among the M possible ones and further breaks the R-symmetry to $\mathbb{Z}_2$: the chiral $\mathbb{Z}_{2M}$ R-symmetry is spontaneously broken on the moduli space. This chiral symmetry breaking is dual to the deformation of the geometry on the gravity side of the correspondence: the U(1) isometry of the conifold (which is a gauge symmetry is supergravity) is spontaneously broken to $\mathbb{Z}_{2M}$ by the non-zero R–R 2-form C_2. The deformation of the geometry further breaks this discrete isometry group to $\mathbb{Z}_2$.

It is the fractional branes, i.e. the D5-branes wrapped on the vanishing 2-cycle of the conifold, which have triggered a deformation of the latter in which the dual 3-cycle in $T^{1,1}$ has acquired a non-zero size, of order $\sqrt{g_s M}$ [KS00a].

8.3.2 Fractional branes of different kinds

Studying the holographic duality induced by the presence of regular and fractional D3-branes at general affine toric CY3 singularities yields a classification of fractional branes according to the low-energy dynamics they trigger, as review in [FHK$^+$06]. Using the correspondence between fractional branes and integer-value assignments to the ZZPs of the geometry as explained in Section 7.5.5, these different kinds of fractional branes can be nicely characterized directly on the toric diagrams of the singularity.

- $\mathcal{N}=2$ **fractional branes** induce flat directions in the moduli space of the corresponding worldvolume theories, on which the dynamics reduces to a four-dimensional $\mathcal{N}=2$ theory and can be described with the associated Seiberg–Witten curve introduced in Section 5.6. In the dual supergravity picture they lead to enhançon-like backgrounds [JPP00, BDVF$^+$01, GKMW01, PRZ01, BBCC09]. The moduli along such a flat direction is a mesonic operator in the corresponding brane tiling which

does not appear in the superpotential. It corresponds to a collection of faces forming a stripe, hence giving rise to a closed loop in the dual quiver, which is the mesonic operator we are speaking of. Only non-isolated affine toric CY3 singularities can host $\mathcal{N} = 2$ fractional branes, in which case there are at least two parallel ZZPs in any brane tiling corresponding to such a non-isolated singularity. The $\mathcal{N} = 2$ fractional brane defined by this pair of ZZPs corresponds to assigning the value $+1$ to one of them and -1 to the other. An example of an $\mathcal{N} = 2$ fractional brane at the affine cone over PdP_4 is displayed in Figure 8.5. The ZZPs are shown in red and blue, and the mesonic operator which is the moduli on the $\mathcal{N} = 2$ flat direction corresponding to this fractional brane is the dashed black oriented path. Geometrically, $\mathcal{N} = 2$ fractional D3-branes arise as D5-branes

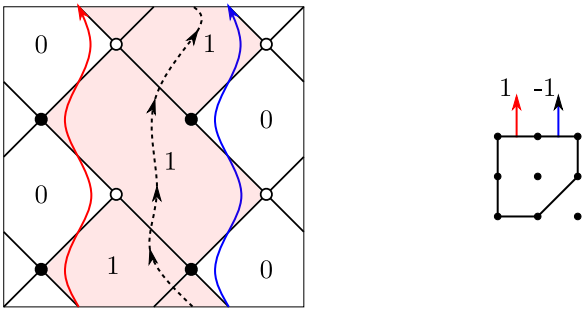


Figure 8.5: An $\mathcal{N} = 2$ fractional brane at PdP_4 and the corresponding value assignment to the ZZPs.

wrapped on a vanishing 2-cycle at the non-isolated singularity which exist at every point along the complex curve of singularities. In the case of non-isolated toric singularities, the latter is always locally of the form $\mathbb{C}^2/\mathbb{Z}_n$ for some $n \geq 2$, which translates into the fact that the corresponding toric diagram has an edge of its boundary with $n - 1$ internal points.

- **Deformation fractional branes** generalize the fractional branes on the conifold we have dealt with: at low energy they trigger a deformation of the geometry in which the 3-cycle in the Sasaki–Einstein manifold dual to the 2-cycle wrapped by these D5-branes wrap, acquires a non-zero minimal size. Versal deformations of affine toric CY3 singularities have been studied in [Alt94]: there is one possible deformation for each sub-web in equilibrium of the (p, q)-web corresponding to the singularity (strictly speaking, this is only the case for isolated singularities). A sub-web in equilibrium is a set of external legs in the (p, q)-web whose outgoing simple vectors sum to zero. The deformation of the geometry can be understood in the $\mathbb{R}^2$ base of the T^2-fibration of the geometry as in Section 7.4: given one such sub-web in equilibrium the deformation corresponds to moving this sub-web in a direction transverse to base plane of the T^2 fibration. Under this deformation a vanishing 3-sphere S^3 at the singularity grows [Mar05]. An example of deformation fractional brane at the singular point of the affine cone over dP_2 is displayed in Figure 8.6. The sub-quiver in a brane tiling corresponding to deformation fractional branes consists of decoupled nodes and closed loops appearing in the superpotential of the dimer model. Many examples of deformation fractional branes have been studied in [FHU05, FHSU06].

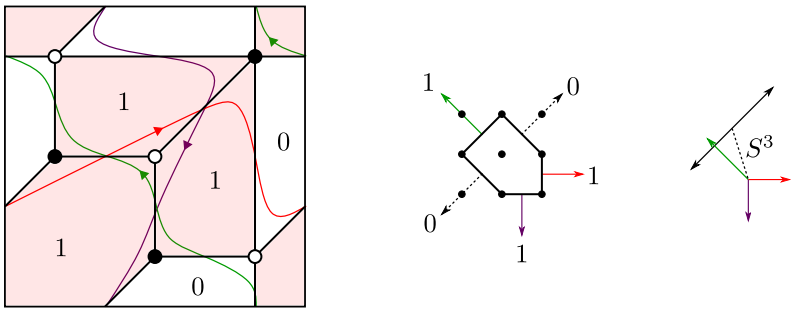


Figure 8.6: A deformation fractional brane at dP_2 and the corresponding value assignment to the ZZPs.

- **DSB fractional branes** (DSB stands for dynamical supersymmetry breaking, as in Section 5.5) are all the fractional branes which are not of the previous two kinds. In general the worldvolume

theory on such fractional branes contain gauge groups of different ranks and bifundamental matter, and the dynamics of at least one of these gauge groups develops an ADS-like superpotential at strong coupling. This dynamical superpotential is runaway and hence if there are classical flat directions in the moduli space, the theory does not admit stable vacua at finite distance in configuration space. The constraints on the ZZP value assignments needed for $\mathcal{N} = 2$ or deformation fractional branes define sublattices of strictly positive codimension in the lattice of ZZP value assignments satisfying the constraints of Equation (7.83), and hence a fractional brane is generically DSB [FHSU06].

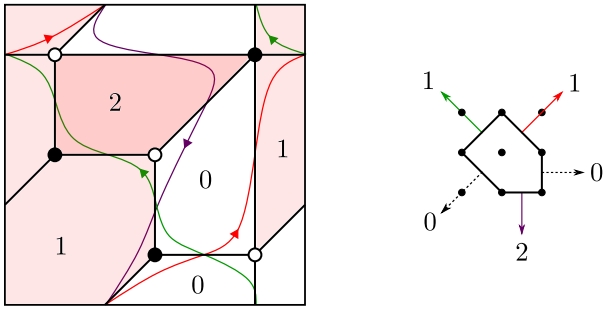


Figure 8.7: A DSB fractional brane at dP_2 and the corresponding value assignment to the ZZPs.

With view towards understand the low-energy dynamics of a gauge theory through the gauge gravity correspondence, it is of interest to embed a given gauge theory in a duality cascade, obtained for example by considering a large number N of regular D3-branes and $M \ll N$ fractional branes of some type at an affine toric CY3 singularity, i.e. a brane tiling with N regular and M fractional branes. The hope to analyze and understand stable[1] dynamical supersymmetry breaking from a dual supergravity perspective has proved very difficult, since when the fractional branes are of DSB type the theory develops runaway directions under the addition of regular branes [BHOP05, FHSU06, BBC05, IS06, BF06, AC07]. It is therefore of interest to construct models exhibiting dynamical supersymmetry breaking on a deformation fractional brane.

The gauge–gravity correspondence induced by regular D3-branes together with different kinds of fractional D3-branes at affine toric CY3 singularities has also been studied [ABB$^+$08], as well as duality cascades and infrared dynamics at orientifolded singularities with or without fractional branes [IY02, AB18]. In general, orientifolds break conformal invariance, just as fractional branes. They are interesting objects to consider with views towards gauge–gravity correspondences, for as discussed at the end of the previous chapter, they enrich a great deal the diversity of gauge theories that can be obtained on the worldvolume of branes at singularities.

[1]Understanding metastable supersymmetry breaking from branes at singularities has also catalyzed much effort, see e.g. [ABFK07a, ABFK07b, ABFK07c].

* * * * * * *

We have discussed the correspondence between type IIB superstrings on $\mathrm{AdS}_5 \times S^5$ and $\mathcal{N} = 4$ super Yang–Mills, as well as two generalizations of great physical significance. One the one hand, one can obtain a huge class of $\mathcal{N} = 1$ AdS–CFT correspondences by considering D3 branes at affine toric CY3 singularities. There are infinite families of the latter for which the explicit Calabi–Yau metric is known, which allows for direct tests of these AdS–CFT correspondences. On the other hand, considering fractional D3-branes together with regular ones breaks the conformal invariance of the world-volume theory. Such non-conformal quiver gauge theories have a rich dynamics, described dually as warped throats solutions of type IIB supergravity.

In the next part of this dissertation, we will address the following question: given some quiver gauge theories of interest, which in our case will be the SU(5) and $3-2$ DSB models, can they possibly be engineered as the worldvolume theory of a deformation fractional brane at a toric affine CY3 singularity? In fact, orientifold CY3 singularities hosting these theories are known since [FHK$^+$07]. However, it was shown more recently that all previously known such completions were actually always spoiling the stable vacuum of the SU(5) and $3-2$ DSB models, ultimately because the corresponding singularities were not isolated. Therefore, we will be interested in a refinement of the question of above: can the SU(5) and $3-2$ DSB models be engineered as the worldvolume theory of a deformation fractional brane at a toric affine CY3 singularity satisfying come constraints (for example, that it is isolated)?

A negative answer to this question would point towards the possibility that these theories, and maybe all DSB models, might be intrinsically incompatible with quantum gravity or at least some string theories. Such a result would have tremendous implications, however it is also probably tremendously hard to obtain in all generality. Nevertheless, a negative answer to our problematic would be a step towards such a swampland theorem.

Conversely, a positive answer would provide an example of an implementation of such DSB models on branes at singularities, and therefore prove that they can be UV-completed in string theory without spoiling their DSB nature. Moreover, the obtained setup would open a window on the possibility of describing dynamical supersymmetry breaking in terms of a dual gravitational theory.

We will be using dimer models extensively, for they are a very convenient tool to describe and work with worldvolume gauge theories of branes at toric affine CY3 singularities.

Part III

Dynamical supersymmetry breaking in string theory : the Octagon

Since the early days of the AdS/CFT correspondence [Mal99, Wit98a, GKP98] and its non-conformal extensions, the possibility of describing, holographically, supersymmetric gauge theories enjoying different IR behaviors has been thoroughly investigated. This has been a rich and lively arena in the field and remarkable results have been obtained in the last two decades. The correspondence between geometry and gauge theory is particularly well understood in the case of $4d$ $\mathcal{N}=1$ gauge theories on D3-branes probing toric Calabi-Yau (CY) 3-folds, for which the map is significantly streamlined by brane tilings (equivalently known as dimer models) [HK05, FHK$^+$06, FHM$^+$06].

Orientifolds [Sag87, PS89, Hor89, DLP89, BS90, BS91] of such singularities are extremely interesting for a variety of reasons. Among them, they expand the possible spectrum [GP96, DM96, FHK$^+$07] (gauge groups and matter fields representations), break conformal invariance [AB18], play an important role in models with non-perturbative effects due to D-brane instantons [ABF$^+$07, BFM07, BCKW09] and are a key ingredient in certain models of phenomenological interest, including ones leading to metastable dynamical supersymmetry (SUSY) breaking [ABFK07b].

While conformal phases, confinement, generation of a mass gap, Coulomb and Higgs-like branches and more generally any supersymmetry preserving dynamics were reproduced in a plethora of examples, not surprisingly (dynamical) supersymmetry breaking has proven to be much harder to achieve. Known examples describe supersymmetry breaking into metastable vacua (see [KPV02, FU06, ABFK07a, ABFK07b] and many other constructions thereafter) or runaway behavior, where the theory breaks supersymmetry dynamically but it does not enjoy a vacuum at finite distance in the space of field VEVs [BHOP05, FHSU06, BBC05, IS06], very much like massless SQCD with a small number of flavors. No models were known, until recently, that enjoy dynamical supersymmetry breaking (DSB) into stable vacua. This has proven to be a harder problem to engineer with D-branes at singularities. This is partly due to the scarcity of known gauge theories that display such a non-supersymmetric vacuum.

Finding models of this kind could be of great relevance both in the context of the gauge/gravity duality and, even more interestingly, in string compactifications. In this latter setup they could be used for model building in GKP-like constructions [GKP02]. Eventually, they might also have an impact on the swampland program [Vaf05, BCV17, Pal19] and recent related conjectures such as [BGVU19].

Recently, a series of papers renewed the interest in models of D-branes at Calabi-Yau (CY) singularities leading to dynamical supersymmetry breaking. This originated from [FHK$^+$07] where an existence proof for a possibly stable DSB model obtained by considering fractional branes at orientifold singularities was given. These results were generalized in [ABMP19], where it was shown that a large class of orientifolds admit fractional D-brane configurations realizing some of the most popular and simple DSB models, namely the incalculable $SU(5)$ [ADS84] and/or $3-2$ [ADS85] models.

In this same work [ABMP19], however, by generalizing previous results of [BGVU19], it was shown that in the decoupling limit [Mal99], in which the DSB fractional D-brane bound state becomes part of a UV complete large N D-brane model and gravity is decoupled, all models display an instability. This instability turned out to have a common, model-independent geometric origin in terms of $\mathcal{N}=2$ fractional branes probing the singularity.[2] More drastically, a no-go theorem was proven in [ABMP19] which implies that whenever $\mathcal{N}=2$ classical flat directions exist at a singularity which admits such DSB models, the quantum behavior of the latter is such that the flat directions are tilted and supersymmetry preserving vacua exist. All of this was mounting evidence for what could be interpreted as the impossibility of engineering stable DSB with D-branes at singularities.

An obvious way to circumvent this no-go theorem and avoid the unwanted slide towards supersymmetric vacua is to look at singularities free of $\mathcal{N}=2$ fractional branes to start with, and see whether stable DSB models of the type above can be engineered there. Or, alternatively, a stronger no-go theorem should exist which excludes such a possibility altogether. This is what we will be concerned with in this part of the manuscript[3]. More precisely, our main goal will be to answer the following question:

Is it possible to get a DSB model, more specifically the $SU(5)$ or the 3-2 models, from D-branes at a Calabi-Yau singularity which is free of any (known) instability?

In more dramatic words, could stable DSB be in the swampland? Quite surprisingly, the answer will be affirmative! In this part of the dissertation we argue that stable DSB is still in the landscape.

[2] $\mathcal{N}=2$ fractional branes arise whenever a Calabi-Yau singularity can be partially resolved to display, locally, a non-isolated $\mathbb{C}^2/\mathbb{Z}_n$ singularity and a Coulomb-like branch associated to it.

[3] In the same vein as in [FHK$^+$07, BGVU19, ABMP19], we will not consider configurations where non-compact flavor branes are added. We note that metastable DSB can be engineered in this way [FU06], and further investigating if stable DSB is possible in these constructions is an interesting problem that we do not address here.

In Chapter 9 we will first carry out a comprehensive investigation that shows that in the minimal realizations of the $SU(5)$ and 3-2 models at orientifolds of singularities, the instability associated to $\mathcal{N}=2$ fractional branes is unavoidable. Remarkably, this result ties the ability to engineer these models to basic geometric features of the underlying singularity: the presence of non-isolated $\mathbb{C}^2/\mathbb{Z}_n$ singularities. This is yet another example of the connection between geometry and features or dynamics of the corresponding quantum field theories, such as e.g. confinement and complex deformations [KS00a] or runaway DSB and the absence of complex deformations [BHOP05, FHSU06]. These general results will then guide our search of models without instabilities. We will show that a simple variant of the $SU(5)$ model, that we dub twin $SU(5)$, can be realized as a local structure which does not directly imply the presence of $\mathcal{N}=2$ fractional branes. This lets one hope for a possible implementation of the twin SU(5) model on a configuration of branes at the orientifold of an isolated toric CY3 singularity. Our analysis, which is done exploiting dimers techniques [FHK$^+$06, FHM$^+$06], relies also on results obtained in [ABF$^+$21a], where a thorough investigation of consistent, anomaly free, D-brane models at orientifold singularities has been performed.

In Chapter 10 we will introduce combinatorial tools known as *triple diagrams* that will be of great help to build dimer models satisfying symmetry constraints and containing the substructure hosting the twin SU(5) model. Using these methods it is possible to construct phases of worldvolume theories on D3-branes at isolated singularities which admit an orientifold projection with a diagonal fixed line in the brane tiling, and with the required substructure. We will present two examples of such dimer models, hence showing in a direct way that the twin SU(5) model can be implemented at isolated singularities. However, these orientifold theories are always plagued with gauge anomalies. This called for a general analysis of gauge anomalies in orientifolds of brane tilings, in order to understand when it is possible to expect to obtain consistent quantum gauge theories in this way.

The ranks of the gauge groups on theories realized at D-branes at a singularity correspond to numbers of (wrapped) D-branes in the configuration. Roughly speaking, the cancellation of local anomalies in the gauge theories correspond to the cancellation of tadpoles in the string construction [LR99, BM00]. When orientifolds are included, in some cases anomaly cancellation can only be achieved upon the addition of non-compact flavor D7-branes (see e.g. [FHK$^+$07, BIMRP14, FRU15, AB18]), which give rise to (anti)fundamental matter, but there are instances where it is possible to cancel the anomalies even in the absence of extra flavors [FHK$^+$07] (see e.g. [PRU00b, BBMR20] for some further examples).

Chapter 11 is devoted to the study of anomalies in gauge theories coming from D-branes at orientifolds of toric singularities, in the *absence* of flavor branes. We will introduce a new geometric algorithm for constructing anomaly-free theories and identify geometric criteria for the existence of such solutions. Remarkably, our results allow us to determine whether an orientifold singularity can admit anomaly-free D-brane gauge theories just by analyzing its geometric structure and avoiding any case-by-case analysis, which has been so far the only known approach for this class of theories. This geometric criterion is therefore a new addition to the list of connections between the geometry of singularities and general properties of the resulting gauge theories, some of which were mentioned above.

The results of Chapter 11 are then used to further refine our search for isolated singularities hosting the twin SU(5) model and without gauge anomalies. Building on the techniques of Chapter 10 we will show in Chapter 12 that the twin $SU(5)$ can be realized by D-branes at an orientifold of a toric CY, the Octagon. This provides a counter-example to what could have been conjectured, namely that DSB models were possible only in singularities admitting $\mathcal{N}=2$ fractional branes, and hence, following the no-go theorem presented in [ABMP19], unstable towards supersymmetric vacua.

While we do not prove nor exclude the existence of other, more involved models sharing the same properties, the example we provide shows that stable DSB can be engineered by brane configurations at CY singularities. Given the implications that this might have in different contexts, including improvements in our understanding of the string landscape and the swampland, it is worth investigating these D-brane constructions further. On a more technical side, the results presented here as well as in [ABF$^+$21a] show, once again, the power of dimer techniques in understanding the properties of D-branes and more generally string theory at CY singularities.

The results presented in this part were obtained during my PhD in collaboration with Riccardo Argurio, Matteo Bertolini, Sebastían Franco, Eduardo García Valdecasas, Shani Meynet and Antoine Pasternak. They were published in [ABF$^+$21b, ABF$^+$21c, ABF$^+$21a] – this is the content of Chapters 9, 11 and 12, while the tools and techniques of Chapter 10 were presented in [Tat21].

Chapter 9

Dynamical Supersymmetry Breaking on branes at toric affine CY3

The study of D-branes at singularities (and in particular affine toric CY3 singularities) provides very fruitful extensions of the original AdS–CFT correspondence, as we have reviewed in Chapter 8. It is particularly interesting to consider fractional branes on top of regular branes, since they lead to rich dynamics at low energies. The DSB models that we have presented in Section 5.5 are fascinating theories with a low-energy dynamics of theoretical and phenomenological interest, and it is natural to wonder whether they can be obtained as the worldvolume theory on a bound state of branes at a singularity.

This question finds echo in the *swampland program*: are DSB models compatible at all with quantum gravity? A full answer to this would have tremendous implications. Even partial answers such as a proof that it is impossible to implement some known DSB models, or on the contrary showing that an explicit model hosting a DSB model exist, would be of great theoretical interest, both in the context of branes at singularities where holographic dualities are at their cleanest, and in the context of true string compactifications, in order to build phenomenological models.

In Section 9.1.1 we present some of the first examples of brane tilings hosting either the SU(5) or the $3-2$ models, that were provided in [FHK$^+$07]. Many more such examples were given in [ABMP19]. However it was shown in [BGVU19] that the UV completion of such DSB models provided by one of these brane tilings actually spoiled the dynamical breaking of supersymmetry, ultimately because the corresponding singularity was not isolated. This was generalized into a no-go theorem in [ABMP19], proving that it is never possible to obtain stable DSB vacua on the worldvolume of bound states of branes at non-isolated singularities. We will review this in Section 9.1.2.

Then, we study possible brane tiling implementations of the SU(5) model in Section 9.2, and of the $3-2$ model in Section 9.3. We will see that all cases imply the presence of $\mathcal{N}=2$ fractional branes at the singularity, i.e. that the singularity is not isolated, but one. This last case is a hexagonal cluster of faces sitting on an orientifold fixed line, which host a twin version of the SU(5) model. Moreover, this hexagonal cluster should belong to a deformation fractional brane and not a DSB fractional brane, in order to ensure the low-energy stability of the theory on the worldvolume of this fractional brane. We can thus wrap-up the results of this chapter as follows:

> The only possible way to embed either the SU(5) or the $3-2$ models in brane tilings, without the instability that the presence of $\mathcal{N}=2$ fractional branes creates, is to have a (consistent) dimer model containing the hexagonal cluster of interest on a deformation fractional brane, such that there exist an orientifold projection of the brane tiling with a fixed line going through the hexagonal cluster. The orientifold of this dimer model with the good fixed line hosts a twin version of the SU(5) DSB model.

This chapter consists of all of [ABF$^+$21b] but the last section, the presentation of which we defer to Chapter 12.

9.1 Introduction

9.1.1 Review of previous implementations of DSB models in brane tilings

The matter fields in orientifolds of dimer models can be either in bifundamental representations of the gauge groups, or in (anti)symmetric representations, i.e. they sit in tensor representations of rank at most 2. If DSB models are to be UV completed in orientifolds of dimer models, it is hence natural to restrict to the study of the SU(5) and $3-2$ models, whose matter fields are in tensor representations of rank at most 2 of the gauge groups.

The first examples of such orientifolds of brane tilings were presented in [FHK+07]. Let us for example consider a brane tiling corresponding to the affine toric CY3 singularity $\mathbb{C}^3/\mathbb{Z}'_6$, where the generator $\zeta = \exp(2\pi i/6)$ of $\mathbb{Z}'_6$ acts on the coordinates of $\mathbb{C}^3$ as

$$\zeta : (z_1, z_2, z_3) \longrightarrow (\zeta z_1, \zeta^2 z_2, \zeta^{-3} z_3) \ . \tag{9.1}$$

We follow the presentation of [ABMP19].

The brane tiling of our interest in depicted in Figure 9.1. It admits an orientifold projection with four fixed points, also displayed on the figure as the four red crosses.

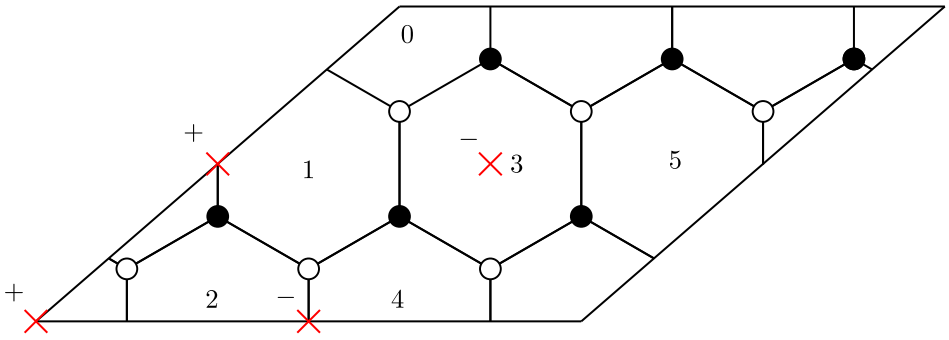


Figure 9.1: The brane tiling of $\mathbb{C}^3/\mathbb{Z}'_6$ hosting the SU(5) DSB model after orientifold.

The orientifold theory described by the brane tiling of Figure 9.1 has gauge group $\mathrm{SO}(N_0)\times\mathrm{SU}(N_1)\times\mathrm{SU}(N_2)\times\mathrm{USp}(N_3)$ and matter content:

$$\begin{aligned} &X_1 = (\square_0, \overline{\square}_1) \ , \quad X_2 = (\square_0, \overline{\square}_2) \ , \quad X_3 = (\square_0, \square_3) \ , \\ &Y_1 = (\square_0, \square_3) \ , \quad Y_2 = (\square_2, \square_3) \ , \quad Z_1 = (\square_1, \square_2) \ , \quad Z_2 = (\overline{\square}_1, \overline{\square}_2) \ , \\ &W = (\square_1, \overline{\square}_2) \ , \quad A = \begin{smallmatrix}\square\\\square\end{smallmatrix}_2 \ , \quad S = \overline{\square\square}_1 \ . \end{aligned} \tag{9.2}$$

The anomaly cancellation conditions for the two unitary gauge groups both read

$$N_0 + N_1 - N_2 - N_3 + 4 = 0 \ . \tag{9.3}$$

The quadruplet $(N_0, N_1, N_2, N_3) = (1, 0, 5, 0)$ is a solution of this anomaly cancellation condition, and the corresponding theory corresponds to the quiver depicted on the left of Figure 9.2.

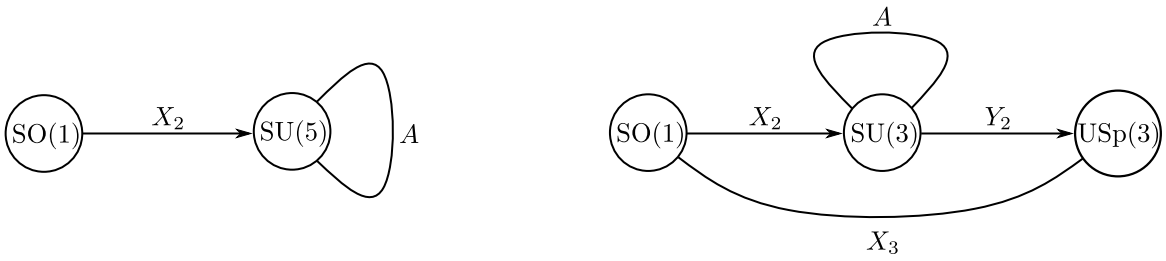


Figure 9.2: Quivers for the SU(5) and $3-2$ DSB models.

Another interesting solution of the anomaly cancellation condition is $(N_0, N_1, N_2, N_3) = (1, 0, 3, 2)$, in which case one finds the quiver on the right of Figure 9.2. Since $\mathrm{USp}(2) = \mathrm{SU}(2)$ and since the

antisymmetric representation of SU(3) is also the antifundamental one, one finds a gauge theory with gauge group $\mathrm{SO}(1) \times \mathrm{SU}(3) \times \mathrm{SU}(2)$, matter content

$$X_2 = (\square_0, \overline{\square}_2) \ , \quad X_3 = (\square_0, \square_3) \ , \quad Y_2 = (\square_2, \square_3) \ , \quad A = \begin{smallmatrix}\square\\\square\end{smallmatrix}_2 = \overline{\square}_2 \ , \tag{9.4}$$

and superpotential

$$W = X_2 Y_2 X_3 \ , \tag{9.5}$$

i.e. it is exactly the $3-2$ DSB model.

Hence the $\mathbb{C}^3/\mathbb{Z}_6'$ singularity with the orientifold projection described above is particularly interesting, since there is an anomaly-free rank assignment which yields the SU(5) DSB model, and another, the $3-2$ DSB model.

The PdP_4 singularity is studied as well in [FHK+07, ABMP19]; it also hosts both the SU(5) and the $3-2$ DSB models. Many more singularities were studied in [ABMP19], and it was shown that an important fraction of them implement either the SU(5) model, or the $3-2$ model, or both. They can be engineered by bound states of fractional D3-branes which can arise at the end of complicated RG-flows (often described by a duality cascade [KS00a]) or on the $\mathcal{N}=4$ Coulomb branch of regular D3-branes, depending on the singularity structure. However, each of the ones hosting at least one of the two DSB models is non-isolated. Equivalently, they admit $\mathcal{N}=2$ fractional branes. The latter are related to a partial resolution of the singularity displaying a non-isolated $\mathbb{C}^2/\mathbb{Z}_n$ singularity and an $\mathcal{N}=2$ Coulomb branch associated to it. We will review in the next section how this destabilizes the DSB vacuum, hence spoiling the hope to construct a stable supersymmetry breaking vacuum on branes at these singularities.

Before that, let us remark that this piece of evidence, together with a exhaustive scan of all dimer models with at most eight faces (as listed in [FHSX17]), led to the conjecture that the SU(5) and $3-2$ DSB models can only exist on branes at non-isolated singularities *without flavor D7-branes*, and furthermore that it is impossible to engineer DSB models in this way [ABMP19]. The goal of this chapter and the next three is to show that this intuition turns out to be wrong, since an explicit dimer model corresponding to an isolated singularity and hosting a twin version of the SU(5) model on a deformation fractional brane was constructed and presented in [ABF+21b, ABF+21c].

Now that we know that such a model can be obtained, it might as well be possible to implement such DSB models in brane tilings more easily if one allows for flavor D7-branes, described for example in [Yam08]. We might as well have embraced this option from the start; considering brane tilings without D7-branes was a working hypothesis more than a physical-motivated assumption. Nevertheless, this parsimonious approach – now that we know that a model without flavor D7-branes exists, lets one hope for less complicated, cleaner geometric duals in the context of gauge–gravity correspondences, if possible at all to tackle such questions.

9.1.2 $\mathcal{N}=2$ Fractional Branes Decay

The swampland program that we alluded to at the end of Section 7.2 has been widely used to constraint phenomenological string models. In particular, applying a refined version of the Weak Gravity Conjecture of [AHMNV07] to the near horizon limits of systems of branes led to the conjectural statement that stable non-supersymmetric AdS vacua are incompatible with Quantum Gravity [OV17]. This was shown to be in agreement with the study of decays of non-supersymmetric AdS vacua in string theory via bubbles of nothing [OS17].

A local statement of this conjecture was proposed in [BGVU19], motivated by the study of the near-horizon limit of branes at singularities. A system of N regular and M fractional D3-branes at a toric affine CY3 singularity Y_6 which is the metric cone over a five-dimensional compact Sasaki–Einstein manifold X_5, yields a near-horizon geometry of a warped form generalizing Equation (8.28):

$$\mathrm{d}s^2 = h^{-1/2}(r)\mathrm{d}x \cdot \mathrm{d}x + h^{1/2}(r)[\mathrm{d}r^2 + r^2 \mathrm{d}s^2_{X_5}] \ , \tag{9.6}$$

at least asymptotically since more drastic behaviors can happen near the tip of the throat, as in [KS00a]. This background is locally AdS in the sense that it is arbitrarily close in r to an $\mathrm{AdS}_5 \times X_5$ background, but with the AdS radius varying in r. The *local anti de Sitter – Weak Gravity Conjecture* (AdS–WGC) of [BGVU19] is that in consistent theories of quantum gravity, stable non-supersymmetric solutions with asymptotics given by local AdS backgrounds defined above are forbidden. In particular, the existence of a *warped throat* with asymptotic metric as in Equation (9.6) and such that near the tip of the throat the

geometry breaks supersymmetry in a stable way[1], translating the stable dynamical breaking of supersymmetry in the dual gauge theory, is forbidden by the local AdS–WGC.

Evidence for the latter was provided by considering known configurations of regular and fractional D3-branes at toric affine CY3 singularities:

- If one considers N regular and M deformation fractional branes at an affine toric CY3 singularity, in the supergravity limit $M \ll N \to \infty$ the classical singularity at the tip of the warped throat is resolved by the growing of the 3-cycle dual to the 2-cycle on which the fractional branes are wrapped as in [KS00a, FHU05]. The background preserves supersymmetry, and hence it does not contradict the AdS–WGC.

- When the M fractional branes are of DSB type, the deformation of near the tip of the throat dual to the strong coupling dynamics of the worldvolume theory cannot preserve supersymmetry, and hence the local AdS–WGC predicts that these warped throats are unstable. Consistently with this conjecture, plenty of evidence has been built showing that the would-be DSB vacuum in these cases is actually unstable [FHSU06, BHOP05, BBC05, IS06, AC07].

A last class of examples considered in [BGVU19] is of particular interest to us, and we now quickly review the argument. Let us consider the dimer model presented in Figure 9.1 and corresponding to the singularity $\mathbb{C}^3/\mathbb{Z}_6'$. We have seen that the rank assignment $(N_0, N_1, N_2, N_3) = (1, 0, 5, 0)$ satisfies the anomaly cancellation condition of Equation (9.3), which can be equivalently stated at the anomaly of compact RR tadpoles at this singularity and with the corresponding choice of fractional branes and orientifold plane. The theory described by this rank assignment is the SU(5) DSB model, which is believed to break supersymmetry dynamically in a vacuum at which the scalar potential is of order $|\Lambda|^4$, where Λ is the strong coupling scale of the SU(5) gauge group.

In order to embed this model in a locally AdS warped throat one adds $N \gg 1$ regular D3-branes at the singularity, so that the gauge group of the theory becomes $\mathrm{SO}(N+1)\times\mathrm{SU}(N)\times\mathrm{USp}(N+5)\times\mathrm{SU}(N)$. The DSB sector of this theory, which contains both the orientifold plane together with $\mathcal{O}(1)$ fractional branes, is sub-leading in $1/N$ and hence the supergravity dual to the large N limit of this theory is expected to be of the form $\mathrm{AdS}_5 \times X_5$ at leading order, where X_5 is an orientifold of the $S^5/\mathbb{Z}_6'$. When sub-leading $1/N$ effects are taken into account, the strong coupling dynamics of the gauge theory triggers a dynamical breaking of supersymmetry, hence yielding a non-supersymmetric locally AdS warped throat. Motivated by the local AdS–WGC, the authors of [BGVU19] have found a new mechanism implying that the DSB vacuum of the low-energy SU(5) theory is actually unstable and hence cannot be used as a way to do de Sitter uplifts in flux compactifications, in place of anti D3-branes as proposed in [RU16a], in the spirit of the construction of [KKLT03].

The moduli space of bound states of regular and fractional D3-branes at singularities always contain a so-called $\mathcal{N} = 4$ *Coulomb branch*, which corresponds to the displacement of regular D3-branes away from the singular point. The bubble-of-nothing instabilities of [OS17] would correspond to instabilities of the supersymmetry-breaking vacuum of the SU(5) model along this $\mathcal{N} = 4$ Coulomb branch, however the running of the holomorphic couplings in the UV $\mathrm{SU}(N+5)$ theory and in the IR SU(5) theories read:

$$\frac{8\pi^2}{g^2_{\mathrm{SU}(N+5)}(\mu)} = \left[3(N+5) - \frac{1}{2}(6N+4)\right] \ln\left(\frac{\mu}{\Lambda_{\mathrm{UV}}}\right) = 13\ln\left(\frac{\mu}{\Lambda_{\mathrm{UV}}}\right) , \tag{9.7}$$

$$\frac{8\pi^2}{g^2_{\mathrm{SU}(5)}(\mu)} = \left[15 - \left(\frac{1}{2}+\frac{3}{2}\right)\right] \ln\left(\frac{\mu}{\Lambda}\right) = 13\ln\left(\frac{\mu}{\Lambda}\right) , \tag{9.8}$$

since the fields do not acquire anomalous dimensions at the superconformal fixed point, because $\mathbb{C}^3/Z_6'$ is an orbifold of $\mathbb{C}^3$, and hence the gauge theory is an orbifold projection of $\mathcal{N} = 4$ SYM.

Assuming that the Higgsing of the gauge theory inducing the breaking $\mathrm{SU}(N+5) \to \mathrm{SU}(5)$ occurs at a scale v, and matching the above two expressions at $\mu = v$, yields $\Lambda = \Lambda_{\mathrm{UV}}$. This implies that the DSB brane sector does not exert any force on the regular branes at the singularity, and hence that the low-energy SU(5) DSB theory is stable along the $\mathcal{N} = 4$ Coulomb branch.

Now, the story is different along the so-called $\mathcal{N} = 2$ *Coulomb branch* of the worldvolume theory on the bound state of branes at the tip of $\mathbb{C}^3/\mathbb{Z}_6'$. This singularity is non-isolated; in particular there is a complex line of $\mathbb{C}^2/\mathbb{Z}_2$ singularities going through the singular point. Fractional $\mathcal{N} = 2$ branes can

[1] The local AdS–WGC does not forbid meta-stable supersymmetry breaking warped throats.

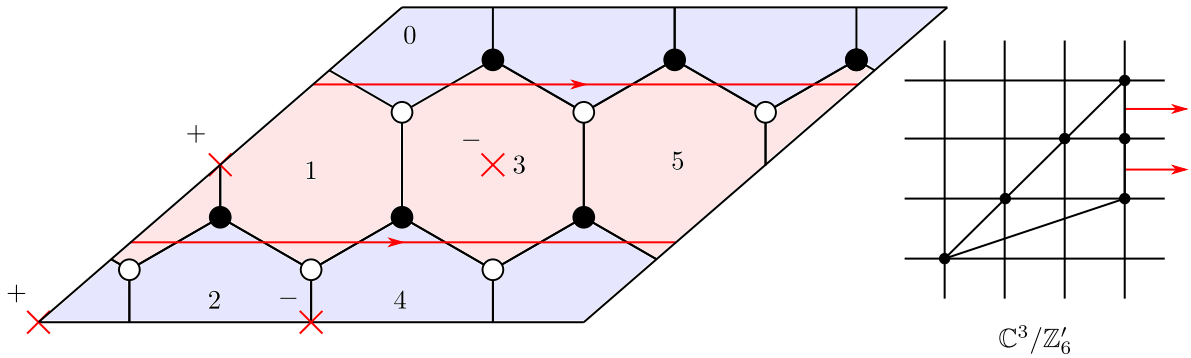


Figure 9.3: Two $\mathcal{N}=2$ strips in the dimer model of the $\mathbb{C}^3/\mathbb{Z}_6'$ singularity.

move along this complex line of singularities, and the displacement along this line is parameterized by the corresponding strip operator in the dimer model, characteristic of $\mathcal{N}=2$ fractional branes.

As we have seen in Chapter 8, $\mathcal{N}=2$ fractional branes correspond to sides of the toric diagram which have lattice points in their interior. A side with n interior points corresponds to a line of $\mathbb{C}^2/\mathbb{Z}_{n+1}$ singularities going through the tip of the Calabi–Yau cone.

In the case of the brane tiling corresponding to the $\mathbb{C}^3/\mathbb{Z}_6'$ singularity as before and displayed once more on the left of Figure 9.3, the two colored strips are $\mathcal{N}=2$ fractional branes. They correspond to the two zig-zap paths drawn in red and reproduced on the toric diagram, on the right of the same figure. These $\mathcal{N}=2$ fractional branes are of particular interest since they are preserved under the orientifold involution, and hence they project to two $\mathcal{N}=2$ fractional branes in the orientifold theory. Each fractional brane corresponds to a mesonic operator represented as a closed path in the quiver dual to the dimer model. Let v and v' be the vacuum expectation values of these mesonic operators. The $\mathcal{N}=4$ Coulomb branch of the theory is parameterized by $v=v'$, while the $\mathcal{N}=2$ Coulomb branch is parameterized by arbitrary values of v and v'.

Assuming that $v \gg v'$, this choice of vacuum expectation values lead to the following Higgsing pattern:

$$\begin{aligned}\mathrm{SO}(N+1)\times\mathrm{SU}(N)\times\mathrm{SU}(N+5)\times\mathrm{USp}(N) &\xrightarrow{v} \mathrm{SO}(1)\times\mathrm{SU}(N)\times\mathrm{SU}(5)\times\mathrm{USp}(N)\\ &\xrightarrow{v'} \mathrm{SO}(1)\times\mathrm{SU}(5)\ .\end{aligned} \tag{9.9}$$

Let $g_{\mathrm{SU}(N+5)}$, $g_{\mathrm{SU}(5)_N}$ and $g_{\mathrm{SU}(5)}$ be the respective holomorphic gauge couplings of the $\mathrm{SU}(N+5)$ gauge group at scales $\mu \gg v$, of the $\mathrm{SU}(5)$ gauge group at scales $v \gg \mu \gg v'$ and of the $\mathrm{SU}(5)$ gauge group at scales $v' \gg \mu$. Let also Λ_{UV}, Λ_N and Λ be the respective strong coupling scales of these gauge groups. The running of these couplings with the scale are:

$$\frac{8\pi^2}{g^2_{\mathrm{SU}(N+5)}(\mu)} = \left[3(N+5)-\frac{1}{2}(6N+4)\right]\ln\left(\frac{\mu}{\Lambda_{\mathrm{UV}}}\right) = 13\ln\left(\frac{\mu}{\Lambda_{\mathrm{UV}}}\right)\ , \tag{9.10}$$

$$\frac{8\pi^2}{g^2_{\mathrm{SU}(5)_N}(\mu)} = \left[15-\frac{1}{2}(4N+4)\right]\ln\left(\frac{\mu}{\Lambda_N}\right) = (13-2N)\ln\left(\frac{\mu}{\Lambda}\right)\ , \tag{9.11}$$

$$\frac{8\pi^2}{g^2_{\mathrm{SU}(5)}(\mu)} = \left[15-\left(\frac{1}{2}+\frac{3}{2}\right)\right]\ln\left(\frac{\mu}{\Lambda}\right) = 13\ln\left(\frac{\mu}{\Lambda}\right)\ , \tag{9.12}$$

Scale matching at $\mu=v$ yields the relation $\Lambda_N^{13-2N} = v^{-2N}\Lambda_{\mathrm{UV}}^{13}$ and at $\mu=v'$, the relation $\Lambda_N^{13-2N} = (v')^{-2N}\Lambda^{13}$. Hence:

$$\Lambda^{13} = \left(\frac{v'}{v}\right)^{2N}\Lambda_{\mathrm{UV}}^{13}\ . \tag{9.13}$$

Contrarily to what happens on the $\mathcal{N}=4$ Coulomb branch, one sees that on the $\mathcal{N}=2$ Coulomb branch the DSB sector exerts a force which tends to minimize Λ by bringing v' to zero while keeping v at a finite value. This effect destabilizes the supersymmetry breaking vacuum. Hence, the UV completion of the SU(5) DSB model via the orientifold of the brane tiling displayed on the left of Figure 9.3 in a warped throat is unstable, which goes along the line with the local AdS–WGC.

The same reasoning holds as one considers the way to obtain the $3-2$ DSB model in the same orientifold of the $\mathbb{C}^3/\mathbb{Z}_6'$ singularity that we have described in the previous section, and it was shown in [ABMP19] that it is in fact completely general. Let us consider a DSB model which emerges in the decoupling limit of a vacuum configuration of a (possibly intricate) system of regular and fractional D-branes, with the vacuum energy depending on the VEVs of the scalar fields. The $\mathcal{N}=4$ Coulomb branch is parameterized by regular branes. If an $\mathcal{N}=2$ fractional brane direction exists, there is in addition an $\mathcal{N}=2$ Coulomb branch. By scale matching, one can show that the energy of the supersymmetry breaking vacuum is related to the strong coupling scale Λ of the SU(5) or SU(3) $\times$ SU(2) gauge groups as follows[2]

$$E_{vac} = \left(\frac{v'}{v}\right)^{\alpha} \Lambda \quad , \quad \alpha \in \mathbb{R} \, , \tag{9.14}$$

where the exponent α is given by a ratio of beta functions and v and v' are the VEVs on the Coulomb branches associated to the $\mathcal{N}=2$ fractional brane and its complement, respectively. Fractional branes are defined modulo regular branes, so that a fractional brane and its complement combine into a regular brane. The case $v=v'$ then corresponds to the $\mathcal{N}=4$ Coulomb branch.

From Equation (9.14), it follows that on the $\mathcal{N}=4$ Coulomb branch the vacuum energy equals Λ and the supersymmetry breaking vacuum is hence preserved. On the $\mathcal{N}=2$ Coulomb branch, instead, where $v \neq v'$, the vacuum energy relaxes to 0, with a moduli space parametrized by v at $v'=0$ or vice-versa, depending on the sign of α. Geometrically, this corresponds to a supersymmetric configuration described by the $\mathcal{N}=2$ fractional branes associated to v located at a finite distance along the non-isolated $\mathbb{C}^2/\mathbb{Z}_n$ singularity describing its Coulomb branch, and their complement sitting at the origin.

The only possibility for evading this decay mechanism of the supersymmetry breaking vacuum is that $\alpha=0$. Using some basic properties of Calabi-Yau's and the fact that fractional branes are described by a non-conformal field theory at low energy, it is shown in that $\alpha \neq 0$ [ABMP19]. The upshot is that all DSB D-brane models constructed in [BGVU19, ABMP19] are actually unstable since, as anticipated, all of them admit $\mathcal{N}=2$ fractional branes. At most they can be metastable.

Let us schematically discuss what occurs to the gauge theory in this process. We denote $G_{\cancel{SUSY}}$ the SUSY breaking model, namely its gauge group (and possibly flavor group), matter fields and interactions. When N regular D3-branes are added, the SUSY breaking sector extends to

$$G_{\cancel{SUSY}} \quad \xrightarrow[\text{branes}]{+\ N \text{ regular}} \quad G_{\cancel{SUSY}+N} \times G'_N \, , \tag{9.15}$$

where $G_{\cancel{SUSY}+N}$ indicates that the ranks of the gauge and flavor groups are increased by N. G'_N denotes the theory associated to the complement. The subindex indicates that all gauge groups in this sector have rank N. In addition, there is matter connecting the $G_{\cancel{SUSY}+N}$ and G'_N sectors. Along the $\mathcal{N}=2$ Coulomb branch, the theory is higgsed down to

$$G_{\cancel{SUSY}+N} \times G'_N \quad \xrightarrow{v \neq v'} \quad G_{\cancel{SUSY}} \times G'_N \, , \tag{9.16}$$

We are left precisely with the original SUSY breaking theory of interest, but now coupled to G'_N. This extension of the theory spoils supersymmetry breaking.

The only way to avoid this decay mechanism is to look for singularities that admit supersymmetry breaking D-brane configurations *and* are free of local $\mathbb{C}^2/\mathbb{Z}_n$ singularities. Whether such geometries exist or not was an open question at the time of [ABMP19]. Answering this question has been one of the main motivations for the work presented in this chapter and the next three. The remainder of this chapter consists of the study of substructures of brane tilings that can possibly lead to the SU(5) or $3-2$ models, and whether they require the presence of $\mathcal{N}=2$ fractional branes, or not.

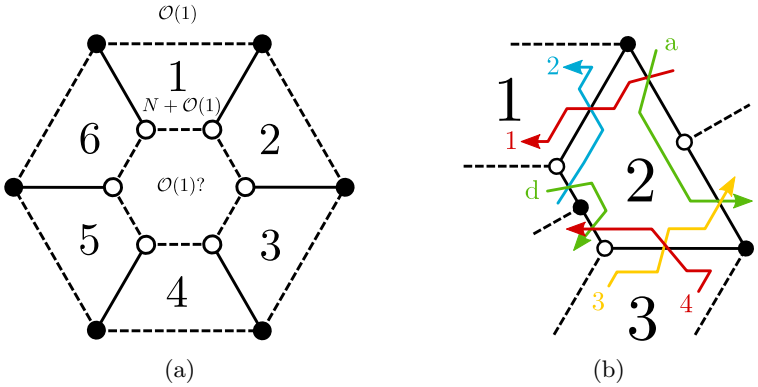
9.1.3 Holes in brane tilings and zig-zag paths

To conclude this introduction we present a technical argument forbidding the presence of holes of reduced rank inside a specific sub-dimer. We will use this result below to rule out different twin models. We rely on ZZP techniques for anomaly cancellation developed in [ABF$^+$21a, But06] and reviewed in Section 7.5.5. One associates a value v_i to every ZZP in the dimer and then assigns an arbitrary rank to a given face

[2] In the 3-2 model, Λ refers to the scale of the gauge group factor, either SU(3) or SU(2), that dominates the dynamics.

in the dimer. The remaining ranks are set by requiring that the rank differences between two adjacent faces m, n obey $N_m - N_n = v_i - v_j$ where i, j are the ZZP separating them.

Consider a ring-shaped sub-dimer of rank $N + \mathcal{O}(1)$. We assume that as we go along it, from one of its faces to another, we only cross edges with identical orientation, see Figure 9.4a. We now show that the region inside the ring, the "hole", is inconsistent if of reduced rank.



(a) (b)

Figure 9.4: (a) Generic ring of rank $N + \mathcal{O}(1)$ surrounded by faces of rank $\mathcal{O}(1)$ with a hole of rank $\mathcal{O}(1)$. (b) Face 2 edges with zig-zag paths.

Consider a face of the ring, as face 2 in Figure 9.4b. The intersections between the ZZP 1, 2, 3 and 4 yield

$$N_1 - N_2 = v_1 - v_2 \sim 0\,, \quad N_2 - N_3 = v_4 - v_3 \sim 0\,, \tag{9.17}$$

where $\sim$ means "equal up to $\mathcal{O}(1)$". Since the hole is supposed to be of rank $\mathcal{O}(1)$, the intersections with Zig-Zags that separate it from the ring give

$$N \sim v_2 - v_d, \quad -N \sim v_d - v_4, \quad \Rightarrow \quad v_2 \sim v_4\,. \tag{9.18}$$

Changing the number of edges between face 2 and the hole can only be done by adding/removing pairs of edges and will not change the fact that

$$v_1 \sim v_2 \sim v_3 \sim v_4 \qquad \text{and} \qquad v_d \sim v_1 - N\,, \tag{9.19}$$

where v_d is understood as any ZZP that comes with the pair of edges added between the hole and face 2. One can repeat the reasoning for every face of the ring and find that its internal edges will be always produced by ZZP $\sim v_1$. This is in contradiction with the presence of ZZP $v_d \sim v_1 - N$ since there are only ZZP $\sim v_1$ entering the hole. It implies that v_d is circular or not present. The first option is forbidden in dimer models and the second spoils the presence of the hole itself. Hence the presence of an anomaly-free hole inside such a ring is inconsistent.

As a comment, let us notice that to reach this conclusion we did not assume anything about the exterior of the ring. If one does not look at the hole but asks that the exterior has a reduced rank, it implies that ZZP v_a on its border, see Figure 9.4b, will satisfy

$$v_a \sim v_1 + N \sim v_3 + N\,, \tag{9.20}$$

and thus we recover the result of Equation (9.18) using Equation (9.17). Again, it can be shown that this result does not depend on the number of edges in contact with the exterior of the ring. The cluster (hexagonal or otherwise) is now viable only with ranks $N + \mathcal{O}(1)$, because it is made only of ZZP $\sim v_1$.

9.2 SU(5) Models

Let us first consider the SU(5) model. This theory has an SU(5) gauge group and one GUT-like chiral family $\square \oplus \overline{\square}$. The presence of the antisymmetric representation implies that if one wants to engineer such a model by D-branes at a CY singularity, an orientifold projection with non-empty fixed locus is necessary. Moreover, one has to consider two gauge groups in order to get the antifundamental representation $\overline{\square}$, which can be generated by either an SU(1) or an SO(1) flavor group [ABMP19].

Using the dimer formalism, there are three classes of orientifolds which possibly lead to chiral fields in antisymmetric representations of the gauge groups, depending on whether they have *fixed points*, *a single fixed line* or *two fixed lines* [FHK+07][3]. We will analyze them in turn.

9.2.1 Fixed Point Orientifolds

Let us remind that fixed point orientifolds are associated to dimers which enjoy a point reflection.

As we now review, we not only need a fixed point on one edge of the SU(5) face, but a second fixed point is needed to avoid anomalies in the face providing the (anti)fundamental matter field.

The first possibility is to directly avoid the anomaly in the flavor group by having it SO or USp. USp is ruled out since it would give always an even number of antifundamentals, hence more than one. We are then left with SO(1).

- SO flavor group

  Figure 9.5 shows the generic structure of a local configuration of a dimer leading to the SU(5) model, including the signs for the two relevant fixed points. The dotted lines and nodes represent a completely general configuration for the rest of the dimer, only constrained by its compatibility with the point reflections. The blue dotted line indicates that it is possible to choose the unit cell such that the two fixed points live on one of the four segments that form its boundary. This comment will be relevant later.

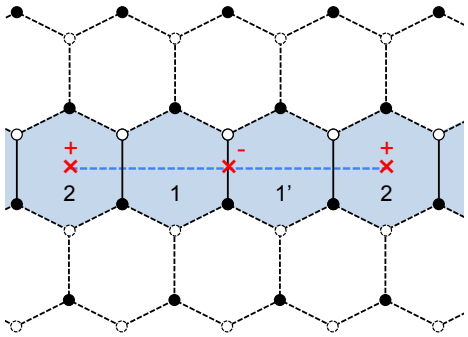


Figure 9.5: Fixed point orientifold realizing the SU(5) model with SO(1) flavor group. The dotted part of the graph indicates the rest of the dimer, which is completely general and not necessarily hexagonal as shown.

  Assigning arbitrary ranks to the gauge groups, N_i for face i in the dimer, the anomaly cancellation conditions (ACC) have a solution in which $N_1 = N_{1'} = 5$, $N_2 = 1$ and the rest of the faces are empty.[4] This choice leads exactly to the SU(5) model. Face 1 becomes the $SU(5)_1$ gauge group. Since face 2 has a fixed point with a positive sign on top of it, becomes the $SO(1)_2$ flavor group.

A second possibility is that the flavor group is of SU type, with its anomaly (when regular branes are added) being canceled by the presence of symmetric matter on a different edge of the face.

- SU flavor group with symmetric

  Figure 9.6 shows the local configuration of a dimer leading to another realization of the SU(5) model in a fixed point orientifold. Once again, the ACC have a solution in which $N_1 = N_{1'} = 5$, $N_2 = N_{2'} = 1$ and the rest of the faces are empty. The resulting theory is the SU(5) model, plus a decoupled singlet corresponding to the symmetric associated to the edge between face 2 and its image.

  Note that the SU(1) group has no anomaly, but the symmetric is necessary to cancel the anomaly when all the ranks are increased by N (corresponding to the addition of N regular D3-branes which

[3] $\mathbb{Z}_2$-involutions of the dimer without fixed loci will be studied in Chapter 13.

[4] Of course whether the ACC of the empty nodes are also satisfied depends on the details of the boundary of the cluster of faces under consideration. This observation also applies to the examples that follow.

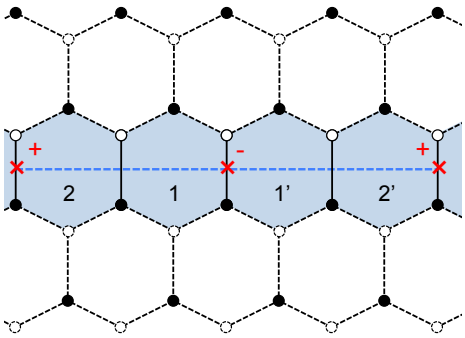


Figure 9.6: Fixed point orientifold realizing the SU(5) model with SU(1) flavor group.

populate the dimer democratically). By construction, the additional (white) faces with rank N will not contribute to the anomaly. In order to cancel the $N+5$ antifundamentals coming from face 1, we need to have a symmetric of SU($N+1$) at face 2. It reduces to a decoupled singlet when $N=0$.

A third possibility is that the flavor group is of SU type, and its anomaly (when regular branes are added) is canceled by 5 fundamentals attached to an SO(5) group. This configuration is shown in Figure 9.7. The low-energy theory of this configuration is an SU(5) model together with a decoupled SO(5) SQCD with one flavor. The latter theory develops an ADS superpotential [ADS85], so that we have a runaway behavior (on top of the DSB of the SU(5) model), and hence no true vacuum. We thus discard this possibility since it is already unstable at this low-energy field theory level.

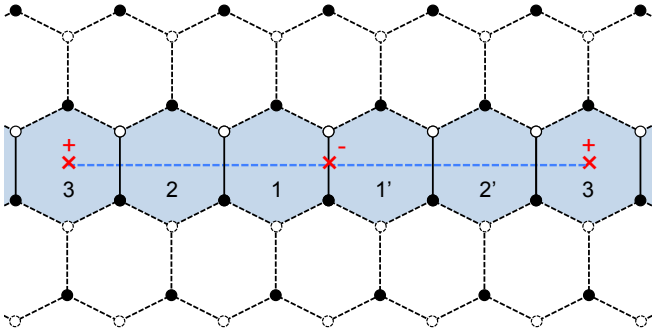


Figure 9.7: Fixed point orientifold realizing the $SU(5)_1$ model with $SU(1)_2$ flavor group and an additional $SO(5)_3$ factor. $SO(5)_3$ develops an ADS superpotential and leads to a runaway behavior.

A fourth possibility is that the flavor group is again of SU type, but now its anomaly is canceled by the presence of a replica of the SU(5) group with its own antisymmetric. We will call this possibility *twin* SU(5) *model*.

- SU flavor group with twin SU(5)

  Figure 9.8 shows the local configuration of a dimer leading to yet another realization of the SU(5) model in a fixed point orientifold. The ACC have a solution in which $N_1 = N_{1'} = 5$, $N_2 = N_{2'} = 1$, $N_3 = N_{3'} = 5$ and the rest of the faces are empty. The resulting theory corresponds to two SU(5) models sharing one and the same SU(1) flavor group which provides their (anti)fundamentals. Since SU(1) is actually empty, and in any case no chiral gauge invariants can be written for each SU(5) model, the twins are effectively decoupled and thus their low-energy dynamics is completely independent.

In principle, we could go on with further possibilities. Indeed, the anomaly of the second SU(5) gauge group at face 3 can be canceled with a fundamental, instead of an antisymmetric. The simplest possibility is that the fundamental is attached to an SO(1) face, however it could also be an SU(1) with a symmetric, or further an SU(1) with 5 antifundamentals given by an SO(5), or another SU(5). The possibilities already discussed above repeat themselves. What is important to notice is that the gauge

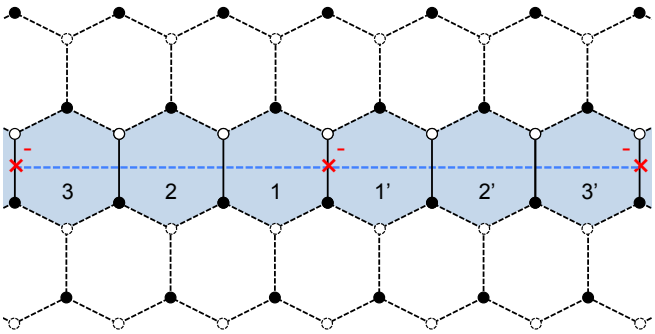


Figure 9.8: Fixed point orientifold realizing the twin SU(5) model.

theory on face 3 would always be an SU(5) with one flavor, hence developing an ADS superpotential and leading to runaway behavior.

We thus conclude that the only possibilities to engineer an SU(5) model, which is stable at low-energies, in a dimer with fixed points are the three bullets above: SO flavor group, SU flavor group with a symmetric and SU flavor group with twin SU(5).

An important remark is that in all the examples above the following holds: there can be a long chain of gauge groups to eventually cancel the anomaly of the initial SU(5) gauge group, but it always ends with an orientifold fixed point.[5] As a consequence, we do not have to look far in order to identify an $\mathcal{N}=2$ fractional brane in these dimers. Remarkably, in all cases the SU(5) model is fully supported on a set of faces that corresponds to an $\mathcal{N}=2$ fractional brane in the parent (i.e., non-orientifolded) theory. From Figure 9.5, Figure 9.6 and Figure 9.8 we see that in all cases the SU(5) model indeed lives on a stripe that gives rise to a gauge invariant not contained in the superpotential. The expectation value of such operator parametrizes the corresponding Coulomb branch.

We conclude that an SU(5) model cannot be obtained for this class of orientifolds if the parent theory does not contain line singularities, i.e. $\mathcal{N}=2$ fractional branes.[6] The previous discussion implies that the no-go theorem in [ABMP19] cannot be avoided for this class of orientifolds.

Let us discuss how the instability explained in Section 9.1.2 is realized in these models in more detail. We start with the model with SO flavor, Figure 9.5. After adding N regular D3-branes, the relevant gauge group becomes

$$\mathrm{SU}(N+5)_1 \times \mathrm{SO}(N+1)_2 \ . \tag{9.21}$$

Let us denote

$$A = \square_1 \quad , \quad \overline{Q} = (\overline{\square}_1, \square_2) \tag{9.22}$$

where A corresponds to the edge in the dimer between face 1 and its orientifold image and $\overline{Q}$ corresponds to the edge between faces 1 and 2. The Coulomb branch is parameterized by the expectation value of the gauge invariant going around the stripe. In principle we can build an $\mathrm{SU}(N+5)_1$ gauge invariant as

$$\phi^{\mathrm{SO}}_{ab} = \overline{Q}^i_a \overline{Q}^j_b A_{ij} \ , \tag{9.23}$$

where i,j are fundamental indices of $\mathrm{SU}(N+5)_1$ and a,b are fundamental indices of $\mathrm{SO}(N+1)_2$. Note that it is in the antisymmetric representation of $\mathrm{SO}(N+1)_2$, hence it does not exist for $N=0$, and it has vanishing trace for $N \geq 1$.

As discussed in [ABMP19], we actually need to go twice around the stripe in order to have a non-vanishing gauge invariant given by

$$\langle \delta^{ac}\delta^{bd}\phi^{\mathrm{SO}}_{ab}\phi^{\mathrm{SO}}_{cd} \rangle \ , \tag{9.24}$$

parametrizing the Coulomb branch. That the gauge invariant still vanishes automatically for $N=0$, is consistent with the fact that the SU(5) model does not have a moduli space and that the additional regular branes are necessary for the instability.

[5] We are ignoring more ramified possibilities. For instance, for an SU(1) flavor at face 2, we could imagine providing the 5 fundamentals from more than one SO gauge group. That would lead to the need of more than one extra fixed point. The other cases can be treated similarly. Thus a more precise statement is that we always need *at least* another fixed point to cancel the anomaly of the SU(5) at face 1.

[6] This result is consistent with an observation made in [RU16b], namely that singularities with deformation branes are incompatible with point projections.

We now consider the case with SU flavor and a symmetric, Figure 9.6. After adding N regular D3-branes, the gauge group becomes

$$\mathrm{SU}(N+5)_1 \times \mathrm{SU}(N+1)_2 \ . \tag{9.25}$$

We denote

$$A = \square_1 \quad , \quad \overline{Q} = (\overline{\square}_1, \square_2) \quad , \quad \overline{S} = \overline{\square\square}_2 \tag{9.26}$$

where now $\overline{S}$ corresponds to the edge between face 2 and its image under the second fixed point. The $\mathrm{SU}(N+5)_1$ gauge invariant is

$$\phi^{\mathrm{SU}}_{ab} = \overline{Q}^i_a \overline{Q}^j_b A_{ij} \ , \tag{9.27}$$

where now a, b are fundamental indices of $\mathrm{SU}(N+1)_2$. It is in the antisymmetric representation of $\mathrm{SU}(N+1)_2$, hence again it does not exist for $N = 0$, and for $N \geq 1$ it cannot be contracted with $\overline{S}^{ab}$ which is symmetric. A non-vanishing gauge invariant is given by

$$\langle \overline{S}^{ac} \overline{S}^{bd} \phi^{\mathrm{SU}}_{ab} \phi^{\mathrm{SU}}_{cd} \rangle \ , \tag{9.28}$$

which now parametrizes the Coulomb branch. The same remarks as in the previous case apply.

Finally, let us discuss the last case of the twin SU(5), where the gauge group becomes

$$\mathrm{SU}(N+5)_1 \times \mathrm{SU}(N+1)_2 \times \mathrm{SU}(N+5)_3 \ . \tag{9.29}$$

We denote

$$A = \square_1 \quad , \quad \overline{Q} = (\overline{\square}_1, \square_2) \quad , \quad \overline{P} = (\overline{\square}_2, \square_3) \quad , \quad \overline{A} = \overline{\square}_3 \tag{9.30}$$

where now $\overline{P}$ corresponds to the edge between faces 2 and 3, and $\overline{A}$ to the edge between face 3 and its image under the second fixed point. The $\mathrm{SU}(N+5)_1$ and $\mathrm{SU}(N+5)_3$ gauge invariants are

$$\phi_{ab} = \overline{Q}^i_a \overline{Q}^j_b A_{ij} \quad , \quad \overline{\phi}^{ab} = \overline{P}^a_\alpha \overline{P}^b_\beta \overline{A}^{\alpha\beta} \ , \tag{9.31}$$

where α, β are fundamental indices of $\mathrm{SU}(N+5)_3$. They are in the antisymmetric and conjugate antisymmetric representation of $\mathrm{SU}(N+1)_2$, respectively. They do not exist for $N = 0$, but for $N \geq 1$ the simplest gauge invariant is given by

$$\langle \phi_{ab} \overline{\phi}^{ab} \rangle \ , \tag{9.32}$$

which parametrizes the Coulomb branch in this case. The same remarks as in the previous cases apply. Further, note that this last case allows for a simpler gauge invariant parametrization of the Coulomb branch because it is the only one where the two fixed points (giving rise to A and $\overline{A}$) have the same sign, see Figure 9.8. In the two previous cases the fixed points have opposite signs, and we have to take the loop twice.

Double $SU(5)$ Models

In some cases, the structure of the dimer is such that it could be possible to use all four fixed points to generate a pair of SU(5) models. Figure 9.9 shows the general structure for a dimer giving rise to two SU(5) models with SO(1) flavor nodes. Other possibilities, for instance two models with SU(1) flavor nodes, an SU(1)/SO(1) combination or two twin SU(5) models, are also feasible. The same logic of previous examples applies to each of the two stripes of blue faces, so we conclude that each of these models contain $\mathcal{N} = 2$ fractional branes and hence are not stable.

The different cases considered so far illustrate the general strategy that we will apply to most of the other models we will be considering. While the DSB models under consideration are relatively simple, we are considering here their embedding into arbitrarily complicated toric singularities. Therefore, establishing the existence of $\mathcal{N} = 2$ fractional branes (which implies the instability of the DSB model) might naively seem an intractable problem since, generically, the majority of the dimer model will be unknown. However, as it occurred in the previous examples, the necessary interplay between the region of the dimer that makes up the DSB model and the orientifold fixed points (or fixed lines, as we will see shortly), implies that we fully know the dimer model along a "short direction" of the unit cell. This is sufficient to identify an $\mathcal{N} = 2$ fractional brane. In even simpler terms, in these cases the DSB models are actually supported on faces of the dimer that define an $\mathcal{N} = 2$ fractional brane. We will see that there is only one specific way to circumvent this argument.

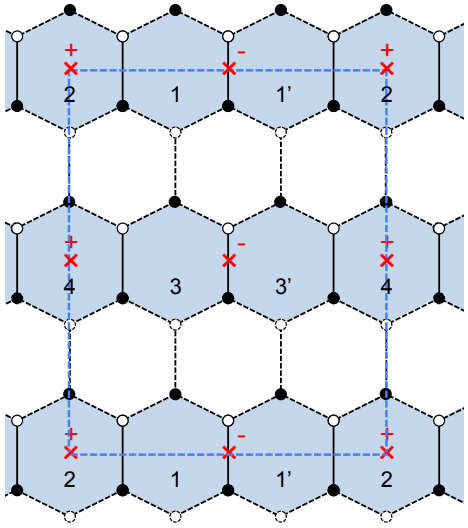


Figure 9.9: General structure of a fixed point orientifold realizing a double SU(5) with SO(1) flavor group model.

9.2.2 Fixed Line Orientifolds

DSB Models between Two Fixed Lines

The cases with two fixed lines are basically identical to the orientifolds considered in the previous section, with the exchange of fixed points for fixed lines. We therefore present them succinctly.

- SO flavor group

  Figure 9.10 shows the local configuration realizing the SU(5) model with SO(1) flavor group, including the signs of the fixed lines. This is achieved by setting $N_1 = N_{1'} = 5$, $N_2 = 1$ and vanishing ranks for all other faces. Since the two lines have opposite signs, this configuration is only possible in orientifolds with two independent fixed lines.

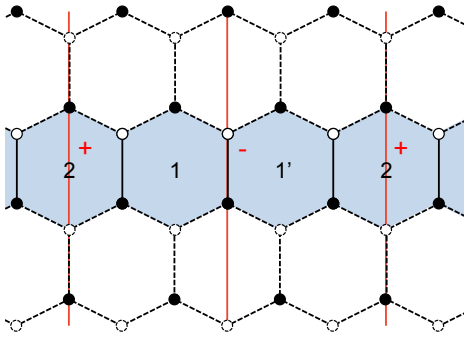


Figure 9.10: Two fixed lines orientifold realizing the SU(5) model with SO(1) flavor group.

- SU flavor group with symmetric

  Figure 9.11 shows the local configuration realizing the SU(5) model with SU(1) flavor group and a symmetric. This corresponds to $N_1 = N_{1'} = 5$, $N_2 = N_{2'} = 1$ and vanishing ranks for all other faces. Since the two lines have opposite signs, this configuration is only possible in orientifolds with two independent fixed lines.

- *SU* flavor group with twin *SU*(5)

  Figure 9.12 shows the local configuration realizing the SU(5) model with SU(1) flavor group and a twin SU(5) model. This corresponds to $N_1 = N_{1'} = 5$, $N_2 = N_{2'} = 1$, $N_3 = N_{3'} = 5$ and vanishing

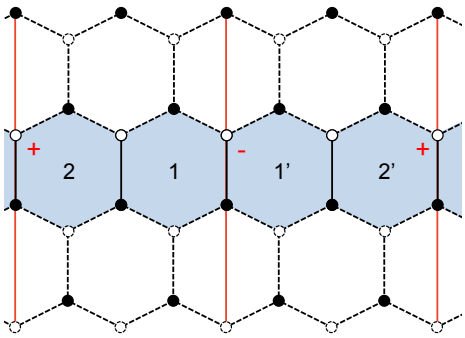


Figure 9.11: Two fixed lines orientifold realizing the SU(5) model with SU(1) flavor group.

ranks for all other faces. In this case the two lines have the same sign, hence it is possible to find this configuration both in orientifolds with two independent fixed lines or with a single diagonal fixed line. Note that in the latter case, we have to consider the situation in which the strip goes from one line to a second one, in a contiguous unit cell.

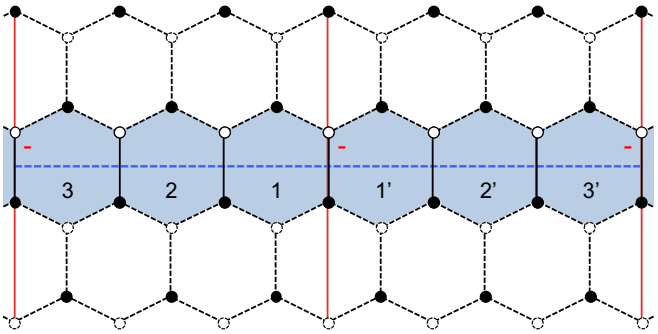


Figure 9.12: Two fixed lines orientifold realizing the twin $SU(5)$ model.

Using the same arguments as for the fixed point orientifolds in Section 9.2.1, we conclude that in all these cases the models are supported on a stripe of faces of the dimer that define an $\mathcal{N} = 2$ fractional brane.

Multiple $SU(5)$ Models

We previously saw that fixed point orientifolds can give rise to double SU(5) models. Similarly, orientifolds with fixed lines can produce multiple SU(5) models, as shown in Figure 9.13. In this case, the number of models is not restricted to two. It is important to note that, unlike in the example shown in the Figure, it is possible for different stripes to use the two fixed lines in different ways, for instance simultaneously leading to models with both SO(1) and SU(1) flavor groups, when the two lines have opposite signs. Once again, our general discussion applies to each individual stripe of blue faces, so we conclude that $\mathcal{N} = 2$ fractional branes exist for each individual stripe and hence the models are not stable.

DSB Models on a Single Fixed Line: the Twin SU(5)

There is one additional way in which an SU(5) model could be engineered. This is when both the projection needed for the antisymmetric of SU(5) and the one for canceling the anomaly due to the antifundamental, are provided by the same fixed line. This could be realized both in orientifolds with a diagonal fixed line, and in orientifolds with two fixed lines. What is important is that only one line is needed to define the relevant cluster of faces.

Importantly, since the orientifold line cannot change sign along the dimer, this possibility is effective only when the two projections have the same sign. Then the only case that fits the bill is the twin SU(5) model, as the one in Figure 9.12.

Basically, the chain of gauge groups represented by faces 1, 2 and 3 has to bend and end on the same line. There are now two possibilities. Either all the black nodes at the bottom of the edges between faces

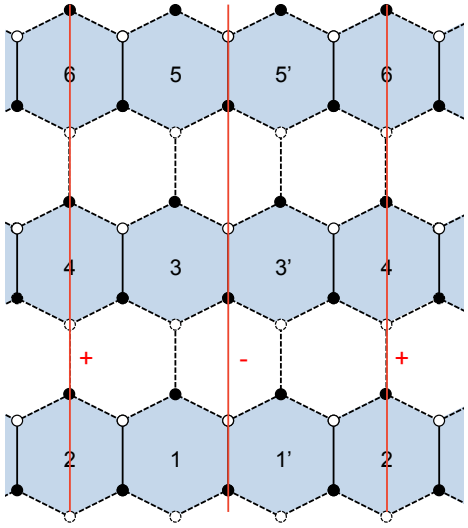


Figure 9.13: An example of the general structure of a portion of a dimer with two fixed lines giving rise to multiple SU(5) models.

1, 2 and 3 are one and the same, or the chain 1-2-3 and their images enclose some (unoccupied) faces of the dimer. The latter case is inconsistent from the dimer point of view, as shown in Section 9.1.3: such a chain cannot be a fractional brane in the parent theory. We are thus left with the former case, which in the dimer corresponds to a *hexagonal cluster of faces* around a node, as depicted in Figure 9.14.

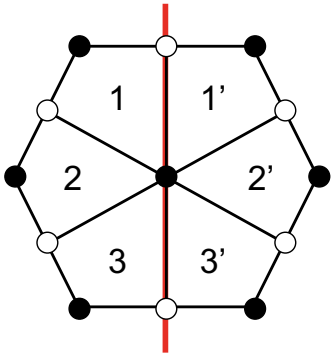


Figure 9.14: The hexagonal cluster with six faces on an orientifold line. All faces are here depicted with four edges, but some of them could have more.

Interestingly, such a collection of faces surrounding a node corresponds to a deformation fractional brane in the classification of [FHSU06]. It is reassuring that unlike in the cases with fixed points, deformation branes are compatible with line orientifolds [RU16b].

The analysis of this case is similar to what we carried out for the twin SU(5) model previously, leading to a gauge group

$$\mathrm{SU}(N+5)_1 \times \mathrm{SU}(N+1)_2 \times \mathrm{SU}(N+5)_3 \ . \tag{9.33}$$

The difference is that now the node at the center of the hexagonal cluster corresponds to a sextic superpotential term. Using the same notation as in Equation (9.30), we have

$$W = \mathrm{tr}\, A\overline{QPAP}^t\overline{Q}^t = \mathrm{tr}\, \phi\overline{\phi} \ . \tag{9.34}$$

For $N = 0$, the superpotential vanishes and we are left with two SU(5) models sharing an SU(1) flavor node, in which both surviving SU(5) factors break supersymmetry dynamically into a stable vacuum. Unlike the other realizations of the twin SU(5) model, in the present one there is no indication that the dimer must contain an $\mathcal{N} = 2$ fractional brane.

Combining the analysis in Section 9.2.1 and Section 9.2.2, we conclude that engineering a single SU(5) DSB model without instabilities at an orientifold of a toric singularity is impossible. Conversely, our analysis implies that engineering a minimal SU(5) model requires non-isolated singularities with curves of $\mathbb{C}^2/\mathbb{Z}_n$ singularities passing through the origin, which in turn result in the instability. This means that the toric diagram must contain internal points on its boundary edges. On the other hand, our analysis shows that an instance of a DSB model, the twin SU(5) model, which is compatible with an orientifold projection with fixed line(s), may exist. We should now understand whether such sub-dimer can actually be embedded into a consistent dimer and, if so, whether such dimer can be free of $\mathcal{N} = 2$ fractional branes. We will investigate these questions in the next three chapters.

9.3 3 − 2 Models

Let us now turn to the $3-2$ model, another prominent example of DSB that was recovered within brane setups at orientifold singularities in [ABMP19]. The model has gauge group SU(3) × SU(2). Its matter content is reminiscent of one SM generation:

$$Q = (\square_3, \overline{\square}_2) \ , \ \overline{U} = \overline{\square}_3 \ , \ \overline{D} = \overline{\square}_3 \ , \ L = \square_2 \ , \tag{9.35}$$

where the subindices indicate the corresponding gauge group in an obvious way. In addition, the theory has the following superpotential

$$W = \overline{D} Q L \ . \tag{9.36}$$

In principle, the above field content (SU gauge groups, (bi)fundamental matter, together with a cubic superpotential) does not seem to require an orientifold projection. As it will become clear in the following, such a projection is nevertheless necessary in order to allow for a fractional brane (i.e. an anomaly free configuration) with the desired ranks for the gauge groups.

9.3.1 General Features

Let us think more carefully about the basic features of the D-brane realization of this model. In this subsection we enumerate all different ways to recover the $3-2$ model from fractional branes at an orientifold singularity. The structure of these models is more intricate than that of the SU(5) model, so it is convenient to draw the corresponding quivers.

The candidate models are presented in Figure 9.15. In the figure, we have kept the ranks of the gauge group general by introducing N_i, $i = 1, \ldots, 4$. These additional integers account for more general configurations of D-branes at the singularity, e.g. the addition of regular or fractional D3-branes, and we posit that anomaly cancellation must hold even in those cases. The $3-2$ model arises when all N_i and the ranks of additional gauge groups, which depend on the specific singularity and are not shown in these quivers, vanish.

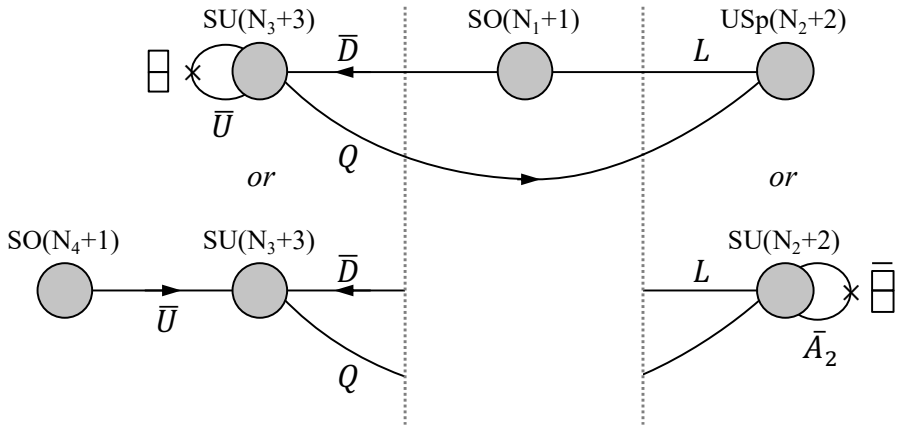


Figure 9.15: Four quivers giving rise to the $3-2$ model when all $N_i = 0$. All these models use three orientifold fixed loci.

For similar reasons as in the case of the SU(5) model, we need at least an additional gauge group factor, which we will call node 1, to serve as a flavor group providing the $\overline{D}$ and L fields. Both $\overline{D}$ and L should be connected to the same node for the superpotential Equation (9.36) to be possible. In dimer terminology, we identify the smallest building block of a $3-2$ model as three faces connected by a trivalent vertex. In this sense $3-2$ model realizations are necessarily more involved than SU(5) model realizations, since the latter only required a building block of two faces.

The quivers in Figure 9.15 should be interpreted as follows. For each of the two endpoints of the quiver, we have presented two possibilities. The two options on the left correspond to realizing $\overline{U}$ as an antisymmetric of node SU(3) or via a fourth gauge group acting as a flavor node. The two options on the right correspond to the fact that node 2 can be either USp(2) or SU(2). All possible combinations of these endpoints realize the desired $3-2$ model, therefore Figure 9.15 accounts for four models.

In principle, the flavor nodes 1 and 4 in Figure 9.15 could be SU or SO. However, if these nodes were of SU type, their ACC in the case of general ranks would require additional nodes, that come to life when regular D-branes are added. Generically, these gauge groups will give rise to new matter fields charged under the nodes of the original quiver. Such fields would contribute to and potentially help in the cancellation of anomalies. However, for N regular D3-branes, it is easy to show that for neither node the anomaly would cancel, as there would still be an imbalance of one or three units for nodes 1 and 4, respectively. In order to cancel the anomalies there are then only two options. The first is to introduce an orientifold projection. It turns out that setting both nodes to be SO is the simplest such option, and without loss of generality we will stick to it in the following. The second option is to compensate the anomaly by a mirror construction. We defer the treatment of the latter possibility to the last subsection.

It is worth noting that in two of the four models described by Figure 9.15, those for which the second gauge group is $\mathrm{SU}(N_2+2)$, we have also introduced an antisymmetric tensor $\bar{A}_2$. This field is necessary for satisfying the ACC for the more general ranks that arise when regular D3-branes are added, as we will see below. It becomes a singlet when $N_2=0$, so it decouples and does not affect the IR physics.

A final option is to get the two antifundamentals of the SU(3), $\overline{U}$ and $\overline{D}$ from the same flavor SO(1) group. However, in order to realize the $3-2$ model, the structure of the dimer model should be such that a $\overline{U}QL$ term is not present in the superpotential. This possibility is then obtained by simply identifying nodes 1 and 4 in Figure 9.15.

We thus reach the conclusion that we need no less than three orientifold projections to realize a $3-2$ model: one for the SO(1) flavor group (thus with a + sign), one for node 2 which is either USp(2) or SU(2) with an antisymmetric (in both cases, with a − sign), and one for node 3, either with an antisymmetric (− sign) or with the SO(1) flavor node 4 (+ sign). Of course some of these projections can be given by the same object, in the case of an orientifold line, provided they require the same sign.[7]

All quivers described by Figure 9.15 are viable as stand-alone gauge theories. However, as for the SU(5) model, we need to verify whether the theories remain anomaly free upon the addition of regular and/or fractional D3-branes. It turns out that the $\mathrm{SO}(N_1+1)\times\mathrm{SU}(N_2+2)\times\mathrm{SU}(N_3+3)\times\mathrm{SO}(N_4+1)$ model does not pass this test, as we show now.

Below we summarize the ACC for each of these models. Our calculations also motivate the choice of the antisymmetric tensor $\overline{A}_2$ to satisfy the ACC. For completeness, we added here as different cases also the two models where node 1 and 4 are identified.

- $\mathrm{SO}(N_1+1)\times\mathrm{USp}(N_2+2)\times\mathrm{SU}(N_3+3)$ with $\square_3$:

$$\textbf{Node 3:}\quad (N_3+3-4)-(N_1+1)+(N_2+2) \;=\; 0\,. \tag{9.37}$$

- $\mathrm{SO}(N_1+1)\times\mathrm{SU}(N_2+2)\times\mathrm{SU}(N_3+3)$ with $\square_3$:

$$\begin{aligned}\textbf{Node 2:}&\quad -(N_2+2-4)+(N_1+1)-(N_3+3) &&= 0\,,\\ \textbf{Node 3:}&\quad (N_3+3-4)-(N_1+1)+(N_2+2) &&= 0\,.\end{aligned} \tag{9.38}$$

  Note that the choice of conjugate representation for the antisymmetric tensor of $SU(N_2+2)$ is fixed by the first equation, in order to satisfy it when all $N_i=0$.

  For these two first models, the ACC reduce to

$$N_1 = N_2 + N_3\,. \tag{9.39}$$

[7] It is worth noting that in all the realizations of the 3-2 model found in [ABMP19], node 3 has an antisymmetric, node 1 is of SO type, while node 2 is USp(2) in the $\mathbb{Z}_6'$ orbifold and in PdP_4, and SU(2) with an antisymmetric in PdP_{3c}, PdP_{4b} and the $\mathbb{Z}_3\times\mathbb{Z}_3$ orbifold.

- $\underline{\mathrm{SO}(N_1+1)\times\mathrm{USp}(N_2+2)\times\mathrm{SU}(N_3+3)\times\mathrm{SO}(N_4+1)}$:

$$\textbf{Node 3:}\quad -(N_1+1)+(N_2+2)-(N_4+1)=0\,. \tag{9.40}$$

In this case, N_3 is not constrained by the ACC, which can be rewritten as

$$N_2=N_1+N_4\,. \tag{9.41}$$

- $\underline{\mathrm{SO}(N_1+1)\times\mathrm{SU}(N_2+2)\times\mathrm{SU}(N_3+3)\times\mathrm{SO}(N_4+1)}$:

$$\begin{array}{rr} \textbf{Node 2:} & -(N_2+2-4)+(N_1+1)-(N_3+3) \;=0\,,\\ \textbf{Node 3:} & -(N_1+1)+(N_2+2)-(N_4+1) \;=0\,. \end{array} \tag{9.42}$$

This translates to the two conditions

$$\begin{array}{rl} N_1= & N_2+N_3\,,\\ N_2= & N_1+N_4\,, \end{array} \tag{9.43}$$

implying $N_3=-N_4$. This in turn sets $N_3=N_4=0$, since all N_i must be positive and potentially large. In principle this issue does not rule out the possible engineering of these models, since the corresponding dimers might give rise to additional gauge groups and fields when regular D3-branes are added, in a way that anomalies are canceled. Assuming however that at least some fractional branes are needed in order to turn on all the ranks of the 3-2 model (i.e. even for $N_i=0$), then such models are excluded.

- $\underline{\mathrm{SO}(N_1+1)\times\mathrm{USp}(N_2+2)\times\mathrm{SU}(N_3+3)\text{ with }2(\overline{\square}_3,\square_1)}$:

$$-2(N_1+1)+(N_2+2)=0\,, \tag{9.44}$$

which is simply

$$N_2=2N_1\,. \tag{9.45}$$

- $\underline{\mathrm{SO}(N_1+1)\times\mathrm{SU}(N_2+2)\times\mathrm{SU}(N_3+3)\text{ with }2(\overline{\square}_3,\square_1)}$:

$$\begin{array}{rr} \textbf{Node 2:} & -(N_2+2-4)+(N_1+1)-(N_3+3) \;=0\,,\\ \textbf{Node 3:} & -2(N_1+1)+(N_2+2) \;=0\,. \end{array} \tag{9.46}$$

This can be simplified into

$$\begin{array}{rl} N_2= & 2N_1\,,\\ N_3= & -N_1\,, \end{array} \tag{9.47}$$

which has no solution beyond $N_i=0$ in the absence of additional ingredients coming from the full dimer.

The results of this appendix can be summarized in the following table:

Gauge groups	ACC
$SO(N_1+1)\times USp(N_2+2)\times SU(N_3+3)$ with $\square_3$	✓
$SO(N_1+1)\times SU(N_2+2)\times SU(N_3+3)$ with $\square_3$	✓
$SO(N_1+1)\times USp(N_2+2)\times SU(N_3+3)\times SO(N_4+4)$	✓
$SO(N_1+1)\times SU(N_2+2)\times SU(N_3+3)\times SO(N_4+4)$	✗
$SO(N_1+1)\times USp(N_2+2)\times SU(N_3+3)$ with $2(\overline{\square}_3,\square_1)$	✓
$SO(N_1+1)\times SU(N_2+2)\times SU(N_3+3)$ with $2(\overline{\square}_3,\square_1)$	✗

In the next subsections we investigate the realization of these models in terms of fixed point and fixed line orientifolds.

9.3.2 Fixed Point Orientifolds

Interestingly, for the purpose of establishing the existence of an $\mathcal{N} = 2$ fractional brane, and hence the instability of the supersymmetry breaking vacuum, it is sufficient to focus on a very small part of all these theories. In particular, all of them contain one of the following two subsectors:

- $\mathrm{SO}(N_1 + 1) \times \mathrm{USp}(N_2 + 2)$.
- $\mathrm{SO}(N_1 + 1) \times \mathrm{SU}(N_2 + 2)$ with the tensor $\overline{A}_2$.

Knowledge of the dimer around gauge groups 1 and 2 will be enough for our purposes. Let us consider the general structure of the dimers associated to these two possibilities.

- $\mathrm{SO}(N_1 + 1) \times \mathrm{USp}(N_2 + 2) \subset 3 - 2$ model

  Figure 9.16 shows the general structure of the relevant part of the dimer model. The edge between faces 1 and 2 represents the L field. Clearly, faces 1 and 2 define a stripe that winds around the unit cell of the parent dimer, giving rise to a gauge invariant that is not in the superpotential. Therefore, they correspond to an $\mathcal{N} = 2$ fractional brane.

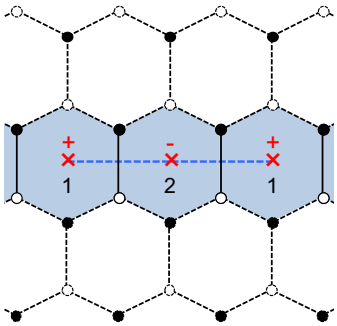


Figure 9.16: A piece of the dimer for a fixed point orientifold realizing the $3 - 2$ model with an $\mathrm{SO}(N_1 + 1) \times \mathrm{USp}(N_2 + 2)$ subsector.

- $\mathrm{SO}(N_1 + 1) \times \mathrm{SU}(N_2 + 2)$ with $\overline{A}_2 \subset 3 - 2$ model

  Figure 9.17 shows the part of the dimer that we are interested in. The edge between faces 1 and 2 corresponds to L, while the one between face 2 and its image gives rise to $\overline{A}_2$. Once again, we see that faces 1, 2 and 2' define an $\mathcal{N} = 2$ fractional brane in the parent dimer. It is interesting to note that this picture is identical to Figure 9.5 for the SU(5) model.

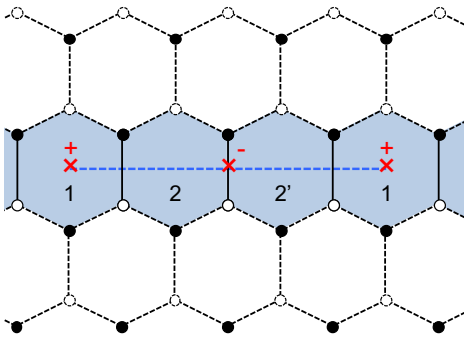


Figure 9.17: A piece of the dimer for a fixed point orientifold realizing the 3-2 model with an $\mathrm{SO}(N_1 + 1) \times \mathrm{SU}(N_2 + 2)$ with $\overline{A}_2$ subsector.

From the previous discussion, we conclude that all realizations of the $3 - 2$ model at fixed point orientifolds suffer from an $\mathcal{N} = 2$ fractional brane instability.

Models with more than one type of $\mathcal{N}=2$ fractional branes

Before moving on, let us consider the models in Figures Figure 9.16 and Figure 9.17 in further detail. As we have already mentioned, in all these cases the portion of the dimer realizing the $3-2$ model involves three fixed points. For concreteness, let us focus on the case in which $\overline{U}$ is an antisymmetric of node 3 and node 2 if of USp type. All other combinations are analogous and lead to the same conclusions. Figure 9.18 shows the general structure of the dimer model. Interestingly, in this case we can identify yet another $\mathcal{N}=2$ fractional brane, in addition to the one covered by our previous analysis. This new fractional brane corresponds to faces 1, 3 and 3' in the parent dimer and is shown in pink in Figure 9.18. We conclude that when sub-dimers as in Figures 9.16 and 9.17 are embedded in a complete dimer model, the corresponding toric singularity has at least two different types of $\mathcal{N}=2$ fractional branes. Explicit models illustrating this phenomenon were constructed in [ABMP19].

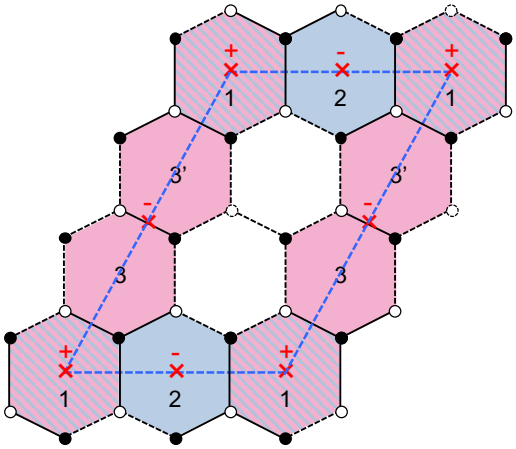


Figure 9.18: General structure of the dimer model for one of the models in Figure 9.15. This model contains two different $\mathcal{N}=2$ fractional branes. They are shown in blue and pink, with the striped face belonging to both of them.

Another interesting fact we would like to notice has to do with the intertwining between SU(5) and $3-2$ models realizations. Figure 9.18 shows that in any such configuration realizing a $3-2$ model, an SU(5) model can also be realized, by simply turning off the rank of node 2, while pumping up the rank of node 3 to SU(5). Even more, $3-2$ model realizations like the one of Figure 9.17 allow for two alternative SU(5) model realizations, the other one being by turning off node 3 and setting node 2 to SU(5), as already noticed when commenting the figure. Multiple explicit examples of this connection can be found in [ABMP19]. The only realization of a $3-2$ model that does not lead directly to a realization of the SU(5) model would be one with USp(2) at node 2 and a node 4 to compensate the anomaly of node 3. Unfortunately, no examples of this exist in the literature, and it is beyond our scope to find one here, as we have in any case shown that it would be afflicted by an $\mathcal{N}=2$ fractional brane instability.

Double $3-2$ Models

It is natural to ask whether fixed point orientifolds can lead to a pair of $3-2$ models. In this case, each of the models should use two of the four fixed points. However, all the models of Figure 9.15 need three different projections, and thus three different fixed points. One could still think about the case where nodes 1 and 4 are identified, where only two identifications are actually required. However in order for node 3 to have two different connections with node 1, the faces corresponding to this $3-2$ model realization end up being spread across all the unit cell, so that again two such models cannot coexist.[8]

9.3.3 Fixed Line Orientifolds

We now consider the realization of the $3-2$ models in orientifolds with fixed lines.

[8] It would be interesting to investigate whether such model can actually be engineered in terms of dimers. Again, since we have already proven that all realizations of the $3-2$ models at fixed point orientifolds are unstable, we do not pursue this challenging question any further.

The analysis in the case in which the $3-2$ model uses two different orientifold fixed lines is identical to the one for fixed points. In particular, it is sufficient to focus on faces 1 and 2. We simply need to replace fixed points by fixed lines in the previous discussion.

- $\mathrm{SO}(N_1+1)\times\mathrm{USp}(N_2+2)\subset 3-2$ model

  Figure 9.19 shows the relevant part of the dimer. We immediately identify an $\mathcal{N}=2$ fractional brane in the parent dimer consisting of faces 1 and 2.

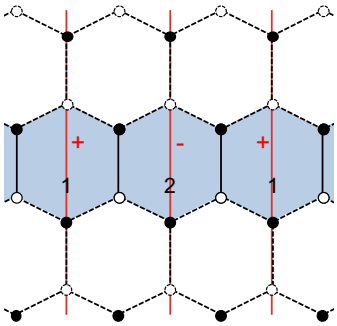


Figure 9.19: A piece of the dimer for an orientifold with two fixed lines realizing the $3-2$ model with an $\mathrm{SO}(N_1+1)\times\mathrm{USp}(N_2+2)$ subsector.

- $\mathrm{SO}(N_1+1)\times\mathrm{SU}(N_2+2)$ with $\overline{A}_2\subset 3-2$ model

  Figure 9.20 shows the part of the dimer that we focus on. Faces 1, 2 and 2' form an $\mathcal{N}=2$ fractional brane in the parent dimer.

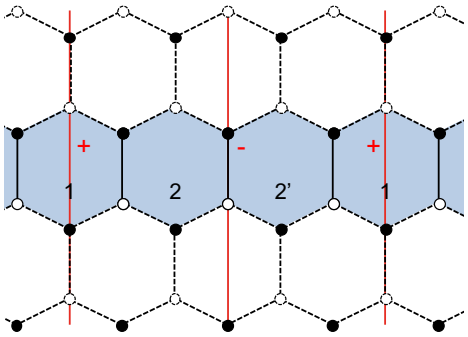


Figure 9.20: A piece of the dimer for an orientifold with two fixed lines realizing the $3-2$ model with an $\mathrm{SO}(N_1+1)\times\mathrm{SU}(N_2+2)$ with $\overline{A}_2$ subsector.

Multiple $3-2$ Models

Orientifolds with fixed lines can in principle give rise to multiple $3-2$ models, stacking them as we did in Figure 9.13 for SU(5). In this case, the projection needed for node 3 can be provided either by the line with a $-$ sign, in case of an antisymmetric, or by the line with a $+$ sign, in case of a flavor node 4. Our previous arguments show that each of these models contain (at least) an $\mathcal{N}=2$ fractional brane and are hence unstable.

Mixed SU(5) — $3-2$ Models

At this point it is interesting to point out that our arguments for multiple models, in the case of fixed lines, indicate that we can also have models that realize a combination of SU(5) and $3-2$ models. Once again, our arguments from Section 9.2 and this section show that each DSB sector would be independently unstable.

9.3.4 Twin $3-2$ models?

We are now left to investigate the possibility that the anomalies of the $3-2$ model are canceled in a twin realization, along the lines of what was done for the SU(5) model in Figures 9.8 and 9.12. Further, we would like to know if there is a realization similar to the one of Figure 9.14, i.e. on a *single* fixed line, which would not automatically imply the presence of $\mathcal{N}=2$ fractional branes.

As already alluded to, we can cancel the anomalies of a node 1 of SU nature, and/or node 4, if in the configuration there is a twin copy of the 3-2 model sharing the SU(1) node. Note that in compensating the anomaly with a twin, it is important that the two models are decoupled. If we were to use the same mechanism to compensate the anomaly of node 2, the non-zero coupling of node 2 itself would couple the twins and alter the low-energy physics of the models (typically destroying the stable supersymmetry breaking dynamics). Hence whatever we do, node 2 will always require a projection. As a consequence, if such twin model is realized in a way that it extends between two different fixed points or fixed lines, by the same arguments used around Figures 9.8 and 9.12, there will be $\mathcal{N}=2$ fractional branes that render the DSB model eventually unstable. We will thus refrain from investigating further the feasibility of such a configuration.

Finally, we would like to see if it is possible to realize a twin $3-2$ model on a single fixed line. Given that node 2 and its twin require a $-$ sign, in principle we have two options. Either both node 3 and its twin have an antisymmetric by ending-up on the same fixed line, or they compensate the anomaly by sharing an SU(1) node 4. It is easy to draw the minimal requirements for the portion of the dimer that would translate these properties, see respectively Figures 9.21 and 9.22.

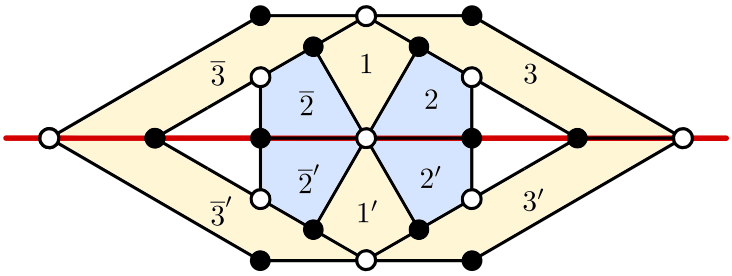


Figure 9.21: A tentative sub-dimer for a twin $3-2$ model where the SU(3) faces have an antisymmetric flavor.

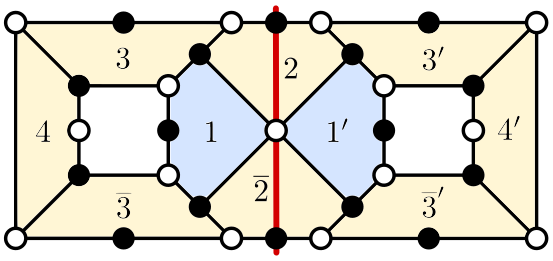


Figure 9.22: A tentative sub-dimer for a twin $3-2$ model where the SU(3) faces share a flavor $\mathrm{SU}(1)_4$.

Naively, these configurations look consistent and one can find a choice of ranks satisfying the ACC. These are the following. For Figure 9.21, $N_3 = N_{3'} = N_{\bar{3}} = N_{\bar{3}'} = M_3+3$, $N_2 = N_{2'} = N_{\bar{2}} = N_{\bar{2}'} = M_2+2$ and $N_1 = N_{1'} = M_2 + M_3 + 1$. For Figure 9.22, $N_3 = N_{3'} = N_{\bar{3}} = N_{\bar{3}'} = M_3 + 3$, $N_1 = N_{1'} = M_1 + 1$, $N_4 = N_{4'} = M_1' + 1$ and $N_2 = N_{\bar{2}} = M_1 + M_1' + 2$.

Assuming that in the parent theory every rank parametrizing the solutions above can be taken independently large, we observe that both situations would imply the existence of a fractional brane described by a ring of faces with equal ranks (up to the usual $\mathcal{O}(1)$ corrections) surrounding a hole. These are obtained by setting $M_2 = 0$ in Figure 9.21, and $M_1 = 0$, $M_1' = M_3$ in Figure 9.22. The ring-shaped would-be fractional brane is depicted in both figures by the yellow-shaded faces. As shown in Section 9.1.3, this is an inconsistent dimer. We conclude that unlike the SU(5) model, there is no way to build a stable twin version of the $3-2$ model on a single orientifold line.

Chapter 10

Inverse algorithm and triple diagrams

In Section 9.2 and Section 9.3 we have shown that the only alternative for an a priori consistent realization of a DSB model which does not automatically imply the presence of an $\mathcal{N} = 2$ fractional brane, and hence is potentially stable in the decoupling limit, is the twin $SU(5)$ living on a single fixed line of an orientifold. The twin $SU(5)$ model is described by the hexagonal cluster depicted in Figure 9.14. Now we want to understand if such cluster can be embedded in a fully consistent dimer and if such dimer can be free of $\mathcal{N} = 2$ fractional branes.

Let us first argue that in the full theory the hexagonal cluster is associated to a fractional brane. Figure 9.14 shows that the ACC are satisfied for $N_1 = N_3 = N + 4$ and $N_2 = N$. Namely, we are free to choose any value of N while all other faces of the dimer sharing an edge with the faces of the hexagonal cluster have vanishing rank. This freedom is associated to the presence of a fractional brane. The twin $SU(5)$ is obtained for $N = 1$, i.e. a single fractional brane.

Now we can ask whether this fractional brane is of deformation or runaway DSB type, in the parent theory (we already know we do not want it to be of $\mathcal{N} = 2$ type). If it were a runaway DSB brane some other regions of the dimer, besides the hexagon, would be populated and the corresponding faces would have ranks with different multiples of N [FHSU06, BBC05]. This is the key ingredient to generate an ADS superpotential and hence a runaway behavior, and this will still be true after orientifolding. Thus a runaway DSB brane in the parent theory, if it survives the orientifold, will still be of runaway type. Populating the dimer with regular branes, the runaway sector will communicate with the twin $SU(5)$ sector, destabilizing the vacuum. The other possibility is that the hexagonal cluster corresponds to a deformation brane in the parent theory and that it survives the orientifold projection. This has no instability in the parent theory, and thus we expect it to remain stable also upon orientifolding.

We are therefore looking for a dimer containing a six-valent node inside a cluster of faces, and hence the corresponding toric diagram must contain at least 6 edges whose associated ZZP are ordered around the relevant node [Gul08, IU10]. Those edges need to be in equilibrium in order for the hexagonal cluster to be on a deformation fractional brane as we have seen in Section 8.3.2, and once removed the rest of the (p, q)-web must be in equilibrium, too. This implies that we need at least two extra ZZP in equilibrium, for a total of eight. Absence of $\mathcal{N} = 2$ fractional branes in the dimer further requires that there cannot be more than one ZZP with a given winding (p, q) of the unit cell. This corresponds to toric diagrams with no more than two consecutive points which are aligned on an external edge.

In summary, we want to understand whether it is possible to find an orientifold of a dimer model satisfying the following conditions.

1. There is no $\mathcal{N} = 2$ fractional brane, or equivalently, the corresponding affine toric CY3 singularity is isolated,
2. The hexagonal cluster shown in Figure 9.14 must line on a fixed line of the orientifold involution,
3. The hexagonal cluster must be isolated on a fractional brane of deformation type,
4. There must exist solutions to the anomaly cancellation conditions in the orientifold theory.

Below, we will refer to these conditions as 1, 2, 3 and 4. Note that none of them holds *a priori*: it could indeed be the case that not the hexagonal cluster of faces of Figure 9.14 itself, but its consistent embedding in a brane tiling, requires $\mathcal{N} = 2$ fractional branes, for example.

In this chapter we will introduce a combinatorial tool called *triple diagram*, introduced in [Thu17], which allows to systematize the fast inverse algorithm, as well as built brane tilings satisfying symmetry constraints and containing particular substructures such as the hexagonal cluster in which we are interested.

Let us first and foremost recall some important characters of brane tilings from Section 7.5, and that we will need below. A *zig-zag path* (ZZP) on Γ is an oriented path of edges of Γ which turns maximally right at each black vertex, and maximally left at each white one. Zig-zag paths can be equivalently represented as strands crossing edges at their middle. A zig-zag path and the corresponding strand are displayed respectively in the middle and on the right of Figure 10.1. Note that this ZZP has a non-trivial winding around the torus: in the basis of homology induced by our choice of fundamental cell, its winding is $(1, 1)$.

A dimer model is *consistent* if 1) there is no topologically trivial ZZP 2) no lift of ZZP in the universal cover $\mathbb{R}^2$ of the torus self-intersects 3) any two distinct lifts of ZZPs in the universal cover of the torus never intersect more than once in the same direction. These technical conditions are equivalent to the existence of (possibly degenerate) isoradial embeddings of Γ in $\mathbb{T}^2$ [HV07, Bro12, IU10]. Recall that we always assume dimer models to be consistent.

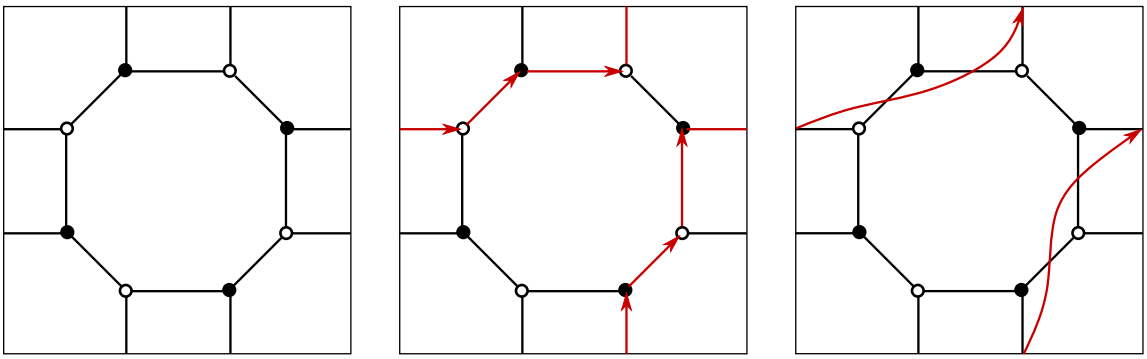


Figure 10.1: A dimer model, a zig-zag path and a zig-zag strand.

Each affine toric CY3 singularity is in one-to-one correspondence with a class of lattice convex polygons (also called *toric diagrams*) up to $\mathrm{SL}_2(\mathbb{Z})$, hence dimer models are also associated with lattice convex polygons up to $\mathrm{SL}_2(\mathbb{Z})$. This last correspondence is purely combinatorial and can be explained without having to invoke non-compact Calabi–Ay threefolds at all, as follows. Since each edge of a dimer model belongs exactly to two ZZPs going in opposite directions, the sum of windings of all ZZPs in a brane tiling is always zero. Ordering the windings according to their slope, one builds the sides of a convex lattice polygon. The choice of a fundamental cell for $\mathbb{T}^2$ translates as the action of $\mathrm{SL}_2(\mathbb{Z})$ on the corresponding polygon, and that the lattice to which the vertices of the polygon belong can be naturally identified with the integral first homology group of the torus $H_1(\mathbb{T}^2, \mathbb{Z}) \simeq \mathbb{Z}^2$.

There are, in general, multiple dimer models corresponding to the same affine toric CY3 singularity. Going from a dimer model to the class of lattice polygons up to $\mathrm{SL}_2(\mathbb{Z})$ is the *forward algorithm* and it is in general not injective; going in the opposite direction and building consistent dimer models from a class of toric diagrams can be done via the *inverse algorithm* which is one-to-many in general [FHKV08b]. Even if the fast-inverse algorithm [HV07, FHKV08b] presented in Section 7.5.2 is a great improvement with respect to the partial resolution techniques used in the first place, it remains a set of instructions which has to be carried out case-by-case with a deft hand. The goal of this chapter is to explain how triple diagrams improve this by providing a loyal fast inverse algorithm.

10.1 Triple diagrams and fast-inverse algorithm

10.1.1 Generalities

Triple diagrams have been introduced by D. Thurston in [Thu17]. A triple diagram is a collection of oriented dimension-one manifolds with boundary, mapped smoothly into a disk (the image of a connected component is called a *strand*) such that:

1. Exactly three strands cross at each intersection point,
2. The endpoints of strands are distinct points on the boundary of the disk and no other point is mapped to the boundary,
3. The orientations of the strands induce consistent orientations on the complementary regions.

Moreover, triple diagrams are considered up to isotopy. An example is shown in Figure 10.2. Note that the orientation of the strands indeed induce consistent orientations for the complementary regions.

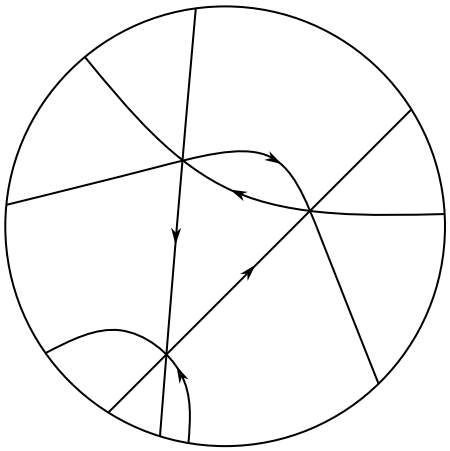


Figure 10.2: A triple diagram consisting of five strands.

Theorem 3 of [Thu17] states that given a disk with $2n$ points on the boundary labeled 'in' or 'out', and such that the labels alternate as one goes around the boundary, any matching (bijection) between 'in' and 'out' points can be realized by a triple diagram. Two examples of triple diagrams realizing a pairing of 5 'in' points labeled $i_1, \dots, i_5$ and 5 'out' points $o_1, \dots, o_5$ on the boundary of the disk (such that for each $j = 1, \dots, 5$, the pairing maps i_j to o_j) are shown in Figure 10.3.

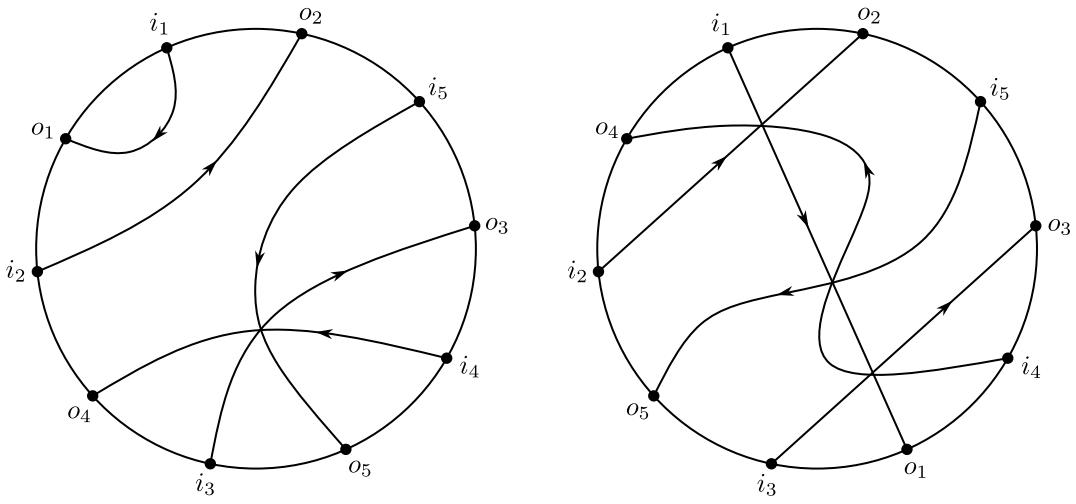


Figure 10.3: Two triple crossing diagrams realizing pairings of 'in' and 'out' points.

From a triple diagram one can always obtain a bipartite graph on the disk as follows. First, one deforms each triple intersection into a small triangle with sides oriented counterclockwise. The fact that triple diagrams are such that the complementary regions have a consistent orientation implies that after this deformation, the complementary regions can be classified in three different types.

- Either the boundary is oriented counterclockwise ; such a region is said to be *black*.
- Either the boundary is oriented clockwise ; such a region is said to be *white*.
- Or the boundary is such that as one goes along it, the orientation of the strands alternate ; such a region is a *face*.

The counterclockwise deformation of the triple diagram on the right of Figure 10.3 is shown on the left of Figure 10.4. White regions are displayed in white, black regions in black white faces are stripped.

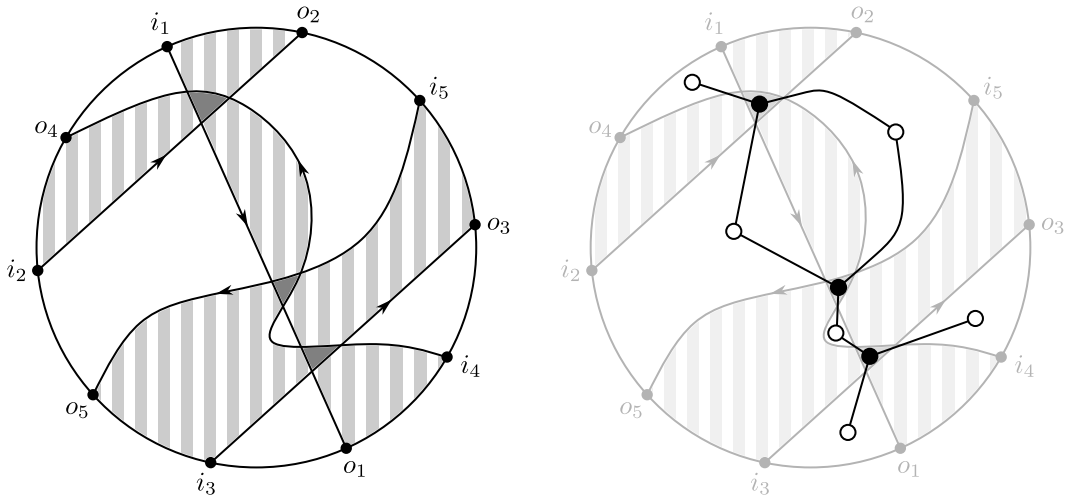


Figure 10.4: Bipartite graphs from triple diagrams.

From this, a bipartite graph is easily obtained by taking one black (resp. white) vertex for each black (resp. white) region, and connecting vertices together according to the adjacency of the regions in the deformed triple diagram. The bipartite graph corresponding to the deformed triple diagram of Figure 10.4 is displayed on the right of the same figure.

There is an elementary transformation of triple diagrams which involves six strands and which is called $2 \leftrightarrow 2$ move [Thu17]. A $2 \leftrightarrow 2$ move is displayed schematically in Figure 10.5. Given a disk with a pairing of 'in' and 'out' points on its boundary, one can restrict oneself to triple diagrams realizing the pairing and with the minimal number of faces. These are called *minimal triple diagrams*. It is proved in [Thu17] that given any two minimal triple diagrams realizing the same pairing, there always exists a finite sequence of $2 \leftrightarrow 2$ moves linking the two together. The $2 \leftrightarrow 2$ moves of triple diagrams correspond to spider moves of bipartite graphs.

10.1.2 Consistent dimer models from triple diagrams

Triple diagrams can be used to systematically construct consistent dimer models from lattice polygons, as reviewed in [GK13]. One considers the simple outgoing normal vectors to the sides of a convex lattice polygon and for each of them one draws a straight line on a torus with winding the coordinates of the vector, such that the induced 'in' and 'out' points on the boundary alternate (it is always possible to do so [GK13]). This is depicted in Figure 10.6 for the toric diagram of the affine cone over dP_1.

After having done this, one needs to keep the boundary data together with the pairing between 'in' and 'out' points only. Forgetting momentarily the identifications between opposite sides of the square, one obtains a topological disk, boundary data and a pairing as in the hypotheses for theorem 3 of [Thu17]. Thus, one can construct a minimal triple diagram realizing the pairing, deform its vertices into tiny counterclockwise triangles, and construct the corresponding bipartite graph. By construction, one is left with a consistent dimer model on a torus that corresponds to the lattice polygon we started with. It

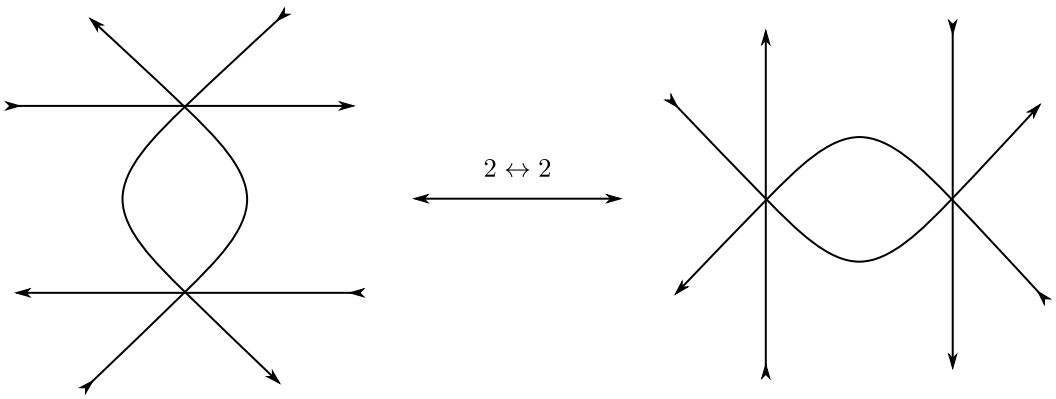


Figure 10.5: A $2 \leftrightarrow 2$ move.

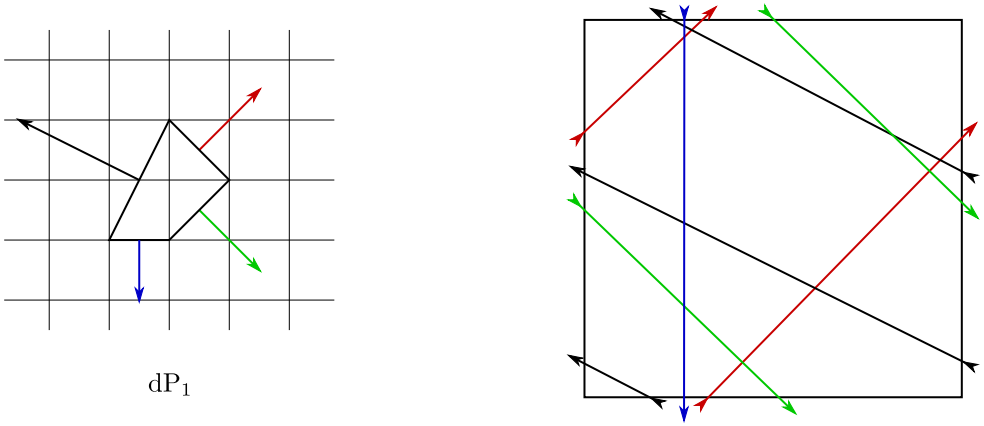


Figure 10.6: Outgoing normal vectors to the side of toric diagrams as geodesics on the torus.

might be the case that there are two-valent vertices in the resulting dimer model, which may as well be removed since they correspond to mass terms in the gauge theory described by the tiling. These two steps for the toric diagram corresponding to the cone over dP_1 are displayed in Figure 10.7. The dimer model obtained on the right of Figure 10.7 famously corresponds to dP_1 [FHV+06], as expected.

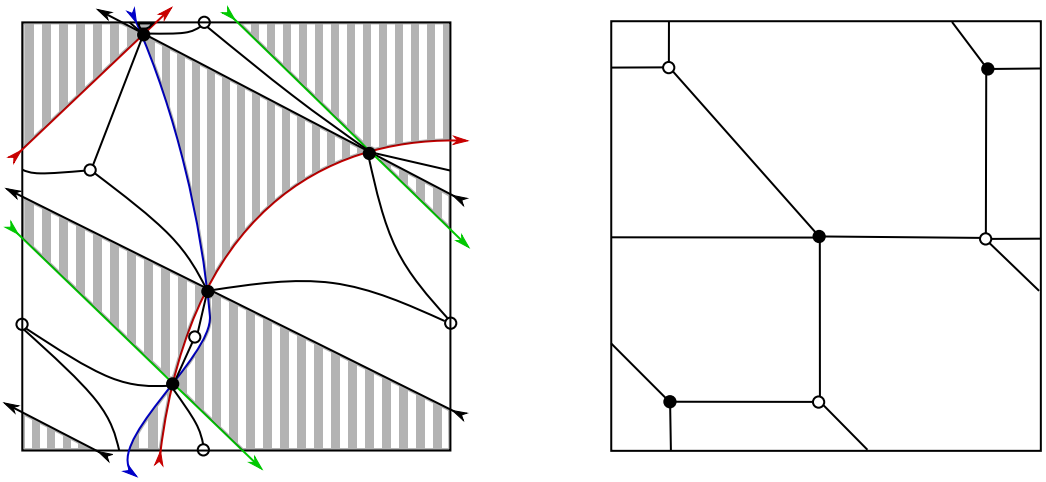


Figure 10.7: Constructing a consistent dimer model for dP_1 with triple diagrams.

This construction is particularly nice since if one starts with a toric diagram that has a particular symmetry, one can hope to preserve this symmetry during the whole process so that one obtains a dimer model also satisfying it. However, we will see in Section 10.3 that it is not always possible to do so.

10.2 Implementation of symmetries and substructures

Let us contemplate back our goal described in the introduction of this chapter: we are trying to embed the hexagonal cluster of faces describing the twin SU(5) model in a (consistent) brane tiling, such that conditions 1 to 4 of page 263 hold. Condition 4 will be analyzed at length in Chapter 11; let us for now concentrate on the first three. Our strategy is as follows.

1. In order to ensure that a dimer model does not contain $\mathcal{N} = 2$ fractional branes, it is enough to start with an isolated singularity, i.e. whose toric diagrams do not have any side containing lattice points in their interior. Doing as in the previous section yields a square with 'in' and 'out' insertion points on the boundary, together with a pairing of these insertions.

2. The brane tiling must contain the hexagonal cluster of our interest on a fixed line of the orientifold involution. We will first analyze the local arrangement of the strands induced by the hexagonal cluster and then embed it in the solid square found in the previous section. Imposing the symmetry of the brane tiling to be constructed under the orientifold involution amounts to constructing a triple diagram in one half of the fundamental cell only, and copying was has been done in the other half, in a symmetric way.

3. Imposing that the hexagonal cluster lies on a deformation fractional brane amounts to choosing wisely the strands defining the cluster. We will see below how to do this explicitly in examples.

10.2.1 The local structure of the hexagonal cluster

The local structure of the strands induced by the 6-valent vertex in the hexagonal cluster is displayed in Figure 10.8. The fixed line of the orientifold involution we need in order for the hexagonal cluster of faces to possibly implement the twin SU(5) model is displayed in dashed red. From the ideas presented in the last section we known that the data of 'in' and 'out' points of the circle surrounding the 6-valent vertex is enough to ensure that it is possible to have the 6-valent vertex inside the disk. Hence we will only keep the boundary data, and try to embed the circle with this data in a square representing the fundamental cell of a dimer model on a torus corresponding to a given class of toric diagrams.

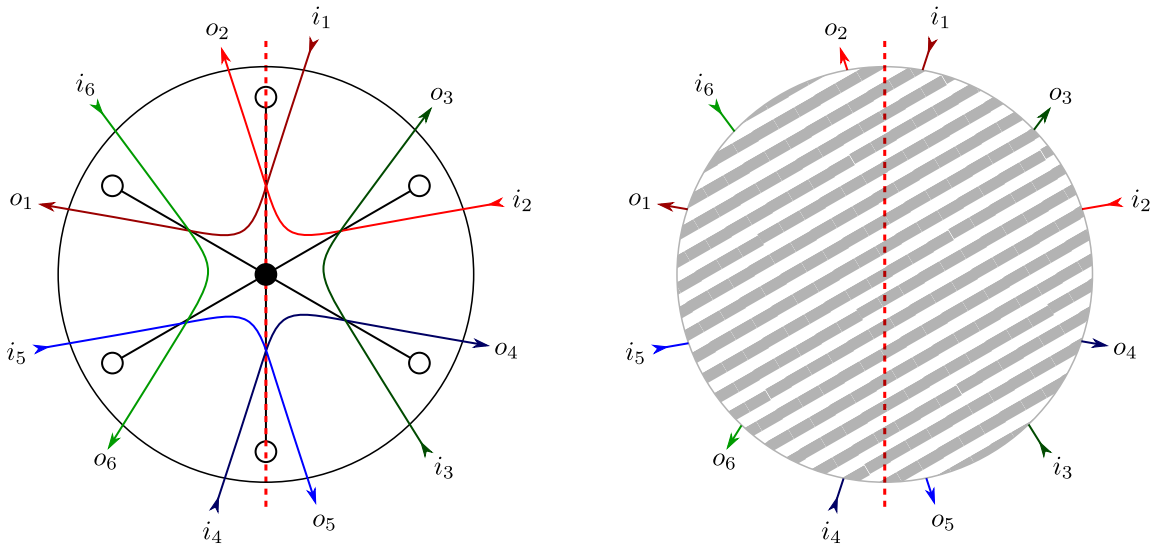


Figure 10.8: The local strand structure induced by a hexagonal vertex.

Note that we have only encoded the structure around the 6-valent vertex in the most parsimonious possible way in Figure 10.8. It might be the case that one needs more, for example to ensure that the hexagonal cluster is isolated on a fractional brane, so that the neighboring faces do not impact the anomaly cancellation conditions on the faces of the (orientifolded) hexagon, which are necessary for the twin SU(5) model to exist. We will refine the local structure we ask for, later on.

10.2.2 An example

The consistency conditions for brane tilings guides us towards considering the six ZZPs appearing in Figure 10.8 to be different. It might seem possible for e.g. the strands $i_1 - o_1$ and $i_4 - o_4$ to be identified in the full brane tiling, however this is likely to be forbidden for the hexagonal cluster to be isolated on a fractional brane. Thus, we will consider this six strands as six different ZZPs.

Since the six faces forming this hexagonal cluster must lie on a deformation brane, we need to start with a toric diagram with strictly more than six sides, and such that the latter form a sub-web in equilibrium, i.e. they can be assembled into a fine lattice polygon on their own.

It appeared experimentally that orientifolds with a single, diagonal fixed line were easier to construct than the ones two horizontal or vertical fixed lines. The first example we are going to present corresponds to a singularity whose toric diagram is displayed on the left of Figure 10.9. Note that since it has no side with lattice points in its interior, the corresponding singularity is isolated and there will be no $\mathcal{N} = 2$ fractional branes in the dimer model.

The first step of the construction consists of drawing straight geodesics on the fundamental cell of a torus, with homology prescribed by the outgoing normal vectors to the sides of the toric diagram, in such a way that it yields a set of 'in' and 'out' points on the boundary of the fundamental cell, such that 'in' and 'out' points alternate. The symmetry we are looking for needs to be implement at this step already: we ensure that the geodesics are placed in a symmetric way with respect to the reflection to the dashed red diagonal line. The result is shown on the right of Figure 10.9, and one can check that the resulting 'in' and 'out' insertions are distinct and alternate, as required.

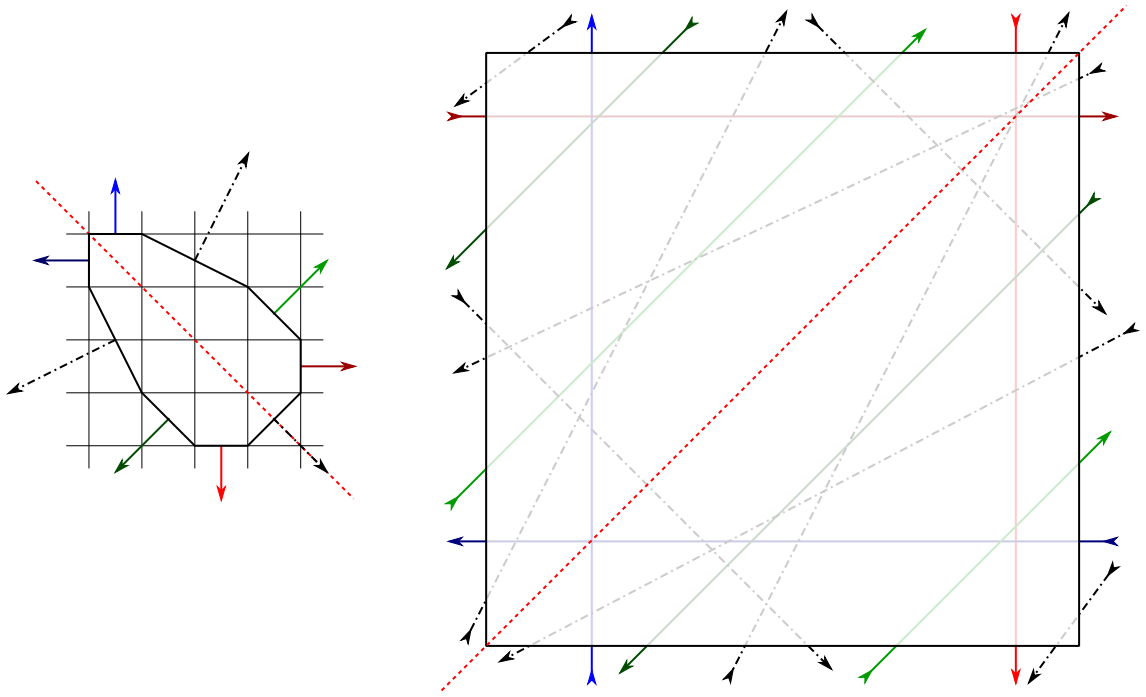


Figure 10.9: Towards a first implementation of the hexagonal cluster.

In the construction described in Figure 10.9 we have anticipated the fact that some of the strands will be the ones defining the hexagonal vertex we are aiming to. As already emphasized, we also need them to be in equilibrium. We have chosen these six strands to be the six plain ones, while the remaining three dotted-dashed ones will not contribute to the six-valent vertex directly.

Now that this is done, we can add the disk on the right of Figure 10.8 inside the fundamental cell, on top of the dashed diagonal red symmetry line. We will make some hypotheses on the local structure close to the 6-valent vertex, so that we refine the diagram on the left of Figure 10.8 to the one on the left of Figure 10.10. One can then embed this circle together with its boundary data in the rectangle on the right of Figure 10.9, and encode in another pairing the way we want the endpoints of all the strands to connect. This is depicted on the right of Figure 10.10.

Each half of the complement of the disk in the square is a topological disk with alternating 'in' and

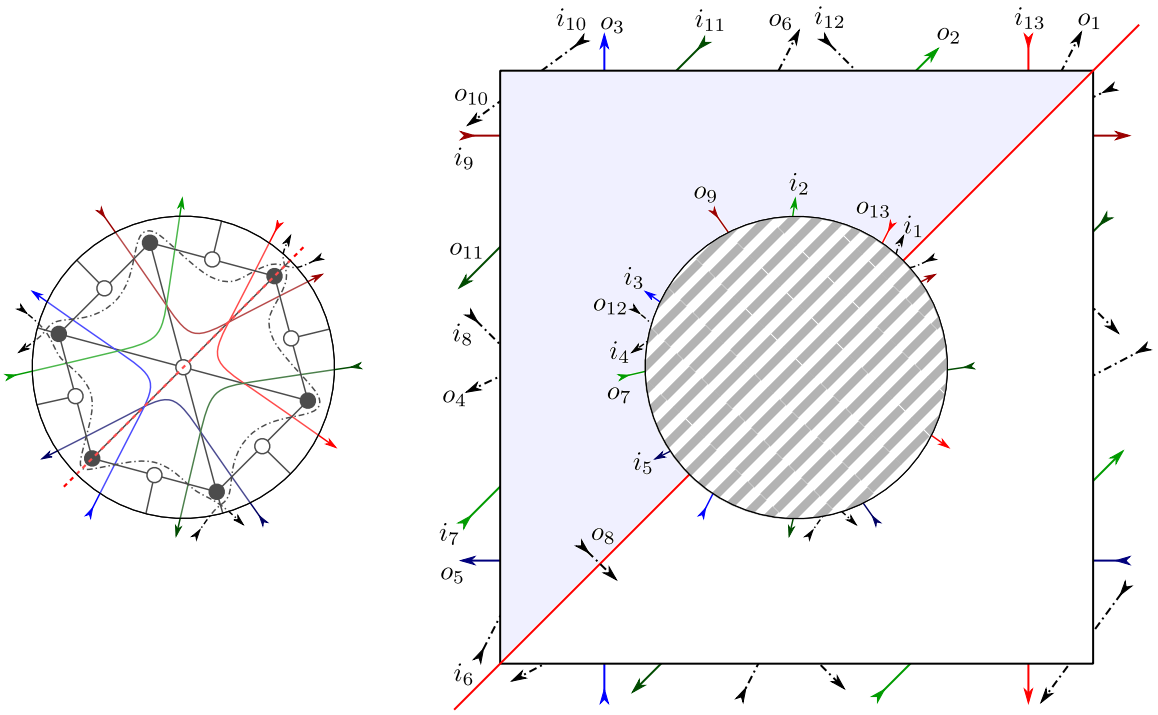


Figure 10.10: The local structure we want to implement (left) and the resulting boundary data (right).

'out' insertions on the boundary, and together with a pairing. Hence one can construct a triple diagram in each of these two regions, in a symmetric way. The resulting triple diagram with the interior of the disk restored, is shown on the left of Figure 10.11. The black, white and shaded regions resulting from a counterclockwise resolution of the vertices are also shown on this figure. One obtains in this way in dimer model on the torus, which is shown on the right of Figure 10.11 under a nice guise, and after the contraction of 2-valent vertices. Some faces are colored in order to emphasize the fact that they do not end on the boundary of the fundamental cell.

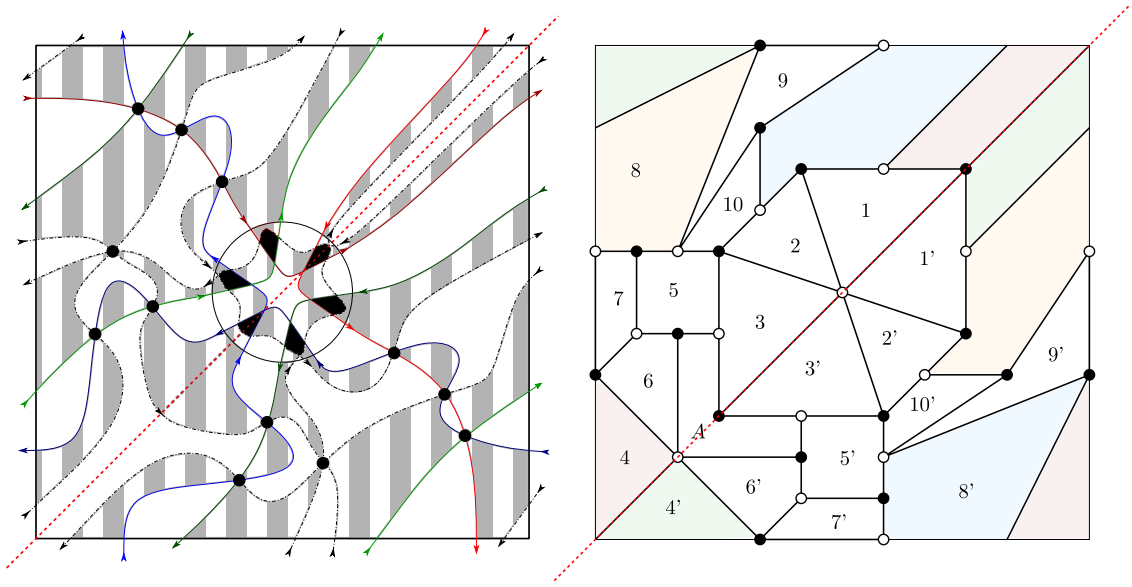


Figure 10.11: The resulting triple diagram and its corresponding dimer model.

This dimer model satisfies the condition 1 of page 263 by construction, as well as the conditions 2 and 3 One can check that this brane tiling is consistent: first of all, it has 21 faces, which is twice the area

of its toric diagram, as expected. Moreover, for some Pfaffian orientation one finds that the determinant of its Kasteleyn matrix is

$$x^{-2}y^{-2}\left(y^4x^4+x^4y^3+x^3y^4-86x^3y^3+57x^3y^2-x^3y+57y^3x^2+555x^2y^2+86x^2y+x^2 \right.$$
$$\left. -y^3x+86y^2x-81y^2x+x+y^2+y\right)\,. \qquad (10.1)$$

and the coefficients corresponding to the vertices of the toric diagram have absolute value 1, which is in agreement with the fact that the dimer model is consistent.

Concerning the condition 3., the hexagonal cluster of faces will lie on a deformation fractional brane if the ZZPs defining its hexagonal node form a sub-web in equilibrium in the toric diagram. In Figure 10.12 the hexagonal cluster of faces is indeed isolated on a deformation fractional brane.

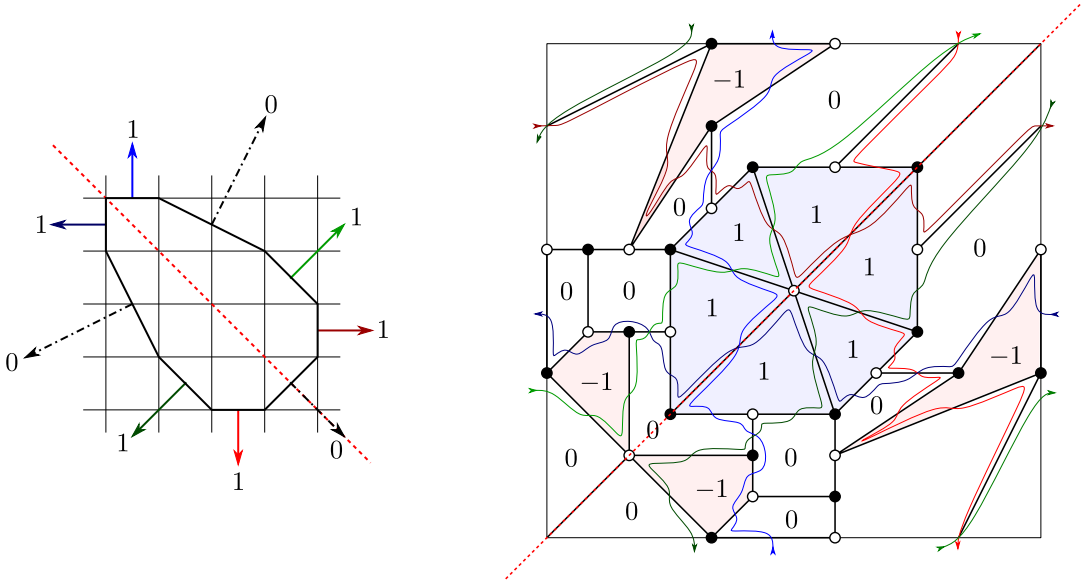


Figure 10.12: The deformation fractional brane on which the hexagonal cluster lies.

Unfortunately, and despite the fact that conditions 1 to 3 are satisfied in this dimer model, condition 4 does not hold: there is no solution to the anomaly cancellation conditions in the orientifold theory, if no flavor D7-branes are added on top of the regular and fractional D3-branes. Since it is our working hypotheses, this model thus has to be discarded. This is more precisely stated by saying there is no solution to the system of affine equations:

$$\begin{aligned}
(N_1-4)-N_4+N_8-N_2 &= 0\\
N_1-N_8+N_{10}-N_3 &= 0\\
-(N_3-4)+N_2-N_5+N_A &= 0\\
N_1-N_8+N_6-(N_4-4) &= 0\\
N_3-N_A+N_6-N_7+N_8-N_{10} &= 0\\
N_A-N_4+N_7-N_5 &= 0\\
N_5-N_6+N_9-N_8 &= 0\\
N_7-N_5+N_9-N_4+N_1-N_2+N_{10} &= 0\\
N_{10}-2N_8+N_7 &= 0\\
N_8-N_2+N_5-N_9 &= 0
\end{aligned} \qquad (10.2)$$

One might think that it would be possible to reach some phase of the orientifold theory which admit solutions to the ACC, after Seiberg-dualizing some faces. However it has not been the case for any of the Seiberg-dual phases for which these ACC were computed, and we will see in the next chapter that this is in fact general for orientifold theories obtained as the quotient under the reflection through a diagonal line in a dimer.

Nevertheless, this model is interesting for it proves in a direct way that it is possible to embed the hexagonal cluster of the twin SU(5) model in a dimer model symmetric with respect to the reflection through a line, and such that this hexagonal cluster of faces is isolated on a deformation fractional brane.

10.2.3 Another example

Another interesting example is the one displayed in Figure 10.13, for the deformation fractional brane on which the six-valent node lies consists only of the hexagonal cluster and isolated faces, all of them having rank 1 (there is no face of rank -1, as was the case in Figure 10.12).

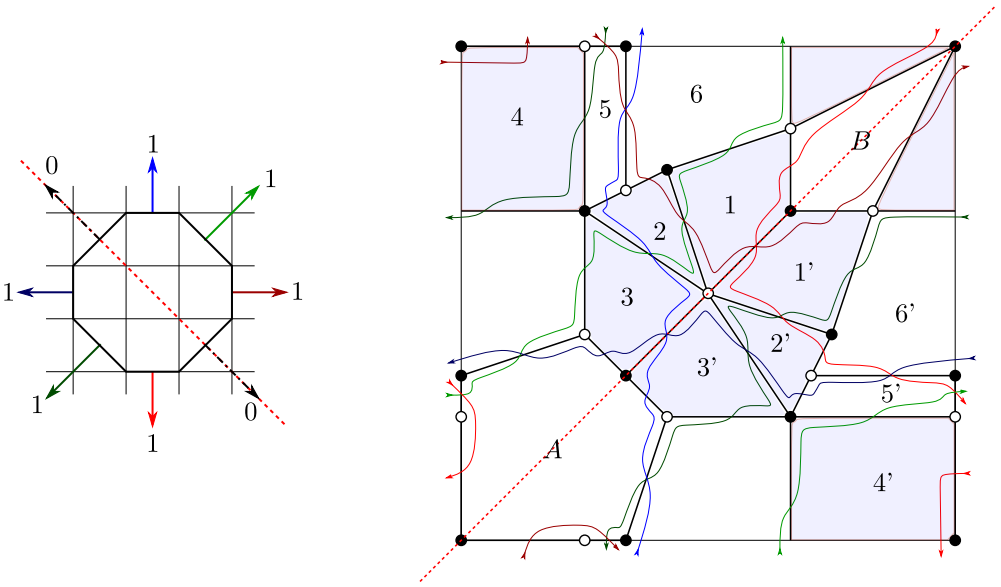


Figure 10.13: Another example

The determinant of the Kasteleyn matrix corresponding to some Pfaffian orientation is:

$$x^{-2}y^{-2}(x^2y+x^3y+xy^2-26x^2y^2+25x^3y^2-x^4y^2-x^4y^2+xy^3+25x^2y^3+26x^3y^3+x^4y^3-x^2y^4+x^3y^4)\ , \tag{10.3}$$

and hence is in agreement with the fact that this dimer model is consistent. It has 14 faces and is minimal since 14 is also twice the area of the octagon displayed on the left of Figure 10.13. This model however is plagued with the same issue as the previous one, namely: the anomaly cancellation conditions in the orientifold theory do not have any solution. They write:

$$\begin{aligned} N_2+4-N_1+N_4-N_7 &= 0\\ N_3-N_1+N_7-N_8 &= 0\\ N_3-4+N_7-N_2-N_5 &= 0\\ N_7-N_8-N_4+N_5 &= 0\\ N_1-N_2+N_8-N_6-N_5+N_3 &= 0\\ N_2-N_6-N_7+N_5 &= 0 \end{aligned} \tag{10.4}$$

One can check directly that they admit no solution indeed.

It would be very interesting to criteria telling when a brane tiling orientifold possibly admits solutions to the anomaly cancellation conditions (without the use of fractional branes), already at the level of the toric diagram. It might very well be possible that this constraint coming from the ACC ultimately forbids the implementations of the hexagonal cluster of the twin SU(5) model with the properties that we are after. This will be the aim of the next chapter.

10.3 Reflection symmetries in dimer models

In this last section we will present a precise result, under the form of an example, which shows that orientifolds of brane tilings with fixed horizontal (equiv. vertical) lines are more severely constrained than the ones with a single diagonal fixed line. This is the reason why the examples we have presented so far consist of diagonal line orientifolds only. Ultimately, these additional constraints come from the implementation of the symmetry itself: when placing ZZP on the square in the first step of the loyal

fast inverse algorithm, imposing the symmetry with respect to fixed horizontal (equiv. vertical) lines often implies ZZP crossings on the boundary of the fundamental cell, while this is not the case for the symmetry with respect to a diagonal fixed line.

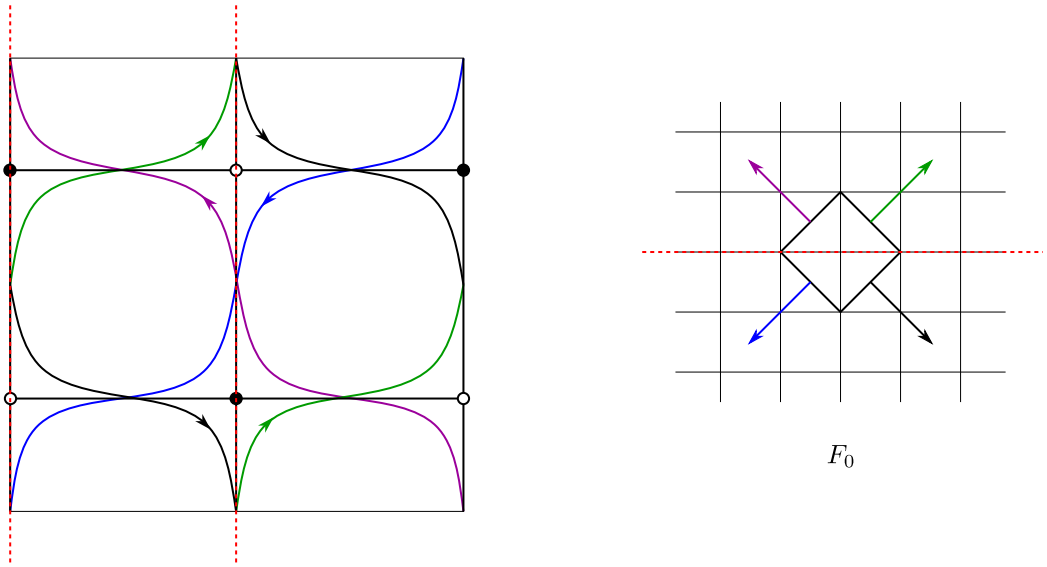


Figure 10.14: The dimer vertical line symmetry leads to a horizontal line symmetry of the toric diagram.

If a brane tiling admits a symmetry for some choice of fundamental cell, the zig-zag paths defined by the dimer model are mapped one to another under this symmetry. Since zig-zag paths are identified with outgoing normal vectors to the sides of the toric diagram, the latter also enjoys a symmetry related to the one of the brane tiling. This has been first studied in [RU16b] for orientifold projections with fixed point and fixed line(s), and we will come back to this while discussing other orientifold projections in Chapter 13. For example, the case of a brane tiling corresponding to the affine cone over the Hirzebruch surface F_0 is displayed in Figure 10.14: the vertical line symmetry of the brane tiling induces a reflection symmetry with respect to a horizontal line in the toric diagram.

One may then wonder: since some symmetry of the toric diagram is a necessary condition for the existence of a corresponding symmetry in a brane tiling, is it also a sufficient condition? In this section we prove that it is not the case when the symmetry in the dimer model is a reflection through a fixed horizontal (equiv. vertical) fixed line.

Let us consider the toric diagram on the left of Figure 10.15, from which it appears that no ZZP is mapped to itself under the reflection with respect to the dashed red line. Hence there must be edges on each symmetry axis in the dimer model, and subsequently it is always possible to choose a symmetric path as the thick black one displayed on the right of Figure 10.15, such that the total number of intersections (counted with signs) between this path and ZZPs in the left half of the cell is zero. Let $n_i \in \mathbb{Z}$, $i = 1, \ldots, 6$ be the number of intersections (with signs) between the black path and the i-th ZZP in the left bluish half of the fundamental cell – and remember that we have chosen the path so that $\sum n_i = 0$. For example, on the schematic tiling on the right of Figure 10.15 one has $n_1 = n_2 = 1$ and $n_4 = n_6 = -1$.

If the dimer model under consideration is symmetric with respect to the dashed red lines, then the black path intersect (with signs) $-n_6$ times the ZZP (1) on the right half of the cell since the ZZPs (1) and (6) are exchanged under the symmetry. Then, from the toric diagram one infers that $n_1 - n_6 = 1 = (1,0) \wedge (1,1)$ since the winding of the black path is $(1,0)$ and hence as it loops around the torus it crosses once (with sign) the ZZPs (1) whose winding is $(1,1)$. Similarly, $n_2 - n_5 = 1$ and $n_3 - n_4 = 1$. Hence $\sum n_i = 2(n_4 + n_5 + n_6) + 3$, which can not be zero since the n_i's are integers, and we are led to the conclusion that there can not be any symmetric dimer model corresponding to this toric diagram and this symmetry. The example of above can be generalized straightforwardly to many other toric diagrams, and answers negatively the question of whether there always exist symmetric dimer models when the toric diagram is compatible with this symmetry. It might be that one needs intrinsically non-perturbative sectors [GEH15] to describe the corresponding orientifolds, if any.

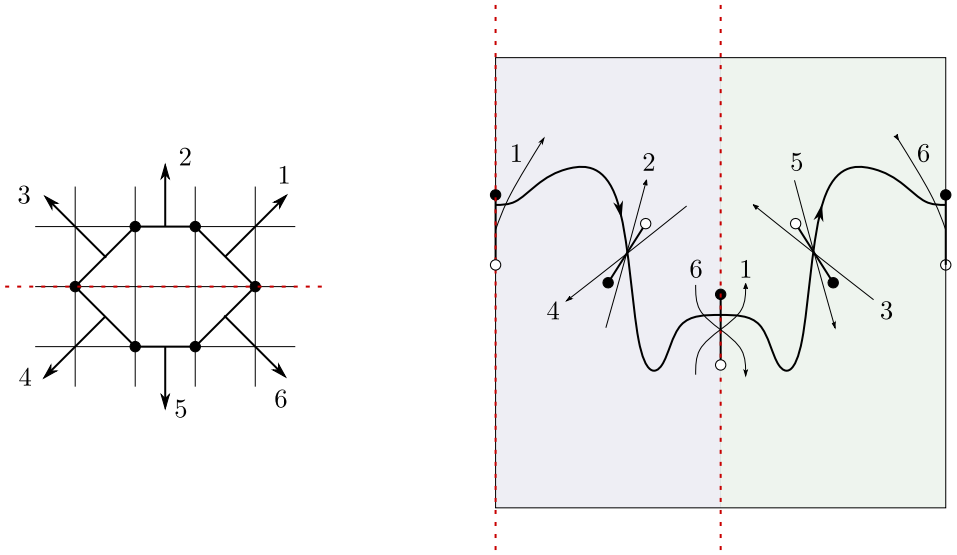


Figure 10.15: From a polygon to a consistent dimer model through triple crossing diagrams.

As one wonders about the existence of constrained dimer models, ZZPs-based arguments provide in general merely *necessary* criteria – such as the one we just derived, whereas the existence theorems for triple point diagrams let one hope for *sufficient* conditions. The knowledge of these would certainly deepen our understanding of orientifolds of affine toric CY3 singularities.

Chapter 11

Dimers orientifolds and anomalies

This chapter is essentially [ABF$^+$21a]. Our presentation is organized as follows. In Section 11.1 we review the dimer model description of gauge theories on D-branes at toric singularities and the role of zig-zag paths in solving anomaly cancellation conditions. In Section 11.2, we consider the more involved case of orientifolds, for which anomaly cancellation conditions generically correspond to non-homogeneous linear systems of equations due to the presence of tensor matter. In Section 11.3, we generalize the algorithm for solving anomaly cancellation conditions based on zig-zag paths to the case of orientifolds. This analysis will lead to the main result of the paper, which we present in Section 11.4: a purely geometric criterion for anomaly cancellation conditions in orientifold field theories just based on the toric data of the singularity. Section 11.5 contains a summary of our results and an outlook.

11.1 Gauge anomalies and dimer models

Brane tilings substantially streamline the connection between the probed toric CY 3-folds and the corresponding quiver theories thanks to powerful combinatorial tools, such as *perfect matchings* and *zig-zag paths* (ZZPs) [FHK$^+$06, HV07, FHKV08a, FV06]. We will be particularly interested in ZZPs, which we recall to be oriented paths on the dimer that go along edges and turn maximally right (resp. left) when they meet a white (resp. black) node. These paths close forming loops with no self-intersections that have non-zero homology around the $\mathbb{T}^2$ in which the diagram is embedded. The two homologies are called mesonic charges, and are associated to two of the three $U(1)$ toric actions of the toric CY 3-fold, the remaining one being the $U(1)_R$ R-symmetry. Remarkably, ZZPs are in one-to-one correspondence with legs in the (p, q) web diagram [HV07, FHKV08a] (equivalently, the outward pointing vectors normal to the external edges of the toric diagram), where their (p, q) labels are exactly the homology charges of the ZZPs.

The ranks N_i of the gauge groups associated to faces in the dimer reflect the configuration of branes at the singularity. These branes include both regular and fractional D3-branes. The latter correspond to higher dimensional branes wrapped on vanishing compact cycles [DDG98]. D-branes source RR tadpoles that must be canceled, this being the geometric counterpart of anomaly cancellation in the dual gauge theory. Tadpole cancellation amounts to cancelling the flux sourced by the branes in compact homology. The configuration $N_i = N$ corresponds to having N regular D3-branes and no fractional branes. This configuration is always tadpole free since the CY is non-compact and the RR flux sourced by regular D3-branes can escape all the way to infinity. Configurations with unequal ranks are obtained by adding fractional branes. The number of independent, anomaly-free fractional brane configurations is in one-to-one correspondence with the number of compact 2-cycles whose Poincaré duals in the CY are non-compact 4-cycles. Indeed, D5 branes wrapped on these (and only these) 2-cycles are tadpole free since the RR flux can again escape to infinity. The number of non-compact 4-cycles, hence of consistent fractional branes, can be related to the number of ZZPs [But06],

$$\#\text{ZZPs} - 3 = \#\text{Fractional Branes} \,. \tag{11.1}$$

Note, in passing, that tadpole cancellation is equivalent to cancellation of non-abelian gauge anomalies even formally for gauge groups with zero rank [Ura01].

In this work we will use anomaly cancellation conditions (ACC) and tadpole cancellation interchangeably. In fact, some $U(1)$ factors can be anomalous and become massive through a generalization of the Green-Schwarz mechanism. Even some of the non-anomalous $U(1)$'s can become massive, see [IRU99]. In the following, we will only consider the anomaly of the non-abelian part of the gauge groups. Henceforth we will often denote them as just $SU(N)$.

Cancellation of anomalies at a given node of the quiver corresponds to having the same number of incoming and outgoing arrows (weighted by the ranks of the nodes at their other endpoints). This is encoded in the (antisymmetric) exchange matrix A defined as $A = \tilde{A} - \tilde{A}^T$, where $\tilde{A}$ is the adjacency matrix of the quiver. The latter is a matrix whose elements $\tilde{A}_{ij}$ count the number of bifundamental chiral superfields $(\overline{\square}_i, \square_j)$ charged under the gauge group i and the gauge group j. With an abuse of language in the following we will call A the adjacency matrix, for simplicity. The matrix A is only sensitive to the chiral content of the theories (e.g. it is zero for the conifold). Cancellation of anomalies amounts to solving the homogeneous system of equations defined by this matrix, that is, finding $\mathrm{Ker}(A)$. Integral vectors in $\mathrm{Ker}(A)$ encode the ranks associated to regular and fractional branes. In this chapter we will investigate how these conditions are modified once orientifold planes are introduced, following [ABF+21a] closely.

11.1.1 Anomalies and Zig-Zag Paths

Throughout this paper, we will find ZZPs to be particularly useful for our purposes. With this in mind, we now review a method for finding anomaly-free rank assignments of dimers based on ZZPs [But06].

We can regard every ZZP as defining an "anomaly-free wall" on the dimer. This is because, due to its definition, every time a ZZP overlaps with a face in the dimer, it does so over exactly a pair of consecutive edges.[1] These two consecutive edges correspond to an incoming and an outgoing arrow in the quiver for the gauge group associated to the face under consideration.[2] This implies that if we add some constant to the ranks of all the faces on one side of the ZZP, the ACC of the faces on the other side of the ZZP do not change, as illustrated in Figure 11.1.

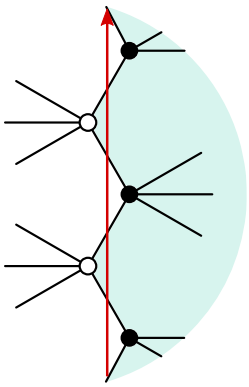


Figure 11.1: A ZZP as an anomaly wall.

With this insight, one recovers the algorithm to construct anomaly-free rank assignments for dimer models outlined in [But06]:

1. The set of ZZPs is given by $\{(p_\Gamma, q_\Gamma) | \Gamma = 1, \ldots, n\}$, where p_Γ and q_Γ are the winding numbers of the ZZP Γ, with respect to a fixed unit cell. To every (p_Γ, q_Γ) assign an integer coefficient v_Γ.

2. Choose one face and assign rank zero to it.

3. In going from face a to an adjacent face b, the rank of the latter will be

$$N_b = N_a + v_\Gamma - v_\Delta \,, \tag{11.2}$$

[1] By overlapping with a face, we mean sharing an edge with it, not just touching it at a node.

[2] More generally, a ZZP might overlap with a given face more than once. Every overlap involves a pair of consecutive edges, so the previous discussion still applies. For an explicit example of this situation, see the yellow ZZP in the dP_1 dimer model shown in Figure 11.2a.

where v_Γ is the coefficient of the ZZP moving to the left with respect to the path from a to b, and v_Δ is the one in the opposite direction. This operation is well defined since we are working on an oriented surface, which implies that we can consistently speak of "right" and "left" of a ZPP.

4. Finally, one must impose two constraints which ensure that the rank assignment is single valued. Consider, for instance, moving along a loop along one of the two cycles of the fundamental cell. When returning to the initial face, the rank should be unchanged. This is granted by imposing

$$\Lambda = \sum_\Gamma v_\Gamma p_\Gamma = 0, \quad M = \sum_\Gamma v_\Gamma q_\Gamma = 0\,. \tag{11.3}$$

We will refer to these two conditions as the Λ and M *topological constraints*.

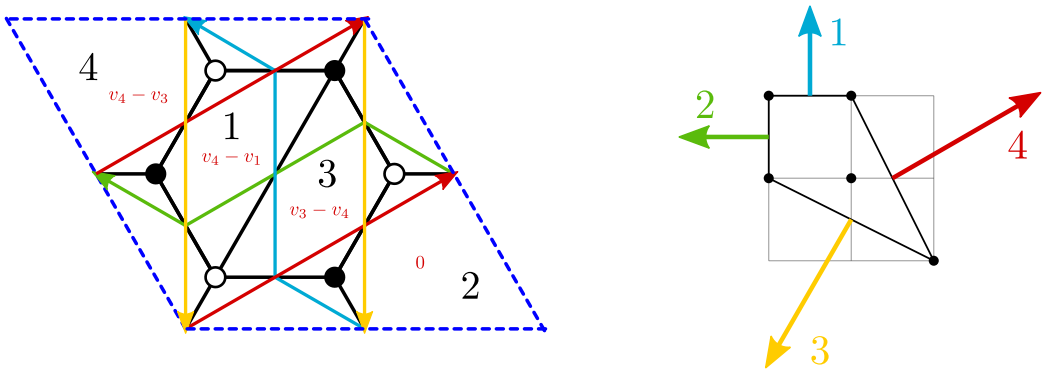
Every choice of values for the v_Γ's consistent with the topological constraints Equation (11.3) gives rise to an anomaly-free rank assignment. Moreover, notice that, by construction, every rank assignment is invariant under a global shift $v_\Gamma \to v_\Gamma + k$. One may use this freedom to fix one of the v_Γ's (equivalently, one of the ZZPs is not independent). There are thus two constraints and one redundancy to be fixed, reproducing the expected number of fractional branes in Equation (11.1). In other words, this construction can account for the most general anomaly-free rank assignment, up to an overall shift of the ranks (i.e. regular branes). Generically, this algorithm can produce negative ranks for some faces, which cannot be directly interpreted as ranks of gauge groups. Of course, one may always add regular branes until all ranks are positive.[3]

11.1.2 Examples

Let us illustrate the algorithm in Section 11.1.1 with two examples, to which we will return when discussing orientifolds.

$\mathbf{dP_1}$

Consider the toric phase of the complex cone over dP_1, or dP_1 for short, which is shown in Figure 11.2. Let us apply the method described above for the computation of the fractional branes.



(a) (b)

Figure 11.2: (a) Dimer diagram for dP_1. We show the ZZPs and the rank assignments coming from them. (b) The toric/web diagram.

[3]It is worth noting that this procedure is closely related to the algorithm for constructing fractional brane rank assignments introduced in [BHK05], in which the difference in the ranks between two nodes in the quiver is proportional to the baryonic $U(1)$ charge of the bifundamental field connecting them, with one independent vector for each baryonic $U(1)$. The relation between the two methods is through the correspondence between baryonic $U(1)$ symmetries and extremal perfect matchings [BZ06] or, equivalently, ZZPs (which correspond to differences between consecutive external perfect matchings). Our procedure is also equivalent to the one for labeling cluster variables associated to plabic (i.e. planar bicolored) graphs using ZZPs [Sco06], and to the even more similar one in the context of cluster integrable systems [GK13] that associates a divisor at infinity on the spectral curve, to each face of the bipartite fat graph under consideration [Foc15].

Let us choose $N_2 = 0$. Applying the algorithm, the other faces are assigned the following ranks:

$$\begin{array}{ccc} N_1 & \leftrightarrow & v_4 - v_1 \\ N_2 & \leftrightarrow & 0 \\ N_3 & \leftrightarrow & v_3 - v_4 \\ N_4 & \leftrightarrow & v_4 - v_3 \end{array} \tag{11.4}$$

From the toric diagram we read the two topological constraints:

$$\left\{ \begin{array}{rcccc} \Lambda & = & -v_2 - v_3 + 2v_4 & = & 0 \\ M & = & v_1 + v_4 - 2v_3 & = & 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rcl} v_1 & = & 2v_3 - v_4 \\ v_2 & = & -v_3 + 2v_4 \end{array} \right. \tag{11.5}$$

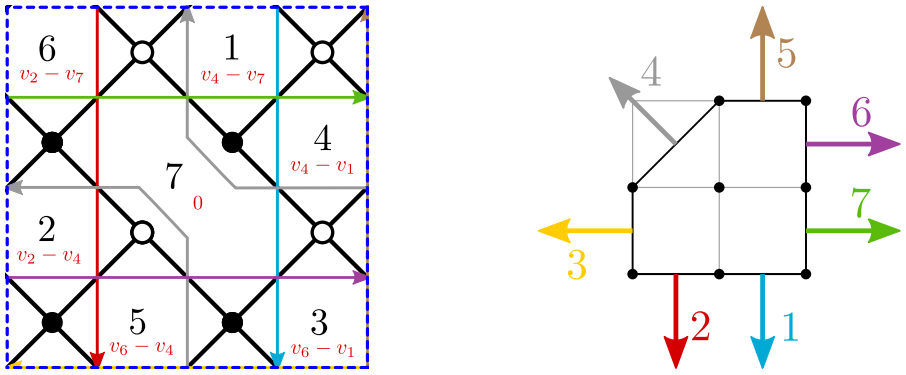
We can further use a global shift of the v_i to set $v_4 = 0$ and find the following rank assignment:

$$\begin{array}{rcrcl} (v_1, v_2, v_3, v_4) & \equiv & \mathbf{v} & = & (2, -1, 1, 0)v_3 \,, \\ (N_1, N_2, N_3, N_4) & \equiv & \mathbf{N} & = & (-2, 0, 1, -1)v_3 \,. \end{array} \tag{11.6}$$

These ranks correspond to a well-known dynamical SUSY breaking fractional brane of dP_1 [BHOP05, FHSU06, BBC05]. We will return to dP_1 in Section 11.3.1.

$\mathbf{PdP_4}$

As a slightly more complicated example, let us study the case of the PdP_4 singularity [FFHH03] in the toric phase considered in [ABMP19] and given in Figure 11.3.



(a) (b)

Figure 11.3: (a) Dimer diagram for PdP_4. We show the ZZPs and the rank assignments coming from them. (b) The toric/web diagram.

From the toric diagram we read:

$$\left\{ \begin{array}{rcl} \Lambda & = & v_6 + v_7 - v_3 - v_4 = 0 \\ M & = & v_5 + v_4 - v_1 - v_2 = 0 \end{array} \right. \Leftrightarrow \left\{ \begin{array}{rcl} v_3 & = & v_6 + v_7 - v_4 \\ v_5 & = & v_1 + v_2 - v_4 \end{array} \right. \tag{11.7}$$

Since a global shift in the v_i does not change the rank assignments, we can impose $v_4 = 0$. We then find the following rank assignment,

$$\mathbf{v} = (v_1, v_2, v_6 + v_7, 0, v_1 + v_2, v_6, v_7) \quad \rightarrow \quad \mathbf{N} = (-v_7, v_2, v_6 - v_1, -v_1, v_6, v_2 - v_7, 0) \,. \tag{11.8}$$

We will return to this example in Section 11.3.1 upon orientifolding it.

11.2 Anomaly Cancellation Conditions in Orientifolds

Determining whether an orientifolded theory admits anomaly-free solutions and, if so, finding them is a relatively straightforward task in a case by case basis. Indeed, writing down the set of anomaly equations

for every gauge group and looking for solutions is not more complicated than for non-orientifolded models. In this section we systematize this calculation, introducing an algorithm for finding anomaly-free solutions in the presence of orientifolds. This, in turn, will allow us to relate the calculation to the one in the unorientifolded theory and, at a later stage, to extend the geometric determination of solutions in terms of zig-zag paths to orientifolds.

In what follows, we will refer to the original, unorientifolded theory as the *mother theory*. Similarly, we will dub the orientifolded theory the *daughter theory*. As described in Section 11.1 finding an anomaly-free rank assignment for the mother theory amounts to finding the kernel of its adjacency matrix (more precisely of the matrix $A = \tilde{A} - \tilde{A}^T$, with $\tilde{A}$ the proper adjacency matrix, see the discussion in Section 11.1). Tensor matter in the daughter theory modifies the ACC, dovetailing the contribution of the O-planes to the RR-charges that must cancel in compact homology. In general, the anomaly/tadpole problem of orientifolded theories corresponds to a non-homogeneous linear system of the form:

$$\bar{A} \cdot N = f \,, \tag{11.9}$$

where $\bar{A}$ is the adjacency matrix of the daughter theory, and f stands for the additional contribution of tensor matter. The difference between two solutions of the system Equation (11.9) is a solution of the corresponding homogeneous one, i.e. it is in the kernel of $\bar{A}$. If one knows a particular solution N_{part} of Equation (11.9), every solution N can be expressed as:

$$N = N_{\text{hom}} + N_{\text{part}} \,, \tag{11.10}$$

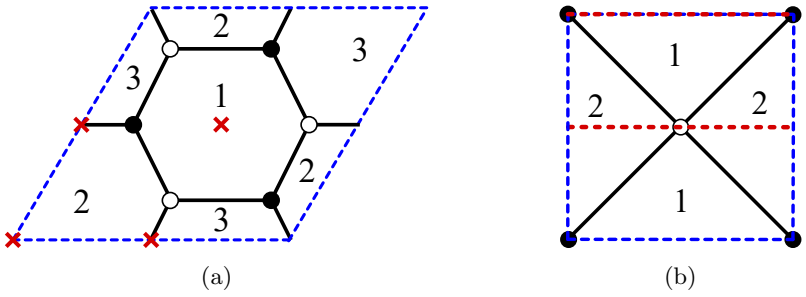
where N_{hom} is a solution of the homogeneous system $\bar{A} \cdot N_{\text{hom}} = 0$.

Remarkably, we will show that whether Equation (11.9) has solutions or not can be directly determined from the toric diagram of the singularity under consideration. In other words, we will establish a geometric criterion for the satisfiability of the ACC in orientifolded theories.

11.2.1 Dimers and Orientifolds

Before proceeding let us recall a few basic features of orientifolds in dimer language from Section 7.6. More details can be found in [FHK+07]. Related works include [IKY08, GEH15, GEH17].

The operation of orientifolding has been studied as a $\mathbb{Z}_2$ involution of the dimer diagram, leaving either fixed points or fixed line(s).[4]. Examples of both possibilities are shown in Figure 11.4. The fixed points or fixed line(s) are assigned signs, which control the projection to the orientifolded theory.



(a) (b)

Figure 11.4: (a) Orientifold of $\mathbb{C}^3/\mathbb{Z}_3$ with fixed points. (b) Orientifold of the conifold with fixed lines.

Fixed Points. In an orientifold of this type, there are four fixed points in a unit cell (see [FHK+07]). In order to preserve SUSY, their four signs must satisfy the so-called *sign rule*: their product must be $(-1)^{n_W/2}$ where n_W is the number of superpotential terms. Faces that sit on top of a fixed point are mapped to themselves and become $SO(N_i)$ or $USp(N_i)$ groups for $+$ and $-$ signs, respectively. Similarly, edges that sit on top of fixed points correspond to self-identified matter fields that are projected to symmetric $\square\square$ (resp. antisymmetric $\begin{smallmatrix}\square\\\square\end{smallmatrix}$) for $+$ (resp. $-$) signs. For faces and edges that are not self-identified, we keep any of the two mirror images in the projected theory. They correspond to $SU(N)$ gauge groups and bifundamental/adjoint matter.

[4]The possibility of $\mathbb{Z}_2$ involutions without fixed loci had not been explored in the literature yet. We will present our findings about this in Chapter 13.

For illustration, let us consider the fixed point orientifold of $\mathbb{C}^3/\mathbb{Z}_3$ shown in Figure 11.4a. The superpotential is the sum of six monomials, so the product of the signs for the four fixed points must be $-$. For concreteness, let us consider the signs $(+,-,-,-)$, where we have ordered them running counter-clockwise starting at the origin. The gauge symmetry and matter content of the orientifolded theory is

$$USp(N_1)_1 \times SU(N_2)_2 \quad \text{with} \quad 2\,\square_2 + \square\square_2 + 3(\square_1, \overline{\square}_2)\ . \tag{11.11}$$

The superpotential results from the projection of the original superpotential (see [FHK+07] for details and examples).

Fixed Lines. Depending on the symmetry of the unit cell, there might be a single diagonal or two parallel fixed lines. We will often refer to the case of two parallel lines as horizontal/vertical lines due to their orientation. In either case, the signs of the fixed lines are unconstrained and can be chosen freely. The rule for projecting faces and fields is the same as with fixed points.

As an example, let us consider the fixed line orientifold of the conifold shown in Figure 11.4b. Faces 1 are 2 are mapped into themselves, so both of them become either symplectic or orthogonal, depending on the signs of the fixed lines. In addition, the matter content consist of two bifundamentals $(\square_1, \square_2)$.[5] As in the previous case, it is also straightforward to determine the projected superpotential.

11.2.2 The Adjacency Matrix of Orientifolded Theories

Consider a toric singularity and a corresponding dimer admitting a $\mathbb{Z}_2$ involution. We can divide the n_g gauge groups of the mother theory into two sets: pairs of faces identified under the involution, and self-identified ones. Therefore, the adjacency matrix of the mother theory, A_{IJ} with $I, J = 1, \ldots, n_g$, can be suitably rearranged as follows:

$$A = \left(\begin{array}{c|c|c} B_{11} & B_{12} & B_{13} \\ \hline B_{21} & B_{22} & B_{23} \\ \hline B_{31} & B_{32} & B_{33} \end{array}\right) \begin{array}{l} \Big\} i \\ \Big\} i+k \\ \Big\} a \end{array} \quad . \tag{11.12}$$

$$\underbrace{\phantom{B_{11}}}_{j} \quad \underbrace{\phantom{B_{12}}}_{j+k} \quad \underbrace{\phantom{B_{13}}}_{b}$$

Here faces $i, j = 1, \ldots, k$ are the surviving ones out of those in the pairs of faces that are mapped into each other (for every pair, we are free to keep any of the two faces). Faces $i+k, j+k$, with $i, j = 1, \ldots, k$, are their images. Finally, the remaining faces $a, b = 1, \ldots, n_g - 2k$ are those that are self-identified. The B matrices are the adjacency matrices between these different subsets. For example, B_{13} is the adjacency matrix between surviving faces and self-identified faces, while B_{23} is the adjacency matrix between the image faces and the self-identified ones. The matrix A is antisymmetric by definition, which in terms of the submatrices B implies that

$$\begin{aligned} &B_{11} = -B_{11}^T\,, \quad B_{22} = -B_{22}^T\,, \quad B_{33} = -B_{33}^T\,, \\ &B_{12} = -B_{21}^T\,, \quad B_{13} = -B_{31}^T\,, \quad B_{23} = -B_{32}^T\,. \end{aligned} \tag{11.13}$$

The $\mathbb{Z}_2$ symmetry of the phase under consideration endows it with further symmetry properties. The

[5] Since the representations of $SO(N)$ and $USp(N)$ are self conjugate, there is no distinction between fundamental and antifundamental representations.

$\mathbb{Z}_2$ projection acts on the bifundamental fields as follows:

$$\begin{array}{ccc} \text{Mother theory} & & \text{Daughter theory} \\ (\overline{\square}_i, \square_j),\ (\overline{\square}_{j+k}, \square_{i+k}) & \to & (\overline{\square}_i, \square_j) \\ (\overline{\square}_i, \square_{j+k}),\ (\overline{\square}_j, \square_{i+k}) & \to & (\overline{\square}_i, \overline{\square}_j) \\ (\overline{\square}_{i+k}, \square_j),\ (\overline{\square}_{j+k}, \square_i) & \to & (\square_i, \square_j) \\ (\overline{\square}_a, \square_i),\ (\overline{\square}_{i+k}, \square_a) & \to & (\square_a, \square_i) \\ (\overline{\square}_i, \square_a),\ (\overline{\square}_a, \square_{i+k}) & \to & (\square_a, \overline{\square}_i) \\ (\overline{\square}_a, \square_b),\ (\overline{\square}_b, \square_a) & \to & (\square_a, \square_b) \end{array} \quad . \tag{11.14}$$

These projections imply that:

$$\begin{aligned} B_{11} = B_{22}^T\,, \quad B_{12} = B_{12}^T\,, \quad B_{21} = B_{21}^T\,, \\ B_{31} = B_{23}^T\,, \quad B_{13} = B_{32}^T, \quad B_{33} = B_{33}^T\,. \end{aligned} \tag{11.15}$$

We can apply Equations (11.13) and (11.15) together to find further relations between the B's,

$$\begin{aligned} B_{11} = -B_{22}\,, \quad B_{12} = -B_{21}\,, \\ B_{13} = -B_{23}\,, \quad B_{31} = -B_{32}\,, \quad B_{33} = 0\,, \end{aligned} \tag{11.16}$$

so that eventually the adjacency matrix is entirely determined by B_{11}, B_{12} and B_{13}:

$$A = \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ -B_{12} & -B_{11} & -B_{13} \\ -B_{13}^T & B_{13}^T & 0 \end{pmatrix}\,. \tag{11.17}$$

In order to illustrate these relations, let us consider the complex cone over PdP_{3b}, as studied in [ABMP19]. The dimer, which is shown in Figure 11.5, admits a $\mathbb{Z}_2$ symmetry with two fixed lines. The

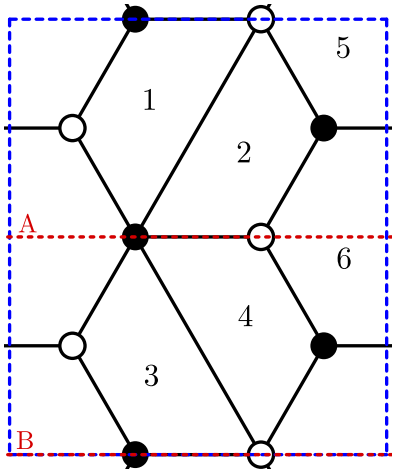


Figure 11.5: Dimer diagram for $PdP3_b$ with two horizontal fixed lines (dotted red).

numbering of the faces has already been chosen such that the adjacency matrix reads

$$A = \left(\begin{array}{cc|cc|cc} 0 & 1 & -1 & 0 & 1 & -1 \\ -1 & 0 & 0 & 1 & 1 & -1 \\ \hline 1 & 0 & 0 & -1 & -1 & 1 \\ 0 & -1 & 1 & 0 & -1 & 1 \\ \hline -1 & -1 & 1 & 1 & 0 & 0 \\ 1 & 1 & -1 & -1 & 0 & 0 \end{array}\right)\,, \tag{11.18}$$

which showcases the general structure in Equations (11.13), (11.15) and (11.16).

Let us now turn our attention to the daughter theory. To compute the ACC for the orientifolded theory we first note that SO/USp groups are automatically anomaly-free and play no role[6]. Further, for the ACC of the non-self-identified faces we have to take into account both the contributions from fields such as $(\overline{\square}_i, \square_j)$, that are counted by B_{11}, and fields such as $(\square_i, \square_j)$ and $(\overline{\square}_i, \overline{\square}_j)$ that are counted by B_{12}, see Equation (11.14). This leads to the homogeneous ACC for the projected theory given by

$$\overline{A} = \left(\begin{array}{c|c} B_{11} + B_{12} & B_{13} \end{array} \right) . \tag{11.19}$$

Applying this to the PdP_{3b} example we get

$$\overline{A} = \left(\begin{array}{cc|cc} -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{array} \right) . \tag{11.20}$$

11.2.3 The Homogeneous Problem

In the previous section, we have constructed the homogeneous part of the ACC for an orientifolded theory. We now show how solutions to the homogeneous problem, namely elements of $\ker(\overline{A})$, are obtained from symmetric rank assignments of the mother theory, which form a subspace of $\ker(A)$. This will allow us to extend the method explained in Section 11.1.1 to the homogeneous problem of orientifolded theories.

We say that a rank assignment of the mother theory N_I^S is symmetric with respect to the $\mathbb{Z}_2$ involution if it satisfies

$$N_i^S = N_{i+k}^S , \quad N_a^S \text{ free} . \tag{11.21}$$

If this rank assignment is anomaly-free in the mother theory (i.e. if it is in the kernel of A), we have

$$A_{IJ} N_J^S = 0 , \tag{11.22}$$

where here and henceforth, summation over repeated indices is understood.

Expanding this equation in terms of the B matrices and exploiting the symmetry properties given in Equation (11.21), it becomes

$$\begin{aligned} (B_{11} + B_{12})_{ij} N_j^S + (B_{13})_{ia} N_a^S &= 0 , \\ (B_{21} + B_{22})_{ij} N_j^S + (B_{23})_{ia} N_a^S &= 0 , \\ (B_{31} + B_{32})_{aj} N_j^S + (B_{33})_{ab} N_b^S &= 0 . \end{aligned} \tag{11.23}$$

From Equation (11.16), we conclude that the first two equations are actually one and the same, while the third equation is trivially satisfied for any symmetric rank assignment. From the first two equations we learn that any symmetric rank assignment N_I^S in the mother theory which satisfies the ACC, defines a solution of the homogenous ACC system of the daughter theory given in Equation (11.9):

$$N_{\text{hom}} = (N_i^S | N_a^S) . \tag{11.24}$$

Equation (11.23) indeed implies that such a vector satisfies:

$$\overline{A} \cdot N_{\text{hom}} = 0 . \tag{11.25}$$

Conversely, if one starts with a vector $(N_i^S | N_a^S)$ satisfying Equation (11.25), the vector $(N_i^S | N_{i+k}^S | N_a^S)$ is a symmetric rank assignment of the mother theory. The definition of $\overline{A}$ in Equation (11.19) implies that the equations in Equation (11.23) hold for $(N_i^S | N_{i+k}^S | N_a^S)$, i.e. that the latter satisfies the ACC of the mother theory. Hence, we have proved the following:

Proposition 11.1. *Rank assignments in the daughter theory which satisfy the homogeneous ACC are in one-to-one correspondence with symmetric rank assignments in the mother theory which satisfy the ACC.*

In the special case where tensors are absent from the daughter theory, the ACC are actually a homogeneous problem and the symmetric rank assignments in the mother theory provide directly the orientifold solutions. The regular brane is such a solution that always exists, and thus guarantees that an orientifold without tensors always admits a non-anomalous solution.

[6]There could have been, however, Witten anomalies à la SU(2) in SO/USp groups. Yet, as discussed in [IKY08], cancellation of local anomalies ensures that the number of fermions is even and the Witten anomaly vanishes.

11.2.4 The Non-Homogeneous Problem

Finding solutions to the ACC in orientifolded theories with tensors is not trivial because their very existence is not guaranteed, since the full system of ACC given in Equation (11.9) has a non-homogeneous part coming from the tensor matter. The Rouché-Capelli theorem gives us a criterion for its solvability: a non-homogeneous system,

$$\bar{A} \cdot N = f \,, \tag{11.26}$$

admits a solution if and only if

$$\text{rank}(\bar{A}) = \text{rank}(\bar{A}|f) \,, \tag{11.27}$$

where $(\bar{A}|f)$ is the matrix obtained by appending the column f to the matrix $\bar{A}$. For us f encodes the contribution to the ACC of the tensor matter, i.e. of the self-identified chiral fields.

In other words, every set of numbers r_i such that

$$r_i \bar{A}_{i\bar{J}} = 0 \tag{11.28}$$

holds for all $\bar{J} = j, a$, must satisfy

$$r_i f_i = 0 \tag{11.29}$$

for the system to be solvable. In this section we show that the coefficients r_i which satisfy Equation (11.28) correspond precisely to the antisymmetric rank assignments of the mother theory.

Suppose that some coefficients r_i satisfying Equation (11.28) exist. Using Equation (11.16) for $\bar{J} = j$, one can show that it implies

$$r_i (B_{11})_{ij} - r_i (B_{21})_{ij} = 0 \,. \tag{11.30}$$

Using Equation (11.16), this is equivalent to

$$r_i (B_{12})_{ij} - r_i (B_{22})_{ij} = 0 \,. \tag{11.31}$$

For $\bar{J} = a$, using Equation (11.16), we find that

$$r_i (B_{13})_{ia} - r_i (B_{23})_{ia} = 0 \,. \tag{11.32}$$

We write

$$N_I^A = (r_i| - r_i|0) \,, \tag{11.33}$$

and equations Equation (11.30) to Equation (11.32) can be expressed as

$$N_I^A A_{IJ} = 0 = A_{JI} N_I^A \,, \tag{11.34}$$

where the second equality merely uses the antisymmetry property of A. Hence, we have proved that any set of r_i satisfying Equation (11.28) defines an antisymmetric rank assignment N_I^A of the mother theory, which satisfies the mother ACC.

Conversely, starting with an antisymmetric rank assignment N_I^A in the mother theory

$$N_i^A = -N_{i+k}^A, \quad N_a^A = 0 \,, \tag{11.35}$$

which satisfies the ACC, one can use equations Equation (11.30) to Equation (11.32) backwards, and thus obtain a set of r_i such that Equation (11.28) holds for all $\bar{J} = j, a$.

Let us emphasize that while symmetric rank assignments in the mother theory are in one-to-one correspondence with solutions of the homogeneous system of ACC in the daughter theory (which by definition form the kernel of $\overline{A}$), the antisymmetric rank assignments in the mother theory correspond rather to the elements of the cokernel of $\overline{A}$, that we will see merely as technical tools. They are useful for determining whether a given daughter theory admits an anomaly-free rank assignment, since the elements in the cokernel of $\overline{A}$ encode the relations between the lines of $\overline{A}$, from which one can row-reduce $\overline{A}$.

Proposition 11.2. *Coefficients of trivial linear combination of lines of $\overline{A}$ are in one-to-one correspondence with the anomaly-free antisymmetric rank assignments in the mother theory.*

To rephrase what we wrote at the beginning of the section, there are anomaly-free rank assignments in the daughter theory if and only if

$$N_i^A f_i = 0 \tag{11.36}$$

for every antisymmetric solution N_I^A of the mother theory, where f is easily computed from the dimer and its orientifold. We call this the "Rouché-Capelli condition."

In general, note that any rank assignment N_I can be split into a symmetric and an antisymmetric component,

$$(N_i|N_{i+k}|N_a) = \frac{1}{2}(N_i + N_{i+k}|N_{i+k} + N_i|2N_a) + \frac{1}{2}(N_i - N_{i+k}|N_{i+k} - N_i|0)\,. \tag{11.37}$$

Both parts are then half-integer valued. Multiplying such a possibly unphysical (in the case it is half integer-valued) rank vector by an even number yields a physical rank vector with the required (anti)symmetry. All of the reasoning of the last two subsections is pure linear algebra, and does not know about the need of integrality for rank assignments, which comes entirely from physics.

11.3 A Zig-Zag Algorithm for Orientifolds

We will now generalize the procedure discussed in Section 11.1.1 to find (anti)symmetric rank assignments in orientifolded theories. The precise details of the algorithm depend on whether the $\mathbb{Z}_2$ involution leaves fixed lines or points. This difference comes from the fact that involutions with fixed lines map nodes to nodes of the same color, while involutions with fixed points map nodes to nodes with opposite color.

We illustrate this difference in Figure 11.6. There we can see that ZZPs around a node make a clockwise or counterclockwise loop. If a node is mapped to a node of the same color it means that the orientation of the loop is preserved, while, in the opposite case, it is reversed.[7] This observation will become crucial when we define (anti)symmetric rank assignment in both the case of fixed lines and points.

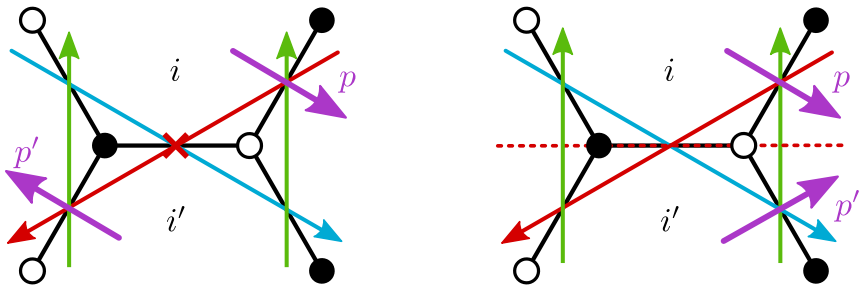


Figure 11.6: The orientifold actions with fixed points (a) and fixed lines (b). p is a path from one face to an adjacent one, and p' its image. In (a) the red and blue ZZPs are self-identified, while the green ones are mapped into each other. In (b), the red and blue ZZPs are mapped into each other, and the green ones are self-identified.

For the forthcoming analysis, we find it useful to introduce the notation $\{\Gamma\} = \{\alpha, \overline{\alpha}, \Gamma\}$ to describe the set of ZZPs: every pair $(\alpha,\ \overline{\alpha})$ corresponds to ZZPs mapped into each other under the orientifold projection, while Γ labels self-identified ZZPs.

11.3.1 Fixed Line Orientifolds

Due to how they act on ZZPs, orientifolds with fixed lines in the dimer correspond to toric diagrams with axes of $\mathbb{Z}_2$ reflection symmetry.[8] Figure 11.7 illustrates the general structure of such axes and

[7] We recall that under both involutions, a dimer is sent to a dimer with all ZZPs going in the opposite direction. The map between ZZPs is understood after additionally reversing the direction of every ZZP, as in [RU16b].

[8] We will refer to such lines of reflection symmetry in the toric diagram as *axes* in order to avoid confusion with the fixed lines in the dimer (which we also call O-lines).

the map between a ZZP and its image in the cases of orientifolds with diagonal and horizontal O-lines (which is analogous to the case with vertical O-lines). Let us elaborate on this kind of figure. Naively, the orientation of the reflection axis in these toric diagrams can be modified by an $SL(2,\mathbb{Z})$ transformation, potentially eliminating the distinction between the diagonal and vertical/horizontal O-line cases. However, the toric diagram after such $SL(2,\mathbb{Z})$ transformation would no longer be symmetric with respect to the axis. Alternatively, we can think about the toric diagrams with reflection axes as coming from specific dimers with fixed lines. In this context, an $SL(2,\mathbb{Z})$ transformation translates into a change of the unit cell of the dimer. But the unit cell is fixed by the specific orientifold under consideration: not any $SL(2,\mathbb{Z})$ transformation is permitted once we have chosen an orientifold identification. In other words, the orientifold obstructs $SL(2,\mathbb{Z})$ transformations.

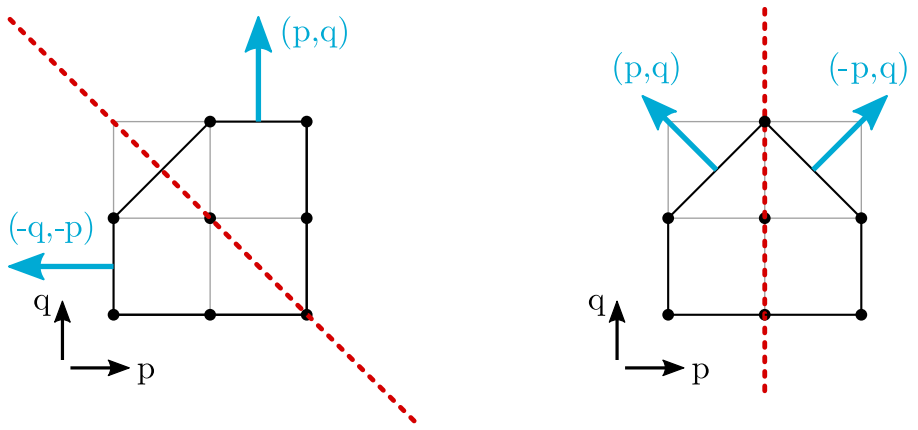


Figure 11.7: The toric diagrams for fixed line orientifolds have an axis of reflection symmetry. The corresponding axes for: (a) diagonal and (b) horizontal O-lines. In both cases we show in blue a generic ZZP and its image.

Symmetric rank assignments. For $\mathbb{Z}_2$ involutions with fixed lines, symmetric rank assignments correspond to symmetric ZZP value assignments:

$$v_\alpha = v_{\overline{\alpha}}, \quad v_\Gamma \text{ free} \tag{11.38}$$

First, recall from Section 11.1.1 that the difference between the ranks of any two faces in the dimer is equal to a sum (with signs) of the values of the ZZPs one crosses as one goes between the two faces. ACC at each face of the dimer ensure that the value of this sum is invariant under smooth (homological) deformations of the path one follows. Furthermore, the topological constraints guarantee that the value of the sum is independent of the homology class of the path on the torus.

Consider two faces i and j and a path p connecting them, and i', j' and p' their respective images under the $\mathbb{Z}_2$ symmetry. Every time p crosses a ZZP α, its image p' crosses α', and these two crossings have the same sign, since the orientation is preserved. From this, it is clear that, if the ZZP value assignment is symmetric, the rank assignment generated by the method in Section 11.1.1 is also symmetric.

Proposition 11.3. *In the case of dimer models with involutions fixing lines, symmetric rank assignments correspond bijectively to symmetric ZZP value assignments (up to the global shift in the values, and such that the topological constraints are satisfied).*

For symmetric value assignments, the topological constraints read:

- Diagonal line ($p_{\bar{\alpha}} = -q_\alpha, q_{\bar{\alpha}} = -p_\alpha$):

$$0 = \Lambda = \sum_\alpha v_\alpha (p_\alpha - q_\alpha) + \frac{1}{2}\sum_\gamma v_\gamma (p_\gamma - q_\gamma) = -M = 0\,. \tag{11.39}$$

- Vertical lines ($p_{\bar{\alpha}} = -p_\alpha, q_{\bar{\alpha}} = q_\alpha$):

$$M = 2\sum_\alpha v_\alpha q_\alpha + \sum_\gamma v_\gamma q_\gamma = 0\,,$$
$$\Lambda = 0\,, \tag{11.40}$$

The case of horizontal lines follows exchanging p_Γ with q_Γ and Λ with M.

We can now compute the total number of symmetric rank assignments. If the dimer under consideration has n ZZPs, symmetric rank assignments correspond to a choice of v_Γ, such that $v_\alpha = v_{\overline{\alpha}}$, and such that topological constraints hold. We also have the freedom to shift the rank of all gauge groups, since regular branes respect the required symmetry. Putting all this together, the number of independent symmetric rank assignments modulo some (possibly half-integer) number of regular branes is

$$\dim(\ker(\overline{A})) - 1 = \frac{1}{2}(n + n_s) - 2\,, \tag{11.41}$$

where n_s is the number of self-identified ZZPs.

Antisymmetric rank assignments. Antisymmetric rank assignments are found in a similar fashion, by imposing the antisymmetry explicitly on the ZZP values, i.e. $v_\Gamma = -v_\Gamma$, or equivalently

$$v_\alpha = -v_{\overline{\alpha}}, \quad v_\Gamma = 0\,. \tag{11.42}$$

This follows from the same reasoning as in the symmetric case: due to the geometric action of the symmetry, it is clear that antisymmetric ZZP value assignments lead to antisymmetric rank assignments in the dimer. Furthermore, if the ZZP value assignment is not antisymmetric *up to a shift*, it is straightforward to see that the rank assignment cannot be antisymmetric either.

In this case there is a subtlety that was not present in the symmetric case. First, the ZZP value method only knows about differences of ranks in the dimer. Equivalently, it only describes anomaly-free rank assignments up to some (half-integer) number of regular branes. The relevant point here is that regular branes are not antisymmetric. Hence, starting from an antisymmetric value assignment, it can well be that the rank assignment one constructs is not antisymmetric per se, but merely antisymmetric after having added some number of regular branes (we will see examples of this later). Then, in the method of Section 11.1.1, a global shift of the ZZP values does not change the resulting rank assignment. The global shift does not preserve antisymmetry, so among the family of value assignments corresponding to a given rank assignment (modulo regular branes), there is a special representative which is an antisymmetric value assignment. Thus instead of focusing on the bijection between the set of antisymmetric rank assignments up to a (half-integer) number of regular branes, and the set of ZZP value assignments which satisfies the topological constraints, and which can be transformed into antisymmetric value assignments thanks to the global shift, one can consider the only representative of such a class of ZZP value assignments, which is antisymmetric. We have proven the following:

Proposition 11.4. *In the case of dimer models with involutions fixing lines, antisymmetric rank assignments correspond bijectively to antisymmetric ZZP value assignments which satisfy the topological constraints.*

When combined with Equation (11.42), the topological constraints $\Lambda = M = 0$ again merge into a single constraint, regardless of the type of fixed line orientifold. The surviving combination however depends on the nature of the fixed lines:

- Diagonal line:

$$\Lambda = \sum_\alpha v_\alpha(p_\alpha + q_\alpha) = -M = 0\,. \tag{11.43}$$

- Vertical lines:

$$\Lambda = 2\sum_\alpha v_\alpha p_\alpha = 0\,,$$
$$M = 0\,. \tag{11.44}$$

For horizontal lines we merely need to exchange p_α with q_α, and Λ with M.

The number of antisymmetric rank assignments is easily computed to be

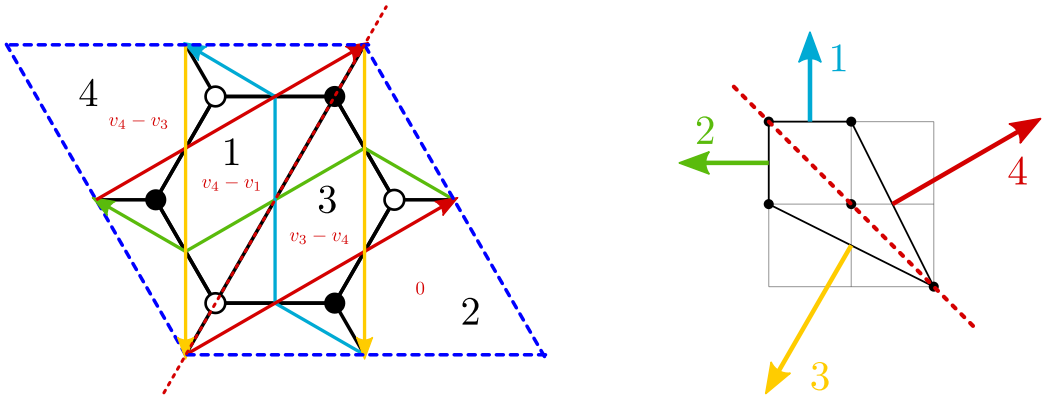
$$\dim(\mathrm{coker}(\overline{A})) = \frac{1}{2}(n - n_s) - 1 \tag{11.45}$$

Adding Equation (11.41) and Equation (11.45), we find that the total number of independent either symmetric or antisymmetric anomaly-free rank assignments is $n-3$, as should be the case since it is the number of anomaly-free rank assignments in the mother theory, up to (half) regular branes.

Below we illustrate these ideas with a few explicit examples, containing both diagonal and vertical/horizontal fixed lines.

No Anomaly-Free Solution: dP_1 with Diagonal Fixed Line

Consider the complex cone over dP_1, which we discussed in Section 11.1.2 as an example of the ZZP method for the mother theory. It admits a $\mathbb{Z}_2$ involution with a fixed line, as shown in Figure 11.8.



(a) (b)

Figure 11.8: (a) Dimer diagram for dP_1 with a diagonal fixed line (dotted red). We show the ZZPs and the rank assignments coming from them. (b) The toric/web diagram with the corresponding symmetry axis.

Adjacency matrices. The adjacency matrix of the mother theory is easily read from the dimer, and it is given by

$$A = \left(\begin{array}{cc|cc} 0 & 2 & -1 & -1 \\ -2 & 0 & -1 & 3 \\ \hline 1 & 1 & 0 & -2 \\ 1 & -3 & 2 & 0 \end{array}\right) \tag{11.46}$$

For concreteness, let us consider the case of a positive O-line. The adjacency matrix for the orientifolded theory is found using Equation (11.19). It is supplemented with the inhomogeneous part and becomes

$$(\bar{A}|f) = \left(\begin{array}{cc|c} -1 & 1 & +4 \\ -3 & 3 & -4 \end{array}\right) . \tag{11.47}$$

We will later discuss how to determine systematically the f_i's. Here it is sufficient to see that since the O-line has a + sign, both tensors are symmetric. The sign of $f_i = \pm 4$ has to be correlated with the sign of the diagonal elements of $\bar{A}$, so that in the ACC we eventually find $\pm(N_i + 4)$ for symmetric tensors and $\pm(N_i - 4)$ for antisymmetric ones (recall that f is on the right hand side of the ACC equations Equation (11.9)).

One may directly solve the simple system Equation (11.47), but we will rather use the algorithm we developed. In the dimer in Figure 11.8a we indicate the linear combination of ZZPs that corresponds to every face (we have chosen face 2 to have rank 0). In Section 11.1.2 we studied the anomaly-free rank assignments in the mother theory and found a one parameter family (besides the regular brane):

$$\mathbf{N} = (-2, 0, 1, -1)v_3 \ , \tag{11.48}$$

which can be decomposed into symmetric and anti-symmetric parts,

$$(-2,0,1,-1)v_3 = -\frac{1}{2}(1,1,1,1)v_3 + \frac{1}{2}(-3,1,3,-1)v_3 \tag{11.49}$$

We see that there is one antisymmetric rank vector $(-3,1,3,-1)$ and no symmetric one (except the regular brane). We now show how to find them directly from the ZZPs.

The antisymmetric rank vector is found by imposing $v_1 = -v_2$ and $v_3 = -v_4$. We cannot use the global shift, since it is not antisymmetric. The periodicity constraints are $\Lambda = M = v_1 - 3v_3 = 0$. We thus find the antisymmetric rank assignment $\mathbf{N} = (-3,1,3,-1)v_3$. In the daughter theory this vector is $\overline{\mathbf{N}} = (-3,1)v_3$. However, it is not in $\ker(\overline{A})$, but in the cokernel. We can use it to row reduce $\overline{A}$ and study whether the linear system $(\overline{A}|f)$ has solutions. Denote $\mathbf{f} = (+4,-4)^T$ the inhomogeneous part of $(\overline{A}|f)$. If $\overline{\mathbf{N}}\cdot\mathbf{f} \neq 0$, the theory is anomalous. This is indeed the case in this example, so we conclude that the daughter theory does not admit an anomaly-free solution.

No Anomaly-Free Solution: PdP_4 with Diagonal Fixed Line

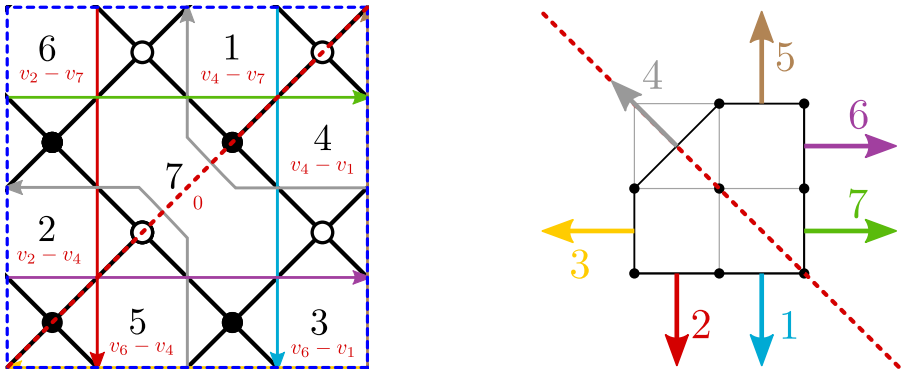
Consider now PdP_4, which we previously discussed in Section 11.1.2. Figure 11.9 shows the dimer and toric diagram for the orientifold under consideration. In Section 11.1.2 we saw that anomaly-free rank assignments of the mother theory are given by:

$$\mathbf{N} = (-v_7, v_2, v_6 - v_1, -v_1, v_6, v_2 - v_7, 0)\,. \tag{11.50}$$

The topological constraints are:

$$\Lambda : v_4 + v_3 = v_6 + v_7\ , \tag{11.51}$$

$$M : v_4 + v_5 = v_1 + v_2\,. \tag{11.52}$$



(a) (b)

Figure 11.9: (a) Dimer diagram for PdP_4 with a diagonal fixed line (dotted red). We show the ZZPs and the rank assignments coming from them. (b) The toric/web diagram with the corresponding symmetry axis.

Adjacency matrices. The adjacency matrices of the mother and daughter theories are:

$$A = \left(\begin{array}{ccc|ccc|c} 0 & 0 & 1 & -1 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 & -1 & 0 & 1 \\ -1 & -1 & 0 & 0 & 0 & 1 & 1 \\ \hline 1 & 1 & 0 & 0 & 0 & -1 & -1 \\ 1 & 1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & 1 & 1 & 0 & -1 \\ \hline -1 & -1 & -1 & 1 & 1 & 1 & 0 \end{array}\right), \qquad (\bar{A}|f) = \left(\begin{array}{ccc|c|c} -1 & -1 & 1 & 1 & -4 \\ -1 & -1 & 1 & 1 & -4 \\ -1 & -1 & 1 & 1 & +4 \end{array}\right), \tag{11.53}$$

where, for concreteness, we have assumed that the sign of the orientifold line is negative.

Symmetric rank assignments. Impose $v_3 = v_5$, $v_2 = v_6$, $v_1 = v_7$. The constraints $M = 0$, $\Lambda = 0$ combine into $v_4 = v_1 + v_2 - v_3$. We can use the global shift freedom to set $v_4 = 0$, which leads to $\mathbf{v^S} = (v_1, v_2, v_1 + v_2, 0, v_1 + v_2, v_2, v_1)$. The resulting symmetric rank assignments in the mother and daughter theories are

$$\begin{aligned} \mathbf{N^S} &= (-v_1, v_2, v_2 - v_1, -v_1, v_2, v_2 - v_1, 0) \\ \overline{\mathbf{N}}^{\mathbf{S}} &= (-v_1, v_2, v_2 - v_1, 0) \ . \end{aligned} \tag{11.54}$$

Note that $\overline{\mathbf{N}}^{\mathbf{S}}$ should be understood as the column vector whose first three elements refer to the faces 1–3 that have an image, while the last refers to the self-identified face 7. When considered as a row vector, one should drop the last element.

Antisymmetric rank assignments. Impose $v_1 = -v_7$, $v_2 = -v_6$, $v_3 = -v_5 = 0$, $v_4 = 0$. We also need to impose the constraint $v_1 + v_2 = -v_3$ with no global shift freedom. We then find a two-parameter family of antisymmetric assignments for the v_Γ, $\mathbf{v^A} = (v_1, v_2, -v_1 - v_2, 0, v_1 + v_2, -v_2, -v_1)$. The corresponding antisymmetric rank assignment is

$$\mathbf{N^A} = (v_1, v_2, -v_1 - v_2, -v_1, -v_2, v_1 + v_2, 0) \ . \tag{11.55}$$

In the daughter theory, this rank assignment gives rise to the two row vectors

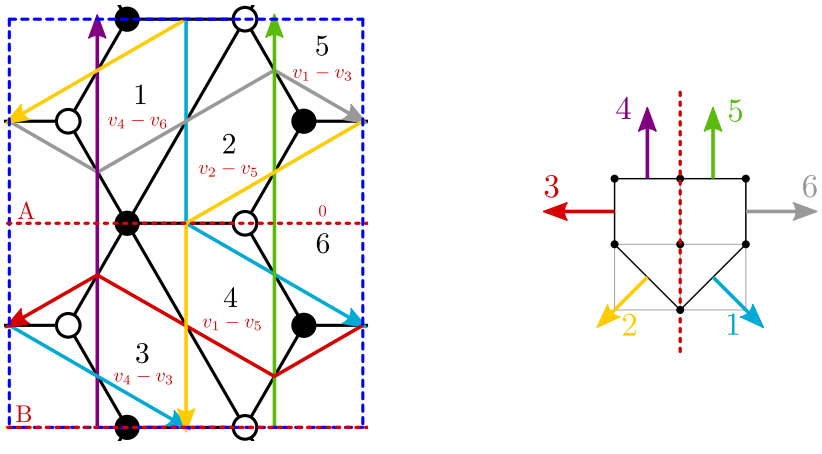
$$\overline{\mathbf{N}}^{\mathbf{A}}_{\mathbf{1}} = (1, 0, -1)v_1, \quad \overline{\mathbf{N}}^{\mathbf{A}}_{\mathbf{2}} = (0, 1, -1)v_2 \tag{11.56}$$

Let us denote by $\mathbf{f} = (-4, -4, 4)^T$ the inhomogeneous part of $(\overline{A}|f)$. We find $\overline{\mathbf{N}}^{\mathbf{A}}_{\mathbf{1}} \cdot \mathbf{f} = -8$ and $\overline{\mathbf{N}}^{\mathbf{A}}_{\mathbf{2}} \cdot \mathbf{f} = -8$. We conclude that anomalies cannot be cancelled in this theory.

This example and the previous one consist of orientifolds with a diagonal fixed line. Both cases turned out to lead to theories in which anomalies cannot be cancelled. In Section 11.4.1 we will present a more detailed general analysis and discuss under which conditions such orientifolds can admit anomaly-free solutions.

An Anomaly-Free Example: PdP_{3b} with Two Fixed Lines

Figure 11.10 shows the dimer and toric diagram for an orientifold of PdP_{3b} with two fixed lines. This theory was studied in [ABMP19], where it was shown that the daughter theory admits an anomaly-free rank assignment if the two O-lines have opposite signs. Note that the horizontal fixed lines in the dimer correspond to a vertical axis of symmetry in the toric diagram.



(a) (b)

Figure 11.10: (a) Dimer diagram for PdP_{3b} with two horizontal fixed lines (dotted red). We show the ZZPs and the rank assignments coming from them. (b) The toric/web diagram with the corresponding symmetry axis.

Adjacency matrices. The adjacency matrices of the mother and daughter theories are:

$$A = \left(\begin{array}{cc|cc|cc} 0 & -1 & 1 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 & -1 & 1 \\ \hline -1 & 0 & 0 & 1 & 1 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 \\ \hline 1 & 1 & -1 & -1 & 0 & 0 \\ -1 & -1 & 1 & 1 & 0 & 0 \end{array}\right), \qquad (\bar{A}|f) = \left(\begin{array}{cc|cc|c} 1 & -1 & -1 & 1 & -4\cdot\mathrm{sign}(B) \\ 1 & -1 & -1 & 1 & +4\cdot\mathrm{sign}(A) \end{array}\right), \quad (11.57)$$

where $\mathrm{sign}(A)$, $\mathrm{sign}(B)$ are the signs of the two O-lines. Let us now turn to the study of symmetric and antisymmetric rank assignments.

Symmetric rank assignments. Let us impose $v_2 = v_1$, $v_6 = v_3$. The constraint $\Lambda = 0$ is trivially satisfied, while $M = 0$ becomes (keeping v_1 and v_3):

$$2v_1 - v_4 = v_5\,. \tag{11.58}$$

Setting $v_4 = 0$, we get

$$\mathbf{v^S} = (v_1, v_1, v_3, 0, 2v_1, v_3)\,, \tag{11.59}$$

giving in turn

$$\mathbf{N^S} = (-v_3, -v_1, -v_3, -v_1, v_1 - v_3, 0)\,. \tag{11.60}$$

Projecting down this vector, we obtain the solutions to the homogeneous problem in the daughter theory.

Antisymmetric rank assignments. We now impose $v_2 = -v_1$, $v_3 = -v_6$, $v_4 = v_5 = 0$. As expected, M is trivially satisfied and one just needs to impose $\Lambda = 0$, which reads $v_3 = v_1$. Remember that the global shift has already been fixed. We then find a one-dimensional family of antisymmetric assignments for the v_Γ:

$$\mathbf{v^A} = (v_1, -v_1, v_1, 0, 0, -v_1)\,. \tag{11.61}$$

The corresponding antisymmetric rank assignment is $\mathbf{N^A} = (v_1, -v_1, -v_1, v_1, 0, 0)$. In the daughter theory this rank assignment gives $\overline{\mathbf{N}}^{\mathbf{A}} = (1, -1)v_1$. One may now use it to row reduce $\overline{A}$. Denote by $\mathbf{f} = (-4\cdot\mathrm{sign}(B), +4\cdot\mathrm{sign}(A))^T$ the inhomogeneous part of $(\overline{A}|f)$. We find $\overline{\mathbf{N}}^{\mathbf{A}}\cdot\mathbf{f} = -4\cdot\mathrm{sign}(B) - 4\cdot\mathrm{sign}(A)$. If $\overline{\mathbf{N}}^{\mathbf{A}}\cdot\mathbf{f} \neq 0$, the theory is anomalous, so we need $\mathrm{sign}(A) = -\mathrm{sign}(B)$, as anticipated.

Anomaly-free rank assignments. As explained in the introduction of the current section, since we have a parametrization of the symmetric rank assignments, we merely need a single solution of the tadpole-cancellation system to write all of them.

Looking at the adjacency matrix of the daughter theory in Equation (11.57) with $\mathrm{sign}(A) = +$ and $\mathrm{sign}(B) = -$, a straightforward solution to the rank assignment is $N_1 = 4$ and $N_2 = N_5 = N_6 = 0$ (in the daughter theory we keep faces $1, 2, 5$ and 6). This gives the following three-parameter family of solutions to the ACC, where we have added $N + v_1 + v_3$ regular branes:

$$\left\{\begin{array}{lcl} N_1 &=& N + v_1 + 4 \\ N_2 &=& N + v_3 \\ N_5 &=& N + 2v_1 \\ N_6 &=& N + v_1 + v_3\ . \end{array}\right. \tag{11.62}$$

11.3.2 Fixed Point Orientifolds

In orientifolds with fixed points, every ZZP is mapped to a ZZP with the same winding numbers [RU16b]. The image of a ZZP can therefore be either itself or another ZZP, if more than one ZZP with the same winding numbers exist.

Contrarily to the cases with fixed lines, in fixed point orientifolds nodes in the dimer are mapped to nodes of the opposite color. In analogy with the case of line orientifolds, let us consider a path p going from a face i to a face j, and its image p' going from the image of i to the image of j. If p crosses a ZZP α, then p' crosses its image α', but since the color of the nodes is inverted in the image, the signs of the crossings are opposite. This implies that a symmetric, respectively antisymmetric, rank assignment

is associated to an antisymmetric, respectively symmetric, value assignment for the ZZP. We therefore have:

Proposition 11.5. *In dimer models with fixed point involutions, symmetric rank assignments up to (half)-regular branes correspond bijectively to antisymmetric ZZP value assignments which satisfy the topological constraints. Similarly, antisymmetric rank assignments correspond bijectively to symmetric ZZP value assignments which satisfy the topological constraints and up to a global shift.*

We have seen that in the cases of fixed point orientifolds, symmetric rank assignments correspond to ZZP value assignments such that:

$$v_\alpha = -v_{\overline{\alpha}}, \quad v_\Gamma = 0\,. \tag{11.63}$$

One can easily verify that the topological constraints are always satisfied by this choice of v_Γ, hence the number of symmetric rank assignment is:

$$\dim(\ker(\overline{A})) = \frac{1}{2}(n - n_s)\,. \tag{11.64}$$

Antisymmetric rank assignments, conversely, correspond to:

$$v_\alpha = v_{\overline{\alpha}}, \quad v_\Gamma = \text{free}\,. \tag{11.65}$$

In this case both topological constraints Λ and M are not trivial:

$$\begin{cases} \Lambda = \sum_\alpha p_\alpha v_\alpha + \sum_{\overline{\alpha}} p_{\overline{\alpha}} v_{\overline{\alpha}} + \sum_\gamma p_\Gamma v_\gamma = 2\sum_\alpha p_\alpha v_\alpha + \sum_\Gamma p_\Gamma v_\Gamma = 0 \\ M = \sum_\alpha q_\alpha v_\alpha + \sum_{\overline{\alpha}} q_{\overline{\alpha}} v_{\overline{\alpha}} + \sum_\gamma q_\Gamma v_\gamma = 2\sum_\alpha q_\alpha v_\alpha + \sum_\gamma q_\Gamma v_\gamma = 0 \end{cases} \tag{11.66}$$

This leads to:

$$\dim(\text{coker}(\overline{A})) = \frac{1}{2}(n + n_s) - 3\,. \tag{11.67}$$

Upon summing the contributions of symmetric and antisymmetric rank assignments, we retrieve the total number of fractional branes, $n - 3$, modulo regular branes.

An Example: PdP_{3b}

Let us return to PdP_{3b}, already studied in Section 11.3.1 but now with fixed points instead of lines. The dimer is shown in Figure 11.11. Note that we have changed the unit cell and face numbering with respect to Figure 11.10 to make it consistent with fixed point reflections.

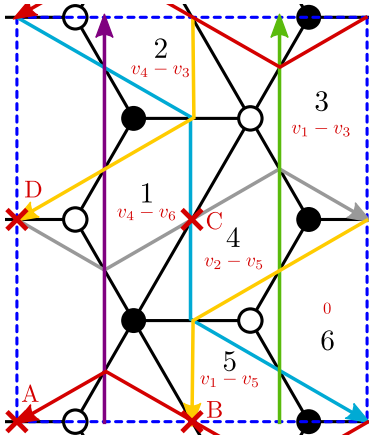


Figure 11.11: Dimer diagram for PdP_{3b} with fixed points. We show the ZZPs and the rank assignments coming from them.

Adjacency matrices. The adjacency matrices of the mother and daughter theories are:

$$A = \left(\begin{array}{ccc|ccc} 0 & 1 & -1 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 & 1 & -1 \\ 1 & -1 & 0 & 1 & -1 & 0 \\ \hline 1 & 0 & -1 & 0 & -1 & 1 \\ 0 & -1 & 1 & 1 & 0 & -1 \\ -1 & 1 & 0 & -1 & 1 & 0 \end{array}\right), \qquad (\bar{A}|f) = \left(\begin{array}{ccc|c} -1 & 1 & 0 & +4\cdot\text{sign}(C) \\ -1 & 1 & 0 & -4\cdot\text{sign}(B) \\ 2 & -2 & 0 & -4\cdot\text{sign}(A)+4\cdot\text{sign}(D) \end{array}\right), \tag{11.68}$$

where sign(A) to sign(D) are the signs of the O-points. Note that the ZZPs 4 and 5 are interchanged by the projection, while all other ZZPs are mapped to themselves. Let us now turn to the study of antisymmetric and symmetric rank assignments.

Symmetric rank assignments. This time we start with antisymmetric ZZP assignments, since for point orientifolds they provide symmetric rank assignments. Let us impose $v_4 = -v_5$, $v_1 = v_2 = v_3 = v_6 = 0$. As already said, the topological constraints are both trivially satisfied. Note that there is no global shift to fix. We obtain a one-parameter family of v_Γ assignments:

$$(0,0,0,1,-1,0)v_4\,. \tag{11.69}$$

The corresponding rank assignment is:

$$\mathbf{N^S} = (1,1,0,1,1,0)v_4\,, \tag{11.70}$$

which is symmetric, as expected. Projecting down this vector, one obtains the solutions to the homogeneous problem in the daughter theory.

Antisymmetric rank assignments. We now turn to symmetric ZZP assignments, responsible for the antisymmetric rank assignments. We only need to impose $v_4 = v_5$. We further fix the global shift by choosing $v_4 = 0$. The topological constraints become:

$$\begin{aligned} \Lambda : v_3 &= v_1 - v_2 + v_6\ , \\ M : v_2 &= -v_1\ . \end{aligned} \tag{11.71}$$

We find a two-dimensional family of symmetric assignments for the v_Γ:

$$(1,-1,2,0,0,0)v_1 + (0,0,1,0,0,1)v_6\,. \tag{11.72}$$

The corresponding antisymmetric rank assignments are:

$$\mathbf{N}^A = (0,-2,-1,-1,1,0)v_1 + (-1,-1,-1,0,0,0)v_6\ . \tag{11.73}$$

Which, up to half regular branes is equal to:

$$\mathbf{N}^A = (1,-3,-1,-1,3,1)\frac{v_1}{2} + (-1,-1,-1,1,1,1)\frac{v_6}{2}\,, \tag{11.74}$$

which is antisymmetric, as expected. Let us split it into two vectors and project them down to the daughter theory to obtain,

$$\overline{\mathbf{N}}_1^A = (1,-3,-1)\frac{v_1}{2}, \quad \overline{\mathbf{N}}_2^A = (-1,-1,-1)\frac{v_6}{2}\ . \tag{11.75}$$

Again, let us use these rank assignments to row reduce $\overline{A}$ by denoting $\mathbf{f} = (+4\cdot\text{sign}(C), -4\cdot\text{sign}(B), -4\cdot\text{sign}(A)+4\cdot\text{sign}(D))^T$. One finds that, for the theory to admit non-anomalous solutions, one must satisfy,

$$\begin{aligned} \overline{\mathbf{N}}_1^A\cdot\mathbf{f} &= \frac{v_1}{2}\left(\text{sign}(C)+3\text{sign}(B)+\text{sign}(A)-\text{sign}(D)\right) = 0\ , \\ \overline{\mathbf{N}}_2^A\cdot\mathbf{f} &= \frac{v_6}{2}\left(-\text{sign}(C)+\text{sign}(B)+\text{sign}(A)-\text{sign}(D)\right) = 0\ . \end{aligned} \tag{11.76}$$

Anomaly-free rank assignments. The solution to Equation (11.76) depends on the sign choices for the four fixed points. Consider for example

$$\text{sign}(A) = \text{sign}(C) = +, \quad \text{sign}(B) = \text{sign}(D) = -\,, \tag{11.77}$$

which is consistent with the sign rule for fixed point orientifolds. In this case, we go back to Equation (11.68) to find a two-parameter family of solutions:

$$\begin{cases} N_1 &= N + v_4 \\ N_2 &= N + v_4 + 4 \\ N_3 &= N\ . \end{cases} \tag{11.78}$$

11.4 General Criteria for Anomaly-Free Orientifolds

In this section we present a general study of the solutions to the non-homogeneous system of ACC of the daughter theory. Remarkably, we can exploit the algorithm of the previous section to determine the existence of such solutions directly from toric data, regardless of the particular phase of the theory. This gives a purely geometric criterion determining whether an orientifolded theory may admit a toric phase with non-anomalous rank assignments.

11.4.1 Diagonal Line Orientifolds

Let us consider orientifolds with a diagonal fixed line. Without loss of generality, we assume that the fixed line has winding numbers $(1,1)$ in the fundamental cell of the dimer. The mapping of ZZPs in this kind of orientifolds has been studied in [RU16b] and we presented a preliminary discussion in Section 11.3.1. The diagonal fixed line in the dimer translates into a reflection symmetry axis in the toric diagram with slope -1, as we already illustrated in Figure 11.7a. This $90°$ rotation of the symmetry axis of the toric diagram with respect to the fixed line in the dimer was explained in [Yam08].

Reflection with respect to the axis of the toric diagram maps a ZZP with winding (p,q), to a ZZP with winding $(-q,-p)$. Figure 11.12 shows an example of a generic toric diagram with a diagonal line orientifold.

- Let l be the number of pairs $\{v_\alpha, v_{\overline{\alpha}}\}$,with $\alpha = 1,..,l$, of ZZPs mapped one to another, which are not parallel to the symmetry axis of the toric diagram.
- Let $l_\parallel$ be the number of self-identified ZZPs $\{v_\gamma\}$ for $\gamma = 1,...,l_\parallel$, which are parallel to the symmetry axis of the toric diagram.

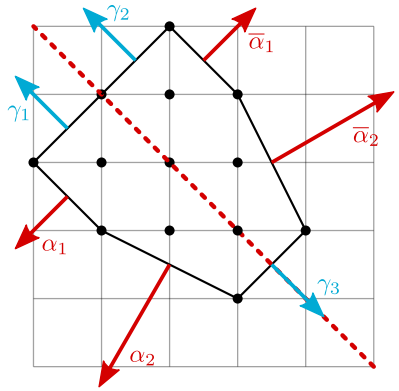


Figure 11.12: A generic toric diagram with a diagonal axis symmetry.

From the previous section, we know how to produce the coefficients of the trivial linear combinations of rows. They are the ranks of the projected SU groups that result from imposing the following conditions on the v_Γ:

$$\begin{aligned} v_\alpha &= -v_{\overline{\alpha}}\,, \\ v_\gamma &= 0 \end{aligned} \tag{11.79}$$

for all α and γ's. The topological constraints Λ and M are given by:

$$\Lambda = \sum_{\alpha,\overline{\alpha}} (v_\alpha p_\alpha + v_{\overline{\alpha}} p_{\overline{\alpha}}) = \sum_\alpha v_\alpha (p_\alpha + q_\alpha) = -M \tag{11.80}$$

where we used $p_{\overline{\alpha}} = -q_\alpha$.

We now recall the Rouché-Capelli theorem: A non-homogeneous linear system has solution *iff* the rank of the associated homogeneous matrix is equal to the rank of the matrix associated to the full system. A trivial linear combination of rows of the homogenous matrix is still trivial when considering the matrix associated to the full system. This can be stated as:

$$\sum_i N_i f_i = 0 \tag{11.81}$$

where f_i is the non-homogeneous contribution to the ACC matrix of the orientifolded theory, coming from the tensor matter.

We now need to derive an expression for N_i in terms of the v_Γ. The Rouché-Capelli theorem tells us that the ACC system admits a solution *iff* Equation (11.81) holds for every value of v_Γ consistent with the topological constraints.

Faces with at Most One Tensor

Let us first focus on the simpler case where every gauge group has at most one tensor field. This result will be easily extended later to cases with more tensors. Consider a face of the mother theory with an edge on top of a fixed line. The rank assignment providing the coefficients for row reduction is given by the condition $N_i = -N_{i+k}$, $v_\alpha = -v_{\overline{\alpha}}$, and the difference between the ranks of two adjacent faces is given by $N_i - N_{i+k} = v_\alpha - v_{\overline{\alpha}}$. Combining these two results, we obtain

$$2N_i = N_i - N_{i+k} \; = \; v_\alpha - v_{\overline{\alpha}} = 2v_\alpha \, . \tag{11.82}$$

Let us now determine the f_i from the toric data. The method we are going to discuss below can be regarded as a generalization to orientifolds of the algorithm for finding the (minimal) matter content of a quiver in terms of basic knowledge of the (p, q) winding numbers of its ZZPs (equivalently of the external legs of the (p, q) web dual to the toric diagram). The intersection number between a given ZZP and the fixed line is

$$\det \begin{pmatrix} p & 1 \\ q & 1 \end{pmatrix} = p - q \, . \tag{11.83}$$

At every such crossing this ZZP, if not self-identified, will intersect its image on the line. The edge on which they cross will produce a tensor or conjugate tensor field, depending on the orientation of the crossing. This is depicted in Figure 11.13.

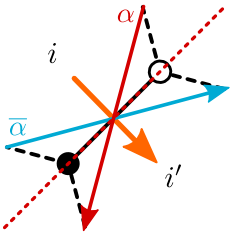


Figure 11.13: Crossing between a ZZP (and its image) over an edge on top of a diagonal fixed line. We show the corresponding bifundamental field in the mother theory.

From the discussion above, it is clear that the non-vanishing components of f_i are exactly those corresponding to the faces with a tensor, for which we have just determined the rank in terms of the ZZP values. Taking into account that the same ZZP can be related to $p_\alpha - q_\alpha$ tensors, this allows us to write Equation (11.81) as

$$\sum_i N_i f_i = (\pm 4) \sum_\alpha v_\alpha (p_\alpha - q_\alpha) = 0 \, , \tag{11.84}$$

where we have factorized the choice of sign for the diagonal O-line.

It is worth noting that the intersection with sign is a topological quantity that counts the minimal number of intersections of the ZZP with the fixed line in the dimer. This is, in fact, a homological invariant. In principle, more intersections are allowed, but they will come in pairs, one with positive and one with negative intersection, as shown in Figure 11.14. When computing the total contribution they cancel, leaving us with Equation (11.84), which does not depend on the particular phase we are considering.

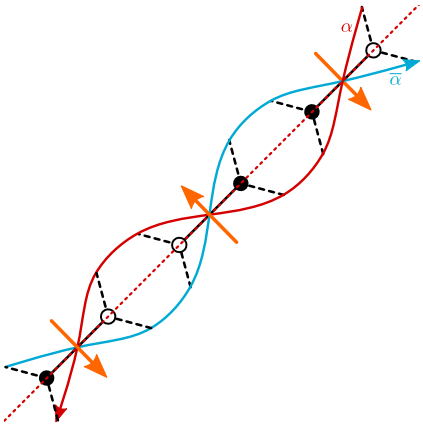


Figure 11.14: When ZPPs are deformed, additional intersections are added in pairs. We show the corresponding bifundamental fields in the mother theory.

We use the topological constraint Equation (11.80) to express the value assigned to v_1, as

$$v_1 = -\frac{1}{p_1 + q_1} \sum_{\alpha \neq 1} v_\alpha (p_\alpha + q_\alpha) \,. \tag{11.85}$$

Plugging this expression into (11.84) and rearranging the terms, we reach the following equality:

$$\sum_{\alpha \neq 1} v_\alpha \left(p_\alpha q_1 - p_1 q_\alpha \right) = 0 \,. \tag{11.86}$$

Then, the Rouché-Capelli theorem can be satisfied for generic v_α *iff*

$$p_\alpha q_1 - p_1 q_\alpha \equiv \det \begin{pmatrix} p_\alpha & p_1 \\ q_\alpha & q_1 \end{pmatrix} = 0 \,, \tag{11.87}$$

which implies that $p_\alpha = p_1, q_\alpha = q_1$ for every α. This implies that the toric diagram has a maximum of 4 edges, 2 of which are orthogonal to the symmetry axis. We dub the corresponding class of toric diagrams the *trapezoids*. An example of such a trapezoid is shown in Figure 11.15. Among trapezoids, we of course include also *triangles*.

Note also that there is a subset of trapezoids for which (11.84) is trivially satisfied. They have $p_\alpha = q_\alpha$ for every α so we refer to them as the *rectangles*, and describe orbifolds of F_0. See Figure 11.16 as an example. We remark that rectangles are the toric diagrams that give rise to line orientifolds without tensors in the spectrum. Thus, we recover the result that the latter always admit a non-anomalous solution.

Preliminary result for diagonal line orientifolds: Unless the toric diagram of the singularity under consideration is a trapezoid, any orientifold theory obtained from a dimer symmetric with respect to its diagonal, *and in which every face has at most one edge along this diagonal*, does not admit anomaly-free solutions.

Faces with Multiple Tensors

Faces with multiple tensors arise in examples as simple as the conifold or $\mathbb{C}^2/\mathbb{Z}_{2n+1}$ orbifolds, upon orientifolding with respect to a diagonal line. We now discuss how the previous discussion is extended

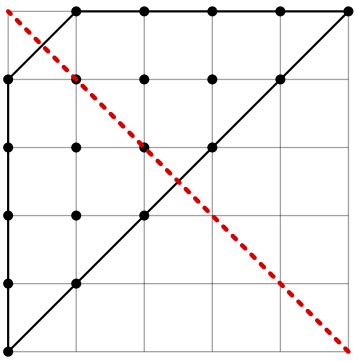


Figure 11.15: An example of a trapezoid for which you can find a non-anomalous diagonal line orientifold.

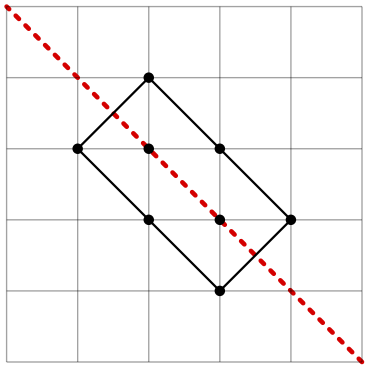


Figure 11.16: An example of a rectangle toric diagram with its diagonal axis of symmetry.

to these cases. We start by considering how multi-tensor faces may be embedded in the dimer. We will see that there are restrictions on the number of tensors a face can have. Moreover, their existence is non-trivial and imposes constraints on the toric diagrams. The analysis of this case is slightly different from the one in the previous section but will lead to the same result.

Interestingly, it is possible to find an upper bound on the number of tensors a face in the dimer can have. Figure 11.17 shows a face with two self-identified edges on the same side of the O-line. If they were adjacent, they would be connected at a 2-valent node, which corresponds to a mass term and then they could be integrated out. Naively, we might imagine that this can be avoided by introducing additional structure between the two edges, which is represented as a blob in Figure 11.17. But the ZZPs generating the edges on the line are the only ones that participate in the blob. In other words, the orange and purple ZZPs in Figure 11.17 must be identified with the blue and green ZZPs, with the precise identification depending on the number of intermediate edges. Therefore, the blob can only correspond to a sequence of edges connected by mass terms. After integrating them out, we are left with either zero or one tensor for an even or odd number of edges, respectively. This implies that a given face can only support more than one tensor in two cases: if they belong to different O-lines or if they belong to the same O-line but are coming from different copies of the unit cell as illustrated in Figure 11.18. In both cases, the previous analysis applies to each instance that the face touches a fixed line, so we conclude that the maximum possible number of tensors at a given face is two. The total number of tensors in the full theory is, however, unrestricted.

From Figure 11.18, we see that there can be three types of ZZPs: ZZPs parallel to the fixed line, which are forbidden since they would have to go through the face with two tensors, spoiling its very existence; ZZPs orthogonal to the fixed line, i.e. self-identified ZZPs, which do not give rise to tensors; finally, ZZPs which intersect in pairs on self-identified edges giving rise to tensors. Thus, the singularity can only have self-identified ZZPs, those of the Γ kind, and at most two couples of ZZPs of the α kind. Moreover, the (p, q) numbers of the latter are also subject to constraints. They cannot cross faces i and i' otherwise than passing by the O-lines, so they can intersect the grey dotted axis in Figure 11.18 at

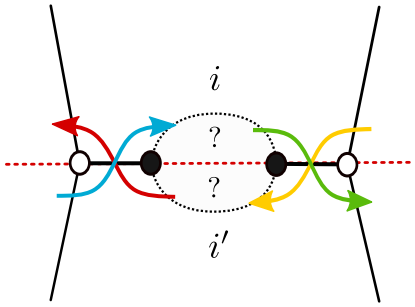


Figure 11.17: Two edges of a given face on a fixed line, separated by a general structure.

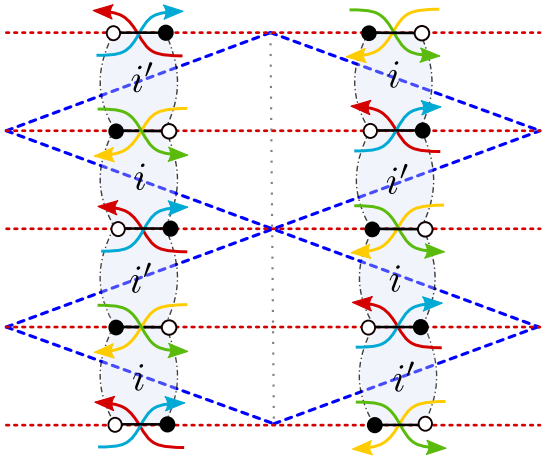


Figure 11.18: Faces with edges on top of the fixed line at different copies of the unit cell.

most twice if only one couple of ZZPs α is involved:

$$|p_\alpha + q_\alpha| \leq 2 \quad \text{for } \alpha = 1\,, \tag{11.88}$$

and once in the case of two couples:

$$|p_\alpha + q_\alpha| = 1 \quad \text{for } \alpha = 1, 2\,. \tag{11.89}$$

Those relations apply both for ZZPs α and $\bar{\alpha}$, for which the sums are respectively negative and positive.

If there is only one couple, the singularity corresponds to a trapezoid as the ones discussed in the previous section. Indeed, we have only one couple of ZZPs of the α kind and the topological constraint imposes $v_1 = 0$ for them, turning the RC condition into a trivial equation.

For two couples, the topological constraints and (11.89) impose

$$v_1 = -v_2\,. \tag{11.90}$$

This is the counterpart of the fact that faces i and i' in Figure 11.18 have to be of opposite ranks, following (11.82). Now, we can write the RC condition allowing faces to support one or two tensors in terms of v_1 only:

$$\sum_i N_i f_i = (\pm 4)(v_1(p_1 - q_1) - v_1(p_2 - q_2)) = 0\,. \tag{11.91}$$

Knowing (11.89), the only solution is $(p_1, q_1) = (p_2, q_2)$ so that we recover trapezoids. Let us note that the last equation considered with (11.90) can be brought to the form of (11.86) for two couples of ZZPs α, hence it is not surprising that a subset of trapezoids appears again as solutions in this context. For instance, the conifold does not provide a non-anomalous diagonal line orientifold while $\mathbb{C}^2/\mathbb{Z}_{2n+1}$ orbifolds do.

We conclude with a general result for diagonal line orientifolds:

Proposition 11.6 (Diagonal line orientifolds)**.** *Unless the toric diagram of the singularity is a trapezoid, any orientifold theory obtained from a dimer with a diagonal O-line is anomalous.*

See Figure 11.19 for more examples.

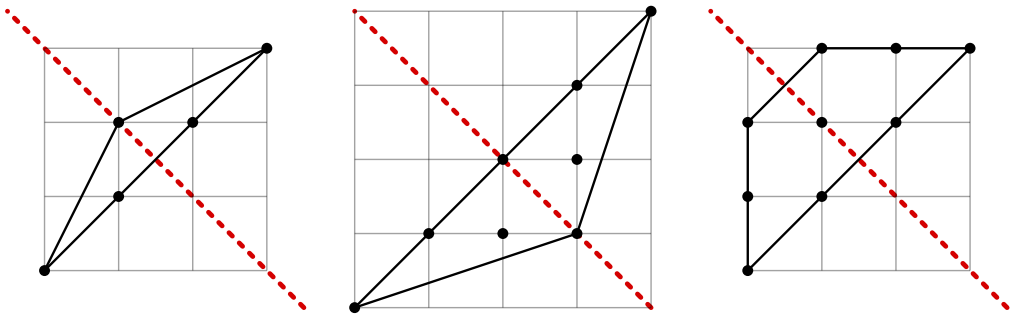


Figure 11.19: Examples of trapezoids, which admit anomaly-free fixed line orientifolds.

11.4.2 Horizontal/Vertical Line Orientifolds

In this section we consider horizontal fixed lines. The case of vertical lines is trivially related by rotation. The reasoning is essentially the same as the one described previously for the case of diagonal lines. This allows us to go fast to the main result for this class of orientifolds. In particular, we will not comment here about rectangles and faces with many tensors since the previous results are easily generalized.

Horizontal symmetry lines in the dimer correspond to a vertical symmetry in the toric diagram. The $\mathbb{Z}_2$ action maps a ZZP with winding (p, q) to a ZZP with winding $(-p, q)$. Again, we distinguish two different types of ZZPs:

- Pairs of ZZPs $\{v_\alpha, v_{\overline{\alpha}}\}$ for $\alpha = 1, ..., l$, where v_α and $v_{\overline{\alpha}}$ are exchanged under the symmetry, thus not parallel to the axis of symmetry.
- Self-identified ZZPs $\{v_\gamma\}$ for $\gamma = 1, ..., l_{\|}$, with winding numbers $(0, 1)$ or $(0, -1)$.

A general illustration of this is depicted in Figure 11.20.

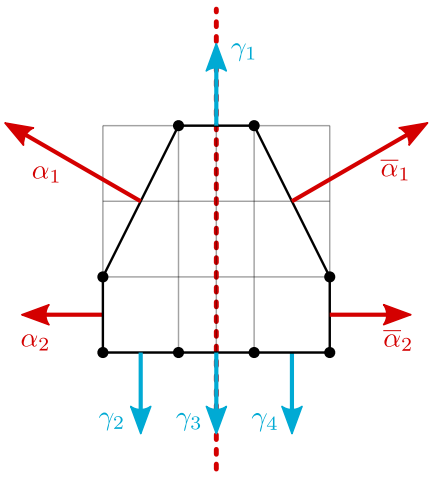


Figure 11.20: A generic singularity with a vertical axis symmetry.

In order to find the antisymmetric solutions to the ACC, we need to look at the antisymmetric value assignments of the ZZPs and impose the topological constraint

$$\Lambda = 2 \sum_{\alpha} v_\alpha p_\alpha = 0 \ . \tag{11.92}$$

Let us now consider the Rouché-Capelli condition. A ZZP of type α with winding numbers (p, q) crosses both fixed lines $-q$ times, counted with sign. The Rouché-Capelli condition can be expressed as

$$\sum_i N_i f_i = -\sum_\alpha v_\alpha q_\alpha (4 \text{ sign}(A) + 4 \text{ sign}(B)) = 0\,, \tag{11.93}$$

where sign(A) and sign(B) indicate the signs of the two fixed lines. Unlike the case of diagonal lines, the Rouché-Capelli condition in (11.93) becomes trivial as soon as sign(A) and sign(B) are different. In that case, the orientifold theory is always anomaly-free.

If the two fixed lines have the same sign, (11.92) allows us to express v_1 in terms of the remaining v_α, as in the case of diagonal lines. Plugging this expression into Equation (11.93) leads to

$$\sum_{\alpha \neq 1} v_\alpha \left(p_\alpha q_1 - p_1 q_\alpha\right) = 0\,. \tag{11.94}$$

With the same analysis of the previous section, we find that singularities with two horizontal lines of the same sign admit a solution to the ACC only if they are trapezoids, just as in the case of diagonal lines. See Figure 11.21 for examples.

Proposition 11.7 (Horizontal/vertical line orientifold)**.** *Toric diagrams symmetric with respect to a horizontal/vertical axis always lead to anomaly-free orientifold theories when the two O-lines have opposite signs. When the signs are the same, instead, in order to yield a non-anomalous orientifold theory the toric diagram of the singularity must be a trapezoid.*

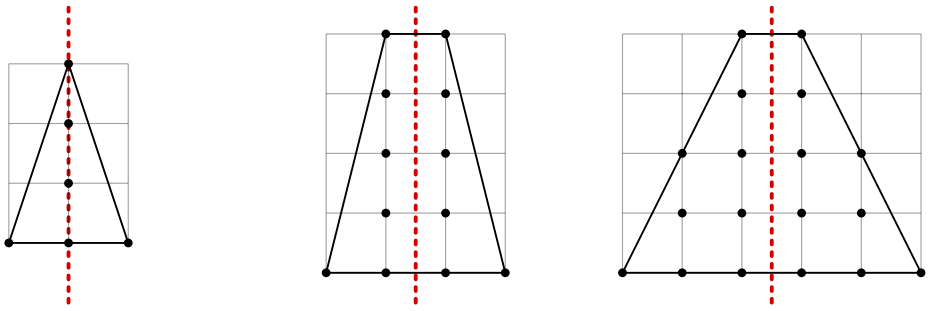


Figure 11.21: Examples of trapezoids, which admit anomaly-free horizontal fixed line orientifolds.

11.4.3 Fixed Point Orientifolds

Finally, we address the case of fixed point orientifolds. We should state right away that the results in this case are less conclusive than for fixed lines. Indeed, one can easily anticipate that having a larger number of signs to fix (at the four fixed points), it will be straightforward to satisfy the ACC just by a wise choice, similarly to the case with horizontal/vertical line. On the other hand, if one sticks with a 'wrong' choice, the restriction on the allowed geometries cannot be characterized as nicely as in the previous cases.

As already explained in Section 11.3, the action of the orientifold on every ZZP is to map it either to itself or to another ZZP with the same winding numbers. We thus divide the ZZPs into two sets:

- Pairs of distinct ZZPs $\{v_\alpha, v_{\overline{\alpha}}\}$ for $\alpha = 1, \cdots, k$ that are exchanged.
- Self-identified ZZPs $\{v_\gamma\}$, for $\gamma = 1, \cdots, l$.

The total number of ZZPs is $n = 2k + l$.

In this kind of orientifolds, tensors arise whenever a pair of self-identified ZZPs intersect over a fixed point. Moreover, a ZZP going through a fixed point necessarily goes through a second fixed point [RU16b]. As a result, it is easy to convince oneself that the number of tensors, if present at all, must be between 2 and 4, and it coincides with the total number of self-identified ZZPs that cross a fixed point.

In order to find the antisymmetric solutions to the ACC, we need to consider symmetric value assignments for the ZZPs, as explained in Section 11.3, subject to the topological constraints

$$\begin{aligned} \Lambda &= 2\sum_\alpha v_\alpha p_\alpha + \sum_\gamma v_\gamma p_\gamma = 0\,, \\ M &= 2\sum_\alpha v_\alpha q_\alpha + \sum_\gamma v_\gamma q_\gamma = 0\,. \end{aligned} \tag{11.95}$$

The RC equation becomes

$$\sum_i N_i f_i = \sum_{\gamma\neq\gamma'} (v_\gamma - v_{\gamma'})(\pm 4) = 0\,. \tag{11.96}$$

where the sum in the middle runs over the tensors. The signs depend on the sign of the fixed points and on the orientations of the self-identified edges. Depending on which of the two faces adjacent to the edge we preserve in the projection, we get tensors or their conjugates, contributing with opposite signs to the ACC.

We recall that the signs of the fixed points, in contrast with fixed lines, are constrained by the *sign rule* [FHK$^+$07]. The rule prescribes that the product of the four signs is $(-1)^{n_W/2}$, with n_W the number of superpotential terms.[9]

We now consider the different possibilities, i.e. $l = 2$, 3 and 4 tensors. Our analysis is general and does not distinguish between faces with single or multiple tensors.

- $l = 2$: In this case we have two tensors, meaning that two ZZPs cross each other on two fixed points. Equation (11.96) reads

$$(v_1 - v_2)(\pm_1 4) \pm (v_1 - v_2)(\pm_2 4) = 0\,, \tag{11.97}$$

  where the $\pm_i$ indicate the signs of the fixed points, while the additional $\pm$ signs depends on whether the tensors are conjugated or not.

  Since only two fixed points are involved in this case, their signs can always be chosen such that this equation is trivially satisfied, while satisfying the sign rule. However, it is interesting to consider whether there are other ways to satisfy this constraint. We can impose $v_1 = v_2$ using the two equations of the topological constraint. Expressing v_1 and v_2 as function of the other v_α's we get

$$\begin{aligned} v_1 &= \frac{2}{p_1 q_2 - p_2 q_1}(p_2 \sum_\alpha v_\alpha q_\alpha - q_2 \sum_\alpha v_\alpha p_\alpha)\,, \\ v_2 &= \frac{2}{p_1 q_2 - p_2 q_1}(q_1 \sum_\alpha v_\alpha p_\alpha - p_1 \sum_\alpha v_\alpha q_\alpha)\,, \end{aligned} \tag{11.98}$$

  where we have assumed $p_1 q_2 - p_2 q_1 \neq 0$. Equating v_1 and v_2, we obtain

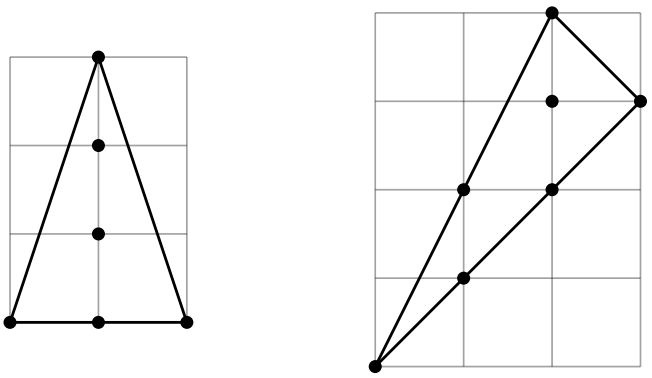
$$\sum_\alpha v_\alpha (p_\alpha (q_1 + q_2) - q_\alpha (p_1 + p_2)) = 0\,. \tag{11.99}$$

  Since this equation must hold for all v_α, the only possibility is that all terms in the summation vanish, thus $p_\alpha(q_1 + q_2) = q_\alpha(p_1 + p_2)$ for all α. Solutions are of the form $p_1 = -p_2$ and $p_\alpha = 0$, up to $SL(2,\mathbb{Z})$ transformations. Those correspond to trapezoids (not necessarily symmetric with respect to any axis) with an even number of ZZPs on each base and only one ZZP on each side.

  If $p_1 q_2 - p_2 q_1 = 0$, it means that $(p_1, q_1) = -(p_2, q_2)$, since the two ZZPs are parallel and, in order to intersect in a consistent way, they must have opposite winding numbers. In this case, the topological constraint imposes $v_1 = v_2$ if $p_\alpha q_{\alpha'} - q_\alpha p_{\alpha'} = 0$ where $\alpha \neq \alpha'$. It means that all non self-identified ZZPs have to be either parallel or anti-parallel to each other. This condition is satisfied by all toric diagrams with the shape of a rectangle or a parallelogram where there is an even number of non self-identified ZZPs. Together with the solutions of the previous paragraph, they constitute a class of trapezoids for which any sign assignment for the fixed points leads to an anomaly free theory when two tensors are involved.

  As an illustration, consider fixed point orientifolds of $\mathbb{C}^3/\mathbb{Z}_6$ with actions (1,1,4) and (1,2,3), whose toric diagrams are shown in Figure 11.22. Both of them admit an orientifold with two tensors. Our analysis implies that only the first one admits tensors with any sign, as it can easily be checked by explicitly solving the ACC.

[9] Generically, it is not known whether the parity of $n_W/2$ can be deduced from the toric diagram.



(a) (b)

Figure 11.22: The toric diagrams for the $\mathbb{C}^3/\mathbb{Z}_6$ orbifolds with actions: (a) (1,1,4) and (b) (1,2,3).

An interesting scenario is when tensors arise from the orientifold projection of adjoints in the mother theory, namely from edges separating self-identified faces. In this case, the ACC of the self-identified gauge group is trivially zero, since it is either SO or USp. In this situation, the two self-identified ZZPs intersect all other ZZPs only once. This can be understood as follows. Let us consider a line passing through the fixed points under consideration. All the non self-identified ZZPs must be parallel to this line, since otherwise their intersections with the line would imply that they go through the self-identified face, which in turn would spoil the fact that it is self-identified. The $\mathbb{C}^2/\mathbb{Z}_{2m}$ orbifolds are examples in this class, see Figure 11.23.

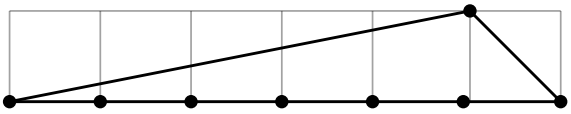


Figure 11.23: The toric diagram of $\mathbb{C}^2/\mathbb{Z}_6$, as an example of the $\mathbb{C}^2/\mathbb{Z}_{2m}$ family.

- $l = 3$: In this case we have three tensors, i.e. three ZZPs intersecting on three fixed points. Equation (11.96) reads

$$(v_1 - v_2)(\pm_1 4) \pm (v_2 - v_3)(\pm_2 4) \pm (v_3 - v_1)(\pm_3 4) = 0 \,. \tag{11.100}$$

Since only three of the fixed points are involved, it is possible to pick their signs such that this equation is trivially satisfied. These choices in turn determine the sign of the fourth fixed point due to the sign rule.

If instead we have a different combination of signs, we end up with an equation of the form

$$v_\gamma - v_{\gamma'} = 0 \,, \tag{11.101}$$

with γ and γ' two of the three ZZPs above. The missing $v_{\gamma''}$ in the previous equation depends on the choice of fixed point signs in Equation (11.100). Therefore, in order to have a solution for all possible fixed point sign assignments we need to impose $v_1 = v_2 = v_3$ with the topological constraint. This means that the ZZPs have winding numbers of the form $(p_1, 0)$, $(-p_1, q_2)$ and $(0, -q_2)$, up to $SL(2,\mathbb{Z})$ transformations. The only solution is $p_1 = q_2 = 1$, corresponding to $\mathbb{C}^3$, i.e. flat space.

A face with multiple tensors imposes constraint(s) of the form $v_1 - v_2 = \pm(v_2 - v_3)$, leading to an RC constraint of the form

$$(v_1 - v_2)(\pm_1 4) \pm (v_1 - v_2)(\pm_2 4) \pm (v_3 - v_1)(\pm_3 4) = 0 \,. \tag{11.102}$$

Again, the existence of solutions depends on the signs of the fixed points. Solutions for generic signs can be obtained only when $v_1 = v_2 = v_3$, i.e. for flat space.

- $l = 4$:

  This case, in contrast with the previous ones, does not always admit a solution to the ACC. The reason for this is that the four fixed points are used, their signs are constrained by the sign rule and we no longer have the freedom of unused fixed points.

  The RC equation can take two different forms, depending on the ZZP intersections:

$$\begin{aligned}&(v_1 - v_2)(\pm_1 4) \pm (v_2 - v_3)(\pm_2 4) \pm (v_3 - v_4)(\pm_3 4) \pm (v_4 - v_1)(\pm_4 4) = 0 \ , \\ &(v_1 - v_2)(\pm_1 4) \pm (v_1 - v_2)(\pm_2 4) \pm (v_3 - v_4)(\pm_3 4) \pm (v_3 - v_4)(\pm_4 4) = 0 \ . \end{aligned} \tag{11.103}$$

  Since the signs of the fixed points are constrained, it is not always possible to trivially solve the RC equation.

  Moreover, it is also impossible to find general non-trivial solutions by using the topological constraint to force some of the v_i to be equal. For the first equation, we need all the v_i to be equal. To do so, we need at least three equations, but the topological constraint provides only two. In the second case, we can impose $v_1 = v_2$ and $v_3 = v_4$ with the following ZZPs: $(1, 0)$, $(-1, 0)$, $(0, 1)$ and $(0, -1)$, which define the conifold singularity. Unfortunately, the conifold gives rise to an RC of the first kind, not of the second one.

To conclude, this partial analysis retained only one toric diagram that can accommodate any signs for its fixed points: flat space. We eventually found some particular trapezoids for which we can freely chose the signs of the tensors when only two are present, but those singularities also allow *in principle* for fixed point orientifolds with four tensors, where our analysis showed its limits. Thus, we cannot say in general that they provide every kind of anomaly-free orientifolds. As an illustrative example, one can check that the orientifold of Figure 11.22a with four tensors does not allow for every combination of signs satisfying the sign rule, although it does with only two tensors.

It would be interesting to investigate further whether it is possible to determine the solvability of the ACC from the toric diagram. We leave this question for future work. In the meantime, orientifolds with four self-identified ZZPs need to be studied in a case by case basis.

11.5 Conclusions

In this paper we studied anomalies in gauge theories living on D-branes probing orientifolds of toric singularities, focusing on pure D3-brane theories, namely without the addition of extra flavors.

We introduced a new, geometric algorithm for finding anomaly-free solutions based on zig-zag paths. The main virtue of this procedure is not so much its practicality over the direct solution of the ACC in explicit examples, but the fact that it allows us to make general statements regarding anomalies directly from geometry. Indeed, we managed to derive stringent no-go theorems that establish the conditions for anomaly-free solutions in these orientifolds. Such results are extremely useful, since until now the cancellation of anomalies in this class of theories was analyzed on a case-by-case basis.

We can summarize our findings as follows, from the most stringent case to the less conclusive one:

- For orientifolds with a fixed diagonal line, for which one has to choose only one sign, we find that only singularities whose toric diagram is a trapezoid with respect to the diagonal axis of symmetry allow for a non-anomalous D-brane gauge theory.
- For orientifolds with fixed horizontal lines, we have two signs to choose. All singularities lead to anomaly-free theories if the two signs are chosen to be opposite to each other. If the singularity has a toric diagram which is a trapezoid with respect to the vertical axis of symmetry, then the theory is non-anomalous also for equal signs.
- For orientifolds with fixed points, there are four signs to choose, up to a constraint on their product. Moreover, the relation between the fixed points in the dimer and the toric diagram of the singularity is less direct. Because of these two facts, it is more difficult to summarize the few instances where a restriction is indeed obtained on the singularities that lead to non-anomalous theories. The particular cases have been detailed in Section 11.4.3.

As an illustration of the power of the ideas introduced in this work, they were exploited in [ABF$^+$21c, ABF$^+$21b] to guide the search of models of D-branes at singularities that display dynamical supersymmetry breaking. Such models necessarily involve orientifolds, but have a potential instability as soon as the singularity allows for a partial resolution which is non-isolated (in D-brane jargon, this translates to the presence of $\mathcal{N} = 2$ fractional branes [BGVU19]). In terms of the toric diagram, this property manifests itself through points within the external edges of diagram, or in other words, parallel ZZPs[10]. It is straightforward to see that toric diagrams that fall in the class of trapezoids always include such points on the boundary, except for few very simple cases (namely F_0 and orbifolds with a toric diagram which is an isosceles triangle with a unit base). As a consequence, if one is to look for fixed line orientifolds which allow for anomaly free D-brane configuration, and with no non-isolated partial resolution, the only option one is left with is horizontal/vertical fixed lines with opposite signs. The octagon singularity [ABF$^+$21c], which is the subject of the next chapter, is the simplest non-trivial singularity that satisfies these requirements.

[10] In [BCH$^+$14, BBMR20] the existence of such flat directions in moduli space was exploited to add relevant mass deformations in dimers with and without orientifolds.

Chapter 12

The Octagon

Let us now come back to our original quest, that is, the study of whether it is possible at all to implement the SU(5) DSB model in brane tilings corresponding to isolated singularities. The ideas presented in the last chapter taught us *a posteriori* why the diagonal line orientifolds presented in Figure 10.11 and Figure 10.13, which incarnated the most parsimonious hope to embed the hexagonal cluster of faces describing the twin SU(5) model in consistent brane tilings corresponding to isolated singularities, were doomed from the start and cannot admit anomaly-free rank assignments without flavor D7-branes. More precisely, the results of Chapter 11 imply that in order to embed the hexagonal cluster of the twin SU(5) model in a brane tiling orientifold without $\mathcal{N} = 2$ fractional branes, one must do so in an orientifold with two vertical or horizontal fixed lines carrying opposite signs.

Despite the result of Section 10.3 which show that orientifolds with two fixed lines are constrained by topological obstructions, it turns out that it is possible to achieve our goal: we will present a dimer model with two vertical fixed lines, corresponding to an isolated singularity, and containing the hexagonal cluster of the twin SU(5) model. Provided the instabilities known as of today are the only ones from which DSB models embedded in warped throats can suffer, this model describes a configuration of fractional branes at a singularity that dynamically breaks supersymmetry in a stable vacuum. The first section of this chapter describes the construction of this distinguished dimer model using the ideas of Chapter 10 as in [Tat21], while the second is the analysis of the model hence obtained, along the lines of [ABF$^+$21c, ABF$^+$21b].

12.1 The rise of the Octagon

We reproduce for convenience the hexagonal cluster of faces corresponding to the twin SU(5) DSB model on the left of Figure 12.1. The local ZZP structure near such a 6-valent vertex is displayed in the middle of the same figure, of which we need only to keep the boundary data, displayed on the right.

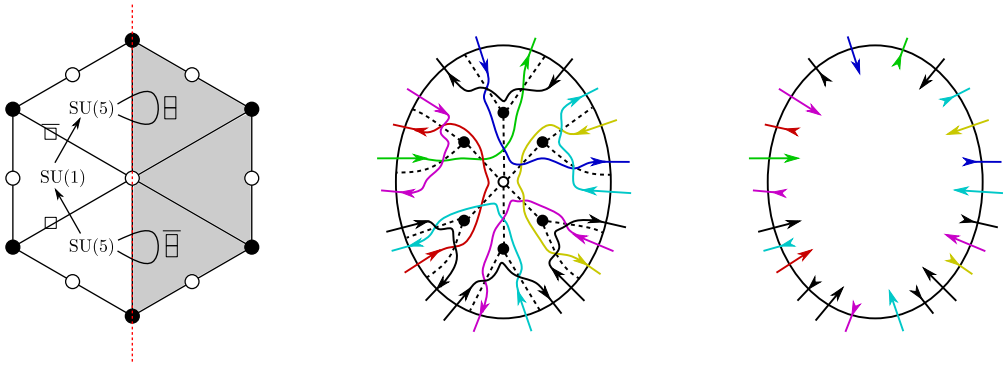


Figure 12.1: The hexagonal cluster hosting the twin SU(5) model.

Our goal is to embed this circle (with in/out points and pairing) inside a consistent dimer model with the constraints listed above. One can show that a polygon that can possibly work must have at least 8 sides; the octagon shown on the left of Figure 12.2 is a parsimonious choice. One places straight lines on

a fundamental cell as before, in a symmetric way with respect to the dashed vertical axis. This is shown in Figure 12.2.

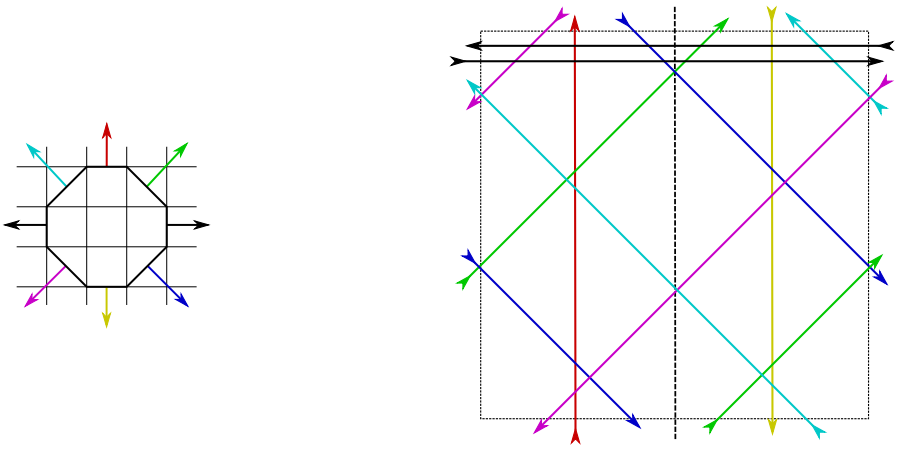


Figure 12.2: The outgoing normal vectors of the octagon as straight geodesics on a torus.

One needs to deform the boundary of the fundamental cell so that endpoints of strands are distinct and such that the 'in' and 'out' insertions alternate. This is shown on the left of Figure 12.3. Keeping the boundary data only, one can insert the circle corresponding to the hexagonal cluster on the symmetry axis in the middle of the cell. The resulting diagram is displayed on the right of Figure 12.3.

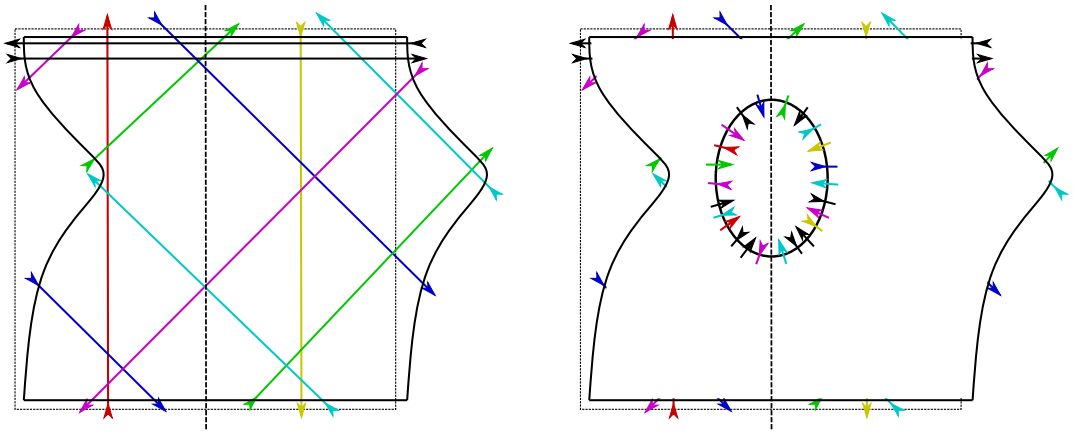


Figure 12.3: Deforming the fundamental cell and embedding the hexagonal cluster.

The bluish and greenish parts of the deformed fundamental cell on the left of Figure 12.4 are topological disks with boundary data and pairing satisfying Thurston's conditions, hence one can construct a symmetric triple crossing diagram realizing the boundary data, shown on the right of the same figure.

Running the algorithm as before yields a dimer model dubbed the Octagon, because of its toric diagram. It is presented under a fancy guise on the right of Figure 12.5.

This dimer model remarkably satisfies all the constraints listed in the introduction of Chapter 10: 1 the octagon singularity is obviously isolated, 2 the hexagonal cluster of the twin SU(5) model does lie on a fixed line of the orientifold projection, 3 it is isolated on a fractional brane (consisting of the faces 1, 2, 3, 7, 12, 13 and 14) which is preserved by the orientifold projection, 4 there exist solutions to the anomaly cancellation conditions in the orientifold theory.

The determinant of the Kasteleyn matrix corresponding to some Pfaffian orientation is

$$\det(K) = w^3z^2 + w^3z + w^2z^3 - 24w^2z^2 + 26w^2z - w^2 + wz^3 + 24wz^2 + 26wz + w - z^2 + z \ . \quad (12.1)$$

One may compute the Newton polygon of the above expression, and it corresponds to the octagon toric diagram, as expected [HK05]. There is a single perfect matching for each of its external points, thus ensuring that the dimer meets a necessary condition of minimality.

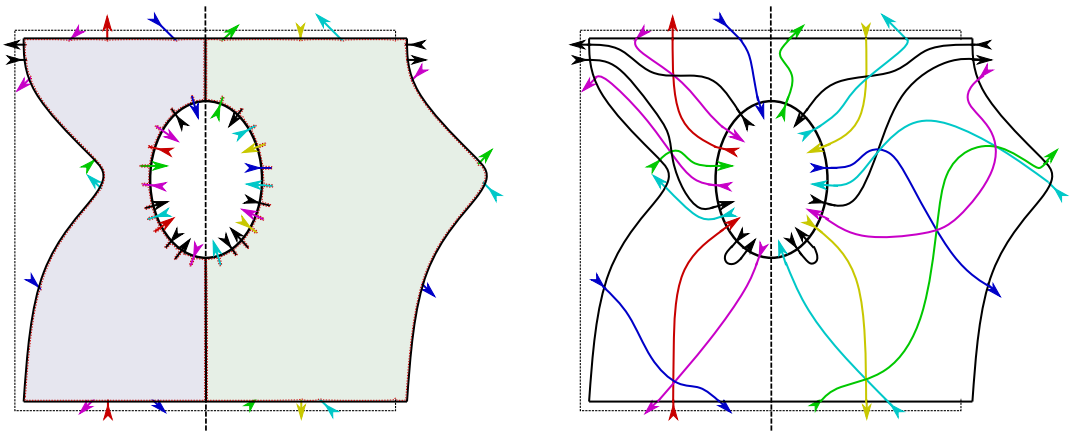


Figure 12.4: From the boundary data to a symmetric triple diagram.

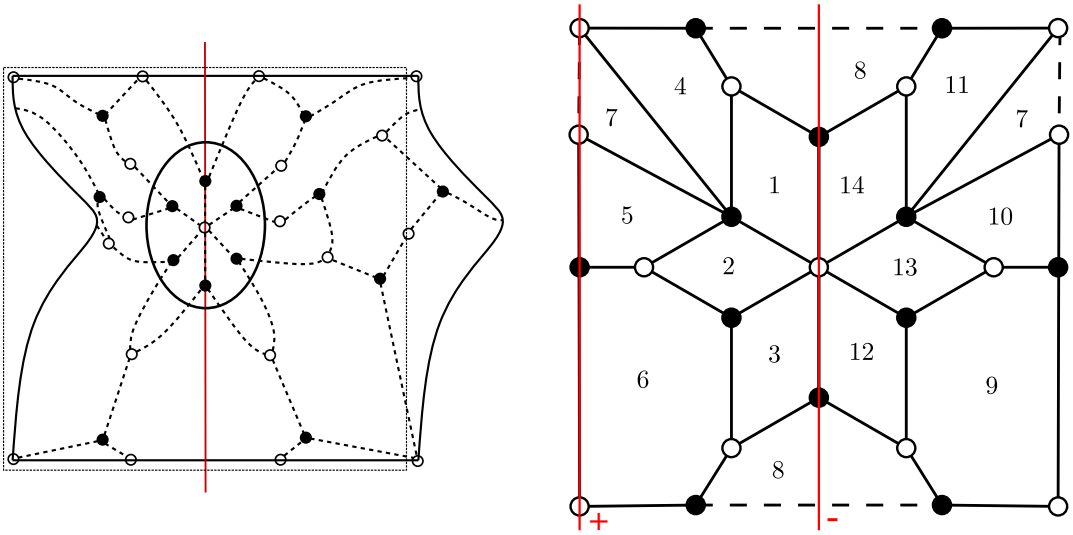


Figure 12.5: The octagon dimer model.

Let us look at the orientifold gauge theory more closely. The orientifold projection identifies faces $(1,\dots,6)$ with faces $(14,\dots,9)$ while faces 7 and 8 are self-identified. Hence, D-branes at such orientifold singularity are described by a matter coupled supersymmetric gauge theory with six SU factors, one SO and one USp factors. The twin $SU(5)$ model is given by the rank assignment $SU(5)_1 \times SU(1)_2 \times SU(5)_3$ with all other faces being empty but face 7 which is a decoupled pure SYM with gauge group $SO(5)$ and hence confines on its own. ACC and self-consistency of such rank assignment follow the general discussion in Section 9.2. This model represents a concrete example of an orientifold singularity which allows DSB by a D-brane bound state which is free of any known instability, in particular the $\mathcal{N}=2$ fractional brane decay channel or the runaway behavior typical of DSB branes. The absence of $\mathcal{N}=2$ fractional branes is clear from the octagon toric diagram, which does not have internal points on boundary edges. This model therefore provides a realization (the first, to our knowledge) of stable DSB with D-branes at CY singularities and suggests for an extension of the string theory landscape as it is currently known.

The Octagon emerges as the simplest possible dimer having all required properties to admit, upon orientifolding, stable DSB D-brane configurations. One might ask whether less minimal models exist which share the same properties. We do not have an answer to this question, yet. Still, dimer techniques have (once again) proven to be a very powerful tool to provide a direct link between geometry and gauge theories dynamics, both in finding no-go theorems, like the one presented in [ABMP19] or the connection between minimal $SU(5)$ and 3-2 models and the presence of $\mathcal{N}=2$ fractional branes established in this paper, as well as in unveiling concrete ways to evade them. Therefore, we cannot exclude that further surprises are possible and generalizations of the Octagon model will eventually be found.

12.2 Analysis of the model

Before performing the orientifold projection, it is straightforward to see that the following rank assignment is anomaly free: faces $1, 2, 3, 7, 12, 13$ and 14 have gauge group $SU(N+M)$, and all the others have gauge group $SU(N)$. Setting $N = 0$, one has only seven $SU(M)$ gauge groups: one isolated Super-Yang-Mills (SYM) on face 7 and the six others forming a loop whose links are bi-fundamentals, together with a sextic superpotential proportional to the only gauge invariant (it is represented by the white dot in the center of the unit cell). This rank assignment corresponds to the worldvolume theory on a stack of M deformation fractional branes of our interest. One can easily see that such a gauge theory eventually leads to a confining behavior just like SYM. This can be naturally UV completed starting from a system of N regular and M fractional D3-branes which trigger a RG-flow that can be described by a duality cascade, similar to [KS00a] and many other examples that were found since then. The effective number of regular branes diminishes along the flow and the deep IR dynamics is described by fractional branes only.

In the presence of the orientifold projection (choosing opposite signs for the two fixed lines), one can see that there is a rank assignment which is anomaly free: $SU(N+M+4)$ for faces 1 and 3, $SU(N+M)$ for face 2, $SO(N+M+4)$ for face 7, $SU(N)$ for faces 4, 5 and 6, and $USp(N)$ for face 8. Setting $N = 0$ we obtain a gauge theory with an isolated $SO(M+4)_7$ SYM theory, which confines on its own, together with a quiver gauge theory based on the group $SU(M+4)_1 \times SU(M)_2 \times SU(M+4)_3$ with matter fields and a superpotential that we proceed to analyze.

The DSB model The gauge theory

$$SU(M+4)_1 \times SU(M)_2 \times SU(M+4)_3 \tag{12.2}$$

has matter content

$$A_1 = \square_1,\ X_{12} = (\overline{\square}_1, \square_2),\ X_{23} = (\overline{\square}_2, \square_3),\ A_3 = \overline{\square}_3 \tag{12.3}$$

and superpotential

$$W = A_1 X_{12} X_{23} A_3 X_{23}^t X_{12}^t\ . \tag{12.4}$$

The superpotential can be interpreted as follows. The gauge invariant $X_{12}^t A_1 X_{12}$ of group 1 and the gauge invariant $X_{23} A_3 X_{23}^t$ of group 3 are respectively in the $\square_2$ and $\overline{\square}_2$ of gauge group 2, with W above providing a bilinear in these two invariants, thus akin to a mass term. It is obvious that the antisymmetrics of $SU(M)_2$ can exist as such only if $M \geq 2$. In this case, one can show that strongly coupled dynamics generates superpotential terms that, together with the tree level one, eventually lead to supersymmetric vacua. For $M = 0$ one gets instead two decoupled theories at faces 1 and 3 both having gauge group $SU(4)$ and one chiral superfield in the antisymmetric, which have a runaway behavior. The case of interest is $M = 1$.

For $M = 1$ node 2 becomes trivial ($SU(1)$ is empty) and, more importantly, the superpotential actually vanishes. Indeed, both nodes 1 and 3 are $SU(5)$ gauge theories with matter in the $\square \oplus \overline{\square}$ representations, and there is no chiral gauge invariant that can be written in this situation [ADS84]. Hence the two gauge theories are effectively decoupled, and their IR behavior can be established independently. Both happen to be the $SU(5)$ model for stable DSB. Since the $SO(5)$ SYM on node 7 just confines, we thus determine that this configuration displays DSB in its vacuum. Quite interestingly, this DSB vacuum may then arise at the bottom of a duality cascade (possibly more complicated with respect to the simpler unorientifolded case, due to the orientifold projection which would modify it, see [AB18]), hence within a stringy UV completed theory.

Stability Is this DSB vacuum stable? There can be different sources of potential instabilities.

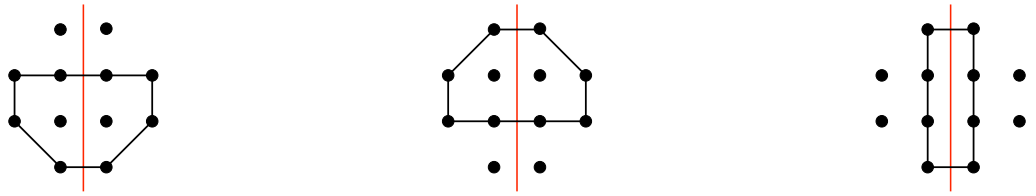
First, one could be concerned about stringy instantons, whose presence may affect the low energy dynamics. Indeed, the D-brane configuration giving rise to the twin $SU(5)$ DSB model, $N = 0, M = 1$, contains both a $USp(0)$ and an $SU(1)$ factor coupling to the $SU(5)$ gauge groups. These are the two instances where contributions to the low-energy effective superpotential are allowed (see [ABF$^+$07] and [Pet08], respectively). However, no such contributions can be generated in our model simply because there are no chiral gauge invariants that can be written which can contribute to the superpotential. We thus conclude that stringy instantons cannot alter the DSB dynamics.

A second source of instability is the one discussed in [BGVU19, ABMP19]. However, since by construction there are no $\mathcal{N} = 2$ fractional branes, there is no vacuum expectation value on which the energy of the DSB vacuum can depend on, or equivalently there is no Coulomb branch along which the energy can slide to zero value.

A third source of instability may come from the $\mathcal{N} = 4$ Coulomb branch represented by regular D3-branes. As in the previously analyzed cases [BGVU19, ABMP19], one can easily show that this is a non-supersymmetric flat direction, essentially because of the conformality of the parent (non-orientifolded, large N) gauge theory. Therefore, there are no supersymmetric vacua along this branch [1].

Lastly, one could in principle wonder whether $\mathcal{N} = 2$ fractional branes could be reintroduced in partial resolutions of the octagon singularity that could arise in some limits, thus potentially destabilizing the DSB vacua or at least making it metastable instead of stable.

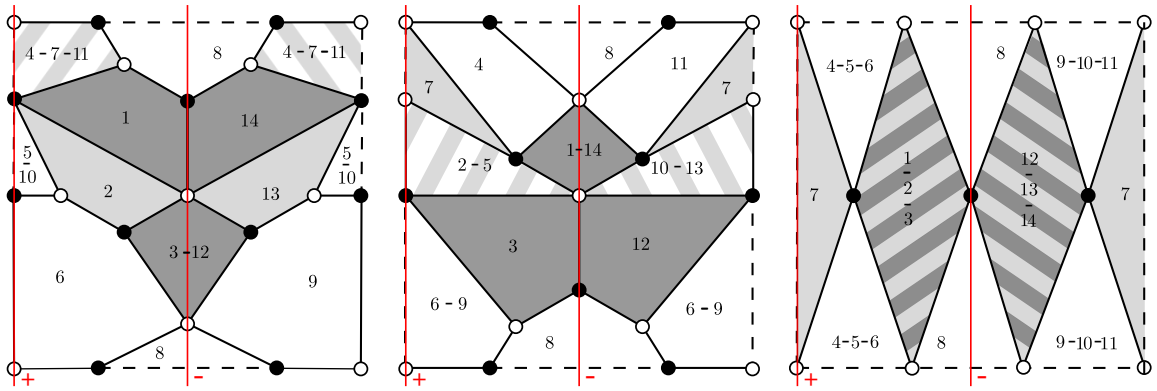
We list in Figure 12.6 the first partial resolutions of the Octagon that allow for the presence of dangerous $\mathcal{N} = 2$ fractional branes and in principle can accommodate the orientifold projection. Following [RU16b], the latter criterion means that the resulting toric diagram has to remain symmetric with respect to the orientifold line. Further partial resolutions consistent with the orientifold projection inexorably lead to orbifolds of the conifold, for which our comments on the case of Figure 12.6c will remain valid.



(a) (b) (c)

Figure 12.6: First partial resolutions of the orientifolded Octagon admitting $\mathcal{N} = 2$ fractional branes.

The corresponding dimer diagrams are obtained following the methods of [GESU06, Gul08] and are presented in Figure 12.7. The algorithm operates the partial resolution by merging some zig-zag paths in the dimer on the right of Figure 12.5. This action is equivalent to assigning a VEV to the edges on which these zig-zag paths cross each other.



(a) (b) (c)

Figure 12.7: Dimer diagrams after partial resolutions.

In the cases of Figure 12.7a and Figure 12.7b, we find that the partial resolution is in obstruction with the very nature of our deformation brane because it implies the fusion of faces of different ranks already at the level of the non-orientifolded theory. This is the gauge theory counterpart of having blown-up cycles wrapped by the deformation brane in the singular geometry, which are forbidden following [GESU06]. In

[1] Flat directions are usually not expected in a non-supersymmetric vacuum. Subleading $1/N$ corrections to anomalous dimensions of matter fields, which could lift such flat direction, are not easily calculable, particularly in a complicated singularity such as the Octagon. However, they should neither change the number of supersymmetric vacua nor modify the behavior of the potential at infinity, at least for sufficiently large N.

the case of Figure 12.7c, the partial resolution is obstructed because it gives a VEV to edges separating faces of ranks that differ by the orientifold charge, for example the edge separating faces 1 and 2.

* * * * * * *

The Octagon dimer model is the first instance, to our knowledge, of a stable DSB configuration of fractional branes. As an existence proof of such configurations, this is enough. However, it is not by chance that this particular singularity has been found, rather one can be led to it by a series of arguments from the previous chapters. The octagon CY3 is most likely the simplest singularity allowing for stable DSB, however more complicated singularities may probably do so as well, though always through the twin $SU(5)$ model. This is the reason why the simplest occurrence of this phenomenon is a singularity corresponding to a quiver with no less than 14 gauge groups.

With this example, we have shown that stable DSB can still be engineered by brane configurations at Calabi-Yau singularities. Of course, this statement relies on the belief that there is no other instability channel for stringy UV completions of DSB models. However even there is, the study of the Octagon will still prove very interesting, since if it is unstable it can only be so through these yet-to-be-discovered instabilities.

Given the remarkable properties of this family of models, we consider it important to study them in further detail.

Part IV

New orientifolds of brane tilings

D-branes at singularities extend the original $\mathcal{N} = 4$ SYM AdS/CFT correspondence [Mal99, Wit98a] to theories without conformal invariance and/or reduced supersymmetry [KW98, KN00, KT00, KS00a]. These setups have enriched our understanding of both QFT and String Theory by providing a geometric understanding of gauge dynamics and dualities while giving tools to tackle brane dynamics through field theory computations [FHH01a, FHH01b, BP01, FHHU01, KS00a, BHOP05, FHSU06, BBC05, IU08, GVTU17]. Furthermore, these setups allow for bottom-up constructions in string phenomenology where most features of the gauge theory depend only on the local features of the compactification [AIQU00, BJL02, VW07]. More recently, warped throats have been used in trying to uplift to de Sitter vacua [FGRU15, RUW15] and in the search for SUSY breaking vacua in quantum gravity [FHSU06, BHOP05, ABFK07b, BGVU19, ABMP19, ABF$^+$21c, ABF$^+$21b].

The correspondence is particularly sharp when one considers $4d$ $\mathcal{N} = 1$ gauge theories arising in D3-branes probing singular, non-compact toric CY_3 varieties. The problem of finding the gauge theory given the CY_3 was solved by brane tilings, also dubbed dimer models [HK05, FHM$^+$06, FHK$^+$06], where all the informations of the gauge theory are encoded in a bipartite graph on a torus.

Besides propagating strings on singular backgrounds, one can consider a particular gauging of a $\mathbb{Z}_2$ isometry of space together with worldsheet parity, that is, an orientifold [PS89, Hor89, DLP89, BS90, BS91]. Its projection on the open string spectrum opens the D-brane/gauge correspondence to new possibilities. For instance, they extend the available gauge and matter fields, which may break conformal invariance, and allow non-perturbative contributions to the superpotential through instantons [IU07, ABF$^+$07, BCKW09]. While direct construction in string theory is in practice only feasible for some orbifold theories, they may be constructed directly in the dimer model [FHK$^+$07] by identifying gauge groups and fields according to a suitable involution of the graph and possibly assigning some signs to the fixed loci in the dimer, corresponding to the different choices in the orientifold projection. This makes possible for toric singularities to be orientifolded. In the same paper, orientifolds were classified in two groups, depending on the involution, those that leave four fixed points and those that leave a single or two fixed lines, in the dimer. Interestingly, these correspond to three of the five possible smooth involutions on the torus [Dug19], the remaining ones corresponding to a shift of the fundamental cell and a glide reflection, i.e. combining a shift and a reflection.

This part of the dissertation presents results towards the study the two last cases, that leave no fixed loci and assess whether they correspond to sensible orientifolds in string theory. It consists of a single chapter that was published as [GVMPT21], in collaboration with Eduardo García-Valdecasas, Shani Meynet and Antoine Pasternak. We will argue that only the glide reflection leads to SUSY preserving orientifolds, while the shift is always breaking it. Moreover, we will show how the orientifold projection corresponding to a glide reflection has remarkable properties. Not only, the projected theory always has a conformal fixed point, but also admits, in some cases, a non-trivial RG-flow described by a cascade of Seiberg dualities, analogous to the one of the conifold [Sei95, KS00a].

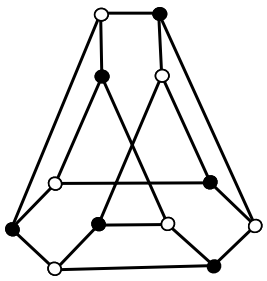


Figure 12.8: The Franklin graph

The orientifold theories described as the quotient of a brane tiling by a glide reflection correspond to bipartite maps on the Klein bottle, thus answering the question of their significance [FV06]. In particular, the Franklin graph depicted above and advertised in [FV06] as one of the simplest example of such a bipartite map on the Klein bottle has a natural interpretation as a glide orientifold of a $\mathbb{C}^3/\mathbb{Z}_{12}$-singularity.

Chapter 13

Dimers in a bottle

The organization of this chapter is as follows. In Section 13.1 we review the basic tools of dimer models and orientifolds, putting them in the context of torus involutions to find the missing cases. In Section 13.2 we describe glide orientifolds starting from orbifolds and describing their general properties. The absence of fixed loci is tackled in Section 13.3, where we understand it to be dual to a pair of opposite sign orientifold planes and use T- and mirror duality to give a global picture. Finally, in Section 13.4 we study the action on the toric geometry through the Zig-Zag paths, allowing the study of fractional branes in the orientifolded theory. A proof of the non-existence of SUSY preserving shift orientifolds is also provided.

13.1 Torus involutions and Orientifolds

13.1.1 Orientifold projections and Dimers

String Theory admits sensible propagation on singular backgrounds. Most notable are orbifolds and orientifolds. These arise as suitable projections on the theory propagating in a smooth background. Such involutions are readily described in dimer models [FHK+07].

Let us focus on orientifold projections in our setup of D3-branes in type IIB, defined by modding out by the action $\Omega R(-1)^{F_L}$, Ω being worldsheet parity, R a geometric $\mathbb{Z}_2$ isometry of the CY_3 and F_L the left-moving fermion number in spacetime. Extended objects are located at the fixed point of the R action, the O-planes. They are non-dynamical objects with a tension and an RR charge as the ones of D-branes. The $\mathbb{Z}_2$ symmetry acts holomorphically on the internal coordinates, and as follows on the Kähler form J and the holomorphic 3-form Ω_3:

$$J \to J \quad \text{and} \quad \Omega_3 \to -\Omega_3\,, \tag{13.1}$$

where the $-$ sign is necessary in order for the O-plane to preserve some common supercharges with the D3-branes. The resulting gauge theory is obtained by looking at the projected open string spectrum. The orientifold projection on Chan-Paton factors is essentially free. Denote by λ the Chan-Paton matrix, the orientifold acts with a unitary matrix γ_Ω:

$$\Omega : \lambda \to \gamma_\Omega \lambda^T \gamma_\Omega^{-1}\,. \tag{13.2}$$

Orientifold projections on D-branes at singularities and their description on dimers were studied in [FHK+07]. In this framework, the orientifold projection corresponds to a $\mathbb{Z}_2$ involution acting on the torus that identifies faces, edges and vertices in an appropriate way, equivalently it can be described as an involution of the brane tiling seen as a bipartite map. The authors studied involutions with fixed loci (see Figure 13.1 for examples) resulting in a set of rules needed to construct the projected theory that we have already described in Section 7.6, and that we now summarize quickly.

1. Self-identified faces project to SO/USp groups, depending on the O-plane charge, $+$ or $-$ respectively. All other faces are identified with their image, merging to one SU group.

2. Every edge on top of a fixed locus becomes a symmetric or antisymmetric tensor (or their conjugate), depending on the O-plane charge, $+$ or $-$ respectively. The remaining edges are identified with their images, merging to bifundamental fields. More concretely, bifundamentals are identified as $(\,\overline{\square}_{\,i},\,\square_{\,j}) \sim (\,\overline{\square}_{\,j'},\,\square_{\,i'}) \to (\,\overline{\square}_{\,i},\,\square_{\,j})$, where i', j' are the images of gauge groups i, j.

(a) (b)

Figure 13.1: (a) Orientifold of $\mathbb{C}^2/\mathbb{Z}_2$ with fixed points. (b) Orientifold of $\mathbb{C}^2/\mathbb{Z}_2$ with fixed lines.

3. The superpotential is found upon projection of the fields.

Before moving on to the next section, we present in detail two examples of orientifold projections.

Fixed Points. In an orientifold of this type, there are four fixed points in a unit cell. In order to preserve SUSY, their signs must satisfy the so-called *sign rule*: their product must be $(-1)^{n_W/2}$ where n_W is the number of superpotential terms.

In the example of Figure 13.1a, we chose the signs $(-+-+)$, starting with the fixed point at the origin of the unit cell and going clockwise. We have that face 1 is identified with face 2, meaning that the resulting theory will have only one gauge group $SU(N)$. The bifundamental fields are identified as follows

$$\begin{gathered} Y_{12} \sim Y_{12} \to \begin{array}{|c|}\hline \\ \hline \\ \hline \end{array}\,, \; X_{21} \sim X_{21} \to \square\square \\ Y_{21} \sim Y_{21} \to \begin{array}{|c|}\hline \\ \hline \\ \hline \end{array}\,, \; X_{12} \sim X_{12} \to \overline{\square\square} \\ Z_{11} \sim Z_{22} \to \text{Adj}\,. \end{gathered}$$

and the superpotential is given by

$$W = X_{12}Y_{21}Z_{11} - X_{21}Y_{12}Z_{11}\,, \tag{13.3}$$

where we implicitly take a trace over gauge indices.

To be sure that this projection preserves some supersymmetry, we need to check the action of the involution on Ω_3. To do so, we compute the mesonic moduli space of our theory, which correspond to the singularity D3-branes are probing. Mesonic operators are given by

$$\begin{array}{ccccccc} x & = & X_{12}X_{21}\,, & y & = & Y_{12}Y_{21} \\ w_1 & = & Y_{12}X_{21}\,, & w_2 & = & Y_{21}X_{12} \\ z_1 & = & Z_{11}\,, & z_2 & = & Z_{22}\,. \end{array} \tag{13.4}$$

F-term equations impose $w_1 = w_2 = w$ and $z_1 = z_2 = z$, and the classical relation between the fields gives $xy = w_1 w_2 = w^2$. Thus, the mesonic moduli space is the symmetric product of N copies of the A_1 singularity, $xy = w^2$, where N is the number of probe D3-branes. The three form, Ω_3, can be easily computed using the Poincaré residue formula:

$$\Omega_3 = \text{Res}\frac{\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}w \wedge \mathrm{d}z}{w^2 - xy} = \frac{\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z}{2w}\,. \tag{13.5}$$

Under the involution, the fields are mapped in the following way

$$\begin{gathered} x \to x\,, \quad y \to y\,, \\ w \to -w\,, \quad z \to z\,, \end{gathered}$$

where the sign taken by a meson is given by the product of the fixed point charges it crosses. The orientifold action on the holomorphic 3-form is thus odd, $\Omega_3 \to -\Omega_3$, meaning that the O-plane in compatible with the supersymmetry charges preserved by the D3-branes. It is easy to see that sign configuration not respecting the sign rule are not supersymmetric.

Fixed Lines. In the example of Figure 13.1b, we have two fixed lines, each one coming with a sign, + or −, which is unconstrained. We chose to assign − to the bottom line and + to the other. The faces are self-identified, leading to a gauge group $USp(N_1) \times SO(N_2)$. The identification of fields gives

$$\begin{aligned} Y_{12} \sim X_{21} \to Q^1_{12} \,, \quad X_{12} \sim Y_{21} \to Q^2_{12} \\ Z_{11} \sim Z_{11} \to \begin{array}{|} \hline \\ \hline \end{array} \,, \quad Z_{22} \sim Z_{22} \to \square\square \,. \end{aligned} \tag{13.6}$$

and the superpotential is given by

$$W = (Q^2_{12}Q^{2T}_{12} - Q^1_{12}Q^{1T}_{12})Z_{11} + (Q^{2T}_{12}Q^2_{12} - Q^{1T}_{12}Q^1_{12})Z_{22} \,. \tag{13.7}$$

The mesons are the same as in the previous example, since the geometry is the same, but the action of the orientifold is different and given by

$$\begin{gathered} x \leftrightarrow y \,, \\ w \to -w \,, \quad z \to -z \,, \end{gathered}$$

where the fixed line exchanges two mesons and introduces a sign to the self-mapped mesons given by the product of the signs of the two fixed lines crossed. We can again see that the SUSY condition is respected

$$\Omega_3 = \frac{\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z}{2w} \to \frac{\mathrm{d}y \wedge \mathrm{d}x \wedge \mathrm{d}z}{2w} = -\Omega_3 \,. \tag{13.8}$$

In particular, we see that the signs of the fixed lines play no role in the last relation.

13.1.2 Torus involutions

There are five inequivalent non-trivial smooth involutions [Dug19], i.e. involutive diffeomorphisms, on a torus[1]. Three of them have a fixed locus and the two others do not. To list all of them we consider a square torus, with complex structure[2] $\tau = i$. We take z as the complex coordinate on the torus, the periodicity condition is $z \sim z + m + ni$, with $m, n \in \mathbb{Z}$. The involutions are given by:

1. Two fixed lines: $z \to \bar{z}$. The fixed loci are two parallel lines located at $\mathrm{Im}(z) = 0, 1/2$ along the real axis. Under this involution the torus is projected to an annulus.

2. Single fixed line: $z \to i\bar{z}$. The fixed line is $\mathrm{Re}(z) = \mathrm{Im}(z)$, corresponding to a diagonal line of the unit cell. The resulting surface is a Moebius strip.

3. Fixed points: $z \to -z$. In this case we have four fixed points, $z = 0, 1/2, i/2$ and $(1 + i)/2$. The resulting topology is that of a sphere with four orbifold half-points.

4. Glide reflection: $z \to \bar{z} + 1/2$. There are no fixed locus. The resulting topology is that of a Klein bottle.

5. Shift: $z \to z + 1/2$. Again, the involution has no fixed locus. The torus is projected to another torus.

As already mentioned, 1, 2 and 3 are involutions with fixed loci correspond to orientifold operations already studied in the literature. In this paper, we will focus on 4, the glide reflection, studying the consistency of such projection and its properties. Regarding involution 5, we will show that the shift is not compatible with the required properties to preserve supersymmetry.

Let us conclude this section with few comments. First, involutions with fixed loci teach us that if the involution is holomorphic, $z \to f(z)$, nodes in the dimer are mapped to nodes of opposite color, while if it is antiholomorphic, $z \to f(\bar{z})$, nodes are mapped to nodes of the same color. This is a requirement from the orientifold mapping of chiral superfields. It gives us a hint for the unexplored involutions. Indeed, we expect 5 to be consistent with an orientifold identification only if nodes are mapped to nodes of the opposite color, while 4 would be consistent only if the mapping is between vertices of the same

[1] They are classified by the topology of their orbit set which is always one of the parabolic 2-orbifolds listed in [Thu02].

[2] We are interested only in smooth involutions, the complex structure doesn't play any role in the analysis, thus we fixed it to a handy value. The use of complex coordinates will be useful for later observations.

color. Second, we stress that the involution should be not only a symmetry for the torus, but also for the embedded dimer model, i.e. an involution symmetry of the bipartite map. In particular, a generic fundamental cell for a dimer model has the shape of a parallelogram. The symmetry may be present in the abstract graph, but in order to be shown explicitly, consider the case of say 2, one has to deform the embedding in such a way that the resulting fundamental cell is now a rhombus, displaying a symmetry with respect to one of the diagonals. From this observation we conclude that in order to display a glide symmetry, the fundamental cell must be a rectangle. Third, a $\mathbb{Z}_2$ glide reflection with diagonal axis is described by the map $z \to i\bar{z} + (1+i)/2$ which has $\mathrm{Re}(z) = \mathrm{Im}(z) + 1/2$ as fixed line, hence they are nothing else than reflections about a diagonal axis. In particular, they do not correspond to a class of smooth involutions not listed above.

Even if we can deform the embedding to make the involution explicit, it is possible that the model can be endowed with extra structures, capturing some physical properties. For example, isoradial embeddings described in [HV07] encode the R-charges of the fields. In this paper, though, we are not interested in these particular cases.

13.2 Glide Orientifolds

In this section we investigate glide reflection orientifolds. We start with orbifold examples, motivating our results in the dimer from the open string projection on the Chan-Paton indices. We also explicitly check that it preserves supersymmetry, in particular, it acts on the CY 3-form as $\Omega_3 \to -\Omega_3$. We extend our results to orbifolds of the conifold, considering the cascade in the presence of deformation fractional branes. Finally, we discuss anomalies, or rather their absence, and conformality in the presence of these orientifolds.

13.2.1 Orbifold $\mathbb{C}^2/\mathbb{Z}_2$

We consider the recipe directly applied in the dimer and then check that it is indeed predicted by open-string computation.

(a) (b)

Figure 13.2: (a) Dimer diagram for the orbifold $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$. The unit cell and the reflection axis are depicted in blue and red respectively. (b) The Klein bottle we obtain with the orientifold projection.

Projection on the dimer model. We present in Figure 13.2 the dimer for the orbifold $\mathbb{C}^2/\mathbb{Z}_2$ where the glide reflection is a combined operation of a horizontal shift by one half of the length of the unit cell followed by a reflection with respect to the dashed red horizontal axis. Nodes are mapped to nodes of the same color, as we want from the analysis in Section 13.1.2. Note that this operation leaves no fixed loci in the unit cell. The projected theory is embedded in a Klein Bottle drawn on one half of the original unit cell, as illustrated in Figure 13.2b.

The edge X_{12} is identified with Y_{12}, X_{21} with Y_{21} and Z_{11} with Z_{22}. Following the rules summarized in Section 13.1, the resulting theory has gauge group $SU(N)_1$ with matter content given by two tensors[3]

[3]The two tensors are of the form ($\overline{\square}_1$, $\overline{\square}_1$) and ($\square_1$, $\square_1$).

and one adjoint field. Note that the tensor fields are not in an irreducible representation, so we split them in their symmetric and antisymmetric parts;

$$\begin{aligned} \mathcal{X}_{S,A} &= \overline{\square\square}_{\,1},\ \overline{\square}_{\,1}\,, \\ \mathcal{Y}_{S,A} &= \square\square_{\,1},\ \square_{\,1}\,, \\ \mathcal{Z} &= \mathrm{Adj}_1\,. \end{aligned} \tag{13.9}$$

The superpotential is obtained by explicitly projecting the original one and keeping half of the terms,

$$W = \mathcal{X}\mathcal{Y}\mathcal{Z}^T - \mathcal{Y}\mathcal{X}\mathcal{Z} = \mathcal{X}_A\mathcal{Y}_S\mathcal{Z} - \mathcal{X}_S\mathcal{Y}_A\mathcal{Z}\,. \tag{13.10}$$

In a SUSY-preserving orientifold in type IIB, the holomorphic 3-form must map to minus itself. This is easy to check by noting that the orientifold action on the mesons is

$$x \leftrightarrow y \qquad w \to w \qquad z \to z\,. \tag{13.11}$$

The action on the 3-form is then

$$\Omega_3 = \frac{\mathrm{d}x\wedge\mathrm{d}y\wedge\mathrm{d}z}{2w} \to \frac{\mathrm{d}y\wedge\mathrm{d}x\wedge\mathrm{d}z}{2w} = -\Omega_3\,. \tag{13.12}$$

It is also clear from the matter content and the first equality of Equation (13.10) that the gauge theory preserves $\mathcal{N}=2$ supersymmetry[4].

It is worth noting that the theory, unlike many examples of projections with fixed loci, is free from any local gauge anomaly, regardless of the gauge group rank. Although this example is rather trivial, we will see that this feature is general and related to tensor fields being absent or coming in pairs, symmetric and antisymmetric, cancelling each other's contribution to the anomaly cancellation conditions (ACC). We also note that the projected theory is actually conformal. Indeed, the β-function of the gauge group can be shown to be zero. The fact that these orientifolds naturally lead to SCFT's will be discussed in Section 13.2.4.

Open string projection. We now consider the orientifold projection on the Chan-Paton indices of the open string spectrum. For D-branes localized on the $\mathbb{C}^2/\mathbb{Z}_2\times\mathbb{C}$ singularity the open string spectrum is obtained by promoting the flat space one to $2N\times 2N$ matrices with a restricted set of non-zero entries:

$$A_\mu = \begin{pmatrix} A_{1\mu} & 0 \\ 0 & A_{2\mu} \end{pmatrix}, \quad \Phi_1 = \begin{pmatrix} 0 & X_{12} \\ X_{21} & 0 \end{pmatrix}, \quad \Phi_2 = \begin{pmatrix} 0 & Y_{12} \\ Y_{21} & 0 \end{pmatrix}, \quad \Phi_3 = \begin{pmatrix} Z_{11} & 0 \\ 0 & Z_{22} \end{pmatrix}, \tag{13.13}$$

where the gauge group is $SU(N)_1\times SU(N)_2$ and matther fields transform in the following representations,

$$X_{ij}, Y_{ij} = (\ \square_{\,i},\ \overline{\square}_{\,j}), \quad Z_{ii} = \mathrm{Adj}_i\,. \tag{13.14}$$

Decomposing the $\mathbb{C}^3$ fields the orbifold superpotential becomes,

$$\begin{aligned} W &= [\Phi_1,\Phi_2]\,\Phi_3 \\ &= X_{12}Y_{21}Z_{11} - Y_{21}X_{12}Z_{22} + X_{21}Y_{12}Z_{22} - Y_{12}X_{21}Z_{11}\,, \end{aligned} \tag{13.15}$$

where an overall trace over gauge indices is understood.

A general orientifold projection on the $\mathbb{C}^3$ fields acts as,

$$A_\mu = -\gamma_\Omega A_\mu^T \gamma_\Omega^{-1}\,, \tag{13.16}$$

$$\Phi_i = R_{ij}\gamma_\Omega \Phi_j^T \gamma_\Omega^{-1}\,, \tag{13.17}$$

where γ_Ω is a $2N\times 2N$ matrix acting on gauge group (Chan-Paton) indices and R_{ij} acts on space indices i,j running from 1 to 3. Different choices for these matrices lead to different orientifold projections. In order to reproduce the glide reflection orientifold, we specifically choose

$$\gamma_\Omega = \begin{pmatrix} 0 & \mathbf{1}_N \\ \mathbf{1}_N & 0 \end{pmatrix}, \quad \text{and} \quad R = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{13.18}$$

[4]The attentive reader might have noticed that this orientifolded theory is identical to the one obtained with fixed points in Section 13.1.1, although the involution acts differently on the coordinates. This is however an artifact of the orbifold $\mathbb{C}^2/\mathbb{Z}_2$ since glide reflections will not provide tensors in general.

so that Φ_1 and Φ_2 coordinates are exchanged by the orientifold. Equation (13.16) translates into

$$A_{1\mu} = -A_{2\mu}^T \,, \tag{13.19}$$

which tells us that the two gauge groups are now identified as one $SU(N)_1$ in the orientifolded theory. Equation (13.17) maps the superfields in the following way:

$$\begin{array}{rcccl} X_{12} & = & Y_{12}^T & \equiv & \mathcal{X}_{A,S} \,, \\ Y_{21} & = & X_{21}^T & \equiv & \mathcal{Y}_{A,S} \,, \\ Z_{11} & = & Z_{22}^T & \equiv & \mathcal{Z} \,. \end{array} \tag{13.20}$$

We recognise the same field content of the theory obtained with the dimer technique. It is easy then to show that we recover the superpotential advertised in Equation (13.10) (up to an irrelevant numerical factor). We thus conclude that the glide reflection on the dimer reproduces the orientifold projection we just computed in string theory.

In the following, we discuss the dimer construction in more involved examples. It is clear that not all dimer models have the required symmetry, and in Section 13.4 we provide a necessary condition for a given toric CY_3 to admit a glide reflection directly from its toric diagram.

13.2.2 More orbifold examples

The previous example has so much symmetry that it could be misleading. Let us start our journey to less symmetric theories by considering $\mathbb{C}^2/\mathbb{Z}_4$, whose dimer model and relevant involution we present in Figure 13.3.

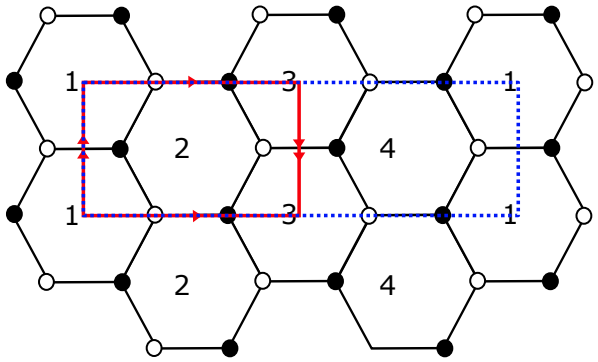


Figure 13.3: Dimer diagram for the orbifold $\mathbb{C}^2/\mathbb{Z}_4$. The unit cell is depicted in blue and we show in red the Klein bottle obtained from the orientifold projection.

From the four initial gauge groups, only two of them are kept after the projection, $SU(N_1)_1 \times SU(N_2)_2$. The surviving fields are

$$\begin{array}{lll} \mathcal{X}_{12} = (\,\overline{\square}_1,\, \square_2)\,, & \mathcal{X}_{21} = (\,\overline{\square}_2,\, \overline{\square}_1)\,, & \mathcal{Y}_{21} = (\,\overline{\square}_2,\, \square_1)\,, \\ \mathcal{Y}_{12} = (\,\square_1,\, \square_2)\,, & \mathcal{Z}_{11} = \mathrm{Adj}_1\,, & \mathcal{Z}_{22} = \mathrm{Adj}_2\,. \end{array} \tag{13.21}$$

and the resulting superpotential is found to be

$$W = \mathcal{X}_{12}\mathcal{Y}_{21}\mathcal{Z}_{11} - \mathcal{Y}_{21}\mathcal{X}_{12}\mathcal{Z}_{22} + \mathcal{X}_{21}\mathcal{Y}_{12}\mathcal{Z}_{22} - \mathcal{Y}_{12}\mathcal{X}_{21}\mathcal{Z}_{11}^T \,. \tag{13.22}$$

Note that despite its similarities with the orbifold $\mathbb{C}^2/\mathbb{Z}_2$ (without orientifold), this model has a different matter content, which cannot be obtained from dimer models.

The perturbative string computation goes as follows. The open sector of strings on the orbifold before

the orientifold projection is:

$$A_\mu = \begin{pmatrix} A_{1\mu} & 0 & 0 & 0 \\ 0 & A_{2\mu} & 0 & 0 \\ 0 & 0 & A_{3\mu} & 0 \\ 0 & 0 & 0 & A_{4\mu} \end{pmatrix}, \quad \Phi_1 = \begin{pmatrix} 0 & X_{12} & 0 & 0 \\ 0 & 0 & X_{23} & 0 \\ 0 & 0 & 0 & X_{34} \\ X_{41} & 0 & 0 & 0 \end{pmatrix},$$
$$\Phi_2 = \begin{pmatrix} 0 & 0 & 0 & Y_{14} \\ Y_{21} & 0 & 0 & 0 \\ 0 & Y_{32} & 0 & 0 \\ 0 & 0 & Y_{43} & 0 \end{pmatrix}, \quad \Phi_3 = \begin{pmatrix} Z_{11} & 0 & 0 & 0 \\ 0 & Z_{22} & 0 & 0 \\ 0 & 0 & Z_{33} & 0 \\ 0 & 0 & 0 & Z_{44} \end{pmatrix}. \tag{13.23}$$

The appropriate orientifold projection, defined as in Equations (13.16) and (13.17), is given by

$$\gamma_\Omega = \begin{pmatrix} 0 & 0 & \mathbf{1}_N & 0 \\ 0 & 0 & 0 & \mathbf{1}_N \\ \mathbf{1}_N & 0 & 0 & 0 \\ 0 & \mathbf{1}_N & 0 & 0 \end{pmatrix}, \quad \text{and} \quad R = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{13.24}$$

It gives the following identification of gauge bosons

$$A_{1\,\mu} = -A_{3\,\mu}^T \quad \text{and} \quad A_{2\,\mu} = -A_{4\,\mu}^T\,, \tag{13.25}$$

the resulting gauge group is $SU(N)_1 \times SU(N)_2$. The matter content follows from

$$\begin{array}{rclcll} X_{12} & = & Y_{43}^T & \equiv & \mathcal{X}_{12} & \in & (\,\overline{\square}_1,\,\square_2\,)\,, \\ X_{23} & = & Y_{14}^T & \equiv & \mathcal{X}_{21} & \in & (\,\overline{\square}_2,\,\overline{\square}_1\,)\,, \\ Y_{21} & = & X_{34}^T & \equiv & \mathcal{Y}_{21} & \in & (\,\overline{\square}_2,\,\square_1\,)\,, \\ Y_{32} & = & X_{41}^T & \equiv & \mathcal{Y}_{12} & \in & (\,\square_1,\,\square_2\,)\,, \\ Z_{11} & = & Z_{33}^T & \equiv & \mathcal{Z}_{11} & \in & \mathrm{Adj}_1\,, \\ Z_{22} & = & Z_{44}^T & \equiv & \mathcal{Z}_{22} & \in & \mathrm{Adj}_2\,. \end{array} \tag{13.26}$$

One can check that the superpotential is the one advertised in Equation (13.22).

The mapping of the mesons is the same as for $\mathbb{C}^2/\mathbb{Z}_2$ so that the holomorphic 3-form transforms as follows:

$$\Omega_3 = \frac{\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z}{4w^3} \to \frac{\mathrm{d}y \wedge \mathrm{d}x \wedge \mathrm{d}z}{4w^3} = -\Omega_3\,, \tag{13.27}$$

and hence suggests that our projection is indeed supersymmetric and the resulting gauge theory preserves $\mathcal{N} = 2$ supersymmetry. Note that the usual orientifold techniques in the dimer, fixed points and line(s), are not able to reproduce it.

Our observations make it clear that any orbifold $\mathbb{C}^2/\mathbb{Z}_{2n} \times \mathbb{C}$ will admit a glide reflection, for any integer n. More general orbifolds, such as $\mathbb{C}^3/\mathbb{Z}_n$ or $\mathbb{C}^3/\mathbb{Z}_p \times \mathbb{Z}_q$, can also enjoy the glide reflections, see an example in Figure 13.6a. In Section 13.4 we will discuss the general geometric condition a singularity should meet in order to admit such orientifold.

$\mathcal{N} = 2$ fractional branes. Let us briefly comment on the fractional branes of the orientifolded theory [FHSU06]. The glide orientifold of $\mathbb{C}^2/\mathbb{Z}_4$ is free of local gauge anomalies for any rank N_1 and N_2. Hence, it has a fractional brane. We find that it is an $\mathcal{N} = 2$ fractional brane corresponding to a subset of the $\mathcal{N} = 2$ fractional branes of the parent theory. In Section 13.4.2 we will discuss this fact in detail.

13.2.3 Conifold-like singularities

As we will explain in Section 13.4, the conifold $\mathcal{C}$ itself does not admit a glide reflection, but conifold-like singularities like its orbifold $\mathcal{C}/\mathbb{Z}_2$ or the zeroth Hirzebruch surface F_0 do. We now study those examples in turn.

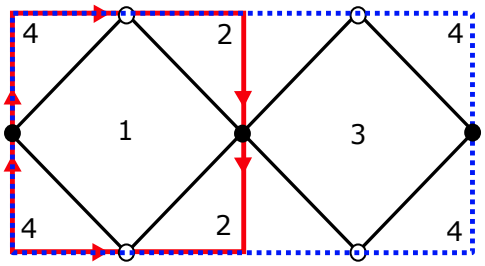


Figure 13.4: Dimer diagram for the orbifold of the conifold $\mathcal{C}/\mathbb{Z}_2$. The unit cell is depicted in blue and we show in red the Klein bottle obtained from the orientifold projection.

Non-chiral orbifold of the conifold $\mathcal{C}/\mathbb{Z}_2$. The dimer model and the glide orientifold of $\mathcal{C}/\mathbb{Z}_2$ are shown in Figure 13.4.

The resulting gauge theory has gauge group $SU(N_1) \times SU(N_2)$ with matter content given by

$$\begin{array}{rclcrcl} A & = & (\,\overline{\square}_1,\,\overline{\square}_2\,)\,, & & B & = & (\,\square_1,\,\square_2\,)\,, \\ C & = & (\,\overline{\square}_1,\,\square_2\,)\,, & & D & = & (\,\square_1,\,\overline{\square}_2\,)\,, \end{array} \tag{13.28}$$

Note in passing that the ACC do not impose any constraint on the ranks, so that N_1 and N_2 may be chosen independently. The superpotential reads

$$W = ABCD - BAC^T D^T\,. \tag{13.29}$$

The more detailed computation runs as follows. Consider a non-chiral orbifold of the conifold such as $\mathcal{C}/(\mathbb{Z}_l \times \mathbb{Z}_m)$. The general action is given by

$$\begin{aligned} &\gamma_g V_{1,2} \gamma_g^{-1} = V_{1,2} \\ &\gamma_g A_1 \gamma_g^{-1} = e^{2\pi i/l} A_1\,, \quad \gamma_g A_2 \gamma_g^{-1} = A_2 \\ &\gamma_g B_1 \gamma_g^{-1} = e^{-2\pi i/l} B_1\,, \quad \gamma_g B_2 \gamma_g^{-1} = B_2\,, \end{aligned} \tag{13.30}$$

and

$$\begin{aligned} &\gamma_g V_{1,2} \gamma_g^{-1} = V_{1,2} \\ &\gamma_g A_1 \gamma_g^{-1} = e^{2\pi i/m} A_1\,, \quad \gamma_g A_2 \gamma_g^{-1} = A_2 \\ &\gamma_g B_1 \gamma_g^{-1} = B_1\,, \quad \gamma_g B_2 \gamma_g^{-1} = e^{-2\pi i/m} B_2\,, \end{aligned} \tag{13.31}$$

where $V_{1,2}$ are the two adjoint vectors related to the gauge groups. In the case of our first example, $\mathcal{C}/\mathbb{Z}_2$, the action gives the following fields

$$\begin{aligned} V_1 &= \begin{pmatrix} V_1 & 0 \\ 0 & V_3 \end{pmatrix}, \qquad V_2 = \begin{pmatrix} V_2 & 0 \\ 0 & V_4 \end{pmatrix}, \qquad A_1 = \begin{pmatrix} 0 & A_{14} \\ A_{32} & 0 \end{pmatrix}, \\ A_2 &= \begin{pmatrix} A_{12} & 0 \\ 0 & A_{34} \end{pmatrix}, \quad B_1 = \begin{pmatrix} B_{21} & 0 \\ 0 & B_{43} \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 & B_{23} \\ B_{41} & 0 \end{pmatrix}, \end{aligned} \tag{13.32}$$

with a superpotential given by

$$W = A_1 B_1 A_2 B_2 - A_1 B_2 A_2 B_1\,. \tag{13.33}$$

We consider the following orientifold projection in order to reproduce the glide projection.

$$\begin{aligned} V_{1,2} &= -\gamma_\Omega V_{1,2}^T \gamma_\Omega^{-1}\,, \\ A_{1,2} &= \gamma_\Omega B_{1,2}^T \gamma_\Omega^{-1}\,, \end{aligned} \tag{13.34}$$

with

$$\gamma_\Omega = \begin{pmatrix} 0 & \mathbf{1}_N \\ \mathbf{1}_N & 0 \end{pmatrix} . \tag{13.35}$$

The action on mesons, $x = (A_1B_1)^2$ $y = (A_2B_2)^2$, $z = A_1B_2$ and $w = A_2B_1$, is

$$\begin{aligned} & x \leftrightarrow y\,, \qquad z \to z\,, \qquad w \to w\,, \\ & \Omega_3 = \frac{\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}z}{2wz^2} \to \Omega_3' = \frac{\mathrm{d}y \wedge \mathrm{d}x \wedge \mathrm{d}z}{2wz^2} = -\Omega_3 \end{aligned} \tag{13.36}$$

which means that the action preserves supersymmetry on the branes.

The gauge group is $SU(N_1)_1 \times SU(N_2)_2$ and matter content given by

$$\begin{array}{rcl} A_{14} = B_{23}^T \equiv & A & = \; (\,\overline{\square}_1,\, \overline{\square}_2\,)\,, \\ B_{41} = A_{32}^T \equiv & B & = \; (\,\square_1,\, \square_2\,)\,, \\ A_{12} = B_{43}^T \equiv & C & = \; (\,\overline{\square}_1,\, \square_2\,)\,, \\ B_{21} = A_{34}^T \equiv & D & = \; (\,\square_1,\, \overline{\square}_2\,)\,, \end{array} \tag{13.37}$$

with superpotential

$$W = ABCD - BAC^TD^T\,. \tag{13.38}$$

Zeroth Hirzebruch surface $\mathbf{F_0}$. We show the dimer model and the glide orientifold of F_0 in Figure 13.5.

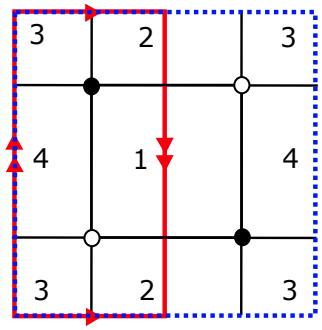


Figure 13.5: Dimer diagram for the Hirzebruch surface F_0. The unit cell is depicted in blue and we show in red the Klein bottle obtained from the orientifold projection.

We take the following actions on the fields

$$\begin{aligned} \gamma_g V_{1,2} \gamma_g^{-1} &= V_{1,2} \\ \gamma_g A_1 \gamma_g^{-1} &= -A_1 \\ \gamma_g A_2 \gamma_g^{-1} &= -A_2 \\ \gamma_g B_1 \gamma_g^{-1} &= B_1 \\ \gamma_g B_2 \gamma_g^{-1} &= B_2\,, \end{aligned} \tag{13.39}$$

leading to

$$\begin{aligned} V_1 &= \begin{pmatrix} V_1 & 0 \\ 0 & V_3 \end{pmatrix}, & V_2 &= \begin{pmatrix} V_2 & 0 \\ 0 & V_4 \end{pmatrix}, & A_1 &= \begin{pmatrix} 0 & A_{14}^1 \\ A_{32}^1 & 0 \end{pmatrix}, \\ A_2 &= \begin{pmatrix} 0 & A_{14}^2 \\ A_{32}^2 & 0 \end{pmatrix}, & B_1 &= \begin{pmatrix} B_{21}^1 & 0 \\ 0 & B_{43}^1 \end{pmatrix}, & B_2 &= \begin{pmatrix} B_{21}^2 & 0 \\ 0 & B_{43}^2 \end{pmatrix}. \end{aligned}$$

The orientifold action maps $1 \to 4$ and $2 \to 3$, it can be summarized as

$$\begin{aligned} V_{1,2} &= -\gamma_\Omega V_{1,2}^T \gamma_\Omega^{-1} \,, \\ A_1 &= \gamma_\Omega A_2^T \gamma_\Omega^{-1} \,, \\ B_1 &= \gamma_\Omega B_2^T \gamma_\Omega^{-1} \,, \end{aligned} \tag{13.40}$$

with

$$\gamma_\Omega = \begin{pmatrix} \mathbf{1}_N & 0 \\ 0 & \mathbf{1}_N \end{pmatrix} . \tag{13.41}$$

The resulting gauge group is $SU(N)_1 \times SU(N)_2$ and the matter content is given by

$$\begin{array}{rcccl} A_{14}^1 = A_{14}^{2T} & \equiv & U_{S,A} & = & \overline{\square\square}_{\,1},\ \overline{\begin{array}{c}\square\\\square\end{array}}_{\,1}\,, \\ A_{32}^2 = A_{32}^{2T} & \equiv & Z_{S,A} & = & \square\square_{\,2},\ \begin{array}{c}\square\\\square\end{array}_{\,2}\,, \\ B_{21}^1 = B_{43}^2 & \equiv & X & = & (\,\square_{\,1},\ \overline{\square}_{\,2})\,, \\ B_{21}^2 = B_{43}^1 & \equiv & Y & = & (\,\square_{\,1},\ \overline{\square}_{\,2})\,, \end{array} \tag{13.42}$$

with the following superpotential,

$$W = X U_S Y^T Z_A - X^T Z_S Y U_A \,. \tag{13.43}$$

In order to compute the action the 3-form, we compute the equations defining the singularity using the geometrical approach described in Section 13.4.3. The singularity is described by the following equations in $\mathbb{C}^9$

$$\begin{aligned} z_1 z_3 = z_2 z_4 = z_0^2 \,, \\ z_1 z_2 = z_5^2 \,, \qquad z_2 z_3 = z_7^2 \,, \\ z_1 z_4 = z_6^2 \,, \qquad z_3 z_4 = z_8^2 \,. \end{aligned} \tag{13.44}$$

The action on mesons is

$$z_1 \leftrightarrow z_2 \,, \qquad z_3 \leftrightarrow z_4 \,, \qquad z_6 \leftrightarrow z_7 \tag{13.45}$$

while all other coordinates are invariant. The action on the 3-form is

$$\Omega_3 = \mathrm{Res} \frac{\mathrm{d}z_1 \wedge \mathrm{d}z_2 \wedge \mathrm{d}z_3 \wedge \mathrm{d}z_4 \wedge \mathrm{d}z_5 \wedge \mathrm{d}z_6 \wedge \mathrm{d}z_7 \wedge \mathrm{d}z_8 \wedge \mathrm{d}z_0}{\prod_i P_i} \to -\Omega_3 \,, \tag{13.46}$$

since the polynomials are invariant and in the numerator we are exchanging three pairs of coordinates, resulting in an overall minus sign.

Deformation fractional branes. We have seen that the $\mathcal{C}/\mathbb{Z}_2$ glide reflection admits fractional brane since the ranks of the two gauge groups may be chosen freely. It is in fact a deformation brane [FHSU06, But06] of the parent theory that survives the orientifold projection, in the precise sense described in [ABF$^+$21a]. A natural question is whether such fractional branes may trigger a non-trivial RG-flow giving rise to a cascade of Seiberg dualities [Sei95, KS00a], and it turns out it does.

For a generic choice of ranks, $SU(N+M)_1 \times SU(N)_2$, one finds that the gauge theory has a non-trivial RG-flow and $SU(N+M)_1$ goes more rapidly to strong coupling as we approach the infrared regime of the theory:

$$\beta_1 = 3M \,, \quad \beta_2 = -3M \,. \tag{13.47}$$

The mesons of the first gauge group are

$$M_1 = BA \quad M_2 = BC \,, \quad M_3 = C^T D^T \quad \text{and} \quad M_4 = DA \,, \tag{13.48}$$

and one thus finds that this gauge theory is Seiberg dual [Sei95] to $SU(N-M)_1 \times SU(N)_2$ with a matter content given by the mesons M_1, M_2, M_3 and M_4 in addition to the following list of bifundamental fields:

$$a = (\,\square_{\,1},\ \square_{\,2})\,, \quad b = (\,\overline{\square}_{\,1},\ \overline{\square}_{\,2})\,, \quad c = (\,\square_{\,1},\ \overline{\square}_{\,2}) \quad \text{and} \quad d = (\,\overline{\square}_{\,1},\ \square_{\,2})\,. \tag{13.49}$$

The superpotential is given by

$$\begin{aligned} W &= M_2M_4 - M_1M_3 + M_1ab + M_2cb + M_3d^Tc^T + M_4ad \\ &= abd^Tc^T - badc \end{aligned} \tag{13.50}$$

where the mesons have been integrated out using F-term relations.

The new gauge theory $SU(N-M)_1 \times SU(N)_2$ ends up with the same matter content and superpotential (up to an overall sign) as the initial $SU(N+M)_1 \times SU(N)_2$. This can be seen easily with the following mapping:

$$A \to b\,, \quad B \to a\,, \quad C \to d \quad \text{and} \quad D \to c\,. \tag{13.51}$$

The M deformation branes thus trigger a cascade of Seiberg dualities à la Klebanov-Strassler [KS00a]. In particular, for N being an integer multiple of M, we expect the cascade flow down to $SU(2M) \times SU(M)$ where the physics should the same as for the deformed conifold. Indeed, we can schematically define baryonic operators $\bar{\mathcal{B}} = \left[AC^T\right]^M$, $\mathcal{B} = \left[BD^T\right]^M$, and a $2M \times 2M$ squared matrix $\mathcal{M}$ in terms of the mesonic operators of Equation (13.48) that should obey a relation of the form

$$\det\mathcal{M} - \bar{\mathcal{B}}\mathcal{B} = \Lambda_{2M}^{4M}\,, \tag{13.52}$$

where Λ_{2M} is the strong coupling scale of $SU(2M)$. Going on the baryonic branch $\bar{\mathcal{B}} = \mathcal{B} = i\Lambda_{2M}^{2M}$, one finds that the mesons decouple, leaving a SYM $SU(M)$ dynamics displaying confinement and chiral symmetry breaking.

We will see later that it is a fact that the orbifolds of the conifold $\mathcal{C}/\mathbb{Z}_m \times \mathbb{Z}_n$ compatible with the glide projection preserve some of their deformation branes. The compatibility of fractional branes of the parent theory with the glide reflection is discussed in Section 13.4.

13.2.4 General properties

As we have seen, and since the glide reflection leaves no fixed loci, we don't expect any self-identified face (i.e. SO or USp gauge group) to show up in the dimer projection. This restricts the number of gauge groups of the parent theory to be even. A further consequence of not having fixed loci is that there are no self-identified bifundamentals, therefore, tensor matter, if present, always comes in antisymmetric-symmetric pairs, cancelling the contributions to the chiral anomaly. This is precisely what happens in the $\mathbb{C}^2/\mathbb{Z}_2$ orbifold, where two edges, charged under two identified groups, are identified, leading to a reducible two index tensor, which splits into the sum of a symmetric and an antisymmetric one. We now see how these facts translates in the absence of non-homogenous terms in the anomaly cancellation conditions, allowing always a solution to the latter, and how such projected theories are actually SCFTs.

Borrowing the notation of [ABF+21a], we know that the ACC matrix of the projected theory is deduced from that of the parent theory. Denote the latter as,

$$A = \left(\begin{array}{c|c|c} B_{11} & B_{12} & B_{13} \\ \hline B_{21} & B_{22} & B_{23} \\ \hline B_{31} & B_{32} & B_{33} \end{array}\right) \begin{array}{l} \Big\} i \\ \Big\} i+k \\ \Big\} a \end{array} \quad , \tag{13.53}$$

$$\underbrace{\phantom{B_{11}}}_{j} \quad \underbrace{\phantom{B_{12}}}_{j+k} \quad \underbrace{\phantom{B_{13}}}_{b}$$

where indices $i, j = 1, \ldots, k$ label the gauge groups surviving the orientifold projection and the corresponding entries represent the anomaly contribution of the field between faces i and j. Indices $i+k$ and $j+k$ represent gauge groups that are identified with i and j under the orientifold action, respectively. The a, b indices label the self identified gauge groups. Finally, the ACC system takes the form

$$A \cdot N = 0\,, \tag{13.54}$$

where N is a vector whose entries, $N_{(j|j+k|a)}$ are the ranks of the corresponding gauge group.

From what we said earlier, we know that $B_{\star 3} = B_{3\star} = B_{33} = 0$, since there are no self-identified gauge groups. Furthermore, we have no net contributions from tensors to the ACC, meaning that there are no non-homogenous terms in the projected theory ACC. From [ABF+21a], we know that the projected ACC can be written as

$$\overline{A} \cdot N = \left(\; B_{11} + B_{12} \; \right) \cdot N = 0 \,. \tag{13.55}$$

It is then easy to see that the all-equal-rank solution in the parent theory is still a solution. Indeed, a general solution for the orientifolded theory has a trivial part, corresponding to a stack of regular branes in the parent theory, and a non-trivial part, corresponding to "symmetric" fractional branes of the parent theory.

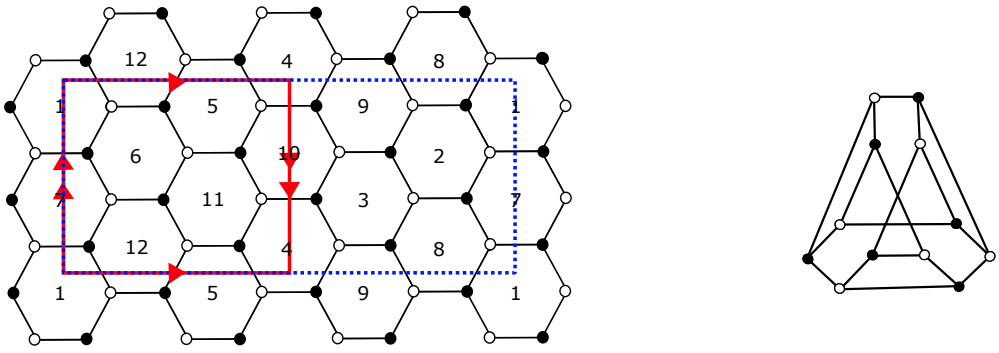
Fixed loci orientifolds have the remarkable property of producing, in general, non-conformal theories. However, this is not true for glide orientifolds. The theory they describe is an SCFT when the ranks of the gauge groups are all the same. This fact can be seen as follows, consider the β-function of the parent theory with N probes D-branes,

$$\beta_{SU(N)_i} = 3N - \sum_{i=1}^{n} \frac{N}{2}(1 - \gamma_i) = 0 \,, \tag{13.56}$$

where γ_i are the anomalous dimensions of the matter fields[5]. From this we can read the β-function of the projected theory whose general form is

$$\beta_{SU(N)_i} = 3N - \sum_{i=1}^{n} \frac{N + b_i}{2}(1 - \gamma_i) \,, \tag{13.57}$$

where the coefficients b_i vanish for fundamental fields and are ± 2 for, respectively, symmetric or anti-symmetric fields. If we assume that the anomalous dimensions of the fields are the same up to $1/N$ corrections and, since all tensors come in pairs of opposite parity, we see that the β-function of the gauge groups of the projected theory vanishes as long as all ranks are equal. This dovetails the fact that a Klein Bottle has zero Euler characteristic and, as explained in [FV06], such surfaces may embed a dimer model describing an SCFT[6]. Franco and Vegh pointed out that the Franklin graph would be a good candidate to be embedded in a Klein Bottle and host a SCFT not embedded in a torus. Indeed, it can be readily found via a glide reflection of $\mathbb{C}^3/\mathbb{Z}_{12}$, see Figure 13.6. This not only confirms their intuition, but it is, to the best or our knowledge, the first instance of such a construction within string theory.



(a) (b)

Figure 13.6: (a) Dimer diagram for the orbifold $\mathbb{C}^3/\mathbb{Z}_{12}$ with action $(1, 5, 6)$. The unit cell and the Klein bottle are depicted in blue and red respectively. (b) The Franklin graph.

[5] We consider Adj fields as couple of anti-fundamentals fields charged under the same gauge group.

[6] Other kind of surfaces obtained from orientifolds with fixed loci were found to accommodate SCFTs in [IY02, AMR20].

13.3 T-duals of the Glide Orientifold

13.3.1 Type IIA picture and the brane tiling

Fixed loci in the dimer have been related to actual orientifold planes in the physical realization of the dimer [FHK+06, IKY08]. In fact, one may consider the D3-branes probing a singularity with an orientifold and track the position of the orientifold in the ambient space to the fixed loci in the dimer through T-dualities. An immediate puzzle arises in the case of glide orientifolds, since there are no fixed loci on the torus, i.e. in the brane tiling. In this section we look at $\mathbb{C}^2/\mathbb{Z}_2$ and argue that these orientifolds, which have 8-dimensional fixed loci in the D3 picture (they are O7-planes), don't have a fixed locus in the tiling in the precise sense of [DP96, Wit98b]. In the latter reference, the shift action is deduced to be T-dual to a pair of opposite charge O-planes on a circle.

Let us again consider N D3-branes at the tip of a singular toric CY_3. As reviewed in the introduction, the dimer presented in Figure 13.2 is physically realized as a web of D5 and NS5-branes. It is obtained by T-duality along two of the three toric cycles of the toric variety. In particular along those corresponding to mesonic symmetries in the field theory, rather than R-symmetry. Focusing on the case at hand, $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$, one may take local coordinates such that x_7, x_9 correspond to the two toric cycles that are to be T-dualized. The D-brane configuration is then as in Table 13.1 which, after two T-duality should become that of Table 13.2. Note that we have avoided including an orientifold plane in the T-dual, as the dimer shift seems to suggest.

After T-duality one finds D5-branes wrapping the dual cycles with local coordinates x'_7, x'_9. These are in turn identified as the coordinates of the torus $\mathbb{T}^2$ where the 5-brane web lives.

	0	1	2	3	4	5	6	7	8	9
$\mathbb{C}^2/\mathbb{Z}_2$					×	×	×	×		
D3	×	×	×	×						
O7	×	×	×	×	×	×			×	×

Table 13.1: D3-branes sitting at the tip of $\mathbb{C}^2/\mathbb{Z}_2$ in the presence of O7-planes.

	0	1	2	3	4	5	6	7′	8	9′
NS5	×	×	×	×			–	– Σ	–	–
D5	×	×	×	×				×		×

Table 13.2: The brane tiling. Σ is the holomorphic curve in the 67′89′-space wrapped by the NS5-brane.

To study the location of the O-plane in the singular geometry, let us introduce the coordinates z_1, z_2 and z_3 of flat space $\mathbb{C}^3$. We define the coordinates of the variety transverse to the D3-branes, $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$, by constructing invariants under the orbifold action:

$$x = z_1^2\,, \quad y = z_2^2, \quad w = z_1 z_2, \quad \text{and} \quad z = z_3\,, \tag{13.58}$$

with the following relation holding,

$$xy = w^2\,. \tag{13.59}$$

As explained in Section 13.2.1, the orientifold action on the dimer implies that it acts on z_1, z_2, z_3 as $z_1 \leftrightarrow z_2$,. In terms of the orbifold invariant coordinates the orientifold action is then,

$$x \leftrightarrow y\,, \quad w \text{ and } z \text{ fixed.} \tag{13.60}$$

Thus, the orientifold plane extends on the surface defined by $x = y = t$, $t^2 = w^2$. From Equation (13.59) we read two toric $U(1)$ isometries of the orbifold:

$$\begin{array}{lllll} U(1)_\alpha: & x \to e^{i\alpha}x\,, & y \to e^{-i\alpha}y\,, & w \to w\,, \\ U(1)_\beta: & x \to e^{i\beta}x\,, & y \to e^{i\beta}y\,, & w \to e^{i\beta}w\,. \end{array} \tag{13.61}$$

We can think about these two isometries as generators of two 1-cycles, α, β. We can introduce local coordinates parametrizing these cycles, defined whenever they are non-singular,

$$\theta_\alpha \equiv \frac{1}{2}\left(\text{Arg}(x) - \text{Arg}(y)\right) \tag{13.62}$$

$$\theta_\beta \equiv \frac{1}{2}\left(\text{Arg}(x) + \text{Arg}(y)\right) \tag{13.63}$$

We can now identify these two coordinates in terms of the coordinates in Table 13.1: $(\theta_\alpha, \theta_\beta) \sim (x_7, x_5)$. The action of the orientifold on these two cycles, T-dual to the physical torus, are just $\theta_\alpha \to -\theta_\alpha$, $\theta_\beta \to \theta_\beta$. We thus learn that the orientifold plane spans x_5 and is located at $x_7 = 0, \pi$. In fact, there are two orientifold planes of opposite charge such that the total flux cancels with no further sources. One may also argue for the signs being opposite by noting the absence of net RR-charges coming from the O-planes in the dimer picture. This can be seen in the absence of SO/USp groups and the corresponding lack of non-homogeneous terms in the ACC, which can be thought of as Gauss law for compact cycles. Quite remarkably, the T-dual of such a cycle with opposite-charge O-planes, is known to be precisely an orientifold acting as a shift on the T-dual cycle. The absence of fixed loci for this action translates into the absence of O-plane in the dual geometry. This is described in [Wit98b] where T-duality acts as a sort of Fourier transform: the O-planes of opposite charge are related to delta function whose transform are constant and opposite, cancelling each other. This interpretation nicely match the Gauss law analogy we presented earlier.

After one T-duality along x_7, the T-dual Type IIA construction is analogous to the ones studied in [PU99, PRU00a, Ura00, ELNS00, FHKU01]. The relevant information is encoded in a cycle x_7' where D4-branes are suspended between two NS5-branes. As we explained before, the orientifold action acts now as a shift, rotating halfway the configuration, see Figure 13.7. This action is consistent with the mapping of gauge groups and matter fields on the dimer model.

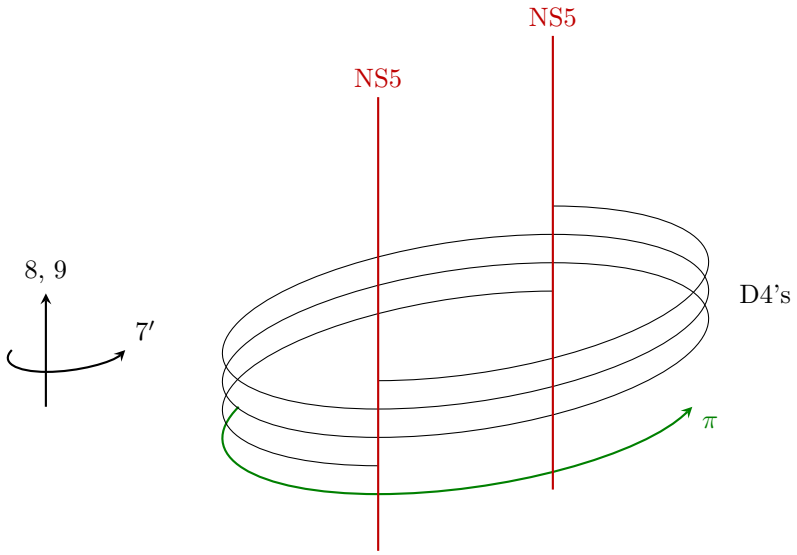


Figure 13.7: Type IIA picture with the orientifold mapping given by a π rotation along x_7' (in green).

Finally, if we further T-dualize along the direction spanned by NS5-branes x_9, we get to the tiling picture. After the last T-duality, the orientifold acts on x_9' as a reflection[7]. Together with the shift on x_7', these actions reproduce the glide reflection that we see on the tiling.

13.3.2 The mirror picture

One further T-duality on the remaining toric cycle brings us to the mirror setup of our starting point [FHKV08a] (D3-branes at singularities). The mirror geometry is fully specified by a Riemann surface Σ_0 in which D6-branes, wrapping 1-cycles, give rise to the appropriate field theory[8]. Gauge fields are

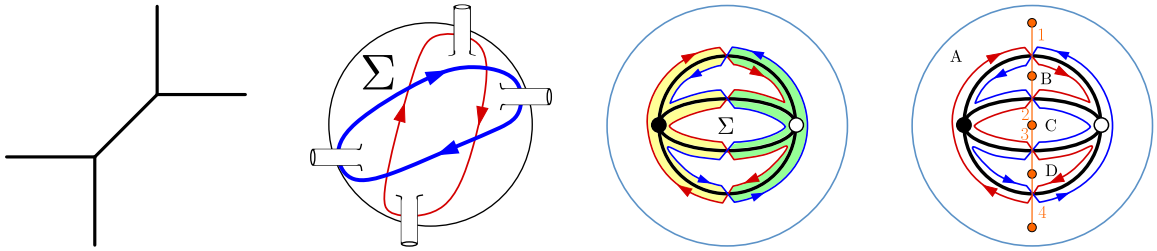
[7] This is a standard fact of orientifolds. Upon T-duality along a direction spanned by the O-plane, an Op-plane is mapped to an O$(p-1)$-plane, with action $\theta \to -\theta$ on the dual cycle.

[8] These 1-cycles are associated to 3-cycles in the full geometry.

associated to different 1-cycles, where D6 are wrapped, matter fields to intersections among these cycles and superpotential terms arise from open string worldsheet instantons supported at disks in this Riemann surface. In fact, this Riemann surface can be seen as the "fattened" version of the web diagram. Furthermore, one can embed the dimer graph on a planar version of it and read immediately both the geometry and the field theory from it. This diagram has been called the shiver. The shiver and the dimer are related by an untwisting procedure [FHKV08a]. The physical interpretation is now as follows.

- Faces in the shiver correspond to ZZPs in the dimer, and represent punctures in Σ_0 with (p, q) charge given by the winding numbers of the ZZP.
- ZZP on the shiver correspond to faces on the dimer, and represent the 1-cycles (Special Lagrangian 3-cycles in the full geometry) where D6-branes wrap. They are hence representing gauge groups.
- ZZP intersections represent brane intersections where open string massless bifundamental fields are located.
- Disk on the surface are euclidean disks in the full geometry where open string worldsheet instantons may arise. These generate the superpotential (albeit non-perturbatively!).

In Figure 13.8 we show the prototypical example, the conifold.



(a) (b) (c) (d)

Figure 13.8: a) Conifold web diagram. b) shows the actual surface describing the mirror geometry. The Zig-Zag paths in blue and red are 1-cycles where D6-branes may wrap. c) Tiling of the mirror Riemann surface Σ_0 giving rise to the conifold gauge theory. d) Orientifold of the conifold. Orange segments and dots denote O-planes and punctures where they end, respectively.

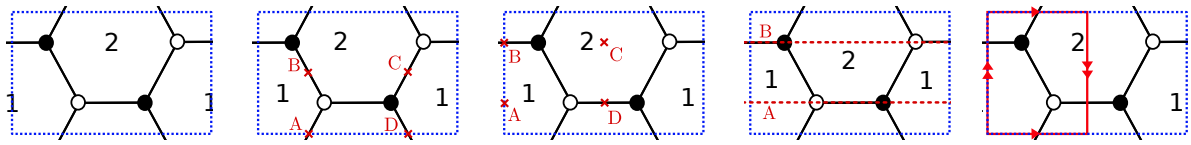
Already in [FHK+07], orientifolds were partially understood in this framework. It was noted that the O6-planes should be viewed as stretching between the different punctures. Different pieces of the O-plane would there be assigned different signs. A full understanding of the available sign choices was not achieved and we will not pursue it here. An orientifold of the conifold is shown in Figure 13.8d.

A simple example: $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$. In this section we explicitly map all orientifolds in the dimer with those in the T-dual and mirror picture. While we are not able to derive matter field projections on the T-dual nor the mirror, we hope to convey a unifying picture. We will focus on $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$ for several reasons. It admits all kinds of orientifolds (fixed points, lines and glide reflections), a simple T-dual set-up and its mirror Σ_0 has genus zero, making it amenable to discussion. Orientifolds in the dimer can be classified in 4 groups, depending on the fixed loci. There are 2 different fixed point orientifolds, shown in Figures 13.9b and 13.9c and a fixed line orientifold, in Figure 13.9d. These were already discussed in [FHK+07]. Different sign assignments, possibly respecting the sign rule, yield the different field contents. Finally, as discussed in Section 13.1, a glide orientifold is realized in the dimer as shown in Figure 13.9e.

A complete classification, up to sign permutations and anomaly cancellation, is shown in Table 13.3.

In Section 13.3.1, we have discussed the setup T-dual to the glide orientifold, shown again in Figure 13.10e. The orientifolds in Figures 13.10b to 13.10d have been discussed in the literature [EJS97, FHKU01, IY02]. One can easily identify these three orientifolds in the dimer setup as Figures 13.9b to 13.9d, respectively. It is worth mentioning that the fixed lines orientifold with the same sign correspond to the O4-plane in the T-dual, while the case with opposite signs is mapped to an O8-plane. Finally, the glide orientifold, Figure 13.9e is identified with Figure 13.10e.

The shiver is shown in Figure 13.11a. It is essentially the same as the T-dual with punctures A and D sitting at the NS5 locations. This was to be expected, since the S^1 in the T-dual is kept in Σ_0. This

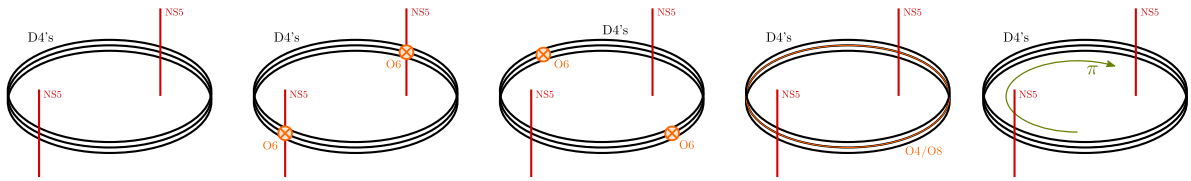


(a) (b) (c) (d) (e)

Figure 13.9: a) Dimer model unit cell of $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$. Three possible orientifold actions with fixed loci are shown in b), c) and d). b) and c) correspond to fixed points, while d) is a fixed line. e) is a glide orientifold with no fixed loci.

O-Type	Figure	Signs	Group	Tensor Content
Fixed Points	Figure 13.9b	$(++++)$	SU	$\text{Adj} + 2\,\square\square + 2\,\overline{\square\square}$
		$(++--)$		$\text{Adj} + \square\square + \overline{\square\square} + \square_{\square} + \overline{\square_{\square}}$
		$(+--+)$		$\text{Adj} + 2\,\square_{\square} + 2\,\overline{\square\square}$
		$(-++-)$		$\text{Adj} + 2\,\square_{\square} + 2\,\square\square$
		$(----)$		$\text{Adj} + 2\,\square_{\square} + 2\,\overline{\square_{\square}}$
	Figure 13.9c	$(++++)$	$SO \times SO$	$\square\square_1 + \square\square_2$
		$(++--)$	$SO \times USp$	$\square\square_1 + \square_{\square 2}$
		$(--++)$	$USp \times SO$	$\square_{\square 1} + \square\square_2$
		$(+-+-)$	$SO \times SO$	$\square_{\square 1} + \square_{\square 2}$
		$(-+-+)$	$USp \times USp$	$\square\square_1 + \square\square_2$
		$(+--+)$	$SO \times USp$	$\square_{\square 1} + \square\square_2$
		$(-++-)$	$USp \times SO$	$\square\square_1 + \square_{\square 2}$
		$(----)$	$USp \times USp$	$\square_{\square 1} + \square_{\square 2}$
Fixed Lines	Figure 13.9d	$(++)$	$SO \times SO$	$\square\square_1 + \square\square_2$
		$(+-)$	$SO \times USp$	$\square_{\square 1} + \square\square_2$
		$(-+)$	$USp \times SO$	$\square\square_1 + \square_{\square 2}$
		$(--)$	$USp \times USp$	$\square_{\square 1} + \square_{\square 2}$
Glide	Figure 13.9e		SU	$\text{Adj} + \overline{\square\square} + \square_{\square} + \square\square + \square_{\square}$

Table 13.3: Different orientifold projections on the dimer.



(a) (b) (c) (d) (e)

Figure 13.10: a) T-dual to $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$. Three possible orientifold actions with fixed loci are shown in b), c) and d). b) and c) correspond to fixed points, while d) is a fixed line. e) is a shift (glide reflection) with no fixed loci.

makes it particularly easy to find the orientifold actions on this surface[9]. The four types are shown in Figure 13.11 with labels matching those of Figures 13.9 and 13.10. Note, in particular, that the glide reflection, in Figure 13.11e consists on a π rotation and a reflection with respect to the dotted orange circle. The total action thus exchanges punctures B and C, for instance. Unlike the other orientifold actions, there are no fixed loci in this case. While the field content can be deduced from the drawings by assigning a sign to every O-plane piece and assigning the matter projection individually (as in the open string computation), it is not clear how to enforce SUSY in this picture. Without further knowledge it is not obvious which projections are SUSY-preserving.

[9]Otherwise one would need to think about the field content on the dimer or the alga map [FHKV08a].

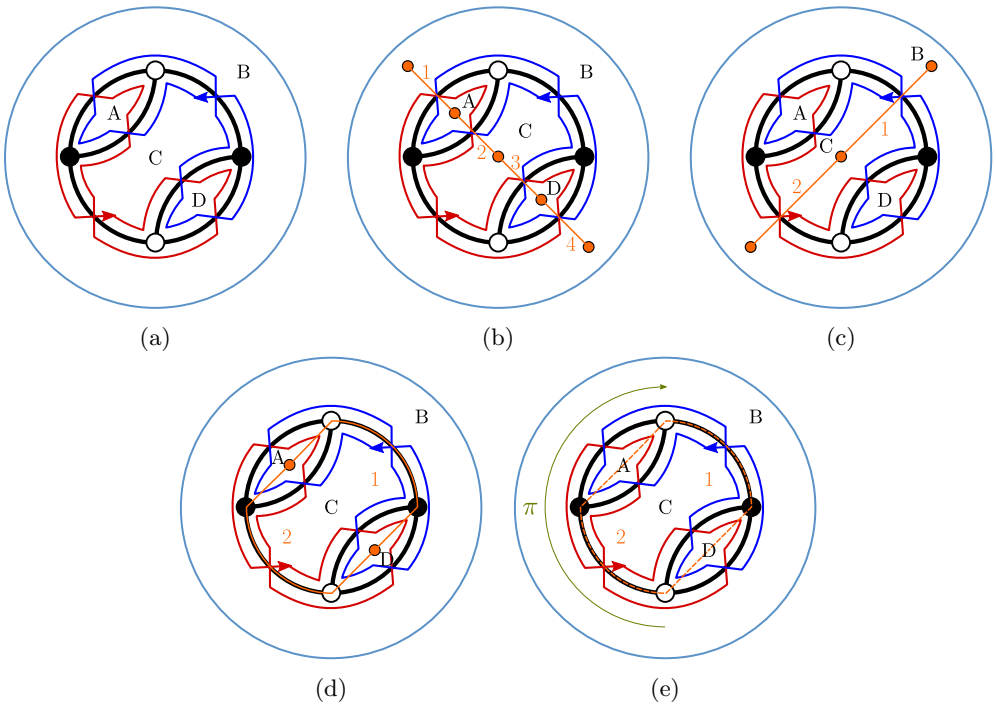


Figure 13.11: a) Tiling of the mirror Riemann surface Σ_0 of $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$. Three possible orientifold planes are shown in b), c) and d). The orientifold plane is shown in orange and its different pieces are labelled using orange numbers. In the dimer b) and c) correspond to fixed points, while d) is a fixed line. e) is a shift orientifold with no fixed loci.

13.4 Involutions and Zig-Zag Paths

In this section, we first develop, in Section 13.4.1, a condition the toric diagram (or equivalently, the ZZP's) of a singularity must satisfy to be compatible with the glide reflection. This enlarges the dictionary between orientifold projections of a given toric singularity and its ZZPs content, as initiated by [RU16b]. Secondly, and with the help of ZZP techniques [But06], we show in Section 13.4.2 how to detect the presence of fractional branes in the orientifolded singularity. Finally, in Section 13.4.3, we give a general proof that the "would be" shift orientifold projection is incompatible with the requirement to preserve SUSY.

13.4.1 Glide Orientifold from the Toric Diagram

A glide reflection can be seen as a combination of a shift and a reflection in the dimer model, even if each of them is not a symmetry per se. Starting from what we learned in our examples and using this simple observation, we can understand how this involution acts on the ZZP content of the toric diagram.

First of all, we notice that the shift and the reflection are performed along the same axis. Consider, for instance, a horizontal shift and axis of reflection as in Figure 13.12. The action of the glide reflection reverts the horizontal component of each ZZP. Actually, the glide reflection leaves no fixed ZZPs, since even those perpendicular to the axis are mapped among themselves because of the shift part of the glide reflection.

Putting the two observations together we can say that: if the glide reflection is composed by a *horizontal shift and a reflection axis*, directed as $(1,0)$ in the dimer, ZZPs are mapped as follows: (p,q) is sent to $(-p,q)$ when $p \neq 0$, while all other ZZPs of the form $(0,\pm 1)$ are mapped to one another, preserving the orientation, meaning that they come in even numbers. In our example of Figure 13.13, the orange $(1,1)$ and purple $(-1,1)$ ZZPs are interchanged. The same is true for the blue and green ZZPs of the $(0,-1)$ type.

Similarly, in order to construct a Klein bottle with a *vertical shift and reflection axis*, the toric diagram should have ZZPs $(1,0)$ and $(-1,0)$ in even numbers, possibly different, and ZZPs (p,q) with $q \neq 0$ paired with ZZPs $(p,-q)$.

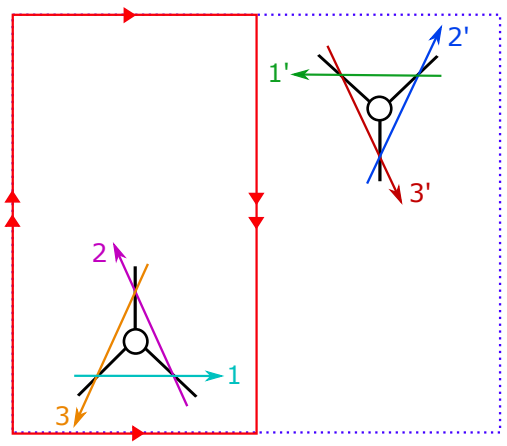


Figure 13.12: The glide orientifold maps together nodes of the same color. The dashed blue line delineates the unit cell of the parent theory, while the red frame represents the orientifold. The ZZPs $1, 2, 3$ are mapped to $1', 2', 3'$ respectively.

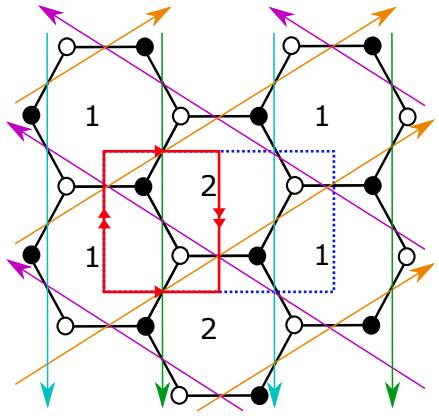
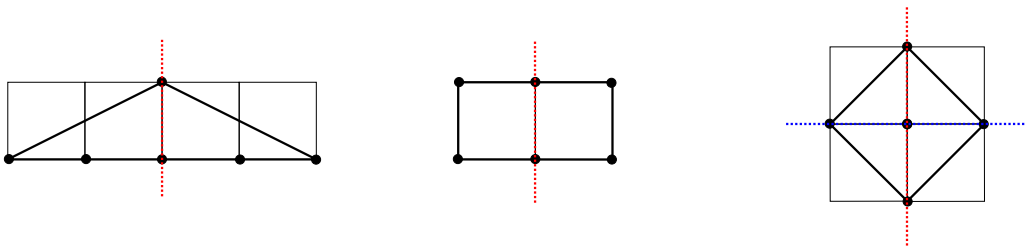


Figure 13.13: The Klein bottle obtained from the dimer of $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$ and the corresponding ZZPs.

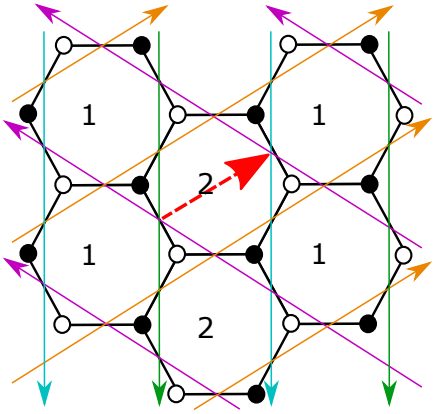
These statements can be summarized by saying that the toric diagram should be symmetric with respect to a vertical or horizontal axis. Moreover, each ZZP has to be mapped to another one, imposing that each kind of ZZP parallel to the axis of reflection in the toric plane should come in even numbers. We show in Figure 13.14 that our examples of Section 13.2 satisfy this criterion.



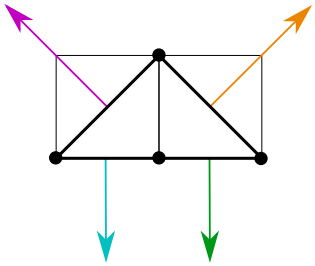
(a) (b) (c)

Figure 13.14: Toric diagrams for singularities that satisfy our necessary criterion to admit one (with an axis of reflection in red) or two (with the second axis of reflection in blue) glide projections: (a) orbifold $\mathbb{C}^2/\mathbb{Z}_4$, (b) conifold-like $\mathcal{C}/\mathbb{Z}_2$, and (c) zeroth Hirzebruch surface F_0.

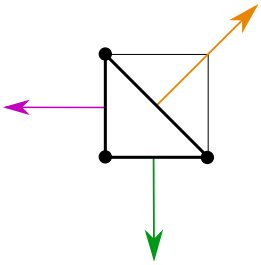
Lastly, an important remark is that this condition may not be satisfied in some of the $SL(2,\mathbb{Z})$ "frames" of the toric diagram, or equivalently, the unit cell in the dimer model may not be symmetric with respect to the glide reflection. Thus, we should state that a generic toric diagram can admit a glide orientifold if it satisfies the conditions above *up to a* $SL(2,\mathbb{Z})$ *action* that can bring its unit cell to a symmetric form with respect to the glide.



(a)



(b)



(c)

Figure 13.15: (a) Dimer diagram of the orbifold $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$. The direction of the shift is given in red, and the ZZPs are illustrated in other colours. (b) Toric diagram for the orbifold $\mathbb{C}^2/\mathbb{Z}_2 \times \mathbb{C}$. (c) Toric diagram for the flat space $\mathbb{C}^3$.

13.4.2 Fractional branes

As already mentioned in Section 13.2.4, these orientifolded theories may admit non-trivial rank assignments, i.e. fractional branes. Their presence can be deduced from the symmetries of the toric diagram and they can be seen as inherited from the "parent" theory. Following [But06, ABF$^+$21a], in what we dub "Butti's Algorithm", we can assign a value v_Γ to each of the n ZZPs of the toric diagram. These values give rise to anomaly free rank assignments, given that they satisfy the following constraints,

$$\begin{cases} \sum_\Gamma v_\Gamma p_\Gamma = 0 \\ \sum_\Gamma v_\Gamma q_\Gamma = 0 \end{cases} , \tag{13.64}$$

where the (p_Γ, q_Γ) are the winding numbers of the ZZP associated to v_Γ.

Since we know how the glide reflection acts on the ZZPs, we may follow the procedure of [ABF$^+$21a] to see which fractional branes survive the projection. As explained there, only *symmetric* fractional branes survive, in the sense that, given two ZZPs v_α and $v_{\bar{\alpha}}$ mapped to each other under the glide reflection, only rank assignments satisfying the following identification survive,

$$v_\alpha = v_{\bar{\alpha}} \, . \tag{13.65}$$

The orientifold projection thus reduces the number of variables v_Γ to the subset of v_α. Moreover, one can check that Equation (13.64) leaves only one non-trivial relation:

$$\sum_\alpha v_\alpha q_\alpha = 0\,. \tag{13.66}$$

Butti's algorithm has a redundancy that allows to perform a global shift on the v_α without affecting the ranks of the gauge groups. Hence, we end up with

$$\#\text{fractional branes} = n/2 - 2 \tag{13.67}$$

in the orientifolded theory.

Butti's algorithm also tells how to construct different kind of fractional branes in the parent theory by specifying a set of v_Γ. We now apply this method to theories with a glide reflection orientifold to see when may $\mathcal{N} = 2$ and deformation fractional branes arise.

- $\mathcal{N} = 2$ fractional branes: The parent theory admits such fractional branes whenever the toric diagram hosts $k > 1$ ZZPs with the same winding numbers, say (p_μ, q_μ). They are turned on whenever only some of these v_μ, among the whole set of ZZPs $\{v_\Gamma\}$, are non-vanishing. Following Equation (13.64), one has

$$\sum_{i=1}^{k} v_{\mu_i} = 0\,, \quad \text{and } v_\nu = 0 \text{ if } (p_\nu, q_\nu) \neq (p_\mu, q_\mu)\,. \tag{13.68}$$

  This condition is compatible with Equation (13.65) only if the k ZZPs are sent to ZZPs with the same winding numbers by the glide reflection, restricting to $(0,1)$ or $(0,-1)$ when (p,q) is mapped to $(-p,q)$. Moreover, k should be a multiple of 4, since for each couple of ZZPs with a symmetric assignment v, we need a second couple with assignment $-v$ in order to satisfy the sum in Equation (13.68). In the examples of Section 13.2, we found that the singularity $\mathbb{C}^2/\mathbb{Z}_4$ satisfies this criterion, see Figure 13.16a.

- Deformation fractional branes: The parent theory will have a deformation fractional brane if there is a subset of m ZZPs in equilibrium $\{v_\sigma\} \subset \{v_\Gamma\}$:

$$\sum_{i=1}^{m} (p_{\sigma_i}, q_{\sigma_i}) = 0\,. \tag{13.69}$$

  The deformation brane is turned on whenever all v_σ have the same non-zero value and all other $v_\tau \notin \{v_\sigma\}$ are vanishing. A glide reflection orientifold theory will have a deformation brane if there is a subset of m ZZPs in equilibrium where each ZZP is accompanied by its image under the glide action, and where m is smaller than n. In the examples of Section 13.2, we found that $\mathcal{C}/\mathbb{Z}_2$ satisfies this criterion while the zeroth Hirzebruch surface F_0 does not, see Figure 13.16b and Figure 13.16c.

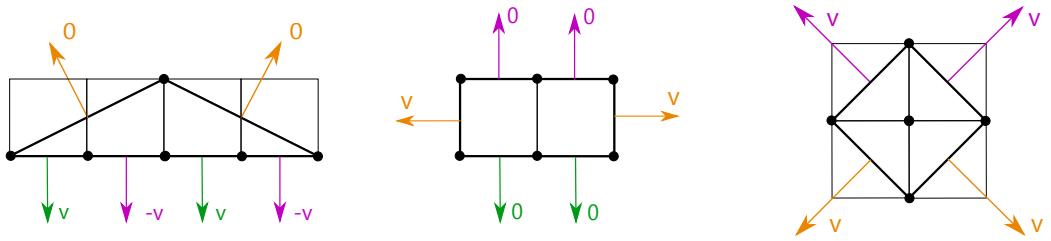
13.4.3 Shift Orientifolds

So far we have only considered orientifolds acting as glide reflections on the dimer. Now we address those acting as a simple shift. We have not discussed these orientifolds earlier because they always break supersymmetry, as we show in the following. In particular, we will see that the holomorphic 3-form Ω_3 is even under such an orientifold action, contradicting the rule of thumb that it should be odd.

As we observed in Section 13.1.2, the shift involution must identify nodes of opposite colors on the dimer, in order to be consistent with the orientifold identification rules. Under such a shift, each ZZP is mapped to a ZZP of opposite winding numbers, $(p,q) \to (-p,-q)$. This can be easily deduced from Figure 13.17.

From the toric diagram, it is possible to obtain the equations defining associated toric variety probed by the D-branes. To do so we need to compute the integer generators of the dual cone to the toric diagram. This procedure is standard in toric geometry and we refer to [CLS11] for all the details. From the lattice vertices on the boundary of the toric diagram (r_i, s_i), we obtain the generators of the cone given by $m_i = (r_i, s_i, 1)$. The dual cone is then given by

$$S^\vee = \{n \in \mathbb{R}^3 | m_i \cdot n \geq 0\}\,, \tag{13.70}$$



(a) (b) (c)

Figure 13.16: Symmetric fractional branes in the parent theory lead to fractional branes in its glide orientifolded version. Couples of ZZPs paired by the glide action are drawn in the same color. (a) $\mathcal{N}=2$ fractional brane in $\mathbb{C}^2/\mathbb{Z}_4$, (b) deformation fractional brane in $\mathcal{C}/\mathbb{Z}_2$, and (c) the zeroth Hirzebruch surface F_0 admits only the regular brane as a symmetric fractional brane.

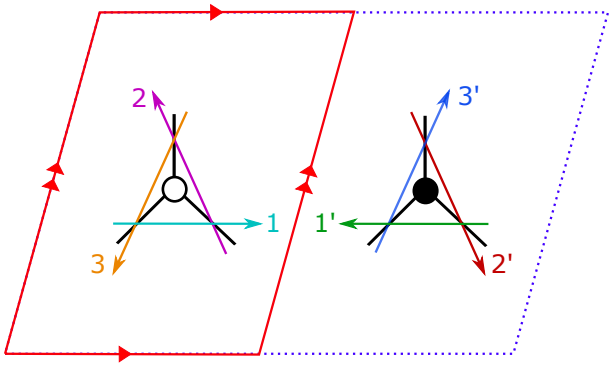


Figure 13.17: The shift orientifold maps white nodes to black nodes, and vice-versa. The dashed blue line delineates the unit cell of the parent theory, while the red frame represents the orientifold. The ZZPs $1,2,3$ are mapped to $1',2',3'$ respectively.

from which it is easy to see that the vectors n are of the form (p,q,a), where (p,q) are the windings of the ZZPs and a is an integer. Indeed, the generators of the dual cone are nothing but the inward pointing vectors, normal to the faces of the cone generated by the m_i. We now need to add the extra generators to span the dual integer cone, $\sigma^\vee = S^\vee \cap \mathbb{Z}^3$. This is achieved by computing linear combinations of the generators with positive rational coefficient and adding all integer vectors we obtain this way. Finally, the equations defining our singularity are given by associating complex coordinates to the generators of the integer dual cone and the relations among them are obtained with the following identification,

$$n_1 + n_2 + \cdots = n_4 + n_5 + \cdots \quad \rightarrow \quad z_1 z_2 \cdots = z_4 z_5 \cdots . \tag{13.71}$$

For example, let us consider the toric diagram of the conifold, which we place in $\mathbb{Z}^2$ as the square with vertices $(0,0),(0,1),(1,1),(0,1)$. The associated cone is

$$\sigma = \left\langle \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\rangle \subset \mathbb{R}^3 , \tag{13.72}$$

and its dual is

$$\sigma^\vee = \langle n_1 = (1,0,0), n_2 = (0,1,0), n_3 = (-1,0,1), n_4 = (0,-1,1) \rangle \subset (\mathbb{R}^3)^* , \tag{13.73}$$

from which it is easy to read the equation defining the singularity:

$$n_1 + n_3 = n_2 + n_4 \quad \rightarrow \quad z_1 z_3 = z_2 z_4 . \tag{13.74}$$

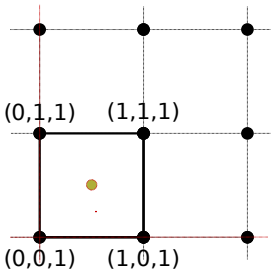


Figure 13.18: The toric diagram of the conifold.

As a second example let us consider the toric diagram of dP_3 and the cone it generates:

$$\sigma = \left\langle \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\rangle \subset \mathbb{R}^3 \,. \tag{13.75}$$

It is dual to $\sigma^\vee$ which is the cone:

$$\begin{aligned} \langle n_1 = (0,1,1), n_2 = (-1,0,1), & n_3 = (-1,-1,1), n_4 = (0,-1,1), \\ & n_5 = (1,0,1), n_6 = (1,1,1), n_0 = (0,0,1) \rangle \ , \end{aligned} \tag{13.76}$$

where we added the vector $n_0 = (0,0,1)$ since $n_1 + n_4 = 2n_0$, meaning that we where missing an integer generator.

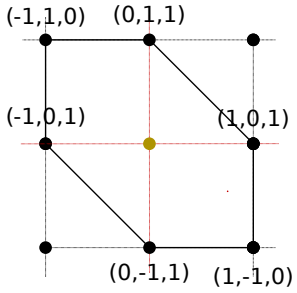


Figure 13.19: The toric diagram of dP_3.

The equations of the variety are

$$\begin{aligned} z_1 z_4 = z_2 z_5 = z_3 z_6 = z_0^2 \\ z_1 z_3 z_5 = z_2 z_4 z_6 \,. \end{aligned} \tag{13.77}$$

We can use the fact that under the shift involution each ZZP is mapped to a ZZP of opposite winding, hence the corresponding toric diagram must be symmetric under the reflection about its center of mass. Such center of mass has, in general, half-integer coordinates (α, β). Under such a reflection, a generic point in the lattice with coordinates $(r, s) \in \mathbb{Z}^2$ is sent to $(2\alpha - r, 2\beta - s)$. Under this operation, the generators of the cone are mapped according to

$$m' = \begin{pmatrix} -1 & 0 & 2\alpha \\ 0 & -1 & 2\beta \\ 0 & 0 & 1 \end{pmatrix} \cdot m \tag{13.78}$$

which maps a generator $m = (a, b, 1)$ to $m' = (2\alpha - r, 2\beta - s, 1)$. The dual cone $S^\vee$ is in turn invariant under the (right) action of that matrix, which acts as

$$n' = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 2\alpha & 2\beta & 1 \end{pmatrix} \cdot n \,, \tag{13.79}$$

or simply,

$$n = (p, q, a) \quad \to \quad n' = (-p, -q, 2\alpha p + 2\beta q + a) \,. \tag{13.80}$$

From these observations, we deduce the following properties:

1. All generators of the dual cone, obtained via Equation (13.70), $n_i = (p_i, q_i, a)$, come paired with another generator $n_i' = (-p_i, -q_i, a')$, for some integer a while a' is obtained via Equation (13.80).

2. Given a generator $n_i = (p_i, q_i, a)$ and its shift image $n_i' = (-p_i, -q_i, a')$, we see that a new integer generator that we were missing can be added $n_0 = (0, 0, 1)$, since

$$n_i + n_i' = (a + a')n_0 \,. \tag{13.81}$$

   This generator is invariant under the shift.

3. All other extra generators come in pairs. Given an extra generator n_l such that

$$n_i + \cdots + n_j' + \cdots = b\, n_l \,, \tag{13.82}$$

   with b integer, by a symmetry argument, we also need to add n_l', since we have[10]

$$n_i' + \cdots + n_j + \cdots = b\, n_l' \,. \tag{13.83}$$

We now rearrange the generators into two sets: the set of n_i with $i = 1, ..., k$ and the set $n_{i+k} = n_i'$ of their images under the shift. Moreover we have n_0 which is the invariant generator. To each generator n_i we associate a complex coordinate z_i. We have $2k + 1$ of them, related by $2k - 2$ relations, that define the toric 3-fold. We divide these relations in two kinds. The k first kind relations are of the form

$$z_i z_{i+k} - z_0^{a+a'} = 0 \,, \tag{13.84}$$

and come from Equation (13.81). We use it to relate every image z_{i+k} to its partner z_i and to z_0. The second kind relations relate all remaining z_i and z_0 together. For example, they may look like

$$z_i z_j z_h^b - z_l z_m z_0^c = 0 \,, \tag{13.85}$$

for some integers b and c.

Under the shift, relations of the first kind are invariant, those of the second kind are not. However, we can build more symmetric expressions for the latter. As we did when going from Equation (13.82) to Equation (13.83), Equation (13.85) becomes, under the shift,

$$z_{i+k} z_{j+k} z_{h+k}^b - z_{l+k} z_{m+k} z_0^c = 0 \,. \tag{13.86}$$

We can now multiply Equation (13.85) by a term $(z_{i+k} z_{j+k} z_{h+k}^b)$ and use the last equation to find

$$(z_i z_j z_h^b)(z_{l+k} z_{m+k}) z_0^c - (z_l z_m)(z_{i+k} z_{j+k} z_{h+k}^b) z_0^c = 0 \,, \tag{13.87}$$

which is now symmetric up to a sign under the shift. We dub these relations the symmetrized second kind relations.

To describe our Calabi-Yau 3-fold, we start with $2k + 1$ variables. From the equations of the first kind we can express all the z_{i+k} in terms of the z_i and z_0, fixing k variables. Then we can use the symmetrized second kind relations to fix $k - 2$ equations, leaving us with only 3 independent variables.

[10] The transformation law in Equation (13.80) acts linearly on Equation (13.82) such that we obtain Equation (13.83).

Now, the non-vanishing holomorphic 3-form Ω_3 is obtained as the Poincaré residue along the CY_3 of the meromorphic $(2k+1)$-form in the ambient space $\mathbb{C}^{2k+1}$:

$$\Omega_3 = \mathrm{Res}\, \frac{\mathrm{d}z_1 \wedge ... \wedge \mathrm{d}z_k \wedge ... \wedge \mathrm{d}z_{2k} \wedge \mathrm{d}z_0}{\left(\prod_{i=1}^{k} P_i \prod_{i=1}^{k-2} Q_i\right)}\,, \tag{13.88}$$

where the P_i are equations of the first kind, while Q_i are of the symmetrized second one.

Under the action of the shift, the numerator of the 3-form is multiplied by $(-1)^k$, since the shift acts on the coordinates exchanging them in pairs. From the denominator we get a factor $(-1)^{k-2}$ coming from the symmetrized second kind equations, cancelling the factor at the numerator and leaving the 3-form invariant. This means that such orientifold projection does not preserve the same supersymmetry as the D3-branes.

Let us finish this section working out an explicit example. In the case of dP_3, one has $k = 3$, and the holomorphic 3-form is the residue of the meromorphic 7-form.

$$\Omega_3 = \mathrm{Res}\, \frac{\mathrm{d}z_1 \wedge ... \wedge \mathrm{d}z_6 \wedge \mathrm{d}z_0}{(z_1 z_4 - z_0^2)(z_2 z_5 - z_0^2)(z_3 z_6 - z_0^2)(z_1 z_3 z_5 - z_2 z_4 z_6)}\,. \tag{13.89}$$

Under the involution, the numerator is multiplied by $(-1)^3$. The first three relations are invariants while for the fourth one takes a minus sign. In the end, Ω_3 is even under the symmetry, and hence there cannot be any supersymmetric shift orientifold of dP_3.

13.5 Conclusion

In this chapter we have studied orientifolds on D3-branes at toric CY_3 singularities using dimer models. We established a classification in terms of smooth involutions of the dimer torus, which allowed us to find the last supersymmetric possibility, the glide reflection orientifold. This possibility may also be reached by directly performing the orientifold projection on the open string spectrum. A last possibility existed, a shift orientifold, but it breaks all supersymmetries, as explicitly argued by studying its action on the holomorphic 3-form. Note that these two cases, not considered before, leave no fixed loci. This exhausts the possible orientifolds acting smoothly on the dimer torus.

Given a toric gauge theory and its associated dimer, one may find the projected theory with the same dictionary as orientifolds with fixed loci. The resulting theories have properties strikingly similar to non-orientifolded theories.

- Unlike orientifold theories with fixed loci, glide reflection orientifolds are guaranteed to satisfy the anomaly cancellation conditions for some rank assignment. In fact, these theories are non-chiral. This fact is non-trivial, see [ABF$^+$21a], and granted by the absence of fixed loci in the glide orientifold that would give raise to tensor matter that could spoil the ACC. From the geometric point of view this boils down to the absence of net RR fluxes sourced by these orientifolds, as there are no fixed loci that can be interpreted as an O-plane. T-duality sheds further light, since the glide orientifold turns to a pair of oppositely charged O-planes on a circle, in the sense of [DP96, Wit98b].

- Again contrary to intuition, these theories are conformal, as shown by explicit computation of the one loop β-function, that vanishes identically.

- Some of these theories admit $\mathcal{N} = 2$ or deformation fractional branes. The latter trigger a cascade of dualities à la Klebanov-Strassler, with a constant step that allows for a UV completion purely in terms of field theory. This is unlike the orientifolds with fixed loci in the literature [AB18] and opens up the possibility of a simple supergravity dual.

- The glide reflection orientifold may be understood in the T-dual and mirror picture, at least for $\mathbb{C}^2/\mathbb{Z}_2$, providing a unifying picture.

This the analysis of orientifolds of brane tilings, or at least those acting as smooth involutions on the torus. However, one may consider other kinds of involutions. For example, involutions not respecting the color mapping presented in Section 13.1.2 or non smooth involutions, can lead to new projections of the tiling, different from the usual orientifold. One may also look for quotients of higher order, in the spirit of what has been done with S-folds [GER16, AT16], and their connection with dimer models. These directions are yet to be explored.

Orientifolds have found extensive use in phenomenological applications by allowing for non-perturbatively generated superpotentials or opening the door to SUSY breaking, for instance. We hope our results may shed light in these and related issues.

Part V

Higher laminations and Hecke TQFTs

Chapter 14

Towards higher laminations

We have seen in Chapter 3 that Thurston's laminations are important characters in the Teichmüller theory of surfaces: in particular, transverse measured laminations provide a spherical compactification of the Teichmüller space to which the natural action of the mapping class group extends. Rational laminations on a surface S naturally correspond to systems of pairwise non-intersecting and non-selfintersecting curves on S with rational weights, up to isotopy. When S is a ciliated surface, rational laminations come in two different flavors depending on what their behaviour at the boundary ∂S of S is: laminations of type $\mathcal{A}$ and of type $\mathcal{X}$. The spaces of rational laminations of type $\mathcal{A}$ and $\mathcal{X}$ on S are respectively denoted $\mathsf{T}^a(S,\mathbb{Q})$ and $\mathsf{T}^x(S,\mathbb{Q})$. Any triangulation Γ of S defines a chart on $\mathsf{T}^a(S,\mathbb{Q})$ and on $\mathsf{T}^x(S,\mathbb{Q})$:

$$\phi_\Gamma : \begin{cases} \mathsf{T}^a(S,\mathbb{Q}) & \longrightarrow & \mathbb{Q}^{|E(\Gamma)|} \\ \mathsf{T}^x(S,\mathbb{Q}) & \longrightarrow & \mathbb{Q}^{|E_i(\Gamma)|} \end{cases} , \tag{14.1}$$

where $E(\Gamma)$ (resp. $E_i(\Gamma)$) is the set of edges (resp. internal edges) of Γ and where $|X|$ denotes the cardinal of any finite set X. Under the flip of an internal edge of Γ, the coordinates describing an $\mathcal{A}$-lamination (resp. $\mathcal{X}$-lamination) transform according to the tropical cluster mutation formulae of type $\mathcal{A}$ (resp. $\mathcal{X}$), and hence rational laminations of type $\mathcal{A}$ (resp. $\mathcal{X}$) are naturally interpreted as the rational tropical points $\mathcal{A}_{S,\mathrm{SL}_2}(\mathbb{Q}^t)$ and $\mathcal{X}_{S,\mathrm{PGL}_2}(\mathbb{Q}^t)$ of the corresponding cluster varieties. The metric completion of $\mathcal{A}_{S,\mathrm{SL}_2}(\mathbb{Q}^t)$ is naturally identified with Thurston's laminations on S [FG06, Corollary 12.1].

A natural subset of each rational lamination space consists of integral laminations, i.e. those laminations whose coordinates are integers in one (equivalently, every) chart. They form the sets $\mathcal{A}_{S,\mathrm{SL}_2}(\mathbb{Z}^t)$ and $\mathcal{X}_{S,\mathrm{PGL}_2}(\mathbb{Z}^t)$. Integral laminations are systems of pairwise non-intersecting and non self-intersecting curves on S up to isotopy, and thus they provide functions on the corresponding Teichmüller spaces. More precisely, there are pairings between laminations of type $\mathcal{A}$ (resp. $\mathcal{X}$) on S and the $\mathcal{X}$- (resp. $\mathcal{A}$-) Teichmüller space of S that we have reviewed in Section 4.2.4. The duality conjectures state that the functions obtained in this way are the generators of the cone of universally positive Laurent polynomials on the dual cluster variety.

The classical Teichmüller spaces $\mathcal{A}_{S,\mathrm{SL}_2}(\mathbb{R}_{>0})$ and $\mathcal{X}_{S,\mathrm{PGL}_2}(\mathbb{R}_{>0})$ of S admit generalizations when SL_2 is replaced by any reductive algebraic split group G. These higher Teichmüller spaces are the $\mathbb{R}_{>0}$-points of cluster varieties: $\mathcal{A}_{S,G}(\mathbb{R}_{>0})$ when G is simply connected (e.g. $G = \mathrm{SL}_N$) and $\mathcal{X}_{S,G}(\mathbb{R}_{>0})$ when G has trivial center (e.g. $G = \mathrm{PGL}_N$). Integral higher laminations can be defined formally as the integral tropical points $\mathcal{A}_{S,G}(\mathbb{Z}^t)$ and $\mathcal{X}_{S,G}(\mathbb{Z}^t)$, and the duality conjectures extend to these cases. However, the generalization of Thurston's rational laminations *as combinatorial objects* remains to be found.

This quest is part of a more general one which aims to the generalization of all the notions that appear in the classical Teichmüller theory – such as complex structures or the mapping class group – to the higher setting. In 2018, Fock and Thomas introduced a notion of *higher complex structure* [FT21], which is defined in terms of the punctual Hilbert scheme of the plane [Tho22, Tho21, Nol22].

The goal of this chapter is to present some thoughts towards such a generalization worked out with Vladimir Fock and Alexander Thomas.

Disclaimer:

Since we do not have a complete picture yet, the reader should be warned that the exposition below is more a cast of the characters that we think should be involved – as well as ideas on how they should interconnect – rather than a fully-fledged pedagogical presentation of a complete theory. The most prominent

novel feature of our approach is the introduction of affine Hecke algebras.

In Section 14.1 we explain how $\mathcal{A}$-laminations parametrize a basis of regular functions on the dual Teichmüller space $\mathcal{X}_{G,S}$. This perspective will guide us through the rest of the chapter. In Section 14.2 we discuss the *hives* of Knutson and Tao, through which Littlewood–Richardson for representations of GL_N and SL_N can be computed, as well as the related *honeycombs*. In Section 14.3 we introduce the Satake correspondence, which motivates the appearance of affine Hecke algebras in higher laminations. We discuss specifically the conjectural combinatorial higher laminations in Section 14.4, which resembles spectral networks that we present briefly in Section 14.5. Eventually, we relate our putative definition of higher laminations to ramified covers of the surface S in Section 14.6.

14.1 $\mathcal{A}$-laminations as a basis of regular functions

14.1.1 Regular functions on moduli stacks of local systems

In order to meaningfully generalize $\mathcal{A}$-laminations to the higher setting, one needs to understand how the algebraic definition of higher laminations can translate to combinatorics. Our starting point is the following result of [FG06, Section 12.5]. Let G be a reductive algebraic group and $R(G)$ the set of isomorphism classes of its finite-dimensional representations. Let $\mathcal{L}(G,S)$ be the moduli space of G-local systems on a punctured surface S, with the notation of Chapter 4. Let also Γ be a triangulation of S.

Definition 14.1. *Let $\overrightarrow{E}(\Gamma)$ be the set of oriented edges of Γ. Then $C(R(G),\Gamma)$ is the set of maps*

$$\lambda : \overrightarrow{E}(\Gamma) \to R(G) \tag{14.2}$$

such that if $\overrightarrow{e}$ and $\overleftarrow{e}$ are the two orientations of the same edge, $\lambda(\overrightarrow{e})$ is contragradient to $\lambda(\overleftarrow{e})$, i.e. for all $g \in G$ one has $\lambda(\overrightarrow{e})(g) = \lambda(\overleftarrow{e})(g^{-1})^*$.

Proposition 14.2. *One has:*

$$\mathcal{O}\left(\mathcal{L}(G,S)\right) = \bigoplus_{\lambda \in \mathcal{C}(R(G),\Gamma)} \bigotimes_{\Delta \in F(\Gamma)} \left(\bigotimes_{\overrightarrow{e} \in \partial\Delta} V_{\lambda(\overrightarrow{e})} \right)^G , \tag{14.3}$$

where $F(\Gamma)$ is the set of faces of Γ and $\overrightarrow{e} \in \partial\Delta$ denotes an edge on the boundary of Δ, oriented counterclockwise.

Proof. The trivalent graph Λ dual to the triangulation Γ is a deformation retract of S, and hence $\mathcal{L}(G,S)$ is isomorphic to the moduli space $\mathcal{L}(G,\Lambda)$ of G-local systems on Λ. Now:

$$\mathcal{O}(\mathcal{L}(G,\Lambda)) \simeq \left(\mathcal{O}(G)^{E(\Lambda)}\right)^{G^{V(\Lambda)}} , \tag{14.4}$$

where $E(\Lambda)$ (resp. $V(\Lambda)$) is the set of edges (resp. vertices) of Λ. This last isomorphism stems from the fact that a G-local system on Λ can be expressed as parallel transport, i.e. the data of elements of G on the edges of Λ, provided an orientation has been chosen for the latter, after having trivialized the local system at the vertices of Λ. Changing the trivialization amounts to an action of $G^{V(\Lambda)}$ on $\mathcal{O}(G)^{E(\Lambda)}$. The Peter–Weyl theorem implies that

$$\mathcal{O}(G) \simeq \bigotimes_{\lambda \in R(G)} V_\lambda \otimes V_\lambda^* . \tag{14.5}$$

Combining this with Equation (14.4) yields

$$\mathcal{O}(\mathcal{L}(G,\Lambda)) \simeq \left(\bigoplus_{\{\lambda : E(\Lambda) \to \widehat{G}\}} \bigotimes_{e \in E(\Lambda)} V_{\lambda(E)} \otimes V_{\lambda(E)}^* \right)^{G^{V(\Lambda)}} . \tag{14.6}$$

Let $\mathcal{C}(R(G),\Lambda)$ be, as in Definition 14.1, the set of maps from oriented edges of Λ to $R(G)$ such that changing the orientation of an edge is accompanied with the replacement of the corresponding representation with the contragradient one. For any $v \in V(\Lambda)$ let $v \leftarrow e$ denote an edge of Λ incident to v and oriented towards v. One can then rewrite Equation (14.6) as:

$$\mathcal{O}(\mathcal{L}(G,\Lambda)) \simeq \bigoplus_{\lambda\in\mathcal{C}(R(G),\Lambda)} \bigotimes_{v\in V(\Lambda)} \left(\bigotimes_{\{v\leftarrow e\}} V_{\lambda(v\leftarrow e)} \right)^G , \tag{14.7}$$

which, when translated in terms of the triangulation Γ, yields the proposition. □

Since Γ is a triangulation, every term which appears in the decomposition of $\mathcal{O}(\mathcal{L}(G,S))$ in Proposition 14.2 is a tensor product of terms of the form

$$(V_a \otimes V_b \otimes V_c)^G , \tag{14.8}$$

where V_a, V_b, V_c are three finite-dimensional irreducible representations of G. The space of such G-invariants is a vector space of finite dimension. Let

$$n_{abc} = \dim(V_a \otimes V_b \otimes V_c)^G . \tag{14.9}$$

Equivalently, n_{abc} is the number of times the irreducible representation V_c^* appears in the product $V_a \otimes V_b$, and one can write

$$V_a \otimes V_b = \bigoplus_c n_{abc}\ V_c^* . \tag{14.10}$$

The n_{abc} are called *Littlewood–Richardson coefficients*. We will introduce *hives*, through which they can be computed, in Section 14.2.

It is proved in [FG06, Section 12.5] that $\mathcal{O}(\mathcal{X}_{G,S})$ is a free $\mathcal{O}(\mathcal{L}(G,S))$-module of rank $|W|^n$ where n is the number of punctures of S. More precisely, the moduli space $\mathcal{X}_{G,S}$ can be naturally seen as a fibered product:

$$\begin{array}{ccc} & \mathcal{X}_{G,S} & \\ \swarrow & & \searrow \\ \mathcal{L}_{G,S} & & H^n \\ \searrow & & \swarrow \\ & (H/W)^n & \end{array} \tag{14.11}$$

where H is a Cartan subgroup of G and $W = \mathrm{Norm}(H)/H$ is the Weyl group. Chevalley restriction implies that $\mathbb{Q}[H]$ is a free $\mathbb{Q}[H]^W$-module of rank $|W|$ and hence

$$\mathcal{O}(\mathcal{X}_{G,S}) = \mathcal{O}(\mathcal{L}_{G,S}) \otimes_{\mathcal{O}((H/W)^n)} \mathcal{O}(H^n) \tag{14.12}$$

is a free $\mathcal{O}(\mathcal{L}_{G,S})$-module of rank $|W|^n$, with a set of generators obtained by pullback of the generators of $\mathbb{Q}[H^n]$ as a free $\mathbb{Q}[(H/W)^n]$-module.

14.1.2 Classical $\mathcal{A}$-laminations for G of type A_1

In this subsection we consider the cases where G is split real of type A_1, i.e. $G = \mathrm{PSL}_2(\mathbb{R})$ in the adjoint form and $G = \mathrm{SL}_2(\mathbb{R})$ in the simply-connected form.

Let us first consider $G = \mathrm{SL}_2(\mathbb{R})$. The isomorphism classes of finite dimensional irreducible representations of $\mathrm{SL}_2(\mathbb{R})$ are in one-to-one correspondence with $\frac{1}{2}\mathbb{Z}_{\geq 0}$, i.e. the elements of the weight lattice $\frac{1}{2}\mathbb{Z}$ in the fundamental Weyl chamber. The irreducible representation with highest weight $\frac{n}{2} \in \frac{1}{2}\mathbb{Z}_{\geq 0}$ is:

$$V_{\frac{n}{2}} = \mathbb{C}_n[x,y] = \mathrm{Vect}\left(x^n, x^{n-1}y, \ldots, xy^{n-1}, y^n\right) , \tag{14.13}$$

where $\mathrm{SL}_2(\mathbb{R})$ acts regularly on the vector space of pairs (x,y), i.e. $V_{\frac{1}{2}}$ is the regular representation.

Let now $a \geq b \in \frac{1}{2}\mathbb{Z}_{\geq 0}$ two highest weights. The *Clebsch–Gordan decomposition* reads:

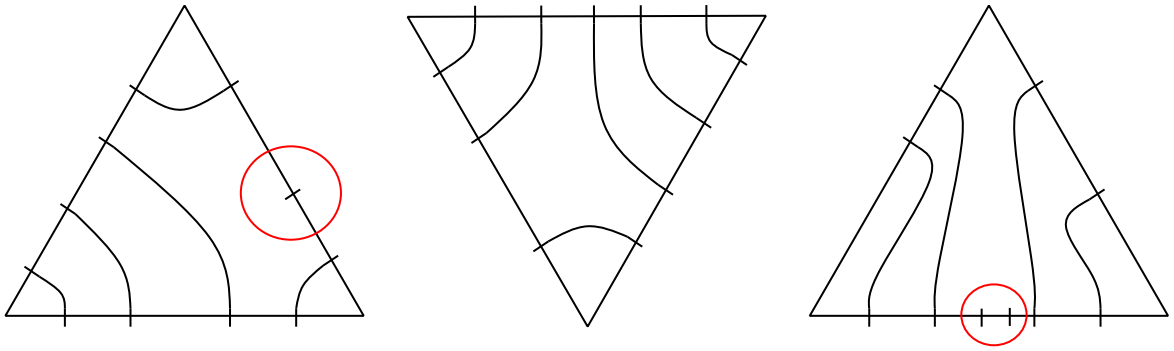
$$V_a \otimes V_b \simeq V_{a-b} \oplus V_{a-b+1} \oplus \cdots \oplus V_{a+b} , \tag{14.14}$$

which can be found in [FH13] for example. For all $a \in \frac{1}{2}\mathbb{Z}_{\geq 0}$ one has $V_a^* = V_a$. Thus if $a, b, c \in \frac{1}{2}\mathbb{Z}_{\geq 0}$:

$$(V_a \otimes V_b \otimes V_c)^{\mathrm{SL}_2(\mathbb{R})} = \begin{cases} 1 & \text{if } a+b+c \in \mathbb{Z} \text{ and } a, b, c \text{ satisfy the triangle inequalities} \\ 0 & \text{otherwise} \end{cases} , \quad (14.15)$$

where a, b, c satisfy the triangle inequalities if $a + b \geq c$, $b + c \geq a$ and $c + a \geq b$.

Recall the reconstruction of classical $\mathcal{A}$-laminations on ciliated surfaces described in Section 3.2.2: consider a triangle Δ of the triangulation Γ of S and $\lambda \in \mathcal{C}(R(G), \Gamma)$. Let V_a, V_b and V_c be the representations that λ associates to the edges on the boundary of Δ oriented counterclockwise, as shown in Figure 14.1. One needs to draw $2a$, $2b$ and $2c$ insertions on the edges of the boundary according to the representation of $\mathrm{SL}_2(\mathbb{R})$ they correspond to. It is possible to connect these insertions with a set of disjoint non-self-intersecting and pairwise non-intersecting curves in the triangle, and such that no insertion is paired with another on the same edge, if and only if $n_{abc} \neq 0$. Moreover, there is a unique way to do it, which translates the fact that $n_{abc} = 1$.



$a + b + c \neq \mathbb{Z}$ OK Triangle inequality not satisfied

Figure 14.1: There is a unique way to construct an $\mathcal{A}$-lamination if and only if $n_{abc} \neq 0$.

Thus each integral classical lamination of type $\mathcal{A}$ in Δ naturally corresponds to a generator of $\mathcal{O}(\mathcal{L}(G, \Delta))$. More generally, one concludes that the elements of $\mathsf{T}^a(S, \mathbb{Z})$ are in one-to-one correspondence with the generators of $\mathcal{O}(\mathcal{L}(G, S))$.

An integral lamination $l \in \mathcal{A}_L(S, \mathbb{Z})$ can be decomposed as

$$l = \sum_\alpha n_\alpha \alpha \ , \quad (14.16)$$

where the sum runs over a set of non-homotopic representatives of the simple closed loops α in S which appear in l, and where n_α is the sum of weights of the curves in l homotopic to α. We have seen in Section 3.3.3 that such an integral $\mathcal{A}$-lamination provides a function on the $\mathcal{X}$-Teichmüller space $\mathcal{T}^x(S)$ of S through the multiplicative pairings:

$$\mathbf{I}_\mathcal{A}(l) = \prod_\alpha \mathrm{Tr}(M_\alpha^{n_\alpha}) \prod_\beta \lambda_\beta^{n_\beta} \ , \quad (14.17)$$

where the first product runs over the curves α not surrounding punctures with M_α is the monodromy along α, whereas the second product runs over the curves β surrounding punctures with λ_β the distinguished eigenvalue of the monodromy along β determined by the framed structure.

In coordinates, an element of $\mathcal{A}_L(S, \mathbb{Z})$ is described by the data of half-integers on the edges of Γ such that the sum of the coordinates on the edges of any triangle of Γ is an integer. Hence:

$$\mathcal{A}_{\mathrm{SL}_2,S}(\mathbb{Z}^t) \subset \mathcal{A}_L(S, \mathbb{Z}) \subset \mathcal{A}_{\mathrm{SL}_2,S}(\frac{1}{2}\mathbb{Z}^t) \ . \quad (14.18)$$

Let $\mathcal{A}_L^0(S, \mathbb{Z}) \subset \mathcal{A}_L(S, \mathbb{Z})$ be the subset consisting of all laminations whose curves (including the special ones bounding punctures) have positive integral weight. An element $l \in \mathcal{A}_L(S, \mathbb{Z})$ belongs to $\mathcal{A}_L^0(S, \mathbb{Z})$ if and only if $n_\alpha > 0$ for all α appearing in the sum in Equation (14.16). Let

$$\mathbf{I}_\mathcal{A}^0(l) = \prod_\alpha \mathrm{Tr}(M_\alpha^{n_\alpha}) \in \mathcal{O}(\mathcal{L}_{\mathrm{SL}_2,S}) \ , \quad (14.19)$$

where M_α is the monodromy along α seen as a function on $\mathcal{L}_{\mathrm{SL}_2,S}$. Let also

$$\mathcal{A}^0_{\mathrm{SL}_2,S}(\mathbb{Z}^t) = A_{\mathrm{SL}_2,S}(\mathbb{Z}^t) \cap \mathcal{A}^0_L(S,\mathbb{Z}) \ . \tag{14.20}$$

Proposition 14.3 (12.2 in [FG06]). *The map $\mathbf{I}^0_{\mathcal{A}}$ provides isomorphisms:*

$$\mathbf{I}^0_{\mathcal{A}} : \mathbb{Q}\{\mathcal{A}^0_L(S,\mathbb{Z})\} \longrightarrow \mathcal{O}(\mathcal{L}_{\mathrm{SL}_2,S}) \ ; \qquad \mathbf{I}^0_{\mathcal{A}} : \mathbb{Q}\{\mathcal{A}^0_{\mathrm{SL}_2,S}(\mathbb{Z}^t)\} \longrightarrow \mathcal{O}(\mathcal{L}_{\mathrm{PGL}_2,S}) \ . \tag{14.21}$$

Proof. Any element in $\mathcal{A}^0_L(S,\mathbb{Z})$ can be described, in coordinates, as a set of non-negative half-integers $\{a_e\}_{e\in E(\Gamma)}$ on the edges of the triangulation Γ of S. Since

$$(V_{a_{e1}} \otimes V_{a_{e2}} \otimes V_{a_{e3}})^{\mathrm{SL}_2(\mathbb{R})} = \begin{cases} 1 & \text{if } a_{e1}+a_{e2}+a_{e3} \in \mathbb{Z} \text{ and triangle inequalities hold} \\ 0 & \text{otherwise} \end{cases} \ , \tag{14.22}$$

the set $\{a_e\}_{e\in E(\Gamma)}$ gives rise to functions on the moduli space $\mathcal{L}_{G,S}$. A first way to do so consists of assigning the representation V_{a_e} to each edge e of Γ, and of considering the tensor product over $F(\Gamma)$ of the invariants $S_{a_{e1},a_{e2},a_{e3}}$ given by Equation (14.22). It follows from Proposition 14.2 that this construction indeed provides a function on $\mathcal{L}_{G,S}$, denoted $S_{\{a_e\}}$ in [FG06]. Moreover it is clear that the collection of $S_{\{a_e\}}$ for all possible sets $\{a_e\}_{e\in E(\Gamma)}$ forms a basis of the functions on $\mathcal{L}_{G,S}$.

Another way to construct a function on $\mathcal{L}_{G,S}$ from the data $\{a_e\}$ consists of assigning the (reducible) representation $V_{1/2}^{\otimes 2a_e}$ to each edge $e \in E(\Gamma)$, and to construct an invariant in $V_{1/2}^{\otimes 2a_{e1}} \otimes V_{1/2}^{\otimes 2a_{e2}} \otimes V_{1/2}^{\otimes 2a_{e3}}$, where $e1, e2$ and $e3$ are the edges on the boundary of a face $\Delta \in F(\Gamma)$, as follows. First, one reconstructs a lamination in Δ corresponding to the boundary data a_{e1}, a_{e2}, a_{e3}, and assign the representation $V_{1/2}$ to each end of the strands of the lamination. There is a unique SL_2-invariant in $V_{1/2} \otimes V_{1/2}$ which can be associated to each strand of the lamination. Taking the tensor product over all strands yields an SL_2-invariant $T_{a_{e1},a_{e2},a_{e3}} \in V_{1/2}^{\otimes 2a_{e1}} \otimes V_{1/2}^{\otimes 2a_{e2}} \otimes V_{1/2}^{\otimes 2a_{e3}}$, and the tensor product of these invariants over all faces of Γ yields a function $T_{\{a_e\}}$ on $\mathcal{L}_{G,S}$ as follows from Proposition 14.2.

The invariants $T_{a_{e1},a_{e2},a_{e3}}$ are linear combinations of the $S_{a_{e1},a_{e2},a_{e3}}$ and vice-versa, and hence the $T_{\{a_e\}}$ are linear combinations of the $S_{\{a_e\}}$ and vice-versa. This implies that the $T_{\{a_e\}}$ forms a basis of functions on $\mathcal{L}_{G,S}$.

A third way to construct a function on $\mathcal{L}_{G,S}$ from $\{a_e\}$ is to reconstruct the lamination corresponding to $\{a_e\}$ and take the product of the $\mathrm{Tr}(M_\alpha)$ over all curves α of the lamination. This yields a function $T'_{\{a_e\}}$:

$$T'_{\{a_e\}} = \prod_\alpha (\mathrm{Tr}(M_\alpha))^{n_\alpha} \ . \tag{14.23}$$

Since $\left((\lambda+\lambda^{-1})^n\right)_{n\in\mathbb{Z}_{\geq 0}}$ and $\left((\lambda^n+\lambda^{-n})\right)_{n\in\mathbb{Z}_{\geq 0}}$ both form a basis of the Laurent polynomials in λ invariant under the involution $\lambda \to \lambda^{-1}$, the functions $\mathbf{I}^0_{\mathcal{A}}(l)$ of Equation (14.19) are linear combinations of the $T'_{\{a_e\}}$ and vice-versa, which proves that the $\mathbf{I}^0_{\mathcal{A}}(l)$ forms a basis of $\mathcal{L}_{G,S}$ as well.

The case of $\mathcal{L}_{\mathrm{PGL}_2,S}$ works similarly, except that the representations of PGL_2 are the V_a for $a \in \mathbb{Z}_{\geq 0}$ only, so that the requirement that $a_{e1}+a_{e2}+a_{e3} \in \mathbb{Z}$ in Equation (14.22) can be dropped. This yields the second isomorphism of the proposition. □

Theorem 14.4 (12.3 in [FG06]).

1. *The functions $\mathbf{I}_{\mathcal{A}}(l)$ when $l \in \mathcal{A}_{\mathrm{SL}_2,S}(\mathbb{Z}^t)$ provide a basis of $\mathcal{O}(\mathcal{X}_{\mathrm{PGL}_2,S})$.*
2. *The functions $\mathbf{I}_{\mathcal{A}}(l)$ when $l \in \mathcal{A}_L(S,\mathbb{Z})$ provide a basis of $\mathcal{O}(\mathcal{X}_{\mathrm{SL}_2,S})$.*

Proof. Note that pairing $\mathbf{I}_{\mathcal{A}}$ of Equation (14.17) differs from $\mathbf{I}^0_{\mathcal{A}}$ in the treatment of the boundary curves only: if β is a curve bounding a boundary component of S and if λ_β is the monodromy along β restricted to the invariant one-dimensional subspace defined by the framing, seen as a function on $\mathcal{X}_{PGL_2,S}$, then $\mathbf{I}^0_{\mathcal{A}}(n\alpha) = \lambda^n + \lambda^{-n}$ whereas $\mathbf{I}_{\mathcal{A}}(n\alpha) = \lambda^n$. The theorem follows from the fact that $\mathbb{Q}[\lambda,\lambda^{-1}]$ is a free $\mathbb{Q}[\lambda,\lambda^{-1}]^{\mathbb{Z}/2\mathbb{Z}}$-module of rank two with generators 1 and λ, where $\mathbb{Z}/2\mathbb{Z}$ acts as $\lambda \to \lambda^{-1}$. Hence the products of $\mathbf{I}^0_{\mathcal{A}}(\alpha)$ over some of the boundary components provide all the generators of the $\mathcal{O}(\mathcal{L}_{\mathrm{PGL}_2,S})$-module $\mathcal{O}(\mathcal{X}_{PGL_2,S})$. Similarly, the products of $\mathbf{I}^0_{\mathcal{A}}(\alpha)$ over some of the boundary components provide all the generators of the $\mathcal{O}(\mathcal{L}_{\mathrm{SL}_2,S})$-module $\mathcal{O}(\mathcal{X}_{SL_2,S})$. □

14.1.3 G-higher $\mathcal{A}$-laminations

One statement of the duality conjectures of [FG06, FG09] is that the algebra of regular functions on an $\mathcal{X}$-cluster variety should have a basis parameterized by the integral tropical points of the corresponding $\mathcal{A}$-cluster variety. While this has been shown not to hold in general in this form [GHK15], the authors of [GHKK18] have provided a refined version of the statement as well as a proof of it. However we will not discuss these issues further, since the aim of this chapter is to explain the general philosophy of one approach to higher laminations in a somewhat informal way. Besides, a full understanding of the details of our approach still requires some work.

Let us therefore assume that G-higher $\mathcal{A}$-laminations on S, i.e. points of $\mathcal{A}_{G,S}(\mathbb{Z}^t)$, parameterize a basis of $\mathcal{O}(\mathcal{X}_{G,S})$. Our main quest is to define precise combinatorial objects generalizing disjoint unions of pairwise non-intersecting and non-self-intersecting curves on S up to equivalence. As we will explain with more details soon, what these combinatorial objects should be is rather clear, however the details of the equivalence relation are not. The analysis of above suggests the following.

1. Since the ring of regular functions on $\mathcal{L}(G,S)$ splits as a direct sum over $\mathcal{C}(R(G),\Gamma)$ (defined in Definition 14.1), the set of higher laminations should itself split as a disjoint union over $\mathcal{C}(R(G),\Gamma)$. Moreover if $\lambda \in \mathcal{C}(R(G),\Gamma)$, the number of distinct laminations corresponding to λ should be

$$\prod_{\Delta \in F(\Gamma)} n_\Delta(\lambda) \ , \tag{14.24}$$

   where $F(\Gamma)$ is the set of faces of Γ and n_Δ is the number $n_{abc}(\lambda)$ corresponding to it. The splitting of Proposition 14.2 further suggests that higher laminations can be entirely constructed in triangles Δ, provided that finite-dimensional irreducible G-representations are associated to the counterclockwise-oriented boundary of Δ, and then glued together as a G-higher lamination on S according to a triangulation Γ of S.

2. The Littlewood–Richardson coefficients for SL_N and PGL_N can be computed via the hives of Knutson and Tao, which we are going to review in Section 14.2. Hives are rational assignments to the vertices of a triangle with an $(N-1)$-subtriangulation satisfying some inequalities generalizing the triangle inequalities of Equation (14.22) that appear in Littlewood–Richardson coefficients for the representations of SL_2 and PGL_2. Moreover, computing a Littlewood–Richardson coefficient n_{abc} amounts to counting the hives with boundary data prescribed by the highest weights a, b and c. We expect n_{abc} non-equivalent G-higher laminations in a triangle Δ with boundary representations V_a, V_b and V_c, and hence there should be a one-to-one correspondence between hives and higher laminations.

3. The Satake correspondence, which we will discuss in Section 14.3, links the ring of finite-dimensional irreducible representations of G to the spherical Hecke algebra of the Langlands dual group G^L. Therefore we expect these spherical Hecke algebras to play a role in the definition of G-higher laminations. Iwahori–Hecke algebras (referred to as Hecke algebras in what follows) are deformations of the group algebras of Coxeter systems (W,S) depending on a parameter q, and as algebras they admit a basis parameterized by the set of simple reflections S, while the multiplication is only constrained by *braid relations* which are the same as in $\mathbb{C}[W]$ and *quadratic relations* which for $s \in S$ read:

$$h_s^2 = (q-1)h_s + q \tag{14.25}$$

   in the so-called standard basis $(h_s)_{s\in S}$. Even if this definition of Hecke algebras does not include spherical Hecke algebras, this points towards the following generalization of Thurston's laminations.

   Conjecture 14.5. *A G-higher lamination is a union of curves on S colored with the positive simple roots of the affine group $\widehat{G}^L$ corresponding to G^L, with vertices of some type determined by the braid and quadratic relations in the (spherical) Hecke algebra of G^L.*

   Two examples of such vertices are shown in Figure 14.2. The edges corresponding to a positive simple root σ_1 (resp. σ_2, σ_3) are shown in red (resp. blue, green). We assume that there is a single edge between the nodes in the corresponding Dynkin diagram associated with σ_1 and σ_2, and a double edge between the ones associated with σ_2 and σ_3.

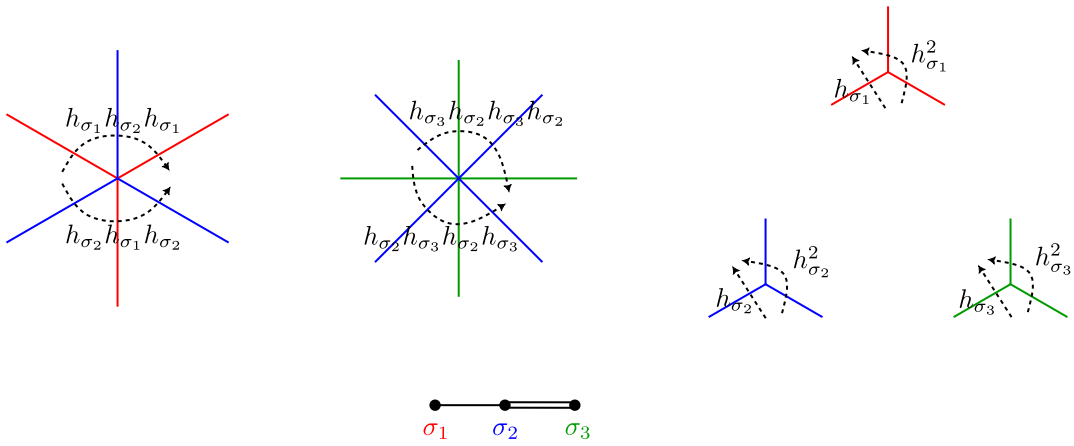


Figure 14.2: Vertices of higher laminations.

The vertices corresponding to the braid relations are such that if one takes a small loop around one of them, the monodromy along it is trivial because it corresponds to a braid relation. The vertices corresponding the the quadratic relations, on the other hand, are specific to Hecke algebras compared to Coxeter group algebras, and encode the fact that the product h_s^2 contains a term proportional to h_s. We will provide more examples of these would-be higher laminations in Section 14.4.

4. The spectral networks of Gaiotto, Moore and Neitzke [GMN13b] are graphs on surfaces of the same kind. They can therefore be used as a guide to construct higher laminations, which despite the similarity with spectral networks cannot be exactly identical: spectral networks encode complex ramified covers of S whereas higher lamination should rather correspond to ramified Lagrangian covers in T^*S. We will discuss spectral networks in Section 14.5.

5. Eventually, the definition of higher laminations of Conjecture 14.5 suggests that there is a link with some ramified covers of the surface S of some sort. We will explore this connection in Section 14.6.

Let us also mention the approach of [Le16] towards higher laminations which involves *affine buildings*. Precisely, a G-higher integral $\mathcal{A}$-lamination, i.e. a point in $\mathcal{A}_{G,S}(\mathbb{Z}^t)$, is associated to a *virtual positive configuration* of points in the affine building of G parameterized by the cyclic set at infinity of the surface S. Le also proposes a similar definition of G-higher integral $\mathcal{X}$-lamination. This is in line with the work of [MS84, Ale08, Par12] which culminate with Parreau's result that spaces of representations of finitely generated groups into a semi-simple group has a compactification in terms of actions on buildings. This is further developed in the groundwork [BP17] which fueled the study of *real spectrum compactifications* of character varieties [Bru88, BIPP21].

There are known links between affine buildings and spherical Hecke algebras [Par06, FKK13], which could bridge the gap between the approach developed in [Le16] and ours.

Transverse though not uncorrelated is the approach of [Xie13], which relies and builds on the ideas of [DMO09, GMN13c, GMN13a]. We will come back to this in Section 14.5. The approach of [Xie13] together with the ideas of [Kup96] and [GS15] inspired in particular [Dou20, DS20] which study a specific basis of the algebra of regular function on the $\mathrm{SL}_3(\mathbb{C})$ character variety of a finite-type surface S.

14.2 Littlewood–Richardson coefficients from hives

14.2.1 The saturation conjecture

Let $\lambda = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, $\mu = \mu_1 \geq \cdots \geq \mu_n \geq 0$ and $\nu = \nu_1 \geq \cdots \geq \nu_n \geq 0$ be three partitions, and let V_λ, V_μ and V_ν be the irreducible representations of $\mathrm{GL}_n(\mathbb{C})$ with respective highest weight λ, μ and ν. An old question in the representation theory of $\mathrm{GL}_n(\mathbb{C})$ is: what conditions need to be satisfied by λ, μ and ν in order for the representation $V_\lambda \otimes V_\mu \otimes V_\nu$ to contain a non-zero $\mathrm{GL}_n(\mathbb{C})$-invariant vector?

Let $n^{\nu}_{\lambda\mu}$ be the Littlewood–Richardson coefficients:

$$V_\lambda \otimes V_\mu = \sum_\nu n^{\nu}_{\lambda\mu} V_\nu \ . \tag{14.26}$$

The question can then be reformulated as: what are the conditions on λ, μ, ν for $n^{\nu}_{\lambda\mu}$ to be strictly positive? Note that in passing from the first formulation of this problem to the second, the representation V_ν needs to be dualized.

The fact that $n^{\nu}_{\lambda\mu} > 0$ can be interpreted combinatorially as the existence of a Littlewood–Richardson skew tableau of shape ν/λ and content μ [LR34], or geometrically as the fact that the intersection of Schubert cells is non-empty (see e.g. [Ful98]). Following [Zel99], let

$$P_r = \{\lambda = (\lambda_1, \dots, \lambda_r) \in \mathbb{Z}^r \mid \lambda_1 \geq \cdots \geq \lambda_r \geq 0\} \tag{14.27}$$

be the semi-group of partitions of length smaller than r, and let

$$LR_r = \left\{(\lambda, \mu, \nu) \in P_r \mid n^{\nu}_{\lambda\mu} > 0\right\} \ . \tag{14.28}$$

The set LR_r is a finitely generated sub-semigroup of $P_r^3 \subset \mathbb{Z}^{3r}$. This fact is as a particular case of a theorem holding for all reductive groups [Éla92], whose proof is attributed to Brion and Knop. Let:

$$\overline{LR_r} = \{(\lambda, \mu, \nu) \in P_r \mid \exists N \in \mathbb{Z}_{>0},\ (N\lambda, N\mu, N\nu) \in LR_r\} \tag{14.29}$$

be the saturation of LR_r. Klyachko provided a description of $\overline{LR_r}$ [Kly98], in terms of inequalities coming from Schubert calculus, of the form:

$$\sum_{k\in K} \lambda_k \leq \sum_{i\in I} \lambda_i + \sum_{j\in J} \lambda_j \ , \tag{14.30}$$

for some triples of subsets $I, J, K \subset \{1, \dots, n\}$ of the same cardinality. The main result is the following theorem:

Theorem 14.6 (Theorems 1.2 and 2.2 in [Kly98])**.**

1. *If $n^{\nu}_{\lambda\mu} > 0$ then λ, μ and ν satisfy the above-mentioned system of linear inequalities,*
2. *Conversely, if λ, μ and ν satisfy the above-mentioned system of linear inequalities then there exists $N \in \mathbb{Z}_{>0}$ such that $n^{N\nu}_{N\lambda\ N\mu} > 0$.*

The *saturation conjecture*, also stated in [Kly98], is that the converse statement in Theorem 14.6 holds in general and not only asymptotically, i.e.:

Conjecture 14.7. *If λ, μ and ν satisfy the above-mentioned system of linear inequalities then $n^{\nu}_{\lambda\mu} > 0$.*

This conjecture was proved by Knutson and Tao in [KT99] using combinatorial tools dubbed tinker-toys, honeycombs and hives. We refer to [Ful98, Zel99, Buc00, Sak12] for more detailed introductions, connections and historical context.

14.2.2 Hives

Hives can be used to determine the Littlewood–Richardson coefficients for representations of $\mathrm{GL}_n(\mathbb{C})$, $\mathrm{SL}_n(\mathbb{C})$ and $\mathrm{PGL}_n(\mathbb{C})$. They encode *Berenstein–Zelevinsky polytopes* [BZ92] in a combinatorial way.

One starts with a triangle with $n+1$ vertices on each side as shown in Figure 14.3, and triangulated in the way depicted on the same figure. This triangle will be denoted Hive_n.

Let us consider the labeling of the vertices of a hive triangle by real numbers. Each elementary rhombus appearing in the hive triangle gives rise to an inequality demanding that the sum of the labels at the obtuse vertices is greater than the sum of the labels at the acute vertices, as depicted in Figure 14.4. A hive is a labeling of the vertices of a hive triangle which satisfies all of these inequalities. A hive is said to be integral if its labeling is integral.

Let us first consider the case of $\mathrm{GL}_n(\mathbb{C})$ following [Buc00]. Let λ, μ, ν be three partitions with $|\lambda| + |\mu| = |\nu|$, and use them to label the vertices on the boundary of a hive as in Figure 14.3. The following theorem was proved by Knutson and Tao [KT99].

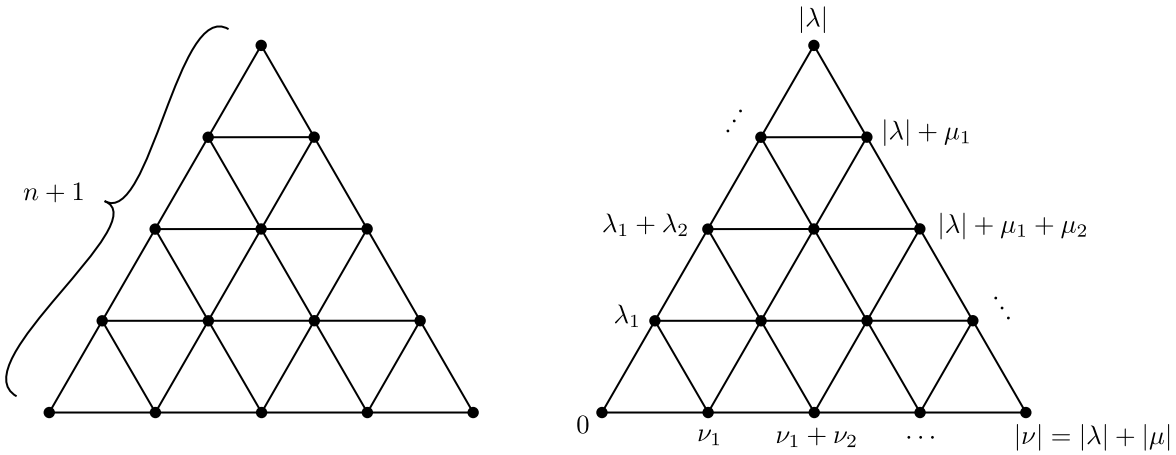
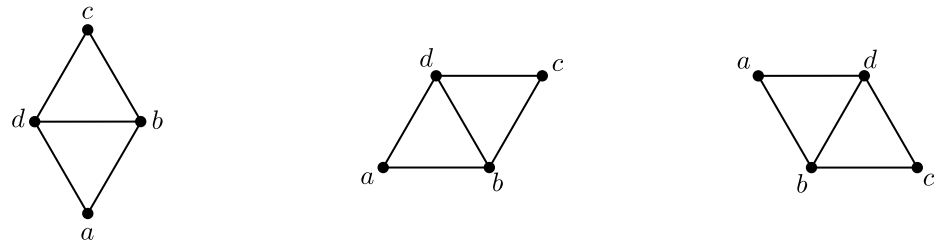


Figure 14.3: The hive triangle Hive_4 (left) and a assignment of border labels (right).



$b + d \geq a + c$ $\qquad$ $b + d \geq a + c$ $\qquad$ $b + d \geq a + c$

Figure 14.4: The rhombus inequalities appearing in hives.

Theorem 14.8. *The Littlewood–Richardson coefficient $c^{\nu}_{\lambda\mu}$ is the number of integral hives with border labels constructed from λ, μ and ν.*

An example is given in Figure 14.5. The partitions $\lambda = \mu = (2, 1)$ and $\nu = (3, 2, 1)$ correspond to irreducible representations of $\mathrm{GL}_3(\mathbb{C})$, respectively denoted $\Lambda_{(2,1)}$ and $\Lambda_{(3,2,1)}$. If one wonders about the multiplicity of $\Lambda_{(3,2,1)}$ in the tensor product $\Lambda_{(2,1)} \otimes \Lambda_{(2,1)}$, Theorem 14.8 implies that it is enough to count the number of integral hives with boundary data determined by λ, μ and ν. There is a single label x which is not fixed by the boundary data, which is the one on the inner vertex of the hive triangle. The right-most rhombus implies that the $x + 6 \leq 5 + 6$ i.e. $x \leq 5$, and the adjacent one with labels $x, 5, 6, 5$ demands that $x + 6 \geq 5 + 5$, i.e. $x \geq 4$. One can easily see that these constraints on x are sharp, and hence this leaves two possibilities: $x = 4$ and $x = 5$, yielding the two hives displayed in Figure 14.5.

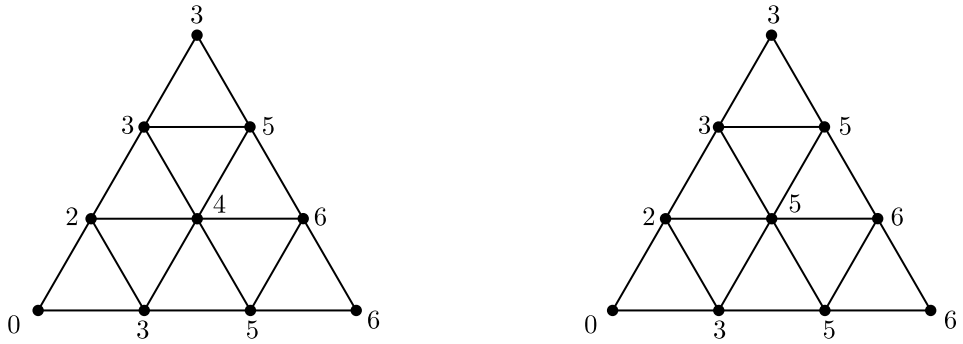


Figure 14.5: Two distinct integral hives on Hive_3 corresponding to the same boundary data.

This result can be extended to compute Littlewood–Richardson coefficients for $\mathrm{SL}_n(\mathbb{C})$ and $\mathrm{PGL}_n(\mathbb{C})$ (actually, for any quotient of $\mathrm{SL}_n(\mathbb{C})$ by a subgroup of its center, even if we are not going to present the method in this very general setting here).

Finite dimensional irreducible representations of $\mathrm{SL}_n(\mathbb{C})$ are labeled by a $(n-1)$-tuple of non-negative integers $(a_1, \dots, a_{n-1})$ encoding their highest weight $(a_1 + \cdots + a_{n-1})L_1 + (a_2 + \cdots + a_{n-1})L_2 + \cdots + a_{n-1}L_{n-1}$ with the notation of [FH13]. The corresponding representation will be denoted $\Gamma_{a_1,\dots,a_{n-1}}$ as in [FH13], so that $\Gamma_{1,0,\dots,0}$ is the fundamental representation V, $\Gamma_{0,1,0,\dots,0}$ is $\wedge^2 V$, and so forth, and so on, so that in general $\Gamma_{a_1,\dots,a_{n-1}}$ is the one appearing in

$$V^{\otimes a_1} \otimes (\wedge^2 V)^{\otimes a_2} \otimes \cdots \otimes (\wedge^{n-1} V)^{\otimes a_{n-1}} \ . \tag{14.31}$$

The Cartan matrix of $\mathfrak{sl}_n(\mathbb{C})$ is:

$$C_{A_{n-1}} = \begin{bmatrix} 2 & -1 & 0 & \dots & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & \dots & 0 \\ 0 & -1 & 2 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \dots & 0 \\ 0 & \dots & 0 & -1 & 2 & -1 & 0 \\ 0 & \dots & \dots & 0 & -1 & 2 & -1 \\ 0 & \dots & \dots & \dots & 0 & -1 & 2 \end{bmatrix} \in \mathcal{M}_{n-1}(\mathbb{Q}) \ , \tag{14.32}$$

and from $(a_1, \dots, a_{n-1})$ we built the following $(n-1)$-tuple:

$$(c_1, \dots, c_{n-1})^T = C_{A_{n-1}}^{-1} \cdot (a_1, \dots, a_{n-1})^T \ . \tag{14.33}$$

For example in the case $n = 3$ the fundamental representation $V = \Gamma_{1,0}$ corresponds to $(c_1, c_2) = (\frac{2}{3}, \frac{1}{3})$, the antisymmetric representation $\wedge^2 V = V^* = \Gamma_{0,1}$, to $(c_1, c_2) = (\frac{1}{3}, \frac{2}{3})$, the adjoint representation is $\Gamma_{1,1}$ and $(c_1, c_2) = (1, 1)$ while the trivial representation $\Gamma_{0,0}$ has $(c_1, c_2) = (0, 0)$.

Computing Littlewood–Richardson coefficients in that case using hives consists of assigning the labels $c_1, \dots c_n$ on the boundary vertices of the hive triangle as follows. Assume that we are interested in computing the multiplicity of $\Gamma_{a_1^\nu,\dots,a_{n-1}^\nu}$ in the tensor product $\Gamma_{a_1^\lambda,\dots,a_{n-1}^\lambda} \otimes \Gamma_{a_1^\mu,\dots,a_{n-1}^\mu}$. One starts by assigning 0 to the vertices of the hive triangle Hive_n. The labels $c_1^\lambda, \dots, c_{n-1}^\lambda$ are assigned to the remaining vertices on the left side of Hive_n clockwise, $c_1^\mu, \dots, c_{n-1}^\mu$ are assigned to the remaining vertices on the right side of Hive_n clockwise, and $c_1^\nu, \dots, c_{n-1}^\nu$ are assigned to the remaining vertices on the bottom side of Hive_n counterclockwise. This is depicted for $n = 4$ in Figure 14.6.

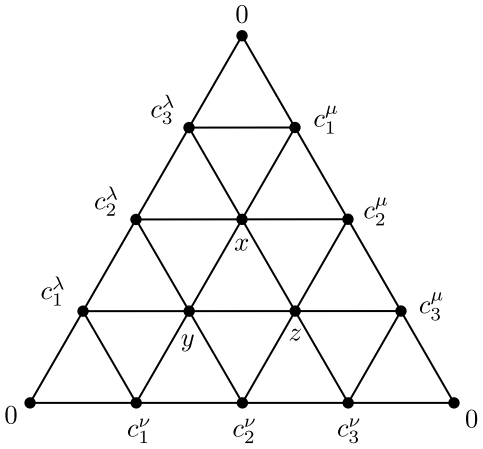


Figure 14.6: Boundary data in the $\mathrm{SL}_n(\mathbb{C})$-case.

.

Note that in this case the border labels are not integers, but belong to $\frac{1}{n}\mathbb{Z}_{\geq 0}$. The result in this case is the following (the specific case of $G = \mathrm{SL}_3$ is discussed in [Dou20]).

Theorem 14.9. *The Littlewood–Richardson coefficient $c_{\lambda\mu}^\nu$ is the number of hives with labels in $\frac{1}{n}\mathbb{Z}_{\geq 0}$, border labels constructed from λ, μ and ν as explained above, and such that for each elementary rhombus $b + d - a - c \in \mathbb{Z}_{\geq 0}$, with the notation of Figure 14.4.*

For example, one can check that the representation $\Gamma_{1,0} \otimes \Gamma_{2,1}$ of $\mathrm{SL}_3(\mathbb{C})$ splits as:

$$\Gamma_{1,0} \otimes \Gamma_{2,1} = \Gamma_{3,1} \oplus \Gamma_{1,2} \oplus \Gamma_{2,0} \tag{14.34}$$

by verifying that there is a single hive with labels in $\frac{1}{3}\mathbb{Z}_{\geq 0}$ with boundary labels $(c_1^\lambda, c_2^\lambda) = (\frac{2}{3}, \frac{1}{3})$, $(c_1^\mu, c_2^\mu) = (\frac{5}{3}, \frac{4}{3})$ and (c_1^ν, c_2^ν) either $(\frac{7}{3}, \frac{5}{3})$, $(\frac{4}{3}, \frac{5}{3})$ or $(\frac{4}{3}, \frac{2}{3})$. These hives are given in Figure 14.7. In each case we display in red a rhombus which imposes that the label in the central vertex is smaller than $\frac{6}{3} = 2$, and in blue a rhombus which imposes that the same label is greater than $\frac{6}{3} = 2$. All other rhombi give rise to equivalent or weaker constraints.

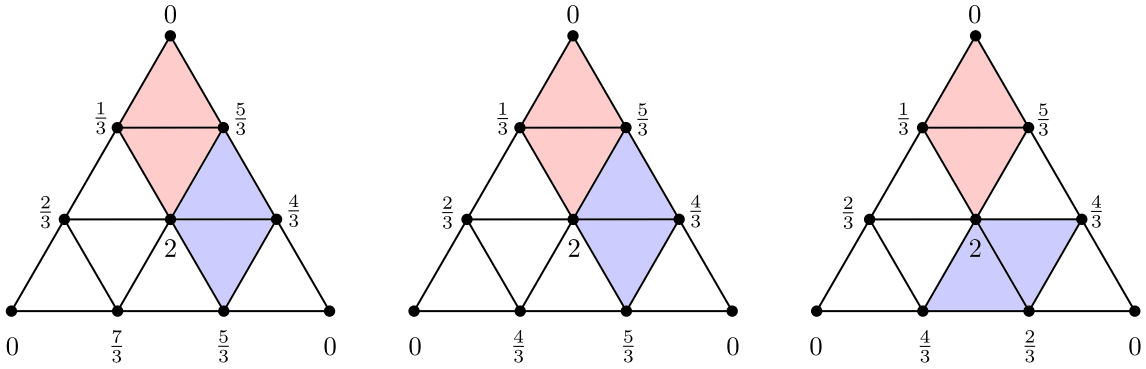


Figure 14.7: The hives appearing in the decomposition $\Gamma_{1,0} \otimes \Gamma_{2,1} = \Gamma_{3,1} \oplus \Gamma_{1,2} \oplus \Gamma_{2,0}$.

.

The rules to compute the Littlewood–Richardson coefficients in the case of $\mathrm{PGL}_n(\mathbb{C})$ are readily obtained by restricting the rules for $\mathrm{SL}_n(\mathbb{C})$ since the highest weights of the irreducible representations of $\mathrm{PGL}_n(\mathbb{C})$ belong to the Weyl chamber in the root lattice of $\mathrm{SL}_n(\mathbb{C})$, which is a subset of its weight lattice.

The hive conditions that we have discussed were shown in [GS15] to correspond to the tropicalization $\mathcal{W}^t$ of a remarkable rational positive function $\mathcal{W}$ on $\mathcal{A}_{G,S}$ dubbed *potential*.

14.2.3 Honeycombs and Horn's conjecture

In this subsection we follow [KT01] – see also the review [Ful98].

The spectrum of an $n \times n$ Hermitian matrix can be written as a weakly decreasing sequence of n real numbers $\lambda = (\lambda_1 \geq \cdots \geq \lambda_n)$. In 1912, Hermann Weyl asked for the conditions on three such spectra λ, μ, ν ensuring the existence of Hermitian matrices H_λ, H_μ and H_ν with respective spectrum λ, μ and ν such that $H_\lambda + H_\mu + H_\nu = 0$ [Wey12].

Some necessary conditions are immediate. For example, taking the trace of $H_\lambda + H_\mu + H_\nu = 0$ yields

$$\sum_{i=1}^{n} \lambda_i + \sum_{j=1}^{n} \mu_j + \sum_{k=1}^{n} \nu_k = 0 \ . \tag{14.35}$$

The sum of the largest eigenvalues of H_λ, H_μ and H_ν also obviously needs to be greater than zero:

$$\nu_1 + \lambda_1 + \mu_1 \geq 0 \ . \tag{14.36}$$

Weyl found a list of such necessary conditions, that are all inequalities of the form of Equation (14.36) apart from the trace condition of Equation (14.35). This fact was explained by Horn [Hor62], who also conjectured an exhaustive list of these inequalities, i.e. a set of necessary and sufficient conditions for Weyl's original problem.

Honeycombs were introduced in [KT99] in order to prove Klyachko's saturation conjecture, which in turn implies Horn's conjecture. We will not explain the proof here and rather refer to [KT99, KT01], however let us now define honeycombs and the statement of Knutson and Tao's main result.

Let us consider the plane

$$\Pi = \left\{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\right\} \subset \mathbb{R}^3, \tag{14.37}$$

with the six *cardinal directions* $(0,1,-1)$, $(-1,1,0)$, $(-1,0,1)$, $(0,-1,1)$, $(1,-1,0)$ and $(1,0,-1)$ drawn pointing Northwest, North, Northeast, Southeast, South and Southwest as in Figure 14.8.

A diagram is defined as a configuration of possibly half-infinite line segments in Π with each edge parallel to one of the six cardinal directions and labeled with a *multiplicity*, i.e. a number in $\mathbb{Z}_{>0}$. Line segments can intersect at *vertices*, and a vertex is a *zero-tension point* if the sum of the vectors corresponding to the cardinal directions of the rays emanating from it, weighted by their multiplicity, is zero.

A *honeycomb* is a diagram with finitely many vertices which all are zero-tension points, and such that the cardinal directions of the semi-infinite lines are Northwest, Northeast or South. An example of a honeycomb is displayed in Figure 14.8, where only multiplicities greater than 2 are shown.

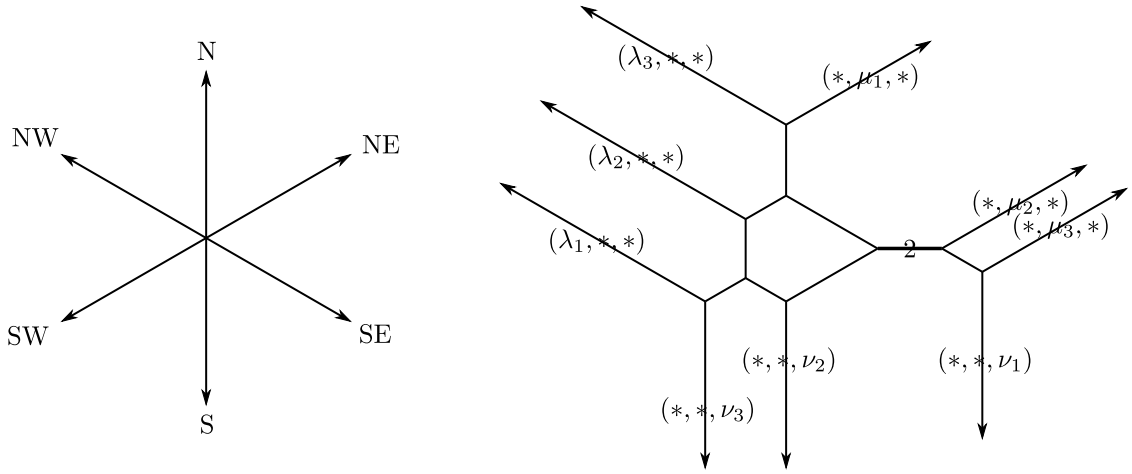


Figure 14.8: The six cardinal directions in Π (left) and a honeycomb (right).

.

Half-infinite lines are aligned with either the NW, NE or S cardinal directions and hence they all correspond to a constant real coordinate. Let them be denoted $\lambda_1,\dots,\lambda_n,\mu_1,\dots,\mu_n$ and $\nu_1,\dots,\nu_n$ as in Figure 14.8.

Theorem 14.10 ([KT01])**.** *There exist Hermitian $n\times n$ matrices H_λ, H_μ and H_ν with respective spectrum λ,μ and ν such that $H_\lambda + H_\mu + H_\nu = 0$ if and only if there exists a honeycomb with boundary values (λ,μ,ν).*

Honeycombs resemble the (p,q)-webs introduced in Section 6.4.2 very much. This calls for a dictionary between five-dimensional quantum field theories with eight supercharges and Horn's conjecture. I am not aware of any research work going in that direction, however such a program might fit very nicely within the ideas developed in [BGH$^+$21]. The next section will hopefully provide another piece of evidence for such an alignment.

14.3 The Satake correspondence and Hecke algebras

14.3.1 Hecke algebras

Generalities. The notion of *Hecke ring* or *Hecke algebra* first appeared in [Shi59] (see also the review [Ver61]) as *l'algèbre des transformations*. Subsequently, Iwahori [Iwa64] and Matsumoto [IM65] generalized this notion, building on Chevalley's construction of finite simple groups from reductive Lie groups [Che55]. In this subsection we follow closely the presentation of [Iwa64].

Roughly speaking, if G is a group and H a subgroup of G, the Hecke algebra $\mathcal{H}(G,H)$ is the algebra of functions on $H\backslash G/H$ with multiplication given by the convolution.

More precisely, one needs to assume that for all $x\in G$, the subgroup H is commensurable with xHx^{-1}, i.e.:

$$[H : H\cap xHx^{-1}] < \infty\ . \tag{14.38}$$

Let $\mathfrak{M}$ be the set of all H-left invariant subsets of G, and let μ be the measure on $\mathfrak{M}$ such that if $A \in \mathfrak{M}$, the measure $\mu(A)$ is the cardinal $|H\backslash A|$. Note that μ is right-invariant, i.e. for all $g \in G$, one has $\mu(Ag) = \mu(A)$. Let now $L(G,H)$ be the set of all complex-valued, $\mathfrak{M}$-measurable and μ-summable functions on G such that:

- for any $g \in G$ and $h, h' \in H$, one has $f(hgh') = f(g)$,
- $\mu(S(f)) < \infty$, where $S(f) = \{x \in G \mid f(x) \neq 0\}$.

The set $L(G,H)$ has naturally the structure of a complex vector space, and moreover it can be endowed with the convolution product:

$$(f_1 \star f_2)(x) = \int_G f_1(xy^{-1}) f_2(y) \mathrm{d}\mu(y) \ , \qquad f_1, f_2 \in L(G,H), \ x \in G. \tag{14.39}$$

The fact that μ is H-right invariant implies that $L(G,H)$ is an associative algebra. The assumption of Equation (14.38) implies that for any $a \in G$, the characteristic function χ_A of the double coset $A = HaH$ is in $L(G,H)$. The set $(\chi_A)_{A \in H\backslash G/H}$ is a base of $L(G,H)$.

Definition 14.11. *The Iwahori–Hecke algebra (or Hecke algebra, for short) $\mathcal{H}(G,H)$ is the free $\mathbb{Z}$-module defined as the $\mathbb{Z}$-submodule of $L(G,H)$ generated by the $(\chi_A)_{A \in H\backslash G/H}$.*

Proposition 14.12 (1.1 in [Iwa64])**.** *The Hecke algebra $\mathcal{H}(G,H)$ is a sub-ring of $L(G,H)$, i.e. it is an algebra. Moreover $\mathcal{H}(G,H) \otimes_{\mathbb{Z}} \mathbb{C} = L(G,H)$.*

Hecke algebras of Chevalley groups over finite fields with respect to a Borel subgroup. The sequel of [Iwa64] mostly revolve around the particular cases of $\mathcal{H}(G(\mathbb{F}_q), B(\mathbb{F}_q))$ where $q = p^\alpha$ is a prime power, $\mathbb{F}_q$ is the finite field with q elements, $G(\mathbb{F}_q)$ is a semi-simple algebraic group over $\mathbb{F}_q$ and $B(\mathbb{F}_q)$ is a Borel subgroup of $G(\mathbb{F}_q)$. Let us denote the corresponding Hecke algebra $\mathcal{H}_q(G,B)$. The main result of [Iwa64] is an explicit description of $\mathcal{H}_q(G,B)$ in terms of generators and relations.

Let $\mathfrak{g}$ be the complex semi-simple Lie algebra corresponding to G and $\mathfrak{h}$ a Cartan subalgebra of $\mathfrak{g}$. Let Δ be the root system corresponding to $(\mathfrak{g}, \mathfrak{h})$ and let $\Pi = \{\alpha_1, \ldots, \alpha_l\}$ be the set of positive simple roots given by the choice of an ordering of the roots of $\mathfrak{g}$. Let θ_{ij} be the angle between α_i and α_j. Let $w_1, \ldots, w_l$ be the reflections in W corresponding to the elements of Π. The Weyl group W of $\mathfrak{g}$ is generated by $w_1, \ldots, w_l$ with the following relations:

$$\begin{cases} w_i^2 = 1 & \forall \ i = 1, \ldots, l \\ (w_i w_j)^k = 1 & \text{if } \theta_{ij} = (k-1)\pi/k \end{cases} . \tag{14.40}$$

Let now H be the Cartan subgroup of G corresponding to $\mathfrak{h}$. The Weyl group W can be naturally expressed as

$$W = \mathrm{Norm}(H)/H \ . \tag{14.41}$$

Let $\omega_1, \ldots, \omega_l$ be representatives of $w_1, \ldots, w_l \in W$ in G. The Hecke algebra $\mathcal{H}_q(G,B)$ is generated by the cosets $S_i = B\omega_i B$ for $i = 1, \ldots, l$ with the following relations:

$$\begin{cases} S_i^2 = (q-1)S_i + q & \forall \ i = 1, \ldots, l \\ (S_i S_j)^k = 1 & \text{if } \theta_{ij} = (k-1)\pi/k \end{cases} . \tag{14.42}$$

Comparing Equation (14.40) and Equation (14.42) makes it clear that the Hecke algebra $\mathcal{H}_q(G,B)$ is an associative, non-commutative (in general) deformation of the group ring $\mathbb{C}[W]$ of the Weyl group W.

The *Bruhat decomposition*

$$W \simeq B\backslash G/B \ , \tag{14.43}$$

makes the link between the Hecke algebras $\mathcal{H}_q(G,B)$ and functions on the Weyl group intuitive. Actually, the Hecke algebras $\mathcal{H}_q(G,B)$ are free $\mathbb{Z}[q^{\pm 1}]$-modules, with a *standard basis* whose elements h_w are in one-to-one correspondence with the elements $w \in W$. Hence every element $h \in \mathcal{H}_q(G,B)$ can be written as:

$$h = \sum_{w \in W} c_w h_w \ , \tag{14.44}$$

where for all $w \in W$, one has $c_w \in \mathbb{Z}[q^{\pm 1}]$.

Let us now briefly discuss a general property of Hecke algebras which will be central in the next chapter.

Definition 14.13 (7.1.7 in [GP00]). *Let A be any associative ring, and let H be a finitely generated free associative A-algebra. A trace function on H is an A-linear map $\tau : H \to A$ such that for all $h, h' \in H$ one has $\tau(hh') = \tau(h'h)$. The set of trace functions on H is an A-module. A trace on H is said to be symmetrizing if the bilinear form*

$$\begin{array}{ccc} H \times H & \to & A \\ (h, h') & \mapsto & \tau(hh') \end{array} \tag{14.45}$$

is non-degenerate. The algebra H endowed with a symmetrizing trace is said to be a symmetric algebra.

Hecke algebras $\mathcal{H}_q(G, B)$ are naturally symmetric algebras, with their *standard* symmetrizing trace coming from the characteristic function on $\{1\} \subset W$:

$$\tau : \begin{array}{ccc} \mathcal{H}_q(G, B) & \to & \mathbb{Z}[q^{\pm 1}] \\ \sum\limits_{w \in W} h_w S_w & \mapsto & h_e \end{array} , \tag{14.46}$$

where $e \in W$ is the neutral element of the group W. Because Hecke algebras are symmetric algebras, their representation theory is well behaved – see [GP00] for example. We will come back to this in the next chapter. Hecke algebras have been generalized in many ways and the representation theory of these generalizations has been studied a lot in the last decades [Chl16]. Let us now present a slight generalization of Iwahori–Hecke algebras which will plays a prominent role in the Satake correspondence.

Affine Hecke algebras. The *affine Hecke algebras* which are the most in the spirit of the Iwahori-Hecke algebras presented in the last paragraph are obtained by replacing the finite field $\mathbb{F}_q$ by the non-Archimedian local field $\mathbb{F}_q((t))$ of formal Laurent series over $\mathbb{F}_q$. In that case the Hecke algebra $\widehat{H}_q(G, B) := \mathcal{H}(G(\mathbb{F}_q((t))), B(\mathbb{F}_q((t))))$ has a set of generators in one-to-one correspondence with the generators of the affine Weyl group $\widehat{W}$ corresponding to G, i.e. with the nodes of the associated affine Dynkin diagram, whereas the relations are again provided by the adjacency in the Dynkin diagram. For example, $\widehat{H}_q(\mathrm{SL}_n, B)$ can be presented as:

$$\left\langle S_0, \dots, S_n \mid S_i^2 = (q-1)q_i + q, \left[\begin{array}{ll} S_i S_j S_i = S_j S_i S_j & \text{for } i, j \in \mathbb{Z}/n\mathbb{Z},\ |j-i| = 1 \\ S_i S_j = S_j S_i & \text{otherwise} \end{array} \right. \right\rangle \tag{14.47}$$

as a $\mathbb{Z}[q^{\pm 1}]$-algebra, to be compared with the affine Dynkin diagram $\widehat{A}_n$ which is a necklace of length $n+1$.

Instead of considering Hecke algebras of a general semi-simple algebraic group with respect to a Borel subgroup, one can study Hecke algebras with respect to a parabolic subgroup P. Let Π^- be the set of negative simple roots opposite to Π, and for all $\Theta \subset \Pi^-$ let P_Θ be the parabolic subgroup of G generated by $\Pi \cup \Theta$, so that $P_\emptyset = B$ and $P_{\Pi^-} = G$. When Θ consists of all elements of Π^- but one, the corresponding parabolic subgroup of G is said to be maximal. The Bruhat decomposition of Equation (14.43) transforms into

$$P_\Theta \backslash G / P_\Theta \simeq W_\Theta \backslash W / W_\Theta , \tag{14.48}$$

where W_Θ is the subgroup of W generated by the positive simple roots whose opposites belong to Θ, and hence the Hecke algebras $\mathcal{H}_q(G, P_\Theta)$ are the space of functions over $W_\Theta \backslash W / W_\Theta$ endowed with the convolution product.

In this affine case, there is a distinguished maximal parabolic subgroup: the *affine* one. In going from a finite-dimensional semi-simple Lie algebra $\mathfrak{g}$ to its affine counterpart $\widehat{\mathfrak{g}}$, one adds a new positive simple root dubbed the *affine root* and completes the root system accordingly. With the notation of above i.e. the positive simple roots being $\Pi = \{\alpha_1, \dots, \alpha_l\}$, let α_0 be the affine root of the affine lie algebra $\widehat{\mathfrak{g}}$ and let $\widehat{\Pi} = \Pi \cup \{\alpha_0\}$. Let $\widehat{\Pi}^- = \{\overline{\alpha_0}, \overline{\alpha_1}, \dots, \overline{\alpha_n}\}$. The choice $\Theta = \{\overline{\alpha_1}, \dots, \overline{\alpha_n}\}$ determines the maximal affine parabolic subgroup $\widehat{P}$ of $\widehat{G}$. If the affine group $\widehat{G}$ is considered as $G(\mathbb{F}_q((t)))$, the maximal affine parabolic $\widehat{P}$ is $G(\mathbb{F}_q[[t]])$, where $\mathbb{F}_q[[t]]$ is the ring of formal power series over $\mathbb{F}_q$.

In the following subsection the Hecke algebras

$$\widehat{\mathcal{H}}_q^{\mathrm{sph}}(G) := \mathcal{H}(G(\mathbb{F}_q((t))), G(\mathbb{F}_q[[t]])) , \tag{14.49}$$

called *spherical Hecke algebras*, will make a natural appearance through the Satake correspondence. They are algebras of functions over $W \backslash \widehat{W} / W$ endowed with the convolution product as above, and satisfy the special property of being commutative.

14.3.2 The Satake isomorphism

The classical *Satake correspondence* introduced in [Sat63] is an isomorphism between the spherical Hecke algebra of a split reductive group G and the representation ring of the (Langlands) dual group G^L.

We follow the presentation of [Gro98]. Let F be a local field and let A be the ring of integers of F; for example $F = \mathbb{F}_q((t))$ is the field of Laurent series in an indeterminate over $\mathbb{F}_q$ and $A = \mathbb{F}_q[[t]]$ is the ring of formal series in an indeterminate t over $\mathbb{F}_q$. Let $\mathbf{G}$ be a connected reductive split algebraic group over F. Let $\mathbf{B}$ be a Borel subgroup of $\mathbf{G}$ and let $\mathbf{T} \subset \mathbf{B}$ a maximal torus, both defined over A. Let $W = \mathrm{Norm}(\mathbf{T})/\mathbf{T}$ be the Weyl group of $\mathbf{G}$. Let

$$X^*(T) = \mathrm{Hom}(\mathbf{T}, \mathbb{G}_m), \tag{14.50}$$
$$X_*(T) = \mathrm{Hom}(\mathbb{G}_m, \mathbf{T}) \tag{14.51}$$

be respectively the character and cocharacter groups of $\mathbf{T}$. The roots of $\mathbf{G}$ are in $X^*(\mathbf{T})$ while the coroots are in $X_*(\mathbf{T})$. A choice of ordering of the set of roots Φ defines the subsets Φ^+ of positive roots and Φ^- of negative roots. Let $\Delta \subset \Phi^+$ be the set of positive simple roots.

Let

$$P^+ = \{\lambda \in X_*(\mathbf{T}) \mid \langle \lambda, \alpha \rangle \geq 0,\ \forall \alpha \in \Delta\} \subset X_*(\mathbf{T}) \tag{14.52}$$

be the positive Weyl chamber defined by Δ. Let G^L be the complex group (Langlands) dual to $\mathbf{G}$. It is the group whose root lattice is the weight lattice of $\mathbf{G}$, and vice-versa. Upon the choice of a maximal torus inside a Borel subgroup in G^L, i.e. $T^L \subset B^L \subset G^L$, one has the isomorphism

$$X^*(T^L) \simeq X_*(\mathbf{T})\ , \tag{14.53}$$

and hence the elements of $P^+ \subset X^*(T^L)$ index the finite-dimensional irreducible representations of G^L. Let $\lambda \in P^+$ and let V_λ be the corresponding irreducible representation of G^L. Let $\chi_\lambda = \mathrm{Tr}\{|V_\lambda\}$ be the character of V_λ. Then

$$\chi_\lambda \in R(G^L) = \mathbb{Z}\left[X^*(T^L)\right]^W\ , \tag{14.54}$$

and $R(G^L)$ is the representation ring of G^L.

The Hecke ring in which we are interested in here is $\mathcal{H}(G, K)$ where $G = \mathbf{G}(F)$ and $K = \mathbf{G}(A)$, as advertised in the previous section. It is the ring of locally constant, compactly support K-biinvariant functions on G, with multiplication given by the convolution on G. A basis of $\mathcal{H}(G, K)$ is provided by the characteristic functions on the elements of $K\backslash G/K$.

Proposition 14.14 (2.10 in [Gro98])**.** *The Hecke ring $\mathcal{H}(G, K)$ is commutative.*

The *Satake transform* is a map

$$\begin{array}{ccc} \mathcal{H}(G,K) & \longrightarrow & R(G^L) \otimes \mathbb{Z}[q^{1/2}, q^{-1/2}] \\ f & \longmapsto & Sf \end{array}\ , \tag{14.55}$$

for the definition of which we refer to [Sat63, Gro98]. The result of our interest in the following.

Proposition 14.15 (3.6 in [Gro98])**.** *The Satake transform yields an isomorphism of rings:*

$$\mathcal{H}(G, K) \otimes \mathbb{Z}[q^{1/2}, q^{-1/2}] \simeq R(G^L) \otimes \mathbb{Z}[q^{1/2}, q^{-1/2}]\ . \tag{14.56}$$

There exists a sheaf-theoretic generalization of the Satake correspondence, dubbed the *geometric Satake correspondence*, introduced in [Gin95].

The philosophy we take out of this correspondence is that the spherical Hecke algebra corresponding to G^L should play a role in the description of G-higher laminations. According to what we have presented in Section 14.1, G-higher laminations correspond to assignments of G-representations to the edges of a triangulation of the surface under consideration. The ring structure on the space of regular functions on $\mathcal{L}_{G,S}$ corresponds to the one on $R(G)$, which is related to the spherical Hecke algebra of G^L.

14.4 Wiring diagrams and higher laminations

14.4.1 Wiring diagrams on the cylinder

Since the elements of $\mathcal{H}(G,K)$ are functions over $W\backslash\widehat{W}/W$, it is of interest to describe the elements of $\widehat{W}$ and of $W\backslash\widehat{W}/W$ in terms that can lead to a generalization of laminations as equivalence classes of curves on S. To this end we will represent the elements of $\widehat{W}$ as a wiring diagram on a cylinder, as in [FM16a]. Let us assume that $G = \mathrm{SL}_N$, so that its affine Weyl group admits the presentation

$$\widehat{W}_{A_{N-1}} = \left\langle \sigma_0, \dots, \sigma_{N-1} \mid \sigma_i^2 = 1,\ (\sigma_i\sigma_j)^3 = 1,\ i,j \in \mathbb{Z}/N\mathbb{Z},\ |i-j| = 1 \right\rangle\ . \tag{14.57}$$

Starting with a horizontal cylinder with N vertices marked on each side and labeled $1,\dots,N$ as in Figure 14.9, one can associate each generator of $\widehat{W}_{A_{N-1}}$ to an elementary wiring diagram in which strands connect the N vertices on the left end of the cylinder to the N vertices on the right end. For every $i = 1,\dots,N-1$, the generator σ_i corresponds to the set of strands which maps i (resp $i+1$) on the left to $i+1$ (resp. i) on the right and any $k \neq i, i+1$ to itself, as shown in Figure 14.9. The generator σ_0 is associated to the affine root of $\widehat{W}_{A_{N-1}}$ and corresponds to the rightmost winding diagram in Figure 14.9 where strands connect each $k \neq 1, N$ to itself, and where 1 and N are exchanged from behind the cylinder.

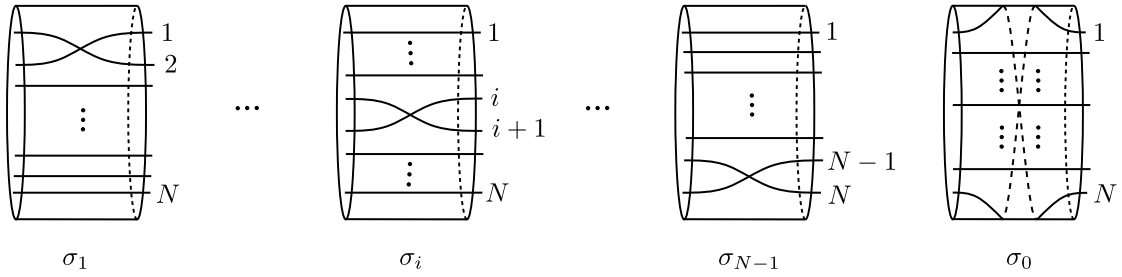


Figure 14.9: The generators of $\widehat{W}_{\mathrm{SL}_N}$ as elementary wiring diagrams.

Any element of $\widehat{W}_{A_{N-1}}$ can be obtained by assembling these elementary cylinders together. Relations in the affine Weyl group are encoded as homotopies in the wiring diagram relatively to the boundaries. For example, the relation $\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2$ is shown as a homotopy equivalence of two wiring diagrams in Figure 14.10.

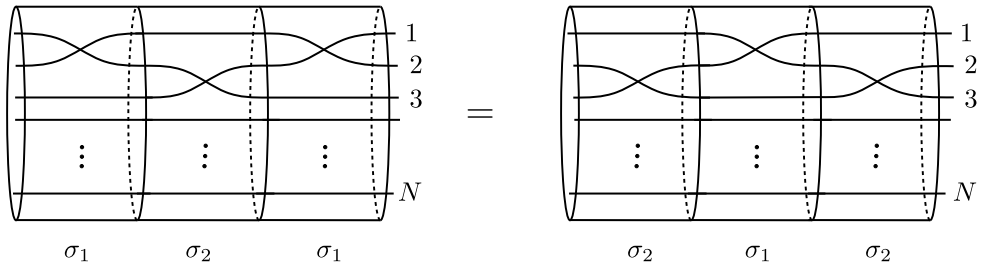


Figure 14.10: Relations in the Weyl group as homotopies of wiring diagrams.

The affine Weyl group $\widehat{W}_{A_{N-1}}$ admits a central co-extension (see [Kac90] and [FM16a] for the specific case we are interested in) $\widehat{W}^{\#}_{A_{N-1}} = \widehat{W}_{A_{N-1}} \rtimes \mathbb{Z}/N\mathbb{Z}$ with presentation:

$$\left\langle \sigma_0, \dots, \sigma_{N-1}, \Lambda \mid \sigma_i^2 = 1,\ (\sigma_i\sigma_j)^3 = 1, \Lambda\sigma_{i+1} = \sigma_i\Lambda,\ \Lambda^N = 1,\ i,j \in \mathbb{Z}/N\mathbb{Z},\ |i-j| = 1 \right\rangle\ . \tag{14.58}$$

The elementary wiring diagram corresponding to Λ maps each $i \in \mathbb{Z}/N\mathbb{Z}$ to $i+1$, as shown on the left of Figure 14.11. The additional relation $\Lambda\sigma_{i+1} = \sigma_i\Lambda$ is also readily described in terms of wiring diagrams (an example is shown in Figure 14.11) while $\Lambda^N = 1$ admits an obvious equivalent as the wiring diagram of Λ^N.

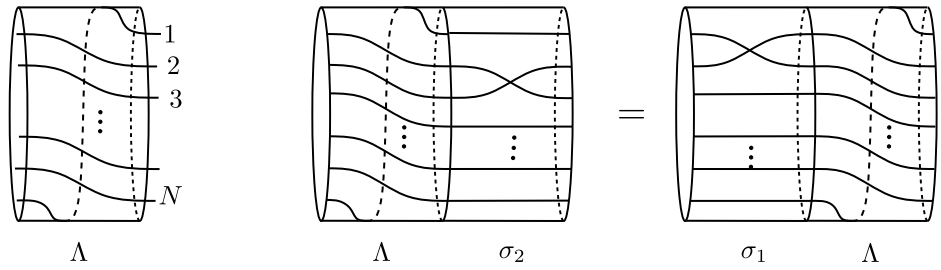


Figure 14.11: The new generator and relations in $W^{\#}_{A_{N-1}}$.

14.4.2 The Satake correspondence revisited

The Satake correspondence introduced in Section 14.3 lets one expect a one-to-one correspondence between wiring diagrams corresponding to the elements of $W\backslash\widehat{W}/W$ and isomorphism classes of irreducible representations of $\mathrm{PGL}_N(\mathbb{C})$ on the one hand, as well as the ones corresponding to the elements of $W\backslash\widehat{W}^{\#}/W$ and the isomorphism classes of irreducible representations of $\mathrm{SL}_N(\mathbb{C})$ on the other hand. This goes as follow.

Isomorphism classes of irreducible representations of $\mathrm{SL}_N(\mathbb{C})$ are labeled by their highest weight $a_1\omega_1 + \cdots + a_{N-1}\omega_{N-1}$, with $\omega_1, \ldots, \omega_{N-1}$ the fundamental weights of $\mathrm{SL}_N(\mathbb{C})$ and $a_1, \ldots, a_{N-1} \in \mathbb{Z}_{\geq 0}$. The root lattice is contained in this weight lattice, and its intersection with the fundamental Weyl chamber consists of the weights $a_1\omega_1 + \cdots + a_{N-1}\omega_{N-1}$ such that $a_1 + 2a_2 + \cdots + (N-1)a_{N-1} \equiv 0\ [N]$.

Consider an arbitrary wiring diagram on the cylinder connecting N points on the left end to N points on the right end, labeled $1, \ldots, N$ as before. This wiring diagram can be considered as an element of $\widehat{W}^{\#}_{A_{N-1}}$. Since we are interested in the elements of $W_{A_{N-1}}\backslash\widehat{W}^{\#}_{A_{N-1}}/W_{A_{N-1}}$ rather than those of $\widehat{W}^{\#}_{A_{N-1}}$, we can append any wiring diagram representing a product of $\sigma_1, \ldots, \sigma_{N-1}$ on the left, and similarly on the right. This freedom always allows us to consider wiring diagrams such that the i-th marked point on the left is linked to the i-th marked point on the right. The strand connecting these two marked points can wind around the cylinder. Let $k_i \in \mathbb{Z}$ be this winding. Eventually, any (equivalence class of a) wiring diagram corresponding to an element of $W_{A_{N-1}}\backslash\widehat{W}^{\#}_{A_{N-1}}/W_{A_{N-1}}$ yields a n-tuple of integers $(k_1, \ldots, k_N) \in \mathbb{Z}^N$. An example of this procedure for $N = 4$ is shown in Figure 14.12.

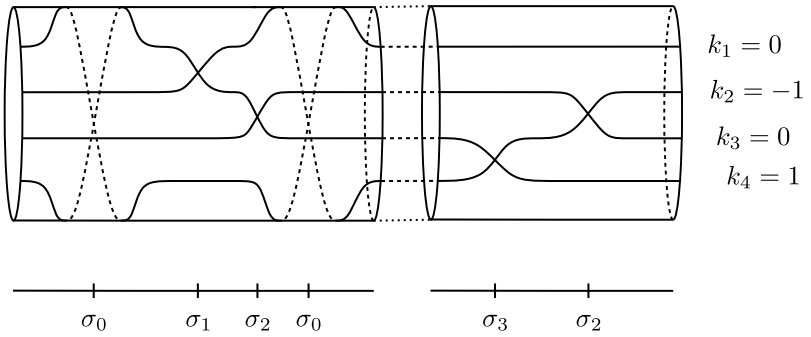


Figure 14.12: From a wiring diagram to a tuple of integers.

Actually, in order to obtain an N-tuple of integers it is enough to consider the quotient set $\widehat{W}^{\#}_{A_{N-1}}/W_{A_{N-1}}$ as we have done in Figure 14.12. Considering the double cosets $W_{A_{N-1}}\backslash\widehat{W}^{\#}_{A_{N-1}}/W_{A_{N-1}}$ allows to further assume that $k_1 \geq k_2 \geq \ldots k_N$. Now one can define

$$(a_1, \ldots, a_{N-1}) = (k_1 - k_2, \ldots, k_{N-1} - k_N) \in \mathbb{Z}^{N-1}_{\geq 0}\ . \tag{14.59}$$

Hence we have obtained a correspondence between equivalence classes of wiring diagrams describing the elements of $W_{A_{N-1}}\backslash\widehat{W}^{\#}_{A_{N-1}}/W_{A_{N-1}}$ and isomorphism classes of irreducible finite-dimensional representations of $\mathrm{SL}_n(\mathbb{C})$.

In order to obtain the representations of $\mathrm{PGL}_N(\mathbb{C})$ only, one needs to consider not all the wiring

diagram but only those such that

$$\sum_{i=1}^{N} k_i \equiv 0 \ [N] \ . \tag{14.60}$$

This follows from the fact that

$$\sum_{i=1}^{N-1} i a_i = \sum_{i=1}^{N-1} i(k_i - k_{i+1}) = \sum_{i=1}^{N} k_i \ . \tag{14.61}$$

Using relations in $\widehat{W}^{\#}_{A_{N-1}}$, any element can be written as a word in $\sigma_0, \dots, \sigma_{N-1}$ times a power of Λ. Since every σ_i contributes equally to the winding of a strand in one direction and of another in the other direction, the power of Λ of a word in $\widehat{W}^{\#}_{A_{N-1}}$ is $\sum_i k_i$. Our claim hence follows from the relation $\Lambda^N = 1$ in $\widehat{W}^{\#}_{A_{N-1}}$. For example, the diagram in Figure 14.12 corresponds to the representation of $\mathrm{PGL}_4(\mathbb{C})$ whose highest weight is

$$\omega_1 + \omega_3 \ . \tag{14.62}$$

The i-th fundamental representation of $\mathrm{SL}_N(\mathbb{C})$ (whose highest weight is ω_i) can be easily obtained by considering a wiring diagram such that $k_1 = \dots = k_i = 1$ and $k_{i+1} = \dots = k_N = 0$. An example for the fundamental representation with highest weight ω_2 for $\mathrm{SL}_4(\mathbb{C})$ is shown in Figure 14.13.

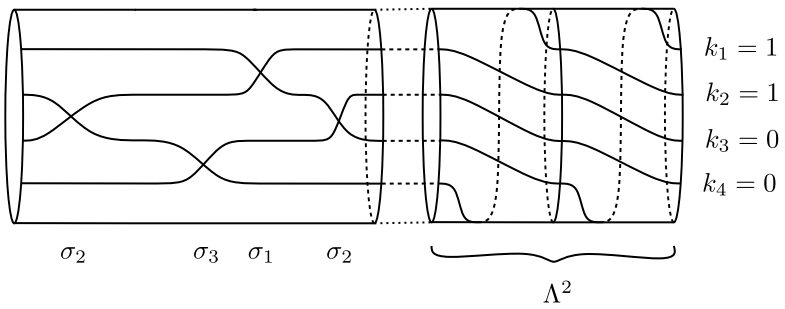


Figure 14.13: A wiring diagram corresponding to the representation ω_2 of $\mathrm{SL}_4(\mathbb{C})$.

Note that the product $\sigma_i\sigma_{i+1}\dots\sigma_{i+j-1}$ maps i to $i+j$, and every $k \in [|i+1, i+j|]$ to $k-1$. Hence, the wiring diagram associated to:

$$(\sigma_i\sigma_{i+1}\dots\sigma_{N-1})(\sigma_{i-1}\sigma_i\dots\sigma_{N-2})(\sigma_{i-2}\dots)(\sigma_{i-3}\dots)\dots(\sigma_1\dots\sigma_{N-i})\Lambda^i \tag{14.63}$$

corresponds to the N-tuple $(k_1, \dots, k_N)$ such that $k_1 = \dots = k_i = 1$ and $k_{i+1} = \dots = k_N = 0$, i.e. to the fundamental representation with highest weight ω_i of $\mathrm{SL}_n(\mathbb{C})$.

14.4.3 Higher laminations in triangles from spherical Hecke algebras

In this subsection we will show how computations in the spherical Hecke algebra give rise to higher laminations in triangles. We will study an example in a detailed way, which hopefully will clarify the philosophy of our construction.

Let us consider the words $\sigma_0\sigma_1\sigma_2\sigma_1$ and $\sigma_1\sigma_2\sigma_1\sigma_0$ in $\widehat{W}_{A_2}$, considered as representatives of two classes in $W_{A_2}\backslash\widehat{W}_{A_2}/W_{A_2} \subset W_{A_2}\backslash\widehat{W}^{\#}_{A_2}/W_{A_2}$. Using the techniques of the previous subsections, one sees that these two words correspond to the representation $\Gamma_{1,1}$ of $\mathrm{SL}_3(\mathbb{C})$, which is also a representation of $\mathrm{PGL}_3(\mathbb{C})$. Let us compute their product in the affine Hecke algebra of $\mathrm{SL}_3(\mathbb{C})$.

$$\begin{aligned}
(h_{\sigma_0}h_{\sigma_1}h_{\sigma_2}h_{\sigma_1})(h_{\sigma_1}h_{\sigma_2}h_{\sigma_1}h_{\sigma_0}) &= (q-1)h_{\sigma_0}h_{\sigma_1}h_{\sigma_2}h_{\sigma_1}h_{\sigma_2}h_{\sigma_1}h_{\sigma_0} + qh_{\sigma_0}h_{\sigma_1}h_{\sigma_2}h_{\sigma_2}h_{\sigma_1}h_{\sigma_0} \\
&= (q-1)h_{\sigma_0}h_{\sigma_1}^2h_{\sigma_2}h_{\sigma_1}^2h_{\sigma_0} + q(q-1)h_{\sigma_0}h_{\sigma_1}h_{\sigma_2}h_{\sigma_1}h_{\sigma_0} + q^2h_{\sigma_0}h_{\sigma_1}^2h_{\sigma_0} \\
&= (q-1)^3h_{\sigma_0}h_{\sigma_1}h_{\sigma_2}h_{\sigma_1}h_{\sigma_0} + q(q-1)^2h_{\sigma_0}h_{\sigma_1}h_{\sigma_2}h_{\sigma_0} + q(q-1)^2h_{\sigma_0}h_{\sigma_2}h_{\sigma_1}h_{\sigma_0} \\
&+ q(q-1)h_{\sigma_0}h_{\sigma_1}h_{\sigma_2}h_{\sigma_1}h_{\sigma_0} + q^2(q-1)h_{\sigma_0}h_{\sigma_1}h_{\sigma_0} + q^3(q-1)h_{\sigma_0} + q^4 \ ,
\end{aligned} \tag{14.64}$$

where the first equality comes from the quadratic relation $h_{\sigma_1}^2 = (q-1)h_{\sigma_1} + q$, while in going from the first line to the second we have used the braid relation $h_{\sigma_2}h_{\sigma_1}h_{\sigma_2} = h_{\sigma_1}h_{\sigma_2}h_{\sigma_1}$ in the first term and the quadratic relation $h_{\sigma_2}^2 = (q-1)h_{\sigma_2} + q$ in the second.

The computation can be described as seven (since there are seven terms in Equation (14.64)) SL_3-higher laminations in a triangle, displayed in Figure 14.14. One starts by assigning the word $\sigma_0\sigma_1\sigma_2\sigma_1$ to the left edge oriented clockwise, and $\sigma_1\sigma_2\sigma_1\sigma_0$ to the right one, also oriented clockwise. In Figure 14.14, red (resp. blue, green) edges correspond to the positive simple root σ_0 (resp. σ_1, σ_2) of SL_3. These diagrams encode the steps of the computation from bottom to top. For example, the first term in Equation (14.64) is obtained by using the quadratic relation for h_{σ_1} first, keeping the term proportional to h_{σ_1}, and then the braid relation $h_{\sigma_2}h_{\sigma_1}h_{\sigma_2} = h_{\sigma_1}h_{\sigma_2}h_{\sigma_1}$. As for the fourth term, it arises after using the quadratic relation for h_{σ_1} first, keeping the term proportional to the unit in the Hecke algebra, and then the quadratic relation for h_{σ_2}, keeping the term proportional to h_{σ_2}. Lastly, note that the power of $(q-1)$ in front of each term equals the number of trivalent vertices in the corresponding diagram, while the power of q can be related to the 'horizontal' curves.

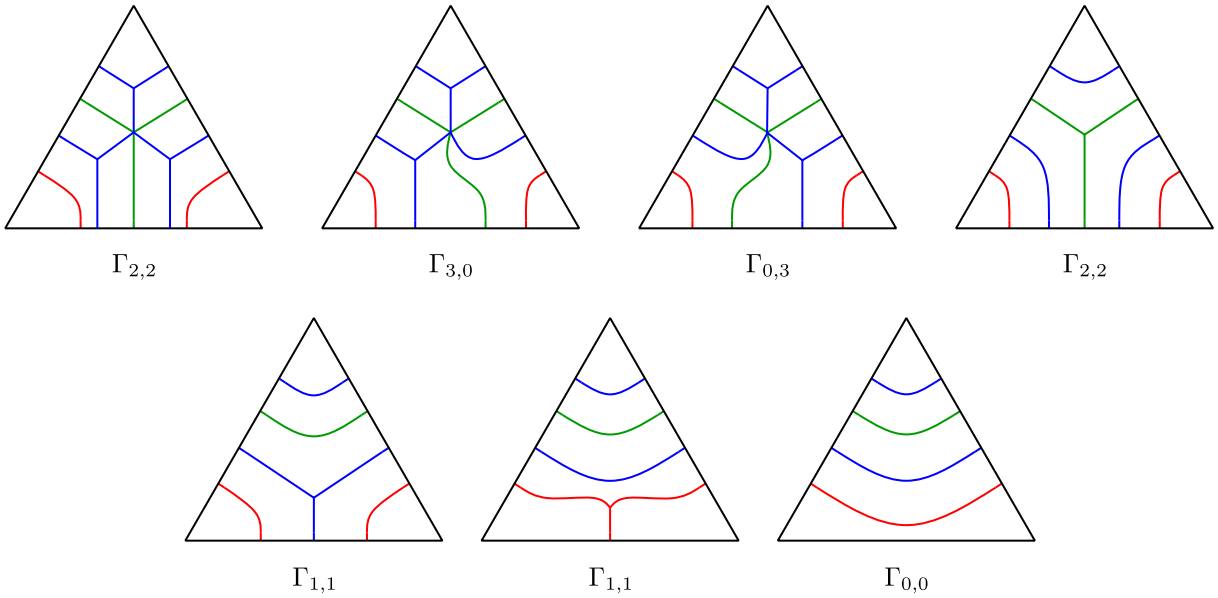


Figure 14.14: A graphical representation of computations in Hecke algebras.

The methods of the previous subsections allow one to associate a representation of SL_3 to each term of the result in Equation (14.64). These respectively correspond to $\Gamma_{2,2}, \Gamma_{3,0}, \Gamma_{0,3}, \Gamma_{2,2}, \Gamma_{1,1}, \Gamma_{1,1}$ and $\Gamma_{0,0}$. This shows that this definition of higher laminations is on the right track: the representation $\Gamma_{1,1} \otimes \Gamma_{1,1}$ splits as

$$\Gamma_{1,1} \otimes \Gamma_{1,1} = \Gamma_{2,2} \oplus \Gamma_{3,0} \oplus \Gamma_{0,3} \oplus \Gamma_{1,1} \oplus \Gamma_{1,1} \oplus \Gamma_{0,0} \ , \tag{14.65}$$

as can be verified using hives, for example. In terms of dimensions this yields

$$64 = 27 + 10 + 10 + 8 + 8 + 1 \ . \tag{14.66}$$

There is one issue however: in the affine Hecke algebra computation we have obtained two terms corresponding to $\Gamma_{2,2}$, while there is only one in the decomposition of $\Gamma_{1,1} \otimes \Gamma_{1,1}$. This raises the question of the equivalence relation to be imposed on such diagrams. Just as classical laminations on a surface S are equivalence classes of systems of curves on S, G-higher laminations should also be defined as equivalence classes of diagrams appearing in Figure 14.14. Isotopy surely is part of the equivalence relation, but there must also be more, as this example demonstrates. Of course the two diagrams corresponding to $\Gamma_{2,2}$ should be equivalent when considered in the spherical Hecke algebra instead of the affine one; however an explicit and complete description of the equivalence relation remains elusive.

Last, let us note that Λ can also be represented as a curve in diagrams in a triangle. However contrarily to the ones corresponding to the σ_i the curve representing Λ needs to be oriented, since $\Lambda \neq \Lambda^{-1}$ in general.

14.4.4 Classical laminations from the spherical Hecke algebra of SL_2

Let us explain very shortly the rough idea of how classical laminations can be retrieved by considering higher laminations as defined above, for $G = \mathrm{SL}_2$.

Consider the SL_2-higher laminations in Figure 14.15. All of them are obtained as describing the terms appearing in the following computation in $\widehat{W}_{A_1}$:

$$\begin{aligned} &h_{\sigma_0}h_{\sigma_1}h_{\sigma_0}h_{\sigma_1}h_{\sigma_0}h_{\sigma_1} \cdot h_{\sigma_1}h_{\sigma_0} = (q-1)h_{\sigma_0}h_{\sigma_1}h_{\sigma_0}h_{\sigma_1}h_{\sigma_0}h_{\sigma_1}h_{\sigma_0} + qh_{\sigma_0}h_{\sigma_1}h_{\sigma_0}h_{\sigma_1}h_{\sigma_0}h_{\sigma_0} \\ =&(q-1)h_{\sigma_0}h_{\sigma_1}h_{\sigma_0}h_{\sigma_1}h_{\sigma_0}h_{\sigma_1}h_{\sigma_0} + q(q-1)h_{\sigma_0}h_{\sigma_1}h_{\sigma_0}h_{\sigma_1}h_{\sigma_0} + q^2h_{\sigma_0}h_{\sigma_1}h_{\sigma_0}h_{\sigma_1} \end{aligned} \tag{14.67}$$

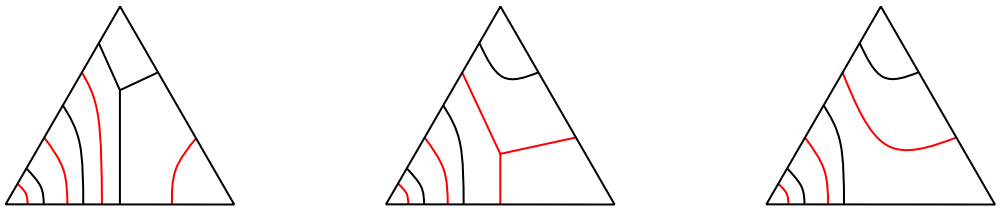


Figure 14.15: Three SL_2-higher laminations.

Removing the singular curve in each case yields three classical $\mathcal{A}$-laminations in a triangle, corresponding to the decomposition

$$V_{3/2} \otimes V_{1/2} = V_2 \oplus V_1 \oplus V_0 \tag{14.68}$$

of SL_2-representations. Note however that according to the correspondence between words $W\backslash W^{\#}/W$ and representations of the Langlands dual group presented in the previous subsections, the diagrams in Figure 14.15 rather correspond to the decomposition

$$V_3 \otimes V_1 = V_4 \oplus V_3 \oplus V_2 \ . \tag{14.69}$$

This discrepancy is likely not uncorrelated with the fact that one needs to consider twice the integral tropical coordinates in the reconstruction of classical $\mathcal{A}$-laminations; it needs to be investigated further.

14.5 Spectral networks

Spectral networks have been introduced by Gaiotto, Moore and Neitzke in their work on BPS counting in theories of class S [GMN10, GMN13c, GMN13a, GMN12, GMN13b, GMN14]. Spectral networks – as well as other combinatorial objects appearing in the theory – resemble higher laminations (as defined in Conjecture 14.5) very much. Spectral networks have been a guideline throughout our quest towards higher laminations, and it is likely that much more can be learned in higher Teichmüller theory from this perspective. The goal of this section is to provide a very brief overview of the physical ideas appearing in this field.

14.5.1 Theories of class S

We have discussed Lagrangian four-dimensional $\mathcal{N} = 2$ theories in Section 5.6 as well as Seiberg–Witten theory. Moreover, we have reviewed in Section 6.4.1 how Seiberg–Witten curves could be constructed by lifting a configuration of D4- and NS5-branes to M-theory.

The low-energy dynamics of a large class of four-dimensional $\mathcal{N} = 2$ theories can be described via brane configurations. For example, superconformal theories described by linear quivers can be obtained by considering K infinite D4-branes intersecting $n+1$ NS5-branes, which yields a theory with $\mathrm{SU}(K)^n$ gauge group and whose matter is encoded in the linear quiver of n nodes with K fundamental hypermultiplets for the first and last simple gauge factors. The Seiberg–Witten (SW) curve in this case is computed in [Wit97]; it is of degree K in v and $n+1$ in t. For instance, when $K = 2$ it can be written as:

$$v^2t^{n+1} + c_1(v^2 - u_1)t^n + \cdots + c_n(v^2 - u_n)t + c_{n+1}v^2 = 0 \ . \tag{14.70}$$

The u_i's parameterize the Coulomb branch of the moduli space, whereas the c_i's encode the gauge couplings of the theory. Following [Gai12], when the hypermultiplets have a zero bare mass one can rewrite the equation of the SW curve as

$$\prod_{a=0}^{n}(t-t_a)v^2 = U(t)t \tag{14.71}$$

where $U(t) \in \mathbb{C}_{n-1}[t]$. The a-th gauge coupling for $a = 1, \dots, n$ then writes

$$\tau_a = \frac{1}{i\pi}\log\frac{t_{a-1}}{t_a} \ , \tag{14.72}$$

while the Seiberg–Witten differential is:

$$\lambda = v\frac{\mathrm{d}t}{t} \ . \tag{14.73}$$

Again as in [Gai12] and before generalizing the description we will obtain, let us consider the case $K = 2$ and $n + 1 = 2$, which corresponds to four-dimensional $\mathcal{N} = 2$ SU(2) gauge theory with $N_f = 4$. The equation of the SW curve in the form of Equation (14.71) reads:

$$(t-1)(t-t_1)v^2 = ut \ . \tag{14.74}$$

Upon the change of variables $v = tx$, this equation becomes

$$t(t-1)(t-t_1)x^2 = u \ , \tag{14.75}$$

while the Seiberg–Witten differential now writes $\lambda = x\mathrm{d}t$. If one further does the replacements

$$t \to \frac{az+b}{cz+d} \ , \qquad x \to (cz+d)^2 x \tag{14.76}$$

for generic $a, b, c, d \in \mathbb{C}$, the equation of the SW curve becomes

$$x^2 = \frac{u}{\Delta_4(z)} \ , \tag{14.77}$$

where $\Delta_4(z)$ is a degree-four polynomial in z. The t_1 appearing in Equation (14.75) is the cross-ratio of its roots.

The SW differential has become $\lambda = x\mathrm{d}z$, and thus z can be considered as a coordinate on a sphere S with four punctures, while x is a coordinate on the fibers of the cotangent bundle of that sphere. In that picture, the differential λ is the restriction of the canonical Liouville 1-form on T^*S to the degree-2 ramified cover of the base S in T^*S defined by Equation (14.77).

This generalizes to arbitrary SU(2) linear quivers with gauge group $\mathrm{SU}(2)^n$, for which the equation of the SW curve can be written as

$$x^2 = \phi_2(z) \ , \tag{14.78}$$

and the SW differential, as $\lambda = x\mathrm{d}z$. One can naturally interpret $\phi_2(z)\mathrm{d}z^2$ as a quadratic differential on the sphere with $n + 3$ punctures $S_{0,n-3}$, which has a simple pole at each puncture of $S_{0,n-3}$. The space of gauge couplings of the theory is identified with the classical Teichmüller space of $S_{0,n-3}$, on which the mapping class group acts as S-duality.

When mass parameters are included, they are encoded in the residues at the simple poles of the Seiberg–Witten differential.

A new class of theories analyzed in [Gai12] are described by *generalized quivers*, constructed from elementary building blocks. For example, generalized SU(2)-quivers are obtained by gluing together copies of the elementary quiver encoding the $\mathcal{N} = 2$ SU(2) gauge theory with four fundamental hypermultiplets, depicted on the left of Figure 14.16. The four flavor SU(2) groups appear in the splitting of the SO(8) flavor symmetry of the SU(2) gauge theory with $N_f = 4$ as:

$$\mathrm{SO}(8) \to \mathrm{SO}(4) \times \mathrm{SO}(4) \simeq \mathrm{SU}(2) \times \mathrm{SU}(2) \times \mathrm{SU}(2) \times \mathrm{SU}(2) \ . \tag{14.79}$$

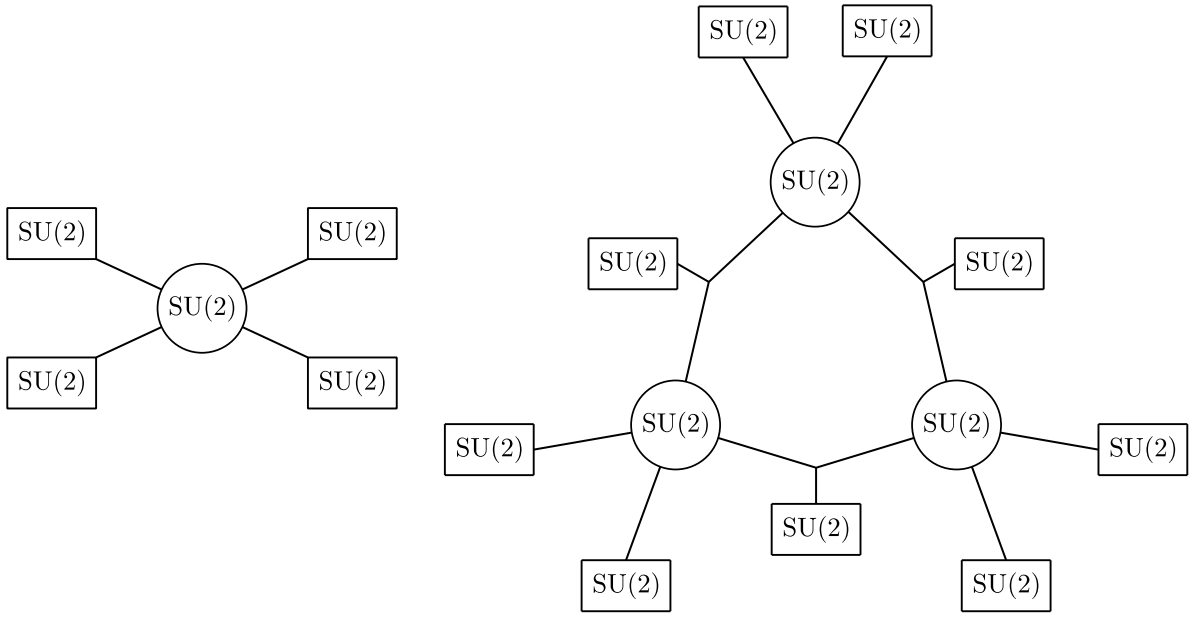


Figure 14.16: Generalized SU(2)-quivers: elementary building block (left) and one example (right).

An example of a generalized SU(2)-quiver is shown on the right of Figure 14.16. These graphs are characterized by the number n of gauge groups and the number g of loops. Under successive S-dualities at single nodes, one can transform a generalized quiver into any other with the same n and g. The family of generalized SU(2) quivers with (n, g) fixed corresponds to a superconformal field theory denoted $\mathcal{T}_{n,g}[A_1]$. Its parameter space of exactly marginal deformations is identified with the Teichmüller space of the punctured surface $S_{g,n}$ of genus g and n punctures, on which the mapping class group acts as S-duality. The weak coupling limits of this theory correspond to degeneration limits of the Riemann surfaces with topology $S_{g,n}$.

These theories can be constructed uniformly in M-theory. One starts by considering a Riemann surface $C = S_{g,n}$ of genus g with n punctures dubbed the *UV curve*, and a stack of two NS5-branes wrapped on it. The worldvolume theory on a stack of two NS5-branes is the *six-dimensional* $\mathcal{N} = (2, 0)$ *superconformal theory of type* A_1, which if appropriately twisted when compactified on $S_{g,n}$ yields a four-dimensional quantum field theory with $\mathcal{N} = 2$ supersymmetry.

The UV curve C is seen as the zero section of the cotangent bundle T^*C. At each puncture one considers a transverse NS5-brane localized on C at the puncture, and sweeping the fiber in T^*C above it. From the point of view of the six-dimensional $\mathcal{N} = (2, 0)$ A_1 SCFT, these transverse NS5-branes appear as codimension-two defects. The Coulomb branch of the corresponding four-dimensional theory corresponds to normalizable deformations of this setup into a single M5-brane wrapping a complex ramified cover of degree two $\Sigma \to C$ in $T^*S_{g,n}$. This cover Σ is determined by the choice of a holomorphic differential $\phi_2(z)\mathrm{d}z^2$ on C, and it is the SW curve of the theory at the point of the Coulomb branch corresponding to $\phi_2(z)$. As before, the SW differential is the restriction of the Liouville 1-form $\lambda = x\mathrm{d}z$ to Σ.

There is a six-dimensional $\mathcal{N} = (2, 0)$ SCFT for each finite-dimensional simply-laced Lie algebra $\mathfrak{g}$, yielding via the same construction a four-dimensional $\mathcal{N} = 2$ theory whose Coulomb branch is $\mathbb{C}^{\mathrm{rk}(\mathfrak{g})}$. These theories are said to be of *class S* (where S stands for six), and denoted $\mathrm{S}[\mathfrak{g}, C, D]$, where D is denotes the defects data, as the transverse NS5-branes of before.

The six-dimensional SCFT corresponding to $\mathfrak{g}$ of type A_K is the worldvolume theory on a stack of $K + 1$ coincident NS5-branes, and hence the theories $\mathrm{S}[A_K, C, D]$ arise as the low-energy limit of a stack of $K + 1$ coincident NS5-branes wrapped on C, together with the set of defects D. This construction is developed thourougly in [GMN13c], in which it is shown that the Coulomb branch of such a theory can be identified with the base of the *Hitchin moduli space* of SU$(K + 1)$-Higgs bundles on C. The whole Hitchin moduli space appears as one compactifies $\mathrm{S}[\mathfrak{g}, C, D]$ further on S^1, as the moduli space of the resulting low-energy three-dimensional supersymmetric quantum field theory [GMN10].

14.5.2 BPS states

BPS states at a point of the Coulomb branch of a theory of class S of type A_K can be described geometrically as M2-branes streching between the different leaves of the SW curve Σ within T^*C [Mik98, GMN13c]. In terms of the six-dimensional theory, the boundaries of M2-branes appear as *strings with variable tension* [KLM$^+$96] – see also [Tac13]. Away from the branching locus of the covering $\Sigma \to C$, these curves are locally labeled by a pair of numbers ij where $i < j \in [|1, K+1|]$. This encodes that the corresponding membrane ends on the i-th and the j-th sheets of Σ. The tension of an ij-string at $p \in C$ is $\lambda(p_i) - \lambda(p_j)$, where p_i and p_j are the i-th and j-th lifts of p.

The central charge of a segment of ij-string extending along a curve $c \subset S_{g,n}$ is

$$Z = \frac{1}{\pi} \int_c \lambda_j - \lambda_j \ , \tag{14.80}$$

where λ_i and λ_j are the restrictions of the Liouville 1-form to the i-th and j-th sheets of Σ respectively, whereas the mass of such a segment is

$$M = \frac{1}{\pi} \int_c |\lambda_j - \lambda_j| \ . \tag{14.81}$$

The BPS bound $M \geq |Z|$ is thus saturated if and only if the phase $e^{i\theta}$ of $\lambda_j - \lambda_j$ is constant along c, in which case $Z = e^{i\theta} M$. BPS strings can begin and end at the punctures of C, branch points of $\Sigma \to C$, and they can also close. Finite mass BPS states in the four-dimensional theory of class S correspond to finite such BPS strings, or finite webs thereof. If a finite BPS string begins and ends at branch points it describes a BPS hypermultiplet, whereas when it is closed it corresponds to a BPS vector multiplet. When lifted to the Seiberg–Witten curve Σ, a web of finite BPS strings becomes a representative of a class in $H^1(\Sigma, \mathbb{Z})$, which encodes its (electric, magnetic, flavor) charges under the low-energy symmetry group $\mathrm{U}(1)^{K-1} \times \mathrm{U}(1)^F$, where F is the rank of the global flavor symmetry group (encoded in D).

The moduli space of the low-energy effective three-dimensional theory obtained by compactifying the theory of class S at hand on S^1 is hyperkähler, with metric the naive *semi-flat metric* modified by quantum corrections. The latter only depend on the four-dimensional BPS states that wind around S^1, and are weighted by the so-called *second helicity supertraces* $\Omega(\gamma; u)$, which are the indices of BPS particles of charge $\gamma \in H^1(\Sigma, \mathbb{Z})$ at a point u of the Coulomb branch. The smoothness of the metric implies that the BPS spectrum must jump along some codimension-one locus on the Coulomb branch [GMN10]: this is the *wall-crossing phenomenon*. Wall-crossing can arise when the complex central charges Z_γ and $Z_{\gamma'}$ of BPS states of charge γ and γ' are aligned: when it is the case, the bound state of charge $\gamma + \gamma'$ is not protected against decay into two BPS particles of respective charge γ and γ'.

Let us once more restrict to the case $N = 1$. Let $\lambda = \lambda_1 - \lambda_2$ and fix some $\theta \in \mathbb{R}/2\pi$. Following [GMN13c], a *WKB curve* with angle θ is a curve on $S_{g,n}$ such that if ∂_t is a tangent vector along c:

$$\lambda \cdot \partial_t \in e^{i\theta} \mathbb{R}^\times \ , \tag{14.82}$$

at all point of c. The union of WKB curves with angle θ forms the *WKB foliation* with angle θ. Near a singularity on C, the local behaviour of the WKB foliation depends only on the sign of $e^{-i\theta} m$, where m is the mass parameter corresponding to this singularity. When $e^{-i\theta} m$ is purely imaginary, the WKB curves with angle θ circle around the singularity while otherwise they spiral into it the latter (see Figure 23 in [GMN13c]).

As we have explained above, finite WKB curves correspond to finite-mass BPS states of the four-dimensional theory, i.e. those states that matter for the computation of the hyperkähler metric on the moduli space of the theory compactified on S^1. Considering first angles θ for which there are no finite WKB curves, it is shown in [GMN13c] that the leaves of the WKB foliation are either asymptotic in both directions to a singular point, or asymptotic in one direction to a branch point and in the other to a singular point. Thus the WKB foliation cuts the surface C in cells, which in turn define the *WKB triangulation* of C. This triangulation has its vertices at the singularities of C – and it is possibly degenerate. The triangulation is supplemented by a decoration, which is canonically defined by the point under consideration of the four-dimensional Coulomb branch.

As θ varies the WKB triangulation does as well. Generically, the edges of the triangulation undergo mere smooth deformations, however there are special values $\theta = \theta_c$ at which the triangulation jumps, and that hints for the presence of BPS states for which $Z = e^{i\theta_c} M$. An example is shown in Figure 14.17,

reproducing Figure 27 in [GMN13c]. The singularities are depicted as black dots, the crosses are branching points of $\Sigma \to C$, and the edges of the WKB triangulation are dashed whereas the leaves of the WKB foliation are shown as thin plain curves. Such a jump corresponds to a *flip* of the WKB triangulation. Two other types of jumps exist, dubbed *pops* and *juggles*.

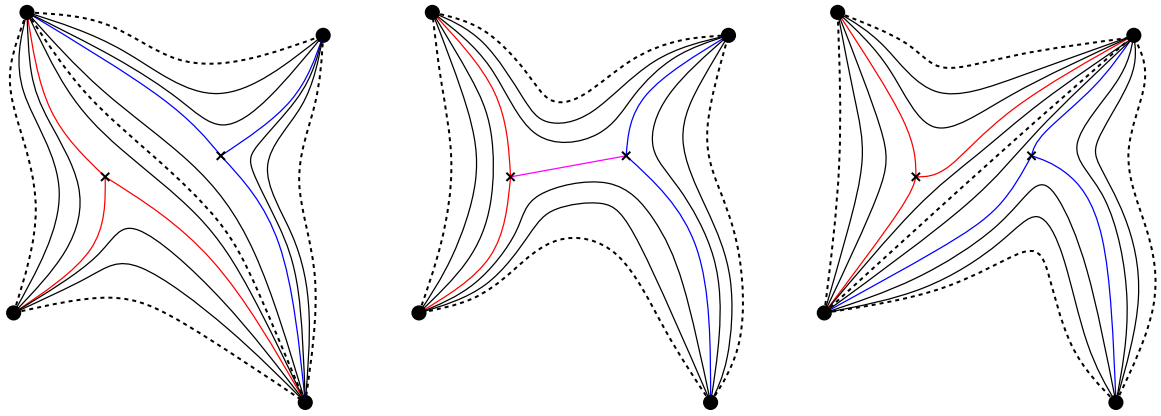


Figure 14.17: A jump in the WKB triangulation corresponding to a BPS hypermultiplet.

Central to the construction of [GMN10, GMN13c] is the family of so-called Darboux coordinates $\mathcal{X}_\gamma^{RH}$ with $\gamma \in H^1(\Sigma, \mathbb{Z})$. They appear in the computation of the hyperkähler metrics on the moduli space of the three-dimensional theory. The construction of [GMN13c] uses Fock–Goncharov coordinates: starting with the WKB triangulation associated with an angle $\theta \in \mathbb{R}/2\pi$, one constructs the Fock-Goncharov $\mathcal{X}$-coordinates $\mathcal{X}_\gamma^\theta$ for the $\mathrm{SL}_2(\mathbb{C})$-connection corresponding to the point on the Coulomb branch under consideration and the (decorated) WKB triangulation. More precisely, the connection

$$\frac{R}{\zeta}\phi + A + R\zeta\overline{\phi} \tag{14.83}$$

is flat if and only if (A, ϕ) is a solution of the Hitchin equations, i.e. corresponds to a point on the Coulomb branch of the theory. Taking $\theta = \arg(\zeta)$, the $\mathcal{X}_\gamma^\theta$ are the Darboux coordinates $\mathcal{X}_\gamma^{RH}$ of interest. This explains the name WKB triangulation: it comes from the fact that it arises as one applies the WKB method to the Higgs field ϕ for small ζ.

Keeping the point on the base of the Hitchin fibration fixed, as one varies θ the WKB triangulation undergoes flips, juggles and pops to which we alluded above. Under flips, Fock–Goncharov coordinates transform according to the cluster mutation formulae of type $\mathcal{X}$, which are a particular case of the Kontsevich–Soibelman wall-crossing formula.

In order to compute the BPS spectrum of a four-dimensional theory of class S, one picks a θ and keep track on how the decorated WKB triangulation varies as θ increases to $\theta + \pi$. It turns out that the two triangulations at θ and at $\theta + \pi$ are related by a mere pop at all vertices, and the corresponding transformation on the Darboux coordinate is uniquely expressed as

$$\mathbf{S} = \prod K_\gamma^{\Omega(\gamma;u)} \ , \tag{14.84}$$

where K_γ are the Kontsevich–Soibelman transformations. Thus, computing $\mathbf{S}$ yields all the information about the BPS spectrum at the point of the Coulomb branch under consideration.

This method allows for example to compute the BPS spectrum of $\mathcal{N} = 2$ SU(2) super Yang–Mills theory with $N_f = 0, 1, 2, 3$ (obtained in [FB96, BF96] via other methods), as well as the Argyres–Douglas SCFTs [AD95] (see [Tac13] for a pedagogical review).

Spectral networks were introduced in [GMN13b], building on the previous work by the same authors. It is a generalization of the picture in type A_1 we have reviewed to the A_K case, for general K.

A spectral network is a union of *walls*, which are oriented curves on the UV curve C labeled by a pair of integers (ij) which, as at the beginning of this subsection, refer to the sheets of the cover $\Sigma \to C$ between which the M2-brane above a given curve is streching[1]. At a simple branch point of the cover

[1] In order to assign well-defined labels to the walls on C, one needs to choose branch cuts so as to trivialize the branched covering $\Sigma \to C$.

where the sheets i and j coincide, three walls of type (ij) emanate. Moreover, the crossing of walls of type (ij) and (jk) generally creates a wall of type (ik). Thus, a spectral network looks very much like the higher laminations proposed in Conjecture 14.5. We refer to [GMN13b] for a precise definition, as well as examples.

An important application of spectral networks of a more mathematical nature is (non)abelianization (see e.g. [GMN13b, HN16, Kid17]). A spectral network provides a canonical recipe to lift paths on C to paths on Σ. Using this, it is possible to construct a GL(1)-connection on Σ from any GL(N)-connection on C (this is called *abelianization*), and a GL(N)-connection on C from any GL(1)-connection on Σ (this is the *non-abelianization*).

14.5.3 Wilson–'t Hooft operators and laminations

The study of supersymmetric *Wilson–'t Hooft line operators* [Mal98, RY01, KW07] was carried out in [DMO09] for the theories described by the generalized SU(2)-quivers introduced in [Gai12] and reviewed in Section 14.5.1. It is shown that there is a one-to-one correspondence between homotopy classes of closed non self-intersecting curves on the UV curve C, and those BPS loop operators. This statement was refined in [GMN13a] where the authors showed that as soon as one allows irregular singularities for the Higgs field on C (as is the case for $\mathcal{N} = 2$ SU(2) super Yang–Mills theories with $N_f = 0, 1, 2, 3$ flavors) one needs to consider not only closed curves but also open curves ending at these singularities. It was conjectured that line operators in theories of class S of type A_1 are in one-to-one correspondence with integral $\mathcal{A}$-laminations.

This was further generalized to the A_N case in [Xie13]. The space of UV line operators is identified with the integral $\mathcal{A}$-laminations on C corresponding to PGL_N, whereas the space of IR line operators is identified with the $\mathcal{X}$-coordinates encoding SL_N-local systems on C. The *canonical map* relates the UV line operators to the IR ones. This picture is very appealing and deserves further study; the intuition from physics is likely to provide hints and guidance towards the understanding of higher laminations as defined in Conjecture 14.5, while conversely it would be interesting to see whether hives, the Satake correspondence and affine Hecke algebras could be of some relevance in the physics of BPS line operators.

14.6 Higher laminations and ramified covers

14.6.1 Higher laminations as Lagrangian surfaces in T^*S

Just as spectral networks are attached to a branched covering $\Sigma \to C$, one can expect integral higher laminations on a surface S to encode some ramified covers of S^2. More precisely, we expect higher laminations to encode Lagrangian ramified covers $\tilde{\Sigma}$ of S in T^*S such that the class of the restriction of the Liouville form λ to $\tilde{\Sigma}$ is in $H^1(\tilde{\Sigma}, \mathbb{Z})$. Let us briefly explain the line of reasoning leading to such a picture. We will restrict to the case of split Lie groups of type A. How to generalize to other types of Lie groups would be the next natural step after understanding precisely the theory for groups of type A.

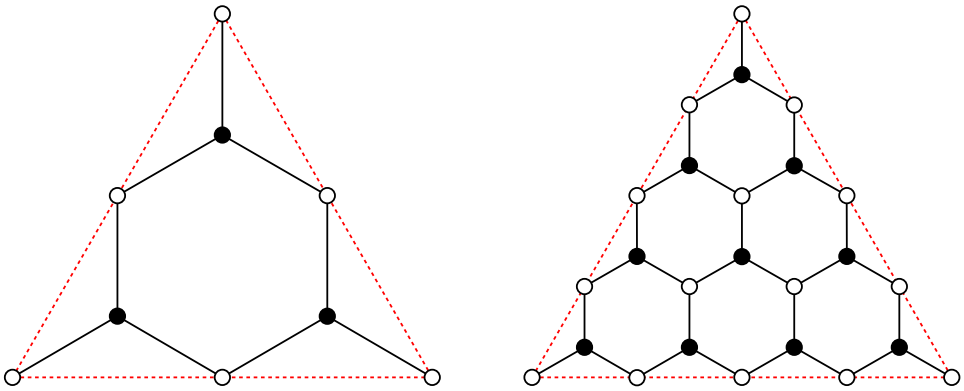


Figure 14.18: The bipartite graphs corresponding to $N = 2$ and $N = 4$.

[2]Since the counting of unramified N-coverings of a surface S involves the symmetric group $\mathfrak{S}_N$ [LZ13], it is plausible that the counting of ramified N-coverings of S involves a Hecke algebra associated to $\mathfrak{S}_N$.

As a starting point we are going to consider once more the algebraic definition of higher $\mathcal{A}$-laminations for a group G of type A_{N-1}, i.e. as the integral tropical points of the cluster $\mathcal{A}$-variety $\mathcal{A}_{S,G}$. Recall how special coordinates on the cluster varieties $\mathcal{A}_{\mathrm{SL}_N,G}$ and $\mathcal{X}_{\mathrm{PGL}_N,G}$ were constructed in Section 4.3: one first chooses a triangulation Γ of S, and construct a bipartite graph Λ_B in each face of Γ. The coordinates are naturally associated with the faces of Λ_B. The example of Λ_B in a triangle for $N = 2$ and $N = 4$ are shown in Figure 14.18: the edges of the triangulation are the dashed red lines.

The G-higher Teichmüller spaces $\mathcal{A}_{S,G}(\mathbb{R}_{>0})$ and $\mathcal{X}_{S,G}(\mathbb{R}_{>0})$ are parameterized by assigning positive real coordinates to the faces of Λ_B and asserting that these change according to the $\mathcal{A}$- (resp $\mathcal{X}$-) mutation rules. Similarly, the G-higher lamination spaces $\mathcal{A}_{S,G}(\mathbb{Z}^t)$ and $\mathcal{X}_{S,G}(\mathbb{Z}^t)$ are parameterized by assigning integral coordinates to the faces of Λ_B and asserting that these change according to the tropical $\mathcal{A}$- (resp $\mathcal{X}$-) mutation rules.

The bipartite graph Λ_B is fat graph, with the latter structure induced by the orientation of the surface S. It can thus be understood as a *dessin d'enfant* [Gro84, JW16], i.e. as a quadruple (E, s_W, s_B, G) where E is the set of edges of Λ_B, s_W (resp. s_B) is the permutation in $\mathfrak{S}(E)$ which maps each edge to the next one in the clockwise (resp. counterclockwise) direction with respect to its white (resp. black) end, and $G = \langle s_W, s_B \rangle < \mathfrak{S}(E)$ is the monodromy group. The *untwisting map*, to which we alluded in Section 7.5, can be conveniently described as a morphism of dessins d'enfants which maps (E, s_W, s_B, G) to (E, s_W, s_B^{-1}, G), i.e. it reverses the cyclic orientation of the edges incident to each black vertex.

The untwisting map changes only the fat structure of Λ_B, and does nothing to the underlying bipartite graph. Whereas the surface corresponding to the dessin d'enfant (E, s_W, s_B, G) is S, the one corresponding to (E, s_W, s_B^{-1}, G) is another surface $\tilde{\Sigma}$ on which Λ_B is also embedded. This new surface $\tilde{\Sigma}$ can naturally be seen as a ramified covering of S with simple ramification points at each black vertex. For example, the untwisting of the bipartite graph in the triangle on the left of Figure 14.18 yields the same bipartite graph embedded in a cylinder, shown on the right of Figure 14.19. The bipartite graph on the cylinder is shown as the thick black lines, while the other two preimages of the one on the triangle appear as thin pale lines and dots. We have added labels on the edges for convenience; the cylinder is a ramified covering of the triangle of degree 3.

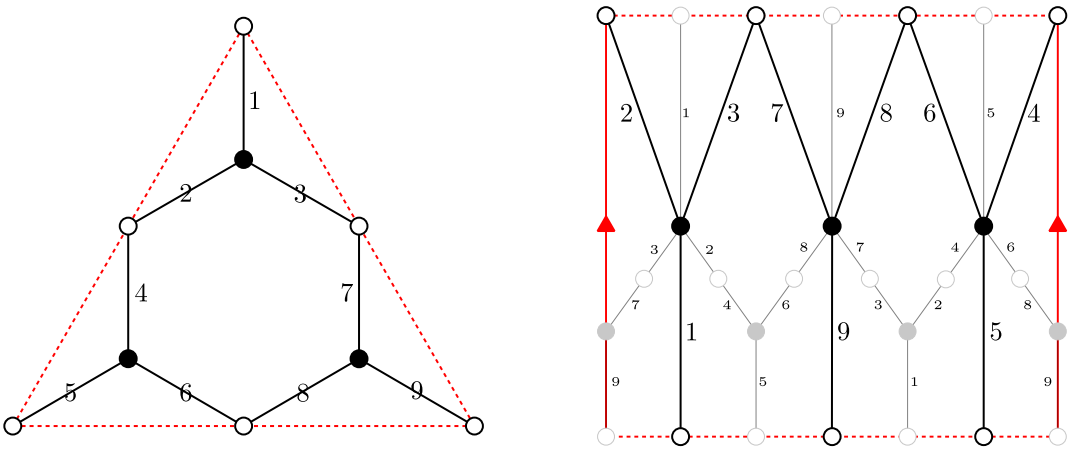


Figure 14.19: An example of the untwisting of a bipartite fat graph.

In general one can show (using the Riemann–Hurwitz formula) that the ramified covering of a triangle obtained by untwisting the bipartite graph in it corresponding to a general $N \in \mathbb{Z}_{\geq 0}$ is of degree $N + 1$. Moreover as one glues triangles together, hence forming the triangulation Γ of the surface S, the sheets of the ramified coverings of each triangle can be identified and hence untwisting the bipartite graph on S defined by Γ and N yields a ramified covering $\tilde{\Sigma}$ of degree $N + 1$ of S.

Note that in the dessin d'enfant approach, the faces of Λ_B are in one-to-one correspondence with the cycles appearing in the decomposition of the permutation xy^{-1} while zig-zag paths are in one-to-one correspondence with the cycles appearing in the decomposition of xy. Therefore, the untwisting map transforms faces into zig-zag paths, and vice-versa. Note that when S has cilia there are some open zig-zag paths, which correspond to open faces in $\Lambda_B \subset \tilde{\Sigma}$, and vice-versa. Each zig-zag path in $\tilde{\Sigma}$ corresponds to some class in the first homology group of $\tilde{\Sigma}$ relative to the boundary, and hence the assignment of

integral coordinates to the faces of $\Lambda_B \subset S$ corresponding to some point in $\mathcal{A}_{S,G}(\mathbb{Z}^t)$ corresponds to the assignment of integral coordinates to those homology classes in $\tilde{\Sigma}$. Because of this, it is natural to conjecture the following.

Conjecture 14.16. *Points of $\mathcal{A}_{S,G}(\mathbb{Z}^t)$ are in one-to-one correspondence with ramified coverings $\tilde{\Sigma}$ of S in T^*S of some sort, such that the restriction of the Liouville 1-form λ to $\tilde{\Sigma}$ is integral, i.e.*

$$\lambda_{|\tilde{\Sigma}} \in H^1(\tilde{\Sigma}, \partial\tilde{\Sigma}, \mathbb{Z}) \ , \tag{14.85}$$

modulo some equivalence relation.

This is the point where our construction steps away from spectral networks slightly: there are not enough complex ramified covers of S to account for all possible higher laminations. This can be explained roughly as follows. Let us consider for simplicity S to be a closed surface of genus g, and let $G = \mathrm{SL}_N(\mathbb{C})$. On the one hand, the moduli space $\mathrm{Hom}(\pi_1 S, G)/G$ of flat G-connections on S has the same dimension as the cotangent space $T^*\mathrm{Hol}_{G,S}$ to the space of holomorphic G-bundles on S, and thus:

$$\dim_{\mathbb{C}} T^*\mathrm{Hol}_{G,S} = \dim_{\mathbb{C}} G \cdot (2g-2) \ . \tag{14.86}$$

On the other hand, assuming that S is endowed with a complex structure as is the case in Hitchin's construction [Hit87, Hit92], the space of $\mathrm{SL}_N(\mathbb{C})$-spectral curves $\Sigma \to S$ in T^*S is parameterized by the Hitchin base:

$$\bigoplus_{i=2}^{n} H^0(S, K^{\otimes i}) \ , \tag{14.87}$$

where K is the canonical bundle on S. As usually done, the Riemann–Roch theorem implies that $\dim_{\mathbb{C}} H^0(S, K^{\otimes i}) = (2m-1)(g-1)$ and hence

$$\dim_{\mathbb{C}} \bigoplus_{i=2}^{n} H^0(S, K^{\otimes i}) = (n^2-1)(g-1) \ . \tag{14.88}$$

Hence one cannot expect the space of complex ramified coverings of S in T^*S to parameterize regular functions on $\mathrm{Hom}(\pi_1 S, G)/G$. However, a related space has the right dimension and hence is a good candidate: a generic $\mathrm{SL}_N(\mathbb{C})$-spectral curve $\tilde{\Sigma}$ on S has genus $2N^2(g-1)+2$, which implies that

$$\dim_{\mathbb{C}}(H^1(\tilde{\Sigma})/H^1(S)) = (n^2-1)(2g-2) \ . \tag{14.89}$$

Thus, it seems of interest to consider the space of surfaces $\tilde{\Sigma}$ in T^*S which are ramified coverings of S of degree N parameterized by $H^1(\tilde{\Sigma})$. It is not the space of $\mathrm{SL}_N(\mathbb{C})$-spectral curves, but rather the space such surfaces which are Lagrangian subspaces of T^*S, because the Lagrangian surfaces are exactly those surfaces to which the restriction of the Liouville 1-form λ is closed. Generators of the ring of regular functions on $\mathrm{Hom}(\pi_1 S, G)/G$ are parameterized by integral laminations, which presumably transposes into an integrality condition on the restriction of λ.

Conjecture 14.16 can therefore be refined as follows:

Conjecture 14.17. *Points of $\mathcal{A}_{S,G}(\mathbb{Z}^t)$ are in one-to-one correspondence with ramified coverings $\tilde{\Sigma}$ of S in T^*S which are Lagrangian subspaces and such that the restriction of the Liouville 1-form λ to $\tilde{\Sigma}$ is integral, i.e.*

$$\lambda_{|\tilde{\Sigma}} \in H^1(\tilde{\Sigma}, \partial\tilde{\Sigma}, \mathbb{Z}) \ , \tag{14.90}$$

modulo Hamiltonian diffeomorphisms preserving the zero section.

14.6.2 From integral Lagrangian surfaces to combinatorial higher laminations

Let us now consider how one may construct a higher lamination on S as defined in Conjecture 14.5 from the data of a Lagrangian cover $\tilde{\Sigma} \subset T^*S$ of S, to which the restriction of the Liouville 1-form is in $H^1(\tilde{\Sigma}, \mathbb{Z})$. Let $b \in \tilde{\Sigma}$ be an arbitrary base-point, and let $z \in S$ be any regular point of the branched covering $\tilde{\Sigma} \longrightarrow S$. Let $z_1, \ldots, z_N$ be the π-preimages of z.

For all $i = 1, \ldots, N$, choose a path γ_i from b to z_i. Then

$$\mu_i = \int_{\gamma_i} \lambda \tag{14.91}$$

depends *a priori* on the choice of γ_i as a real number. However since $\lambda_{|\tilde{\Sigma}} \in H^1(\tilde{\Sigma}, \mathbb{Z})$, if one considers instead μ_i as an angle in $\mathbb{R}/\mathbb{Z} \simeq S^1$ after composing by the projection $\mathbb{R} \longrightarrow S^1$, it is independent of the choice of γ_i. This is schematically depicted in Figure 14.20.

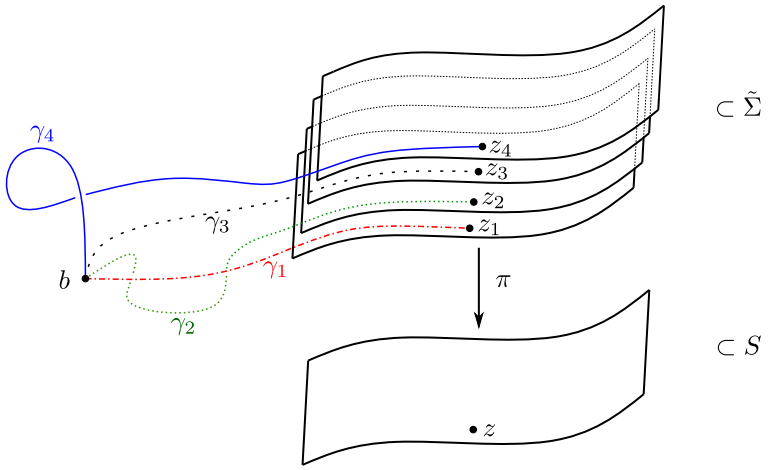


Figure 14.20: The construction of a higher lamination from an integral Lagrangian covering of S.

Hence this construction yields N maps $\mu_i : S \backslash \{\text{branching locus}\} \longrightarrow S^1$, which can be considered as N global sections of a topologically trivial S^1-bundle P over S, as depicted in Figure 14.21. This perspective connects with the wiring diagrams introduced in Section 14.4. The difference between SL_N and PGL_N appears here as in Section 14.4: in the first case one has

$$\sum_{i=1}^{N} \mu_i = 0 \in S^1 \ , \tag{14.92}$$

whereas there is no restriction in the second.

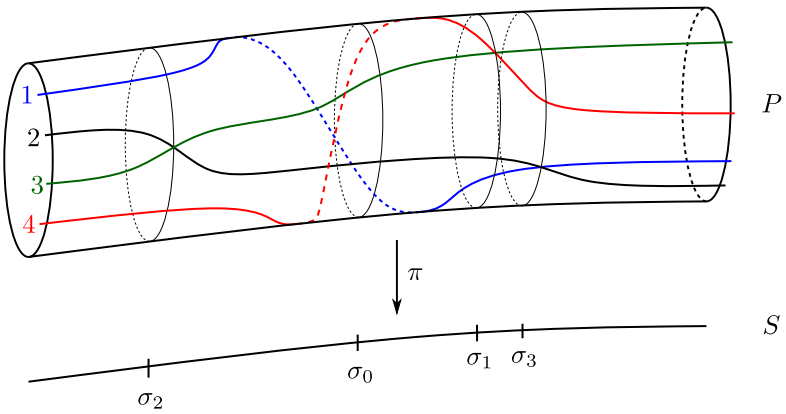


Figure 14.21: The circle bundle over $\mathbb{S}^1$ together with its sections.

Definition 14.18. *Let $i, j \in [\![1, N]\!]$. A regular point $z \in S$ belongs to an (ij)-branch of the higher lamination corresponding to $(\tilde{\Sigma}, \lambda_{|\tilde{\Sigma}})$ if $\mu_i(z) = \mu_j(z)$.*

This is of course very reminiscent of spectral networks, and the union of all branches constructed in this way is a higher lamination as defined in Conjecture 14.5. Just as spectral networks, one can expect that near a branch point of S, a higher lamination consists of three edges with the same label emanating from it.

The notion of (ij) branch for the higher lamination only makes sense when labels can be assigned to the sheets of the covering Σ globally, for example after the choice of branch cuts on S. However, one may choose a global section σ of the S^1-bundle so that the edges of the laminations can always be assigned simple reflections $\sigma_1, \ldots, \sigma_{N-1}$ as in [FM16a] as Section 14.4. This is for example shown in Figure 14.21.

* * * * * * *

We have presented a conjectural definition of higher laminations on a ciliated surface S in Conjecture 14.5, which generalizes Thurston's integral laminations to higher-Teichmüller spaces. The definition is based on the interpretation of integral $\mathcal{A}$-laminations as generators of the ring of regular functions on the corresponding $\mathcal{X}$-cluster variety. The structure of that ring, investigated in Section 14.1, involves Littlewood–Richardson coefficients which can be computed via the hives introduced in Section 14.2. Moreover, the Satake correspondence presented in Section 14.3 points towards the introduction of spherical Hecke algebras. In general, Hecke algebras are deformations of the group ring of Coxeter groups, which is interpreted as evidence for a correspondence between higher laminations and ramified covers of S, in the spirit of spectral networks reviewed in Section 14.5. These ramified covers should presumably be the Lagrangian surfaces $\tilde{\Sigma}$ in T^*S such that the restriction of the Liouville 1-form is in $H^1(\tilde{\Sigma}, \mathbb{Z})$. From such a ramified cover over S and an integral 1-form on it, one can construct a higher lamination on S.

In Figure 14.22, the plain arrows represent steps of the construction which are rather well understood, whereas the dashed arrow corresponds to a correspondence which still has to be worked out. Moreover, an important question which remains to be addressed is the description of the equivalence relations which must appear in the space of Lagrangian covers, as well as in the space of higher laminations as defined in Conjecture 14.5. Further studies of the aforementioned objects should hopefully guide us towards such an understanding.

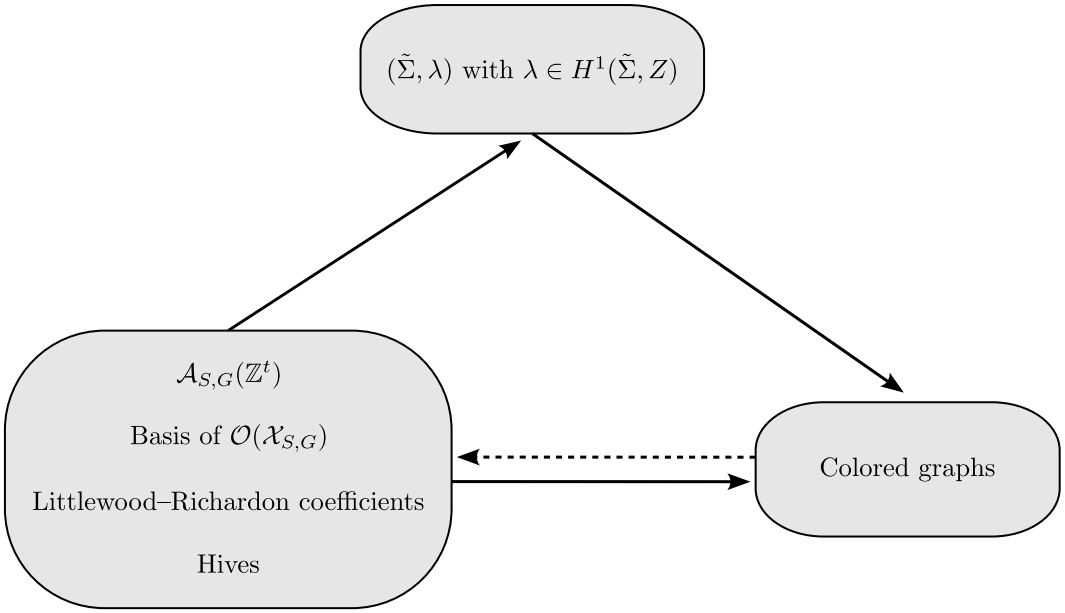


Figure 14.22: The general structure of our approach towards higher laminations.

Chapter 15

Topological Quantum Field Theories from Hecke algebras

Higher laminations as defined in the previous chapter are associated with spherical Hecke algebras. More precisely, given a triangulation Γ of a surface S, an integral G-higher lamination on S should correspond to the assignment of elements of the basis of the Hecke algebra to each edge of Γ, while what happens inside each face of Γ should relate to computations in the spherical Hecke algebra. The goal of this chapter is to study a kind of toy-model of higher laminations, in which the spherical Hecke algebra is replaced by the Hecke algebra corresponding to a finite Coxeter system. These toy-models are simpler than the higher laminations of the previous chapter because the latter type of Hecke algebra is finite dimensional, contrarily to spherical Hecke algebras; however this comes at the cost of loosing commutativity.

This framework interestingly leads to the construction of two-dimensional non-commutative topological quantum field theories (TQFTs) associated to Hecke algebras of finite Coxeter systems. These TQFTs associate an invariant to each ciliated surface, which is a Laurent polynomial for punctured surfaces. There is a graphical way to compute the invariant using minimal colored graphs, in the spirit of the higher laminations defined in Conjecture 14.5. We give explicit formulas in terms of the Schur elements of the Hecke algebra and prove positivity properties for the invariants when the Coxeter group is of classical type, or one of the exceptional types H_3, E_6 and E_7. The TQFTs are of open-closed nature, in that the boundary of ciliated surfaces consist of a disjoint union of circles and segments. The invariant that they associate to a ciliated surface depends on the labels on the boundary segments in a non-commutative way, which is a specificity of this construction. This chapter is a copy of the article [FTT21].

15.1 Introduction

Iwahori–Hecke algebras (referred to as *Hecke algebras* in the sequel) are remarkable associative non-commutative deformations of Coxeter groups depending on a parameter q. We use these Hecke algebras to construct topological invariants of ciliated surfaces. This construction behaves nicely under gluing along the boundary of the surfaces and hence defines a topological quantum field theory (TQFT).

The origin of this construction comes from the study of character varieties of surface groups. The attempt is to generalize Thurston's laminations which can be identified with a basis of the function space of the $\mathrm{SL}_2(\mathbb{C})$-character variety. It turns out that one is led to consider affine Hecke algebras. The question arose what happens if one uses a finite Hecke algebra. This led to the TQFT and the graphical calculus for Hecke algebras presented in this paper.

Two-dimensional TQFTs can be classified by algebraic objects. Closed TQFTs are in bijection with Frobenius algbras, and open-closed TQFTs with so-called Cardy–Frobenius algebras [LP08, AN07]. The Hecke algebra is a particular example of a Cardy–Frobenius algebra. We see the main contribution of our paper in the natural way the TQFT arises, in the diagrammatic calculus and in the positivity properties of the constructed TQFT.

Let G be a finite-dimensional simple Lie group over $\mathbb{F}_q$, where $q = p^\alpha$ is the power of a prime number. Let H be a Cartan subgroup, B a corresponding Borel subgroup and $W \simeq \mathrm{Norm}(H)/H$ the Weyl group. As review in Section 14.3, the Hecke algebra $\mathcal{H}_G^q$ is the algebra of functions on G invariant under left

and right shift by B, or equivalently the algebra of functions of the double coset $B\backslash G/B$, with product given by the convolution divided by the order of B. Since as a set the Weyl group W is in bijection with $B\backslash G/B$, the Hecke algebra $\mathcal{H}_G^q$ is a deformation of the group algebra $\mathbb{C}[W]$.

The structure constants in the corresponding Hecke algebra $\mathcal{H}_G^q$ are related to the counting of $\mathbb{F}_q$-points of flag varieties associated to the corresponding algebraic group (see e.g. [Cur88]). In a judiciously chosen basis the (modified) structure constants of $\mathcal{H}_G^q$ are Laurent polynomial in q invariant under cyclic permutations of their indices. Subsequently, one can associate such a structure constant to an oriented topological triangle with oriented sides labeled with the corresponding basis elements of $\mathcal{H}_G^q$, without the need to fix an initial edge. This construction extends naturally to the Hecke algebras corresponding to finite Coxeter systems (W, S) which are not Weyl groups.

Now one can consider gluing two triangles along one of their edges if its orientation and labels coincide. Summing over all possible labels for the edge along which the gluing is performed, that is over all basis vectors of the Hecke algebra, yields a Laurent polynomial that only depends on the labels on the exterior edges. Repeating this procedure one obtains a way to associate a Laurent polynomial $P \in \mathbb{Z}[v^{\pm 1}]$ to any triangulated ciliated surface Σ^1 whose boundary is labeled by elements of the Weyl group W:

$$P_{\Sigma,W}(v) = \sum \prod_f c_f(v) \ , \tag{15.1}$$

where the sum runs over all labelings of the inner edges of the triangulation by elements of W, the product runs over all faces f, and $c_f(v) \in \mathbb{Z}[v^{\pm 1}]$ denotes the structure constant associated to a triangle f (see Section 15.3 for the precise definition).

The definition of the polynomial P as a state sum uses a triangulation of the surface. Associativity in the Hecke algebra gives (see Theorem 15.18):

Theorem 15.1. *The polynomial invariant is independent of the triangulation. Hence it is a topological invariant of the ciliated surface.*

In Section 15.4, we introduce a diagrammatic way to compute the product in the Hecke algebra when W is a Weyl group. From this we get a graphical calculus of our invariants using graphs with edges labeled by simple reflections in W. These graphs are still called higher laminations, in a slight abuse of notation. They emerge from the interpretation of the structure constants in the Hecke algebra in terms of configurations of triples of flags. Figure 15.1 gives two examples of graphs for $W = \mathfrak{S}_3$.

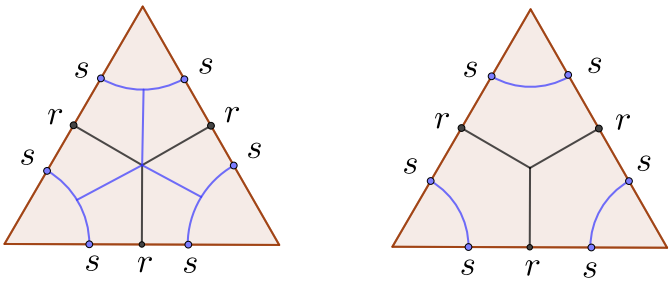


Figure 15.1: Examples of graphs for $W = \mathfrak{S}_3$

Counting higher laminations with some weight gives a diagrammatic way to compute the polynomial invariant (see Theorem 15.41):

Theorem 15.2. *The polynomial invariant for a ciliated surface Σ and a Weyl group W is given by*

$$P_{\Sigma,W}(Q) = \sum_{\Gamma} Q^{\mathrm{ram}(\Gamma)}$$

where the sum runs over all higher laminations Γ of type W and $Q^{\mathrm{ram}(\Gamma)}$ is the weight associated to Γ.

As a consequence we get a symmetry in the invariant that for closed surfaces (see Proposition 15.43):

[1] In this chapter we denote ciliated surfaces using the letter Σ instead of S in order to avoid confusion with the set S of simple reflections of the Coxeter system (W, S).

Corollary 15.3. *For a closed surface, the invariant is a polynomial in q, where $Q = q^{1/2} - q^{-1/2}$, invariant under the transformation $q \mapsto q^{-1}$.*

In Section 15.5, we extend our construction to ciliated surfaces whose boundary is labeled by elements of the Hecke algebra. Our main result is then the following (see Theorem 15.46):

Theorem 15.4. *Our construction satisfies the axioms of a 2-dimensional[2] TQFT as defined in [Ati88].*

Hence we construct a family of TQFTs, one for each finite Coxeter system. The interesting feature of these TQFTs is that they are non-commutative. This ultimately comes from the fact that the cobordisms we consider are ciliated surfaces, whose boundary is a disjoint union of segments joining adjacent cilia.

Using the gluing property of a TQFT, we can decompose ciliated surfaces into elementary parts (marked tori and marked cylinders). The center of the Hecke algebra plays a prominent role. Since the structure of the latter is well understood we can derive an explicit expression for the invariant in terms of the Schur elements of the Hecke algebra (see Theorem 15.58):

Theorem 15.5. *For a ciliated surface Σ of genus g with k marked points and n boundary components equipped with labels $(h_1, ..., h_n) \in \mathcal{H}^n$, the polynomial invariant is:*

$$P_{\Sigma,W}(q) = \sum_{\lambda} (\dim V_\lambda)^k s_\lambda(q)^{2g-2+k+n} \chi_\lambda(h_1) \cdots \chi_\lambda(h_n) \tag{15.2}$$

where the sum is taken over all irreducible representations V_λ of the Hecke algebra and where s_λ are the corresponding Schur elements.

From these explicit expressions and a thorough analysis of the Schur elements, carried out in Section 15.8, we can derive positivity properties of the invariant (Theorem 15.60 and Corollary 15.63):

Theorem 15.6. *The polynomial invariant has positive coefficients for all Coxeter groups of classical type and in the exceptional types H_3, E_6 and E_7. For all other types, the polynomial can have negative coefficients.*

In type A and for boundary labels with positive coefficients in the Kazhdan–Lusztig basis, the polynomial invariant has positive coefficients.

Notations. We write *Hecke algebra* for an Iwahori–Hecke algebra for a finite Coxeter system. When we speak about graphs, we always mean fat graphs, i.e. graphs embedded in some surface (planar graphs for example). We say that a polynomial is *positive* if all its coefficients are positive. Furthermore, we use the following notations:

$\Sigma_{g,k}$	ciliated surface (see Section 15.2.3)
(W, S)	Coxeter system with simple reflections S
$\mathcal{H}_{(W,S)}, \mathcal{H}$	Iwahori–Hecke algebra of (W, S)
v, q, Q	formal parameters in the Hecke algebra linked by $q = v^{-2}$ and $Q = q^{1/2} - q^{-1/2}$

15.2 Preliminaries

We recall the definition of the Hecke algebra of a Coxeter system and introduce its standard and Kazhdan–Lusztig bases following [EW16, Lib19], and then set our conventions and notations for ciliated surfaces as in section 2 of [FG07].

15.2.1 Hecke algebras and their standard basis

Definition 15.7. *Let S be a finite set, and for all $s, t \in S$ let $m_{st} \in \mathbb{N} \cup \{\infty\}$ with $m_{ss} = 1$. The associated Coxeter group is defined as*

$$W = \langle s \in S \mid (st)^{m_{st}} = \mathrm{id} \ \ \forall s, t \in S \rangle \ . \tag{15.3}$$

[2] In the notation of Atiyah, we construct a one-dimensional TQFT. In modern language we speak about a d-dimensional TQFT for d being the dimension of the bordisms, so 2 dimensions in our case.

Let $l : W \to \mathbb{N}_{\geq 0}$ be the length function which to each element of W associates the minimal length of its representatives, and let $\geq$ be the Bruhat order on W, i.e. $w \geq w'$ if w' has a representative of minimal length which is a subword of a representative of minimal length of w.

Let v be an indeterminate. The *Hecke algebra* $\mathcal{H}_{(W,S)}$ corresponding to (W, S) as defined in [Iwa64] is the associative $\mathbb{Z}[v^{\pm 1}]$-algebra with generators $\{h_s\}_{s\in S}$ and relations of two types, the quadratic ones:

$$h_s^2 = (v^{-1} - v)h_s + 1 \tag{15.4}$$

for all $s \in S$, and the braid ones:

$$h_s h_r \cdots = h_r h_s \ldots \tag{15.5}$$

for all $s, t \in S$ such that $m_{st} < \infty$, in which case there are m_{st} terms on each side.

Except when the explicit subscript (W, S) is needed for clarity or in explicit examples, we will drop it in what follows and simply refer to the Hecke algebra corresponding to (W, S) as $\mathcal{H}$.

Remark 15.8. *An equivalent, but different, way to describe the Hecke algebra is to use generators $(T_s)_{s\in S}$ which satisfy the same braid relations and where the quadratic relation is*

$$T_s^2 = (q - 1)T_s + q \ . \tag{15.6}$$

for another indeterminate q. Apart from Section 15.4.1, we always use the the "normalized" version of the Hecke algebra where the quadratic relations take the form of Equation (15.4) with $v = q^{-1/2}$.

Let $\bar{w} = s_1 \ldots s_k$ be a reduced expression for some $w \in W$ in terms of $s_1, \ldots, s_k \in S$, and set:

$$h_{\bar{w}} = h_{s_1} \ldots h_{s_k} \ . \tag{15.7}$$

By the famous result of Matsumoto [Mat64] that every reduced expression for w can be obtained from $\bar{w}$ using braid relations only, the element $h_{\bar{w}}$ does not depend on the choice of a reduced expression for w and one can define $h_w := h_{\bar{w}}$. Let also $h_e := 1$.

Lemma 15.9. *The set $\{h_w\}_{w\in W}$ is a basis of $\mathcal{H}$ as a free $\mathbb{Z}[v^{\pm 1}]$-module, called the standard basis of $\mathcal{H}$.*

Let $s \in S$ and $w \in W$. The multiplication in $\mathcal{H}$ can be rewritten as:

$$h_s h_w = \begin{cases} h_{sw} & \text{if} \quad w \leq sw \\ (v^{-1} - v)h_w + h_{sw} & \text{if} \quad sw \leq w \end{cases} \ . \tag{15.8}$$

Example 15.10. *The Coxeter system of type A_1 is $(\mathfrak{S}_2, \{s\})$, where $\mathfrak{S}_2$ is the group of permutations of a set of two elements, and s is its generating involution. The corresponding Hecke algebra $\mathcal{H}_{(\mathfrak{S}_2,\{s\})}$ has basis (h_e, h_s) as $\mathbb{Z}[v^{\pm 1}]$-module, and the relations $h_s h_e = h_e h_s$ and $h_s^2 = (v^{-1} - v)h_s + 1$ hold.*

Example 15.11. *The Coxeter system of type A_2 is $(\mathfrak{S}_3, \{s, t\})$, where s and t are two transpositions generating $\mathfrak{S}_3$. The standard basis of the Hecke algebra $\mathcal{H}_{(\mathfrak{S}_3,\{s,t\})}$ as a $\mathbb{Z}[v^{\pm 1}]$-module is $(h_e, h_s, h_t, h_{st}, h_{ts}, h_{sts})$. However as a $\mathbb{Z}[v^{\pm 1}]$-algebra, $\mathcal{H}_{(\mathfrak{S}_3,\{s,t\})}$ is generated by h_e, h_s and h_t only, where h_e is in the center and with the quadratic relations of Equation (15.4) for h_s and h_t, and the braid relation:*

$$h_s h_t h_s = h_t h_s h_t \ . \tag{15.9}$$

Example 15.12. *The Coxeter system of type B_2 is $(\mathbb{D}_4, \{s, t\})$, where s and t are two transpositions which generate the dihedral group $\mathbb{D}_4$. The standard basis of the corresponding Hecke algebra $\mathcal{H}_{(\mathbb{D}_4,\{s,t\})}$ is $(h_e, h_s, h_t, h_{st}, h_{ts}, h_{sts}, h_{tst}, h_{stst})$ as a $\mathbb{Z}[v^{\pm 1}]$-module. The multiplication is such that h_e commutes with h_s and h_t, both h_s and h_t satisfy the quadratic relation of Equation (15.4), and there is the following braid relation:*

$$h_s h_t h_s h_t = h_t h_s h_t h_s \ . \tag{15.10}$$

15.2.2 The Kazhdan–Lusztig basis

Let $s \in S$. One can check easily that the inverse of h_s is $h_s + v - v^{-1} \in \mathcal{H}$. Since the set of all h_s for $s \in S$ generates $\mathcal{H}$ as a $\mathbb{Z}[v^{\pm 1}]$-algebra, it follows that for every $w \in W$, the corresponding h_w also admits an inverse.

The morphism of $\mathbb{Z}$-modules defined by

$$\iota : \left\{ \begin{array}{ccc} \mathcal{H} & \to & \mathcal{H} \\ v & \mapsto & v^{-1} \\ h_w & \mapsto & (h_{w^{-1}})^{-1} \end{array} \right. \tag{15.11}$$

is a ring automorphism of $\mathcal{H}$. A cornerstone of Kazhdan–Lusztig theory is the following theorem (see [KL79, Theorem 1.1]):

Theorem 15.13 (Kazhdan–Lusztig)**.** *For all $w \in W$ there exists a unique ι-self-dual element of $\mathcal{H}$ of the form*

$$b_w = h_w + \sum_{z \leq w} h_{z,w} h_z \tag{15.12}$$

where $h_{z,w} \in v\mathbb{Z}[v]$. Moreover the set $\{b_w\}_{w \in W}$ is a basis of $\mathcal{H}$ as $\mathbb{Z}[v^{\pm 1}]$-module.

The *Kazhdan–Lusztig polynomial* $p_{z,w}$ is defined as:

$$p_{z,w} = v^{l(w)-l(z)} h_{z,w} \ . \tag{15.13}$$

Example 15.14. *For $s \in S$, the element $b_s = h_s + v \in \mathcal{H}$ is always self-dual under ι and moreover it is of the form of Equation* (15.12)*. It is clear that $b_e = h_e = 1$. Hence the Kazhdan–Lusztig basis of $\mathcal{H}_{(\mathfrak{S}_2, \{s\})}$ is the pair $(b_e = 1, b_s)$.*

Positivity. The Kazhdan–Lusztig basis enjoys numerous positivity properties which have been shown to be a consequence of a combinatorial Hodge theory in the category of Soergel bimodules [Wil16]. For example the Kazhdan–Lusztig polynomials $p_{z,w}$ of Equation (15.13) are positive (meaning that $p_{z,w} \in \mathbb{Z}_{\geq 0}[v]$), and the structure constants of the Hecke algebra are positive when expressed in the Kazhdan–Lusztig basis: if one sets

$$b_x b_y = \sum {\mu_{xy}}^z b_z \ , \tag{15.14}$$

then ${\mu_{xy}}^z \in \mathbb{Z}_{\geq 0}[v^{\pm 1}]$.

15.2.3 Ciliated surfaces

A **ciliated surface** is an oriented topological surface obtained as follows: remove n disjoint open disks labeled $1, \ldots, n$ out of the oriented surface of genus g with k punctures, for $g, k, n \geq 0$. On the i-th boundary circle add $p_i \geq 1$ marked points called cilia, for $i = 1, \ldots, n$.

Topologically, a ciliated surface is determined by its genus, the integer k and the set $\{p_1, ..., p_n\}$. We will denote such a surface $\Sigma_{g,k,\{p_1,\ldots,p_n\}}$ in general, and simplify the writing to $\Sigma_{g,k}$ when it has no cilium, and to $\Sigma_{g,\{p_1,\ldots,p_n\}}$ when it has no puncture. Let $c = \sum p_i$ be the total number of cilia.

Figure 15.2: The ciliated surface $\Sigma_{2,3,\{2,2,1,1,3\}}$.

In the sequel we will only consider ciliated surfaces such that $k+n \geq 1$, and such that when $g=0$ either $k+n \geq 3$ or $k+n=2$ and $c \geq 1$ or $k+n=1$ and $c \geq 3$, for such ciliated surfaces can be triangulated. A *triangulation* T of Σ is a decomposition of Σ into triangles such that every vertex of a triangle is either a cilium or a puncture. Edges of T belonging to the boundary of Σ are said to be external and the others, internal. Let $\#F(\mathsf{T})$, $\#E(\mathsf{T})$, $\#E_0(\mathsf{T})=c$ and $\#V(\mathsf{T})=k+c$ be respectively the number of faces, edges (external and internal), internal edges and vertices of T. The Euler characteristic of the closure $\overline{\Sigma}$ is

$$\#F(\mathsf{T}) - \#E(\mathsf{T}) + \#V(\mathsf{T}) = 2-2g-n \ , \tag{15.15}$$

and since T is a triangulation:

$$3\#F(\mathsf{T}) = 2\#E(\mathsf{T}) - \#E_0(\mathsf{T}) \ . \tag{15.16}$$

From Equation (15.15) and Equation (15.16) one deduces that:

$$\begin{array}{l} \#E(\mathsf{T}) = 6g-6+2c+3(k+n) \\ \#F(\mathsf{T}) = 4g-4+c+2(k+n) \end{array} \ . \tag{15.17}$$

Except when $g=0$ and $(k,n)=(0,n)$ or $(k,n)=(1,n)$ the number of triangulations of a ciliated surface is infinite. However one can always reach any triangulation from a reference one in a finite number of *flips* which consist of replacing the diagonal of a quadrilateral formed by two adjacent triangles with the other diagonal.

In what follows we will speak of the *boundary of a ciliated surface* to refer to the disjoint union of the boundary segments connecting two adjacent cilia.

15.3 Definition of the polynomial and first properties

We present our construction which associates Laurent polynomials to ciliated surfaces in a pedestrian way. A more abstract viewpoint comes in Section 15.5.

15.3.1 On the standard structure constants in Hecke algebras

Let (W,S) be a Coxeter system and $\mathcal{H}$ the corresponding Hecke algebra. In Section 15.2.1 we introduced the standard basis $\{h_w\}_{w\in W}$ of $\mathcal{H}$ as a free $\mathbb{Z}[v^{\pm 1}]$-module. Let $\mathcal{H}^*$ be the free $\mathbb{Z}[v^{\pm 1}]$-module dual to $\mathcal{H}$ with standard dual basis $\{h^w\}_{w\in W}$.

By definition, the structure constants ${c_{xy}}^z$ are given by

$${c_{xy}}^z = h^z(h_x \cdot h_y) \ \in \mathbb{Z}[v^{\pm 1}] \tag{15.18}$$

for $x,y,z \in W$. Let us set:

$$c_{xyz} := {c_{xy}}^{z^{-1}} \ . \tag{15.19}$$

The notation of Equation (15.19) can be understood through the standard trace in the Hecke algebra. A **trace** on $\mathcal{H}$ is a $\mathbb{Z}[v^{\pm 1}]$-linear map $\mathrm{tr} : \mathcal{H} \to \mathbb{Z}[v^{\pm 1}]$ such that $\mathrm{tr}\Big(hh'\Big) = \mathrm{tr}\Big(h'h\Big)$ for all $h,h' \in \mathcal{H}$. A trace is said to be *symmetrizing* if the map $h' \mapsto \mathrm{tr}(hh')$ is non-degenerate for all $h \neq 0$. The map

$$\mathrm{tr}\left(\sum_{w\in W} c_w h_w\right) = h^e\left(\sum_{w\in W} c_w h_w\right) = c_e \tag{15.20}$$

is a symmetrizing trace on $\mathcal{H}$ called the *standard trace* [GP00, Proposition 8.1.1]. It is easy to see that the standard trace on $\mathcal{H}$ satisfies:

$$\mathrm{tr}(h_x h_y) = \delta_{x,y^{-1}} \ . \tag{15.21}$$

Proposition 15.15. *For all $x,y,z \in W$, one has*

$$c_{xyz} = \mathrm{tr}(h_x h_y h_z) \ . \tag{15.22}$$

In particular, this implies that c_{xyz} is cyclically symmetric: $c_{xyz} = c_{yzx} = c_{zxy}$.

Proof. By definition of the structure constants one has:

$$h_x h_y = \sum_{z'} c_{xyz'} h_{z'^{-1}} \ . \tag{15.23}$$

Multiplying by h_z, taking the trace and using Equation (15.21) yields:

$$\mathrm{tr}(h_x h_y h_z) = \sum_{z'} c_{xyz'} \delta_{z,z'} = c_{xyz} \ , \tag{15.24}$$

which concludes the proof. □

15.3.2 An invariant for surfaces with punctures

Triangles and structure constants. Consider a triangle with oriented edges, labeled with elements $x, y, z \in W$. Fix also an orientation of the triangle. We associate to the triangle the structure constant $c_{x^a y^b z^c} \in \mathbb{Z}[v^{\pm 1}]$ where a is 1 (respectively, -1) if the orientation of the edge labeled by x is induced by the triangles orientation (respectively, if not), and *mutatis mutandis* for the two other edges. Figure 15.3 gives two examples.

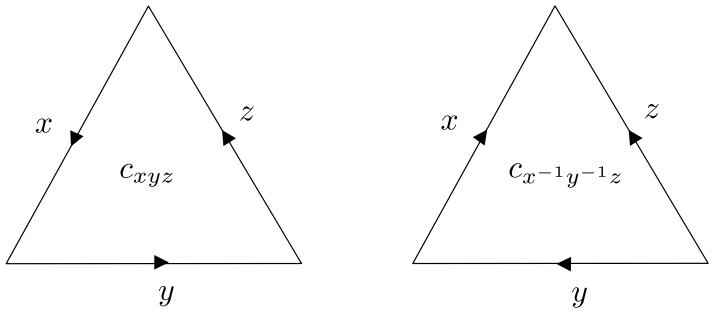


Figure 15.3: We associate c_{xyz} to the left-most triangle and $c_{x^{-1}y^{-1}z}$ to the right-most one.

Note that by Proposition 15.15 the quantity associated to the triangle is well-defined. In addition, it does not depend on the chosen orientation of the triangle:

Proposition 15.16. *The structure constants satisfy $c_{xyz} = c_{z^{-1}y^{-1}x^{-1}}$. Thus, the associated quantity to a triangle is independent of its orientation.*

Proof. Consider the following map σ on $\mathcal{H}$, defined on the standard basis by

$$\sigma(h_w) = h_{w^{-1}} \tag{15.25}$$

and extended by linearity. We claim that σ is an anti-involution of $\mathcal{H}$, i.e. that $\sigma(h_a h_b) = \sigma(h_b)\sigma(h_a)\ \forall a, b \in W$. Indeed it is enough to check this on the basic relations: the quadratic relation is invariant under σ, and for $w = sw'$ where $s \in S$ and w' is a reduced word, we have

$$\sigma(h_s h_{w'}) = \sigma(h_w) = h_{w^{-1}} = h_{w'^{-1}} h_s = \sigma(h_{w'})\sigma(h_s) \ . \tag{15.26}$$

By definition, we have

$$h_x h_y = \sum_{z \in W} c_{xyz} h_{z^{-1}} \ . \tag{15.27}$$

Applying σ gives

$$h_{y^{-1}} h_{x^{-1}} = \sum_{z \in W} c_{xyz} h_z \ . \tag{15.28}$$

Hence $c_{xyz} = c_{y^{-1}x^{-1}z^{-1}}$. We conclude by cyclicity of c_{xyz}. □

Gluing triangles into ciliated surfaces. Let T be a triangulation of the oriented topological surface $\Sigma_{g,k}$ of genus g and with $k \geq 1$ punctures.

Definition 15.17. *Let $P_{g,k,W,\mathsf{T}}$ be the Laurent polynomial:*

$$P_{g,k,W,\mathsf{T}}(v) = \sum \prod_{f \in \mathsf{T}} c_f(v) \in \mathbb{Z}[v^{\pm 1}] \ , \tag{15.29}$$

where the sum runs over all possible labelings of the edges of T by elements of W and the product over all faces f of T, and where $c_f(v)$ is the structure constant associated to the face f of T as before.

More generally, we can define such a polynomial for any ciliated surface with triangulation T, as soon as each boundary component is labeled with an element of W. The sum in Equation (15.29) *runs in this case over all possible labelings of internal edges of T.*

In order to compute $P_{g,k,W,\mathsf{T}}$ one has to choose an orientation for the internal edges of T so that the $c_f(v)$ are well defined. Since we are summing over all possible labels of the edges, the polynomial $P_{g,k,W,\mathsf{T}}$ does not depend on these choices since changing the orientation amounts to replace $w \in W$ by its inverse.

The main point of our construction is the following:

Theorem 15.18. *The polynomial invariant does not dependent on the triangulation T, hence it is a topological invariant of the ciliated surface. Further, it is preserved under reversing the orientation of the surface (and thus inverting all the boundary data).*

We denote the invariant by $P_{\Sigma,W}$ for a ciliated surface Σ or by $P_{g,k,W}$ in case of a punctured surface $\Sigma_{g,k}$. The theorem follows from the *associativity* of the product in the Hecke algebra.

Proof. The second part is a direct consequence of Proposition 15.16 and the definition of $P_{g,k,W}$.

For the first part, let $x, y, z \in W$. Then by definition:

$$h_x h_y = \sum_{w \in W} c_{xyw} h_{w^{-1}} \ , \tag{15.30}$$

hence

$$\sum_{w,v \in W} c_{xyw} c_{w^{-1}zv} h_{v^{-1}} = (h_x h_y) h_z = h_x (h_y h_z) = \sum_{t,v \in W} c_{yzt} c_{xt^{-1}v} h_{v^{-1}} \ , \tag{15.31}$$

which implies

$$\sum_{w \in W} c_{xyw} c_{w^{-1}zv} = \sum_{t \in W} c_{yzt} c_{xt^{-1}v} \tag{15.32}$$

for all $x, y, z, v \in W$. Equation (15.32) can be described graphically as in Figure 15.4.

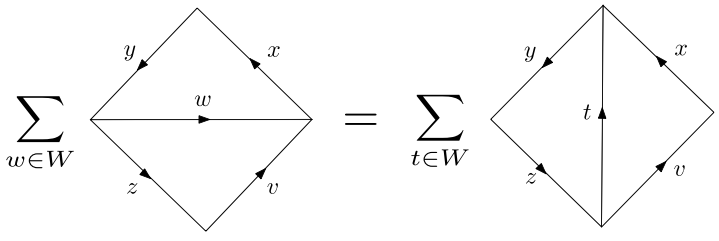


Figure 15.4: A consequence of associativity in Hecke algebras.

Now since any two triangulations of $\Sigma_{g,k}$ can be related via a sequence of flips, Equation (15.32) implies the proposition. □

Example 15.19. *Let us give some examples of $P_{g,k,W}$. Later on we will come back to them and see how to compute them.*

- $P_{0,3,\mathfrak{S}_2}(v) = P_{1,1,\mathfrak{S}_2}(v) = v^2 + 2 + v^{-2}$.
- $P_{0,4,\mathfrak{S}_2}(v) = v^4 + 2v^2 + 2 + 2v^{-2} + v^{-4}$.

- $P_{0,3,\mathfrak{S}_3}(v) = v^6 + 2v^4 + 10v^2 + 10 + 10v^{-2} + 2v^{-4} + v^{-6}$.
- $P_{1,1,\mathfrak{S}_3}(v) = v^6 + 2v^4 + 4v^2 + 4 + 4v^{-2} + 2v^{-4} + v^{-6}$.
- $P_{\triangle_{x,y,z},W}(v) = c_{xyz}(v)$ *where* $\triangle_{x,y,z}$ *stands for the triangle seen as the ciliated surface* $\Sigma_{0,0,\{3\}}$ *with labels* $x, y, z \in W$ *on the exterior edges assigned counterclockwisely.*

We give additional examples in Section 15.7 where we explain how to use Sage and the package CHEVIE of Gap3 to compute these polynomials.

There are three remarkable observations to be done for punctured surfaces from these examples: the polynomials are functions of $v^{-2} = q$, they are invariant under $q \mapsto q^{-1}$ and have positive coefficients. The first two observations are actually properties that we will prove in Proposition 15.43, while we will analyze the positivity in Section 15.6.4.

Remark 15.20. *Jumping ahead a bit, since for punctured surfaces the invariant polynomial only depends on* $v^{-2} = q$ *we will use the variable* q *in this case instead of* v *to lighten the notation. For surfaces with cilia we keep the use of* v*.*

15.3.3 Gluing surfaces

The polynomial invariant behaves nicely under gluing of two surfaces Σ_1 and Σ_2 with same boundary data D as in Figure 15.5. Schematically:

$$P(\Sigma_1 \cup \Sigma_2) = \sum_D P(\Sigma_1, D) P(\Sigma_2, D) \ , \tag{15.33}$$

where $P(\Sigma, D)$ denotes the polynomial invariant for the surface Σ with boundary data D.

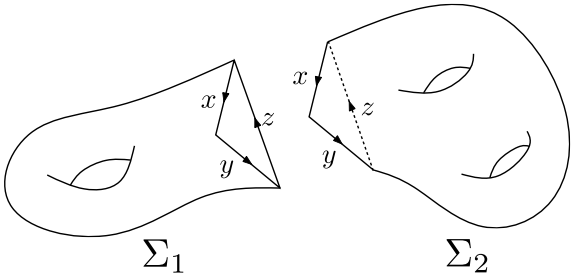


Figure 15.5: Gluing ciliated surfaces.

We can use Equation (15.33) to compute recursively the polynomial. For $\Sigma_1 = \Sigma_{0,0,3}$ a triangle and $\Sigma_2 = \Sigma_{g,k,\{3,\dots\}}$ a surface with a triangle boundary (that is, a circle component of its boundary with three cilia), Equation (15.33) becomes:

$$P\left(\Sigma_{g,k,\{\dots\}}, D\right) = \sum_{x,y,z \in W} c_{xyz} P\left(\Sigma_{g,k,\{3,\dots\}}, D \cup \{x, y, z\}\right) \ , \tag{15.34}$$

where D is the boundary data of the ciliated surface $\Sigma_{g,k,\{\dots\}}$ which may have non-empty boundary, and $D \cup \{x, y, z\}$ is the boundary data of $\Sigma_{g,k,\{3,\dots\}}$. Note that we used Proposition 15.16.

For a ciliated surface with a triangle boundary carrying the labels x, y and z, the gluing can be used to reduce the number of punctures. To do so, we use for Σ_2 a quadrilateral with boundary z, x, x'^{-1}, z'^{-1}. The surface after gluing is depicted in Figure 15.6. The polynomial invariant of this quadrilateral is given by $\sum_w c_{xx'^{-1}w} c_{zw^{-1}z'^{-1}}$. Hence we get from the gluing property:

$$P\left(\Sigma_{g,k+1} \backslash \triangle_{x,y,z}\right) = \sum_{x',z',w \in W} c_{xx'^{-1}w} c_{zw^{-1}z'^{-1}} P\left(\Sigma_{g,k} \backslash \triangle_{x',y,z'}\right) \ , \tag{15.35}$$

with the obvious generalizations to the general cases.

Let us now study our invariants for $q = 1$ since the Hecke algebra specializes to the group algebra $\mathbb{C}[W]$ in that case.

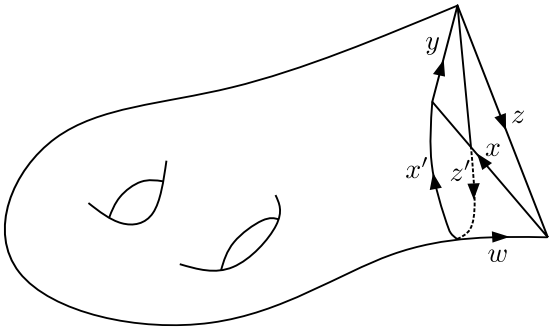


Figure 15.6: Reducing the number of punctures.

15.3.4 Invariants of punctured surfaces at $q = 1$

For $q = 1$ the Hecke algebra specializes to the group algebra $\mathbb{C}[W]$.

Proposition 15.21. *For a punctured surface $\Sigma_{g,k}$ the value at $q = 1$ of the polynomial is:*

$$P_{g,k,W}(1) = (\#W)^{k-1} \times \#\{\textit{solutions in } W \textit{ to } \prod_{i=1}^{g}[a_i, b_i] = 1\} \ . \tag{15.36}$$

In particular $P_{0,k,W}(1) = (\#W)^{k-1}$ which is in accordance with Example 15.19.

Remark 15.22. *The right-most term in the right-hand-side of Equation* (15.36) *can be expressed in terms of the characters of W using Frobenius' formula (see [LZ13, theorem A.1.10 in the Appendix by Don Zagier]). This yields:*

$$P_{g,k,W}(1) = (\#W)^{2g-2+k} \times \sum_{\chi} \frac{1}{\chi(1)^{2g-2}} \ . \tag{15.37}$$

Notice the similarity to Burnside's formula for Hurwitz numbers (see [Gun16, Theorem 1.3] and the original paper [Hur91]).

Our strategy to prove Proposition 15.21 is to develop the surface $\Sigma_{g,k}$ as a $4g$-gon and to count explicitly the contributions.

Proof. We start with the case $k = 1$ and $g > 0$. The surface $\Sigma_{g,1}$ is obtained by gluing the sides of a $4g$-gon as described on the left of Figure 15.7 in the case $g = 2$. We also show a triangulation T. Note that the $4g$ vertices of the polygon correspond to a single point in $\Sigma_{g,1}$ which is the puncture.

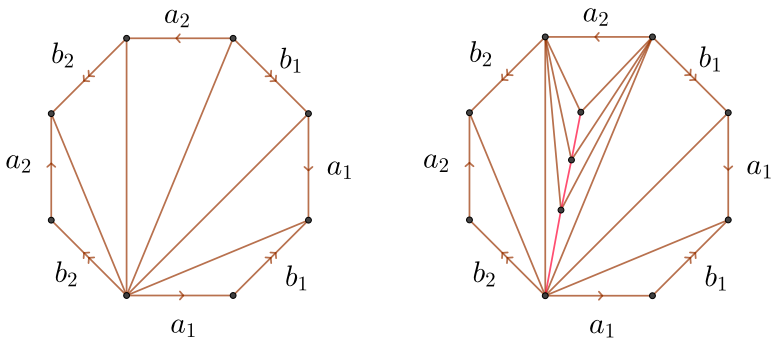


Figure 15.7: Gluing of $4g$-gon

The specializations of the structure constants c_{xyz} of the Hecke algebra at $q = 1$ are equal to 1 if $xyz = 1$ in W and to zero otherwise. In order to get non-zero contributions to the polynomial invariant

specialized at $q = 1$, the label on a side of a face f of T is determined by the labels on the two other sides of f. Let now $a_1, b_1, ..., a_g, b_g \in W$ be the labels assigned to the edges of the $4g$-gon. Let T be a triangulation of the $4g$-gon whose inner edges are all incident to a fixed vertex as on the left of Figure 15.7. The labels on the inner edges of T are completely determined by the a_i and b_i, otherwise the corresponding contribution to the invariant vanishes.

There is the following additional constraint:

$$\prod_{i=1}^{g}[a_i, b_i] = 1 \ , \tag{15.38}$$

which can be understood as the consequence of the fact that the products of labels assigned to each triangle whose boundary is oriented counterclockwise has to be 1 if one considers all the triangles glued together along edges as on the left of Figure 15.7.

Every solution of Equation (15.38) has contribution 1 to the polynomial, and all other choices of a_i and b_i do not contribute. Therefore we get the proposition in the case $k = 1$.

For $k > 1$, we add in our picture $k-1$ marked points in the middle of the polygon. We complete to a triangulation as shown on the right of Figure 15.7. In particular there is a path of length $k-1$ connecting all k punctures shown in red in Figure 15.7. To each edge of this path, we associate a new label $x_i \in W$. One easily checks that these x_i together with the a_j and b_j from above uniquely determines the labels assigned to each edge of the triangulation. The only relation is Equation (15.38) from above. Having fixed the data on the boundary of the $4g$-gon we have $(\#W)^{k-1}$ free choices which contribute by 1 to our polynomial.

Eventually for $g = 0$ we represent $\Sigma_{0,k}$ as the gluing of a $2k-2$-gon with boundary $a_1, a_1^{-1}, a_2, a_2^{-1}, ..., a_{k-1}, a_{k-1}^{-1}$ with a triangulation similar as the ones of Figure 15.7. The labels on the inner edges are determined by the boundary data and the constraint on the boundary is always satisfied. So we get $P_{0,k,W}(1) = (\#W)^{k-1}$. □

15.4 Graphical calculus

In this section, we present a diagrammatic interpretation of our polynomial. In particular we give a graphical way to multiply elements in the Hecke algebra. In this section, we work with a Coxeter system (W, S) associated to a Weyl group.

We start from the observation that the structure constants in the Hecke algebra are linked to configurations of flags. From that, we define graphs with labeled edges which count the contributions to our polynomial.

15.4.1 Structure constants and flag counting

There is a well-known link between the structure constants in the Hecke algebra and triples of flags (see e.g. [Cur88, Proposition 2.2]). We present a short way to get this link.

In Section 15.2.1, we introduced the Hecke algebra using generators and relations. However the original definition by Iwahori [Iwa64] is geometric. Consider a finite field $\mathbb{F}_q$, a simple Lie group G over $\mathbb{F}_q$ and fix a Borel subgroup B of G. The Hecke algebra $\mathcal{H}_G^q$ is the algebra of functions on G invariant under left and right shift by B. Multiplication is given by the convolution divided by the order of the Borel subgroup $\#B$. Note that the double quotient $B\backslash G/B$ is in bijection with W. This is why the Hecke algebra has a basis $(T_w)_{w\in W}$ parameterized by the Weyl group. In this geometric approach the generators $(T_s)_{s\in S}$ satisfy the braid relations and the quadratic relation reads:

$$T_s^2 = (q-1)T_s + q \ . \tag{15.39}$$

The structure constants ${C_{xy}}^z$ are defined by

$$T_x T_y = \sum_{z\in W} {C_{xy}}^z(q) T_z \ . \tag{15.40}$$

From the definition of the convolution product, it follows that

$${C_{xy}}^z(q) = \#\{h \in x \mid h^{-1}g \in y\}/\#B \ , \tag{15.41}$$

where we interpret x, y and z as elements of $W \cong B\backslash G/B$ and where g is any representative of $z \in B\backslash G/B$.

The quotient G/B is called the *flag variety* of G. For $G = \mathrm{GL}_n(\mathbb{C})$ this is the space of complete flags in $\mathbb{C}^n$. Recall the isomorphism $G\backslash(G/B)^2 \cong B\backslash G/B$ given by $h \mapsto (F_0, hF_0)$ where $F_0 \in G/B$ is the class of the unit element in G. Combining this isomorphism with $B\backslash G/B \cong W$, we see that the relative positions of two flags is described by the Weyl group. Using this isomorphism we can rewrite the formula for the structure constants as

$$\begin{aligned} {C_{xy}}^z(q) &= \#\{h \in G \mid (F_0, hF_0) = x, (F_0, h^{-1}gF_0) = y\}/\#B \\ &= \#\{h \in G \mid (F_0, hF_0) = x, (hF_0, gF_0) = y\}/\#B \ . \end{aligned} \tag{15.42}$$

Taking into account that B acts freely on the right on the set $\{h \in x \mid h^{-1}g \in y\}$ and denoting gF_0 by F_2 and hF_0 by F_1, we finally get:

Proposition 15.23. *The structure constants ${C_{xy}}^z(q)$ count the configuration of three flags with prescribed relative positions:*

$${C_{xy}}^z(q) = \#\{F_1 \in G/B \mid (F_0, F_1) = x, (F_1, F_2) = y\} \ , \tag{15.43}$$

where $(F_0, F_2) = z$.

Example 15.24. *Let $G = \mathrm{SL}(2)$. Then $G/B = \mathbb{P}^1, W = \{1, s \mid s^2 = 1\}$, ${C_{11}}^1 = 1, {C_{11}}^s = {C_{s1}}^1 = {C_{1s}}^1 = 0$ and ${C_{1s}}^s = {C_{s1}}^s = 1$ obviously. To compute ${C_{ss}}^1(q)$, choose $F_0 = F_2 = \infty$. Then ${C_{ss}}^1(q) = \#\{p \in \mathbb{P}^1(\mathbb{F}_q) \mid p \neq \infty\} = q$. To compute ${C_{ss}}^s(q)$, choose $F_0 = 0$ and $F_2 = \infty$. Then ${C_{ss}}^s(q) = \#\{p \in \mathbb{P}^1(\mathbb{F}_q) \mid p \neq 0, \infty\} = q - 1$.*

15.4.2 Finite higher laminations for triangles

In this subsection, we present a graphical way to compute the product in the Hecke algebra in the standard basis. Our graphs are very similar to the ones that appear in the article [EW16] by Elias–Williamson, in the context of the categorification of the Hecke algebras in terms of Soergel bimodules.

Definition

We use the normalized version of the Hecke algebra where the quadratic relation reads $h_s^2 = 1 + Qh_s$ where $Q = v - v^{-1} = q^{1/2} - q^{-1/2}$.

The idea is to use the interpretation of the structure constants as triples of flags from the previous subsection. The transition between two flags with given relative position is decomposed into elementary moves. This corresponds to a decomposition of a product into elementary ones involving simple elements only. For example the quadratic relation is graphically given by the following picture:

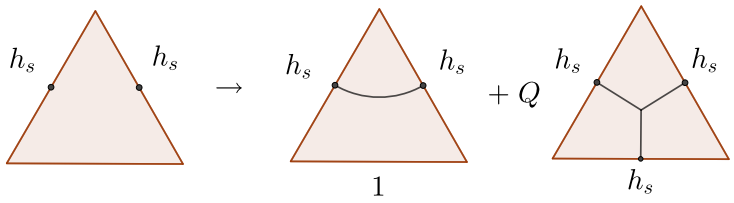


Figure 15.8: Graphical multiplication in Hecke algebra

More generally for $x, y \in W$ the product $h_x h_y$ is represented as a formal sum of graphs with edges labeled by simple reflections. First one has to fix reduced expressions of x and y, using the simple reflections. Each simple element h_s is represented graphically by an edge labeled by $s \in S$. Elementary computations in $\mathcal{H}$ are either braid or quadratic relations, hence the vertices of the graph are of two types:

- *trivalent* with the three edges carrying the same label, called *ramification point*,
- *of braid type* at the crossing of edges carrying the labels s and t such that $(st)^m = e$ in W. In that case there are m incident edges of type s and m of type t which alternate (the cases $m = 2$ and $m = 3$ are drawn in the middle and on the right of Figure 15.9).

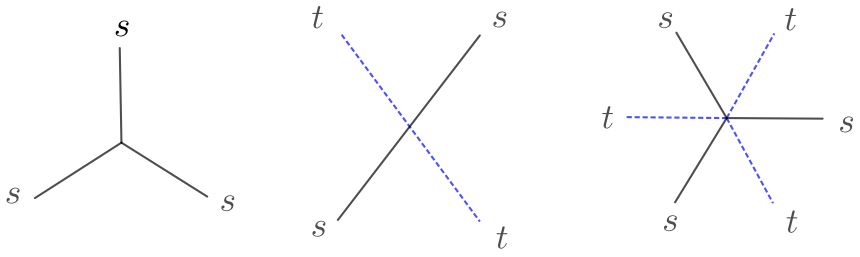


Figure 15.9: Vertex types

Definition 15.25. *A finite higher lamination of type (W, S) (in short, higher lamination) on a triangle t is an equivalence class of planar graphs $\Gamma \subset t$ with edges labeled by elements in S, and satisfying the following criteria:*

1. *each vertex is either trivalent or of braid type,*
2. *the edges of Γ intersect the boundary of t transversally,*
3. *reading the labels along an edge of t gives a reduced word in W,*
4. *the graph is minimal in the sense described below.*

The equivalence class is generated by isotopy and the relations (1.1), (2.1), (2.2), (3.1) and (3.2) described below in Section 15.4.2.

Remark 15.26. *We call these graphs finite higher laminations because they generalize rational bounded measured laminations as described in [FG07]. The rough idea is that one recovers rational bounded measured laminations from the higher laminations for the affine Weyl group $\widehat{A}_1 = \langle s, t \mid s^2 = t^2 = 1 \rangle$ after removing the singular leaves. Details of this correspondence and higher laminations corresponding to affine Coxeter systems will be discussed in details in a forthcoming publication. Aspects of higher laminations as tropical points of higher Teichmüller spaces and generalizations of rational measured laminations have been studied in [Xie13] and [Le16].*

Minimality. Let Γ be an S-labeled graph on an oriented triangle t satisfying the first three conditions of Definition 15.25, and let x, y, z be the elements of W corresponding to the reduced words on the sides of t. A configuration of flags on $t \backslash \Gamma$, one flag for each face, is called **valid** if the relative position of two adjacent flags is given by the simple element which appears as the label on the edge between the two corresponding faces.

Definition 15.27. *The graph Γ is minimal if for all triple of flags, one flag for each vertex of t, there is at most one valid configuration of flags on $t \backslash \Gamma$ which extends the triple of flags.*

An example of a non-minimal graph is given in Figure 15.10: the flag corresponding to the inner component of $t \backslash \Gamma$ is not uniquely determined by the flags around.

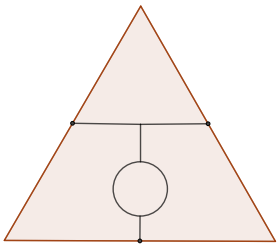


Figure 15.10: Example of a non-minimal graph

Remark 15.28. *A simple criterion to decide whether a graph is minimal or not is still lacking. In particular, we would like to find a minimality criterion for general Coxeter systems (W, S) for which flags are not defined.*

Relations

Given a triple of flags, there might be several minimal labeled graphs realizing the configuration. In other words, the decomposition into elementary moves between flags is not unique. For example, given two flags in relative position $w \in W$ any reduced expression for w is a working decomposition into elementary moves. Hence we have to quotient out by relations in order to count each triple of flags only once in Proposition 15.23. We describe here a set of relations that we conjecture to be complete.

There are relations involving one, two and three different labels (also called "colors"). The one-color relation is given by the following picture:

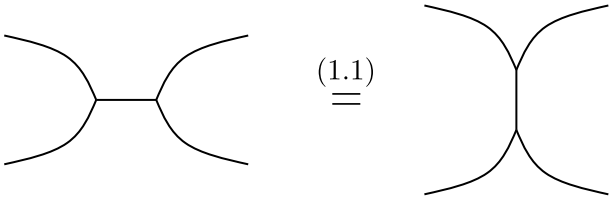


Figure 15.11: One-color relation

The two-color relations are of two kinds: the first states that one can simplify two neighbor vertices when they are of the same braid type, as shown on the left of Figure 15.12 for $m = 3$ (for $m = 2$ this gives the second Reidemeister move). The second relation describes how to glide a trivalent vertex through a vertex of braid type and is shown on the center (resp. right) of Figure 15.12 for $m = 2$ (resp. $m = 3$).

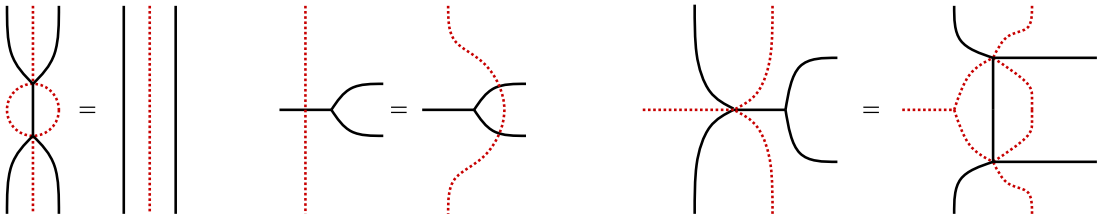


Figure 15.12: Two-color relation

Eventually there is a three-color relation for each parabolic subgroup of rank 3 in W, displayed in [EW16, Section 5]. Two examples are presented in Figure 15.13: the upper-one is the relation corresponding to a subgroup of type $A_1 \times A_1 \times A_1$ (it is the third Reidemeister move) and the lower-one corresponds to a subgroup of type A_3.

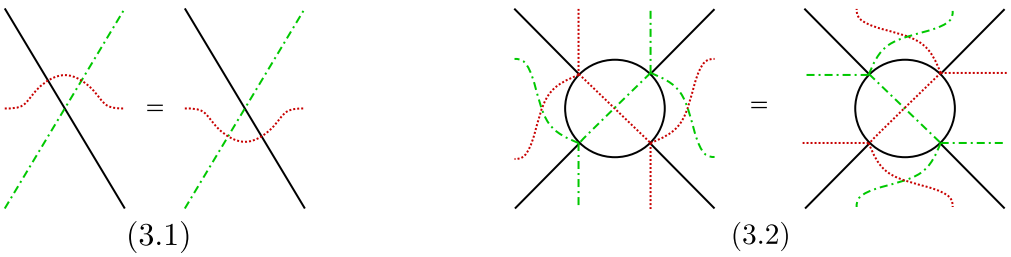


Figure 15.13: Three-color relation

Remark 15.29. *All the relations of [EW16] which do not imply a loose end are relations for our higher laminations. It seems there is a link between the graphical calculus in the Hecke algebra we introduced*

and the graphical calculus of Elias–Williamson describing morphisms between Bott–Samelson bimodules, even we do not understand the correspondence yet.

Figure 15.14 shows the equivalence of two seemingly different graphs through these relations.

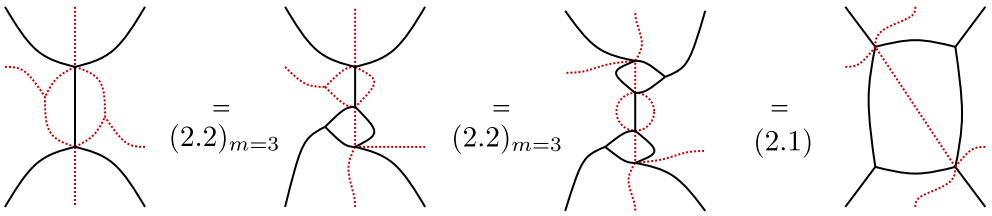


Figure 15.14: An example of applying relations

The two-color relation (2.1) and all three-color relations give a complete list of relations for reduced expressions in W: two reduced expressions of an element of W can be related through a finite number of these relations [Ron09, 2.5]. This is enough to show that the latter are all the relations we need in the case of higher laminations without ramification points.

In general, we conjecture the following:

Conjecture 15.30. *The relations of above are complete: two minimal labeled graphs corresponding to the same triple of flags on the vertices of the triangle can be related through a finite sequence of them.*

The conjecture seems reasonable since we only have to look for relations involving ramification points. Our one-color relation describes how two ramification points interact and the two-color relation (2.2) describes the interaction between a ramification point and a vertex of braid type. It does not seem too presumptuous to expect that these are the only cases one needs to consider.

Existence

We show that we can associate a set of representatives Γ of higher laminations to any multiplication in the Hecke algebra. The non-trivial part is to show minimality, that is, the existence of a unique configuration of flags on the connected components of $t\backslash\Gamma$ (henceforth called the *faces* of Γ).

Let Γ be a representative of a higher lamination. Recall that an assignment of flags to the faces of Γ is *valid* if the relative position of any two adjacent flags is the label of the edge between the corresponding faces.

We start with some easy results, the proof of which is left as an exercise for the interested reader.

Lemma 15.31. *Let $v, w \in W$ such that vw is not reduced. Then there are reduced expressions for v and w of the form $v = v's$ and $w = sw'$ where $s \in S$ is a simple reflection.*

Corollary 15.32. *For $s, t \in S$, consider a reduced expression for $w \in W$ of the form $w = w'u$ where u is a word in the letters s and t with maximal length. Denote by $\bar{u}$ the word obtained from u by exchanging s and t. Then $w'\bar{u}$ is also reduced.*

The two other lemmas concern the extension of flag configurations.

Lemma 15.33. *Given two flags in opposite faces with respect to a vertex a braid type, and whose relative position is compatible with the local structure around the vertex, there is a unique way to associate a valid configuration of flags around the vertex extending the initial data.*

In type A this can be verified easily by a computation in $\mathbb{P}^2$.

Lemma 15.34. *Given two flags F_0 and F_1 in relative position w and a reduced word $w = s_{i_1} \cdots s_{i_k}$, there is a unique sequence of flags $F_0 = F_{i_1}, F_{i_2}, ..., F_{i_{k+1}} = F_1$ such that $(F_{i_l}, F_{i_{l+1}}) = s_{i_l}$ for all $l = 1, ..., k$.*

Any product $h_x h_y$ in $\mathcal{H}$ of elements of the standard basis can be decomposed into elementary moves which are either a quadratic relation or a braid relation, and we have diagrams for these. By juxtaposition we get a collection of graphs corresponding to the product $h_x h_y$.

Proposition 15.35. *The graphs obtained from expressing $h_x h_y$ in the standard basis of $\mathcal{H}$ are representatives of higher laminations.*

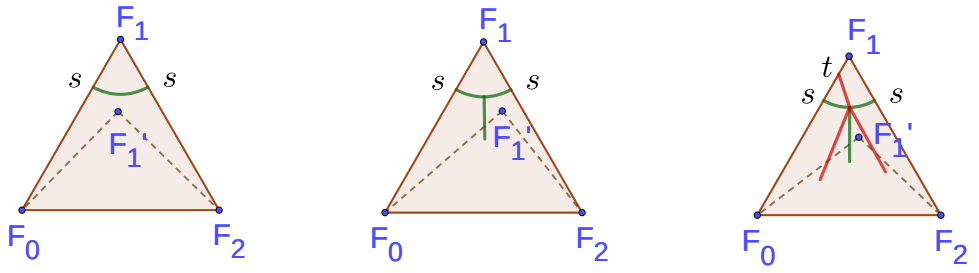


Figure 15.15: Local structure around F_1. Left: simple edge. Middle: ramification point. Right: braid vertex.

Proof. Let $x, y \in W$ and h_x, h_y be the corresponding elements in $\mathcal{H}$. The latter can be written as a product of the h_{s_i} corresponding to the elements in S. Let us fix such a decomposition for h_x and h_y. As one computes the product $h_x h_y$ in $\mathcal{H}$, one uses braid relations and quadratic relations until all the terms one obtains are products of h_{s_i} corresponding to reduced words in W. Each quadratic relation increments the number of terms in the expression by one. Let us consider one of the terminal terms and assume it corresponds to h_z for $z \in W$. Let Γ be the corresponding graph on a triangle t whose sides are respectively labeled by the reduced words corresponding to y, x and z^{-1} as one reads the boundary counterclockwise.

By Proposition 15.23, we can associate a triple of flags (F_0, F_1, F_2) to the vertices of t such that $(F_2, F_1) = y$, $(F_1, F_0) = x$ and $(F_2, F_0) = z$ in $G\backslash(G/B)^2$.

The only non-trivial fact to check is that Γ is minimal, i.e. that we can extend the triple of flags to a unique valid configuration. We prove this by induction on the number of faces of Γ. The initialisation with one face is trivial since it corresponds to $1 \times 1 = 1$.

In general, by Lemma 15.34 we have flags assigned to all boundary faces. We can assume that the local configuration around the vertex with flag F_1 is given by one of the three cases shown in Figure 15.15. This is because away from that vertex, we can only apply braid relations (since x and y are reduced) which amount to choose another reduced expression for x or y.

Case 1: There is a simple edge next to F_1.

The region just this edge F_1 already has an associated flag F_1' since it is a boundary face. The restriction of Γ to the triangle (F_0, F_1', F_2) has strictly less faces and we can choose the boundary of this new triangle so that it is still a higher lamination: first one can clearly assume that Γ intersects the boundary of the new triangle transversely and the sides chosen close to the sides of the original triangle so that the words assigned to them are the words on the sides of the original triangle with the letter corresponding to the edge between F_1 and F_1' removed. Thus we can apply the induction hypothesis.

Case 2: There is a ramification point next to F_1.

As in case 1, the two flags below F_1 are determined by Lemma 15.34. Let F_1' be one of these flags. The restriction of Γ to (F_0, F_1', F_2) has strictly less faces, we can again assume that it is a higher lamination and hence apply the induction hypothesis.

Case 3: There is a braid vertex next to F_1.

Since all boundary faces are already assigned a flag, there are two opposite faces around the braid vertex carrying a a flag and this assignment is consistent with the local structure of Γ. By Lemma 15.33, we can uniquely extend this configuration around all the regions around the braid vertex. Let F_1' be the flag below the braid vertex as shown on the right of Figure 15.15. Let us show that we can apply the induction hypothesis to the restriction of Γ to the triangle formed by (F_0, F_1', F_2). The only non-trivial fact to check is that the expressions induced on the edges (F_0, F_1') and (F_1', F_2) are reduced. By changing the reduced expression for x or y, we can assume that the number of incoming edges in the braid vertex is maximal. The fact that the restriction of Γ is reduced then follows from Corollary 15.32. □

Given a representative Γ of a higher lamination one wonders how many triples of flags modulo G correspond to it. Let $\mathrm{ram}(\Gamma)$ be the number of trivalent vertices of Γ and recall that we denoted $l : W \to \mathbb{N}$ the Bruhat length on (W, S).

Proposition 15.36. *Let Γ be a representative of a higher lamination on a triangle t inducing the labels $x, y, z \in W$ on the sides, and F_0, F_2 two flags in relative position z. The number of flags F_1 such that (F_0, F_1, F_2) extends to a valid configuration on $t\backslash\Gamma$ is:*

$$Q^{\mathrm{ram}(\Gamma)} q^{1/2(l(x)+l(y)-l(z))} \ . \tag{15.44}$$

The proof is similar to the one for Proposition 15.35, with a distinction of the same three cases from Figure 15.15. Note that by Proposition 15.35, any contribution from a triple of flags appears at most once.

Proof. We reason by induction on the number of faces of Γ. The proposition is true for the empty graph.

Case 1: There is a simple edge next to F_1.

The restriction Γ' of Γ to the triangle (F_0, F_1', F_2) is a representative of a higher lamination with strictly less faces. We have $\mathrm{ram}(\Gamma') = \mathrm{ram}(\Gamma)$, $z' = z$, $l(x') = l(x) - 1$ and $l(y') = l(y) - 1$. The only restriction on F_1 is its relative position to F_1' which is a simple reflection. If F_1' is fixed this are q flags satisfying this constraint. Using the induction hypothesis, the number of possible F_1's is:

$$q \times Q^{\mathrm{ram}(\Gamma')} q^{1/2(l(x')+l(y')-l(z'))} = Q^{\mathrm{ram}(\Gamma)} q^{1/2(l(x)+l(y)-l(z))} \ . \tag{15.45}$$

Case 2: There is a ramification point next to F_1.

Let F_1' be a flag in one of the regions below F_1 (say on the boundary of y). The restriction Γ' of Γ to the triangle (F_0, F_1', F_2) is a representative of a higher lamination with strictly less faces. We have $\mathrm{ram}(\Gamma') = \mathrm{ram}(\Gamma) - 1$, $z' = z$, $x' = x$ and $l(y') = l(y) - 1$. If F_1' is fixed the, by Lemma 15.34 the flag F_1'' next to F_1' is uniquely determined. The only restriction on F_1 is its relative position to F_1' and F_1''. Since these are the same simple reflection, there are $q - 1$ possible F_1's. Using the induction hypothesis and the fact that $Q = q^{1/2} - q^{-1/2}$, the number of possible F_1's is:

$$(q-1) \times Q^{\mathrm{ram}(\Gamma')} q^{1/2(l(x')+l(y')-l(z'))} = Q^{\mathrm{ram}(\Gamma)} q^{1/2(l(x)+l(y)-l(z))} \ . \tag{15.46}$$

Case 3: There is a braid vertex next to F_1.

Let F_1' be the flag assigned to the face of Γ below F_1 as shown in Figure 15.15. The restriction Γ' of Γ to the triangle (F_0, F_1', F_2) is a representative of a higher lamination with strictly less faces (as follows from the same reasoning as in the proof of Proposition 15.35). Furthermore, all parameters $\mathrm{ram}(\Gamma'), l(x'), l(y')$ and $l(z')$ are the same as the ones corresponding to Γ. If we know the flag F_1', by Lemma 15.34 this fixes all boundary flags of the triangle (F_0, F_1', F_2). In particular, this gives two flags in opposite regions around the braid vertex. By Lemma 15.33, this determines uniquely F_1. We conclude by the induction hypothesis. □

A consequence of this last proposition is that the number of ramification points $\mathrm{ram}(\Gamma)$ does not depend on the representative of the higher lamination. One can easily check that all the relations of Section 15.4.2 preserve the number of ramification points.

Product in the Hecke algebra

We now describe the graphical computation of a product $h_x h_y$ in the Hecke algebra, for some $x, y \in W$.

Let us choose reduced expressions for x and y and write them on the two upper sides of the triangle following the counterclockwise orientation of the boundary. For each triple of flags appearing in the product as described in Proposition 15.23 we choose a corresponding graph Γ. By Proposition 15.35, these graphs represent higher laminations. On the last edge of the triangle we read a reduced expression for some element $h_z(\Gamma) \in \mathcal{H}$.

Theorem 15.37. *We have*

$$h_x h_y = \sum_{\Gamma} Q^{\mathrm{ram}(\Gamma)} h_z(\Gamma) \ , \tag{15.47}$$

where the sum runs over all isotopy classes of graphs Γ coming from triples of flags associated to $h_x h_y$. Assuming Conjecture 15.30, the sum can be taken over all higher laminations inducing x and y on the two upper sides of the triangle.

Proof. Recall the structure constants ${C_{xy}}^z$ of the Hecke algebra with quadratic relation $T_s^2 = (q-1)T_s + q$. Combining Proposition 15.23 with Proposition 15.36, we get

$$\begin{aligned} {C_{xy}}^z &= \#\{F_1 \in G/B \mid (F_0, F_1) = x, (F_1, F_2) = y\} \\ &= \sum_{\Gamma_z} Q^{\mathrm{ram}(\Gamma_z)} q^{1/2(l(x)+l(y)-l(z))} \ , \end{aligned} \tag{15.48}$$

where the sum is taken over all graphs Γ_z which induce z on the third side of the triangle.

Let us now relate the structure constants of $\mathcal{H}$ in the basis $(T_w)_{w\in W}$ to those in the standard basis $(h_w)_{w\in W}$. The two basis are linked by $T_w = q^{l(w)/2} h_w$. Hence:

$${c_{xy}}^z = q^{-1/2(l(x)+l(y)-l(z))} {C_{xy}}^z \ . \tag{15.49}$$

Therefore, we get:

$$h_x h_y = \sum_{z\in W} {c_{xy}}^z h_z = \sum_{z\in W} \sum_{\Gamma_z} Q^{\mathrm{ram}(\Gamma_z)} h_z = \sum_{\Gamma} Q^{\mathrm{ram}(\Gamma)} h_z(\Gamma) \ . \tag{15.50}$$

Assuming Conjecture 15.30 we can uniquely associate a higher lamination to each triple of flags appearing in $h_x h_y$. □

Let us give a concrete example of the graphical interpretation of a product in the Hecke algebra $\mathcal{H}_{(\mathfrak{S}_3,\{s,t\})}$.

Example 15.38. *Let us multiply h_{sts} with h_{st} in $\mathcal{H}_{(\mathfrak{S}_3,\{s,t\})}$. The direct computation reads:*

$$\begin{aligned} h_{sts} h_{st} &= h_s h_t h_s^2 h_t \\ &= h_s h_t^2 + Q h_s h_t h_s h_t \\ &= h_s + Q h_s h_t + Q h_s^2 h_t h_s \\ &= h_s + Q h_{st} + Q h_{ts} + Q^2 h_{sts} \ , \end{aligned} \tag{15.51}$$

and it corresponds to the graphs of Figure 15.16.

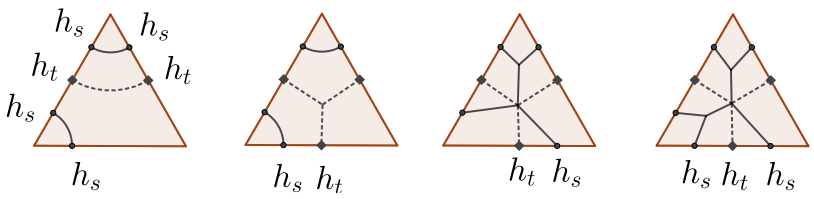


Figure 15.16: The graphical analogue of the product $h_{sts}h_{st}$.

Remark 15.39. *To a higher lamination Γ on a triangle t with set R of ramification points, one can associate a monodromy map $\pi_1(t\backslash R) \to W$ in the following way. To any based loop $\gamma \in \pi_1(t\backslash R)$ which intersects Γ transversely one associates the product of all labels of $\gamma \cap \Gamma$ following the orientation of γ. It is easy to check that it only depends on the homotopy class of γ.*

For higher lamination in type A_n, it is then possible to associate to Γ an n-sheeted cover with simple ramification points at R and a trivialization over each connected component of $t\backslash\Gamma$, such that the transition between two adjacent regions is given by the label on the separating edge. This is why we call the trivalent vertices of Γ ramification points.

15.4.3 Higher laminations for ciliated surfaces

We now define higher laminations on surfaces, which provides a diagrammatic way to compute our polynomial invariant. This viewpoint gives a direct proof of the invariance under $q \mapsto q^{-1}$ for a closed surface (this is Proposition 15.43 below).

Let us consider a ciliated surface Σ where each boundary component is labeled by an element of the Weyl group, and fix a triangulation T of Σ. A *higher lamination* Γ on Σ is the gluing of representatives of higher laminations on all triangles of T (their boundary data has of course to coincide).

Remark 15.40. *It should be possible to define higher laminations without using a triangulation as an equivalence class of minimal labeled graphs inducing the elements of W on the boundary of Σ, mimicking the definition for triangles. The non-trivial point is to define minimality.*

Let us see how to compute our polynomial using the graphical calculus. Draw all possible higher laminations on Σ compatible with the boundary data. Recall that $\mathrm{ram}(\Gamma)$ is the number of trivalent vertices of Γ. As in Theorem 15.37 we fix representatives of the higher laminations associated to triples of flags. Assuming Conjecture 15.30, this choice is irrelevant.

Theorem 15.41. *The polynomial invariant for a ciliated surface Σ and a Weyl group W is given by*

$$P_{\Sigma,W}(Q) = \sum_{\Gamma} Q^{\mathrm{ram}(\Gamma)} \ , \tag{15.52}$$

where the sum runs over all higher laminations of type W.

Proof. Let us consider a triangulation T of Σ and a higher lamination Γ. The latter allows to associate elements of the standard basis h_w to each edge of T. By Theorem 15.37 we know that for a triangle t the contribution of Γ to the structure constant associated to t is given by Q^{m_t} where m_t is the number of ramification points in the triangle. We conclude by the definition of the polynomial invariant:

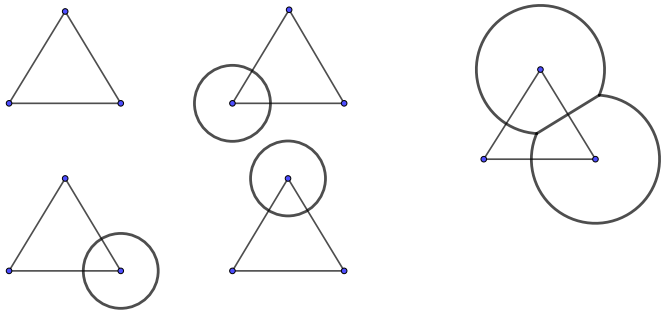
$$\begin{aligned} P_{\Sigma,W}(Q) &= \sum_e \prod_t c_{xyz}(Q) = \sum_e \prod_t \sum_{\Gamma_e} Q^{\mathrm{ram}(\Gamma|_t)} \\ &= \sum_e \sum_{\Gamma_e} Q^{\mathrm{ram}(\Gamma)} \\ &= \sum_{\Gamma} Q^{\mathrm{ram}(\Gamma)} \ , \end{aligned} \tag{15.53}$$

where $\sum_e$ is the sum over all possible labels of the edges of T by elements of W, $\prod_t$ is the product over all faces t of T and $\prod_{\Gamma_e}$ is the product over all higher laminations compatible with the labels on the edges of T. □

Let us see how this works in a simple case:

Example 15.42. *Let us consider the sphere with three holes $\Sigma_{0,3}$ and the Hecke algebra associated with the Coxeter group $\mathfrak{S}_2$. The surface $\Sigma_{0,3}$ is drawn as a triangle in the plane (together with a point at infinity one gets the sphere). We look for all possible higher laminations.*

The following pictures are possible:



Note that the higher lamination which goes around two vertices of the triangle is the same as the circle around the third vertex, since we are on a sphere.

This shows that our polynomial is given by $4 + Q^2 = q + 2 + q^{-1}$.

Proposition 15.43. *For a punctured surface $\Sigma_{g,k}$, the invariant $P_{g,k,W}$ is a polynomial in $q = v^{-2}$. Furthermore, it is invariant under the transformation $q \mapsto q^{-1}$.*

Proof. Using the graphical calculus, we have seen in Theorem 15.41 that

$$P_{g,k,W}(Q) = \sum_{\Gamma} Q^{\mathrm{ram}(\Gamma)} \ , \tag{15.54}$$

where the sum runs over all higher laminations.

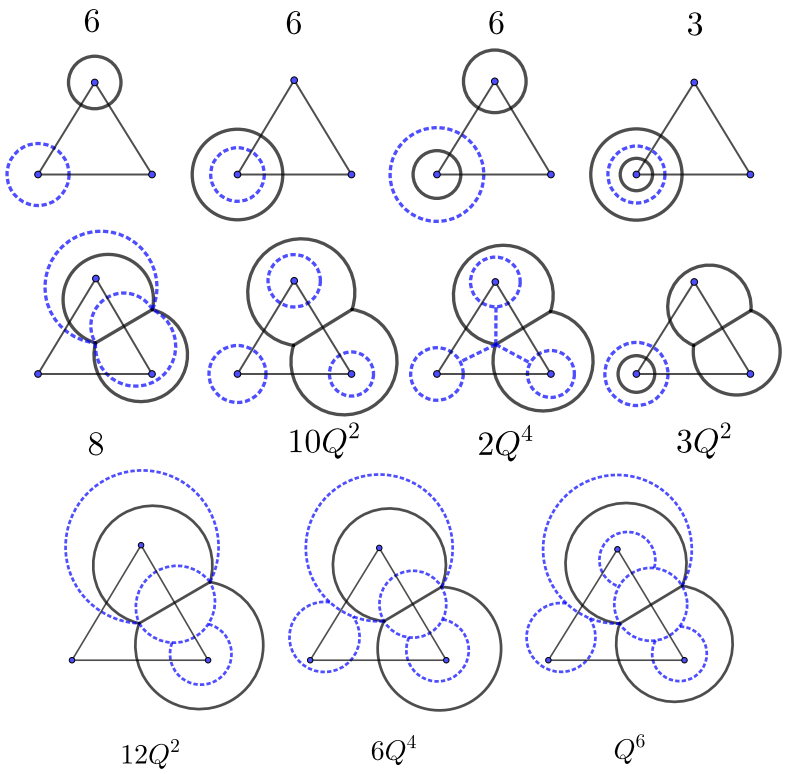
For a punctured surface, the only vertices of odd degree in a higher lamination are the ramification points. Hence there is an even number of them. Therefore our polynomial is given by a polynomial expression in $Q^2 = q - 2 + q^{-1}$ which is both a polynomial in q and invariant under $q \mapsto q^{-1}$. □

Using the graphical calculus, we can compute the first example of an invariant for a Hecke algebra corresponding to a Coxeter system other than $\mathfrak{S}_2$.

Example 15.44. *Let us consider $\Sigma_{0,3}$ with $\mathfrak{S}_3$.*

If we use only one color, then we are reduced to the case $\mathfrak{S}_2$, so the contributions from one color higher laminations to our polynomial are given by $2(4 + Q^2) - 1$ (we have to subtract one in order to count the empty higher lamination only once).

Using both colors, here are the possible pictures with their contributions:



The contribution of a picture is given by Q to the power the number of ramification points, times the multiplicity. For example, the fifth picture has no ramification point and multiplicity 8 since for each of the three dashed loops there are two possibilities (going around a vertex of the triangle or not). For the sixth picture the dashed circles can be there or not with the only requirement that there must be at least one otherwise it is a one-color diagram, hence there are $8 - 1 = 7$ possibilities. Then colors can be exchanged, but this gives only 3 new possibilities.

Adding up all the terms, we get

$$36 + 27Q^2 + 8Q^4 + Q^6 = q^3 + 2q^2 + 10q + 10 + 10q^{-1} + 2q^{-2} + q^{-3}.$$

You see the weakness of the graphical calculus: there is no easy way to check whether all the possible higher laminations have been found.

Remark 15.45. *In Remark 15.39, we have seen that higher laminations are linked to ramified covers of special type (some marking and a minimality condition). Our polynomial counts these ramified covers.*

15.5 Hecke topological quantum field theory

In this section we construct the invariants introduced above in a more intrinsic way. Given a Coxeter system (W, S), we construct a 2-dimensional topological quantum field theory which associates a copy of the Hecke algebra $\mathcal{H}$ or its dual $\mathcal{H}^*$ to topological segments, while ciliated surfaces play the role of cobordisms. Punctured surfaces define elements of the base ring $\mathbb{Z}[v^{\pm 1}]$ which are nothing else than the invariants of Section 15.3.2.

15.5.1 Triangle invariants and gluing revisited

We redefine our construction in more intrinsic terms, showing in particular its independence from any choice of basis in $\mathcal{H}$ and to allow boundary labels to be any elements in $\mathcal{H}$ and not only those of the standard basis.

First, we notice that the non-degenerate pairing given by the trace gives a canonical isomorphism i between $\mathcal{H}$ and $\mathcal{H}^*$. In the standard basis this is given by

$$i : \sum_{w\in W} c_w h_w \mapsto \sum_{w\in W} c_w h_{(w^{-1})} \ . \tag{15.55}$$

The isomorphism i is independent of a basis. To see this, take any basis $(A_w)_{w\in W}$ of the A-module $\mathcal{H}$. Denote by (A^w) the trace-dual basis, i.e. such that $\mathrm{tr}(A_x A^y) = \delta_x^y$ for all $x, y \in W$. Then

$$i : \sum_{w\in W} c_w A_w \mapsto \sum_{w\in W} c_w A^w \tag{15.56}$$

which is in accordance with Equation (15.55) since $h^w = h_{w^{-1}}$.

Second, we can redefine our construction for a triangle, with oriented edges. Choose an orientation of the triangle. To the triangle, we associate the tensor $c \in \mathcal{H}^a \otimes \mathcal{H}^b \otimes \mathcal{H}^c$ which comes from the multiplication in $\mathcal{H}$. Here, $a, b, c \in \{1, *\}$ depending on whether the orientation of the edge is in accordance with the orientation of the triangle or not. Whenever the orientation of an edge is not consistent with the orientation of the triangle, we can use the isomorphism i. For example, in Figure 15.17 we get a tensor $c \in \mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}^*$ which to $h, h' \in \mathcal{H}$ and $g \in \mathcal{H}^*$ associates

$$c(h, h', g) = g(hh') = \mathrm{tr}(hh' i(g)) \ .$$

Since $c_{xyz} = \mathrm{tr}(h_x h_y h_z)$ we get the link to our initial construction in Section 15.3.1.

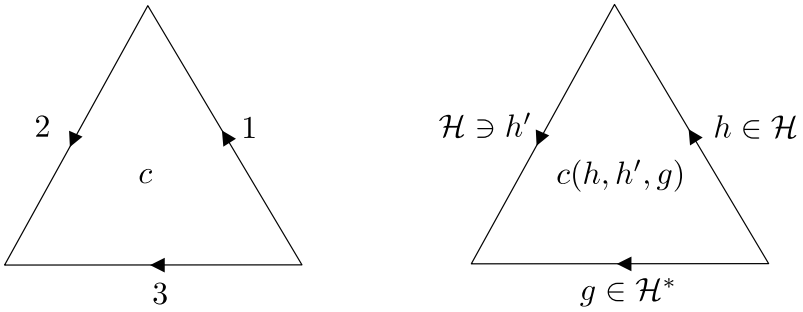


Figure 15.17: Assigning a tensor to triangles.

Finally, when we glue two triangles along an edge with opposite orientations, we use the natural pairing between $\mathcal{H}^*$ and $\mathcal{H}$. Consider for example the two triangles drawn on the left of Figure 15.18. The upper-one corresponds to the tensor $c_u = c \in \mathcal{H}_1^* \otimes \mathcal{H}_2^* \otimes \mathcal{H}_3^*$ which is the one of Equation (15.19), and the lower-one, to the corresponding tensor $c_d \in \mathcal{H}_4 \otimes \mathcal{H}_5^* \otimes \mathcal{H}_6^*$. The indices tell to which edge the copies of $\mathcal{H}$ or $\mathcal{H}^*$ correspond. The edge 3 is positively oriented with respect to the upper triangle hence is associated a copy of $\mathcal{H}$ while the edge 4 is negatively oriented with respect to the lower one and is associated a copy of $\mathcal{H}^*$. Hence we can glue the edges 3 and 4 together.

It is clear that the gluing does not depend on the orientation of the edge (since the isomorphism i can be used to change this orientation). Since our redefinition is a linear extention of the construction of Section 15.3, the gluing does not depend on which diagonal of the quadrilateral we choose (by Theorem 15.18).

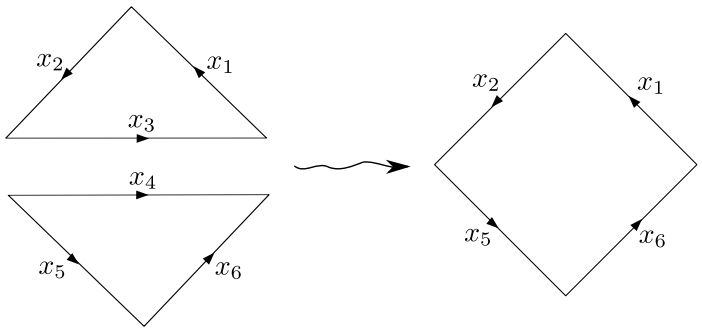


Figure 15.18: Gluing two triangles along an edge.

One can glue arbitrary ciliated surfaces together along edges in the same way. Since any ciliated surface of our interest admits a triangulation, it is assigned a tensor invariant by gluing the invariants corresponding to the faces of its triangulation. This tensor does not depend on the triangulation since any two triangulations can be related by a finite sequence of flips, hence it is a topological invariant of the ciliated surface.

15.5.2 Hecke topological quantum field theories

First we describe our category $\mathcal{C}$ of cobordisms: the objects are disjoint unions of oriented segments, and the morphisms from an object A to another object B are the ciliated surfaces whose boundary is the disjoint union of B and A with the orientation reversed.

Let us consider the functor

$$F : \mathcal{C} \to \mathbb{Z}[v^{\pm 1}] - \text{Mod} \tag{15.57}$$

which associates a copy of $\mathcal{H}$ to the positively oriented segment, a copy of $\mathcal{H}^*$ to the negatively oriented segment, and the proper tensor product of copies of $\mathcal{H}$ and $\mathcal{H}^*$ to a disjoint union of oriented segments. For A and B two objects in $\mathcal{C}$ and a ciliated surface $\Sigma \in \text{Hom}(A, B)$, the morphism $F(\Sigma)$ is the tensor invariant constructed in Section 15.5.1 which is indeed in $F(\partial\Sigma)$. By Theorem 15.18, we know that this tensor only depends on Σ and not on a triangulation.

Theorem 15.46. *For any finite Coxeter system (W, S) the functor F satisfies the axioms of a topological quantum field theory listed in [Ati88].*

Proof. Our construction makes clear that F is invariant under orientation preserving diffeomorphisms of ciliated surfaces and their boundary, that it is involutory, and that it is multiplicative. □

Let Σ be a ciliated surface without cilia and hence without boundary. Then $F(\Sigma) \in \mathbb{Z}[v^{\pm 1}]$, that is, it a Laurent polynomial in v. It clearly coincides with the invariant for punctured surfaces defined in Section 15.3.

15.5.3 Invariants of n-gons

Since our construction is a TQFT, i.e. behaves well under gluing, we can decompose a ciliated surface into elementary parts. Representing a surface as the gluing of the edges of a polygon, we get simple expressions for the invariant.

Consider an arbitrary basis $(C_w)_{w\in W}$ of the A-module $\mathcal{H}$ and let $(C^w)_{w\in W}$ the trace-dual basis in $\mathcal{H}$. The tensor associated to a polygon is simply given by

Proposition 15.47. *Consider a polygon with n edges, all oriented in the same way. Then the tensor c_n associated to the polygon is given by*

$$c_n = \text{tr}(C^{w_1} \cdots C^{w_n}) C_{w_1} \otimes \cdots \otimes C_{w_n} \ . \tag{15.58}$$

Note that this is equivalent to: the tensor of a polygon associates to $(h_1, ..., h_n) \in \mathcal{H}^n$ (one label for each edge) the scalar $\text{tr}(h_1 \cdots h_n) \in \mathbb{Z}[v^{\pm 1}]$.

Proof. We use induction on n. For $n = 3$ the formula is true by the definition of our invariant. For the passage from n to $n+1$, decompose a $(n+1)$-gon into one triangle and an n-gon (see Figure 15.19. We then get:

$$\begin{aligned} c_{n+1} &= \left\langle \mathrm{tr}\Big(C^{w_1}\cdots C^{w'_n}\Big)\, C_{w_1}\otimes\cdots\otimes C_{w'_n}, \mathrm{tr}\big(C^{w_n}C^{w_{n+1}}C_{w'_n}\big)\, C_{w_n}\otimes C_{w_{n+1}}\otimes C^{w'_n}\right\rangle \\ &= \mathrm{tr}\Big(C^{w_1}\cdots C^{w'_n}\Big)\,\mathrm{tr}\big(C^{w_n}C^{w_{n+1}}C_{w''_n}\big)\, C^{w'_n}(C_{w'_n})\, C_{w_1}\otimes\cdots\otimes C_{w_n}\otimes C_{w_{n+1}} \\ &= \mathrm{tr}(C^{w_1}\cdots C^{w_n}C^{w_{n+1}})\, C_{w_1}\otimes\cdots\otimes C_{w_n}\otimes C_{w_{n+1}}\ , \end{aligned}$$

where we used the induction hypothesis in the first line, the gluing formula in the second, and the contraction of indices property of tensors in the last line. □

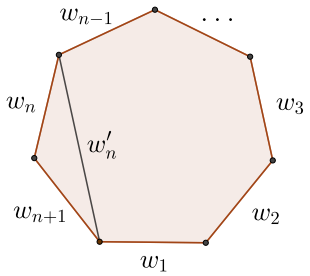


Figure 15.19: Decomposition of a polygon.

As a consequence, we get a simple formula for the invariant of a ciliated surface without boundary components (only marked points):

Corollary 15.48. *The polynomial invariant for a surface $\Sigma_{g,k}$ (where $k \geq 1$) is given by*

$$P_{g,k,W} = \mathrm{tr}\ (C_w C^w)^{k-1}(C_x C_y C^x C^y)^g\ . \tag{15.59}$$

Proof. A surface $\Sigma_{g,k}$ can be obtained as the gluing of a polygon with $4g+k-1$ edges (like in Figure 15.7 for $k = 1$). The first $4g$ edges add handles, while the last $k-1$ edges add cones (so a marked point). By Proposition 15.47, the gluing property and contraction, we get the result. □

Remark 15.49. *There are lots of possible expressions, all equivalent, for a given surface $\Sigma_{g,k}$, one for each gluing of a polygon giving $\Sigma_{g,k}$.*

Example 15.50. *Consider the sphere with three punctures, represented by the gluing of two triangles as in Figure 15.20. Reading the picture yields a map $\mathbb{Z}[v^{\pm 1}] \to \mathbb{Z}[v^{\pm 1}]$ given by*

$$1 \mapsto \mathrm{tr}(C^x C^y C^z)\,\mathrm{tr}(C_x C_z C_y) = \mathrm{tr}(C^x X^y C_y C_x)\ . \tag{15.60}$$

This gives the same result as representing the sphere as the gluing of a square.

15.6 Explicit expression and positivity

In this section, we explicitly compute the polynomial invariant for ciliated surfaces and determine in which cases positivity hold. The key observation is that for a punctured surface, the polynomial is the trace of a central element in the Hecke algebra $\mathcal{H}$. The structure of the center can be understood using the Wedderburn decomposition of $\mathcal{H}$. The polynomial is a sum of powers of Schur elements multiplied by irreducible characters.

15.6.1 Schur elements and Wedderburn decomposition

The Hecke algebra with its standard trace is a specific example of a symmetric algebra, that is, an algebra with a non-degenerate trace function. There is a general theory of these algebras which we present briefly here. In particular we expose some general facts about Schur elements and the Wedderburn decomposition for Hecke algebras. Main references are [GP00], Chapter 7 and 8 of [Chl16] and the article [Neu06].

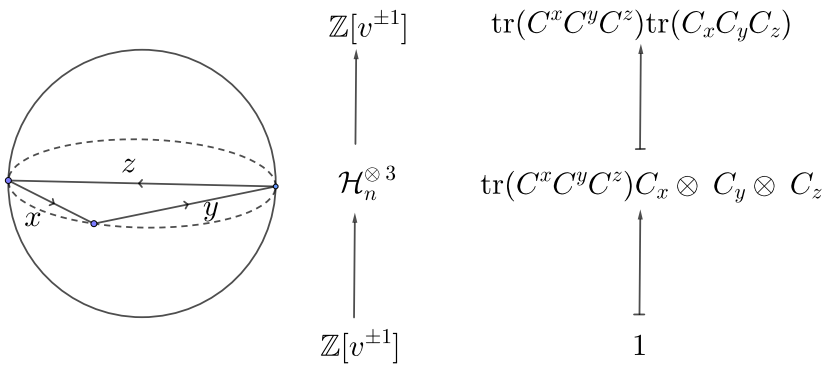


Figure 15.20: Reading off the polynomial

Symmetric algebras. Fix a commutative integral domain A. A *symmetric algebra* H is an A-algebra, which is free and finitely generated over A, equipped with a symmetric trace $\tau : H \to A$. The fact that τ is a trace means that $\tau(hh') = \tau(h'h)\ \forall h, h' \in H$, and that it is symmetric, that $\cdot \to \tau(\cdot h)$ is non-degenerate for all $h \in H$. The trace gives an isomorphism $H \cong H^* = \mathrm{Hom}_A(H, A)$ via $h \in H \mapsto \tau(h\ .\ )$. Denote by B the associated quadratic form $B(h_1, h_2) = \tau(h_1 h_2)$.

Denote by $T(H) \simeq (H/[H, H])^*$ the space of traces on H. Note that the space of traces $T(H)$ is canonically isomorphic to the center $Z(H)$ of H (see for example [GP00, Proposition 7.1.7]).

We further assume H semi-simple. Note that if $A = \mathbb{R}$ and B is positive definite, the algebra H is automatically semi-simple since an orthogonal complement to a left ideal of H is a right ideal and *vice versa*.

Denote by $\mathrm{Irr}(H)$ the set of irreducible representations of H and for $\lambda \in \mathrm{Irr}(H)$ denote by the same letter the map $\lambda : H \to \mathrm{End}(\mathrm{V}_\lambda)$. Then the semi-simplicity of H implies that

$$H = \bigoplus_{\lambda \in \mathrm{Irr}(H)} \mathrm{End}(\mathrm{V}_\lambda)\ . \tag{15.61}$$

This is the Artin–Wedderburn decomposition, or a version of the Peter–Weyl theorem in this setting.

For any representation λ of H one can associate a trace called a character denoted by $\chi_\lambda \in T(H)$, defined by $\chi_\lambda(h) = \mathrm{tr}\lambda(h)$ and a corresponding element Z_λ of the center $Z(H)$. The sets $\{\chi_\lambda\}$ and $\{Z_\lambda\}$ for $\lambda \in \mathrm{Irr}(H)$ form orthogonal bases in the spaces $T(H)$ and $Z(H)$, respectively. Denote by s_λ the inverses the coefficients of the decomposition of unity in H with respect to the base of the Z_λ:

$$\textstyle\sum_\lambda \frac{1}{s_\lambda} Z_\lambda = 1\ . \tag{15.62}$$

The s_λ are called *Schur elements*. From Equation (15.62) we immediately get

$$\tau = \textstyle\sum_\lambda \frac{1}{s_\lambda} \chi_\lambda\ . \tag{15.63}$$

Further, by Equation (15.61), we see that Z_λ acts on V_λ as $s_\lambda\,\mathrm{id}$ (which is the usual definition of the Schur elements). The elements Z_λ can be computed as follows:

Proposition 15.51. *Let $(C_w)_{w\in W}$ be a basis of H and let (C^w) be its trace-dual basis in H. The central element Z_λ can be computed by*

$$Z_\lambda = \sum_{w \in W} \chi_\lambda(C_w) C^w\ . \tag{15.64}$$

Proof. For $h \in H$, we compute (using Einstein summation convention):

$$\tau(Z_\lambda h) = \chi_\lambda(h) = \chi_\lambda(\tau(C^w h) C_w) = \tau(C^w h)\chi_\lambda(C_w) = \tau(\chi_\lambda(C_w) C^w h) \tag{15.65}$$

where we used that $h = \tau(C^w h) C_w$ by definition of C^w. Hence the proposition follows by the non-degeneracy of τ. □

There is another definition of Schur elements which will be useful in the sequel. Let $\varphi : V \to V'$ be an A-morphism between right H-modules. Let $I(\varphi) : V \to V'$ be defined by:

$$I(\varphi) \cdot v = \sum_w \varphi(vC_w)C^w \ . \tag{15.66}$$

The morphism $I(\varphi)$ does not depend on the choice of basis (C_w) and it is a morphism of H-modules [GP00, Lemma 7.1.10].

For $V = V' = V_\lambda$ an irreducible representation, we have [GP00, Theorem 7.2.1]:

Proposition 15.52. *Let $\varphi \in \mathrm{End}(V_\lambda)$. Then:*

$$I(\varphi) = s_\lambda \,\mathrm{tr}(\varphi)\,\mathrm{id} \ . \tag{15.67}$$

Hecke algebras. Now, we specialize to the Hecke algebra $(\mathcal{H}, \mathrm{tr})$ with its standard trace. By [GU89], the Hecke algebra is semisimple over the localized ring $A = \mathbb{Z}[q^{\pm 1}]/P(q)$ where P is the Poincaré polynomial of $\mathcal{H}$. The Artin–Wedderburn theorem implies that

$$\mathcal{H} \simeq \bigoplus_{\lambda \in \mathrm{Irr}(\mathcal{H})} \mathrm{End}(V_\lambda) \ . \tag{15.68}$$

Remark 15.53. *For Hecke algebras, the irreducible representations are all inside so-called left cell representations. For type A, the left cell representations are all irreducible. The article [Neu06] by Neunhöffer describes explicitly the adapted basis for type A, i.e. the matrix elements for each factor* $\mathrm{End}(V_\lambda)$*. We will use these cell representations only for Proposition 15.62 below.*

The decomposition of Equation (15.68) implies that the center of the Hecke algebra is given by diagonal matrices.

Proposition 15.54. *The elements $(Z_\lambda)_{\lambda \in \mathrm{Irr}(\mathcal{H})}$ form a basis of the center $Z(\mathcal{H})$ such that:*

$$Z_\lambda Z_\mu = \delta_{\lambda,\mu} s_\lambda Z_\lambda \ \ \forall\, \lambda, \mu \in \mathrm{Irr}(\mathcal{H}) \ . \tag{15.69}$$

Proof. Since the characters form a basis of the space of trace functions, the Z_λ's form a basis of the center $Z(H)$. The Wedderburn decomposition implies that $Z_\lambda Z_\mu = 0$ for $\lambda \neq \mu$. Now, for all $h \in H$:

$$\mathrm{tr}\big(Z_\lambda^2 h\big) = \chi_\lambda(Z_\lambda h) = \chi_\lambda(s_\lambda h) = \mathrm{tr}(s_\lambda Z_\lambda h) \ . \tag{15.70}$$

Hence $Z_\lambda^2 = s_\lambda Z_\lambda$. □

There is a special symmetry in the Schur elements of Hecke algebras. Let γ be the $\mathbb{Z}[v^{\pm 1}]$-algebra homomorphism on $\mathcal{H}$ given by $\gamma(h_s) = -qh_s^{-1}$. For $\lambda \in \mathrm{Irr}(\mathcal{H})$ let λ^* be the composition $\lambda \circ \gamma$ called the *dual representation*. Proposition 9.4.3 in [GP00] states that:

Proposition 15.55. *For $\lambda \in \mathrm{Irr}(\mathcal{H})$, we have*

$$s_{\lambda^*}(q) = s_\lambda(q^{-1}).$$

15.6.2 Central elements

Recall the expression of the invariant for punctured surfaces in terms of the standard trace of Equation (15.59):

$$P_{g,k,W} = \mathrm{tr}\ (C_w C^w)^{k-1}(C_x C_y C^x C^y)^g.$$

For surfaces with boundary labeled by element in $\mathcal{H}$, terms of the form $C_w h C^w$ arise.

Theorem 15.56. *The elements of the form $(C_w C^w)^{k-1}(C_x C_y C^x C^y)^g$ and $C_w h C^w$ (for $h \in \mathcal{H}$) are in the center of the Hecke algebra.*

Proof. For the first statement, let $s = (C_w C^w)^{k-1}(C_x C_y C^x C^y)^g$. By the non-degeneracy of the trace, it is sufficient to show that

$$\mathrm{tr}(sh_1h_2) = \mathrm{tr}(h_1sh_2) \;\forall\; h_1, h_2 \in \mathcal{H}. \tag{15.71}$$

The expression $\mathrm{tr}(sh_1h_2)$ is the invariant of a surface with two boundary components labeled by h_1 and h_2. The same holds for $\mathrm{tr}(h_1sh_2) = \mathrm{tr}(sh_2h_1)$ with h_1 and h_2 exchanged.

A π-rotation of the surface as shown in Figure 15.21 exchanges h_1 and h_2. Since it does not change the topology, the invariants of the two surfaces coincide. This implies Equation (15.71) and thus the theorem.

The second statement is analogous. The surface corresponding to the invariant

$$\mathrm{tr}\, C_w h C^w h_1 h_2 \tag{15.72}$$

is a cylinder with one boundary labeled by h and another boundary with two cilia and labels h_1 and h_2. Again a π-rotation exchanges h_1 and h_2. □

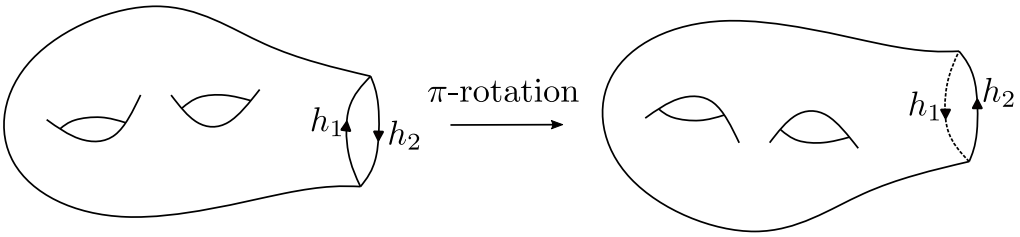


Figure 15.21: Exchanging h_1 and h_2.

Note that the central elements $C_w C^w$ and $C_x C_y C^x C^y$ can be seen as Casimir elements of order 2 and 4 for the Hecke algebra.

From Proposition 15.54 we know that the Schur elements $(Z_\lambda)_{\lambda\in\mathrm{Irr}(\mathcal{H})}$ form a basis of the center $Z(\mathcal{H})$ of the Hecke algebra. We can determine the decomposition of the two building blocks $C_w h C^w$ and $C_x C_y C^x C^y$ of expressions in the form of Equation (15.59) in this basis:

Proposition 15.57. *For $h \in \mathcal{H}$, we have:*

$$C_w h C^w = \textstyle\sum_\lambda \chi_\lambda(h) Z_\lambda \ , \tag{15.73}$$

and

$$C_x C_y C^x C^y = \textstyle\sum_\lambda s_\lambda Z_\lambda \ . \tag{15.74}$$

Note that in particular (for $h = 1$), we get

$$C_w C^w = \textstyle\sum_\lambda \dim(V_\lambda) Z_\lambda \ . \tag{15.75}$$

Proof. For the first assertion, take $V = V' = V_\lambda$ and $\varphi_h \in \mathrm{End}(V_\lambda)$, the action induced by right multiplication by h. Let $h' \in V_\lambda$ and write $C_x h C^x = \sum_\mu e_\mu Z_\mu$. On the one hand we have

$$h' C_x h C^x = I(\varphi_h).h' = s_\lambda \,\mathrm{tr}(\varphi_h) h' = s_\lambda \chi_\lambda(h) h' \tag{15.76}$$

thanks to Proposition 15.52, and on the other

$$h' C_x h C^x = \textstyle\sum_\mu e_\mu Z_\mu h' = e_\lambda s_\lambda h' \tag{15.77}$$

since $Z_\lambda h' = s_\lambda h'$ and 0 for other values of μ. Comparing coefficients we get $e_\lambda = \chi_\lambda(h)$.

For the second assertion, consider the map $\varphi_{\lambda,x} : V_\lambda \to V_\lambda$ given by

$$\varphi_{\lambda,x}(v) = vC^x \ . \tag{15.78}$$

Thus:

$$I(\varphi_{\lambda,x}).v = \sum_y \varphi_{\lambda,x}(vC_y)C^y = \sum_y vC_y C^x C^y \ , \tag{15.79}$$

and by Proposition 15.52:

$$I(\varphi_{\lambda,x}).v = s_\lambda \operatorname{tr}(\varphi_{\lambda,x})v = s_\lambda\, \chi_\lambda(C^x)v \tag{15.80}$$

where we have used the definition of the character χ_λ. Hence:

$$\sum_y vC_yC^xC^y = s_\lambda\, \chi_\lambda(C^x)\, v \text{ for } v \in V_\lambda. \tag{15.81}$$

Recall the Wedderburn decomposition

$$\mathcal{H} \cong \bigoplus_\lambda \operatorname{End}(V_\lambda)\ . \tag{15.82}$$

Let (B_x) be an adapted basis of $\mathcal{H}$ given by the matrix coefficients in the factors of the decomposition, and (B^x) its trace-dual basis in $\mathcal{H}$. Let $\lambda(x)$ be the unique λ such that $B_x \in V_\lambda$.

Since $B_xB_y = 0$ whenever $\lambda(x) \neq \lambda(y)$, we know that $B^x \in \operatorname{Span}(B_y \mid \lambda(y) = \lambda(x))$, and hence $B^xB_y = 0$ for $\lambda(x) \neq \lambda(y)$. Hence $\chi_\lambda(A^x) = 0$ if $\lambda \neq \lambda(x)$.

Equation (15.81) implies that:

$$\sum_y B_xB_yB^xB^y = s_{\lambda(x)}\chi_{\lambda(x)}(B^x)B_x\ , \tag{15.83}$$

and eventually:

$$\begin{aligned}\sum_{x,y} B_xB_yB^xB^y &= \sum_x s_{\lambda(x)}\chi_{\lambda(x)}(B^x)B_x \\ &= \sum_\lambda s_\lambda \sum_{x\,|\,\lambda(x)=\lambda} \chi_\lambda(B^x)B_x \\ &= \sum_\lambda s_\lambda \sum_x \chi_\lambda(B^x)B_x \\ &= \sum_\lambda s_\lambda Z_\lambda\end{aligned} \tag{15.84}$$

where we used that $\chi_\lambda(B^x) = 0$ if $\lambda \neq \lambda(x)$ and that $Z_\lambda = \sum_x \chi_\lambda(B^x)B_x$ from Proposition 15.51. This concludes the proof. □

15.6.3 Explicit expression

We are now ready to compute the invariants for ciliated surfaces. We start with punctured surfaces.

Theorem 15.58. *The polynomial invariant corresponding to a punctured surface is given by*

$$P_{g,k,W}(q) = \sum_\lambda (\dim V_\lambda)^k s_\lambda(q)^{2g-2+k}\ . \tag{15.85}$$

Proof. We have that:

$$\begin{aligned}P_{g,k,W}(q) &= \operatorname{tr}\left((C_wC^w)^{k-1}\,(C_xC_yC^xC^y)^g\right) \\ &= \operatorname{tr}\left(\textstyle\sum_\lambda \dim V_\lambda Z_\lambda\right)^{k-1}\left(\sum_\lambda s_\lambda Z_\lambda\right)^g \\ &= \operatorname{tr}\sum_\lambda (\dim V_\lambda)^{k-1} s_\lambda^{2g-2+k} Z_\lambda \\ &= \sum_\lambda (\dim V_\lambda)^k s_\lambda^{2g-2+k}\ ,\end{aligned} \tag{15.86}$$

where the first equality is Equation (15.59), the second one comes from Proposition 15.57, the third one from Proposition 15.54, and in the last one we used $\operatorname{tr} Z_\lambda = \chi_\lambda(1) = \dim V_\lambda$. □

For example, let us consider the case $W = \mathfrak{S}_2$. There are two Schur elements in $\mathcal{H}_{\mathfrak{S}_2}$, respectively $s_1 = 1 + q$ and $s_2 = 1 + q^{-1}$. Hence the invariant corresponding to a genus g surface with k punctures is given by

$$P_{g,k,\mathfrak{S}_2} = (1+q)^{2g-2+k} + (1+q^{-1})^{2g-2+k} \ . \tag{15.87}$$

Looking at the explicit expression in Theorem 15.58, we observe several phenomena:

- We can evaluate the expression at $k = 0$, although we have no definition of our invariant for closed surfaces. In view of Remark 15.39 this is not surprising since our polynomial seems to count special ramified coverings over Σ.

- Specializing to $q = 1$ and using $s_\lambda(1) = |W|(\dim V_\lambda)^{-1}$ (in type A this is the *hook length formula*), we recover the result of Equation (15.37).

- The invariance under $q \mapsto q^{-1}$ of our polynomial can be obtained from the explicit expression of Theorem 15.58 and the duality property of Schur elements of Proposition 15.55. Interestingly, the coefficients of each Schur element are symmetric as can be observed by a case-by-case study (see Proposition 15.68). It would be nice to recover this property using the TQFTs developed in this paper.

Let us turn to the case of a ciliated surface $\Sigma_{g,k,\{p_1,\dots,p_n\}}$ of genus g, $k \geq 1$ punctures and boundary components with $p_i \geq 1$ cilia. Let $h_i \in \mathcal{H}$ be the product of the elements along the i-th boundary component following the orientation of Σ (h_i *per se* is not well-defined but its conjugacy class is and it is enough to write Theorem 15.59 unambiguously since characters are class functions).

Theorem 15.59. *For a ciliated surface $\Sigma_{g,k,\{p_1,\dots,p_n\}}$ with labels $(h_1, \dots, h_n) \in \mathcal{H}^n$, the polynomial invariant is:*

$$P_{\Sigma,W}(q) = \sum_\lambda (\dim V_\lambda)^k s_\lambda(q)^{2g-2+k+n} \chi_\lambda(h_1) \cdots \chi_\lambda(h_n) \ . \tag{15.88}$$

Proof. We can suppose $n \geq 1$ since the punctures case was already treated in Theorem 15.58. There is a polygonal gluing yielding the ciliated surface Σ in the following way: start with a disc with one cilia on the boundary. Then glue g handles to it, giving a term $(C_x C_y C^x C^y)^g$, then add k punctures, giving a term $(C_w C^w)^k$, add other boundary components with labels h_2 to h_n, giving a term $\prod_{i=2}^n C_w h_i C^w$, and finally add the label h_1 to the initial boundary circle.

Hence, we get:

$$\begin{aligned} P_{\Sigma,W}(q) &= \mathrm{tr}\ (C_x C_y C^x C^y)^g (C_w C^w)^k \textstyle\prod_{i=2}^n (C_w h_i C^w)\ h_1 \\ &= \mathrm{tr}\ (\textstyle\sum_\lambda s_\lambda^{2g-1} Z_\lambda)(\sum_\mu (\dim V_\mu)^k s_\mu^{k-1} Z_\mu) \prod_{i=2}^n (\sum_{\lambda_i} \chi_{\lambda_i}(h_i) Z_{\lambda_i})\ h_1 \\ &= \sum_\lambda (\dim V_\lambda)^k s_\lambda^{2g-2+k+n} \textstyle\prod_{i=2}^n \chi_\lambda(h_i)\ \mathrm{tr}(Z_\lambda h_1) \\ &= \sum_\lambda (\dim V_\lambda)^k s_\lambda^{2g-2+k+n} \textstyle\prod_{i=1}^n \chi_\lambda(h_i) \ , \end{aligned} \tag{15.89}$$

where we used Equation (15.59), Proposition 15.54 and the equality $\mathrm{tr}(Z_\lambda h_1) = \chi_\lambda(h_1)$. □

Two remarks on the explicit expression:

- A puncture is equivalent to a boundary component labeled by $1 \in \mathcal{H}$, since $\dim(V_\lambda) = \chi_\lambda(1)$.

- The compatibility of Theorem 15.59 with the gluing property of Equation (15.33) is ensured by the orthogonality of the irreducible characters (see [GP00, Corollary 7.2.4]):

$$\chi_\lambda(C_w)\chi_\mu(C^w) = \delta_{\lambda,\mu} s_\lambda \dim(V_\lambda) \ . \tag{15.90}$$

15.6.4 Positivity properties

We now turn to the positivity properties of our invariants. For punctured surfaces this reduces to the study of positivity properties of the Schur elements. For a ciliated surface, the characters of elements in the Kazhdan–Lusztig basis appear.

In Section 15.8 we carry out the study of positivity properties of Schur elements, using a formula of [Chl16]. We obtain the following theorem:

Theorem 15.60. *The following positivity properties hold for punctures surfaces:*

1. *The polynomial invariant $P_{g,k,W}(q)$ has positive coefficients for all classical W and for the exceptional types H_3, E_6 and E_7.*

2. *For all other types, i.e. $I_2(m)$ for $m \geq 5$, H_4, F_4 and E_8, the invariants may have negative coefficients.*

The first part (apart from H_3) is a direct consequence of the positivity of Schur elements, the second is a case-by-case study. We are thankful to Sebastian Manecke for his contribution to the case H_3.

Proof. By Theorem 15.67, the Schur elements have positive coefficients in all classical types and for E_6 and E_7. Hence Theorem 15.58 implies the positivity of the coefficients of the invariants for punctured surfaces in those cases.

Let us now study the case H_3. An explicit computation shows that there are only two Schur elements with negative coefficients which only differ by a shift by q^5. In other words, we can write these two elements as some Laurent polynomial P and q^5P. The corresponding irreducible representations of $\mathcal{H}_{H_3}$ are 3-dimensional. There are two other Schur elements whose representation is 3-dimensional. These are of the form Q and q^5Q for some other Laurent polynomial Q. More explicitly let $a = \cos\left(\frac{2\pi}{5}\right)$ and $b = \cos\left(\frac{4\pi}{5}\right)$. Then:

$$\begin{aligned} P(q) &= (2-2b)q + (6-6b) + (7-2b)q^{-1} - 10aq^{-2} \\ &\quad - 10aq^{-3} + (7-2b)q^{-4} + (6-6b)q^{-5} + (2-2b)q^{-6} \\ Q(q) &= (2-2a)q + (6-6a) + (7-2a)q^{-1} - 10bq^{-2} \\ &\quad - 10bq^{-3} + (7-2a)q^{-4} + (6-6a)q^{-5} + (2-2a)q^{-6}\ . \end{aligned} \tag{15.91}$$

We are going to show that $P^l + Q^l$ *has positive coefficients for all* l. This implies the positivity of the coefficients of the invariants corresponding to H_3 by the explicit formula of Theorem 15.58.

We note that the coefficients of Q are positive, and so are those of Q^l for all $l \geq 1$. We prove by induction $l \mapsto l+5$ that P^l has positive coefficients for $l \geq 5$. This is checked by direct computation for P^5 to P^9. Then $P^{l+5} = P^5P^l$ gives the induction heredity. For $l \in \{1,2,3,4\}$ one checks explicitly that $P^l + Q^l$ is positive.

To prove the second part, we note by explicit computation that in types G_2, F_4, H_4 and E_8, the polynomial has negative coefficients for $g = 0$ and $k = 3$.

Consider now type $I_2(m)$ with $m = 2l+1$ odd. Then by [GP00, Theorem 8.3.4], the Schur elements are given by

$$\begin{aligned} s_0 &= 1 + 2q + 2q^2 + \ldots + 2q^{m-1} + q^m \\ s_{l+1} &= 1 + 2q^{-1} + \ldots + 2q^{1-m} + q^{-m} \\ s_j &= \frac{m}{2 - 2\cos\left(\frac{2\pi j}{m}\right)}\left(q - 2\cos\left(\frac{2\pi j}{m}\right) + q^{-1}\right) \quad \text{for} \quad 1 \leq j \leq l \end{aligned} \tag{15.92}$$

where s_0 and s_{l+1} correspond to 1-dimensional representations and all others to representations which are of dimension 2. Hence the constant term of $P_{0,3,W}(q)$ is given by

$$2 - 8m\sum_{j=1}^{l} \frac{\cos\left(\frac{2\pi j}{m}\right)}{1 - \cos\left(\frac{2\pi j}{m}\right)} = 2 - 8m\left(-l + \frac{1}{2}\sum_{j=1}^{l} \frac{1}{\sin^2\left(\frac{\pi j}{m}\right)}\right)\ . \tag{15.93}$$

For $l \geq 5$ we have $\frac{\pi^2}{m^2} < \frac{2}{5l}$ (by analysis of the roots of a quadratic polynomial in l). Using $\sin(x) \leq x$ we obtain:

$$-l + \frac{1}{2}\sum_{j=1}^{l} \frac{1}{\sin^2(\frac{\pi j}{m})} > -l + \frac{1}{2}\frac{1}{\sin^2(\frac{\pi}{m})} > \frac{l}{4} \ . \tag{15.94}$$

Thus, the constant term is smaller than $2 - 2ml$ which is strictly negative for $l \geq 5$. For $l = 3$ and $l = 4$, explicit computations also show negative constant terms. For $l = 2$, one checks that for $g = 0, k = 4$ there are negative coefficients in the polynomial.

Consider type $I_2(m)$ with $m = 2l$ even. Then again by [GP00] Theorem 8.3.4, in addition to the Schur elements given above (for $1 \leq j \leq l-1$ for the 2-dimensional representations), there are two others whose corresponding representation is 1-dimensional given by:

$$s_{\varepsilon_1} = s_{\varepsilon_2} = l(q + 2 + q^{-1}) \ . \tag{15.95}$$

Hence the constant term of $P_{0,3,W}(q)$ is given by

$$2 + 2m - 8m\left(-(l-1) + \sum_{j=1}^{l-1} \frac{1}{2\sin^2(\frac{\pi j}{m})}\right) \ . \tag{15.96}$$

The same techniques as for m odd apply and give that this constant term is negative for $l \geq 5$. For $l = 3, 4$, explicit computations show that the constant term is still negative. This concludes for the cases $I_2(m)$ with $m \geq 5$. □

Remark 15.61. *Recall that the invariants are traces of elements of the form*

$$(C_w C^w)^{k-1}(C_x C_y C^x C^y)^g \ . \tag{15.97}$$

These expressions are independent of the choice of basis (C_w), and hence we may choose the Kazhdan–Lusztig basis (b_w) to do the computations. Despite all the known positivity results concerning this very special basis of the Hecke algebra, we are unable to deduce from that the positivity of our invariants because of the appearance of the dual basis (b^w). Theorem 15.60 can be seen as a family of positivity properties in the center of the Hecke algebra indexed by punctured surfaces.

Let us analyze the case of a general ciliated surface with boundary components labeled by elements of the Kazhdan–Lusztig basis (KL basis for short in the sequel).

Proposition 15.62. *In type A, all irreducible characters evaluated at elements of the Kazhdan–Lusztig basis are positive.*

Proof. Fix an irreducible representation with character χ_λ. In type A, we know that this representation is a cell representation for some left cell Λ. Further by [Neu06, Eq. 4] , we have

$$\chi_\lambda(b_w) = \sum_{x \in \Lambda} \operatorname{tr} b_w b_x b^x \ . \tag{15.98}$$

Since the structure constants in the KL basis are positive, all $\operatorname{tr} b_x b^x b_w$ are positive. □

Let $\mathcal{H}_{\geq 0}$ denote the set of elements in the Hecke algebra which have non-negative coordinates in the KL basis.

Corollary 15.63. *In type A, the invariant corresponding to a ciliated surface with boundary labels in $\mathcal{H}_{\geq 0}$, has positive coefficients.*

Proof. Let α be the product of the labels on the boundary following the latter according to the orientation induced by the surface. Since $\mathcal{H}_{\geq 0}$ is stable under product, we get $\alpha \in \mathcal{H}_{\geq 0}$. In type A both the Schur elements and the characters $\chi_\lambda(b_w)$ are positive, and we conclude with the explicit expression of Theorem 15.59. □

For all other types the characters $\chi_\lambda(b_w)$ can have negative coefficients and hence the invariant can have negative coefficients as well. It is for example the case for $g = 2, k = 1, W = B_2$ and $\alpha = b_{rsr}$.

Using the link between the standard trace and the expansion of an element in a given basis, we get:

Corollary 15.64. *In type A, any expression of the form $(C_wC^w)^k(C_xC_yC^xC^y)^g$ has positive coefficients in the dual Kazhdan–Lusztig basis (b^w).*

Proof. The coefficient of $g = (C_wC^w)^h(C_xC_yC^xC^y)^g$ along b^w when expressed in the KL basis is $\operatorname{tr} b_w g$ which is positive by the previous corollary. □

Example 15.65. *For $W = \mathfrak{S}_2$, the Hecke algebra is commutative and the invariant is given by $\operatorname{tr}\big((C_1C^1 + C_sC^s)^m\big)$ in any basis (C_1, C_s). The dual KL basis is $(b^1, b^s) = (h_1 - q^{-1/2}h_s, h_s)$. Using induction, one easily checks that*

$$\operatorname{tr}\big((b_1b^1 + b_sb^s)^m\big) = ((1+q)^{m-1} + (1+q^{-1})^{m-1})b^1 + q^{(m-1)/2}(q^{1/2} + q^{-1/2})^m b^s \ . \tag{15.99}$$

Each coefficient is indeed positive. Note that the first coefficient is $P_{0,m+1,\mathfrak{S}_2}(q)$.

If we use labels with positive coefficients in the basis adapted to the Wedderburn decomposition, we get:

Proposition 15.66. *Consider a ciliated surface with labeled boundary, such that the product of the labels on each boundary component has positive coefficients in the Wedderburn-adapted basis. Then the polynomial invariant is positive.*

Proof. This follows from the explicit expression in Theorem 15.59 and the fact that the characters evaluated at the Wedderburn adapted basis are non-negative. □

15.7 Computation using Sage and CHEVIE

In this section, we describe two ways to compute our invariants by computer, the first using Sage [S+21] and the second using CHEVIE, a package of Gap3 (see [Mic15] and [GHL+96]).

15.7.1 Sage

We recommend the online platform SageMathCell where you can perform computations using Sage without any installation. To compute the polynomial invariant, we use Equation (15.59). For example, a possible code for computing the polynomial $P_{1,3,A_3}$ is:

```
k = 3
g = 1
R.<v> = LaurentPolynomialRing(QQ)
H = IwahoriHeckeAlgebra('A3', v, -1/v)
W = H.coxeter_group()
T = H.T();
S = (sum(T(i)*T(i.inverse()) for i in W))**(k-1)*(sum
(sum(T(i)*T(j)*T(i.inverse())*T(j.inverse()) for j in W)
for i in W))**g
```

Using the parameters v and $-1/v$ for H corresponds to the normalized version of the Hecke algebra in which $(h_s + v)(h_s - 1/v) = 0$. An H.T(w) for w in the Coxeter group is the standard basis element of H corresponding to w, while H.Cp(w) is the element in the KL basis associated with w.

This code yields the sum $(h_wh^w)^{k-1}(h_xh_yh^xh^y)^g$. However we are interested in its trace, hence the $P_{g,k,W}$ is the coefficient of 1 in the result, in which one can replace v^{-2} by q. Here are some examples of computations:

- $P_{0,3,\mathfrak{S}_3}(q) = q^3 + 2q^2 + 10q + 10 + 10q^{-1} + 2q^{-2} + q^{-3}$.
- $P_{1,1,\mathfrak{S}_3}(q) = q^3 + 2q^2 + 4q + 4 + 4q^{-1} + 2q^{-2} + q^{-3}$.
- $P_{0,4,\mathfrak{S}_3}(q) = q^6 + 4q^5 + 8q^4 + 10q^3 + 24q^2 + 36q + 50 + ...$
- $P_{0,3,\mathfrak{S}_4}(q) = q^6 + 3q^5 + 5q^4 + 33q^3 + 67q^2 + 108q + 142 + ...$
- $P_{0,3,G_2}(q) = q^6 + 2q^5 + 2q^4 + 2q^3 + 2q^2 + 72q - 18 + ...$

For a surface with one boundary component labeled by h one needs to multiply the big sum by h, and expand the result in the standard basis (to read of the constant term). Here is an example for type A_3 and h an element in the KL-basis:

```
Cp = H.Cp()
r,s,t = W.simple_reflections()
T(Cp(r*s*t*r)*S)
```

15.7.2 CHEVIE

CHEVIE is a package of Gap3 (not included in Gap4). We recommend the installation from the webpage of Jean Michel. The advantage of CHEVIE is that it knows the Schur elements and characters. So we can use the explicit expression from Theorem 15.58 to compute our polynomial.

Here is a code computing $P_{0,3,E_8}$:

```
g:=0;;
k:=3;;
W:= CoxeterGroup("E",8);;
v:=X(Cyclotomics);; v.name:="v";;
H:=Hecke(W,[[v,-v^-1]]);;
T:=Basis(H,"T");;
Cp:=Basis(H,"C'");;
schur:=SchurElements(H);;
list:=[1..Length(schur)];;
dim:=HeckeCharValues(T());;
Sum(list,i->dim[i]^k*schur[i]^(2*g-2+k));
```

Using the explicit formula from Theorem 15.58, we get the following general formula for type A_2 where $m = 2g - 2 + k$:

$$P_{g,k,\mathfrak{S}_3}(q) = (1 + 2q + 2q^2 + q^3)^m + (1 + 2q^{-1} + 2q^{-2} + q^{-3})^m + 2^k(q + 1 + q^{-1})^m.$$

For type G_2:

$$\begin{aligned} P_{g,k,G_2}(q) = {} & (1 + 2q + ... + 2q^5 + q^6)^m + (1 + 2q^{-1} + ... + 2q^{-5} + q^{-6})^m \\ & + 2(3q^{-1} + 6 + 3q)^m + 2^k(6q - 6 + 6q^{-1})^m + 2^k(2q + 2 + 2q^{-1})^m. \end{aligned}$$

For a surface with one boundary component, we can compute the polynomial using Theorem 15.59: one only need to change the last row of the code above to:

```
h:=Cp(1);;
Sum(list,i->dim[i]^(k-1)*schur[i]^(2*g-2+k)*HeckeCharValues(h)[i]);
```

One can of course change the value of h at will.

15.8 Positivity for Schur elements

We conclude this chapter with a detailed study of the coefficients of Schur elements associated to a Iwahori–Hecke algebra and their positivity. The main tool is an explicit formula for the Schur elements using generalized hook lengths from [Chl16] (see in particular Example 2.5 for more references).

Theorem 15.67. *The Schur elements $s_\lambda(q)$ have positive coefficients for all Coxeter groups of classical type and for the exceptional types E_6 and E_7.*

Proof. For the exceptional types E_6 and E_7 an explicit computation using CHEVIE shows the positivity of the Schur elements, while in all other exceptional types there are Schur elements with negative coefficients.

For the classical types we use the explicit formula from Theorem 4.3 in [Chl16] which was first published in [CJ12]. We need to introduce some notations to state the formula. For classical types, an irreducible representation λ is described by a set of Young diagrams $(\lambda^{(0)}, ..., \lambda^{(l-1)})$ also called a

multipartition (for Weyl groups we have $l = 1$ or $l = 2$). The *generalized hook length* at $(i, j) \in \lambda$ with respect to two Young diagrams λ and μ is

$$h_{i,j}^{\lambda,\mu} = \lambda_i - i + \mu'_j - j + 1 \ , \tag{15.100}$$

where λ_i denotes the length of the i-th row in λ and μ' is the conjugated Young diagram. For $\lambda = \mu$ this gives the usual hook length.

We can now state the explicit formula. The Schur element indexed by a multipartition $\lambda = (\lambda^{(0)}, ..., \lambda^{(l-1)})$ of the integer n is given by

$$s_\lambda(q) = (-1)^{n(l-1)} q^{-N(\bar{\lambda})} \prod_{0 \le s \le l-1} \prod_{(i,j) \in \lambda^{(s)}} \left([h_{i,j}^{\lambda^{(s)},\lambda^{(s)}}]_q \prod_{0 \le t \le l-1, t \ne s} (q^{h_{i,j}^{\lambda^{(s)},\lambda^{(t)}}} Q_s Q_t^{-1} - 1) \right) \ . \tag{15.101}$$

Note that we have not introduced the notation $N(\bar{\lambda})$ because we can ignore this part of the formula for our purposes.

Type A corresponds to $l = 1$ hence the formula simplifies to:

$$s_\lambda(q) = q^{-N(\bar{\lambda})} \prod_{i,j \in \lambda} [h_{i,j}^{\lambda}]_q \ . \tag{15.102}$$

The Schur element is thus a product of *quantized* hook lengths (modulo a shift). These quantum integers are all positive, and hence is the Schur element $s_\lambda(q)$.

Type B corresponds to $l = 2, Q_0 = q$ and $Q_1 = -1$. An irreducible representation is parameterized by a pair of Young diagrams (λ, μ), which yields:

$$s_{\lambda,\mu}(q) = q^{-N(\lambda \cup \mu)} \prod_{(i,j) \in \lambda} [h_{i,j}^{\lambda,\lambda}]_q (q^{2+\lambda_i - i + \mu'_j - j} + 1) \times \prod_{(i,j) \in \mu} [h_{i,j}^{\mu,\mu}]_q (q^{\mu_i - i + \lambda'_j - j} + 1) \ . \tag{15.103}$$

Since the quantum integers are positive, the Schur element again has positive coefficients.

To obtain the results for type D, we have to use a link to type B which is established in [Chl09, Part 2.3] using Clifford theory. The result is that we can reduce type D to $l = 2$ with parameters $Q_0 = q$ and $Q_1 = -q$. An irreducible representation is given by an unordered pair of Young diagrams (λ, μ). If $\lambda \ne \mu$ we get:

$$s_{\lambda,\mu}(q) = \frac{1}{2} q^{-N(\lambda \cup \mu)} \prod_{(i,j) \in \lambda} [h_{i,j}^{\lambda,\lambda}]_q (q^{\lambda_i - i + \mu'_j - j + 1} + 1) \times \prod_{(i,j) \in \mu} [h_{i,j}^{\mu,\mu}]_q (q^{\mu_i - i + \lambda'_j - j + 1} + 1) \ . \tag{15.104}$$

For $\lambda = \mu$ we get

$$s_{\lambda,\lambda}(q) = q^{-N(\lambda \cup \lambda)} \prod_{(i,j) \in \lambda} [h_{i,j}^{\lambda,\lambda}]_q^2 (q^{\lambda_i - i + \mu'_j - j + 1} + 1)^2 \ . \tag{15.105}$$

In both cases the expressions show the positivity of the Schur elements. □

Apart from the positivity, the coefficients of the Schur elements for Coxeter groups satisfy other interesting properties. Call a sequence of integers $(a_1, a_2, ..., a_n)$ *symmetric* if $a_i = a_{n-i}$ and call it *log-concave* if $a_i^2 \ge a_{i-1} a_{i+1}$ (taking the logarithm gives precisely the condition for a concave function).

Proposition 15.68. *For any Coxeter group, the coefficients of the Schur elements are symmetric. In type A_n, the coefficients are in addition log-concave.*

The proof of the first part relies on a general formula for Schur elements in terms of cyclotomic polynomials which are symmetric. The second part relies on Chlouveraki–Jacon's formula and the observation that log-concavity is multiplicative.

Proof. In [Chl16], Formula (2.2) gives the following expression of Schur elements:

$$s_\lambda = \xi_\lambda q^{-a_\lambda} \prod_{\Phi \in \mathrm{Cyc}_\lambda} \Phi(q^{n_{\lambda,\Phi}}) \ , \tag{15.106}$$

where $\xi_\lambda \in \mathbb{R}$, $n_{\lambda,\Phi} \in \mathbb{Z}_{>0}$ and Cyc_λ is a set of cyclotomic polynomials. This formula comes from a case-by-case study.

Since we are interested in the symmetry of the sequence of coefficients, we can ignore the prefactor $\xi_\lambda q^{-a_\lambda}$. Furthermore, cyclotomic polynomials are known to be symmetric. Finally a product of symmetric polynomials is still symmetric. This proves the first part.

In type A_n, we have seen above in Equation (15.102) that

$$s_\lambda(q) = q^{-N(\bar{\lambda})} \prod_{i,j \in \lambda} [h^\lambda_{i,j}]_q \ . \tag{15.107}$$

The sequence of coefficients in a quantum integer is $(1, 1, ..., 1)$ which is positive and log-concave. By Proposition 2 in [Sta89], the product of positive log-concave sequences preserves these properties, thus the coefficients of the Schur elements are log-concave. □

15.9 Perspectives

Our construction can be generalized in different seemingly promising ways, that we list and comment briefly.

Generalization to other symmetric algebras. Our construction of a 2-dimensional quantum field theory for which cobordisms are ciliated surfaces seems to work for any symmetric finitely generated algebra over *nice* rings, for example for cyclotomic Hecke algebras and Yokonuma–Hecke algebras [Chl16]. We believe that the explicit expressions of 15.58 and 15.59 stay true, but properties like the invariance under $q \mapsto q^{-1}$, positivity or the interpretation via counting of higher laminations might not hold in general.

Generalization to affine Hecke algebras. As already underlined in the introduction and in Remark 15.26, our original motivation for the present work was to study the space of functions over character varieties and its canonical basis: higher laminations. The Satake correspondence is expected to play a role in this story. Since it identifies the spherical affine Hecke algebra corresponding to an algebraic reductive Lie group G with the space of representations of the Langlands dual $G^\vee$, the generalization of the TQFTs constructed above to spherical affine Hecke algebras should be related to the corresponding character varieties. Since those symmetric algebras are not finitely generated anymore, the construction of the TQFTs must involve some regularization of the infinite sums which then appear in the gluing process. Higher laminations, which would generalize finite higher laminations to this new setup at least in the case of affine Hecke algebras, would be very similar to the spectral networks of [GMN13b], calling for a physical interpretation of our construction and possibly an understanding in terms of BPS states counting in $4d$ $\mathcal{N} = 2$ theories of class S.

Categorification. Hecke algebras are categorified by Soergel bimodules [Soe07]. Can our construction also be expressed in term of these bimodules? Whereas it was tempting to look for such a categorified version of our TQFTs to explain the positivity properties of the invariants we observed at the very beginning of our study, we have seen that the positivity does not hold in all cases. We still expect a possible categorification - maybe only for type A - which would explain the positivity properties of the Schur elements of the Hecke algebras as described in Section 15.8, or the positivity properties of the invariants corresponsing to punctured surfaces.

Chapter 16

Symmetric rings and TQFTs

In this chapter, we formalize the correspondence between symmetric rings and open-closed topological quantum field theories. In particular, we will show how to construct a TQFT from a symmetric ring, and conversely how a symmetric ring can be retrieved from a TQFT.

16.1 Introduction

16.1.1 Hecke algebras

Iwahori–Hecke algebras, referred to as *Hecke algebras* in the sequel, are associative non-commutative deformations of Coxeter groups depending on a parameter q. They appear naturally in the representation theory of simple Lie groups defined over finite fields $\mathbb{F}_q$.

Geometrically, let G be a simple Lie group defined over $\mathbb{F}_q$ and B a Borel subgroup of G. The Hecke algebra $\mathcal{H}(G,q)$ corresponding to G and q is the algebra of functions on the double coset $B\backslash G/B$ with the convolution product.

More algebraically, let (W,S) be a Coxeter system and $s,t \in S$ two simple reflections. Let $m_{st} \in \mathbb{N} \cup \{\infty\}$ be the order of st in W. The Hecke algebra $\mathcal{H}(W,q)$ admits the following presentation as $\mathbb{Z}[q^{\pm 1}]$-algebra:

$$H(W,q) = \left\langle (T_s)_{s\in S} \;\middle|\; \begin{array}{ll} (T_s)^2 = (q-1)T_s + q & \forall s \in S \\ (T_sT_t)^{m_{st}} = 1 & \forall s \neq t \in S \end{array} \right\rangle . \tag{16.1}$$

Setting $q = 1$ yields the group algebra of W. For any reduced expression $w = s_1 \cdots s_n$ of a $w \in W$, one sets $T_w = T_{s_1} \cdots T_{s_n}$. Matsumoto's lemma implies that T_w does not depend on the reduced expression for w. The Hecke algebra $H(W,q)$ is then generated by $(T_w)_{w\in W}$ as $\mathbb{Z}[q^{\pm 1}]$-module. The two definitions of Hecke algebras coincide when one takes W to be the Weyl group of G.

16.1.2 Symmetric rings

A ring H over an integral domain A (we will be interested in cases where A is the ring of Laurent polynomials $\mathbb{Z}[v^{\pm 1}]$ or a ring containing it) is said to be *symmetric* if it is free as a module over A and endowed with a *symmetrizing trace* $\tau \in H^*$. A trace function on H is a element $\tau \in H^*$ such that for all $h, h' \in H$, $\tau(hh') = \tau(h'h)$. A trace function is said to be symmetrizing if the bilinear form $(h,h') \to \tau(hh')$ is non-degenerate. Equivalently, $\tau \in H^*$ is a symmetrizing trace if its coproduct $\Delta\tau \in H^* \otimes H^*$ is symmetric and invertible.

Let us denote the inverse of $\Delta\tau$ by $C = C_u \otimes C^u \in H \otimes H$, where $\{C_u\}$ is any basis of H and $\{C^u\}$ is its dual with respect to $\Delta\tau$. Here and in the sequel we will omit the summation sign over repeated indices. The fact that C is the inverse of $\Delta\tau$ can be written as:

$$\forall h \in H,\ \tau(hC_u)C^u = h\ , \tag{16.2}$$

which is equivalent to

$$\forall h_1, h_2 \in H,\ \tau(h_1C_u)\tau(h_2C^u) = \tau(h_1h_2)\ . \tag{16.3}$$

Here and below we denote by h or h_i arbitrary elements of H and by z or z_i arbitrary elements of it center Z.

Using C we can define two canonical elements of H – the so called *Euler element* $E = m(C) = C_u C^u$ and the *double Euler element* $G = m(C^2) = C_u C_v \otimes C^u C^v$, where $m : H \otimes H \to H$ is the multiplication map. The element C is an analog of the quadratic Casimir for the semi-simple Lie algebras. Its main property is given by the

Proposition 16.1. *For any $X \in H \otimes H$*

$$XC = CX', \tag{16.4}$$

where we have denoted by $X' \in H \otimes H$ the element X with the tensor factors interchanged.

Proof. Let h be an arbitrary element of H. Introduce the matrix $X^u_w = \tau(hC_w C^u) = \tau(C^u h C_w)$. The identity (16.3) implies that for any $h \in H$ we have $h = \tau(hC^u)C_u = \tau(hC_u)C^u$. Applying (16.2) to hC_w and to $C^w h$ we get $hC_w = X^u_w C_u$ and $C^u h = X^u_w C^w$. It implies that $(h \otimes 1)C = hC_w \otimes C^w = (X^u_w C_u) \otimes C^w = C_u \otimes (X^u_w C^w) = C_u \otimes C^w h = C(1 \otimes h)$.

We show similarly that $(1 \otimes h)C = C(h \otimes 1)$ and therefore that $(h_1 \otimes h_2)C = C(h_2 \otimes h_1)$. □

This proposition has several useful consequences:

Corollary 16.2. *$C_u h C^u \in Z(H)$ for any $h \in H$ where $Z(H)$ denotes the center of H.*

Proof. For any $h_1 \in H$ we have $h_1 C_u h C^u = h_1 m((1 \otimes h)C) = m((h_1 \otimes h)C) = m(C(h \otimes h_1)) = m(C(h \otimes 1))h_1 = C_u h C^u h_1$. □

This corollary implies that $\iota^* : h \mapsto C_u h C^u$ maps to the center and that the map $h \mapsto C_u h C^u E^{-1}$ is a projection $H \to Z$. Of course the latter is defined only if E is invertible.

Remark 16.3. *By [Moo15, Section 14], we know that E is invertible if and only if H is semisimple.*

Corollary 16.4. *$C^2 \in Z \otimes Z$.*

Proof. $C^2 = C_u C_v \otimes C^u C^v = C_v \otimes C^u C^v C_u \in H \otimes Z$ according to the previous corollary. Similarly we can show that $C^2 \in Z \otimes H$. □

If the Euler element E is invertible we can define the trace on Z canonically induced by τ as $\tau^\circ(z) = \tau(zE^{-1})$ ($\tau^\circ \in Z^*$). The inverse element $C^\circ \in Z \otimes Z$ is given by $C^\circ = C^2$. Indeed,

$$\tau(z_1 C_u C_v E^{-1})\tau(z_2 C^u C^v E^{-1}) = \tau(z_1 C^v E^{-1} z_2 C_v E^{-1}) = \tau(z_1 z_2 E^{-1}) \tag{16.5}$$

Corollary 16.5. *Let U and V be two left H-modules. The A-linear maps $\mathrm{Hom}_A(U, V)$ form a bimodule over H. Then for any $\phi \in \mathrm{Hom}_A(U, V)$ we have $C_u \phi C^u \in \mathrm{Hom}_H(U, V)$.*

Proof. The proof is almost the same as for the previous corollary. Abusing notation denote left and right action maps $m : H \otimes \mathrm{Hom}_A(U, V) \to \mathrm{Hom}_A(U, V)$ and $m : \mathrm{Hom}_A(U, V) \otimes H \to \mathrm{Hom}_A(U, V)$ by the same letter as the multiplication map. For arbitrary $h \in H$ we have $hC_u \phi C^u = hm((1 \otimes \phi)C) = m((h \otimes \phi)C) = m((1 \otimes \phi)C(1 \otimes h)) = m(C(\phi \otimes 1))h = C_u \phi C^u h$. □

Corollary 16.6. *For any $h_1, h_2 \in H$ we have $C_u h_1 h_2 C^u = C_u h_2 h_1 C^u$.*

This follows from the fact that:

$$\begin{aligned} C_u h_1 h_2 C^u &= m((1 \otimes h_1 h_2)C) \\ &= m((1 \otimes h_1)C(h_2 \otimes 1)) = C_u h_2 h_1 C^u \ . \end{aligned} \tag{16.6}$$

It implies in particular that the kernel of the map $\iota^* : h \mapsto C_u h C^u$ coincides with the commutator subalgebra $[H, H]$ and thus $H = Z \oplus [H, H]$ as an A-module.

16.1.3 Topological Quantum Field Theories

Here we will introduce the notion of the open-closed topological quantum field theory (TQFT) following essentially [MS06b] and [AN07] in a slightly restrictive way which is sufficient for our needs.

As in Section 15.2.3, we define a *ciliated surface* to be an oriented compact two-dimensional smooth surface with boundary and with marked points inside the surface as well as on the boundary. In order to distinguish the two kinds of marked points we will call the former *punctures* and the latter *cilia*. The connected components of the boundary of the underlying surface are called *holes*.

The boundary $\partial\Sigma$ of a ciliated surface Σ is the collection of oriented intervals and circles obtained by removing cilia from the boundary of the underlying surface. We do not include punctures in the boundary. The boundary of a ciliated surface has an orientation induced by the one of the surface. Given two connected components α and β of $\partial\Sigma$ of the same type one can form another ciliated surface by gluing them together along an orientation-reversing diffeomorphism $\alpha \to \beta$. We denote the resulting surface by $\Sigma/(\alpha \sim \beta)$. Observe that if α and β have one or two common ends, the resulting surface have one or two additional punctures.

An (open-closed) TQFT over an integral domain A (in our case $A = \mathbb{Z}[q^{\pm 1}]$ and its localizations) is a functor from a category of cobordisms whose objects are oriented circles and segments, and whose morphisms are ciliated surfaces, to the category of A-modules. It associates a free A-module $H(X)$ to any collection X of oriented circles and intervals and an element $c_\Sigma \in H(\partial\Sigma)^*$ considered as a linear form on $H(\partial\Sigma)$, to every ciliated surface. It must satisfy the following axioms:

1. *Tensor structure :* If X_1 and X_2 are two disjoint collections of open circles and intervals, $H(X_1 \sqcup X_2) = H(X_1) \otimes H(X_2)$. This implies that for any collection X the space $H(X)$ is a tensor product of the modules $H := H(\text{interval})$ denoted by H and $Z := H(\text{circle})$. If Σ_1 and Σ_2 are two disjoint ciliated surfaces, it must also be the case that $c_{\Sigma_1 \sqcup \Sigma_2} = c_{\Sigma_1} \otimes c_{\Sigma_2}$.

2. *Non-degeneracy:* The forms corresponding to a disc with two cilia $c_{\bullet\bullet} \in H^* \otimes H^*$, the annulus $c_{\circ} \in Z^* \otimes Z^*$ and the annulus with one cilium $c_{\bullet\circ} \in H^* \otimes Z^*$ are of maximal rank over A.

3. *Gluing :* Denote by $C \in H \otimes H$ the inverse of $c_{\bullet\bullet}$ and by $C^\circ \in Z \otimes Z$ the inverse of $c_{\circ}$. If α and β are two intervals in the boundary of Σ than $c_{\Sigma/\alpha\sim\beta} = \langle c_\Sigma, C\rangle_{\alpha\beta}$. Similarly if α and β are two circles, than $c_{\Sigma/\alpha\sim\beta} = \langle c_\Sigma, C^\circ\rangle_{\alpha\beta}$. Here by $\langle\cdot,\cdot\rangle_{\alpha\beta}$ we mean the substitution of the second argument into the first to the places α and β.

4. *Functoriality :* The correspondence $X \to H(X)$ and $\Sigma \to c_\Sigma$ is functorial with respect to the diffeomorphisms. Namely, let $S(X)$ is the group of (homotopy classes of orientation-preserving) diffeomorphisms acting on X. This group acts on $H(X)$ permuting the tensor factors. We require that the subgroup of $S(\partial\Sigma)$ of classes of diffeomorphisms extensible to Σ leaves c_Σ invariant.

16.2 TQFTs and symmetric rings

16.2.1 From a TQFT to a symmetric ring

Given a TQFT one can define the structure of an associative ring with unity and trace on the module H.

For this purpose it is convenient to extend the functor $\Sigma \to c_\Sigma$ to ciliated surfaces whose boundary components are oriented but not necessarily positively, i.e. according to the orientation of Σ. Denote the positively and negatively oriented part of the boundary by $\partial^+\Sigma$ and $\partial^-\Sigma$, respectively. For such surfaces let $c_\Sigma \in H(\partial^+\Sigma)^* \otimes H(\partial^-\Sigma)$ be defined from the form $c_{\tilde\Sigma}$ where $\tilde\Sigma$ is the same ciliated surface as Σ except that its boundary is entirely positively oriented, by applying the isomorphism $C : H^* \to H$ and $C^\circ : Z^* \to Z$ to the factors of $c_{\tilde\Sigma}$ corresponding to the negatively oriented components of the boundary of Σ. The gluing axiom 3 implies that gluing boundary components with opposite orientations amounts just to canonical pairing of the corresponding tensor factors. Such forms $c_\Sigma \in H(\partial^+\Sigma)^* \otimes H(\partial^-\Sigma)$ can equivalently be considered as maps $H(\partial^+\Sigma) \to H(\partial^-\Sigma)$.

Pictorially we mark negatively oriented boundary components by thick lines. When it is convenient to do so, we will also replace the forms $c_\Sigma(\cdots)$ by the picture of Σ with arguments attached to positively oriented (thin) boundary components. The structure of the unital ring with trace on H is given by:

1. The trace is given by $\tau = c_{\bullet\circ} \in \mathrm{Hom}_A(H, A)$.

2. The unity is given by $1 = c_{\bigcirc} \in H$.

3. The product is $c_{\Upsilon} \in \mathrm{Hom}_A(H \otimes H, H)$.

One can verify following for example [AN07] that the properties of the symmetric ring are satisfied. The associativity of the produt follows from the fact that there are two ways of cutting a square into triangles:

$$(h_1h_2)h_3 = \quad = h_1(h_2h_3)\ . \tag{16.7}$$

By construction $c_{\bigcirc} \in \mathrm{Hom}_A(H, H)$ is the identity map, and hence we can verify that $c_{\bigcirc} \in H$ is indeed the unity, since: $h \cdot 1 =$ $= h$. Similarly the identity $\tau(h_1h_2) =$ $= \tau(h_2h_1)$ proves the trace property.

16.2.2 From a symmetric ring to a TQFT

Conversely, given a symmetric algebra (H, τ) over a commutative ring A one can construct a TQFT provided the Euler element E is invertible.

To define a form c_Σ for an arbitrary ciliated surface Σ let us describe its value on a decoration of boundary segments of Σ by elements of H and boundary circles with elements of Z. Let the *perimeter* P of a hole in Σ be the corresponding element of Z if it has no cilium, and the expression $\mathsf{P} = C_u h_1 \dots h_n C^u \in Z$, where $h_1, \dots, h_n$ are the elements of H attached to the segments between the cilia and taken in the cyclic order, otherwise. It follows from Corollary 16.2 that the perimeter is always in the center of H and Corollary 16.6 implies that it does not depend on the choice of the starting segment. Now let us define the form c_Σ by the explicit formula:

$$c_\Sigma(\text{decoration}) = \prod_{\gamma \in \text{ connected components of } \Sigma} \tau\left(G^{g_\gamma} E^{k_\gamma - 1} \prod_{i \text{ hole in } \gamma} \mathsf{P}_i\right)\ , \tag{16.8}$$

where g_γ, and k_γ are respectively the genus and the number of punctures of the connected component γ of Σ, each P_i is the perimeter of the corresponding hole, while E and G are respectively the Euler element and double Euler element of (H, τ).

If we multiply the trace by a constant $\tau \mapsto \kappa\tau$ all forms are multiplied by a power of κ called the *degree* which is easy to compute. Obviously $\deg \tau = 1$, $\deg C = \deg E = -1$, $\deg G = -2$, $\deg \mathsf{P} = 0$ for an empty hole $\deg \mathsf{P} = -1$ otherwise. One can easily deduce from (16.8) that the degree of an arbitrary form is $\deg c_\Sigma = \chi(\Sigma)$, where by $\chi(\Sigma)$ we mean the Euler characteristic of the surface with marked points removed and empty holes filled by disks.

In order to prove that the formula (16.8) indeed defines a TQFT we merely need to check the gluing axiom since the others are obviously satisfied. This, in turn, is an easy consequence of the formulas (16.3) and (16.5). For example, let Σ be the connected surface with one hole and perimeter $\mathsf{P} = C_u h_1 \cdots h_n C^u$, and let us glue two non-adjacent intervals, say the first and the k-th. Then

$$\begin{aligned}\langle c_\Sigma, C_{\alpha\beta}\rangle &= \tau(G^\gamma E^{k-1} C_u C_v h_2 \cdots h_{k-1} C^v h_{k+1} \cdots h_n C^u) \\ &= \tau(G^\gamma E^{k-1} C_u h_{k+1} \cdots h_n C^u C_v h_2 \cdots h_{k-1} C^v) = c_{\Sigma/\alpha\sim\beta}.\end{aligned} \tag{16.9}$$

The second equality follows from the fact that any perimeter belongs to the center.

Similarly, let us glue the 0-th interval of the hole of a connected surface with one hole with the n-th interval of the hole of another connected surface with one hole, such that the perimeters of the single hole of each surface are respectively $\mathsf{P}_1 = C_u h_0 \cdots h_n C^u$ and $\mathsf{P}_2 = C_u h_{k+1} \cdots h_n C^u$:

$$\begin{aligned}\langle c_\Sigma, C_{\alpha\beta}\rangle &= \tau(G^{\gamma_1} E^{k_1 - 1} C_u C_v h_2 \cdots h_k C^u)\tau(G^{\gamma_2} E^{k_2 - 1} C_w h_{k+1} \cdots h_{n-1} C^v C^w) \\ &= \tau(G^{\gamma_1} E^{k_1 - 1} C_u C^w G^{\gamma_2} E^{k_2 - 1} C_w h_{k+1} \cdots h_{n-1} h_1 \cdots h_k C^u) \\ &= \tau(G^{\gamma_1 + \gamma_2} E^{k_1 + k_2 - 1} C_u h_1 \cdots h_{n-1} C^u) = c_{\Sigma/\alpha\sim\beta},\end{aligned} \tag{16.10}$$

where the third equality follows from Equation (16.3). The remaining cases are easily obtained in the same way.

Example 1: Let H be a matrix ring $\mathrm{End}\, V$, where V is a free A-module of rank n. Let the symmetrizing trace be $\tau = s^{-1}\mathrm{tr}$, where tr is the standard matrix trace and s some invertible element of the ring A. The center Z of H can be identified with A. Given a basis of V one can choose the set of matrices $(C_u)_{u=(i,j)\in[|1,n|]^2}$ as the basis of H, where $C_{(i,j)} = e^i{}_j$ is the matrix with only zero components but a 1 at the i-th row and j-th column. Then the elements of the dual basis $(C^u)_{u\in[|1,n|]^2}$ are $C^u = se^j{}_i$ for $u = (i,j)$. The bilinear element C can be written as

$$C = s\sum_{i,j} e^i{}_j \otimes e^j{}_i \ . \tag{16.11}$$

The map to the center is

$$C_u h C^u = se^j{}_i h e^i{}_j = s\,\mathrm{tr}(h)\mathrm{Id}_V \ , \tag{16.12}$$

where Id_V is the identity $n\times n$ matrix (the unit in H), and thus the Euler element is $E = (s\dim V)\,\mathrm{Id}_V$. The double Euler element is

$$G = s^2 e^i{}_j e^k{}_l e^j{}_i e^l{}_k = s^2\delta^k{}_l \mathrm{Id}_V e^l{}_k = s^2\mathrm{Id}_V \ , \tag{16.13}$$

and perimeters are $\mathsf{P}_i = s\ \mathrm{tr}\,(h_i)\,\mathrm{Id}_V$, where $h_i = h_{i,1}\cdots h_{i,n_i}$ is the product of the labels attached between the cilia in a cyclic order. Therefore the forms for connected surfaces are given by

$$c_\Sigma = s^{2g-2+k+m}(\dim V)^k \prod_{i=1}^m \mathrm{tr}\,(h_i)\,. \tag{16.14}$$

Example 2: Let H be a direct sum of matrix algebras:

$$H = \bigoplus_\lambda \mathrm{End}_A\, V_\lambda \ , \tag{16.15}$$

with the symmetrizing trace $\tau = \sum s_\lambda^{-1}\,\mathrm{tr}_\lambda$. The Euler element is

$$E = \bigoplus_\lambda s_\lambda \dim V_\lambda\ \mathrm{Id}_\lambda \ , \tag{16.16}$$

and therefore we see that the coefficients s_λ are determined by the eigenvalues of the Euler element E. The perimeters are

$$\mathsf{P}_i = \bigoplus_\lambda s_\lambda\,\mathrm{tr}_\lambda(h_i)\ \mathrm{Id}_\lambda \ , \tag{16.17}$$

where $h_i = h_{i,1}\cdots h_{i,n_i}$ as in the previous example. Hence the forms are given by

$$c_\Sigma = \sum_\lambda s_\lambda^{2g-2+k+m}(\dim V_\lambda)^k \prod_{i=1}^m \mathrm{tr}_\lambda(h_i). \tag{16.18}$$

16.2.3 Application to the Hecke ring

Let H be the Hecke algebra corresponding to the Coxeter system (W,S). There is a symmetrizing trace τ on H called the *canonical trace* since it corresponds to the canonical trace on the group algebra of the finite group W when $q=1$. It is the map $H\to\mathbb{Z}[q^{\pm1}]$ given by $\sum_{w\in W} c_w T_w \mapsto c_e$, where e denotes the neutral element of W. The corresponding bilinear form C is $C = \sum q^{-l(w)} T_w \otimes T_{w^{-1}}$ and is well defined even over $\mathbb{Z}[q^{\pm1}]$. The map $H\to Z$ given by $h\mapsto T_w h T_{w^{-1}}$ is the deformation of the summation over the group acting by conjugation. The Euler element is thus given by

$$E = \sum_w q^{-l(q)} T_w T_{w^{-1}} \ . \tag{16.19}$$

Its value in the trivial representation (the one where $T_w \mapsto q^{l(w)}$) is $P(q) = \sum_w q^{l(w)}$. The polynomial $P(q)$ is called the *Poincaré polynomial* of the Hecke algebra. It can be considered as a q-analog of the order of the group W. When W is a Weyl group the Poincaré polynomial of H can be written as

$P(q) = \prod_i [e_i]_q$ where the e_i are the exponents of the corresponding Lie algebra, and where we used the standard notation for q-numbers $[n]_q = (1-q^n)/(1-q)$.

Given an irreducible representation of the Hecke algebra $\rho_\lambda : H \to \operatorname{End} V_\lambda$ where V_λ is a free A-module, one can define its character $\chi_\lambda \in H^*$ by $\chi_\lambda(h) = \operatorname{tr}(\rho_\lambda(h))$. The *generic degree* $d_\lambda(q)$ is defined by the identity $d_\lambda = P(q) \dim V_\lambda \ \rho_\lambda(E^{-1})$. The generic degrees are polynomials in q and for the Weyl groups they are q-analogs of the dimensions of the irreducible representations given by q-analogues of the hook length formulas. They are listed for all Weyl groups in [Car85] and for remaining Coxeter groups in [AL82]. We will also use the Laurent polynomials $s_\lambda(q)$ called *Schur elements* given by $s_\lambda(q) = P(q)/d_\lambda(q) = \dim V_\lambda \ \rho_\lambda(E)$.

The Hecke algebra is not semi-simple over $\mathbb{Z}[q^{\pm 1}]$ but it becomes semi-simple over the localized ring $A = \mathbb{Z}[q^{\pm 1}]/P(q)$ [GU89]. Therefore according to the Artin–Wedderburn theorem the corresponding Hecke algebra can be decomposed as

$$(H, \tau) = \bigoplus_\lambda (\operatorname{End}_A V_\lambda, s_\lambda^{-1} \operatorname{tr}_\lambda), \tag{16.20}$$

where λ runs over the set $\operatorname{Irr}(W)$ of irreducible representations of the ring H which coincide with the set of irreducible representations of the Coxeter group W. Observe that since $s_\lambda(q) = P(q)/d_\lambda(q)$ the Schur elements s_λ are invertible in the ring A.

Comparing with the examples above we deduce the forms c_Σ for the Hecke algebra in the cases of connected surfaces:

$$c_\Sigma = \sum_{\lambda \in \operatorname{Irr}(W)} s_\lambda^{2g-2+k+m} (\dim V_\lambda)^k \prod_{i=1}^m \chi_\lambda(h_i). \tag{16.21}$$

Example 3: Let us consider a few particular cases of c_Σ corresponding to surfaces Σ without boundary. First, when Σ is a sphere, one has:

$$c_{\bigcirc} = \tau(E^{-1}) = P^{-2} \sum_\lambda d_\lambda^2 \ . \tag{16.22}$$

It is the only form which is not a Laurent polynomial in q. When Σ is a sphere with one puncture:

$$c_{\bullet} = \tau(1) = 1 = \sum_\lambda s_\lambda^{-1} \dim V_\lambda = P^{-1} \sum_\lambda d_\lambda \dim V_\lambda \ , \tag{16.23}$$

from which one deduces the identity $\sum_\lambda d_\lambda \dim V_\lambda = P$. Adding more punctures on the sphere yields:

$$c_{\bullet\bullet} = \tau(E) = \sum_\lambda (\dim V_\lambda)^2 = \dim H \ , \tag{16.24}$$

$$c_{\bullet\bullet\bullet} = \tau(E^2) = \sum_\lambda s_\lambda (\dim V_\lambda)^3 \ , \tag{16.25}$$

Now, in the cases of the torus and the torus with one puncture one obtains:

$$c_{\text{torus}} = \tau(GE^{-1}) = \sum_\lambda 1 = \dim Z \ , \tag{16.26}$$

$$c_{\text{torus with one puncture}} = \tau(G) = \sum_\lambda s_\lambda \dim V_\lambda \ . \tag{16.27}$$

The Poincaré polynomial of the Hecke ring A_1 is $P = 1+q$ and the generic degrees are $d_{\square\square} = 1$, $d_{\square} = q$, therefore the Schur elements are, respectively $s_{\square\square} = 1+q$, $s_{\square} = 1+q^{-1}$. Therefore we get $c_{\bigcirc} = \frac{1+q^2}{(1+q)^2}$, $c_{\bullet} = 1$, $c_{\bullet\bullet\bullet} = c_{\text{torus with one puncture}} = q+2+q^{-1}$, $c_{\bullet\bullet} = c_{\text{torus}} = 2$.

The Poincaré polynomial of the Hecke ring A_2 is $P = [3]_q! = 1+2q+2q^2+q^3$ and the generic degrees are $d_{\square\square\square} = 1$, $d_{\square\!\square} = q^3$, $d_{\square\!\!\square\!\!\square} = q+q^2$ which yields the Schur elements $s_{\square\square\square} = q^3+2q^2+2q+1$, $s_{\square\!\square} = 1+2q^{-1}+2q^{-2}+q^{-3}$, $s_{\square\!\!\square\!\!\square} = q+1+q^{-1}$. Therefore:

$c_{\bigcirc} = \frac{1+q^2+2q^3+q^4+q^6}{(1+q)^2(1+q+q^2)^2}$, $c_{\bullet} = 1$, $c_{\bullet\bullet\bullet} = q^3+2q^2+10q+10+10q^{-1}+2q^{-2}+q^{-3}$, $c_{\text{torus with one puncture}} = q^3+2q^2+4q+4+4q^{-1}+2q^{-2}+q^{-3}$, $c_{\bullet\bullet} = 6$, $c_{\text{torus}} = 3$.

Index

(p,q)-fivrebrane, 180
(p,q)-string, 180
(p,q)-web, 187, 207, 354
$6d$ $\mathcal{N}=(2,0)$ SCFTs, 364
G-Higgs bundle, 134
G-local system, 133, 136
$\mathcal{A}$-torus of a seed, 129
Θ-positivity, 143
$\mathcal{X}$-torus of a seed, 129
$\mathcal{N}=2$ Coulomb branch, 244
$\mathcal{N}=2$ fractional brane, 233
$\mathcal{N}=4$ Coulomb branch, 244
a-maximization, 209
't Hooft anomaly matching condition, 158
't Hooft coupling, 227

abelianization, 367
AdS–CFT dictionary, 228
affine building, 349
affine flag, 136
affine Hecke algebra, 356
affine maximal parabolic subgroup, 356
affine root, 356
affine toric variety, 199
Affleck–Dine–Seiberg superpotential, 162, 235
amoeba projection, 188
anomalous dimension, 157
anomaly cancellation conditions, 198
Anosov representation, 134, 137
antisymmetric rank assignement, 283
associated flag bundle, 136

Bekenstein-Hawking entropy, 225
Berenstein–Zelevinsky polytope, 350
bipartite graph, 208
black hole evaporation, 225
black hole information paradox, 225
boundary of a ciliated surface, 89
boundary of AdS, 228
bounded measured lamination, 115
BPS bound, 170
BPS state, 170
brane tiling, 208
Bruhat decomposition, 355
bulk, 227, 228

Calabi–Yau manifold, 191, 192
calculable DSB model, 166
central charges a and c, 158
Chan–Paton factor, 177
character variety, 125
chiral anomaly, 158
chiral superfield, 154
ciliated surface, 88, 409
cilium, 88, 409
classical moduli space of vacuua, 156
Clebsch–Gordan decomposition, 345
cluster, 63
cluster $\mathcal{A}$-coordinate, 129
cluster $\mathcal{A}$-mutation, 106, 130
cluster $\mathcal{X}$-mutation, 98, 130
cluster $\mathcal{X}$-coordinate, 129
cluster algebra of finite type, 75
cluster chart, 76
cluster ensemble, 125, 128
cluster modular group, 130
cluster modular groupoid, 130
cluster variable, 63
cluster variety, 125, 128
combinatorial type of a triangulation, 90
complex Grassmannian, 62
configuration space, 135
conformal window of SQCD, 164
conifold, 201, 229
consistent brane tiling, 210, 264
contraction of 2-valent nodes, 71, 213
convex core, 94
Coulomb branch, 169
covariantly constant spinor, 192
Cox homogeneous coordinates, 200
critical dimension, 175

D-brane, 177
D-term, 155
D-term equation, 156
dark matter, 151
daughter theory, 279
decorated Teichmüller space, 88
decoupling limit, 227
deformation fractional brane, 234
degenerate torus fibration, 206

Dehn twist, 88
del Pezzo surface, 200
dessin d'enfant, 368
dimer model, 208
Dirac–Born–Infeld action, 177
disk model of $\mathbb{H}$, 83
domain of discontinuity, 108
double Euler element, 408
DSB fractional brane, 234
dual canonical basis, 61
duality cascade, 232
duality conjectures, 107
Dynkin label, 157

Einstein frame, 176
elliptic isometry of $\mathbb{H}$, 84
Euler element, 408
exchange matrix, 67, 129
extended cluster, 63
extended exchange matrix, 67
extremal black-brane, 178
extremal generators of a cone, 204, 216

F-term, 155
F-term equation, 69, 156
fast inverse algorithm, 213
fat graph, 69, 218, 226, 368, 375
Fayet–Illiopoulos model, 160
Fayet–Illiopoulos term, 155
Fenchel–Nielsen coordinates, 88
flip, 63, 90, 366
flux quantization, 194
forward algorithm, 212, 264
fractional brane, 197, 219, 231
Fricke embedding, 87
frozen variable, 63
frozen vertex, 66
Fuchsian group, 86
Fuchsian model, 86
fundamental group of the groupoid, 92, 126
fundamental groupoid, 92
fundamental meson, 216

gauge engineering, 174
gauge–gravity duality, 231
Gauged Linear Sigma Model, 202
generalized SU(N)-quiver, 363
generalized Maslov index, 144
geodesic lamination, 108
geometric engineering, 191
glide reflection, 317
global symmetries in brane tilings, 217
good train track, 111
groupoid, 92, 126
GSO projection, 175

Haag–Lopuszanski–Sohnius theorem, 154
half-plane model of $\mathbb{H}$, 83
Hanany–Witten transition, 183, 233
Hawking temperature, 225
Hecke algebra, 354
height of a perfect matching, 211
Hermitian Lie group of tube type, 143
Hermitian symmetric space of tube type, 143
hexagonal cluster of the twin SU(5) model, 254
hidden sector, 164
hierarchy problem, 151
Higgs branch, 169
Higgs bundle, 133, 364
Higgs field, 133
higher complex structure, 343
Hilbert basis, 203, 216
Hirzebruch surface, 200
Hitchin equation, 134
Hitchin fibration, 134
Hitchin moduli space, 134, 364
Hitchin section, 134
hive, 350
Hodge diamond of a compact CY3, 192
hole of a ciliated surface, 88, 409
holographic principle, 226
holomorphic 3-form of a CY3, 315
holomorphic gauge coupling, 157
honeycomb, 353
Horn's conjecture, 353
horocycle, 85
hyperbolic isometry of $\mathbb{H}$, 84
hyperbolic metric, 86
hyperbolic surface, 86
hypermultiplet, 168

inverse algorithm, 212, 264
isothermal coordinates, 86

Jacobian algebra (quiver with potential), 69
Jacobian ideal (quiver with potential), 69

Kasteleyn matrix, 210
Kazhdan–Lusztig basis, 377
Kazhdan–Lusztig polynomial, 377
KKLT construction, 194
Kleinian group, 108
Klyachko's saturation conjecture, 350
Konishi anomaly, 167

landscape of string vacua, 194
Langlands dual, 59
Langlands dual group, 357
lattice of characters, 126
lattice of cocharacters, 126
Lie group of Hermitian type, 143
limit set of a Kleinian group, 107
Littlewood–Richardson coefficient, 345, 350
local AdS–WGC, 243

local mirror symmetry, 214
log-canonical Poisson structure, 99, 129

M2-brane, 182
M5-brane, 182
Macaulay2, 203
magical $\mathfrak{sl}_2$-triple, 146
mapping class group, 85
marked complex structure, 85
Markov equation, 80
Markov number, 80
Maslov dequantization, 73, 188
master space, 69
maximal representation, 125, 143
McKay correspondence, 215
mirror curve, 214
mirror symmetry, 193
Montonen–Olive duality, 171
mother theory, 279
multiplicative group scheme, 126
mutable vertex, 66
mutation, 65, 66, 130

Nelson–Seiberg criterion, 166
non-abelianization, 367
non-calulable DSB model, 166
non-negative unipotent sub-semigroup, 135
non-renormalization theorems, 156
normalized partition function, 217
NS5-brane, 179
NSVZ beta function, 157

O'Raifeartaigh model, 160
open brane, 181
orientifold, 220
orientifold plane, 221

parabolic isometry of $\mathbb{H}$, 84
partial resolution, 211
perfect matching, 210
Pfaffian orientation, 211
physical gauge coupling, 157
Pick theorem, 218
Plücker coordinate, 62
Plücker embedding, 62
Poincaré polynomial, 411
Polonyi model, 159
polyhedral cone, 198
polystable Higgs bundle, 134
positive G-local system, 137
positive atlas, 126
positive chart, 126
positive framed G-local system, 137
positive rational function, 126
positive rational map, 126
positive scheme, 126
positive space, 126
positive triple of flags, 135
positive variety, 125
principal affine bundle, 140
pseudo-line arrangement, 64
pseudo-moduli space of vacua, 161
pseudo-modulus, 161

quantum torus algebra, 131
quiver, 65, 197
quiver of finite mutation type, 67

R–R tadpole, 198
R-symmetry, 153
Raymond–Neveu–Schwarz formalism, 174
real spectrum compactification, 349
real superfield, 154
Regge slope, 175
resolution of toric singularities, 205
RG cascade, 232
ring of universally Laurent polynomials, 127

S-duality, 180
Sasaki–Einstein manifold, 229
Satake correspondence, 357
scalar potential, 156
Schur element, 396
second helicity supertrace, 365
seed, 72, 128
seed cluster transformation, 130
seed pattern, 72
Seiberg duality, 163
Seiberg–Witten curve, 171, 362
Seiberg–Witten differential, 363
semifield, 72, 127
shear coordinates, 94
shift involution, 317
Shilov boundary, 143
skew-symmetrizable matrix, 68, 128
smooth involution of T^2, 221
smooth surface of finite type, 86
smooth torus involution, 317
special curve, 114
spherical Hecke algebra, 356
spider move, 70, 213, 218
split algebraic torus, 126
stable Higgs bundle, 134
standard basis of a Hecke algebra, 376
string compactification, 191
string frame, 176
string revolutions, 173
strings with variable tension, 365
supercharge, 153
superconformal theory, 157
superpotential, 69
superspace, 154
swampland program, 194, 241
symmetric algebra, 356, 396, 407

symmetric rank assignement, 282
symmetrizing trace, 356, 378, 396, 407

T-duality, 179
target space, 173
Teichmüller metric, 87
Teichmüller spaces with holes, 88
theory of class S, 364
Toledo invariant, 143
topological constraints Λ and M, 277
toric duality, 212
toric fan, 199
totally positive $n \times n$ matrix, 64
totally positive Grassmannian, 62
totally positive unipotent semigroup, 135
totally positive variety, 125
train track, 111
trapezoid (toric diagram), 295
triple diagram, 264
tropical semifield, 72
twin SU(5) model, 253
twisted flag configuration, 102

unbounded measured lamination, 117
universally positive Laurent polynomial, 122, 127
untwisting map, 214, 368
urban renewal, 213
UV curve, 171, 364

vector superfield, 154, 168, 171

wall-crossing phenomenon, 365
warped throat, 243
Weil–Peterson metric, 87
Wess–Zumino gauge, 155
Wilson–'t Hooft line, 367
Witten index, 167
WKB foliation, 365
WKB triangulation, 365
world-sheet, 173
worldvolume of a brane, 177

zig-zag path, 210, 264, 368
ZZP value assignement, 276

Bibliography

[AABC+18] Marta Aparicio-Arroyo, Steven Bradlow, Brian Collier, Oscar García-Prada, Peter B Gothen, and André Oliveira. Exotic components of $\mathrm{SO}(p,q)$-surface group representations, and their Higgs bundle avatars. *Comptes Rendus Mathematique*, 356(6):666–673, 2018.

[AABC+19] Marta Aparicio-Arroyo, Steven Bradlow, Brian Collier, Oscar García-Prada, Peter B Gothen, and André Oliveira. $\mathrm{SO}(p,q)$-Higgs bundles and Higher Teichmüller components. *Inventiones mathematicae*, 218(1):197–299, 2019.

[AB83] Michael Atiyah and Raoult Bott. The Yang–Mills equations over Riemann surfaces. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 308(1505):523–615, 1983.

[AB18] Riccardo Argurio and Matteo Bertolini. Orientifolds and duality cascades: confinement before the wall. *JHEP*, 02:149, 2018.

[ABB+08] Riccardo Argurio, Francesco Benini, Matteo Bertolini, Cyril Closset, and Stefano Cremonesi. Gauge/gravity duality and the interplay of various fractional branes. *Phys. Rev. D*, 78:046008, 2008.

[ABF+07] Riccardo Argurio, Matteo Bertolini, Gabriele Ferretti, Alberto Lerda, and Christoffer Petersson. Stringy instantons at orbifold singularities. *JHEP*, 06:067, 2007.

[ABF+21a] Riccardo Argurio, Matteo Bertolini, Sebastián Franco, Eduardo García-Valdecasas, Shani Meynet, Antoine Pasternak, and Valdo Tatitscheff. Dimers, Orientifolds and Anomalies. *JHEP*, 02:153, 2021.

[ABF+21b] Riccardo Argurio, Matteo Bertolini, Sebastián Franco, Eduardo García-Valdecasas, Shani Meynet, Antoine Pasternak, and Valdo Tatitscheff. Dimers, Orientifolds and Stability of Supersymmetry Breaking Vacua. *JHEP*, 01:061, 2021.

[ABF+21c] Riccardo Argurio, Matteo Bertolini, Sebastián Franco, Eduardo García-Valdecasas, Shani Meynet, Antoine Pasternak, and Valdo Tatitscheff. The Octagon and the Non-Supersymmetric String Landscape. *Phys. Lett. B*, 815:136153, 2021.

[ABFK07a] Riccardo Argurio, Matteo Bertolini, Sebastián Franco, and Shamit Kachru. Gauge/gravity duality and meta-stable dynamical supersymmetry breaking. *JHEP*, 01:083, 2007.

[ABFK07b] Riccardo Argurio, Matteo Bertolini, Sebastián Franco, and Shamit Kachru. Meta-stable vacua and D-branes at the conifold. *JHEP*, 06:017, 2007.

[ABFK07c] Riccardo Argurio, Matteo Bertolini, Sebastián Federico Franco, and Shamit Kachru. Metastable supersymmetry breaking and gauge/gravity duality. *Fortsch. Phys.*, 55:644–648, 2007.

[ABJM08] Ofer Aharony, Oren Bergman, Daniel Louis Jafferis, and Juan M. Maldacena. N=6 superconformal Chern-Simons-matter theories, M2-branes and their gravity duals. *JHEP*, 10:091, 2008.

[ABMP19] Riccardo Argurio, Matteo Bertolini, Shani Meynet, and Antoine Pasternak. On supersymmetry breaking vacua from D-branes at orientifold singularities. *JHEP*, 12:145, 2019.

[AC07] Riccardo Argurio and Cyril Closset. A Quiver of Many Runaways. *JHEP*, 09:080, 2007.

[AD95] Philip C. Argyres and Michael R. Douglas. New phenomena in SU(3) supersymmetric gauge theory. *Nucl. Phys. B*, 448:93–126, 1995.

[Adl69] Stephen L. Adler. Axial vector vertex in spinor electrodynamics. *Phys. Rev.*, 177:2426–2438, 1969.

[ADS84] Ian Affleck, Michael Dine, and Nathan Seiberg. Dynamical Supersymmetry Breaking in Chiral Theories. *Phys. Lett. B*, 137:187, 1984.

[ADS85] Ian Affleck, Michael Dine, and Nathan Seiberg. Dynamical Supersymmetry Breaking in Four-Dimensions and Its Phenomenological Implications. *Nucl. Phys. B*, 256:557–599, 1985.

[AFM12] Antonio Amariti, Davide Forcella, and Alberto Mariotti. Integrability on the Master Space. *JHEP*, 06:053, 2012.

[AGM+00] Ofer Aharony, Steven S. Gubser, Juan M. Maldacena, Hirosi Ooguri, and Yaron Oz. Large N field theories, string theory and gravity. *Phys. Rept.*, 323:183–386, 2000.

[AH97] Ofer Aharony and Amihay Hanany. Branes, superpotentials and superconformal fixed points. *Nucl. Phys. B*, 504:239–271, 1997.

[AHBC+16] Nima Arkani-Hamed, Jacob L. Bourjaily, Freddy Cachazo, Alexander B. Goncharov, Alexander Postnikov, and Jaroslav Trnka. *Grassmannian Geometry of Scattering Amplitudes*. Cambridge University Press, 4 2016.

[AHD05] Nima Arkani-Hamed and Savas Dimopoulos. Supersymmetric unification without low energy supersymmetry and signatures for fine-tuning at the LHC. *JHEP*, 06:073, 2005.

[AHK98] Ofer Aharony, Amihay Hanany, and Barak Kol. Webs of (p,q) five-branes, five-dimensional field theories and grid diagrams. *JHEP*, 01:002, 1998.

[AHMNV07] Nima Arkani-Hamed, Lubos Motl, Alberto Nicolis, and Cumrun Vafa. The String landscape, black holes and gravity as the weakest force. *JHEP*, 06:060, 2007.

[AIQU00] Gerardo Aldazabal, Luis E. Ibanez, Fernando Quevedo, and Angel M. Uranga. D-branes at singularities: A Bottom up approach to the string embedding of the standard model. *JHEP*, 08:002, 2000.

[AKMV05] Mina Aganagic, Albrecht Klemm, Marcos Marino, and Cumrun Vafa. The Topological vertex. *Commun. Math. Phys.*, 254:425–478, 2005.

[AL82] Dean Alvis and George Lusztig. The representations and generic degrees of the Hecke algeba of type H4. *Journal für die reine und angewandte Mathematik*, 1982.

[AL01] Paul S. Aspinwall and Albion E. Lawrence. Derived categories and zero-brane stability. *JHEP*, 08:004, 2001.

[Ale08] Daniele Alessandrini. Tropicalization of group representations. *Algebraic & Geometric Topology*, 8(1):279–307, 2008.

[Alt94] Klaus Altmann. The versal deformation of an isolated toric Gorenstein singularity. *arXiv preprint*, 1994. `https://arxiv.org/abs/alg-geom/9403004`.

[AMR20] Andrea Antinucci, Salvo Mancani, and Fabio Riccioni. Infrared duality in unoriented Pseudo del Pezzo. *Phys. Lett. B*, 811:135902, 2020.

[AN07] Andrei V. Alexeevski and Sergey N. Natanzon. Hurwitz numbers for regular coverings of surfaces by seamed surfaces and Cardy-Frobenius algebras of finite groups. *arXiv preprint*, 2007. arXiv:0709.3601.

[And87] Tsuyoshi Ando. Totally positive matrices. *Linear algebra and its applications*, 90:165–219, 1987.

[Arg01] Philip C. Argyres. An Introduction to Global Supersymmetry. `https://homepages.uc.edu/~argyrepc/cu661-gr-SUSY/susy2001.pdf`, 2001.

[Asp04] Paul S. Aspinwall. D-branes on Calabi-Yau manifolds. In *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 2003): Recent Trends in String Theory*, pages 1–152, 3 2004.

[AT16] Ofer Aharony and Yuji Tachikawa. S-folds and 4d N=3 superconformal field theories. *JHEP*, 06:044, 2016.

[Ati88] Michael F. Atiyah. Topological quantum field theory. *Publications Mathématiques de l'IHÉS*, 68:175–186, 1988.

[Bat94] Victor V. Batyrev. Dual polyhedra and mirror symmetry for Calabi-Yau hypersurfaces in toric varieties. *J. Alg. Geom.*, 3:493–545, 1994.

[BBC05] Matteo Bertolini, Francesco Bigazzi, and Aldo L. Cotrone. Supersymmetry breaking at the end of a cascade of Seiberg dualities. *Phys. Rev. D*, 72:061902, 2005.

[BBCC09] Francesco Benini, Matteo Bertolini, Cyril Closset, and Stefano Cremonesi. The N=2 cascade revisited and the enhancon bearings. *Phys. Rev. D*, 79:066012, 2009.

[BBFS09] Mladen Bestvina, Kenneth Bromberg, Koji Fujiwara, and Juan Souto. Shearing coordinates and convexity of length functions on Teichmüller space. *arXiv preprint*, 2009. `https://arxiv.org/abs/0902.0829`.

[BBMR20] Massimo Bianchi, Davide Bufalini, Salvo Mancani, and Fabio Riccioni. Mass deformations of unoriented quiver theories. *JHEP*, 07:015, 2020.

[BBS06] Katrin Becker, Melanie Becker, and John H. Schwarz. *String theory and M-theory: A modern introduction*. Cambridge University Press, 12 2006.

[BCG+20] Antoine Bourget, Santiago Cabrera, Julius F. Grimminger, Amihay Hanany, and Zhenghao Zhong. Brane Webs and Magnetic Quivers for SQCD. *JHEP*, 03:176, 2020.

[BCGP+21] Steve Bradlow, Brian Collier, Oscar Garcia-Prada, Peter Gothen, and André Oliveira. A general Cayley correspondence and higher Teichmüller spaces. *arXiv preprint*, 2021. `https://arxiv.org/abs/2101.09377`.

[BCH+14] Massimo Bianchi, Stefano Cremonesi, Amihay Hanany, Jose Francisco Morales, Daniel Ricci Pacifici, and Rak-Kyeong Seong. Mass-deformed Brane Tilings. *JHEP*, 10:027, 2014.

[BCKW09] Ralph Blumenhagen, Mirjam Cvetic, Shamit Kachru, and Timo Weigand. D-Brane Instantons in Type II Orientifolds. *Ann. Rev. Nucl. Part. Sci.*, 59:269–296, 2009.

[BCV17] T. Daniel Brennan, Federico Carta, and Cumrun Vafa. The String Landscape, the Swampland, and the Missing Corner. *PoS*, TASI2017:015, 2017.

[BD88] Tom Banks and Lance J. Dixon. Constraints on String Vacua with Space-Time Supersymmetry. *Nucl. Phys. B*, 307:93–108, 1988.

[BD02] David Berenstein and Michael R. Douglas. Seiberg duality for quiver gauge theories. *arXiv preprint*, 2002. `https://arxiv.org/abs/hep-th/0207027`.

[BDVF+01] Matteo Bertolini, Paolo Di Vecchia, Marialuisia Frau, Alberto Lerda, Raffaele Marotta, and Igor Pesando. Fractional D-branes and their gauge duals. *JHEP*, 02:014, 2001.

[Bek73] Jacob D. Bekenstein. Black holes and entropy. *Phys. Rev. D*, 7:2333–2346, 1973.

[Bek74] Jacob D. Bekenstein. Generalized second law of thermodynamics in black hole physics. *Phys. Rev. D*, 9:3292–3300, 1974.

[Ber] Matteo Bertolini. On branes, quivers and geometry. Private notes.

[Ber03] Matteo Bertolini. Four lectures on the gauge/gravity correspondence. *Int. J. Mod. Phys. A*, 18:5647–5712, 2003.

[Ber15] Matteo Bertolini. Lectures on supersymmetry. *Lecture notes given at SISSA*, 2015.

[BF96] Adel Bilal and Frank Ferrari. Curves of marginal stability, and weak and strong coupling BPS spectra in N=2 supersymmetric QCD. *Nucl. Phys. B*, 480:589–622, 1996.

[BF06] Andrea Brini and Davide Forcella. Comments on the non-conformal gauge theories dual to Y**p,q manifolds. *JHEP*, 06:050, 2006.

[BFH+05] Sergio Benvenuti, Sebastián Franco, Amihay Hanany, Dario Martelli, and James Sparks. An Infinite family of superconformal quiver gauge theories with Sasaki-Einstein duals. *JHEP*, 06:064, 2005.

[BFM07] Massimo Bianchi, Francesco Fucito, and Jose F. Morales. D-brane instantons on the T**6 / Z(3) orientifold. *JHEP*, 07:038, 2007.

[BGH+21] Antoine Bourget, Julius F. Grimminger, Amihay Hanany, Marcus Sperling, and Zhenghao Zhong. Branes, Quivers, and the Affine Grassmannian. *arXiv preprint*, 2021. `https://arxiv.org/abs/2102.06190`.

[BGPG07] Steven B. Bradlow, Oscar García-Prada, and Peter B. Gothen. What is... a Higgs bundle. *Notices of the AMS*, 54(8), 2007.

[BGVU19] Ginevra Buratti, Eduardo García-Valdecasas, and Angel M. Uranga. Supersymmetry Breaking Warped Throats and the Weak Gravity Conjecture. *JHEP*, 04:111, 2019.

[BHHP20] Jiakang Bao, Yang-Hui He, Edward Hirst, and Stephen Pietromonaco. Lectures on the Calabi-Yau Landscape. *arXiv preprint*, 2020. `https://arxiv.org/abs/2001.01212`.

[BHK05] Sergio Benvenuti, Amihay Hanany, and Pavlos Kazakopoulos. The Toric phases of the Y**p,q quivers. *JHEP*, 07:021, 2005.

[BHOP05] David Berenstein, Christopher. P. Herzog, Peter Ouyang, and Samuel Pinansky. Supersymmetry breaking from a Calabi-Yau singularity. *JHEP*, 09:084, 2005.

[BIMRP14] Massimo Bianchi, Gianluca Inverso, Jose Francisco Morales, and Daniel Ricci Pacifici. Unoriented Quivers with Flavour. *JHEP*, 01:128, 2014.

[BIPP21] Marc Burger, Alessandra Iozzi, Anne Parreau, and Maria Beatrice Pozzetti. The real spectrum compactification of character varieties: characterizations and applications. *Comptes Rendus. Mathématique*, 359(4):439–463, 2021.

[BIW10] Marc Burger, Alessandra Iozzi, and Anna Wienhard. Surface group representations with maximal Toledo invariant. *Annals of mathematics*, pages 517–566, 2010.

[BJ69] John S. Bell and Roman Jackiw. A PCAC puzzle: $\pi^0 \to \gamma\gamma$ in the σ model. *Nuovo Cim. A*, 60:47–61, 1969.

[BJL02] David Berenstein, Vishnu Jejjala, and Robert G. Leigh. The Standard model on a D-brane. *Phys. Rev. Lett.*, 88:071602, 2002.

[BKR99] Tom Bridgeland, Alastair King, and Miles Reid. Mukai implies mckay. *arXiv preprint*, 1999. `https://arxiv.org/abs/math/9908027v1`.

[BKV98] Michael Bershadsky, Zurab Kakushadze, and Cumrun Vafa. String expansion as large N expansion of gauge theories. *Nucl. Phys. B*, 523:59–72, 1998.

[BM60] Armand Borel and James C Moore. Homology theory for locally compact spaces. *Michigan Mathematical Journal*, 7(2):137–159, 1960.

[BM00] Massimo Bianchi and Jose F. Morales. Anomalies & tadpoles. *JHEP*, 03:030, 2000.

[BP00] Raphael Bousso and Joseph Polchinski. Quantization of four form fluxes and dynamical neutralization of the cosmological constant. *JHEP*, 06:006, 2000.

[BP01] Chris E. Beasley and M. Ronen Plesser. Toric duality is Seiberg duality. *JHEP*, 12:001, 2001.

[BP17] Marc Burger and Maria Beatrice Pozzetti. Maximal representations, non-Archimedean Siegel spaces, and buildings. *Geometry & Topology*, 21(6):3539–3599, 2017.

[BP21] Jonas Beyrer and Beatrice Pozzetti. Positive surface group representations in $\mathsf{PO}(p,q)$. *arXiv preprint*, 2021. `https://arxiv.org/abs/2106.14725`.

[Bro12] Nathan Broomhead. *Dimer models and Calabi-Yau algebras*, volume 1011. American Mathematical Soc., 2012.

[Bru88] Gregory W. Brumfiel. The real spectrum compactification of Teichmüller space. *Contem. Math*, 74:51–75, 1988.

[BS90] Massimo Bianchi and Augusto Sagnotti. On the systematics of open string theories. *Phys. Lett. B*, 247:517–524, 1990.

[BS91] Massimo Bianchi and Augusto Sagnotti. Twist symmetry and open string Wilson lines. *Nucl. Phys. B*, 361:519–538, 1991.

[BS11] Tom Banks and Nathan Seiberg. Symmetries and Strings in Field Theory and Gravity. *Phys. Rev. D*, 83:084019, 2011.

[Buc00] Anders S. Buch. The saturation conjecture (after A. Knutson and T. Tao), with an appendix by William Fulton. *ENSEIGNEMENT MATHEMATIQUE*, 46(1/2):43–60, 2000.

[Bur99] Yuri M. Burman. Triangulations of surfaces with boundary and the homotopy principle for functions without critical points. *Annals of Global Analysis and Geometry*, 17(3):221–238, 1999.

[Bus87] Thomas H. Buscher. A Symmetry of the String Background Field Equations. *Phys. Lett. B*, 194:59–62, 1987.

[Bus88] Thomas H. Buscher. Path Integral Derivation of Quantum Duality in Nonlinear Sigma Models. *Phys. Lett. B*, 201:466–472, 1988.

[But06] Agostino Butti. Deformations of Toric Singularities and Fractional Branes. *JHEP*, 10:080, 2006.

[BZ92] Arkady D. Berenstein and Andrei V. Zelevinsky. Triple Multiplicities for $sl(r+1)$ and the Spectrum of the Exterior Algebra of the Adjoint Representation. *Journal of Algebraic Combinatorics*, 1(1):7–22, 1992.

[BZ05] Agostino Butti and Alberto Zaffaroni. R-charges from toric diagrams and the equivalence of a-maximization and Z-minimization. *JHEP*, 11:019, 2005.

[BZ06] Agostino Butti and Alberto Zaffaroni. From toric geometry to quiver gauge theory: The Equivalence of a-maximization and Z-minimization. *Fortsch. Phys.*, 54:309–316, 2006.

[Car85] Roger W. Carter. Finite groups of Lie type: Conjugacy classes and complex characters. *Pure Appl. Math.*, 44, 1985.

[CdlO90] Philip Candelas and Xenia C. de la Ossa. Comments on Conifolds. *Nucl. Phys. B*, 342:246–268, 1990.

[CG97] Neil Chriss and Victor Ginzburg. *Representation theory and complex geometry*, volume 42. Springer, 1997.

[Che55] Claude Chevalley. Sur certains groupes simples. *Tohoku Mathematical Journal, Second Series*, 7(1-2):14–66, 1955.

[Chl09] Maria Chlouveraki. *Blocks and families for cyclotomic Hecke algebras*. Springer, 2009.

[Chl16] Maria Chlouveraki. Hecke algebras, generalisations and representation theory. `https://hal.archives-ouvertes.fr/tel-01411063/document`, 2016. Habilitation thesis.

[CHS19] Santiago Cabrera, Amihay Hanany, and Marcus Sperling. Magnetic quivers, Higgs branches, and 6d N=(1,0) theories. *JHEP*, 06:071, 2019. [Erratum: JHEP 07, 137 (2019)].

[CJ12] Maria Chlouveraki and Nicolas Jacon. Schur elements for the Ariki–Koike algebra and applications. *Journal of Algebraic Combinatorics*, 35(2):291–311, 2012.

[Cle07] Jean-Louis Clerc. An invariant for triples in the Shilov boundary of a bounded symmetric domain. *Communications in Analysis and Geometry*, 15(1):147–174, 2007.

[Clo09] Cyril Closset. Toric geometry and local Calabi-Yau varieties: An Introduction to toric geometry (for physicists). *arXiv preprint*, 2009. `https://arxiv.org/abs/0901.3695`.

[CLPP05] Mirjam Cvetic, Hong Lu, Don N. Page, and Christopher N. Pope. New Einstein-Sasaki spaces in five and higher dimensions. *Phys. Rev. Lett.*, 95:071101, 2005.

[CLPP09] Mirjam Cvetic, Hong Lu, Don N. Page, and Christopher N. Pope. New Einstein-Sasaki and Einstein spaces from Kerr-de Sitter. *JHEP*, 07:082, 2009.

[CLS90] Philip Candelas, Monika Lynker, and Rolf Schimmrigk. Calabi-Yau Manifolds in Weighted P(4). *Nucl. Phys. B*, 341:383–402, 1990.

[CLS11] David A. Cox, John B. Little, and Henry K. Schenck. *Toric varieties*, volume 124. American Mathematical Soc., 2011.

[CM67] Sidney R. Coleman and Jeffrey Mandula. All Possible Symmetries of the S Matrix. *Phys. Rev.*, 159:1251–1256, 1967.

[Col80] Sidney R. Coleman. 1/N. In *17th International School of Subnuclear Physics: Pointlike Structures Inside and Outside Hadrons*, 3 1980.

[Col85] Sidney Coleman. *Aspects of Symmetry: Selected Erice Lectures*. Cambridge University Press, Cambridge, U.K., 1985.

[Cor88] Kevin Corlette. Flat g-bundles with canonical metrics. *Journal of differential geometry*, 28(3):361–382, 1988.

[Cor93] Kevin Corlette. Non abelian Hodge theory. In *Proc. Symp. Pure Math*, volume 54 (2), pages 125–140, 1993.

[Cox93] David A. Cox. The Homogeneous coordinate ring of a toric variety, revised version. *arXiv preprint*, 1993. `https://arxiv.org/abs/alg-geom/9210008v2`.

[Cry73] Colin W. Cryer. The LU-factorization of totally positive matrices. *Linear Algebra and its Applications*, 7(1):83–92, 1973.

[Cry76] Colin W. Cryer. Some properties of totally positive matrices. *Linear Algebra and Its Applications*, 15(1):1–25, 1976.

[Cur88] Charles W. Curtis. Representations of Hecke algebras. *Astérisque*, 168:13–60, 1988.

[Dab97] Atish Dabholkar. Lectures on orientifolds and duality. In *ICTP Summer School in High-Energy Physics and Cosmology*, pages 128–191, 6 1997.

[DDG98] Duiliu-Emanuel Diaconescu, Michael R. Douglas, and Jaume Gomis. Fractional branes and wrapped branes. *JHEP*, 02:013, 1998.

[Del06] Pierre Deligne. *Équations différentielles à points singuliers réguliers*, volume 163. Springer, 2006.

[DGM97] Michael R. Douglas, Brian R. Greene, and David R. Morrison. Orbifold resolution by D-branes. *Nucl. Phys. B*, 506:84–106, 1997.

[DHMT09] John Davey, Amihay Hanany, Noppadol Mekareeya, and Giuseppe Torri. Brane Tilings, M2-branes and Chern-Simons Theories. In *49th Cracow School of Theoretical Physics: Non-perturbative Gravity and Quantum Chromodynamics*, 10 2009.

[DHVW85] Lance J. Dixon, Jeffrey A. Harvey, Cumrun Vafa, and Edward Witten. Strings on Orbifolds. *Nucl. Phys. B*, 261:678–686, 1985.

[DHVW86] Lance J. Dixon, Jeffrey A. Harvey, Cumrun Vafa, and Edward Witten. Strings on Orbifolds. 2. *Nucl. Phys. B*, 274:285–314, 1986.

[Dia01] Duiliu-Emanuel Diaconescu. Enhanced D-brane Categories from String Field Theory. *JHEP*, 06:016, 2001.

[DLP89] Jin Dai, Robert G. Leigh, and Joseph Polchinski. New Connections Between String Theories. *Mod. Phys. Lett. A*, 4:2073–2083, 1989.

[DM96] Michael R. Douglas and Gregory W. Moore. D-branes, quivers, and ALE instantons. *arXiv preprint*, 1996. `https://arxiv.org/abs/hep-th/9603167`.

[DMO09] Nadav Drukker, David R. Morrison, and Takuya Okuda. Loop operators and S-duality from curves on Riemann surfaces. *JHEP*, 09:031, 2009.

[DOS05] Michael Dine, Dugan O'Neil, and Zheng Sun. Branches of the landscape. *JHEP*, 07:014, 2005.

[Dou00] Michael R. Douglas. Topics in D geometry. *Class. Quant. Grav.*, 17:1057–1070, 2000.

[Dou01] Michael R. Douglas. D-branes, categories and N=1 supersymmetry. *J. Math. Phys.*, 42:2818–2843, 2001.

[Dou03] Michael R. Douglas. The Statistics of string / M theory vacua. *JHEP*, 05:046, 2003.

[Dou20] Douglas, Daniel C. and Sun, Zhe. Tropical Fock-Goncharov coordinates for SL_3-webs on surfaces I: construction. *arXiv preprint arXiv:2011.01768*, 2020.

[DP96] Atish Dabholkar and Jaemo Park. Strings on orientifolds. *Nucl. Phys. B*, 477:701–714, 1996.

[DS20] Daniel C. Douglas and Zhe Sun. Tropical Fock-Goncharov coordinates for SL_3-webs on surfaces II: naturality. *arXiv preprint arXiv:2012.14202*, 2020.

[Duf68] Richard James Duffin. Potential theory on a rhombic lattice. *Journal of Combinatorial Theory*, 5(3):258–272, 1968.

[Dug19] Daniel Dugger. Involutions on surfaces. *Journal of Homotopy and Related Structures*, 14(4):919–992, 2019.

[DVR18] Ulf H. Danielsson and Thomas Van Riet. What if string theory has no de Sitter vacua? *Int. J. Mod. Phys. D*, 27(12):1830007, 2018.

[DWZ08] Harm Derksen, Jerzy Weyman, and Andrei Zelevinsky. Quivers with potentials and their representations I: Mutations. *Selecta Mathematica*, 14(1):59–119, 2008.

[EFS12] Richard Eager, Sebastián Franco, and Kevin Schaeffer. Dimer Models and Integrable Systems. *JHEP*, 06:106, 2012.

[EGK+97a] S. Elitzur, A. Giveon, D. Kutasov, E. Rabinovici, and A. Schwimmer. Brane dynamics and N=1 supersymmetric gauge theory. *Nucl. Phys. B*, 505:202–250, 1997.

[EGK97b] Shmuel Elitzur, Amit Giveon, and David Kutasov. Branes and N=1 duality in string theory. *Phys. Lett. B*, 400:269–274, 1997.

[EJL98] Eduardo Eyras, Bert Janssen, and Yolanda Lozano. Five-branes, K-K monopoles and T-duality. *Nucl. Phys. B*, 531:275–301, 1998.

[EJS97] Nick J. Evans, Clifford V. Johnson, and Alfred D. Shapere. Orientifolds, branes, and duality of 4-D gauge theories. *Nucl. Phys. B*, 505:251–271, 1997.

[Éla92] Alexander G Élashvili. Invariant algebras. *Advances in Soviet Math*, 8:57–64, 1992.

[ELNS00] Isabel P. Ennes, Carlos Lozano, Stephen G. Naculich, and Howard J. Schnitzer. Elliptic models, type IIB orientifolds and the AdS / CFT correspondence. *Nucl. Phys. B*, 591:195–226, 2000.

[EW16] Ben Elias and Geordie Williamson. Soergel calculus. *Representation Theory of the American Mathematical Society*, 20(12):295–374, 2016. arXiv:math/1309.0865.

[FB96] Frank Ferrari and Adel Bilal. The Strong coupling spectrum of the Seiberg-Witten theory. *Nucl. Phys. B*, 469:387–402, 1996.

[FFHH03] Bo Feng, Sebastián Franco, Amihay Hanany, and Yang-Hui He. UnHiggsing the del Pezzo. *JHEP*, 08:058, 2003.

[FG06] Vladimir V. Fock and Alexander B. Goncharov. Moduli spaces of local systems and higher Teichmüller theory. *Publications Mathématiques de l'IHÉS*, 103:1–211, 2006. ArXiv:math/0311149.

[FG07] Vladimir V. Fock and Alexander B. Goncharov. Dual Teichmüller and lamination spaces. *Handbook of Teichmüller theory*, 1(11):647–684, 2007.

[FG09] Vladimir V. Fock and Alexander B. Goncharov. Cluster ensembles, quantization and the dilogarithm. *Annales scientifiques de l'École Normale Supérieure*, Ser. 4, 42(6):865–930, 2009. ArXiv:math/0311245.

[FGH12] Sebastián Franco, Daniele Galloni, and Yang-Hui He. Towards the Continuous Limit of Cluster Integrable Systems. *JHEP*, 09:020, 2012.

[FGRU15] Sebastián Franco, Daniele Galloni, Ander Retolaza, and Angel Uranga. On axion monodromy inflation in warped throats. *JHEP*, 02:086, 2015.

[FH13] William Fulton and Joe Harris. *Representation theory: a first course*, volume 129. Springer Science & Business Media, 2013.

[FHH01a] Bo Feng, Amihay Hanany, and Yang-Hui He. D-brane gauge theories from toric singularities and toric duality. *Nucl. Phys. B*, 595:165–200, 2001.

[FHH01b] Bo Feng, Amihay Hanany, and Yang-Hui He. Phase structure of D-brane gauge theories and toric duality. *JHEP*, 08:040, 2001.

[FHHU01] Bo Feng, Amihay Hanany, Yang-Hui He, and Angel M. Uranga. Toric duality as Seiberg duality and brane diamonds. *JHEP*, 12:035, 2001.

[FHK+06] Sebastián Franco, Amihay Hanany, Kristian D. Kennaway, David Vegh, and Brian Wecht. Brane dimers and quiver gauge theories. *JHEP*, 01:096, 2006.

[FHK+07] Sebastián Franco, Amihay Hanany, Daniel Krefl, Jaemo Park, Angel M. Uranga, and David Vegh. Dimers and orientifolds. *JHEP*, 09:075, 2007.

[FHKU01] Bo Feng, Yang-Hui He, Andreas Karch, and Angel M. Uranga. Orientifold dual for stuck NS5-branes. *JHEP*, 06:065, 2001.

[FHKV08a] Bo Feng, Yang-Hui He, Kristian D. Kennaway, and Cumrun Vafa. Dimer models from mirror symmetry and quivering amoebae. *Adv. Theor. Math. Phys.*, 12(3):489–545, 2008.

[FHKV08b] Bo Feng, Yang-Hui He, Kristian D Kennaway, and Cumrun Vafa. Dimer models from mirror symmetry and quivering amoebae. *Advances in Theoretical and Mathematical Physics*, 12(3):489–545, 2008.

[FHM+06] Sebastián Franco, Amihay Hanany, Dario Martelli, James Sparks, David Vegh, and Brian Wecht. Gauge theories from toric geometry and brane tilings. *JHEP*, 01:128, 2006.

[FHMn16] Sebastián Franco, Yasuyuki Hatsuda, and Marcos Mariño. Exact quantization conditions for cluster integrable systems. *J. Stat. Mech.*, 1606(6):063107, 2016.

[FHSU06] Sebastián Franco, Amihay Hanany, Fouad Saad, and Angel M. Uranga. Fractional branes and dynamical supersymmetry breaking. *JHEP*, 01:011, 2006.

[FHSX17] Sebastián Franco, Yang-Hui He, Chuang Sun, and Yan Xiao. A Comprehensive Survey of Brane Tilings. *Int. J. Mod. Phys. A*, 32(23n24):1750142, 2017.

[FHU05] Sebastián Franco, Amihay Hanany, and Angel M. Uranga. Multi-flux warped throats and cascading gauge theories. *JHEP*, 09:028, 2005.

[FHV+06] Sebastián Franco, Amihay Hanany, David Vegh, Brian Wecht, and Kristian D Kennaway. Brane dimers and quiver gauge theories. *Journal of High Energy Physics*, 2006(01):096, 2006.

[FKK13] Bruce Fontaine, Joel Kamnitzer, and Greg Kuperberg. Buildings, spiders, and geometric Satake. *Compositio Mathematica*, 149(11):1871–1912, 2013.

[FM11] Benson Farb and Dan Margalit. *A primer on mapping class groups*. Princeton university press, 2011.

[FM16a] Vladimir V. Fock and Andrey Marshakov. Loop groups, clusters, dimers and integrable systems. In *Geometry and quantization of moduli spaces*, pages 1–65. Springer, 2016.

[FM16b] Vladimir V. Fock and Andrey Marshakov. *Loop Groups, Clusters, Dimers and Integrable Systems*, pages 1–65. Springer International Publishing, 2016.

[FNP+81] Willy Fischler, Hans Peter Nilles, Joseph Polchinski, Stuart Raby, and Leonard Susskind. Vanishing Renormalization of the D Term in Supersymmetric U(1) Theories. *Phys. Rev. Lett.*, 47:757, 1981.

[Foc97] Vladimir V. Fock. Dual Teichmüller spaces. *arXiv preprint dg-ga/9702018*, 1997.

[Foc15] Vladimir V. Fock. Inverse spectral problem for GK integrable system, 2015. arXiv:1503.00289.

[Fom10] Sergey Fomin. Total positivity and cluster algebras. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*, pages 125–145. World Scientific, 2010.

[Fra11] Sebastián Franco. Dimer Models, Integrable Systems and Quantum Teichmuller Space. *JHEP*, 09:057, 2011.

[FRU15] Sebastián Franco, Ander Retolaza, and Angel Uranga. D-brane Instantons as Gauge Instantons in Orientifolds of Chiral Quiver Theories. *JHEP*, 11:165, 2015.

[FST12] Anna Felikson, Michel Shapiro, and Pavel Tumarkin. Skew-symmetric cluster algebras of finite mutation type. *J. Eur. Math. Soc.*, 14:1135–1180, 2012. ArXiv:math/0811.1703.

[FT21] Vladimir Fock and Alexander Thomas. Higher complex structures. *International Mathematics Research Notices*, 2021(20):15873–15893, 2021.

[FTT21] Vladimir Fock, Valdo Tatitscheff, and Alexander Thomas. Topological quantum field theories from Hecke algebras. *arXiv preprint*, 5 2021. arXiv:2105.09622.

[FU06] Sebastián Franco and Angel M. Uranga. Dynamical SUSY breaking at meta-stable minima from D-branes at obstructed geometries. *JHEP*, 06:031, 2006.

[Fuj84] Kazuo Fujikawa. On the Evaluation of Chiral Anomaly in Gauge Theories with Gamma(5) Couplings. *Phys. Rev. D*, 29:285, 1984.

[Ful98] William Fulton. Eigenvalues of sums of hermitian matrices. *Séminaire Bourbaki*, 40:255–269, 1998.

[Ful16] William Fulton. *Introduction to Toric Varieties.(AM-131), Volume 131*. Princeton university press, 2016.

[FV06] Sebastián Franco and David Vegh. Moduli spaces of gauge theories from dimer models: Proof of the correspondence. *JHEP*, 11:054, 2006.

[FWZ06] Sergey Fomin, Lauren Williams, and Andrey Zelevinsky. Introduction to Cluster Algebras. Chapter 1-3. *arXiv preprint*, 2006. `https://arxiv.org/abs/math/0602259`.

[FZ99a] Sergey Fomin and Andrei Zelevinsky. Double Bruhat cells and total positivity. *Journal of the American Mathematical Society*, 12(2):335–380, 1999.

[FZ99b] Sergey Fomin and Andrei Zelevinsky. Total positivity: tests and parametrizations. *arXiv preprint math/9912128*, 1999.

[FZ02] Sergey Fomin and Andrei Zelevinsky. Cluster algebras I: Foundations. *J. Amer. Math. Soc.*, 15:497–529, 2002. ArXiv:math/0104151.

[FZ03] Sergey Fomin and Andrei Zelevinsky. Cluster algebras II: Finite type classification. *Invent. math.*, 154:63–121, 2003. ArXiv:math/0208229.

[FZ05] Sergey Fomin and Andrei Zelevinsky. Cluster algebras III: Upper bounds and double Bruhat cells. *Duke Math. J.*, 126:1–52, 2005. ArXiv:math/0305434.

[FZ07] Sergey Fomin and Andrei Zelevinsky. Cluster algebras IV: Coefficients. *Compositio Mathematica*, 143:112–164, 2007. ArXiv:math/0602259.

[Gai12] Davide Gaiotto. N=2 dualities. *JHEP*, 08:034, 2012.

[GEH15] Iñaki García-Etxebarria and Ben Heidenreich. Strongly coupled phases of $\mathcal{N} = 1$ S-duality. *JHEP*, 09:032, 2015.

[GEH17] Iñaki García-Etxebarria and Ben Heidenreich. S-duality in $\mathcal{N} = 1$ orientifold SCFTs. *Fortsch. Phys.*, 65(3-4):1700013, 2017.

[GER16] Inaki García-Etxebarria and Diego Regalado. $\mathcal{N} = 3$ four dimensional field theories. *JHEP*, 03:083, 2016.

[GESU06] Inaki Garcia-Etxebarria, Fouad Saad, and Angel M. Uranga. Quiver gauge theories at resolved and deformed singularities using dimers. *JHEP*, 06:055, 2006.

[GHK15] Mark Gross, Paul Hacking, and Sean Keel. Birational geometry of cluster algebras. *Algebraic geometry*, 2(2):137–175, 2015.

[GHKK18] Mark Gross, Paul Hacking, Sean Keel, and Maxim Kontsevich. Canonical bases for cluster algebras. *Journal of the American Mathematical Society*, 31(2):497–608, 2018.

[GHL+96] Meinolf Geck, Gerhard Hiss, Frank Lübeck, Gunter Malle, and Götz Pfeiffer. CHEVIE-A system for computing and processing generic character tables. *Applicable Algebra in Engineering, Communication and Computing*, 7(3):175–210, 1996.

[Gin95] Victor Ginzburg. Perverse sheaves on a loop group and Langlands' duality. *arXiv preprint*, 1995. `https://arxiv.org/abs/alg-geom/9511007`.

[GK02] Félix R. Gantmacher and Mark G. Krein. *Oscillation matrices and kernels and small vibrations of mechanical systems: revised edition*. American Mathematical Society, 2002.

[GK11] A. B. Goncharov and R. Kenyon. Dimers and cluster integrable systems. *arXiv preprint*, 2011. `https://arxiv.org/abs/1107.5588`.

[GK13] Alexander B Goncharov and Richard Kenyon. Dimers and cluster integrable systems. In *Annales scientifiques de l'École Normale Supérieure*, volume 46, pages 747–813, 2013.

[GKMW01] Jerome P. Gauntlett, Nakwoo Kim, Dario Martelli, and Daniel Waldram. Wrapped five-branes and N=2 superYang-Mills theory. *Phys. Rev. D*, 64:106008, 2001.

[GKP98] Steven S. Gubser, Igor R. Klebanov, and Alexander M. Polyakov. Gauge theory correlators from noncritical string theory. *Phys. Lett. B*, 428:105–114, 1998.

[GKP02] Steven B. Giddings, Shamit Kachru, and Joseph Polchinski. Hierarchies from fluxes in string compactifications. *Phys. Rev. D*, 66:106006, 2002.

[GKT97] Steven S. Gubser, Igor R. Klebanov, and Arkady A. Tseytlin. String theory and classical absorption by three-branes. *Nucl. Phys. B*, 499:217–240, 1997.

[GLW21] Olivier Guichard, François Labourie, and Anna Wienhard. Positivity and representations of surface groups. *arXiv preprint*, 2021. `https://arxiv.org/abs/2106.14584`.

[GMN10] Davide Gaiotto, Gregory W. Moore, and Andrew Neitzke. Four-dimensional wall-crossing via three-dimensional field theory. *Commun. Math. Phys.*, 299:163–224, 2010.

[GMN12] Davide Gaiotto, Gregory W. Moore, and Andrew Neitzke. Wall-Crossing in Coupled 2d-4d Systems. *JHEP*, 12:082, 2012.

[GMN13a] Davide Gaiotto, Gregory W. Moore, and Andrew Neitzke. Framed BPS States. *Adv. Theor. Math. Phys.*, 17(2):241–397, 2013.

[GMN13b] Davide Gaiotto, Gregory W. Moore, and Andrew Neitzke. Spectral networks. *Annales Henri Poincare*, 14:1643–1731, 2013.

[GMN13c] Davide Gaiotto, Gregory W. Moore, and Andrew Neitzke. Wall-crossing, hitchin systems, and the wkb approximation. *Advances in Mathematics*, 234:239–403, 2013.

[GMN14] Davide Gaiotto, Gregory W. Moore, and Andrew Neitzke. Spectral Networks and Snakes. *Annales Henri Poincare*, 15:61–141, 2014.

[GMSW04a] Jerome P. Gauntlett, Dario Martelli, James Sparks, and Daniel Waldram. Sasaki-Einstein metrics on S**2 x S**3. *Adv. Theor. Math. Phys.*, 8(4):711–734, 2004.

[GMSW04b] Jerome P. Gauntlett, Dario Martelli, James Sparks, and Daniel Waldram. Supersymmetric AdS(5) solutions of M theory. *Class. Quant. Grav.*, 21:4335–4366, 2004.

[GMSW04c] Jerome P. Gauntlett, Dario Martelli, James F. Sparks, and Daniel Waldram. A New infinite class of Sasaki-Einstein manifolds. *Adv. Theor. Math. Phys.*, 8(6):987–1000, 2004.

[Gor94] Vladimir V. Gorbatsevich. *Lie groups and Lie algebras III: Structure of Lie groups and Lie algebras*, volume 41. Springer Science & Business Media, 1994.

[GP96] Eric G. Gimon and Joseph Polchinski. Consistency Conditions for Orientifolds and D-Manifolds. *Phys. Rev. D*, 54:1667–1676, 1996.

[GP00] Meinolf Geck and Götz Pfeiffer. *Characters of finite Coxeter groups and Iwahori-Hecke algebras*. Number 21 in London Math. Soc. Monographs, New Series. Oxford University Press, 2000.

[Gro84] Alexandre Grothendieck. *Esquisse d'un programme*. Chez l'auteur, 1984.

[Gro98] Benedict H. Gross. On the Satake isomorphism. *London Mathematical Society Lecture Note Series*, pages 223–238, 1998.

[GRR15] Alberto García-Raboso and Steven Rayan. Introduction to non-abelian Hodge theory. In *Calabi-Yau varieties: arithmetic, geometry and physics*, pages 131–171. Springer, 2015.

[GS] Daniel R. Grayson and Michael E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at `http://www.math.uiuc.edu/Macaulay2/`.

[GS84] Michael B. Green and John H. Schwarz. Anomaly Cancellation in Supersymmetric D=10 Gauge Theory and Superstring Theory. *Phys. Lett. B*, 149:117–122, 1984.

[GS93] Branko Grünbaum and Geoffrey C. Shephard. Pick's theorem. *The American Mathematical Monthly*, 100(2):150–161, 1993.

[GS15] Alexander Goncharov and Linhui Shen. Geometry of canonical bases and mirror symmetry. *Inventiones mathematicae*, 202(2):487–633, 2015.

[GSO77] Ferdinando Gliozzi, Joel Scherk, and David I. Olive. Supersymmetry, Supergravity Theories and the Dual Spinor Model. *Nucl. Phys. B*, 122:253–290, 1977.

[GSR79] Marcus T. Grisaru, W. Siegel, and M. Rocek. Improved Methods for Supergraphs. *Nucl. Phys. B*, 159:429, 1979.

[GU89] Akihiko Gyoja and Katsuhiro Uno. On the semisimplicity of Hecke algebras. *Journal of the Mathematical Society of Japan*, 41(1):75–79, 1989. Project-Euclid:04110075.

[Gui08] Olivier Guichard. Composantes de Hitchin et représentations hyperconvexes de groupes de surface. *Journal of Differential Geometry*, 80(3):391–431, 2008.

[Gul08] Daniel R. Gulotta. Properly ordered dimers, R-charges, and an efficient inverse algorithm. *JHEP*, 10:014, 2008.

[Gun16] Sam Gunningham. Spin Hurwitz numbers and topological quantum field theory. *Geometry & Topology*, 20(4):1859–1907, 2016.

[GVMPT21] Eduardo García-Valdecasas, Shani Meynet, Antoine Pasternak, and Valdo Tatitscheff. Dimers in a Bottle. *JHEP*, 04:274, 2021.

[GVTU17] Eduardo García-Valdecasas Tenreiro and Angel Uranga. Backreacting D-brane instantons on branes at singularities. *JHEP*, 08:061, 2017.

[GW18] Olivier Guichard and Anna Wienhard. Positivity and higher Teichmüller theory. *arXiv preprint*, 2018. `https://arxiv.org/abs/1802.02833`.

[GZ86] Israel M. Gelfand and Andrei Zelevinsky. Canonical basis in irreducible representations of gl3 and its applications. *Group theoretical methods in physics*, 2:127–146, 1986.

[Har05] Jeffrey A. Harvey. TASI 2003 lectures on anomalies. *arXiv preprint*, 2005. `https://arxiv.org/abs/hep-th/0509097`.

[Har16] Daniel Harlow. Jerusalem Lectures on Black Holes and Quantum Information. *Rev. Mod. Phys.*, 88:015002, 2016.

[Har18] Daniel Harlow. TASI Lectures on the Emergence of Bulk Physics in AdS/CFT. *PoS*, TASI2017:002, 2018.

[Hat88] Allen E Hatcher. Measured lamination spaces for surfaces, from the topological viewpoint. *Topology and its Applications*, 30(1):63–88, 1988.

[Hat91] Allen Hatcher. On triangulations of surfaces. *Topology and its Applications*, 40(2):189–194, 1991.

[Haw75] Stephen W. Hawking. Particle Creation by Black Holes. *Commun. Math. Phys.*, 43:199–220, 1975. [Erratum: Commun.Math.Phys. 46, 206 (1976)].

[Haw76] Stephen W. Hawking. Breakdown of Predictability in Gravitational Collapse. *Phys. Rev. D*, 14:2460–2473, 1976.

[Hem02] Raymond Hemmecke. On the computation of hilbert bases of cones. In *Mathematical software*, pages 307–317. World Scientific, 2002.

[Hil90] David Hilbert. Über die Theorie der algebraischen Formen. *Mathematische annalen*, 36(4):473–534, 1890.

[Hit87] Nigel J. Hitchin. The self-duality equations on a Riemann surface. *Proceedings of the London Mathematical Society*, 3(1):59–126, 1987.

[Hit92] Nigel J. Hitchin. Lie groups and Teichmüller space. *Topology*, 31(3):449–473, 1992.

[Hit20] Nigel J. Hitchin. Michael Atiyah: Geometry and Physics. *Literature and History of Mathematical Science, CMSA Series in Mathematics, International Press, Boston*, 2020.

[HIV00] Kentaro Hori, Amer Iqbal, and Cumrun Vafa. D-branes and mirror symmetry. *arXiv preprint*, 2000. `https://arxiv.org/abs/hep-th/0005247`.

[HK05] Amihay Hanany and Kristian D. Kennaway. Dimer models and toric diagrams. *arXiv preprint*, 2005. `https://arxiv.org/abs/hep-th/0503149`.

[HLS75] Rudolf Haag, Jan T. Lopuszanski, and Martin Sohnius. All Possible Generators of Supersymmetries of the S Matrix. *Nucl. Phys. B*, 88:257, 1975.

[HM16] Yasuyuki Hatsuda and Marcos Marino. Exact quantization conditions for the relativistic Toda lattice. *JHEP*, 05:133, 2016.

[HN16] Lotte Hollands and Andrew Neitzke. Spectral Networks and Fenchel–Nielsen Coordinates. *Lett. Math. Phys.*, 106(6):811–877, 2016.

[Hor62] Alfred Horn. Eigenvalues of sums of hermitian matrices. *Pacific Journal of Mathematics*, 12(1):225–241, 1962.

[Hor89] Petr Horava. Strings on World Sheet Orbifolds. *Nucl. Phys. B*, 327:461–484, 1989.

[HU98] Amihay Hanany and Angel M. Uranga. Brane boxes and branes on singularities. *JHEP*, 05:013, 1998.

[Hua20] Pengfei Huang. Non-Abelian Hodge theory and related topics. *SIGMA*, 16, 2020.

[Hur91] Adolf Hurwitz. Über die Nullstellen der hypergeometrischen Reihe. *Mathematische Annalen*, 38(3):452–458, 1891.

[HV00] Kentaro Hori and Cumrun Vafa. Mirror symmetry. *arXiv preprint*, 2000. `https://arxiv.org/abs/hep-th/0002222v3`.

[HV07] Amihay Hanany and David Vegh. Quivers, tilings, branes and rhombi. *JHEP*, 10:029, 2007.

[HVZ09] Amihay Hanany, David Vegh, and Alberto Zaffaroni. Brane Tilings and M2 Branes. *JHEP*, 03:012, 2009.

[HW97a] Amihay Hanany and Edward Witten. Type IIB superstrings, BPS monopoles, and three-dimensional gauge dynamics. *Nucl. Phys. B*, 492:152–190, 1997.

[HW97b] Martin Henk and Robert Weismantel. The height of minimal Hilbert bases. *Results in Mathematics*, 32:298–303, 1997.

[HZ08] Amihay Hanany and Alberto Zaffaroni. Tilings, Chern-Simons Theories and M2 Branes. *JHEP*, 10:111, 2008.

[IKY08] Yosuke Imamura, Keisuke Kimura, and Masahito Yamazaki. Anomalies and O-plane charges in orientifolded brane tilings. *JHEP*, 03:058, 2008.

[IM65] Nagayoshi Iwahori and Hideya Matsumoto. On some Bruhat decomposition and the structure of the Hecke rings of p-adic Chevalley groups. *Publications Mathématiques de l'IHÉS*, 25:5–48, 1965.

[IRU99] Luis E. Ibanez, Raúl Rabadán, and Angel M. Uranga. Anomalous U(1)'s in type I and type IIB D = 4, N=1 string vacua. *Nucl. Phys. B*, 542:112–138, 1999.

[IS06] Kenneth A. Intriligator and Nathan Seiberg. The Runaway quiver. *JHEP*, 02:031, 2006.

[IS07] Kenneth A. Intriligator and Nathan Seiberg. Lectures on Supersymmetry Breaking. *Class. Quant. Grav.*, 24:S741–S772, 2007.

[IT12] Yoichi Imayoshi and Masahiko Taniguchi. *An introduction to Teichmüller spaces*. Springer Science & Business Media, 2012.

[IU07] Luis E. Ibanez and Angel M. Uranga. Neutrino Majorana Masses from String Theory Instanton Effects. *JHEP*, 03:052, 2007.

[IU08] Luis E. Ibanez and Angel M. Uranga. Instanton induced open string superpotentials and branes at singularities. *JHEP*, 02:103, 2008.

[IU10] Akira Ishii and Kazushi Ueda. A note on consistency conditions on dimer models, 2010.

[IU12] Luis E. Ibanez and Angel M. Uranga. *String theory and particle physics: An introduction to string phenomenology*. Cambridge University Press, 2 2012.

[Ive12] Birger Iversen. *Cohomology of sheaves*. Springer Science & Business Media, 2012.

[IVNO08] Amer Iqbal, Cumrun Vafa, Nikita Nekrasov, and Andrei Okounkov. Quantum foam and topological strings. *Journal of High Energy Physics*, 2008(04):011, 2008.

[IW03] Kenneth A. Intriligator and Brian Wecht. The Exact superconformal R symmetry maximizes a. *Nucl. Phys. B*, 667:183–200, 2003.

[Iwa64] Nagayoshi Iwahori. On the structure of the Hecke ring of a Chevalley group over a finite field. *J. Fac. Sci. Univ. Tokyo Sec. 1A*, 10:215–236, 1964.

[IY02] Shin'ichi Imai and Takashi Yokono. Comments on orientifold projection in the conifold and SOxUSp duality cascade. *Phys. Rev. D*, 65:066007, 2002.

[JPP00] Clifford V. Johnson, Amanda W. Peet, and Joseph Polchinski. Gauge theory and the excision of repulson singularities. *Phys. Rev. D*, 61:086001, 2000.

[JQ93] Arthur Jaffe and Frank Quinn. "Theoretical mathematics": toward a cultural synthesis of mathematics and theoretical physics. *Bulletin of the American Mathematical Society*, 29(1):1–13, 1993.

[JW16] Gareth A. Jones and Jürgen Wolfart. *Dessins d'enfants on Riemann surfaces*. Springer, 2016.

[Kac77] Victor G. Kac. Lie superalgebras. *Advances in mathematics*, 26(1):8–96, 1977.

[Kac90] Victor G. Kac. *Infinite-dimensional Lie algebras*. Cambridge university press, 1990.

[Kak98] Zurab Kakushadze. Gauge theories from orientifolds and large N limit. *Nucl. Phys. B*, 529:157–179, 1998.

[Ken03] Richard Kenyon. An introduction to the dimer model. *arXiv preprint*, 2003. `https://arxiv.org/abs/math/0310326`.

[Ken07] Kristian D. Kennaway. Brane Tilings. *Int. J. Mod. Phys. A*, 22:2977–3038, 2007.

[Kid17] Omar Kidwai. *Spectral networks, abelianization, and opers*. PhD thesis, Oxford U., Inst. Math., 2017.

[KKLT03] Shamit Kachru, Renata Kallosh, Andrei D. Linde, and Sandip P. Trivedi. De Sitter vacua in string theory. *Phys. Rev. D*, 68:046005, 2003.

[KL79] David Kazhdan and George Lusztig. Representations of Coxeter groups and Hecke algebras. *Inventiones mathematicae*, 53(2):165–184, 1979.

[Kle97] Igor R. Klebanov. World volume approach to absorption by nondilatonic branes. *Nucl. Phys. B*, 496:231–242, 1997.

[KLM+96] Albrecht Klemm, Wolfgang Lerche, Peter Mayr, Cumrun Vafa, and Nicholas P. Warner. Selfdual strings and N=2 supersymmetric field theory. *Nucl. Phys. B*, 477:746–766, 1996.

[Kly98] Alexander A. Klyachko. Stable bundles, representation theory and hermitian operators. *Selecta Mathematica, New Series*, 4(3):419–445, 1998.

[KN79] George Kempf and Linda Ness. The length of vectors in representation spaces. In *Algebraic geometry*, pages 233–243. Springer, 1979.

[KN00] Igor R. Klebanov and Nikita A. Nekrasov. Gravity duals of fractional branes and logarithmic RG flow. *Nucl. Phys. B*, 574:263–274, 2000.

[Kon95] Maxim Kontsevich. Homological algebra of mirror symmetry. In *Proceedings of the International Congress of Mathematicians*, pages 120–139. Birkhäuser Basel, 1995.

[KPV02] Shamit Kachru, John Pearson, and Herman L. Verlinde. Brane / flux annihilation and the string dual of a nonsupersymmetric field theory. *JHEP*, 06:021, 2002.

[KS98] Shamit Kachru and Eva Silverstein. 4-D conformal theories and strings on orbifolds. *Phys. Rev. Lett.*, 80:4855–4858, 1998.

[KS00a] Igor R. Klebanov and Matthew J. Strassler. Supergravity and a confining gauge theory: Duality cascades and χ-SB resolution of naked singularities. *JHEP*, 08:052, 2000.

[KS00b] Maximilian Kreuzer and Harald Skarke. Complete classification of reflexive polyhedra in four-dimensions. *Adv. Theor. Math. Phys.*, 4:1209–1230, 2000.

[KT99] Allen Knutson and Terence Tao. The honeycomb model of $\mathrm{GL}_n(\mathbb{C})$ tensor products I: Proof of the saturation conjecture. *Journal of the American Mathematical Society*, 12(4):1055–1090, 1999.

[KT00] Igor R. Klebanov and Arkady A. Tseytlin. Gravity duals of supersymmetric SU(N) x SU(N+M) gauge theories. *Nucl. Phys. B*, 578:123–138, 2000.

[KT01] Allen Knutson and Terence Tao. Honeycombs and sums of hermitian matrices. *Notices Amer. Math. Soc*, 48(2), 2001.

[Kup96] Greg Kuperberg. Spiders for rank-2 Lie algebras. *Communications in mathematical physics*, 180(1):109–151, 1996.

[KW98] Igor R. Klebanov and Edward Witten. Superconformal field theory on three-branes at a Calabi-Yau singularity. *Nucl. Phys. B*, 536:199–218, 1998.

[KW07] Anton Kapustin and Edward Witten. Electric-Magnetic Duality And The Geometric Langlands Program. *Commun. Num. Theor. Phys.*, 1:1–236, 2007.

[Lab04] François Labourie. Anosov flows, surface groups and curves in projective space. *arXiv preprint math/0401230*, 2004.

[Laz01] Calin I. Lazaroiu. Generalized complexes and string field theory. *JHEP*, 06:052, 2001.

[Le16] Ian Le. Higher laminations and affine buildings. *Geometry & Topology*, 20(3):1673–1735, 2016. arXiv:math/1209.0812.

[Li93] Jun Li. The space of surface group representations. *Manuscripta Mathematica*, 78(1):223–243, 1993.

[Lib19] Nicolas Libedinsky. Gentle introduction to Soergel bimodules I: the basics. *Sao Paulo Journal of Mathematical Sciences*, 13(2):499–538, 2019. arXiv:math/1702.00039.

[LNV98] Albion E. Lawrence, Nikita Nekrasov, and Cumrun Vafa. On conformal field theories in four-dimensions. *Nucl. Phys. B*, 533:199–209, 1998.

[Loe55] Charles Loewner. On totally positive matrices. *Mathematische Zeitschrift*, 63(1):338–340, 1955.

[LP08] Aaron D. Lauda and Hendryk Pfeiffer. Open–closed strings: Two-dimensional extended TQFTs and Frobenius algebras. *Topology and its Applications*, 155(7):623–666, 2008.

[LR34] Dudley E. Littlewood and Archibald R. Richardson. Group characters and algebra. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 233(721-730):99–141, 1934.

[LR99] Robert G. Leigh and Moshe Rozali. Brane boxes, anomalies, bending and tadpoles. *Phys. Rev. D*, 59:026004, 1999.

[LT96] Markus A. Luty and Washington Taylor. Varieties of vacua in classical supersymmetric gauge theories. *Phys. Rev. D*, 53:3399–3405, 1996.

[Lus90] George Lusztig. Canonical bases arising from quantized enveloping algebras. *Journal of the American Mathematical Society*, 3(2):447–498, 1990.

[Lus94] George Lusztig. Total positivity in reductive groups. In *Lie theory and geometry*, pages 531–568. Springer, 1994.

[Lus10] George Lusztig. *Introduction to quantum groups*. Springer Science & Business Media, 2010.

[Lut05] Markus A. Luty. 2004 TASI lectures on supersymmetry breaking. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics in $D \geqq 4$*, pages 495–582, 9 2005.

[LV98] Naichung C. Leung and Cumrun Vafa. Branes and toric geometry. *Adv. Theor. Math. Phys.*, 2:91–118, 1998.

[LZ13] Sergei Lando and Alexander Zvonkin. *Graphs on surfaces and their applications*, volume 141. Springer Science & Business Media, 2013.

[Mal98] Juan M. Maldacena. Wilson loops in large N field theories. *Phys. Rev. Lett.*, 80:4859–4862, 1998.

[Mal99] Juan M. Maldacena. The Large N limit of superconformal field theories and supergravity. *Int. J. Theor. Phys.*, 38:1113–1133, 1999.

[Mar05] Marcos Marino. Chern-Simons theory and topological strings. *Rev. Mod. Phys.*, 77:675–720, 2005.

[Mar14] Robert J Marsh. Lecture notes on cluster algebras. *AMC*, 10:12, 2014.

[Mat64] Hideya Matsumoto. Générateurs et relations des groupes de Weyl généralisés. *Comptes rendus hebdomadaires de l'académie des sciences*, 258(13):3419, 1964.

[McK80] John McKay. Graphs, singularities, and finite groups. In *Proc. Symp. Pure Math*, volume 37, page 183, 1980.

[Mer01] Christian Mercat. Discrete Riemann surfaces and the Ising model. *Communications in Mathematical Physics*, 218(1):177–216, 2001.

[Mic15] Jean Michel. The development version of the CHEVIE package of GAP3. *Journal of Algebra*, 435:308–336, 2015.

[Mik98] A. Mikhailov. BPS states and minimal surfaces. *Nucl. Phys. B*, 533:243–274, 1998.

[MN01] Juan M. Maldacena and Carlos Nunez. Supergravity description of field theories on curved manifolds and a no go theorem. *Int. J. Mod. Phys. A*, 16:822–855, 2001.

[MO77] Claus Montonen and David I. Olive. Magnetic Monopoles as Gauge Particles? *Phys. Lett. B*, 72:117–120, 1977.

[Moo15] Gregory Moore. A few Remarks on Topological Field Theory. `http://www.physics.rutgers.edu/~gmoore/695Fall2015/TopologicalFieldTheory.pdf`, 2015. Accessed: 2022-01-02.

[Mos88] Lee Mosher. Tiling the projective foliation space of a punctured surface. *Transactions of the American Mathematical Society*, pages 1–70, 1988.

[MP99] David R. Morrison and M. Ronen Plesser. Nonspherical horizons. 1. *Adv. Theor. Math. Phys.*, 3:1–81, 1999.

[MS84] John W Morgan and Peter B Shalen. Valuations, trees, and degenerations of hyperbolic structures, i. *Annals of Mathematics*, 120(3):401–476, 1984.

[MS05] Dario Martelli and James Sparks. Toric Sasaki-Einstein metrics on S**2 x S**3. *Phys. Lett. B*, 621:208–212, 2005.

[MS06a] Dario Martelli and James Sparks. Toric geometry, Sasaki-Einstein manifolds and a new infinite class of AdS/CFT duals. *Commun. Math. Phys.*, 262:51–89, 2006.

[MS06b] Gregory W. Moore and Graeme Segal. D-branes and K-theory in 2D topological field theory. *arXiv preprint*, 2006. `https://arxiv.org/pdf/hep-th/0609042.pdf`.

[MSY06] Dario Martelli, James Sparks, and Shing-Tung Yau. The Geometric dual of a-maximisation for Toric Sasaki-Einstein manifolds. *Commun. Math. Phys.*, 268:39–65, 2006.

[Mur95] Hitoshi Murayama. Studying noncalculable models of dynamical supersymmetry breaking. *Phys. Lett. B*, 355:187–192, 1995.

[MV84] Yannick Meurice and Gabriele Veneziano. Susy vacua versus chiral fermions. *Phys. Lett. B*, 141:69–72, 1984.

[MV19] Jacob McNamara and Cumrun Vafa. Cobordism Classes and the Swampland. *arXiv preprint*, 2019. `https://arxiv.org/abs/1909.10355`.

[MV21] Miguel Montero and Cumrun Vafa. Cobordism Conjecture, Anomalies, and the String Lamppost Principle. *JHEP*, 01:063, 2021.

[Nek03] Nikita A. Nekrasov. Seiberg-Witten prepotential from instanton counting. *Adv. Theor. Math. Phys.*, 7(5):831–864, 2003.

[Neu06] Max Neunhöffer. Kazhdan–Lusztig basis, Wedderburn decomposition, and Lusztig's homomorphism for Iwahori–Hecke algebras. *Journal of Algebra*, 303(1):430–446, 2006.

[Nol22] Alexander Nolte. Canonical Maps from Spaces of Higher Complex Structures to Hitchin Components. *arXiv preprint*, 2022. arXiv:2204.04732.

[NS94] Ann E. Nelson and Nathan Seiberg. R symmetry breaking versus supersymmetry breaking. *Nucl. Phys. B*, 416:46–62, 1994.

[NSVZ86] Victor A. Novikov, Mikhail A. Shifman, Arkady I. Vainshtein, and Valentin I. Zakharov. x. *Phys. Lett. B*, 166:329–333, 1986.

[OOSV18] Georges Obied, Hirosi Ooguri, Lev Spodyneiko, and Cumrun Vafa. De Sitter Space and the Swampland. *arXiv preprint*, 2018. `https://arxiv.org/abs/1806.08362v3`.

[OS17] Hirosi Ooguri and Lev Spodyneiko. New Kaluza-Klein instantons and the decay of AdS vacua. *Phys. Rev. D*, 96(2):026016, 2017.

[OV07] Hirosi Ooguri and Cumrun Vafa. On the Geometry of the String Landscape and the Swampland. *Nucl. Phys. B*, 766:21–33, 2007.

[OV17] Hirosi Ooguri and Cumrun Vafa. Non-supersymmetric AdS and the Swampland. *Adv. Theor. Math. Phys.*, 21:1787–1801, 2017.

[Pal13] Frederic Palesi. Introduction to positive representations and Fock-Goncharov coordinates. hal-01218570f, 2013.

[Pal19] Eran Palti. The Swampland: Introduction and Review. *Fortsch. Phys.*, 67(6):1900037, 2019.

[Par06] James Parkinson. Buildings and Hecke algebras. *Journal of Algebra*, 297(1):1–49, 2006.

[Par12] Anne Parreau. Compactification d'espaces de représentations de groupes de type fini. *Mathematische Zeitschrift*, 272(1):51–86, 2012.

[Pen87] Robert C Penner. The decorated Teichmüller space of punctured surfaces. *Communications in Mathematical Physics*, 113(2):299–339, 1987.

[Pen92] Robert C Penner. Weil-Petersson volumes. *Journal of Differential Geometry*, 35(3):559–608, 1992.

[Pet08] Christoffer Petersson. Superpotentials From Stringy Instantons Without Orientifolds. *JHEP*, 05:078, 2008.

[PH92] Robert C Penner and John L Harer. *Combinatorics of train tracks*. Number 125 in Annals of Mathematics Studies. Princeton University Press, 1992.

[Pol98] Alexander M. Polyakov. String theory and quark confinement. *Nucl. Phys. B Proc. Suppl.*, 68:1–8, 1998.

[Pol04] Joseph Polchinski. Monopoles, duality, and string theory. *Int. J. Mod. Phys. A*, 19S1:145–156, 2004.

[Pos06] Alexander Postnikov. Total positivity, Grassmannians, and networks. *arXiv preprint math/0609764*, 2006.

[PRU00a] Jaemo Park, Raúl Rabadán, and Angel M. Uranga. N=1 type IIA brane configurations, chirality and T-duality. *Nucl. Phys. B*, 570:3–37, 2000.

[PRU00b] Jaemo Park, Raúl Rabadán, and Angel M. Uranga. Orientifolding the conifold. *Nucl. Phys. B*, 570:38–80, 2000.

[PRZ01] Michela Petrini, Rodolfo Russo, and Alberto Zaffaroni. N=2 gauge theories and systems with fractional branes. *Nucl. Phys. B*, 608:145–161, 2001.

[PS89] Gianfranco Pradisi and Augusto Sagnotti. Open String Orbifolds. *Phys. Lett. B*, 216:59–67, 1989.

[PT98] Erich Poppitz and Sandip P. Trivedi. Dynamical supersymmetry breaking. *Ann. Rev. Nucl. Part. Sci.*, 48:307–350, 1998.

[PU99] Jaemo Park and Angel M. Uranga. A Note on superconformal N=2 theories and orientifolds. *Nucl. Phys. B*, 542:139–156, 1999.

[Ron09] Mark Ronan. *Lectures on buildings: updated and revised*. University of Chicago Press, 2009.

[RS99a] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.

[RS99b] Lisa Randall and Raman Sundrum. An Alternative to compactification. *Phys. Rev. Lett.*, 83:4690–4693, 1999.

[RU16a] Ander Retolaza and Angel Uranga. De Sitter Uplift with Dynamical Susy Breaking. *JHEP*, 04:137, 2016.

[RU16b] Ander Retolaza and Angel Uranga. Orientifolds of Warped Throats from Toric Calabi-Yau Singularities. *JHEP*, 07:135, 2016.

[RUW15] Ander Retolaza, Angel M. Uranga, and Alexander Westphal. Bifid Throats for Axion Monodromy Inflation. *JHEP*, 07:099, 2015.

[RY01] Soo-Jong Rey and Jung-Tay Yee. Macroscopic strings as heavy quarks in large N gauge theory and anti-de Sitter supergravity. *Eur. Phys. J. C*, 22:379–394, 2001.

[S+21] W. A. Stein et al. *Sage Mathematics Software (Version 10.15.7)*. The Sage Development Team, 2021. `http://www.sagemath.org`.

[Sag87] Augusto Sagnotti. Open Strings and their Symmetry Groups. In *NATO Advanced Summer Institute on Nonperturbative Quantum Field Theory (Cargese Summer Institute)*, pages 0521–528, 9 1987.

[Sak12] Serkan Sakar. *Schubert calculus, adjoint representation and moment polytopes*. PhD thesis, bilkent university, 2012.

[Sat63] Ichirô Satake. Theory of spherical functions on reductive algebraic groups over p-adic fields. *Publications Mathématiques de l'IHÉS*, 18:5–69, 1963.

[Sch98] Alexander Schrijver. *Theory of linear and integer programming*. John Wiley & Sons, 1998.

[Sch99] Martin Schmaltz. Duality of nonsupersymmetric large N gauge theories. *Phys. Rev. D*, 59:105018, 1999.

[Sco06] Joshua S. Scott. Grassmannians and cluster algebras. *Proc. London Math. Soc. (3)*, 92(2):345–380, 2006.

[Sei93] Nathan Seiberg. Naturalness versus supersymmetric nonrenormalization theorems. *Phys. Lett. B*, 318:469–475, 1993.

[Sei94] Nathan Seiberg. Exact results on the space of vacua of four-dimensional SUSY gauge theories. *Phys. Rev. D*, 49:6857–6863, 1994.

[Sei95] Nathan Seiberg. Electric-Magnetic duality in supersymmetric nonAbelian gauge theories. *Nucl. Phys. B*, 435:129–146, 1995.

[Shi59] Goro Shimura. Sur les intégrales attachées aux formes automorphes. *Journal of the Mathematical Society of Japan*, 11(4):291–311, 1959.

[Sim90] Carlos T. Simpson. Nonabelian Hodge theory. In *Proceedings of the International Congress of Mathematicians*, volume 1, pages 747–756, 1990.

[Sim92] Carlos T. Simpson. Higgs bundles and local systems. *Publications Mathématiques de l'IHÉS*, 75:5–95, 1992.

[Soe07] Wolfgang Soergel. Kazhdan–Lusztig-Polynome und unzerlegbare Bimoduln über Polynomringen. *Journal of the Institute of Mathematics of Jussieu*, 6(3):501–525, 2007.

[Spa11] James Sparks. Sasaki-Einstein Manifolds. *Surveys Diff. Geom.*, 16:265–324, 2011.

[SS00] Yael Shadmi and Yuri Shirman. Dynamical supersymmetry breaking. *Rev. Mod. Phys.*, 72:25–64, 2000.

[Sta89] Richard P. Stanley. Log-concave and Unimodal Sequences in Algebra, Combinatorics, and Geometry. *Ann. New York Acad. Sci*, 576(1):500–535, 1989.

[Str96] Andrew Strominger. Open p-branes. *Phys. Lett. B*, 383:44–47, 1996.

[Str05] Matthew J. Strassler. The Duality cascade. In *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 2003): Recent Trends in String Theory*, pages 419–510, 5 2005.

[Sus95] Leonard Susskind. The World as a hologram. *J. Math. Phys.*, 36:6377–6396, 1995.

[Sus05] Leonard Susskind. *The cosmic landscape: String theory and the illusion of intelligent design*. Back Bay Books, 2005.

[Sus08] Leonard Susskind. *The black hole war: My battle with Stephen Hawking to make the world safe for quantum mechanics*. Hachette UK, 2008.

[SW94a] Nathan Seiberg and Edward Witten. Electric-Magnetic duality, monopole condensation, and confinement in N=2 supersymmetric Yang-Mills theory. *Nucl. Phys. B*, 426:19–52, 1994. [Erratum: Nucl.Phys.B 430, 485–486 (1994)].

[SW94b] Nathan Seiberg and Edward Witten. Monopoles, duality and chiral symmetry breaking in N=2 supersymmetric QCD. *Nucl. Phys. B*, 431:484–550, 1994.

[SYZ96] Andrew Strominger, Shing-Tung Yau, and Eric Zaslow. Mirror symmetry is T-duality. *Nucl. Phys. B*, 479:243–259, 1996.

[Tac13] Yuji Tachikawa. *N=2 supersymmetric dynamics for pedestrians.* Lecture Notes in Physics. Springer Charm, 12 2013.

[Tat21] Valdo Tatitscheff. Inverse algorithm and triple point diagrams. In *Nankai Symposium on Mathematical Dialogues: In celebration of S.S.Chern's 110th anniversary*, 11 2021.

[tH74] Gerard 't Hooft. A Planar Diagram Theory for Strong Interactions. *Nucl. Phys. B*, 72:461, 1974.

[tH80] Gerard 't Hooft. Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking. *NATO Sci. Ser. B*, 59:135–157, 1980.

[tH93] Gerard 't Hooft. Dimensional reduction in quantum gravity. *Conf. Proc. C*, 930308:284–296, 1993.

[Tho21] Alexander Thomas. Differential operators on surfaces and rational WKB method. *arXiv preprint*, 2021. arXiv:2111.07946.

[Tho22] Alexander Thomas. Generalized punctual Hilbert schemes and $\mathfrak{g}$-complex structures. *International Journal of Mathematics*, 33(01):2250004, 2022.

[Thu88] William P. Thurston. On the geometry and dynamics of diffeomorphisms of surfaces. *Bulletin of the American mathematical society*, 19(2):417–431, 1988.

[Thu98] William P. Thurston. Minimal stretch maps between hyperbolic surfaces. *arXiv preprint math/9801039*, 1998.

[Thu02] William P. Thurston. The Geometry and Topology of Three-Manifolds. Notes `http://library.msri.org/books/gt3m/`, 2002.

[Thu17] Dylan P Thurston. From dominoes to hexagons. In *Proceedings of the 2014 Maui and 2015 Qinhuangdao Conferences in Honour of Vaughan FR Jones' 60th Birthday*, pages 399–414. Australian National University, Mathematical Sciences Institute, 2017.

[TM79] William P. Thurston and John W. Milnor. The geometry and topology of three-manifolds, 1979.

[Ura00] Angel M. Uranga. A New orientifold of C**2 / Z(N) and six-dimensional RG fixed points. *Nucl. Phys. B*, 577:73–87, 2000.

[Ura01] Angel M. Uranga. D-brane probes, RR tadpole cancellation and K theory charge. *Nucl. Phys. B*, 598:225–246, 2001.

[Vaf05] Cumrun Vafa. The String landscape and the swampland. *arXiv preprint*, 2005. `https://arxiv.org/abs/hep-th/0509212`.

[Ver61] Jean-Louis Verdier. Sur les intégrales attachées aux forms automorphes. *Séminaire Bourbaki*, 6:149–175, 1961.

[VW07] Herman Verlinde and Martijn Wijnholt. Building the standard model on a D3-brane. *JHEP*, 01:106, 2007.

[Wei87] Steven Weinberg. Anthropic Bound on the Cosmological Constant. *Phys. Rev. Lett.*, 59:2607, 1987.

[Wei13] Steven Weinberg. *The quantum theory of fields. Vol. 3: Supersymmetry.* Cambridge University Press, 6 2013.

[Wen16] Richard Wentworth. Higgs bundles and local systems on Riemann surfaces. In *Geometry and quantization of moduli spaces*, pages 165–219. Springer, 2016.

[Wey12] Hermann Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, 1912.

[Whi52] Anne M. Whitney. A reduction theorem for totally positive matrices. *Journal d'Analyse Mathématique*, 2(1):88–92, 1952.

[Wil16] Geordie Williamson. The Hodge theory of the Hecke category. *arXiv preprint arXiv:1610.06246*, 2016. arXiv:math/1610.06246.

[Wit81] Edward Witten. Dynamical Breaking of Supersymmetry. *Nucl. Phys. B*, 188:513, 1981.

[Wit82] Edward Witten. Constraints on Supersymmetry Breaking. *Nucl. Phys. B*, 202:253, 1982.

[Wit93] Edward Witten. Phases of N=2 theories in two-dimensions. *Nucl. Phys. B*, 403:159–222, 1993.

[Wit95] Edward Witten. String theory dynamics in various dimensions. *Nucl. Phys. B*, 443:85–126, 1995.

[Wit97] Edward Witten. Solutions of four-dimensional field theories via M theory. *Nucl. Phys. B*, 500:3–42, 1997.

[Wit98a] Edward Witten. Anti-de Sitter space and holography. *Adv. Theor. Math. Phys.*, 2:253–291, 1998.

[Wit98b] Edward Witten. Toroidal compactification without vector structure. *JHEP*, 02:006, 1998.

[Xie13] Dan Xie. Higher laminations, webs and N=2 line operators. *arXiv preprint*, 2013. `https://arxiv.org/abs/1304.2390`.

[Yam08] Masahito Yamazaki. Brane Tilings and Their Applications. *Fortsch. Phys.*, 56:555–686, 2008.

[Yon74] Tamiaki Yoneya. Connection of Dual Models to Electrodynamics and Gravidynamics. *Prog. Theor. Phys.*, 51:1907–1920, 1974.

[Z$^+$20] Piotr A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

[Zel99] Andrei Zelevinsky. Littlewood–Richardson semigroups. *New perspectives in algebraic combinatorics (Berkeley, CA, 1996–97), Math. Sci. Res. Inst. Publ*, 38:337–345, 1999.

Les algèbres et variétés amassées se manifestent naturellement dans divers champs de la physique mathématique, comme la théorie de Teichmüller de rang supérieur et l'étude des modèles de dimères.

D'une part, on étudie la généralisation des laminations de Thurston aux espaces de Teichmüller de rang supérieur correspondant à des groupes réels déployés. Cela conduit notamment à l'introduction de théories topologiques des champs quantiques liées aux algèbres de Iwahori–Hecke des groupes de Coxeter finis. Celles-là associent un polynôme de Laurent entier à chaque surface épointée de type fini.

D'autre part, les modèles de dimères nous permettent de prouver l'existence d'une complétion ultraviolette stable du modèle SU(5) de brisure dynamique de supersymétrie. On dérive de plus des résultats généraux quant à l'existence d'anomalies de jauge sur des D-branes transverses à des orientifolds de singularités Calabi–Yau affines toriques. Enfin, on donne un sens physique aux modèles de dimères sur la bouteille de Klein.

Par ailleurs, les deux parties introductives de ce manuscrit présentent de manière pédagogique la théorie de Teichmüller de rang supérieur de Fock et Goncharov, puis les modèles de dimères en théorie des cordes ainsi que leur emploi dans l'étude des correspondances holographiques.

Cluster algebras and varieties naturally appear in various fields of mathematical physics, such as higher Teichmüller theory and dimer models – known as brane tilings in the context of string theory.

On the first hand, we study the generalisation of Thurston's laminations to higher Teichmüller spaces in the real split case. This guides us towards introducing topological quantum field theories associated with the Iwahori–Hecke algebras of finite Coxeter groups. Those assign a Laurent polynomial with integer coefficients to each punctured surface of finite type.

On the other hand, we use dimer models to prove the existence of a stable ultraviolet completion of the dynamical supersymmetry breaking SU(5) model. Moreover, we derive general results on the existence of gauge anomalies in the worldvolume of D-branes at orientifolds of affine toric Calabi–Yau singularities. Lastly, we provide a physical interpretation of brane tilings on the Klein bottle.

Besides, the two preliminary parts of this dissertation are pedogogical invitations first to Fock and Goncharov's higher Teichmüller theory and then to the use of dimer models in string theory and in holography.

INSTITUT DE RECHERCHE MATHÉMATIQUE AVANCÉE
UMR 7501
Université de Strasbourg
CNRS
IRMA, UMR 7501
7 rue René Descartes
F-67000 STRASBOURG

Tél. 03 68 85 01 29

irma.math.unistra.fr
irma@math.unistra.fr

IRMA 2022/004
http://tel.archives-ouvertes.fr/tel-03766040