Le moddèle de Potts bidimensionnel

Jesper Lykke Jacobsen

Université de Paris XI — U.F.R. des sciences d'Orsay

Habilitation à diriger des recherches

Spécialité :

Physique Théorique

présentée par

Jesper Lykke JACOBSEN

Laboratoire de Physique Théorique et Modèles Statistiques,
Université de Paris-Sud, Bâtiment 100, 91405 Orsay Cedex

Sujet :

Le modèle de Potts bidimensionnel

Soutenue le 29 novembre 2002 devant le jury composé de

Edouard BREZIN	rapporteur
Henk HILHORST	président
Giuseppe MUSSARDO	rapporteur
Bernard NIENHUIS	rapporteur
Jean-Bernard ZUBER	examinateur

Table des matières

Avant-propos

Ce mémoire fait le point de mes travaux sur le modèle de Potts et le modèle O(n). J'étudie ces modèles en deux dimensions, où des méthodes performantes sont disponibles : théorie conforme, intégrabilité, matrices aléatoires, matrices de transfert. Nous verrons que, en deux dimensions, le modèle de Potts et le modèle O(n) sont étroitement liés. C'est pourquoi l'intitulé du mémoire fait seulement référence à l'un d'entre eux.

La rédaction m'a mené à relire un nombre de références clefs. Cette lecture m'a parfois guidé vers d'autres travaux intéressants que je ne connaissais pas auparavant. Je me suis alors rendu compte qu'il existe encore de nombreux thèmes à approfondir dans ce domaine. J'espère que ma future recherche saura en profiter.

Je souhaite ici remercier les membres du jury qui ont bien voulu participer à la soutenance de cette thèse. Mes pensées vont aussi vers mes collaborateurs et vers les collègues avec lesquels j'ai eu des conversations fructueuses au cours des années. Enfin, un grand merci à Gisèle Bannier qui a apporté une aide précieuse en relisant les textes en français.

Publications

Un certain nombre de publications sont attachées à chaque chapitre de ce mémoire. Leurs références d'origine sont les suivantes :

Chapitre 2 :
- J. L. Jacobsen, *Duality relations for M coupled Potts models*, cond-mat/0001367.
  Physical Review E **62**, R1–4 (2000).

Chapitre 3 :
- J. L. Jacobsen, J. Salas et A. D. Sokal, *Transfer matrices and partition-function zeros for antiferromagnetic Potts models. III. Triangular-lattice chromatic polynomial*, cond-mat/0204587.
  Journal of Statistical Physics (soumis).

Chapitre 4 :
- S. C. Chang, J. L. Jacobsen, J. Salas et R. Shrock, *Exact Potts model partition functions for strips of the triangular lattice*, cond-mat/0211623.
  Journal of Statistical Physics (soumis).

Chapitre 5 :
- J. L. Jacobsen et J. Kondev, *Field theory of compact polymers on the square lattice*, cond-mat/9804048.
  Nuclear Physics B **532** [FS], 635–688 (1998).
- J. L. Jacobsen et J. Kondev, *Conformal field theory of the Flory model of protein melting*, cond-mat/0209247.
  Physical Review E (soumis).
- Vl. S. Dotsenko, J. L. Jacobsen et M. Picco, *Classification of conformal field theories based on Coulomb gases. Application to loop models*, hep-th/0105287.
  Nuclear Physics B **618**, 523–550 (2001).

Chapitre 6 :
- P. Di Francesco, E. Guitter et J. L. Jacobsen, *Exact meander asymptotics : a numerical check*, cond-mat/0003008.
  Nuclear Physics B **580** [FS], 757–795 (2000).

- J. L. Jacobsen et P. Zinn-Justin, *A transfer matrix approach to the enumeration of colored links*, math-ph/0104009.
  Journal of Knot Theory and its Ramifications **10**, 1233–1267 (2001).
- J. L. Jacobsen, N. Read et H. Saleur, *Dense loops, supersymmetry, and Goldstone phases in two dimensions*, cond-mat/0205033.
  Physical Review Letters (soumis).

Chapitre 7 :

- J. L. Cardy et J. L. Jacobsen, *Critical behavior of random bond Potts models*, cond-mat/9705038.
  Physical Review Letters **79**, 4063–4066 (1997).
- J. L. Jacobsen et J. L. Cardy, *Critical behaviour of random bond Potts models : A transfer matrix study*, cond-mat/9711279.
  Nuclear Physics B **515** [FS], 701–742 (1998).
- J. L. Jacobsen et M. Picco, *Large-q asymptotics of the random bond Potts model*, cond-mat/9910071.
  Physical Review E **61**, R13–16 (2000).
- J. L. Jacobsen, *Multiscaling of energy correlations in the random bond Potts model*, cond-mat/9912304.
  Physical Review E **61**, R6060–6062 (2000).

Chapitre 8 :

- Vl. S. Dotsenko, J. L. Jacobsen, M.-A. Lewis et M. Picco, *Coupled Potts models : Self-duality and fixed point structure*, cond-mat/9812227.
  Nuclear Physics B **546** [FS], 505–557 (1999).
- Vl. S. Dotsenko, J. L. Jacobsen, X. S. Nguyen et R. Santachiara, *Universality of coupled Potts models*, cond-mat/0112120.
  Nuclear Physics B **631**, 426–446 (2002).

Chapitre 1

Introduction

1.1 Le modèle de Potts

Le modèle de Potts est sans doute parmi les modèles les plus étudiés dans la littérature de la physique théorique. Nous devons son introduction à C. Domb, qui, au début des années 1950, proposa à son thésard R. B. Potts d'étudier une généralisation du célèbre modèle d'Ising où chaque spin peut être dans un nombre arbitraire d'états discrets. Pourtant, le modèle suggéré par Domb est maintenant connu comme le modèle Z_N, ou bien le modèle d'horloge, alors que le modèle qui nous intéresse ici ne fut mentionné que dans une petite remarque vers la fin de l'article de Potts [142].

Sa formulation précise est simple : Soit $G = (V, E)$ un graphe avec des sommets V et des arêtes E, et associons à chaque $v \in V$ une variable discrète (dite spin) pouvant prendre q valeurs distinctes, $\sigma_v = 1, 2, \ldots, q$. Le hamiltonien du modèle de Potts à q états sur G vaut alors

$$-\beta\mathcal{H} = K \sum_{e_{ij} \in E} \delta(\sigma_i, \sigma_j), \tag{1.1}$$

où $\beta = 1/kT$ est la température inverse, K la constante de couplage et e_{ij} l'arête qui relie les sommets i et j. La fonction delta de Kronecker $\delta(x, y)$ vaut 1 si $x = y$, sinon elle vaut 0. Le but devient alors de calculer la fonction de partition

$$Z = \sum_{\sigma_i = 1}^{q} \exp(-\beta\mathcal{H}), \tag{1.2}$$

et plus généralement de trouver les propriétés thermodynamiques du modèle comme fonction de K et de q.

L'intérêt principal du modèle de Potts est qu'il démontre (au moins dans le domaine ferromagnétique, $K > 0$) une compétition entre l'énergie et l'entropie, avec une phase ordonnée (ferromagnétique) à basse température et une phase désordonnée (paramagnétique) à haute température. On s'attend alors à ce qu'il y ait, pour un choix de G approprié avec $|V| \to \infty$, une ou plusieurs transitions de phase comme fonction des paramètres (q, K). Ensuite on s'interroge sur la localisation et la nature exacte de ces transitions ainsi que sur leurs dépendances de G.

Le modèle de base permet plusieurs variantes intéressantes. Typiquement, G est pris comme un réseau régulier en dimension d, mais, on peut également considérer G comme un arbre ou moyenner sur un ensemble de réseaux aléatoires. Le couplage K peut être complété soit par d'autres interactions entre plusieurs spins, soit par des champs externes. On peut aussi considérer les couplages comme des variables aléatoires sur lesquelles il faut moyenner au niveau des quantités thermodynamiques.

1.2 Bref aperçu historique

Il existe une littérature très riche sur le modèle de Potts, notamment après l'explosion d'activité de recherche qu'il a engendrée à partir de 1970. Parmi les techniques très utilisées pendant les années 70, citons la transformation de dualité, le développement en séries, le groupe de renormalisation, les simulations numériques, et avant tout l'intégrabilité sous la forme de l'ansatz de Bethe [163, 13]. Ces approches ont permis la construction des diagrammes de phase, la localisation exacte des certains points critiques,

l'évaluation exacte de l'énergie libre en certains cas particuliers, ainsi que l'estimation d'exposants critiques.

Comme très souvent dans la physique statistique, le cas de deux dimensions ($d = 2$) s'est avéré un champ de travail particulièrement fructueux. D'autres méthodes spécialisées à ce cas, à savoir les approches d'ondes de spin et de gaz de Coulomb, vinrent s'ajouter à celles évoquées précédemment autour de 1980. Ce développement a notamment permis de calculer de manière exacte (mais non rigoureuse) les valeurs de plusieurs exposants critiques en fonction de q.

Tous ces résultats furent résumés en 1982 dans l'excellente revue de Wu [163], et la même année apparut la monographie de Baxter [13] qui demeure la référence idéale sur l'approche d'intégrabilité à ce jour. La technique du gaz de Coulomb fut très clairement exposée par Nienhuis [136]. Le lecteur cherchant un aperçu du bilan au début des années 80, ainsi que des références précises aux travaux d'origine peut se tourner vers ces trois œuvres.

Le but de ce mémoire est d'exposer quelques développements postérieurs dans le domaine du modèle de Potts et de certains autres modèles, tels que le modèle O(n), qui sont étroitement liés à celui-ci. Sans une précision explicite du contraire, nous allons désormais considérer uniquement le cas bidimensionnel ($d = 2$). La présentation prend comme point de départ quelques aspects du sujet général qui rentrent dans les intérêts de recherche de l'auteur et auxquels il a parfois contribué. En effet, à l'heure actuelle, l'activité qui se déroule autour du modèle de Potts a pris une telle ampleur[1] qu'une revue exhaustive sur le sujet ne serait probablement envisageable que sous forme d'une collaboration entre plusieurs auteurs.

1.3 La théorie conforme de champs

Une large partie de cette recherche a été déclenchée par l'apparition de la théorie conforme de champs qui fit son entrée dans la littérature physique en 1984 [17]. Cette technique est basée sur une observation faite par Polyakov en 1970 [141]. Rappellons que l'hypothèse qui est au cœur de l'approche du groupe de renormalisation est l'invariance sous un changement d'échelle. Cette approche a été appliquée avec beaucoup de succès aux systèmes ayant une transition de phase de second ordre. Polyakov argumente donc que si, en plus, les interactions sont de courte portée, cette invariance *globale* sous changement d'échelle peut être promue en une symétrie *locale*.

En $d = 2$, les conséquences de cette observation sont considérables, puisque les générateurs infinitésimaux L_n d'une telle symétrie sont en nombre infini. L'élaboration mathématique [17] montre alors l'existence d'une infinité de contraintes sur la théorie de champs dont la manifestation algébrique est l'algèbre de Virasoro

$$[L_n, L_m] = (n-m)L_{n+m} + \frac{c}{12} n(n^2-1)\delta(n+m,0). \tag{1.3}$$

La constante c, dite la charge centrale, joue ici le rôle d'un paramètre qui caractérise complétement la théorie de champs sous-jacente.

La construction de la théorie de représentation de cet algèbre permet de fixer la valeur de c et des exposants critiques des opérateurs physiques. Il est naturel d'imposer deux contraintes de nature physique sur les représentations acceptables :

- *unitarité* (la norme d'un opérateur physique est positive) et
- *minimalité* (le nombre d'opérateurs physiques est fini).

Les modèles $\mathcal{M}_p$ satisfaisant ces critères, dits modèles minimaux, seront alors spécifiés par un entier $p = 3, 4, 5, \ldots$. Leur charge centrale

$$c(p) = 1 - \frac{6}{p(p+1)} \tag{1.4}$$

ainsi que les exposants critiques (poids conformes) de leurs opérateurs physiques

$$h_{r,s}(p) = \frac{[(p+1)r - ps]^2 - 1}{4p(p+1)} \qquad (1 \le s < r < p) \tag{1.5}$$

sont connus pour chaque valeur de p [69]. (La formule (1.5) est souvent appelée le tableau de Kac.) En particulier, les cas $p = 3$ et $p = 5$ de ce classement correspondent respectivement aux cas $q = 2$ et $q = 3$

[1] Une recherche sur les archives électroniques `http ://fr.arxiv.org` démontre l'apparition d'environ un article par semaine ayant le mot "Potts" dans son titre.

du modèle de Potts ferromagnétique au point critique. Les cas $p = 4$ et $p = 6$ sont associés aux modèles tricritiques correspondants. Une exposition plus détaillée peut être trouvée dans [62].

1.4 Contenu de ce mémoire

Le reste de ce mémoire va présenter quelques développements plus récents et, notamment, leur rapport avec la théorie conforme.

Nous commençons, dans le Chapitre 2, par l'introduction de quelques transformations utiles sur la fonction de partition (1.2), transformations qui permettront de concevoir q comme un nombre complexe arbitraire, plutôt que comme un nombre entier d'états. Ces transformations sont ensuite utilisées pour établir des transformations de dualité pour une certaine classe de fonctions de corrélation du modèle de Potts et pour la fonction de partition de plusieurs modèles de Potts couplés entre eux.

Dans le Chapitre 3, nous entamons une étude du modèle de Potts dans le régime antiferromagnétique. En particulier, la limite $T \to 0$ permet d'interpréter le modèle comme un problème de coloriage des sommets V. Nous discutons le diagramme de phase de ce problème en fonction de q complexe. Ensuite nous continuons, dans le Chapitre 4, par une discussion du diagramme de phase ainsi que les proprietés critiques associées dans l'espace (q, e^K) pour deux choix de G, les réseaux carré et triangulaire.

Le problème correspondant de coloriage d'arêtes (coloriage de Tait) est exposé dans le Chapitre 5. Une suite de transformations exactes établie le lien avec le modèle O(n) à $T \to 0$ dont nous trouvons de nouvelles phases. Ces phases s'interprètent comme des modèles de polymères compacts qui présentent un intérêt dans l'étude des protéines confinées. A titre d'application concrète, nous résolvons le modèle de Flory pour la fusion d'une protéine.

Nous discutons ensuite, dans le Chapitre 6, des possibilités de définir des modèles de Potts ou des modèles O(n) sur des réseaux aléatoires planaires. Plusieurs de ces possibilités mènent à des modèles combinatoires intéressants, permettant notamment d'énumérer les classes topologiques des intersections entre deux courbes autoévitantes fermées (méandres) et des nœuds alternés.

Dans le Chapitre 7, nous considérons un autre type de désordre, où les couplages du modèle de Potts sont des variables aléatoires gelées. Ce désordre permet, dans certains cas, de modifier la nature de la transition de phase. Par la technique des répliques, ce problème est étroitement lié à celui de modèles couplés que nous étudions dans le Chapitre 8. Le rôle du désordre est de coupler les répliques entre elles par leurs opérateurs de densité d'énergie. Ce type de couplage rend possible un développement perturbatif autour de $q = 2$. Nous confrontons ce développement avec des résultats numériques, et nous discutons les perspectives d'un éventuel traitement analytique exact.

Chapitre 2

Dualité

Depuis l'introduction du modèle de Potts [142], il était clair que sa fonction de partition admet une transformation de dualité. De manière générale, la dualité remplace le graphe (planaire, car $d = 2$) sur lequel est défini le modèle par son dual. Elle est donc particulièrement utile si ces deux graphes coïncident, comme c'est le cas pour le réseau carré. La supposition de l'unicité de la transition de phase permet alors immédiatement de déterminer la température critique. Un aperçu détaillé des résultats obtenus par cette technique peut être trouvé dans [151, 163].

Il est possible de dériver la relation de dualité de plusieurs manières différentes [163], mais, la plus simple passe par une réécriture de (1.2) due à Kasteleyn et Fortuin [103]. Comme cette réécriture va jouer un rôle important dans ce mémoire, nous allons la présenter en détail dans le paragraphe suivant.

Des relations de dualité existent aussi pour des fonctions de corrélation de n spins. Le cas $n = 2$ est un résultat classique qui relie les opérateurs de spin à des opérateurs de désordre. Or, son utilité pratique ne semble avoir été exploitée que plus récemment. En effet, cette relation est un outil performant dans l'étude numérique des propriétés magnétiques du modèle de Potts [49, 51], même avec du désordre [82]. Sur le côté analytique, elle a été utilisée dans [31, 96]. Plus généralement, Wu et collaborateurs ont trouvé des relations de dualité pour des corrélateurs à n spins, mais à condition que tous les spins se situent autour d'une seule face du réseau [164, 81, 165, 125]. Il paraît difficile de lever cette condition [107, 108].

Finalement, la dualité s'applique aussi au cas de N modèles de Potts couplés entre eux à travers leurs opérateurs d'énergie. Le cas $N = 2$ est un résultat ancien [43], obtenu en utilisant la méthode de transformation de Fourier sur réseau [162]. Cette même méthode appliquée au cas $N = 3$ devient très compliquée [49], et doit être abandonnée pour des valeurs de N plus élevées. Cependant, en généralisant la transformation de Kasteleyn-Fortuin au cas de plusieurs modèles, les calculs se simplifient rendant possible la généralisation à N arbitraire [89, 51].

2.1 Modèle d'amas

Nous allons exposer la transformation de Kasteleyn-Fortuin pour la fonction de partition

$$Z = \sum_{\{\sigma\}} \prod_{\langle ij \rangle \subseteq E} \exp\left[K_{ij}\delta(\sigma_i, \sigma_j)\right], \tag{2.1}$$

où nous permettons des constantes de couplage K_{ij} distinctes pour chaque arête de G. Comme $\delta(\sigma_i, \sigma_j)$ ne prend que deux valeurs, nous avons l'identité

$$\exp\left[K_{ij}\delta(\sigma_i, \sigma_j)\right] = 1 + u_{ij}\delta(\sigma_i, \sigma_j) \tag{2.2}$$

avec $u_{ij} = \exp(K_{ij}) - 1$. Dans le développement du produit, nous associons un sous-ensemble E' de E à chaque terme selon la règle suivante : une arête $\langle ij \rangle \subseteq E$ fait partie de E' si, et seulement si, nous avons pris le facteur $u_{ij}\delta(\sigma_i, \sigma_j)$. Après la sommation $\sum_{\{\sigma\}}$ nous trouverons alors

$$Z = \sum_{E' \subseteq E} \left(\prod_{\langle ij \rangle \in E'} u_{ij} \right) q^{C(E')}, \tag{2.3}$$

où $C(E')$ est le nombre de composantes connexes dans le sous-graphe induit par E'.

Dans le cas homogène, $u_{ij} = u$,

$$Z(u,q) = \sum_{E' \subseteq E} u^{|E'|} q^{C(E')} \tag{2.4}$$

est connu comme le *polynôme dichromatique* dans la théorie de graphes (après un changement de variables non linéaire). Dans ce contexte, il fut introduit par Whitney en 1932 [161], alors que le cas particulier $u = -1$ (dit *polynôme chromatique*) entra dans la littérature mathématique déjà en 1912 [25]. Nous allons étudier le polynôme chromatique en détail dans le Chapitre 3.

L'avantage du polynôme dichromatique est qu'il permet, par prolongement analytique, de définir le modèle de Potts pour q non entier. Cette remarque sera cruciale par la suite. La représentation (2.3) de Z, due à Kasteleyn et Fortuin [103], est souvent appelée le *modèle d'amas*.

Il est également possible de définir un modèle d'amas pour le cas où l'interaction spin-spin est complétée par un champ magnétique [27]. L'idée est d'augmenter G par un spin "imaginaire" fixe, $\sigma_0 = 1$, qui sera relié à tout autre spin par une arête supplémentaire. Un alternatif plus adapté à des simulations numériques a été développé dans [82].

Une autre représentation utile [13] de Z, dite *modèle de boucles*, est obtenue de (2.3) en utilisant la relation d'Euler :

$$Z = q^{|V|/2} \sum_{E' \subseteq E} \left(\prod_{\langle ij \rangle \in E'} \frac{u_{ij}}{\sqrt{q}} \right) q^{L(E')/2}. \tag{2.5}$$

Ici, $L(E')$ est le nombre de boucles autoévitantes fermées sur le graphe G_{m} médial à G. Plus précisement, G_{m} est le graphe dont les sommets bisectent les arêtes de G ; deux sommets du graphe médial sont reliés par une arête, si leurs arêtes correspondantes en G sont incidentes à un même sommet et à une même face. Par convention, chaque arête de G_{m} est couverte par une boucle ; à chaque sommet, deux boucles sont tangentes. Enfin, une arête $\langle ij \rangle \in V$ est coupée par deux segments de boucles sur G_{m} si $\langle ij \rangle \notin V'$, et par aucun segment si $\langle ij \rangle \in V'$.

La Figure 2.1 montre l'application de ces concepts à un cas concret.

La représentation en modèle de boucles est capitale pour la solution exacte du modèle de Potts [13], car elle permet de distribuer les poids $\sqrt{q}$ par boucle de manière *locale* [11, 13]. La même construction apparaît dans le développement de haute température du modèle O(n) [134] et ses généralisation plus récentes [7, 112, 83]. En effet, le modèle de boucles est le lien formel entre les modèles de Potts et O(n).

Dans la suite, nous allons montrer que la représentation en modèle d'amas (ou de boucles) est également possible quand plusieurs modèles de Potts sont couplés par leur densité d'énergie [89, 51].

2.2 Dualité de Z

A partir de la représentation (2.3), il est facile de déduire la transformation de dualité de Z. Stipulons d'abord la dualité entre le terme $E' = E$ avec le poids

$$Z_\bullet(\{u_{ij}\}) = q \prod_{\langle ij \rangle \in E} u_{ij} \tag{2.6}$$

et le terme $E' = \emptyset$ avec le poids

$$Z^*_\circ(\{u^*_{ij}\}) = q^{|V|^*}. \tag{2.7}$$

Le nombre de spins sur le réseau dual est fixé par la relation d'Euler, $|E| = |V| + |V^*| - 2$.

Etablir la dualité entre les fonctions de partition complètes, Z et Z^*, équivaut à montrer que tous leurs termes, reliés deux à deux par une règle de dualité, ont le même poids relatif par rapport aux états de référence. Par état de référence, nous entendons $Z_\bullet$ pour le cas de Z, et $Z^*_\circ$ pour le cas de Z^*. La règle de dualité est la suivante :

- une configuration E' contribuant à Z est duale à une configuration E'^* contribuant à Z^*, où E'^* est le complément de E' par rapport à E (modulo la bijection qui relie chaque arête dans E à son arête coupante dans E^*).

En particulier, l'élimination d'une arête de la configuration E (poids relatif $1/u_{ij}$) correspond à l'ajout d'une arête à la configuration duale (poids relatif u^*_{ij}/q). Les couplages u_{ij} et u^*_{ij} doivent alors être reliés par

$$u_{ij}u^*_{ij} = q. \tag{2.8}$$

L'élimination successive de plusieurs arêtes de E peut impliquer la création d'un nouvel amas (composante connexe) portant un facteur supplémentaire de q. Réciproquement, selon la règle de dualité, ceci entraîne la création d'un cycle sur le graphe dual, ce qui donne encore une fois un facteur supplémentaire de q. Comme tout sous-ensemble $E' \subseteq E$ peut être construit par de telles éliminations d'arêtes, nous avons donc prouvé la transformation de dualité [162]

$$Z(\{u_{ij}\}) = qCZ^*(\{u^*_{ij}\}), \tag{2.9}$$

où $C = q^{-|V^*|}\prod_{\langle ij\rangle} u_{ij}$ est une constante.

Pour voir comment la dualité permet de fixer la température critique, considérons à titre d'exemple le cas où G est le réseau carré infini et $u_{ij} = u$. On a alors $G = G^*$ et (2.8) relie le *même* modèle de Potts à deux températures différentes. Si la transition de phase est unique, elle doit donc être située au point autodual, c'est-à-dire à $u = \pm\sqrt{q}$. La branche $u = \sqrt{q}$ correspond effectivement au point critique ferromagnétique, comme le montre la solution exacte de Baxter [13]. Or, cette méthode ne permet pas de déterminer deux autres branches critiques $u = -2\pm\sqrt{4-q}$ qui sont mutuellement duales et correspondent aux points critiques antiferromagnétiques [14]. Nous allons étudier les propriétés antiferromagnétiques du modèle de Potts dans les Chapitres 3 et 4.

Il est également possible d'utiliser (2.8) pour déterminer le point critique du modèle avec des couplages aléatoires ferromagnétiques [109, 82].

2.3 Dualité du corrélateur à deux spins

Afin de discuter la dualité des fonctions de corrélation, définissons d'abord le paramètre d'ordre local

$$M_a(r) = \delta(\sigma_r, a) - \frac{1}{q} \tag{2.10}$$

avec composantes $a = 1, 2, \ldots, q$. Dans la phase de haute température, toutes ses composantes sont nulles. En revanche, dans la phase ordonnée (à basse température), la symétrie Z_q est spontanément brisée et une des composantes, disons $a = 1$, possède une espérance positive.

Un simple calcul montre que le corrélateur spin-spin

$$G_{aa}(r_1, r_2) = \langle M_a(r_1)M_a(r_2)\rangle \tag{2.11}$$

est proportionnel à la probabilité que les deux points $r_1, r_2 \in V$ résident dans le même amas. Soit $S(r_1, r_2)$ un chemin sur G qui relie les sommets r_1 et r_2. Selon la règle de dualité, il y a une correspondance biunivoque entre [les configurations où r_1 et r_2 appartiennent au même amas] et [les configurations duales où aucun cycle sur les amas duaux ne coupe $S(r_1, r_2)$ un nombre impair de fois].

L'interprétation de l'objet $S(r_1, r_2)$, dit coupure (ou *seam* en anglais), est plus simple dans le contexte du modèle de boucles : $S(r_1, r_2)$ n'est coupé par aucune boucle un nombre impair de fois. Encore un autre point de vue est que $G_{aa}(r_1, r_2)$ est proportionnel au rapport Z'/Z, où Z' dessine la fonction de partition calculée en présence d'un opérateur qui décale les états de Potts

$$\sigma \to (\sigma + 1) \bmod q \tag{2.12}$$

le long de chaque arête qui traverse $S(r_1, r_2)$. Ces conditions au bord sont souvent appelées tordues (ou *twisted* en anglais). La ligne de défauts (2.12) peut être vue comme l'insertion d'un opérateur, dit opérateur de désordre, aux points r_1 et r_2. Dans ce sens, la dualité transforme le corrélateur spin-spin en un corrélateur désordre-désordre.

Cette construction est illustrée sur la Figure 2.1. A travers une matrice de transfert appropriée [82], elle permet le calcul numérique de la fonction de partition contrainte Z' dans une géométrie cylindrique.

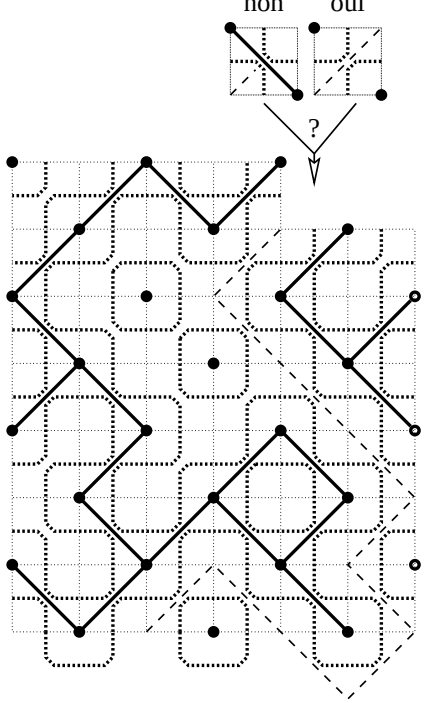


FIG. 2.1 – Illustration des transformations de dualité pour le cas où G est une partie du réseau carré (incliné d'un angle $\pi/4$) avec des conditions aux bords périodiques dans la direction horizontale. Les cercles noirs sont les sommets V, et les lignes noires montrent la représentation d'amas. Les lignes pointillées indiquent la position des boucles correspondantes. Les boucles vivent sur le réseau médial $G_{\rm m}$ et elles sont reflétées à la fois par les amas et par les amas duaux (non montrés). Nous montrons également le calcul de la fonction de partition Z' dans la présence d'une coupure $S(r_1, r_2)$ (ligne brisée).

Comme autre application analogue, nous pouvons mentionner les calculs de gaz de Coulomb avec une charge à l'infini [46, 47, 83]. En effet, le concept de coupure facilite le calcul de l'exposant magnétique.

Il est également possible de pondérer les boucles qui traversent la coupure un nombre impair de fois par une constante non nulle, mais différente de celle de $\sqrt{q}$ utilisée pour les autres boucles. Cette possibilité a été exploitée dans [31] pour des calculs de gaz de Coulomb, tandis que [96] traite le cas supersymétrique.

2.4 Corrélateurs de n spins de bord

Les considérations de la section précédente se généralisent au cas des fonctions de corrélation de n spins, pour peu que tous ces spins soient autour d'une même face du réseau. Par la suite, nous supposons que le plongement de G dans le plan est tel que cette face entoure le point à l'infini.

A titre d'exemple, nous allons reproduire le calcul pour $n = 2$ dans une formulation [164] qui se généralise vers n plus élevé. Le même calcul pour $n = 3$ peut être trouvé dans [164]. Pour $n \geq 4$, la méthode doit être complétée [81] avec des considérations supplémentaires [165] (voir ci-dessous).

2.4.1 Méthode diagrammatique

Soit G un graphe planaire avec $n = 2$ sommets marqués (i et j) sur le bord extérieur ; voir la Figure 2.2. La probabilité que $\{\sigma_i, \sigma_j\}$ soient dans les états $\{\sigma, \sigma'\}$ vaut

$$P_2(\sigma, \sigma') = \langle \delta(\sigma_i, \sigma)\delta(\sigma_j, \sigma') \rangle \,. \tag{2.13}$$

En termes de la fonction de partition restreinte $Z_{\sigma\sigma'} = Z(\sigma_i = \sigma, \sigma_j = \sigma')$, nous avons l'égalité évidente

$$Z = qZ_{11} + q(q-1)Z_{12}. \tag{2.14}$$

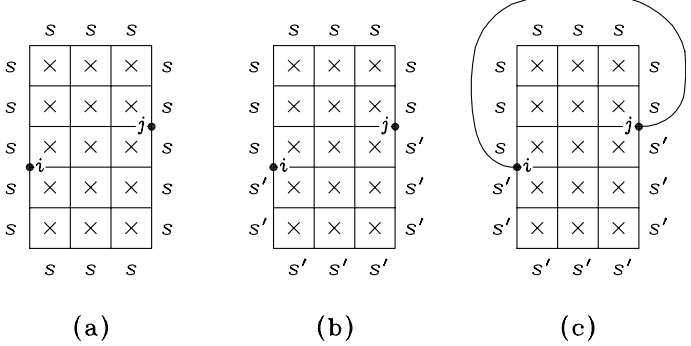


FIG. 2.2 – Calcul du corrélateur de $n = 2$ spins de bord [164]. Les croix dénotent les spins duaux sur les faces intérieures. a) Tous les spins au bord interagissent avec un même spin s sur la face extérieure. b) Deux spins distincts s et s' sur la face extérieure. c) Graphe auxiliaire.

Pour le calcul de la fonction de partition duale, nous imaginons que les spins adjacents à la face extérieure interagissent avec *deux* spins, s et s', localisés sur cette face. Avec une notation analogue à celle introduite ci-dessus, $Z^*_{\sigma\sigma'} = Z^*(s = \sigma, s' = \sigma')$. Pour l'instant, $s = s'$ et nous avons $Z^* = qZ^*_{11}$. L'équation (2.9) produit alors la relation

$$Z_{11} + (q-1)Z_{12} = qCZ^*_{11}. \tag{2.15}$$

Soit maintenant G' le graphe auxiliaire obtenu de G en reliant i et j par une arête supplémentaire de couplage K. Comme cette nouvelle arête sépare s et s', les fonctions de partition du réseau modifié valent

$$\tilde{Z} = q\mathrm{e}^K Z_{11} + q(q-1)Z_{12}, \qquad \tilde{Z}^* = q\mathrm{e}^{K^*} Z^*_{11} + q(q-1)Z^*_{12}. \tag{2.16}$$

La dualité (2.9) donne alors une seconde relation :

$$\mathrm{e}^K Z_{11} + (q-1)Z_{12} = C(\mathrm{e}^K - 1)[\mathrm{e}^{K^*} Z^*_{11} + (q-1)Z^*_{12}]. \tag{2.17}$$

On peut maintenant résoudre (2.15) et (2.17) :

$$Z_{11} = C[Z^*_{11} + (q-1)Z^*_{12}], \qquad Z_{12} = C[Z^*_{11} - Z^*_{12}]. \tag{2.18}$$

Remarquons que ce résultat ne dépend pas de K. Le résultat pour la probabilité P_2 vaut finalement

$$P_2(\sigma, \sigma') = \frac{1}{q^2}\left\{1 + [q\delta(\sigma, \sigma') - 1]\,\frac{Z^*_{12}}{Z^*_{11}}\right\}. \tag{2.19}$$

2.4.2 Règles de sommation

Ce calcul se généralise sans problème au cas $n = 3$, où le nombre de fonctions de partition restreintes vaut cinq : Z_{111}, Z_{211}, Z_{121}, Z_{112} et Z_{123}. Le nombre de graphes auxiliaires vaut également cinq, chaque graphe donnant une équation, et l'on peut déduire les quantités Z_{ijk} [164].

Pourtant, pour $n \geq 4$, le nombre c_n de graphes auxiliaires est inférieur au nombre b_n de $Z_{ijkl\ldots}$ [81]. En effet, les b_n s'interprètent comme le nombre de partitions d'un ensemble de n éléments, tandis que les c_n sont les nombres de Catalan. Ces derniers sont soumis à des contraintes de planarité. Par exemple, pour $n = 4$ la quantité Z_{1212} ne correspond à aucun graphe auxiliaire *planaire*.

Suivant cette remarque, Wu et Huang [165] ont montré que le problème redevient soluble à cause de l'existence de $b_n - c_n$ règles de sommation parmi les quantités $Z_{ijkl\ldots}$.

Considérons, à titre d'exemple, le cas $n = 4$, avec $b_4 = 15$ et $c_4 = 14$. La quantité "non planaire" Z_{1212} s'écrit alors comme une somme de trois termes :

$$Z_{1212} = T_1 + T_2 + T_3, \tag{2.20}$$

où T_1 regroupe les contributions dans lesquelles i et k appartiennent à un même amas. De même, T_2 dénote les contributions où j et l appartiennent à un même amas, et T_3 représente les contributions où tous les quatre spins de bord appartiennent à des amas différents.

Le même type de raisonnement donne les relations $Z_{1213} = T_1 + T_3$, $Z_{2131} = T_2 + T_3$ et $Z_{1234} = T_3$. Ce qui prouve la règle de sommation

$$Z_{1212} = Z_{1213} + Z_{2131} - Z_{1234} \tag{2.21}$$

qui fournit la quinzième équation manquante.

De manière générale, la planarité des amas de Kasteleyn-Fortuin assure l'existence d'une règle de sommation pour chaque $Z_{ijkl\ldots}$ "non planaire". Ces règles sont ensuite déduites du principe d'inclusion-exclusion [25]. Chaque identité ainsi trouvée est distincte à cause de l'unicité du développement en amas.

Le relations de dualité pour les $Z_{ijkl\ldots}$ dans le cas de n général ont une forme remarquablement simple. Cette forme a été conjecturée dans [165] et prouvée dans [125]. Une formulation de ces résultats dans le cadre de la théorie de graphes a été donnée dans [166].

2.4.3 Cas de plusieurs faces

La question qui s'impose maintenant est de savoir si ces résultats peuvent être étendus au cas où les n spins se trouvent autour de plusieurs faces différentes du réseau (et, bien sûr, généralement à des positions complètement arbitraires). En particulier, le cas $n = 4$ de deux spins autour d'une face et deux spins autour d'une autre face a un grand intérêt pratique : il permettrait de trouver le comportement du corrélateur énergie-énergie sous la transformation de dualité [81]. Des résultats de ce type pourraient servir, par exemple, à construire le secteur énergétique de la matrice de transfert en s'appuyant sur des considérations topologiques. Une application au modèle de Potts désordonné où l'absence d'automoyennement rend très délicat l'évaluation numérique des exposants énergétiques [82, 88] paraît particulièrement désirable [81].

Cette question fut examinée par King [107]. Il trouva que la dualité ne permet pas de déterminer la transformation de toutes les fonctions de partition restreintes individuellement, mais seulement de quelques combinaisons linéaires. Une généralisation à plus de deux faces de ce type de résultat est également possible [108]. Or, à l'heure actuelle, l'approche de [108] ne retrouve pas toutes les relations données dans [107] pour le cas de deux faces.

2.5 Couplage de plusieurs modèles de Potts

La transformation de dualité s'applique également au cas de N modèles de Potts couplés au travers de leurs opérateurs d'énergie. Le hamiltonien pour le cas le plus simple, $N = 2$, s'écrit

$$-\beta H = \sum_{\langle ij \rangle} \left[K_{\bullet\circ}\delta(\sigma_i^{(1)}, \sigma_j^{(1)}) + K_{\circ\bullet}\delta(\sigma_i^{(2)}, \sigma_j^{(2)}) + K_{\bullet\bullet}\delta(\sigma_i^{(1)}, \sigma_j^{(1)})\delta(\sigma_i^{(2)}, \sigma_j^{(2)}) \right], \tag{2.22}$$

où $\sigma_i^{(k)}$ dénotent les spins de Potts dans le modèle numéro $k = 1, 2$. Ce modèle fut d'abord étudié dans [43].

La technique utilisée à l'époque était celle d'une transformation de Fourier sur réseau [162]. Elle s'applique au cas où une symétrie (ici de type Z_q) permet d'écrire les poids de Boltzmann sous forme d'une matrice cyclique. Or, cette technique devient d'une grande complexité si le nombre de modèles augmente. En particulier, $N = 3$ est un exercice assez compliqué [49].

Dans l'article joint à ce chapitre, nous montrons comment le cas de N général devient faisable si le système de modèles couplés est reformulé en modèle d'amas où en modèle de boucles. En termes des variables qui généralisent les u_{ij} dans (2.3) ou (2.5), les relations de dualité sont d'une simplicité consternante. Nous considérons ici le cas où le couplage entre les modèles réalise une symétrie S_N ; le cas de couplages arbitraires est traité dans la Section 3 de [51], attachée au Chapitre 8. Nous remettrons également à ce chapitre la motivation physique pour étudier des modèles couplés.

2.6 Article "Duality relations for M coupled Potts models"

Nous établierons des transformations de dualité explicites pour des systèmes de M modèles de Potts à q états couplés à travers leurs densités locales d'énergie. Ceci généralise des résultats connus pour $M = 1, 2, 3$. L'espace M-dimensionnel de constantes de couplages contient une sous-variété autoduale de dimension $D_M = \lfloor M/2 \rfloor$. Pour le cas $M = 4$, la variation de la charge centrale effective le long de la surface autoduale est étudiée par des techniques numériques de matrice de transfert. Nous mettons en évidence l'existence d'un ensemble de points critiques qui correspondent à des théories conformes avec un algèbre étendu de symétrie S_M.

PHYSICAL REVIEW E

STATISTICAL PHYSICS, PLASMAS, FLUIDS, AND RELATED INTERDISCIPLINARY TOPICS

THIRD SERIES, VOLUME 62, NUMBER 1 PART A JULY 2000

RAPID COMMUNICATIONS

The Rapid Communications section is intended for the accelerated publication of important new results. Since manuscripts submitted to this section are given priority treatment both in the editorial office and in production, authors should explain in their submittal letter why the work justifies this special handling. A Rapid Communication should be no longer than 4 printed pages and must be accompanied by an abstract. Page proofs are sent to authors.

Duality relations for M coupled Potts models

Jesper Lykke Jacobsen
LPTMS, Bâtiment 100, Université Paris–Sud, F-91405 Orsay, France
(Received 21 April 2000)

We establish explicit duality transformations for systems of M q-state Potts models coupled through their local energy density, generalizing known results for $M=1,2,3$. The M-dimensional space of coupling constants contains a self-dual submanifold of dimension $D_M=[M/2]$. For the case $M=4$, the variation of the effective central charge along the self-dual surface is investigated by numerical transfer matrix techniques. Evidence is given for the existence of a family of critical points, corresponding to conformal field theories with an extended S_M symmetry algebra.

PACS number(s): 05.50.+q, 05.70.Jk, 75.10.Hk

For several decades, the q-state Potts model has been used to model ferromagnetic materials [1], and an impressive number of results are known about it, especially in two dimensions [2–4]. More recently, its random-bond counterpart has attracted considerable attention [5], primarily because it permits one to study how quenched randomness coupling to the local energy density can modify the nature of a phase transition.

But despite the remarkable successes of conformal invariance applied to pure two-dimensional systems, the amount of analytical results on the random-bond Potts model is rather scarce. Usually the disorder is dealt with by introducing M replicas of the original model, with mutual energy-energy interactions, and taking the limit $M\to 0$. The price to be paid is, however, that the resulting system loses many of the properties (such as unitarity) that lie at the heart of conventional conformal field theory [6,7].

Very recently, an alternative approach was suggested by Dotsenko *et al.* [8]. These authors point out that the perturbative renormalization group [6] (effectively an expansion around the Ising model in the small parameter $\varepsilon=q-2$) predicts the existence of a nontrivial infrared fixed point at interlayer coupling $g_*\propto -\varepsilon/(M-2)+\mathcal{O}(\varepsilon^2)$, so that the regions $M<2$ and $M>2$ are somehow dual upon changing the sign of the coupling constant[1] [9]. More interestingly, for $M=3$ they identify the exact lattice realization of a critical theory with exponents consistent with those of the perturbative treatment, and they conjecture that this generalizes to any integer $M\geqslant 3$. Their proposal is then to study this class of coupled models, which are now unitary by definition, and only take the limit $M\to 0$ once the exact expressions for the various critical exponents have been worked out. One could hope to attack this task by means of extended conformal field theory, thus combining the Z_q symmetry of the spin variable by a non-Abelian S_M symmetry upon permuting the replicas.

Clearly, a first step in this direction is to identify the lattice models corresponding to this series of critical theories, parametrized by the integer $M\geqslant 3$. For $M=3$ this was achieved [8] by working out the duality relations for M coupled Potts models on the square lattice, within the M-dimensional space of coupling constants giving rise to S_M symmetric interactions amongst the lattice energy operators

[1]The case $M=2$ is special: For $q=2$ (the Ashkin-Teller model) the coupling presents a marginal perturbation, giving rise to a halfline of critical points along which the critical exponents vary continuously [3]. On the other hand, for $q>2$ where the perturbation is relevant, the model is still integrable, but now presents a mass generation leading to noncritical behavior [9].

of the replicas. Studying numerically the variation of the effective central charge [10] along the resulting self-dual line, using a very powerful transfer matrix technique, the critical point was unambiguously identified with one of the endpoints of that line.

Unfortunately, it was hard to see how such duality relations could be extended to the case of general M. The calculations in Ref. [8] relied on a particular version [11] of the method of lattice Fourier transforms [12], already employed for $M=2$ two decades ago [13]. Though perfectly adapted to the case of linear combinations of cosinoidal interactions within a single (vector) Potts model [12], this approach led to increasingly complicated algebra when several coupled models were considered. Moreover, it seemed impossible to recast the end results in a reasonably simple form for larger M.

In the present Rapid Communication we wish to assess whether such a scenario of a unique critical point with an extended S_M symmetry can indeed be expected to persist in the general case of $M\geqslant 3$ symmetrically coupled models. We explicitly work out the duality transformations for any M, and show that they can be stated in a very simple form [Eq. (9)] after redefining the coupling constants.

The lattice identification of the $M=3$ critical point in Ref. [8] crucially relied on the existence of a *one-parameter* self-dual manifold, permitting only two possible directions of the initial flow away from the decoupling fixed point. We find in general a richer structure with an $[M/2]$-dimensional self-dual manifold. Nonetheless, from a numerical study of the case $M=4$ we end up concluding that the uniqueness of the nontrivial fixed point can be expected to persist, since the decoupling fixed point acts as a saddle point of the effective central charge.

Consider then a system of M identical planar lattices, stacked on top of one another. On each lattice site i, and for each layer $\mu=1,2,\ldots,M$, we define a Potts spin $\sigma_i^{(\mu)}$ that can be in any of $q=2,3,\ldots$ distinct states. The layers interact by means of the reduced Hamiltonian

$$\mathcal{H}=\sum_{\langle ij\rangle}\mathcal{H}_{ij}, \tag{1}$$

where $\langle ij\rangle$ denotes the set of lattice edges, and an S_M symmetric nearest-neighbor interaction is defined as

$$\mathcal{H}_{ij}=-\sum_{m=1}^{M}K_m\sum_{\mu_1\neq\mu_2\neq\cdots\mu_m}{}'\ \prod_{l=1}^{m}\delta(\sigma_i^{(\mu_l)},\sigma_j^{(\mu_l)}). \tag{2}$$

By definition the primed summation runs over the $\binom{M}{m}$ terms for which the indices $1\leqslant\mu_l\leqslant M$ with $l=1,2,\ldots,m$ are all different, and $\delta(x,y)=1$ if $x=y$ and zero otherwise.

For $M=1$ the model thus defined reduces to the conventional Potts model, while for $M=2$ it is identical to the Ashkin-Teller-like model considered in Ref. [13], where the Potts models of either layer are coupled through their local energy density. For $M>2$, additional multienergy interactions between several layers have been added, since such interactions are generated by the duality transformations, as we shall soon see. However, from the point of view of conformal field theory these supplementary interactions are irrelevant in the continuum limit. The case $M=3$ was discussed in Ref. [8].

By means of a generalized Kasteleyn-Fortuin transformation [14] the local Boltzmann weights can be recast as

$$\exp(-\mathcal{H}_{ij})=\prod_{m=1}^{M}\prod_{\mu_1\neq\mu_2\neq\cdots\mu_m}{}'\times\left[1+(e^{K_m}-1)\prod_{l=1}^{m}\delta(\sigma_i^{(\mu_l)},\sigma_j^{(\mu_l)})\right]. \tag{3}$$

In analogy with the case of $M=1$, the products can now be expanded so as to transform the original Potts model into its associated random cluster model. To this end we note that Eq. (3) can be rewritten in the form

$$\exp(-\mathcal{H}_{ij})=b_0+\sum_{m=1}^{M}b_m\sum_{\mu_1\neq\mu_2\neq\cdots\mu_m}{}'\ \prod_{l=1}^{m}\delta(\sigma_i^{(\mu_l)},\sigma_j^{(\mu_l)}), \tag{4}$$

defining the coefficients $\{b_m\}_{m=0}^{M}$. The latter can be related to the physical coupling constants $\{K_m\}_{m=1}^{M}$ by evaluating Eqs. (3) and (4) in the situation where precisely m out of the M distinct Kronecker δ-functions are nonzero. Clearly, in this case Eq. (3) is equal to e^{J_m}, where

$$J_m=\sum_{k=1}^{m}\binom{m}{k}K_k \tag{5}$$

for $m\geqslant 1$, and we set $J_0=K_0=0$. On the other hand, we find from Eq. (4) that this must be equated to $\Sigma_{k=0}^{m}\binom{m}{k}b_k$. This set of $M+1$ equations can be solved for the b_k by recursion, considering in turn the cases $m=0,1,\ldots,M$. After some algebra, the edge weights b_k (for $k\geqslant 0$) are then found as

$$b_k=\sum_{m=0}^{k}(-1)^{m+k}\binom{k}{m}e^{J_m}. \tag{6}$$

The partition function in the spin representation

$$Z=\sum_{\{\sigma\}}\prod_{\langle ij\rangle}\exp(-\mathcal{H}_{ij}) \tag{7}$$

can now be transformed into the random cluster representation as follows. First, insert Eq. (4) on the right-hand side of the above equation, and imagine expanding the product over the lattice edges $\langle ij\rangle$. To each term in the resulting sum we associate an edge coloring $\mathcal{G}$ of the M-fold replicated lattice, where an edge (ij) in layer m is considered to be colored (occupied) if the term contains the factor $\delta(\sigma_i^{(m)},\sigma_j^{(m)})$, and uncolored (empty) if it does not. (In this language, the couplings J_k correspond to the local energy density summed over all possible permutations of precisely k simultaneously colored edges.)

The summation over the spin variables $\{\sigma\}$ is now trivially performed, yielding a factor of q for each connected component (cluster) in the coloring graph. Keeping track of the prefactors multiplying the δ-functions, using Eq. (4), we conclude that

$$Z=\sum_{\mathcal{G}}\prod_{m=1}^{M}q^{C_m}b_m^{B_m}, \tag{8}$$

where C_m is the number of clusters in the mth layer, and B_m is the number of occurencies in $\mathcal{G}$ of a situation where precisely m $(0 \leqslant m \leqslant M)$ edges placed on top of one another have been simultaneously colored.

It is worth noticing that the random cluster description of the model has the advantage that q only enters as a parameter. By analytic continuation one can thus give meaning to a noninteger number of states. The price to be paid is that the C_m are, *a priori*, nonlocal quantities.

In terms of the edge variables b_m the duality transformation of the partition function is easily worked out. For simplicity we shall assume that the couplings constants $\{K_m\}$ are identical between all nearest-neighbor pairs of spins, the generalization to an arbitrary inhomogeneous distribution of couplings being trivial. By analogy with the case $M=1$, a given coloring configuration $\mathcal{G}$ is taken to be dual to a coloring configuration $\tilde{\mathcal{G}}$ of the dual lattice obtained by applying the following duality rule: *Each colored edge intersects an uncolored dual edge*, and vice versa. In particular, the demand that the configuration $\mathcal{G}_{\text{full}}$ with all lattice edges colored be dual to the configuration $\mathcal{G}_{\text{empty}}$ with no colored (dual) edge fixes the constant entering the duality transformation. Indeed, from Eq. (8), we find that $\mathcal{G}_{\text{full}}$ has weight $q^M b_M^E$, where E is the total number of lattice edges, and $\mathcal{G}_{\text{empty}}$ is weighted by $q^{MF}\tilde{b}_0^E$, where F is the number of faces, including the exterior one. We thus seek for a duality transformation of the form $q^{MF}\tilde{b}_0^E Z(\{b_m\}) = q^M b_M^E \tilde{Z}(\{\tilde{b}_m\})$, where for any configuration $\mathcal{G}$ the edge weights must transform so as to keep the same relative weight between $\mathcal{G}$ and $\mathcal{G}_{\text{full}}$ as between $\tilde{\mathcal{G}}$ and $\mathcal{G}_{\text{empty}}$.

An arbitrary coloring configuration $\mathcal{G}$ entering Eq. (8) can be generated by applying a finite number of changes to $\mathcal{G}_{\text{full}}$, in which an edge of weight b_M is changed into an edge of weight b_m for some $m=0,1,\ldots,M-1$. By such a change, in general, a number $k \leqslant M-m$ of pivotal bonds are removed from the coloring graph, thus creating k new clusters, and the weight relative to that of $\mathcal{G}_{\text{full}}$ will change by $q^k b_m / b_M$. On the other hand, in the dual configuration $\tilde{\mathcal{G}}$ a number $M-m-k$ of clusters will be lost, since each of the k new clusters mentioned above will be accompanied by the formation of a *loop* in $\tilde{\mathcal{G}}$. The weight change relative to $\mathcal{G}_{\text{empty}}$ therefore amounts to $\tilde{b}_{M-m}/(\tilde{b}_0 q^{M-m-k})$. Comparing these two changes we see that the factors of q^k cancel nicely, and after a change of variables $m \to M-m$ the duality transformation takes the simple form

$$\tilde{b}_m = \frac{q^m b_{M-m}}{b_M} \quad \text{for } m=0,1,\ldots,M, \tag{9}$$

the relation with $m=0$ being trivial.

Self-dual solutions can be found by imposing $\tilde{b}_m = b_m$. However, this gives rise to only $[(M+1)/2]$ independent equations

$$b_{M-m} = q^{M/2-m} b_m \quad \text{for } m=0,1,\ldots,\left[\frac{M-1}{2}\right], \tag{10}$$

and the M-dimensional parameter space $\{b_m\}_{m=1}^M$, or $\{K_m\}_{m=1}^M$, thus has a self-dual submanifold of dimension $D_M=[M/2]$. In particular, the ordinary Potts model ($M=1$) has a unique self-dual point, while for $M=2$ [13] and $M=3$ [8] one has a line of self-dual solutions.

Our main result is constituted by Eqs. (5) and (6) relating the physical coupling constants $\{K_m\}$ to the edge weights $\{b_m\}$, in conjunction with Eqs. (9) and (10) giving the explicit (self-)duality relations in terms of the latter.

Since the interaction energies entering Eq. (3) are invariant under a simultaneous shift of all Potts spins, an alternative way of establishing the duality transformations procedes by Fourier transformation of the energy gaps [11]. This method was used in Refs. [13] and [8] to work out the cases $M=2$ and $M=3$, respectively. However, as M increases, this procedure very quickly becomes quite involved. To better appreciate the ease of the present approach, let us briefly pause to see how the parametrizations of the self-dual lines for $M=2,3$, expressed in terms of the couplings $\{K_m\}$, can be reproduced in a most expedient manner.

For $M=2$, Eq. (10) gives $b_2=q$, where from Eqs. (5) and (6) $b_2 = e^{2K_1+K_2} - 2e^{K_1} + 1$. Thus,

$$e^{K_2} = \frac{2e^{K_1}+(q-1)}{e^{2K_1}}, \tag{11}$$

in accordance with Ref. [13]. Similarly, for $M=3$ one has $b_1 = q b_2 / b_3 = b_2/\sqrt{q}$, with $b_1 = e^{K_1}-1$, b_2 as before, and $b_3 = e^{3K_1+3K_2+K_3} - 3e^{2K_1+K_2} + 3e^{K_1} - 1$. This immediately leads to the result given in Ref. [8],

$$\begin{aligned} e^{K_2} &= \frac{(2+\sqrt{q})e^{K_1} - (1+\sqrt{q})}{e^{2K_1}}, \\ e^{K_3} &= \frac{3(e^{K_1}-1)(1+\sqrt{q}) + q^{3/2}+1}{[(2+\sqrt{q})e^{K_1} - (1+\sqrt{q})t]^3} e^{3K_1}. \end{aligned} \tag{12}$$

Returning now to the general case, we notice that the self-dual manifold always contains two special points for which the behavior of the M coupled models can be related to that of a single Potts model. At the first such point,

$$b_m = q^{m/2} \quad \text{for } m=0,1,\ldots,\left[\frac{M}{2}\right], \tag{13}$$

one has $K_1 = \log(1+\sqrt{q})$ and $K_m = 0$ for $m=2,3,\ldots,M$, whence the M models simply decouple. The other point

$$b_m = \delta(m,0) \quad \text{for } m=0,1,\ldots,\left[\frac{M}{2}\right] \tag{14}$$

corresponds to $K_m=0$ for $m=1,2,\ldots,M-1$ and $K_M = \log(1+q^{M/2})$, when the resulting model is equivalent to a single q^M-state Potts model. Evidently, for $M=1$ these two special points coincide.

Specializing now to the case of a regular two-dimensional lattice, it is well-known that at the two special points the model undergoes a phase transition, which is continuous if the effective number of states (q or q^M as the case may be) is $\leqslant 4$ [15]. In Ref. [8] the question was raised whether one in general can identify further nontrivial critical theories on the self-dual manifolds. In particular, it was argued that for

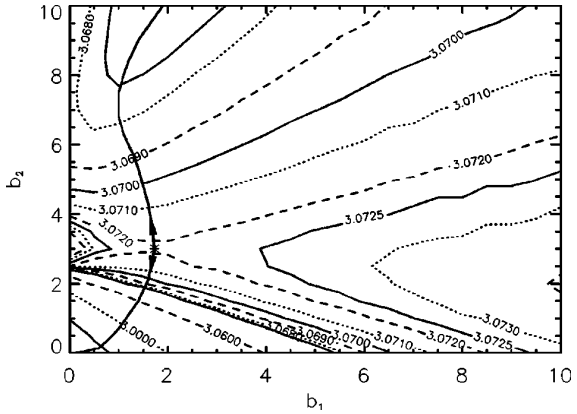


FIG. 1. Contour plot of the effective central charge $c(6,8)$ along the self-dual surface (b_1,b_2) for four coupled three-state Potts models. The decoupled fixed point is shown as an asterisk, and renormalization group flow lines are sketched as a guide to the eye.

$M=3$ there is indeed such a point, supposedly corresponding to a conformal field theory with an extended S_3 symmetry.

To get an indication whether such results can be expected to generalize also to higher values of M, we have numerically computed the effective central charge of $M=4$ coupled models along the two-dimensional self-dual surface. We were able to diagonalize the transfer matrix for strips of width $L=4,6,8$ lattice constants in the equivalent loop model. Technical details of the simulations have been reported in Ref. [8]. Relating the specific free energy $f_0(L)$ to the leading eigenvalue of the transfer matrix in the standard way, two estimates of the effective central charge, $c(4,6)$ and $c(6,8)$, were then obtained by fitting data for two consecutive strip widths according to [16]

$$f_0(L)=f_0(\infty)-\frac{\pi c}{6L^2}+\cdots. \tag{15}$$

A contour plot of $c(6,8)$, based on a grid of 21×21 parameter values for (b_1,b_2), is shown in Fig. 1. The data for $c(4,6)$ look qualitatively similar, but are less accurate due to finite-size effects. We should stress that even though the absolute values of $c(6,8)$ are some 4% below what one would expect in the $L\to\infty$ limit, the *variations* in c are supposed to be reproduced much more accurately [8]. On the figure $q=3$, but other values of q in the range $2<q\leqslant4$ lead to similar results.

According to Zamolodchikov's c-theorem [10], a system initially in the vicinity of the decoupled fixed point $(b_1,b_2)=(\sqrt{q},q)$, shown as an asterisk on the figure, will start flowing downhill in this central charge landscape. Figure 1 very clearly indicates that the decoupled fixed point acts as a saddle point, and there are thus only two possibilities for the direction of the initial flow.

The first of these will take the system to the stable fixed point at the origin which trivially corresponds to one self-dual q^4-state Potts model. For $q=3$ this leads to the generation of a finite correlation length, consistent with $c_{\text{eff}}=0$ in the limit of an infinitely large system. As expected, the flow starts out in the b_2 direction, meaning that it is the energy-energy coupling between layers (K_2) rather than the spin-spin coupling within each layer (K_1) that controls the initial flow.

More interestingly, if the system is started out in the *opposite* direction (i.e., with K_2 slightly positive) it will flow towards a third nontrivial fixed point, for which the edge weights tend to infinity in some definite ratios. (Exactly what these ratios are is difficult to estimate, given that the asymptotic flow direction exhibits finite-size effects.) Seemingly, at this point the central charge is only slightly lower than at the decoupled fixed point, as predicted by the perturbative renormalization group [8]. From the numerical data we would estimate the drop in the central charge as roughly $\Delta c=0.01$–0.02, in good agreement with the perturbative treatment which predicts $\Delta c=0.0168+\mathcal{O}(\varepsilon^5)$ [8].

All of these facts are in agreement with the conjectures put forward in Ref. [8], and in particular one would think that this third fixed point corresponds to a conformal field theory with a non-Abelian extended S_4 symmetry.

Finally, the numerics for $q=2$ (four coupled Ising models) is less conclusive, and we cannot rule out the possibility of a more involved fixed point structure. In particular, a $c=2$ theory is not only obtainable by decoupling the four models, but also by a pairwise coupling into two mutually decoupled four-state Potts (or Ashkin-Teller) models. Indeed, a similar phenomenon has already been observed for the case of *three* coupled Ising models [8].

The author is indebted to M. Picco for some very useful discussions.

[1] R. B. Potts, Proc. Cambridge Philos. Soc. **48**, 106 (1952).
[2] F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).
[3] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, New York, 1982).
[4] H. Saleur, Nucl. Phys. B **360** [FS], 219 (1991).
[5] J. Cardy, Physica A **263**, 215 (1999).
[6] A. W. W. Ludwig and J. L. Cardy, Nucl. Phys. B **285** [FS], 687 (1987).
[7] A. W. W. Ludwig, Nucl. Phys. B **330**, 639 (1990).
[8] Vl. S. Dotsenko, J. L. Jacobsen, M.-A. Lewis, and M. Picco, Nucl. Phys. B **546** [FS], 505 (1999).
[9] I. Vaysburd, Nucl. Phys. B **446**, 387 (1995).
[10] A. B. Zamolodchikov, Pis'ma Zh. Éksp. Teor. Fiz. **43**, 565 (1986) [JETP Lett. **43**, 730 (1986)].
[11] F. Y. Wu and Y. K. Wang, J. Math. Phys. **17**, 439 (1976).
[12] R. Savit, Rev. Mod. Phys. **52**, 453 (1980).
[13] E. Domany and E. K. Riedel, Phys. Rev. B **19**, 5817 (1979).
[14] P. W. Kasteleyn and C. M. Fortuin, J. Phys. Soc. Jpn. (suppl.) **26**, 11 (1969).
[15] R. J. Baxter, J. Phys. C **6**, L445 (1973).
[16] H. Blöte, J. L. Cardy, and M. P. Nightingale, Phys. Rev. Lett. **56**, 742 (1986); I. Affleck, *ibid.* **56**, 746 (1986).

Chapitre 3

Coloriage de sommets

3.1 Le polynôme chromatique

Il est très intéressant de prendre la limite $u \to -1$ du polynôme dichromatique (2.4) : cette limite correspond à la limite $K \to -\infty$ du modèle de Potts, c'est-à-dire au modèle antiferromagnétique à température nulle. Dans cette limite, la fonction de partition est connue sous le nom de *polynôme chromatique* et elle fut d'abord considérée par Birkhoff [25].

L'intérêt du polynôme chromatique $Z(-1,q)$ est double. Premièrement, ainsi que nous l'indique la nomenclature, $Z(-1,q)$ s'interprète de manière combinatoire comme le nombre de coloriages des sommets V du graphe G par q couleurs différentes. Plus précisement, si deux sommets $i,j \in V$ sont reliés par une arête $e_{ij} \in E$, alors ces deux sommets doivent porter des couleurs différentes, $\sigma_i \neq \sigma_j$. Deuxièmement, c'est bien connu que des contraintes géométriques de ce type sont susceptibles d'introduire de la frustration dans le modèle. Par frustration, nous entendons l'impossibilité du système de satisfaire simultanément toutes les contraintes qu'on lui impose. En même temps, la rigidité de la contrainte de q-coloriabilité fait que si l'on cherche à adapter l'état du système aux contraintes en un endroit défini, on va souvent engendrer des conflits avec ces mêmes contraintes en un autre endroit. Autrement dit, le système possède des corrélations effectives de longue portée et il est susceptible d'être dans un état critique.

Le polynôme chromatique a fait l'objet d'un grand nombre d'études dans la littérature mathématique. Le lecteur intéressé pourra se tourner vers les revues classiques [144, 145]. Une excellente bibliographie a été dressée par Chia [38].

3.2 Frustration géométrique : un exemple

L'exemple le plus connu (et le plus ancien) de la frustration géométrique est sans doute le modèle d'Ising sur le réseau triangulaire [155, 26, 135]. Bien sûr, on a $Z(u=-1,q=2)=0$ puisque ce réseau n'est pas bicoloriable. Mais il existe toujours un nombre macroscopique d'états fondamentaux [160] qui sont caractérisés par le fait que, autour de chaque face triangulaire, les trois spins ne prennent pas tous la même valeur. Autrement dit, précisement une des trois arêtes entourant chaque face est frustrée (puisque les deux spins qu'elle relie prennent la même valeur).

Il est facile de voir que, dans la limite continue, la classe d'universalité du modèle ainsi défini est différente de celle du modèle d'Ising ferromagnétique [26, 135]. En effet, si on efface les arêtes frustrées, le réseau est transformé en un pavage du plan par trois types de losanges ; voir la Figure 3.1. Ces losanges peuvent être vus comme la projection bidimensionnelle d'un empilement de boîtes cubiques en trois dimensions. Le modèle se transforme donc en un modèle d'interface avec une hauteur scalaire. On peut se convaincre que ce modèle est régi par la théorie conforme d'un champ bosonique libre. En particulier, la charge centrale vaut $c=1$. Ceci est en contraste avec le cas ferromagnétique [137] qui est équivalent à un fermion libre [153], ce qui donne $c=\frac{1}{2}$ [62].

Une autre transformation exacte met en évidence la présence de corrélations effectives de longue portée. Dessinons les arêtes du réseau hexagonal qui sont duaux aux arêtes non frustrées du réseau triangulaire. Ces arêtes forment un gaz de boucles compactes sur le réseau hexagonal : chaque sommet est traversé par exactement une boucle. La Figure 3.1 montre que ces boucles peuvent être très longues : dans le Chapitre 5 nous montrerons que la distribution des longueurs est en effet une loi de puissance.

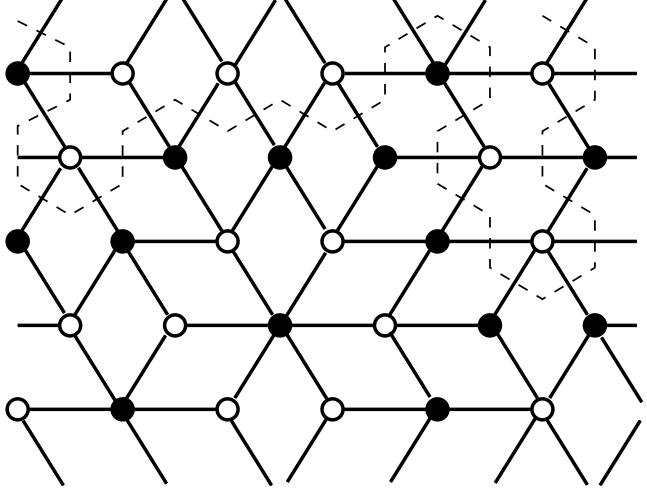


FIG. 3.1 – Coloriage des sommets du réseau triangulaire avec deux couleurs (• et ∘). En permettant une arête frustrée (•—• ou ∘—∘) par face, les arêtes non frustrées (trait continu) définissent un pavage du plan par trois types de losanges. Les duaux des arêtes frustrées (exemple en trait brisé) forment des boucles compactes sur le réseau hexagonal.

Le fait que $Z(u=-1,q=2)=0$ est un exemple d'un zéro de la fonction de partition en un nombre de Beraha, $q=B_4=2$. Des zéros de ce type vont jouer un rôle important par la suite. En quelque sorte, on pourrait considérer que "l'accident" que Z s'annule exactement à $q=2$ cache le vrai comportement critique du système. Ce comportement serait donc plutôt visible en s'approchant à $q=2$ dans le sens d'une limite, $q\to 2$. Nous allons adopter ce point de vue dans le Chapitre 4, où nous verrons que la charge centrale dans la limite $q\to 2$ vaut $c=-\frac{25}{2}\neq 1$.

3.3 Zéros de la fonction de partition

L'étude du rôle des zéros de la fonction de partition pour le comportement critique d'un système fut initiée dans deux publications classiques par Lee et Yang [167, 120]. Leur idée était de considérer l'énergie libre $F(L)$ d'un système thermodynamique de taille caractéristique L comme une fonction d'un paramètre p prenant des valeurs *complexes*. Dans l'exemple traité par Lee et Yang, ce paramètre est le champ magnétique du modèle d'Ising [120], mais en principe on pourrait traiter n'importe quel paramètre physique d'un système thermodynamique arbitraire de la même manière.

Pour tout L fini, $F(L)$ est une fonction analytique de la restriction de p à l'axe réel. Pourtant, ceci n'est plus forcément vrai dans la limite thermodynamique $L\to\infty$. Notamment, il est possible de développer des discontinuités, ou d'autres types de singularités, dont l'interprétation physique est l'existence d'une transition de phase à une certaine valeur critique p_c réelle. En utilisant un résultat pour les fonctions d'une variable complexe, Lee et Yang argumentèrent que, déjà pour L finie, cette transition doit laisser des signatures dans le plan de p complexe.

Plus précisement, si une région infinitésimale autour d'un point p_0 sur l'axe réel est dépourvue de singularités pour tout L fini, alors $F(L\to\infty)$ est analytique en p_0. Dans le cas contraire, des zéros de $Z(L)$ dans le plan de p complexe vont s'approcher à p_0 lorsque L augmente afin de "pincer" l'axe réel dans la limite $L\to\infty$.

De nombreux auteurs, beaucoup trop nombreux pour être mentionnés ici, ont poursuivi ce programme pour une grande variété de modèles. Une vague d'activité a notamment été déclenchée par la possibilité de calculer $Z(L)$ exactement (typiquement sous forme d'un polynôme dans le paramètre p) à l'aide d'un ordinateur. Pour la partie de l'histoire liée au polynôme chromatique, on peut consulter les références de [147].

3.4 Les nombres de Beraha

Beraha observa [18] que le polynôme chromatique d'un graphe planaire a souvent des zéros très proches à un ou plusieurs des nombres

$$B_n = 4\cos^2(\pi/n), \qquad n = 2, 3, \ldots \tag{3.1}$$

dits *nombres de Beraha*. Pourtant, on peut prouver [147] qu'aucun nombre de Beraha non entier, à la possible exception de B_{10} pour lequel la preuve ne marche pas, agit comme zéro pour un graphe fini. Par contre, il a été prouvé [19, 20] que certains graphes en forme de rubans de largeur fixe et de longueur L variable fournissent des zéros qui convergent vers quelques B_n non entiers quand $L \to \infty$.

L'idée d'origine derrière les travaux de Beraha et collaborateurs fut probablement de trouver un graphe planaire fini avec un zéro chromatique à $q = 4$. Une telle découverte aurait contredit la Conjecture de Quatre Couleurs [75] : l'ironie de l'histoire était qu'entre la date de soumission (1976) de [19] et sa publication (1979), cette conjecture avait été promue en théorème [3, 4] ! Néanmoins, Beraha et Kahane observèrent que certains familles de rubans ont des zéros chromatiques qui s'accumulent à $q = 4$ lorsque $L \to \infty$.

Le rôle des nombres de Beraha dans le contexte d'une théorie de champs fut illuminé par Saleur [150].

Récemment, une série de travaux [147, 90, 95] fut initiée afin de trouver les zéros chromatiques dans le plan de q complexe pour certaines familles de rubans, directement dans la limite $L \to \infty$. Nous allons en donner quelques résultats après une petite digression technique.

3.5 Outils techniques

Le polynôme chromatique d'un ruban G_L de longueur L peut être trouvé à partir de sa matrice de transfert :

$$Z_{G_L}(-1, q) = \mathrm{tr}\left[A(q)T(q)^L\right] = \sum_{k=1}^{M} \alpha_k(q)\lambda_k(q)^L. \tag{3.2}$$

Ici, $A(q)$ est une matrice qui implante les conditions au bord libres aux extrémités du ruban. $\alpha_k(q)$ et $\lambda_k(q)$ sont respectivement les amplitudes et les valeurs propres de la matrice de transfert $T(q)$ de dimension M.

Il n'est pas évident que le polynôme (di)chromatique puisse être calculé à partir d'une matrice de transfert, puisque les amas de Kasteleyn-Fortuin sont *a priori* des objets non locaux. Pourtant, une telle construction est tout à fait possible : elle fut discutée pour le cas chromatique dans [24] et pour le cas dichromatique dans [27].

Les points d'accumulation lorsque $L \to \infty$ des zéros chromatiques forment soit des courbes continues, soit des points isolés dans le plan de q complexe. Appelons une valeur propre $\lambda_k(q)$ *dominante* si $|\lambda_k(q)| \geq |\lambda_l(q)|$ pour tout $l = 1, 2, \ldots, M$. Par le théorème de Beraha-Kahane-Weiss [19, 20, 154], les points d'accumulation z sont alors caractérisés par, soit

- l'existence d'une unique valeur propre dominante dont l'amplitude correspondante s'annule en z,
- soit l'existence d'au moins deux valeurs propres dominantes en z.

La première condition correspond au cas des zéros isolés et la deuxième au cas de courbes continues.

Une manière de tracer les courbes continues passe par le calcul, pour une grille de valeurs de $\theta \in [0, \pi[$, des racines de la résultante [147]

$$R_\theta(q) = \mathrm{Res}_\lambda\left[P(\lambda, q), P(\mathrm{e}^{i\theta}\lambda, q)\right], \tag{3.3}$$

où $P(\lambda, q) = \det[\lambda I - T(q)]$ est le polynôme caractéristique de $T(q)$. Les extrémités de la courbe correspondent à une double racine de $P(\lambda, q)$ et seront alors trouvées pour $\theta = 0$.

On peut argumenter [147] que la convergence des zéros vers leurs points d'accumulation lorsque $L \to \infty$ possède des effets de taille finie d'ordre $1/L$ pour un point à l'intérieur d'une courbe et d'ordre $1/L^2$ pour un point d'extrémité. Remarquons aussi que les courbes peuvent former des bifurcations lorsque plus de deux valeurs propres deviennent dominantes en un même point.

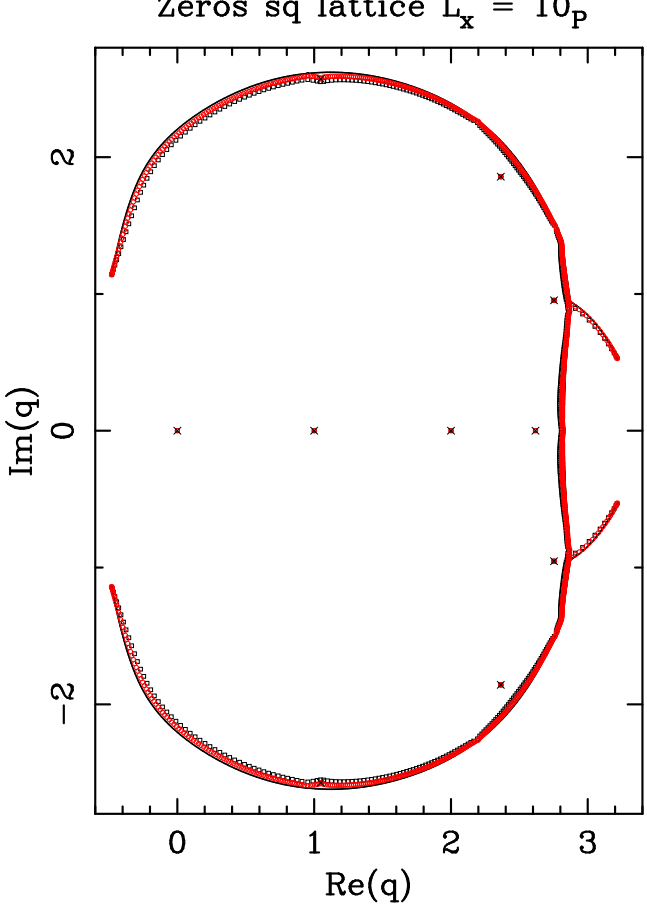


FIG. 3.2 – Zéros chromatiques des rubans du réseau carré de tailles 10×50 (carrés) et 10×100 (losanges). Les conditions au bord sont libres dans la direction longitudinale et périodiques dans la direction transversale. Les points d'accumulation des zéros forment des points isolés (croix) ainsi qu'une courbe continue. Une inspection minutieuse de cette courbe montre qu'elle possède dix composantes connexes et deux points de bifurcation.

Finalement, les zéros isolés peuvent être identifiés à partir du résultat simple :

$$\det D = \prod_{k=1}^{M} \alpha_k(q) \prod_{1 \le i \le j \le M} [\lambda_j(q) - \lambda_i(q)]^2 , \tag{3.4}$$

où D est une matrice $M \times M$ avec entrées $D_{ij} = \sum_{k=1}^{M} \alpha_k(q) [\lambda_k(q)]^{i+j-2}$.

3.6 Structure des zéros chromatiques

Les courbes $\mathcal{B}_m$ des points d'accumulation des zéros chromatiques ont été calculées pour des rubans des réseaux carré [147, 90] et triangulaire [95] de taille $m \times \infty$ pour tout $m \le 11$. Les effets de différents types de conditions au bord dans la direction de m ont également été étudiés. (Pour les références aux articles d'origine pour quelques valeurs de m moins importantes, voir [147, 90, 95].)

Pour ces deux réseaux, on trouve que les $\mathcal{B}_m$ coupent l'axe réel en un point $q_0(m)$ qui s'interprète comme un point de non analyticité de l'énergie libre. Lorsque $m \to \infty$, on a $q_0(m) \to q_0$, où le nombre q_0 dépend du réseau. Numériquement, il semblerait que $q_0 = 3$ pour le réseau carré [90] : la Figure 3.2 montre $\mathcal{B}_{10}$ pour ce réseau.

Les $\mathcal{B}_m$ possèdent aussi une paire de branches ouvertes qui s'approchent à l'axe des q réels lorsque m augmente. Il est raisonnable de supposer que ces branches coupent l'axe réel en un point critique q_c dans la limite $m \to \infty$. Numériquement, il semblerait que $q_c = 3$ pour le réseau carré et $q_c = 4$ pour le réseau triangulaire. Ce scénario est d'autant plus crédible qu'on peut se convaincre de la criticité de ces deux modèles :

- Le polynôme chromatique sur le réseau carré avec $q = 3$ est équivalent au modèle de six vertex [122] avec $c = 1$.
- Le cas du réseau triangulaire avec $q = 4$ est équivalent au tricoloriage du réseau Kagomé [78, 131] avec $c = 2$.

Le rôle important de q_0 est souligné par l'observation empirique que seuls les nombres de Beraha avec $B_n \leq q_0$ sont des points d'accumulation des zéros chromatiques [16, 149, 147, 90, 95]. Autrement dit, la phase dont l'intersection avec l'axe des q réels est l'intervalle $[q_0, q_c]$ est distincte de la "phase de Beraha" (qui, en fait, est techniquement une phase de Berker-Kadanoff, comme nous l'expliquerons dans le Chapitre 4).

De l'information supplémentaire sur le cas du réseau triangulaire est disponible directement dans la limite thermodynamique ($m \to \infty$), où Baxter [15, 16] a trouvé une solution exacte. En particulier, la position de la courbe $\mathcal{B}_\infty$ est donnée implicitement par un système d'équations faisant intervenir des produits infinis. Or, ces produits convergent très lentement quand q s'approche de l'axe réel, rendant difficile ne serait-ce qu'une estimation numérique du point q_0. Ce point est discuté en détail dans l'article [95] attaché à ce chapitre, où nous montrons que la valeur correcte n'est pas celle donnée par Baxter [16] mais plutôt

$$q_0(\text{triangulaire}) = B_{12} = 2 + \sqrt{3}. \tag{3.5}$$

Pour le réseau carré, la limite thermodynamique n'a pas été résolue. Une explication possible est que la courbe $\mathcal{B}_m$ a tendance à se fractionner en une multitude de composantes connexes quand m augmente [147, 90] : sur la Figure 3.2 on peut en distinguer dix. Paradoxalement, c'est justement pour le réseau carré que le diagramme de phase complète dans le plan (u, q) est le mieux compris au niveau de la théorie de champs [149, 150]. Nous y reviendrons lors du prochain chapitre où nous discuterons également les classes d'universalité des polynômes chromatiques.

3.7 Article "Triangular lattice chromatic polynomial"

Nous étudions le polynôme chromatique $P_G(q)$ des rubans $m \times n$ du réseau triangulaire avec largeur $m \leq 12_{\rm P}, 9_{\rm F}$ (avec des conditions au bord périodiques et libres respectivement) et avec longueur n arbitraire (avec des conditions au bord longitudinales libres). Le polynôme chromatique est la limite à température nulle de la fonction de partition du modèle de Potts antiferromagnétique à q états. Nous calculons la matrice de transfert pour de tels rubans dans la représentation de Fortuin-Kasteleyn et nous obtenons les ensembles correspondants de points d'accumulation des zéros chromatiques dans le plan de q complexe, dans la limite $n \to \infty$. Nous recalculons la courbe limite obtenue par Baxter dans la limite thermodynamique $m, n \to \infty$ et nous trouvons de nouvelles propriétés intéressantes avec de possibles conséquences physiques. Finalement, nous analysons les points d'accumulation isolés et leur relation avec les nombres de Beraha.

Transfer Matrices and Partition-Function Zeros for Antiferromagnetic Potts Models

III. Triangular-lattice chromatic polynomial

Jesper Lykke Jacobsen
Laboratoire de Physique Théorique et Modèles Statistiques
Université Paris-Sud
Bâtiment 100
F-91405 Orsay, FRANCE
JACOBSEN@IPNO.IN2P3.FR

Jesús Salas
Departamento de Física Teórica
Facultad de Ciencias, Universidad de Zaragoza
Zaragoza 50009, SPAIN
JESUS@MELKWEG.UNIZAR.ES

Alan D. Sokal
Department of Physics
New York University
4 Washington Place
New York, NY 10003 USA
SOKAL@NYU.EDU

April 24, 2002

Abstract

We study the chromatic polynomial $P_G(q)$ for $m \times n$ triangular-lattice strips of widths $m \le 12_{\rm P}, 9_{\rm F}$ (with periodic or free transverse boundary conditions, respectively) and arbitrary lengths n (with free longitudinal boundary conditions). The chromatic polynomial gives the zero-temperature limit of the partition function for the q-state Potts antiferromagnet. We compute the transfer matrix for such strips in the Fortuin–Kasteleyn representation and obtain the corresponding accumulation sets of chromatic zeros in the complex q-plane in the limit $n \to \infty$. We recompute the limiting curve obtained by Baxter in the thermodynamic limit $m, n \to \infty$ and find new interesting features with possible physical consequences. Finally, we analyze the isolated limiting points and their relation with the Beraha numbers.

Key Words: Chromatic polynomial; chromatic root; antiferromagnetic Potts model; triangular lattice; transfer matrix; Fortuin–Kasteleyn representation; Beraha–Kahane–Weiss theorem; Beraha numbers.

1 Introduction

The antiferromagnetic q-state Potts model [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14] exhibits unusual behavior not found in ferromagnets. Indeed, its phase diagram and critical behavior depend crucially on the lattice structure, unlike the ferromagnetic case where the concept of universality applies. Specifically, for each lattice $\mathcal{L}$ there exists a number $q_c(\mathcal{L})$ such that for all $q > q_c$ the model is disordered at any temperature, including zero temperature [11]. Exactly at $q = q_c(\mathcal{L})$ the system is disordered at all positive temperatures and has a zero-temperature critical point. The zero-temperature limit of the antiferromagnetic Potts model is particularly interesting because its partition function on a finite graph G coincides with the chromatic polynomial $P_G(q)$, which counts the number of ways of coloring the vertices of G using q colors subject to the constraint that adjacent vertices always receive different colors [15].

In Refs. [16, 17] we undertook a study of the zeros of the chromatic polynomial when the parameter q is allowed to take complex values (see [16] for detailed references to the previous literature). In those papers we studied strips of the square lattice; here we extend that work to the triangular lattice. The triangular-lattice case is of particular interest because the path-breaking work of Baxter [18, 19] provides a conjectured exact solution in the thermodynamic limit.

The study of the complex zeros of the chromatic polynomial is inspired by the Yang–Lee picture of phase transitions [20]. We study families of graphs G_n for which the chromatic polynomial can be expressed via a transfer matrix of fixed size $M \times M$:

$$P_{G_n}(q) = \mathrm{tr}[A(q)\, T(q)^n] \tag{1.1a}$$

$$= \sum_{k=1}^{M} \alpha_k(q)\, \lambda_k(q)^n \;, \tag{1.1b}$$

where the transfer matrix $T(q)$ and the boundary-condition matrix $A(q)$ are polynomials in q, so that the eigenvalues $\{\lambda_k\}$ of T and the amplitudes $\{\alpha_k\}$ are algebraic functions of q. Rather than using $T(q)$ to compute the zeros of the chromatic polynomial for a *finite* strip $m \times n$, we have focussed on the direct calculation of their accumulation points in the limit $n \to \infty$, i.e. for the case of an *semi-infinite* strip [21, 22, 23, 24, 16, 17]. According to the Beraha–Kahane–Weiss theorem [25, 26, 27], the accumulation points of zeros when $n \to \infty$ can either be isolated limiting points (when the amplitude associated to the dominant eigenvalue vanishes, or when all eigenvalues vanish simultaneously) or belong to a limiting curve $\mathcal{B}$ (when two dominant eigenvalues cross in modulus). By studying the limiting curves for different values of the strip width m, we hope to learn new features of the thermodynamic limit $m \to \infty$.

To determine the isolated limiting points, we shall take advantage of the following simple result [22]:

$$\det D = \prod_{k=1}^{M} \alpha_k \prod_{1 \le i \le j \le M} (\lambda_j - \lambda_i)^2 \;, \tag{1.2}$$

where D is the $M \times M$ matrix with entries $D_{ij} = \sum_{k=1}^{M} \alpha_k (\lambda_k)^{i+j-2} = P_{G_{i+j-2}}$.

An important feature of the limiting curve $\mathcal{B}$ is that it typically crosses the positive real axis at a point $q_0(m)$.[1] Physically, $q_0(m)$ corresponds to a point of non-analyticity of the ground-state degeneracy per site [23, 28]. As the strip width m grows, this crossing point $q_0(m)$ increases and presumably tends to a limiting value $q_0(\infty)$. On the other hand, as $m \to \infty$ the curve $\mathcal{B} = \mathcal{B}_m$ presumably tends to a thermodynamic-limit curve $\mathcal{B}_\infty$. We define q_c to be the largest value where $\mathcal{B}_\infty$ crosses the real q-axis: please note that q_c may or may not correspond to crossings of the real axis for any finite m. In general the value $q_0(\infty)$ is smaller than q_c, although for some lattices they may coincide (this depends on the shape of the curves $\mathcal{B}_m$ and $\mathcal{B}_\infty$). Indeed, in Ref. [17] evidence was presented suggesting that for the square lattice $q_0(\infty) = q_c = 3$.

The crucial role of q_0 is further emphasized by studying the relation between chromatic polynomials and the so-called *Beraha numbers*

$$B_n \;=\; 4\cos^2\frac{\pi}{n} \;=\; 2 + 2\cos\frac{2\pi}{n} \qquad \text{for } n = 2, 3, \ldots \; . \tag{1.3}$$

It has been found in a number of cases [29, 21, 22] that chromatic roots tend to accumulate at some of the Beraha numbers. In the case of the *square lattice*, we have found empirically [16, 17] that on a strip of width m with either free or periodic transverse boundary conditions[2], there is at least one vanishing amplitude $\alpha_i(q)$ at each of the first m Beraha numbers $B_2, \ldots, B_{m+1}$ (but not higher ones). It thus appears that in the limit $m \to \infty$ all the Beraha numbers will be zeros of some amplitude.[3] Moreover, we found that the vanishing amplitude corresponds to the eigenvalue obtained by analytic continuation in q from the one that is dominant at small real q, in agreement with a conjecture of Baxter [19, p. 5255]. Thus, the first few Beraha numbers — namely, those (up to at most B_{m+1}) that lie below the point $q_0(m)$ — correspond to the vanishing of a dominant amplitude and hence, by the Beraha–Kahane–Weiss theorem, to a limit point of chromatic roots, while the remaining Beraha numbers do not. As the strip width m grows, $q_0(m)$ tends to $q_0(\infty)$, and the limiting points of chromatic roots are thus constrained to be the points $B_2, B_3, \ldots, B_p < q_0(\infty)$. This scenario for the accumulation of chromatic roots at *some* of the Beraha numbers was set forth by Baxter [19] and elaborated by Saleur [8]; further references can be found in [16].

---

[1] If there is more than one such crossing, we define $q_0(m)$ to be the *smallest* such crossing. When no such crossing occurs, the limiting curve often includes a pair of complex-conjugate endpoints rather close to the positive real q-axis. In these cases, we define $q_0(m)$ to be the point closest to that axis with positive imaginary part.

[2] Let m (resp. n) denote the number of sites in the transverse (resp. longitudinal) direction of the strip, and let F (resp. P) denote free (resp. periodic) boundary conditions in a given direction. Then we use the terminology: free ($m_{\rm F} \times n_{\rm F}$), cylindrical ($m_{\rm P} \times n_{\rm F}$), cyclic ($m_{\rm F} \times n_{\rm P}$), and toroidal ($m_{\rm P} \times n_{\rm P}$). In this paper we consider free and cylindrical boundary conditions, as well as a new type of boundary condition that we shall call "zig-zag" (Section 5).

[3] As we shall see, a similar (but not identical) statement appears to hold true also for the triangular lattice: see Section 7.2.

In the present publication we shall be concerned with the antiferromagnetic Potts model on the *triangular lattice*. For this case, Baxter and collaborators [30, 18, 19] have determined the exact free energy (among other quantities) on two special curves in the (q,v)-plane:

$$v^3 + 3v^2 - q = 0 \tag{1.4}$$

$$v = -1 \tag{1.5}$$

The uppermost branch ($v \geq 0$) of curve (1.4) is known to correspond to the ferromagnetic critical point [30, 3], and Baxter [18] initially conjectured (following a hint of Nienhuis [31]) that the zero-temperature antiferromagnetic model (1.5) is critical in the interval $0 \leq q \leq 4$. This prediction is known to be correct for $q = 2$ [32, 33, 34] and is believed to be correct also for $q = 4$ [5, 35, 36]. On the other hand, for $q = 3$ the conjecture contradicts the rigorous result [37], based on Pirogov-Sinai theory, that there is a low-temperature phase with long-range order and small correlation length.[4] In any case, for $q > 4$ we expect that the triangular-lattice Potts model is noncritical even at zero temperature; this has recently been confirmed by Monte Carlo simulation of the models with $q = 5, 6$ [38]. We therefore expect that for the triangular lattice $q_c = 4$.

For the model (1.5), Baxter [18] used a Bethe Ansatz to compute three different expressions $\lambda_i(q)$ $[i = 1, 2, 3]$ that he argued correspond to the dominant eigenvalues of the transfer matrix in different regions $\mathcal{D}_i$ of the complex q-plane; in a second paper [19] he provided corrected estimates for the precise locations of $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$. Using these formulae, he determined the value of $q_0(\infty)$ as

$$q_0(\infty, \mathrm{tri}) \approx 3.81967 \qquad \text{(Baxter)} \tag{1.6}$$

An important outcome of the present paper (see Section 6) is that the "phase diagram" of the triangular-lattice model is actually more complicated than what Baxter found. In particular, it now appears that the correct value of $q_0(\infty, \mathrm{tri})$ is slightly smaller than Baxter's value (1.6), the corrected value being

$$q_0(\infty, \mathrm{tri}) = B_{12} = 2 + \sqrt{3} \approx 3.73205 \qquad \text{(this paper)} \tag{1.7}$$

As a consequence we conjecture that the isolated limiting chromatic roots of the infinite-size triangular lattice are $B_2, \ldots, B_{11}$ and possibly B_{12}, rather than $B_2, \ldots, B_{14}$ as conjectured by Baxter.

To study the approach to the thermodynamic limit, we have computed the transfer matrix for triangular-lattice strips of widths $2 \leq m \leq 9$ with free boundary conditions and $2 \leq m \leq 12$ with cylindrical boundary conditions, and we have determined the corresponding limiting curves $\mathcal{B}$. These results also serve as a testing ground for the general conjectures on Beraha numbers as stated above (see Ref. [16] for further details).

---

[4] A Monte Carlo study of the $q = 3$ model found strong evidence for a first-order transition to an ordered phase at $\beta J \approx -1.594$ [10].

Previous studies using a similar approach have been made notably by Shrock and collaborators. In particular, they have considered triangular-lattice strips of width $m \le 5$ for free and cylindrical boundary conditions [24, 41, 42, 43, 44].[5] They have also considered other boundary conditions for the same lattice [44, 45, 46, 47].[6] Generalizations to nonzero temperature for several boundary conditions have been carried out in Refs. [48, 49]. Finally, Refs. [47, 50] discuss some general structural properties of the Potts-model partition function and chromatic polynomial on square-lattice and triangular-lattice strips.

This paper is laid out as follows: In Section 2 we discuss some brief preliminaries. In Section 3 we give our numerical results for free transverse boundary conditions, and in Section 4 for periodic transverse (cylindrical) boundary conditions. We have also examined a third type of boundary conditions, called "zig-zag", which we introduce and motivate in Section 5. In Section 6 we analyze Baxter's [18, 19] exact solution for the thermodynamic limit and revisit his phase diagram. Finally, in Section 7 we present our conclusions.

2 Preliminaries

The general theory of the transfer-matrix method used here has been explained in Ref. [16], and the implementational details of an improved algorithm have been given in Ref. [17]. Suffice it here to mention that we have used the Fortuin–Kasteleyn representation [39, 40] of the Potts model in the computation of the transfer matrix; therefore, all quantities are expressed as polynomials in q.

To compute the limiting curves $\mathcal{B}$ we have used two different techniques: the resultant method and a direct-search method. These techniques have been described in [16, Section 4.1], and we use here the same conventions and notation.

Let us briefly mention a few improvements/additions to our methodology:

Computation of T points. We have adopted an improved method for locating T points, based on applying a numerical minimization algorithm (e.g. MATHEMATICA's `FindMinimum`) to the function

$$F(q) \;=\; (|\lambda_1(q)| - |\lambda_2(q)|)^2 \,+\, (|\lambda_1(q)| - |\lambda_3(q)|)^2 \,+\, (|\lambda_2(q)| - |\lambda_3(q)|)^2 \;, \qquad (2.1)$$

where $\lambda_1, \lambda_2, \lambda_3$ are the three eigenvalues of the transfer matrix of largest modulus. At any given q these eigenvalues can easily be computed numerically by finding the roots of the characteristic polynomial of the transfer matrix. Using this method, we are able to locate T points far more precisely than in our previous work.

Fixed zeros. When q is an integer and the graph G is not q-colorable, we have $P_G(q) = 0$. For this reason, certain small integers q can be "fixed" zeros of the

[5] The case $m = 4$ with cylindrical boundary conditions was first done by Beraha and Kahane [21]. The case $m = 5$ with cylindrical boundary conditions was first done by Beraha, Kahane and Weiss [22].

[6] The case $2_{\rm F} \times n_{\rm P}$ (i.e., cyclic boundary conditions) was first done by Beraha, Kahane and Weiss [22].

zero-temperature partition function, independent of the strip length n. In particular, $q = 0, 1$ are roots for all widths $m \geq 2$ and lengths $n \geq 1$. Furthermore, $q = 2$ is a root for all triangular-lattice strips of widths $m \geq 2$ and lengths $n \geq 2$, because the triangular lattice is not bipartite. Finally, $q = 3$ is a root for all *cylindrical* triangular-lattice strips of widths that are not multiples of 3 (with lengths $n \geq 2$), because these graphs are not 3-colorable.

It is interesting to see how these behaviors come about from the point of view of the transfer-matrix formalism. The partition function on a lattice of length n has the form

$$Z_n \;=\; \sum_{k=1}^{M} \alpha_k(q)\, \lambda_k(q)^{n-1} \;, \tag{2.2}$$

where the $\{\lambda_k\}$ are the eigenvalues of the transfer matrix and the $\{\alpha_k\}$ are the corresponding amplitudes. A particular value q can then be a "fixed" zero of Z for any of three reasons:

1) All the amplitudes α_k vanish at q. Then $Z_n(q) = 0$ for all $n \geq 1$.

2) All the eigenvalues λ_k vanish at q. Then $Z_n(q) = 0$ for all $n \geq 2$.

3) "Mixed case": Neither all the amplitudes nor all the eigenvalues vanish at q, but for each k either α_k or λ_k vanishes at q (or both). Then $Z_n(q) = 0$ for all $n \geq 2$.

As we shall see, the points $q = 0$ and $q = 1$ will be fixed roots belonging to Case 1: all the amplitudes vanish due to an overall prefactor $q(q-1)$. The point $q = 2$ will be a fixed root belonging sometimes to Case 1 and sometimes to Case 3 (and to Case 2 when $m = 2$ for all boundary conditions). For cylindrical strips where the width is not a multiple of 3, the point $q = 3$ will be a fixed root belonging to Case 2 for $m = 4_{\rm P}$ and to Case 3 for $m \geq 5_{\rm P}$. We shall endeavor to explain in each case the mechanism underlying the fixed zeros; these results will be summarized in Section 7.3.

Computation of isolated limiting points. To find the isolated limiting points, we first compute symbolically the determinant $\det D(q)$ defined by (1.2); this determinant is a polynomial in q with integer coefficients, typically of very high degree. We then compute numerically the zeros of this polynomial, using the MPSolve 2.1.1 package [51, 52]; these zeros correspond to points q where at least one amplitude $\alpha_k(q)$ vanishes. Finally, we test numerically each of these zeros to see whether the amplitude corresponding to the *dominant* eigenvalue is vanishing; if it is, then the point in question is an isolated limiting point. This method is guaranteed to discover *all* of the isolated limiting points. We shall not bother to report here all the zeros of $\det D(q)$, but only (a) the isolated limiting points and (b) the Beraha numbers B_n that are zeros of some subdominant amplitude. Please note that whenever B_n is a zero of $\det D(q)$, so are all the primitive generalized Beraha numbers

$$B_n^{(k)} \;=\; 4\cos^2\frac{k\pi}{n} \;=\; 2 + 2\cos\frac{2\pi k}{n} \tag{2.3}$$

where k is relatively prime to n, since they have the same minimal polynomial $p_n(q)$ [16, Section 2.3].

Unfortunately, in some cases the matrix $D(q)$ is so large that we have been unable to compute symbolically its determinant. In these cases, it is more convenient to compute numerically the eigenvalues $\{\lambda_j\}$ and their corresponding amplitudes $\{\alpha_j\}$ and check (a) whether any of the amplitudes vanish and (b) whether the amplitude $\alpha^\star$ associated to the dominant eigenvalue vanishes. We have restricted our search to certain "candidate" values of q (or neighborhoods in the complex q-plane), namely (a) the Beraha numbers B_n for $n \le 50$, and (b) any real or complex values of $q \notin \mathcal{B}$ where zeros of Z_n seem to be accumulating as n gets large. When there is a exact candidate (such as the Beraha numbers B_n), we have computed the amplitudes with high-precision arithmetic (200 digits of precision at least). We considered that an amplitude is zero when its absolute value is less than (for instance) 10^{-190}. When we do not have an exact candidate, we tried to minimize the dominant amplitude $|\alpha^\star|$ around a region where zeros of Z_n tend to accumulate as n grows. This situation occurred only for the strips of widths $8_{\rm Z}$ and $10_{\rm Z}$ ("zig-zag" boundary conditions, see Section 5). The condition $|\alpha^\star| \lesssim 10^{-52}$ holds for all the cases reported here.

In all the cases where we are unable to compute $\det D(q)$ symbolically, we are able to assert that certain points are indeed isolated limiting points, but we cannot claim with confidence that we have found *all* of the isolated limiting points.

3 Numerical Results for the Triangular-Lattice Chromatic Polynomial: Free Boundary Conditions

We have computed the transfer matrix $\mathsf{T}(m_{\rm F})$ and the limiting curves $\mathcal{B}$ for triangular-lattice strips of widths $2 \le m \le 9$ with free boundary conditions in both directions. We also write L_x as a synonym for the strip width m.

As explained in Ref. [16], the chromatic polynomial for this family of strip lattices can be written as

$$Z(m_{\rm F} \times n_{\rm F}) \;=\; \mathbf{u}^{\rm T} \mathsf{H} \mathsf{T}(m_{\rm F})^{n-1} \mathbf{v}_{\rm id} \tag{3.1}$$

where $\mathbf{u}$ and $\mathbf{v}_{\rm id}$ are certain vectors, and $\mathsf{T}(m_{\rm F}) = \mathsf{V}\mathsf{H}$ is the transfer matrix. Here H (resp. V) corresponds to adding one new layer of horizontal (resp. vertical and diagonal) bonds: see Figure 1(a). The matrices H, V and $\mathsf{T}(m_{\rm F})$ act on the space of connectivities of sites in the top layer, whose basis elements $\mathbf{v}_{\mathcal{P}}$ are indexed by partitions $\mathcal{P}$ of the single-layer vertex set $\{1, 2, \ldots, m\}$. In particular, $\mathbf{v}_{\rm id} = \mathbf{v}_{\{\{1\},\{2\},\ldots,\{m\}\}}$. Since the strip lattices we are dealing with are planar, only *non-crossing* partitions $\mathcal{P}$ can occur.

In the particular case of the chromatic polynomial (i.e. the *zero-temperature* antiferromagnet), the horizontal operator H is a projection ($\mathsf{H}^2 = \mathsf{H}$), and we can work in its image subspace by using the modified transfer matrix $\mathsf{T}'(m_{\rm F}) = \mathsf{H}\mathsf{V}\mathsf{H}$ in place

of $\mathsf{T}(m_{\rm F}) = \mathsf{VH}$, and using the basis vectors

$$\mathbf{w}_{\mathcal{P}} \;=\; \mathsf{H}\mathbf{v}_{\mathcal{P}} \tag{3.2}$$

in place of $\mathbf{v}_{\mathcal{P}}$. Then we can rewrite (3.1) as

$$Z(m_{\rm F} \times n_{\rm F}) \;=\; \mathbf{u}^{\rm T}\mathsf{T}'(m_{\rm F})^{n-1}\mathbf{w}_{\rm id} \tag{3.3}$$

where $\mathbf{w}_{\rm id} = \mathsf{H}\mathbf{v}_{\rm id}$. Please note that $\mathbf{w}_{\mathcal{P}} = \mathsf{H}\mathbf{v}_{\mathcal{P}} = 0$ for any partition $\mathcal{P}$ that includes nearest-neighbor sites in the same block, so we can ignore all such partitions. The dimension of the transfer matrix $\mathsf{T}'(m_{\rm F})$ is therefore equal to the number of non-crossing non-nearest-neighbor partitions of the set $\{1, 2, \ldots, m\}$, which is given by the Motzkin number M_{m-1} [16]. To simplify the notation, we will drop the prime in $\mathsf{T}'(m_{\rm F})$ and denote the basis elements $\mathbf{w}_{\mathcal{P}}$ by a shorthand using Kronecker delta functions: for instance, $\mathbf{w}_{\{\{1,3\},\{2\},\{4,6\},\{5\}\}}$ will be written $\delta_{13}\delta_{46}$. We denote the set of basis elements for a given strip as $\mathbf{P} = \{\mathbf{w}_{\mathcal{P}}\}$. For instance, the basis for $m = 3$ is $\mathbf{P} = \{1, \delta_{13}\}$.

We have checked the self-consistency of our finite-lattice results using the trivial identity

$$Z(m_{\rm F} \times n_{\rm F}) \;=\; Z(n_{\rm F} \times m_{\rm F}) \tag{3.4}$$

for all pairs $2 \le m, n \le 9$.

3.1 $L_x = 2_{\rm F}$

This case is trivial, as the transfer matrix is one-dimensional:

$$Z(2_{\rm F} \times n_{\rm F}) = q(q-1)(q-2)^{2(n-1)} \tag{3.5}$$

Since there is only one eigenvalue, there is obviously no crossing, hence $\mathcal{B} = \varnothing$. However, there are zeros for all n at $q = 0, 1$ and for $n \ge 2$ at $q = 2$. The fixed zeros at $q = 0, 1$ arise from a vanishing amplitude, and the fixed zero at $q = 2$ arises from a vanishing eigenvalue.

3.2 $L_x = 3_{\rm F}$

The transfer matrix is two-dimensional. In the basis $\mathbf{P} = \{1, \delta_{13}\}$ it can be written as

$$T(3_{\rm F}) \;=\; \begin{pmatrix} q^3 - 7q^2 + 17q - 14 & q^2 - 6q + 9 \\ -q + 2 & q - 3 \end{pmatrix} \tag{3.6}$$

and the partition function is equal to

$$Z(3_{\rm F} \times n_{\rm F}) \;=\; q(q-1) \begin{pmatrix} q-1 \\ 1 \end{pmatrix}^{\rm T} \cdot T(3_{\rm F})^{n-1} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{3.7}$$

The limiting curve $\mathcal{B}$ (see Figure 2) contains three disconnected pieces and it crosses the real axis at $q_0 \approx 2.5698402910$. There are six endpoints:

$$q \approx 1.2047381150 \pm 1.1596169599\, i \tag{3.8a}$$
$$q \approx 2.3930361082 \pm 0.2538745688\, i \tag{3.8b}$$
$$q \approx 3.4022257768 \pm 0.5865084714\, i \tag{3.8c}$$

These results were previously obtained by Roček *et al.* [24].

The determinant $\det D(q)$ has the form

$$\det D(q) = q^2(q-1)^2(q-2)^3 \tag{3.9}$$

Thus, it vanishes at the first three Beraha numbers $q = 0, 1, 2$. At those points the dominant amplitude vanishes, hence they are isolated limiting points.

In fact, the partition function vanishes at $q = 0, 1$ for all n, and at $q = 2$ for all $n \geq 2$. Obviously, at $q = 0, 1$ both amplitudes vanish, due to the prefactor $q(q-1)$ in (3.7). [This happens for *all* strips of width $m \geq 2$; we will henceforth call these zeros "trivial".] At the fixed zero $q = 2$, there is one nonzero eigenvalue ($\lambda^\star = -1$) with a vanishing amplitude and one zero eigenvalue with a nonvanishing amplitude; we are therefore in Case 3 described in Section 2. The fourth real zero (see Table 1) converges at an approximate $1/n$ rate to the value $q_0 \approx 2.5698402910$.

3.3 $L_x = 4_{\mathrm{F}}$

The transfer matrix is four-dimensional. In the basis $\mathbf{P} = \{1, \delta_{13}, \delta_{24}, \delta_{14}\}$, it takes the form

$$T(4_{\mathrm{F}}) = \begin{pmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ -q^2+5q-6 & q^2-5q+6 & -q+3 & -2(q-3) \\ -q^2+5q-6 & q^2-5q+6 & q^2-6q+9 & q^2-8q+15 \\ q-2 & q^2-5q+6 & -q+3 & q^2-7q+13 \end{pmatrix} \tag{3.10}$$

where

$$T_{11} = q^4 - 10q^3 + 39q^2 - 70q + 48 \tag{3.11a}$$
$$T_{12} = q^3 - 9q^2 + 26q - 24 \tag{3.11b}$$
$$T_{13} = q^3 - 9q^2 + 28q - 30 \tag{3.11c}$$
$$T_{14} = q^3 - 10q^2 + 36q - 45 \tag{3.11d}$$

The partition function is equal to

$$Z(4_{\mathrm{F}} \times n_{\mathrm{F}}) = q(q-1) \begin{pmatrix} (q-1)^2 \\ q-1 \\ q-1 \\ q-2 \end{pmatrix}^{\mathrm{T}} \cdot T(4_{\mathrm{F}})^{n-1} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{3.12}$$

The limiting curve $\mathcal{B}$ (see Figure 3) contains two complex-conjugate disconnected pieces that do not cross the real axis. The closest points to the real axis are $q_0 \approx 2.7592502040 \pm 0.1544431251\,i$. There are six endpoints:

$$q \approx 0.8164709452 \pm 1.2804094073\,i \tag{3.13a}$$
$$q \approx 2.7592502040 \pm 0.1544431251\,i \tag{3.13b}$$
$$q \approx 3.6398304896 \pm 0.5986827987\,i \tag{3.13c}$$

There are T points at $q \approx 3.3341785562 \pm 0.8829730283\,i$. These results were previously obtained by Roček *et al.* [24, 41].

The determinant $\det D(q)$ has the form

$$\det D(q) = -q^4(q-1)^4(q-2)^{13}(q^2-3q+1)(q-3)^6(q^4-11q^3+46q^2-86q+61)^2 \tag{3.14}$$

We recognize, as factors in (3.14), the first five minimal polynomials $p_k(q)$ for the Beraha numbers B_k [16, Table 1]; therefore $\det D(q)$ vanishes at the first five Beraha numbers $q = 0, 1, 2, B_5, 3$. The dominant amplitude vanishes at $q = 0, 1, 2, B_5$, so these are isolated limiting points. At $q = 3$ only two subdominant amplitudes vanish, so this is not an isolated limiting point. Similarly, all of the zeros of the last factor in (3.14) correspond to the vanishing of subdominant amplitudes only, so none of them is an isolated limiting point.

In fact, the partition function vanishes at $q = 0, 1$ for all n, and at $q = 2$ for all $n \geq 2$. The fixed zeros at $q = 0, 1$ are trivial. At $q = 2$, there are three zero eigenvalues and a unique leading eigenvalue $\lambda^\star = 4$ with zero amplitude. Notice that the transfer matrix is not diagonal for $q = 2$: there is a 2×2 nontrivial Jordan block corresponding to $\lambda = 0$ and whose contribution to the partition function is zero for all n. The amplitude corresponding to the other $\lambda = 0$ eigenvalue is 2; we are therefore in Case 3 described in Section 2. Finally, the fourth real zero converges exponentially fast to B_5 (see Table 1).

Please note that for this strip there is a vanishing subdominant amplitude at $q = B_k$ for a Beraha number k *greater than* $m + 1$ (namely, $B_6 = 3$). As we shall see, this occurs frequently for the triangular lattice, and contrasts with the behavior observed for the square lattice [16, 17].

3.4 $L_x = 5_{\mathrm{F}}$

The transfer matrix is nine-dimensional; it can be found in the MATHEMATICA file `transfer3.m` available as part of the electronic version of this paper in the `cond-mat` archive. This strip has been previously studied by Shrock and Tsai [44]; but they did not compute the limiting curve.

The limiting curve $\mathcal{B}$ is connected (see Figure 4). It crosses the real axis at $q_0 = 3$. There are six endpoints:

$$q \approx 0.5586170364 \pm 1.2816149610\,i \tag{3.15a}$$
$$q \approx 3.0474871745 \pm 0.8171660680\,i \tag{3.15b}$$
$$q \approx 3.7782975917 \pm 0.5699779858\,i \tag{3.15c}$$

The topology of the limiting curve is rather involved. It has 12 T points: $q \approx 3.1572589261 \pm 0.7951215102\,i$, $q \approx 3.1251751109 \pm 0.8152460413\,i$, $q \approx 3.2093444343 \pm 0.9296294663\,i$, $q \approx 3.3452062643 \pm 0.9758086833\,i$, $q \approx 3.3248793469 \pm 0.9987588766\,i$, and $q \approx 3.2492362818 \pm 1.1185073809\,i$. These points define four closed regions. The first five T points are the vertices of two complex-conjugate regions which look approximately like rectangular bands. The third, fifth and sixth T points above define two complex-conjugate triangular-like regions.

The determinant $\det D(q)$ is given by

$$\det D(q) \;=\; q^9(q-1)^9(q-2)^{66}(q^2-3q+1)^4(q-3)^{61}P(q)^2 \tag{3.16}$$

where the polynomial $P(q)$ can be found in the MATHEMATICA file `transfer3.m`. The factors appearing in $\det D(q)$ are the first five polynomials $p_k(q)$ given in [16, Table 1]; therefore $\det D(q)$ vanishes at $q = 0, 1, 2, B_5, 3$. The dominant amplitude vanishes at the first four of them (but not at $q = 3$), so that $q = 0, 1, 2, B_5$ are isolated limiting points. All of the zeros of $P(q)$ correspond to the vanishing of subdominant amplitudes only, so none of them is an isolated limiting point.

The first two real zeros $q = 0, 1$ are trivial ones. At $q = 2$ there are two nonzero eigenvalues with zero amplitudes and seven zero eigenvalues. The transfer matrix is not diagonalizable: there are two Jondan blocks of dimension 3 and 2 respectively associated to the eigenvalue $\lambda = 0$. The contribution of these blocks to the partition function vanishes for all n. In addition, the amplitude corresponding to the other two zero eigenvalues are 0 and 2; we are therefore in Case 3 described in Section 2. The fourth real zero converges exponentially fast to the value $q = B_5$ (see Table 1); and the fifth real zero converges at an approximate $1/n$ rate to the value $q_0 = 3$. This agrees with the fact that $q = 3$ is a regular limiting point and not an isolated limiting point.

3.5 $L_x = 6_{\mathrm{F}}$

The transfer matrix is 21-dimensional; it can be found in the MATHEMATICA file `transfer3.m`.

The limiting curve $\mathcal{B}$ is connected (see Figure 5). It crosses the real axis at $q_0 \approx 3.1609256737$. There are six endpoints:

$$q \;\approx\; 0.3796307748 \pm 1.2450702104\,i \tag{3.17a}$$

$$q \;\approx\; 2.9641235697 \pm 1.1179839989\,i \tag{3.17b}$$

$$q \;\approx\; 3.8664092416 \pm 0.5329463088\,i \tag{3.17c}$$

There are four T points at $q \approx 3.3081144403 \pm 1.2171494282\,i$, and $q \approx 3.5005856709 \pm 0.9442298756\,i$.

The determinant $\det D(q)$ is given by

$$\det D(q) \;=\; q^{21}(q-1)^{21}(q-2)^{360}(q^2-3q+1)^{13}(q-3)^{503}(q^3-5q^2+6q-1)P(q)^2 \tag{3.18}$$

where the polynomial $P(q)$ can be found in the MATHEMATICA file `transfer3.m`. The factors appearing in $\det D(q)$ are the first six polynomials $p_k(q)$ given in [16,

Table 1]. Thus, $\det D(q)$ vanishes at $q = 0, 1, 2, B_5, 3, B_7$. The dominant amplitude vanishes at the first five of them (but not at B_7), so that $q = 0, 1, 2, B_5, 3$ are isolated limiting points. All of the zeros of $P(q)$ correspond to the vanishing of subdominant amplitudes only, so none of them is an isolated limiting point.

The first two real zeros $q = 0, 1$ are trivial ones. At $q = 2$ we get six nonzero eigenvalues with zero amplitudes and 15 zero eigenvalues. We find again that the transfer matrix is not diagonalizable for $q = 2$: there are five nontrivial Jordan blocks (one of dimension 3 and four of dimension 2) corresponding to the eigenvalue $\lambda = 0$, and whose contribution to the partition function always vanishes. The amplitudes of the other four zero eigenvalues are (0,0,0,2). This combination seems to be the generic case for free boundary conditions: all amplitudes vanish except one corresponding to a zero eigenvalue. When the transfer matrix is not diagonalizable, then the nontrivial Jordan blocks correspond to $\lambda = 0$ and their contribution is always zero. The fourth and fifth real zeros converge exponentially fast to the values $q = B_5$ and $q = 3$, respectively (see Table 1); the sixth real zero converges at an approximate $1/n$ rate to the value $q_0 \approx 3.1609256737$.

3.6 $L_x = 7_{\mathrm{F}}$

The transfer matrix is 51-dimensional; it can be found in the MATHEMATICA file `transfer3.m`. In this case we have been unable to compute symbolically the resultant, hence the computation of the limiting curve has been performed using the direct-search method.

The limiting curve $\mathcal{B}$ is connected (see Figure 6). It crosses the real axis at $q_0 \approx 3.2764013231$. There are six endpoints:

$$q \approx 0.250538 \pm 1.196864\, i \tag{3.19a}$$

$$q \approx 3.925804 \pm 0.496672\, i \tag{3.19b}$$

$$q \approx 2.878928 \pm 1.343851\, i \tag{3.19c}$$

There are four T points at $q \approx 3.6146786603 \pm 0.9081562491\, i$ and $q \approx 3.2704141478 \pm 1.5194310419\, i$.

We have been unable to compute the determinant $\det D(q)$. However, we computed the amplitudes numerically at each of the Beraha numbers B_n up to B_{50} and determined in particular whether it is an isolated limiting point or not. As always, $q = 0, 1$ are trivial isolated limiting points where all the amplitudes vanish. The dominant amplitude also vanishes at $q = 2, B_5, 3, B_7$, so they are isolated limiting points too. Finally, a subdominant amplitude vanishes at $q = B_8$.

The first two real zeros $q = 0, 1$ are trivial ones. At $q = 2$, all amplitudes vanish except one corresponding to a zero eigenvalue. In particular, there are 12 nonzero eigenvalues with zero amplitudes, 8 zero eigenvalues with zero amplitudes, and one zero eigenvalue with nonzero amplitude. We also find 12 nontrivial Jordan blocks corresponding to $\lambda = 0$, whose contibution to the partition function is always zero. The fourth, fifth and sixth real zeros converge exponentially fast to the values B_5, 3 and B_7, respectively (see Table 2); however, the convergence to $B_7 \approx 3.246979603717$

is slowed by its nearness to the regular limiting point $q_0 \approx 3.2764013231$. For lengths $n \gtrsim 77$, a seventh real zero appears: it converges (at an approximate $1/n$ rate) to q_0.

3.7 $L_x = 8_{\mathrm{F}}$

The transfer matrix is 127-dimensional; it can be found in the MATHEMATICA file `transfer3.m`. Again we have used the direct-search method to locate the points of the limiting curve $\mathcal{B}$.

The limiting curve $\mathcal{B}$ is connected (see Figure 7). It crosses the real axis at $q_0 \approx 3.3610599515$. There are six endpoints:

$$q \approx 0.154432 \pm 1.146669\, i \tag{3.20a}$$

$$q \approx 2.793496 \pm 1.521468\, i \tag{3.20b}$$

$$q \approx 3.967566 \pm 0.463648\, i \tag{3.20c}$$

There are four T points at $q \approx 3.2555859898 \pm 1.7000353877\, i$ and $q \approx 3.6703287722 \pm 0.8845072864\, i$.

We were unable to compute the determinant $\det D(q)$. However, we computed the amplitudes numerically at each of the Beraha numbers B_n up to B_{50} and determined in particular whether it is an isolated limiting point or not. As always, $q = 0, 1$ are trivial isolated limiting points where all the amplitudes vanish. The dominant amplitude vanishes also at $q = 2, B_5, 3, B_7$, so they are isolated limiting points too. Finally, some subdominant amplitudes vanish at $q = B_8, B_9$; they are not isolated limiting points.

The first two real zeros $q = 0, 1$ are trivial ones. At the third real zero $q = 2$, all amplitudes vanish except one corresponding to a zero eigenvalue. At $q = 2$ the transfer matrix is not diagonalizable: we find 30 nontrivial Jordan blocks corresponding to $\lambda = 0$, but their contribution to the partition function vanishes for all $n \geq 1$. We also get 33 nonzero eigenvalues with zero amplitudes, 20 zero eigenvalues with zero amplitudes, and one zero eigenvalue with nonzero amplitude. The fourth, fifth and sixth real zeros converge exponentially fast to the values B_5, 3 and B_7, respectively (see Table 2). We also expect a seventh real zero converging (at an approximate $1/n$ rate) to $q_0 \approx 3.3610599515$. Such zero does not appear up to lengths $n = 96$ (see Table 2). We would need to go to very large n to observe this additional zero.

3.8 $L_x = 9_{\mathrm{F}}$

The transfer matrix is 323-dimensional; it can be found in the MATHEMATICA file `transfer3.m`. The size of this transfer matrix prevented us from computing the limiting curve $\mathcal{B}$. However, we were able to compute the point where the limiting curve crosses the real axis: it is $q_0 \approx 3.4251304673$. In Figure 8 we show the zeros of Z for the finite lattices $9_{\mathrm{F}} \times 45_{\mathrm{F}}$ and $9_{\mathrm{F}} \times 90_{\mathrm{F}}$.

The first two real zeros $q = 0, 1$ are trivial ones. We have checked numerically that the dominant amplitude vanishes at $q = 2, B_5, 3, B_7, B_8$ (and at no other Beraha numbers up to B_{50}); these are therefore isolated limiting points. By inspection of

Figure 8 we do not find any candidate for a non-Beraha real isolated limiting point or for a complex isolated limiting point. A subdominant amplitude vanishes for $q = B_9$ and B_{10}.

At $q = 2$ the transfer matrix is not diagonalizable: we find 76 nontrivial Jordan blocks (up to dimension 5×5) corresponding to $\lambda = 0$; none of these blocks contribute to the partition function for any $n \geq 1$. We also get 74 nonzero eigenvalues with zero amplitudes, 50 zero eigenvalues with zero amplitudes, and one zero eigenvalue with nonzero amplitude. The fourth, fifth, and sixth real zeros converge exponentially fast to the values B_5, 3 and B_7, respectively (see Table 2). We expect that the seventh real zero will converge exponentially fast to $B_8 = 2 + \sqrt{2}$ and that there will be (for large enough length n) an additional real zero that will eventually converge to the value $q \approx 3.4251304673 > B_8$ at an approximate $1/n$ rate. However, we would probably need to go to very large lengths n in order to observe this behavior.

4 Numerical Results for the Triangular-Lattice Chromatic Polynomial: Cylindrical Boundary Conditions

We have also computed the transfer matrix $\mathsf{T}(m_{\mathrm{P}})$ and the limiting curves $\mathcal{B}$ for triangular-lattice strips of widths $2 \leq m \equiv L_x \leq 12$ with cylindrical boundary conditions, i.e. periodic b.c. in the transverse direction and free b.c. in the longitudinal direction.

The partition function can be written analogously to (3.3) as

$$Z(m_{\mathrm{P}} \times n_{\mathrm{F}}) = \mathbf{u}^{\mathrm{T}} \mathsf{T}(m_{\mathrm{P}})^{n-1} \mathbf{w}_{\mathrm{id}} \tag{4.1}$$

where $\mathsf{T}(m_{\mathrm{P}}) = \mathsf{HVH}$. (See [16, Section 3.1] for some remarks about the actual computation of this transfer matrix.) Since $\mathsf{T}(m_{\mathrm{P}})$ commutes with translations (due to the periodic b.c.) and the vectors $\mathbf{u}$ and $\mathbf{w}_{\mathrm{id}}$ are translation-invariant, we can restrict attention to the translation-invariant subspace. The dimension of $\mathsf{T}(m_{\mathrm{P}})$ is therefore given by the number of equivalence classes modulo translation of non-crossing non-nearest-neighbor partitions. This number is denoted by TriCyl(m) in [16, Table 2]. A general analytic expression for TriCyl(m) is not known, although such a formula has been obtained for prime values of m [53, Theorem 3].

We have checked our results for widths $2 \leq m_{\mathrm{P}} \leq 12$ and lengths $n_{\mathrm{F}} = 2, 3, 4$ by comparing to the results of Beraha–Kahane–Weiss [22] (resp. Chang–Shrock [44]) for *width* $n_{\mathrm{F}} = 2$ (resp. $n_{\mathrm{F}} = 3, 4$) and *length* m_{P} (i.e. cyclic boundary conditions), using the trivial identity

$$Z(m_{\mathrm{P}} \times n_{\mathrm{F}}) \;=\; Z(n_{\mathrm{F}} \times m_{\mathrm{P}}) \;. \tag{4.2}$$

This is a highly non-trivial check of the correctness of our results.

The triangular lattice with cylindrical boundary conditions possesses a curious reflection symmetry that we shall now explain. Note first that the triangular lattice is

not invariant under reflection in the transverse direction[7], since reflection changes the direction of the diagonal bonds. Nevertheless, *in the translation-invariant subspace* the transfer matrix does commute with reflection, because by translating the upper row (of a pair of rows) by one unit, one can change $\diagdown\!\diagup$ into $\diagup\!\diagdown$, thereby converting the triangular lattice into a reflected triangular lattice! Because the transfer matrix (in the translation-invariant subspace) commutes with reflection, we can pass to a new basis consisting of connectivities that are either even or odd under reflection.[8] In this new basis, the transfer matrix $\mathsf{T}(m_{\rm P})$ is block-diagonal:

$$\mathsf{T}(m_{\rm P}) \;=\; \begin{pmatrix} \mathsf{T}_+(m_{\rm P}) & 0 \\ 0 & \mathsf{T}_-(m_{\rm P}) \end{pmatrix} \;, \tag{4.3}$$

where T_+ (resp. T_-) corresponds to the reflection-even (resp. reflection-odd) subspace. Furthermore, the reflection-odd components of the start vector $\mathbf{w}_{\rm id} = \mathsf{H}\mathbf{v}_{\rm id}$ are identically zero, since both $\mathbf{v}_{\rm id}$ and H are manifestly reflection-invariant. Likewise, the reflection-odd components of the final vector $\mathbf{u}$ are identically zero, since the definition of $\mathbf{u}$ involves only H and not V. Therefore (for either of these two reasons), the amplitudes $\alpha_k(q)$ corresponding to the reflection-odd subspace are all identically vanishing; these eigenvalues make no contribution whatsoever to the partition function. It follows that we can work entirely within the reflection-invariant subspace, which has dimension SqCyl(m) [16, Table 2]. For strip widths $m \geq 8$, SqCyl(m) is strictly smaller than TriCyl(m), so that the matrix T_- is nontrivial.

4.1 $L_x = 2_{\rm P}$

This case is trivial, as the transfer matrix is one-dimensional:

$$Z(2_{\rm P} \times n_{\rm F}) \;=\; q(q-1)[(q-2)(q-3)]^{n-1} \;. \tag{4.4}$$

Please note that the triangular-lattice strip $2_{\rm P} \times n_{\rm F}$ is *not* equivalent to the strip $2_{\rm F} \times n_{\rm F}$ for any length $n \geq 2$.

Since there is only one eigenvalue, there is obviously no crossing, hence $\mathcal{B} = \varnothing$. However, there are zeros for all n at $q = 0, 1$ and for $n \geq 2$ at $q = 2, 3$.

---

[7] More precisely, if the strip width is even, one can choose to reflect either through a pair of lattice sites or through a pair of bisectors (sites with half-integer coordinates); the two choices differ by a subsequent translation. If the strip width is odd, then every reflection axis passes through one lattice site and one bisector.

[8] Let us call $\{v_j\}_{j=1}^M$ the connectivity basis in the translation-invariant subspace. Some of these basis elements are invariant under reflection; the rest can be grouped into pairs (v_α, v_β) that map into each other under reflection. A basis for the reflection-even (i.e. reflection-invariant) subspace is then given by the basis elements in the first set together with the combinations $v_\alpha + v_\beta$ from the second set. A basis for the reflection-odd subspace is given by the combinations $v_\alpha - v_\beta$ from the second set.

4.2 $L_x = 3_{\rm P}$

This case is also trivial, as the transfer matrix is again one-dimensional:

$$Z(3_{\rm P} \times n_{\rm F}) \;=\; q(q-1)(q-2)(q^3 - 9q^2 + 29q - 32)^{n-1} \,. \tag{4.5}$$

Again $\mathcal{B} = \varnothing$, and the amplitude vanishes at $q = 0, 1, 2$ (which are the first three Beraha numbers). For $n \geq 2$ there are additional fixed zeros at $q \approx 2.5466023485$ and $q \approx 3.2266988258 \pm 1.4677115087\,i$, where the eigenvalue vanishes. This strip was studied by Roček *et al.* [41].

4.3 $L_x = 4_{\rm P}$

The transfer matrix is two-dimensional. In the basis $\mathbf{P} = \{1, \delta_{13} + \delta_{24}\}$ it can be written as

$$T(4_{\rm P}) \;=\; \begin{pmatrix} q^4 - 12q^3 + 58q^2 - 135q + 126 & 2(q^3 - 10q^2 + 33q - 36) \\ -2(q^2 - 7q + 12) & 2(q^2 - 6q + 9) \end{pmatrix} \tag{4.6}$$

and the partition function is equal to

$$Z(4_{\rm P} \times n_{\rm F}) \;=\; q(q-1) \begin{pmatrix} q^2 - 3q + 3 \\ 2(q-1) \end{pmatrix}^{\rm T} \cdot T(4_{\rm P})^{n-1} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{4.7}$$

The limiting curve $\mathcal{B}$ (see Figure 9) contains three pieces: two complex-conjugate arcs and a self-conjugate loop. In addition, at the point $q = 3$ both eigenvalues vanish simultaneously (i.e. the transfer matrix $T(4_{\rm P})$ itself vanishes); this is a special (degenerate) species of isolated limiting point [27]. (One of the amplitudes does not vanish at $q = 3$, but that is irrelevant.) The self-conjugate loop-like component crosses the real axis at $q_0 \approx 3.4814056002$ and $q = 4$. There are four endpoints:

$$q \;\approx\; 1.3705340683 \pm 2.7508526144\,i \tag{4.8a}$$

$$q \;\approx\; 3.6294659317 \pm 0.6958422780\,i \tag{4.8b}$$

This limiting curve was first obtained the pioneering paper of Beraha and Kahane [21]. They drew the important conclusion that $q = 4$ is a limiting point of (complex) chromatic roots for the sequence $4_{\rm P} \times n_{\rm F}$ of *planar* graphs — hence the wonderful title of their paper, "Is the Four-Color Conjecture Almost False?"[9]

The determinant $\det D(q)$ has the form

$$\det D(q) = 8q^2(q-1)^2(q-2)(q^2 - 3q + 1)(q-3)^2(q-4)^2 \tag{4.9}$$

We recognize the first five minimal polynomials $p_k(q)$ given in [16, Table 1]. Hence, the determinant vanishes at the first five Beraha numbers $q = 0, 1, 2, B_5, 3$. It also

[9] Ironically, by the time that this article was published, the Four-Color Conjecture had become the Four-Color Theorem. The Beraha–Kahane article was submitted in 1976 but not published until 1979.

vanishes at $q = 4$, which corresponds to B_∞. The dominant amplitude vanishes only at $q = 0, 1, 2, B_5$; these values of q correspond to isolated limiting points.

The fixed zeros at $q = 0, 1$ are trivial ones. At the fixed zero $q = 2$, there is one nonzero eigenvalue ($\lambda^\star = 10$) with a vanishing amplitude and one zero eigenvalue with a nonvanishing amplitude ($\alpha = 2$); we are therefore in Case 3 described in Section 2. The fourth real zero converges exponentially fast to B_5 (see Table 3), in agreement with the fact that this is an isolated limiting point. The fifth real zero, at $q = 3$, is a fixed zero where both eigenvalues vanish (one of the amplitudes is 0 and the other is 18); we are therefore on Case 2 described in Section 2. The fact that $q = 3$ is a fixed zero is due to the width not being a multiple of 3. Finally, $q = 4$ is a crossing point where both eigenvalues take the value $\lambda = 2$ and one of the amplitudes vanishes. The sixth real zero in Table 3 converges at an approximate $1/n$ rate to the value $q_0 \approx 3.4814056002$.

For this strip there is a vanishing subdominant amplitude at $q = B_6 = 3$, which is *greater than* $B_{m+1} = B_5$, in contrast with the behavior observed for the square lattice [16, 17].

4.4 $L_x = 5_{\rm P}$

The transfer matrix is two-dimensional. In the basis $\mathbf{P} = \{1, \delta_{13} + \text{perm.}\}$ it can be written as

$$T(5_{\rm P}) \;=\; \begin{pmatrix} T_{11} & 5(q^4 - 14q^3 + 76q^2 - 187q + 174) \\ -q^3 + 11q^2 - 43q + 58 & 3q^3 - 35q^2 + 132q - 162 \end{pmatrix} \tag{4.10}$$

where

$$T_{11} \;=\; q^5 - 15q^4 + 95q^3 - 320q^2 + 579q - 452 \;, \tag{4.11}$$

and the partition function is equal to

$$Z(5_{\rm P} \times n_{\rm F}) \;=\; q(q-1)(q-2) \begin{pmatrix} q^2 - 2q + 2 \\ 5(q-1) \end{pmatrix}^{\rm T} \cdot T(5_{\rm P})^{n-1} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{4.12}$$

The limiting curve $\mathcal{B}$ (see Figure 10) contains three disconnected pieces: two complex-conjugate arcs and a self-conjugate loop-like arc. This latter piece crosses the real axis at $q_0 \approx 3.2072219810$ and at $q \approx 3.2847747616$. There are four endpoints:

$$q \;\approx\; 0.4772525688 \pm 2.5694937945\, i \tag{4.13a}$$

$$q \;\approx\; 3.5227474312 \pm 0.1876729035\, i \tag{4.13b}$$

The limiting curve was first obtained in [44].

The determinant $\det D(q)$ has the form

$$\det D(q) = 5q^2(q-1)^2(q-2)^2(q^2 - 3q + 1)(q-3)(q^3 - 11q^2 + 43q - 58)^2 \tag{4.14}$$

We recognize the first five minimal polynomials $p_k(q)$ given in [16, Table 1]. Hence, the determinant vanishes at the first five Beraha numbers $q = 0, 1, 2, B_5, 3$. Furthermore,

the dominant amplitude vanishes at all these points; hence they are isolated limiting points. The last factor of the determinant does not provide additional isolated limiting points.

The fixed zeros at $q = 0, 1, 2$ are trivial ones where all amplitudes vanish due to the prefactor $q(q-1)(q-2)$ in (4.12). The fourth real zero converges exponentially fast to B_5 (see Table 3). The fifth real zero, at $q = 3$, is a fixed zero where there is one nonzero eigenvalue ($\lambda^\star = -2$) with a vanishing amplitude and one zero eigenvalue with a nonvanishing amplitude ($\alpha = 30$); we are therefore in Case 3 described in Section 2. The fact that $q = 3$ is a fixed zero is due to the width not being a multiple of 3. Finally, the sixth real zero in Table 3 converges at an approximate $1/n$ rate to the value $q_0 \approx 3.2072219810$.

4.5 $L_x = 6_{\rm P}$

The transfer matrix is five-dimensional; it can be found in the MATHEMATICA file `transfer3.m`.

The limiting curve $\mathcal{B}$ is connected (see Figure 11). It crosses the real axis at $q_0 \approx 3.2524186216$. There are four endpoints:

$$q \approx 0.0207708231 \pm 2.2756742729\, i \tag{4.15a}$$

$$q \approx 4.2838551928 \pm 0.6111544521\, i \tag{4.15b}$$

There are T points at $q \approx 3.9766954928 \pm 0.9167681670\, i$.

The determinant $\det D(q)$ has the form

$$\begin{aligned} \det D(q) \;=\; & 1769472 q^5 (q-1)^5 (q-2)^6 (q^2-3q+1)^3 (q-3)^{15} (q^3-5q^2+6q-1) \\ & \times (q^2-5q+5)^2 (q-4)^8 P(q)^2 \end{aligned} \tag{4.16}$$

where the polynomial $P(q)$ can be found in the file `transfer3.m`. The first six factors in (4.16) are the first six minimal polynomials given in [16, Table 1]; hence the determinant vanishes at the first six Beraha numbers $q = 0, 1, 2, B_5, 3, B_7$. It also vanishes at the Beraha number B_{10}, whose minimal polynomial is $q^2 - 5q + 5$, and at $q = 4$. The dominant amplitude vanishes only at the first six Beraha numbers, so these values are the only isolated limiting points. The polynomial $P(q)$ does not provide additional isolated limiting points.

The first two zeros $q = 0, 1$ are trivial ones. At $q = 2$ we have three nonzero eigenvalues with vanishing amplitudes, one zero eigenvalue with zero amplitude, and one zero eigenvalue with nonzero amplitude; we therefore fall in Case 3 of Section 2. The fourth and fifth real zeros converge exponentially fast to $q = B_5, 3$ (see Table 3). The next Beraha number $B_7 \approx 3.2469796037$ is very close to the value $q_0 \approx 3.2524186216$ where the limiting curve $\mathcal{B}$ crosses the real axis. This explains why the convergence rate to the sixth real zero in Table 3 is not as fast as expected (empirically the convergence is roughly $\sim n^{-1.9}$); but we expect that it will be ultimately exponential (for very large n). We also expect a seventh real zero for large enough n; this zero is expected to converge (at an approximate $1/n$ rate) to the value $q_0 \approx 3.2524186216$. We would need to go to very large n to observe this additional zero.

Finally, for this strip there is a vanishing subdominant amplitude at $q = B_{10}$, which is *greater than* $B_{m+1} = B_7$, in contrast with the behavior observed for the square lattice [16, 17].

4.6 $L_x = 7_{\rm P}$

The transfer matrix is six-dimensional; it can be found in the MATHEMATICA file `transfer3.m`.

The limiting curve $\mathcal{B}$ is connected (see Figure 12). It crosses the real axis at $q_0 \approx 3.4790022937$ and $q \approx 3.6798199576$. It enters the half-plane $\mathrm{Re}(q) < 0$, and there are four endpoints

$$q \approx -0.2279183274 \pm 2.0134503491\, i \tag{4.17a}$$

$$q \approx \quad 3.9930118897 \pm 0.6273386181\, i \tag{4.17b}$$

There are four T points at $q \approx 3.6222949363 \pm 0.1398555812\, i$ and $q \approx 3.9816630253 \pm 0.8993269516\, i$. Finally, it is worth noticing that the limiting curve encloses a small region around $3.479 \lesssim \mathrm{Re}(q) \lesssim 3.680$ and $|\mathrm{Im}(q)| \lesssim 0.14$.

The determinant $\det D(q)$ has the form

$$\begin{aligned} \det D(q) \;=\; & 68841472 q^6 (q-1)^6 (q-2)^6 (q^2-3q+1)^4 (q-3)^{21} (q^3-5q^2+6q-1) \\ & \times (q^2-4q+2)(q^2-5q+5)^2 P(q)^2 \end{aligned} \tag{4.18}$$

where the polynomial $P(q)$ can be found in the file `transfer3.m`. The first seven factors in (4.18) are precisely the first minimal polynomials given in [16, Table 1]. The next one (q^2-5q+5) is the tenth minimal polynomial in [16, Table 1]. Hence, the determinant $\det D(q)$ vanishes at the Beraha numbers $q = 0, 1, 2, B_5, 3, B_7, B_8, B_{10}$. However, the dominant amplitude vanishes only at the first seven Beraha numbers $q = 0, 1, 2, B_5, 3, B_7, B_8$; these values are the only isolated limiting points for this strip. The polynomial $P(q)$ does not provide additional isolated limiting points.

The first three zeros $q = 0, 1, 2$ are trivial ones. The fifth real zero, at $q = 3$, is a fixed zero where there are 2 nonzero eigenvalues with zero amplitudes, one zero eigenvalue with zero amplitude, and 3 zero eigenvalues with nonzero amplitudes. The fourth, sixth and seventh real zeros converge exponentially fast to $q = B_5, B_7, B_8$ (see Table 3). The eighth real zero seems to converge at an approximate $1/n$ rate to the value $q_0 \approx 3.4790022937$.

In contrast with the behavior observed for the square lattice, we find a vanishing subdominant amplitude at $q = B_{10}$, which is greater than $B_{m+1} = B_8$.

4.7 $L_x = 8_{\rm P}$

The transfer matrix $T(8_{\rm P})$ is 15-dimensional. As discussed at the beginning of this section, the transfer matrix can be brought into block-diagonal form

$$T(8_{\rm P}) \;=\; \begin{pmatrix} T_+(8_{\rm P}) & 0 \\ 0 & T_-(8_{\rm P}) \end{pmatrix} \tag{4.19}$$

where the block T_+ (resp. T_-) is 14-dimensional (resp. 1-dimensional) and corresponds to the subspace of reflection-invariant (resp. reflection-odd) connectivities. Moreover, the amplitude corresponding to the reflection-odd eigenvalue is identically vanishing; this eigenvalue therefore makes no contribution to the partition function. The reduced transfer matrix $T_+(8_P)$ can be found in the MATHEMATICA file `transfer3.m`. The one-dimensional block is $T_-(8_P) = -q^3 + 6q^2 - 8q - 3$.

There are, however, two further curious features for which we have, as yet, no explanation. First of all, we find *another* eigenvalue $\lambda = -q^3 + 6q^2 - 8q - 3$, this time inside the reflection-invariant subspace. Secondly (and even more mysteriously), this eigenvalue too has an identically vanishing amplitude. (We have checked this fact numerically.) The pair of eigenvalues $\lambda = -q^3+6q^2-8q-3$ can be observed explicitly by forming the characteristic polynomial of the transfer matrix, which can be factored as[10]

$$\det[T(8_P) - \lambda \mathbf{1}] \;=\; (\lambda + q^3 - 6q^2 + 8q + 3)^2 \, Q_2(q,\lambda) \tag{4.20}$$

where $Q_2(q,\lambda)$ is a polynomial in q and λ (it is obviously of degree 13 in λ). Unfortunately, we have been unable to find a further change of basis to make $T_+(8_P)$ block-diagonal and thereby bring out explicitly the eigenvalue $\lambda = -q^3+6q^2-8q-3$ lying inside that subspace.

In order to compute the limiting curve $\mathcal{B}$ we have mainly used the resultant method. The existence of a double eigenvalue $\lambda = -q^3 + 6q^2 - 8q - 3$ in the full transfer matrix $T(8_P)$ makes the resultant identically zero for $\theta = 0$. However, this problem does not arise if we consider the reduced matrix $T_+(8_P)$. Nevertheless, the existence of an identically zero amplitude within the reduced subspace makes the computation of the limiting curve $\mathcal{B}$ not completely straightforward, as only those eigenvalues with non-identically-vanishing amplitudes should be taken into account in computing $\mathcal{B}$. A simple solution is devised by noting that the resultant method uses only the characteristic polynomial of the transfer matrix. Therefore, we can drop the factor $(\lambda + q^3 - 6q^2 + 8q + 3)^2$ in (4.20) and compute the resultant using the polynomial $Q_2(q,\lambda)$. In this way, we obtain a nonzero resultant, into which the zero-amplitude eigenvalues do not enter.[11]

The limiting curve $\mathcal{B}$ is connected (see Figure 13). It crosses the real axis at $q_0 \approx 3.5147694243$. It enters the half-plane $\mathrm{Re}(q) < 0$, and there are four endpoints

$$q \;\approx\; -0.3713655472 \pm 1.7983425919\, i \tag{4.21a}$$

[10] In most of the previously studied cases with cylindrical boundary conditions (namely, triangular-lattice strips of widths $4_P \le m \le 7_P$ and square-lattice strips of widths $4_P \le m \le 9_P$), the characteristic polynomial associated to the transfer matrix cannot be factored as in (4.20). In other words, none of the eigenvalues λ is a polynomial in q. The cases with $m \le 3_P$ are trivial as the transfer matrix is one-dimensional: there is a single eigenvalue, which is indeed a polynomial in q.

[11] We also checked — though this is not relevant to computing $\mathcal{B}$ for the boundary conditions being considered here — that the zero-amplitude eigenvalue $\lambda = -q^3 + 6q^2 - 8q - 3$ is not dominant at any of the zeros of our resultant. If the zero-amplitude eigenvalue $\lambda = -q^3 + 6q^2 - 8q - 3$ were in fact dominant somewhere in the complex q-plane, then by modifying the top and bottom endgraphs (as shown in [41]) it might be possible to make that eigenvalue (in either the reflection-even or reflection-odd sector or both) contribute to the chromatic polynomial and thereby obtain a *different* limiting curve $\mathcal{B}$ for the different choice of endgraphs.

$$q \approx 4.0496984440 \pm 0.7359317819\,i \tag{4.21b}$$

There are cusp-like structures around $q \approx 4.04 \pm 0.74\,i$. A closer look shows that these structures are in fact T points located at $q \approx 4.0428606488 \pm 0.7417105390\,i$, from which there emerge very short branches terminating at the endpoints $q \approx 4.0496984440 \pm 0.7359317819\,i$.

We form a matrix $D(q)$ of dimension 13 rather than 15; in this way we can avoid the two identically vanishing amplitudes.[12] Its determinant has the form

$$\begin{aligned} \det D(q) \;=\; & \text{const.} \times q^{13}(q-1)^{13}(q-2)^{24}(q^2-3q+1)^{10}(q-3)^{191} \\ & \times (q^3-5q^2+6q-1)^4 (q^2-4q+2)(q^3-6q^2+9q-1) \\ & \times (q^2-5q+5)^{14}(q^3-7q^2+14q-7)^2(q-4)^{54} \\ & \times (12q^6-196q^5+1355q^4-5126q^3+11337q^2-14086q+7755) \\ & \times P_1(q)^2 \end{aligned} \tag{4.22}$$

where the polynomial $P_1(q)$ can be found in the file `transfer3.m`. The first nine factors in (4.22) are precisely the first minimal polynomials given in [16, Table 1]; therefore, $\det D(q)$ vanishes at the Beraha numbers $q = B_2, \ldots, B_{10}$. The next factor is the square of the polynomial $q^3 - 7q^2 + 14q - 7$, which is $p_{14}(q)$ [16, Table 1], so that $\det D(q)$ also vanishes at $q = B_{14}$. Finally, the determinant also vanishes at $q = 4 = B_\infty$. Unlike what we have seen for cylindrical strips of smaller width, in this case the remaining part of $\det D(q)$ is *not* the square of a polynomial with integer coefficients; rather, there is the additional degree-6 factor preceding $P_1(q)^2$. The dominant amplitude vanishes only at the first seven Beraha numbers $q = 0, 1, 2, B_5, 3, B_7, B_8$, so these values are the only isolated limiting points for this strip. The degree-6 factor and the polynomial $P_1(q)$ do not provide additional isolated limiting points.

The first two zeros $q = 0, 1$ are trivial ones. At $q = 2, 3$, all amplitudes vanish except for a few corresponding to zero eigenvalues; we are thus in Case 3 of Section 2. At $q = 2$, there are 10 nonzero eigenvalues with zero amplitudes, 3 zero eigenvalues with zero amplitudes, and one zero eigenvalue with nonzero amplitude. At $q = 3$ the transfer matrix is not diagonalizable: there are 2 nonzero eigenvalues with zero amplitudes, one 2×2 nontrivial Jordan block corresponding to $\lambda = 0$ with no contribution at all to the partition function for all n, and 10 zero eigenvalues with nonzero amplitudes. The fourth, sixth and seventh real zeros converge exponentially fast to $q = B_5, B_7, B_8$ (see Table 3). The eighth real zero seems to converge at an approximate $1/n$ rate to the value $q_0 \approx 3.5147694243$.

Finally, we find vanishing subdominant amplitudes at $q = B_{10}, B_{14}$, which are greater than $B_{m+1} = B_9$.

---

[12] If we were to form a 15-dimensional or 14-dimensional matrix $D(q)$, its determinant would be identically zero. This tells us (in case we did not know it already) that two of the amplitudes are identically zero.

4.8 $L_x = 9_{\rm P}$

The transfer matrix $T(9_{\rm P})$ is 28-dimensional. It can be brought into block-diagonal form

$$T(9_{\rm P}) \;=\; \begin{pmatrix} T_+(9_{\rm P}) & 0 \\ 0 & T_-(9_{\rm P}) \end{pmatrix} \;, \tag{4.23}$$

where the block T_+ (resp. T_-) is 22-dimensional (resp. 6-dimensional) and corresponds to the subspace of reflection-invariant (resp. reflection-odd) connectivities. Moreover, all the amplitudes corresponding to the reflection-odd subspace are identically vanishing; this subspace therefore makes no contribution to the partition function. The reduced transfer matrix $T_+(9_{\rm P})$ can be found in the MATHEMATICA file `transfer3.m`.

Mysteriously, all of the eigenvalues in the reflection-odd subspace have "copies" in the reflection-even subspace. This can be seen by computing the characteristic polynomial associated to the transfer matrix $T(9_{\rm P})$, which can be factored as follows:

$$\det[T(9_{\rm P}) - \lambda \mathbf{1}] \;=\; Q_1(q,\lambda)^2 \, Q_2(q,\lambda) \tag{4.24}$$

where $Q_1(q,\lambda)$ and $Q_2(q,\lambda)$ are polynomials in q and λ; here Q_1 (resp. Q_2) is of degree 6 (resp. 16) in λ. More specifically, Q_1 (resp. $Q_1 Q_2$) is the characteristic polynomial of T_- (resp. T_+); the fact that Q_1 appears as a factor in the characteristic polynomial of T_+ is direct proof of the just-mentioned "copying" of eigenvalues. Even more mysteriously, our numerical checks suggest that all the eigenvalues coming from Q_1 have identically zero amplitudes — not only in the reflection-odd subspace (where this is well understood) but also in the reflection-invariant subspace. We thus find 6 pairs of equal eigenvalues with identically vanishing amplitudes.

In order to be able to use the resultant method in this case, we proceed as in the previous subsection: we drop the polynomial $Q_1(q,\lambda)^2$ (which contains the zero-amplitude eigenvalues) from the characteristic polynomial associated to $T(9_{\rm P})$ and compute the resultant with the polynomial $Q_2(q,\lambda)$ (which contains the eigenvalues with nonzero amplitudes). We have computed the points with $\theta = 0$ with the resultant method; the points with other values of θ have been computed using the direct-search method.

The limiting curve $\mathcal{B}$ is connected (see Figure 14). It crosses the real axis at $q_0 \approx 3.5270636990$. It enters the half-plane $\mathrm{Re}(q) < 0$, and there are four endpoints

$$q \;\approx\; -0.4576020413 \pm 1.6238415411\, i \tag{4.25a}$$

$$q \;\approx\; 4.2828643197 \pm 0.3823491910\, i \tag{4.25b}$$

There are T points located at $q \approx 4.0160853030 \pm 0.7870153859\, i$.

We form a matrix $D(q)$ of dimension 16 rather than 28, in order to avoid the 12 identically zero amplitudes. Its determinant has the form

$$\begin{aligned} \det D(q) \;=\; & \mathrm{const} \times q^{16}(q-1)^{16}(q-2)^{16}(q^2-3q+1)^{13}(q-3)^{173} \\ & \times (q^3-5q^2+6q-1)^5 (q^2-4q+2)^4 (q^3-6q^2+9q-1) \\ & \times (q^2-5q+5)^7 (q^3-7q^2+14q-7)^2 P(q) \;, \end{aligned} \tag{4.26}$$

where the polynomial $P(q)$ can be found in the file `transfer3.m`. We find the same "Beraha factors" as in the triangular-lattice strip of width $8_{\rm P}$ (Section 4.7). Thus, $\det D(q)$ vanishes at the Beraha numbers $q = B_2, \ldots, B_{10}, B_{14}$. The polynomial $P(q)$ is not the square of any polynomial with integer coefficients; rather it can be written as $P_1(q)P_2(q)^2$ where P_1 and P_2 are integer-coefficient polynomials. The dominant amplitude vanishes only at the first seven Beraha numbers $q = 0, 1, 2, B_5, 3, B_7, B_8$; thus, these values are the only isolated limiting points for this strip. The polynomial $P(q)$ does not provide additional isolated limiting points.

The first three zeros $q = 0, 1, 2$ are trivial ones (i.e., all amplitudes vanish identically). The fourth, fifth, sixth and seventh real zeros converge exponentially fast to $q = B_5, 3, B_7, B_8$ (see Table 4). The eighth real zero seems to converge at an approximate $1/n$ rate to the value $q_0 \approx 3.5270636990$.

We again find a vanishing subdominant amplitude at $q = B_{14}$, which is greater than $B_{m+1} = B_{10}$.

4.9 $L_x = 10_{\rm P}$

The transfer matrix $T(10_{\rm P})$ is 67-dimensional. It can be brought into block-diagonal form

$$T(10_{\rm P}) \;=\; \begin{pmatrix} T_+(10_{\rm P}) & 0 \\ 0 & T_-(10_{\rm P}) \end{pmatrix} \;, \tag{4.27}$$

where the block T_+ (resp. T_-) is 51-dimensional (resp. 16-dimensional) and corresponds to the subspace of reflection-invariant (resp. reflection-odd) connectivities. All the amplitudes corresponding to the reflection-odd subspace are identically vanishing; this subspace therefore makes no contribution to the partition function. The reduced transfer matrix $T_+(10_{\rm P})$ can be found in the MATHEMATICA file `transfer3.m`.

The characteristic polynomial of the transfer matrix $T(10_{\rm P})$ obviously factors as $\det[T(10_{\rm P}) - \lambda \mathbf{1}] = Q_1(q,\lambda)\, Q(q,\lambda)$, where Q_1 (resp. Q) is the characteristic polynomial of T_- (resp. T_+). Numerically we have found, once again, that all the eigenvalues in the reflection-odd subspace have "copies" in the reflection-even subspace. Therefore, the polynomial $Q(q,\lambda)$ should have $Q_1(q,\lambda)$ as a factor, so that

$$\det[T(10_{\rm P}) - \lambda \mathbf{1}] \;=\; Q_1(q,\lambda)^2\, Q_2(q,\lambda) \tag{4.28}$$

where $Q_2(q,\lambda)$ is a polynomial of degree 35 in λ. Unfortunately, we have been unable to compute the characteristic polynomial $Q(q,\lambda)$ of the reduced transfer matrix $T_+(10_{\rm P})$ and verify the conjectured factorization (4.28).

Once again, we have found numerically that the "copied" eigenvalues have identically vanishing amplitudes. We thus find 16 pairs of equal eigenvalues with identically vanishing amplitudes.

We have used the direct-search method in the computation of $\mathcal{B}$. The limiting curve $\mathcal{B}$ is connected (see Figure 15). It crosses the real axis at $q_0 \approx 3.6348299654$. It enters the half-plane $\mathrm{Re}(q) < 0$, and there are four endpoints

$$q \;\approx\; -0.510807 \pm 1.481233\, i \tag{4.29a}$$

$$q \;\approx\; 4.113231 \pm 0.492835\, i \tag{4.29b}$$

There are T points located at $q \approx 4.0632619066 \pm 0.8803786140\,i$.

We have numerically checked that the dominant amplitude vanishes at the first nine Beraha numbers $q = 0, 1, 2, B_5, 3, B_7, B_8, B_9, B_{10}$ (and at no others); therefore, these values are the only isolated limiting points for this strip. We have found no evidence of complex isolated limiting points from the zeros of the finite-length strips.

The first two zeros $q = 0, 1$ are trivial ones. At $q = 2, 3$ we are in Case 3 described in Section 2. At $q = 2$, there are 38 nonzero eigenvalues with zero amplitudes, 12 zero eigenvalues with zero amplitudes and one zero eigenvalue with a nonzero amplitude. At $q = 3$ the transfer matrix is not diagonalizable: we find 4 nonzero eigenvalues with zero amplitudes, one 2×2 nontrivial Jordan block corresponding to $\lambda = -3$ which does not contribute to the partition function for any n, 40 zero eigenvalues with nonzero amplitudes, and two nontrivial Jordan blocks (of dimensions 3 and 2 respectively) corresponding to $\lambda = 0$. The contribution of these later blocks is zero except for $n = 1$. The fourth, sixth, seventh, eight and ninth real zeros converge exponentially fast to $q = B_5, B_7, B_8, B_9, B_{10}$ (see Table 4). The tenth real zero seems to converge at an approximate $1/n$ rate to the value $q_0 \approx 3.6348299654$.

Finally, we have also checked that there are vanishing amplitudes (in addition to the trivial 32 identically zero amplitudes) for $q = B_{11}$, B_{14}, and B_{18}. In all these cases, the vanishing amplitude corresponds to a subdominant eigenvalue; thus, none of these points is an isolated limiting point. Please note that the last two values (namely, B_{14} and B_{18}) are greater than $B_{m+1} = B_{11}$.

4.10 $L_x = 11_{\rm P}$

The transfer matrix $T(11_{\rm P})$ is 145-dimensional. It can be brought into block-diagonal form

$$T(11_{\rm P}) \;=\; \begin{pmatrix} T_+(11_{\rm P}) & 0 \\ 0 & T_-(11_{\rm P}) \end{pmatrix} \;, \tag{4.30}$$

where the block T_+ (resp. T_-) is 95-dimensional (resp. 50-dimensional) and corresponds to the subspace of reflection-invariant (resp. reflection-odd) connectivities. All the amplitudes corresponding to the reflection-odd subspace are identically vanishing; this subspace therefore makes no contribution to the partition function. The reduced transfer matrix $T_+(11_{\rm P})$ can be found in the MATHEMATICA file `transfer3.m`.

The characteristic polynomial of the transfer matrix $T(11_{\rm P})$ obviously factors as $\det[T(11_{\rm P}) - \lambda \mathbf{1}] = Q_1(q,\lambda)\, Q(q,\lambda)$, where Q_1 (resp. Q) is the characteristic polynomial of T_- (resp. T_+). Numerically we have found, once again, that all the eigenvalues in the reflection-odd subspace have "copies" in the reflection-even subspace. Therefore, the polynomial $Q(q,\lambda)$ should have $Q_1(q,\lambda)$ as a factor, so that

$$\det[T(11_{\rm P}) - \lambda \mathbf{1}] \;=\; Q_1(q,\lambda)^2\, Q_2(q,\lambda) \tag{4.31}$$

where $Q_2(q,\lambda)$ is a polynomial of degree 45 in λ. Unfortunately, we have been unable to compute the characteristic polynomial $Q(q,\lambda)$ of the reduced transfer matrix $T_+(11_{\rm P})$ and verify the conjectured factorization (4.31).

Once again, we have found numerically that the "copied" eigenvalues have identically vanishing amplitudes. We thus find 50 pairs of equal eigenvalues with identically vanishing amplitudes.

As in the preceding subsection, we used the direct-search method in the computation of $\mathcal{B}$. This curve crosses the real q-axis at $q_0 \approx 3.6441399017$ (see Figure 16). It enters the half-plane $\mathrm{Re}(q) < 0$, and there are four endpoints

$$q \approx -0.543988 \pm 1.363241\, i \tag{4.32a}$$

$$q \approx 4.156093 \pm 0.529420\, i \tag{4.32b}$$

There are T points located at $q \approx 4.0425923021 \pm 0.6927608569\, i$.

We have numerically checked that the dominant amplitude vanishes at the first nine Beraha numbers $q = 0, 1, 2, B_5, 3, B_7, B_8, B_9, B_{10}$ (and at no others), so that these values are the only real isolated limiting points for this strip. We inspected the zeros of the finite-length strips and found no evidence of complex isolated limiting points.

The first three zeros $q = 0, 1, 2$ are trivial ones, as all amplitude vanish. At $q = 3$ the transfer matrix is not diagonalizable. There are four nontrivial Jordan blocks corresponding to four nonzero eigenvalues. Furthermore, the contribution of these Jordan blocks to the partition function vanishes for all $n \geq 1$. We also find 8 nonzero eigenvalues with zero amplitudes, 30 zero eigenvalues with nonzero amplitudes, and 30 zero eigenvalues with zero amplitudes. The fourth, sixth, seventh, eighth and ninth real zeros converge exponentially fast to $q = B_5, B_7, B_8, B_9, B_{10}$ (see Table 4). The tenth real zero seems to converge at an approximate $1/n$ rate to the value $q_0 \approx 3.6441399017$.

Finally, we have also checked that there are vanishing amplitudes (in addition to the trivial 100 identically zero amplitudes) for $q = B_{11}, B_{12}, B_{14}, B_{18}$. In all these cases, the vanishing amplitude is subdominant; thus, none of these points is an isolated limiting point. Again, the values B_{14} and B_{18} are greater than $B_{m+1} = B_{12}$.

4.11 $L_x = 12_{\mathrm{P}}$

The transfer matrix $T(12_{\mathrm{P}})$ is 368-dimensional. It can be brought into block-diagonal form

$$T(12_{\mathrm{P}}) = \begin{pmatrix} T_+(12_{\mathrm{P}}) & 0 \\ 0 & T_-(12_{\mathrm{P}}) \end{pmatrix} , \tag{4.33}$$

where the block T_+ (resp. T_-) is 232-dimensional (resp. 136-dimensional) and corresponds to the subspace of reflection-invariant (resp. reflection-odd) connectivities. All the amplitudes corresponding to the reflection-odd subspace are identically vanishing; this subspace therefore makes no contribution to the partition function. The reduced transfer matrix $T_+(12_{\mathrm{P}})$ can be found in the MATHEMATICA file `transfer3.m`.

The characteristic polynomial of the transfer matrix $T(12_{\mathrm{P}})$ obviously factors as $\det[T(12_{\mathrm{P}}) - \lambda \mathbf{1}] = Q_1(q,\lambda)\, Q(q,\lambda)$, where Q_1 (resp. Q) is the characteristic polynomial of T_- (resp. T_+). Numerically we have found, once again, that all the eigenvalues

in the reflection-odd subspace have "copies" in the reflection-even subspace. Therefore, the polynomial $Q(q,\lambda)$ should have $Q_1(q,\lambda)$ as a factor, so that

$$\det[T(12_{\rm P}) - \lambda \mathbf{1}] \;=\; Q_1(q,\lambda)^2\, Q_2(q,\lambda) \tag{4.34}$$

where $Q_2(q,\lambda)$ is a polynomial of degree 96 in λ. Unfortunately, we have been unable to compute the characteristic polynomial $Q(q,\lambda)$ of the reduced transfer matrix $T_+(12_{\rm P})$ and verify the conjectured factorization (4.34).

Once again, we have found numerically that the "copied" eigenvalues have identically vanishing amplitudes. We thus find 136 pairs of equal eigenvalues with identically vanishing amplitudes.

Due to the large dimension of the transfer matrix, we have been unable to compute the limiting curve. However, we have managed using the direct-search method to compute the point where $\mathcal{B}$ crosses the real q-axis: $q_0 \approx 3.6431658979$. We have also computed the position of the pair of complex-conjugate T points that are obvious in Figure 16; the result is $q \approx 4.05713658 \pm 0.73432479\,i$.

We have numerically checked that the dominant amplitude vanishes at the first nine Beraha numbers $q = 0, 1, 2, B_5, 3, B_7, B_8, B_9$, and B_{10}. These values are the only isolated limiting points for this strip. We have found no evidence of complex isolated limiting points by inspecting the zeros of the finite-length strips.

The first two zeros $q = 0, 1$ are trivial ones. At $q = 2$ there are 164 nonzero eigenvalues with zero amplitudes, 2 zero eigenvalues with nonzero amplitudes, and 66 zero eigenvalues with zero amplitudes. The convergence to $q = B_5, 3, B_7$, B_8, B_9, B_{10} is exponentially fast (see Table 4). The tenth real zero seems to converge at an approximate $1/n$ rate to the value $q_0 \approx 3.6431658979$.

Finally, we have also checked that there are vanishing amplitudes (in addition to the trivial 272 identically zero amplitudes) for $q = B_{11}, B_{12}, B_{13}, B_{14}, B_{18}, B_{22}$. In all these cases, the vanishing amplitude is subdominant; thus, none of these points is an isolated limiting point.

In this case we find three vanishing subdominant amplitudes (B_{14}, B_{18}, and B_{22}) that are greater than the value $B_{m+1} = B_{13}$.

5 Numerical Results for the Triangular-Lattice Chromatic Polynomial: Zig–Zag Boundary Conditions

Until now, we have built up the triangular lattice by transferring along a direction that is perpendicular to one of the principal directions of the lattice. An alternative choice, of course, would be to transfer along a direction that is *parallel* to a principal direction. When periodic boundary conditions are imposed across the strip, these two constructions are inequivalent: they yield different finite graphs. We shall refer to this alternative construction, with periodic boundary conditions in the transversal direction and free boundary conditions in the longitudinal direction, as "zig–zag"

boundary conditions, and it will be denoted by the subscript Z. Note that the lattice width m must be *even*.

For zig–zag boundary conditions, the transfer matrix is not given by the formulae of Ref. [16]. Rather, as is evident from Figure 1(b), the transfer matrix now takes the following form

$$\mathsf{T}(m_{\rm Z}) = \mathsf{H}\mathsf{V}_{\rm even}\mathsf{H}\mathsf{V}_{\rm odd} \;, \tag{5.1}$$

where $\mathsf{V}_{\rm even}$ (resp. $\mathsf{V}_{\rm odd}$) is the product of the matrices associated to the vertical bonds located at even (resp. odd) sites of the lattice. With this definition, all the formulae applied in the previous sections hold.

Our original motivation for introducing this new construction was the following: It is clear from Sections 3 and 4 that the limiting curves for the strips with free and cylindrical boundary conditions differ qualitatively by the existence of a small additional inward-pointing branch for the case of free b.c., which is absent for cylindrical b.c. In the limit of infinite width, one might wonder whether this branch extends to $q = 2$, as the triangular-lattice Ising model is known to have a zero-temperature critical point [32]. We found it interesting to examine whether we can recover such a branch by imposing zig–zag boundary conditions. The answer turns out to be negative; but it seems to us that zig–zag b.c. are interesting in their own right, irrespective of this initial motivation.

We have computed the transfer matrix $T(m_{\rm Z})$ and the limiting curves $\mathcal{B}$ for triangular-lattice strips of even widths $2 \le m \equiv L_x \le 10$. It is interesting to note that the trick discussed in [16, Sections 3.1 and 3.3] for the standard construction of a cylindrical triangular-lattice strip is not necessary here. On the other hand, the dimension of the transfer matrix $T(m_{\rm Z})$ is in general different from TriCyl(m), because the invariances are different: for zig-zag boundary conditions, the system is invariant under translations of *even* (but not odd) length and under reflections.

5.1 $L_x = 2_{\rm Z}$

This case is trivial, as the transfer matrix is one-dimensional:

$$Z(2_{\rm Z} \times n_{\rm F}) = q(q-1)(q-2)^{2(n-1)} \tag{5.2}$$

Please note that the strip $2_{\rm Z} \times n_{\rm F}$ is equivalent to $2_{\rm F} \times n_{\rm F}$. Since there is only one eigenvalue, there is obviously no crossing, hence $\mathcal{B} = \varnothing$. However, there are zeros for all n at $q = 0, 1$ and for $n \ge 2$ at $q = 2$.

5.2 $L_x = 4_{\rm Z}$

The transfer matrix is three-dimensional. In the basis $\mathbf{P} = \{1, \delta_{13}, \delta_{24}\}$ it can be written as

$$T(4_{\rm Z}) \;=\; \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ 2q-5 & q^2-4q+4 & 1 \\ T_{31} & q^2-4q+4 & 3q^2-17q+25 \end{pmatrix} \tag{5.3}$$

where

$$T_{11} = q^4 - 12q^3 + 56q^2 - 121q + 101 \tag{5.4a}$$
$$T_{12} = q^3 - 8q^2 + 20q - 16 \tag{5.4b}$$
$$T_{13} = q^3 - 10q^2 + 34q - 40 \tag{5.4c}$$
$$T_{31} = 2q^3 - 17q^2 + 50q - 50 \tag{5.4d}$$

and the partition function is equal to

$$Z(4_{\rm P} \times n_{\rm Z}) = q(q-1) \begin{pmatrix} q^2-3q+3 \\ q-1 \\ q-1 \end{pmatrix}^{\rm T} \cdot T(4_{\rm Z})^{n-1} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \tag{5.5}$$

The limiting curve $\mathcal{B}$ (see Figure 18) contains two complex-conjugate arcs, which do not cross the real q-axis. There are four endpoints:

$$q \approx 2.0991442518 \pm 2.5589234827\, i \tag{5.6a}$$
$$q \approx 2.7371672817 \pm 0.1723332852\, i \tag{5.6b}$$

We have found a complex-conjugate pair of double zeros of the resultant for $\theta = 0$ (see [16, Section 4.1.1]) at $q \approx 3.7718445063 \pm 1.1151425080\, i$. At these values the limiting curve appears at first glance to be singular (see Figure 18). However, a closer look reveals that this is *not* the case; in fact, the limiting curve is perfectly analytic around these two points.[13]

The determinant $\det D(q)$ has the form

$$\begin{aligned} \det D(q) &= -q^3(q-1)^3(q-2)^2(q^2-3q+1)(q-3)^2(2q-5)^4 \\ &\quad \times (q^3 - 10q^2 + 34q - 38)^2 \end{aligned} \tag{5.7}$$

We recognize the first five minimal polynomials $p_k(q)$ given in [16, Table 1]. Hence, the determinant vanishes at the first five Beraha numbers $q = 0, 1, 2, B_5, 3$. The dominant amplitude vanishes at all these points except at $q = 3$; therefore, $q = 0, 1, 2, B_5$ are isolated limiting points. It is interesting to note that the dominant amplitude vanishes also at $q = 5/2$, so that this too is an isolated limiting point. This is the first time we have found a real isolated limiting point that is not a Beraha number. (For the square lattice, we did not find any non-Beraha real isolated limiting point [16, 17].) There are no additional isolated limiting points coming from the last factor $q^3 - 10q^2 + 34q - 38$, as the vanishing amplitudes in question all correspond to subdominant eigenvalues.

[13] In the notation of [16, Section 4.2], the characteristic polynomial of $T(4_{\rm Z})$ can be expanded as

$$P(\lambda, q) = c(\lambda - \lambda_0)^2 + d(q - q_0)^2 + e(q - q_0)(\lambda - \lambda_0) + \ldots$$

around the points $q_0 \approx 3.7718445063 \pm 1.1151425080\, i$ and the dominant (double) eigenvalue $\lambda_0 = -0.5237532362 + 4.5580089825\, i$ (the linear terms in $\lambda - \lambda_0$ and $q - q_0$ vanish). This expansion leads to analytic eigenvalues $\lambda_\pm(q)$ around $q = q_0$ and to an analytic equimodular locus around $q = q_0$.

The first two real zeros $q = 0, 1$ are trivial as all the amplitudes vanish. The third real zero $q = 2$ belongs to Case 3 of Section 2: the two nonzero eigenvalues have zero amplitude, and there is an additional zero eigenvalue with a nonzero amplitude. Finally, the fourth and fifth real zeros converge exponentially fast to the values $5/2$ and B_5, respectively. In summary, we find five isolated limiting points $q = 0, 1, 2, 5/2$ and B_5. This is in agreement with Table 5.

Please note that there is a vanishing subdominant amplitude at $q = B_6 = 3$. This value is greater than $B_{m+1} = B_5$, in contrast with the observed behavior for the square lattice [16, 17].

5.3 $L_x = 6_{\mathrm{Z}}$

The transfer matrix is seven-dimensional; it can be found in the MATHEMATICA file `transfer3.m`.

The limiting curve $\mathcal{B}$ is connected (see Figure 19). It crosses the real axis at $q \approx 3.1752579126$. There are four endpoints:

$$q \approx 0.3618461880 \pm 2.5093731708\, i \tag{5.8a}$$

$$q \approx 4.2589504182 \pm 0.7015734543\, i \tag{5.8b}$$

There are T points at $q \approx 3.8395346820 \pm 1.1149959335\, i$.

The determinant $\det D(q)$ has the form

$$\det D(q) = 81q^7(q-1)^7(q-2)^{18}(q^2-3q+1)^4(q-3)^{44}(q^3-5q^2+6q-1)P(q)^2 \tag{5.9}$$

where the polynomial $P(q)$ can be found in the file `transfer3.m`. The first six polynomial are the first six minimal polynomials given in [16, Table 1]; hence the determinant vanishes at the first six Beraha numbers $q = 0, 1, 2, B_5, 3, B_7$. The dominant amplitude vanishes at the first five Beraha numbers $q = 0, 1, 2, B_5, 3$ as well as at three of the zeros of $P(q)$, namely $q \approx 2.7226328355$ and $q \approx 3.6696077451 \pm 0.9506864736\, i$. This is the first triangular-lattice strip where we find complex isolated limiting points. In the square-lattice case, complex isolated limiting points were quite common: we found such limiting points for $L \geq 6$ with both free and cylindrical boundary conditions [16, 17].

The first two real zeros $q = 0, 1$ are trivial ones. The third real zero $q = 2$ falls in Case 3 of Section 2: there are four nonzero eigenvalues with zero amplitudes, two zero eigenvalues with nonzero amplitudes, and one zero eigenvalue with zero amplitude. The convergence of the fourth, fifth and sixth real zeros to their corresponding limiting values (namely, $B_5, 2.7226328355$, and 3) is exponentially fast, as shown in Table 5. Finally, the seventh real zero converges at an approximate $1/n$ rate to the value $q \approx 3.1752579126$.

In summary, we find six real isolated limiting points at $q = 0, 1, 2, B_5, 2.7226328355, 3$ and B_7 (see Table 5) and two complex isolated limiting points at $q \approx 3.6696077451 \pm 0.9506864736\, i$. However, since the complex isolated limiting points are very near $\mathcal{B}$, it is very difficult to observe the convergence to them as distinct from the convergence to $\mathcal{B}$ (see Figure 19).

5.4 $L_x = 8_{\rm Z}$

The transfer matrix is 24-dimensional; it can be found in the MATHEMATICA file `transfer3.m`.

The limiting curve $\mathcal{B}$ is connected (see Figure 20). It crosses the real axis at $q \approx 3.3941047539$. There are four endpoints:

$$q \approx -0.2143469947 \pm 2.0301412598\, i \tag{5.10a}$$

$$q \approx 4.2899063418 \pm 0.5046183096\, i \tag{5.10b}$$

There are T points at $q \approx 4.0055796610 \pm 0.8830638824\, i$.

In this case we were unable to compute the determinant $\det D(q)$. However, we checked numerically whether any of the amplitudes vanishes at the Beraha numbers B_n, and if this occurs, whether the vanishing amplitude is the leading one or not. We have made this check up to $n = 50$. We have found that at least one amplitude vanishes at the Beraha numbers $q = 0, 1, 2, B_5, 3, B_7, B_8, B_9$. The dominant amplitude vanishes only at the first six (namely, $q = 0, 1, 2, B_5, 3, B_7$), so that these latter numbers are isolated limiting points (see Table 5). In Table 5 we also notice an additional isolated zero at $q \approx 2.8214204955$. We have numerically confirmed that this point is indeed an isolated limiting point by minimizing the absolute value of the dominant amplitude $\alpha^\star$ in a neighborhood of that point.

The first two real zeros $q = 0, 1$ are trivial ones. At $q = 2$, there are 13 nonzero eigenvalues with zero amplitudes, 2 zero eigenvalues with nonzero amplitudes, and 9 zero eigenvalues with zero amplitudes. Finally, the other real zeros in Table 5 converge exponentially fast to their corresponding limiting values (namely, $B_5, 2.8214204955, 3$, and B_7). We expect, for sufficiently larger lengths n, an additional real zero larger than B_7 and converging to $q \approx 3.3941047539$; but we apparently need to go beyond $n = 80$ to see it. In summary, there are seven real isolated limiting points at $q = 0$, 1, 2, B_5, 2.8214204955, 3, and B_7.

By minimizing the absolute value of the dominant amplitude $\alpha^\star$, we have found a pair of complex-conjugate isolated limiting points at $q \approx 3.8327415674 \pm 0.73050211595\, i$ (See Figure 20). Again, we are not sure that we have found all complex isolated limiting points for this strip.

5.5 $L_x = 10_{\rm Z}$

The transfer matrix is 87-dimensional; it can be found in the MATHEMATICA file `transfer3.m`.

In this case we were unable to compute the limiting curve. However, we were able to compute the value of q where that curve crosses the real axis: $q \approx 3.5204366907$.

The matrix $D(q)$ is too large for us to compute its determinant. Instead, we have checked numerically the eigenvalues and amplitudes of the transfer matrix at the Beraha numbers B_n with $2 \le n \le 50$. We have found that at least one amplitude vanishes at the Beraha numbers $q = 0, 1, 2, B_5, 3, B_7, B_8, B_9, B_{10}, B_{11}$. The dominant amplitude vanishes only at the first seven (namely, $q = 0, 1, 2, B_5, 3, B_7, B_8$), so that these latter numbers are isolated limiting points (see Table 5). In Table 5 we also

notice two additional isolated zeros at $q \approx 2.8737312493$ and $q \approx 3.3831285312$. We have numerically checked that in both cases the leading amplitudes vanish, so they too are isolated limiting points.

The first two real zeros $q = 0, 1$ are trivial. At $q = 2$, there are 35 nonzero eigenvalues with zero amplitudes, 31 zero eigenvalues with zero amplitudes, and 21 zero eigenvalues with nonzero emplitudes. The convergence of the fourth through ninth real zeros to their corresponding limiting values (namely, $B_5, 2.8737312493, 3, B_7, 3.3831285312, B_8$) is exponentially fast as shown in Table 5. Finally, the tenth real zero converges at an approximate $1/n$ rate to the value $q \approx 3.5204366907$.

In summary, there are nine real isolated limiting points at $q = 0$, 1, 2, B_5, 2.8737312493, 3, B_7, 3.3831285312, and B_8. We have been unable to say whether or not there are any complex isolated limiting points; but we do not see any candidate in Figure 21.

6 Thermodynamic Limit

In this section we will review the Bethe-Ansatz solution found by Baxter [18, 19] for the thermodynamic limit of the zero-temperature triangular-lattice Potts antiferromagnet, and carefully recalculate the limiting curve $\mathcal{B}_\infty$ where the chromatic roots are expected to accumulate. The resulting picture will be substantially similar to that set forth by Baxter [19], but with a few important qualitative differences.

6.1 Baxter's solution

In terms of the variables x and θ defined by

$$q \;=\; 2 - x - x^{-1} \;=\; 2 + 2\cos\theta \tag{6.1}$$

with $|x| < 1$ and $0 < \mathrm{Re}\,\theta < \pi$, Baxter defined three functions (eigenvalues) $g_i(q)$ as follows:

$$g_1(q) \;=\; -\frac{1}{x}\prod_{j=1}^{\infty} \frac{(1-x^{6j-3})(1-x^{6j-2})^2(1-x^{6j-1})}{(1-x^{6j-5})(1-x^{6j-4})(1-x^{6j})(1-x^{6j+1})} \tag{6.2a}$$

$$\log g_2(q) \;=\; \int_{-\infty}^{\infty} \mathrm{d}k\, \frac{\sinh k\theta}{2k} \left(\frac{\sinh[k(\pi-2\theta)/2]}{(2\cosh k\theta - 1)\sinh(\pi k/2)} - \frac{\cosh[k(\pi-2\theta)/2]}{(2\cosh k\theta + 1)\cosh(\pi k/2)} \right) \tag{6.2b}$$

$$\log g_3(q) \;=\; \int_{-\infty}^{\infty} \mathrm{d}k\, \frac{\sinh k\theta\,[\sinh k(\pi-\theta)]}{k \sinh \pi k\,[2\cosh k(\pi-\theta) - 1]} \tag{6.2c}$$

These formulae were obtained in [18], but the corresponding ranges of validity were established only in [19]. In particular, Baxter found that the complex q-plane can be divided into three domains $\mathcal{D}_i$ $[i = 1, 2, 3]$, in each of which the dominant eigenvalue

is g_i. According to Baxter [19], the intersections of these regions with the real axis are as follows:

$$\mathcal{D}_1 \cap \mathbb{R} = \{q > 4\} \cup \{q < 0\} \tag{6.3a}$$
$$\mathcal{D}_2 \cap \mathbb{R} = \{q_0 < q < 4\} \tag{6.3b}$$
$$\mathcal{D}_3 \cap \mathbb{R} = \{0 < q < q_0\} \tag{6.3c}$$

Baxter therefore determined the parameter q_0 by solving the equation

$$g_2(q_0) = g_3(q_0) \tag{6.4}$$

via Newton's method and numerical integration of (6.2b)/(6.2c), and found $q_0 \approx 3.81967$. We have refined this computation using the same method, and find

$$q_0(\text{Baxter}) \approx 3.819671731239719\ . \tag{6.5}$$

This point is labelled F in Figures 24 and 26 ($\theta_{\text{F}} \approx 0.427907971348122$). We have also plotted the eigenvalues g_2 and g_3 over the entire range $0 < \theta < \pi$ and verified that there is only one crossing point, namely (6.5). However, as we will argue later, the intersections $\mathcal{D}_i \cap \mathbb{R}$ are in fact more complicated than what is claimed in (6.3), so that (6.5) is *not* in fact the correct value of q_0.

In order to obtain the limiting curve $\mathcal{B}_\infty$ in the complex q-plane, Baxter [19] took advantage of the following simpler expressions for the ratios of eigenvalues:

$$\frac{g_2(q)}{g_1(q)} = \prod_{j=1}^{\infty} \left(\frac{1 - \omega p^{2j-1}}{1 - \omega p^{2j}} \right)^3 \frac{1 - p^{6j}}{1 - p^{6j-3}} \tag{6.6a}$$

$$\frac{g_3(q)}{g_1(q)} = \prod_{j=1}^{\infty} \left(\frac{1 + \omega^2 y^j}{1 - \omega^2 y^j} \right)^3 \frac{1 - y^{3j}}{1 + y^{3j}} \tag{6.6b}$$

where

$$p = -e^{i\pi^2/3\theta} \tag{6.7a}$$
$$y = e^{-2i\pi^2/3(\pi-\theta)} \tag{6.7b}$$
$$\omega = e^{2\pi i/3} \tag{6.7c}$$

In these equations we require that $|p| < 1$ and $|y| < 1$ so that the products converge; this corresponds to $\operatorname{Im} \theta < 0$.

As Baxter [19] noted, the parameter θ enters into the products (6.6a,b) only via p or y, respectively, and these two variables are invariant under the transformations

$$\frac{\pi}{\theta} \to \frac{\pi}{\theta} + 6k \qquad \Longrightarrow \qquad p \to p \tag{6.8a}$$

$$\frac{\pi}{\pi - \theta} \to \frac{\pi}{\pi - \theta} + 3k' \qquad \Longrightarrow \qquad y \to y \tag{6.8b}$$

for any integers k, k'. Thus, each solution of $|g_2/g_1| = 1$ in the complex p-plane corresponds to an infinite family of solutions in the complex θ-plane: these can be thought of as a "primary" solution (namely, the one with largest $\mathrm{Re}\,\theta$ contained in the physical region $0 < \mathrm{Re}\,\theta < \pi$) and its "images" under the transformation (6.8a) with $k \geq 1$. As $k \to \infty$ these "image" curves converge to $\theta = 0$ ($q = 4$). Likewise, each solution of $|g_3/g_1| = 1$ in the complex y-plane corresponds to an infinite family of solutions in the complex θ-plane: a "primary" solution (the one with smallest $\mathrm{Re}\,\theta$ contained in the physical region) and its "images" under the transformation (6.8b) with $k' \geq 1$. As $k' \to \infty$ these "image" curves converge to $\theta = \pi$ ($q = 0$). It is important to note that the transformations (6.8) do *not* tell anything about the dominant or subdominant character of the equimodular curve at the transformed value of θ; this property has to be checked by other means.

Important Remark. Neither the eigenvalues (6.2b,c) nor the eigenvalue ratios (6.6a,b) or $g_3/g_1 = (6.6a)/(6.6b)$ — nor even their absolute values — are invariant under the transformation $\theta \to \theta + 2\pi k''$ that leaves q invariant. Therefore, we *must* require $0 < \mathrm{Re}\,\theta < \pi$ when using these formulae.

Remarks. 1. The formula $q = 2 - x - x^{-1}$ maps the disc $|x| < 1$ one-to-one onto the q-plane cut along the interval $[0, 4]$. Therefore (as Baxter observed [19]), when q is real and $q > 4$ or $q < 0$, the parameter x is also real, as is the eigenvalue g_1 defined in (6.2a).

2. The formula $q = 2 + 2\cos\theta$ maps the strip $0 < \mathrm{Re}\,\theta < \pi$ one-to-one onto the q-plane cut along the intervals $(-\infty, 0]$ and $[4, \infty)$. In particular, when q is real and $0 < q < 4$, the parameter θ is real ($0 < \theta < \pi$), as are the eigenvalues g_2 and g_3 defined in (6.2b)/(6.2c).

3. The definitions (6.2a), (6.6a) and(6.6b) cannot be applied directly *on* the real q-axis for $0 < q < 4$; rather, one must consider a limit in which $\mathrm{Im}\,\theta \uparrow 0$ and hence $|x|, |p|, |y| \uparrow 1$.

6.2 Computation of limiting curves

Our goal is to compute the locus of points where two or more of the eigenvalues g_i are equimodular, and to determine at each such point whether these equimodular eigenvalues are dominant or subdominant. We carry out this procedure as follows:

1) We use (6.6a) to compute the locus $|g_2/g_1| = 1$ in the complex p-plane (Figure 22); we then transform the resulting plot to the complex θ-plane using (6.7a). As noted above, each curve in the p-plane corresponds to an infinite family of curves in the θ-plane. Along each of these latter curves, we compute $|g_3/g_1|$ and classify the curve (or portions of it) as dominant or subdominant.

2) We use (6.6b) to compute the locus $|g_3/g_1| = 1$ in the complex y-plane (Figure 23); we then transform the resulting plot to the complex θ-plane using (6.7b). Each curve in the y-plane corresponds to an infinite family of curves in the θ-plane. Along each of these latter curves, we compute $|g_2/g_1|$ and classify the curve (or portions of it) as dominant or subdominant.

3) We use (6.6a,b) to compute the locus $|g_3/g_2| = 1$ directly in the complex θ-plane and to determine dominance or subdominance.

4) We combine the three families of equimodular curves into a single θ-plane plot (Figure 24).

5) Finally, we transform the resulting curves to the complex q-plane using (6.1). The resulting "phase diagram" is shown in Figure 25; a detailed view near the point $q = 4$ is shown in Figure 26.

Despite the explicit formulae (6.6), these computations are far from straightforward, due to the slow convergence of the products when $|p|$ or $|y|$ is near 1 (i.e. when q is near the interval $0 < q < 4$ of the real axis) and to the need for very high numerical precision in intermediate stages of the computation. We discuss these technical points in the Appendix.

The equimodular curves $|g_2/g_1| = 1$ in the complex p-plane are shown in Figure 22. Each equimodular curve C_n has exactly two endpoints. As we approach the circle $|p| = 1$, more smaller equimodular curves appear. In order to disentangle the larger curves from these new smaller curves, we have followed each equimodular curve carefully as it approaches the $|p| = 1$ limit. In Figure 22 we have shown all equimodular curves that intersect the circle $|p| = 0.99$.

The principal feature is a curve C_1 running from point G ($p = -1$, $\theta = \pi/6$ and images) via the origin ($p = 0$, $\theta = -i0$) to point H ($p = -i$, $\theta = 2\pi/3$ and images). The next-longest curve (C_2) runs from point I ($p = i$, $\theta = 2\pi/9$ and images) to point J ($p = e^{3\pi i/7}$, $\theta = 7\pi/30$ and images). The third-longest curve (C_3) runs from point K ($p = e^{-\pi i/5}$, $\theta = 5\pi/12$ and images) to point L ($p = e^{-\pi i/4}$, $\theta = 4\pi/9$ and images). Some further equimodular curves and their endpoints are shown on Figure 22 and enumerated in Table 8. In this table we have first shown the curves C_n for which both endpoints are well-determined (see below). Then we have listed some other well-determined endpoints whose counterparts could not be estimated with sufficient accuracy; these points are grouped into the category "Others".

It is curious that all these endpoints appear to lie at $p = e^{i\varphi}$ where φ is a rational multiple of π (with a small denominator). In order to test this conjecture, we have performed detailed fits as follows: For each endpoint we first obtained ten nearby equimodular points p with $|p| = 0.990, 0.991, \ldots, 0.999$. Then we performed a least-squares fit of the data[14] using the polynomial Ansatz

$$\frac{\operatorname{Arg} p}{\pi} \;=\; \frac{\operatorname{Arg} p_{\mathrm{Fit}}}{\pi} + \sum_{k=1}^{8} a_k (1 - |p|)^k \tag{6.9}$$

in order to estimate the phase $\varphi = \operatorname{Arg} p_{\mathrm{Fit}}$ at the endpoint. If the equimodular curve is smooth close to $|p| = 1$, this Ansatz is expected to work well. We have chosen an

[14] We have made the computations with data truncated to eight decimal digits, hence the error of the input data is 10^{-8}. Note, however, that the deviations from (6.9) are *not* statistical flcutuations; rather, they are "corrections to scaling", i.e. due to neglected higher-order terms in (6.9). A large value of the χ^2 is thus a signal that we need to include higher-order terms in our Ansatz.

eighth-order polynomial in order to take into account as many data points as possible while allowing a little freedom (we have one degree of freedom in the fits). As a check, we have repeated this computation with lower-degree polynomial Ansätze and dropping the data with the smallest values of $|p|$ (in order to have at least one degree of freedom in the fit). We have used the stability of the estimates for $\mathrm{Arg}\, p_{\mathrm{Fit}}$ as a guideline to decide whether a fit is good or not (see below).

We next asked whether the estimated value of $\mathrm{Arg}\, p_{\mathrm{Fit}}/\pi$ is or is not close to a rational number with a small denominator. We have used the following criterion: the real number x is "close" to the rational number m/n whenever the

$$\text{"discrepancy"} \;\equiv\; |nx - m| \tag{6.10}$$

is sufficiently small. Since every real number x and integer n have a "discrepancy" of at most $1/2$ (for a suitable choice of m), one must insist that the "discrepancy" be much smaller than $1/2$ in order to have good evidence that $x = m/n$. We have chosen to accept that the number x is close enough to the rational number m/n only when the "discrepancy" is ≤ 0.001.

We have been able to fit the data corresponding to 42 endpoints. (We have numerically located additional endpoints, but we have not included here those endpoints that correspond to very small equimodular curves close to the circle $|p| = 1$.) Of these 42 endpoints, we have obtained a reasonably good fit for the 32 points displayed in Table 8: here a "reasonably good fit" is defined as one for which the "discrepancy" is $\leq 10^{-3}$ and for which the values of $\mathrm{Arg}\, p_{\mathrm{Fit}}$ and the "discrepancy" are stable under changes in the degree of the polynomial Ansatz (6.9). It is interesting to note that there are 15 endpoints satisfying the the stronger condition "discrepancy" $\leq 10^{-5}$.

We have found several smaller equimodular curves that are hard to see on Figure 22, and we suspect that they too have rational endpoints. Indeed, we would not be surprised to learn that such rational endpoints are *dense* in the unit circle (though we have no idea how to prove this conjecture).

The equimodular curves $|g_3/g_1| = 1$ in the complex y-plane are shown in Figure 23. Again, each equimodular curve D_n has exactly two endpoints, and smaller equimodular curves appear as we approach the $|y| = 1$ limit. We have shown in Figure 22 all equimodular curves that intersect the circle $|y| = 0.99$.

The principal feature is again a curve D_1 running from point G ($y = e^{-4\pi i/5}$, $\theta = \pi/6$ and images) via the origin ($p = 0$, $\theta = \pi - i0$) to point H ($y = 1$, $\theta = 2\pi/3$ and images). The next-longest curve (D_2) runs from point K ($y = e^{6\pi i/7}$, $\theta = 5\pi/12$ and images) to point L ($y = e^{4\pi i/5}$, $\theta = 4\pi/9$ and images). A much shorter curve (D_3) runs from point I ($y = e^{-6\pi i/7}$, $\theta = 2\pi/9$ and images) to point J ($p = e^{-20\pi i/23}$, $\theta = 7\pi/30$ and images). Finally, a small curve (D_4) runs from point G′ ($y = e^{-8\pi i/11}$, $\theta = \pi/12$ and images) to point H′ ($y = e^{-10\pi i/13}$, $\theta = 2\pi/15$ and images); we call these points G′ and H′ because they correspond to θ values that are images of the points G and H under the p-plane transformation (6.8a) [though they are not images in the y-plane].

Once again, all the endpoints appear to lie at $y = e^{i\psi}$ where ψ is a rational multiple of π with small denominator. Even more curiously, many of these endpoints

correspond to θ values that are also observed as endpoints in the $|g_2/g_1| = 1$ plot (e.g., all points with a label in Tables 8 and 9). We performed fits to those endpoints in the same manner as just explained for the p-plane. We found 44 endpoints, and of these we obtained reasonably good fits for the 36 endpoints displayed in Table 9. All these points satisfy the criterion "discrepancy" ≤ 0.001, and both $\operatorname{Arg} p_{\text{Fit}}$ and the "discrepancy" are rather stable as we vary the polynomial Ansatz in the fits (6.9). Of these 36 endpoints, we find 13 satisfying "discrepancy" $\leq 10^{-5}$.

We have also found more smaller equimodular curves that are hard to see on Figure 23; we conjecture that they have rational endpoints and we wonder whether these rational endpoints are dense in the unit circle.

Remark. It is important to note that the two main points in Figures 24–25 (namely, G and H) are extremely well determined in both the p- and y-planes. In particular, their "discrepancies" range from 10^{-6} down to 2×10^{-8}. Thus, we can trust that the values of θ for these two points are given *exactly* by $\theta_G = \pi/6$ and $\theta_H = 2\pi/3$ (see Tables 8–9).

6.3 Summary of θ-plane phase diagram

Let us now describe the resulting zero-temperature "phase diagram" in the complex θ-plane (Figure 24) and discuss the agreements and discrepancies with respect to Baxter [19]. For simplicity we have labelled the points by the same letters as in Figure 5 of [19]. The portions of curves where the equimodular eigenvalues are dominant (resp. subdominant) are depicted in black (resp. pink).

The curve A–B (resp. B–C) corresponds to the dominant equimodularity of g_3 (resp. g_2) and g_1. These two curves together with the real θ axis enclose the rest of the dominant curves. The position of these points is

$$\theta_{\text{A}} = \pi \tag{6.11a}$$

$$\theta_{\text{B}} \approx 0.508588719845180 - 0.625516375803391\, i \tag{6.11b}$$

$$\theta_{\text{C}} = 0 \tag{6.11c}$$

The point B is triply equimodular (i.e. a T point), so that three equimodular curves cross there:

1) C–B–H, which corresponds to $|g_1| = |g_2|$ (it is dominant along C–B and subdominant along B–H); and

2) A–B–Q–G, which corresponds to $|g_1| = |g_3|$ (it is dominant along A–B, subdominant along B–Q, and dominant again along Q–G);

3) R–B–Q–F, which corresponds to $|g_2| = |g_3|$ (it is subdominant along R–B, dominant along B–Q, and subdominant again along Q–F). Point R corresponds to $\theta = -i\infty$.

This last result contradicts [19], where the entire curve B–Q–F is claimed to be dominant. The position of these points is[15]

$$\theta_{\rm Q} \approx 0.440568708859061 - 0.235993788540783\,i \tag{6.12a}$$

$$\theta_{\rm F} \approx 0.427907971348122 \tag{6.12b}$$

The point Q is also triply equimodular, so that three equimodular curves cross there. Two of them have just been discussed: A–B–Q–G and R–B–Q–F. The third is C–Q–G, which corresponds to $|g_1| = |g_2|$; it is dominant along C–Q and subdominant along Q–G. Please note that the subdominant curve Q–G lies always to the right of (but very close to) the dominant curve Q–G.

There are infinitely many equimodular curves $|g_1| = |g_2|$ terminating at point C ($q = 4$) and converging to it: they are images under (6.8a) of the two curves C–B–H and C–Q–G. The dominant eigenvalue alternates between g_1 and g_2 as these curves are crossed. For simplicity, we have shown in Figure 24 only the first few of these image curves. The endpoints of these curves can be obtained easily using the values of $\theta_G = \pi/6$ and $\theta_H = 2\pi/3$ (see Table 8) and transformation (6.8a). The result is

$$\theta_{G,k} = \frac{\pi}{6(1+k)} \tag{6.13a}$$

$$\theta_{H,k} = \frac{2\pi}{3(1+4k)} \tag{6.13b}$$

According to (1.3)/(6.1), they correspond to Beraha numbers: $\theta_{G,k}$ is B_{12+12k}, and $\theta_{H,k}$ is B_{3+12k}.

Likewise, there are infinitely many equimodular curves $|g_1| = |g_3|$ terminating at point A ($q = 0$) and converging to it: they are images under (6.8b) of the two curves A–B–Q–G and A–H. The dominant eigenvalue alternates between g_1 and g_3 as these curves are crossed. Once again, we have shown in Figure 24 only the first few of these image curves. The endpoints of these curves can be obtained easily using the values of $\theta_G = \pi/6$ and $\theta_H = 2\pi/3$ (see Table 9) and transformation (6.8b). The result is

$$\theta_{G,k'} = \frac{1+15k'}{6+15k'}\pi \tag{6.14a}$$

$$\theta_{H,k'} = \frac{2+3k'}{3+3k'}\pi \tag{6.14b}$$

In this case, none of them corresponds to a Beraha number.

There are also many small equimodular curves lying near the real θ-axis, some of which are dominant; they arise from the curves I–J, K–L, etc. in Figures 22 and 23. The number of these curves grows rapidly as $|\operatorname{Im}\theta| \to 0$, so we cannot possibly compute all of them; moreover, the computation becomes increasingly difficult as $|\operatorname{Im}\theta| \to 0$, since $|p|$ and $|y|$ are tending to 1. We have, in any case, shown in

[15] Despite appearances, the point F in Figure 24 is *not* triply equimodular: the dominant equimodular curve $|g_1| = |g_2|$ meets the real axis at H′ ($\theta = 2\pi/15 \approx 0.41888$), i.e. slightly *below* point F; this is discussed below at (6.13b) ff. This splitting is somewhat more visible on Figure 26.

Figure 24 all those curves in the range $0.2 \le \mathrm{Re}\,\theta \le 2.3$ that intersect the half-plane $\mathrm{Im}\,\theta \le -0.01$. Again, one can obtain the endpoints of these curves by applying transformations (6.8) to the corresponding θ values in Tables 8 and 9. These endpoints are not in general Beraha numbers. Only point I ($\theta_I = 2\pi/9$) and its transformed values under (6.8a),

$$\theta_{I,k} = \frac{2\pi}{9 + 12k} , \tag{6.15}$$

are Beraha numbers (namely, B_{9+12k}).

It follows from Figure 24 that the point q_0 should not be identified with F (as Baxter [19] did), but rather with the point G at position

$$\theta_{\mathrm{G}} = \frac{\pi}{6} \approx 0.523598775598299 \tag{6.16}$$

(see Tables 8 and 9).

6.4 Summary of q-plane phase diagram

In Figure 25 we show the above "phase diagram" in the q-plane (for clarity, only the dominant equimodular curves have been depicted). This is quite similar to Figure 5 of [19], except for four issues:

1) The "phase diagram" around point C ($q = 4$) is richer than the one found by Baxter (see Figure 26 for a detailed plot of this region). The largest components of the region $\mathcal{D}_2$ (where g_2 is dominant) are bounded by C–B–Q–C and its complex-conjugate counterpart C–E–Q′–C. The points B and Q take the values

$$q_{\mathrm{B}} \approx 4.099903170634857 + 0.649694690705481\, i \tag{6.17a}$$

$$q_{\mathrm{Q}} \approx 3.859627688708099 + 0.203154495450945\, i \tag{6.17b}$$

However, there are additional components of $\mathcal{D}_1$ and $\mathcal{D}_2$ near point C; indeed, as we approach point C the dominant eigenvalue alternates between g_1 and g_2. Thus, g_1 is dominant on the region lying between C–Q–G–Q′–C and C–B_{15}–C; g_2 is dominant on the region between the curve C–B_{15}–C and the next curve we find towards C, etc. Baxter found [via the transformation (6.8a)] only half of the curves around point C, namely, those curves that are images of C–B–H and past through the Beraha numbers B_{3+12k}. (In Figure 26 we have shown two of these curves, corresponding to B_{15} and B_{27}.)

2) If we define q_0 as the point on the real q-axis where the region $\mathcal{D}_3$ ends, then the above discussion implies that q_0 is not given by $q_{\mathrm{F}} \approx 3.819671731239719$ as Baxter believed [cf. (6.5)], but rather by

$$q_0 = q_{\mathrm{G}} = B_{12} = 2 + \sqrt{3} \approx 3.732050807568877 \tag{6.18}$$

[cf. (6.16)].

3) The "phase diagram" around point A ($q = 0$) is also richer than the one found by Baxter. The analytic structure is similar to the one already discussed for the point

C, except for the fact that now the dominant eigenvalue alternates between g_1 and g_3. Again, Baxter found [via (6.8b)] only some of the equimodular curves in this region. The physical meaning of these new curves is not clear to us.

4) We also find many new dominant equimodular curves lying very close to the real q-axis between points A and C. Some of them (lying between H and G) have been depicted in Figure 25. These curves were missed by Baxter and their physical meaning is unclear.

On Figure 25 we have also superposed the limiting curve for $L = 11_{\rm P}$ (see Figure 16).[16] This curve lies quite close to the main parts of the infinite-strip-width limiting curve obtained in this section. The point q_0 for the $L = 11_{\rm P}$ strip (as well as for the rest of the strips considered in this paper: see Table 6) lies below both Baxter's prediction (6.5) and our somewhat lower prediction (6.18). So our results in Sections 3–5 are compatible with both predictions; unfortunately our strips are not yet wide enough to distinguish between them. On the other hand, it is precisely around points A and C that the finite-strip limiting curve is not defined, so our transfer-matrix results do not give any clue as to whether the additional dominant equimodular curves we have found there can be neglected (as Baxter did) or not.

Another way of discovering whether the true value of q_0 is given by point F (6.5) or by point G (6.18) is to consider the isolated limiting points of wide triangular-lattice strips. We expect that all *real* isolated limiting points are smaller than q_0. Thus, if we find any isolated limiting point larger than (6.18) [and smaller than (6.5)], then we should conclude that our prediction is false. In Tables 1, 3, 4 and 5 we do not find any such zero. The largest real zero we have found is $\approx 3.6345747 < q_{\rm G} = B_{12} \approx 3.7320508\ldots$. On the other hand, for free and cylindrical boundary conditions, all real isolated limiting points are expected to be Beraha numbers. Thus, if q_0 were given by F, as Baxter predicts, then the largest real isolated limiting point would be $B_{14} \approx 3.801938$. On the other hand, if q_0 is given by G, as we predict, then the largest real isolated limiting point would be $B_{11} \approx 3.682507$ (or perhaps $B_{12} \approx 3.732051$). In Figure 26 we have also depicted the position of the Beraha numbers $B_{11}, \ldots, B_{16}$ to make easier the comparison with the two alternative values of q_0.

7 Summary and Outlook

7.1 Behavior of dominant-eigenvalue-crossing curves $\mathcal{B}$

In this paper we have computed the transfer matrix for triangular-lattice strips of width $3 \le L_x \le 9$ for free boundary conditions, $4 \le L_x \le 12$ for cylindrical boundary conditions, and $L_x = 4, 6, 8, 10$ for "zig–zag" boundary conditions. The transfer matrix allows the computation of the chromatic zeros for strips of arbitrary length L_y. As the length L_y tends to infinity (for fixed width L_x), the chromatic zeros accumulate along certain curves (limiting curves $\mathcal{B}$) and around certain points (isolated limiting points) according to the Beraha–Kahane–Weiss theorem [25, 26, 27].

[16] $L = 12_{\rm P}$ would be substantially similar, had we been able to compute the limiting curve for it (compare Figures 16 and 17).

For all the above strips except $L_x = 10_{\rm F}, 12_{\rm P}, 10_{\rm Z}$, we have been able to compute the limiting curves $\mathcal{B}$. The exact computation of all the isolated limiting points has been carried out for $L_x \le 6_{\rm F}$, $L_x \le 9_{\rm P}$ and $L_x \le 6_{\rm Z}$; for the larger strips we were able to check that certain values of q are isolated limiting points, but we cannot be certain that we have found all of them. By studying the behavior of the limiting curves and isolated limiting points as a function of the strip width L_x (and boundary conditions), we hope to shed light on the thermodynamic limit $L_x, L_y \to \infty$.

The basic properties of both limiting curves and isolated limiting points are summarized in Table 6. In all cases the identity

$$\begin{aligned} \text{endpoints} \;\; = \;\; & (2 \times \text{components}) + (2 \times \text{double points}) + (\text{T points}) \\ & - (2 \times \text{enclosed regions}) \end{aligned} \tag{7.1}$$

holds. This identity can be derived by simple topological/graph-theoretic arguments.

By inspection of Table 6, we observe some regularities when L_x becomes large. For all three boundary conditions, the curve $\mathcal{B}$ appears to become connected ($\#\mathrm{C} = 1$) when L_x is large enough. For all L_x, the number of endpoints ($\#\mathrm{E}$) is 6 for free boundary conditions and 4 for the other two boundary conditions. (Note, however, that for $L_x \ge 7_{\rm F}$, $L_x \ge 10_{\rm P}$ and $L_x = 8_{\rm Z}$ our counts on the number of endpoints are only lower bounds: we may have missed some.) When L_x is large enough, the number of T points ($\#\mathrm{T}$) is 4 for free boundary conditions and 2 for the other boundary conditions. We have found no evidence of double points for any of the strips considered. Finally, the number of enclosed regions is zero except for $L_x = 5_{\rm F}$, $L_x \le 5_{\rm P}$ and $L_x = 7_{\rm P}$. These regularities are in sharp contrast with the square-lattice case [16, 17], where the number of connected components and endpoints seems to grow with the strip width. In particular, we have not found in the triangular-lattice limiting curves any trace of the small gaps and bulb-like regions that are so common in the square-lattice case. It therefore seems that the thermodynamic limit may be achieved in a smoother way for the triangular lattice than for the square lattice. Finally, it is worth mentioning that in all cases except $L_x = 4_{\rm F}$ and $4_{\rm Z}$, the limiting curve $\mathcal{B}$ crosses the real q-axis, thus defining q_0. By contrast, for the square lattice, q_0 is well-defined only for odd widths; for even widths with both free and cylindrical boundary conditions, we found either that $\mathcal{B}$ fails to intersect the real axis or that it contains a segment of the real axis passing through a double point [16, 17].

Let us also note that, as in the square-lattice case [16, 17], we find chromatic zeros with $\mathrm{Re}\, q < 0$. Indeed, for $L_x \ge 7_{\rm P}$ and $L_x \ge 8_{\rm Z}$ we find that the limiting curve $\mathcal{B}$ intersects the half-plane $\mathrm{Re}\, q < 0$. For free boundary conditions, none of our limiting curves ($L_x \le 8_{\rm F}$) reach this half-plane; but from Table 6 we can see that $\min \mathrm{Re}\, q$ is decreasing and we expect that it will be < 0 for $L_x \ge 10_{\rm F}$ (and possibly already for $9_{\rm F}$).

The regularities exhibited by the limiting curves become clearer when we superpose them all (with fixed boundary conditions). This is done in Figures 27, 28 and 29 for free, cylindrical and zig-zag b.c., respectively. We find an overall behavior similar to that found for the square lattice [16, 17]. For free boundary conditions (Figure 27), we find a monotonic behavior with the width L_x: both the leftmost arcs and the rightmost arcs move outwards as L_x is increased (see also the columns labelled $\min \mathrm{Re}\, q$

and $\max \mathrm{Re}\, q$ in Table 6). The value of q_0 (or $\mathrm{Re}\, q_0$ for $L_x = 4_{\mathrm{F}}$) is also monotonically increasing in L_x. The overall shape of the limiting curves is similar to the expected limiting curve in the thermodynamic limit (Figure 25). We expect that as L_x grows, the leftmost endpoints will tend towards $q = 0$, while the rightmost endpoints will go to $q = q_c(\mathrm{tri}) = 4$. The crossing point q_0 will eventually go to either point F [cf. (6.5)] or point G [cf. (6.18)] in Figure 25. Unfortunately, our numerical data are not good enough to tell unambiguously the true limit. There is additional one feature of the limiting curves with free boundary conditions that does *not* correspond to any feature of the predicted thermodynamic-limit curve: namely, a pair of small complex-conjugate branches emerging from T points and pointing inwards. From Figure 27, it seems that the size of these branches does not go to zero as L_x is increased (at least up to 8 or 9); rather their size stays more or less constant. We are unable to say whether these branches will get shorter for larger L_x and ultimately disappear in the limit $L_x \to \infty$.

In Figure 28 we superpose all the limiting curves with cylindrical boundary conditions. As in the square-lattice case, the behavior of the leftmost part of these curves seems to be monotonic: the arcs move outwards as L_x is increased. In particular, $\min \mathrm{Re}\, q$ decreases monotonically with the strip width (see Table 6). However, the behavior on the right side of the plot is clearly not monotonic: there are differences depending on the quantity L_x mod 3. This is to be expected, since with periodic boundary conditions in the transversal direction, strip widths that are not multiples of 3 are somewhat *unnatural* as they introduce frustration in the 3-state Potts antiferromagnet. Thus, the dependence on L_x in the interval $3 \lesssim \mathrm{Re}\, q \lesssim 4.5$ is not a surprise (the same feature is present in the square-lattice case, where we find an even-odd dependence on the limiting curves [16, 17]). For fixed values of L_x mod 3, we find that q_0 is monotonic in L_x: for $L_x = 1$ mod 3 it decreases, while for $L_x = 0$ or 2 mod 3 it increases. The shape of the limiting curves for $L_x = 0$ mod 3 is very similar to the expected thermodynamic limit (Figure 25), while in the other cases we find stronger finite-size effects that we expect to disappear in the limit $L_x \to \infty$. Finally, it is worth mentioning the absence of the extra branches that appear for free boundary conditions.

In Figure 29 we superpose the limiting curves for "zig-zag" boundary conditions. In this case we only have three curves, so we are unable to extract any definitive conclusion. We can only confirm the monotonic behavior on the leftmost side of the curves and the absence of any extra branches. Again, the overall shape is similar to the expected thermodynamic limit depicted in Figure 25.

If we compare the limiting curves for different boundary conditions, we see that the thermodynamic limit is achieved faster for cylindrical and "zig-zag" boundary conditions than for free boundary conditions (due to existence of "surface" effects in the later). This is similar to the behavior observed in the square-lattice case.

Remark. In the computation of the limiting curves $\mathcal{B}$ for triangular-lattice strips with *cylindrical* boundary conditions, we have found a curious behavior involving identically vanishing amplitudes. (This is important, because the eigenvalues corresponding to the identically vanishing amplitudes must be *excluded* from the compu-

tation of the equimodular curves. For square-lattice strips, by contrast, we have not observed any identically vanishing amplitudes [16, 17].) As explained in the introduction to Section 4, the transfer matrix can be written (after a change of basis) in the block-diagonal form

$$\mathsf{T}(m_{\mathrm{P}}) \;=\; \begin{pmatrix} \mathsf{T}_+(m_{\mathrm{P}}) & 0 \\ 0 & \mathsf{T}_-(m_{\mathrm{P}}) \end{pmatrix} , \tag{7.2}$$

where the matrix $\mathsf{T}_+(m_{\mathrm{P}})$ lives on the subspace of reflection-invariant connectivities and has dimension SqCyl(m) [i.e., the dimension of the transfer matrix for a square-lattice strip of width m with cylindrical boundary conditions], while the matrix $\mathsf{T}_-(m_{\mathrm{P}})$ lives on the subspace of reflection-odd connectivities and has dimension TriCyl(m) $-$ SqCyl(m). For $m \ge 8_{\mathrm{P}}$ we have TriCyl(m) $>$ SqCyl(m) and this decomposition becomes nontrivial. Now, simple symmetry arguments (see Section 4) explain why all the eigenvalues in the reflection-odd subspace should have identically vanishing amplitudes. But what is curious and mysterious is that each of these eigenvalues has an identical "partner" in the reflection-even subspace, also with identically vanishing amplitude. This means that the characteristic polynomial associated to the transfer matrix $\mathsf{T}(m_{\mathrm{P}})$ can be factored as

$$\det[\mathsf{T}(m_{\mathrm{P}}) - \lambda \mathbf{1}] \;=\; Q_1(q,\lambda)^2 \, Q_2(q,\lambda) \,, \tag{7.3}$$

where the zero-amplitude eigenvalues are those coming from the factor $Q_1(q,\lambda)^2$. In particular, the number of eigenvalues with zero amplitude (# VA) is always even, and it equals twice the dimension of the reflection-odd subspace:

$$\#\mathrm{VA}(m) \;=\; 2 \times [\mathrm{TriCyl}(m) - \mathrm{SqCyl}(m)] \tag{7.4}$$

(see Table 7). This is, at any rate, what we have found for $8_{\mathrm{P}} \le m \le 12_{\mathrm{P}}$ (see Sections 4.7–4.11); we conjecture that it holds for larger widths as well. It follows that the effective dimension of the transfer matrix TriCyl$'(m)$ is given by

$$\mathrm{TriCyl}'(m) \;=\; 2\mathrm{SqCyl}(m) - \mathrm{TriCyl}(m) \,. \tag{7.5}$$

Numerical values for all these quantities are displayed in Table 7; of course, the values of # VA and TriCyl$'(m)$ for $m = 13, 14$ are *conjectures*.

Numerical values of TriCyl(m) and SqCyl(m) were first reported in [16, Table 2]. An analytic formula of TriCyl(m) for prime values of m has been obtained in [53, Theorem 3]. This paper also contains a conjecture for SqCyl(m) with prime $m \ge 3$ [53, Conjecture 2]. Finally, an analytic formula for TriCyl$'(m)$ has been conjectured for arbitrary values of m [53, Conjecture 1].

7.2 Behavior of amplitudes and the Beraha conjecture

Let us now discuss the isolated limiting points and the role of the Beraha numbers in the triangular-lattice strips. Our results show that the number of isolated limiting

points is a non-decreasing function of the strip width L_x (for each boundary condition), at least up to the maximum L_x we have been able to investigate. For free and cylindrical boundary conditions we did not find any complex isolated limiting points (see Table 6). For "zig-zag" boundary conditions we find a pair of complex-conjugate isolated limiting points for $L_x = 6_{\rm Z}$, and we have evidence of the existence of another pair of complex-conjugate isolated limiting points for $L_x = 8_{\rm Z}$.

Concerning the real isolated limiting points, most of them are Beraha numbers (1.3). It is only for "zig-zag" boundary conditions that we find real isolated limiting points that are *not* Beraha numbers: for $L_x = 4_{\rm Z}$ we find $q = 5/2$; for $L_x = 6_{\rm Z}$, $q \approx 2.722633$; for $L_x = 8_{\rm Z}$, $q \approx 2.821420$; and for $L_x = 10_{\rm Z}$, we find two such points, $q \approx 2.873731$ and $q \approx 3.383129$. It is not clear to us how these non-Beraha real isolated limiting points behave as $L_x \to \infty$, e.g., whether their number is bounded or unbounded.

For all the lattices we have studied, we observed empirically that there is at least one vanishing amplitude $\alpha_i(q)$ at each of the Beraha numbers up to B_{L+1} (see Table 10). It is reasonable to conjecture that this holds for all L (in agreement with a similar conjecture for the square lattice [16, Conjecture 7.1]):

Conjecture 7.1 *For a triangular-lattice strip of width L with free, cylindrical or "zig-zag" boundary conditions, at each Beraha number $q = B_2, \ldots, B_{L+1}$ there is at least one vanishing amplitude $\alpha_i(q)$. That is,* $\det D(q) = 0$ *for* $q = B_2, \ldots, B_{L+1}$.

In contrast with the square-lattice case [16, Conjectures 7.2 and 7.3], however, we find that there is a vanishing amplitude [hence $\det D(q) = 0$] also at some Beraha numbers *larger* than B_{L+1}. Indeed, we find examples for each boundary condition (see Table 10):

- For free boundary conditions, $q = B_6$ is a zero of $\det D(q)$ for $L = 4_{\rm F}$.
- For cylindrical boundary conditions, $q = B_6$ is a zero of $\det D(q)$ for $L = 4_{\rm P}$; $q = B_{10}$ is a zero for $L = 6_{\rm P}, 7_{\rm P}, 8_{\rm P}$; $q = B_{14}$ is a zero for $L = 8_{\rm P}, 9_{\rm P}, 10_{\rm P}, 11_{\rm P}, 12_{\rm P}$; $q = B_{18}$ is a zero for $L = 10_{\rm P}, 11_{\rm P}, 12_{\rm P}$; and finally, $q = B_{22}$ is a zero for $L = 12_{\rm P}$.
- For "zig-zag" boundary conditions, $q = B_6$ is a zero of $\det D(q)$ for $L = 4_{\rm Z}$.

We have systematically checked all Beraha numbers up to B_{50} to make this list.

Please note that for free and "zig-zag" boundary conditions, we have found only *one* case each in which a Beraha number beyond B_{L+1} is a zero of $\det D(q)$, namely the relatively small value of $L = 4$. It is conceivable that for all larger L there are no such Beraha zeros, as is conjectured for all L for the square lattice [16, Conjecture 7.2]. However, this conjecture clearly cannot be true for cylindrical boundary conditions. There is presumably some pattern that tells us *which* Beraha numbers $q = B_k$ with $k > L+1$ can be zeros of $\det D(q)$. Thus far only a few Beraha numbers $(B_6, B_{10}, B_{14}, B_{18}, B_{22})$ have appeared on that list. Indeed, we conjecture that the pattern is the following:

Conjecture 7.2 *For a triangular-lattice strip of width L with cylindrical boundary conditions, the Beraha numbers where* $\det D(q)$ *vanishes is given by the union of the sets* $\{B_2, B_3, ..., B_{L+1}\}$ *and* $\{B_{4k-2} \mid k = 1, 2, ..., \lfloor L/2 \rfloor\}$*, the upper limit on* k *being the integer part of* $L/2$*.*

7.3 Nature of the fixed zeros

In Section 2 we discussed the "fixed" zeros that occur at small integers q (here $q = 0, 1, 2, 3$) when the graph fails to be q-colorable. From the point of view of the transfer-matrix formalism, these fixed zeros can arise in either of three ways:

1) All the amplitudes α_k vanish at q. Then $Z_n(q) = 0$ for all lengths $n \geq 1$.

2) All the eigenvalues λ_k vanish at q. Then $Z_n(q) = 0$ for all $n \geq 2$.

3) "Mixed case": Neither all the amplitudes nor all the eigenvalues vanish at q, but for each k either α_k or λ_k vanishes at q (or both). Then $Z_n(q) = 0$ for all $n \geq 2$.

Let us now summarize what we have found concerning the nature of these fixed zeros for triangular-lattice strips:

$\boldsymbol{q = 0, 1.}$ At $q = 0, 1$ all the amplitudes vanish, due to the prefactor $q(q-1)$ in the left vector $\mathbf{u}$. These points therefore belong to Case 1.

$\boldsymbol{q = 2.}$ At $q = 2$ the behavior depends on the boundary conditions and on the strip width L_x:

- Free boundary conditions:
  - $L_x = 2_{\rm F}$: The one eigenvalue vanishes at $q = 2$ (Case 2).
  - $L_x = 3_{\rm F}$: There is one nonzero eigenvalue with a zero amplitude, and one zero eigenvalue with a nonzero amplitude (Case 3).
  - $L_x = 4_{\rm F}$: There is at least one nonzero eigenvalue with a zero amplitude and exactly one zero eigenvalue with a nonzero amplitude (Case 3). The transfer matrix at $q = 2$ is not diagonalizable: it has a nontrivial Jordan block corresponding to $\lambda = 0$.
  - $L_x \geq 5_{\rm F}$: There is at least one nonzero eigenvalue with a zero amplitude, at least one zero eigenvalue with a zero amplitude, and exactly one zero eigenvalue with a nonzero amplitude (Case 3). We also find that for all $L_x \geq 4_{\rm F}$ the transfer matrix at $q = 2$ is not diagonalizable, i.e. it has nontrivial Jordan blocks (all corresponding to eigenvalue $\lambda = 0$).
- Cylindrical boundary conditions:
  - L_x odd: All the amplitudes vanish, due to the prefactor $q(q-1)(q-2)$ in the left vector $\mathbf{u}$ (Case 1).

- $L_x = 2_{\rm P}$: The one eigenvalue vanishes at $q = 2$ (Case 2).
- $L_x = 4_{\rm P}$: There is one nonzero eigenvalue with a zero amplitude, and one zero eigenvalue with a nonzero amplitude (Case 3).
- L_x even $\geq 6_{\rm P}$: There is at least one nonzero eigenvalue with a zero amplitude, at least one zero eigenvalue with a zero amplitude, and at least one zero eigenvalue with a nonzero amplitude (Case 3).

- Zig-zag boundary conditions:
  - $L_x = 2_{\rm Z}$: This is identical to $L_x = 2_{\rm F}$ (Case 2).
  - $L_x = 4_{\rm Z}$: There are two nonzero eigenvalues with zero amplitudes, and one zero eigenvalue with a nonzero amplitude (Case 3).
  - L_x even $\geq 6_{\rm Z}$: There is at least one nonzero eigenvalue with a zero amplitude, at least one zero eigenvalue with a zero amplitude, and at least one zero eigenvalue with a nonzero amplitude (Case 3).

***q* = 3.** The point $q = 3$ is a fixed zero only for cylindrical boundary conditions with strip widths L_x that are not a multiple of 3. There are two distinct situations:

- $L_x = 4_{\rm P}$: Both eigenvalues vanish, so that the whole transfer matrix vanishes (Case 2).
- $L_x = 5_{\rm P}$: There is one nonzero eigenvalue with a zero amplitude, and one zero eigenvalue with a nonzero amplitude (Case 3).
- $L_x = 7_{\rm P}, 8_{\rm P}, 10_{\rm P}, 11_{\rm P}$: There is at least one nonzero eigenvalue with zero amplitude, at least one zero eigenvalue with a nonzero amplitude, and at least one zero eigenvalue with a zero amplitude [or nontrivial Jordan block corresponding to $\lambda = 0$ with no contribution to the partition function for any $n \geq 1$] (Case 3). We also find that for $L_x \geq 8_{\rm P}$ the transfer matrix at $q = 3$ is not diagonalizable, i.e. it has nontrivial Jordan blocks.

Acknowledgments

We wish to thank Dario Bini for supplying us the MPSolve 2.1.1 package [51, 52] and for many discussions about its use; George Andrews and Mireille Bousquet-Mélou for useful suggestions concerning the numerical computation of the products (6.6); Hubert Saleur for emphasizing the importance of the Beraha numbers; Norman Weiss for suggesting that we study the resultant; and Robert Shrock for many helpful conversations throughout the course of this work.

The authors' research was supported in part by U.S. National Science Foundation grants PHY–9900769 (J.S. and A.D.S.) and PHY–0099393 (A.D.S.) and CICyT (Spain) grant FPA2000-1252 (J.S.).

A Numerical Computation of $\prod_{n=1}^{\infty}(1-tx^n)$

In this appendix we discuss briefly some of the technical issues involved in the numerical computation of Baxter's products (6.6). Everything can be expressed in terms of the function

$$R(t,x) \;=\; \prod_{n=1}^{\infty}(1-tx^n)\,, \tag{A.1}$$

which is defined for complex t and x satisfying $|x|<1$ and was first studied by Euler [54]. Here we need the cases $t=\pm 1$ and $t=\pm$ a cube root of unity. A more detailed discussion, including proofs, can be found in [55].

The numerical computation of $R(t,x)$ clearly becomes delicate when $|x| \uparrow 1$. In particular, direct use of the product (A.1) gives an algorithm that is only "linearly convergent", i.e. the number of significant digits in the answer grows linearly with the number of terms taken. Moreover, the constant of proportionality in this relation is proportional to $1-|x|$, and thus deteriorates linearly as $|x| \uparrow 1$. Finally, there is severe loss of numerical precision when multiplying numbers that are very near 1. An alternative approach can be based on the representation

$$\log R(t,x) \;=\; -\sum_{k=1}^{\infty}\frac{t^k}{k}\,\frac{x^k}{1-x^k}\,, \tag{A.2}$$

which is valid whenever $|x|<1$ and $|tx|<1$. This sum is again only linearly convergent, but the problem of loss of numerical precision is alleviated by use of the logarithm.

A much more efficient algorithm can be based on the identity

$$R(t,x) \;=\; \sum_{n=0}^{\infty}\frac{(-t)^n x^{n(n+1)/2}}{(1-x)(1-x^2)\cdots(1-x^n)} \tag{A.3}$$

due to Euler.[17] Because of the $x^{n(n+1)/2}$ factor in the numerator, this algorithm is "quadratically convergent":

Proposition A.1 ([55]) *Define*

$$a_n \;=\; \frac{(-t)^n x^{n(n+1)/2}}{(1-x)(1-x^2)\cdots(1-x^n)}\,. \tag{A.4}$$

Then, for $|t|\le 1$ and $|x| \le e^{-\gamma}$ $(\gamma>0)$, we have

(a) $\Delta_N \;\equiv\; \left|\sum_{n=N}^{\infty} a_n\right| \;\le\; \dfrac{e^{\pi^2/6\gamma - N(N+1)\gamma/2}}{1-e^{-(N+1)\gamma}}$

[17] For a proof of (A.3), see e.g. [56, p. 19, Corollary 2.2], [57, p. 34, Lemma 4(a)] or [58, pp. 22–23].

(b) $\delta_N \equiv \dfrac{\left|\sum_{n=N}^{\infty} a_n\right|}{|R(t,x)|} \le \dfrac{e^{\pi^2/3\gamma - N(N+1)\gamma/2}}{1 - e^{-(N+1)\gamma}}$

Corollary A.2 ([55]) *Let $K \ge 0$, and suppose that $|t| \le 1$ and $|x| \le e^{-\gamma}$ $(\gamma > 0)$.*

(a) If $N \ge \sqrt{\dfrac{\pi^2}{3\gamma^2} + \dfrac{2K}{\gamma}}$, then $\Delta_N \le e^{-K}$.

(b) If $N \ge \sqrt{\dfrac{2\pi^2}{3\gamma^2} + \dfrac{2K}{\gamma}}$, then $\delta_N \le e^{-K}$.

It turns out [55] that the *a priori* bound of Proposition A.1(b) is asymptotically within 9.1% of being sharp when $t = 1$, $0 < x = e^{-\gamma} < 1$ and $N \gg 1/\gamma$ (moreover, in this case it is asymptotically sharp as $\gamma \downarrow 0$). But since this bound is overly pessimistic in other cases, it is of some value to provide an *a posteriori* bound on the truncation error that is more realistic, when $x \notin (0,1)$, than the *a priori* bound. Here is such a bound, which can be used a stopping criterion in the numerical algorithm:

Proposition A.3 ([55]) *Let $|t| \le 1$, $|x| \le e^{-\gamma}$ $(\gamma > 0)$ and $N > (\log 2)/\gamma$. Then:*

(a) $\Delta_N \equiv \left|\sum_{n=N}^{\infty} a_n\right| \le |a_{N-1}| \dfrac{e^{-N\gamma}}{1 - 2e^{-N\gamma}}$

(b) $\delta'_N \equiv \dfrac{\left|\sum_{n=N}^{\infty} a_n\right|}{|S_N|} \le \dfrac{|a_{N-1}|}{|S_N|} \dfrac{e^{-N\gamma}}{1 - 2e^{-N\gamma}}$ *where* $S_N \equiv \sum_{n=0}^{N-1} a_n$

[Note also that $\delta_N \le \delta'_N/(1 - \delta'_N)$.]

In particular, if $N \ge (\log 3)/\gamma$, we have $\Delta_N \le |a_{N-1}|$ and $\delta'_N \le |a_{N-1}|/|S_N|$.

Let us conclude by making some brief remarks about the numerical precision that is required in intermediate stages of the calculation based on (A.3). It turns out [55] that the largest term $\max_n |a_n|$ can be as large in magnitude as $e^{\pi^2/12\gamma}$ (and is indeed of this order when $0 < x < 1$), while the answer $R(t,x)$ can be as small in magnitude as $e^{-\pi^2/6\gamma}$ (and is indeed of this order when $t = 1$ and $0 < x < 1$). It is therefore necessary to maintain, in intermediate stages of the calculation, approximately $(\pi^2/4\gamma)/\log 10 \approx 1.07/\gamma$ digits of working precision beyond the number of significant digits desired in the final answer.

We used all three algorithms — the product (A.1), the logarithmic sum (A.2) and the quadratically convergent sum (A.3) — and carefully cross-checked the value of $R(t,x)$; we also verified numerically the error bounds of Proposition A.1, Corollary A.2 and Proposition A.3. In order to guarantee that the roundoff error is under control, we performed all computations using MATHEMATICA with a working precision of at least 100 digits and often much more (increasing the working precision until the answer is independent of the precision used).

References

[1] F.Y. Wu, Rev. Mod. Phys. **54**, 235 (1982); **55**, 315 (E) (1983).

[2] F.Y. Wu, J. Appl. Phys. **55**, 2421 (1984).

[3] R.J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, London–New York, 1982).

[4] P.P. Martin, *Potts Models and Related Problems in Statistical Mechanics*. (World Scientific, Singapore, 1991).

[5] R.J. Baxter, J. Math. Phys. **11**, 784 (1970).

[6] R.J. Baxter, Proc. Roy. Soc. London A **383**, 43 (1982).

[7] J.-S. Wang, R.H. Swendsen and R. Kotecký, Phys. Rev. B **42**, 2465 (1990).

[8] H. Saleur, Commun. Math. Phys. **132**, 657 (1990).

[9] H. Saleur, Nucl. Phys. B **360**, 219 (1991).

[10] J. Adler, A. Brandt, W. Janke and S. Shmulyan, J. Phys. A **28**, 5117 (1995).

[11] J. Salas and A.D. Sokal, J. Stat. Phys. **86**, 551 (1997), cond-mat/9603068.

[12] J. Salas and A.D. Sokal, J. Stat. Phys. **92**, 729 (1998), cond-mat/9801079.

[13] S.J. Ferreira and A.D. Sokal, J. Stat. Phys. **96**, 461 (1999), cond-mat/9811345.

[14] J. Cardy, J.L. Jacobsen and A.D. Sokal, J. Stat. Phys. **105**, 25 (2001), cond-mat/0101197.

[15] R.C. Read and W.T. Tutte, in *Selected Topics in Graph Theory 3*, ed. L.W. Beineke and R.J. Wilson (Academic Press, London, 1988).

[16] J. Salas and A.D. Sokal, J. Stat. Phys. **104**, 609 (2001), cond-mat/0004330.

[17] J.L. Jacobsen and J. Salas, J. Stat. Phys. **104**, 701 (2001), cond-mat/0011456.

[18] R.J. Baxter, J. Phys. A **19**, 2821 (1986).

[19] R.J. Baxter, J. Phys. A **20**, 5241 (1987).

[20] C.N. Yang and T.D. Lee, Phys. Rev. **87**, 404 (1952).

[21] S. Beraha and J. Kahane, J. Combin. Theory B **27**, 1 (1979).

[22] S. Beraha, J. Kahane and N.J. Weiss, J. Combin. Theory B **28**, 52 (1980).

[23] R. Shrock and S.-H. Tsai, Phys. Rev. E **55**, 5165 (1997), cond-mat/9612249.

[24] M. Roček, R. Shrock and S.-H. Tsai, Physica A **252**, 505 (1998), cond-mat/9712148.

[25] S. Beraha, J. Kahane and N.J. Weiss, Proc. Nat. Acad. Sci. USA **72**, 4209 (1975).

[26] S. Beraha, J. Kahane and N.J. Weiss, in *Studies in Foundations and Combinatorics* (Advances in Mathematics Supplementary Studies, Vol. 1), ed. G.-C. Rota (Academic Press, New York, 1978).

[27] A.D. Sokal, Chromatic roots are dense in the whole complex plane, cond-mat/0101197.

[28] R. Shrock and S.-H. Tsai, Phys. Rev. E **56**, 1342 (1997), cond-mat/9703249.

[29] S. Beraha, unpublished, circa 1974.

[30] R.J. Baxter, H.N.V. Temperley and S.E. Ashley, Proc. Roy. Soc. London A **358**, 535 (1978).

[31] B. Nienhuis, Phys. Rev. Lett. **49**, 1062 (1982).

[32] J. Stephenson, J. Math. Phys. **5**, 1009 (1964).

[33] H.W.J. Blöte and H.J. Hilhorst, J. Phys. A **15**, L631 (1982).

[34] B. Nienhuis, H.J. Hilhorst and H.W.J. Blöte, J. Phys. A **17**, 3559 (1984).

[35] C.L. Henley, private communications.

[36] J. Salas and A.D. Sokal, unpublished.

[37] A.C.D. van Enter, R. Fernández and A.D. Sokal, unpublished (1996).

[38] J. Salas and A.D. Sokal, in preparation.

[39] P.W. Kasteleyn and C.M. Fortuin, J. Phys. Soc. Japan **26** (Suppl.), 11 (1969).

[40] C.M. Fortuin and P.W. Kasteleyn, Physica **57**, 536 (1972).

[41] M. Roček, R. Shrock and S.-H. Tsai, Physica A **259**, 367 (1998), cond-mat/9807106.

[42] R. Shrock and S.-H. Tsai, Phys. Rev. E **58**, 4332 (1998), cond-mat/9808057.

[43] R. Shrock, Discrete Math. **231**, 421 (2001), cond-mat/9908307.

[44] S.-C. Chang and R. Shrock, Annals Phys. **290**, 124 (2001), cond-mat/0004129.

[45] R. Shrock and S.-H. Tsai, Physica A **275**, 429 (2000), cond-mat/9907403.

[46] S.-C. Chang and R. Shrock, Physica A **292**, 307 (2001), cond-mat/0007491.

[47] S.-C. Chang and R. Shrock, Physica A **296**, 131 (2001), cond-mat/0005232.

[48] S.-C. Chang and R. Shrock, Physica A **286**, 189 (2000), cond-mat/0004181.

[49] S.-C. Chang, J. Salas and R. Shrock, in preparation.

[50] S.-C. Chang and R. Shrock, General structural results for Potts model partition functions on lattice strips, cond-mat/0201223.

[51] D.A. Bini and G. Fiorentino, Numerical computation of polynomial roots: MPSolve – Version 2.0. FRISCO report (1998). Available at `http://www.dm.unipi.it/pages/bini/public_html/papers/mpsolve.ps.Z`. Software package available at `http://www.dm.unipi.it/pages/bini/public_html/software/mps2.tar.gz`.

[52] D.A. Bini and G. Fiorentino, Numer. Algorithms **23**, 127 (2000).

[53] S.-C. Chang, J. Salas and R. Shrock, J. Stat. Phys. **107**, 1207 (2002), cond-mat/0108144.

[54] L. Euler, *Introduction to Analysis of the Infinite* [*Introductio in Analysin Infinitorum*, 1748], 2 vols., translated by John D. Blanton (Springer-Verlag, New York, 1988/1990).

[55] A.D. Sokal, Numerical computation of $\prod_{n=1}^{\infty}(1-tx^n)$, in preparation.

[56] G.E. Andrews, *The Theory of Partitions* (Cambridge University Press, Cambridge, 1998).

[57] M.I. Knopp, *Modular Functions in Analytic Number Theory* (Markham, Chicago, 1970).

[58] R. Remmert, *Classical Topics in Complex Function Theory* (Springer-Verlag, New York–Berlin–Heidelberg, 1998).

Lattice	4th Zero	5th Zero	6th Zero
$3_F \times 3_F$			
$3_F \times 6_F$	2.552816126636		
$3_F \times 9_F$			
$3_F \times 12_F$	2.562658027317		
$3_F \times 15_F$			
$3_F \times 18_F$	2.565287184975		
$3_F \times 21_F$			
$3_F \times 24_F$	2.566507072062		
$3_F \times 27_F$			
$3_F \times 30_F$	2.567211365497		
$4_F \times 4_F$	2.604661945742		
$4_F \times 8_F$	2.618028652707		
$4_F \times 12_F$	2.618033986251		
$4_F \times 16_F$	2.618033988749		
$4_F \times 20_F$	2.618033988750		
$4_F \times 24_F$	2.618033988750		
$4_F \times 28_F$	2.618033988750		
$4_F \times 32_F$	2.618033988750		
$4_F \times 36_F$	2.618033988750		
$4_F \times 40_F$	2.618033988750		
$5_F \times 5_F$	2.618161303055	2.795370504128	
$5_F \times 10_F$	2.618033988749		
$5_F \times 15_F$	2.618033988750	2.947523648832	
$5_F \times 20_F$	2.618033988750		
$5_F \times 25_F$	2.618033988750	2.968180058756	
$5_F \times 30_F$	2.618033988750		
$5_F \times 35_F$	2.618033988750	2.976760450197	
$5_F \times 40_F$	2.618033988750		
$5_F \times 45_F$	2.618033988750	2.981534673779	
$5_F \times 50_F$	2.618033988750		
$6_F \times 6_F$	2.618033979731		
$6_F \times 12_F$	2.618033988750	3.001429148693	3.054848659601
$6_F \times 18_F$	2.618033988750	3.000001523178	3.100527321592
$6_F \times 24_F$	2.618033988750	3.000000001785	3.118151997375
$6_F \times 30_F$	2.618033988750	3.000000000002	3.127749140385
$6_F \times 36_F$	2.618033988750	3.000000000000	3.133811079422
$6_F \times 42_F$	2.618033988750	3.000000000000	3.137993327670
$6_F \times 48_F$	2.618033988750	3.000000000000	3.141054810628
$6_F \times 54_F$	2.618033988750	3.000000000000	3.143393623378
$6_F \times 60_F$	2.618033988750	3.000000000000	3.145239011028
Beraha	2.618033988750	3	3.246979603717

Table 1: Real zeros of the chromatic polynomials of finite triangular-lattice strips with free boundary conditions in both directions, to 12 decimal places. A blank means that the zero in question is absent. The first three real zeros $q = 0, 1, 2$ are exact on all lattices. "Beraha" indicates the Beraha numbers $B_5 = (3 + \sqrt{5})/2$, $B_6 = 3$, and B_7.

Lattice	4th Zero	5th Zero	6th Zero	7th Zero
$7_F \times 7_F$	2.618033988750	2.978584823651		
$7_F \times 14_F$	2.618033988750	3.000000029690	3.160410975706	
$7_F \times 21_F$	2.618033988750	3.000000000000		
$7_F \times 28_F$	2.618033988750	3.000000000000	3.218685236695	
$7_F \times 35_F$	2.618033988750	3.000000000000		
$7_F \times 42_F$	2.618033988750	3.000000000000	3.236121891966	
$7_F \times 49_F$	2.618033988750	3.000000000000		
$7_F \times 56_F$	2.618033988750	3.000000000000	3.243833695579	
$7_F \times 63_F$	2.618033988750	3.000000000000		
$7_F \times 70_F$	2.618033988750	3.000000000000	3.246633282347	
$7_F \times 77_F$	2.618033988750	3.000000000000	3.247059872523	3.254369173708
$7_F \times 84_F$	2.618033988750	3.000000000000	3.246965843358	
$7_F \times 91_F$	2.618033988750	3.000000000000	3.246982133140	3.258435734303
$8_F \times 8_F$	2.618033988750	3.000359693703	3.095706393163	
$8_F \times 16_F$	2.618033988750	3.000000000000	3.229632685380	
$8_F \times 24_F$	2.618033988750	3.000000000000	3.246928323759	
$8_F \times 32_F$	2.618033988750	3.000000000000	3.246979586275	
$8_F \times 40_F$	2.618033988750	3.000000000000	3.246979603712	
$8_F \times 48_F$	2.618033988750	3.000000000000	3.246979603717	
$8_F \times 56_F$	2.618033988750	3.000000000000	3.246979603717	
$8_F \times 64_F$	2.618033988750	3.000000000000	3.246979603717	
$8_F \times 72_F$	2.618033988750	3.000000000000	3.246979603717	
$8_F \times 80_F$	2.618033988750	3.000000000000	3.246979603717	
$8_F \times 88_F$	2.618033988750	3.000000000000	3.246979603717	
$8_F \times 96_F$	2.618033988750	3.000000000000	3.246979603717	
$9_F \times 9_F$	2.618033988750	2.999999518372		
$9_F \times 18_F$	2.618033988750	3.000000000000	3.246969773686	
$9_F \times 27_F$	2.618033988750	3.000000000000	3.246979603720	3.342943823308
$9_F \times 36_F$	2.618033988750	3.000000000000	3.246979603717	
$9_F \times 45_F$	2.618033988750	3.000000000000	3.246979603717	3.374646284957
$9_F \times 54_F$	2.618033988750	3.000000000000	3.246979603717	
$9_F \times 63_F$	2.618033988750	3.000000000000	3.246979603717	3.387946181123
$9_F \times 72_F$	2.618033988750	3.000000000000	3.246979603717	
$9_F \times 81_F$	2.618033988750	3.000000000000	3.246979603717	3.395349738491
$9_F \times 90_F$	2.618033988750	3.000000000000	3.246979603717	
Beraha	2.618033988750	3	3.246979603717	3.414213562373

Table 2: Real zeros of the chromatic polynomials of finite triangular-lattice strips with free boundary conditions in both directions, to 12 decimal places. We use the same notation as in Table 1.

Lattice	4th Zero	5th Zero	6th Zero	7th Zero	8th Zero
$4_P \times 4_F$	2.617986010522	3	3.465246100723		
$4_P \times 8_F$	2.618033988740	3	3.475055224065		
$4_P \times 12_F$	2.618033988750	3	3.477452996799		
$4_P \times 16_F$	2.618033988750	3	3.478536268722		
$4_P \times 20_F$	2.618033988750	3	3.479153472532		
$4_P \times 24_F$	2.618033988750	3	3.479552148708		
$4_P \times 28_F$	2.618033988750	3	3.479830901859		
$4_P \times 32_F$	2.618033988750	3	3.480036768366		
$4_P \times 36_F$	2.618033988750	3	3.480195030232		
$4_P \times 40_F$	2.618033988750	3	3.480320488501		
$5_P \times 5_F$	2.618033990394	3			
$5_P \times 10_F$	2.618033988750	3	3.196843987850		
$5_P \times 15_F$	2.618033988750	3			
$5_P \times 20_F$	2.618033988750	3	3.202699178454		
$5_P \times 25_F$	2.618033988750	3			
$5_P \times 30_F$	2.618033988750	3	3.204333275156		
$5_P \times 35_F$	2.618033988750	3			
$5_P \times 40_F$	2.618033988750	3	3.205100311429		
$5_P \times 45_F$	2.618033988750	3			
$5_P \times 50_F$	2.618033988750	3	3.205545558020		
$6_P \times 6_F$	2.618033988750	3.001033705947	3.125892136302		
$6_P \times 12_F$	2.618033988750	3.000000003803	3.198900652620		
$6_P \times 18_F$	2.618033988750	3.000000000000	3.217111179820		
$6_P \times 24_F$	2.618033988750	3.000000000000	3.225649637432		
$6_P \times 30_F$	2.618033988750	3.000000000000	3.230657835149		
$6_P \times 36_F$	2.618033988750	3.000000000000	3.233968503481		
$6_P \times 42_F$	2.618033988750	3.000000000000	3.236327213212		
$6_P \times 48_F$	2.618033988750	3.000000000000	3.238096251767		
$6_P \times 54_F$	2.618033988750	3.000000000000	3.239473538415		
$6_P \times 60_F$	2.618033988750	3.000000000000	3.240576619481		
$6_P \times 66_F$	2.618033988750	3.000000000000	3.241479828709		
$6_P \times 72_F$	2.618033988750	3.000000000000	3.242232528364		
$6_P \times 78_F$	2.618033988750	3.000000000000	3.242868805497		
$6_P \times 84_F$	2.618033988750	3.000000000000	3.243412961909		
$6_P \times 90_F$	2.618033988750	3.000000000000	3.243882786313		
$7_P \times 7_F$	2.618033988750	3	3.247001348628	3.404690481534	
$7_P \times 14_F$	2.618033988750	3	3.246979603718	3.414217072295	3.458917430738
$7_P \times 21_F$	2.618033988750	3	3.246979603717	3.414213561735	
$7_P \times 28_F$	2.618033988750	3	3.246979603717	3.414213562373	3.470544903913
$7_P \times 35_F$	2.618033988750	3	3.246979603717	3.414213562373	
$7_P \times 42_F$	2.618033988750	3	3.246979603717	3.414213562373	3.473634831556
$7_P \times 49_F$	2.618033988750	3	3.246979603717	3.414213562373	
$7_P \times 56_F$	2.618033988750	3	3.246979603717	3.414213562373	3.475070205361
$7_P \times 63_F$	2.618033988750	3	3.246979603717	3.414213562373	
$7_P \times 70_F$	2.618033988750	3	3.246979603717	3.414213562373	3.475899672990
$8_P \times 8_F$	2.618033988750	3	3.246979601854		
$8_P \times 16_F$	2.618033988750	3	3.246979603717	3.414214415195	3.472683999084
$8_P \times 24_F$	2.618033988750	3	3.246979603717	3.414213562387	3.488644630018
$8_P \times 32_F$	2.618033988750	3	3.246979603717	3.414213562373	3.495735217349
$8_P \times 40_F$	2.618033988750	3	3.246979603717	3.414213562373	3.499773262291
$8_P \times 48_F$	2.618033988750	3	3.246979603717	3.414213562373	3.502387969424
$8_P \times 56_F$	2.618033988750	3	3.246979603717	3.414213562373	3.504221641913
$8_P \times 64_F$	2.618033988750	3	3.246979603717	3.414213562373	3.505579831565
$8_P \times 72_F$	2.618033988750	3	3.246979603717	3.414213562373	3.506626776159
$8_P \times 80_F$	2.618033988750	3	3.246979603717	3.414213562373	3.507458740757
Beraha	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238

Table 3: Real zeros of the chromatic polynomials of finite triangular-lattice strips with periodic boundary conditions in the transverse direction and free boundary conditions in the longitudinal direction, to 12 decimal places. A blank means that the zero in question is absent. The first three real zeros $q = 0, 1, 2$ are exact on all lattices. "Beraha" indicates the Beraha numbers $B_5 = (3+\sqrt{5})/2$, $B_6 = 3$, B_7, $B_8 = 2+\sqrt{2}$, and B_9.

Lattice	4th Zero	5th Zero	6th Zero	7th Zero	8th Zero	9th Zero	10th Zero
$9_P \times 9_F$	2.618033988750	3.000000000000	3.246980644227	3.382733076359			
$9_P \times 18_F$	2.618033988750	3.000000000000	3.246979603717	3.414215827400	3.467483864312		
$9_P \times 27_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562359			
$9_P \times 36_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.499429426359		
$9_P \times 45_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373			
$9_P \times 54_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.508825024982		
$9_P \times 63_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373			
$9_P \times 72_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.513393802382		
$9_P \times 81_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373			
$9_P \times 90_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.516109505154		
$10_P \times 10_F$	2.618033988750	3	3.246979603717	3.414213601215	3.522072913706		
$10_P \times 20_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088885496		
$10_P \times 30_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238		
$10_P \times 40_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238		
$10_P \times 50_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238		
$10_P \times 60_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.618274945403	3.620352727045
$10_P \times 70_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.618041117772	3.623157797032
$10_P \times 80_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.618034257877	3.624885448640
$10_P \times 90_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.618033998995	3.626138707957
$10_P \times 100_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.618033989140	3.627101535574
$11_P \times 11_F$	2.618033988750	3	3.246979603717	3.414213539527			
$11_P \times 22_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088885575		
$11_P \times 33_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.608601511861	
$11_P \times 44_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238		
$11_P \times 55_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.617975980728	
$11_P \times 66_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.618034451624	3.627344614702
$11_P \times 77_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.618033985241	
$11_P \times 88_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.618033988777	3.631979435582
$11_P \times 99_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.618033988750	
$11_P \times 110_F$	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.618033988750	3.634574709990
$12_P \times 12_F$	2.618033988750	3.000000000000	3.246979603717	3.414213593041	3.511032635472		
$12_P \times 24_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.532088885001		
$12_P \times 36_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.532088886238		
$12_P \times 48_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.532088886238		
$12_P \times 60_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.532088886238		
$12_P \times 72_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.532088886238	3.618040035384	3.624320958404
$12_P \times 84_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.532088886238	3.618034035926	3.627419917635
$12_P \times 96_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.532088886238	3.618033989120	3.629588218978
$12_P \times 108_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.532088886238	3.618033988753	3.631215401061
$12_P \times 120_F$	2.618033988750	3.000000000000	3.246979603717	3.414213562373	3.532088886238	3.618033988750	3.632487726562
Beraha	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.618033988750	3.682507065662

Table 4: Real zeros of the chromatic polynomials of finite triangular-lattice strips with periodic boundary conditions in the transverse direction and free boundary conditions in the longitudinal direction, to 12 decimal places. We use the same notation as in Table 3.

Lattice	4th Zero	5th Zero	6th Zero	7th Zero	8th Zero	9th Zero	10th Zero
$4_Z \times 4_F$	2.485072022789	2.527537649962	2.596617094656				
$4_Z \times 8_F$	2.499965989337	2.500034085574	2.618031965217				
$4_Z \times 12_F$	2.499999937358	2.500000062643	2.618033988527				
$4_Z \times 16_F$	2.499999999885	2.500000000115	2.618033988750				
$4_Z \times 20_F$	2.500000000000	2.500000000000	2.618033988750				
$4_Z \times 24_F$	2.500000000000	2.500000000000	2.618033988750				
$4_Z \times 28_F$	2.500000000000	2.618033988750					
$4_Z \times 32_F$	2.500000000000	2.618033988750					
$4_Z \times 36_F$	2.500000000000	2.618033988750					
$4_Z \times 40_F$	2.500000000000	2.618033988750					
$6_Z \times 6_F$	2.618033988528						
$6_Z \times 12_F$	2.618033988750	3.000017186720	3.117917986708				
$6_Z \times 18_F$	2.618033988750	3.000000004191	3.141107899326				
$6_Z \times 24_F$	2.618033988750	3.000000000001	3.150834657646				
$6_Z \times 30_F$	2.618033988750	3.000000000000	3.156227017803				
$6_Z \times 36_F$	2.618033988750	2.722632835458	3.000000000000	3.159661924115			
$6_Z \times 42_F$	2.618033988750	2.722632835458	3.000000000000	3.162043675850			
$6_Z \times 48_F$	2.618033988750	2.722632835458	3.000000000000	3.163793040154			
$6_Z \times 54_F$	2.618033988750	2.722632835458	3.000000000000	3.165132700997			
$6_Z \times 60_F$	2.618033988750	2.722632835458	3.000000000000	3.166191662980			
$8_Z \times 8_F$	2.618033988750	3.000000844168	3.203925019292				
$8_Z \times 16_F$	2.618033988750	3.000000000000	3.246976356780				
$8_Z \times 24_F$	2.618033988750	3.000000000000	3.246979603696				
$8_Z \times 32_F$	2.618033988750	2.821420495535	3.000000000000	3.246979603717			
$8_Z \times 40_F$	2.618033988750	2.821420495535	3.000000000000	3.246979603717			
$8_Z \times 48_F$	2.618033988750	2.821420495535	3.000000000000	3.246979603717			
$8_Z \times 56_F$	2.618033988750	2.821420495535	3.000000000000	3.246979603717			
$8_Z \times 64_F$	2.618033988750	2.821420495535	3.000000000000	3.246979603717			
$8_Z \times 72_F$	2.618033988750	2.821420495535	3.000000000000	3.246979603717			
$8_Z \times 80_F$	2.618033988750	2.821420495535	3.000000000000	3.246979603717			
$10_Z \times 10_F$	2.618033988750	3.000000000000	3.246953122227				
$10_Z \times 20_F$	2.618033988750	3.000000000000	3.246979603717	3.416320582746	3.428202969384		
$10_Z \times 30_F$	2.618033988750	3.000000000000	3.246979603717	3.414213564771	3.470075808656		
$10_Z \times 40_F$	2.618033988750	2.873731249334	3.000000000000	3.246979603717	3.414213562373	3.484585415185	
$10_Z \times 50_F$	2.618033988750	2.873731249334	3.000000000000	3.246979603717	3.414213562373	3.492475877808	
$10_Z \times 60_F$	2.618033988750	2.873731249334	3.000000000000	3.246979603717	3.414213562373	3.497477600415	
$10_Z \times 70_F$	2.618033988750	2.873731249334	3.000000000000	3.246979603717	3.383128531235	3.414213562373	3.500942791087
$10_Z \times 80_F$	2.618033988750	2.873731249334	3.000000000000	3.246979603717	3.383128531235	3.414213562373	3.503489509874
$10_Z \times 90_F$	2.618033988750	2.873731249334	3.000000000000	3.246979603717	3.383128531235	3.414213562373	3.505442176204
$10_Z \times 100_F$	2.618033988750	2.873731249334	3.000000000000	3.246979603717	3.383128531235	3.414213562373	3.506987965042
Beraha	2.618033988750	3	3.246979603717	3.414213562373	3.532088886238	3.618033988750	3.682507065662

Table 5: Real zeros of the chromatic polynomials of finite triangular-lattice strips with zig-zag boundary conditions, to 12 decimal places. We use the same notation as in Table 3.

	Eigenvalue-Crossing Curves $\mathcal{B}$								Isolated Points	
Lattice	# C	# E	# T	# D	# ER	$\min \mathrm{Re}\, q$	q_0	$\max \mathrm{Re}\, q$	# RI	# CI
2_F									2	0
3_F	3	6	0	0	0	1.20474	2.56984	3.40223	3	0
4_F	2	6	2	0	0	0.81647	$2.75925 \pm 0.15444\, i^*$	3.63983	4	0
5_F	1	6	12	0	4	0.55862	3	3.77830	4	0
6_F	1	6	4	0	0	0.37963	3.16093	3.86641	5	0
7_F	$1^\dagger$	$6^\dagger$	$4^\dagger$	0	$0^\dagger$	0.25054	3.27640	3.92580	6	0
8_F	$1^\dagger$	$6^\dagger$	$4^\dagger$	0	$0^\dagger$	0.13343	3.36106	3.96756	6	0
9_F							3.42513		7	0
2_P									2	0
3_P									3	0
4_P	3	4	0	0	1	1.37053	3.48141	4	5	0
5_P	3	4	0	0	1	0.47725	3.20722	3.87699	5	0
6_P	1	4	2	0	0	0.02077	3.25242	4.28386	6	0
7_P	1	4	4	0	1	−0.22792	3.47900	3.99964	7	0
8_P	1	4	2	0	0	−0.37137	3.51477	4.04970	7	0
9_P	1	4	2	0	0	−0.45760	3.52706	4.28286	7	0
10_P	$1^\dagger$	$4^\dagger$	$2^\dagger$	0	$0^\dagger$	−0.51081	3.63483	4.12341	9	0
11_P	$1^\dagger$	$4^\dagger$	$2^\dagger$	0	$0^\dagger$	−0.54399	3.64414	4.15609	9	0
12_P			$2^\dagger$				3.64317		9	0
2_Z									2	0
4_Z	2	4	0	0	0	2.09914	$2.73717 \pm 0.17233\, i^*$	4.00485	5	0
6_Z	1	4	2	0	0	0.36185	3.17526	4.25895	6	1
8_Z	$1^\dagger$	$4^\dagger$	$2^\dagger$	0	$0^\dagger$	−0.21435	3.39410	4.28991	7	$1^\dagger$
10_Z							3.52044		9	

Table 6: Summary of qualitative results for the eigenvalue-crossing curves $\mathcal{B}$ and for the isolated limiting points of zeros. For each triangular-lattice strip considered in this paper, we give the number of connected components of $\mathcal{B}$ (# C), the number of endpoints (# E), the number of T points (# T), the number of double points (# D), and the number of enclosed regions (# ER); we also give the minimum value of $\mathrm{Re}\, q$ on $\mathcal{B}$, the smallest value q_0 where $\mathcal{B}$ intersects the real axis (* denotes an almost-crossing), and the maximum value of $\mathrm{Re}\, q$ on $\mathcal{B}$. We also report the number of real isolated limiting points of zeros (# RI) and the number of complex-conjugate pairs of isolated limiting points (# CI). The symbol † indicates uncertain results.

m	TriCyl(m)	# VA	TriCyl$'$(m)	SqCyl(m)
1	1	0	1	1
2	1	0	1	1
3	1	0	1	1
4	2	0	2	2
5	2	0	2	2
6	5	0	5	5
7	6	0	6	6
8	15	2	13	14
9	28	12	16	22
10	67	32	35	51
11	145	100	45	95
12	368	272	96	232
13	870	742*	126*	498
14	2211	1940*	267*	1239

Table 7: Transfer-matrix dimensions for a triangular-lattice strip of width m and cylindrical boundary conditions. For each value of the strip width m we give the dimension of the transfer matrix [TriCyl(m)], the number of vanishing amplitudes (# VA), and the *effective* dimension of the transfer matrix [TriCyl$'$(m) = TriCyl(m) − # VA]. For comparison, we also give the dimensionality of the transfer matrix for a square-lattice strip of width m and cylindrical boundary conditions [SqCyl(m)]. The values of TriCyl(m) and SqCyl(m) were obtained in [16, 53]. An asterisk denotes *conjectured* results.

Curve	Point	$\mathrm{Arg}\, p_{\mathrm{Fit}}/\pi$	$\mathrm{Arg}\, p/\pi$	"discrepancy"	θ/π
C_1	G	1.0000000(3)	1	0.00000005	1/6
	H	−0.4999999(3)	−1/2	0.0000001	2/3
C_2	I	0.5000001(3)	1/2	0.0000002	2/9
	J	0.4285703(3)	3/7	0.000008	7/30
C_3	L	−0.2500003(3)	−1/4	0.000001	4/9
	K	−0.1999993(3)	−1/5	0.000004	5/12
C_4	P	0.2727311(7)	3/11	0.00004	11/42
	O	0.2499995(3)	1/4	0.000002	4/15
C_5	M	−0.1428584(3)	−1/7	0.000009	7/18
	N	−0.1249980(3)	−1/8	0.00002	8/21
C_7	T	−0.3846221(7)	−5/13	0.00009	13/24
	S	−0.3749980(3)	−3/8	0.00002	8/15
C_8		0.1250023(3)	1/8	0.00002	8/27
		0.1199957(7)	3/25	0.0001	25/84
C_9	V	0.2000006(3)	1/5	0.000003	5/18
	U	0.1875171(7)	3/16	0.0003	16/57
C_{13}		0.1034699(7)	3/29	0.0006	29/96
		0.0999991(7)	1/10	0.000009	10/33
C_{14}		0.1500210(7)	3/20	0.0004	20/69
		0.1428558(3)	1/7	0.00001	7/24
C_{17}		−0.4285727(3)	−3/7	0.000009	7/12
		−0.4230929(3)	−11/26	0.0004	26/45
C_{18}		−0.1000025(3)	−1/10	0.00003	10/27
		−0.0909045(7)	−1/11	0.00005	11/30
C_{19}		−0.0769285(7)	−1/13	0.00007	13/36
		−0.0714257(7)	−1/14	0.00004	14/39
Others		−0.7500002(3)	−3/4	0.00000009	4/27
		−0.7727176(7)	−17/22	0.0002	22/147
		−0.5999993(3)	−3/5	0.000004	5/6
		0.7499995(3)	3/4	0.000002	4/21
		0.6000006(3)	3/5	0.000003	5/24
		0.2999986(7)	3/10	0.00001	10/39

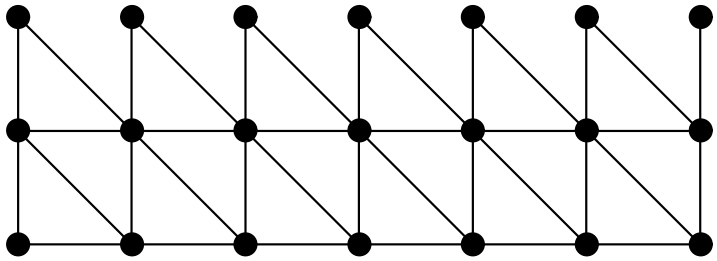
Table 8: Endpoints in the complex p-plane with $|p| = 1$ of the curves where $|g_2/g_1| = 1$. For each endpoint we show the "Curve" to which it belongs (see Figure 22), the estimated value of its phase $\mathrm{Arg}\, p_{\mathrm{Fit}}$ (see text), the conjectured exact value $\mathrm{Arg}\, p$, the "discrepancy" (6.10), and the corresponding "primary" θ value. For some selected values we also include a label ("Point"). When we have a curve for which one endpoint is well-determined and the other is not, we include the former point in the category "Others" (see text).

Curve	Point	Arg $y_{\rm Fit}/\pi$	Arg y/π	"discrepancy"	θ/π
D_1	G	$-0.8000002(3)$	$-4/5$	0.000001	1/6
	H	0.0000000(6)	0	0.00000002	2/3
D_2	K	0.8571435(3)	6/7	0.000005	5/12
	L	0.7999997(3)	4/5	0.000002	4/9
D_3	J	$-0.8695494(7)$	$-20/23$	0.0004	7/30
	I	$-0.8571420(1)$	$-6/7$	0.000006	2/9
D_4		$-0.7692308(7)$	$-10/13$	0.0000004	2/15
		$-0.7272743(3)$	$-8/11$	0.00002	1/12
D_5		$-0.5263122(7)$	$-10/19$	0.00007	53/72
		$-0.4827684(7)$	$-14/29$	0.0003	79/108
D_6		$-0.4615359(7)$	$-6/13$	0.00003	35/48
		$-0.4000003(3)$	$-2/5$	0.000001	13/18
D_7		$-0.5714279(3)$	$-4/7$	0.000004	20/27
		$-0.5454561(3)$	$-6/11$	0.00002	31/42
D_9	N	0.9230794(3)	12/13	0.00003	8/21
	M	0.9090890(3)	10/11	0.00002	7/18
D_{11}	U	$-0.9268596(7)$	$-38/41$	0.001	16/57
	V	$-0.9230785(7)$	$-12/13$	0.00002	5/18
D_{13}		0.2857149(7)	2/7	0.000004	11/18
		0.2816705(7)	20/71	0.001	112/183
D_{14}		$-0.7200434(7)$	$-18/25$	0.001	2/27
		$-0.7058815(7)$	$-12/17$	0.00001	1/18
D_{15}	P	$-0.9032084(7)$	$-28/31$	0.0005	11/42
	O	$-0.9090929(3)$	$-10/11$	0.00002	4/15
Others		$-0.4705939(7)$	$-8/17$	0.0001	46/63
		$-0.4800291(7)$	$-12/25$	0.0007	68/93
		$-0.5161460(7)$	$-16/31$	0.0005	86/117
		$-0.5217586(3)$	$-12/23$	0.0004	64/87
		$-0.6153817(3)$	$-8/13$	0.00004	38/51
		$-0.6315679(7)$	$-12/19$	0.0002	56/75
		$-0.6956730(7)$	$-16/23$	0.0005	1/24
	S	0.5714292(3)	4/7	0.000004	8/15
		0.5599911(7)	14/25	0.0002	29/54
		0.5516847(7)	16/29	0.001	34/63
	T	0.5454538(7)	6/11	0.000008	13/24
		0.3999997(3)	2/5	0.000002	7/12

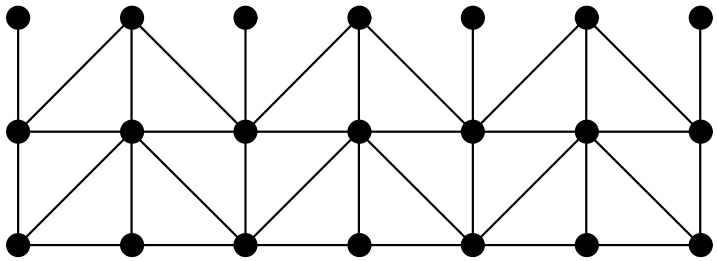
Table 9: Endpoints in the complex y-plane with $|y| = 1$ of the curves where $|g_3/g_1| = 1$. For each endpoint we show the "Curve" to which it belongs (see Figure 23), the estimated value of its phase Arg $y_{\rm Fit}$ (see text), the conjectured exact value Arg y, the "discrepancy" (6.10), and the corresponding "primary" θ value. For some selected values we also include a label ("Point"). When we have a curve for which one endpoint is well-determined and the other is not, we include the former point in the category "Others" (see text).

L	Beraha numbers
$2_{\rm F}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3$
$3_{\rm F}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4$
$4_{\rm F}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ B_6$
$5_{\rm F}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ B_6$
$6_{\rm F}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ B_7$
$7_{\rm F}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ \boldsymbol{B}_7\ B_8$
$8_{\rm F}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ \boldsymbol{B}_7\ B_8\ B_9$
$9_{\rm F}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ \boldsymbol{B}_7\ \boldsymbol{B}_8\ B_9\ B_{10}$
$2_{\rm P}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3$
$3_{\rm P}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4$
$4_{\rm P}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ B_6$
$5_{\rm P}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6$
$6_{\rm P}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ \boldsymbol{B}_7\ B_{10}$
$7_{\rm P}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ \boldsymbol{B}_7\ \boldsymbol{B}_8\ B_{10}$
$8_{\rm P}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ \boldsymbol{B}_7\ \boldsymbol{B}_8\ B_9\ B_{10}\ B_{14}$
$9_{\rm P}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ \boldsymbol{B}_7\ \boldsymbol{B}_8\ B_9\ B_{10}\ B_{14}$
$10_{\rm P}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ \boldsymbol{B}_7\ \boldsymbol{B}_8\ \boldsymbol{B}_9\ \boldsymbol{B}_{10}\ B_{11}\ B_{14}\ B_{18}$
$11_{\rm P}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ \boldsymbol{B}_7\ \boldsymbol{B}_8\ \boldsymbol{B}_9\ \boldsymbol{B}_{10}\ B_{11}\ B_{12}\ B_{14}\ B_{18}$
$12_{\rm P}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ \boldsymbol{B}_7\ \boldsymbol{B}_8\ \boldsymbol{B}_9\ \boldsymbol{B}_{10}\ B_{11}\ B_{12}\ B_{13}\ B_{14}\ B_{18}\ B_{22}$
$2_{\rm Z}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3$
$4_{\rm Z}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ B_6$
$6_{\rm Z}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ B_7$
$8_{\rm Z}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ \boldsymbol{B}_7\ B_8\ B_9$
$10_{\rm Z}$	$\boldsymbol{B}_2\ \boldsymbol{B}_3\ \boldsymbol{B}_4\ \boldsymbol{B}_5\ \boldsymbol{B}_6\ \boldsymbol{B}_7\ \boldsymbol{B}_8\ B_9\ B_{10}\ B_{11}$

Table 10: Beraha numbers B_n that are zeros of $\det D(q)$. Those shown in boldface (resp. normal face) correspond to the vanishing of a dominant (resp. subdominant) amplitude.



(a)



(b)

Figure 1: Two ways of building a triangular-lattice strip using a transfer-matrix approach. (a) Standard method (see e.g. Ref. [16]). (b) Alternative method (called "zig-zag" boundary conditions).

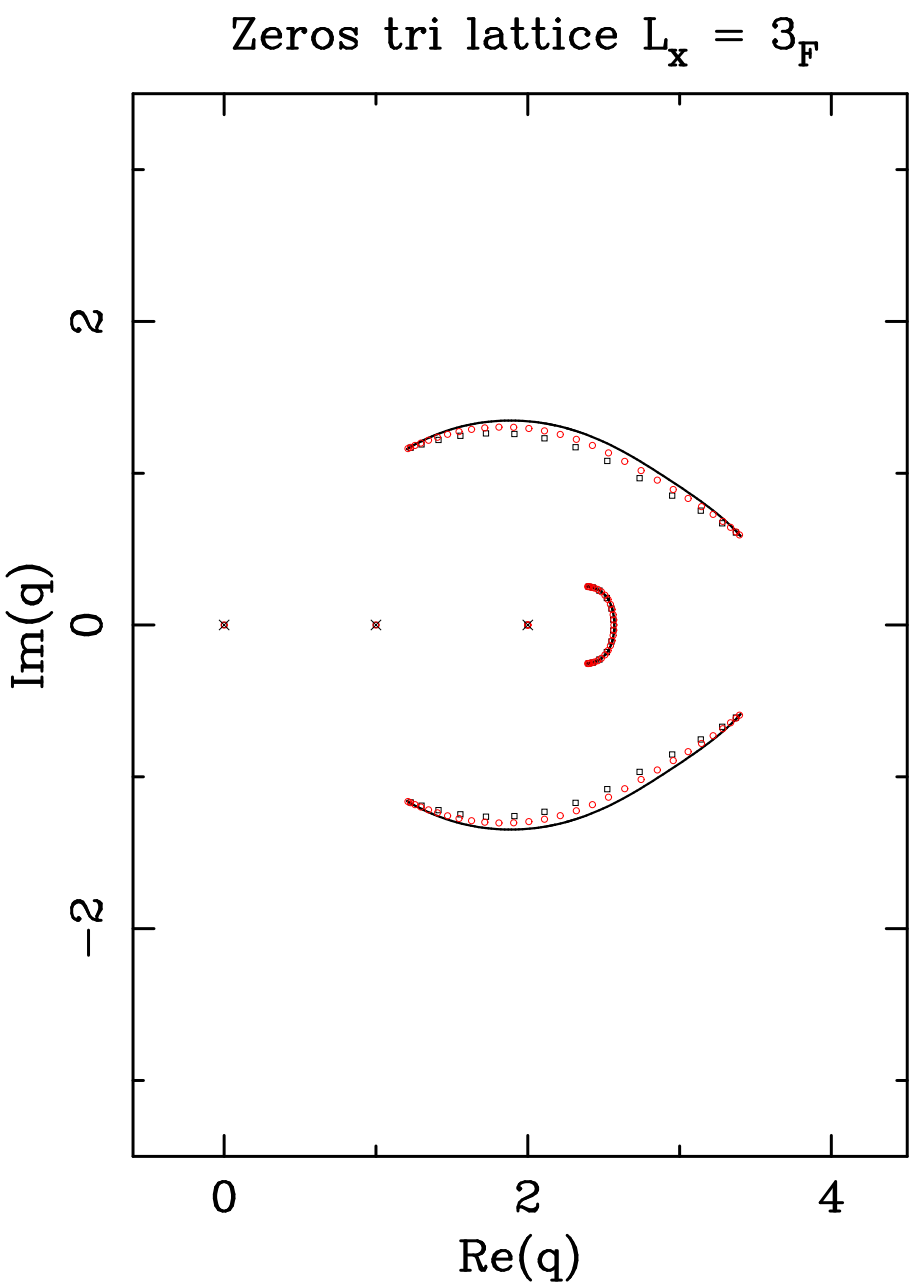


Figure 2: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $3_F \times 15_F$ (squares), $3_F \times 30_F$ (circles) and $3_F \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

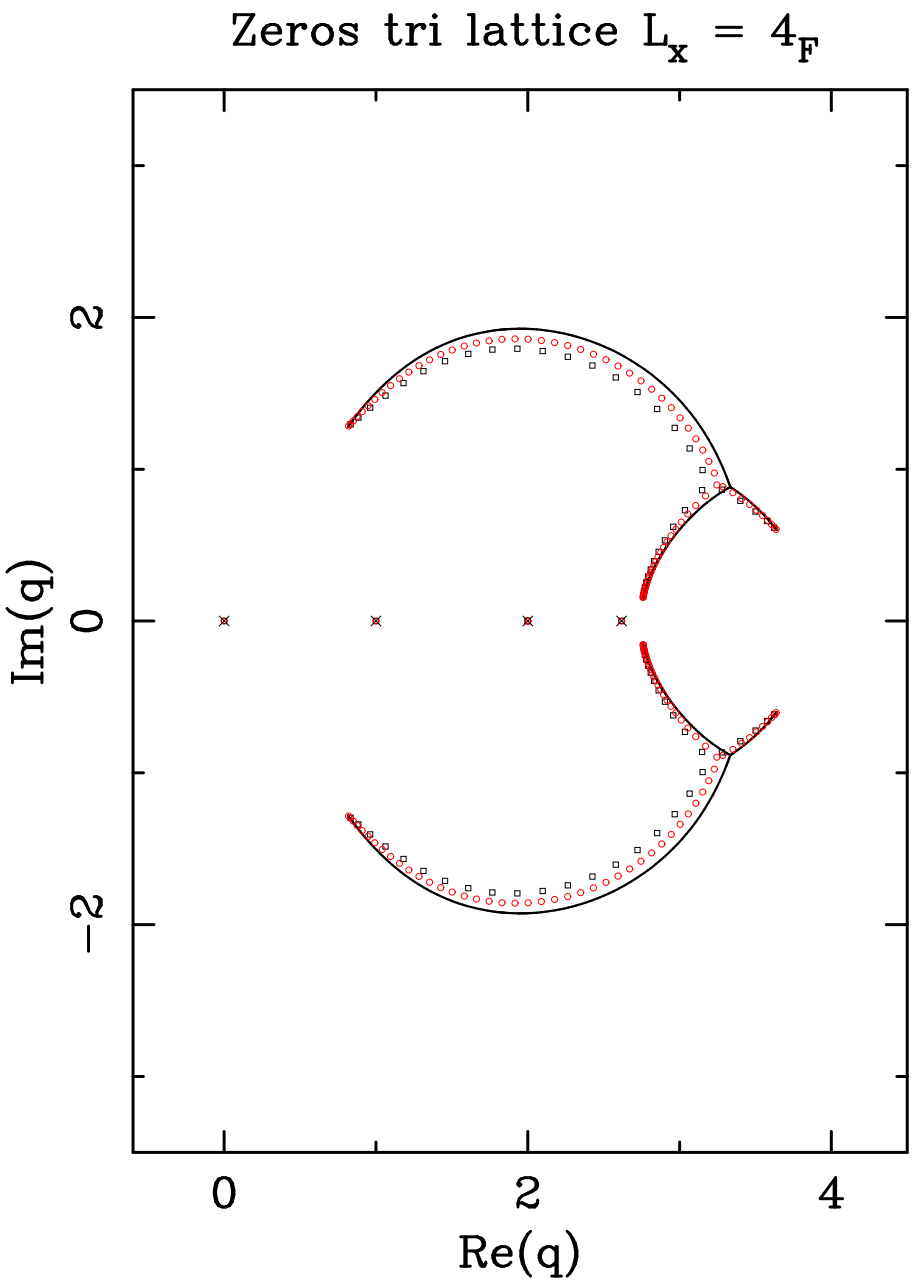


Figure 3: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $4_F \times 20_F$ (squares), $4_F \times 40_F$ (circles) and $4_F \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

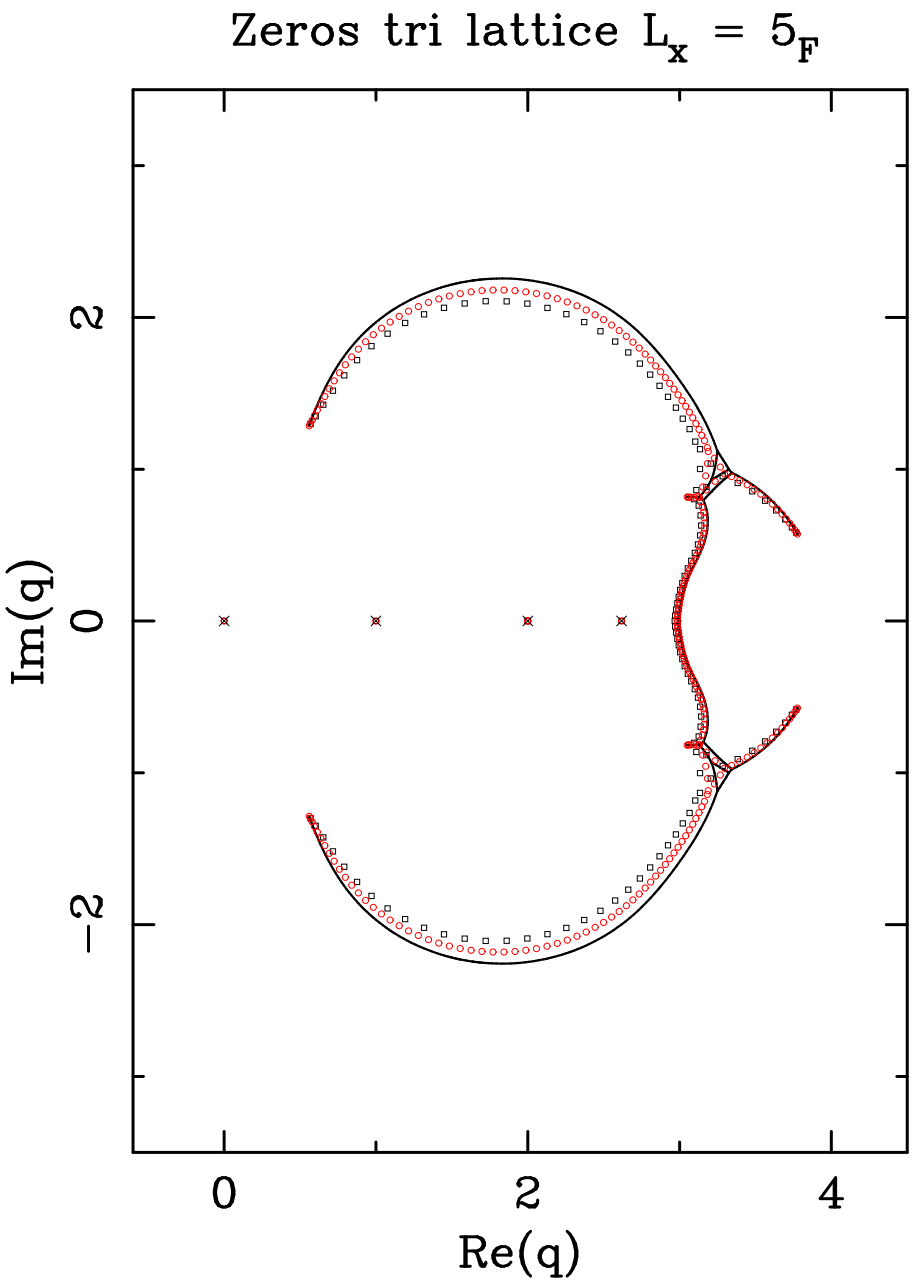


Figure 4: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $5_F \times 25_F$ (squares), $5_F \times 50_F$ (circles) and $5_F \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

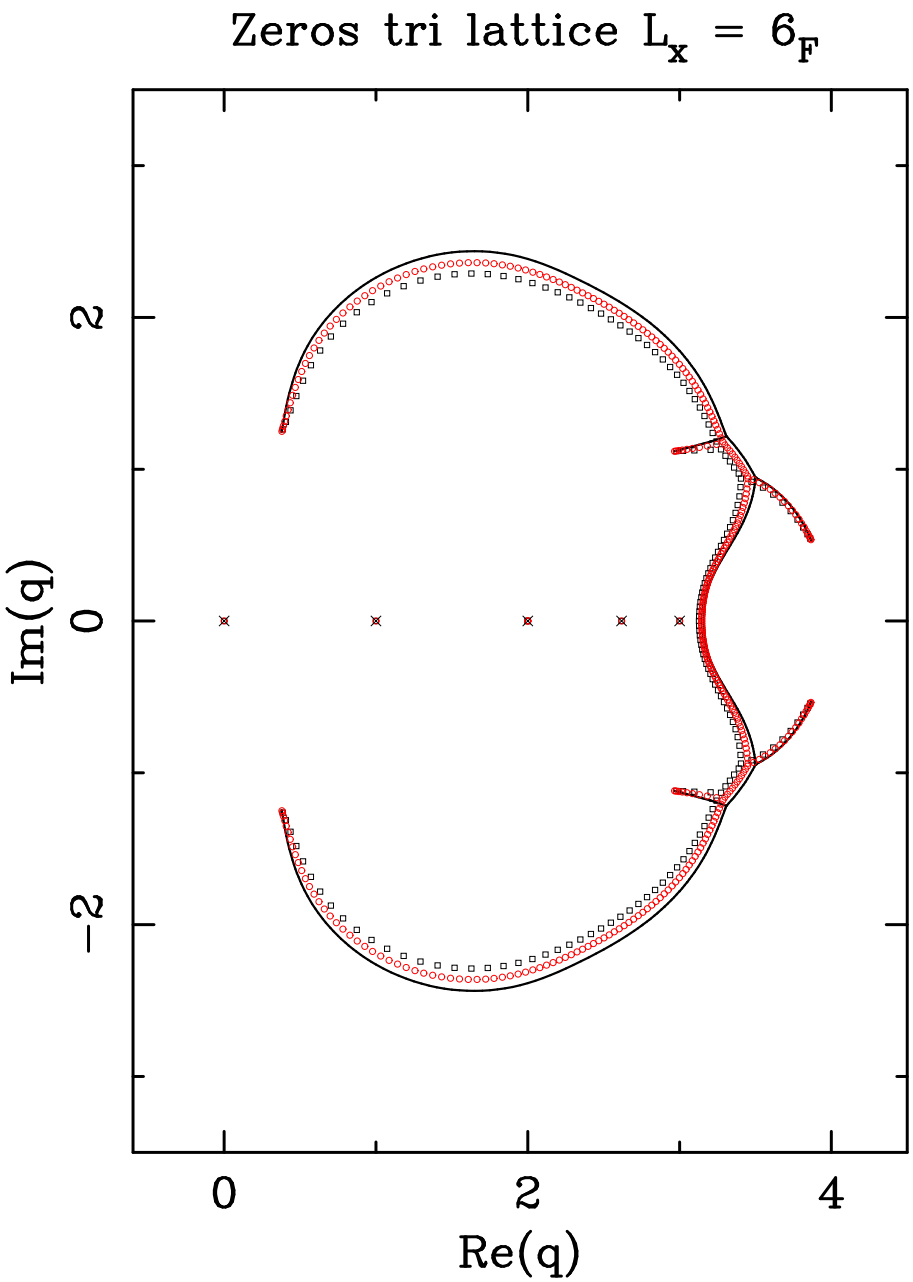


Figure 5: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $6_F \times 30_F$ (squares), $6_F \times 60_F$ (circles) and $6_F \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

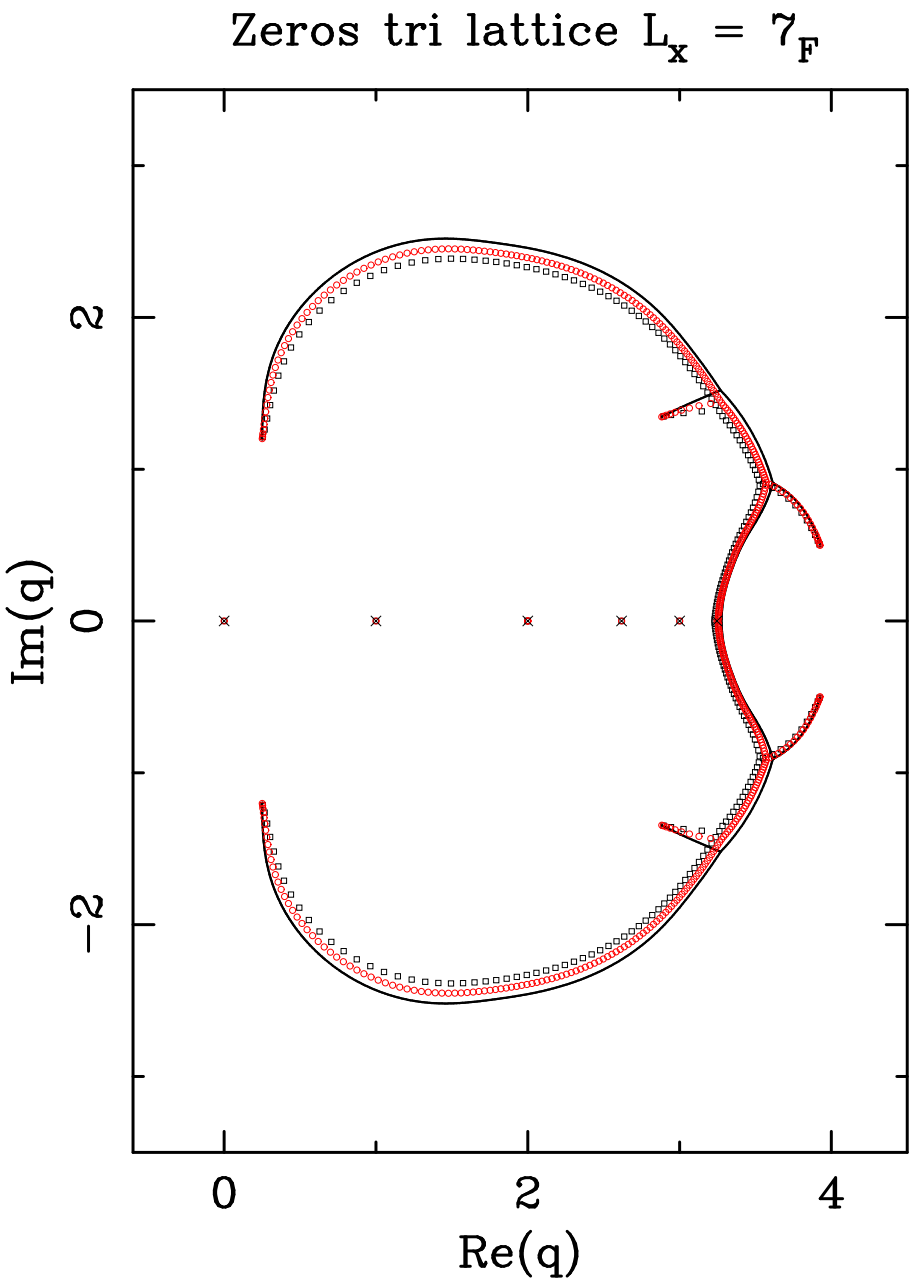


Figure 6: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $7_F \times 35_F$ (squares), $7_F \times 70_F$ (circles) and $7_F \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

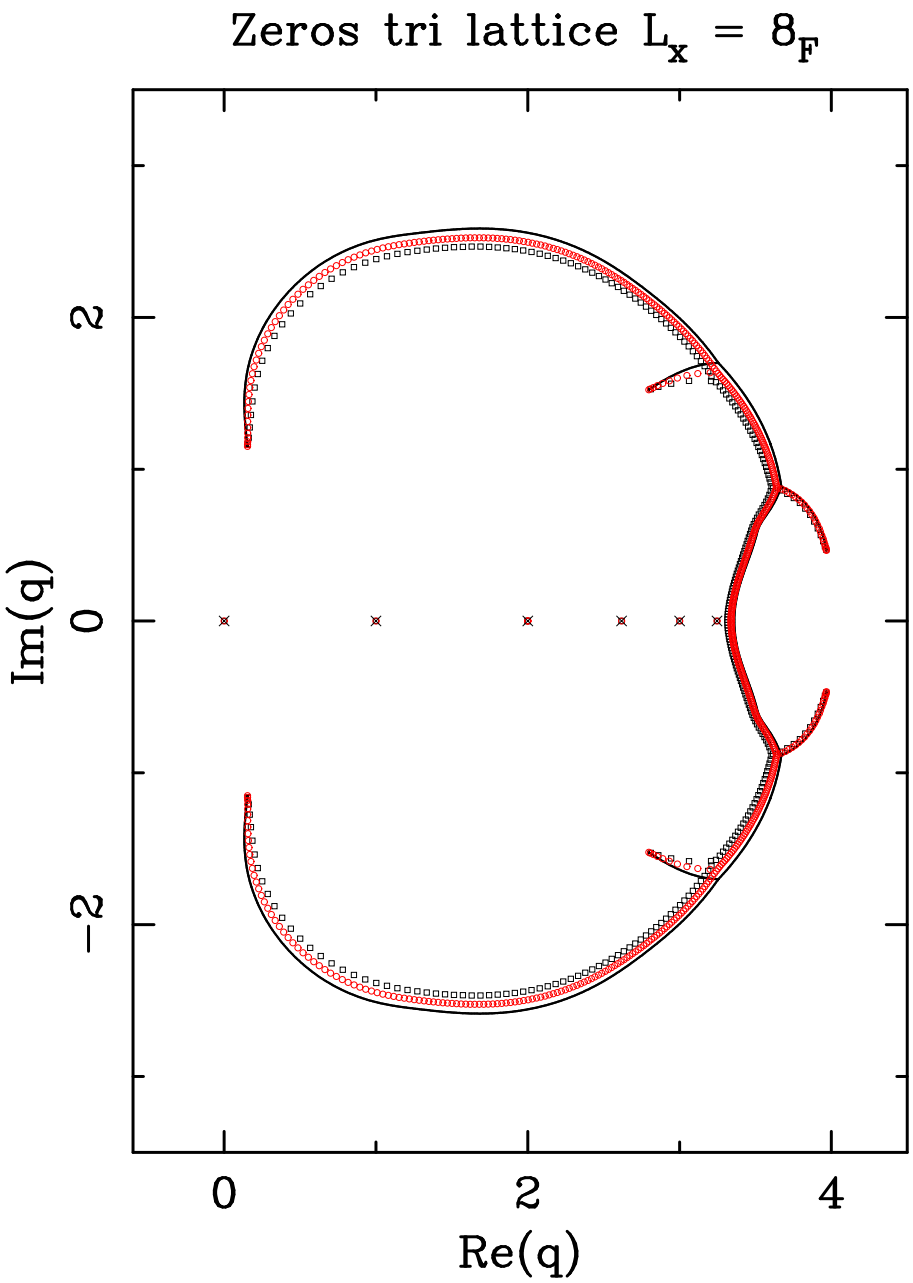


Figure 7: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $8_F \times 40_F$ (squares), $8_F \times 80_F$ (circles) and $8_F \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

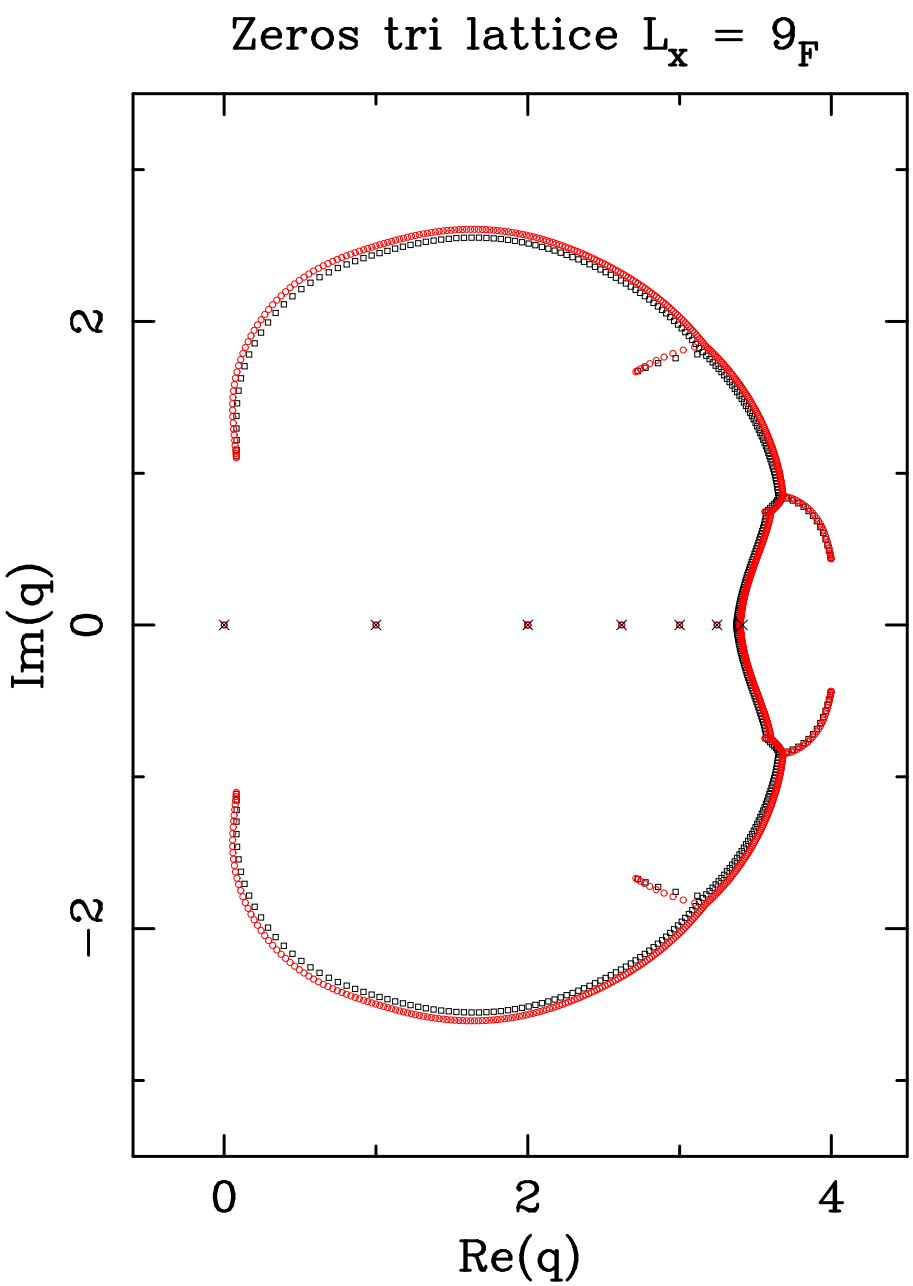


Figure 8: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $9_F \times 45_F$ (squares) and $9_F \times 90_F$ (circles). The isolated limiting points are denoted by a $\times$.

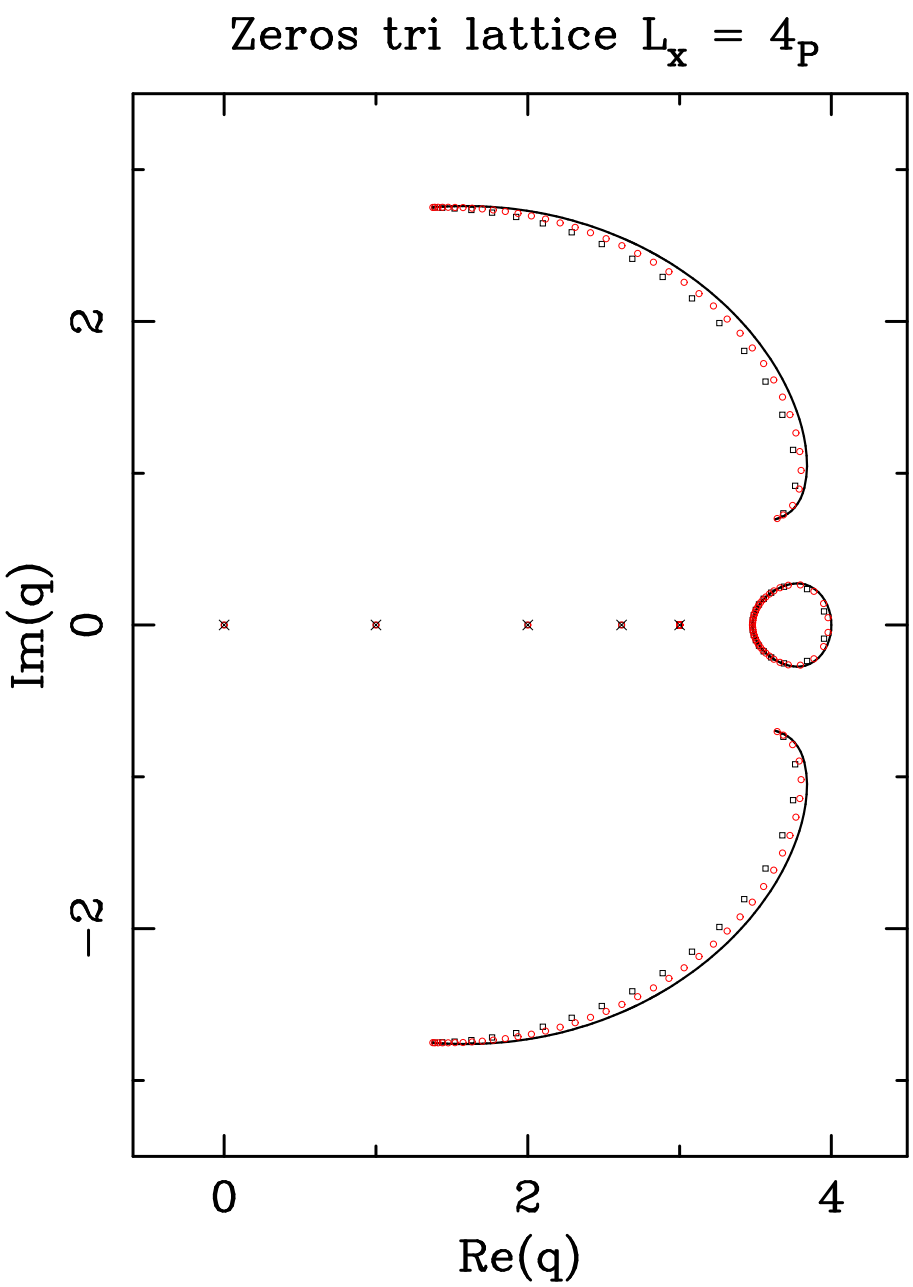


Figure 9: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $4_P \times 20_F$ (squares), $4_P \times 40_F$ (circles) and $4_P \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

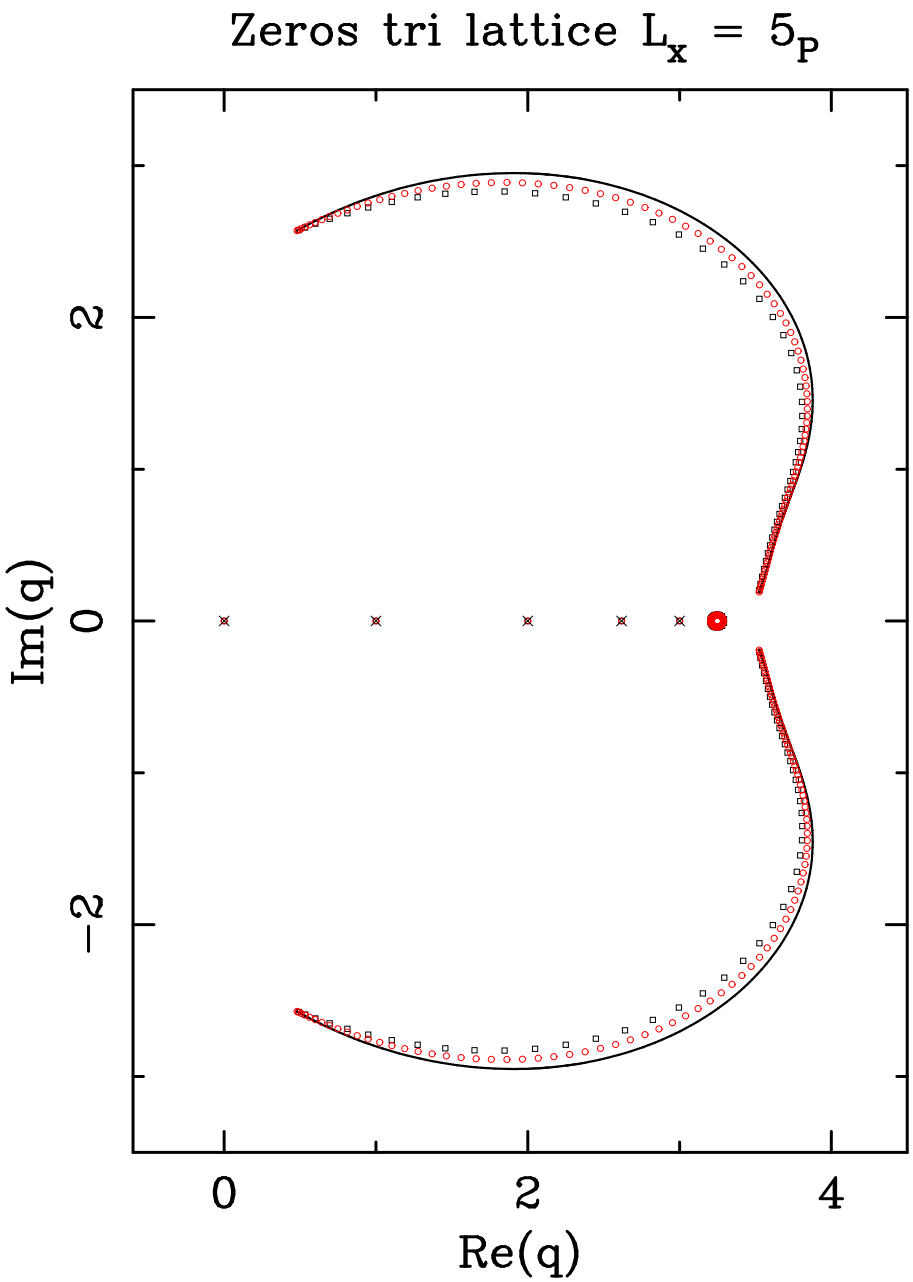


Figure 10: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $5_P \times 25_F$ (squares), $5_P \times 50_F$ (circles) and $5_P \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

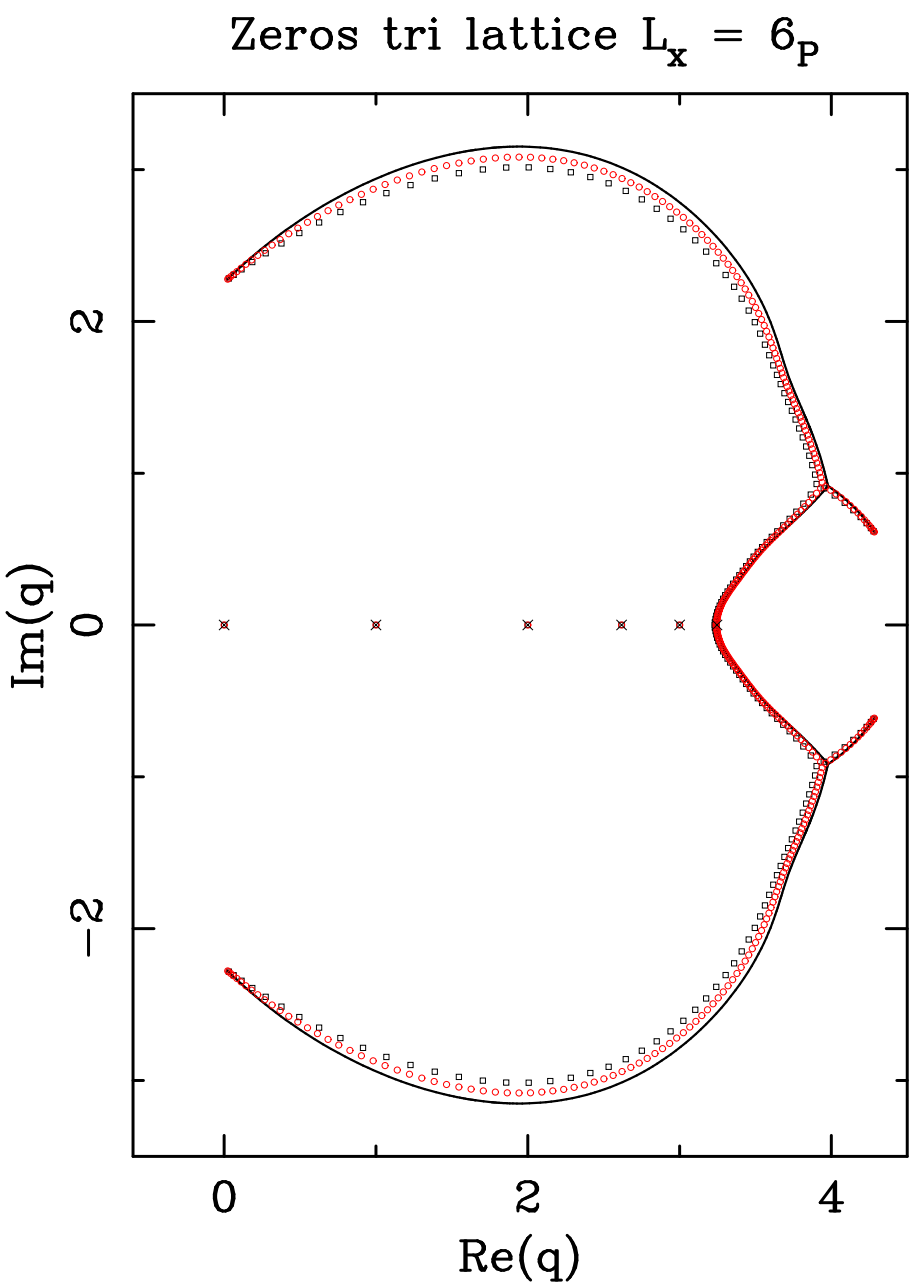


Figure 11: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $6_P \times 30_F$ (squares), $6_P \times 60_F$ (circles) and $6_P \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

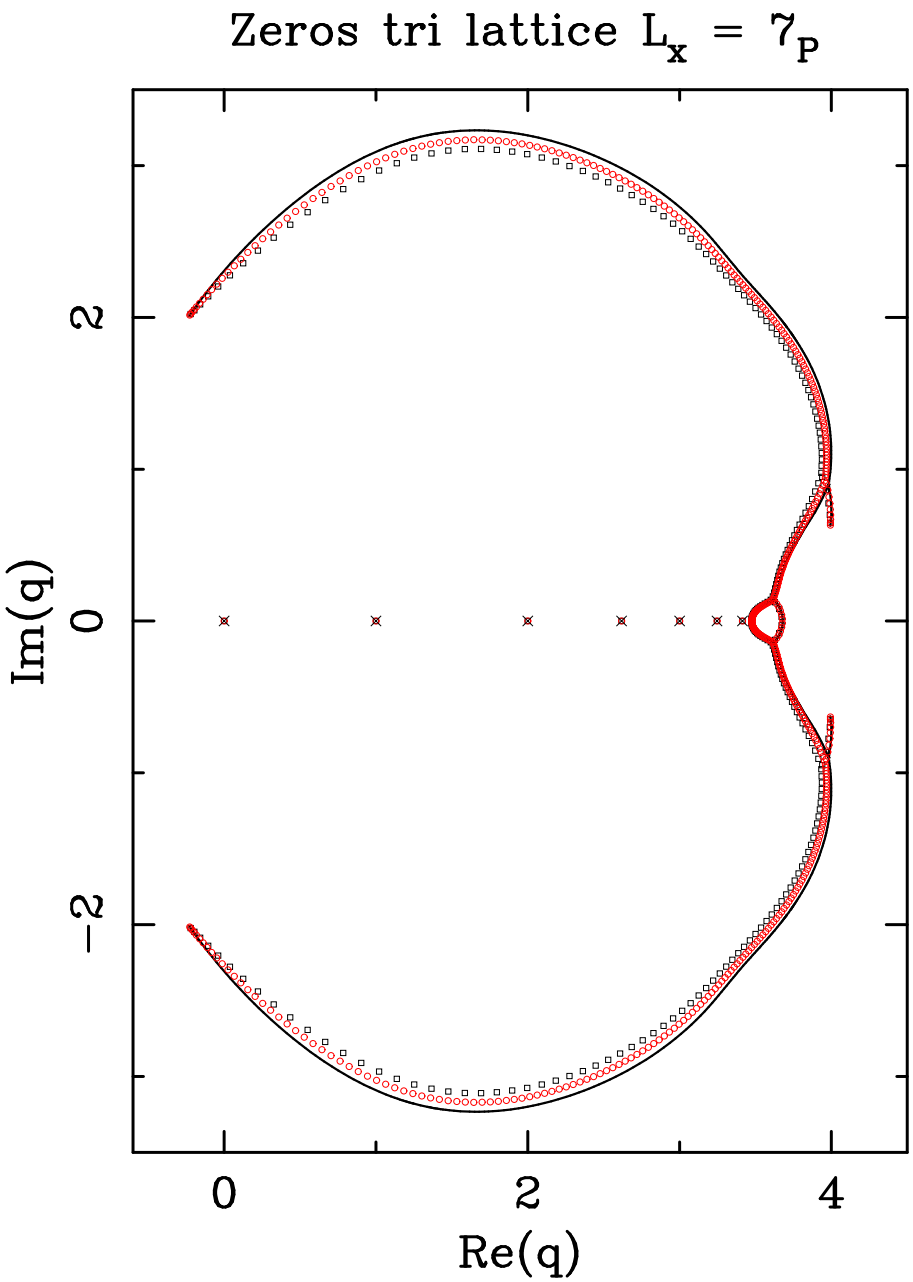


Figure 12: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $7_P \times 35_F$ (squares), $7_P \times 70_F$ (circles) and $7_P \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

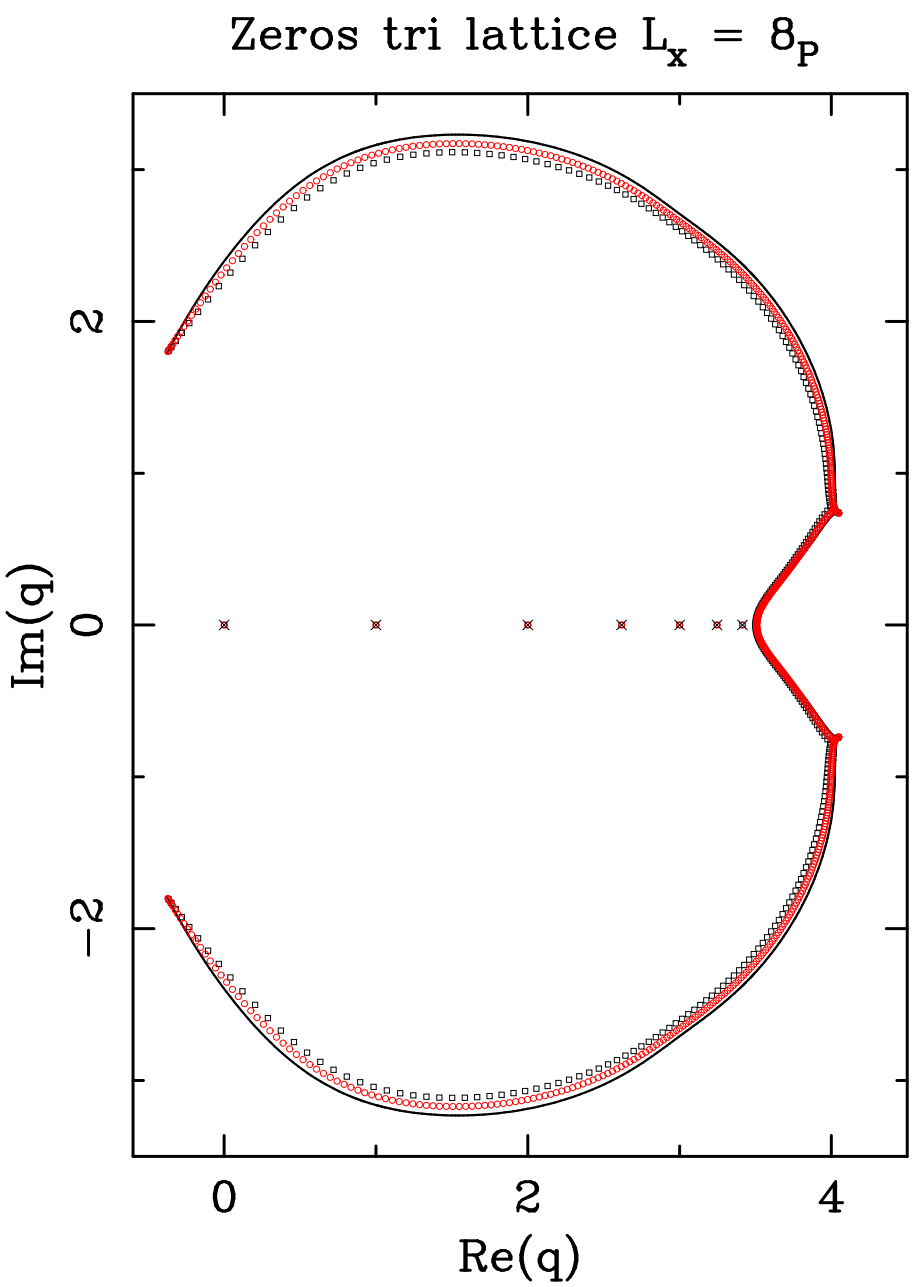


Figure 13: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $8_P \times 40_F$ (squares), $8_P \times 80_F$ (circles) and $8_P \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

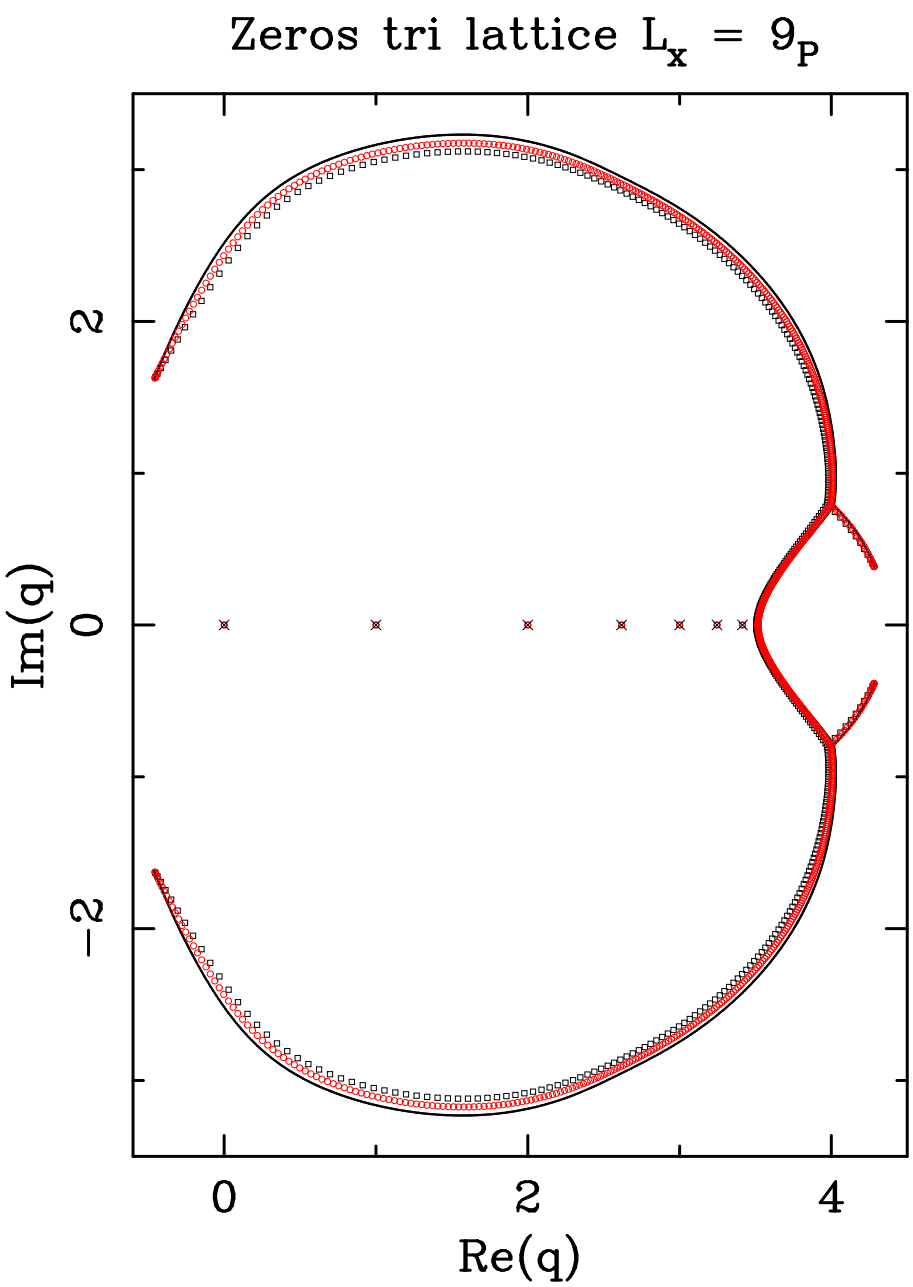


Figure 14: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $9_P \times 45_F$ (squares), $9_P \times 90_F$ (circles) and $9_P \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

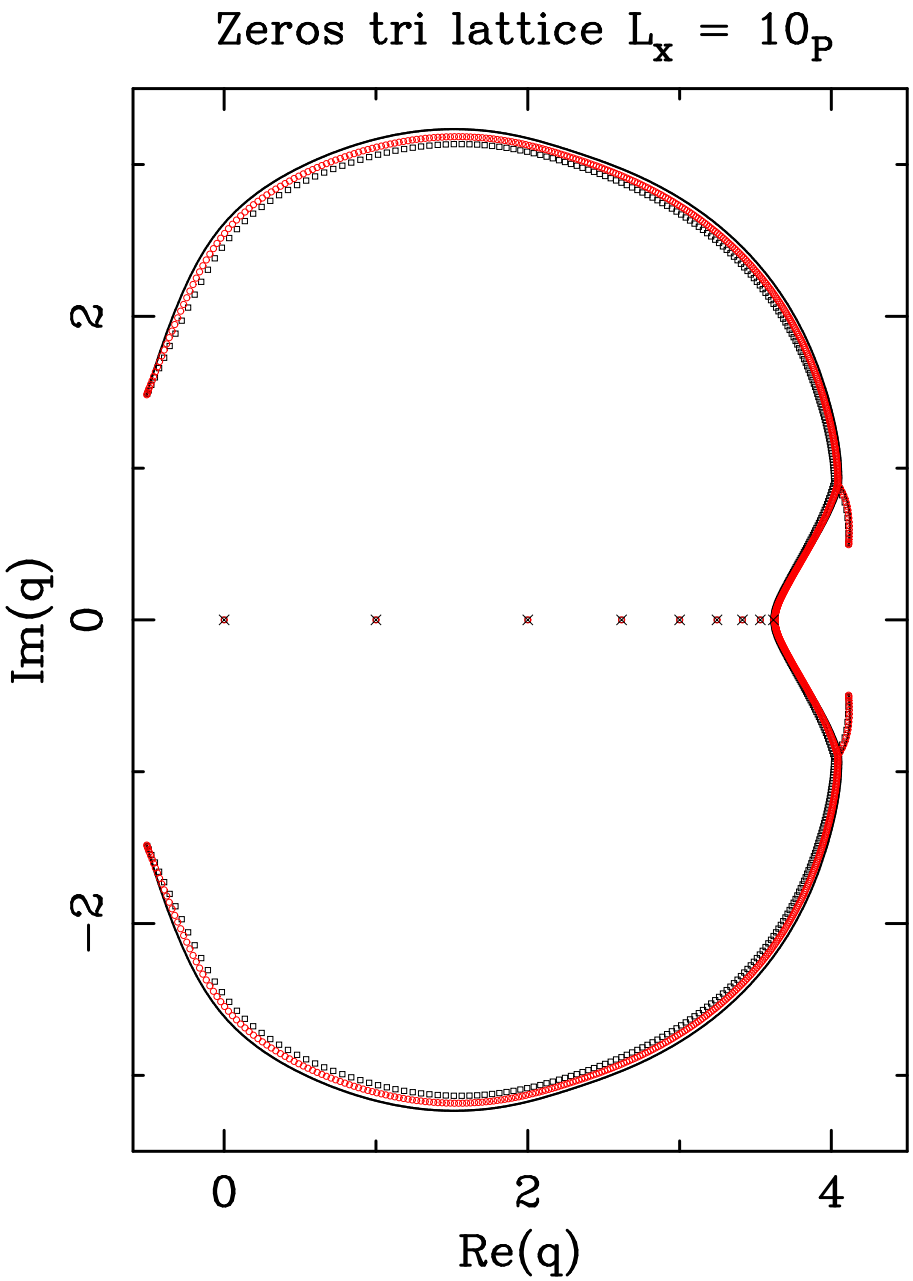


Figure 15: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $10_P \times 50_F$ (squares), $10_P \times 100_F$ (circles) and $10_P \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

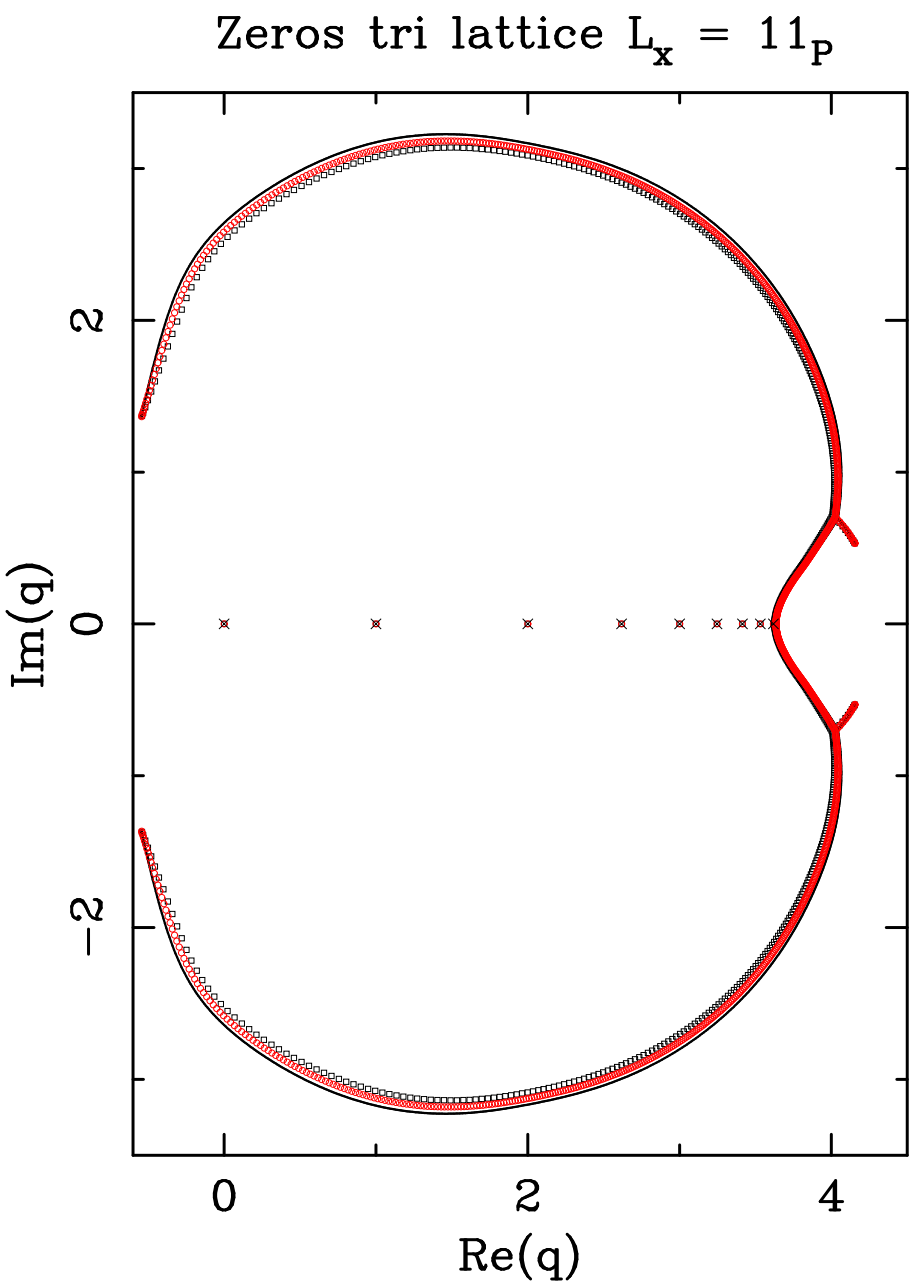


Figure 16: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $11_P \times 55_F$ (squares), $11_P \times 110_F$ (circles) and $11_P \times \infty_F$ (solid line). The isolated limiting points are denoted by a $\times$.

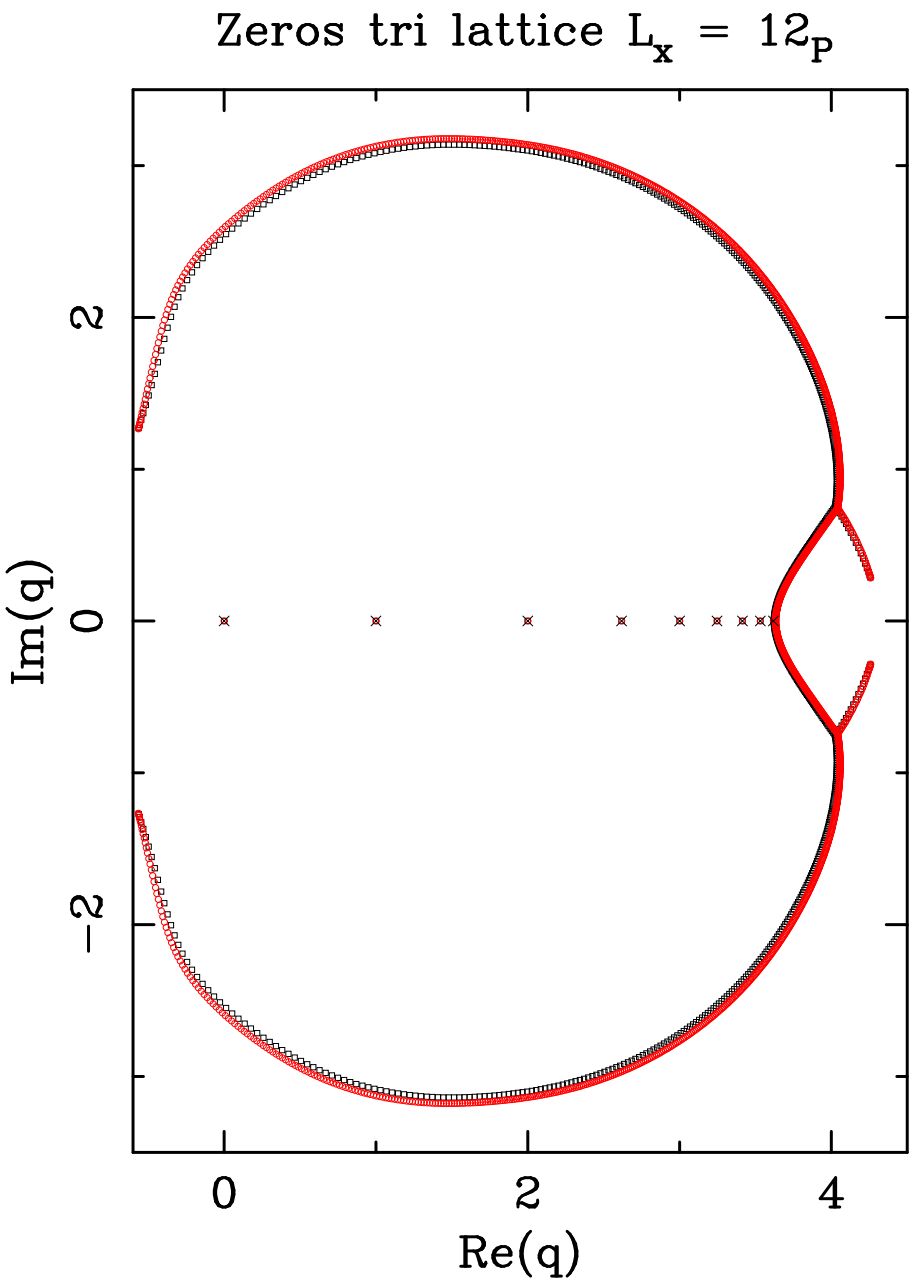


Figure 17: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $12_P \times 60_F$ (squares) and $12_P \times 120_F$ (circles). The isolated limiting points are denoted by a $\times$.

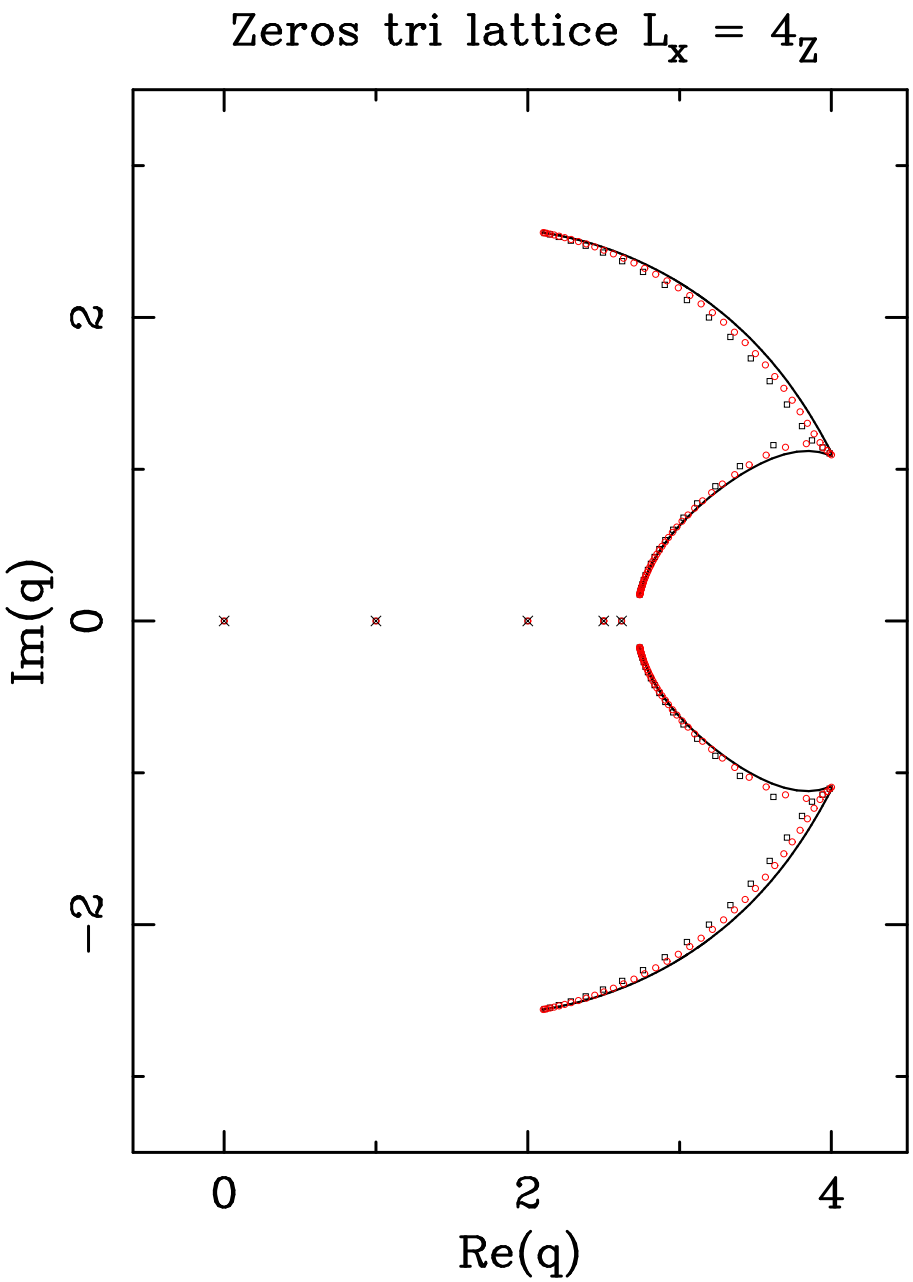


Figure 18: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $4_P \times 20_Z$ (squares), $4_P \times 40_Z$ (circles) and $4_P \times \infty_Z$ (solid line). The isolated limiting points are denoted by a $\times$.

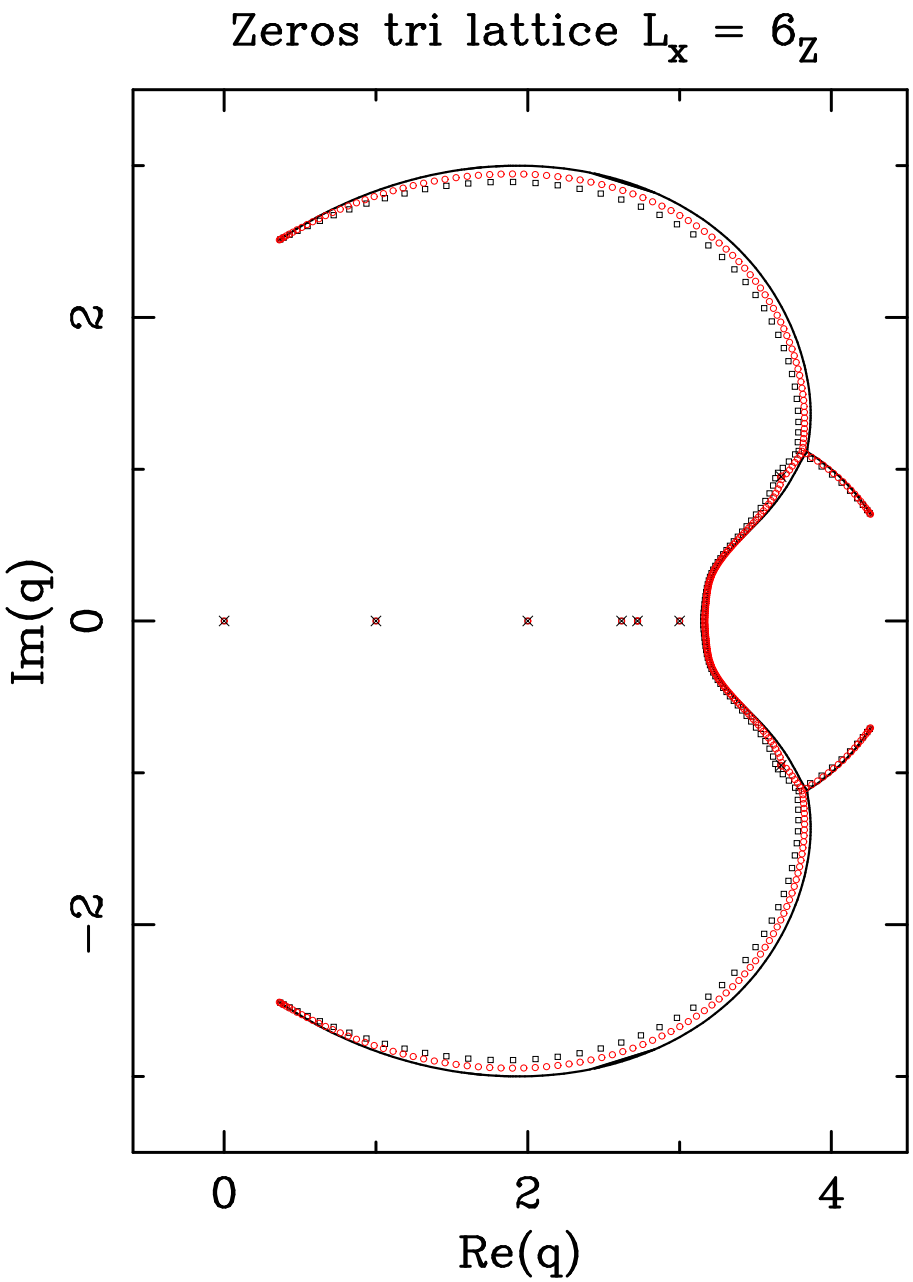


Figure 19: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $6_P \times 30_Z$ (squares), $6_P \times 60_Z$ (circles) and $6_P \times \infty_Z$ (solid line). The isolated limiting points are denoted by a $\times$.

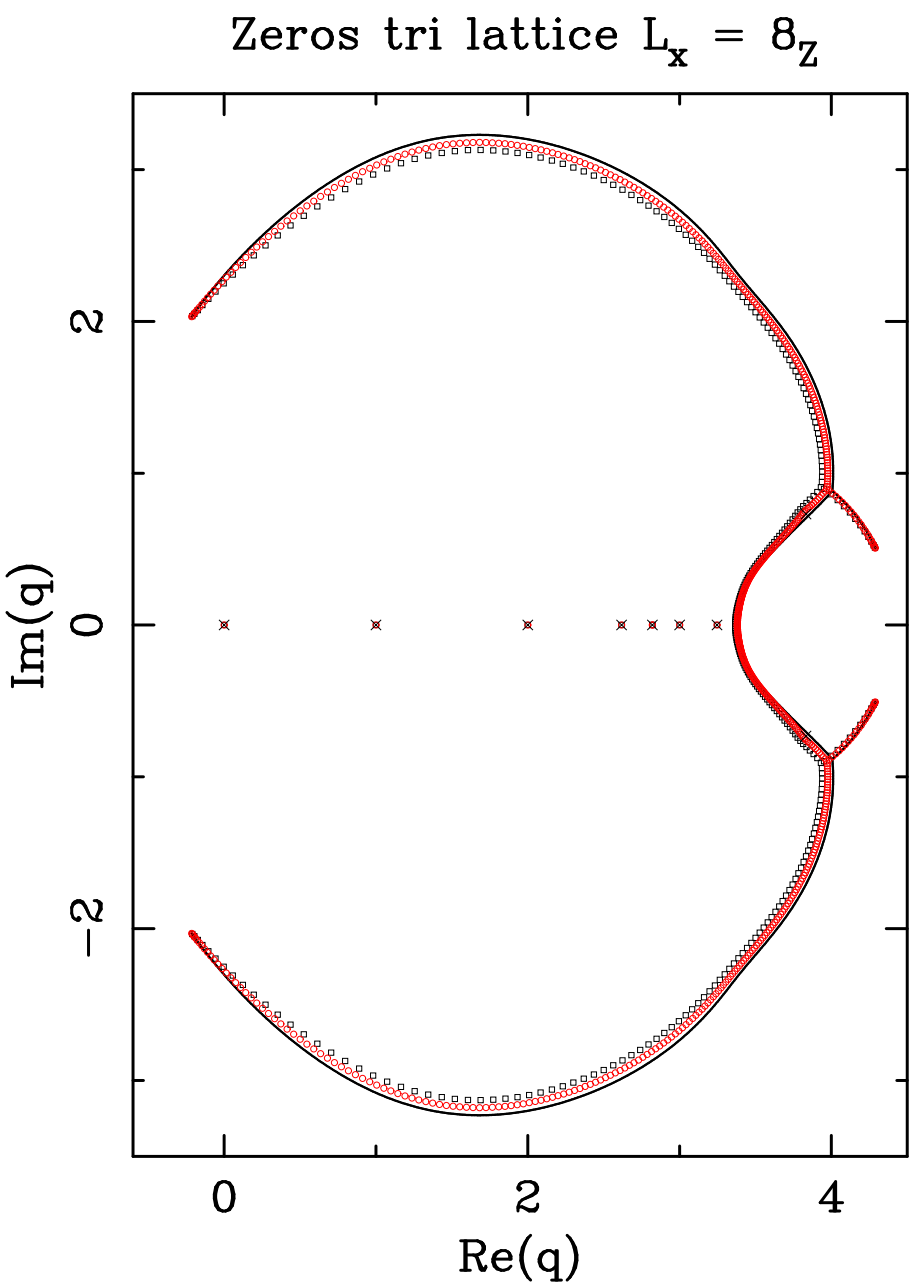


Figure 20: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $8_P \times 40_Z$ (squares), $8_P \times 80_Z$ (circles) and $8_P \times \infty_Z$ (solid line). The isolated limiting points are denoted by a $\times$.

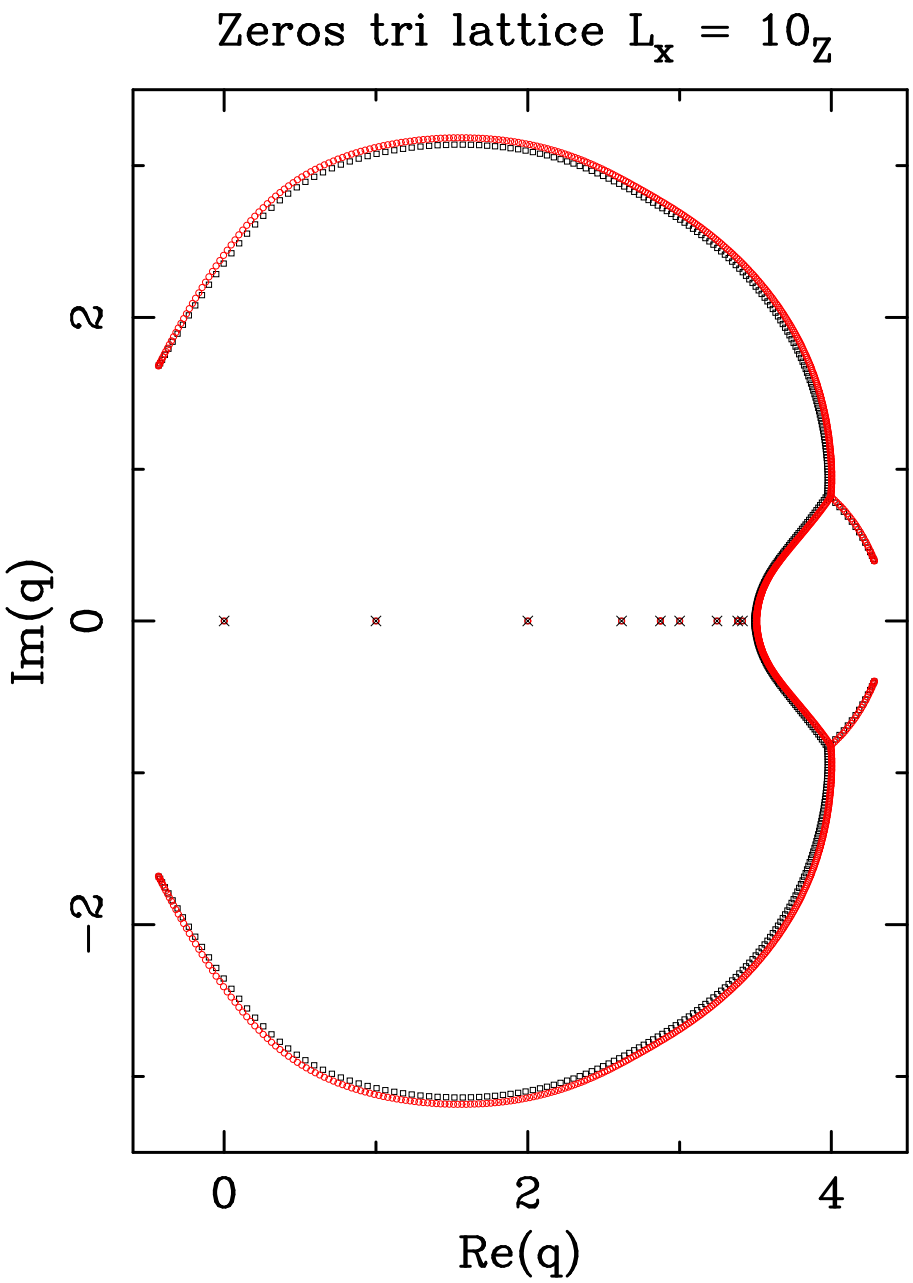


Figure 21: Zeros of the partition function of the q-state Potts antiferromagnet on the triangular lattices $10_P \times 50_Z$ (squares) and $10_P \times 80_Z$ (circles). The isolated limiting points are denoted by a $\times$.

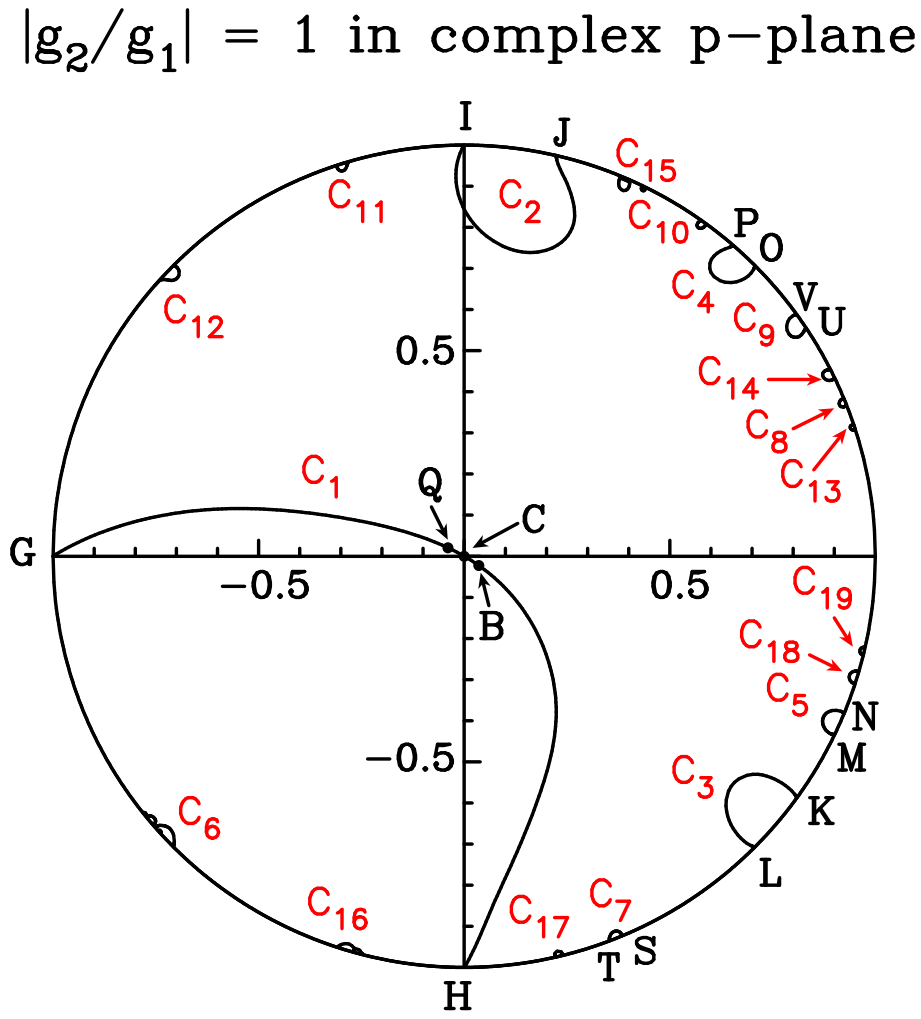


Figure 22: Equimodular curves $|g_2/g_1| = 1$ in the complex p-plane. Several important points are labelled G, H, I, J, ... (see text).

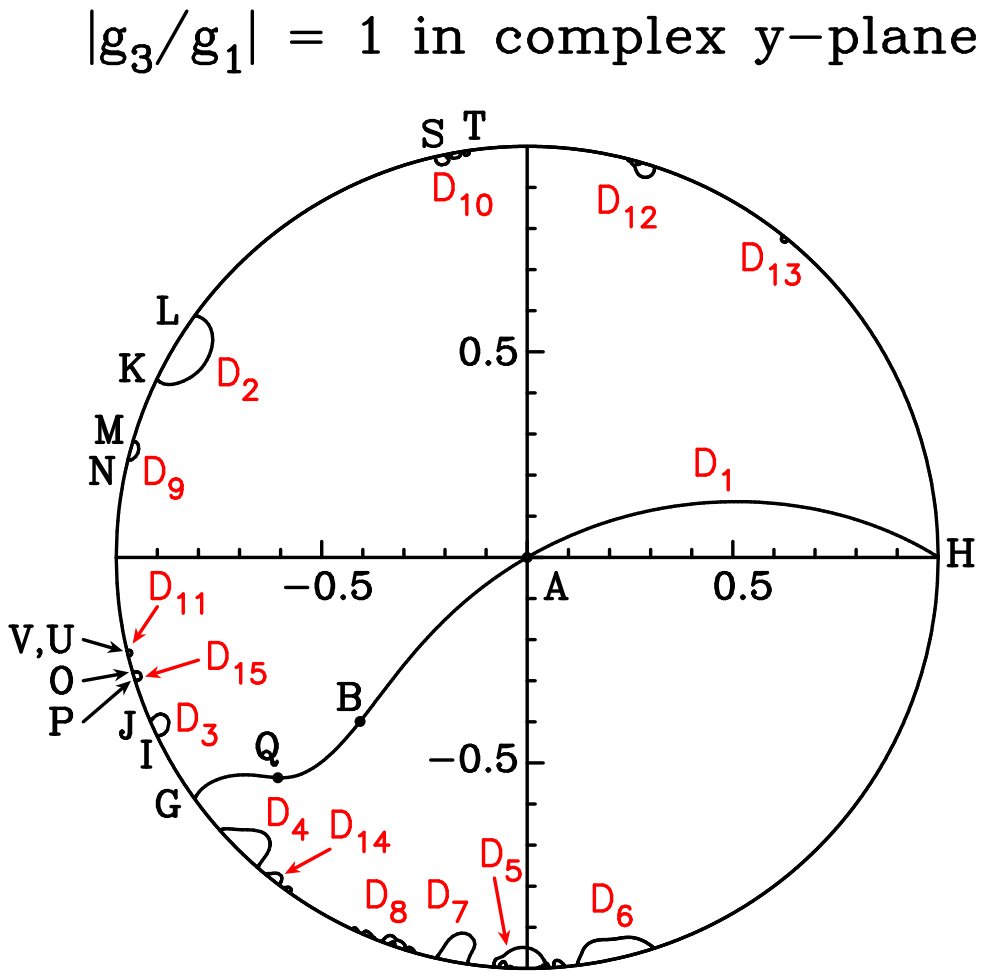


Figure 23: Equimodular curves $|g_3/g_1| = 1$ in the complex y-plane. Several important points are labelled G, H, I, J, ... (see text).

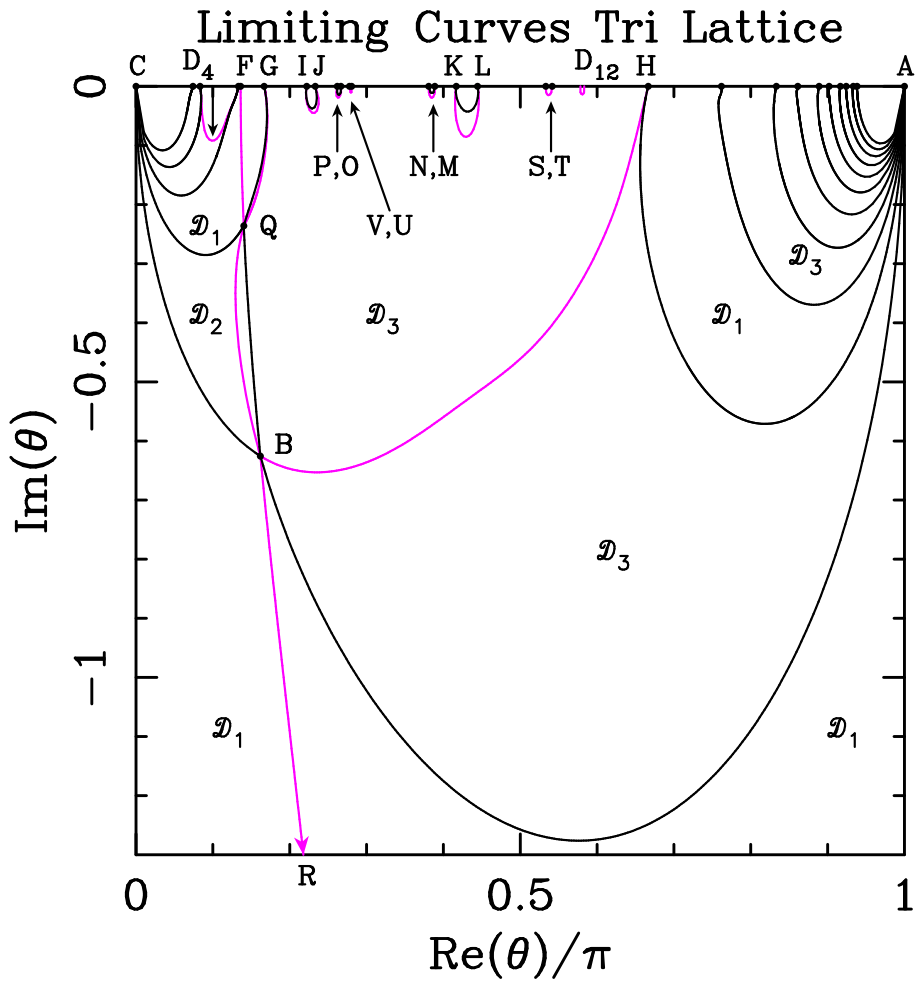


Figure 24: Equimodular curves for the eigenvalues g_i in the complex θ-plane. The portions of curves where the equimodular eigenvalues are dominant (resp. subdominant) are depicted in black (resp. pink). The eigenvalue g_i is dominant in each region labelled $\mathcal{D}_i$. Several important points are labelled A, B, C, ... (see text). To facilitate comparison with Baxter's results [19, Fig. 5], we have used the same labelling of points wherever possible.

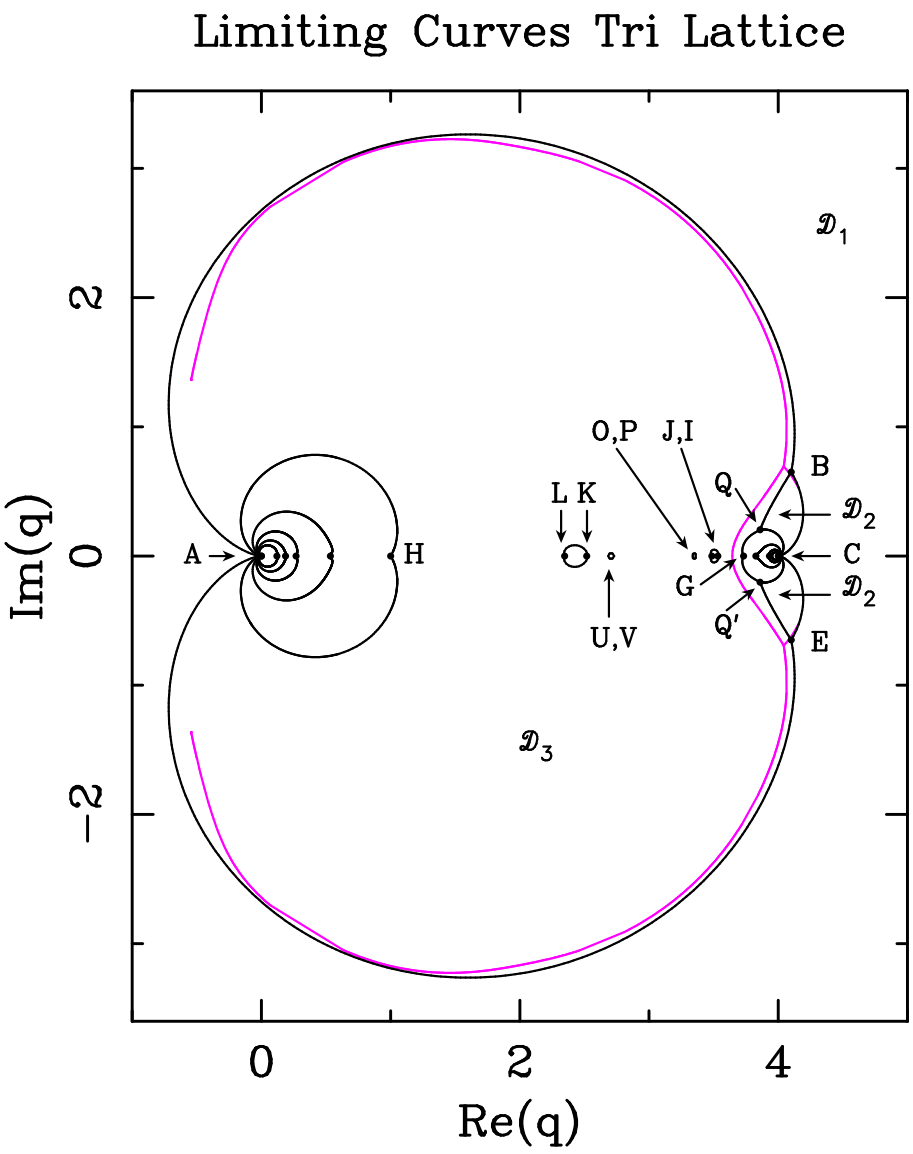


Figure 25: Dominant equimodular curves for the eigenvalues g_i in the complex q-plane (in black). For comparison, we show (in pink) the limiting curve $\mathcal{B}$ for the strip $L_x = 11_{\rm P}$. To facilitate comparison with Baxter's results [19, Fig. 5], we have used the same labelling of points wherever possible.

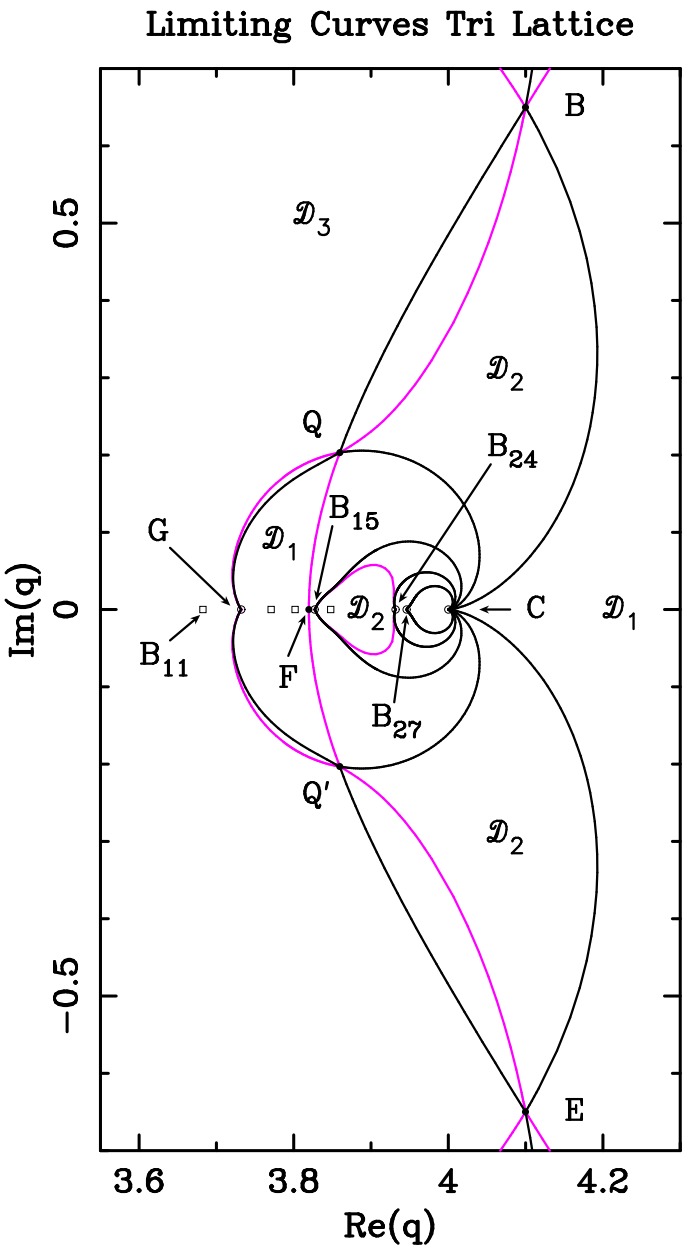


Figure 26: Detail of Figure 25 around the point $q = 4$. We depict the dominant (resp. subdominant) equimodular curves in black (resp. pink). The solid circles ($\bullet$) denote special points discussed in the text, while the squares ($\square$) and empty circles ($\circ$) denote the Beraha numbers $q = B_{11}, \ldots, B_{16}$, B_{24}, B_{27} and $B_\infty = 4$ (point C). The empty circles denote those Beraha numbers which belong to any equimodular curve. We denote by $\mathcal{D}_i$ the regions where the eigenvalue g_i is dominant.

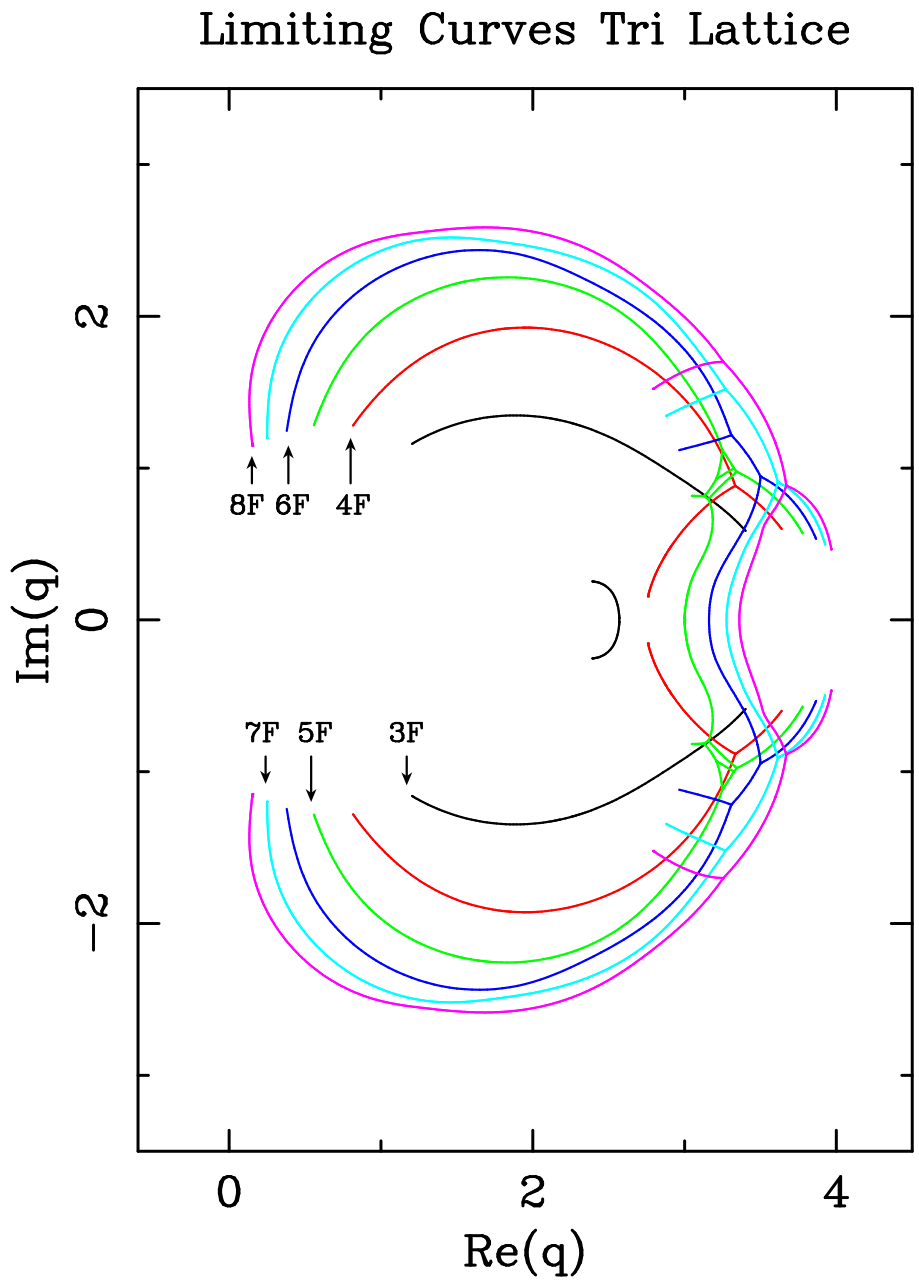


Figure 27: Limiting curves for the triangular-lattice strips $L_{\rm F} \times \infty_{\rm F}$ with $3 \le L \le 8$.

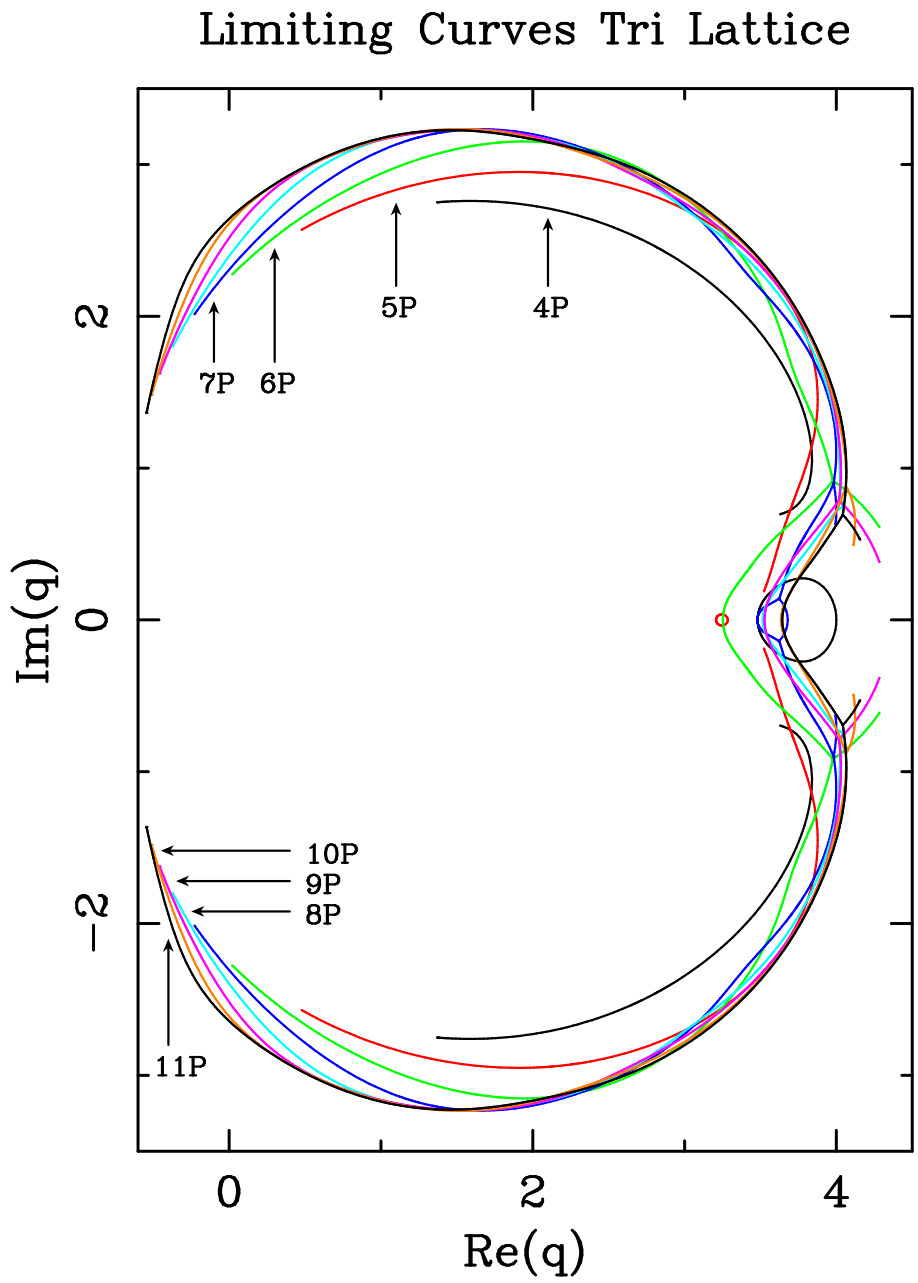


Figure 28: Limiting curves for the triangular-lattice strips $L_{\rm P} \times \infty_{\rm F}$ with $4 \le L \le 11$.

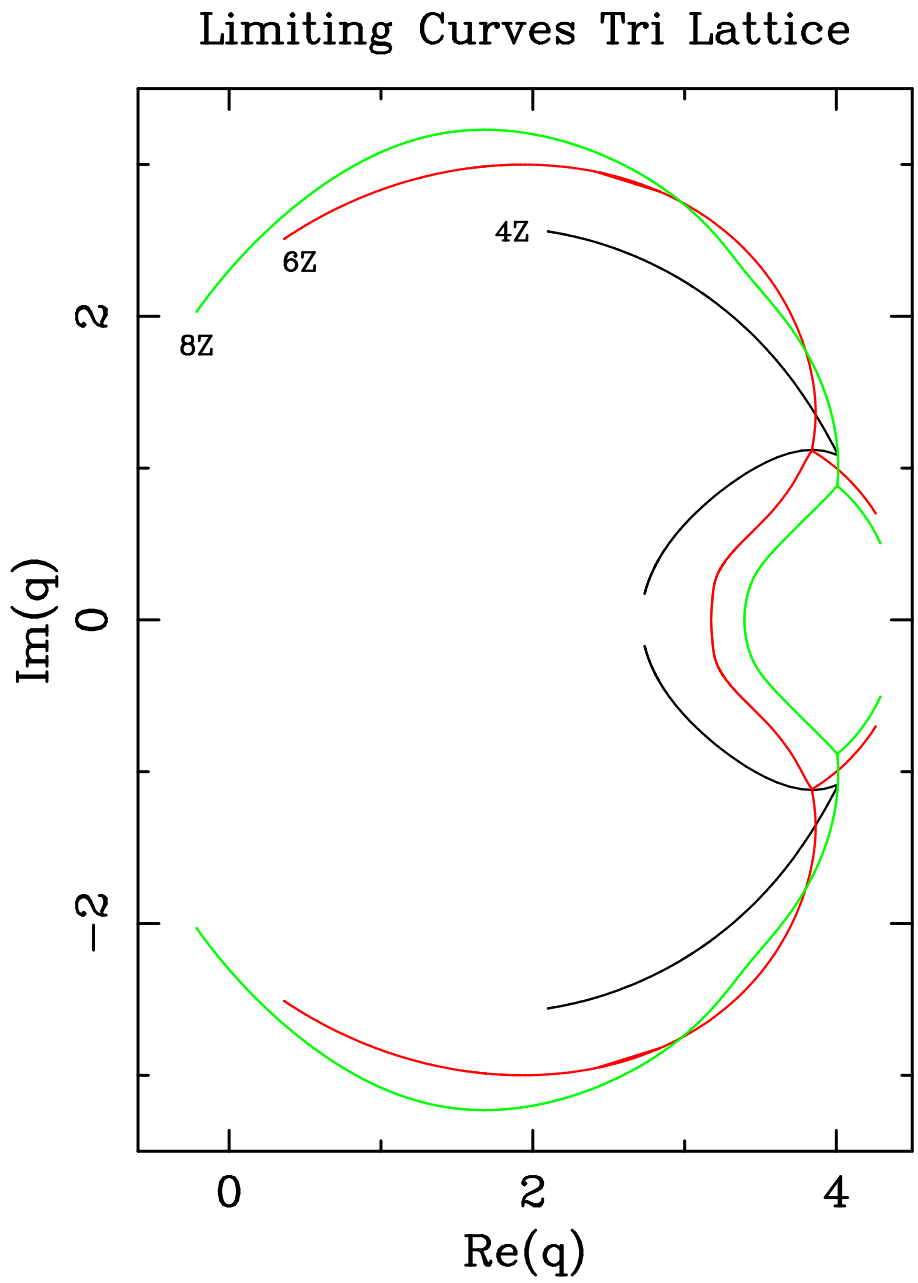


Figure 29: Limiting curves for the triangular-lattice strips $L_{\rm Z} \times \infty_{\rm F}$ with $L = 4, 6, 8$.

Chapitre 4

Diagramme de phase

Le diagramme de phase du modèle de Potts possède une structure très riche, en particulier dans le domaine antiferromagnétique. En vue de la formulation du modèle en tant que modèle d'amas (2.4) il est naturel de paramétrer le diagramme de phase par (q, u) pour des valeurs de $u = \mathrm{e}^K - 1$ quelconques, bien que seul $u \geq -1$ corresponde à des températures physiques dans la formulation de spins (1.1).

Dans le Chapitre 3, nous avons vu que la dépendance du réseau est cruciale, au moins à température nulle ($u = -1$). Afin de mettre en évidence d'éventuels effets non universels, nous allons désormais considérer deux réseaux différents : les réseaux carré et triangulaire. Le réseau hexagonal présente également un intérêt pratique, mais il est lié au réseau triangulaire par dualité (voir le Chapitre 2).

En particulier, nous allons nous intéresser aux lignes critiques dans le plan (q, u). Il est bien connu que ces lignes coïncident souvent avec les valeurs des paramètres pour lesquelles le modèle est intégrable : en effet, l'invariance conforme et l'intégrabilité sont tous les deux liés à l'existence d'un nombre infini de lois de conservation.

Pour le réseau carré, Baxter [13, 14] a déterminé l'énergie libre (ainsi que d'autres quantités d'intérêt physique) le long des courbes

$$u = \pm\sqrt{q}, \tag{4.1}$$
$$u = -2 \pm \sqrt{4-q}. \tag{4.2}$$

Pour le réseau triangulaire, les courbes intégrables sont [12, 15, 16]

$$u^3 + 3u^2 = q, \tag{4.3}$$
$$u = -1. \tag{4.4}$$

Dans le deux cas, il est bien établi que la branche avec $u \geq 0$ décrit la transition entre les phases ferromagnétique ($u \gg 1$) et paramagnétique ($u \ll 1$). Ces transitions sont de premier ordre (avec une chaleur latente non nulle) si $q > 4$ et de second ordre si $0 \leq q \leq 4$ [10]. Les transitions de second ordre sont *universelles* dans la mesure que les exposants critiques ne dépendent que de q et non pas du réseau : en posant

$$q = 4\cos^2\left(\frac{\pi}{p+1}\right) \tag{4.5}$$

on retrouve exactement les modèles minimaux unitaires (1.4)–(1.5). Cette universalité est par ailleurs confirmée par des calculs de type gaz de Coulomb [136].

Dans ce chapitre, nous sommes plutôt concernés par le comportement antiferromagnétique. Une étude intéressante du réseau carré fut initiée par Saleur [150], mais de nombreuses questions demeurent ouvertes, en particulier en ce qui concerne le réseau triangulaire. Nous allons également nous appuyer sur des résultats numériques [34, 98].

Bien sûr, on s'attend à ce que les résultats dépendent de manière cruciale du choix des conditions au bord. Sauf indication contraire, nous allons adopter des conditions au bord cylindriques car ce choix facilite le contact avec la théorie conforme et avec le formalisme de matrice de transfert. Des conditions au bord ouvertes donneraient des résultats similaires, car les deux possibilités (ouverte/cylindrique) sont plongeables dans le plan. Par contre, la situation serait très différente pour des conditions au bord périodiques dans la direction de transfert (géométrie toroïdale/ruban de Möbius/bouteille de Klein) :

par exemple, le Théorème de Quatre Couleurs [3, 4] pour le coloriage de graphes planaires devient un théorème de sept couleurs sur le tore [146].

4.1 Réseau carré

L'approche d'intégrabilité pour le modèle de Potts est basée sur une réécriture de la fonction de partition faisant intervenir l'algèbre de Temperley-Lieb [126]. Un grand pas en avant se fit autour de 1990 quand il fut réalisé que le modèle possède également la symétrie du groupe quantique $\mathrm{U}_q\mathrm{sl}(2)$ pour toute valeur de u [149, 150]. Plus précisement, la formulation en modèle de boucles (2.5) permet une redistribution locale des poids q : ceci définit un modèle de vertex. C'est la matrice de transfert de ce modèle de vertex qui commute avec les générateurs de $\mathrm{U}_q\mathrm{sl}(2)$ [149].

4.1.1 Phase de Berker-Kadanoff

En évoquant une théorie de champ moyen, Berker et Kadanoff [23] prédirent que pour une dimension d suffisamment élevée il doit y avoir une région antiferromagnétique avec des exposants qui ne dépendent pas de la température. Le point de terminaison d'une telle région serait un point critique antiferromagnétique. Or, pendant longtemps une telle phase ne fut pas observée en dimension finie.

Avec l'approche d'intégrabilité basée sur $\mathrm{U}_q\mathrm{sl}(2)$ il est possible d'identifier la théorie conforme décrivant la ligne critique $u = -\sqrt{q}$ pour $0 \le q \le 4$ [150]. Avec la paramétrisation $q = B_n$ (3.1), la charge centrale et les exposants critiques sont

$$c = 1 - \frac{6(n-1)^2}{n}, \tag{4.6}$$

$$h_{r,s} = \frac{(nr-s)^2 - (n-1)^2}{4n}. \tag{4.7}$$

En particulier, l'exposant thermique s'identifie comme $x_T = 2h_{21} = 3n/2 - 1$. On notera que $x_T > 2$ partout : la température est donc une perturbation non pertinente le long de la ligne $u = -\sqrt{q}$ et il s'agit bien d'une phase de Berker-Kadanoff. Des calculs numériques de matrice de transfert confirment ces valeurs de c et de x_T [98].

A partir d'une interprétation basée sur les flots sous le groupe de renormalisation, il est naturel de s'attendre à ce que le domaine d'attraction de la ligne $u = -\sqrt{q}$ soit compris entre les courbes (4.2). Nous allons revenir à ce point dans un instant.

4.1.2 Rôle des nombres de Beraha

Saleur [149] argumenta que le long de la courbe $u = -\sqrt{q}$, l'amplitude de la valeur propre dominante s'annule quand $q = B_n$ (3.1) avec n positif entier. En particulier, les expressions (4.6)–(4.7) ne sont plus applicables. Cet argument est encore basé sur la symétrie $\mathrm{U}_q\mathrm{sl}(2)$, mais avec quelques hypothèses supplémentaires sur la structure des valeurs propres. Par le théorème de Beraha-Kahane-Weiss [19, 20, 154], on déduit l'existence d'un point d'accumulation des zéros de Z en $q = B_n$.

Pourtant, ce résultat n'implique pas nécessairement l'annulation de Z à $T = 0$ dans la limite thermodynamique. Pour clarifier le lien avec la limite chromatique, Saleur évoqua encore une supposition : que la structure des valeurs propres ne change pas dans le domaine d'attraction des points fixes $u = -\sqrt{q}$. Comme nous l'avons vu, on s'attend à ce que ce domaine soit limité par $u = -2 \pm \sqrt{4-q}$. Par conséquence, dans la plage de températures comprise entre ces deux courbes, l'amplitude de la valeur propre dominante devrait s'annuler à tout nombre de Beraha, $q = B_n$ avec $n = 1, 2, 3, \ldots$.

Un support numérique pour cette prédiction peut être trouvé dans [33], où les auteurs calculent les courbes limites $\mathcal{B}$ (discutées lors du Chapitre 3) dans le plan de u complexe. Pour plusieures valeurs de q, on voit que ces courbes ont tendence à couvrir un segment sur l'axe de u réel, centré environ à $u = -\sqrt{q}$. Cependant, les largeurs des rubans traités dans [33] sont assez faibles, et il est difficile de se former une idée claire du comportement dans la limite thermodynamique.

Il est intéressant de constater que la ligne de température nulle n'intersecte la phase de Berker-Kadanoff que pour $0 \le q < 3$. Si le scénario proposé par Saleur est juste, cela veut dire que seuls les

nombres de Beraha $B_n < 3$ (c'est-à-dire, $n = 2, 3, 4, 5$) sont des point d'accumulation des zéros chromatiques. C'est effectivement ce qu'on observe numériquement [147, 90]. Pourtant, une étude minutieuse de la matrice de transfert [147, 90] montre aussi que *tout* nombre de Beraha est associé à une amplitude nulle dans la limite thermodynamique : or, cette amplitude est seulement dominante pour $q < 3$. Ce rôle "caché" des B_n est encore assez mal compris.

Saleur [149, 150] a suggéré que le mécanisme faisant intervenir les nombres de Beraha pourrait être universel. En effet, la limite anisotropique des réseaux triangulaire et hexagonal possède le même hamiltonien que la limite anisotropique du réseau carré. Par contre, les points fixes qui contrôlent la phase Berker-Kadanoff sont moins bien compris pour ces réseaux.

4.1.3 Lignes parafermioniques

Quand on traverse les bords (4.2) de la phase de Berker-Kadanoff, beaucoup de niveaux se croisent dans la matrice de transfert [150]. Exactement sur les lignes (4.2), Saleur a trouvé encore un opérateur qui commute avec la matrice de transfert du modèle de vertex : il s'agit d'un opérateur de conjugaison de charge C, ayant la propriété $C^2 = I$. Comme candidat pour une théorie conforme avec la symétrie $\mathrm{U}_q\mathrm{sl}(2)$ et un opérateur de conjugaison, Saleur proposa les théories parafermioniques Z_k construites par Zamolodchikov et Fateev [168, 71].

La charge centrale et les dimensions des opérateurs physiques des parafermions sont [168]

$$c = 2 - \frac{6}{k+2}, \tag{4.8}$$

$$D_j = \frac{j(j+1)}{k+2}, \quad j = 0, 1, \ldots, k/2, \tag{4.9}$$

$$d_r = \frac{r(k-r)}{2k(k+2)}, \quad r = 1, 2, \ldots, k-1. \tag{4.10}$$

Les dimensions D_j des opérateurs énergétiques et les dimensions d_r des opérateurs de spin sont celles des champs primaires par rapport à l'algèbre de Virasoro. La construction de la théorie de représentation des modules dégénérés montre que seulement un sous-ensemble de ces champs sont primaires par rapport à l'algèbre chirale parafermionique.

Avec la paramétrisation habituelle, $q = B_n$, la connaissance préalable des dimensions de certains opérateurs a conduit Saleur [150] à proposer la relation $k = n - 2$. Ici, $n \in [2, \infty[$ prend des valeurs réelles quelconques. Cette identification est bien vérifiée numériquement au niveau de la charge centrale [98]. Par contre, les prédictions [150]

$$x_H = 2d_{k/2}, \tag{4.11}$$

$$x_T = 2D_1 \tag{4.12}$$

pour les dimensions physiques des opérateurs d'aimantation et d'énergie sont contredites par un certain nombre d'observations. Premièrement, la valeur $x_T = 2/3$ à $q = 3$ ne coïncide pas avec le résultat exact $x_T = 3/2$ du modèle de six vertex, obtenu en supposant que les excitations dominantes à faible température sont de type tourbillon-antitourbillon [32]. La valeur $x_T = 3/2$ est aussi trouvée par des calculs numériques de type matrice de transfert [98]. Deuxièmement, la proposition pour x_H est (en partie) basée sur le résultat exact $x_H = 1/16$ du modèle d'Ising *ferromagnétique*. Certes, les modèles d'Ising ferromagnétique et antiferromagnétique sont équivalents dans le secteur de l'identité par une réflexion des spins sur un des deux sous-réseaux. Par contre, cette équivalence n'est pas assurée dans le secteur impair. La construction topologique de celui-ci au niveau de la matrice de transfert montre que, en effet, la dimension x_H est *négative* [98] pour tout $q \in [0, 4]$.

En conclusion, le rapport entre la ligne antiferromagnétique et la théorie parafermionique n'est pas encore compris au niveau du contenu en opérateurs.

La présence de la symétrie discrète Z_k paraît assez mystérieuse, comme l'avait déjà remarqué Saleur [150]. Une explication possible pour le cas de $q = 3$ ($k = 4$) a été avancée dans [83]. Pourtant, il se peut que la symétrie Z_k soit réalisée de manière non linéaire (dans un sens à préciser), ce qui rendrait son interprétation physique difficile.

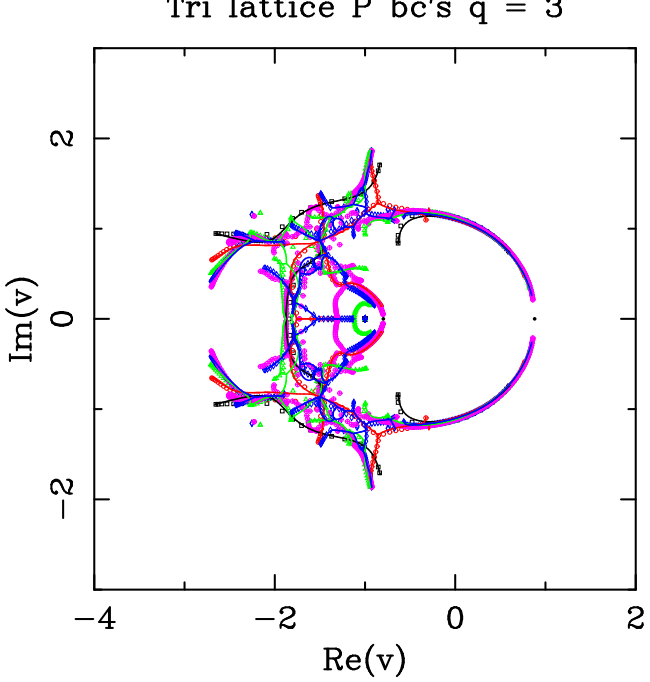


FIG. 4.1 – Points d'accumulation des zéros du polynôme dichromatique avec $q = 3$ sur des rubans du réseau triangulaire. Les conditions au bord sont cylindriques. Chaque couleur correspond à une largeur du ruban : $L = 2$ (noir), $L = 3$ (rouge), $L = 4$ (vert), $L = 5$ (bleu), $L = 6$ (rose).

4.1.4 Température infinie

Il est intéressant de constater qu'on peut s'approcher à la limite de température infinie ($u \to 0$) de plusieurs manières inéquivalentes. Le long de $u = \pm\sqrt{q}$, on a une théorie critique avec $c \to -2$. Le long de $u = 0$, la théorie est bien sûr non critique avec, formellement, $c = 0$. Finalement, en s'approchant à $q = 0$ le long de $u = -2 + \sqrt{4 - q}$, la conjecture parafermionique donne $c \to -1$.

Cette dernière limite correspond à des forêts sur le réseau carré, où chaque arête porte un poids $w = -1/4$ [150]. Par le flot de renormalisation, on peut comprendre ce point comme une transition de phase entre les régimes $w < -1/4$ avec $c = -2$ et $w > -1/4$ qui est non critique. Cette interprétation est confirmée par le calcul des courbes limites $\mathcal{B}$ [98].

4.2 Réseau triangulaire

Dans la section précédente, nous avons déjà fait allusion à d'autres réseaux que le réseau carré. Bien que certains aspects de l'approche d'intégrabilité [149, 150] aient des traits universels, le rôle de la phase de Berker-Kadanoff n'est pas tout aussi bien compris pour le cas du réseau triangulaire dont nous allons discuter dans ce paragraphe.

On peut facilement se convaincre qu'un diagramme de phase composé des trois branches de la courbe (4.3) ne peut être complet, au moins si l'on suppose que ces branches sont des points fixes pour le groupe de renormalisation. Car, hormis ces branches il existe trois branches triviales de points fixes attractifs non critiques à $u = -\infty$, à $u = 0$ et à $u = +\infty$. Pour que les flots sous renormalisation soient consistants il faut donc un nombre *pair* de branches non triviales.

Une possibilité serait de supposer que la limite chromatique (4.4) soit encore une ligne de points fixes. Pourtant, cette solution nécessiterait une inversion de la nature (attractive/répulsive) des deux branches qui se rencontrent en $(u, q) = (-1, 2)$. Ceci ne serait pas très crédible. Il paraît plus probable qu'une autre ligne, dont la paramétrisation exacte n'est pas connue pour l'instant, émerge du point $(u, q) = (-1, 4)$ qui est un point critique avec $c = 2$ [78, 131].

Pour éclaircir ces points, nous avons entrepris une étude numérique du diagramme de phase [34, 98]. Les résultats principaux sont les suivants :

- Les trois branches de (4.3) sont respectivement dans la classe d'universalité de la transition ferromagnétique (branche supérieure, $u > 0$), de la phase de Berker-Kadanoff (branche intermédiaire) et de la théorie parafermionique (branche inférieure).
- Le domaine d'attraction de la phase de Berker-Kadanoff est borné d'en bas par la branche inférieure de (4.3) et d'en haut par une ligne de non analyticité de l'énergie libre. Cette ligne est le prolongement, pour $u \neq -1$, du point $(u, q) = (-1, q_0)$ déterminé par (3.5) qu'elle relie à $(u, q) = (0, 0)$. Les transitions correspondantes sont du premier ordre.
- Le point critique à $(u, q) = (-1, 4)$ fait en effet partie d'une ligne de transitions de second ordre. Cette ligne se termine à son intersection avec la ligne de non analyticité (voir ci-dessus), dans un point tricritique situé à environ $q_{\rm tc} \approx 3.5$. Les charges centrales des théories correspondantes sont numériquement proches de 2.
- Les courbes limites $\mathcal{B}$ dans le plan de u complexe sont d'une complexité considérable, notamment dans le domaine antiferromagnétique : voir la Figure 4.1. Elles possèdent des branches qui pincent l'axe réel au niveau de la branche inférieure de (4.3) et de la ligne de non analyticité.

De plus amples détails sont donnés dans l'article [34] qui est attaché à ce chapitre.

4.3 Article "Exact Potts model partition functions"

Nous présentons des calculs exacts de la fonction de partition $Z(G,q,v)$ du modèle de Potts à q états et avec une variable de température v. Le modèle est défini sur des rubans G à n sommets du réseau triangulaire ayant une largeur transversale L variable et une longueur m arbitrairement élevée. Nous imposons des conditions au bord longitudinales libres et des conditions au bord transversales soit libres, soit périodiques. La fonction de partition a alors la forme $Z(G,q,v) = \sum_{j=1}^{N_{Z,G,\lambda}} c_{Z,G,j}(\lambda_{Z,G,j})^{m-1}$. Nous donnons des expressions de $N_{Z,G,j}$ et de sa spécialisation au cas $v = -1$ pour L arbitraire. L'énergie libre est calculée exactement dans la limite de rubans infiniment longs et nous discutons la limite thermodynamique. Nous démontrons comment l'énergie interne d'un ruban avec des conditions aux bords cylindriques est reliée aux propriétés critiques du modèle de Potts sur le réseau triangulaire infini. Dans le cas des valeurs complexes arbitraires de q et de v nous déterminons la courbe singulière $\mathcal{B}$: cette courbe est formée des points d'accumulation des zéros de la fonction de partition dans la limite $m \to \infty$, soit dans le plan de q pour v fixé, soit dans le plan de v pour q fixé.

Exact Potts Model Partition Functions for Strips of the Triangular Lattice

Shu-Chiuan Chang
C. N. Yang Institute for Theoretical Physics
State University of New York
Stony Brook, N. Y. 11794-3840
USA
after 9/10/2002: Department of Applied Physics:
Faculty of Science
Tokyo University of Science
1-3 Kagurazaka, Shinjuku-ku
Tokyo 162-8601
JAPAN
chang@rs.kagu.tus.ac.jp

Jesper Lykke Jacobsen
Laboratoire de Physique Théorique et Modèles Statistiques
Université Paris-Sud
Bâtiment 100
F-91405 Orsay
FRANCE
jacobsen@ipno.in2p3.fr

Jesús Salas
Departamento de Física Teórica
Facultad de Ciencias, Universidad de Zaragoza
Zaragoza 50009
SPAIN
jesus@melkweg.unizar.es

Robert Shrock
C. N. Yang Institute for Theoretical Physics
State University of New York
Stony Brook, N. Y. 11794-3840
USA
robert.shrock@sunysb.edu

November 26, 2002

Abstract

We present exact calculations of the Potts model partition function $Z(G,q,v)$ for arbitrary q and temperature-like variable v on n-vertex strip graphs G of the triangular lattice for a variety of transverse widths equal to L vertices and for arbitrarily great length equal to m vertices, with free longitudinal boundary conditions and free and periodic transverse boundary conditions. These have

the form $Z(G,q,v) = \sum_{j=1}^{N_{Z,G,\lambda}} c_{Z,G,j} (\lambda_{Z,G,j})^{m-1}$. We give general formulas for $N_{Z,G,j}$ and its specialization to $v = -1$ for arbitrary L. The free energy is calculated exactly for the infinite-length limit of the graphs, and the thermodynamics is discussed. It is shown how the internal energy calculated for the case of cylindrical boundary conditions is connected with critical quantities for the Potts model on the infinite triangular lattice. Considering the full generalization to arbitrary complex q and v, we determine the singular locus $\mathcal{B}$, arising as the accumulation set of partition function zeros as $m \to \infty$, in the q plane for fixed v and in the v plane for fixed q.

Key Words: Potts model, triangular lattice, exact solutions, transfer matrix, Fortuin-Kasteleyn representation, Tutte polynomial.

1 Introduction

The q-state Potts model has served as a valuable model for the study of phase transitions and critical phenomena [1, 2]. In this paper we present some theorems on structural properties of Potts model partition functions on triangular-lattice strips of arbitrary width equal to L vertices and arbitrarily great length equal to m vertices. We also report exact calculations of Potts model partition functions for a number of triangular-lattice strips of various widths and arbitrarily great lengths. Using these results, we consider the limit of infinite length. For this limit we calculate thermodynamic functions and determine the loci in the complex q and temperature planes where the free energy is non-analytic. These loci arise as the continuous accumulation sets of partition-function zeros. This work is an extension to the triangular lattice of our earlier study for the square lattice [3].

Consider a graph $G = (V, E)$, defined by its vertex set V and edge set E. Denote the number of vertices and edges as $|V| \equiv n$ and $|E|$, respectively. For technical simplicity, we restrict to connected loopless graphs. On this graph G, at temperature T, the Potts model is defined by the partition function

$$Z(G, q, v) = \sum_{\{\sigma_n\}} e^{-\beta \mathcal{H}} \tag{1.1}$$

with the (zero-field) Hamiltonian

$$\mathcal{H} = -J \sum_{\langle ij \rangle} \delta_{\sigma_i \sigma_j} \tag{1.2}$$

where $\sigma_i = 1, \ldots, q$ are the spin variables on each vertex $i \in V$; $\beta = (k_B T)^{-1}$; and $\langle ij \rangle \in E$ denotes pairs of adjacent vertices. We use the notation

$$K = \beta J \ , \quad a = e^K \ , \quad v = a - 1 \tag{1.3}$$

so that the physical ranges are (i) $a \ge 1$, i.e., $v \ge 0$ corresponding to $\infty \ge T \ge 0$ for the Potts ferromagnet, and (ii) $0 \le a \le 1$, i.e., $-1 \le v \le 0$, corresponding to $0 \le T \le \infty$ for the Potts antiferromagnet. One defines the (reduced) free energy per site $f = -\beta F$, where F is the actual free energy, via

$$f(\{G\}, q, v) = \lim_{n \to \infty} \ln[Z(G, q, v)^{1/n}] \tag{1.4}$$

where we use the symbol $\{G\}$ to denote $\lim_{n \to \infty} G$ for a given family of graphs G.

For our results in this paper we shall consider two types of boundary conditions: free and cylindrical. Here, free boundary conditions mean free in both the transverse and longitudinal directions (the latter being the one that is varied for a fixed width), while cylindrical boundary conditions mean periodic in the transverse direction and free in the longitudinal direction. For free (resp. cylindrical) boundary conditions, we have studied strips of widths $2 \le L \le 6$ (resp. $2 \le L \le 9$). As noted, we shall also give exact results valid for these strips with arbitrary width and length.

There are several motivations for this work. Clearly, new exact calculations of Potts model partition functions are of value in their own right. In addition, these calculations can give insight into the complex-temperature phase diagram of the two-dimensional (2D) Potts model on a particular lattice. This is useful, since the 2D Potts model has never been solved except in the $q = 2$ Ising case. From a mathematical point of view, the partition function of the Potts model on a graph G is equivalent to the Tutte polynomial on the same graph G (see below). Thus, we can extract very useful combinatorial information on the graph G.

Let $G' = (V, E')$ be a spanning subgraph of G, i.e. a subgraph having the same vertex set V and an edge set $E' \subseteq E$. Then $Z(G, q, v)$ can be written as the sum [4, 5, 6]

$$Z(G, q, v) = \sum_{G' \subseteq G} q^{k(G')} v^{|E'|} \tag{1.5}$$

where $k(G')$ denotes the number of connected components of G'. The formula (1.5) enables one to generalize q from $\mathbb{Z}_+$ to $\mathbb{R}_+$ (keeping v in its physical range), and it also shows that $Z(G, q, v)$ is a polynomial in q and v (equivalently, a).

The Potts model partition function on a graph G is essentially equivalent to the Tutte polynomial [7, 8, 9] and Whitney rank polynomial [10, 2, 11, 12, 13, 14]. Here the Tutte polynomial of an arbitrary graph $G = (V, E)$ is

$$T(G, x, y) = \sum_{G' \subseteq G} (x-1)^{k(G')-k(G)} (y-1)^{c(G')} \tag{1.6}$$

where G' again denotes a spanning subgraph of G and $c(G')$ denotes the number of independent circuits in G', satisfying $c(G') = |E'| + k(G') - |V|$. Since we only consider connected graphs G, we have $k(G) = 1$. From (1.5) and (1.6), it follows that the Potts model partition function $Z(G, q, v)$ is related to the Tutte polynomial $T(G, x, y)$ according to

$$Z(G, q, v) = (x-1)^{k(G)} (y-1)^{|V|} T(G, x, y) \tag{1.7}$$

where

$$x = 1 + \frac{q}{v} \tag{1.8a}$$

$$y = a = v + 1 \tag{1.8b}$$

so that

$$q = (x-1)(y-1) \tag{1.9}$$

Previous exact calculations of Potts model partition functions for arbitrary q and v on lattice strips and/or studies of their properties include [3, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26]; a related early study of chromatic and Tutte polynomials for recursive families of graphs is [27].

Various special cases of the Potts model partition function are of interest. One special case is the zero-temperature limit of the Potts antiferromagnet, i.e., $v = -1$. For sufficiently large q, on a given lattice or graph G, this exhibits nonzero ground

state entropy S_0 (without frustration). This is equivalent to a ground state degeneracy per site (vertex), $W > 1$, since $S_0 = k_B \ln W$. The $T = 0$ partition function of the q-state Potts antiferromagnet on G satisfies

$$Z(G, q, -1) = P(G, q) \tag{1.10}$$

where $P(G, q)$ is the chromatic polynomial (in q) expressing the number of ways of coloring the vertices of the graph G with q colors such that no two adjacent vertices have the same color [4, 12, 28, 29]. The minimum number of colors necessary for this coloring is the chromatic number of G, denoted $\chi(G)$. We have

$$W(\{G\}, q) = \lim_{n \to \infty} P(G, q)^{1/n} \tag{1.11}$$

In the context of our current work we recall that the chromatic number for the 2D triangular lattice is $\chi(\mathrm{tri}) = 3$. This chromatic number also applies to strips of the triangular lattice with free longitudinal boundary conditions and free transverse boundary conditions. For the triangular-lattice strips with cylindrical (i.e., free longitudinal and periodic transverse) boundary conditions, $\chi = 3$ if the width $L = 0$ mod 3 and $\chi = 4$ if $L = 1$ or 2 mod 3. References to papers on the special case $v = -1$ are given, e.g., in [17, 25, 30, 3].

Using the formula (1.5) for $Z(G, q, v)$, one can generalize q from $\mathbb{Z}_+$ not just to $\mathbb{R}_+$ but to $\mathbb{C}$ and v from its physical ferromagnetic and antiferromagnetic ranges $0 \le v \le \infty$ and $-1 \le v \le 0$ to $v \in \mathbb{C}$. A subset of the zeros of Z in the two-complex dimensional space $\mathbb{C}^2$ defined by the pair of variables (q, v) can form an accumulation set in the $n \to \infty$ limit, denoted $\mathcal{B}$, which is the continuous locus of points where the free energy is nonanalytic. This locus is determined as the solution to a certain $\{G\}$-dependent equation. For a given value of v, one can consider this locus in the q plane, and we denote it as $\mathcal{B}_q(\{G\}, v)$. In the special case $v = -1$ where the partition function is equal to the chromatic polynomial, the zeros in q are the chromatic zeros, and $\mathcal{B}_q(\{G\}, v = -1)$ is their continuous accumulation set in the $n \to \infty$ limit. With the exact Potts partition function for arbitrary temperature, one can study $\mathcal{B}_q$ for $v \ne -1$ and, for a given value of q, one can study the continuous accumulation set of the zeros of $Z(G, q, v)$ in the v plane (complex-temperature or Fisher zeros [31] - other early references include [32, 33, 34]). This set will be denoted $\mathcal{B}_v(\{G\}, q)$.

2 General Results for Recursive Families of Graphs

A recursive family of graphs is one in which one constructs successive members of the family in a recursive manner starting from an initial member. Recursive families of graphs that are of particular interest here are strips of regular lattices of a given width L vertices (with free or cylindrical boundary conditions) and arbitrarily great length m vertices (with free boundary conditions).

A general form for the Potts model partition function for the strip graphs considered here is [17]

$$Z(G, q, v) = \sum_{j=1}^{N_{Z,G,\lambda}} c_{G,j} (\lambda_{G,j})^{m-1} \tag{2.1}$$

where the coefficients $c_{G,j}$ and corresponding terms $\lambda_{G,j}$, as well as the total number $N_{Z,G,\lambda}$ of these terms, depend on the type of recursive graph G (width and boundary conditions) but not on its length. (In [17], a slightly different labelling convention was used so that $\lambda_{G,j}^m$ rather than $\lambda_{G,j}^{m-1}$ appeared in the summand of eq. (2.1).) In the special case $v = -1$ where Z reduces to the chromatic polynomial (zero-temperature Potts antiferromagnet), eq. (2.1) reduces to the form [35]

$$P(G,q) = \sum_{j=1}^{N_{P,G,\lambda}} c_{G,j} (\lambda_{P,G,j})^{m-1} \tag{2.2}$$

For the lattice strips of interest here, we define the following explicit notation. Let $N_{Z,\mathrm{sq},\mathrm{BC}_t\ \mathrm{BC}_\ell,L,\lambda}$ denote the total number of λ's for the square-lattice strip with the transverse (t) and longitudinal (ℓ) boundary conditions BC_t and BC_ℓ of width L. Henceforth where no confusion will result, we shall suppress the λ subscript. The explicit labels are $N_{Z,\mathrm{sq},\mathrm{FF},L}$ and $N_{Z,\mathrm{tri},\mathrm{FF},L}$ for the strips of the square and triangular lattices with free boundary conditions, and $N_{Z,\mathrm{sq},\mathrm{PF},L}$ and $N_{Z,\mathrm{tri},\mathrm{PF},L}$ for the strips of these respective lattices with cylindrical boundary conditions.

For the lattice strip graphs of interest here we can express the partition function via a transfer matrix T (in the Fortuin-Kasteleyn representation) of fixed size $M \times M$:

$$Z(G,q,v) = \mathrm{tr}\left[A(q,v) \cdot \mathsf{T}(q,v)^{m-1}\right] \tag{2.3}$$

which then yields the form (2.1). Since the transfer matrix T and the boundary-condition matrix A are polynomials in q and v, it follows that the eigenvalues $\{\lambda_k\}$ of the transfer matrix and the coefficients $c_{G,j}$ are algebraic functions of q and v.

One of the basic structural properties of the Potts model partition function on a given strip is the number of different eigenvalues of the transfer matrix (in the Fortuin-Kasteleyn representation), $N_{Z,G,\mathrm{BC},L}$, in eq. (2.1). In [3], in addition to proving various formulas for these numbers for certain strip graphs, we presented a conjecture (denoted Conjecture 3 in [3] and given as Theorem 4.3.5 by one of us (S.-C.C.) in [36]). Using extensions of the sort of reasoning employed in [3, 36], we have now succeeded in proving this conjecture. We have

Theorem 2.1 *For arbitrary L,*

$$2N_{Z,\mathrm{sq},\mathrm{PF},L} - N_{Z,\mathrm{tri},\mathrm{PF},L} = \begin{cases} N_{Z,\mathrm{sqtri},\mathrm{FP},\frac{L}{2}} & \text{for even } L \\ \frac{1}{2} N_{Z,\mathrm{sqtri},\mathrm{FP},\frac{L+1}{2}} & \text{for odd } L \end{cases} \tag{2.4}$$

where the quantity $N_{Z,\mathrm{sqtri},\mathrm{FP},L}$ is given by [20]:

$$N_{Z,\mathrm{sqtri},\mathrm{FP},L} = \binom{2L}{L} \tag{2.5}$$

Proof We recall first that the quantity $2N_{Z,\mathrm{sq},\mathrm{FF},L} - N_{Z,\mathrm{tri},\mathrm{FF},L}$ discussed in [3] gives the number of non-crossing partitions for a transverse slice of these two respective strips (which is the path graph T_L) such that these partitions are symmetric under

reflection about the longitudinal axis. The quantity $2N_{Z,\mathrm{sq,PF},L} - N_{Z,\mathrm{tri,PF},L}$ gives the corresponding number of non-crossing partitions for a transverse slice (which is the circuit graph C_L) of the two respective cylindrical strips such that these partitions are symmetric under reflections about the longitudinal axis and rotations around this axis (the latter being included since there is no special azimuthal direction). We shall prove eq. (2.4) for odd L first and then for even L.

For odd L, let $n = (L+1)/2$, and denote $2N_{Z,\mathrm{sq,PF},L=2n-1} - N_{Z,\mathrm{tri,PF},L=2n-1}$ as X_n for simplicity. Consider a transverse slice with periodic boundary conditions. Since this is topologically invariant under rotations around the longitudinal axis, we can label one vertex as 1 and other vertices $2, 3, \ldots, n, n', (n-1)', \ldots, 2'$, in a counterclockwise manner (relative to a specified longitudinal direction), and consider the reflection symmetry with respect to the longitudinal axis passing through vertex 1. In order to classify the types of colorings of the vertices, we shall introduce diagrams consisting of the L vertices on this transverse slice. In this context, we shall refer to two vertices as being "connected" if these have the same color and shall denote this by using the Kronecker delta function.

The sets $\mathbf{P}_{X_n}$ of partitions for $n = 1, 2, 3$ that are invariant under this reflection symmetry are: $\mathbf{P}_{X_1} = \{1\}$, $\mathbf{P}_{X_2} = \{1, \delta_{2,2'}, \delta_{1,2}\delta_{1,2'} = \delta_{1,2,2'}\}$, and

$$\begin{aligned}\mathbf{P}_{X_3} \;=\; \{&1, \delta_{2,2'},\ \delta_{3,3'},\ \delta_{1,2,2'},\ \delta_{1,3,3'},\ \delta_{2,3}\delta_{2',3'},\ \delta_{2,2'}\delta_{3,3'},\\ &\delta_{2,3,2',3'},\ \delta_{1,2,2'}\delta_{3,3'},\ \delta_{1,2,3,2',3'}\}\end{aligned} \tag{2.6}$$

We can classify these partitions into cases that have m vertices on one side of the slice (including vertex 1) connected to at least one other vertex (on the same side or the other side, indicated by the primes above), where $0 \le m \le n$. For $m = 0$, this is the partition 1, that is the identity partition, defined as the one in which all blocks are "singletons", i.e., there are no connections in the sense given above. For $m = 1$, there is only one possibility: $\delta_{x_1,x_1'}$, where $2 \le x_1 \le n$. For $m = 2$, let us denote the connected vertices as x_1, x_2 and, with no loss of generality, take $x_1 < x_2$; then there are the partitions δ_{x_1,x_2}, $\delta_{x_1,x_2,x_1'}$, and $\delta_{x_1,x_1'}\delta_{x_2,x_2'}$. For the first case, x_1 can be vertex 1, but for the second and third cases, $2 \le x_1 \le n$ since x_1 and x_1' are different vertices. The corresponding partitions $\delta_{x_1',x_2'}$ for the first and second cases can be obtained by reflection symmetry from the ones that we have listed and hence, for simplicity, are not shown. With no loss of generality, we take $x_1 < x_2 < \ldots < x_m$. For $m = 3$, the partitions are $\delta_{x_1,x_2}\delta_{x_3,x_3'}$, δ_{x_1,x_2,x_3}, $\delta_{x_1,x_2,x_1'}\delta_{x_3,x_3'}$, $\delta_{x_1,x_2,x_3,x_1'}$, $\delta_{x_1,x_1'}\delta_{x_2,x_3}$, $\delta_{x_1,x_1'}\delta_{x_2,x_3,x_2'}$, $\delta_{x_1,x_1'}\delta_{x_2,x_2'}\delta_{x_3,x_3'}$. Having given these illustrations of the specific partitions for $0 \le m \le 3$, we next proceed to the general case.

The partitions that have m vertices on one side of the slice connected to as least one other vertex (on the same or opposite side) can be classified further. Let us denote a_m as the number of cases where the vertex x_1 has only the connection $\delta_{x_1,x_1'}$ for $1 \le m$. In these cases, x_1 cannot be vertex 1, and the number of the partitions for each $m \le n-1$ is $\binom{n-1}{m}$. The last three partitions of $m = 3$ given above are examples of these cases. We denote b_m as the number of cases where the vertex x_1 has connection to at least one unprimed vertex with or without $\delta_{x_1,x_1'}$ for $2 \le m$. The first and second partitions for $m = 3$ are examples of these cases without $\delta_{x_1,x_1'}$, and

the third and fourth partitions for $m = 3$ are examples of these cases with $\delta_{x_1,x_1'}$. Notice that while x_1 cannot be vertex 1 for the cases with $\delta_{x_1,x_1'}$, which have the number of partitions $\binom{n-1}{m}$ for each $m \leq n-1$, x_1 can be vertex 1 for the cases without $\delta_{x_1,x_1'}$, and the number of partitions for each $m \leq n$ is $\binom{n}{m}$. Therefore,

$$X_n = 1 + \sum_{m=1}^{n-1} a_m \binom{n-1}{m} + \sum_{m=2}^{n-1} b_m \binom{n-1}{m} + \sum_{m=2}^{n} b_m \binom{n}{m} \tag{2.7}$$

Next, we shall obtain expressions for a_m and b_m. The cases for a_m can be obtained from all the cases with $m-1$ vertices by changing x_i to x_{i+1} for $1 \leq i \leq m-1$ and adding $\delta_{x_1,x_1'}$, i.e., a_m is the same as the total number of the cases with $m-1$ vertices,

$$a_m = a_{m-1} + 2b_{m-1} \tag{2.8}$$

Now consider the cases for b_m without $\delta_{x_1,x_1'}$. These can be further divided into two possibilities: the cases with δ_{x_1,x_2} and the cases with δ_{x_1,x_i} where $2 < i \leq m$. Denote the numbers of these two possibilities as d_m and e_m, respectively. Clearly,

$$b_m = d_m + e_m \tag{2.9}$$

The cases for d_m can be obtained from the cases for b_{m-1} by changing x_i to x_{i+1} for $1 \leq i \leq m-1$ and adding δ_{x_1,x_2}, and from all the cases with $m-2$ vertices, where the number is a_{m-1}, by changing x_i to x_{i+2} for $1 \leq i \leq m-2$ and adding δ_{x_1,x_2}. Therefore,

$$d_m = a_{m-1} + b_{m-1} \tag{2.10}$$

The cases for e_m can be obtained from the cases for d_{m-2} by changing x_i to x_{i+2} for $2 \leq i \leq m-2$ and adding δ_{x_2,x_3}, from the cases for d_{m-3} by changing x_i to x_{i+3} for $2 \leq i \leq m-3$ and adding δ_{x_2,x_3,x_4}, etc. In general, the cases for e_m are obtained from the cases for $d_{m-\ell}$ by changing x_i to $x_{i+\ell}$ for $2 \leq i \leq m-\ell$ and $2 \leq \ell \leq m-2$, and adding a set of ℓ vertices, where all the vertices in this set must have at least one connection to at least one other vertex in this set without reflection or rotation symmetry. The number of all possible connections of this set with ℓ vertices is the Riordan number (given as sequence A005043 in [37]), which will be denoted as r_ℓ and may be defined via the generating function [38]

$$R(z) = \frac{1 + z - (1 - 2z - 3z^2)^{1/2}}{2z(1+z)} = \sum_{\ell=0}^{\infty} r_\ell z^\ell \tag{2.11}$$

Therefore,

$$e_m = \sum_{\ell=2}^{m-2} d_{m-\ell}\, r_\ell \tag{2.12}$$

We list the first few numerical values of a_m, b_m, d_m, e_m, r_m in Table 1. In terms of generating functions

$$A(z) \;\; = \;\; a_1 z + a_2 z^2 + a_3 z^3 + \ldots$$

$$
\begin{aligned}
B(z) &= b_2 z^2 + b_3 z^3 + b_4 z^4 + \ldots \\
D(z) &= d_2 z^2 + d_3 z^3 + d_4 z^4 + \ldots \\
E(z) &= e_4 z^4 + e_5 z^5 + e_6 z^6 + \ldots \\
R(z) &= 1 + r_2 z^2 + r_3 z^3 + r_4 z^4 + \ldots
\end{aligned} \tag{2.13}
$$

(recalling the zero entries in Table 1), we can re-express eqs. (2.8) to (2.12) as

$$
\begin{aligned}
\frac{A(z) - a_1 z}{z} &= A(z) + 2B(z) \\
B(z) &= D(z) + E(z) \\
\frac{D(z)}{z} &= A(z) + B(z) \\
E(z) &= D(z)(R(z) - 1)
\end{aligned} \tag{2.14}
$$

We find

$$
A(z) = \frac{z}{\sqrt{1 - 2z - 3z^2}} \tag{2.15}
$$

The coefficients in the expansion of $A(z)$ in eq. (2.13) are, up to a shift, the central trinomial coefficients (given as sequence A002426 in [37]), i.e., for each value $1 \leq m$ the largest coefficient of $(1 + z + z^2)^{m-1}$. Next,

$$
B(z) = \frac{1}{2}\Big[\frac{1 - z}{\sqrt{1 - 2z - 3z^2}} - 1\Big] \tag{2.16}
$$

The coefficients in the expansion of $B(z)$ in (2.13) are given by the coefficients for the next-to-central column in the expansion of $(1 + z + z^2)^{m-1}$ for $1 \leq m$ (listed as sequence A005717 in [37]). For $D(z)$ we have the closed form

$$
D(z) = \frac{z}{2}\Big[\sqrt{\frac{1 + z}{1 - 3z}} - 1\Big] \tag{2.17}
$$

The coefficients in the expansion of $D(z)$ in (2.13) are the numbers of directed animals of size $m - 1$ on the square lattice for $1 \leq m$ (given as sequence A005773 in [37]). Finally, we have

$$
E(z) = z\frac{1 - 3z + (1 - 2z)\sqrt{1 - 2z - 3z^2}}{1 - 2z - 3z^2 + (3z - 1)\sqrt{1 - 2z - 3z^2}} - 1 + z \tag{2.18}
$$

The coefficients in the expansion of $E(z)$ in (2.13) are the coefficients forming the second column from the center, in a tabular format, in the expansion of $(1+z+z^2)^{m-2}$ for $2 \leq m$ (given as sequence A014531 in [37]).

Recall the binomial transformation for two sequences of numbers $[s_0, s_1, s_2, \ldots]$, $[t_0, t_1, t_2, \ldots]$ with generating functions $S(z) = \sum_{n=0}^{\infty} s_n z^n$ and $T(z) = \sum_{n=0}^{\infty} t_n z^n$. If these sequences have the relation

$$t_n = \sum_{m=0}^{n} \binom{n}{m} s_m \ , \tag{2.19}$$

then [39]

$$T(z) = \frac{1}{1-z} S(\frac{z}{1-z}) \tag{2.20}$$

In our case, if the generating function of X_n is $X(z) = \sum_{n=1}^{\infty} X_n z^n$, then we can combine eqs. (2.7), (2.15) and (2.16) to get

$$\begin{aligned} X(z) &= \frac{1}{1-z} - 1 + \frac{z}{1-z} A\left(\frac{z}{1-z}\right) + \frac{z}{1-z} B\left(\frac{z}{1-z}\right) + \frac{1}{1-z} B\left(\frac{z}{1-z}\right) \\ &= \frac{1}{2}\Big[\frac{1}{\sqrt{1-4z}} - 1\Big] \end{aligned} \tag{2.21}$$

This has the expansion

$$X(z) = \sum_{n=1}^{\infty} \binom{2n-1}{n} z^n \tag{2.22}$$

Note that $\binom{2n-1}{n} = \frac{1}{2}\binom{2n}{n} = \frac{1}{2}\binom{L+1}{(L+1)/2}$ (given as sequence A001700 in [37]).

Let us next proceed to consider the number of non-crossing partitions for a slice of the transverse vertices which has periodic boundary conditions and reflection symmetry for even L. Denote $n = L/2$, and $2N_{Z,\mathrm{sq,PF},L=2n} - N_{Z,\mathrm{tri,PF},L=2n}$ by Y_n for simplicity. There are two possibilities: the reflection axis does not go through any vertex or goes through two vertices. These possibilities will be denoted as type I and and type II partitions, respectively, and the number of partitions of each of these two types as Y_n^{I} and Y_n^{II}. For type I partitions, label the vertices on one side of the reflection axis as 1, 2,..., n and the corresponding reflected vertices as $1', 2', \ldots, n'$ on the other side. For type II partitions, label the vertices as 1, 2, ..., n, $n+1$, n', ..., $2'$, where vertices 1 and $n+1$ are on the reflection axis. The sets $\mathbf{P}_{Y_n^{\mathrm{I}}}$ of type I partitions for $n = 1, 2$ having this reflection symmetry are: $\mathbf{P}_{Y_1^{\mathrm{I}}} = \{1, \delta_{1,1'}\}$ and $\mathbf{P}_{Y_2^{\mathrm{I}}} = \{1, \delta_{1,1'}, \delta_{2,2'}, \delta_{1,2}\delta_{1',2'}, \delta_{1,2,1'}, \delta_{1,1'}\delta_{2,2'}\}$. The sets $\mathbf{P}_{Y_n^{\mathrm{II}}}$ of type II partitions for $n = 1, 2$ having this reflection symmetry are: $\mathbf{P}_{Y_1^{\mathrm{II}}} = \{1, \delta_{1,2}\}$ and $\mathbf{P}_{Y_2^{\mathrm{II}}} = \{1, \delta_{2,2'}, \delta_{1,3}, \delta_{1,2,2'}, \delta_{2,3,2'}, \delta_{1,2,3,2',3'}\}$. We notice that partitions 1, i.e. identity and $\delta_{1,2,\ldots,L}$ (i.e., a unique block) are both contained in the type I and type II classes of partitions. Since we have rotational symmetry, the partitions which are not symmetric with respect to the central axis perpendicular to the reflection axis are counted twice in either type I or type II classes of partitions. A similar statement applies to the partitions which are symmetric with respect to the axis perpendicular to the reflection axis if L is a multiple of 4. If L is not a multiple of 4, the partitions that are symmetric with respect to the perpendicular axis are counted once in both type I and type II classes. Therefore, $2Y_n$ is the sum of all possible partitions in these two classes of partitions.

We again classify these partitions into cases which have m vertices on one side of the slice connected to at least one other vertex (including vertex 1 and vertex $n+1$ for type II partitions). It is clear that $0 \le m \le n$ for type I partitions and $0 \le m \le n+1$ for type II partitions. Let us consider type I first. For $m = 0$, there is the partition 1 (identity). For $m = 1$, there is only one possibility: $\delta_{x_1,x_1'}$, where $1 \le x_1 \le n$. For $m = 2$, there are δ_{x_1,x_2}, $\delta_{x_1,x_2,x_1'}$ and $\delta_{x_1,x_1'}\delta_{x_2,x_2'}$, where $x_1 < x_2$. The corresponding $\delta_{x_1',x_2'}$ for the first and second cases are not shown for simplicity. We shall again take $x_1 < x_2 < \ldots < x_m$. The number of the cases with m vertices connected is $a_m + 2b_m = a_{m+1}$ as discussed above eq. (2.8), and the number of the partitions is $\binom{n}{m}$ for each $0 \le m \le n$. Therefore,

$$Y_n^{\rm I} = \sum_{m=0}^{n} a_{m+1} \binom{n}{m} \quad \text{for } 1 \le n \tag{2.23}$$

Let the generating function of $Y_n^{\rm I}$ be $Y^{\rm I}(z) = \sum_{n=1}^{\infty} Y_n^{\rm I} z^n$. Using eq. (2.15) and modifying eq. (2.20), we have

$$Y^{\rm I}(z) = \frac{1}{1-z}\left[\frac{A(\frac{z}{1-z})}{\frac{z}{1-z}}\right] - a_1 = \frac{1}{\sqrt{1-4z}} - 1 \tag{2.24}$$

For type II partitions, neither vertex 1 nor vertex $n+1$ has a corresponding symmetric partner, $1'$ and $(n+1)'$, respectively. We have to classify the cases where x_1 is only connected to at least one other unprimed vertex (in these cases, x_1 can be vertex 1), the number of which cases was denoted b_m earlier, into two possibilities: the cases where x_m is only connected to at least one other unprimed vertex (so that x_m can be vertex $n+1$), and the cases where x_m is connected to x_m' (so that x_m cannot be vertex $n+1$) among other possible connections. The number of partitions for these two possibilities will be denoted as f_m for $2 \le m$ and h_m for $3 \le m$, respectively. The partitions where x_1 is connected to x_1', i.e., $\delta_{x_1,x_1'}$, among other possible connections can also be classified into two possibilities: the cases where x_m is only connected to other unprimed vertices, and the cases where x_m is connected to x_m' among other possible connections. The number of partitions for these two possibilities will be denoted as h_m' for $3 \le m$ and i_m for $1 \le m$, respectively. Notice that $h_m' = h_m$ because of the reflection symmetry. We shall need another set of partitions where x_1 is connected to x_m with possible connection to other unprimed vertices, and all the other vertices are connected to at least one other unprimed vertex; we denote the number of partitions for them as j_m for $2 \le m$. If we add an additional connection $\delta_{x_1,x_1'}$ to these partitions, and denote the number of partitions as k_m, then $k_1 = 1$ and $k_m = j_m$ for $2 \le m$.

For the cases for f_m, x_1 and x_m can be vertices 1 and $n+1$, respectively, so the number of the partitions is $\binom{n+1}{m}$ for each m. For the cases for h_m, x_1 can be vertex 1 but x_m cannot be vertex $n+1$, so the number of the partitions is $\binom{n}{m}$ for each m. For the cases for i_m, neither x_1 can be vertex 1 nor x_m can be vertex $n+1$, so the

number of the partitions is $\binom{n-1}{m}$ for each m. Therefore,

$$Y_n^{\rm II} = 1 + \sum_{m=2}^{n+1} f_m \binom{n+1}{m} + 2\sum_{m=3}^{n} h_m \binom{n}{m} + \sum_{m=1}^{n-1} i_m \binom{n-1}{m} \tag{2.25}$$

We next obtain expressions for f_m, h_m and i_m. From the definitions,

$$f_m + h_m = b_m \tag{2.26}$$

and

$$h'_m + i_m = a_m + b_m = d_{m+1} \ , \tag{2.27}$$

where we use eq. (2.10). The cases for f_m include all the cases for j_m, plus the cases for both f_{m-2} and h_{m-2} with the additional connection δ_{x_{m-1},x_m}, plus the cases for both f_{m-3} and h_{m-3} with the additional connection $\delta_{x_{m-2},x_{m-1},x_m}$, etc. We have

$$f_m = j_m + \sum_{\ell=2}^{m-2} j_\ell \ [f_{m-\ell} + h_{m-\ell}] = j_m + \sum_{\ell=2}^{m-2} j_\ell \ b_{m-\ell} \tag{2.28}$$

The cases for h_m can be obtained by adding δ_{x_m,x'_m} to the cases for f_{m-1} and h_{m-1}, and adding $\delta_{x_{m-1},x_m,x'_m}$ to the cases for f_{m-2} and h_{m-2}, etc. We have

$$h_m = \sum_{\ell=1}^{m-2} k_\ell \ [f_{m-\ell} + h_{m-\ell}] = \sum_{\ell=1}^{m-2} k_\ell \ b_{m-\ell} \tag{2.29}$$

The cases for h'_m can be obtained by adding δ_{x_{m-1},x_m} to the cases for h'_{m-2} and i_{m-2}, and adding $\delta_{x_{m-2},x_{m-1},x_m}$ to the cases for h'_{m-3} and i_{m-3}, etc. We have

$$h'_m = \sum_{\ell=2}^{m-1} j_\ell \ [h'_{m-\ell} + i_{m-\ell}] = \sum_{\ell=2}^{m-1} j_\ell \ d_{m-\ell+1} \ , \tag{2.30}$$

which should be equal to h_m, as mentioned before. The cases for i_m include all the cases for k_m, plus the cases for both h'_{m-1} and i_{m-1} with additional connection δ_{x_m,x'_m}, plus the cases for both h'_{m-2} and i_{m-2} with additional connection $\delta_{x_{m-1},x_m,x'_m}$, etc. We have

$$i_m = k_m + \sum_{\ell=1}^{m-1} k_\ell \ [h'_{m-\ell} + i_{m-\ell}] = k_m + \sum_{\ell=1}^{m-1} k_\ell \ d_{m-\ell+1} \tag{2.31}$$

The cases for j_m can be obtained from the cases for r_m by removing the cases for r_{m-2} with additional δ_{x_{m-1},x_m}, removing the cases for r_{m-3} with additional $\delta_{x_{m-2},x_{m-1},x_m}$, etc. We have

$$j_m = r_m - \sum_{\ell=2}^{m-2} j_\ell \ r_{m-\ell} \tag{2.32}$$

We list the first few numerical values of f_m, h_m, i_m, j_m, k_m in Table 2. In terms of the generating functions

$$F(z) \ = \ f_2 z^2 + f_3 z^3 + f_4 z^4 + \ldots$$

$$
\begin{aligned}
H(z) &= H(z)' = h_3 z^3 + h_4 z^4 + h_5 z^5 + \ldots \\
I(z) &= i_1 z + i_2 z^2 + i_3 z^3 + \ldots \\
J(z) &= j_2 z^2 + j_3 z^3 + j_4 z^4 + \ldots \\
K(z) &= k_1 z + k_2 z^2 + k_3 z^3 + \ldots
\end{aligned} \tag{2.33}
$$

we can reexpress eqs. (2.26) to (2.32) as

$$
\begin{aligned}
F(z) + H(z) &= B(z) \\
H(z) + I(z) &= \frac{D(z)}{z} \\
F(z) &= J(z) + B(z)J(z) \\
H(z) &= B(z)K(z) = \frac{D(z)}{z} J(z) \\
I(z) &= K(z) + \frac{D(z)}{z} K(z) \\
J(z) &= R(z) - 1 - (R(z) - 1)J(z) = (R(z) - 1)(1 - J(z))
\end{aligned} \tag{2.34}
$$

We find

$$
J(z) = \frac{R(z) - 1}{R(z)} = \frac{1}{2}(1 - z - \sqrt{1 - 2z - 3z^2}) \ , \tag{2.35}
$$

which is essentially the Motzkin number [40, 41] (given as sequence A001006 in [37]). This can be understood since the cases for j_m are in one-to-one correspondence with the non-crossing, non-nearest-neighbor partitions of $m-1$ vertices with free boundary condition, as given in eq. (2.1.1) of [3], i.e., $j_m = M_{m-2}$. We also have

$$
K(z) = z + J(z) = \frac{1}{2}(1 + z - \sqrt{1 - 2z - 3z^2}) \tag{2.36}
$$

The first five equations in eq. (2.34) are redundant, and can be solved to have

$$
F(z) = \frac{z^2}{\sqrt{1 - 2z - 3z^2}} = zA(z) \ , \tag{2.37}
$$

i.e., $f_m = a_{m-1}$, which is the largest coefficient of $(1 + z + z^2)^{m-2}$ for $2 \leq m$. Next, we have

$$
H(z) = \frac{1}{2}\left[\left(\frac{1 - 2z}{1 - 3z}\right)\sqrt{1 - 2z - 3z^2} - 1\right] \tag{2.38}
$$

The coefficients in the expansion of $H(z)$ are the numbers of directed animals of size $m-2$ on the square lattice with the first two quadrants (which can grow in right, left and up directions) for $2 \leq m$ (given as sequence A005774 in [37]). Finally,

$$
I(z) = z\sqrt{\frac{1 + z}{1 - 3z}} = 2D(z) + z \ , \tag{2.39}
$$

which is given as sequence A0025565 in [37].

Denote the generating function of $Y_n^{\rm II}$ as $Y^{\rm II}(z) = \sum_{n=1}^\infty Y_n^{\rm II} z^n$. Combining eqs. (2.25), (2.37), (2.38), (2.39), and using the binomial transformation (2.20), we obtain the relation

$$\begin{aligned} Y^{\rm II}(z) &= \frac{1}{1-z} - 1 + \frac{1}{z(1-z)} F(\frac{z}{1-z}) + \frac{1}{1-z} H(\frac{z}{1-z}) + \frac{z}{1-z} I(\frac{z}{1-z}) \\ &= \frac{1}{1-z} - 1 + \frac{z}{(1-z)^2\sqrt{1-4z}} + \frac{1}{2(1-z)^2}\Big[\frac{1-3z}{\sqrt{1-4z}} + z - 1\Big] \\ &\quad + \frac{z^2}{(1-z)^2\sqrt{1-4z}} \\ &= \frac{1}{\sqrt{1-4z}} - 1 \ , \end{aligned} \tag{2.40}$$

which is the same as the $Y^{\rm I}(z)$ in eq. (2.24). We finally have

$$Y(z) = \frac{Y^{\rm I}(z) + Y^{\rm II}(z)}{2} = \frac{1}{\sqrt{1-4z}} - 1 \ , \tag{2.41}$$

which is the generating function of $\binom{2n}{n} = \binom{L}{L/2}$, i.e., the central binomial coefficients given as sequence A000984 in [37].

Eqs. (2.21) and (2.41) prove the theorem for odd and even L, respectively. $\square$

Using extensions of the methods in [3, 36], we have also succeeded in proving Conjecture 4 in [3] (given as Theorem 4.3.7 in [36]) for the number of λ's for the triangular lattice with cylindrical boundary condition for the Potts model partition function:

Theorem 2.2 *For arbitrary L,*

$$N_{Z,{\rm tri},{\rm PF},L} = \frac{1}{L}\Big[C_L + \sum_{d|L;\ 1\le d<L} \phi(L/d) \binom{2d}{d}\Big] \tag{2.42}$$

where $d|L$ means that d divides L and $\phi(n)$ is the Euler function, equal to the number of positive integers not exceeding the positive integer n and relatively prime to n.

Proof As shown in [19], $N_{Z,{\rm tri},{\rm FF},L} = C_L$ is the number of non-crossing partitions of a set of L vertices with free boundary conditions. $LN_{Z,{\rm tri},{\rm PF},L} - N_{Z,{\rm tri},{\rm FF},L}$ is equal to $2(L-1)$ for prime L, as shown in [3, 36], since all the partitions of $N_{Z,{\rm tri},{\rm PF},L}$ have periodicity L for prime L except for the partitions 1 (i.e., all blocks being singletons) and $\delta_{1,2,\ldots,L}$ (i.e., a unique block) which have periodicity 1. Consider a general L and assume that L has the factor d (so that for prime L, d can only be 1 or L). Denote the number of partitions which have periodicity d modulo rotations as $2\alpha_d$. Then

$$LN_{Z,{\rm tri},{\rm PF},L} - N_{Z,{\rm tri},{\rm FF},L} = \sum_{d|L} 2\alpha_d (L-d) \tag{2.43}$$

where d are all the positive integers that divide L.

There are two kinds of partitions among $2\alpha_d$, each of which contains α_d specific partitions. The first kind of partitions is defined by the condition that each set of d adjacent vertices does not have connection with any other vertex in the complementary subset of vertices. Let us label the vertices in one set of d vertices as 1,2,...,d. For $d = 1$, the set of partition is $\{1\}$. For $d = 2$, the set of partition is $\{\delta_{1,2}\}$. For $d = 3$, the set of partitions is $\{\delta_{1,2}, \delta_{1,3}, \delta_{1,2,3}\}$. The second kind of partitions is defined by the condition that all of the first vertices of each of the sets of d vertices are connected to each other. For $d = 1$, the set of partition is $\{\delta_{1,1'}\}$, where $1'$ is the first vertex of the adjacent set of d vertices. For $d = 2$, the set of partition is $\{\delta_{1,1'}\}$. For $d = 3$, the set of partitions is $\{\delta_{1,1'}, \delta_{1,1'}\delta_{2,3}, \delta_{1,1',3}\}$. Since there is a one-to-one correspondence between these two kinds of partitions, let us only consider the second kind of the partitions which always have the connection $\delta_{1,1'}$. Denote α'_d as the number of partitions which has periodicity less than or equal to d modulo rotations. This can be written in term of α_d as

$$\alpha'_d = \sum_{d'|d} \alpha_{d'} \tag{2.44}$$

For the set of d vertices, if vertex 1 does not connect to any other vertex except $1'$, the first vertex of an adjacent set, then the number of partitions is equal to $N_{Z,\mathrm{tri},\mathrm{FF},d-1} = C_{d-1}$. If vertex 1 has a connection to vertex b for $1 < b < d$ in addition to vertex $1'$, then the number of partitions is $C_{b-2}C_{d-b}$. The connection of vertex 1 to vertex $d-b+2$ is equivalent to the connection of vertex 1 to vertex b under rotation symmetry, and should not be considered again. Of course, vertex 1 can have connection to more than one vertices in addition to $\delta_{1,1'}$. Therefore, to calculate α'_d, we have to partition d first. For each partition of d, e.g., $[x_1, x_2, \ldots, x_i]$ with $x_1 + x_2 + \ldots + x_i = d$, where i is the numbers of vertices in the set of d vertices connected to vertex 1 (including vertex 1), we multiply all the corresponding C_{x_i-1}'s by the number of different combinations of this partition modulo rotations (thus [2,2,1,1] is the same as [2,1,1,2] and [1,1,2,2], but different from [2,1,2,1]). As an example, for $d = 5$, we have the partitions $5 = 4+1 = 3+2 = 3+1+1 = 2+2+1 = 2+1+1+1 = 1+1+1+1+1$, and therefore

$$\begin{aligned}
\alpha'_5 &= C_4 + \frac{1}{2}C_3C_0\binom{2}{1} + \frac{1}{2}C_2C_1\binom{2}{1} + \frac{1}{3}C_2C_0^2\binom{3}{1} + \frac{1}{3}C_1^2C_0\binom{3}{2} \\
&\quad + \frac{1}{4}C_1C_0^3\binom{4}{1} + C_0^5\binom{5}{5} \\
&= 14 + 5\times 1 + 2\times 1 + 2\times 1^2 + 1^2\times 1 + 1\times 1^3 + 1^5 = 26
\end{aligned} \tag{2.45}$$

The factors $\frac{1}{2}$, $\frac{1}{3}$, and $\frac{1}{4}$ are included because of the equivalence under rotations. The numbers α'_d for small d are listed in Table 3.

We find that α'_d is the same as the rooted planar trees with d edges with the property that rotations about the root vertex yield equivalent trees, given as sequence A003239 in [37]. The reason can be explained as follows. Define a "planted" tree as a rooted tree with the property that the root vertex has degree one (where the degree of a vertex is the number of edges connected to it). Now we can construct planted

trees from subtrees of the rooted planar tree that are each connected to the root vertex. Note that the number of planted tree with d edges is C_{d-1} [38]. The sequence A003239 also gives the number of necklaces with a total of $2d$ beads where d beads have one color and the other d beads have another color. The explicit formula is

$$\alpha'_d = \frac{1}{2d} \sum_{d'|d} \phi(d/d') \binom{2d'}{d'} , \tag{2.46}$$

where $\phi(n)$ was defined above after eq. (2.42).

Now the number of partitions that have periodicity equal to d modulo rotations, α_d, is the Möbius transformation [39] of α'_d, given by

$$\alpha_d = \sum_{d'|d} \mu(d/d') \alpha'_{d'} \tag{2.47}$$

where $\mu(n)$ is the Möbius function, defined as -1 if n is prime, 0 if n has a square factor, and 1 for other n. We find that α_d is listed as sequence A022553 in [37], which is the Lyndon words containing a total of $2d$ letters with d letters of one type, and the other d letters of another. The values of α_d for small d are listed in Table 3. One has the explicit formula

$$\alpha_d = \frac{1}{2d} \sum_{d'|d} \mu(d/d') \binom{2d'}{d'} , \tag{2.48}$$

i.e., $2d\alpha_d$, is the Möbius transformation of $\binom{2d}{d}$, or equivalently,

$$\binom{2d}{d} = \sum_{d'|d} 2d' \alpha_{d'} \tag{2.49}$$

Therefore, the total number of partitions that have periodicity d modulo rotations is

$$2\alpha_d = \frac{1}{d} \sum_{d'|d} \mu(d/d') \binom{2d'}{d'} \tag{2.50}$$

This is given as sequence A060165 in [37].

We finally have

$$\begin{aligned} LN_{Z,\mathrm{tri,PF},L} - N_{Z,\mathrm{tri,FF},L} &= \sum_{d|L} 2\alpha_d (L-d) \\ &= \sum_{d|L} 2L\alpha_d - \binom{2L}{L} \\ &= \sum_{d|L} \frac{L}{d} \sum_{d'|d} \mu(d/d') \binom{2d'}{d'} - \binom{2L}{L} \\ &= \sum_{d'|L} \sum_{d|L; d'|d} \frac{L}{d} \mu(d/d') \binom{2d'}{d'} - \binom{2L}{L} \end{aligned}$$

$$= \sum_{d'|L} \phi(L/d') \binom{2d'}{d'} - \binom{2L}{L}$$
$$= \sum_{d|L;\ 1\le d<L} \phi(L/d) \binom{2d}{d} \tag{2.51}$$

and the theorem follows. □

Theorems 2.1 and 2.2 imply a simple corollary which gives an exact formula for $N_{Z,\mathrm{sq},\mathrm{PF},L}$:

Corollary 2.1 *For arbitrary L,*

$$N_{Z,\mathrm{sq},\mathrm{PF},L} = \frac{1}{2}\left[N_{Z,tri,PF,L} + \binom{L}{[L/2]}\right] \tag{2.52}$$

Proof For even L is trivial. For odd L, we only have to notice that

$$(1/2)\binom{L+1}{(L+1)/2} = \frac{L(L-1)\cdots\frac{L+3}{2}}{(\frac{L-1}{2})!} = \binom{L}{(L-1)/2} = \binom{L}{[L/2]} \tag{2.53}$$

This proves the corollary. □

One of us (S.-C.C.) has presented a conjecture, Conjecture 4.2.2 in [36], which generalizes Theorem 3 in [3] (i.e., Theorem 4.2.6 in [36]). We restate this conjecture here. Consider the number of λ's in the chromatic polynomial for a strip of the triangular lattice with cylindrical boundary conditions:

Conjecture 2.1 *For arbitrary L,*

$$N_{P,\mathrm{tri},\mathrm{PF},L} = \frac{1}{L}\Big[d_L + \sum_{d|L;\ 1\le d<L} \phi(L/d) t_d\Big] \tag{2.54}$$

where $t_d = a_{d+1}$ is the largest coefficient in the expansion of $(1+z+z^2)^d$, i.e., the central trinomial coefficient (given as sequence A002426 in [37]), and d_L is essentially the Riordan number r_L (given as sequence A005043 in [37])

$$d_L = \begin{cases} 1 & L = 1 \\ r_L & L \ge 2 \end{cases} \tag{2.55}$$

where the reader is cautioned not to confuse the d_L in eq. (2.54) with the different quantity d_m in eqs. (2.9) and (2.10). We can motivate this conjecture as follows. We know that d_L is the number of non-crossing non-nearest-neighbor partitions of a set of L vertices with periodic boundary conditions, as shown in [19]. The number $LN_{P,\mathrm{tri},\mathrm{PF},L} - d_L$ is equal to $L-1$ for prime L, as shown in [3, 36], since all the partitions of $N_{P,\mathrm{tri},\mathrm{PF},L}$ have periodicity L for prime L except for the partitions 1 (i.e., all blocks being singletons) which have periodicity 1. Consider a general L having

the factor d, and denote the number of non-nearest-neighbor partitions that have periodicity d modulo rotations as β_d. Then

$$LN_{P,\mathrm{tri,PF},L} - d_L = \sum_{d|L} \beta_d (L-d) \tag{2.56}$$

There are two kinds of partitions among β_d; we denote the number of these as β_d^{I} and β_d^{II}, respectively. The first kind of partitions is defined by the condition that each set of d adjacent vertices does not have connection to any other vertices in other sets. Label the d vertices as 1,2,...,d. For $d = 1$, the set of partitions consists of just $\{1\}$. There is no non-nearest-neighbor partition for $d = 2$. For $d = 3$, the set of partitions is $\{\delta_{1,3}\}$. For $d = 4$, the set of partitions is $\{\delta_{1,4}, \delta_{1,3}\}$. The second kind of partitions is defined by the condition that all of the first vertices of each set of d vertices are connected to each other. There is no non-nearest-neighbor partition for $d = 1$. For $d = 2$, the set of partitions is comprised of $\{\delta_{1,1'}\}$, where $1'$ is the first vertex of an adjacent set of d vertices. For $d = 3$, the set of partitions is $\{\delta_{1,1'}\}$. For $d = 4$, the set of partitions is $\{\delta_{1,1'}, \delta_{1,1'}\delta_{2,4}\}$. Notice that for $d = 4$, $\{\delta_{1,1',3}\}$ is the same as $\{\delta_{1,1'}\}$ for $d = 2$, and should not be included to avoid double-counting.

Consider the first kind of partitions. If the first vertex 1 is connected to the last vertex d among other possible connections with other vertices in the set of d vertices, the number of these partitions is r_{d-1}, the Riordan number (given as sequence A005043 in [37]). This can be seen by identifying the vertices 1 and d to form a circuit with $d-1$ vertices [19]. In addition, the set of d vertices can also be partitioned into several parts where the first and the last vertices of each part are connected, but parts are not connected to each other. To calculate β_d^{I}, we partition d, and apply the same transformation as illustrated in Theorem 2.2 (above Table 3) on r_{d-1}, then apply a Möbius transformation. The values of β_d^{I} for small d are listed in Table 4.

Consider the second kind of partitions. By an argument similar to that given in the proof of Theorem 2.2, if vertex 1 is not connected to any other vertex except $1'$ among the set of d vertices, then the number of partitions is just $N_{P,\mathrm{tri,FF},d-1} = M_{d-2}$. In addition, the vertex 1 can also be connected to other vertices among the set of d vertices. To calculate β_d^{II}, we again partition d, and apply the same transformation, as illustrated in Theorem 2.2 on M_{d-2}. We then apply a Möbius transformation. The values of β_d^{II} for small d are listed in Table 4. We find that β_d^{I} and β_d^{II} are the same up to $d = 10$, except for $d = 1$ and 2. The values of $\beta_d = \beta_d^{\mathrm{I}} + \beta_d^{\mathrm{II}}$ are also listed in Table 4 for small d. Although β_d is not listed in [37], we find that the following relation holds up to $d = 10$, and is similar to eq. (2.50):

$$\beta_d = \frac{1}{d} \sum_{d'|d} \mu(d/d') t_d \tag{2.57}$$

If this is correct for arbitrary d, then the conjecture is proved.

Combining Conjecture 2.1 and Conjecture 1 in [3], we have

Conjecture 2.2 *For arbitrary L,*

$$N_{P,\mathrm{sq},\mathrm{PF},L} = \begin{cases} \frac{1}{2}\left[N_{P,\mathrm{tri},\mathrm{PF},L} + \frac{1}{2}N_{P,\mathrm{sqtri},\mathrm{FP},\frac{L}{2}}\right] & \text{for even } L \\ \frac{1}{2}\left[N_{P,\mathrm{tri},\mathrm{PF},L} + \frac{1}{4}N_{P,\mathrm{sqtri},\mathrm{FP},\frac{L+1}{2}} - \frac{1}{2}r_{\frac{L-1}{2}}\right] & \text{for odd } L \geq 3 \end{cases} \tag{2.58}$$

where $N_{P,\mathrm{sqtri},\mathrm{FP},L}$ was given in [20], and r_n is the Riordan number (given as sequence A005043 in [37]).

3 Potts Model Partition Functions for Triangular-lattice Strips with Free Boundary Conditions

The Potts model partition function $Z(G,q,v)$ for a triangular-lattice strip of width L and length m with free boundary conditions is given by

$$Z(L_{\mathrm{F}} \times m_{\mathrm{F}}, q, v) = \mathbf{v}_{\mathrm{id}}^{\mathrm{T}} \cdot \mathsf{H} \cdot \mathsf{T}^{m-1} \cdot \mathbf{u} \tag{3.1a}$$

$$= \mathbf{w}_{\mathrm{id}}^{\mathrm{T}} \cdot \mathsf{T}^{m-1} \cdot \mathbf{u} \tag{3.1b}$$

where $\mathbf{w}^{\mathrm{T}} = \mathbf{v}_{\mathrm{id}}^{\mathrm{T}} \cdot \mathsf{H}$. Hereafter we shall follow the notation and the computational methods developed in [3, 19, 42, 30] (For chromatic polynomials, a related matrix formulation has been discussed in [43, 44, 45, 46].) Concerning notation, no confusion should result between the vertex set V, the variable v, and the vector $\mathbf{v}$. Here $\mathsf{T} = \mathsf{V} \cdot \mathsf{H}$ is the transfer matrix, and H (resp. V) corresponds to adding one more layer of horizontal (resp. vertical) bonds. The matrices T, V, and H act on the space connectivities of sites on the top layer, whose basis elements are indexed $\mathbf{v}_{\mathcal{P}}$ are indexed by partitions $\mathcal{P}$ of the single-layer vertex set $\{1,\ldots,L\}$. In particular, $\mathbf{v}_{\mathrm{id}} = \mathbf{v}_{\{\{1\},\{2\},\ldots,\{L\}\}}$.

To simplify the notation, we shall denote the elements of the basis $\mathbf{v}_{\mathcal{P}}$ by a shorthand using Kronecker delta functions: for instance $\mathbf{v}_{\{\{1,3\},\{2\},\{4,5\}\}}$ will be written $\delta_{1,3}\delta_{4,5}$. We denote the set of basis elements for a given strip as $\mathbf{P} = \{\mathbf{v}_{\mathcal{P}}\}$.

As we are dealing with planar lattices, only non-crossing partitions occur. The number of such partitions is given by the Catalan numbers

$$C_n = \frac{1}{n+1}\left(\begin{array}{c} 2n \\ n \end{array}\right) \tag{3.2}$$

In the triangular-lattice strip with free boundary conditions there is no additional symmetry that allows us to reduce the number of partitions; thus the dimension of the transfer matrix for width L is [19]

$$N_{Z,\mathrm{tri},\mathrm{FF},L} = C_L \tag{3.3}$$

We have obtained the transfer matrices $\mathsf{T}(L_{\mathrm{F}})$ and the vectors $\mathbf{w}_{\mathrm{id}}$ and $\mathbf{u}$ using symbolic computation with MATHEMATICA as in [19, 3]. We have double-checked these results using a different program written in PERL.

An equivalent way to present a general formula for the partition function is via a generating function. Labelling a lattice strip of a given type and width as G_m, with m the length, one has

$$\Gamma(G,q,v,z) = \sum_{m=0}^{\infty} z^m Z(G_m,q,v) \tag{3.4}$$

where $\Gamma(G,q,v,z)$ is a rational function

$$\Gamma(G,q,v,z) = \frac{\mathcal{N}(G,q,v,z)}{\mathcal{D}(G,q,v,z)} \tag{3.5}$$

with

$$\mathcal{N}(G,q,v,z) = \sum_{j=0}^{\deg_z(\mathcal{N})} A_{G,j} z^j \tag{3.6}$$

$$\mathcal{D}(G,q,v,z) = 1 + \sum_{j=1}^{N_{Z,\mathrm{tri},\mathrm{BC},L}} b_{G,j} z^j = \prod_{j=1}^{N_{Z,\mathrm{tri},\mathrm{BC},L}} (1-\lambda_{G,j} z) \tag{3.7}$$

where the subscript BC denotes the boundary conditions. In the transfer-matrix formalism, the $\lambda_{G,j}$'s in the denominator of the generating function, eq. (3.7), are the eigenvalues of T.

Strips of the triangular lattice are well-defined for widths $L \geq 2$. The partition function $Z(G,q,v)$ has been calculated (for arbitrary q, v, and m) for the strip with $L = 2_{\mathrm{F}}$ in [18] by two of us using a systematic iterative application of the deletion-contraction theorem. $Z(G,q,v)$ was also studied for arbitrary q and v and zeros calculated for $L = 3, 4$ and various lengths in [16]. Here we shall review the $L = 2$ case in the context of the transfer-matrix method. We note that for $L \geq 4_{\mathrm{F}}$, the expressions for $\mathsf{T}(L_{\mathrm{F}})$, $\mathbf{w}_{\mathrm{id}}(L_{\mathrm{F}})$ and $\mathbf{u}(L_{\mathrm{F}})$ are so lengthy that we cannot include them here. They are available from the authors on request and in the MATHEMATICA file `transfer_Tutte_tri.m` which is available with the electronic version of this paper in the `cond-mat` archive at `http://www.lanl.gov`.

3.1 $L = 2_{\mathrm{F}}$

The number of elements in the basis is equal to $C_2 = 2$: $\mathbf{P} = \{1, \delta_{1,2}\}$. In this basis, the transfer matrices and the other relevant quantities are given by

$$\mathsf{T} = \begin{pmatrix} q^2+4qv+5v^2+v^3 & (1+v)(q+3v+v^2) \\ v^2(q+3v+v^2) & v^2(1+v)(2+v) \end{pmatrix} \tag{3.8a}$$

$$\mathbf{w}_{\mathrm{id}}^{\mathrm{T}} = q\,(q+v, 1+v) \tag{3.8b}$$

$$\mathbf{u}^{\mathrm{T}} = (1,0) \tag{3.8c}$$

Because certain expressions recur in transfer matrices for wider strips, it is convenient to re-express (3.8) in terms of these expressions; we have

$$\mathsf{T} = \begin{pmatrix} T_0 & D_1 E_3 \\ v^2 E_3 & v^2 D_1 D_2 \end{pmatrix} \tag{3.9a}$$

$$\mathbf{w}_{\mathrm{id}}^{\mathrm{T}} = q\,(F_1, D_1) \tag{3.9b}$$

where

$$D_k = v + k \tag{3.10a}$$
$$F_k = q + kv \tag{3.10b}$$
$$E_k = q^2 + kv + v^2 \tag{3.10c}$$
$$T_0 = q^2 + 4qv + 5v^2 + v^3 \tag{3.10d}$$

In terms of this transfer matrix and these vectors one calculates the partition function $Z(G_m, q, v)$ for the strip with a given length m via eq. (3.1). Equivalently, one can calculate the partition function using a generating function, and this was the way in which the results were presented in [18], with

$$\mathcal{N} = \prod_{j=1}^{2} (1 - \lambda_{tf2,j} z) \tag{3.11}$$

and

$$\lambda_{tf2,(1,2)} = \frac{1}{2}\Big[T_{S12} \pm (q + 3v + v^2)\sqrt{R_{S12}}\ \Big] \tag{3.12}$$

where

$$T_{S12} = v^4 + 4v^3 + 7v^2 + 4qv + q^2 \tag{3.13}$$

and

$$R_{S12} = q^2 + 2qv - 2qv^2 + 5v^2 + 2v^3 + v^4 \tag{3.14}$$

The product of these eigenvalues, which is the determinant of T, is

$$\det(\mathsf{T}) = v^2(1+v)^2(v+q)^2 = v^2 D_1^2 F_1^2 \tag{3.15}$$

The vanishing of this determinant at $v = -1$ and $v = -q$ occur because in each case one of the two eigenvalues is absent for, respectively, the chromatic and flow polynomials [47]. Analogous formulas can be given for $\det(\mathsf{T})$ for higher values of L; we omit these for brevity.

3.2 $L = 3_{\mathrm{F}}$

The number of elements in the basis is equal to $C_3 = 5$: $\mathbf{P} = \{1, \delta_{1,2}, \delta_{1,3}, \delta_{2,3}, \delta_{1,2,3}\}$. In this basis, the transfer matrices and the other relevant quantities are given by

$$\mathsf{T} = \begin{pmatrix} T_1 & D_1F_3E_3 & T_2 & D_1T_3 & D_1^2T_4 \\ v^2F_3E_3 & v^2D_1D_2F_2 & v^2T_5 & v^2D_1T_4 & v^2D_1^2D_2 \\ v^3E_3 & v^2D_1E_3 & v^2T_5 & v^3D_1D_2 & v^2D_1^2D_2 \\ v^2E_3E_4 & v^2D_1D_2E_3 & v^2T_6 & v^2D_1D_2E_4 & v^2D_1^2D_2^2 \\ v^4D_3E_3 & v^3D_1D_2E_3 & v^3T_7 & v^4D_1D_2D_3 & v^3D_1^2D_2^2 \end{pmatrix} \tag{3.16a}$$
$$\mathbf{w}_{\mathrm{id}}^{\mathrm{T}} = q\left(F_1^2, D_1F_1, E_2, D_1F_1, D_1^2\right) \tag{3.16b}$$
$$\mathbf{u}^{\mathrm{T}} = (1,0,0,0,0) \tag{3.16c}$$

where the T_k are shorthand notations used in this section, defined as

$$T_1 = q^3 + 7q^2v + 19qv^2 + 19v^3 + 2qv^3 + 5v^4 \tag{3.17a}$$
$$T_2 = q^2 + 7qv + 16v^2 + qv^2 + 9v^3 + 2v^4 \tag{3.17b}$$
$$T_3 = q^2 + 6qv + 11v^2 + qv^2 + 4v^3 \tag{3.17c}$$
$$T_4 = q + 5v + 2v^2 \tag{3.17d}$$
$$T_5 = q + 6v + 4v^2 + v^3 \tag{3.17e}$$
$$T_6 = 2q + 13v + 13v^2 + 6v^3 + v^4 \tag{3.17f}$$
$$T_7 = q + 12v + 13v^2 + 6v^3 + v^4 \tag{3.17g}$$

4 Potts Model Partition Functions for Triangular-lattice Strips with Cylindrical Boundary Conditions

The Potts model partition function $Z(G, q, v)$ for a triangular-lattice strip of width L vertices and length m vertices with cylindrical boundary conditions can be written as

$$Z(L_{\mathrm{P}} \times m_{\mathrm{F}}, q, v) = \mathbf{v}_{\mathrm{id}}^{\mathrm{T}} \cdot \mathsf{H} \cdot \mathsf{T}^{m-1} \cdot \mathbf{u} \tag{4.1a}$$
$$= \mathbf{w}_{\mathrm{id}}^{\mathrm{T}} \cdot \mathsf{T}^{m-1} \cdot \mathbf{u} \tag{4.1b}$$

where again $\mathbf{w}_{\mathrm{id}}^{\mathrm{T}} = \mathbf{v}_{\mathrm{id}}^{\mathrm{T}} \cdot \mathsf{H}$.

In the computation of the transfer matrix for a triangular-lattice strip with cylindrical boundary conditions there is a technical complication in order to treat correctly the last diagonal bond joining columns L and 1 [19, Section 3]. Instead of considering a triangular-lattice strip of width L and cylindrical boundary conditions, we start with a strip of width $L+1$ and free boundary conditions. The parameter v is the same for all edges, except for the vertical edges corresponding to column $L+1$, where it takes the value $v = 0$. After performing the computation, we identify columns 1 and $L+1$. This procedure implies that there are double vertical edges (with parameters v and 0) connecting sites on column 1; but their net contribution is v, as expected.

We obtain in this way a transfer matrix of dimension C_{L+1}. This matrix can be simplified by noting that there are many zero eigenvalues. Let us denote by $\{\mathbf{v}_j^{(s)}\}$ (resp. $\{\mathbf{v}_j^{(n)}\}$) the basis elements corresponding to $L+1$ being (resp. not being) a singleton. The number of elements of $\{\mathbf{v}_j^{(s)}\}$ is C_L; hence the cardinality of $\{\mathbf{v}_j^{(n)}\}$ is $C_{L+1} - C_L$. The zero eigenvalues are associated with certain eigenvectors of the form $\mathbf{v}_1 - \mathbf{v}_2$ where $\mathbf{v}_1 \in \{\mathbf{v}_j^{(s)}\}$ and $\mathbf{v}_2 \in \{\mathbf{v}_j^{(n)}\}$. Let us make the following change of variable

$$\mathbf{v}'_i = \begin{cases} \mathbf{v}_i^{(s)} - \mathbf{v}_j^{(n)} & i = 1, \ldots, C_{L+1} - C_L \\ \mathbf{v}_i^{(s)} & i = C_{L+1} - C_L + 1, \ldots, C_{L+1} \end{cases} \tag{4.2}$$

where the first elements corresponds to eigenvectors with zero eigenvalues, and the last C_L elements are those of the basis $\{\mathbf{v}_j^{(s)}\}$. Then, the transfer matrix takes the simple form

$$\mathsf{T}' = \begin{pmatrix} 0 & \mathsf{T}^{(n)} \\ 0 & \mathsf{T}^{(s)} \end{pmatrix} \tag{4.3}$$

where $\mathsf{T}^{(s)}$ (resp. $T^{(n)}$) is a matrix of dimension C_L (resp. $C_{L+1} - C_L$). In this new basis the vectors $\mathbf{w}_{\rm id}$ and $\mathbf{u}$ take the form

$$\mathbf{w}'_{\rm id} = \left(0, \mathbf{w}_{\rm id}^{(s)}\right) \tag{4.4a}$$

$$\mathbf{u}' = \left(0, \mathbf{u}^{(s)}\right) \tag{4.4b}$$

Thus, we can write the partition function in terms of $\mathsf{T}^{(s)}$, $\mathbf{w}^{(s)}$ and $\mathbf{u}^{(s)}$ alone

$$Z(L_{\rm P} \times m_{\rm F}, (q,v) = \mathbf{w}_{\rm id}^{(s)\ {\rm T}} \cdot [\mathsf{T}^{(s)}]^{m-1} \cdot \mathbf{u}^{(s)} \tag{4.5}$$

So far we have a transfer matrix with the same dimension as for free boundary conditions (namely, $N_{Z,{\rm tri,PF},L} = C_L$). We can reduce even more the dimension of the transfer matrix by noting that cylindrical boundary conditions introduce an extra symmetry (i.e. translations along the transverse direction). In particular, we can make a further change of basis in the subspace $\{\mathbf{v}_j^{(s)}\}$ so that

$$\mathbf{v}_i^{(s)\prime} = \begin{cases} \mathbf{v}_i^{(s,n)} & i = 1, \ldots, C_L - N_{Z,{\rm tri,PF},L} \\ \mathbf{v}_i^{(s,t)} & i = C_L - N_{Z,{\rm tri,PF},L} + 1, \ldots, C_L \end{cases} \tag{4.6}$$

where the last $N_{Z,{\rm tri,PF},L}$ (resp. the first $C_L - N_{Z,{\rm tri,PF},L}$) elements $\{\mathbf{v}_i^{(s,t)}\}$ (resp. $\{\mathbf{v}_i^{(s,n)}\}$) are translational-invariant (resp. non-translational-invariant) combinations of the original vectors $\{\mathbf{v}_i^{(s)}\}$. In this new basis, the transfer matrix $\mathsf{T}^{(s)}$ takes a block diagonal form

$$\mathsf{T}^{(s)\prime} = \begin{pmatrix} \mathsf{T}^{(s,n)} & 0 \\ 0 & \mathsf{T}^{(s,t)} \end{pmatrix} \tag{4.7}$$

and the vectors $\mathbf{w}_{\rm id}^{(s)}$ and $\mathbf{u}^{(s)}$ take the form

$$\mathbf{w}_{\rm id}^{(s)\prime} = \left(0, \mathbf{w}_{\rm id}^{(s,t)}\right) \tag{4.8a}$$

$$\mathbf{u}(s)' = \left(0, \mathbf{u}^{(s,t)}\right) \tag{4.8b}$$

The partition function can be computed using the transfer matrix $\mathsf{T}^{(s,t)}$

$$Z(L_{\rm P} \times m_{\rm F}, (q,v) = \mathbf{w}_{\rm id}^{(s,t)\ {\rm T}} \cdot [\mathsf{T}^{(s,t)}]^{m-1} \cdot \mathbf{u}^{(s,t)} \tag{4.9}$$

The dimension of $\mathsf{T}^{(s,t)}$ will be denoted by $N_{Z,{\rm tri,PF},L}$ and is given by Theorem 2.2 (See also Table 5).

For $L \geq 6$ there is a further simplification, which is also present in the chromatic-polynomial case [30]: in the translation-invariant subspace the transfer matrix does

commute with the reflection operation. Thus, we can pass to a new basis consisting of connectivities that are either even or odd under reflection. In this new basis, the transfer matrix $\mathsf{T}^{(s,t)}$ has the block diagonal form

$$\mathsf{T}^{(s,t)\prime} = \begin{pmatrix} \mathsf{T}_{-}^{(s,t)} & 0 \\ 0 & \mathsf{T}_{+}^{(s,t)} \end{pmatrix} \tag{4.10}$$

and the vectors $\mathbf{w}_{\mathrm{id}}^{(s,t)}$ and $\mathbf{u}^{(s,t)}$ take the form

$$\mathbf{w}_{\mathrm{id}}^{(s,t)\prime} = \left(0, \mathbf{w}_{\mathrm{id},+}^{(s,t)}\right) \tag{4.11a}$$

$$\mathbf{u}(s,t)' = \left(0, \mathbf{u}_{+}^{(s,t)}\right) \tag{4.11b}$$

Thus, we can compute the partition function using the transfer matrix $\mathsf{T}^{(s,t)\prime}$

$$Z(L_{\mathrm{P}} \times m_{\mathrm{F}}, (q,v) = \mathbf{w}_{\mathrm{id},+}^{(s,t)\ \mathrm{T}} \cdot [\mathsf{T}_{+}^{(s,t)}]^{m-1} \cdot \mathbf{u}_{+}^{(s,t)} \tag{4.12}$$

In what follows, we shall drop the superindices $^{(st)}$ and the subindex $_+$ to simplify the notation. The dimension of the transfer matrix $\mathsf{T}_{+}^{(s,t)} \equiv \mathsf{T}$ corresponds to the number of partition classes of the numbers $\{1, 2, \ldots, L\}$ that are invariant under translations and reflections. Thus, this number is precisely $N_{Z,\mathrm{sq,PF},L}$ and is given in terms of L by eq. (2.52) (See Table 5 for some numerical values).

We have obtained the transfer matrices $\mathsf{T}(L_{\mathrm{P}})$ and the vectors $\mathbf{w}_{\mathrm{id}}(L_{\mathrm{P}})$ and $\mathbf{u}(L_{\mathrm{P}})$ using symbolic computation with MATHEMATICA for $L \leq 5_{\mathrm{P}}$. For $L \geq 6_{\mathrm{P}}$, we have used two different programs (one written in C and the other one in PERL), which were also used to double-check the results for $L \leq 5_{\mathrm{P}}$.

For $L \geq 5_{\mathrm{P}}$, the expressions for $\mathsf{T}(L_{\mathrm{P}})$ and the vectors $\mathbf{w}_{\mathrm{id}}(L_{\mathrm{P}})$ and $\mathbf{u}(L_{\mathrm{P}})$ can be found in the MATHEMATICA file `transfer_Tutte_tri.m` which is available with the electronic version of this paper in the `cond-mat` archive at `http://www.lanl.gov`.

4.1 $L = 2_{\mathrm{P}}$

The number of elements in the basis is two, with $\mathbf{P} = \{1, \delta_{1,2}\}$; the transfer matrices and the other relevant quantities are given by

$$\mathsf{T} = \begin{pmatrix} q^2 + 6qv + 12v^2 + qv^2 + 8v^3 + 2v^4 & D_1^2(q + 4v + 2v^2) \\ v^2(2q + 12v + 13v^2 + 6v^3 + v^4) & v^2 D_1^2 D_2^2 \end{pmatrix} \tag{4.13a}$$

$$\mathbf{w}_{\mathrm{id}}^{\mathrm{T}} = q\left(E_2, D_1^2\right) \tag{4.13b}$$

$$\mathbf{u}^{\mathrm{T}} = (1, 0) \tag{4.13c}$$

As before, we have used the shorthand notation introduced above (3.10a)/(3.10c).

4.2 $L = 3_{\mathrm{P}}$

The number of (translational-invariant) elements in the basis is three: $\mathbf{P} = \{1, \delta_{1,2} + \delta_{1,3} + \delta_{2,3}, \delta_{1,2,3}\}$. The transfer matrices and the vectors $\mathbf{w}_{\mathrm{id}}$ and $\mathbf{u}$ are

given by

$$\mathsf{T} = \begin{pmatrix} T_{11} & 3D_1T_{12} & D_1^3T_{13} \\ v^2T_{21} & v^2D_1T_{22} & v^2D_1^3D_2^2 \\ v^4D_3T_{31} & 3v^3D_1D_2T_{32} & v^3D_1^3D_2^3 \end{pmatrix} \tag{4.14a}$$

$$\mathbf{w}_{\rm id}^{\rm T} = q\left(V_1, 3D_1E_2, D_1^3\right) \tag{4.14b}$$

$$\mathbf{u}^{\rm T} = (1,0,0) \tag{4.14c}$$

where the factors D_k, E_k are given in (3.10a)/(3.10c), and

$$V_1 = q^2 + 3qv + 3v^2 + 2v^3 \tag{4.15a}$$

$$T_{11} = q^3 + 9q^2v + 33qv^2 + 4qv^3 + 50v^3 + 21v^4 + 3v^5 \tag{4.15b}$$

$$T_{12} = q^2 + 8qv + 20v^2 + 2qv^2 + 14v^3 + 3v^4 \tag{4.15c}$$

$$T_{13} = q + 6v + 3v^2 \tag{4.15d}$$

$$T_{21} = q^2 + 10qv + 30v^2 + 2qv^2 + 22v^3 + 7v^4 + v^5 \tag{4.15e}$$

$$T_{22} = 6q + 38v + 2qv + 42v^2 + 18v^3 + 3v^4 \tag{4.15f}$$

$$T_{31} = 3q + 18v + 15v^2 + 6v^3 + v^4 \tag{4.15g}$$

$$T_{32} = q + 12v + 13v^2 + 6v^3 + v^4 \tag{4.15h}$$

4.3 $L = 4_{\rm P}$

The number of (translational-invariant) elements in the basis is six: $\mathbf{P} = \{1, \delta_{1,2} + \delta_{2,3} + \delta_{3,4} + \delta_{1,4}, \delta_{1,3} + \delta_{2,4}, \delta_{1,2,3} + \delta_{1,2,4} + \delta_{1,3,4} + \delta_{2,3,4}, \delta_{1,2,3,4}, \delta_{1,4}\delta_{2,3} + \delta_{1,2}\delta_{3,4}\}$. The transfer matrix is given by

$$\mathsf{T} = \tag{4.16}$$
$$\begin{pmatrix} T_{11} & 4D_1T_{12} & 2E_4T_{13} & 4D_1^2T_{14} & D_1^4T_{15} & 2D_1^2T_{16} \\ v^2T_{21} & 2v^2D_1T_{22} & v^2T_{23} & v^2D_1^2T_{24} & v^2D_1^4D_2^2 & 2v^2D_1^2D_2T_{26} \\ v^3T_{31} & 2v^2D_1T_{32} & 4v^2T_{34}^2 & 4v^2D_1^2D_2T_{34} & v^2D_1^4D_2^2 & v^2D_1^2T_{36} \\ v^4D_3T_{41} & v^3D_1T_{42} & 2v^3T_{34}T_{54} & 2v^3D_1^2D_2T_{44} & v^3D_1^4D_2^3 & v^3D_1^2D_2T_{46} \\ v^6D_3^2T_{51} & 4v^5D_1D_2D_3T_{52} & 2v^4T_{54}^2 & 4v^4D_1^2D_2^2T_{54} & v^4D_1^4D_2^4 & 2v^5D_1^2D_2^2D_3T_{56} \\ v^4T_{62}^2 & 4v^4D_1D_2T_{62} & v^4T_{63} & 2v^4D_1^2D_2^2 & 0 & 2v^4D_1^2D_2^2 \end{pmatrix}$$

where the factors $D_k(v)$ and E_k are defined in (3.10a)/(3.10c); the T_{ij} are given by

$$T_{11} = q^4 + 12q^3v + 62q^2v^2 + 164qv^3 + 4q^2v^3 + 192v^4 + 29qv^4 + 72v^5 + 6v^6 \tag{4.17a}$$

$$T_{12} = q^3 + 11q^2v + 47qv^2 + q^2v^2 + 80v^3 + 11qv^3 + 40v^4 + 5v^5 \tag{4.17b}$$

$$T_{13} = q^2 + 8qv + 24v^2 + qv^2 + 16v^3 + 4v^4 \tag{4.17c}$$

$$T_{14} = q^2 + 10qv + 28v^2 + 3qv^2 + 20v^3 + 4v^4 \tag{4.17d}$$

$$T_{15} = q + 8v + 4v^2 \tag{4.17e}$$

$$T_{16} = q^2 + 10qv + 32v^2 + 3qv^2 + 24v^3 + 5v^4 \tag{4.17f}$$

$$
\begin{aligned}
T_{21} &= q^3 + 12q^2v + 54qv^2 + 2q^2v^2 + 96v^3 + 18qv^3 + 59v^4 + qv^4 \\
&\quad +12v^5 + v^6 \qquad (4.17g) \\
T_{22} &= 3q^2 + 28qv + q^2v + 80v^2 + 13qv^2 + 65v^3 + qv^3 + 18v^4 + 2v^5 \qquad (4.17h) \\
T_{23} &= 3q^2 + 34qv + 96v^2 + 23qv^2 + 120v^3 + 8qv^3 + 67v^4 + qv^4 \\
&\quad +18v^5 + 2v^6 \qquad (4.17i) \\
T_{24} &= 10q + 56v + 6qv + 63v^2 + qv^2 + 26v^3 + 4v^4 \qquad (4.17j) \\
T_{26} &= q + 8v + 5v^2 + v^3 \qquad (4.17k) \\
T_{31} &= 2q^2 + 18qv + 48v^2 + 4qv^2 + 31v^3 + 8v^4 + v^5 \qquad (4.17l) \\
T_{32} &= q^2 + 14qv + 52v^2 + 4qv^2 + 43v^3 + 14v^4 + 2v^5 \qquad (4.17m) \\
T_{34} &= q + 6v + 4v^2 + v^3 \qquad (4.17n) \\
T_{36} &= 2q + 24v + 27v^2 + 12v^3 + 2v^4 \qquad (4.17o) \\
T_{41} &= q^2 + 11qv + 36v^2 + 2qv^2 + 24v^3 + 7v^4 + v^5 \qquad (4.17p) \\
T_{42} &= 2q^2 + 43qv + q^2v + 216v^2 + 28qv^2 + 270v^3 + 4qv^3 + 138v^4 \\
&\quad +36v^5 + 4v^6 \qquad (4.17q) \\
T_{44} &= 3q + 24v + qv + 27v^2 + 12v^3 + 2v^4 \qquad (4.17r) \\
T_{46} &= q + 24v + 27v^2 + 12v^3 + 2v^4 \qquad (4.17s) \\
T_{51} &= 4q + 24v + 17v^2 + 6v^3 + v^4 \qquad (4.17t) \\
T_{52} &= 2q + 18v + 15v^2 + 6v^3 + v^4 \qquad (4.17u) \\
T_{54} &= q + 12v + 13v^2 + 6v^3 + v^4 \qquad (4.17v) \\
T_{56} &= 4 + 3v + v^2 \qquad (4.17w) \\
T_{62} &= q + 6v + 2v^2 \qquad (4.17x) \\
T_{63} &= 3q + 24v + 26v^2 + 12v^3 + 2v^4 \qquad (4.17y)
\end{aligned}
$$

The vectors $\mathbf{w}_{\rm id}$ and $\mathbf{u}$ are given by

$$
\begin{aligned}
\mathbf{w}_{\rm id}^{\rm T} &= q\left(V_1, 4D_1V_2, 2E_2^2, 4D_1^2E_2, D_1^4, 2D_1^2E_2\right) \qquad (4.18a) \\
\mathbf{u}^{\rm T} &= (1,0,0,0,0,0) \qquad (4.18b)
\end{aligned}
$$

where

$$
\begin{aligned}
V_1 &= q^3 + 4q^2v + 6qv^2 + 4v^3 + v^4 \qquad (4.19a) \\
V_2 &= q^2 + 3qv + 3v^2 + v^3 \qquad (4.19b)
\end{aligned}
$$

5 Partition Function Zeros in the q Plane

In this section we shall present results for zeros and continuous accumulation sets $\mathcal{B}_q$ (in the q-plane) for the partition function of the Potts antiferromagnet on triangular-lattice strips of widths $L \le 5$ with free or cylindrical boundary conditions.

In Figure 1 we show the partition-function zeros in the q-plane for strips of sizes $L_{\rm F} \times (10L)_{\rm F}$ with $2 \le L \le 5$ and free boundary conditions. We also show the corresponding limiting curves $\mathcal{B}_q(L,v)$ for the infinite-length limit. Figure 1(a) displays the

zeros for $v = -1$ (i.e., the chromatic zeros [48, 51, 30]), and Figures 1(b)–(d) display the corresponding zeros for the non-zero temperatures $v = -0.75$ (b), $v = -0.5$ (c), and $v = -0.25$ (d). The case $L = 2$ was studied in [18] and the cases $L = 3, 4$ were studied in (Figs. 3.15, 3.22 of) [16]) for arbitrary temperature, and our zeros are in agreement with these earlier works.

The corresponding partition-function zeros and accumulation sets for triangular-lattice strips with cylindrical boundary conditions are shown in Figure 2. Again, Figure 2(a)–(d) displays the zeros for $v = -1$ (a), $v = -0.75$ (b), $v = -0.5$ (c), and $v = -0.25$ (d). The cases $L = 3, 4$ were studied previously in (Figs. 3.19, 3.25 of) [16] for arbitrary temperature, and our zeros are in agreement with this work.

The case $v = -1$, which is the zero-temperature Potts antiferromagnet (chromatic polynomial) has been previously studied in [48, 50, 51, 19, 30] for the free longitudinal boundary conditions and in [52, 53, 51, 54, 20] for periodic longitudinal boundary conditions. For the case of free longitudinal boundary conditions [30] contains results for $L \le 9_{\rm F}$ and $L \le 12_{\rm P}$. Our Figures 1(a) and 2(a) include calculations up to $L = 5$ for comparison with other values of v. Although some curves, such as those for cylindrical boundary conditions, may enclose regions, the curves do not enclose regions containing the origin. One observes that for either type of transverse boundary condition, as the width L increases, the left-hand arc endpoints move slowly toward the origin. When this was observed in earlier work for several different lattice strips, [48], it motivated the suggestion that in the limit $L \to \infty$ for strips with free longitudinal boundary conditions, the limiting $\mathcal{B}_q$ would separate the q plane into regions including a curve passing through $q = 0$ [48, 15]. The specific calculation of $\mathcal{B}_q(v = -1)$ in the limit $L \to \infty$ reported by Baxter [46] has this feature. (For critical comments on certain features of Baxter's results, see the detailed discussion in [30].) The property that $\mathcal{B}_q$ separates the q plane into regions with one of the curves on $\mathcal{B}_q$ passing through the origin is also observed for lattice strips with finite width L if one imposes periodic longitudinal boundary conditions [55, 56, 52, 53, 57, 58, 51, 54, 59].

In making inferences about possible $L \to \infty$ characteristics of the continuous accumulation set of zeros $\mathcal{B}(G_s, L \times \infty)$ in the q or v plane for infinite-length, width-L lattice strip graphs of type G_s, one should recall that, in general, $\lim_{L\to\infty} \mathcal{B}(G_s, L \times \infty)$ is different from the continuous accumulation set of the zeros of the partition function for the usual 2D thermodynamic limit defined by starting with an $L_y \times L_x$ section of a regular lattice and letting L_x and L_y both approach infinity with L_y/L_x a finite nonzero constant. This type of noncommutativity was encountered in previous studies of $\mathcal{B}$ for the Potts model free energy on infinite-length, finite-width strips with periodic longitudinal boundary conditions [17, 18, 23, 22]; for these strips, $\mathcal{B}$ is noncompact in the v plane, reflecting the fact that the Potts model has a ferromagnetic critical point only at $T = 0$ (i.e. $K = \infty$, hence $v = \infty$) for any width L, no matter how great, whereas for the 2D lattice defined in the thermodynamic limit, it has a ferromagnetic critical point at a finite temperature, so $\mathcal{B}$ is compact in the v plane. Noncommutativity of this type was also found in studies of $\mathcal{B}_q$. For example, in calculations of $\mathcal{B}_q$ for infinite-length strips of the triangular lattice with cyclic boundary conditions, it was found that this locus always passes through $q = 2$ [52, 53, 51, 54], whereas, in contrast, the locus found in [46] for the infinite-

width limit of strips with cylindrical boundary conditions does not pass through $q = 2$. Similarly, in calculations of $\mathcal{B}_q$ for infinite-length strips of the square lattice with cyclic boundary conditions, it was found that this locus always passes through $q = 2$, whereas in calculations of $\mathcal{B}_q$ for infinite-length strips of the square lattice with cylindrical boundary conditions [50, 19, 59, 42], it was found that $\mathcal{B}$ does not pass through $q = 2$, strongly suggesting that this difference will persist in the limit $L_y \to \infty$.

At nonzero temperature for the antiferromagnet, as represented in our figures by the range $-0.75 \le v \le -0.25$, the partition-function zeros in the q-plane have a different shape: as L increases, the limiting curves $\mathcal{B}_q(L, v)$ tend to a bean-shaped curve or set of arcs, open on the left, without substantial protruding branches, in contrast to many of the $v = -1$ curves. For a given value of v in the range considered, as the width L increases, the curve envelope moves outward somewhat and the arc endpoints on the left move slowly toward $q = 0$. This behavior is consistent with the hypotheses that for a given v, as $L \to \infty$, (i) $\mathcal{B}_q$ would approach a limiting locus as $L \to \infty$ and (ii) this locus would separate the q plane into different regions, with a curve passing through $q = 0$ as well as a maximal real value, $q_c(v)$. This is qualitatively the same type of behavior that was found earlier for the square-lattice strips [17, 23, 3]. In particular, our results are consistent with the inference that as $L \to \infty$, $\mathcal{B}_q$ for $v = -1$ would pass through $q_c(v = -1) = 4$, corresponding to the property that the $q = 4$ Potts antiferromagnet has a zero-temperature critical point on the (infinite) triangular lattice [60].

For a given L, as v increases from -1 to 0, i.e., as the inverse temperature β decreases from infinity to 0 for the antiferromagnet, the zeros and the limiting curve contract to a point at $q = 0$. This is an elementary consequence of the fact that these lattice strips have fixed maximal vertex degree and as the parameter $K = \beta J$ approaches zero, the spin-spin interaction term in $\mathcal{H}$, eq. (1.2), vanishes, so that the sum over states just counts all q possible spin states independently at each vertex, and $Z(G, q, v)$ approaches the value $Z(G, q, 0) = q^n$. An upper bound on the magnitudes of the zeros is given by the following theorem by Sokal:

Theorem 5.1 *(Sokal [61]) Let $G = \{V, E\}$ be a loopless finite undirected graph of maximum degree $\le r$, equipped with complex edge weights $\{v_e\}_{e \in E}$ satisfying $|1 + v_e| \le 1$ for all e. Let $|v|_{\max} = \max_{e \in E} |v_e|$. Then all the zeros of $Z(G, q, \{v_e\})$ lie in the disc $|q| < C(r) v_{\max}$ with $C(r) \le 7.963907r$.*

This is a loose bound; for all of the strips with cylindrical boundary conditions and for the free strips with widths $L \ge 3$, the maximal degree is $r = 6$, so that the above theorem implies that $|q| \lesssim 47.8 |v|_{\max}$. Thus, for example, for $v = -1$, this reads $|q| \lesssim 47.8$, whereas in fact $|q| \lesssim 4$ for free boundary conditions and $|q| \lesssim 4.5$ for cylindrical boundary conditions. A general feature is that the limiting curves and associated zeros tend to be located mostly in the $\mathrm{Re}(q) \ge$ half plane.

As noted, it is evident in Figures 1 and 2 that as L increases, the accumulation set $\mathcal{B}_q(L, v)$ moves outwards. As expected, the convergence to the limit $L \to \infty$ seems to be faster with cylindrical boundary conditions, as there are no surface effects when the length is made infinite.

One can also plot $\mathcal{B}_q$ for the ferromagnetic region $0 \le v \le \infty$ (e.g., Figs. 2 and 13 of [18]). Although we have not included these plots here, we note that an elementary Peierls argument shows that the q-state Potts ferromagnet on infinite-length, finite-width strips has no finite-temperature phase transition and associated magnetic long range order. Hence, for this model $\mathcal{B}_q$ does not cross the positive real q axis for $0 \le v < \infty$.

6 Partition Function Zeros in the v Plane

6.1 General

In this section we shall present results for zeros and continuous accumulation sets $\mathcal{B}_v$ (in the v-plane) for the partition function of the Potts antiferromagnet on triangular-lattice strips of widths $L \le 5$ and free or cylindrical boundary conditions. Our results hold for arbitrarily great length L and for any real or complex value of q; they thus complement calculations of the Potts model partition function for fixed positive integer values of q on sections of the triangular lattice [62, 63, 64, 65]. We shall focus here on integer values of q, since these are the most relevant from a physical point of view. We recall the possible noncommutativity in the definition of the free energy for certain integer values of q (see eqs. (2.10), (2.11) of [17] or (1.17) of [18]):

$$\lim_{n\to\infty} \lim_{q\to q_s} Z(G,q,v)^{1/n} \ne \lim_{q\to q_s} \lim_{n\to\infty} Z(G,q,v)^{1/n} \tag{6.1}$$

As discussed in [17], because of this noncommutativity, the formal definition (1.4) is, in general, insufficient to define the free energy f at these special points q_s; it is necessary to specify the order of the limits that one uses in eq. (6.1). We denote the two definitions using different orders of limits as f_{qn} and f_{nq}:

$$f_{nq}(\{G\},q,v) = \lim_{n\to\infty} \lim_{q\to q_s} n^{-1} \ln Z(G,q,v) \tag{6.2}$$

$$f_{qn}(\{G\},q,v) = \lim_{q\to q_s} \lim_{n\to\infty} n^{-1} \ln Z(G,q,v) \ . \tag{6.3}$$

As a consequence of the noncommutativity (6.1), it follows that for the special set of points $q = q_s$ one must distinguish between (i) $(\mathcal{B}_v(\{G\},q_s))_{nq}$, the continuous accumulation set of the zeros of $Z(G,q,v)$ obtained by first setting $q = q_s$ and then taking $n \to \infty$, and (ii) $(\mathcal{B}_v(\{G\},q_s))_{qn}$, the continuous accumulation set of the zeros of $Z(G,q,v)$ obtained by first taking $n \to \infty$, and then taking $q \to q_s$. For these special points (cf. eq. (2.12) of [17]),

$$(\mathcal{B}_v(\{G\},q_s))_{nq} \ne (\mathcal{B}_v(\{G\},q_s))_{qn} \tag{6.4}$$

Here this noncommutativity will be relevant for $q = 0$ and $q = 1$.

In Figures 3–7 we show the partition-functions zeros in the v-plane (for a fixed value of q) for strips of sizes $L_{\rm F} \times (10L)_{\rm F}$ with $2 \le L \le 5$ and free boundary conditions. We also show the corresponding limiting curves $\mathcal{B}_v(L,q)$ for the limit of infinite strip

length. For simplicity, we have displayed each value of L on a different plot: $L = 2$ (a), $L = 3$ (b), $L = 4$ (c), and $L = 5$ (d). The corresponding partition-function zeros and accumulation sets for triangular-lattice strips with cylindrical boundary conditions are shown in Figures 8–12 with the same notation as for the former figures. Complex-temperature phase diagrams and associated partition function zeros were given in [18] for $L = 2$ for free and periodic longitudinal boundary conditions and free transverse boundary conditions. Results for $L = 4_{\rm F}$ and $L = 4_{\rm P}$ at zero and finite temperatures were given before in [16]. Our present calculations are in agreement with, and extend, this previous work.

On the infinite triangular lattice (defined via the 2D thermodynamic limit as given above), the phase transition point separating the paramagnetic (PM) and ferromagnetic (FM) phases is determined as the (unique) real positive solution of the equation [67]

$$v^3 + 3v^2 - q = 0 \tag{6.5}$$

In previous studies such as [17, 18], it has been found that although infinite-length, finite-width strips are quasi-one-dimensional systems and hence the Potts model has no physical finite-temperature transition on such systems, some aspects of the complex-temperature phase diagram have close connections with those on the (infinite) triangular lattice. We shall discuss some of these connections below.

6.2 $q = 0$

From the cluster representation of $Z(G, q, v)$, eq. (1.5), it follows that this partition function has an overall factor of $q^{k(G)}$, where $k(G)$ denotes the number of components of G, i.e., an overall factor of q for a connected graph. Hence, $Z(G, q = 0, v) = 0$. In the transfer matrix formalism, this is evident from the overall factor of q coming from the vector $\mathbf{w}_{\rm id}$. However, if we first take the limit $n \to \infty$ to define $\mathcal{B}$ for $q \neq 0$ and then let $q \to 0$ or, equivalently, extract the factor q from the left vector $\mathbf{w}_{\rm id}$, we obtain a nontrivial locus, namely $(\mathcal{B}_v(\{G\}, 0)_{qn}$. This is a consequence of the noncommutativity (6.4) for $q = 0$.

With the second order of limits or the equivalent removal of the factor of q in Z, we obtain the locus $\mathcal{B}_v(q = 0)$ shown in Figures 3 (free boundary conditions) and 8 (cylindrical boundary conditions). The accumulation set $\mathcal{B}_v(L, q = 0)$ seems to converge to a roughly circular curve. We see in Figures 3 and 8 that the limiting curves cross the real v-axis at $v \approx -3$. We note the interesting feature that this is a root of eq. (6.5) for $q = 0$. For the case of cylindrical boundary conditions, $\mathcal{B}_v(q = 0)$ includes a small line segment on the real axis near $v = -3$, with a length that decreases as L increases. As L increases, the arc endpoints on the upper and lower right move toward the real axis. It is possible that these could pinch this axis at $v = 0$ as $L \to \infty$, corresponding to the other root of (6.5) for $q = 0$.

6.3 $q = 1$

For $q = 1$, the spin-spin interaction in (1.2) always has the Kronecker delta function equal to unity, and hence the Potts model partition function is trivially given

by

$$Z(G, q = 1, v) = e^{K|E|} = (1 + v)^{|E|} \tag{6.6}$$

where $|E|$ is the number of edges in the graph G. This has a single zero at $v = -1$. But again, one encounters the noncommutativity (6.4) for $q = 1$. It is interesting to analyze this in terms of the transfer matrix formalism. At this value of q, both the transfer matrix and the left vector $\mathbf{w}_{\rm id}$ are non-trivial. There is thus a cancellation of terms that yields the result (6.6). The case $L = 2_{\rm P}$ is the simplest one to analyze: the eigenvalues and coefficients for $q = 1$ are given by

$$\lambda_1(1, v) = 2v^2 \; ; \qquad c_1(1, 0) = 0 \tag{6.7a}$$

$$\lambda_2(1, v) = (1 + v)^6 \; ; \quad c_2(1, 0) = (1 + v)^2 \tag{6.7b}$$

Thus, only the second eigenvalue contributes to the partition function, and it gives the expected result $Z(2_{\rm P} \times m_{\rm F}, q = 1, v) = (1 + v)^{6m-4}$. For $L = 2_{\rm F}$ we obtain:

$$\lambda_1(1, v) = v^2 \; ; \qquad c_1(1, 0) = 0 \tag{6.8a}$$

$$\lambda_2(1, v) = (1 + v)^4 \; ; \quad c_2(1, 0) = (1 + v) \tag{6.8b}$$

giving rise to $Z(2_{\rm F} \times m_{\rm F}, q = 1, v) = (1 + v)^{4m-3}$. The case $L = 3_{\rm P}$ is similar: there is a single eigenvalue $\lambda_1(1, v) = (1 + v)^9$ with a non-zero coefficient $c_1(1, v) = (1 + v)^3$, and the other two eigenvalues [which are the roots of $x^2 - xv^2(3v^3 + 13v^2 + 24v + 3) + v^5(1 + v)(6 + 5v + 6v^2)$] have identically zero coefficients $c_{2,3}(1, v) = 0$. Thus, the partition function takes the form $Z(3_{\rm P} \times m_{\rm F}, q = 1, v) = (1 + v)^{9m-6} = (1 + v)^{|E|}$.

In general, we conclude that at $q = 1$ only the eigenvalue $\lambda = (1 + v)^{3L_{\rm P}}$ (resp. $\lambda = (1 + v)^{3L_{\rm F}-2}$) contributes to the partition function for cylindrical (resp. free) boundary conditions, and its coefficient is $c = (1 + v)^{L_{\rm P}}$ (resp. $c = (1 + v)^{L_{\rm F}-1}$). The other eigenvalues do not contribute as they have zero coefficients. This is the analogue of what was found for the strips with cyclic boundary conditions, where the various λ's fall into sets $\lambda_{G,d,j}$ such that all of the λ's with a fixed d have a unique coefficient which is a polynomial of degree d in q given by [17, 18, 20]

$$c^{(d)} = U_{2d}\Big(\frac{\sqrt{q}}{2}\Big) = \sum_{j=0}^{d} (-1)^j \binom{2d - j}{j} q^{d-j} \tag{6.9}$$

where $U_\nu(z)$ is the Chebyshev polynomial of the second kind. These coefficients vanish at certain values of q, which means that if one evaluates the partition function first at these values and then takes the limit $n \to \infty$, the corresponding λ's will not contribute to Z, while if one takes $n \to \infty$ first, calculates the free energy and the locus $\mathcal{B}_{qn}$, and then sets q equal to one of these values, the λ's will, in general contribute. In particular, we recall (eq. (2.18) of [20]) that if $q = 1$, then $c^{(d)}$ vanishes if $d = 1 \bmod 3$. Thus, we see similar manifestations of the noncommutativity (6.4) for strips with free and periodic boundary conditions.

In our present case, in order to obtain $\mathcal{B}_{qn}$, we have computed the Tutte-polynomial zeros and the corresponding limiting curves for $q = 0.999$ (see Figures 4 and 9). The accumulation sets $\mathcal{B}_v(L, q = 0.999)$ for $L = 2_{\rm F}$ to $L = 5_{\rm F}$ consists of arcs that come

close to forming an almost closed bean-shaped curve, with an involution on the right that deepens as L increases. Solving the $q = 1$ special case of eq. (6.5) yields the roots $v = -2.879385..$, $v = -0.652703...$, and $v = 0.5320888...$ The locus $\mathcal{B}_v$ crosses the real v axis at two points, and our results are consistent with the inference that as $L \to \infty$, these two crossing points are the first two roots listed above. For free boundary conditions with width $L = 3, 4, 5$, $\mathcal{B}_{qn}$ exhibits a small involution on the left. A noteworthy feature of this locus is that it is relatively smooth, without the prongs that tend to occur for the other values of q discussed here.

6.4 $q = 2$

The zeros and accumulation sets for $q = 2$ are displayed in Figures 5 and 10 for free and cylindrical boundary conditions. Fig. 5(a) contains the same information as Fig. 4 of [18] (which is plotted in a different temperature variable, a^{-1}). The finite-size effects for the accumulation sets $\mathcal{B}_v(L, q = 2)$ are noticeably larger for free, in comparison with cylindrical, boundary conditions, as expected. In the latter case, the curves $\mathcal{B}_v(L, q = 2)$ for $L = 2, 3, 4, 5$ fall very approximately one on top of the preceding one. For cylindrical boundary conditions, we see that the curve $\mathcal{B}_v(L, q = 2)$ is symmetric under the replacement $a \to -a$. The reason for this is that in this case all of the vertices except for the end-vertices, which constitute a vanishingly small fraction in the limit of infinite length, are equivalent (i.e., the graph is r-regular) and have even degree r. In general, this property applies for the complex-temperature phase diagram of the $q = 2$ (Ising) special case of the Potts model for an infinite lattice where the coordination number is even [68, 69, 70]. Our strips with free transverse boundary conditions are not r-regular graphs because the vertices on the upper and lower sides have a different degree than those in the interior. Because of this, the $\mathcal{B}_v$ in this case does not have the $a \to -a$ symmetry. From previous work [17, 18] one knows that the loci $\mathcal{B}_v$ are different for strips with free or periodic transverse boundary conditions and free longitudinal boundary conditions, on the one hand, and free or periodic transverse boundary conditions and periodic (or twisted periodic) longitudinal boundary conditions. One anticipates, however, that in the limit of infinite width, the subset of the complex-temperature phase diagram that is relevant to real physical thermodynamics will be independent of the boundary conditions used to obtain the 2D thermodynamic limit.

In the 2D thermodynamic limit, one knows the complex-temperature phase diagram exactly for the $q = 2$ (Ising) case. (This isomorphism involves the redefinition of the spin-spin exchange constant $J_{\rm Potts} = 2J_{\rm Ising}$ and hence $K_{\rm Potts} = 2K_{\rm Ising}$, where $K_{\rm Potts}$ is denoted simply K here.) The simplest way to portray the complex-temperature phase diagram is in the a^2 or $u = a^{-2}$ plane since this automatically incorporates the $a \to -a$ symmetry noted above. In the u plane, the complex-temperature phase diagram, with boundaries given by $\mathcal{B}_u$, is (see Fig. 1(a) of [70] which is equivalent, by duality to the complex-temperature phase diagram for the honeycomb lattice given as Fig. 2 in [72])

$$\mathcal{B}_u : \quad \{|u + \frac{1}{3}| = \frac{2}{3}\} \quad \cup \quad \{-\infty \le u \le -\frac{1}{3}\} \tag{6.10}$$

i.e., the union of a circle centered at $u = -1/3$ with radius $2/3$ and the semi-infinite line segment extending leftward from $u = -1/3$ along the real u axis. In the a^{-1} plane (Fig. 1(b) of [70], related by duality to Fig. 3 of [72]), $\mathcal{B}$ is the union of a vertically elongated oval crossing the real axis at $\pm 1/\sqrt{3}$, the imaginary a^{-1} axis at $\pm i$, and two semi-infinite line segments extending from $i/\sqrt{3}$ to $i\infty$ and from $-i/\sqrt{3}$ to $-i\infty$ along the imaginary axis. Equivalently, in the a plane, $\mathcal{B}$ is the union of a horizontally elongated oval crossing the real a axis at $\pm\sqrt{3}$ and the imaginary a axis at $\pm i$, and a line segment along the imaginary axis extending between $\sqrt{3}i$ and $-\sqrt{3}i$. The locus $\mathcal{B}_v$ in the v plane is obtained from this by translation by one unit, since $v = a - 1$. This locus separates the complex v plane into three phases: (i) the paramagnetic phase, including the infinite-temperature point $v = 0$, where the S_q symmetry is realized explicitly (S_q being the symmetric group on q numbers, the symmetry group of the Hamiltonian), (ii) the ferromagnetic phase, including the real interval $v_c(q = 2) \le v \le \infty$ where the S_q symmetry is spontaneously broken by the existence of a nonzero magnetization, and (iii) an unphysical phase (denoted "O" for "other" in [70]) including the point $v = -2$. Here

$$a_c(q = 2) = v_c(q = 2) + 1 = \sqrt{3} \tag{6.11}$$

is the physical critical point separating the PM and FM phases (for a review of the Ising model on the triangular lattice, see, e.g., [73]) . These physical PM and FM phases have complex-temperature extensions off the real v axis. The PM and O phases are separated by the subset of the vertical line segment extending between $a = i$ and $a = -i$; this line segment terminates at the points $a = \pm\sqrt{3}i$. Because of the maximal frustration, there is no antiferromagnetic phase at finite temperature. The presence of a zero-temperature critical point in the 2D Ising antiferromagnet [74] is manifested by the fact that $\mathcal{B}_v$ passes through $v = -1$, i.e., $a = 0$ (as part of the above-mentioned vertical line segment). The complex-temperature phase boundary $\mathcal{B}_v$ crosses the real v axis at $v = \sqrt{3} - 1$, separating the FM and PM phases, at $v = -1$, separating the PM and O phases, and at $v = -1 - \sqrt{3}$, separating the O and (complex-temperature analytic continuation of the) FM phases. In [18] the $\mathcal{B}$ for an infinite-length free or cyclic strip with width $L = 2$ were compared with this 2D phase diagram. These three points, $v = -1, -1 \pm \sqrt{3}$, are the three roots of the $q = 2$ special case of eq. (6.5).

Using our exact results, we can compare our loci $\mathcal{B}_v$ for a wide variety of widths and either free or periodic transverse boundary conditions with the known complex-temperature phase diagram for the Ising model on the infinite 2D triangular lattice. This comparison is simplest for the case of cylindrical boundary conditions, so we concentrate on these results. For the finite values of L that we have considered, $\mathcal{B}_v$ has the form of two complex-conjugate arcs that cross two complex-conjugate line segments on the imaginary axis at $v = -1 \pm i$. One sees that as L increases, the endpoints of the arcs move down toward the real axis, as do the endpoints of the line segments. As $L \to \infty$, we expect that these arc endpoints will close, forming the above-mentioned horizontally elongated oval and vertical line segment extending from $v = -1 + \sqrt{3}i$ to $v = -1 - \sqrt{3}i$ that constitute the complex-temperature phase boundaries $\mathcal{B}$ for the Ising model on the infinite triangular lattice.

6.5 $q = 3$

In contrast to the $q = 2$ case, the free energy of the q-state Potts model has not been calculated exactly for $q \geq 3$ on any 2D (or higher-dimensional) lattice and hence the complex-temperature phase diagrams are not known exactly. The $q = 3$ special case of eq. (6.5) has the root

$$v_{PM-FM,q=3} \equiv v_c(q=3) = -1 + \cos(2\pi/9) + \sqrt{3}\sin(2\pi/9) = 0.879385... \quad (6.12)$$

corresponding to the physical PM-FM phase transition point, and two other roots at the complex-temperature values

$$v = -1 + \cos(2\pi/9) - \sqrt{3}\sin(2\pi/9) = -1.347296... \quad (6.13)$$

and

$$v = -1 - 2\cos(2\pi/9) = -2.532089... \quad (6.14)$$

Discussions of the complex-temperature solutions of eq. (6.5) and their connections with the complex-temperature phase diagram have been given in [62, 63, 64, 65]. A number of studies involving exact calculation of the partition function for various q values on large sections of the triangular lattice have been performed [62, 63, 64, 65]. (There have also been many studies calculating zeros for the Potts model with $q \geq 3$ on the square lattice; see [3] for references to these works.)

The zeros and accumulation sets for $q = 3$ are displayed in Figures 6 and 11 for free and cylindrical boundary conditions. We expect that the pair of complex-conjugate endpoints in this regime will eventually converge to the ferromagnetic critical point $v_c(q = 3)$ as $L \to \infty$. However, obviously, an infinite-length strip of finite width L is a quasi-one-dimensional system, so the Potts model has no physical finite-temperature phase transition on such a strip for any finite L.

In the antiferromagnetic regime $-1 \leq v < 0$, we observe noticeable finite-size effects even with cylindrical boundary conditions. In this regime, we also observe a complex-conjugate pair of endpoints with small value of $\mathrm{Im}(v)$ that, as $L \to \infty$, are expected to approach the real v axis at the transition point separating the paramagnetic and antiferromagnetic (AFM) phases of the $q = 3$ Potts antiferromagnet on the infinite triangular lattice. Monte Carlo and series analyses [75, 76, 77, 78] have yielded the conclusion that the PM-AFM transition in the $q = 3$ Potts antiferromagnet on the triangular lattice is weakly first-order. A high-accuracy determination of the location of the PM-AFM transition temperature T was obtained in [78] by means of Monte Carlo simulations: $T = 0.62731 \pm 0.00006$, or equivalently

$$v_{PM-AFM,q=3} = -0.79691 \pm 0.00003 \quad (6.15)$$

We shall improve this estimate below.

Finally, in the complex-temperature interval $v < -1$, the finite-size and boundary condition effects are evidently very strong. Because of this, in previous work, a combination of partition-function zeros and analyses of low-temperature series expansions was used [65]; these enable one at least to locate some points on the complex-temperature phase boundary. As regards the infinite 2D triangular lattice, because

of a duality relation, the complete physical temperature interval $0 \le T \le \infty$, i.e., $0 \le a \le 1$ of the q-state Potts antiferromagnet on the honeycomb lattice is mapped to the complex-temperature interval $-\infty \le v \le -q$) on the triangular lattice (and vice versa) [63]. As was noted in [63], it follows that because the $q = 3$ Potts antiferromagnet on the honeycomb lattice is disordered for all temperatures, including $T = 0$, the free energy for this model on the triangular lattice is analytic in the interval $-\infty < v \le -3$, and hence no part of the complex-temperature phase boundary $\mathcal{B}_v$ can cross the negative real axis in this interval. In particular, one anticipates that as $L \to \infty$ for the infinite-length, width-L strips, the left-most arcs on $\mathcal{B}_v$ will not close and pinch the negative real axis in this interval. Our calculations of $\mathcal{B}_v$ are consistent with the inference that as $L \to \infty$, this locus crosses the real axis at the points (6.13) and (6.14), although there are significant differences between the loci obtained with free and cylindrical boundary conditions.

We also observe certain line segments on the real v axis in the complex-temperature region. We note that massless phases with algebraic decays of correlation functions have been suggested for the Potts model on the (infinite) square lattice at real values of v and q in the intervals $-2 - \sqrt{4-q} \le v \le -2 + \sqrt{4-q}$ with $q \in (0,4)$ and $q \ne B_r = 4\cos^2(\pi/r)$ [66]. However, the correspondence of these suggestions with our results is not clear; for example, the above interval suggested in [66] shrinks to zero as $q \to 4$, but we observe clear line segments on the real v axis for $q = 4$ for both free and cylindrical boundary conditions (see Figs. 7 and 12). A possible physical subset of the above range of v given in [66] would be the antiferromagnet interval $-1 \le v < 0$. However, the condition that $q \ne B_r$ excludes all of the integral values of q in the indicated range (recall that $B_2 = 0$, $B_3 = 1$, $B_4 = 2$, $B_6 = 3$, and $B_1 = 4$). The claim in [66] is thus complicated by the fact that although it is possible formally to define the Potts model partition function $Z(G,q,v)$ using (1.5) for real positive non-integral q for the antiferromagnetic case, $-1 \le v < 0$, here the model does not satisfy the usual statistical mechanical requirement that the partition function is positive, and hence does not, in general, admit a Gibbs measure [71, 17]. This leads to pathologies that preclude a physical interpretation, such as negative partition function, negative specific heat, and non-existence of a $|V| \to \infty$ limit for thermodynamic functions that is independent of boundary conditions [71, 17, 18]. As regards the connection with the locus $\mathcal{B}$, a signal of a massless phase would be a line segment on $\mathcal{B}$ on the real v axis for fixed q or the real q axis for fixed v. For the zero-temperature Potts antiferromagnet, i.e., chromatic polynomial, $v = -1$, these phases would thus occur in the intervals between the Beraha numbers, $0 < q < 1$, $1 < q < 2$, $2 < q < (1/2)(3+\sqrt{5})$, and so forth. However, it has been proved that there are no real zeros of a chromatic polynomial in the intervals $-\infty < q < 0$, $0 < q < 1$, and $1 < q \le 32/27$ [79, 80]. Since $\mathcal{B}$ forms as an accumulation set of zeros, this makes it difficult to see how there could be a line segments in these intervals, in particular, the intervals $0 < q < 1$ and $1 < q \le 32/27$. Again, it is not clear how to reconcile the results of these theorems with a suggestion that there would be massless phases with associated real line segments on $\mathcal{B}$ in these intervals.

6.6 $q = 4$

For $q = 4$, eq. (6.5) has the physical root

$$v_{PM-FM,q=4} \equiv v_c(q = 4) = 1 \tag{6.16}$$

corresponding to the PM-FM phase transition point and a double root at the complex-temperature point

$$v = -2 \tag{6.17}$$

The zeros and accumulation sets for $q = 4$ are displayed in Figures 7 and 12. One observes the approach of the right-most complex-conjugate arcs to the real axis as L increases, i.e. the approach to the PM-FM critical point in this case. For a given L, the approach to the exactly known value $v_{PM-FM,q=4} = 1$ in eq. (6.16) is closer for cylindrical versus free boundary conditions, as is anticipated since the former minimize boundary effects. The $q = 4$ Potts antiferromagnet on the triangular lattice has a zero-temperature critical point, so that $v = -1$ is on $\mathcal{B}_v$ [60] (this is not a root of eq. (6.5)). For $L \ge 2$ for free boundary conditions and for $L \ge 2$ for cylindrical boundary conditions, we see how a pair of complex-conjugate arc endpoints approaches the real axis as L increases, consistent with the inference that these would pinch at $v = -1$ in the $L \to \infty$ limit. For both free and cylindrical boundary conditions and various values of L, one sees that $\mathcal{B}_v$ contains an intersection point at the complex-temperature value $v = -2$, in agreement with the expectation from eq. (6.17). The fact that the $q = 4$ Potts antiferromagnet is disordered on the honeycomb lattice for all temperatures T including $T = 0$ implies that $\mathcal{B}_v$ does not cross the negative real axis in the interval $-\infty < v \le -4$ [63]. In particular, this implies that as $L \to \infty$, the leftmost arc endpoints on $\mathcal{B}_v$ in the figures do not move down to pinch the negative real axis in this interval $-\infty < v \le -4$, provided that the limit $L \to \infty$ of these infinite-length, width-L strips commutes with the 2D thermodynamic limit for the triangular lattice as regards this aspect of the complex-temperature phase diagram.

We do not show plots for $q \ge 5$, but recall that the Potts antiferromagnet is expected to be disordered (with exponential decay of spin-spin correlation functions) even at $T = 0$ on the triangular lattice. This can be proved rigorously for $q \ge 11$ as a slight improvement of the result that q-state Potts antiferromagnet is disordered at all temperatures on a lattice with coordination number r if $q > 2r$ [71]. The property that the Potts antiferromagnet is disordered at all T on the triangular lattice for $q \ge 5$ is reflected in the property that $\mathcal{B}_v$ does not pass through $v = -1$.

7 Internal Energy and Specific Heat

The partition function (1.1) can be used to derive the free-energy density $f(G, q, v)$

$$f(G, q, v) = \frac{1}{|V|} \log Z(G, q, v) \tag{7.1}$$

for finite $|V|$, with the $V \to \infty$ limit having been defined in eq. (1.4) above. The internal energy E and the specific heat C are derived in the usual way from the free

energy as

$$E(G,q,v) = -J\frac{\partial f}{\partial K} = -J(v+1)\frac{\partial f}{\partial v} \tag{7.2}$$

and

$$C = \frac{\partial E}{\partial T} = k_B K^2 (v+1)\left[\frac{\partial f}{\partial v} + (v+1)\frac{\partial^2 f}{\partial v^2}\right] \equiv k_B K^2 C_H \tag{7.3}$$

Henceforth, for convenience, we shall use a definition of E without the factor $-J$ in (7.2), and we shall use the dimensionless function C_H in discussions of the specific heat. Let us suppose that G is a triangular-lattice strip graph of size $L \times m$. In the limit $m \to \infty$, since only the dominant eigenvalue $\lambda_d(q,v)$ of the transfer matrix contributes to the free energy, one has

$$f(q,v;L) = \frac{1}{L}\log\lambda_d \tag{7.4a}$$

$$E(q,v;L) = \frac{1}{L}\Lambda_1 \tag{7.4b}$$

$$C_H(q,v;L) = \frac{1}{L}\left[\Lambda_2 + \Lambda_1 - \Lambda_1^2\right] \tag{7.4c}$$

where the $\Lambda_i(q,v;L)$ are defined by

$$\Lambda_i(q,v;L) = \frac{(1+v)^i}{\lambda_d(q,v;L)}\frac{\partial^i \lambda_d}{\partial v^i} \tag{7.5}$$

The infinite-length limits of the triangular-strips considered here are quasi-one-dimensional systems with analytic free energies at all temperatures. Hence the dominant eigenvalue λ_d is the same on the whole semi-axis $\mathrm{Im}(v) = 0$, $\mathrm{Re}(v) \geq -1$. Furthermore, as discussed in [17], the free energy and its derivatives with respect to the temperature are independent of the longitudinal boundary conditions in the limit $m \to \infty$ (although they depend on the transverse boundary conditions).

In Figure 13 we have plotted the internal energy E (7.2), the specific heat C_H (7.3), and the Binder cumulant U_4 (7.11) (see below) for $q = 3$ on a triangular-lattice strip of width L and infinite length with cylindrical boundary conditions.

The behavior of the energy for the triangular-lattice strips with cylindrical boundary conditions is interesting: the curves cross each others close to the critical value v_c in the ferromagnetic regime. In particular, for $q = 2$ we find that all curves cross at $v_c(2) = \sqrt{3} - 1 \approx 0.7320508...$ (See Table 7). For $q = 3, 4$ we find that the crossings are close to the respective PM-FM critical points $v_c(3)$ given exactly in eq. (6.12) ($= 0.8793852...$) and $v_c(4) = 1$, but they do not coincide precisely with these critical points (see Table 8). The reason for the above behavior is the following: the triangular-lattice Potts model on a triangular lattice at a given value of the temperature Boltzmann variable v is related by duality to the hexagonal-lattice Potts model at a different temperature variable

$$v' = \frac{q}{v} \tag{7.6}$$

This relation is exact when the original triangular-lattice is defined on an infinitely long cylinder of width L. In the Ising case $q = 2$, there is an additional transformation (namely, the star-triangle transformation) that maps back the hexagonal-lattice Potts model onto a triangular-lattice Potts model at temperature variable $v''(v)$. This transformation allows us to compute the critical temperature v_c (i.e., v_c is the unique fixed point of the equation $v''(v) = v$) and the critical energy $E_c = E(q, v_c; L)$. When we perform the computation of E_c in the limit $m \to \infty$, all finite-size corrections disappear, so that [81]

$$E_c(q = 2; L) \equiv E(q = 2, v_c; L) = E_c(q = 2; \infty) = \frac{5}{2} \tag{7.7}$$

However, for $q \neq 2$ there is no star-triangle transformation, and this implies the existence of corrections to scaling:

$$E_c(q; L) \equiv E(q, v_c; L) = E_c(q; \infty) + \sum_{k=1}^{\infty} \frac{A_k}{L^{\omega_k}} , \qquad q \neq 2 \tag{7.8}$$

where the parameters $\{\omega_k(q)\}$ are correction-to-scaling exponents that depend in general on the value of q.

One can obtain a pseudo-critical temperature $v_E \equiv v_E(q, L, L')$ by solving the equation

$$E(q, v_E, L) = E(q, v_E, L') \tag{7.9}$$

When $L, L' \to \infty$, we expect that this quantity will converge to the true critical value $v_c(q)$. This method has been employed in the the literature to locate critical points for several statistical-mechanical systems [82, 83, and references therein].

Another pseudo-critical temperature can be obtained by looking at the point $v_C \equiv v_C(q, L)$ where the specific heat $C(q, v; L)$ attains a maximum value. This value differs from the bulk critical value $v_c(q)$ by finite-size-corrections of order $\sim L^{-1/\nu}$ [84].

We can also consider higher derivatives of the free energy with respect to K. In particular, the quantity $\bar{u}_4$ is the fourth derivative of the free energy with respect to K and, in the limit $m \to \infty$ it can be written as

$$\bar{u}_4(q, v; L) = \frac{\partial^4 f}{\partial K^4} \tag{7.10}$$

Instead of $\bar{u}_4$, it is more useful to deal with the phenomenological quantity U_4 (also called the Binder cumulant [85]) defined as

$$U_4(q, v; L) = \frac{1}{L^2} \frac{\bar{u}_4}{C^2} \tag{7.11}$$

A plot of this quantity for $q = 3$ is given in Figure 13(c). A third pseudo-critical temperature $v_U \equiv v_U(q, L)$ can be defined as the value at which the Binder cumulant (7.11) attains a minimum value. Again, this estimate is expect to differ from the bulk critical temperature $v_c(q)$ by terms of order $L^{-1/\nu}$ [84].

The computation of the transfer matrices for triangular-lattice strips of width L allows us to compute these three pseudo-critical temperatures (namely, $v_E(q, L, L')$,

$v_C(q,L)$, and $v_U(q,L)$). We have computed these estimates for several values of $1/2 \leq q \leq 4$ in the ferromagnetic regime for strips with cylindrical boundary conditions (See Tables 6–8).[1,2] We expect a finite-size behavior of these three pseudo-critical temperatures of the type

$$v_Q(q;L) = v_c(q) + \sum_{k=1}^{\infty} \frac{A_{Q,k}}{L^{\omega_k}} \tag{7.12}$$

where $Q = E, C, U$, and the $\{A_{Q,k}, \omega_k\}$ are correction-to-scaling amplitudes and exponents, respectively. These quantities depend in general on q. We can try to estimate the critical temperature $v_c(q)$ by fitting the data to the above Ansatz (7.12) with only the leading correction-to-scaling term included [3]

$$v_Q(q;L) = v_c(q) + A_Q L^{-\Delta_Q} \tag{7.13}$$

The data listed in Tables 6–8 does not have any statistical error; there is only a truncation error ($\sim 10^{-10}$) due to the truncation of the data to ten decimal places. We have performed the fits in the following way: for a given value of the width $L_{\min}$, we take the three data points with $L = L_{\min}, L_{\min}+1, L_{\min}+2$ and perform a standard least-squares fit (notice that there are no degrees of freedom and hence, $\chi^2 = 0$). The variation of the estimates for v_c with the value of $L_{\min}$ will give us an estimate of its error bar.

Although the exact phase transition temperature for the q-state Potts ferromagnet is known, it is of some interest to compare these pseudo-critical temperature comparisons since we do not know, *a priori*, which estimate v_Q will give the most accurate results. In addition, we do not know the values of the correction-to-scaling exponents Δ_Q as (a) we do not know whether some of the dominant amplitude will vanish and (b) we expect that the estimates for Δ_Q will be an effective exponent, as the widths are not very large and we are dropping all subdominant corrections to scaling. It is thus very important to perform a comprehensive check of the method before studying a phase transition whose critical temperature is not known (for instance, the 3-state triangular-lattice Potts antiferromagnet; see below).

The exact phase transition temperature is quoted on the rows labelled "E" in Tables 6–8. It is clear that the estimates coming from the pseudo-critical temperature v_E are better by far than the other two; and the estimate v_U is more accurate than v_C. More precisely, the difference between the extrapolated value for v_c and the exact value is $\sim 10^{-7}$ for v_E for all the $1.5 \leq q \leq 4$ values considered. However, for v_U the

[1] The blank entries in Table 6 correspond to values for which the estimates belong to the antiferromagnetic regime.

[2] To compute the values at $q = 1$, we have computed all quantities at $q = 0.999$ and $q = 1.001$, and then taken their mean value. In practice, this procedure has only been done for v_E, as there are no estimates for v_C for $q < 1$, and the estimates for v_U have a very different finite-size behavior when $q \to 1^{\pm}$.

[3] We have included an explicit dependence on the observable $Q = E, C, U$ in both the amplitude and the exponent Δ because they are both effective quantities (See below).

discrepancy is of order $\sim 10^{-5}$ for $q \gtrsim 1.5$ and $\sim 10^{-2}$ to 10^{-4} for $q \lesssim 1$. Finally, for v_C the discrepancy is $\sim 10^{-4}$ for $q \gtrsim 2$, and $\sim 10^{-3}$ for $1 \lesssim q \lesssim 1.5$. In conclusion, we can establish the position of the ferromagnetic critical temperature $v_c(q)$ (over the whole range of values of q) by using the pseudo-critical temperature $v_E(L, L')$. The results for v_C and v_U are at least two orders of magnitudes worse, and the accuracy also depends on the value of q: it worsens for $q \lesssim 2$ for v_C, and for $q \lesssim 1.5$ for v_U.

Let us finish this discussion by quoting the estimates for Δ_Q. Indeed, the values of Δ_E are larger than the rest. In fact, $\Delta_E \approx 3.57$ for $q \gtrsim 2.5$, and it decreases from $\Delta_E \approx 3.41$ at $q = 1.5$ down to $\Delta_E \approx 2.94$ at $q = 0.5$. The values of Δ_U increases smoothly with q. In the interval where it works better $1.5 \leq q \leq 4$ it goes from $\Delta_E \approx 1.89$ to $\Delta_E \approx 2.95$. However, when it does not work so well (namely, $0.5 \leq q \leq 1.001$) it increases from $\Delta_E \approx 0.41$ to $\Delta_E \approx 1.56$. Finally, the specific-heat exponent grows from $\Delta_C \approx 1.60$ to $\Delta_C \approx 2.52$ along the interval $2 \leq q \leq 4$ (where it gives the better estimates), and takes the values $\Delta_C \approx 1.30$ at $q = 1.5$ and $\Delta_C \approx 0.86$ at $q = 1.001$.

One can try to extend the previous analysis to the antiferromagnetic regime. In Table 9 we show the estimates for the critical temperature of the 3-state Potts antiferromagnet using strips with cylindrical boundary conditions and widths that are multiples of 3. This constraint is due to the (mod 3)-oscillations that appear in antiferromagnets: in Figure 13 we clearly observe such oscillations in the antiferromagnetic regime. Thus, we keep only the data with $L = 3, 6, 9, 12$ that is expected to be closer to the thermodynamic limit. (When L is not a multiple of 3, the corresponding triangular-lattice strip with cylindrical boundary conditions is not tripartite, unlike those strips with L a multiple of 3 or the infinite triangular lattice; recall our earlier discussion of the chromatic number for these strips.) The value quoted on the row labelled "MC" comes from the Monte-Carlo study by Adler *et al.* [78]. For $L = 12$ we just list the estimate from the energy crossing since this is superior to the other two estimates, and this gives the value

$$v_c(q = 3) = -0.796927(20) \tag{7.14}$$

The error bar quoted in (7.14) was roughly estimated by comparing the above result to the value of $v_c(q = 3)$ obtained by fitting the data points $v_E(L, L')$ with $L = 3, 6, 9$ and $L' = L + 3$, namely, $v_c(q = 3) = -0.796907$. This is indeed a very conservative estimate for this error bar. Our results in (7.14) is in agreement with, and more accurate than, the estimate from [78] listed above in eq. (6.15).

Acknowledgment

The research of R.S. was supported in part by the NSF grant PHY-9722101.

References

[1] R. B. Potts, Proc. Camb. Phil. Soc. **48**, 106 (1952).

[2] F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).

[3] S.-C. Chang, J. Salas, and R. Shrock, J. Stat. Phys. **107** 1207-1253 (2002), cond-mat/0108144.

[4] G. D. Birkhoff, Ann. Math. **14**, 42 (1912).

[5] P. W. Kasteleyn and C. M. Fortuin, J. Phys. Soc. Jpn. **26**, 11 (1969) (Suppl.).

[6] C. M. Fortuin and P. W. Kasteleyn, Physica **57**, 536 (1972).

[7] W. T. Tutte, Can. J. Math. **6**, 80 (1954).

[8] W. T. Tutte, J. Combin. Theory **2**, 301 (1967).

[9] W. T. Tutte, "Chromials", in Lecture Notes in Math. v. 411 (1974) 243; *Graph Theory*, vol. 21 of Encyclopedia of Mathematics and Applications (Addison-Wesley, Menlo Park, 1984).

[10] H. Whitney, Bull. Amer. Math. Soc. **38**, 572 (1932).

[11] F. Y. Wu, J. Stat. Phys. **52**, 99 (1988).

[12] N. L. Biggs, *Algebraic Graph Theory* (2nd ed., Cambridge Univ. Press, Cambridge, 1993).

[13] D. J. A. Welsh, *Complexity: Knots, Colourings, and Counting*, London Math. Soc. Lect. Note Ser. 186 (Cambridge University Press, Cambridge, 1993).

[14] B. Bollobás, *Modern Graph Theory* (Springer, New York, 1998).

[15] R. Shrock, in the *Proceedings of the 1999 British Combinatorial Conference, BCC99* (July, 1999), Discrete Math. **231**, 421 (2001), cond-mat/9908387.

[16] H. Kluepfel, Stony Brook thesis "The q-State Potts Model: Partition Functions and Their Zeros in the Complex Temperature and q Planes" (July, 1999); H. Kluepfel and R. Shrock, unpublished.

[17] R. Shrock, Physica A **283**, 388 (2000), cond-mat/0001389.

[18] S.-C. Chang and R. Shrock, Physica **A 286**, 189 (2000), cond-mat/0004181.

[19] J. Salas and A. Sokal, J. Stat. Phys., **104**, 609 (2001), cond-mat/0004330.

[20] S.-C. Chang and R. Shrock, Physica A **296**, 131 (2001), cond-mat/0005232.

[21] S.-C. Chang and R. Shrock, Int. J. Mod. Phys. B **15**, 443 (2001), cond-mat/0007505.

[22] S.-C. Chang and R. Shrock, Physica A **296**, 183 (2001) cond-mat/0008477.

[23] S.-C. Chang and R. Shrock, Physica A **296**, 234 (2001) cond-mat/0011503.

[24] S.-Y. Kim and R. Creswick, Phys. Rev. **E63**, 066107 (2001) cond-mat/0102090.

[25] S.-C. Chang and R. Shrock, Physica A **301**, 301 (2001) cond-mat/0106607.

[26] S.-C. Chang and R. Shrock, Phys. Rev. E **64**, 066116 cond-mat/0107012 (2001).

[27] N. L. Biggs, R. M. Damerell, and D. A. Sands, J. Combin. Theory B **12**, 123 (1972).

[28] R. C. Read, J. Combin. Theory **4**, 52 (1968).

[29] R. C. Read and W. T. Tutte, "Chromatic Polynomials", in *Selected Topics in Graph Theory, 3*, eds. L. W. Beineke and R. J. Wilson (Academic Press, New York, 1988.).

[30] J.L. Jacobsen, J. Salas, and A.D. Sokal, Transfer matrices and partition-function zeros for antiferromagnetic Potts model. III. Triangular-lattice chromatic polynomial. cond-mat/0204587.

[31] M. E. Fisher, *Lectures in Theoretical Physics* (Univ. of Colorado Press, 1965), vol. 7C, p. 1.

[32] R. Abe, Prog. Theor. Phys. **38**, 322 (1967).

[33] S. Katsura, Prog. Theor. Phys. **38**, 1415 (1967).

[34] S. Ono, Y. Karaki, M. Suzuki, and C. Kawabata, J. Phys. Soc. Jpn. **25**, 54 (1968).

[35] S. Beraha, J. Kahane, and N. Weiss, J. Combin. Theory B **27**, 1 (1979); *ibid.* **28**, 52 (1980).

[36] S.-C. Chang, "Exact results for q-state Potts model partition functions", Ph.D. thesis, State Univ. of New York at Stony Brook (May, 2002).

[37] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences.* Published electronically at `http://www.research.att.com/~njas/sequences/` .

[38] R. P. Stanley, *Enumerative Combinatorics* (Cambridge University Press, Cambridge, 1999), v. 2.

[39] M. Bernstein and N. J. A. Sloane, Linear Algebra and Its Applications (Seidel Festschrift).

[40] T. Motzkin, Bull. Amer. Math. Soc. **54**, 352 (1948).

[41] R. Donaghey and L. W. Shapiro, J. Combin. Theory, A **23**, 291 (1977).

[42] J.L. Jacobsen and J. Salas, J. Stat. Phys. **104**, 701 (2001), cond-mat/0011456.

[43] N. L. Biggs, Bull. London Math. Soc. **9**, 54 (1977).

[44] N. L. Biggs, Combin. Theory B **82**, 19 (2001).

[45] R. J. Baxter, J. Phys. A **19**, 2821 (1986).

[46] R. J. Baxter, J. Phys. A **20**, 5241 (1987).

[47] S.-C. Chang and R. Shrock, Flow Polynomials and their Asymptotic Limits for Lattice Strip Graphs, cond-mat/0205424.

[48] M. Roček, R. Shrock, and S.-H. Tsai, Physica **A252**, 505 (1998)., cond-mat/9712148

[49] R. Shrock and S.-H. Tsai, Physica **A259**, 315 (1998), cond-mat/9807105

[50] M. Roček, R. Shrock, and S.-H. Tsai, Physica **A259**, 367 (1998), 9807106.

[51] S.-C. Chang and R. Shrock, Ann. Phys. **290**, 124 (2001), cond-mat/0004129.

[52] R. Shrock and S.-H. Tsai, Phys. Rev. **E60**, 3512 (1999), cond-mat/9910377.

[53] R. Shrock and S.-H. Tsai, Physica A **275**, 429 (2000), cond-mat/9907403.

[54] S.-C. Chang and R. Shrock, Physica A **292**, 307 (2001), cond-mat/0007491.

[55] R. Shrock and S.-H. Tsai, Phys. Rev. **E55**, 5165 (1997), cond-mat/9612249.

[56] R. Shrock and S.-H. Tsai, J. Phys. A Letts. **32**, L195 (1999), cond-mat/9903233.

[57] R. Shrock, Phys. Lett. **A261**, 57 (1999), cond-mat/9908323.

[58] N. L. Biggs and R. Shrock, J. Phys. A (Letts) **32**, L489 (1999), cond-mat/0001407.

[59] S.-C. Chang and R. Shrock, Physica A **290**, 402 (2001), cond-mat/0004161.

[60] R. J. Baxter, J. Math. Phys. **11**, 784 (1970).

[61] A.D. Sokal, Combin. Prob. Comput. **10**, 41 (2001), cond-mat/9904146.

[62] P. P. Martin and J.-M. Maillard, J. Phys. A **19**, L547 (1986).

[63] H. Feldmann, R. Shrock, and S.-H. Tsai, J. Phys. A (Lett.) **30**, L663 (1997), cond-mat/9710018.

[64] H. Feldmann, R. Shrock, and S.-H. Tsai, J.Phys. Rev. **E57**, 1335 (1998), cond-mat/9711058.

[65] H. Feldmann, A. J. Guttmann, I. Jensen, R. Shrock, and S.-H. Tsai, J. Phys. A **31**, 2287 (1998), cond-mat/9801305.

[66] H. Saleur, Commun. Math. Phys. **132**, 657 (1990); Nucl. Phys. **B360**, 219-263 (1991).

[67] D. Kim and R. Joseph, J. Phys. C **7**, L167 (1974).

[68] V. Matveev and R. Shrock, J. Phys. A **28**, 1557 (1995), hep-lat/9408020.

[69] V. Matveev and R. Shrock, J. Phys. A **28**, 5235 (1995), hep-lat/9503005.

[70] V. Matveev and R. Shrock, J. Phys. A **29**, 803 (1996), hep-lat/9411023.

[71] J. Salas and A. Sokal, J. Stat. Phys. **86**, 551 (1997), cond-mat/9603068; A. Sokal, private communication.

[72] R. Abe, T. Dotera, and T. Ogawa, Prog. Theor. Phys. **85**, 509 (1991).

[73] Domb, C. 1960 *Adv. in Phys.* **9** 149.

[74] J. Stephenson, J. Math. Phys. **5**, 1009 (1964) and references therein.

[75] G. S. Grest, J. Phys. A **14**, L217 (1981).

[76] Y. Saito, J. Phys. A **15**, 1885 (1982).

[77] E. G. Enting and F. Y. Wu, J. Stat. Phys. **28**, 351 (1982).

[78] J. Adler, A. Brandt, W. Janke, and S. Shmulyian, J. Phys. A **28**, 5117 (1995).

[79] B. Jackson, Combin. Prob. Comput. **2**, 325 (1993).

[80] V. Thomassen, Combin. Prob. Comput. **6**, 497 (1997).

[81] J. Salas, J. Phys. A **35**, 1833 (2002), cond-mat/0110287.

[82] M.A. Yurishchev, Nucl. Phys. B (Proc. Suppl.) **83-84**, 727 (2000), hep-lat/9908019.

[83] M.A. Yurishchev, Nucl. Phys. B (Proc. Suppl.) **106**, 917 (2002), hep-lat/0109025.

[84] M.N. Barber, in *Phase Transitions and Critical Phenomena*, Vol. 8, edited by C. Domb and J.L. Lebowitz (Academic Press, New York, 1983).

[85] M.S.S. Challa, D.P. Landau and K. Binder, Phys. Rev. B **34**, 1841 (1986).

m	1	2	3	4	5	6	7	8	9	10
a_m	1	1	3	7	19	51	141	393	1107	3139
b_m	0	1	2	6	16	45	126	357	1016	2907
d_m	0	1	2	5	13	35	96	267	750	2123
e_m	0	0	0	1	3	10	30	90	266	784
r_m	0	1	1	3	6	15	36	91	232	603

Table 1: Some numerical values of a_m, b_m, d_m, e_m, r_m

m	1	2	3	4	5	6	7	8	9	10
f_m	0	1	1	3	7	19	51	141	393	1107
h_m	0	0	1	3	9	26	75	216	623	1800
i_m	1	2	4	10	26	70	192	534	1500	4246
j_m	0	1	1	2	4	9	21	51	127	323
k_m	1	1	1	2	4	9	21	51	127	323

Table 2: Numbers of f_m, h_m, i_m, j_m, k_m

d	1	2	3	4	5	6	7	8	9	10
C_{d-1}	1	1	2	5	14	42	132	429	1430	4862
α'_d	1	2	4	10	26	80	246	810	2704	9252
α_d	1	1	3	8	25	75	245	800	2700	9225

Table 3: Numbers of $C_{d-1}, \alpha'_d, \alpha_d$

d	1	2	3	4	5	6	7	8	9	10
β_d^{I}	1	0	1	2	5	11	28	68	174	445
β_d^{II}	0	1	1	2	5	11	28	68	174	445
β_d	1	1	2	4	10	22	56	136	348	890
t_d	1	3	7	19	51	141	393	1107	3139	8953

Table 4: Numbers of β_d^{I}, β_d^{II}, β_d, and t_d.

L	$N_{Z,\mathrm{tri,FF},L}$	C_{L+1}	$N_{Z,\mathrm{tri,PF},L}$	$N_{Z,\mathrm{sq,PF},L}$
1	1	2	1	1
2	2	5	2	2
3	5	14	3	3
4	14	42	6	6
5	42	132	10	10
6	132	429	28	24
7	429	1430	63	49
8	1430	4862	190	130
9	4862	16796	546	336
10	16796	58786	1708	980
11	58786	208012	5346	2904
12	208012	742900	17428	9176

Table 5: Dimensions of the transfer matrix for triangular-lattice strips. For each strip width L we give the dimension of the transfer matrix for free boundary conditions $N_{Z,\mathrm{tri,FF},L}$ (which is equal to the Catalan number C_L), the dimension of the full transfer matrix for cylindrical boundary conditions (which is C_{L+1}), the dimension for cylindrical boundary conditions when translational symmetry is taken into account $N_{Z,\mathrm{tri,PF},L}$, and the dimension when we project onto the subspace of reflection-invariant connectivities $N_{Z,\mathrm{sq,PF},L}$.

q	L	v_C	v_U	L'	v_E
0.5	2			3	0.3855829942
	3			4	0.3847515915
	4		0.0362505003	5	0.3845315269
	5		0.0683846033	6	0.3844522368
	6		0.0922653853	7	0.3844168074
	7		0.1110484508	8	0.3843985957
	8		0.1263790456	9	0.3843882849
	9		0.1392308072		
	∞		0.402		0.38436686
	E	0.3843671526	0.3843671526		0.3843671526
0.999	2			3	0.5326299901
	3			4	0.5320721491
	4		0.0855663158	5	0.5319334539
	5		0.1493120749	6	0.5318861979
	6		0.1946816994	7	0.5318661552
	7		0.2291028542	8	0.5318563347
	8		0.2562939069	9	0.5318510121
	9		0.2784080729		
	∞		0.558		0.53184119
	E	0.5318414075	0.5318414075		0.5318414075
1	2			3	0.5328765341
	3			4	0.5323192792
	4			5	0.5321807427
	5			6	0.5321335445
	6			7	0.5321135278
	7			8	0.5321037207
	8			9	0.5320984057
	∞				0.53208860
	E	0.5320888862	0.5320888862		0.5320888862
1.001	2	0.1911168937	0.0890417228	3	0.5331230781
	3	0.3006378613	0.2852335683	4	0.5325664093
	4	0.3532531983	0.3750722572	5	0.5324280315
	5	0.3855204044	0.4214503869	6	0.5323808911
	6	0.4076019935	0.4490378514	7	0.5323609004
	7	0.4237541586	0.4669772403	8	0.5323511067
	8	0.4361202804	0.4793897234	9	0.5323457993
	9	0.4459098965	0.4883826060		
	∞	0.5378	0.53288		0.53233601
	E	0.5323362257	0.5323362257		0.5323362257

Table 6: Pseudo-critical temperatures for the q-state Potts model in the ferromagnetic regime. For each value of $q \lesssim 1$ and the strip width L, we show the pseudo-critical temperatures v_C, v_U, and v_E. At $v = v_C$ [resp. $v = v_U$], the specific heat $C_H(q, v; L)$ [resp. the cumulant $U_4(q, v; L)$] attains its maximum [resp. minimum] value. At $v = v_E$ the energies for widths L and $L' = L + 1$ cross. The rows labelled $L = \infty$ show the extrapolated data to $L = \infty$ (see text), and the rows labelled "E" show the exact values of the critical temperature $v_c(q)$.

q	L	v_C	v_U	L'	v_E
1.5	2	0.3528174426	0.3641204223	3	0.6421567042
	3	0.4761242418	0.5117472589	4	0.6418879337
	4	0.5283927381	0.5662745790	5	0.6418236445
	5	0.5571932786	0.5922393905	6	0.6418023976
	6	0.5752485686	0.6066812044	7	0.6417936360
	7	0.5875189961	0.6155591541	8	0.6417894549
	8	0.5963385107	0.6214146331	9	0.6417872438
	9	0.6029456311	0.6254843739		
	∞	0.6430	0.6417897		0.64178341
	E	0.6417835275	0.6417835275		0.6417835275
2	2	0.4923479329	0.5368701014	3	0.7320508076
	3	0.6090497052	0.6477325016	4	0.7320508076
	4	0.6545851407	0.6863163796	5	0.7320508076
	5	0.6779639930	0.7036554781	6	0.7320508076
	6	0.6917883697	0.7128237907	7	0.7320508076
	7	0.7007283606	0.7182229040	8	0.7320508076
	8	0.7068826013	0.7216553263	9	0.7320508076
	9	0.7113202081	0.7239660420		
	∞	0.73271	0.732008		0.7320508076
	E	0.7320508076	0.7320508076		0.7320508076
2.5	2	0.6069444365	0.6576959827	3	0.8097033821
	3	0.7143925462	0.7495237824	4	0.8099490872
	4	0.7537340192	0.7792175664	5	0.8100049902
	5	0.7728306915	0.7918221561	6	0.8100228881
	6	0.7835975327	0.7981861988	7	0.8100300791
	7	0.7902789716	0.8017934770	8	0.8100334325
	8	0.7947142805	0.8040139149	9	0.8100351693
	9	0.7978101827	0.8054677154		
	∞	0.81046	0.809995		0.81003803
	E	0.8100379292	0.8100379292		0.8100379292

Table 7: Pseudo-critical temperatures for the q-state Potts model in the ferromagnetic regime. For each value of $1.5 \leq q \leq 2.5$ and the strip width L, we show the pseudo-critical temperatures v_C, v_U, and v_E. Notation is as in Table 6.

q	L	v_C	v_U	L'	v_E
3	2	0.7037350269	0.7548212317	3	0.8787491090
	3	0.8024796990	0.8336040111	4	0.8792191863
	4	0.8367131720	0.8573852148	5	0.8793241938
	5	0.8525340980	0.8669589648	6	0.8793575050
	6	0.8610840803	0.8715920290	7	0.8793708114
	7	0.8661968447	0.8741280057	8	0.8793769918
	8	0.8694810441	0.8756436301	9	0.8793801836
	9	0.8717067404	0.8766110378		
	∞	0.87967	0.879353		0.87938540
	E	0.8793852416	0.8793852416		0.8793852416
3.5	2	0.7878934565	0.8376620405	3	0.9413319636
	3	0.8789024847	0.9064866067	4	0.9420076334
	4	0.9089394979	0.9259778215	5	0.9421562000
	5	0.9222116633	0.9334431103	6	0.9422030026
	6	0.9291089227	0.9369143229	7	0.9422216338
	7	0.9330936401	0.9387526207	8	0.9422302741
	8	0.9355757822	0.9398209283	9	0.9422347354
	9	0.9372119953	0.9404864726		
	∞	0.94244	0.942221		0.94224203
	E	0.9422418510	0.9422418510		0.9422418510
4	2	0.8627238740	0.9107070432	3	0.9988399935
	3	0.9469004865	0.9714927384	4	0.9997049072
	4	0.9734479683	0.9877041563	5	0.9998924497
	5	0.9846953640	0.9936232441	6	0.9999512055
	6	0.9903282704	0.9962710802	7	0.9999745475
	7	0.9934777281	0.9976288578	8	0.9999853718
	8	0.9953829235	0.9983964745	9	0.9999909679
	9	0.9966059950	0.9988633354		
	∞	1.00015	0.999988		1.00000014
	E	1.0000000000	1.0000000000		1.0000000000

Table 8: Pseudo-critical temperatures for the q-state Potts model in the ferromagnetic regime. For each value of $3 \leq q \leq 4$ and the strip width L, we show the pseudo-critical temperatures v_C, v_U, and v_E. Notation is as in Table 6.

q	L	v_C	v_U	L'	v_E
3	3	-0.7537129688	-0.7746989054	6	-0.7984897326
	6	-0.7916555922	-0.7942230532	9	-0.7971540641
	9	-0.7952008646	-0.7959971452	12	-0.7969905288
	12			15	-0.7969527708
	∞	-0.79660	-0.79668		-0.796927(20)
	MC	-0.79691(3)	-0.79691(3)		-0.79691(3)

Table 9: Pseudo-critical temperatures for the 3-state Potts model in the antiferromagnetic regime. For each value of the strip width $L = 3k$, we show the pseudo-critical temperatures v_C, v_U, and v_E computed on strips with cylindrical boundary conditions. Notation is as in Table 6, and the row labelled "MC" shows the Monte Carlo estimate for $v_c(q)$.

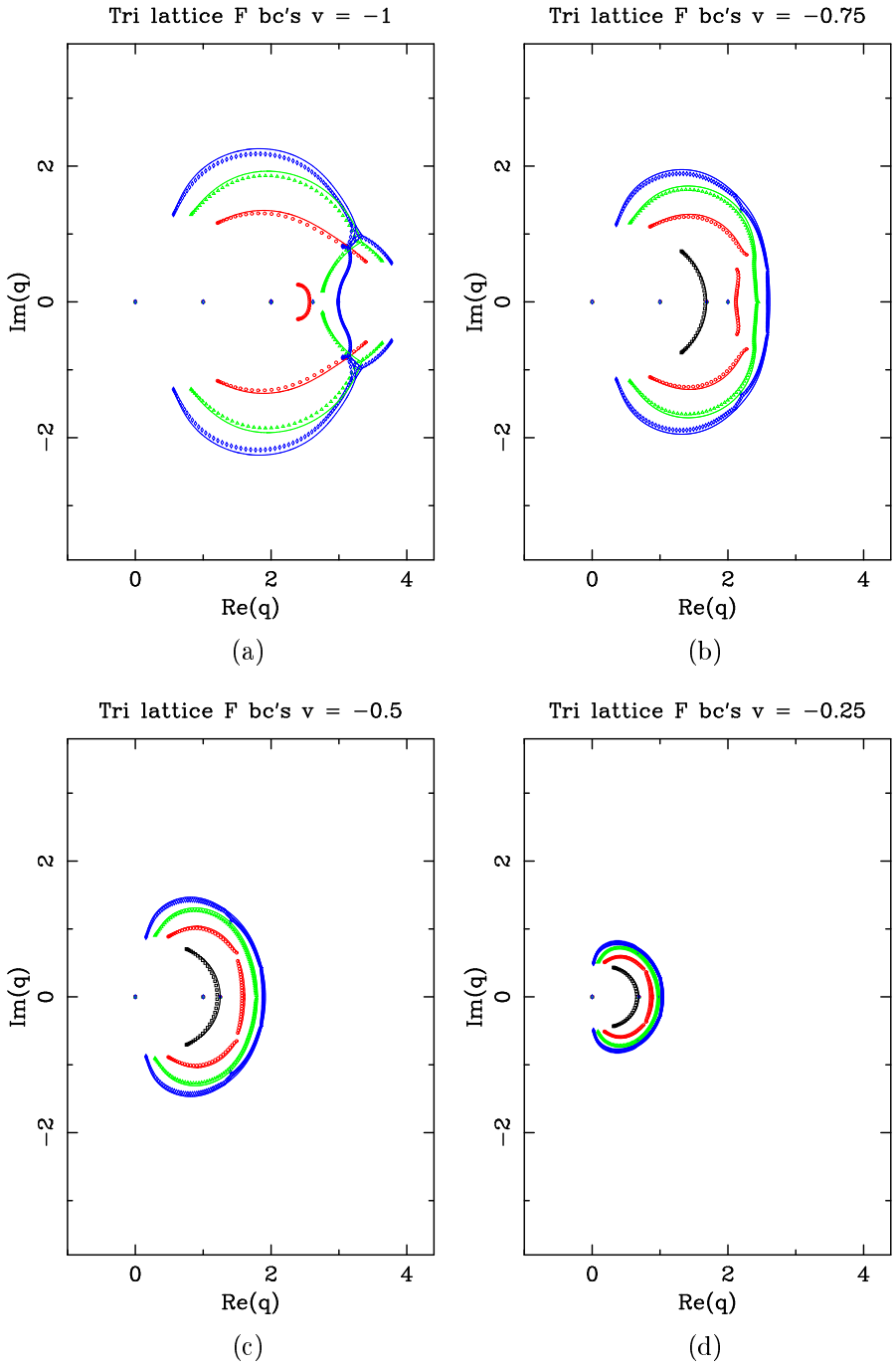


Figure 1: Limiting curves forming the singular locus $\mathcal{B}_q$ for the Potts model free energy for (a) $v = -1$, (b) $v = -3/4$, (c) $v = -1/2$, and (d) $v = -1/4$ on strips with free boundary conditions and several widths L: 2 (black), 3 (red), 4 (green), and 5 (blue). We also show the partition-function zeros for the strips $L_{\rm F} \times (10L)_{\rm F}$ for the same values of L: 2 ($\Box$, black), 3 ($\circ$, red), 4 ($\triangle$, green), and 5 ($\Diamond$, blue).

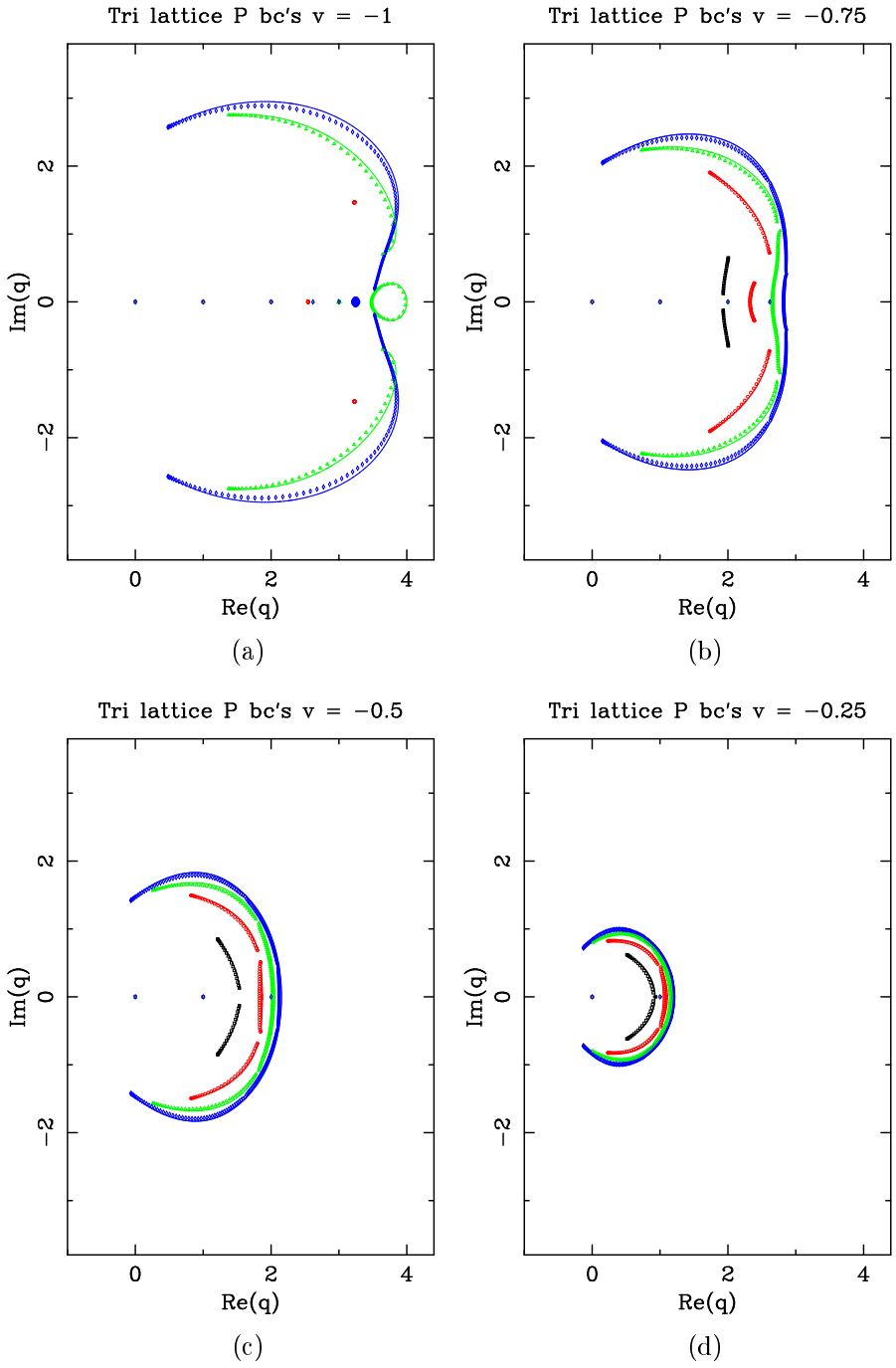


Figure 2: Limiting curves forming the singular locus $\mathcal{B}_q$ for the Potts model free energy for (a) $v = -1$, (b) $v = -3/4$, (c) $v = -1/2$, and (d) $v = -1/4$ on strips with cylindrical boundary conditions and several widths L: 2 (black), 3 (red), 4 (green), and 5 (blue). We also show the partition-function zeros for the strips $L_{\rm P} \times (10L)_{\rm F}$ for the same values of L. The symbols are as in Figure 1.

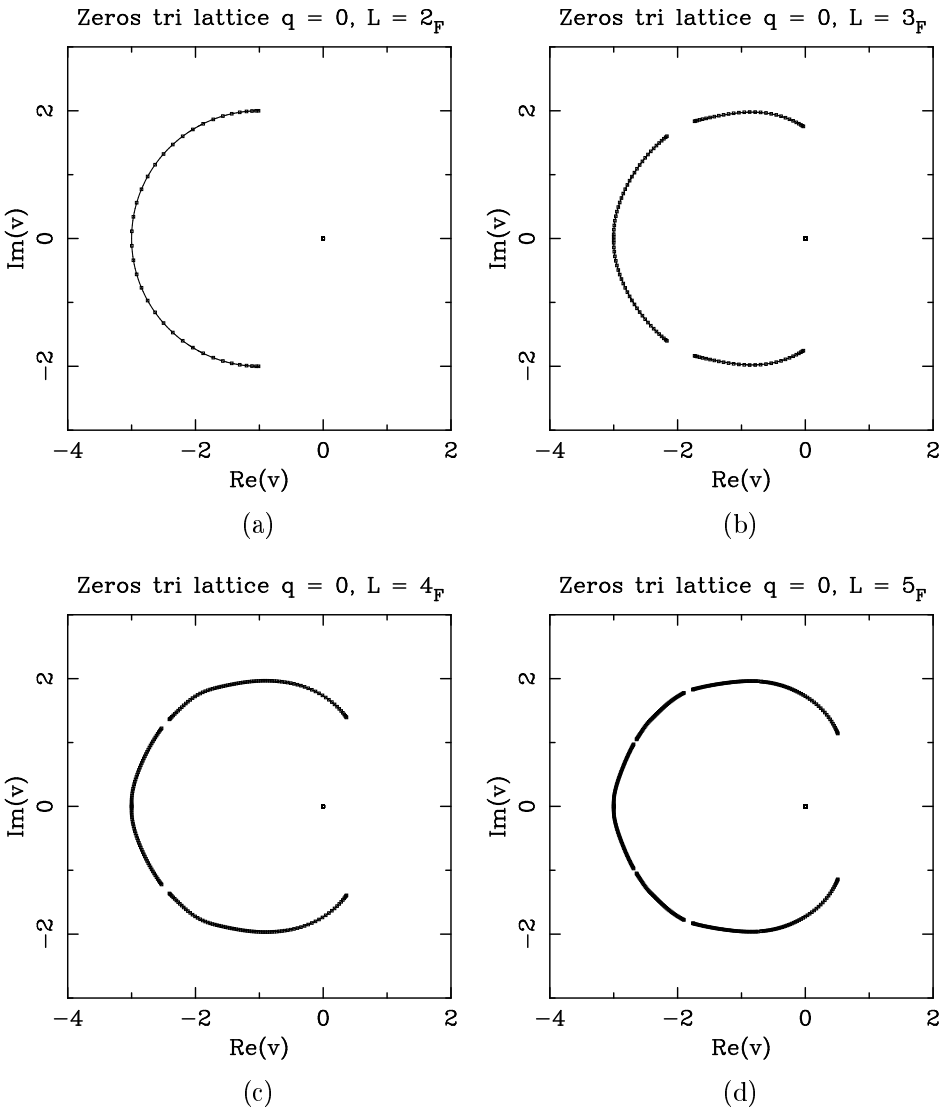


Figure 3: Limiting curves forming the singular locus $\mathcal{B}$, in the v plane, for the free energy, defined with the order f_{qn}, of the Potts model for $q = 0$ on the $L_{\rm F} \times \infty_{\rm F}$ triangular-lattice strips with (a) $L = 2$, (b) $L = 3$, (c) $L = 4$, and (d) $L = 5$. We also show the zeros of $Z(G, 0, v)/q$ corresponding to the strips $L_{\rm F} \times (10L)_{\rm F}$ for each value of L.

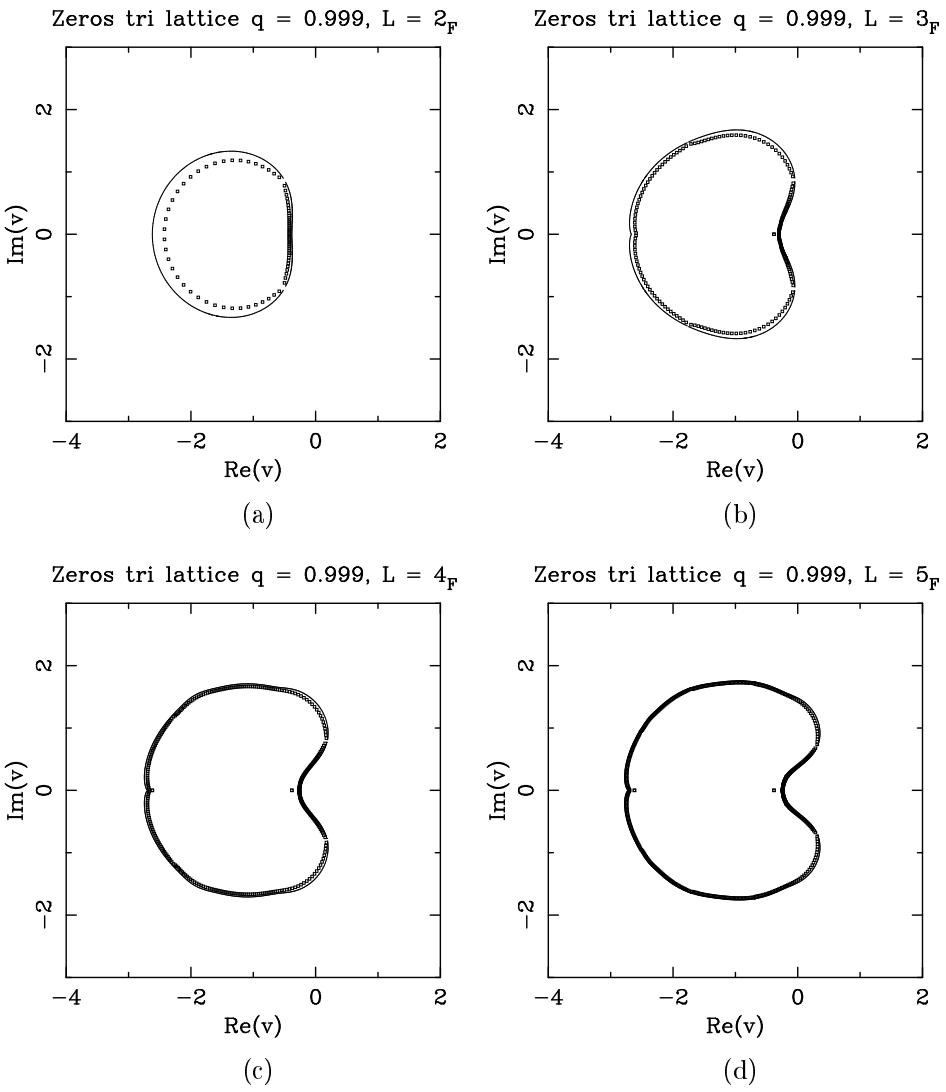


Figure 4: Limiting curves forming the singular locus $\mathcal{B}$, in the v plane, for the free energy of the Potts model for $q = 0.999$ on the $L_F \times \infty_F$ triangular-lattice strips with (a) $L = 2$, (b) $L = 3$, (c) $L = 4$, and (d) $L = 5$. This are essentially equivalent to the limiting curves for f_{qn} at $q = 1$. We also show the partition-function zeros corresponding to the strips $L_F \times (10L)_F$ for each value of L.

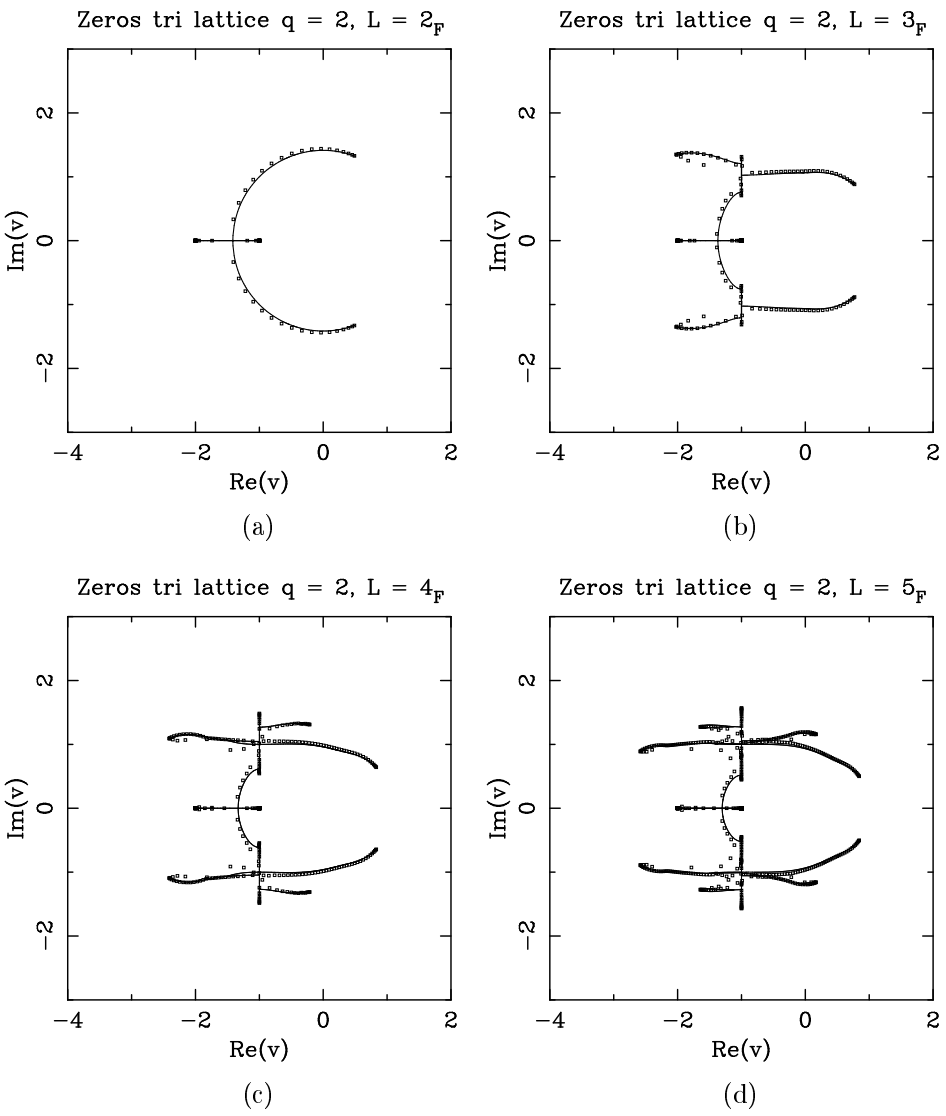


Figure 5: Limiting curves forming the singular locus $\mathcal{B}$, in the v plane, for the free energy of the Potts model for $q = 2$ on the $L_{\rm F} \times \infty_{\rm F}$ triangular-lattice strips with (a) $L = 2$, (b) $L = 3$, (c) $L = 4$, and (d) $L = 5$. We also show the partition-function zeros corresponding to the strips $L_{\rm F} \times (10L)_{\rm F}$ for each value of L.

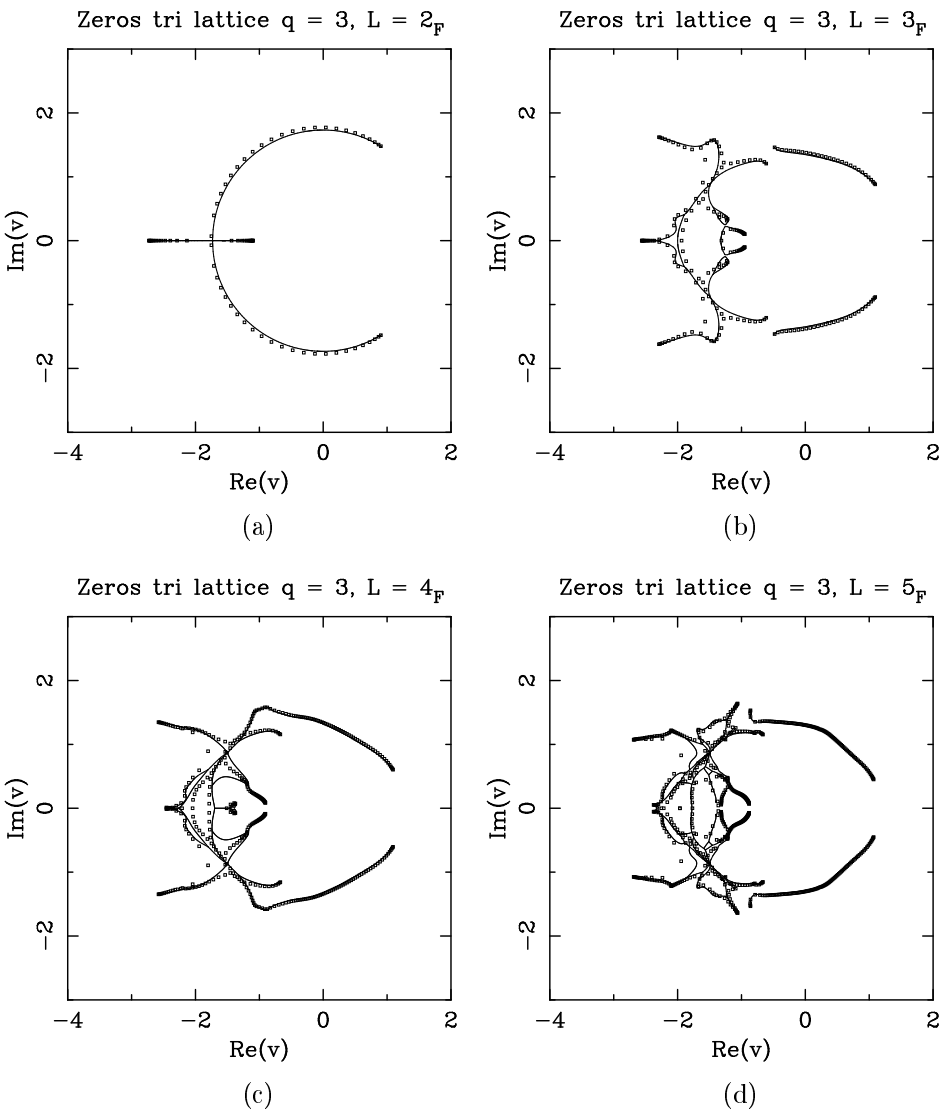


Figure 6: Limiting curves forming the singular locus $\mathcal{B}$, in the v plane, for the free energy of the Potts model for $q = 3$ on the $L_{\rm F} \times \infty_{\rm F}$ triangular-lattice strips with (a) $L = 2$, (b) $L = 3$, (c) $L = 4$, and (d) $L = 5$. We also show the partition-function zeros corresponding to the strips $L_{\rm F} \times (10L)_{\rm F}$ for each value of L.

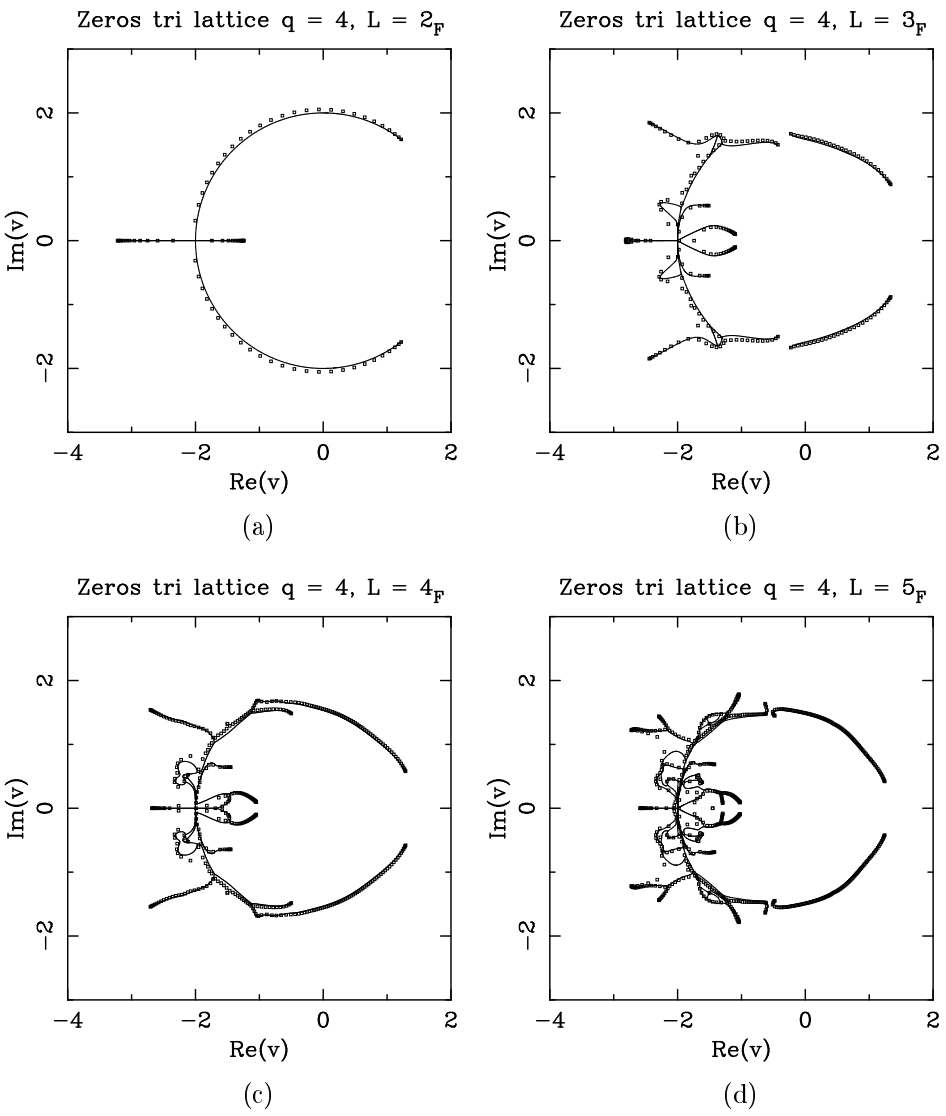


Figure 7: Limiting curves forming the singular locus $\mathcal{B}$, in the v plane, for the free energy of the Potts model for $q = 4$ on the $L_{\rm F} \times \infty_{\rm F}$ triangular-lattice strips with (a) $L = 2$, (b) $L = 3$, (c) $L = 4$, and (d) $L = 5$. We also show the partition-function zeros corresponding to the strips $L_{\rm F} \times (10L)_{\rm F}$ for each value of L.

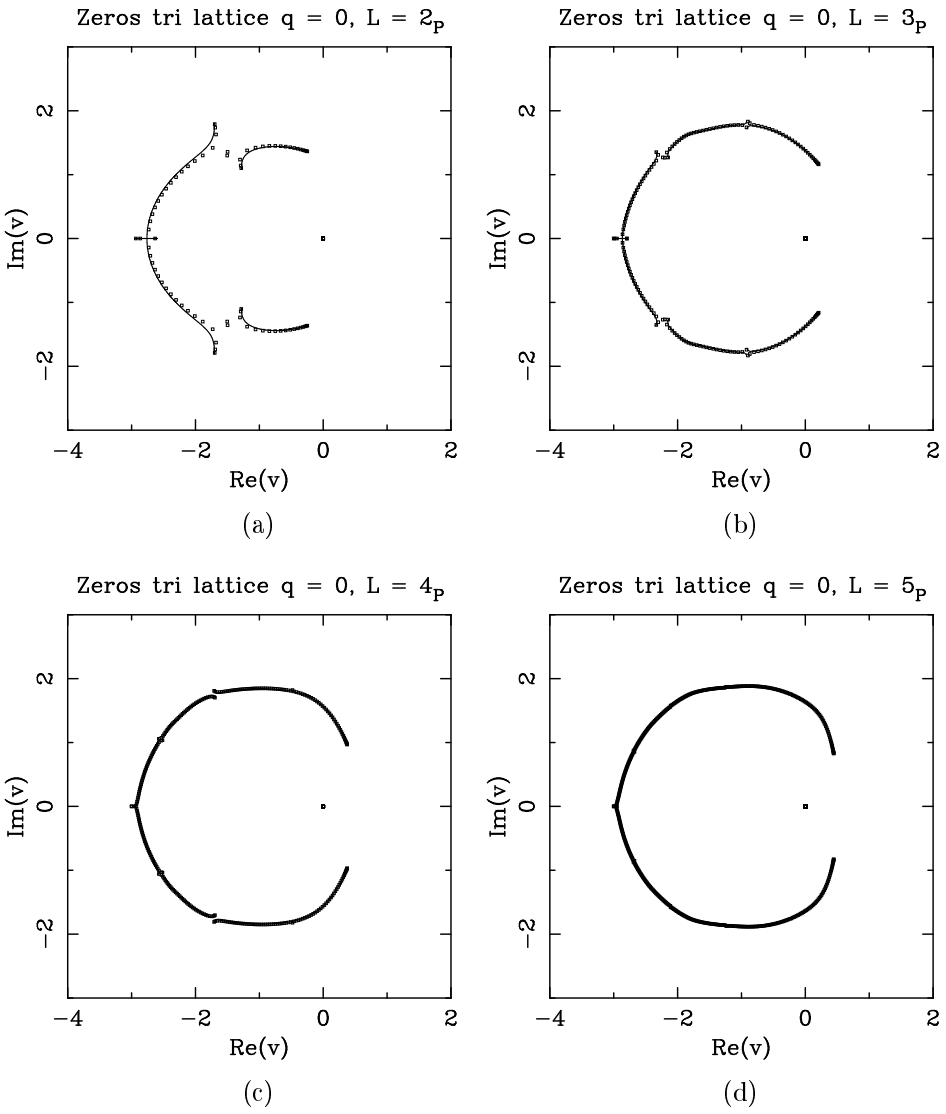


Figure 8: Limiting curves forming the singular locus $\mathcal{B}$, in the v plane, for the free energy of the Potts model, defined with the order f_{qn}, for $q = 0$ on the $L_{\rm P} \times \infty_{\rm F}$ triangular-lattice strips with (a) $L = 2$, (b) $L = 3$, (c) $L = 4$, and (d) $L = 5$. We also show the zeros of $Z(G, 0, v)/q$ corresponding to the strips $L_{\rm P} \times (10L)_{\rm F}$ for each value of L.

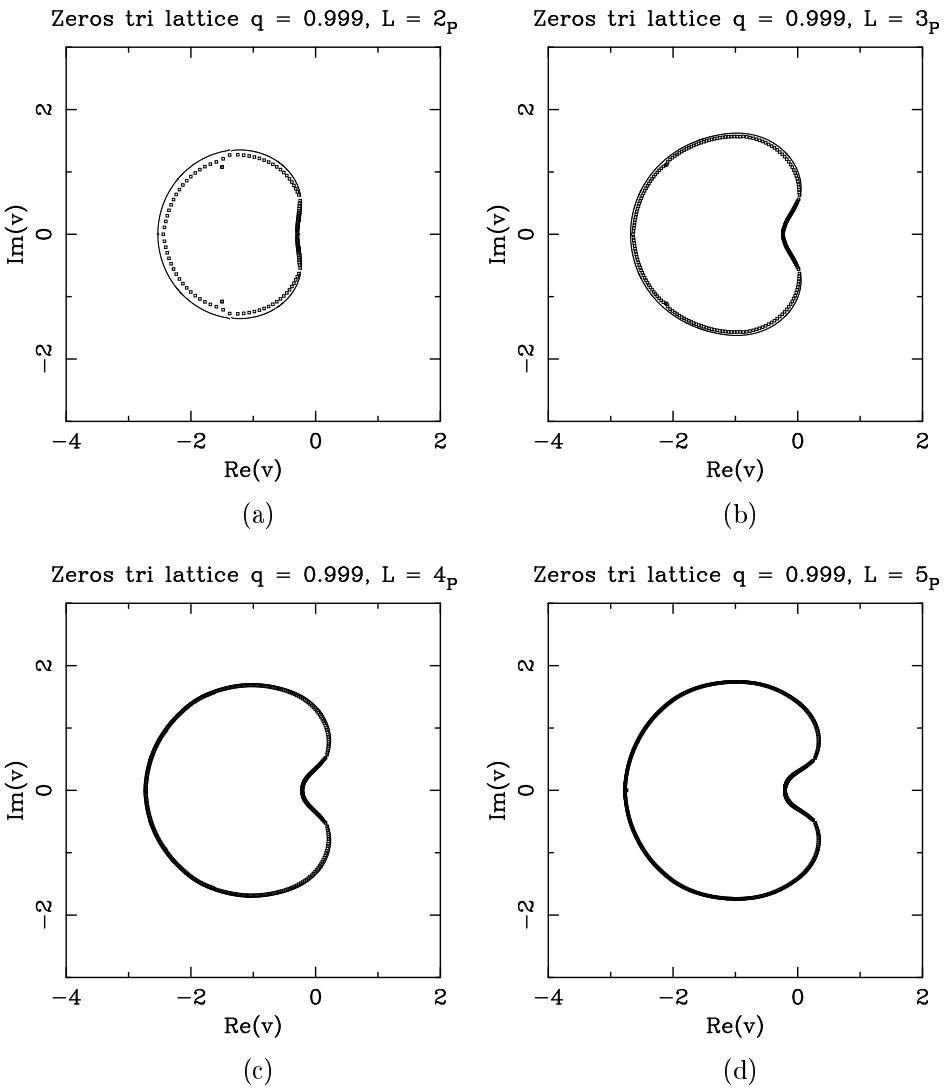


Figure 9: Limiting curves forming the singular locus $\mathcal{B}$, in the v plane, for the free energy of the Potts model for $q = 0.999$ on the $L_{\rm P} \times \infty_{\rm F}$ triangular-lattice strips with (a) $L = 2$, (b) $L = 3$, (c) $L = 4$, and (d) $L = 5$. These are essentially equivalent to the limiting curves for f_{qn} at $q = 1$. We also show the partition-function zeros corresponding to the strips $L_{\rm P} \times (10L)_{\rm F}$ for each value of L.

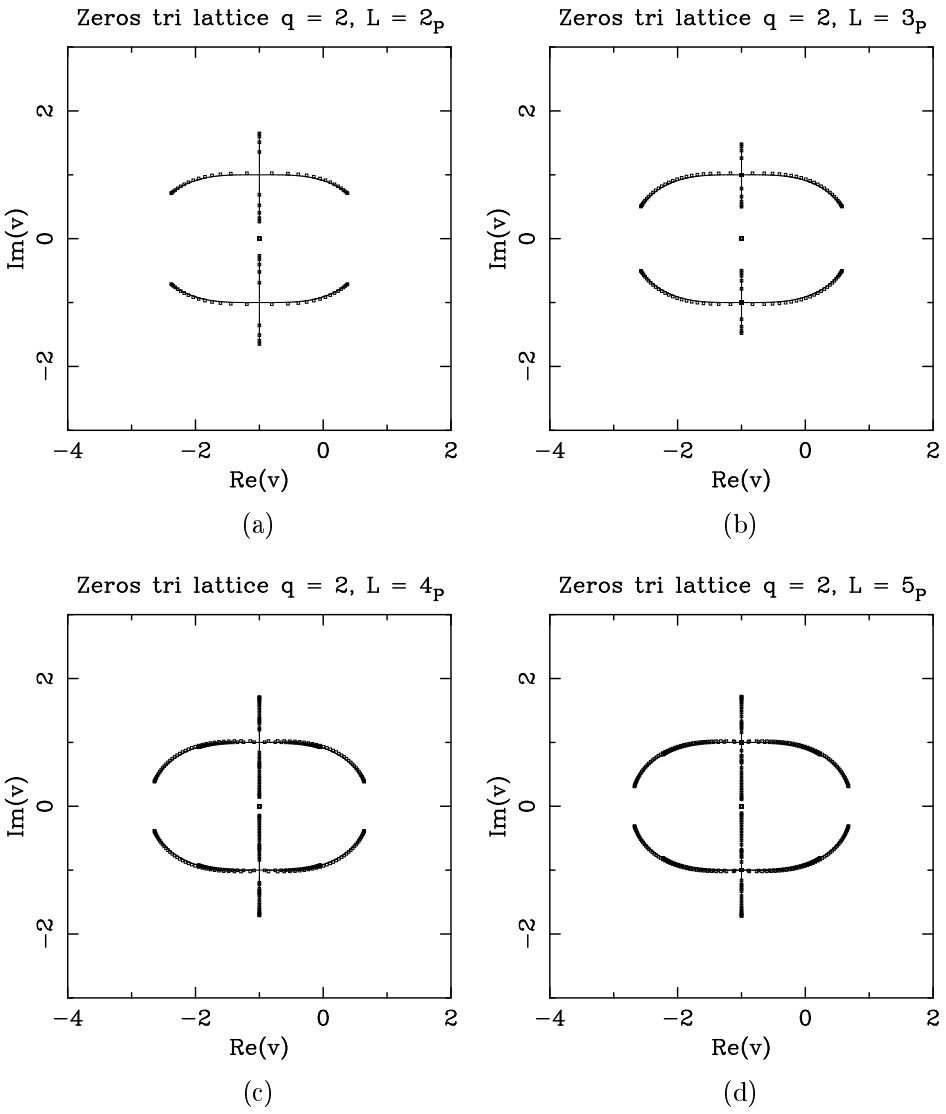


Figure 10: Limiting curves forming the singular locus $\mathcal{B}$, in the v plane, for the free energy of the Potts model for $q = 2$ on the $L_{\rm P} \times \infty_{\rm F}$ triangular-lattice strips with (a) $L = 2$, (b) $L = 3$, (c) $L = 4$, and (d) $L = 5$. We also show the partition-function zeros corresponding to the strips $L_{\rm P} \times (10L)_{\rm F}$ for each value of L.

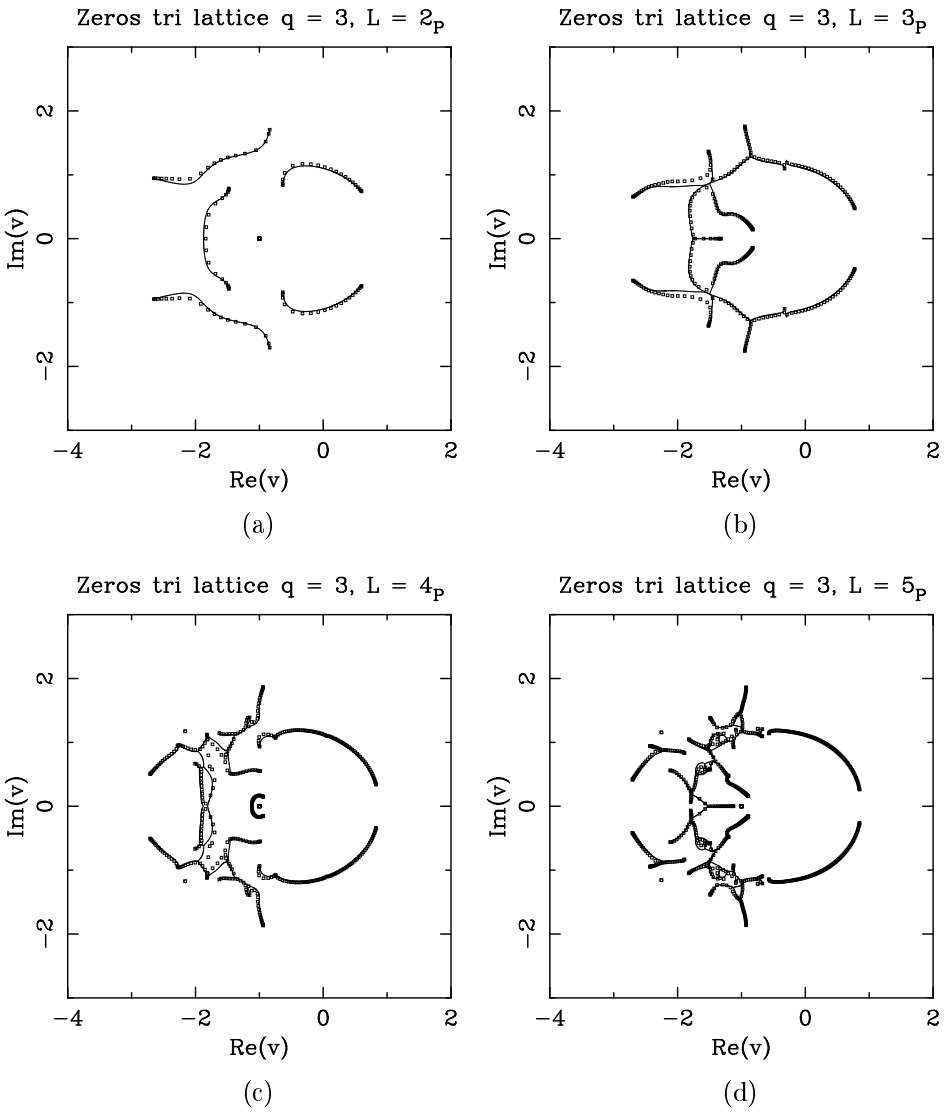


Figure 11: Limiting curves forming the singular locus $\mathcal{B}$, in the v plane, for the free energy of the Potts model for $q = 3$ on the $L_{\rm P} \times \infty_{\rm F}$ triangular-lattice strips with (a) $L = 2$, (b) $L = 3$, (c) $L = 4$, and (d) $L = 5$. We also show the partition-function zeros corresponding to the strips $L_{\rm P} \times (10L)_{\rm F}$ for each value of L.

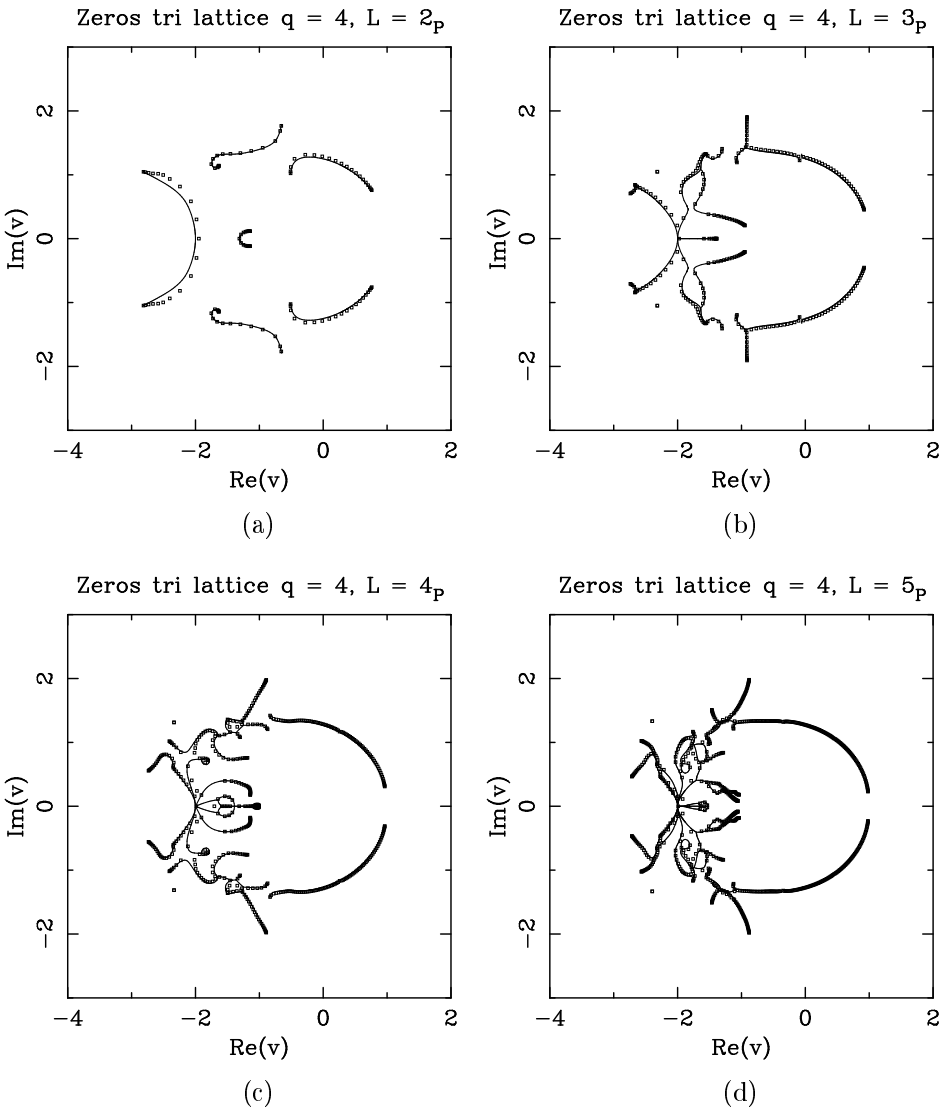


Figure 12: Limiting curves forming the singular locus $\mathcal{B}$, in the v plane, for the free energy of the Potts model for $q = 4$ on the $L_{\rm P} \times \infty_{\rm F}$ triangular-lattice strips with (a) $L = 2$, (b) $L = 3$, (c) $L = 4$, and (d) $L = 5$. We also show the partition-function zeros corresponding to the strips $L_{\rm P} \times (10L)_{\rm F}$ for each value of L.

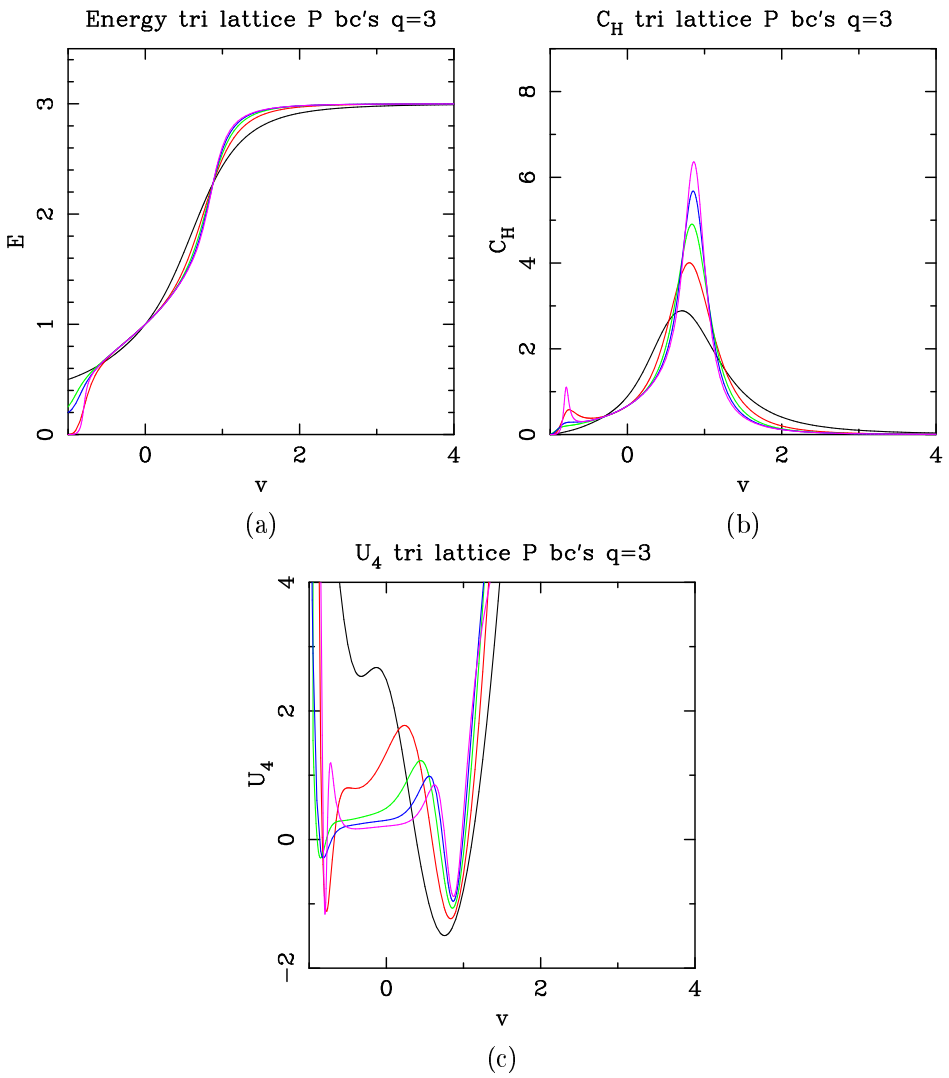


Figure 13: Thermodynamic observables for the 3-state Potts model on triangular-lattice strips of sizes $L_{\rm P} \times \infty_{\rm F}$. We show (a) the energy density E, (b) the specific heat C_H, and (c) the Binder cumulant U_4 as a function of the temperature-like parameter v for several strip widths L: 2 (black), 3 (red), 4 (green), 5 (blue), and 6 (pink). $-1 \le v < 0$ corresponds to the antiferromagnetic regime, while $v > 0$ to the ferromagnetic one.

Chapitre 5

Coloriage d'arêtes : polymères compacts

La non universalité mise en évidence par le problème de coloriage de sommets (voir le Chapitre 3) soulève également la question de savoir si d'autres types de coloriages peuvent mener à des problèmes intéressants. Par exemple, on pourrait colorier les faces plutôt que les sommets du graphe, avec la contrainte que chaque arête sépare deux faces de couleurs différentes. Or, par dualité, ceci équivaut à colorier les sommets du graphe dual ce qui nous ramenè au problème étudié précédemment. Une autre possibilité, plus intéressante, de colorier les faces serait d'exiger la différence des couleurs de toutes les faces qui touchent un même sommet : par dualité, ceci correspond à un coloriage de sommets avec une contrainte sur les faces. Il paraît que ce problème a été assez peu étudié, sauf pour le cas du réseau triangulaire [163, 13].

Le point de départ du présent chapitre est encore une variante du problème de coloriage : on attribue maintenant une couleur à chaque arête avec la contrainte que toutes les arêtes incidentes à un même sommet aient des couleurs différentes. Les configurations qui en résultent sont connues commes des *coloriages de Tait* [158]. On s'attend à ce que ce problème soit plus intéressant pour des graphes de valence constante et avec un nombre de couleurs (supérieure ou) égale à celle-ci.

Bien évidemment, un problème de coloriage d'arêtes équivaut toujours à un problème de coloriage de sommets d'un graphe associé. Par exemple, le tricoloriage des arêtes du réseau hexagonal est équivalent au tricoloriage des sommets du réseau Kagomé ; ce problème a $c = 2$ [78, 131]. Pourtant, le graphe associé n'est pas toujours planaire comme le montre l'exemple du tétracoloriage des arêtes du réseau carré.

Lors de l'étude du coloriage de sommets, nous avons utilisé le paramètre q du polynôme chromatique afin de définir une déformation intéressante du problème de coloriage proprement dit. Cette déformation faisait appel à des objets non locaux : les amas de Kasteleyn-Fortuin [103]. Le coloriage d'arêtes ouvre des possibilités similaires, comme nous le verrons dans un instant, où les objets non locaux sont des boucles autoévitantes compactes.

Cette extension du problème constitue sa motivation physique, car les boucles servent à modéliser les conformations de polymères compacts. Ce type de polymères bidimensionnels a été étudié dans des expériences récentes de microscopie fluorescente : dans ces expériences, des molécules d'ADN sont adsorbées sur des bicouches lipides, ce qui confine leurs espace de phase à deux dimensions [130].

Au niveau formel, les boucles fournissent un autre lien entre le modèle de Potts et le modèle $\mathrm{O}(n)$.

5.1 Tricoloriage du réseau hexagonal

A partir d'un tricoloriage des arêtes du réseau hexagonal, il est possible de définir une configuration de boucles fermées autoévitantes en effaçant toutes les arêtes avec une couleur donnée. Ces boucles sont des cycles le long desquels les deux couleurs restantes alternent. Une boucle donnée est donc invariante sous une permutation de ces deux couleurs le long de sa trajectoire. En particulier, si tous les tricoloriages ont un poids 1, alors chaque boucle porte un poids de 2.

Pourtant, les boucles définies de cette manière ne sont pas les boucles intervenant dans le développement de haute température du modèle $\mathrm{O}(n)$. Par construction, l'ensemble de boucles passe par la totalité de sommets du réseau. Dans le modèle $\mathrm{O}(n)$, par contre, des sommets incidents à aucune boucle apparaissent avec un poids t, où t peut être interprété comme la température [134].

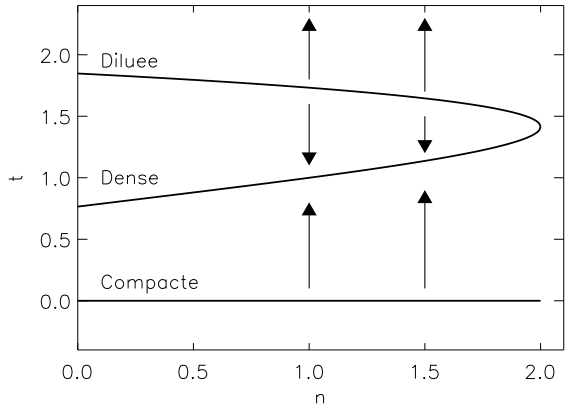


FIG. 5.1 – Diagramme de phase du modèle O(n) sur le réseau hexagonal. Les flêches représentent les flots sous le groupe de renormalisation.

Plus précisement, le modèle O(n) est défini par le hamiltonien

$$-\beta H = \sum_{\langle ij \rangle} K \vec{S}_i \cdot \vec{S}_j \tag{5.1}$$

où $\vec{S}_i$ est un vecteur de n composantes ayant la normalisation $\vec{S}_i \cdot \vec{S}_i = n$. Le modèle de boucles associé apparaît par une troncature du développement de haute température :

$$\exp(-\beta H) = \prod_{\langle ij \rangle} \left(1 + K \vec{S}_i \cdot \vec{S}_j\right). \tag{5.2}$$

Le célèbre travail de Nienhuis [134] est basé sur le développement graphique de cette expression qui prend la forme du modèle de boucles énoncé ci-dessus.

Nous allons discuter la validité de cette troncature du développement de haute température dans le Chapitre 6. Pour l'instant, la fonction de partition du modèle O(n) vaut donc

$$Z = \sum_{\mathcal{G}} t^V n^L \tag{5.3}$$

où V est le nombre de sommets incidents à aucune boucle, L représente le nombre de boucles fermées et $\mathcal{G}$ dénote les configurations de boucles sur le réseau hexagonal. Il est facile de se convaincre que, pour ce choix de réseau, V est toujours un nombre pair. On a donc $Z = Z(t^2)$, et le modèle est invariant sous la transformation $t \to -t$.

Nienhuis a montré que le modèle (5.3) est soluble pour [134]

$$t^2 = 2 \pm \sqrt{2 - n} \tag{5.4}$$

avec une transition de second ordre quand $-2 \leq n \leq 2$. La branche "+" (dite phase diluée) est le point critique proprement dit, tandis que la branche "−" (dite phase dense) est une ligne de points fixes où la température est une perturbation non pertinente. (Le lecteur ne manquera pas d'apprécier l'analogie avec la phase de Berker-Kadanoff discutée lors du Chapitre 4.) Les résultats de Nienhuis ont été rendus rigoureux par une solution du type ansatz de Bethe pour l'énergie libre [15] : l'extension de cette solution à taille finie [6] montre que la charge centrale dans la phase dense est identique à celle des modèles minimaux (1.4), à condition de mettre $q = n^2$ dans la paramétrisation (4.5). Ceci fournit encore un lien entre le modèle de Potts et le modèle O(n). Il existe également des coïncidences frappantes au niveau des exposants critiques [54].

Le diagramme de phase, avec les flots du groupe de renormalisation, est montré dans la Figure 5.1. L'invariance sous $t \to -t$ se traduit par le fait que la limite $t \to 0$ constitue une autre branche de points fixes [28] : le modèle de tricoloriage est le point $n = 2$ sur cette ligne. (Le point $n = 1$ a déjà été discuté dans le Chapitre 2 : c'est le modèle d'Ising antiferromagnétique sur le réseau triangulaire.) En vue de la

nature des conformations des boucles, nous allons appeler cette branche la *phase compacte* du modèle $\mathrm{O}(n)$.

Pour compléter le diagramme de phase, remarquons que la ligne $n = 2$, $0 < t^2 < 2$ est également critique (phase de Berezinskii-Kosterlitz-Thouless, ou BKT) [22, 116] et que la région $n > 2$ contient un flot vers un point critique à $n = \infty$ qui est dans la classe d'universalité du modèle de Potts à trois états [74].

Nous allons désormais nous concentrer sur la phase compacte, $t = 0$. La compacité des boucles se traduit par une loi de conservation de lignes qui permit à Baxter de résoudre rigoureusement le modèle pour $n > 2$ [9]. Dans ce régime, le modèle est non critique et toutes les boucles restent de longueur finie dans la limite thermodynamique. Quand $n \to 2$, la longueur de la plus grande boucle (qui peut s'interpréter comme la longueur de corrélation) diverge. Par contre, la longueur moyenne d'une boucle prise au hasard reste finie [86] : elle vaut en fait 18. Exactement à $n = 2$ l'énergie libre a une singularité essentielle [9], ce qui signale une transition de phase.

Pour $-2 \leq n \leq 2$, le modèle est critique. Ses exposants critiques, différents de ceux des phases dense et diluée discutées ci-dessus, furent d'abord trouvés de manière numérique [28] avant qu'une solution exacte ne fût établie [7]. L'ensemble de ces exposants, ainsi qu'un exposant nouveau qui décrit l'influence d'une température infinitésimale, fut ensuite retrouvé par Kondev et collaborateurs par une approche très différente [112]. Bien que non rigoureuse, cette approche est intéressante pour plusieurs raisons :

- Elle construit explicitement la théorie conforme qui décrit les fluctuations critiques des boucles. En particulier, les objets géométriques seront associés à des opérateurs dans une théorie de champs dite théorie de Liouville.
- Elle permet de résoudre d'autres modèles où la technique de l'ansatz de Bethe ne convient pas ou est très difficile à appliquer.

Par la suite, nous allons illustrer les ingrédients de cette approche appliquée à la phase compacte du modèle $\mathrm{O}(n)$ sur le réseau hexagonal.

5.2 Théorie de Liouville

Nous allons suivre en partie la présentation de [112] mais avec l'ajout de quelques simplifications postérieures [114, 83].

La construction de la théorie de Liouville passe par un certain nombre de transformations exactes ainsi que quelques suppositions de nature physique qui permettront de faire le contact avec la théorie de champs.

5.2.1 Modèle d'interface

L'idée de base est de décrire les fluctuations critiques des boucles au travers de celles d'un modèle d'interface associé. Ce modèle fait intervenir une variable de hauteur définie sur les faces du réseau où vivent les boucles. La variation de la hauteur entre deux faces voisines est déterminée par l'état de l'arête qui les sépare. Par contre, la valeur absolue des hauteurs n'a aucun sens pour l'instant : il y a une symétrie sous une translation globale de l'interface.

Nous avons vu que chaque configuration de L boucles est compatible avec 2^L tricoloriages. Plus précisement, la spécification d'un tricoloriage permet non seulement de définir les boucles mais aussi d'attribuer une orientation à chacune d'entre elles. Cette possibilité sera pleinement exploitée dans le paragraphe suivant : à présent, nous en tirons la conséquence que les états de l'interface (modulo une translation globale) doivent être en correspondance biunivoque avec les tricoloriages.

Soit $\mathbf{A}$, $\mathbf{B}$ et $\mathbf{C}$ les trois états (couleurs) possibles d'une arête. Pour que la variable de hauteur soit bien définie quand on entoure un sommet, il faut que la somme des différences de hauteur s'annule :

$$\mathbf{A} + \mathbf{B} + \mathbf{C} = \mathbf{0}. \tag{5.5}$$

L'interprétation la plus libérale qui respecte cette unique relation parmi trois quantités consiste en ce qu'on permet aux hauteurs de prendre des valeurs vectorielles dans un espace bidimensionnel. Plus généralement, pour satisfaire une relation du type (5.5) parmi Q couleurs différents, tout en traitant

les couleurs différentes de manière symétrique, les vecteurs associés doivent former un hypertétraèdre en $Q-1$ dimensions : dans le cas $Q=3$ discuté ici, il s'agit tout simplement d'un triangle équilatéral.

Il est utile de comparer cette construction avec celle qu'on trouverait pour le modèle $\mathrm{O}(n)$ usuel. Soit $\mathbf{A}$ et $\mathbf{B}$ les deux couleurs choisies qui alternent le long de la trajectoire d'une boucle. Alors, le modèle $\mathrm{O}(n)$ à température non nulle comprend des sommets dont les trois arêtes incidentes sont toutes dans l'état $\mathbf{C}$, avec un poids $t \neq 0$. Ceci donne la contrainte supplémentaire $3\mathbf{C}=0$, ou bien $\mathbf{A}+\mathbf{B}=\mathbf{0}$: on peut alors se contenter d'une hauteur scalaire. Cette observation permet d'expliquer de façon intuitive la différence des classes d'universalité des deux modèles.

5.2.2 Redistribution locale des poids

Le prochain pas est de redistribuer le poids n par boucle de manière locale : cette redistribution est nécessaire afin de définir une théorie de champs locale.

Nous allons décomposer le poids n de la manière suivante [134]

$$n = 2\cos(\pi e_0) = \exp(\mathrm{i}\pi e_0) + \exp(-\mathrm{i}\pi e_0). \tag{5.6}$$

La bijection entre les tricoloriages et les configurations de boucles orientées permet alors d'attribuer un poids $\exp(\mathrm{i}\pi e_0)$ à une boucle chirale (orientée contrairement au sens des aiguilles d'une montre) et un poids $\exp(-\mathrm{i}\pi e_0)$ à une boucle antichirale. Comme, en plus, une boucle chirale sur le réseau hexagonal tourne nécessairement six fois plus à gauche qu'à droite, ces poids peuvent être redistribués localement en attribuant un poids $\exp(\mathrm{i}\pi e_0/6)$ à chaque sommet où une boucle tourne à gauche et un poids $\exp(-\mathrm{i}\pi e_0/6)$ à chaque sommet où une boucle tourne à droite. Soit $\lambda(\mathbf{x})$ la valeur de ce poids en un sommet situé au point $\mathbf{x}$. La fonction de partition prend alors la forme locale

$$Z = \sum_{\mathcal{G}'} \prod_{\mathbf{x}} \lambda(\mathbf{x}), \tag{5.7}$$

où $\mathcal{G}'$ représente les configurations de boucles compactes orientées.

C'est le modèle de vertex correspondant qui forme le point de départ pour l'approche d'ansatz de Bethe [7].

5.2.3 Limite continue

Le passage à la limite continue du modèle d'interface passe par une construction particulière qui a été expliquée en [111] et élaborée en [112, 83, 99]. Par la suite, nous allons en évoquer les points essentiels.

Le changement minimal d'un tricoloriage donné consiste en la permutation de deux couleurs (pas nécessairement $\mathbf{A}$ et $\mathbf{B}$) le long d'une trajectoire fermée le long de laquelle les deux alternent. Les coloriages qui permettent un maximum de tels changements sont caractérisés par le fait que toutes les boucles de deux couleurs alternées ont la longueur minimale de six arêtes : dans le modèle d'interface, les états correspondants sont macroscopiquement plats. Par une raisonnement entropique, on s'attend alors à ce que l'interface fasse des fluctuations bornées autour de domaines d'états plats.

Deux états plats peuvent être liés par une transformation qui permute les couleurs des arêtes autour d'un tiers des faces. Il est naturel de considérer deux états ainsi liés comme des voisins dans un espace vectoriel où chaque état plat est repéré par un point qui correspond à sa hauteur moyenne. En faisant subir à un état plat donné un certain nombre de transformations de permutation des couleurs il est possible de revenir au même état plat, mais situé en un autre point dans l'espace des hauteurs. Les translations correspondantes constituent un réseau $\mathcal{R}$ appelé *réseau de répétition*.

Dans cette construction, l'interface n'est plus invariante par une translation quelconque, mais seulement par des translations qui appartiennent à $\mathcal{R}$. Dans la limite continue, la hauteur $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^2$ doit alors être compactifiée par rapport à l'espace $\mathbb{R}^2/\mathcal{R}$. Comme cet espace n'est pas simplement connexe, la construction du gaz de Coulomb associée au modèle d'interface induit la possibilité d'avoir des défauts topologiques de type tourbillon avec une charge $\mathbf{m}$ (dite charge *magnétique*) qui prend ses valeurs en $\mathcal{R}$. Inversement, des charges $\mathbf{e} \in \mathcal{R}^*$ dites *électriques* seront associées aux opérateurs de vertex (aussi appelés ondes de spins) $\exp(\mathbf{ie}\cdot\mathbf{h})$.

5.2.4 Théorie de champs

Dans la limite continue, il est possible de définir une fonction de partition modifiée $Z_>$ qui décrit correctement les fluctuations de longue portée présentes dans Z. Par la théorie de champs, $Z_>$ est relié à une intégrale fonctionnelle

$$Z_> = \int \mathcal{D}\mathbf{h}\, \exp(-S[\mathbf{h}]), \tag{5.8}$$

où S est l'action euclidienne d'une théorie de Liouville aux couplages imaginaires. L'action contient trois termes

$$S = S_\mathrm{E} + S_\mathrm{B} + S_\mathrm{L} \tag{5.9}$$

dits respectivement le terme d'élasticité, le terme au bord et le terme de Liouville.

- Le terme d'élasticité S_E garantie la consistance du modèle d'interface car il limite les fluctuations entropiques autour des états plats. Il prend la forme

$$S_\mathrm{E} = \frac{1}{2}\int \mathrm{d}^2\mathbf{x}\, K_{\alpha\beta}\boldsymbol{\partial}h^\alpha \cdot \boldsymbol{\partial}h^\beta, \tag{5.10}$$

  où $\boldsymbol{\partial} = (\partial_1, \partial_2)$ dénote le gradient. L'objet $\mathtt{K} = K_{\alpha\beta}$ est un tenseur d'élasticité agissant sur les composantes des hauteurs. En général, les symétries du modèle entraîne l'annulation d'un certain nombre de composantes de $K_{\alpha\beta}$: ici, c'est un tenseur diagonal 2×2.
- Le terme au bord S_B attribue un poids n aux boucles qui sentent le bord du système. Le plus aisé pour l'application de la théorie conforme est de définir le modèle sur un cylindre : les boucles concernées par S_B seront alors celles qui possèdent un enroulement par rapport au cylindre. Pour de telles boucles, la redistribution locale des poids expliqueé ci-dessus ne marche pas, car leurs nombres de virages à droite et à gauche sont égaux. Ce problème est résolu en mettant

$$S_\mathrm{B} = \mathrm{i}\int \mathrm{d}^2\mathbf{x}\, \mathbf{e}_0 \cdot [\mathbf{h}(+\infty) - \mathbf{h}(-\infty)], \tag{5.11}$$

  c'est-à-dire en couplant la différence de hauteur entre les deux extrémités du cylindre à une *charge de vide* $\mathbf{e}_0$. Autrement dit, S_B correspond à l'insertion de deux opérateurs de vertex $\exp(\pm\mathrm{i}\mathbf{e}_0 \cdot \mathbf{h})$ aux bords du système. La charge $\mathbf{e}_0$ est choisie de manière à donner aux boucles enroulantes le poids désiré, n.
- Le terme de Liouville S_L implante les poids locaux $\lambda(\mathbf{x})$ de (5.7). Par l'analyse de Fourier, il peut être développé sur des opérateurs de vertex dont on ne retient que les plus pertinents (dans le sens d'une théorie conforme). Comme le poids d'une boucle est invariant par rapport à la renormalisation, il faut exiger la *marginalité* de ces opérateurs [114] : ils servent alors à écranter le gaz de Coulomb [46, 47]. Cette partie de la construction est cruciale, car l'ansatz de marginalité permet de relier les constantes élastiques $K_{\alpha\beta}$ au poids n.

La théorie de Liouville (5.9) est une théorie conforme exactement soluble. En particulier, les exposants critiques $x(\mathbf{e}, \mathbf{m})$ des opérateurs ayant une charge électromagnétique $(\mathbf{e}, \mathbf{m})$ ainsi que la charge centrale c sont exactement connus [46, 47] :

$$x(\mathbf{e}, \mathbf{m}) = \frac{1}{4\pi}\left[\mathbf{e}\cdot\mathtt{K}^{-1}\cdot(\mathbf{e} - 2\mathbf{e}_0) + \mathbf{m}\cdot\mathtt{K}\cdot\mathbf{m}\right], \tag{5.12}$$

$$c = \dim(\mathbf{h}) + 12x(\mathbf{e}_0, \mathbf{0}). \tag{5.13}$$

Remarquez en particulier la dépendance de la dimensionalité de l'espace de hauteurs : cette dépendance traduit, dans le langage de la théorie de champs, le fait que la phase compacte possède plus de degrés de liberté que les phases dense et diluée.

5.2.5 Exposants critiques

La formule (5.12) permet de calculer un ensemble d'exposants critiques liés aux propriétés conformationnelles des boucles.

En particulier, il est fructueux de considérer l'insertion de deux défauts topologiques de charge magnétique $\pm\mathbf{m}$ dans le modèle d'interface. Par exemple, le choix $\mathbf{m} = 3\mathbf{C}$ permet d'obtenir l'exposant

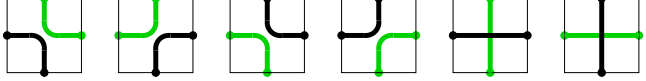


FIG. 5.2 – Les six configurations des boucles autour d'un sommet dans le modèle de boucles compactes sur le réseau carré.

thermique $x_t = x(\mathbf{0}, 3\mathbf{C})$ qui caractérise l'effet d'une température infinitésimale sur la phase compacte [112].

D'autres choix de $\mathbf{m}$ correspondent à une situation où un nombre k de cordes ouvertes se propagent au milieu des boucles, entre les deux points d'insertion. Par exemple, pour $n \to 0$, $x_1 = x(\mathbf{e}_0, \mathbf{C} - \mathbf{B})$ donne le rapport entre le nombre de chemins et de cycles hamiltoniens sur le réseau hexagonal avec N sommets :

$$\frac{\#\text{chemins}}{\#\text{cycles}} \sim N^{\gamma}. \tag{5.14}$$

L'exposant $\gamma = 1 - x_1$ caractérise aussi l'interaction entropique entre les deux extrémités d'un chemin hamiltonien : l'interaction est répulsive si $\gamma > 1$ et attractive si $\gamma < 1$. La solution explicite [7, 112] montre que le réseau hexagonal fournit un cas marginal : $\gamma = 1$, ce qui veut dire qu'un chemin hamiltonien se comporte comme prédit par la théorie du champ moyen [138, 79].

5.3 Tétracoloriage du réseau carré

Il est possible d'adapter cette construction de boucles compactes au réseau carré [111]. Ce cas est particulièrement intéressant puisque des calculs numériques [8] demontrèrent que l'exposant γ est en désaccord avec la théorie du champ moyen [138, 79]. La solution exacte [115, 83] fournit ensuite le résultat $\gamma = 117/112$, en excellent accord avec les calculs numériques [8, 115].

Le modèle sur le réseau carré fait intervenir *deux* types de boucles qui peuvent se rencontrer à un sommet en six manières différentes : voir la Figure 5.2. Remarquez que les deux types de boucles peuvent se traverser : cette propriété sera cruciale pour la discussion des méandres lors du Chapitre 6.

La solution du modèle de boucles compactes sur le réseau carré fut donnée dans l'article [83], attaché à ce chapitre.

5.4 Autres réseaux

Vu que les boucles compactes, et en particulier les cycles hamiltoniens, ont des propriétés conformationnelles différentes sur les réseaux hexagonal et carré, on pourrait être tempté d'adapter la construction de la théorie de Liouville à d'autres réseaux. Pourtant, pour que le lien entre des configurations de coloriage et de boucles orientées puisse être fait, il faut que le réseau soit biparti et de valence constante. Notamment, les réseaux bipartis ont la propriété que tous leurs cycles sont de longueur paire [158], ce qui permet de réaliser une boucle comme une trajectoire de deux couleurs alternées.

Ces critères ne sont satisfait que par un nombre de réseaux réguliers (techniquement : de type archimèdéen ou k-uniforme) très restreint [72]. Parmi ceux-ci on trouve le réseau (4.8^2) qui est un pavage régulier du plan par des carrés et des octagones. Or, une étude détaillee montre que le modèle O(n) défini sur ce réseau n'a pas de point fixe à température nulle [85].

Une autre possibilité est d'étudier des modèles de boucles compactes sur des réseaux orientés. Dans un tel modèle, les arêtes du réseau sont équippées d'une flèche de direction fixe. Seules sont autorisées les boucles le long desquelles l'orientation de toutes les flèches est constante. Par exemple, le modèle de boucles compactes sur le réseau Manhattan (qui est un réseau carré avec une orientation particulière [102]) est équivalent au modèle de Potts sur un réseau associé [53] : ce réseau est simplement le réseau carré non orienté. Le résultat pour les chemins hamiltoniens (limite $n \to 0$) met en évidence un comportement du type champ moyen [138, 79] : $\gamma = 1$.

Le réseau Manhattan a aussi été étudié dans le cas qui présente deux types de boucles, comme indiqué sur la Figure 5.2 [68]. Ce modèle est équivalent à deux modèles de Potts définis respectivement sur un réseau carré et sur son dual. Les deux modèles sont couplés par une interaction de quatre spins et

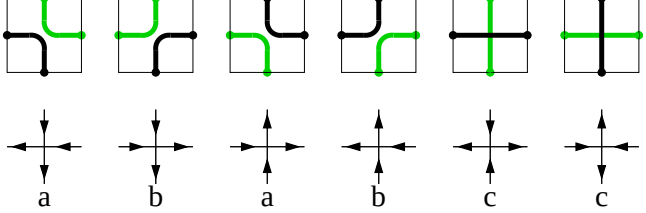


FIG. 5.3 – Transformation entre le modèle de deux types de boucles compactes et le modèle à six sommets sur le réseau carré. Sur le sous-réseau impair il faut invertir l'orientation des flèches par rapport à la figure.

possède un point intégrable avec $c = 1$ quand la fugacité de toutes les boucles vaut $\sqrt{2}$. Des simulations numériques indiquent que le couplage de quatre spins est non pertinent quand $c \leq 1$: on trouve alors le comportement de deux modèles O(n) découplés, chaque modèle étant dans la phase dense. Par contre, dans le domaine de fugacités qui aurait $c > 1$ le couplage est pertinent et le modèle devient massif [68].

5.5 Modèle de Flory

Le modèle de deux types de boucles compactes sur le réseau carré est équivalent au modèle à six sommets quand la fugacité de toutes les boucles vaut un : voir la Figure 5.3. Il devient alors naturel de définir un modèle plus général où, en plus des fugacités des boucles, l'on met des poids locaux a, b et c comme indiqué sur la figure. Pour le cas isotrope on peut mettre $a = b = 1$ sans perte de généralité : pour éviter la confusion avec la charge centrale, nous allons désormais noter $w_{\mathrm{X}} = c$.

Ce modèle n'est rien d'autre que le modèle de Flory pour la fusion d'une protéine [60, 148]. La modélisation d'une protéine par un polymère compact est justifiée par le fait que les acides aminés d'une protéine sont, soit hydrophobes, soit polaires : Après mise en contact de la protéine avec de l'eau dans une cellule, cette protéine forme alors des structures compactes afin d'éviter le contact de l'eau avec ses parties hydrophobes. Dans le modèle de Flory, la rigidité de la protéine est représentée sur le réseau carré par un poids w_{X} à chaque sommet où deux monomères consécutifs sont alignés.

On s'attend à ce qu'il y ait une transition de phase entre le domaine de grand w_{X} (basse température), où la protéine cherche à s'aligner le plus possible en formant de la structure secondaire, et le domaine de petit $w_{\mathrm{X}} \approx 1$ (haute température), où la protéine est dans une phase fondue : les fluctuations thermiques détruisent alors la structure secondaire.

La théorie de Liouville de [83] peut être généralisée afin d'étudier l'effet d'un poids $w_{\mathrm{X}} \neq 1$. En particulier, il est possible de calculer les exposants critiques de manière exacte au point de fusion [99]. On trouve $\gamma = 11/16$, ce qui signifie une attraction assez forte entre les extrémités du polymère : c'est effectivement le mécanisme physique prévu par analogie avec l'effondrement d'un polymère au point theta.

Toute la phase de haute température est critique, avec des exposants critiques qui dépendent de la température de manière non universelle. Cette dépendance peut être étudiée numériquement avec une grande précision [99].

D'autres généralisations du modèle permet d'étudier la transition vers la phase non compacte ainsi que le lien avec une généralisation du modèle à huit sommets. Nous reportons l'étude de ces sujets à l'article [99] qui est attaché à ce chapitre.

5.6 Questions de classification

Au niveau formel, la structure des vecteurs d'écran du modèle de deux types de boucles compactes [83] présente des similitudes frappantes avec la théorie conforme étendue, basée sur l'algèbre de Lie WA_3. On s'interroge alors pour savoir si le modèle de boucles possède une symétrie étendue, ce qui permettrait éventuellement une étude systématique de son contenu d'opérateurs. De tels résultats seraient très utiles, voir indispensables, afin de coupler ce modèle de boucles au désordre. Le modèle désordonné serait un bon candidat pour la résolution du modèle dit HP qui décrit la physique d'un hétéropolymère compact.

Dans [50], le dernier article attaché à ce chapitre, nous montrons que le modèle de boucles rentre en effet dans une classification plus générale des gaz de Coulomb que celle basée sur les algèbres de Lie.

5.7 Article "Field theory of compact polymers on the square lattice"

A partir d'un modèle de deux types de boucles compactes sur le réseau carré, nous dérivons des résultats exacts pour la statistique conformationnelle des polymères compacts. Ce modèle de boucles possède une *variété bidimensionnelle* de points fixes critiques, chaque point étant caractérisé par une infinité d'exposants critiques géométriques. Nous calculons ces exposants *exactement* en transformant le modèle de boucles en un modèle d'interface, dont la limite thermodynamique est une théorie de champs de type Liouville. Les formules pour la charge centrale et les premiers exposants critiques sont comparées à des résultats numériques de matrice de transfert : l'accord est excellent. Les polymères compacts sont identifiés par un point particulier dans le diagramme de phase du modèle de boucles. La valeur de l'exposant conformationnel $\gamma = 117/112$, calculée ici pour la première fois, est en contradiction avec la théorie de champ moyen. Des polymères compacts en interaction sont décrits par une *ligne* de points fixes le long de laquelle γ varie continûment.

ELSEVIER

Nuclear Physics B 532 [FS] (1998) 635-688

NUCLEAR PHYSICS B

Field theory of compact polymers on the square lattice

Jesper Lykke Jacobsen [a,b,1], Jané Kondev [c,d,2]

[a] *Somerville College and Department of Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK*

[b] *Institute of Physics and Astronomy, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark*

[c] *Institute for Advanced Study, Olden Lane, Princeton, NJ 08540,USA*

[d] *Department of Physics, Princeton University, Princeton, NJ 08540, USA*

Received 6 May 1998; accepted 26 June 1998

Abstract

Exact results for conformational statistics of compact polymers are derived from the two-flavour fully packed loop model on the square lattice. This loop model exhibits a *two-dimensional manifold* of critical fixed points each one characterised by an infinite set of geometrical scaling dimensions. We calculate these dimensions *exactly* by mapping the loop model to an interface model whose scaling limit is described by a Liouville field theory. The formulae for the central charge and the first few scaling dimensions are compared to numerical transfer matrix results and excellent agreement is found. Compact polymers are identified with a particular point in the phase diagram of the loop model, and the non-mean field value of the conformational exponent $\gamma = 117/112$ is calculated for the first time. Interacting compact polymers are described by a *line* of fixed points along which γ varies continuously. © 1998 Elsevier Science B.V.

PACS: 05.50.+q; 11.25.Hf; 64.60.Ak; 64.60.Fr

1. Introduction

Lattice models of loops have emerged as an important paradigm in two-dimensional critical phenomena. They allow for a determination of the scaling properties of different

[1] Present address: Laboratoire de Physique Statistique, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05, France. E-mail: lykke@lps.ens.fr; lykke@dfi.aau.dk.

[2] E-mail: janek@ias.edu

PII S0550-3213(98)00571-9

types of random walks which are used to model conformations of different phases of polymers [1]. For instance, the solution of the O(n) loop model has led to exact results for conformational exponents of swollen and dense polymers [2], as well as polymers at the theta point [3]. The theta point is the tricritical point which governs the transition between the swollen and the collapsed phase of polymers in solution [1]. Examples of conformational exponents are γ, which describes the scaling of the number of polymer conformations with the number of monomers $\mathcal{N}$, and ν, for the scaling of the linear size of a polymer, as measured by the radius of gyration, with $\mathcal{N}$. Here we calculate for the first time the exact value of γ for polymers on the square lattice, in the *compact phase*. Compact polymers completely fill the lattice and are of direct relevance to statistical studies of protein folding [4,5].

Further motivation for studying loop models comes from the Fortuin-Kasteleyn construction which maps many discrete spin models (e.g., Q-state Potts) to random cluster models. Since cluster boundaries in two dimensions form loops this naturally leads to a loop model representation. This random geometrical description of two-dimensional lattice models then provides a setting in which a general theory of their scaling limits can be sought. It is one of the goals of this paper to outline a specific proposal for such a theory in the form of an effective field theory of fluctuating loops. This field theory is constructed following the Coulomb gas recipe [6] with some important new ingredients added [7]. It describes the fluctuations of a random surface for which the loops are contour lines.

Scaling limits of many (but not all) two-dimensional lattice models are described by conformally invariant field theories [8,9]. This observation has led to exact results for critical exponents and other universal quantities, and to a classification of critical points based on their symmetry properties with respect to the group of conformal transformations. An obvious question which is often difficult to answer is: given a particular lattice model how does one *construct* the conformal field theory of its scaling limit? Loop models provide examples for which the scaling limit can be constructed in a *physically* transparent way. This is accomplished by mapping a loop model to an interface model, where the loops are simply equal-height contours. An explicit coarse graining procedure is then implemented for the height model, and it leads to a well-known conformal field theory – the *Liouville field theory*.

Interesting examples of loop models are also provided by one-dimensional quantum models, spin chains in particular, where loops appear as world lines of the spin. This mapping of spins to loops has recently been used to formulate very efficient numerical schemes for simulating spin chains and ladders. These *loop algorithms* allow one to simulate much bigger system sizes and lower temperatures than by using more traditional algorithms with local updates [10]. The loop representation of quantum spin chains also gives an illuminating stochastic-geometrical view of their quantum fluctuations [11]. For example, the spin-spin correlation function is related to the probability that two points on the space-time lattice belong to the same loop. This insight might lead to a *practical* theory of plateau transitions in the Integer Quantum Hall Effect, i.e. one that would allow for a calculation of the correlation length exponent and other universal quantities

which have been measured in experiments. Namely, the Chalker–Coddington network model [12], which is believed to be in the same universality class as the plateau transitions, was recently mapped to an $SU(n \to 0)$ quantum spin chain [13]. It remains to be seen if this spin chain has a tractable loop-model representation.

In the bigger picture, loop models are of interest as simple examples where the fundamental constituents are non-local, extended objects as opposed to point-like objects such as particles and spins. Fluctuating geometries of this sort are used to model flux lines in superconductors, domain walls in magnets, and crystalline interfaces, to name a few experimentally relevant systems.

The extended nature of loops turns out to have profound consequences when one attempts to write down an effective continuum description of these models, say, following Landau's dictum of expanding the free energy (Euclidean action) in powers of the order parameter and its derivatives. Namely, terms which are geometrical in origin and non-perturbative in nature, and hence cannot be inferred from symmetry arguments alone, appear in the action. On the other hand, exactly *because* these geometrical terms are present the values of the *effective* coupling constants of the field theory are completely determined, a rather remarkable occurrence.

Usually in an effective description provided by a field theory, coupling constants are phenomenological parameters fixed by auxiliary information about observable quantities, such as the response functions or the related correlation functions. The Coulomb gas approach to two-dimensional critical phenomena is an example of an effective theory wherein the electromagnetic coupling constant (i.e. the "magnitude of the unit charge") is determined from an exact solution of the model; typically it suffices to calculate the exact value of a single critical exponent. Our construction of an effective field theory of loop models closely parallels the Coulomb gas method with the important difference that the coupling constants are determined without recourse to any exact information about the model. For the model at hand no such information is available anyway, and moreover there are indications that the model is not exactly solvable [14]. On one level our theory can be viewed as a trick that allows one to calculate critical exponents in two-dimensional loop models without doing the "hard work" of exactly solving the model. On a deeper level it shows that lattice models of loops lead to continuum theories that are *geometrical* in nature, i.e. devoid of any couplings that depend on the microscopic details.

Here we study in detail the two-flavour fully packed loop (FPL^2) model on the square lattice. This is a statistical model which describes two flavours of loops that occupy the bonds of the square lattice, subject to certain close packing constraints to which we shall return shortly. The phase diagram of this model is described by two variables, n_b and n_g, which are the loop fugacities of the two flavours; see Fig. 1. The phase diagram of the FPL^2 model has three important features that we wish to emphasize from the outset:

(i) For loop fugacities that fall into the region $0 \leqslant n_\mathrm{b}, n_\mathrm{g} \leqslant 2$ of the phase diagram the model is critical, i.e. it exhibits a power-law distribution of loop sizes. The novel feature is that every point in the critical region defines a *different* universality class characterised by an infinite set of geometrical critical exponents. All previously studied

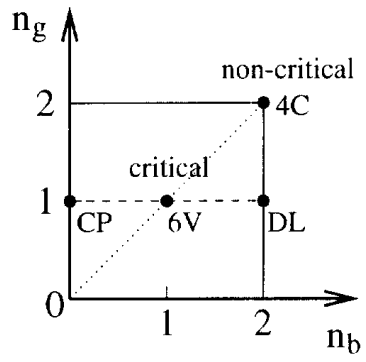


Fig. 1. Phase diagram of the two-flavour fully packed loop model on the square lattice. The loop model is critical for loop fugacities $0 \leqslant n_b, n_g \leqslant 2$. Particular points in the critical phase map to previously studied models: 6V – equal weighted six-vertex model [16], DL – dimer loop model [15], 4C – four-colouring model [28]. The dashed line is the fully packed loop model studied numerically in Ref. [14]; the point CP along this line corresponds to the problem of compact polymers. Finally, the dotted line is the loop model for which an effective Liouville field theory was constructed in Ref. [7].

loop models (e.g., Q-state Potts, O(n) models) exhibit a *line* of fixed points.

(ii) The effective field theory of the FPL^2 model in the critical region describes a fluctuating two-dimensional interface in five dimensions, which is characterised by *three* elastic constants. We calculate these three couplings exactly as a function of the two-loop fugacities. It is important to note that all previously solved loop models are characterised by a single elastic constant.

(iii) From the field theory of the FPL^2 model we calculate for the first time *exact* results for the conformational exponents of compact polymers on the square lattice. Furthermore, a particular line of fixed points in the phase diagram of the FPL^2 model can be identified with *interacting* compact polymers ($n_b = 0, n_g \leqslant 2$). We find that along this line the exponent γ changes continuously, whilst ν stays constant.

The organisation of the paper is as follows. In Section 2 we review the scaling theory of compact polymers which provides our main motivation for introducing the two-flavour fully packed loop model on the square lattice in Section 3. The rest of the paper is devoted to the study of this model using field theoretical techniques and numerical transfer matrix calculations.

The FPL^2 model is mapped to an interface model in Section 4. For the interface model we construct the scaling limit in terms of a Liouville field theory, in Section 5. In Sections 6 and 7 we make use of the field theory to calculate the central charge and the infinite set of geometrical exponents associated with loops, in the critical region of the loop model. A short description of the non-critical region based on the field theory is given next in Section 8.

Following the field theoretical treatment of the FPL^2 model, in Sections 9 and 10 we describe the construction of transfer matrices for different boundary conditions. They are used to determine the central charge, the first few geometrical exponents, and the residual entropy; the numerical results are in excellent agreement with the theoretical predictions. Finally, in Section 11, we present some general observations regarding compact polymers

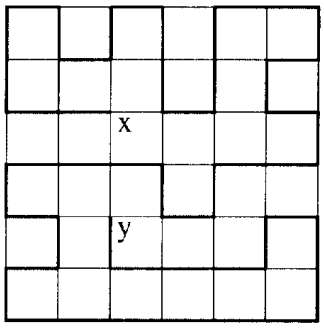


Fig. 2. Compact polymer on the square lattice; x and y are the positions of the chain ends.

and the Coulomb gas description of conformal field theories. We also comment on the dimer-loop model [15] and the three-state Potts antiferromagnet [16], in light of our solution of the fully packed loop model on the square lattice. The appendices are reserved for the calculation of scaling dimensions of operators in the Liouville field theory and the enumeration of connectivities which are used for constructing the transfer matrices.

2. Compact polymers

Compact polymers, or Hamiltonian walks, are self-avoiding random walks that visit *all* the sites of the underlying lattice; see Fig. 2. They have been used as simple models of polymer melts [17] and appear in statistical studies of protein folding [4,5]. Unlike dilute and dense polymers whose scaling properties were calculated exactly from the O(n) loop model [18], compact polymers defied a similar treatment until recently. Numerical transfer matrix calculations [19], a Bethe-ansatz solution [20], and a Coulomb gas theory [21] of the fully packed loop model on the *honeycomb* lattice, all conclude that compact polymers define a new universality class of critical behaviour. Here we study compact polymers on the *square* lattice. We calculate exact scaling exponents and find them to be distinct from the honeycomb case. This was first reported in Ref. [14] on the basis of numerical transfer matrix results.

The lattice dependence of critical properties distinguishes the compact polymer problem from its dilute and dense counterparts in a crucial way. It places them into the class of geometrically frustrated critical systems.[3] A physically relevant measure of frustration for compact polymers is the number of *contacts* per monomer. Contacts are realised by monomer pairs where the two monomers are nearest neighbours on the lattice but are not adjacent along the polymer chain. In lattice models of proteins hydrophobic interactions among the amino acids occur at contacts [4,5]. For the square model studied here the number of contacts per monomer is *two*, whilst on the honeycomb lattice it is *one*.

[3] Another example is the antiferromagnetic three-state Potts model which has a zero-temperature critical point on the square [16] and the Kagomé [22] lattices characterised by different critical exponents.

In order to study the scaling properties of compact polymers we focus our attention on the two most widely studied conformational exponents ν and γ. If R is the radius of gyration of the polymer then

$$R \sim \mathcal{N}^{\nu}, \tag{2.1}$$

where $\mathcal{N}$ is the number of monomers. Since compact polymers visit all the sites of a lattice, they are space-filling and we conclude that $\nu = 1/2$. This simple result will serve as an important check of our field theoretical calculations where it will be recovered.

In order to define the conformational exponent γ we introduce $C(\mathcal{N})$, the number of compact polymers (Hamiltonian walks) on a square lattice with $\mathcal{N}$ sites. Since a compact polymer fills the lattice, boundary conditions (free, periodic, etc.) play an important rôle. Following Saleur and Duplantier [23], we define γ in a way that is insensitive to the boundaries. Namely, if we introduce the quantity $C_{\circ}(\mathcal{N})$, the number of compact-polymer *rings*, then we can expect

$$\frac{C(\mathcal{N})}{C_{\circ}(\mathcal{N})} \sim \mathcal{N}^{\gamma}, \tag{2.2}$$

where γ does not depend on the choice of boundary conditions. Therefore, in order to calculate γ we need to solve the hard combinatorial problem of counting the number of open and closed compact polymers on the square lattice. Following de Gennes we do this by mapping the counting problem to the calculation of a correlation function in a particular statistical model at the critical point.

Consider the quantity $Z(\mathbf{x},\mathbf{y};\mathcal{N})$, the number of compact polymer conformations that start at the vertex $\mathbf{x}$ of the $\sqrt{\mathcal{N}} \times \sqrt{\mathcal{N}}$ square lattice, and end at $\mathbf{y}$ (see Fig. 2); we consider the limit $1 \ll |\mathbf{x}-\mathbf{y}| \ll \sqrt{\mathcal{N}}$, where $\mathbf{x}$ and $\mathbf{y}$ are chosen far from the boundaries of the lattice. For this quantity we can write down the scaling form [23]

$$Z(\mathbf{x},\mathbf{y};\mathcal{N}) = C_{\circ}(\mathcal{N})\,|\mathbf{x}-\mathbf{y}|^{-2x_1} f\left(\frac{|\mathbf{x}-\mathbf{y}|}{\mathcal{N}^{1/2}}\right), \tag{2.3}$$

where $f(u)$ is a scaling function with the property $f(u) \to$ const. as $u \to 0$, and x_1 is a geometrical exponent related to γ. Integrating $Z(\mathbf{x},\mathbf{y};\mathcal{N})$ over all end-points $\mathbf{y}$ and comparing the result to Eq. (2.2), the scaling relation

$$\gamma = 1 - x_1 \tag{2.4}$$

follows.

To calculate the geometrical exponent x_1 we introduce the two-flavour fully packed loop model on the square lattice. The fact that we need *two* loop flavours follows from the simple observation that the bonds not covered by the compact polymer also form loops whose number is *unconstrained*. For the loop model we then construct an effective field theory in which $Z(\mathbf{x},\mathbf{y};\mathcal{N})$ becomes a two-point correlation function. The asymptotics of this function can be calculated exactly and we find $x_1 = -5/112$, from which

$$\gamma = 117/112 = 1.0446\ldots \tag{2.5}$$

follows. This is to be compared to the mean-field theory value $\gamma_{\rm MF} = 1$ [24], which is also the result obtained for compact polymers on the honeycomb lattice [20].

The conformational exponent γ was measured directly from enumerations of conformations of chains with lengths up to 30 in Ref. [4], and the value $\gamma = 1.01(5)$ was reported. More recently, from a numerical transfer matrix study of the fully packed loop model on the square lattice the geometrical exponent $x_1 = -0.0444(1)$ was determined [14], in excellent agreement with the exact result.

Another quantity of interest is the connective constant κ which determines the leading, exponential with system size scaling of the number of compact polymers [25]

$$C(\mathcal{N}) \sim \kappa^{\mathcal{N}} \kappa_{\rm s}^{\mathcal{N}^{(d-1)/d}} \mathcal{N}^{\gamma-1}. \tag{2.6}$$

Here $\kappa_{\rm s}$ is the surface connective constant; it appears due to the space-filling nature of compact polymers. Both the value $\kappa = 1.475(15)$ found in Ref. [4], and the estimate $\kappa \simeq 1.472$ obtained from transfer matrix calculations similar to ours [26], seem in favour of the mean-field result $\kappa_{\rm MF} = 4/e = 1.4715\ldots$ [24].[4] In Section 10.4 we report the very accurate numerical value

$$\kappa = 1.472801(10), \tag{2.7}$$

which shows that the connective constant for compact polymers also deviates slightly from the mean-field result.

For the remainder of the paper we elaborate on the calculation of γ for compact polymers, in the process unveiling an extremely rich phase diagram of the associated loop model. As remarked earlier, it contains a two-dimensional region of fixed points, which we characterise in detail by calculating the central charge and the geometrical exponents associated with loops for each point on the critical manifold.

3. Two-flavour loop model

The two-flavour fully packed loop model on the square lattice was introduced in Ref. [28] as the loop representation of the four-colouring model [29]. It is the natural generalisation of the fully packed loop model on the honeycomb lattice, which is the loop representation of the three-colouring model [21]. In general, a q-colouring model on a q-fold coordinated lattice is given by edge colourings of the lattice with q different colours; an edge colouring of a graph is one where no two bonds that share a common vertex are coloured equally. The colouring model is mapped to a loop model by choosing $[q/2]$ colour pairs; each pair defines strings of alternating colour that necessarily form loops (unless they terminate at the boundary). In this way we end up with a loop model with $[q/2]$ flavours of loops.

[4] Very recently the field theory of Ref. [24] has been improved [27] yielding, however, unchanged values for $\gamma_{\rm MF}$ and $\kappa_{\rm MF}$.

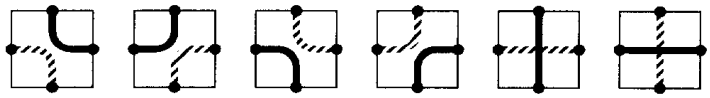


Fig. 3. The six vertex configurations of the FPL^2 model that are allowed by the fully packing constraint. Black and grey loop segments are shown here as solid and hatched lines respectively. Each vertex is adjacent to four edges, here shown as dots, that are referred to as "dangling" if they are not connected to an edge of a neighbouring vertex. Note that the two rightmost vertices explicitly permit the two flavours to cross.

To define the FPL^2 model we first specify the allowed loop configurations $\mathcal{G}$. In $\mathcal{G}$ every bond of the square lattice belongs to one and only one loop of either flavour, and loops of the same flavour are not allowed to cross. Representing the two flavours by solid (black) and hatched (grey) line segments respectively this fully packing constraint allows each vertex of the square lattice to have one of the six appearances depicted in Fig. 3. Each loop is assigned a fugacity depending on its flavour: n_{b} for black loops and n_{g} for grey loops. The partition function of the FPL^2 model is then

$$Z = \sum_{\mathcal{G}} n_{\mathrm{b}}^{N_{\mathrm{b}}} n_{\mathrm{g}}^{N_{\mathrm{g}}}. \tag{3.1}$$

The fully packed loop model of Batchelor et al. [14] is obtained by setting the loop fugacity of the grey loops to unity. In the limit $n_{\mathrm{b}} \to 0$ we recover the compact polymer problem.

If we define a restricted partition function of the FPL^2 model, to which only configurations with a single black loop segment propagating between points $\mathbf{x}$ and $\mathbf{y}$ contribute, then $Z(\mathbf{x},\mathbf{y};\mathcal{N})$ in Eq. (2.3) is obtained in the limit $n_{\mathrm{b}} \to 0$, $n_{\mathrm{g}} \to 1$. The first limit discards all configurations with black loops present, leaving only the black Hamiltonian walk (compact polymer) between $\mathbf{x}$ and $\mathbf{y}$, whilst the second ensures that all walks are weighted equally. We could also consider weighting different Hamiltonian walks differently by setting $n_{\mathrm{g}} \neq 1$. This situation can be interpreted as describing interacting compact polymers, and, as will be shown later, it leads to a continuously varying exponent γ. A similar property of interacting oriented polymers in the swollen phase was suggested by Cardy from a field theoretical calculation [30]. Recent numerical studies of the interacting oriented self-avoiding walk by Trovato and Seno [31], though, seem to be at odds with Cardy's prediction of an exponent γ that varies continuously with the interaction strength.

Some idea of the phase diagram of the FPL^2 model as a function of n_{b} and n_{g} can be obtained by examining the extreme limits of the loop fugacities. Namely, for $n_{\mathrm{b}}, n_{\mathrm{g}} \to \infty$ all loops have the minimum length of four, i.e. they each surround a single plaquette of the square lattice. There are no large loops in the system and the model is non-critical, or in other words, the average loop length is finite. On the other hand, in the critical phase of the loop model, which is the subject of this paper, in a typical configuration one finds loops of all sizes characterised by a power-law distribution. This leads to an average loop length which diverges with the system size. Such is the case in the other extreme limit of loop fugacities, $n_{\mathrm{b}}, n_{\mathrm{g}} \to 0$, when the loops cover the whole lattice.

Other previously studied models that are particular points in the phase diagram of the

FPL^2 model are the four-colouring model, the dimer loop model, and the equal weighted six-vertex model; see Fig. 1. For $(n_b, n_g) = (2,2)$ the loop fugacity of each loop can be evenly (1+1) distributed among the two ways of colouring the bonds occupied by the loop with two colours in an alternating fashion: **ABAB**... for black loops and **CDCD**... for grey loops. This is then the symmetric four-colouring model (**A**, **B**, **C**, and **D** are the colours) studied by Baxter [32]. In the dimer loop model black and white dimers are placed on the square lattice so that every vertex is covered by one of each [15]. If we identify the dimer covered bonds with the black loops then this model is mapped to the $(n_b, n_g) = (2,1)$ FPL^2 model. And finally $(n_b, n_g) = (1,1)$ constitutes the equal-weighted six-vertex model [33], the allowed vertices being those of Fig. 3.

4. Height representation

The critical phase of the FPL^2 model can be described in terms of an effective field theory, following the general procedure discussed in Ref. [6]. The idea is to think of loops as contours of a scalar field, which we refer to as the height. Depending on the loop model in question the height can have one or more components. If the number of components is $D_\perp$ then the effective field theory of the loop model describes a fluctuating two-dimensional interface in $D_\perp + 2$ dimensions.

To introduce the heights we first map the loop model to an *oriented* loop model, as shown in Fig. 4. The orientation of every loop is chosen randomly and independently. Every non-oriented loop configuration is thus transformed into an oriented one ($\mathcal{G}'$); the number of oriented configurations that correspond to the same non-oriented loop configuration is simply $2^{N_b+N_g}$.

Next, for each loop we redistribute its weight (fugacity), n_b or n_g depending on whether it is black or grey, between the two possible orientations. For the black loops we do this by assigning to, say, the clockwise orientation the *phase factor* $\exp(i\pi e_b)$, and the opposite phase, $\exp(-i\pi e_b)$, to a counter-clockwise oriented black loop. Similarly for grey loops the clockwise oriented ones are assigned a weight $\exp(i\pi e_g)$ whilst the counter-clockwise loops are weighted with $\exp(-i\pi e_g)$. The loop fugacities are related to the newly introduced parameters e_b and e_g by

$$\begin{aligned} n_b &= 2\cos(\pi e_b), \\ n_g &= 2\cos(\pi e_g), \end{aligned} \tag{4.1}$$

since the partition function of the original (non-oriented) model, as given by Eq. (3.1), must be recovered by independently summing over the two possible orientations for each loop. Note that for $0 \leqslant n_b, n_g \leqslant 2$ the parameters e_b and e_g are real, whilst for $n_b, n_g > 2$ they are purely imaginary. As discussed in more detail in Section 8 this is the crucial property that leads to a critical state of the loop model in the former and a non-critical one in the latter case.

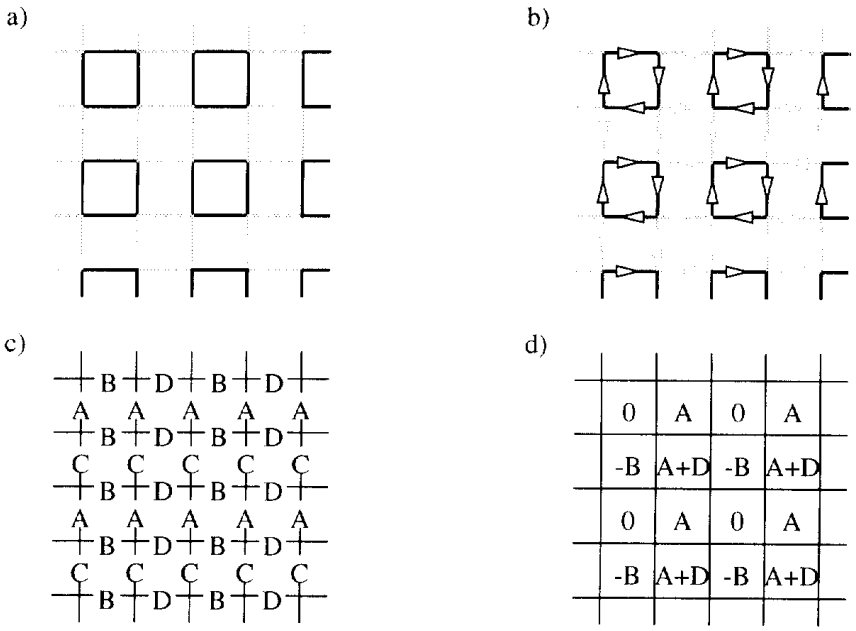


Fig. 4. Mapping of the FPL^2 model to an interface model. (a) $\rightarrow$ (b): Transform the loop configuration into an *oriented* loop configuration by choosing the orientation of each loop independently and randomly. (b) $\rightarrow$ (c): Every bond in the oriented loop configuration is in one of four states, depending on its flavour and direction; these four states are represented by three-vectors $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and $\mathbf{D}$. (c) $\rightarrow$ (d): The microscopic height $\mathbf{z}$ of the interface model changes from plaquette to neighbouring plaquette by $\mathbf{A}, \mathbf{B}, \mathbf{C}$, or $\mathbf{D}$ depending on the state of the bond between the two plaquettes. The change in $\mathbf{z}$ is positive going clockwise around even vertices and counterclockwise around odd ones.

Now that the loops are oriented we can interpret them as contours of a height field; the orientation is necessary as it determines the direction of increasing height. The systematic construction of the *microscopic heights* sets out from the observation that every bond of the square lattice is in one of four possible states: it can be coloured black or grey, and oriented from an even to an odd site, or from odd to even. "Even" and "odd" refer here to the two sublattices of the bipartite square lattice; every even site is surrounded by four nearest neighbouring odd sites, and *vice versa*.

The four possible bond-states are represented by four vectors – which are the colours in the four-colouring representation – $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$; see Fig. 4c. The microscopic heights $\{\mathbf{z}\}$ are defined on the dual lattice and the change in height when going from one plaquette centre to the next is given by $\mathbf{A}, \mathbf{B}, \mathbf{C}$ or $\mathbf{D}$, depending on the state of the bond which is crossed; see Fig. 4d. For the height to be uniquely defined the four vectors must satisfy the constraint $\mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D} = \mathbf{0}$. This means that the microscopic heights live in a *three-dimensional* vector space, which we take to be $\mathbb{Z}^3$. In other words, the oriented FPL^2 model maps to a model of a two-dimensional interface in five spatial dimensions.

By reasons of symmetry the four vectors are chosen so as to point from the centre to the vertices of a regular tetrahedron. With a suitable choice of coordinates they are represented by three-vectors:

$$\mathbf{A} = (-1,+1,+1),$$
$$\mathbf{B} = (+1,+1,-1),$$
$$\mathbf{C} = (-1,-1,-1),$$
$$\mathbf{D} = (+1,-1,+1). \tag{4.2}$$

This is the same normalisation as the one used in Ref. [28].

Mapping the loop model to on oriented loop model also allows for a *local* redistribution of the loop weights. This is important since it leads to a local field theory for the heights. As we shall find out shortly, though local, this field theory is somewhat unconventional due to the non-local, extended nature of the fundamental microscopic objects it purports to describe.

To redistribute the phase factors associated with oriented loops we assign a phase $\exp(-i\pi e_b/4)$ to a vertex of the square lattice if a black loop makes a left turn at that vertex, the opposite phase $\exp(+i\pi e_b/4)$ if it makes a right turn, and the weight 1 if it continues straight. The total vertex weight $\lambda(\mathbf{x})$ is a product of the phase factor originating from the black loop and an equivalent one from the grey loop passing through the same vertex $\mathbf{x}$. The partition function of the FPL^2 model, Eq. (3.1), can now be rewritten as a sum over oriented loop configurations (i.e. colouring configurations)

$$Z = \sum_{\mathcal{G}'} \prod_{\mathbf{x}} \lambda(\mathbf{x}). \tag{4.3}$$

Once the height at a single point is fixed $\mathcal{G}'$ is in a one-to-one correspondence with the configurations of the microscopic heights, and the summand in the above equation is the appropriate weight. In the critical phase of the FPL^2 model the interface described by Eq. (4.3) is *rough*, and the field theory is constructed so as to correctly reproduce its long-wavelength fluctuations.

4.1. Spectrum of electromagnetic charges

The mapping from oriented loop configurations, which are equivalent to edge colourings, to microscopic height configurations is one to many. In particular, two height configurations corresponding to the same edge colouring can have their heights shifted with respect to each other by a *global* shift $\mathbf{m} \in \mathcal{R}$. The set $\mathcal{R}$ forms a three-dimensional Bravais lattice, i.e. it is closed under integral linear combinations, and its elements are the *magnetic* charges in the Coulomb gas representation of the FPL^2 model. The lattice reciprocal to the lattice of magnetic charges, $\mathcal{R}^*$, defines the *electric* charges $\mathbf{e} \in \mathcal{R}^*$, with the property $\mathbf{e} \cdot \mathbf{m} = 2\pi m, m \in \mathbb{Z}$.

The construction of the lattice $\mathcal{R}$ for the FPL^2 model follows the usual prescription for height models, and has been carried out in detail in Ref. [28]. For the sake of completeness we outline this construction below.

It is convenient to first identify the *flat* states (also referred to as the *ideal* states), i.e. those colouring states which minimise the variance of the microscopic height $\mathbf{z}$. From the

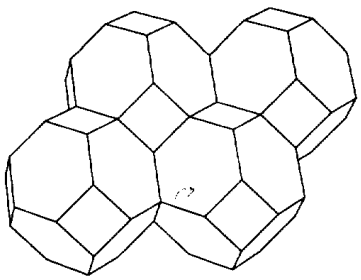


Fig. 5. The ideal state graph of the FPL^2 model in the four-colouring representation.

height mapping described above it follows that these states have all of their plaquettes coloured with two colours only; an example is shown in Fig. 4c. This leads to a colouring state that is periodic, with the same 2×2 colouring pattern repeated throughout the lattice. There are twenty four flat/ideal states for the colouring representation of the FPL^2 model, corresponding to the number of permutations of four different colours. Namely, an ideal state is completely specified by listing the colours of the bonds around a single site (say the origin), starting from the left horizontal bond and proceeding clockwise. To each flat state we assign a *coarse grained height* $\mathbf{h} = \langle \mathbf{z} \rangle$, which is the average microscopic height over a 2×2 unit cell of the colouring.

The flat states form a three-dimensional graph, which we refer to as the ideal state graph, $\mathcal{I}$. Namely, starting from any ideal state four other ideal states can be reached by exchanging a pair of colours that form a plaquette. For example, by exchanging the colours **A** and **B** in Fig. 4c all the **ABAB** plaquettes are turned into **BABA** plaquettes to give a new ideal state. Under these plaquette flips only the microscopic heights at the centres of the affected plaquettes are changed. In this way the ideal states form a four-fold coordinated graph in height space, where each vertex is indexed by a colour permutation, and its position in $\mathbb{R}^3$ is given by the coarse grained height $\mathbf{h}$. Bonds are associated with transpositions of two colours; they lie along the direction defined by the difference of the two colour vectors, and have a length of $1/\sqrt{2}$ if the normalisation in Eq. (4.2) is chosen.

The ideal state graph is a tiling of $\mathbb{R}^3$ with truncated octahedra; this regular polyhedron is better known as the Wigner–Seitz cell of a body-centred cubic (bcc) lattice (see Fig. 5). A single truncated octahedron in $\mathcal{I}$ has twenty four vertices corresponding to the twenty four different ideal states. The set of vertices in $\mathcal{I}$ representing the same ideal state form the *repeat lattice* $\mathcal{R}$, which is face-centred cubic (fcc) with a conventional cubic cell of side 4.

To obtain the continuum description of the FPL^2 model we coarse grain the microscopic height over domains of ideal states. This gives rise to the coarse grained height $\mathbf{h}$ which we can consider to be *compactified* on $\mathbb{R}^3/\mathcal{R}$. The phase space of the height is not simply connected, thus allowing for topological defects (vortices) with topological charges that take their values in $\mathcal{R}$ [34]. These defects are associated with *magnetic* charges in the Coulomb gas representation of the FPL^2 model. *Electric* charges on the other hand are associated with vertex operators $\exp(i\mathbf{e} \cdot \mathbf{h})$. If we take the height to live

in $\mathbb{R}^3/\mathcal{R}$ then vertex operators are well defined only for values of the electric charge $\mathbf{e} \in \mathcal{R}^*$. $\mathcal{R}^*$ is the lattice *reciprocal* to the lattice of magnetic charges $\mathcal{R}$, and it is a body-centred cubic (bcc) lattice with a conventional cubic cell of side π.

5. Construction of the field theory

An effective field theory of the FPL^2 model should describe large scale properties of loops. The kind of questions we expect it to answer are ones that do not refer to the microscopic details of the lattice model. For example, from the effective field theory we shall calculate the asymptotics of the probability that two points lie on the same loop, when the separation between the points is large compared to the lattice spacing. From this and related quantities the conformational exponents of compact polymers can be extracted.

The field theory of the FPL^2 model is defined by the Euclidean action for the coarse grained height $\mathbf{h}$. Consider a typical configuration of the oriented FPL^2 model which is equivalent to the colouring model. It consists of domains of ideal states. To each ideal state domain we assign a coarse grained height, defined earlier as the average microscopic height over the domain. In the continuum limit we assume that this height is a smoothly varying function of the basal plane coordinates (x^1, x^2). The partition function that takes into account only the large scale fluctuations of the height can be written as a functional integral,

$$Z_> = \int \mathcal{D}\mathbf{h} \exp(-S[\mathbf{h}]), \tag{5.1}$$

where S is the Euclidean action of a Liouville field theory with imaginary couplings [7]. The Liouville action contains three terms,

$$S = S_E + S_B + S_L. \tag{5.2}$$

Each one has a concrete geometrical interpretation in the FPL^2 model, which we describe next.

5.1. Elastic term

The first term in the effective action for the FPL^2 model describes the elastic fluctuations of the interface. It gives less weight to configurations that deviate from the flat states, by penalising finite gradients of the height. This term is entropic in origin. Namely, in order to change the colour of a particular bond in the four-colouring representation of the loop model, say $\mathbf{C} \to \mathbf{B}$, all the **C**'s and **B**'s have to be interchanged along the **CB** loop which contains the chosen bond. This transformation we call a *loop flip*; see Fig. 6. The ideal states *maximise* the number of loops of alternating colour and consequently they have the largest entropy of loop flips.

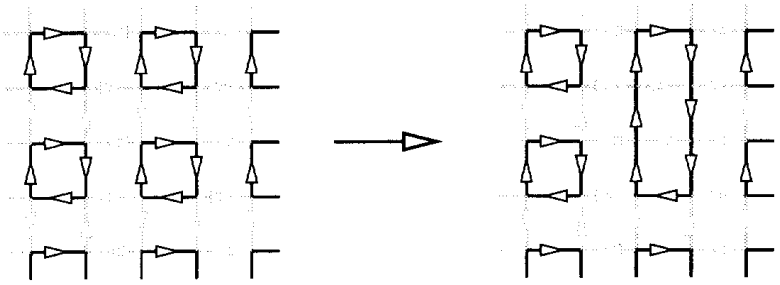


Fig. 6. A loop flip changes one oriented loop configuration into another. Here the bond states **C** and **B** are exchanged along a single **BC** plaquette (cf. Fig. 4c).

In its most general form the elastic term in the effective action can be written as a gradient expansion,

$$S_{\mathrm{E}} = \frac{1}{2} \int d^2\mathbf{x}\, K^{ij}_{\alpha\beta} \partial_i h^\alpha \partial_j h^\beta, \tag{5.3}$$

where higher powers of the height gradients and higher derivatives of the height are less relevant at large scales. The stiffness tensor $K^{ij}_{\alpha\beta}$ nominally has 36 components; the indices $i, j = 1, 2$ are for the basal plane coordinates, whilst $\alpha, \beta = 1, 2, 3$ label the three components of the height. Summation over repeated indices is assumed throughout.

The number of independent non-zero components of the stiffness tensor (i.e. elastic constants) is actually only *three*, once all the symmetries of the FPL^2 model are taken into account. The relevant symmetry transformations, that is the ones that become the symmetries of the effective action, are the ones that leave the weights of oriented loop configurations unchanged. First, there are the lattice symmetries, translations and rotations, which cut the number of independent elastic constants down to six. The terms that are allowed in S_{E} are scalars under rotations in the basal plane $\{(x^1, x^2)\}$, and they are necessarily of the form $\boldsymbol{\partial} h^\alpha \cdot \boldsymbol{\partial} h^\beta$, where $\boldsymbol{\partial} = (\partial_1, \partial_2)$ is the usual gradient. Second, the FPL^2 model possesses colour symmetries,

$$\mathbf{A} \leftrightarrow \mathbf{B} : e_{\mathrm{b}} \leftrightarrow -e_{\mathrm{b}} \text{ and } z_1 \leftrightarrow z_3 \tag{5.4}$$

and

$$\mathbf{C} \leftrightarrow \mathbf{D} : e_{\mathrm{g}} \leftrightarrow -e_{\mathrm{g}} \text{ and } z_1 \leftrightarrow -z_3, \tag{5.5}$$

which interchange the colours and at the same time transform the microscopic heights. Taking into account the colour symmetries the elastic contribution to the action takes on the form

$$S_{\mathrm{E}} = \frac{1}{2} \int d^2\mathbf{x} \left\{ K_{11} [(\boldsymbol{\partial} h^1)^2 + (\boldsymbol{\partial} h^3)^2] + 2K_{13} (\boldsymbol{\partial} h^1 \cdot \boldsymbol{\partial} h^3) + K_{22} (\boldsymbol{\partial} h^2)^2 \right\}. \tag{5.6}$$

Furthermore, by introducing a change of coordinates in height space,

$$H^1 = \tfrac{1}{2}(h^1 - h^3), \qquad H^2 = h^2, \qquad H^3 = \tfrac{1}{2}(h^1 + h^3) \tag{5.7}$$

S_E becomes diagonal,

$$S_E = \frac{1}{2}\int d^2\mathbf{x}\, g_\alpha (\partial H^\alpha)^2. \tag{5.8}$$

The three coupling constants g_α ($\alpha = 1,2,3$) are linearly related to the three elastic constants,

$$g_1 = 2(K_{11} - K_{13}), \qquad g_2 = K_{22}, \qquad g_3 = 2(K_{11} + K_{13}). \tag{5.9}$$

The appearance of *three* elastic constants is rather intriguing from the viewpoint of loop models that have been solved previously. The Q-state Potts, the O(n), and the honeycomb FPL models are all characterised by a *single* coupling constant, which has been determined case by case from their exact solutions. Below we shall show that all three couplings in Eq. (5.8) can be calculated exactly from the *loop ansatz* introduced in Ref. [7].[5] The ansatz states that the operator which enforces the complex weights assigned to oriented loops is *marginal* in the renormalisation group sense. This property of the field theory is intimately related to the random geometry of loops; we elaborate on this important point in Section 5.3.2.

5.2. Boundary term

The mapping of the loop model to an oriented loop model with local complex weights $\lambda(\mathbf{x})$ (Eq. (4.3)) fails for loops that experience the boundary. For example, if we define the FPL^2 model on a cylinder, loops that wind around the cylinder will not be weighted properly. The winding loop has an equal number of left and right turns and hence it will be assigned a weight one. Summing over the two orientations gives a weight two, and not the correct n_b or n_g, depending on the flavour. To correctly weight these loops one introduces a boundary term into the effective action,

$$S_B = \frac{i}{4\pi}\int d^2\mathbf{x}\, (\mathbf{e}_0 \cdot \mathbf{h})\widetilde{\mathcal{R}}; \tag{5.10}$$

$\widetilde{\mathcal{R}}$ is the scalar curvature and $\mathbf{e}_0$ is the *background* electric charge, which is to be determined. Since we are only concerned with the situation where the lattice on which the FPL^2 model is defined is flat, the scalar curvature vanishes everywhere except at the boundary.

To determine $\mathbf{e}_0$ we consider the FPL^2 model on the cylinder. The scalar curvature of the cylinder is proportional to the difference of two delta functions situated at the two far ends of the cylinder: $\widetilde{\mathcal{R}} = 4\pi[\delta(+\infty) - \delta(-\infty)]$. Therefore S_B has the effect of placing vertex operators $\exp(\pm i\mathbf{e}_0 \cdot \mathbf{h})$ at $x^2 = \pm\infty$; here x^2 is the coordinate along the length of the cylinder. These vertex operators assign an additional weight $\exp(i\mathbf{e}_0 \cdot (\mathbf{h}(+\infty) - \mathbf{h}(-\infty))$ to oriented loop configurations on the cylinder. Now, in

[5] The coupling constant g for all the loop models known to date can be calculated using this method, therefore dispensing with the need for an exact solution.

order for $\mathbf{h}(+\infty) - \mathbf{h}(-\infty)$ to be non-zero there must be at least a single winding loop present. If this winding loop is black, then the height difference is **A** or **B** depending on its orientation; similarly if the loop is grey the height difference is **C** or **D**. Furthermore, if the background charge is chosen so as to satisfy

$$\begin{aligned} \mathbf{e}_0 \cdot \mathbf{A} &= \pi e_\mathrm{b}, & \mathbf{e}_0 \cdot \mathbf{B} &= -\pi e_\mathrm{b}, \\ \mathbf{e}_0 \cdot \mathbf{C} &= \pi e_\mathrm{g}, & \mathbf{e}_0 \cdot \mathbf{D} &= -\pi e_\mathrm{g}, \end{aligned} \tag{5.11}$$

then the winding loops will be assigned their proper weights. This is again seen by summing over the two possible orientations of the winding loop. In the normalisation chosen for the colour vectors, Eq. (4.2), the unique solution of the system of linear equations in Eq. (5.11) is

$$\mathbf{e}_0 = -\frac{\pi}{2}(e_\mathrm{g} + e_\mathrm{b}, 0, e_\mathrm{g} - e_\mathrm{b}). \tag{5.12}$$

This calculation of the *vector* background charge generalises the scalar case studied previously [6].

5.3. Liouville potential

The elastic term and the boundary term make up the usual Coulomb gas approach to two-dimensional critical phenomena. Recently we have argued that this description is incomplete and that an extra term S_L must be added to the effective action. To see this consider a large loop in the bulk, one that does not experience the boundary. Without the extra term this loop would be weighted exclusively by the bulk term S_E. There are two problems with this: S_E is real whilst an oriented loop should be weighted by a complex phase, and, S_E does not distinguish between the two orientations of a loop which are assigned different weights. We conclude that an extra *bulk* term is necessary!

The most general form of a bulk term is

$$S_\mathrm{L} = \int d^2\mathbf{x}\, w[\mathbf{h}(\mathbf{x})], \tag{5.13}$$

where $\exp(-w[\mathbf{h}(\mathbf{x})])$ is the scaling limit of $\lambda(\mathbf{x})$ in Eq. (4.3). In this sense S_L is energetic in origin, as opposed to S_E, which we argued in Section 5.1 accounts for the entropy of edge colourings.

Microscopically, the vertex weight λ can be written in terms of the colours of the bonds around the particular vertex as $\lambda = \exp(-w)$ where

$$\begin{aligned} w(\mathbf{B},\mathbf{C},\mathbf{A},\mathbf{D}) &= 0, \\ w(\mathbf{B},\mathbf{D},\mathbf{A},\mathbf{C}) &= 0, \\ w(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}) &= \mp i\frac{\pi}{4}(e_\mathrm{g} + e_\mathrm{b}), \\ w(\mathbf{B},\mathbf{A},\mathbf{C},\mathbf{D}) &= \mp i\frac{\pi}{4}(e_\mathrm{g} - e_\mathrm{b}), \end{aligned}$$

$$w(\mathbf{A},\mathbf{B},\mathbf{D},\mathbf{C}) = \mp i\frac{\pi}{4}(e_b - e_g),$$
$$w(\mathbf{B},\mathbf{A},\mathbf{D},\mathbf{C}) = \mp i\frac{\pi}{4}(-e_b - e_g); \tag{5.14}$$

the top sign is for even vertices whilst the bottom sign applies to odd vertices of the square lattice. Here we adopt the notation $(\sigma_1,\sigma_2,\sigma_3,\sigma_4)$ for the ordering of the colours around a vertex by listing the colours clockwise from the leftmost bond. The operator w is completely specified by the values it takes on the six edge colourings listed above since it does not change under cyclic permutations of its arguments.

By explicitly going through the six colour configurations listed above it is easily checked that

$$w(\mathbf{x}) = \frac{i}{16}\,\mathbf{e}_0 \cdot \mathbf{Q}(\mathbf{x}), \tag{5.15}$$

where the cross-staggered operator [28] is defined by

$$\mathbf{Q}(\mathbf{x}) = \pm[\sigma_1(\mathbf{x}) - \sigma_3(\mathbf{x})] \times [\sigma_2(\mathbf{x}) - \sigma_4(\mathbf{x})]. \tag{5.16}$$

Since $\mathbf{Q}(\mathbf{x})$ is manifestly invariant under 90° rotations of the colours around $\mathbf{x}$, (5.15) is seen to hold true for any distribution of the colours around a given vertex.

In order to find the coarse grained version of $w(\mathbf{x})$ we express it as a function of the height field $\mathbf{h}(\mathbf{x})$. First note that the microscopic operator $w(\mathbf{x})$ is *uniform* in each of the ideal states of the four colouring model. As such it defines a function on the ideal state graph $w(\mathbf{h})$, where $\mathbf{h} \in \mathcal{I}$ is the coarse grained height. Furthermore, it is a periodic function of $\mathbf{h}$ and it can therefore be written as a Fourier sum:

$$w(\mathbf{h}) = \sum_{\mathbf{e}\in\mathcal{R}_w^*} \tilde{w}_\mathbf{e} \exp(i\mathbf{e}\cdot\mathbf{h}). \tag{5.17}$$

The electric charges appearing in the sum take their values in the sub-lattice $\mathcal{R}_w^* \subset \mathcal{R}^*$, which is the lattice reciprocal to the lattice of *periods* of $w(\mathbf{h})$. In the continuum limit the coarse-grained height $\mathbf{h}$ is promoted into the height field $\mathbf{h}(\mathbf{x})$, and the scaling limit of the operator w is obtained by replacing $\mathbf{h}$ by $\mathbf{h}(\mathbf{x})$ in Eq. (5.17). Therefore, $w[\mathbf{h}(\mathbf{x})]$ is a sum of vertex operators,

$$w[\mathbf{h}(\mathbf{x})] = \sum_{\mathbf{e}\in\mathcal{R}_w^*} \tilde{w}_\mathbf{e} \exp(i\mathbf{e}\cdot\mathbf{h}(\mathbf{x})), \tag{5.18}$$

of which only the most relevant one(s) are kept in the effective action. Since the relevance of an operator is determined by its scaling dimension we turn to this calculation next.

5.3.1. *Dimensions of charge operators*

In the Coulomb gas formalism operators are associated with either electric or magnetic charges. Electric operators are vertex operators $\exp(i\mathbf{e}\cdot\mathbf{h})$ and they appear as the scaling limits of microscopic operators in the FPL^2 model that can be expressed as local functions of the colours; the loop-weight operator is one example.

Magnetic operators on the other hand cannot be expressed as local functions of the height but can be thought of as a constraint on the height field that generates a topological defect of strength $\mathbf{m}$. If $\mathbf{x}$ is the position of the defect core then the net height increase around any loop that encloses $\mathbf{x}$ is $\mathbf{m}$ (assuming no other defects are encircled). Geometrical exponents for loops in the FPL^2 model are given by dimensions of electric and magnetic operators in the associated Coulomb gas.

For an operator that has total electromagnetic charge $(\mathbf{e}, \mathbf{m})$, where $\mathbf{e} = (e_1, e_2, e_3)$ and $\mathbf{m} = (m^1, m^2, m^3)$, the scaling dimension is the sum of the electric and magnetic dimensions,[6]

$$2x(\mathbf{e}, \mathbf{m}) = \frac{1}{2\pi}\left[\frac{1}{g_\alpha}E_\alpha(E_\alpha - 2E_{0\alpha}) + g_\alpha(M^\alpha)^2\right], \tag{5.19}$$

where

$$E_1 = e_1 - e_3, \qquad E_2 = e_2, \qquad E_3 = e_1 + e_3 \tag{5.20}$$

and

$$M^1 = \tfrac{1}{2}(m^1 - m^3), \qquad M^2 = m^2, \qquad M^3 = \tfrac{1}{2}(m^1 + m^3) \tag{5.21}$$

are the electric and magnetic charge vectors in the basis in which the elastic term in the action is diagonal. Since the magnetic charges are given by height differences they must transform according to Eq. (5.7), whilst the electric charges transform in a dual fashion (cf. their appearance in the vertex operators).

5.3.2. Loop ansatz

With the dimension formula in hand, we can settle the issue of the most relevant operators appearing in the Fourier expansion of $w(\mathbf{h})$; see Eq. (5.18). There are twelve vertex operators to choose from corresponding to the twelve (110)-type vectors in the bcc lattice $\mathcal{R}^*$; these are the shortest vectors in the lattice $\mathcal{R}_w^*$. To find which of these electric charges minimise $x(\mathbf{e}, \mathbf{0})$ (Eq. (5.19)) it is convenient to first consider the simpler case of the FPL^2 model for $n_b = n_g$.

For the FPL^2 model with equal fugacities for the black and grey loops the effective action is considerably simplified. Namely, in this case the cyclic permutation of the colours,

$$(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}) \leftrightarrow (\mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{A}) : (z_1, z_2, z_3) \leftrightarrow (-z_1, z_3, -z_2) \tag{5.22}$$

does not change the vertex weight λ, and is thus an additional symmetry of the action S. This symmetry implies that $K_{13} = 0$ and $K_{22} = K_{11}$ in Eq. (5.6). Consequently there is only one elastic constant, $K \equiv K_{11}$. This then simplifies the formula for the dimension of an electromagnetic charge,

$$2x(\mathbf{e}, \mathbf{m}) = \frac{1}{2\pi K}\mathbf{e} \cdot (\mathbf{e} - 2\mathbf{e}_0) + \frac{K}{2\pi}\mathbf{m}^2, \tag{5.23}$$

[6] The derivation of Eq. (5.19) is an exercise in Gaussian integration and is reviewed in Appendix A.

where from Eq. (5.12) it follows that the background charge in this case has only one non-zero component, $\mathbf{e}_0 = -\pi(e_b, 0, 0)$. Now it is a simple matter to check that of the twelve (110)-type vectors in the lattice of electric charges $\mathcal{R}^*$, the four charges

$$\begin{aligned}
\mathbf{e}^{(1)} &= (-\pi, 0, +\pi), \\
\mathbf{e}^{(2)} &= (-\pi, 0, -\pi), \\
\mathbf{e}^{(3)} &= (-\pi, +\pi, 0), \\
\mathbf{e}^{(4)} &= (-\pi, -\pi, 0)
\end{aligned} \tag{5.24}$$

are degenerate in dimension and they minimise $2x(\mathbf{e}, \mathbf{0})$. These are therefore the electric charges of the vertex operators that are kept in the action.

Now we turn to the *loop ansatz* which states that the operator $w(\mathbf{h})$ is exactly marginal in the renormalisation group sense. This is the statement that the loop weight does not renormalise at large scales. The geometrical meaning of this becomes obvious when one realises that the number of loops inside a domain of size ρ, whose linear size is *comparable* to ρ, is thermodynamically conjugate to the loop weight at scale ρ. Thus the loop ansatz states that the number of large loops does not grow with scale (more precisely it is sufficient to assume that it does not grow faster than any power of the scale). The analogous statement can be proven rigorously for critical percolation where it is the source of hyperscaling [35].

The assumption that there is of order one loop at every scale is linked to the variance of the height difference between two points in the basal plane, separated by a macroscopic distance $|\mathbf{x}|$. Namely, if we assume that when going from one point to the other there is of order one contour loop that is crossed at every *scale*, and further assuming that the directions of these contours are independent from scale to scale, it follows from the law of large numbers that the variance of the height difference grows as the number of contours crossed, that is as $\log(|\mathbf{x}|)$. This, of course, is nothing but the large $|\mathbf{x}|$ behaviour of $\langle (H^\alpha(\mathbf{x}) - H^\alpha(\mathbf{0}))^2 \rangle$ calculated in the Gaussian model of Eq. (5.8).

The loop ansatz, or in other words the marginality hypothesis for the loop weight operator, simply translates into a statement about its scaling dimension:

$$x(\mathbf{e}^{(i)}, \mathbf{0}) = 2, \quad i = 1, 2, 3, 4. \tag{5.25}$$

This, using the dimension formula Eq. (5.23), leads to a formula for the single elastic constant K.

In the general case $n_b \neq n_g$, the scaling dimensions of the four electric charges identified above are

$$\begin{aligned}
x(\mathbf{e}^{(1)}, 0) &= \pi \frac{1 - e_b}{g_1}, \\
x(\mathbf{e}^{(2)}, 0) &= \pi \frac{1 - e_g}{g_3}, \\
x(\mathbf{e}^{(3)}, 0) = x(\mathbf{e}^{(4)}, 0) &= \frac{\pi}{4} \left(\frac{1 - 2e_b}{g_1} + \frac{1}{g_2} + \frac{1 - 2e_g}{g_3} \right);
\end{aligned} \tag{5.26}$$

the last two remain degenerate in dimension. The dimensions of the first two charges are also equal due to the "duality" transformation of the FPL^2 model which exchanges the two flavours, $n_b \leftrightarrow n_g$. This transforms the microscopic heights $z_2 \to -z_2$ and $z_3 \to -z_3$ (and similarly for the appropriate components of the height field). Furthermore, the elastic constants K_{11} and K_{22} in Eq. (5.6) are unchanged, whilst $K_{13} \to -K_{13}$. Finally, from Eq. (5.9) it follows that the duality transformation exchanges the couplings $g_1 \leftrightarrow g_3$ thus rendering $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$ degenerate in dimension, as the FPL^2 model is self-dual.

Unlike the case of $n_b = n_g$, the loop ansatz in the general case requires that at least *two* of the electric charges $\mathbf{e}^{(i)}$ $(i = 1, 2, 3, 4)$ remain marginal, thus enforcing the non-renormalisability of the two fugacities n_b and n_g. If we now further *assume* that these charges are unrelated by the "duality" transformation described above, it follows that in fact all four are marginal. The three couplings are then simply calculated by setting the right-hand sides of Eq. (5.26) equal to 2. We find

$$\begin{aligned} g_1 &= \frac{\pi}{2}(1 - e_b), \\ g_3 &= \frac{\pi}{2}(1 - e_g), \\ \frac{1}{g_2} &= \frac{1}{g_1} + \frac{1}{g_3}. \end{aligned} \tag{5.27}$$

One final comment is in order. The relation $1/g_2 = 1/g_1 + 1/g_3$ comes as somewhat of a surprise, as it was not anticipated on symmetry grounds. Of course, since a particular point in the critical region of the FPL^2 model is determined by two parameters, n_b and n_g, one relation between the three couplings is to be expected. It is therefore an interesting open question whether a critical loop model can be constructed in which g_2 would be unconstrained.[7]

With the values of the couplings g_1, g_2, and g_3 in hand, as well as the formula for the scaling dimensions of charged operators, Eq. (5.19), we are fully equipped to calculate critical exponents of the FPL^2 model. In particular, in the next section we calculate the formulae for the central charge and the geometrical exponents associated with loops as a function of the loop fugacities, n_b and n_g, for the whole critical region of the model.

6. Central charge

We now turn to the calculation of the central charge in the critical region, $0 \leqslant n_b, n_g \leqslant 2$. Exactly at the point $(n_b, n_g) = (2, 2)$ the background charge vanishes, $\mathbf{e}_0 = \mathbf{0}$, and the action consists only of the elastic term S_E given by Eq. (5.8). Since this is then simply a theory of three free massless bosonic fields we conclude that, in this case, $c = 3$ [28].

For a general value of the background charge this generalises to [36]

$$c = 3 + 12x(\mathbf{e}_0, \mathbf{0}). \tag{6.1}$$

[7] This possibility was suggested to us by D. Huse.

One way to rationalise the factor of 12 is to compare the coefficients of the finite-size corrections in the well-known formulae [37,38]

$$f_0(\infty) - f_0(L) = \frac{\pi c}{6L^2} + \ldots, \tag{6.2}$$

$$f_i(L) - f_0(L) = \frac{2\pi x_i}{L^2} + \ldots, \tag{6.3}$$

where $f_{0,i}(L)$ is the free energy density on a cylinder of circumference L, the subscript 0 referring to the vacuum and i to the case when an operator of scaling dimension x_i is inserted. The physical meaning of Eq. (6.1) is that the presence of the background charge – $+\mathbf{e}_0$ and $-\mathbf{e}_0$ at the two ends of the cylinder – lowers the free energy and with it the central charge.

Now using the dimension formula, Eq. (5.19), and inserting the values of the couplings g_α from Eq. (5.27), we arrive at

$$c = 3 - 6\left(\frac{e_b^2}{1-e_b} + \frac{e_g^2}{1-e_g}\right), \tag{6.4}$$

where we recall that $n_b = 2\cos(\pi e_b)$ and similarly for n_g. In Table 1 the numerically calculated values of the central charge are compared to the above formula, and excellent agreement is found.

7. Geometrical scaling dimensions

7.1. Two-string dimension

In addition to the central charge, the Coulomb gas representation of the loop model provided by the Liouville field theory, Eq. (5.2), allows for the evaluation of various geometrical scaling dimensions. As an example of such a quantity, consider the probability $G_2(r)$ that two points separated by a distance r lie on the same, say, black loop. In the critical phase we expect this probability to decay as $G_2(r) \sim r^{-2x_2}$, which defines the scaling dimension x_2. Since a black loop is represented as a sequence of alternating **A**- and **B**-coloured edges it follows from the colouring constraint that the microscopic heights **z** just outside this loop differ by integer multiples of **C** and **D** only. In other words, a black loop is a *contour* loop for the component of the height along the direction perpendicular to both **C** and **D**, i.e. the $(1,0,-1)$ direction in height space. Similarly the grey loops are contour loops for the height component along the $(1,0,1)$ direction.

It has been argued that the scaling dimension governing the probability that two points belong to the same contour loop of a random Gaussian surface equals $1/2$, independent of the stiffness [39]. Thus, for $(n_b, n_g) = (2,2)$ when $\mathbf{e}_0 = \mathbf{0}$ and the effective field theory is Gaussian, we expect $x_2 = 1/2$. For other values of the fugacities the Gaussian theory is modified by the background charge and the same argument cannot be made.

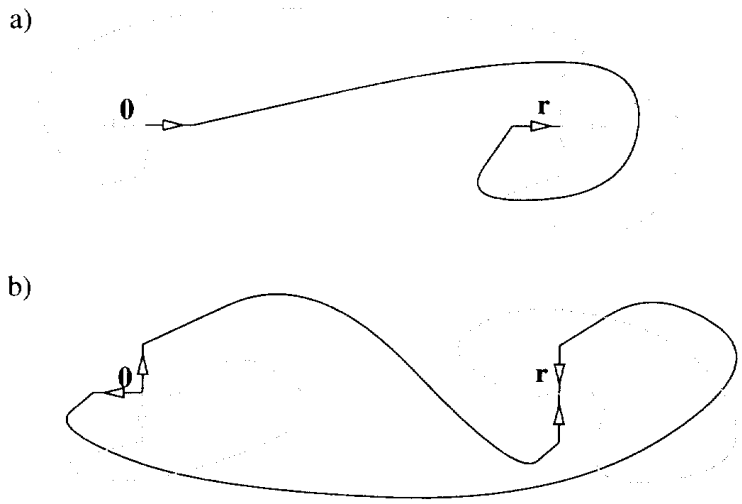


Fig. 7. Defect configurations used to calculate geometrical exponents x_1 (a) and x_2 (b) in the FPL^2 model. In (a) there is a single oriented black loop segment and a single oriented grey loop segment propagating from **0** to **r**, whilst in (b) there are two oriented black loop segments between **0** and **r**.

A more illuminating way of making contact with the interface representation is to view $G_2(r)$ as a two-string correlation function associated with defect configurations where two black strings emanating from the origin annihilate one another at a distant point **r**; see Fig. 7b. This can be accomplished by rewriting $G_2(r)$ as $Z(r)/Z$, where Z is the partition function defined by Eq. (4.3), and $Z(r)$ is similarly defined but with the summation restricted to those configurations $\mathcal{G}'_{\mathbf{r}}$ where an oriented black loop passes through the points **0** and **r**. Now consider reversing the direction of one half of the loop, so that instead of having one oriented loop passing through **0** and **r** we have two oriented loop segments directed from **0** to **r** [6]. This corresponds to the introduction of defect configurations at these two points, where we have violated the edge-colouring constraint. At **0** we find a $(\mathbf{C},\mathbf{D},\mathbf{A},\mathbf{A})$ configuration of colours which in the height language corresponds to a vortex of strength

$$\mathbf{m}_2 = \mathbf{A} - \mathbf{B} = (-2,0,2). \tag{7.1}$$

The strength of the vortex (its Burgers charge) at **0** is calculated as the total height change around **0**. Similarly, at **r** we have the corresponding antivortex $(\mathbf{B},\mathbf{B},\mathbf{C},\mathbf{D})$ of strength $-\mathbf{m}_2$ as illustrated in Fig. 7b.

In order to calculate x_2 for general values of the loop fugacities we have to take into account the effect of the complex phase factors associated with oriented loops. Namely, when one or more, say, black strings are associated with a vortex–antivortex configuration, spurious phase factors $\exp(\pm i\pi e_{\mathrm{b}})$ will arise whenever a black loop segment winds around one of the vortex cores [6]; for example, in Fig. 7b one of the two black strings winds once around point **r**. The spurious winding phase can be removed by inserting the vertex operator $\exp(i\mathbf{e}_{\mathrm{b}}\cdot\mathbf{h})$ at the positions of both vortex cores. Since a black loop has alternating **A** and **B** colours the electric charge $\mathbf{e}_{\mathrm{b}}$ must satisfy

$$\begin{aligned} \mathbf{e}_b \cdot \mathbf{A} &= \pi e_b, & \mathbf{e}_b \cdot \mathbf{B} &= -\pi e_b, \\ \mathbf{e}_b \cdot \mathbf{C} &= 0, & \mathbf{e}_b \cdot \mathbf{D} &= 0. \end{aligned} \tag{7.2}$$

Similarly, if there are grey strings propagating between two vertices the spurious phase factors associated with winding configurations are corrected with vertex operators whose electric charge $\mathbf{e}_g$ is determined by

$$\begin{aligned} \mathbf{e}_g \cdot \mathbf{A} &= 0, & \mathbf{e}_g \cdot \mathbf{B} &= 0, \\ \mathbf{e}_g \cdot \mathbf{C} &= \pi e_g, & \mathbf{e}_g \cdot \mathbf{D} &= -\pi e_g. \end{aligned} \tag{7.3}$$

Using Eq. (4.2) for the colour-vectors we find

$$\mathbf{e}_b = -\frac{\pi}{2}(e_b, 0, -e_b), \qquad \mathbf{e}_g = -\frac{\pi}{2}(e_g, 0, e_g). \tag{7.4}$$

Going back to the two-string operator we conclude that it has total electromagnetic charge $(\mathbf{e}_b, \mathbf{m}_2)$.

Finally, from the general expression for the scaling dimension of an electromagnetic operator, Eq. (5.19), it follows that

$$2x_2 = 2x(\mathbf{e}_b, \mathbf{m}_2) = (1 - e_b) - \frac{e_b^2}{1 - e_b}. \tag{7.5}$$

In Table 5 exact values of x_2 calculated from this formula are compared to numerical results, and excellent agreement is found.

Interestingly the expression for x_2 is *independent* of e_g, i.e. it is not affected by the fugacity of grey loops. This observation conforms to our understanding of the scaling of compact polymers. The compact polymer problem is recovered in the limit $n_b \to 0$ in which case there is a single black loop on the lattice. Since the loop fills space its Hausdorff dimension is necessarily $D = 2$. Scaling tells us that [40]

$$D = 2 - x_2, \tag{7.6}$$

from which the result $x_2 = 0$ follows, *independent* of the fugacity of the grey loops. The fact that our formula reproduces this simple result in the $n_b = 0$ ($e_b = 1/2$) case provides a non-trivial check on its validity.

7.2. *One-string dimension*

The scaling dimension x_1, corresponding to one black and one grey string propagating between two points on the lattice, can be computed in a way that is completely analogous to the case of two black strings discussed above. (Note that the fully packing constraint ensures that if there is a single black string between two points then these points are also connected by a grey string; see Fig. 7a.) Choosing one point on the even sublattice and the other on the odd leads to the appearance of the defect configuration $(\mathbf{A}, \mathbf{C}, \mathbf{C}, \mathbf{D})$ on both sites of the square lattice. These in turn correspond to vortices in the height representation with topological charges $\pm\mathbf{m}_1$, where

$$\mathbf{m}_1 = \mathbf{C} - \mathbf{B} = (-2, -2, 0). \tag{7.7}$$

Since strings of both flavours are now present the compensating electric charge is $\mathbf{e}_b + \mathbf{e}_g = \mathbf{e}_0$. Hence

$$2x_1 = 2x(\mathbf{e}_0, \mathbf{m}_1) = \frac{1}{4}\left[(1-e_b) + (1-e_g)\right] + \frac{(1-e_b)(1-e_g)}{(1-e_b)+(1-e_g)} - \left[\frac{e_b^2}{1-e_b} + \frac{e_g^2}{1-e_g}\right]. \tag{7.8}$$

There are of course several different ways of choosing the defect configurations (in this case, eight), but it should hardly come as a surprise that they all lead to the same expression for the scaling dimension.

Unlike x_2, x_1 depends on both loop fugacities. Going back to our original motivation, the compact polymer problem ($n_b = 0 \Rightarrow e_b = 1/2$), x_1 determines the value of the conformational exponent $\gamma = 1 - x_1$, which describes the scaling of the number of compact polymers with the number of monomers. We see that depending on e_g there will be a continuum of γ's. How do we interpret this?

First note that the problem of counting the number of conformations of a single compact polymer is the case $n_g = 1$ ($e_g = 1/3$) which simply assigns equal weights to all conformations. Using Eq. (7.8) this choice leads to $x_1 = -5/112$ and to the result $\gamma = 117/112$ advertised in the abstract. Changing n_g (e_g) away from $n_g = 1$, on the other hand, has the effect of favouring certain compact polymer conformations over others depending on the number of loops formed by the uncovered (grey) bonds. In this sense the weight assigned to grey loops can be thought of as an interaction between the monomers of the compact polymer, albeit a peculiar non-local one. Nonetheless, it is interesting that this interaction changes the scaling properties of the compact polymer leading to a continuously varying exponent γ (more on this in Section 11).

7.3. *Many-string dimensions*

The dimensions x_1 and x_2 given above are contained in a more general set of string dimensions x_{s_b,s_g} governing the probability of having s_b black loop segments and s_g grey loop segments propagating between two points on the lattice [18]. More precisely, we consider two microscopic regions centred around points separated by a macroscopic distance, one region being the source and the other the sink of the oriented loop segments. Since the defect configurations obtained by violations of the edge colouring constraint must necessarily give rise to an *even* number of strings we will only consider the case when $s_b + s_g$ is even.

Consider first the case $s_b = 2k_b$ and $s_g = 2k_g$. The appropriate magnetic charge is obtained by combining k_b vortices with charge $\mathbf{A} - \mathbf{B} = (-2, 0, 2)$, and k_g vortices with charge $\mathbf{C} - \mathbf{D} = (-2, 0, -2)$. The defect with charge $\mathbf{A} - \mathbf{B}$ acts as a source of two black segments, whilst $\mathbf{C} - \mathbf{D}$ is associated with two grey loop segments. We also need to introduce the electric charge $\mathbf{e}_b + \mathbf{e}_g$ to compensate for the extra winding phase

associated with the black and grey loop segments. The total electromagnetic charge is therefore

$$[\mathbf{e}_{2k_b,2k_g}, \mathbf{m}_{2k_b,2k_g}] = [\mathbf{e}_b(1-\delta_{k_b,0}) + \mathbf{e}_g(1-\delta_{k_g,0}), -2(k_b+k_g, 0, k_g-k_b)], \tag{7.9}$$

and from the dimension formula, Eq. (5.19), we find

$$2x_{2k_b,2k_g} = (1-e_b)k_b^2 + (1-e_g)k_g^2 - \frac{e_b^2}{1-e_b}(1-\delta_{k_b,0}) - \frac{e_g^2}{1-e_g}(1-\delta_{k_g,0}). \tag{7.10}$$

This formula generalises Eq. (7.5).

Similarly, for $s_b = 2k_b - 1$ and $s_g = 2k_g - 1$ the electromagnetic charge is

$$[\mathbf{e}_{2k_b-1,2k_g-1}, \mathbf{m}_{2k_b-1,2k_g-1}] = [\mathbf{e}_0, -2(k_b+k_g-1, 1, k_g-k_b)]; \tag{7.11}$$

the magnetic charge is obtained by combining $k_b - 1$ defects of charge $\mathbf{A} - \mathbf{B}$, $k_g - 1$ defects of charge $\mathbf{C} - \mathbf{D}$, and a single defect of charge $\mathbf{C} - \mathbf{B}$ which produces the remaining single black and grey strings originating from the same vertex. The scaling dimension is found to be

$$2x_{2k_b-1,2k_g-1} = \frac{1}{4}\left[(1-e_b)(2k_b-1)^2 + (1-e_g)(2k_g-1)^2\right] + \frac{(1-e_b)(1-e_g)}{(1-e_b)+(1-e_g)} - \left[\frac{e_b^2}{1-e_b} + \frac{e_g^2}{1-e_g}\right]. \tag{7.12}$$

This generalises the expression given in Ref. [7] and correctly reduces to Eq. (7.8) for $k_b, k_g = 1$.

7.4. Thermal dimension

We now turn our attention to the thermal scaling dimension. The FPL^2 model can be thought of as the zero-temperature limit of a more general model where we allow for thermal excitations that violate the close packing constraint. In this sense the temperature variable is thermodynamically conjugate to the constraint that every vertex be visited by (say) a black loop. An appropriate defect configuration for computing x_T within the FPL^2 model is therefore $(\mathbf{C}, \mathbf{D}, \mathbf{C}, \mathbf{D})$. This is a vortex of strength

$$\mathbf{m}_T = 2(\mathbf{C} + \mathbf{D}) = (0, -4, 0), \tag{7.13}$$

and since no strings terminating in the bulk are generated there is no compensating electric charge. The scaling dimension is then

$$2x_T = 2x(0, \mathbf{m}_T) = 4\frac{(1-e_b)(1-e_g)}{(1-e_b)+(1-e_g)}. \tag{7.14}$$

The exact values of x_T quoted in Table 2 are calculated using this formula. The numerical results are in excellent agreement.

7.5. Boundary-string dimensions

The simplest example of a string operator that cannot be accessed within the formalism presented above is that of one black and no grey strings propagating between two vertices of the square lattice. Since this configuration has an odd number of strings connecting two sites of the lattice these two sites necessarily reside on the *boundary*.

If we define the FPL^2 model on the cylinder, as will be the case when we construct its transfer matrix in Section 9, a single black string can be enforced to run along the length of the cylinder if its circumference is chosen *odd*. Taking our cue from the formulae derived above for the bulk string operators we *guess* the formula

$$2X = \frac{1}{4} + \frac{1-e_{\mathrm{b}}}{4} - \frac{e_{\mathrm{b}}^2}{1-e_{\mathrm{b}}} \tag{7.15}$$

from the numerical results shown in Table 3. X is the scaling dimension of the boundary operator which corresponds to a single black (or grey) string.

The Coulomb gas interpretation of the second and third term in Eq. (7.15) is rather apparent when one compares them with Eq. (7.5). The second term can be rationalised as coming from a magnetic charge $(-1,0,1)$ which is half the charge $\mathbf{m}_2$ in Eq. (7.1), associated with two black strings; this is saying that we have a partial dislocation generated at the boundary. The third term is due to the compensating electric charge $\mathbf{e}_{\mathrm{b}}$ for a single black string, same as in the two-string case.

The first, constant term does not have an immediate interpretation. A possible scenario is that it is due to the boundary condition imposed on the height by virtue of having a cylinder of odd circumference. Namely, a translation along the periodic coordinate by an amount equal to the circumference (L) exchanges an even site for an odd site (and *vice versa*) resulting in a transformation of the height: $\mathbf{h}(x^1,x^2) = \mathrm{P}\mathbf{h}(x^1+L,x^2)$. Since $\mathrm{P}^2 = 1$ this boundary condition can be thought of as the insertion of a twist operator into the partition function. The twist operator has dimension $1/8$ regardless of the stiffness of the interface [41].

The above considerations permit us to calculate the scaling dimension for the general case of an odd number of strings. For definiteness we consider the case of $s_{\mathrm{b}} = 2k_{\mathrm{b}} - 1$ and $s_{\mathrm{g}} = 2k_{\mathrm{g}}$. The magnetic charge pertaining to this situation is found by combining $2k_{\mathrm{b}} - 1$ defects of charge $\frac{1}{2}(\mathbf{A}-\mathbf{B})$ with k_{g} defects of charge $\mathbf{C}-\mathbf{D}$, totalling

$$[\mathbf{e}_{2k_{\mathrm{b}}-1,2k_{\mathrm{g}}}, \mathbf{m}_{2k_{\mathrm{b}}-1,2k_{\mathrm{g}}}] = [\mathbf{e}_{\mathrm{b}} + \mathbf{e}_{\mathrm{g}}(1-\delta_{k_{\mathrm{g}},0}), (1-2k_{\mathrm{b}}-2k_{\mathrm{g}}, 0, 2k_{\mathrm{b}}-2k_{\mathrm{g}}-1)]. \tag{7.16}$$

Taking into account the contribution from the twist operator, i.e. adding $1/4$ to the result obtained from Eq. (5.19), the scaling dimension is then

$$2x_{2k_{\mathrm{b}}-1,2k_{\mathrm{g}}} = \frac{1}{4} + \frac{1}{4}\left[(1-e_{\mathrm{b}})(2k_{\mathrm{b}}-1)^2 + (1-e_{\mathrm{g}})(2k_{\mathrm{g}})^2\right] - \left[\frac{e_{\mathrm{b}}^2}{1-e_{\mathrm{b}}} + \frac{e_{\mathrm{g}}^2}{1-e_{\mathrm{g}}}(1-\delta_{k_{\mathrm{g}},0})\right]. \tag{7.17}$$

7.6. Complete spectrum of string dimensions

Finally, the results of Eqs. (7.10), (7.12) and (7.17) can be combined into a single equation for the scaling dimension of a string operator that corresponds to s_b black loop segments and s_g grey loop segments:

$$2x_{s_b,s_g} = \frac{1}{4}\delta^{(2)}_{s_b+s_g,1} + \frac{1}{4}\left[(1-e_b)s_b^2 + (1-e_g)s_g^2\right] - \left[\frac{e_b^2}{1-e_b}(1-\delta_{s_b,0}) + \frac{e_g^2}{1-e_g}(1-\delta_{s_g,0})\right] + \delta^{(2)}_{s_b,1}\delta^{(2)}_{s_g,1}\frac{(1-e_b)(1-e_g)}{(1-e_b)+(1-e_g)}, \tag{7.18}$$

where we have defined $\delta^{(2)}_{i,j} \equiv \delta_{i=j(\mathrm{mod}\ 2)}$.

8. Termination of critical behaviour

In the preceding sections we have developed an effective description of the critical phase of the FPL^2 model in the form of a field theory. This theory has to break down at large values of the loop fugacity since in this case a typical state of the model will consist of small loops only, i.e. a power-law distribution of loop sizes will be absent. That this indeed happens can be seen from the Liouville field theory itself as it carries the seeds of its own demise.

The mapping of the loop model to an oriented loop model for $n_b, n_g \leqslant 2$ works equally well for $n_b > 2$ or $n_g > 2$. From Eq. (4.1) it follows that in the latter case at least one of the parameters, e_b or e_g, will be pure imaginary. This affects the Liouville potential which for $n_b > 2$ or $n_g > 2$ becomes a *relevant* perturbation to the (modified) Gaussian action $S_E + S_B$.

To understand how this comes about we consider the simple case provided by the $n_b = n_g$ FPL^2 model, discussed in Section 5.3.2. Namely, as we increase the value of the loop fugacity we expect small loops to be favoured and the stiffness K of the interface to grow. In the critical phase this is offset by the decrease in the background charge in a way that leaves the Liouville potential marginal. Now when the loop fugacity exceeds 2 the background charge $\mathbf{e}_0 = -\pi(e_b, 0, 0)$ becomes pure imaginary and the dimension of the Liouville potential

$$x_L = \frac{\pi}{2}\frac{1-e_b}{K} \tag{8.1}$$

can no longer stay marginal; here $x_L \equiv x(\mathbf{e}^{(i)}, \mathbf{0})$, where the charges $\mathbf{e}^{(i)}$ are given in Eq. (5.24), and their scaling dimensions are calculated from Eq. (5.23). In fact, assuming that the stiffness K continues to increase with the loop fugacity for $n_b = n_g > 2$, x_L turns complex with a real part that is *smaller* than two, rendering the Liouville potential *relevant*.

If we make the usual assumption of no intervening fixed points, the relevant Liouville potential will generate a finite correlation length and the loop model will no longer be critical. The correlation length has the physical interpretation of the average size of a loop in the system. This scenario has been confirmed for the fully packed loop model on the honeycomb lattice, from the Bethe ansatz solution of this model [42].

A different view of the non-critical region of the FPL^2 model is provided by the locking potential $V(\mathbf{h})$. Namely, the discrete nature of the microscopic heights can be taken into account in the field theory by a negative potential in height space that is peaked around the flat, ideal states. As such, this potential is uniform on the ideal state graph and can therefore be expanded in a Fourier series. Examination of the most relevant vertex operators in this series [28] reveals that they are the same as the ones for the loop-weight (Liouville) potential, $w(\mathbf{h})$. Therefore, just like $w(\mathbf{h})$, the locking potential in the non-critical region of the phase diagram is a *relevant* perturbation. Thus, it will lock the interface in one of the ideal, flat states. In this flat phase the height fluctuations are bounded (as opposed to being logarithmically divergent) which is just another way of saying that large contour loops are exponentially suppressed. On the other hand, in the critical region of the FPL^2 model the locking potential is marginal as it would be for an interface model *at* the roughening transition [43]. This might indicate that the whole critical region of the FPL^2 model can be understood as a manifold of essential singularities in some more general model, as was the case for the honeycomb FPL model [21,44].

As stated in Sections 6 and 7, our results for the central charge and a number of the geometrical scaling dimensions have been very accurately confirmed by transfer matrix calculations. Before turning to a discussion of our numerical results we describe the particular representation of the transfer matrix used to obtain them.

9. Construction of the transfer matrix

To construct the transfer matrix for the FPL^2 model on a cylinder of circumference L we write the partition function as

$$Z^{(M)} = \sum_{\mathcal{G}_M} n_{\mathrm{b}}^{N_{\mathrm{b}}} n_{\mathrm{g}}^{N_{\mathrm{g}}}, \tag{9.1}$$

where the length of the cylinder M has been explicitly indicated. Periodic boundary conditions are imposed in the horizontal direction, whereas the bottom and the top row of the cylinder have open boundary conditions and hence terminate in L dangling edges. We recall that the restriction of the summation to the set of fully packed graph configurations $\mathcal{G}_M$ implies that locally the vertices are constrained to have one of the six appearances shown in Fig. 3. In the first four possible vertices the loop segments do not cross, whilst in the last two vertices the two flavours intersect. The global constraint that

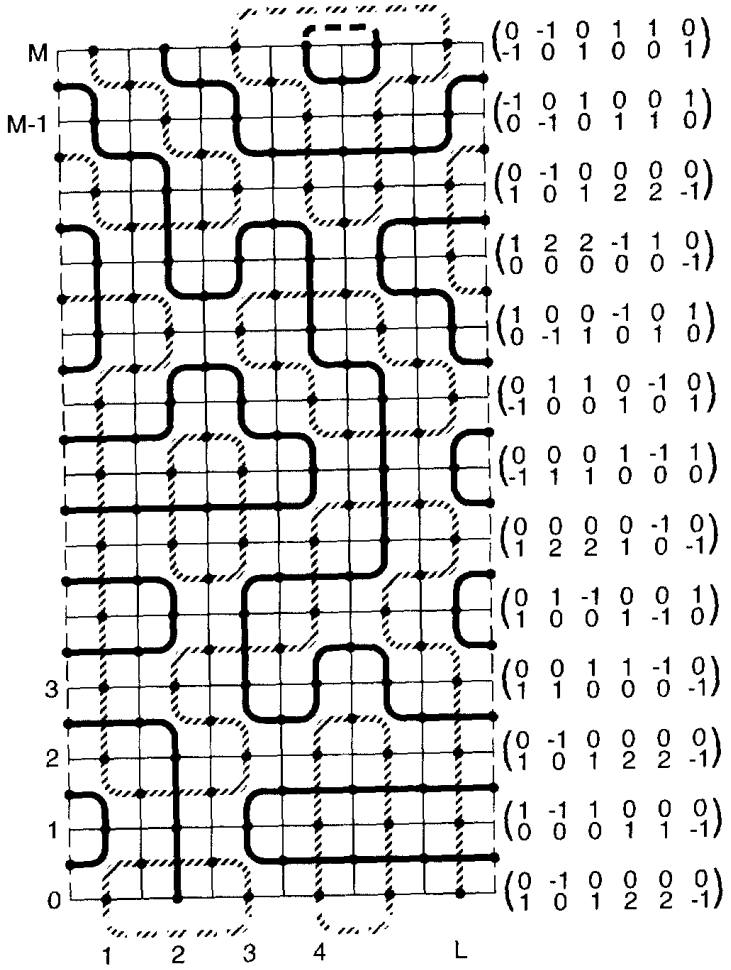


Fig. 8. A typical loop configuration for $L = 6$ and $M = 12$. The dashed lines along the left and the right boundaries illustrate the periodic boundary conditions. Horizontally the vertices are numbered from 1 to L, whilst vertically the rows are labelled by the number of the vertex immediately below them. This particular configuration is constrained to having precisely one string of each flavour spanning the length of the cylinder, and hence it contributes to the geometrical exponent x_1. To the right we show the index representation of the connectivity state pertaining to each row (see Appendix B for details). Any valid configuration can be interpreted as a "jigsaw puzzle" assembled from the six "pieces" shown in Fig. 3. Note that when laying down the first row of this puzzle it must be stipulated how the dangling edges of row 0, which are not part of a string, are pairwise interconnected below that row. These implicit connections as well as their counterparts in row M have been depicted by dashed loop segments.

all loops be closed in the limit of an infinite system means that loop segments cannot terminate in the bulk but only at the dangling edges in the top and bottom rows.

A typical loop configuration for a cylinder with $L = 6$ and $M = 12$ is shown in Fig. 8. The horizontal numbering pertains to the vertices, whilst in the vertical direction it is more convenient to label each row by the number of the vertex immediately below it. Accordingly the labels 0 and M refer to the bottom and the top row of dangling edges respectively. We shall soon see that the inclusion in $\mathcal{G}_M$ of one or more strings running between the dangling edges of row 0 and M helps us access the geometrical exponents of the model. In particular, the configuration of Fig. 8 having one such string

of each flavour furnishes a contribution to the scaling dimension x_1 which determines the conformational exponent $\gamma = 1 - x_1$.

9.1. Connectivity basis

The construction of a transfer matrix (TM) for Eq. (9.1) appears to be obstructed by the non-locality of N_i ($i = \mathrm{b}, \mathrm{g}$). The key to solving this problem is to write the TMs in a basis of *connectivity states* comprising information about how the dangling ends of row M are pairwise interconnected in the preceding rows and, if strings are present, information about the positions of such strings. In addition the connectivity states must keep track of the particular flavour of any loop or string segment terminating in row M. Our construction generalises the work of Blöte and Nightingale for the Q-state Potts model [45,46] and that of Blöte and Nienhuis for the O(n) model [47] to take the extra flavour information into account, and our notation is consistent with that of these authors.

It is essential to be able to represent a given connectivity state both in an *index representation* giving direct access to the flavour and connectedness information just mentioned, and in a *number representation* assigning an integer in the range $1, 2, \ldots, C_L^{(s_\mathrm{b}, s_\mathrm{g})}$ to the state under consideration. The latter representation enables us to enumerate the entries of the TM, whilst the former allows us to determine the number of loop closures when going from one connectivity state to another and hence the value of a particular entry in the TM. Here $C_L^{(s_\mathrm{b}, s_\mathrm{g})}$ is the number of distinct connectivity states for a cylinder of width L accommodating s_i strings of flavour $i = \mathrm{b}, \mathrm{g}$. The construction of these two representations, the mapping between them, and the evaluation of the $C_L^{(s_\mathrm{b}, s_\mathrm{g})}$ for $(s_\mathrm{b}, s_\mathrm{g}) = (0,0)$, $(1,0)$, $(1,1)$ and $(2,0)$ is deferred to Appendix B.

Designating the connectivity states by Greek letters we can write the partition function as a sum of restricted partition functions

$$Z^{(M)} = \sum_\beta Z_\beta^{(M)} = \sum_\beta \sum_{\mathcal{G}_M} \delta\big(\beta, \phi(\mathcal{G}_M)\big) n_\mathrm{b}^{N_\mathrm{b}} n_\mathrm{g}^{N_\mathrm{g}}, \tag{9.2}$$

where $\phi(\mathcal{G}_M)$ is the connectivity of the L dangling edges of row M, and $\delta(i,j)$ is the Kronecker delta. Now consider appending another row to the cylinder, giving us a total graph configuration $\mathcal{G}_{M+1} = \mathcal{G}_M \cup \mathcal{G}'$. Evidently the connectivity of the dangling edges of row $M+1$ is determined solely by that of the preceding row and by the appended subgraph $\mathcal{G}'$

$$\phi(\mathcal{G}_{M+1}) = \psi\big(\phi(\mathcal{G}_M), \mathcal{G}'\big). \tag{9.3}$$

Letting N_i' denote the number of loop closures induced by $\mathcal{G}'$ we arrive at the relation

$$\begin{aligned} Z_\alpha^{(M+1)} &= \sum_{\mathcal{G}_{M+1}} \delta\big(\alpha, \phi(\mathcal{G}_{M+1})\big) n_\mathrm{b}^{N_\mathrm{b}+N_\mathrm{b}'} n_\mathrm{g}^{N_\mathrm{g}+N_\mathrm{g}'} \\ &= \sum_\beta \sum_{\mathcal{G}_M} \delta\big(\beta, \phi(\mathcal{G}_M)\big) n_\mathrm{b}^{N_\mathrm{b}} n_\mathrm{g}^{N_\mathrm{g}} \sum_{\mathcal{G}' | \mathcal{G}_M} \delta\big(\alpha, \psi(\phi(\mathcal{G}_M), \mathcal{G}')\big) n_\mathrm{b}^{N_\mathrm{b}'} n_\mathrm{g}^{N_\mathrm{g}'} \end{aligned}$$

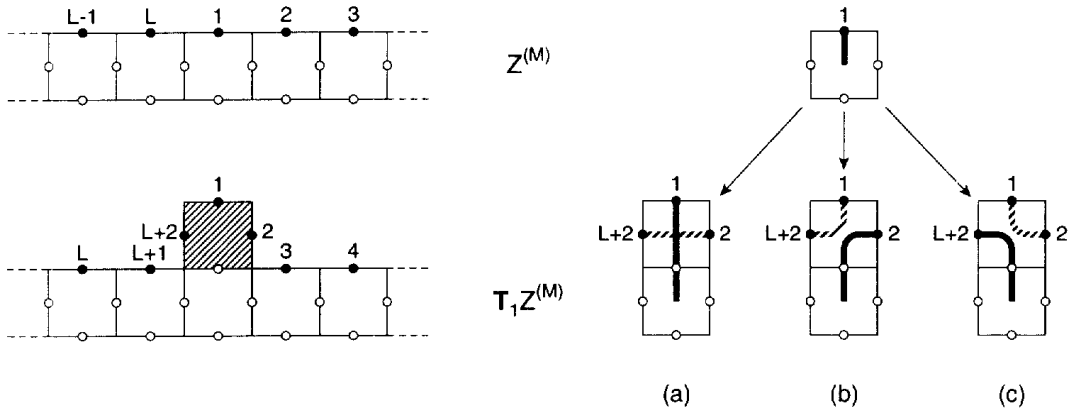


Fig. 9. Adding the first vertex of the $(M+1)$th row increases the number of dangling edges from L to $L+2$. The labelling of the "active" edges (filled circles) before and after addition of the new vertex (shaded) is as shown in the left part of the figure. The part of the lattice relevant for determining which of the vertices of Fig. 3 fit onto a given connectivity of row M, has been depicted in the right part of the figure. This information constitutes the *vertex rules*, and is explained in the text.

$$= \sum_{\beta} T_{\alpha\beta} Z_{\beta}^{(M)}, \tag{9.4}$$

where the transfer matrix is defined by

$$T_{\alpha\beta} = \sum_{\mathcal{G}'|\mathcal{G}_M} \delta\big(\alpha, \psi(\phi(\mathcal{G}_M), \mathcal{G}')\big) n_{\mathrm{b}}^{N'_{\mathrm{b}}} n_{\mathrm{g}}^{N'_{\mathrm{g}}}. \tag{9.5}$$

The notation $\mathcal{G}'|\mathcal{G}_M$ means that the summation is constrained to those subgraphs $\mathcal{G}'$ that fit the dangling edges of $\mathcal{G}_M$.

9.2. *Single-vertex decomposition*

A quintessential step in the practical implementation of the TM is its decomposition into matrices each corresponding to the addition of a single vertex,

$$\mathbf{T} = \mathbf{T}_L \cdot \mathbf{T}_{L-1} \cdot \ldots \cdot \mathbf{T}_1. \tag{9.6}$$

Here the single-vertex matrix $\mathbf{T}_i$, which adds the vertex at horizontal position i of the new row, has the advantage of being *sparse*, and we shall soon see that it has at most three non-zero entries per column. This property leads to a dramatic reduction of the time and storage requirements for the calculations.

As was the case in the O(n) model [47], a minor complication arises due to the fact that the addition of the first vertex of a new row increases the number of dangling edges from L to $L+2$. This is illustrated in the left part of Fig. 9. Upon addition of further vertices the number of dangling edges is kept fixed at $L+2$, until the Lth vertex completes the row, and we are back at L dangling edges. Thus the dimensions of the single-vertex matrices are $C_{L+2} \times C_L$ for $\mathbf{T}_1$, $C_{L+2} \times C_{L+2}$ for $\mathbf{T}_2, \ldots, \mathbf{T}_{L-1}$, and $C_L \times C_{L+2}$ for $\mathbf{T}_L$.

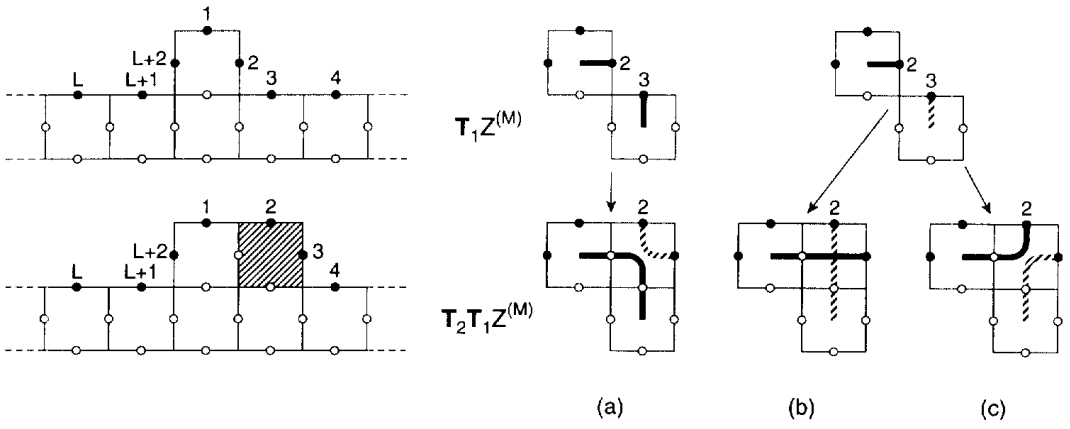


Fig. 10. Addition of subsequent vertices keeps the number of dangling edges fixed at $L + 2$. In the left part of the figure the system is shown before and after the addition of the second vertex (shaded). Vertex rules are displayed to the right. Situation (a) allows for the possibility of a black loop closure.

In Fig. 9 we illustrate the action of $\mathbf{T}_1$ on $Z^{(M)}$ in detail. To ensure that row $M + 1$, when completed, will have the same labels on its dangling edges as was the case in the preceding row, the solid dots illustrating the "active" dangling edges must be relabelled as shown in the lower left part of the figure. Shown to the right are the three possible choices of vertices fitting onto a black loop segment terminating at the dangling end 1 of $Z^{(M)}$. There are thus three non-zero entries in each column of $\mathbf{T}_1$. Since no loop closures of either flavour can be induced ($N'_{\mathrm{b}} = N'_{\mathrm{g}} = 0$ in Eq. (9.5)) all these entries are unity. Similar considerations hold true when the loop segment to be fitted is grey, and the vertex rules can be read off from the figure by interchanging the two flavours.

When acting with any one of the subsequent single-vertex TMs $\mathbf{T}_2, \ldots, \mathbf{T}_{L-1}$ the situation is as depicted in Fig. 10 for the case of $\mathbf{T}_2$. As the number of dangling edges is kept fixed no relabelling is needed, apart from the translation of labels 2 and 3 up on top of the newly added vertex. The vertex rules for the case where edge 2 of $\mathbf{T}_1 Z^{(M)}$ is black are shown in the right part of the figure; similar rules for the case where it is grey can be obtained by permuting the two flavours.

In situation (a) only one vertex fits onto the two dangling edges. The column of $\mathbf{T}_2$ determined by the number representation of the connectivity pertaining to the $L + 2$ dangling ends that are active in the upper part of the figure thus has only one non-zero entry. Its value is either n_{b} or 1 depending on whether a black loop closure is induced ($i_2^{\mathrm{b}} = i_3^{\mathrm{b}}$) or not ($i_2^{\mathrm{b}} \neq i_3^{\mathrm{b}}$). In the index representation of the new connectivity state $i_2^{\mathrm{g}} = i_3^{\mathrm{g}}$ is set equal to a positive integer not assumed by any other i_k^{g}. The new values of the black indices depend on whether a loop closure is induced or not. In the former case we simply set $i_2^{\mathrm{b}} = i_3^{\mathrm{b}} = 0$. In the latter, the two left-over black partners must be mutually connected before assigning $i_2^{\mathrm{b}} = i_3^{\mathrm{b}} = 0$.

Situations (b) and (c) correspond to two entries of each column of $\mathbf{T}_2$ taking the value unity, the others being zero. Since loop closures are out of the question the handling of these cases is simple. In (b) the two flavours cross, and the indices of sites 2 and 3 are interchanged. Case (c) is even simpler: it corresponds to a diagonal entry

in $\mathbf{T}_2$.

When strings are present a few modifications of the above rules are necessary. In situation (a), if one of i_2^{b} and i_3^{b} equals -1 and the other is positive, the left-over partner to the non-string black segment must be made the new string. And if both i_2^{b} and i_3^{b} equal -1 the corresponding entry of $\mathbf{T}_2$ must be forced to zero, since two strings cannot be allowed to annihilate.

Finally, consider closing the $(M+1)$th row through the action of $\mathbf{T}_L$, as depicted in Fig. 11. The labels $L+1$ and $L+2$ now disappear, and as far as the labelling goes the system is back in its original state. Each column of $\mathbf{T}_L$ has at most one non-zero entry per column, as witnessed by the vertex rules displayed in the right part of the figure. Once again, only half of the vertex rules are shown, and the other half is found by interchanging the two flavours.

In situation (a) no vertex of Fig. 3 can fit onto the three dangling edges at positions L, $L+1$ and $L+2$. The corresponding entry of $\mathbf{T}_L$ must therefore be forced to zero. Situations (b), (c) and (d) leave us to determine whether, for a given connectivity of the $L+2$ dangling edges, a black loop closure occurs or not. The handling in terms of the index representation is exactly as described above.

10. Numerical results

10.1. Central charge

The reduced free energy per vertex in the limit $M \to \infty$ of an infinitely long cylinder is given by

$$f_0^{(0,0)}(L) = \lim_{M\to\infty} \frac{1}{LM} \ln \operatorname{Tr} Z^{(M)} = -\frac{1}{L} \ln \lambda_0^{(0,0)}, \tag{10.1}$$

where $\lambda_0^{(s_{\mathrm{b}},s_{\mathrm{g}})}$ is the largest eigenvalue of $\mathbf{T}^{(s_{\mathrm{b}},s_{\mathrm{g}})}$. The partition function for a cylinder of length M is found by iterating the no-string TM

$$Z^{(M)} = \left(\mathbf{T}^{(0,0)}\right)^M Z^{(0)}. \tag{10.2}$$

It is well known that conformal invariance relates the amplitude of the $1/L^2$ corrections to $f_0^{(0,0)}(\infty)$ to the central charge c [37]. A further (non-universal) $1/L^4$ correction due to the operator $T\overline{T}$, where T denotes the stress tensor, must also be present in any conformally invariant system [48]. It is therefore found in a number of cases [45,49,50] that fits of the form

$$f_0^{(0,0)}(L) = f_0^{(0,0)}(\infty) - \frac{\pi c}{6L^2} + \frac{A}{L^4} \tag{10.3}$$

yield very rapidly converging estimates for c. An efficient application of Eq. (10.3) is to determine c from parabolic least-squares fits of the finite-size data against $1/L^2$ [49,50].

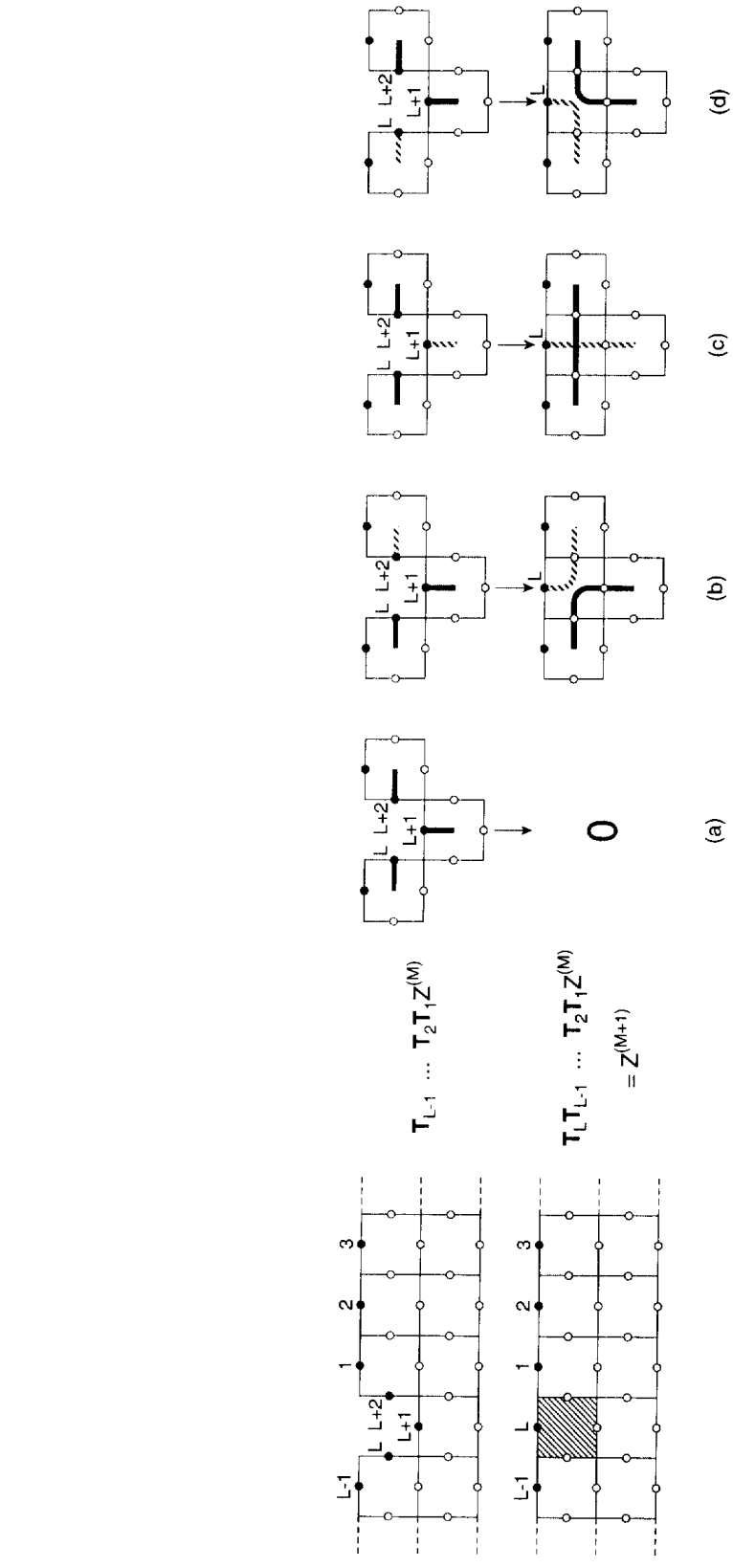


Fig. 11. Completing the $(M+1)$th row by adding the Lth vertex (shaded) brings the number of dangling edges back to L. The labelling is now consistent with that of the preceding rows. Vertex rules, shown to the right, now include a disallowed configuration. Namely, in situation (a) none of the vertices of Fig. 3 fit in, and the corresponding entry of the transfer matrix must be forced to zero. Situations (b), (c) and (d) offer various possibilities for a black loop closure.

Table 1
Estimants $c(L_0, L_{\max})$ for the central charge are obtained from parabolic least-squares fits against $1/L^2$ using the numerical data for $L_0 \leqslant L \leqslant L_{\max}$. The extrapolation in L_0 is described in the text

n_b	n_g	$c(4,14)$	$c(6,14)$	$c(8,14)$	$c(10,14)$	Extrapolation	Exact
0.0	0.0	−2.8943	−2.8861	−2.9220	−2.9514	−3.0037	−3.0000
0.5	0.0	−1.8528	−1.7641	−1.7716	−1.7873	−1.8152	−1.8197
0.5	0.5	−0.7295	−0.6249	−0.6159	−0.6220	−0.6328	−0.6395
1.0	0.0	−1.0012	−0.9542	−0.9636	−0.9761	−0.9983	−1.0000
1.0	0.5	0.1341	0.1877	0.1924	0.1895	0.1843	0.1803
1.0	1.0	0.9918	0.9969	0.9986	0.9999	1.0004	1.0000
1.5	0.0	−0.3765	−0.3669	−0.3817	−0.3923	−0.4111	−0.4124
1.5	0.5	0.7652	0.7746	0.7729	0.7715	0.7690	0.7678
1.5	1.0	1.6215	1.5818	1.5778	1.5806	1.5856	1.5876
1.5	1.5	2.2541	2.1691	2.1581	2.1627	2.1709	2.1751
2.0	0.0	0.0706	0.0549	0.0342	0.0235	−0.0019	0.0000
2.0	0.5	1.2209	1.1937	1.1868	1.1861	1.1849	1.1803
2.0	1.0	2.0792	2.0002	1.9899	1.9937	2.0005	2.0000
2.0	1.5	2.7139	2.5919	2.5737	2.5781	2.5859	2.5876
2.0	2.0	3.1629	3.0121	2.9885	2.9936	3.0027	3.0000

In Table 1 the results of such fits including the data points for $L_0 \leqslant L \leqslant L_{\max}$ are shown as a function of L_0. Numerically we were able to access $L_{\max} = 14$, in which case the largest single-vertex TMs have dimension $\sim 7 \times 10^6$ (see Table B.1). The extrapolation of the estimants $c(L_0, L_{\max})$ to the limit of infinite L_0 is assumed to take the form of a power law,

$$c(L_0, L_{\max}) = c + kL_0^{-p}, \tag{10.4}$$

at least within an asymptotic regime of large enough L_0. As is evident from Table 1 the last three estimants usually exhibit monotonicity, thus allowing us to fix the constants c, k and p. When this was not the case, or whenever the power p thus obtained was too small to produce a reliable extrapolation, the Ising-like value $p = 2$ was used by default to extrapolate the last two estimants. An error bar for this type of fit can be estimated from the variation among the individual estimants. The extrapolants are invariably in excellent agreement with our analytical results, the relative deviation being typically of the order 10^{-3}.

The results for c are shown for all integer and half-integer values of $n_i \in [0, 2]$. Because of the symmetric appearance of the two flavours in Eq. (9.1) only $n_b \geqslant n_g$ need be considered. For either $n_b = 1$ or $n_g = 1$ the FPL^2 model reduces to the simpler FPL model earlier considered by Batchelor et al. [14], and for $n_b = n_g$ we recover another special case recently investigated by one of us [7].

10.2. Thermal scaling dimension

A further prediction of conformal invariance is that the finite-size scaling of the first gap in the eigenvalue spectrum of $\mathbf{T}^{(0,0)}$ is related to the thermal scaling dimension [38]

Table 2
The thermal scaling dimension x_T. The extrapolation of the estimants $x_T(L_0, L_{\max})$ is described in the text. For comparison we also show the numerical data for the case of either n_b or n_g being unity [14]. Due to level crossing the values of x_T for $(n_b, n_g) = (0,0)$ and $(2,2)$ are found from the gap $f_2^{(0,0)}(L) - f_0^{(0,0)}(L)$ rather than from $f_1^{(0,0)}(L) - f_0^{(0,0)}(L)$

n_b	n_g	$x_T(4,14)$	$x_T(6,14)$	$x_T(8,14)$	$x_T(10,14)$	Extrapolation	Ref. [14]	Exact
0.0	0.0	0.5712	0.5280	0.5121	0.5060	0.4987		0.5000
0.5	0.0	0.5704	0.5535	0.5452	0.5417	0.5366		0.5372
0.5	0.5	0.5916	0.5882	0.5845	0.5825	0.5789		0.5804
1.0	0.0	0.5826	0.5798	0.5765	0.5748	0.5708	0.573 (1)	0.5714
1.0	0.5	0.6204	0.6227	0.6218	0.6211	0.6199	0.6200 (5)	0.6206
1.0	1.0	0.66368	0.66600	0.66642	0.66654	0.66663	0.6666 (1)	0.66667
1.5	0.0	0.5965	0.6053	0.6060	0.6058	0.6054		0.6063
1.5	0.5	0.6493	0.6559	0.6574	0.6578	0.6585		0.6619
1.5	1.0	0.7782	0.7094	0.7108	0.7115	0.7130	0.713 (1)	0.7146
1.5	1.5	0.8950	0.7657	0.7674	0.7684	0.7702		0.7699
2.0	0.0	0.6167	0.6295	0.6338	0.6349	0.6356		0.6667
2.0	0.5	0.7481	0.6878	0.6913	0.6927	0.6945		0.7345
2.0	1.0	0.8741	0.7566	0.7552	0.7565	0.7588	0.76 (1)	0.8000
2.0	1.5	0.9436	0.8755	0.8284	0.8303	0.8337		0.8702
2.0	2.0	0.9996	0.9850	0.9400	0.9200	0.8876		1.0000

$$f_1^{(0,0)}(L) - f_0^{(0,0)}(L) = \frac{2\pi x_T}{L^2} + \ldots, \tag{10.5}$$

where $f_1^{(0,0)}$ is found from the next-largest eigenvalue of $\mathbf{T}^{(0,0)}$ through $f_1^{(0,0)} = -(1/L)\ln\lambda_1^{(0,0)}$. These computations were also carried through for even L up to $L_{\max} = 14$. In this case as well the convergence of the estimants can be considerably sped up by including a $1/L^4$ term in Eq. (10.5) and performing parabolic least-squares fits versus $1/L^2$.

The results for x_T as displayed in Table 2 again agree with those of the previously studied special cases [14,7]. The data for $(n_b, n_g) = (0,0)$ merit a special comment. Monitoring the three leading eigenvalues $\lambda_0^{(0,0)}$, $\lambda_1^{(0,0)}$ and $\lambda_2^{(0,0)}$ as a function of n for $n_b = n_g \equiv n$ we found that $\lambda_1^{(0,0)}$ and $\lambda_2^{(0,0)}$ are exactly degenerate for all n down to $n \sim 0.20$. Hereafter $\lambda_1^{(0,0)}$ splits off from $\lambda_2^{(0,0)}$ and eventually becomes degenerate with $\lambda_0^{(0,0)}$ at $n = 0$. Because of this level crossing it thus seems very likely that near $(n_1, n_2) = (0,0)$ the thermal eigenvalue should be related to the gap $f_2^{(0,0)}(L) - f_0^{(0,0)}(L)$. Comparison with the exactly known result $x_T = 1/2$ [7] confirms this suspicion. A similar comment holds true near $(n_b, n_g) = (2,2)$, and again we find fair agreement with the exact result if we apply Eq. (10.5) to $\lambda_2^{(0,0)}$, and not to $\lambda_1^{(0,0)}$ (which in this case becomes two-fold degenerate).

For $n_b < 2$ the extrapolants are again in excellent ($\sim 10^{-3}$ or better) agreement with our analytical results. For $n_b = 2$ the slower convergence can be attributed to logarithmic corrections [51] arising from an enhanced number of marginal vertex operators. Indeed, of the twelve vertex operators corresponding to the shortest vectors in $\mathcal{R}_w^*$, Eq. (5.18), seven stay marginal when either $n_b < 2$ or $n_g < 2$. In the general case, when both

$n_b < 2$ and $n_g < 2$, there are only four marginal vertex operators; this is the loop ansatz, Eq. (5.25).

10.3. Dimensions of string operators

We now turn our attention to the determination of the scaling dimensions associated with one or more strings spanning the length of the cylinder. The presence of one black string corresponds to a height mismatch in the ideal states, and the relevant scaling dimension X is therefore that of a *twist-like operator* [41]. We have calculated the leading eigenvalue of $\mathbf{T}^{(1,0)}$ for odd system sizes up to $L_{\max} = 13$ and determined the corresponding estimants $c_{\text{odd}}(L_0, L_{\max})$ by the usual parabolic fits to $f_0^{(1,0)}(L)$, cf. Eq. (10.3). Estimants $X(L_0, L_{\max})$ are then defined by

$$X(L_0, L_{\max}) = \frac{c - c_{\text{odd}}(L_0, L_{\max})}{12}, \tag{10.6}$$

where the factor of 12 originates from a comparison of Eq. (10.3) with Eq. (10.5). For the central charge c of an even-sized system we use our analytical results, Eq. (6.4).

These estimants and their extrapolations are found in Table 3. Note that we can no longer limit the parameter values by $n_b \geqslant n_g$, as the condition $(s_b, s_g) = (1, 0)$ treats the two flavours asymmetrically. In the case of the FPL model ($n_b = 1$) it was found [14] that X was independent of n_g. It is evident from our numerical data that this n_g-independence in fact pertains to all $n_b \in [0, 2]$. Final results for X as a function of n_b have therefore been computed by averaging the available extrapolated scaling dimensions over n_g. For $n_b = 1$ the agreement with the result $X \approx 1/8$ found by Batchelor et al. [14] is excellent. Furthermore we are able to conjecture the general formula, Eq. (7.15), for X as a function of the loop fugacities.

When $(s_b, s_g) = (1, 1)$ the parity of L must again be even, and we can make parabolic fits for the gap $f_0^{(1,1)}(L) - f_0^{(0,0)}(L)$, as in Eq. (10.5), without taking resort to the less accurate method of fitting for two central charges separately as above. The corresponding universal amplitude is identified with the scaling dimension x_1. The results, now for $L_{\max} = 12$, are shown in Table 4, and our values for the scaling dimension are once again in agreement with the analytical results, apart from $n_g = 2$ where logarithmic corrections are the most likely source of systematic errors [51].

Finally, the results for x_2 as obtained from parabolic fits for the gap $f_0^{(2,0)}(L) - f_0^{(0,0)}(L)$ are shown in Table 5. Again we have $L_{\max} = 12$. Just like in the case of X we find the extrapolated values of x_2 to be independent of n_g, and final results are obtained by averaging over this parameter.

10.4. Entropy

Apart from the various universal quantities, such as the central charge and the scaling dimensions, the transfer matrices also provide numerical values for the residual entropy per vertex, $s = f_0(\infty)$. In the limit $n_g \to 0$ of compact polymers this quantity is of

Table 3
Estimants $X(L_0, L_{\max})$ for the scaling dimension of the twist operator along with their extrapolations to the infinite-system limit. For $n_b = 1$ the value $X = 1/8$ was previously found to be independent of n_g [14]. It is evident that this n_g-independence holds for any value of n_b, and in accordance herewith our final result is obtained by averaging the various extrapolants over n_g

n_b	n_g	$X(3,13)$	$X(5,13)$	$X(7,13)$	$X(9,13)$	Extrapolation	Result	Exact
0.0	0.0	−0.05586	−0.06109	−0.06203	−0.06232	−0.06257	−0.06269 (31)	−0.06250
0.0	0.5	−0.06080	−0.06197	−0.06220	−0.06233	−0.06253		−0.06250
0.0	1.0	−0.06043	−0.06198	−0.06221	−0.06233	−0.06250		−0.06250
0.0	1.5	−0.05869	−0.06156	−0.06215	−0.06233	−0.06259		−0.06250
0.0	2.0	−0.05804	−0.06190	−0.06297	−0.06316	−0.06324		−0.06250
0.5	0.0	0.04674	0.04538	0.04558	0.04569	0.04587	0.04583 (16)	0.04591
0.5	0.5	0.04572	0.04585	0.04589	0.04588	0.04588		0.04591
0.5	1.0	0.04643	0.04622	0.04614	0.04607	0.04595		0.04591
0.5	1.5	0.04781	0.04675	0.04638	0.04622	0.04590		0.04591
0.5	2.0	0.04828	0.04664	0.04593	0.04573	0.04555		0.04591
1.0	0.0	0.11895	0.12278	0.12398	0.12438	0.12501	0.12497 (8)	0.12500
1.0	0.5	0.12346	0.12422	0.12458	0.12470	0.12489		0.12500
1.0	1.0	0.12465	0.12485	0.12495	0.12496	0.12498		0.12500
1.0	1.5	0.12584	0.12540	0.12529	0.12521	0.12508		0.12500
1.0	2.0	0.12652	0.12549	0.12513	0.12501	0.12490		0.12500
1.5	0.0	0.17106	0.18253	0.18453	0.18536	0.18662	0.18663 (25)	0.18687
1.5	0.5	0.18283	0.18468	0.18553	0.18585	0.18633		0.18687
1.5	1.0	0.18515	0.18588	0.18620	0.18632	0.18646		0.18687
1.5	1.5	0.18684	0.18684	0.18687	0.18684	0.18680		0.18687
1.5	2.0	0.18878	0.18796	0.18759	0.18741	0.18696		0.18687
2.0	0.0	0.2076	0.2296	0.2321	0.2340	0.2369	0.2392 (27)	0.2500
2.0	0.5	0.2283	0.2323	0.2342	0.2351	0.2371		0.2500
2.0	1.0	0.2325	0.2347	0.2358	0.2363	0.2383		0.2500
2.0	1.5	0.2357	0.2372	0.2379	0.2383	0.2400		0.2500
2.0	2.0	0.2402	0.2413	0.2417	0.2420	0.2435		0.2500

interest to the protein folding community, due to the fact that native conformations of all globular proteins are compact [5].

Using our knowledge of the exact form of the finite-size corrections of order $1/L^2$, Eq. (10.3), we have obtained very accurate extrapolations to the limit of an infinite system.[8] After subtracting the $1/L^2$ correction a series of estimants $s(L, L_{\max})$ may be obtained by fitting the residual size dependence to a pure $1/L^4$ form. The remaining L-dependence of these estimants turns out to be well accounted for by a further $1/L^4$ fit, and in this way we arrive at a final value for s. The error bar on the final value can be estimated as its deviation from the most accurate extrapolant, $s(L_{\max} - 2, L_{\max})$.

The most accurate results are quite naturally found by employing this procedure on $f_0^{(0,0)}(L)$, and they are shown in Table 6. Results obtained by extrapolating the free energies for other sectors of the transfer matrix containing strings are consistent herewith

[8] The logarithmic corrections to the free energy implied by the $\mathcal{N}^{\gamma-1}$ term in Eq. (2.6) does not pertain to the cylindrical geometry implicit in our transfer matrix calculations. A similar remark applies to the surface term $\kappa_s \mathcal{N}^{(d-1)/d}$.

Table 4
Scaling dimension x_1, corresponding to one string of each flavour

n_b	n_g	$x_1(4,12)$	$x_1(6,12)$	$x_1(8,12)$	Extrapolation	Ref. [14]	Exact
0.0	0.0	−0.2433	−0.2447	−0.2470	−0.2500		−0.2500
0.5	0.0	−0.1328	−0.1295	−0.1303	−0.1313		−0.1323
0.5	0.5	−0.01713	−0.01228	−0.01217	−0.01217		−0.0131
1.0	0.0	−0.0440	−0.0423	−0.0430	−0.0439	−0.0444 (1)	−0.0446
1.0	0.5	0.0737	0.0763	0.0764	0.0765	0.0750 (3)	0.0761
1.0	1.0	0.16608	0.16646	0.16657	0.16663	0.1667 (1)	0.16667
1.5	0.0	0.0267	0.0271	0.0264	0.0255		0.0260
1.5	0.5	0.1466	0.1472	0.1472	0.1472		0.1483
1.5	1.0	0.2411	0.2395	0.2395	0.2394	0.242 (2)	0.2405
1.5	1.5	0.3196	0.3159	0.3156	0.3156		0.3162
2.0	0.0	0.0845	0.0848	0.0844	0.0839		0.1042
2.0	0.5	0.2070	0.2067	0.2071	0.2076		0.2295
2.0	1.0	0.3048	0.3021	0.3024	0.3028	0.307 (2)	0.3250
2.0	1.5	0.3882	0.3841	0.3842	0.3843		0.4044
2.0	2.0	0.4640	0.4618	0.4635	0.4657		0.5000

Table 5
Scaling dimension x_2, corresponding to two black strings

n_b	n_g	$x_2(4,12)$	$x_2(6,12)$	$x_2(8,12)$	Extrapolation	Result	Exact
0.0	0.0	0.0000	0.0000	0.0000	0.0000	0.0000 (0)	0.0000
0.0	0.5	0.0000	0.0000	0.0000	0.0000		0.0000
0.0	1.0	0.0000	0.0000	0.0000	0.0000		0.0000
0.0	1.5	0.0000	0.0000	0.0000	0.0000		0.0000
0.0	2.0	0.0000	0.0000	0.0000	0.0000		0.0000
0.5	0.0	0.1279	0.1355	0.1372	0.1389	0.1386 (2)	0.1386
0.5	0.5	0.1365	0.1371	0.1378	0.1387		0.1386
0.5	1.0	0.1377	0.1374	0.1379	0.1385		0.1386
0.5	1.5	0.1383	0.1375	0.1379	0.1384		0.1386
0.5	2.0	0.1392	0.1376	0.1379	0.1383		0.1386
1.0	0.0	0.2333	0.2447	0.2472	0.2504	0.2495 (5)	0.2500
1.0	0.5	0.2488	0.2477	0.2484	0.2493		0.2500
1.0	1.0	0.2514	0.2487	0.2490	0.2494		0.2500
1.0	1.5	0.2538	0.2497	0.2495	0.2492		0.2500
1.0	2.0	0.2573	0.2512	0.2504	0.2494		0.2500
1.5	0.0	0.3197	0.3377	0.3416	0.3466	0.3487 (26)	0.3506
1.5	0.5	0.3429	0.3425	0.3443	0.3466		0.3506
1.5	1.0	0.3486	0.3457	0.3466	0.3478		0.3506
1.5	1.5	0.3548	0.3497	0.3496	0.3494		0.3506
1.5	2.0	0.3636	0.3561	0.3547	0.3529		0.3506
2.0	0.0	0.3920	0.4202	0.4268	0.4353	0.446 (12)	0.5000
2.0	0.5	0.4244	0.4277	0.4323	0.4382		0.5000
2.0	1.0	0.4346	0.4348	0.4382	0.4426		0.5000
2.0	1.5	0.4468	0.4452	0.4474	0.4502		0.5000
2.0	2.0	0.4640	0.4618	0.4635	0.4657		0.5000

Table 6
Residual entropy s, obtained by extrapolating $f_0^{(0,0)}(L)$ to the infinite-system limit

n_b	n_g	$f_0^{(0,0)}(4)$	$f_0^{(0,0)}(6)$	$f_0^{(0,0)}(8)$	$f_0^{(0,0)}(10)$	$f_0^{(0,0)}(12)$	$f_0^{(0,0)}(14)$	s
0.0	0.0	0.17328680	0.28881133	0.31784496	0.32923359	0.33490107	0.33815371	0.346575 (14)
0.5	0.0	0.26740000	0.33317928	0.35057672	0.35745438	0.36088114	0.36284872	0.367950 (9)
0.5	0.5	0.35063553	0.37668215	0.38371283	0.38639599	0.38769210	0.38842126	0.390258 (3)
1.0	0.0	0.32923947	0.36764369	0.37752555	0.38137032	0.38327066	0.38435762	0.387166 (7)
1.0	0.5	0.40772622	0.41103439	0.41126990	0.41111188	0.41095017	0.41082815	0.410405 (2)
1.0	1.0	0.46298939	0.44576535	0.43960110	0.43671524	0.43513763	0.43418273	0.4315233 (4)
1.5	0.0	0.37601935	0.39599984	0.40063320	0.40233073	0.40314475	0.40360330	0.404771 (5)
1.5	0.5	0.45180855	0.43964968	0.43509788	0.43291844	0.43171625	0.43098591	0.4289459 (10)
1.5	1.0	0.50624745	0.47501911	0.46431698	0.45948057	0.45688890	0.45533728	0.4510742 (17)
1.5	1.5	0.54930614	0.50513652	0.49006459	0.48331974	0.47972832	0.47758588	0.471726 (2)
2.0	0.0	0.41389271	0.42018005	0.42097629	0.42111147	0.42113891	0.42114428	0.421145 (6)
2.0	0.5	0.48795109	0.46429984	0.45622604	0.45257417	0.45061901	0.44945033	0.4462607 (10)
2.0	1.0	0.54202495	0.50046092	0.48641918	0.48016010	0.47683419	0.47485271	0.4694505 (18)
2.0	1.5	0.58515036	0.53158535	0.51333087	0.50520022	0.50088581	0.49831761	0.491323 (3)
2.0	2.0	0.62122666	0.55918707	0.53795845	0.52850379	0.52348906	0.52050483	0.5123870 (19)

but have error bars that are roughly 10 times larger. If the fugacity of one of the strings equals two the error bars are even larger, which is to be anticipated from the fact that logarithmic corrections to the scaling dimensions are larger than similar corrections to the central charge [51].

In the special case of the equal-weighted six-vertex model, $(n_b, n_g) = (1, 1)$, our value for s is in excellent agreement with the exact result due to Lieb [52],

$$s(1,1) = \frac{3}{2}\ln\left(\frac{4}{3}\right) \simeq 0.4315231\ldots, \tag{10.7}$$

and in the limit of two mutually excluding Hamiltonian walks, $(n_b, n_g) = (0, 0)$, we are able to conjecture the result

$$s(0,0) = \frac{1}{2}\ln(2) \simeq 0.3465735\ldots \tag{10.8}$$

In fact, after having made this conjecture we discovered that the numerical values of $f_0^{(1,1)}(L)$, i.e. the free energy per site in the sector where we enforce one string of each flavour, are *independent* of L for $4 \leqslant L \leqslant 12$, and equal to $\frac{1}{2}\ln(2)$ with full 16-digit machine precision. Since the free energy per site in the thermodynamic limit is unchanged by the introduction of a string defect, this observation lends credibility to the correctness of the above conjecture.

Our result in the compact polymer limit merits special attention. Traditionally the entropy is quoted in terms of the so-called connective constant $\kappa = e^{s(1,0)}$; see Eq. (2.6). Early approximations due to Flory [53] and Huggins [54] yielded respectively

$$\kappa_{\text{Flory}} = \frac{z-1}{e} \simeq 1.104 \tag{10.9}$$

and

$$\kappa_{\text{Huggins}} = (z-1)\left(1-\frac{2}{z}\right)^{z/2-1} = \frac{3}{2}. \tag{10.10}$$

Here $z = 4$ is the coordination number of the square lattice. More recently, $\kappa \simeq 1.472$ was found from transfer matrix calculations [26] and $\kappa = 1.475(15)$ by exhaustive computer enumeration of short-chain configurations [4]. Both these results are very close to the mean-field value $\kappa_{\text{MF}} = z/e = 1.4715\ldots$ [24], and it is tempting to conclude that conformations of compact polymers are in fact described by mean-field theory [4]. However, our result

$$\kappa = 1.472801(10) \tag{10.11}$$

demonstrates that this is not the case.

11. Discussion

From the construction of the effective field theory of the FPL^2 model some rather general conclusions regarding the scaling of compact polymers, and the relation between loop models and conformal field theory can be drawn. It also provides new insights into the three-state Potts antiferromagnet and the dimer loop model, which are identified with specific points in the phase diagram of the FPL^2 model. We conclude the paper with a discussion of these topics.

11.1. Compact polymers

One of the main motivations for studying fully packed loop models is provided by compact polymers, their scaling properties in particular. Just like polymers in the dilute and dense phase, compact polymers form a critical geometrical system characterised by conformational exponents γ and ν. The exponent γ relates the number of conformations of the polymer to the number of monomers; see Section 2 for details. The other conformational exponent (ν) relates the linear size of the polymer to the number of monomers. For compact structures it has the trivial value $1/2$ since these polymers are space filling.

Prior to our work, exact results have been obtained for compact polymers on the Manhattan [55] and the honeycomb [20] lattice, and the mean-field value $\gamma = 1$ was found in both cases. This value of γ indicates that the two ends of the compact polymer are independent at large distances. This follows from the scaling relation $x_1 = 1 - \gamma = 0$, where the one-string dimension x_1 describes the probability $G_1(r) \sim r^{-2x_1}$ that the two chain ends are separated by a distance r. In this regard the scaling of compact polymers on the Manhattan and the honeycomb lattices is equivalent to that of ideal chains. Ideal chain configurations are described by simple random walks for which each step is independent of the previous one.

Here we have calculated the exact conformational exponent $\gamma = 117/112$ for compact polymers on the square lattice. The fact that $\gamma > 1$ is tantamount to an effective *repulsion* between the ends of the chain, indicating non-ideal behaviour. Indeed, the fact that the connective constant κ in Eq. (10.11) is larger than its mean-field value indicates that the origin of this repulsion is *entropic*. Earlier numerical studies of this problem utilising direct enumerations of chain conformations have failed to see any deviation from the ideal chain result $\gamma_{\mathrm{MF}} = 1$ [4]; we can attribute this to the fact that the actual difference is indeed very small ($\gamma - \gamma_{\mathrm{MF}} = 5/112$) and below the numerical accuracy previously achieved. The same comment can be made for the connective constant.

Another interesting aspect of compact polymers is that their scaling properties are lattice dependent. This is in contrast to the dilute and dense case which are described by conformational exponents that do not depend on the lattice type (e.g., honeycomb versus square). As remarked earlier this "lack of universality" is due to a kind of geometrical frustration that arises from the fully packing constraint imposed on the loop models which are employed in studies of compact polymers.

Finally, the field theory solution of the FPL^2 model uncovered a property of compact polymers that, to our knowledge, was not previously anticipated. The fact that there is a whole line of critical points in this loop model in the Hamiltonian walk limit ($n_{\mathrm{b}} \to 0$) indicates a continuum of universality classes described by compact polymers on the square lattice. In particular the exponent γ can be changed continuously by adjusting the fugacity of the loops uncovered by the polymer. The loop weight of the uncovered (grey) loops can be thought of as an effective interaction amongst the monomers, albeit a non-local one. A similar effect of interactions on directed self-avoiding walks was discovered by Cardy [30] from a field theoretical analysis of the problem. The existence of a continuously varying γ in this case was recently challenged by numerical results [31].

11.2. Relation to other models

The FPL^2 model is a loop model which exhibits a two-dimensional manifold of fixed points in its phase diagram. Certain points in the critical region map to previously studied lattice models and here we comment on the relevance of our results for these models.

11.2.1. Dimer loop model

The dimer loop model studied by Raghavan et al. [15] is the $n_{\mathrm{b}} = 2$, $n_{\mathrm{g}} = 1$ FPL^2 model; see Fig. 1. The dimer loop model is defined by placing black and white dimers on the square lattice so that every vertex of the lattice is covered by exactly one black and one white dimer. Every such configuration is given equal weight. The mapping to the FPL^2 model is achieved by identifying the bonds covered by dimers as making up the black loops, whilst the uncovered bonds form the grey loops. The original motivation for studying this dimer problem is that it leads to a height model with a two-component height; cf. the traditional dimer model which is described by a single component height.

Performing Monte Carlo simulations of the dimer loop model Raghavan et al. reached the conclusion that one of the two height components is rough whilst the other one is "anomalously smooth", i.e. its structure function decays at small wave-vectors $\mathbf{q}$ slower than $1/\mathbf{q}^2$; a $1/\mathbf{q}^2$ dependence is to be expected in a Gaussian field theory.

In the light of our results we would conclude that the dimer loop model is critical with a central charge $c = 2$. This follows from Eq. (6.4) for $n_b = 2$ and $n_g = 1$. The two components of the height found by Raghavan et al. should therefore both be rough, each contributing one to the central charge ($c = 1 + 1$). Furthermore, we believe that the observed anomalous behaviour of one of the heights can be attributed to the fact that this model is exactly at the boundary of the critical region of the FPL^2 model. We observe a similar effect in our numerical transfer matrix results which show largest deviations from the proposed exact formulae for loop fugacities at the critical-region boundary. The culprit might be logarithmic corrections due to the presence of marginal operators. To check this hypothesis and reconcile it with the fact that no such effects are seen in Monte Carlo simulations of the four-colouring model [28] ($n_b = n_g = 2$), which is also at the boundary of the critical region, simulations of the dimer loop model for larger system sizes would be welcome.

11.2.2. Three-state Potts antiferromagnet

The critical ground state of three-state Potts antiferromagnet maps to the equal-weighted six-vertex model [16] which is the $n_b = n_g = 1$ point in the critical region of the FPL^2 model; see Fig. 1. Along the line $n_b = n_g$ the colouring representation of the FPL^2 model has the additional symmetry with respect to cyclic permutations of the four colours; see Section 5.3.2. This explains the origins of the $\mathbb{Z}_4$ symmetry found by Saleur for the *three*-state Potts antiferromagnet [56].

11.2.3. Folding model

The folding model of the square-diagonal lattice recently investigated by Di Francesco [57] maps onto a constrained version of the $(n_b, n_g) = (2, 2)$ FPL^2 model. The constraint consists in allowing only the vertices 1, 3, 5 and 6 of Fig. 3 for sites on the even sublattice, and similarly vertices 2, 4, 5 and 6 on the odd sublattice.

We have modified our transfer matrices to take this constraint into account. Our result for the folding entropy, $s = 0.4604(4)$, is in complete agreement with Ref. [57].[9] Interestingly enough the finite-size scaling of the gaps in the eigenvalue spectrum seems to indicate that the model is not critical for general values of the loop fugacities. From the field theory of the FPL^2 model we should be able to understand why the constraint imposed by the folding model leads to a relevant perturbation which takes the system away from criticality. This we leave as an interesting open question. Incidentally, the situation is very reminiscent of the reformulation of the Q-state Potts model in terms of

[9] Our normalisation is "per vertex" whilst that of Di Francesco is "per triangle". Accordingly we find twice his result.

a staggered vertex model. Only at the critical point are the vertex weights on the even and odd sublattices identical, thus allowing for an exact solution of the model [32].

11.3. Conformal field theory

The Liouville field theory proposed for the effective theory of the FPL^2 model in the critical region is conformally invariant. Each point in the critical phase is characterised by the central charge and the scaling dimensions of primary fields, which are associated with electric and magnetic charges in the Coulomb gas. For generic values of the loop fugacities the background charge $\mathbf{e}_0$ is not commensurate with the electric charges that make up the lattice $\mathcal{R}^*$. This implies that amongst the electric operators there will be many (an infinite number, in fact) that have negative dimensions, signaling the non-unitary nature of the conformal field theory. Non-unitary CFTs appear in many other critical geometrical models, critical percolation being the best known example.

Liouville field theory provides the Euclidean action for the Coulomb gas description of conformal field theories proposed by Dotsenko and Fateev [36]. As such it contains the so-called screening charges which are the vertex operators that make up the Liouville potential. In the original formulation these charges were introduced on formal grounds so as to ensure the existence of non-vanishing four-point correlation functions in the theory. In order for these vertex operators not to disrupt the conformal symmetry of the modified Gaussian model (the modification is the addition of the boundary term to the gradient-square action) they are necessarily marginal, i.e. their scaling dimension is two.

Here we have found a physical interpretation of the screening charges. Their rôle in loop models is to ensure that the number of large loops from scale to scale stays of order one; this translates into the statement that the loop fugacities do not flow under the action of the renormalisation group.

The fact that we have a concrete physical interpretation of the screening charges directly leads to the calculation of the elastic constants in the Liouville field theory. In the traditional Coulomb gas approach these coupling constants are calculated by comparing with formulae derived from an exact solution of the model. Once these constants are known marginal vertex operators that play the rôle of screening charges can be written down. Our construction basically reverses this procedure, and by doing so *makes no reference to an exact solution.*

Finally, we end with a speculatory note concerning the prospects of solving the FPL^2 model via Bethe ansatz. Namely, all loop models to date have been solved by this method after mapping them to a vertex model, following a procedure analogous to the one outlined in Section 4. This does not seem to work for the FPL^2 model, at least not along the $n_g = 1$ line [14]. Why this is so is an interesting open question.

One possibility is that the *full* FPL^2 model needs to be considered as opposed to the FPL model studied by Batchelor *et al.* for which $n_g = 1$ is fixed. A more intriguing possibility is that a Bethe ansatz solution might be hindered (or made more difficult) by the non-trivial elasticity displayed by the FPL^2 model in its interface representation.

This statement we base solely on the observation that all previously solved loop models are simple as interface models in the sense that the height fluctuations are described by a single elastic constant. For the FPL^2 model, as described in Section 5, the stiffness tensor consists of three independent components. Whether indeed the interface representation of the loop model has any bearing on its Bethe ansatz solvability remains to be seen.

Acknowledgements

Useful discussions with M. Aizenman, J.L. Cardy, B. Duplantier, D.S. Fisher, C.L. Henley, T. Hwa, T. Prellberg, T. Spencer, F.Y. Wu, and C. Zeng are gratefully acknowledged. The authors would like to thank the Institute for Theoretical Physics at Santa Barbara, where this collaboration was initiated, for its warm hospitality. This research was supported in part by the Engineering and Physical Sciences Research Council under Grant GR/J78327, and by the National Science Foundation under Grant PHY94-07194, and DMS 9304580.

Appendix A. Dimensions of electric and magnetic operators

We calculate the scaling dimensions of electric and magnetic operators in the Coulomb gas theory described by the action

$$S_{\mathrm{CG}} = \frac{1}{2}\int d^2\mathbf{x}\, g_\alpha (\boldsymbol{\partial} H^\alpha)^2 + \frac{i}{4\pi}\int d^2\mathbf{x}\, (\mathbf{E}_0 \cdot \mathbf{H})\widetilde{\mathcal{R}}, \tag{A.1}$$

where $\widetilde{\mathcal{R}}$ is the scalar curvature. We are interested in the situation when the height field is defined on a flat surface, in which case $\widetilde{\mathcal{R}}$ is zero everywhere except at the boundaries.

A.1. Electric charges

The scaling dimension $x(\mathbf{E})$ of the electric-type operator $\exp(i\mathbf{E}\cdot\mathbf{H}(\mathbf{x}))$ follows from the two-point function

$$\left\langle \mathrm{e}^{i\mathbf{E}\cdot\mathbf{H}(\mathbf{x})}\mathrm{e}^{-i(\mathbf{E}-2\mathbf{E}_0)\cdot\mathbf{H}(\mathbf{y})}\right\rangle \sim |\mathbf{x}-\mathbf{y}|^{-2x(\mathbf{E})}, \tag{A.2}$$

where the expectation value is with respect to the measure defined by the action S_{CG}. The extra electric charge $2\mathbf{E}_0$ appears due to the charged boundary conditions enforced by the curvature term in the Coulomb gas action, Eq. (A.1).

We break up the calculation into two parts. First we calculate the two-point function, Eq. (A.2), in the absence of the background charge ($\mathbf{E}_0 = 0$). We make use of the property of Gaussian integrals,

$$\left\langle \mathrm{e}^{i\mathbf{E}\cdot\mathbf{H}(\mathbf{x})}\mathrm{e}^{-i\mathbf{E}\cdot\mathbf{H}(\mathbf{y})}\right\rangle = \exp\left(-\tfrac{1}{2}(E_\alpha)^2\left\langle (H^\alpha(\mathbf{x}) - H^\alpha(\mathbf{y}))^2\right\rangle\right), \tag{A.3}$$

and of the known propagator for the massless scalar field in two dimensions (where we have dropped the regulators at large and small distances),

$$\left\langle (H^\alpha(\mathbf{x}) - H^\alpha(\mathbf{y}))^2 \right\rangle = \frac{1}{\pi g_\alpha} \ln|\mathbf{x}-\mathbf{y}|. \tag{A.4}$$

Combining the above two equations and comparing the result with Eq. (A.2), we find

$$2x_{\mathrm{e}}^{(0)}(\mathbf{E}) = \frac{1}{2\pi g_\alpha}(E_\alpha)^2; \tag{A.5}$$

the superscript (0) is there to remind us that this formula is valid only for $\mathbf{E}_0 = 0$.

This result for the two-point function can be rewritten as

$$\left\langle \mathrm{e}^{i\mathbf{E}\cdot\mathbf{H}(\mathbf{x})} \mathrm{e}^{-i\mathbf{E}\cdot\mathbf{H}(\mathbf{y})} \right\rangle = \exp[\mathcal{E}_{\mathbf{E}}^{(0)}(\mathbf{x},\mathbf{y})], \tag{A.6}$$

where

$$\mathcal{E}_{\mathbf{E}}^{(0)}(\mathbf{x},\mathbf{y}) = -\frac{1}{2\pi g_\alpha}(E_\alpha)^2 \ln|\mathbf{x}-\mathbf{y}| \tag{A.7}$$

is the energy of two (vector) electric charges interacting via the two-dimensional Coulomb force; in this language S_{E} is the energy of the electrostatic field set up by the electric charges $\pm\mathbf{E}$, expressed in terms of the electrostatic potential $\mathbf{h}$. This seemingly trivial rewriting makes the calculation of $x(\mathbf{E})$, the electric dimension in the presence of a background charge, physically transparent.

To properly take into account the curvature term we define the height field over a disc of radius R, instead of the infinite plane, keeping in mind that at the end of the calculation we need to take the limit $R \to \infty$. In the case of the disc $\widetilde{\mathcal{R}} = 8\pi\delta(R)$, and the curvature term introduces a charge $-2\mathbf{E}_0$ at the disc boundary. Therefore, the vacuum of the modified Coulomb gas must contain a *floating charge* $+2\mathbf{E}_0$ in the disc interior, and the electrostatic energy of this charged vacuum is $\mathcal{E}_{2\mathbf{E}_0}^{(0)}(0,R) = -4\mathbf{E}_{0\alpha}^2 \ln(R)/2\pi g_\alpha$. Now, to find the scaling dimension of a vertex operator of charge $\mathbf{E}$, we imagine placing charges $+\mathbf{E}$ and $-\mathbf{E}$ at points $\mathbf{x}$ and $\mathbf{y}$ in the disc interior, and we calculate the total electrostatic energy with respect to the charged vacuum. The floating charge will coalesce with that of the oppositely charged vertex operator, $-\mathbf{E}$. Using Coulombs law, Eq. (A.7), we then calculate the interaction energy of charges $+\mathbf{E}$ at $\mathbf{x}$, $-\mathbf{E}+2\mathbf{E}_0$ at $\mathbf{y}$, and $-2\mathbf{E}_0$ at R, keeping in mind $R \gg |\mathbf{x}-\mathbf{y}|$. The final result

$$\mathcal{E}_{\mathbf{E}}(\mathbf{x},\mathbf{y}) = -\frac{1}{2\pi g_\alpha} E_\alpha (E_\alpha - 2E_{0\alpha}) \ln|\mathbf{x}-\mathbf{y}| \tag{A.8}$$

is obtained after the energy of the charged vacuum is subtracted. Now it is a simple matter to read off the scaling dimension as the negative coefficient in front of the logarithm,

$$2x(\mathbf{E}) = \frac{1}{2\pi g_\alpha} E_\alpha (E_\alpha - 2E_{0\alpha}). \tag{A.9}$$

This result can be derived in a more rigorous fashion by constructing the stress-energy tensor for the field theory S_{CG} and calculating its operator product with the vertex operator $\exp(i\mathbf{E}\cdot\mathbf{H})$ [36].

A.2. Magnetic charge

To calculate the magnetic dimension $x(\mathbf{M})$ we consider the ratio of partition functions,

$$Z_{>\mathbf{M}}(\mathbf{x},\mathbf{y})/Z_{>} \sim |\mathbf{x}-\mathbf{y}|^{-2x(\mathbf{M})}. \tag{A.10}$$

$Z_{>\mathbf{M}}(\mathbf{x},\mathbf{y})$ is the sum (path integral) over height configurations where a vortex and an antivortex, of topological charge $\pm\mathbf{M}$, are placed at positions $\mathbf{x}$ and $\mathbf{y}$ of the basal plane, whilst $Z_{>}$ is the unconstrained sum:

$$Z_{>} = \int \mathcal{D}\mathbf{H} \exp\left(-\frac{1}{2}\int d^2\mathbf{x}\, g_\alpha (\boldsymbol{\partial} H^\alpha)^2\right). \tag{A.11}$$

Here we have dropped the curvature term since it does not affect correlation functions of magnetic operators.

We can use the electrostatic analogy once again. Namely, we consider the interaction energy between two topological defects, $\mathcal{E}_{\mathbf{M}}(\mathbf{x},\mathbf{y})$. Since $Z_{>}$ is a Gaussian path integral, it follows that

$$Z_{>\mathbf{M}}(\mathbf{x},\mathbf{y})/Z_{>} = \exp[\mathcal{E}_{\mathbf{M}}(\mathbf{x},\mathbf{y})], \tag{A.12}$$

where

$$-\mathcal{E}_{\mathbf{M}}(\mathbf{x},\mathbf{y}) = \frac{g_\alpha}{2\pi}(M^\alpha)^2 \ln|\mathbf{x}-\mathbf{y}|. \tag{A.13}$$

The above interaction energy is calculated as the Gaussian action of the classical configuration of the height field, $\mathbf{h}_{\mathrm{c}}$. $\mathbf{h}_{\mathrm{c}}$ solves the classical equations of motion (Laplace's equation) with boundary conditions dictated by the presence of topological defects at $\mathbf{x}$ and $\mathbf{y}$ [34]. The scaling dimension of a magnetic-type operator is then the coefficient in front of the logarithm in Eq. (A.13),

$$2x(\mathbf{M}) = \frac{g_\alpha}{2\pi}(M^\alpha)^2. \tag{A.14}$$

Appendix B. Enumeration of the connectivities

The implementation of the transfer matrix (TM) for the FPL^2 model on a cylinder of width L and length M requires an enumeration of the possible connectivity states of the L points on the dangling edges of row M. Each of these L points can either

(i) be connected by $\mathcal{G}_M$ to one of the dangling edges of row 0 through a string of flavour $i = \mathrm{b}, \mathrm{g}$, or

(ii) be connected by $\mathcal{G}_M$ to one and only one other point in row M through a loop segment of flavour $i = \mathrm{b}, \mathrm{g}$.

A suitable representation of this information is furnished by a double state vector

$$\begin{pmatrix} i_1^{\mathrm{b}} i_2^{\mathrm{b}} i_3^{\mathrm{b}} \dots i_L^{\mathrm{b}} \\ i_1^{\mathrm{g}} i_2^{\mathrm{g}} i_3^{\mathrm{g}} \dots i_L^{\mathrm{g}} \end{pmatrix}, \tag{B.1}$$

which we shall refer to as the *index representation*. The indices i_k^{b} $(k = 1, 2, \dots, L)$ are defined as follows:

(i) $i_k^{\mathrm{b}} = i_l^{\mathrm{b}}$ is a (non-unique) positive integer if and only if points k and l are interconnected through a black string.

(ii) $i_k^{\mathrm{b}} = 0$ if and only if point k touches a grey string or loop segment.

(iii) $i_k^{\mathrm{b}} = -1$ if and only if point k is connected to a dangling edge of row 0 through a black string.

A similar definition is true for the indices i_k^{g} provided that one reads "grey" instead of "black" and *vice versa*. Two index representations are said to be identical if they are so up to the arbitrariness of the choice of positive integers. Also note that if $i_k^{\mathrm{b}} \neq 0$ we have $i_k^{\mathrm{g}} = 0$ and conversely.

A restriction on those indices that take positive values follows from the fact that loops of the same flavour are not allowed to intersect. Namely, if $j < k < l < m$ the equalities $i_j^{\mathrm{b}} = i_l^{\mathrm{b}}$ and $i_k^{\mathrm{b}} = i_m^{\mathrm{b}}$ cannot both be true. So in addition to being pairwise these connectivities are also *well-nested* [45]. The same is true for the grey indices, whereas there are no such restrictions when both flavours are involved. Indeed, connectivity states with $i_j^{\mathrm{b}} = i_l^{\mathrm{b}}$ and $i_k^{\mathrm{g}} = i_m^{\mathrm{g}}$ are explicitly allowed by the last two vertices shown in Fig. 3.

In practice we are only interested in the first few eigenvalues of TMs having a definite number of strings of each flavour. The relevant sectors of the TM are denoted $\mathbf{T}^{(s_{\mathrm{b}}, s_{\mathrm{g}})}$, where s_i is the number of strings of flavour $i = \mathrm{b}, \mathrm{g}$. The fully packing constraint means that we can only examine system sizes L that have the same parity as $s_{\mathrm{b}} + s_{\mathrm{g}}$. The various sectors have different physical interpretations and each requires a different enumeration of the connectivity states. Since the two flavours enter at an equal footing in the partition function, Eq. (9.1), we only need consider $s_{\mathrm{b}} \geqslant s_{\mathrm{g}}$. The $\mathbf{T}^{(0,0)}$ sector contains information about the free energy and the energy-like correlation length. The geometrical scaling dimensions x_1 and x_2 can be obtained from the $\mathbf{T}^{(1,1)}$ and the $\mathbf{T}^{(2,0)}$ sectors respectively. Finally the sector $\mathbf{T}^{(1,0)}$ gives the scaling dimension of the twist-like operator.

Whilst the index representation contains all information necessary for determining the value of a given entry in the TM it is obviously not suitable for labelling the entries. We therefore need another representation, the so-called *number representation*, in which the connectivities are labelled by the integers $1, 2, \dots, C_L^{(s_{\mathrm{b}}, s_{\mathrm{g}})}$, where $C_L^{(s_{\mathrm{b}}, s_{\mathrm{g}})}$ is the number of different connectivity states in the relevant sector. The practical implementation of the TMs relies on the mapping from the index to the number representation and its inverse.

We shall now consider, one by one, the various sectors of the TM.

B.1. $\mathbf{T}^{(0,0)}$ sector

When no strings are present all the L dangling edges of row M are pairwise connected with either a black or a grey loop segment. In particular L must be even. For any particular connectivity we can then decompose L as $L = 2p_\mathrm{b} + 2p_\mathrm{g}$, where $p_i \geqslant 0$ is the number of *pairs* of dangling edges covered by a flavour i loop segment. Since loops of different flavours are allowed to cross (see Fig. 3) the total number of connectivities is

$$C_L^{(0,0)} = \sum_{L=2p_\mathrm{b}+2p_\mathrm{g}} \binom{L}{2p_\mathrm{b}} c_{p_\mathrm{b}} c_{p_\mathrm{g}}, \tag{B.2}$$

where c_p is the number of pairwise well-nested connectivities of $2p$ points. The c_p's were first considered in the context of the Potts model [45], but were also found to play a central rôle in the TM formulation of the O(n) model [47]. We shall now briefly recall how they are evaluated.

Consider a well-nested pairwise connectivity of $2p$ points given by the index representation $(i_1 i_2 \ldots i_{2p})$. A recursion relation follows from observing that $i_1 = i_{2k}$ for precisely one integer $k \geqslant 1$. According to the well-nestedness criterion the sub-sequences $(i_2 i_3 \ldots i_{2k-1})$ and $(i_{2k+1} i_{2k+2} \ldots i_{2p})$ are both well nested, and indices occurring in one of them do not occur in the other. Hence for $p \geqslant 1$

$$c_p = \sum_{k=1}^{p} c_{k-1} c_{p-k}, \tag{B.3}$$

and $c_0 = 1$. By means of the generating function $P(x) = \sum_{p=0}^{\infty} c_p x^p$ it is readily shown [47] that

$$c_p = \frac{(2p)!}{p!(p+1)!}, \tag{B.4}$$

and that asymptotically $c_p \sim 4^p$.

Using Eqs. (B.2) and (B.4) we can now compute explicit values for the $C_L^{(0,0)}$. These are shown for $2 \leqslant L \leqslant 16$ in Table B.1.

For obvious reasons we shall call the function

$$\rho(i_1 i_2 \ldots i_{2p}) = k \tag{B.5}$$

defined by $i_1 = i_{2k}$ the *cut function* of the index representation $(i_1 i_2 \ldots i_{2p})$. A complete ordering of the well-nested sequences is now induced by applying the cut function first to the whole sequence, then recursively to its right and finally to its left part [45,47]. Accordingly, the mapping from the index to the number representation for a well-nested one-flavour connectivity is accomplished by

$$\sigma(i_1 i_2 \ldots i_{2p}) = \begin{cases} 1 & \text{if } p \leqslant 1 \\ \displaystyle\sum_{l=1}^{k-1} c_{l-1} c_{p-l} + \sigma(i_2 \ldots i_{2k-1}) & \\ \quad + [\sigma(i_{2k+1} \ldots i_{2p}) - 1] c_{k-1} & \text{otherwise,} \end{cases} \tag{B.6}$$

Table B.1
The number $C_L^{(s_b,s_g)}$ of FPL^2 connectivity states for L dangling edges accommodating s_i strings of flavour i = b, g. Only values of L with the same parity as $s_b + s_g$ are shown. When more than one string of any flavour is present further restrictions than the well-nestedness criterion apply, as described in the text. Accordingly the number $\tilde{C}_L^{(2,0)}$ is merely a useful upper limit on the true $C_L^{(2,0)}$. The efficiency of writing the TMs in the connectivity basis can be appreciated by comparing $C_L^{(0,0)}$ to 4^L, the latter being the dimensions of the TM written in the conventional colour basis, where every dangling end is labelled independently by **A**, **B**, **C** or **D**

L	4^L	$C_L^{(0,0)}$	$C_L^{(1,1)}$	$\tilde{C}_L^{(2,0)}$	L	$C_L^{(1,0)}$
2	16	2	2	1	1	1
4	256	10	24	12	3	6
6	4096	70	300	150	5	50
8	65 536	588	3920	1960	7	490
10	1 048 576	5544	52 920	26 460	9	5292
12	16 777 216	56 628	731 808	365 904	11	60 984
14	268 435 456	613 470	10 306 296	5 153 148	13	736 164
16	4 294 967 296	6 952 660	147 232 800	73 616 400	15	9 202 050

where the c_p are given by Eq. (B.4).

To give a complete specification of the connectivity of any *one* flavour in the state (B.1) we need to keep track of the positions of those indices that are zero. For a fixed number of z zero indices this is accomplished by the lexicographic ordering

$$\psi(i_1 i_2 \ldots i_L) = \begin{cases} 1 & \text{if } L = 1 \text{ or } z = L \\ \psi(i_2 i_3 \ldots i_L) & \text{if } i_1 \neq 0 \\ \binom{L-1}{z} + \psi(i_2 i_3 \ldots i_L) & \text{if } i_1 = 0, \end{cases} \tag{B.7}$$

assigning the lowest value to the sequence with all the zeros accumulated to the right.

The number representation of the two-flavour state (B.1) is now obtained by first ordering according to the number of indices i_k^b being zero, then lexicographically ordering the positions of these zero indices, and finally using the ordering (B.6), first on the well-nested subsequence of non-zero black indices and then on the corresponding grey subsequence. More precisely, the mapping from the index to the number representation in the $(s_b, s_g) = (0,0)$ sector is given by

$$\phi^{(0,0)}\begin{pmatrix} \mathbf{i}^b \\ \mathbf{i}^g \end{pmatrix} = \sum_{k=p_b+1}^{L/2} \binom{L}{2k} c_k c_{L/2-k} + [\psi(\mathbf{i}^b) - 1] c_{p_b} c_{p_g} + [\sigma(\tilde{\mathbf{i}}^b) - 1] c_{p_b} + \sigma(\tilde{\mathbf{i}}^g), \tag{B.8}$$

where $\mathbf{i}^b = (i_1^b i_2^b \ldots i_L^b)$ denotes the sequence of black indices and $\tilde{\mathbf{i}}^b$ the subsequence of the p_b pairs of non-zero indices (and, of course, similarly for the grey flavour).

The inversion of Eq. (B.8), so as to furnish a mapping from the number to the index representation, is straightforward if we know how to invert the functions σ and ψ. Details on this have already been given in Ref. [50].

B.2. $\mathbf{T}^{(1,0)}$ sector

In the case of one black string spanning the length of the cylinder the number of dangling edges in row M can be written as $L = 2p_b + 2p_g + 1$, where the p_i have the same meaning as above. In particular L must be odd.

The presence of *one* string of either flavour does not impose any additional restrictions on the connectivity states of the subsequence of positive indices of that flavour. Indeed, if the position of the string is given by $i_r^b = -1$ the non-zero subsequence of $(i_{r+1}^b \dots i_L^b i_1^b \dots i_{r-1}^b)$ is still well nested, and the arguments given above apply. The number of connectivity states is therefore found by multiplying the L possible positions of the string by the number of $(s_b, s_g) = (0,0)$ states of the remaining $L-1$ points

$$C_L^{(1,0)} = LC_{L-1}^{(0,0)}. \tag{B.9}$$

Explicit values are shown in Table B.1.

Similarly the mapping from the index to the number representation is found by first ordering after the position r of the string, and then after the value of $\phi^{(0,0)}$ taken of the remaining indices

$$\phi^{(1,0)}\begin{pmatrix} i_1^b\, i_2^b \dots i_L^b \\ i_1^g\, i_2^g \dots i_L^g \end{pmatrix} = (r-1)C_{L-1}^{(0,0)} + \phi^{(0,0)}\begin{pmatrix} i_1^b \dots i_{r-1}^b\, i_{r+1}^b \dots i_L^b \\ i_1^g \dots i_{r-1}^g\, i_{r+1}^g \dots i_L^g \end{pmatrix}. \tag{B.10}$$

B.3. $\mathbf{T}^{(1,1)}$ sector

When one string of each flavour is present $L = 2p_b + 2p_g + 2$ must be even, and again it suffices to augment the considerations from the $\mathbf{T}^{(0,0)}$ case by some book-keeping as to the positions of the two strings. Explicit values of

$$C_L^{(1,1)} = L(L-1)C_{L-2}^{(0,0)} \tag{B.11}$$

are shown in Table B.1.

Letting r_i denote the position of the string of flavour $i = b, g$ we find that

$$\begin{aligned}\phi^{(2,0)}\begin{pmatrix} \mathbf{i}^b \\ \mathbf{i}^g \end{pmatrix} &= [\,(r_b-1)(L-1) + (|r_g - r_b| - 1)\,]C_{L-2}^{(0,0)} \\ &\quad + \phi^{(0,0)}\begin{pmatrix} i_1^b \dots i_{r_b-1}^b i_{r_b+1}^b \dots i_{r_g-1}^b i_{r_g+1}^b \dots i_L^b \\ i_1^g \dots i_{r_b-1}^g i_{r_b+1}^g \dots i_{r_g-1}^g i_{r_g+1}^g \dots i_L^g \end{pmatrix}\end{aligned} \tag{B.12}$$

is the desired mapping from the index to the number representation.

A possible configuration of the system for $(s_b, s_g) = (1,1)$ is illustrated in Fig. 8, where the index representation of the connectivity state for each completed row is shown to the right of the figure.

B.4. $\mathbf{T}^{(2,0)}$ sector

Considering now the case of two black strings, it appears that the number of connectivity states for L even is given by

$$\tilde{C}_L^{(2,0)} = \binom{L}{2} C_{L-2}^{(0,0)}, \tag{B.13}$$

where we have simply divided Eq. (B.11) by 2 to take into account the indistinguishability of two strings of the *same* flavour. This is however not quite true, since for $L \geqslant 4$ the number (B.13) includes certain disallowed basis states. For $L = 4$ these are

$$\begin{pmatrix} -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{B.14}$$

The reason why these states are not valid is that, by definition of the allowed vertices (see Fig. 3), black loop segments cannot cross a black string. In general, therefore, any configuration where a positive black index is positionally separated from its "partner" by one or more black strings, *both* to the right and to the left (since periodic boundary conditions are assumed throughout), is not a valid one, even though the positive indices of each flavour satisfy the well-nestedness criterion. Accordingly, the true $C_L^{(2,0)}$ is less than the $\tilde{C}_L^{(2,0)}$ of Eq. (B.13).

We have not found it worthwhile to pursue the solution of this complication, since the numbers $\tilde{C}_L^{(2,0)}$ are already less than the $C_L^{(1,1)}$, and we need to diagonalise the transfer matrices $\mathbf{T}^{(1,1)}$ and $\mathbf{T}^{(2,0)}$ for the same values of L in order to determine the scaling dimensions x_1 and x_2 with the same numerical precision. Instead we found it efficient to construct all the $\tilde{C}_L^{(2,0)}$ basis states, list the number representations of those that are disallowed, and force the corresponding entries of $\mathbf{T}^{(2,0)}$ to zero.

With this proviso the mapping from the index to the number representation is

$$\begin{aligned} \phi^{(2,0)} \begin{pmatrix} \mathbf{i}^{\mathrm{b}} \\ \mathbf{i}^{\mathrm{g}} \end{pmatrix} &= [\psi(\mathbf{i}^{\mathrm{b}} + \mathbf{1}) - 1] C_{L-2}^{(0,0)} \\ &\quad + \phi^{(0,0)} \begin{pmatrix} i_1^{\mathrm{b}} \cdots i_{r_1-1}^{\mathrm{b}} i_{r_1+1}^{\mathrm{b}} \cdots i_{r_2-1}^{\mathrm{b}} i_{r_2+1}^{\mathrm{b}} \cdots i_L^{\mathrm{b}} \\ i_1^{\mathrm{g}} \cdots i_{r_1-1}^{\mathrm{g}} i_{r_1+1}^{\mathrm{g}} \cdots i_{r_2-1}^{\mathrm{g}} i_{r_2+1}^{\mathrm{g}} \cdots i_L^{\mathrm{g}} \end{pmatrix}, \end{aligned} \tag{B.15}$$

where r_1 and r_2 are the positions of the two black strings, and $\psi(\mathbf{i}^{\mathrm{b}} + \mathbf{1})$ means that we should lexicographically order the positions of the black indices that are -1.

References

[1] P.-G. de Gennes, Scaling Concepts in Polymer Physics (Cornell University Press, Ithaca, 1979).
[2] B. Nienhuis, Phys. Rev. Lett. 49 (1982) 1062.
[3] B. Duplantier and H. Saleur, Phys. Rev. Lett. 59 (1987) 539.
[4] C.J. Camacho and D. Thirumalai, Phys. Rev. Lett. 71 (1993) 2505.

[5] H.S. Chan and K.A. Dill, Macromolecules 22 (1989) 4559.
[6] B. Nienhuis, *in* Phase Transitions and Critical Phenomena, ed. C. Domb and J.L. Lebowitz, Vol. 11 (Academic, London, 1987).
[7] J. Kondev, Phys. Rev. Lett. 78 (1997) 4320.
[8] A.A. Belavin, A.M. Polyakov and A.B. Zamolodchikov, Nucl. Phys. B 241 (1984) 333.
[9] Conformal Invariance and Applications to Statistical Mechanics, ed. C. Itzykson, H. Saleur and J.-B. Zuber (World Scientific, Singapore, 1988).
[10] H.G. Evertz, The loop algorithm, cond-mat/9707221, and references therein.
[11] M. Aizenman and B. Nachtergaele, Commun. Math. Phys. 164 (1994) 17.
[12] J.T. Chalker and P.D. Coddington, J. Phys. C 21 (1988) 2665.
[13] D.-H. Lee, Phys. Rev. B 50 (1994) 10788.
[14] M.T. Batchelor, H.W.J. Blöte, B. Nienhuis and C.M. Yang, J. Phys. A 29 (1996) L399.
[15] R. Raghavan, C.L. Henley and S.L. Arouh, J. Stat. Phys. 86 (1997) 517.
[16] H. Park and M. Widom, Phys. Rev. Lett. 63 (1989) 1193.
[17] J. Bascle, T. Garel and H. Orland, J. Phys. A: Math. Gen. 25 (1992) L1323.
[18] B. Duplantier, *in* Fundamental Problems in Statistical Mechanics VII, ed. H. van Beijeren (North-Holland, Amsterdam, 1990).
[19] H.W.J. Blöte and B. Nienhuis, Phys. Rev. Lett. 72 (1994) 1372.
[20] M.T. Batchelor, J. Suzuki and C.M. Yung, Phys. Rev. Lett. 73 2646 (1994).
[21] J. Kondev, J. deGier and B. Nienhuis, J. Phys. A 29 (1996) 6489.
[22] D.A. Huse and A.D. Rutenberg, Phys. Rev. B 45 (1992) 7536.
[23] B. Duplantier and H. Saleur, Nucl. Phys. B 290 (1987) 291.
[24] H. Orland, C. Itzykson and C. de Dominicis, J. Phys. (Paris) 46 (1985) L353.
[25] A.L. Owczarek, T. Prellberg and R. Brak, Phys. Rev. Lett. 70 (1993) 951.
[26] T.G. Schmalz, G.E. Hite and D.J. Klein, J. Phys. A: Math. Gen. 17 (1984) 445.
[27] S. Higuchi, Phys. Rev. E 58 (1998) 128.
[28] J. Kondev and C.L. Henley, Phys. Rev. B 52 (1995) 6628.
[29] N. Read, *in* Proceedings of the Kagomé Workshop, ed. P. Chandra (NEC Laboratories, Princeton, 1992).
[30] J.L. Cardy, Nucl. Phys. B 419 (1994) 411.
[31] A. Trovato and F. Seno, Phys. Rev. E 56 (1997) 131.
[32] R.J. Baxter, Exactly Solved Models in Statistical Mechanics (Academic Press, New York, 1982).
[33] J. Cardy, *in* Fields, Strings and Critical Phenomena, ed. E. Brezin and J. Zinn-Justin (North-Holland, Amsterdam, 1990).
[34] P.M. Chaikin and T.C. Lubensky, Principles of condensed matter physics, Ch. 9 (Cambridge Univ. Press, Cambridge, 1995)
[35] M. Aizenman, Nucl. Phys. B 485 (1997) 551.
[36] Vl.S. Dotsenko and V.A. Fateev, Nucl. Phys. B 240 (1984) 312; B 251 (1985) 691.
[37] H.W. Blöte, J.L. Cardy and M.P. Nightingale, Phys. Rev. Lett. 56 (1986) 742;
I. Affleck, Phys. Rev. Lett. 56 (1986) 746.
[38] J.L. Cardy, J. Phys. A 16 (1983) L355.
[39] J. Kondev and C.L. Henley, Phys. Rev. Lett. 74 (1995) 4580.
[40] H. Saleur and B. Duplantier, Phys. Rev. Lett. 58 (1987) 2325.
[41] P. Ginsparg, *in* Fields, Strings, and Critical Phenomena, ed. E. Brézin and J. Zinn-Justin (North-Holland, Amsterdam, 1989).
[42] A. Kast, J. Phys. A: Math. Gen. 29 (1996) 7041.
[43] S.T. Chui and J.D. Weeks, Phys. Rev. B 14 (1976) 4978;
J.V. José, L.P. Kadanoff, S. Kirkpatrick and D.R. Nelson, Phys. Rev. B 16 (1977) 1217.
[44] R.J. Baxter, J. Math. Phys. 11 (1970) 784.
[45] H.W.J. Blöte and M.P. Nightingale, Physica A 112 (1982) 405.
[46] J. Cardy and J.L. Jacobsen, Phys. Rev. Lett. 79 (1997) 4063.
[47] H.W.J. Blöte and B. Nienhuis, J. Phys. A 22 (1989) 1415.
[48] J.L. Cardy, *in* Phase Transitions and Critical Phenomena, Vol. 11, ed. C. Domb and J.L. Lebowitz (Academic Press, London, 1987).
[49] S.L.A. de Queiroz, Phys. Rev. E 51 (1995) 1030.
[50] J.L. Jacobsen and J. Cardy, Nucl. Phys. B 515 (1998) 701.
[51] J.L. Cardy, J. Phys. A: Math. Gen. 19 (1986) L1093.

[52] E.H. Lieb, Phys. Rev. 162 (1967) 162.
[53] P.J. Flory, J. Chem. Phys. 10 (1942) 51.
[54] M.L. Huggins, J. Phys. Chem. 46 (1942) 151.
[55] P.W. Kasteleyn, Physica 29 (1963) 1329;
B. Duplantier, J. Stat. Phys. 49 (1987) 411.
[56] H. Saleur, Nucl. Phys. B 360 (1991) 219.
[57] P. Di Francesco, Nucl. Phys. B 525 (1998) 507.

5.8 Article "Conformal field theory of the Flory model"

Nous étudions la limite d'échelle d'un modèle de boucles compactes en deux dimensions dont les boucles possèdent une rigidité de courbure. Cette limite d'échelle est décrite par une théorie conforme à trois paramètres que nous caractérisons par sa représentation de gaz de Coulomb. Un choix de deux de ces trois paramètres reproduit la ligne critique du *modèle à six sommets* qui est exactement soluble. Un autre choix correspond au *modèle de Flory* de la fusion de protéines. Nous calculons les valeurs exactes de la charge centrale et des exposants critiques du modèle de Flory, au point de fusion. Nous montrons que la fusion de protéines, comme décrite par le modèle de Flory, est une transition de phase de *second* ordre, en contradiction avec la théorie de champ moyen [138]. Les résultats de la théorie conforme sont comparés avec des calculs numériques du type matrice de transfert.

Conformal field theory of the Flory model of polymer melting

Jesper Lykke Jacobsen
Laboratoire de Physique Théorique et Modèles Statistiques
Université Paris-Sud
Bâtiment 100, F-91405 Orsay, FRANCE
JACOBSEN@IPNO.IN2P3.FR

Jané Kondev
Physics Department, MS057, Brandeis University
Waltham, MA 02454, USA
KONDEV@BRANDEIS.EDU
(Dated: September 23, 2002)

We study the scaling limit of a fully packed loop model in two dimensions, where the loops are endowed with a bending rigidity. The scaling limit is described by a *three-parameter* family of conformal field theories, which we characterize via its Coulomb-gas representation. One choice for two of the three parameters reproduces the critical line of the exactly solvable six-vertex model, while another corresponds to the Flory model of polymer melting. Exact central charge and critical exponents are calculated for polymer melting in two dimensions. Contrary to predictions from mean-field theory we show that polymer melting, as described by the Flory model, is *continuous*. We test our field theoretical results against numerical transfer matrix calculations.

I. INTRODUCTION

Over the years, polymers physics has greatly benefited from studies of lattice models. One persistent theme has been the use of lattice models to uncover universal properties of chain molecules. An example is provided by the scaling exponents which characterize the statistical properties of polymer conformations, in the limit of very long chains [1]. For polymer chains confined to live in two dimensions, exact values of exponents were calculated by Nienhuis [2] using the self-avoiding walk on the honeycomb lattice. The predicted value of the swelling exponent, which relates the linear size of the polymer to the number of monomers, was directly measured in recent fluorescence microscopy studies of DNA absorbed on a lipid bilayer [3].

Here we turn to the problem of polymer melting, which deals with a possible phase transition induced by the competition between chain entropy and bending rigidity. Bending rigidity determines the persistence length of the polymer. This is the distance over which the relative orientations of two chain segments are decorrelated due to thermal fluctuations. The long chain limit mentioned in the previous paragraph is obtained when the polymer length is much greater than its persistence length.

It is important to point out that the effect of finite bending rigidity depends crucially on the steric constraints imposed on the polymer by its interactions with the solvent. For example, in the presence of a good solvent the polymer is in a "dilute" phase. Typical chain conformations are swollen with empty space between the monomers filled by solvent molecules. On the lattice, the dilute phase is characterized by a vanishing fraction of sites occupied by monomers. In this phase, the bending rigidity simply increases the persistence length of the polymer, and it does not lead to a phase transition. This can be verified analytically in two dimensions, within the framework of Nienhuis' self-avoiding walk model [4, 5].

The picture changes considerably when the polymer is in a "compact" phase, with the monomers occupying all the available space. Such a situation is relevant, for instance, when modelling the conformations of globular proteins [6]. Compactness in this case follows from the interaction between hydrophobic amino-acids and the solvent (water), which leads to the expulsion of the solvent from the bulk of the protein. The simplest way to model this effect is to enforce compactness as a global, steric constraint on the polymer configurations [6]. Within this compact phase, one expects a phase transition from a disordered melt to an ordered crystal as the stiffness of the polymer is increased.

To study this melting transition, in 1956 Flory introduced a lattice model [7]. Flory's model, in its simplest formulation, consists of a single chain, described by a self-avoiding walk on the square lattice, endowed with a bending rigidity. To describe the melted phase the chain is taken to be maximally compact, filling all the sites of the square lattice; see Fig. 1. The resistance to bending is modelled by an energy penalty for making 90° turns.

In the Flory model, at infinite temperature the entropy dominates and the polymer will exhibit a finite density of bends, as in Fig. 1a. As the temperature is lowered to zero all the bends are expelled from the bulk and their density goes to zero, as in Fig. 1b. The nature of the transition from the high temperature melt to the low temperature crystal has been debated over the years [8]. Here we show that the melting transition is *continuous* and calculate exact values of scaling exponents at the transition.

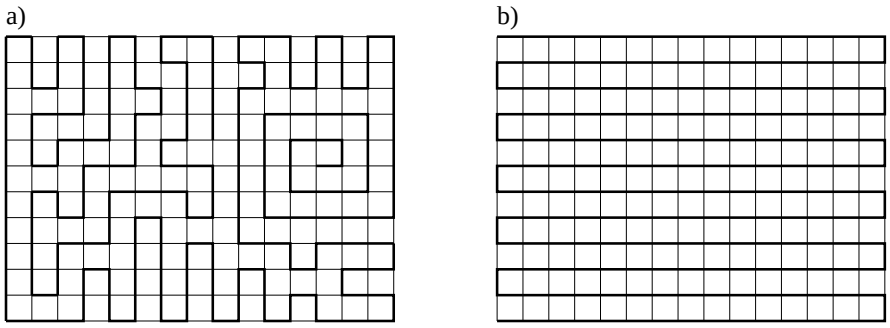


FIG. 1: Compact polymer configurations on an 11×15 square lattice: a) Typical configuration in the melt phase, and b) zero-temperature crystalline state, in which the number of bends is minimum.

In his original paper, Flory [7] proposed a mean-field treatment which predicts a first order transition. According to [7], the density of bends goes to zero at the transition and the chain entropy vanishes. This prediction of a first order transition with a vanishing entropy was challenged by Nagle [9]. Namely, he showed that the exactly solvable six-vertex model maps to a related polymer model which differs from Flory's by the presence of polymer loops of all sizes. Applying Flory's mean-field approximation to this model leads once again to the prediction of a first order melting transition. However, as Nagle pointed out, this is at odds with the exact solution of the six-vertex model [10] which predicts a continuous, infinite order transition. This observation makes it questionable that the Flory approach is valid in the original model as well. In fact, a few years later Gujrati and Goldstein [11] proved that the polymer entropy in Flory's model stays finite all the way down to zero temperature when it finally vanishes. However, the order of the transition still remained an unresolved question.

Monte Carlo simulations of Baumgartner and Yoon [12], where they allowed for many chains and a finite density of empty sites, showed a first order melting transition. Soon thereafter Saleur [13], using a transfer matrix approach, presented numerical evidence of a continuous transition, similar to the one found in the six-vertex model. More recently, Bascle, Garel and Orland [14] proposed an improved mean-field treatment of the Flory model, which does not suffer from the problem of a vanishing entropy at the transition. It also predicts a first order transition.

Here we show that polymer melting is continuous, as originally argued by Saleur [13], by making use of a particular model, the *semiflexible loop* (SFL) *model*, and its height representation. Furthermore we calculate the central charge and exact scaling exponents at the transition. These results are checked against detailed numerical transfer matrix computations.

The SFL loop model can be thought of as a "loop generalization" of the so-called F-model [9], in which suitably defined loops carry additional Boltzmann weights. The F-model is a special case of the six-vertex model [10], in which all vertices carry equal weights. This connection will serve as the motivation for introducing a more general model, the *generalized six-vertex model*, in which the general (zero-field) six-vertex model is endowed with extra loop weights. We shall finally introduce a similarly generalized version of the eight-vertex model [10]. Its interest from a polymer point of view is that it allows for a unified description of semiflexible lattice polymers in a variety of phases: compact, dense and dilute. Furthermore it allows us to discuss the effect of vacancies on the polymer melting transition.

The paper is organized as follows. In the next section we introduce the SFL model, which, in the limit of zero loop weight, gives the Flory model of polymer melting, and we discuss its phase diagram. In Sec. III we discuss the height representation of the loop model and how it leads to a conformal field theory in the scaling limit. We make use of the field theory in Sec. IV to calculate the central charge and scaling exponents, which we check against numerical transfer matrix computations in Sec. V. In Sec. VI we propose a phase diagram for the generalized six-vertex and eight-vertex models. We end with a discussion of the scaling of semiflexible compact polymers, and we argue that the generalized eight-vertex model furnishes a rather complete description of non-compact semiflexible polymers. An appendix is reserved for a detailed discussion of the construction of the transfer matrices.

II. SEMIFLEXIBLE LOOP MODEL

Here we define the SFL model, and give a rough sketch of its phase diagram based on the limits of weak and strong bending rigidity. The fact that the SFL model reduces to the F-model in the limit of unit loop fugacity [9], plays an important role in guiding our intuition about the loop model. It also provides an exactly solvable line in the phase

diagram, against which the field theoretical and numerical results can be checked.

A. Definition of the model

The semiflexible fully packed loop model on the square lattice (the "semiflexible loop model", or SFL for short) is defined by filling the square lattice with loops drawn along the lattice edges. Allowed loop configurations satisfy two constraints:

- Self avoidance — loops are not allowed to cross, and
- Full packing — every site is visited by exactly one loop.

On the square lattice with periodic boundary conditions, edges that are not covered by loops also form loops, as there are two unoccupied edges associated with every site of the lattice. These we refer to as "ghost loops".

Given the configurations of the semiflexible loop model, the Boltzmann weights are defined in the following way. Every real loop is given weight $n_{\rm b}$, and every ghost loop has weight $n_{\rm g}$. (In all the figures the real and ghost loops are shown as black and gray respectively, whence the subscripts b and g.) The parameters $n_{\rm b}$ and $n_{\rm g}$ act as fugacities of the two loop flavors, and as such they control the average number of loops of each flavor [32]. They can be varied independently as the number of ghost loops is not fixed by the number of real loops [15]. Furthermore, a weight $w_{\rm X}$ is assigned to each vertex of the lattice at which the real and ghost loops cross. For $w_{\rm X} > 1$ this has the effect of disfavoring vertices at which the loop makes a 90° bend, or, in other words, the loops are semiflexible. The partition function of the semiflexible loop model is

$$Z = \sum_{\mathcal{G}} n_{\rm b}^{N_{\rm b}} n_{\rm g}^{N_{\rm g}} w_{\rm X}^{V} \ , \tag{1}$$

where the sum runs over all allowed loop configurations $\mathcal{G}$. $N_{\rm b}$ and $N_{\rm g}$ are the number of real and ghost loops, respectively, while V is the number of crossing vertices; these are the two rightmost vertices in Fig. 3. In the limit $n_{\rm b} \to 0$, with $n_{\rm g} = 1$, we recover the Flory model: $Z/n_{\rm b}$ counts compact polymer loops each weighed by $w_{\rm X}^{V}$.

The semiflexible loop model can be thought of as the generalization of the FPL^2 model introduced in Ref. [15]. The FPL^2 model is given by the partition function, Eq. (1), with $w_{\rm X} = 1$. It has a critical phase for $|n_{\rm b}|, |n_{\rm g}| \leq 2$, characterized by a power law distribution of loop sizes. For other values of the loop weights the model is non-critical with a distribution of loop sizes cut off at a finite value (fixed by the correlation length). Below we will show that the vertex weight $w_{\rm X}$, for each point in the critical phase of the FPL^2 model, produces a line of fixed points which terminates in a Kosterlitz-Thouless transition.

B. Qualitative phase diagram

Rough, qualitative features of the phase diagram of the semiflexible loop model can be deduced from the limits of zero and infinite bending rigidity. The motivation for developing a precise theory of the phase diagram, as mentioned in the introduction, stems from the interest in the $n_{\rm b} \to 0$, $n_{\rm g} = 1$ case, which is the Flory model of polymer melting. We are also motivated by the relation of the SFL model to the integrable six-vertex model, and its generalizations.

1. Flory model

In the Flory limit of the SFL model, the $w_{\rm X} = 1$ point is the compact polymer problem, which we have studied previously [15]. Here one is concerned with enumerating all self-avoiding walks that visit every site of the lattice. We have shown that compact polymers on the square lattice are a critical geometry characterized by non-mean-field scaling exponents which can be calculated exactly from a field theory.

As $w_{\rm X}$ is increased away from one, we are dealing with a compact polymer with a bending rigidity. In the limit $w_{\rm X} \to \infty$ we arrive at a frozen phase in which the density of vertices at which the polymer bends goes to zero. This is the polymer crystal. At an intermediate weight $w_{\rm X} = w_{\rm X}^{\rm c}$ ($1 < w_{\rm X}^{\rm c} < \infty$) there will be a melting transition. One of the important unresolved problems is the nature of this transition. Here we construct an effective field theory of the Flory model and show that the melting transition is *continuous*.

Another interesting issue is the region of $0 \leq w_{\rm X} < 1$. As $w_{\rm X} \to 0$ straight-going vertices are completely suppressed, and with appropriate boundary conditions the only allowed configurations are those of a checkerboard pattern of small

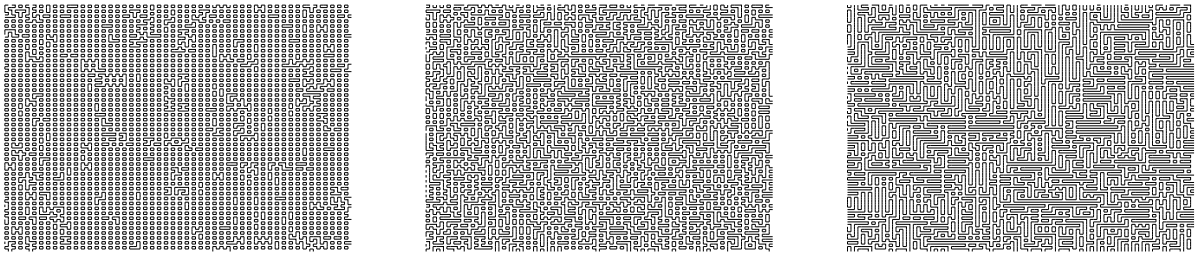


FIG. 2: Typical configurations in the SFL model with $n_{\rm b} = n_{\rm g} = 2$ and bending rigidity parameter $w_{\rm X} = 1/4$ (left panel), $w_{\rm X} = 1$ (middle) and $w_{\rm X} = 4$ (right). We shall show that the left and middle panels correspond to critical melt states, while the right panel is a non-critical crystalline state. In the latter, domains of non-zero staggered polarization (see Sec. II B 2) are clearly visible.

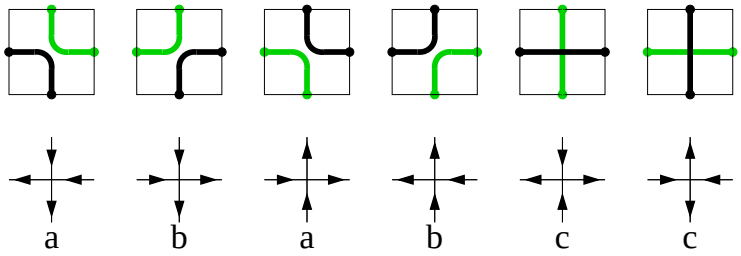


FIG. 3: Correspondence between the vertices of the six-vertex model and the FPL^2 model (here shown for an even vertex; at odd vertices the arrows are reversed).

loops, each loop having its minimal length of four. If the Flory limit ($n_{\rm b} \to 0$) is taken before the $w_{\rm X} \to 0$ limit, there has to be a number of straight-going vertices at the boundary, the dominant configurations being those of a single wiggly line. In any case, the $w_{\rm X} \to 0$ limit is again a crystalline phase of zero entropy. We shall however argue below that the corresponding crystallization transition is located at $w_{\rm X} = 0$ and is thus rather uninteresting.

A qualitative idea of the physics underlying the phase diagram of the SFL model can be obtained by looking at some typical configurations for various values of $w_{\rm X}$; see Fig. 2. The images were obtained by performing Monte Carlo simulations on a square lattice of size 100×100 with toroidal boundary conditions. For technical reasons [16] we take $n_{\rm b} = n_{\rm g} = 2$ and no loops of non-contractible topology are allowed. (Further details on the algorithm used for these simulations can be found in [16].)

2. *Six-vertex model*

Before turning our sights to the semiflexible loop model it is instructive to review exact results for the (zero-field) six-vertex (6V) model. The 6V model corresponds to the $n_{\rm b} = n_{\rm g} = 1$ line in the phase diagram of the SFL model. The mapping between the two is simple: at even (odd) vertices the edges covered by the real loops are identified with arrows pointing out (in), while the edges covered by the ghost loops correspond to arrows pointing in (out); see Fig. 3. The appropriate six-vertex weights are: $a = b = 1$ and $c = w_{\rm X}$ [33].

In the 6V model there is an order-disorder transition as a function of the vertex weight $w_{\rm X}$. In the ordered state, which is obtained for $w_{\rm X} \to \infty$, all the vertices are of the c-variety (cf. Fig. 3). The order parameter is the staggered polarization, which in the loop language can be expressed as the difference between the number of horizontal and the number of vertical loop-covered edges, per site [13]. The exact solution of the six-vertex model predicts a continuous (infinite-order) transition occurring at $w_{\rm X} = 2$ [10]. The disordered phase for $0 < w_{\rm X} < 2$ is critical with an infinite correlation length and power-law correlations. Below we will show that there is an analogous transition in the semiflexible loop model, as $w_{\rm X}$ is varied, for all values of $|n_{\rm b}|, |n_{\rm g}| \leq 2$, including the Flory case ($n_{\rm b} \to 0, n_{\rm g} = 1$). In the Flory model this was observed previously by Saleur in numerical transfer matrix computations [13]. For the

critical phase of the model we shall construct an effective field theory using the interface representation of the loop model. This leads to exact (but non-rigorous) results for the central charge and scaling dimensions, which we confirm via numerical transfer matrix calculations.

III. FIELD THEORY CONSTRUCTION

To construct a field theory for the critical phase of the semiflexible loop model (SFL) we make use of the height representation of the fully packed loop model on the square lattice (FPL^2). This was already described in detail in our previous work [15], and here it is briefly reviewed for completeness. The main effect of the vertex weight $w_{\rm X}$ on the field theory is to renormalize one of its coupling (elastic) constants. This does not change the central charge, but leads to continuously varying scaling dimensions for a specific subset of operators, which we identify. These results are confirmed by our numerical transfer matrix computations.

A. Height map

The height mapping is defined on the space of oriented loop configurations $\{\mathcal{G}'\}$. We associate $2^{N_{\rm g}+N_{\rm b}}$ oriented loop configuration $\mathcal{G}'$ with each loop configuration $\mathcal{G}$ of the SFL model by independently orienting every real and every ghost loop clockwise or counterclockwise.

The Boltzmann weight of an oriented loop is $\exp(i\phi)$, where the phase $\phi = \pm\pi e_{\rm b}$ for clockwise (counterclockwise) oriented real loops, and $\phi = \pm\pi e_{\rm g}$ for the two orientations of the ghost loops. To recover $n_{\rm b}$ and $n_{\rm g}$ for the loop weights, after summing over the two possible orientations we must set

$$\begin{aligned} n_{\rm b} &= 2\cos(\pi e_b) \\ n_{\rm g} &= 2\cos(\pi e_g) \ . \end{aligned} \tag{2}$$

This particular partition of the loop weights between the two orientations has the advantage of allowing the loop weights to be distributed among all the vertices that the loop visits, thus rendering the weights local. This is achieved by assigning the phase $\pi e_{\rm b}/4$ ($\pi e_{\rm g}/4$) to every vertex at which the oriented real (ghost) loop makes a right turn, and the opposite phase for left turns. The fact that for every closed loop on the square lattice the difference between the number of left and right turns is ± 4, is what makes these vertex weights work. The total vertex weight is then given by the product of phase factors when the loops bend, while a weight $w_{\rm X}$ is assigned to vertices at which the loops do not bend.

Turning back to the height map, we define microscopic heights $\mathbf{h}(\mathbf{x})$ on the lattice $\{\mathbf{x}\}$ dual to the square lattice on which the loops are defined. Once the height at the origin is fixed, the heights on all the other vertices of the dual lattice are uniquely specified by the oriented loop configuration. Namely, the height difference between nearest neighbor vertices of the dual lattice is $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ or $\mathbf{D}$, depending on the state of the edge that separates them. The four height-difference vectors, also referred to as "colors", are associated with the four possible states of any given edge, which can be either covered by a real or a ghost loop, with one of two possible loop orientations. Real loops are formed by alternating cycles of $\mathbf{A}$ and $\mathbf{B}$ colored edges, while the $\mathbf{C}$ and $\mathbf{D}$ colored edges are ones visited by the ghost loops. Note that the difference between an $\mathbf{ABAB}\cdots$ and a $\mathbf{BABA}\cdots$ cycle encodes the orientation of the corresponding (real) loop.

The fully packing constraint and the requirement that the height be unique (i.e., the sum of height differences along any closed lattice path must be zero), imposes a single algebraic constraint on the four colors: $\mathbf{A}+\mathbf{B}+\mathbf{C}+\mathbf{D}=0$. It follows that only three of the four vectors are linearly independent. A convenient choice that respects the symmetries between the four colors is to let the corresponding vectors point from the center to the vertices of a regular tetrahedron:

$$\begin{aligned} \mathbf{A} &= (-1,+1,+1)\ , \quad & \mathbf{B} &= (+1,+1,-1)\ , \\ \mathbf{C} &= (-1,-1,-1)\ , \quad & \mathbf{D} &= (+1,-1,+1)\ . \end{aligned} \tag{3}$$

The effective field theory for the SFL model describes the fluctuations of the coarse-grained heights which retains only the long-wavelength (much larger than the lattice spacing) Fourier modes of the microscopic heights.

B. Effective field theory: $w_{\rm X} = 1$

For $w_{\rm X} = 1$ we have the familiar case of the fully packed loop model on the square lattice. Its effective field theory was discusses in a previous publication [15] and here it is reviewed for completeness.

The partition function of the loop model in the height representation can be written as a path integral over the coarse grained heights with the (dimensionless) action:

$$S = S_{\rm E} + S_{\rm B} + S_{\rm L} \ . \tag{4}$$

This action only takes into account the long-wavelength fluctuations of the microscopic height. The three terms in the action are of different origin.

The elastic term,

$$S_{\rm E} = \frac{1}{2}\int {\rm d}^2{\bf x}\ \left\{K_{11}[(\boldsymbol{\partial} h^1)^2 + (\boldsymbol{\partial} h^3)^2] + 2K_{13}(\boldsymbol{\partial} h^1 \cdot \boldsymbol{\partial} h^3) + K_{22}(\boldsymbol{\partial} h^2)^2\right\} \ , \tag{5}$$

accounts for the height fluctuations due to the entropy of fully packing the square lattice with oriented loops. Equivalently, this is the entropy of edge coloring the square lattice with four different colors. The elastic term favors oriented loop configurations that minimize the variance of the microscopic height; these are the macroscopically flat states. In terms of the color degrees of freedom the flat states have the property that the four edges of each elementary plaquette are colored by two colors only.

The particular form of the matrix of elastic constants, $\mathbf{K}$, is fixed by the lattice symmetries and symmetries associated with permuting the colors $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$. The elastic constants K_{ij} are functions of the loop fugacity. For the $w_{\rm X} = 1$ case they were calculated in Ref. [15] using the loop ansatz [17], which allows one to identify the marginal screening charges [18]. For the FPL^2 model there are four screening charges:

$$\begin{aligned}
{\bf e}^{(1)} &= (-\pi, 0, +\pi) \\
{\bf e}^{(2)} &= (-\pi, 0, -\pi) \\
{\bf e}^{(3)} &= (-\pi, +\pi, 0) \\
{\bf e}^{(4)} &= (-\pi, -\pi, 0) \ .
\end{aligned} \tag{6}$$

These electric charges are associated with the most relevant vertex operators appearing in the Fourier expansion of the operator conjugate to the loop weight (see Eq. (12), below).

Demanding that all four charges have scaling dimension equal to two gives (using Eq. 16)

$$\begin{aligned}
K_{11} &= \frac{\pi}{8}(2 - e_g - e_b) \\
K_{13} &= \frac{\pi}{8}(e_b - e_g) \\
K_{22} &= \frac{\pi}{2}\frac{(1-e_b)(1-e_g)}{2 - e_b - e_g}
\end{aligned} \tag{7}$$

for the elastic constants of the FPL^2 model; $e_{\rm b}$ and $e_{\rm g}$ satisfy Eq. (2) and take their values on the interval $[0, 1/2]$. Below we will argue that the effect of $w_{\rm X} \neq 1$ is to change the value of the elastic constant K_{22} while leaving the other two unchanged.

The boundary term in the action,

$$S_{\rm B} = \frac{\rm i}{4\pi}\int {\rm d}^2{\bf x}\ ({\bf e}_0 \cdot {\bf h}({\bf x}))\rho({\bf x}) \ , \tag{8}$$

enforces the correct weight of topologically non-trivial loops. If the oriented loop model is defined with periodic boundary conditions along one direction (i.e., on a cylinder) these would be the loops that completely wind around the cylinder [34]. On a cylinder the scalar curvature, ρ, is non-zero only at the two boundaries at infinity. $S_{\rm B}$ has the effect of placing background electric charges $\pm{\bf e}_0$ at the two boundaries, where the identification

$${\bf e}_0 = -\frac{\pi}{2}(e_{\rm g} + e_{\rm b}, 0, e_{\rm g} - e_{\rm b}) \tag{9}$$

comes about by demanding that the oriented winding loops be assigned correct phase factors, $\exp(\pm {\rm i}\pi e_{\rm b})$ or $\exp(\pm {\rm i}\pi e_{\rm g})$ [15].

The third term, called the Liouville term,

$$S_{\rm L} = \int {\rm d}^2{\bf x}\ w[{\bf h}({\bf x})] \ , \tag{10}$$

owes its existence to the complex weights associated with oriented loops in the bulk. The local redistribution of the loop weights made in Sec. III A leads to complex vertex weights, which in turn depend only on the colors of the four edges around the vertex. If we write the vertex weight as $\exp(-w)$ then

$$\begin{aligned} w(\mathbf{B},\mathbf{C},\mathbf{A},\mathbf{D}) &= 0, \\ w(\mathbf{B},\mathbf{D},\mathbf{A},\mathbf{C}) &= 0, \\ w(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}) &= \mp \mathrm{i}\frac{\pi}{4}(e_{\mathrm{g}}+e_{\mathrm{b}}), \\ w(\mathbf{B},\mathbf{A},\mathbf{C},\mathbf{D}) &= \mp \mathrm{i}\frac{\pi}{4}(e_{\mathrm{g}}-e_{\mathrm{b}}), \\ w(\mathbf{A},\mathbf{B},\mathbf{D},\mathbf{C}) &= \mp \mathrm{i}\frac{\pi}{4}(e_{\mathrm{b}}-e_{\mathrm{g}}), \\ w(\mathbf{B},\mathbf{A},\mathbf{D},\mathbf{C}) &= \mp \mathrm{i}\frac{\pi}{4}(-e_{\mathrm{b}}-e_{\mathrm{g}}) \ ; \end{aligned} \tag{11}$$

the top (bottom) sign is for even (odd) vertices, and the colors are listed in order, starting from the left-most edge and proceeding clockwise around the vertex. The weight operator w is invariant under cyclic permutations of the colors and it is a periodic function of the heights around a vertex. In the scaling limit the vertex weights give rise to the operator $w[\mathbf{h}(\mathbf{x})]$ in Eq. (10) which can be written as a Fourier series

$$w[\mathbf{h}(\mathbf{x})] = \sum_{\mathbf{e}\in\mathcal{R}_w^*} \tilde{w}_{\mathbf{e}} \exp(\mathrm{i}\mathbf{e}\cdot\mathbf{h}(\mathbf{x})) \ . \tag{12}$$

The electric charges $\mathbf{e}$ appearing in the Fourier expansion are dictated by the lattice of periodicities $\mathcal{R}_w$ of the operator $w[\mathbf{h}]$; $\mathcal{R}_w^*$ is the reciprocal lattice. $\mathcal{R}_w$ is determined by inspection of the values the loop weight operator takes on the flat states: vectors in $\mathcal{R}_w$ connect flat states on which the loop weight operator takes identical values. The most relevant charges in $\mathcal{R}_w^*$ are the four given in Eq. (6). We identify them with the screening charges [18] of the Coulomb gas. This is the content of the loop ansatz introduced in Ref. [17].

C. Effective field theory: $w_{\mathrm{X}} \neq 1$

For the SFL model, when $w_{\mathrm{X}} \neq 1$, the Liouville term in Eq. (4) is modified, while the elastic and the boundary terms are unchanged. The number of marginal screening charges appearing in Eq. (12) is reduced from four to two, and the loop ansatz fixes the values of K_{11} and K_{13} only. They do not depend on the value of w_{X} and are given by the $w_{\mathrm{X}} = 1$ formulae, Eq. (7). K_{22}, on the other hand, is a non-universal function of w_{X}. Below we present arguments for this scenario, which is supported by exact results available in the 6V case (i.e., for $n_{\mathrm{b}} = n_{\mathrm{g}} = 1$), and by our numerical transfer matrix calculations described in Sec. V below.

The new vertex weight w_{X} changes the value of w in Eq. (11) from 0 to $-\ln w_{\mathrm{X}}$ for the vertex states $(\mathbf{B},\mathbf{C},\mathbf{A},\mathbf{D})$, $(\mathbf{B},\mathbf{D},\mathbf{A},\mathbf{C})$, and six other related to these two by cyclic permutations of the colors. The weights of the other 16 vertex states are unchanged. We consider the consequences of this change on the effective field theory.

In the height representation of the SFL model, the change in vertex weight corresponds to adding

$$S_{\mathrm{X}} = \int \mathrm{d}^2\mathbf{x} \ X[\mathbf{h}(\mathbf{x})] \ , \tag{13}$$

to the action. The X operator takes the value $\ln w_{\mathrm{X}}$ on the flat states made up of $(\mathbf{B},\mathbf{C},\mathbf{A},\mathbf{D})$ or $(\mathbf{B},\mathbf{D},\mathbf{A},\mathbf{C})$ type vertices, and vanishes on all the others. By inspection of the graph of flat states we find that the lattice of periodicities for the operator X, $\mathcal{R}_X$, is the span of $(1,0,-1)$, $(1,0,1)$ and $(0,1,0)$; these are the height difference vectors between the flat states in the support of X. This observation implies that $X[\mathbf{h}]$ can be expanded in a Fourier series over electric charges that live in the dual lattice $\mathcal{R}_X^*$ which is the span of $(\pi,0,\pi)$, $(\pi,0,-\pi)$ and $(0,2\pi,0)$.

If we consider the effect of S_{X} as a perturbation on the action of the FPL^2 model the electric charges $(0,\pm 2\pi,0)$ play a special role. Namely, the operator product expansion of $\exp\left(i(0,2\pi,0)\cdot\mathbf{h}\right)$ and $\exp\left(-i(0,2\pi,0)\cdot\mathbf{h}\right)$ contains the $(\boldsymbol{\partial} h^2)^2$ operator, and therefore leads to the renormalization of K_{22} [21]. This follows from the fact that the background charge, Eq. (9), has a vanishing second component. On the other hand, for charges $\mathbf{e}$ with non-zero first or third component, the effect of the background charge is that the operator product expansion of $\exp\left(i\mathbf{e}\cdot\mathbf{h}\right)$ and $\exp\left(-i\mathbf{e}\cdot\mathbf{h}\right)$ does not contain $\boldsymbol{\partial} h^i \cdot \boldsymbol{\partial} h^j$ operators and therefore does not lead to the renormalization of the elastic constants K_{ij}.

The Coulomb gas representation of the height model provides a clear physical picture of the effect of $S_{\rm X}$ on the critical action of the FPL^2 model. For $w_{\rm X}=1$ the dimension of the $(0,\pm 2\pi,0)$ charges follows from Eq. (16),

$$x_X = \frac{(2\pi)^2}{4\pi K_{22}} = 2\left(\frac{1}{1-e_b}+\frac{1}{1-e_g}\right). \tag{14}$$

It is greater than 2 in the whole critical region of the FPL^2 model. These charges are therefore irrelevant in the renormalization group sense. In the Coulomb gas picture the $(0,\pm 2\pi,0)$ charges appear as bound pairs of neutral dipoles. Increasing $w_{\rm X}$ will have the effect of increasing the bare fugacity of these dipoles which will in turn increase the value of the coupling K_{22} appearing in the effective field theory. Formally, this can be seen in perturbation theory making use of the operator product expansion [21]. Physically, the renormalization of K_{22} can be understood as the screening effect of dipoles. The dipoles lower the Coulomb energy between two electric test-charges having a non-zero second component, corresponding to an increase in the value of K_{22} which plays the role of a dielectric constant.

At a critical value $w_{\rm X}^{\rm c}$ there will be a Kosterlitz-Thouless type transition of the SFL model into a flat state with a vanishing density of vertices at which the polymer bends. At the transition the $(0,\pm 2\pi,0)$ charges are marginal, i.e., their scaling dimension is equal to 2. Using Eq. (14) this observation gives rise to the prediction for the critical value of K_{22}:

$$K_{22}(w_{\rm X}^{\rm c}) = \frac{\pi}{2}\ . \tag{15}$$

For values of $w_{\rm X}$ smaller than $w_{\rm X}^{\rm c}$, K_{22} will be a *non-universal* function of $w_{\rm X}$. In the $n_{\rm b}=n_{\rm g}=1$ case, the formula $K_{22}=\arcsin(w_{\rm X}/2)$ follows from the exact solution of the 6V model [10]. The critical value of the vertex weight is $w_{\rm X}^{\rm c}({\rm 6V})=2$ and $K_{22}(2)=\pi/2$ is in agreement with Eq. (15). For other values of $n_{\rm b}$ and $n_{\rm g}$ our numerical transfer matrix calculations are in good agreement with Eq. (15).

The introduction of the vertex weight $w_{\rm X}$ also has an effect on the screening charges, Eq. (6), that appear in the Liouville part of the action. First consider the $n_{\rm b}=n_{\rm g}$ case of the SFL model. Due to the presence of the $w_{\rm X}$ term cyclic permutations of the four colors around a vertex are no longer a symmetry of the vertex weight. Therefore, unlike the $w_{\rm X}=1$ case [15], there are now two independent elastic constants, K_{22} and K_{11} appearing in $S_{\rm E}$. $K_{13}=0$ follows from the remaining Z_2 symmetry of the vertex weights which are invariant under *two* cyclic permutations, such as $(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D})\to(\mathbf{C},\mathbf{D},\mathbf{A},\mathbf{B})$. The deduced structure of the elasticity matrix implies that the four electric charges in Eq. (6) are no longer degenerate in dimension for arbitrary $w_{\rm X}$. Since the dimensions of $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$ are independent of K_{22} they are identified as the two screening charges tied to the non-renormalizability of the loop weights [15]. As in the FPL^2 model we then assume that these two charges remain marginal when $n_b\neq n_g$. Using the dimension formula, Eq. (16), this then fixes the values of the two elastic constants, K_{11} and K_{13}, to the values quoted in Eq. (7).

Finally, it is interesting to look at some extreme limits of K_{22} in view of the effective field theory. Consider first the limit $K_{22}\to\infty$ in which height fluctuations in the second height component are completely suppressed. (As we are outside the critical phase, we are here referring to the bare value of the coupling.) Clearly, height fluctuations must always be present in the microscopic four-coloring model, but it is nevertheless instructive to look for the states that minimize the fluctuations of h^2. From the choice of the color vectors, Eq. (3), it is not difficult to see that on the four sites of $\{\mathbf{x}\}$ surrounding a given vertex, h^2 fluctuates by two units for the first four vertices of Fig. 3 and by one unit for the last two vertices. All vertices must therefore be of the c-type, corresponding to the limit $w_{\rm X}\to\infty$. Thus, $K_{22}\to\infty$ as $w_{\rm X}\to\infty$.

Conversely, as $K_{22}\to 0$ the fluctuations in h^2 become unbounded and the effective field theory loses its consistency (since it was based on the assumption that the interfacial entropy is due to bounded fluctuations around the macroscopically flat states). However, the argument given above indicates that a small value of K_{22} should correspond to a small number of straight-going vertices in the loop model. Thus, we would conjecture that $K_{22}\to 0$ as $w_{\rm X}\to 0$. This expectation is confirmed by the exact result for the 6V case [10] and also by extrapolation of our numerical results for $K_{22}(w_{\rm X})$ in the Flory case.

Apart from these limiting values, we would of course expect K_{22} to be a monotonically increasing function of $w_{\rm X}$ throughout the critical phase.

In the next section we compute the central charge and the scaling dimensions of various operators in the semiflexible loop model from its effective field theory. We identify quantities that depend on the non-universal elastic constant K_{22}; these are then predicted to vary continuously with $w_{\rm X}$.

IV. OPERATORS AND SCALING DIMENSIONS

The effective field theory of the semiflexible loop model describes a Coulomb gas of electric and magnetic charges in the presence of background and screening charges. The magnetic charges $\mathbf{m}$ are vectors in $\mathcal{R}$ which is the lattice

of periodicities of the graph of flat states, while the electric charges $\mathbf{e}$ take their values in the reciprocal lattice $\mathcal{R}^*$ [15]. With the normalization adopted for the vectors $\mathbf{A}$ through $\mathbf{D}$, Eq. (3), $\mathcal{R}$ is a face-centered cubic lattice whose conventional cubic cell has sides of length 4, while is $\mathcal{R}^*$ is a body-centered cubic lattice whose conventional cubic cell has sides of length π.

The scaling dimension of an operator which has total electromagnetic charge $(\mathbf{e}, \mathbf{m})$ is the sum of its electric and magnetic dimensions, and it is a function of the elastic constants and the background charge [18]:

$$x(\mathbf{e}, \mathbf{m}) = \frac{1}{4\pi} \left[(\mathbf{e}\mathbf{K}^{-1}) \cdot (\mathbf{e} - 2\mathbf{e}_0) + (\mathbf{m}\mathbf{K}) \cdot \mathbf{m} \right] \ . \tag{16}$$

$\mathbf{K}$ is the 3×3 matrix of elastic constants and $\mathbf{K}^{-1}$ is its inverse.

From Eq. (16) and the form of $\mathbf{K}$ (Eq. (5)) and $\mathbf{e}_0$ (Eq. (9)) it immediately follows that operators whose electric and magnetic charges both have a vanishing second component will have a K_{22}-independent scaling dimension. The scaling dimension in this case is independent of w_{X} and equal to its known value at $w_{\mathrm{X}} = 1$ [15]. Operators with $\mathbf{e}$ and $\mathbf{m}$ charges whose second components are not both zero will, on the other hand, have a scaling dimension that varies continuously with w_{X}. These predictions are confirmed by our numerical results.

A. Central charge

The central charge of the SFL model follows from its critical action. The three height components (bosonic free fields) each contribute one unit to the central charge while the contribution from the background charge is $12x(\mathbf{e}_0, 0)$. Using Eq. (16) for $x(\mathbf{e}_0, 0)$, Eq. (9) for the background charge $\mathbf{e}_0$, and the calculated values of the elastic constants K_{11} and K_{13}, Eq. (7), we find

$$c = 3 - 6 \left(\frac{e_{\mathrm{b}}^2}{1 - e_{\mathrm{b}}} + \frac{e_{\mathrm{g}}^2}{1 - e_{\mathrm{g}}} \right), \tag{17}$$

independent of the unknown value of K_{22}.

For the 6V model, which corresponds to the $e_{\mathrm{b}} = e_{\mathrm{g}} = 1/3$ line in the SFL model, the above formula gives $c = 1$ for the central charge along the critical line. This result also follows directly from the exact solution of the 6V model.

For the Flory model of polymer melting, which is the $e_{\mathrm{b}} = 1/2$, $e_{\mathrm{g}} = 1/3$ case, the predicted central charge is $c = -1$. This value is confirmed by our numerical transfer matrix calculations (see Sec. V).

B. Thermal operator

The SFL model can be thought of as the zero-temperature limit of a more general model where we allow for thermal excitations that violate the fully packing constraint. Violations of the constraint lead to vertices with the four adjacent edges colored $(\mathbf{C}, \mathbf{D}, \mathbf{C}, \mathbf{D})$. In the height representation such a vertex is identified with a topological defect (screw dislocation) whose charge, i.e., the sum of height differences around the vertex, is

$$\mathbf{m}_T = 2(\mathbf{C} + \mathbf{D}) = (0, -4, 0). \tag{18}$$

Other vertices which have no $\mathbf{A}$ or $\mathbf{B}$ colored edges are possible, but they have a larger magnetic charge and are hence less relevant.

In the Coulomb gas picture a topological defect corresponds to a magnetic charge. Therefore, the thermal dimension can be calculated using Eq. (16), and we find

$$x_T = x(0, \mathbf{m}_T) = \frac{4}{\pi} K_{22} \ . \tag{19}$$

We make use of this equation below as it allows us to determine the unknown elastic constant K_{22} from a measurement of the thermal scaling dimension. Once this elastic constant is known, scaling dimensions of all electromagnetic operators can be calculated from Eq. (16).

C. String operators

A particularly important set of operators in any loop model are the string operators. Their two-point function is defined as the probability of having the small neighborhoods around two fixed points on the lattice, which are separated by a large distance, connected by $s_{\rm b}$ real loop segments and $s_{\rm g}$ ghost loop segments. For simplicity, we shall require $s_{\rm b}$ and $s_{\rm g}$ to be either both even or both odd; $s_{\rm b}+s_{\rm g}$ odd requires L to be odd which produces a twist in the height, as discussed in Ref. [15]. In the height representation these string configurations are mapped to two topological defects, one serving as the source and the other as the sink of oriented loop segments. When the oriented loop segments wind around the defect points they are assigned spurious phase factors by the vertex weights; these phase factors can however be compensated by introducing appropriate electric charges at the positions of the defects [22].

In the case $s_{\rm b}=2k_{\rm b}$ and $s_{\rm g}=2k_{\rm g}$, i.e., when the number of real and ghost strings are both even, the electric and magnetic charge of the corresponding string operator are [15]

$$\begin{aligned} \mathbf{e}_{2k_{\rm b},2k_{\rm g}} &= -\frac{\pi}{2}(e_{\rm b},0,-e_{\rm b})(1-\delta_{k_{\rm b},0})-\frac{\pi}{2}(e_{\rm g},0,e_{\rm g})(1-\delta_{k_{\rm g},0}) \\ \mathbf{m}_{2k_{\rm b},2k_{\rm g}} &= -2(k_{\rm b}+k_{\rm g},0,k_{\rm g}-k_{\rm b})\ . \end{aligned} \tag{20}$$

Since the charges have vanishing second component their dimension is independent of K_{22} and constant along the whole critical line $w_{\rm X}\le w_{\rm X}^{\rm c}$. The value of the string dimension follows from Eq. (16),

$$x_{2k_{\rm b},2k_{\rm g}}=\frac{1}{2}\left[(1-e_{\rm b})k_{\rm b}^2+(1-e_{\rm g})k_{\rm g}^2-\frac{e_{\rm b}^2}{1-e_{\rm b}}(1-\delta_{k_{\rm b},0})-\frac{e_{\rm g}^2}{1-e_{\rm g}}(1-\delta_{k_{\rm g},0})\right] \tag{21}$$

and is identical to that of the FPL^2 model [15]. Our numerical simulations confirm that even string dimensions are constant along the critical line.

In the odd string case, when $s_{\rm b}=2k_{\rm b}-1$ and $s_{\rm g}=2k_{\rm g}-1$, the electric and magnetic charge are [15]

$$\begin{aligned} \mathbf{e}_{2k_{\rm b}-1,2k_{\rm g}-1} &= -\frac{\pi}{2}(e_{\rm g}+e_{\rm b},0,e_{\rm g}-e_{\rm b}) \\ \mathbf{m}_{2k_{\rm b}-1,2k_{\rm g}-1} &= -2(k_{\rm b}+k_{\rm g}-1,1,k_{\rm g}-k_{\rm b})\ . \end{aligned} \tag{22}$$

Notably the magnetic charge has a non-vanishing second component. Using Eq. (16) we calculate

$$x_{2k_{\rm b}-1,2k_{\rm g}-1}=\frac{K_{22}}{\pi}+\frac{1}{8}\left[(1-e_{\rm b})(2k_{\rm b}-1)^2+(1-e_{\rm g})(2k_{\rm g}-1)^2\right]-\frac{1}{2}\left[\frac{e_{\rm b}^2}{1-e_{\rm b}}+\frac{e_{\rm g}^2}{1-e_{\rm g}}\right] \tag{23}$$

for the odd string dimension. It depends on the value of K_{22} and will therefore vary continuously with $w_{\rm X}$. At the melting transition the exponents are exactly known from Eq. (15). This is confirmed by our numerical transfer matrix results, which we describe next.

V. TRANSFER MATRIX RESULTS

To check the correctness of our field theoretical predictions, we have numerically diagonalized the transfer matrix of the semi-flexible loop model (and of its various generalizations, to be discussed below) defined on semi-infinite cylinders of even widths L ranging from 4 to 14.

The existence of a transfer matrix may not be a priori obvious, since the Boltzmann weights depend on the number of loops, which is a non-local quantity. We have however already shown in an earlier publication [15] how this is resolved by working in a basis of states that contains non-local information about how loop segments are interconnected at a given stage of the computations. In that paper, it was also shown that the full transfer matrix contains various sectors, the leading eigenvalues of which provide finite-size estimations of the free energy and of the various critical exponents, using the standard CFT relations [24, 25]

$$f_0(L) = f_0(\infty)-\frac{\pi c}{6L^2}+\cdots \tag{24}$$

$$f_k(L)-f_0(L) = \frac{2\pi x_k}{L^2}+\cdots \tag{25}$$

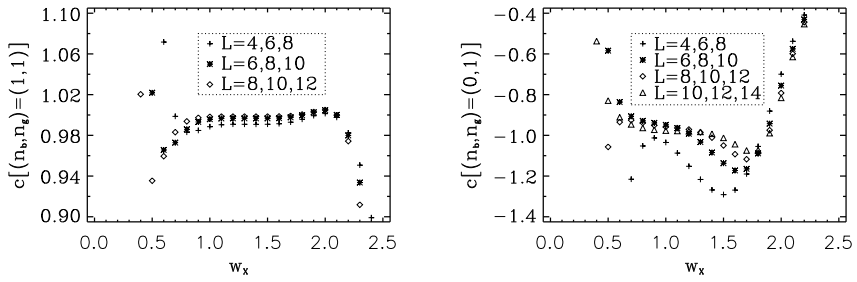


FIG. 4: Central charge c as a function of the bending rigidity $w_{\rm X}$ for the 6V model (left) and the semiflexible loop model (right). We show three-point fits for different system sizes, as indicated on the legends.

Here, the label k refers either to a higher eigenvalue in the sector to which f_0 belongs, or to the leading eigenvalue in another sector characterized by some topological defect of charge k.

To access critical exponents, we shall mainly be concerned with topological defects that consist in enforcing that a certain number of strings of either flavor propagate along the length direction of the cylinder. These give rise to a two-parameter family of critical exponents $x_{s_{\rm b},s_{\rm g}}$ corresponding to $s_{\rm b}$ real strings and $s_{\rm g}$ ghost strings. The corresponding topological charges, Eq. (20) and Eq. (22), take the form of three-dimensional electromagnetic vector charges. In the transfer matrix calculations, each of these topological sectors is associated with a different state space. The difficulty of precisely characterizing these spaces limited our previous approach [15] to at most two strings. In Appendix A we present an algorithm that explicitly constructs the required state spaces for any $(s_{\rm b}, s_{\rm g})$, based on an iterative procedure and hashing techniques.

By inspection of the eigenstates produced by our previous algorithm [15], it turns out that many of the basis states carry zero weight. One would then expect that identical results can be obtained more efficiently by working in a basis in which such states have been eliminated from the outset. We defer the technical details of how this can be done to Appendix A. It is also shown how the block-diagonalization scheme can be carried even further, by exploiting various conservation laws that are most easily understood from the analogy between the SFL model and the six-vertex model. One important consequence is that the constrained free energy $f_T(L)$ that is linked to the thermal scaling dimension can now be obtained as a leading eigenvalue, rather than as the second eigenvalue in the stringless sector. This considerably improves the efficiency of the computations.

Finally, the matrix elements need some modification in order to take into account the bending rigidity parameter $w_{\rm X}$. This is readily done, without any modification of the basis states, since $w_{\rm X}$ is a purely local quantity.

Before turning to our numerical results, we should mention that we have submitted our transfer matrices to several tests, in order to verify their correctness:

- For $w_{\rm X} = 1$, all numerically determined string dimensions $x_{s_{\rm b},s_{\rm g}}$ with $s_{\rm b} + s_{\rm g} = 2$ or 4 agree to at least three significant digits with their exact values in the cases $(n_{\rm b}, n_{\rm g}) = (1,1)$ [10] and $(n_{\rm b}, n_{\rm g}) = (0,1)$ [15].
- All eigenvalues found for the FPL^2 model agree with those obtained from our previous algorithm [15].
- For the six-vertex model [10], we have compared the extrapolated bulk free energy with Baxter's exact expression.
- Again for the six-vertex model, we find excellent agreement with the exact formulae $x_{1,1} = \frac{K_{22}}{\pi}$ and $x_T = \frac{4K_{22}}{\pi}$, where $K_{22} = \arcsin(w_{\rm X}/2)$ is the elastic constant.
- We have also found agreement with the first few terms in diagrammatic expansions around various limits of infinite fugacities.

A. Central charge

A crucial prediction of our field theory is that, for given values of the loop fugacities $(n_{\rm b}, n_{\rm g})$, the central charge of the SFL model should be independent of $w_{\rm X}$, as long as the latter is constrained to the critical regime, $0 < w_{\rm X} \le w_{\rm X}^{\rm c}$.

In Fig. 4 we show the effective central charge as a function of $w_{\rm X}$, in the cases $(n_{\rm b}, n_{\rm g}) = (1,1)$ and $(0,1)$. The result $c = 1$ for $0 < w_{\rm X} \le 2$ is well established for the 6V model, but the plot for this case is still useful as it gives

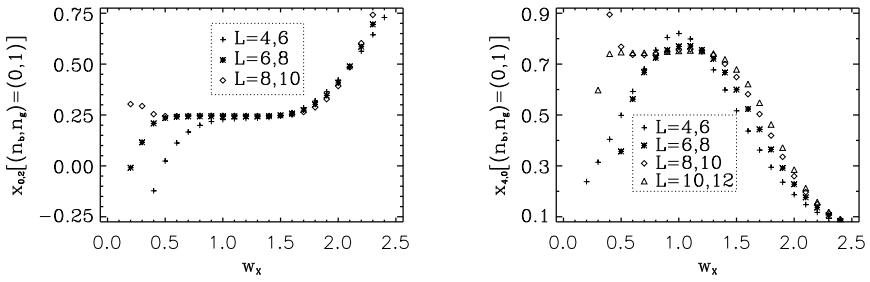


FIG. 5: Scaling dimensions x_{s_b,s_g} with even s_b and s_g. The left and right panels show respectively $x_{0,2}$ and $x_{4,0}$. System sizes used in the two-point fits are indicated on the legends.

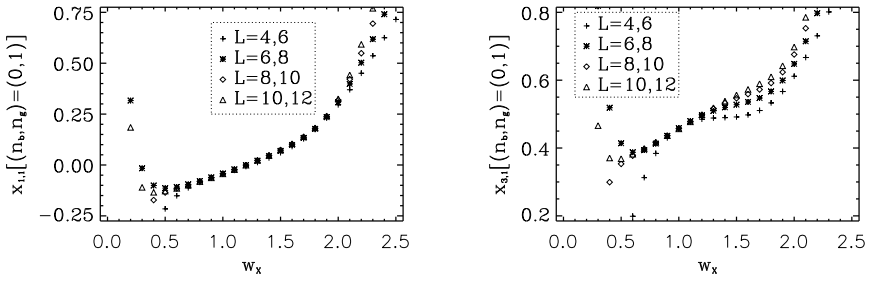


FIG. 6: Scaling dimensions x_{s_b,s_g} with odd s_b and s_g. We show $x_{1,1}$ (left) and $x_{3,1}$ (right) for the SFL model.

us some guidance as to what finite-size effects to expect. In particular, note that these become more pronounced when w_X is small, and that the termination of critical behavior at $w_X = 2$ is clearly signalled by the finite-size data's levelling off. Another effect is that while for $w_X < 2$ the distance between successive estimates decreases with system size, for $w_X > 2$ we observe this distance to increase.

Although finite-size effects play a more important role in the $(n_b, n_g) = (0, 1)$ case, the general picture is quite similar. The figure leaves little doubt that $c = -1$ for $0 < w_X \leq w_X^c$. We also obtain a first rough estimate $w_X^c = 1.95 \pm 0.15$, not far away from the 6V-model value.

Here, and elsewhere, we mainly show fits in which the convergence of the relations Eqs. (24) and (25) has been accelerated through the inclusion of a non-universal $1/L^4$ correction, as predicted by conformal invariance.

B. String dimensions

We next turn to the computation of the magnetic-type scaling dimensions x_{s_b,s_g} describing the scaling of the operator that inserts s_b real strings and s_g ghost strings. To study these, the width L of the strip must have the same parity as $s_b + s_g$. For simplicity we shall limit ourselves to the case of even L. There are then two classes of exponents: Those in which s_b and s_g are both even, and those in which they are both odd. The field theory predicts that the former should stay constant on the critical line, parameterized by w_X, while the latter are expected to vary continuously as functions of w_X; see Sec. IV C.

On Fig. 5 we show two examples of exponents with s_b, s_g even, within the SFL model $[(n_b, n_g) = (0, 1)]$ with varying w_X. They correspond respectively to the insertion of two ghost strings $(x_{0,2})$ and of four real strings $(x_{4,0})$. From the figure it should be evident that $x_{0,2} = \frac{1}{4}$ and $x_{4,0} = \frac{3}{4}$ are constant throughout the critical phase. In the latter case the finite-size variations are quite pronounced, as might have been anticipated given the higher number of strings. Careful observation of the distance between subsequent finite-size points however strongly suggests that the variation will eventually die away.

Examples of exponents with s_b, s_g odd are given on Fig. 6. In both cases, $x_{1,1}$ and $x_{3,1}$, the convergence to

w_X	x_T	$x_{1,1}$		$x_{3,1}$		$x_{1,3}$	
		CFT	Num.	CFT	Num.	CFT	Num.
0.4	0.141(7)						
0.5	0.207(5)						
0.6	0.275(4)	-0.117	-0.115	0.383	0.385	0.549	0.548
0.7	0.346(3)	-0.100	-0.100	0.400	0.402	0.567	0.565
0.8	0.413(2)	-0.087	-0.085	0.413	0.418	0.580	0.582
0.9	0.4913(7)	-0.065	-0.066	0.435	0.436	0.601	0.601
1	$\frac{4}{7}$	$-\frac{5}{112}$		$\frac{51}{112}$		$\frac{209}{336}$	
1.1	0.6525(10)	-0.026	-0.025	0.474	0.475	0.641	0.643
1.2	0.7429(5)	-0.002	-0.002	0.498	0.498	0.665	0.665
1.3	0.8365(3)	0.022	0.022	0.522	0.521	0.688	0.688
1.4	0.9374(1)	0.047	0.048	0.547	0.546	0.713	0.715
1.5	1.0490(1)	0.075	0.076	0.575	0.576	0.742	0.75
1.6	1.1769(7)	0.107	0.11	0.607	0.62	0.774	0.80
1.7	1.333(2)	0.148	0.15	0.648	0.68	0.815	0.90
1.8	1.541(5)	0.202	0.20	0.702	0.7	0.869	0.9
1.9	1.861(8)	0.29	0.3	0.79	0.8	0.95	1.0
w_X^c	2	$\frac{5}{16}$		$\frac{13}{16}$		$\frac{47}{48}$	

TABLE I: Thermal exponent x_T measured for varying values of w_X. The corresponding values of the scaling dimensions $x_{1,1}$, $x_{3,1}$ and $x_{1,3}$ are compared with their directly measured counterparts.

monotonically increasing functions of w_X is clearly brought out. Also note the agreement with the exact results for $w_X = 1$, which read respectively $x_{1,1} = -\frac{5}{112} \simeq -0.0446$ and $x_{3,1} = \frac{51}{112} \simeq 0.455$ [15].

C. Thermal scaling dimension

As described in Appendix A, the thermal scaling dimension is linked to the gap between transfer matrix sectors in which there is an even (respectively an odd) number of flavor crossings in the basis states. Because of the relation

$$K_{22} = \frac{\pi}{4} x_T, \tag{26}$$

measuring this gap gives a direct means of accessing the elastic constant associated with the second height component in the field theory.

According to the field theory, K_{22} is a non-universal function of w_X, and once it is known the values of all the other critical exponents follow. This suggests the following numerical check of the field theoretic scenario: For several values of w_X, we measure x_T from the transfer matrix, and use it to determine K_{22}. We then compute the predictions for the various other scaling dimensions (the x_{s_b,s_g}) from the field theory, Eq. (16), by use of the numerically determined value of K_{22}, and compare them with values measured directly from the transfer matrices.

The result of this verification is shown in Table I. The values for x_T are based on transfer matrices for strips up to size $L = 14$, here extrapolated to the limit $L \to \infty$. The agreement between the CFT predictions and numerics is in general excellent. Note however, that the precision deteriorates whenever w_X approaches zero or w_X^c, its critical values.

Based on these data we can refine our estimate for the location of the melting transition:

$$w_X^c = 1.92 \pm 0.02. \tag{27}$$

In Fig. 7 we compare our numerical results for the curve $x_T(w_X)$ in the SFL case with the exactly known result of the 6V-model, $x_T = \frac{4}{\pi} \arcsin(w_X/2)$ [10]. Although the functional forms are quite reminiscent, we have unfortunately not been able to conjecture a convincing exact expression in the SFL case.

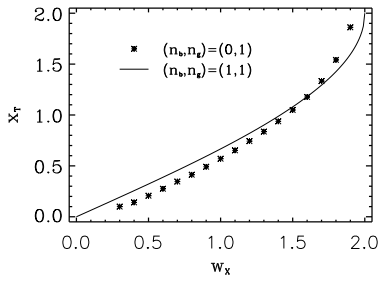


FIG. 7: Thermal scaling dimension x_T versus bending rigidity w_X in the Flory model (symbols), as compared to the exact result of the six-vertex model (line).

VI. PHASE DIAGRAM

A. Generalized six-vertex model

Given the one-to-one correspondence between the six vertex configurations in the FPL^2 model and the six arrow configurations in the six-vertex model (see Fig. 3), it is natural to define a generalized six-vertex model in which the standard arrow weights are supplemented by the non-local loop weights n_b, n_g of the FPL^2 model.

Until now we have only considered the isotropic case of $a = b$ (see Fig. 3. Let us briefly recall the effect of taking $a \neq b$ in the six-vertex model [10]. Define the parameters w and μ by

$$\Delta = \frac{a^2 + b^2 - w_X^2}{2ab} = -\cos\mu, \qquad 0 < \mu < \pi \tag{28}$$

$$\frac{a}{b} = \frac{\exp(i\mu) - \exp(iw)}{\exp(i\mu + iw) - 1}, \qquad -\mu < w < \mu. \tag{29}$$

Then, taking $a \neq b$ corresponds to twisting the usual square lattice into a rhombus, defined by the anisotropy angle [26]

$$\theta = \frac{\pi}{2}(1 + \frac{w}{\mu}). \tag{30}$$

All this means is that the central charge and the critical exponents, when measured in the usual way from a transfer matrix, get multiplied by a geometrical factor of $\sin(\theta)$.

In Fig. 8 we plot the effective central charge of the SFL model with $b = w_X = 1$ and varying a against the variable $\tau = \theta/\pi$, defined in terms of the above 6V expressions. By the word "effective" we mean that we do not correct for the lattice distortion, the effect of which can then be read off from the graph. If the effect of the anisotropy were the same as in the 6V model, the plot should just look like the function $-\sin(\pi\tau)$, since the SFL model has (real) central charge $c = -1$. Clearly, this is not the case, and so the non-locality of the loop weights has a non-trivial effect on the anisotropy factor. We leave this as an interesting open question.

B. Generalized eight-vertex model

It is also of interest to consider the loop generalization of the eight-vertex model. In terms of the loops there are two different ways of resolving the vertices that act as sources or sinks of the eight-vertex arrows, and so we are led to consider the ten-vertex model defined by Fig. 9. In addition to the local weights which are shown on the figure, we assign the usual non-local loop weights n_b and n_g.

For simplicity, we shall disregard the effects of anisotropy, and thus only two types of local weights are of interest. The first is the weight w_X of having the two loop flavors cross, same as in the SFL model. The second is a contact interaction w_c, assigned to the vertices where two loop segments of the same flavor touch one another. One may consider letting it depend on the flavor index, but in order to stay close to the definition of the conventional eight-vertex model we shall here take the contact interaction to be flavor independent.

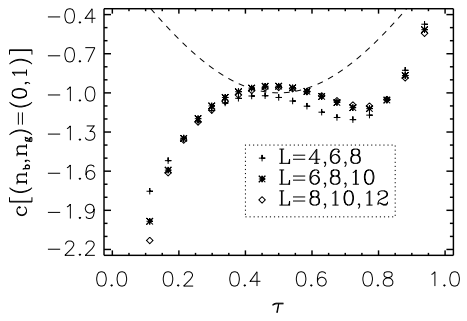


FIG. 8: Anisotropy effects in the generalized six-vertex model with $(n_{\rm b}, n_{\rm g}) = (0, 1)$. The symbols show the effective central charge for various system sizes. For comparison, the dashed line shows the function $-\sin(\pi\tau)$, which would have been the exact result if the anisotropy had had the same effect as in the six-vertex model.

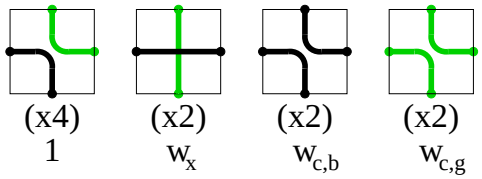


FIG. 9: Vertices defining the generalized eight-vertex model, along with their corresponding multiplicities and local weights.

The motivation for the contact interaction is to be able to exclude the loops of a given flavor from any number of lattice vertices. As this violates the compactness constraint, we expect the conclusions of our earlier paper on the transition from the compact to the dense phase [23] to apply: A non-zero value of $w_{\rm c}$ should induce a flow towards a phase where the two loop flavors decouple, and the critical properties are just those of two non-interacting O(n) models (with $n = n_{\rm b}$ and $n_{\rm g}$ respectively) in the low-temperature (dense) phase.

A detailed numerical study of the behavior of the effective central charge in the parameter space $(w_{\rm X}, w_{\rm c})$ has led us to suggest that the phase diagram of the generalized eight-vertex model is as shown on Fig. 10.

For $w_{\rm c} = 0$, the model reduces to the SFL model, and so below the melting point M (i.e., for $w_{\rm X} < w_{\rm X}^{\rm c}$ with, very roughly, $w_{\rm X}^{\rm c} \approx 2$) we have a line of critical points along which critical exponents that depend on the second height

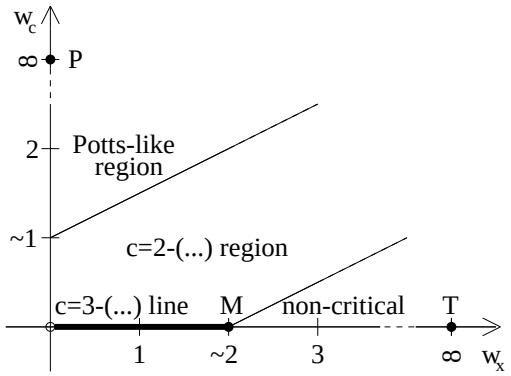


FIG. 10: Proposed phase diagram of the generalized eight-vertex model.

component vary continuously, while the central charge

$$c(n_{\rm b}, n_{\rm g}) = 3 - \frac{6e_{\rm b}^2}{1-e_{\rm b}} - \frac{6e_{\rm g}^2}{1-e_{\rm g}} \tag{31}$$

is constant. The end-point of the SFL line, with $w_{\rm X} = \infty$, is a trivial attractive fixed point T, favoring configurations in which all loops go straight in the bulk (they are necessarily reflected at the boundaries enjoying free boundary conditions). The point T is believed to govern the theories to the right of M (i.e., with $w_{\rm X} > w_{\rm X}^{\rm c}$), including a portion of the phase diagram with non-zero but small $w_{\rm c}$ (see Fig. 10).

Moving away from the critical line of the SFL model, towards positive values of the contact interaction, we observe numerically that the central charge drops abruptly by one unit, and stays constant as a function of $w_{\rm c}$ up to some finite critical value $w_{\rm c}^{\rm c}$ that depends on $w_{\rm X}$. This is the dense phase of the DPL^2 model [23] with central charge

$$c(n_{\rm b}, n_{\rm g}) = \left(1 - \frac{6e_{\rm b}^2}{1-e_{\rm b}}\right) + \left(1 - \frac{6e_{\rm g}^2}{1-e_{\rm g}}\right). \tag{32}$$

Here, the two loop flavors decouple, and critical exponents are just the sum of the critical exponents for two non-interacting O(n) models (with $n = n_{\rm b}$ and $n_{\rm g}$) in the dense phase. We have verified numerically this prediction for the exponent $x_{1,1}$ for a number of different loop fugacities. We have also observed numerically that the critical exponents do not depend on $w_{\rm X}$ throughout the dense phase. This confirms the expectations that in non-compact phases the only effect of the bending rigidity is to renormalize the persistence length of the polymer, as already discussed in the introduction.

Finally, for $w_{\rm c}$ large enough, the numerically evaluated central charges suggest that the models flow into an attractive fixed point P situated at $(w_{\rm X}, w_{\rm c}) = (0, \infty)$. Here, only contact-type vertices are allowed, and since the different loop flavors can no longer coexist, the partition function at P becomes simply a sum, $Z = Z_{\rm b} + Z_{\rm g}$, where Z_k involves only contact vertices of flavor k (with $k = {\rm b}, {\rm g}$). But clearly Z_k is just the loop-model representation [10] of a self-dual Potts model with $q_k = (n_k)^2$ states. It is intuitively clear (and explicitly brought out by the exact solution [10]) that the free energy of the q-state Potts model is an increasing function of q. Therefore, the sum $Z = Z_{\rm b} + Z_{\rm g}$ will be dominated by the term with the largest value of q. Thus, the point P has central charge

$$c(n_{\rm b}, n_{\rm g}) = \max\left(1 - \frac{6e_{\rm b}^2}{1-e_{\rm b}}, 1 - \frac{6e_{\rm g}^2}{1-e_{\rm g}}\right), \tag{33}$$

and, by the usual identification of the critical Potts model with the dense phase of the O($n = \sqrt{q}$) model, the critical exponents are simply those of a single O($\max(n_{\rm b}, n_{\rm g})$) model in the dense phase.

We would expect that only this large-$w_{\rm c}$ portion of the phase diagram gets modified by letting the contact interaction be flavor dependent. Let us recall that in the conventional O(n) model [2] with a finite *positive* vacancy fugacity $w_{\rm c}$ the critical behavior of the loops is described by either of *two* critical branches. The first branch, known as the *dense* branch [2, 28], is attractive in $w_{\rm c}$ and as such controls the entire domain of low $w_{\rm c}$. Its central charge is the one referred to above:

$$c = 1 - 6\tilde{e}^2/(1-\tilde{e}) \tag{34}$$

in the usual parameterization $n = 2\cos(\pi\tilde{e})$. The second branch, known as the *dilute* branch [2, 27], is repulsive in $w_{\rm c}$ and as such requires $w_{\rm c}$ to be tuned to a particular n-dependent critical value. In other words, the fugacity of a vacancy can tune the O(n) model to its critical point. The central charge of the dilute phase is

$$c = 1 - 6\tilde{e}^2/(1+\tilde{e}) \ , \tag{35}$$

using the same parameterization as above.

In particular, in the DPL^2 model [23] the two loop flavors act as decoupled O(n) models, and depending on the fugacities of the two flavors of vacancies each of the models can reside in either the dense or the dilute phase, giving a total of four different phases. We expect this conclusion to hold true in the generalized eight-vertex model (i.e., with an added bending rigidity $w_{\rm X}$). Note that only when $n_{\rm b} = n_{\rm g}$ can we simultaneously take the two O(n) models to their critical point by tuning a vacancy fugacity $w_{\rm c}$ which is common for the two loop flavors. In the general case, when $n_{\rm b} \neq n_{\rm g}$ we would need two distinct parameters, $w_{\rm c,b}$ and $w_{\rm c,g}$, as indicated on Fig. 9. Presumably this would lead to a richer phase diagram, with critical lines corresponding to dense-dilute, dilute-dense and dilute-dilute behavior of the two O(n) models.

Let us return to the phase diagram shown in Fig. 10. In the special case of the eight-vertex model, the two oblique transition lines shown on Fig. 10 are known to be of slope 1/2 [10]. Actually, they are just images of the line OM under certain exact symmetries of the eight-vertex model [10]. Thus, they have again $c = 1$, whereas the "bulk" of the phase diagram is non-critical.

These two features can be accounted for within the framework of the generalized eight-vertex model with $(n_{\rm b}, n_{\rm g}) = (1, 1)$. First, note that our field theory predicts that the region on Fig. 10 which is limited by the two oblique lines and the coordinate axes is actually *critical* with central charge $c = 0$ (dense phase); this is obtained by setting $e_{\rm b} = e_{\rm g} = 1/3$ in Eq. (32). This is not in contradiction with the exact result [10] that this same region is non-critical within the eight-vertex model. Namely, the generalized eight-vertex model is embedded in a much larger Hilbert space. More precisely, our statement is that the first and third height components possess critical fluctuations and constitute a $c = 0$ theory, even though the second height component is massive. This scenario is brought out very clearly by the numerics, as we observe the leading transfer matrix eigenvalues in the sectors determining the free energy and x_T to *coincide* within the concerned region. Thus, x_T and K_{22} vanish identically, cf. Eq. (26).

Second, the field theory also accounts for the fact that, within the 8V model, the two oblique lines are critical with $c = 1$. Namely, we claim that they simply correspond to dilute-phase behavior within the generalized 8V model. More precisely, since $n_{\rm b} = n_{\rm g}$ the two decoupled O(n) models must be driven to their critical points simultaneously by tuning the common parameter $w_{\rm c,b} = w_{\rm c,g} \equiv w_{\rm c}$. Setting $\tilde{e} = 1/3$ in Eq. (35) gives a contribution of $c = 1/2$ for each of the models, whence $c_{\rm total} = 1/2 + 1/2 = 1$ as expected.

Finally, in the 8V model, the part of the $w_{\rm c}$ axis with $0 < w_{\rm c} < 1$ constitutes a further image of the line OM under an exact symmetry. We believe this to be "accidental" in the sense that we have seen no sign of a finite interval of the $w_{\rm c}$ being critical within the generalized 8V model with other values of the fugacities.

Taking a common contact parameter $w_{\rm c,b} = w_{\rm c,g}$ for the generalized 8V model with $n_{\rm b} \neq n_{\rm g}$ destroys the criticality of the two oblique lines of Fig. 10. They still act as transition lines in the sense that they separate the basins of attraction of the dense phase and the points P and T respectively. However, the transition is now expected to be a first order one. This is confirmed by our numerical results for the $(n_{\rm b}, n_{\rm g}) = (0, 1)$ case which show that the effective central charge develops violent finite-size effects upon approach of the transition lines. Further support for this scenario is furnished by Monte Carlo simulations [12] where a finite concentration of empty sites was shown to lead to a first order transition.

The oblique lines in Fig. 10 are expected to move away from their exactly known 8V positions when we vary the loop fugacities away from the trivial values $(n_{\rm b} = n_{\rm g} = 1)$. Some evidence for this is already available from our determination of the melting point M in the Flory case; see Eq. (27). In general, we have been able to numerically determine the position of the uppermost line from the transfer matrix spectra. Recall from the discussion near Eq. (26) that the coupling K_{22} can be linked to the gap between the leading eigenvalues in two topologically characterized transfer matrix sectors. By scanning through $w_{\rm c}$ at fixed $w_{\rm X}$ we have observed (at least in the Flory case) that these two eigenvalues become degenerate as soon as $w_{\rm c}$ moves away from zero (even at a value as small as $w_{\rm c} \simeq 10^{-6}$). This degeneracy eventually disappears when there is a level crossing in the groundstate sector of the transfer matrix, at some finite $w_{\rm c}$.

We have measured the position of this level crossing as a function of system size and extrapolated it to the thermodynamic limit. To test the reliability of the method, we have first applied it to the $(n_{\rm b}, n_{\rm g}) = (1, 1)$ case. Our final estimate $w_{\rm c} = 1.52 \pm 0.02$ at fixed $w_{\rm X} = 1$ is in good agreement with the exact result $w_{\rm c} = 3/2$ [10]. The same method applied to the Flory case, $(n_{\rm b}, n_{\rm g}) = (0, 1)$, yields $w_{\rm c} = 1.4294 \pm 0.0005$ at $w_{\rm X} = 1$ and $w_{\rm c} = 1.958 \pm 0.005$ at $w_{\rm X} = 2$. These values are clearly different from those predicted by the 8V model.

In conclusion, we believe that it would be most interesting to study the generalized eight-vertex model in more detail, using the exact techniques of integrable systems. In particular, it is conceivable that the present treatment misses some subtle exceptional points in the phase diagram.

VII. DISCUSSION

The semiflexible loop model was defined as a generalization of the two-flavor fully packed loop model on the square lattice, by introducing a vertex weight associated with vertices at which the loop does not undergo a 90° turn. We have proposed an effective field theory of the semiflexible loop model based on its height representation. This leads to exact results for the Flory model of polymer melting in two dimensions. Furthermore we have shown that the loop model provides a generalization of the eight-vertex model with an interesting phase diagram. Here we comment further on these two main results.

A. Scaling of semiflexible compact polymers

Polymers configurations are random and as such they are described by probability distributions. Their critical nature, in the long chain limit, is revealed by the fact that these distributions have scaling forms characterized by universal exponents. The simplest distribution is the probability $p(r,l)$ that the end-to-end distance equals r for a polymer of contour length l. In the scaling limit, when r is much greater than the lattice spacing and much less than the radius of gyration of the polymer, we have [1]

$$p(r,l) = r^{\theta} f(r/l^{\nu}) \ . \tag{36}$$

Here f is a scaling function, ν is the " swelling exponent", and θ the "cyclization exponent".

For semiflexible compact polymers, which correspond to the Flory model with $w_{\rm X} \leq w_{\rm X}^{\rm c}$, the swelling exponent is $\nu = 1/2$. This is an exact result which simply follows from the fact that compact polymers are space filling, regardless of $w_{\rm X}$. Furthermore, the swelling exponent can be related to the string dimension $x_{2,0}$ through the scaling law $\nu = (2 - x_{2,0})^{-1}$ [15]. Then replacing $e_{\rm b} = 1/2$ and $e_{\rm g} = 1/3$ in Eq. (21) gives $x_{2,0} = 0$ and $\nu = 1/2$, for all values of $w_{\rm X}$. This calculation then serves as a non-trivial check on the field theory.

The cyclization exponent is related to the scaling dimension associated with one real and one ghost loop segment: $\theta = -2x_{1,1}$ [15]. From Eq. (23) it follows that θ will vary continuously as the polymer is made stiffer by increasing $w_{\rm X}$. At the melting transition the exact result for θ follows from the computed value of the critical elastic constant, Eq. (15),

$$\theta^{\rm c} = -\frac{5}{8} \ . \tag{37}$$

The negative value implies that at the transition (and slightly below it) the two ends of the polymer feel an effective *attraction*. This is surprising as the naive expectation is that the two ends of a polymer will feel an effective repulsion due to the self-avoiding constraint. For stiff compact polymers this naive expectation is not met. Whether this will persist in three dimensions is an interesting open question.

B. Generalized eight-vertex model

The generalized 8V model gives a quite detailed modelization of two-dimensional lattice polymers. It possesses the following features:

- Steric constraints (self-avoidance and connectedness of the polymer chains) are modelled exactly;
- Possibility of introducing polydispersity, by taking $n_{\rm b}$ away from zero;
- A bending rigidity parameter $w_{\rm X}$ allows to control the transition between a melt and a crystalline phase;
- A contact interaction parameter $w_{\rm c}$ (or alternatively a fugacity of a vacancy) controls the transitions between compact, dense, dilute, Potts-like and non-critical phases;
- Possibility of introducing non-local interactions (although of a peculiar form), by taking $n_{\rm g}$ away from one.

The phase diagram of a somewhat similar model was studied in the Bethe approximation by Lise, Maritan, and Pelizzola [29]. However, in the compact limit the results of these authors are equivalent to Flory's mean-field treatment, as they do not take into account the non-local features of the polymers. We have here treated the excluded-volume effects in an exact manner. On the other hand, the model of Ref. [29] includes an additional feature:

- A contact interaction between non-consecutive monomers that are nearest neighbors on the lattice allows to drive the model to tricriticality, i.e., to access the theta-point physics.

This interaction is not present in the generalized 8V-model. If we were to include it, we would need the contact interaction to be flavor dependent (the authors of Ref. [29] do not consider what we refer to as ghost loops). However, we do not believe that anything new can be learnt from such a generalization. First, the contact interaction is redundant in the compact phase ($w_{\rm c} = 0$), as the number of contacts is constant (actually maximal) in any fully packed configuration. Second, in the non-compact phases ($w_{\rm c} \neq 0$) our field theory predicts a decoupling into two independent O(n) models. One would expect the flavor-dependent contact interactions to act independently on the two decoupled models, and the problem essentially reduces to that of the theta-point physics of a standard O(n) model [30].

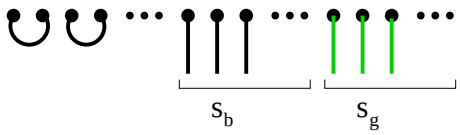


FIG. 11: Reference state used for generating the sector with s_k flavor-k strings ($k = \mathrm{b}, \mathrm{g}$).

FIG. 12: Reference state used for generating the thermal sector.

We leave it as an interesting question whether the generalized 8V model can be tackled using the methods of integrable systems. From Fig. 9 it can obviously be formulated as a forty-vertex model (taking into account the loop orientations) with complex vertex weights. To our knowledge, such a model has not been studied previously. If one could solve it, it would be particularly interesting to work out the exact expression of the coupling constant $K_{22}(w_{\mathrm{X}})$ as a function of the loop fugacities n_{b} and n_{g}. Also, it is conceivable that our field theoretical approach has missed some exceptional critical points in the phase diagram.

C. Order of the melting transition

In this paper we have established that the order of the melting transition within the Flory model is second order, as first suggested by Saleur [13]. We have also explained how the introduction of a finite density of vacancies may lead to a first order transition, as observed in Monte Carlo simulations [12]. This combined scenario settles a long controversy in the literature [8].

APPENDIX A: CONSTRUCTION OF THE TRANSFER MATRICES

The transfer matrix construction of Ref. [15] relied on an explicit bijection between the set of allowed connectivity states $\mathcal{C}$ and the set of integers $Z_{|\mathcal{C}|} = \{1, 2, 3, \ldots, |\mathcal{C}|\}$. However, in many cases it is difficult to furnish an *a priori* characterization of the set of allowed basis states and its cardinality. Moreover, some of the states utilized in [15] were found to carry zero weight in the leading eigenvectors of the corresponding sectors of the transfer matrix, and so one should think that it would be possible to eliminate them from the outset.

To remedy this situation it is preferable to use another approach. Without prior knowledge of the state space, the latter is explicitly generated by acting with the transfer matrix $\mathcal{T}$ on a reference state $|v_0\rangle$ which is known to belong to the image of $\mathcal{T}$ in the concerned sector. In this way, a certain number of image states is generated, which can be inserted in an appropriate data structure using hashing techniques [31]. One then acts with $\mathcal{T}$ on these states, generating a new list of states, and continues in this way until no new states are generated. The resulting list is the complete state space of $\mathcal{T}$ in the concerned sector.

It remains to find an appropriate reference state $|v_0\rangle$ for each physically interesting sector of $\mathcal{T}$. For the sector $(s_{\mathrm{b}}, s_{\mathrm{g}})$ in which s_k flavor-k strings ($k = \mathrm{b}, \mathrm{g}$) span the length of the cylinder generated upon action of $\mathcal{T}$, the reference state can be chosen as shown in Fig. 11. This state simply consists of $(L - s_{\mathrm{b}} - s_{\mathrm{g}})/2$ real arches followed by s_{b} real strings and s_{g} ghost strings. L must of course have the same parity as $s_{\mathrm{b}} + s_{\mathrm{g}}$.

Choosing the sector corresponding to the thermal scaling dimension is a little less obvious. A useful observation is made by exploiting the correspondence with states of the six-vertex model, as depicted in Fig. 3. In a given row, let N_k^{even} and N_k^{odd} be the number of flavor-k loop segments residing on even and odd vertical edges, respectively. Then define

$$Q = \left(N_{\mathrm{b}}^{\mathrm{even}} - N_{\mathrm{b}}^{\mathrm{odd}}\right) - \left(N_{\mathrm{g}}^{\mathrm{even}} - N_{\mathrm{g}}^{\mathrm{odd}}\right). \tag{A1}$$

By inspection of Fig. 3 it is seen that Q is nothing but the vertical flux of arrows within a given row. By the ice rule, Q is a conserved quantity and can thus be used to label a sector of $\mathcal{T}$.

The reference state of Fig. 11 with $(s_{\mathrm{b}}, s_{\mathrm{g}}) = (0, 0)$ is seen to have $Q = 0$. The first excited state with no strings has $Q = \pm 4$ and is depicted on Fig. 12. Its first four sites are occupied by mutually penetrating real and ghost arches,

L	$(0,0)$	Thermal	$(2,0)$	$(1,1)$	$(4,0)$	$(3,1)$	$(2,2)$
2	2	—	1	1	—	—	—
4	8	1	8	12	1	2	2
6	46	12	69	141	15	42	72
8	332	124	664	1720	196	684	1056
10	2784	1280	6960	21760	2520	10320	14800
12	25888	13605	77664	283584	32565	151500	205920
14	259382	149604	907837		425019		

TABLE II: Sizes of various sectors of the SFL model transfer matrix defined on a cylinder of width L with periodic boundary conditions. The symbol $(s_\mathrm{b}, s_\mathrm{g})$ labels the sector in which s_k flavor-k strings ($k = \mathrm{b}, \mathrm{g}$) run along the length of the cylinder. The sectors $(0,0)$ and 'Thermal' have no strings, but the parity of the number of flavor-crossings in the basis states is fixed to be even and odd, respectively.

followed by $(L-4)/2$ simple real arches. In general, for any given L, states with $Q = \pm 4q$ exist for $q = 0, 1, \ldots, \lfloor L/4 \rfloor$. The number of states in the k'th sector is just

$$\sum_{n=q}^{L/2-q} \binom{L/2}{n+q} \binom{L/2}{n-q} C_{L/2-n} C_n, \tag{A2}$$

where $C_n = \frac{(2n)!}{n!(n+1)!}$ are the Catalan numbers. Using a sum rule on the binomial coefficients, it is easily seen that the total number of states without strings, summed over the sector index q, reads simply

$$\sum_{n=0}^{L/2} \binom{L}{2n} C_{L/2-n} C_n. \tag{A3}$$

This is nothing but the dimension of the state space used in [15].

From entropic reasons it is fairly obvious that the free energy belongs to the sector $q = 0$. We are now going to argue that the thermal scaling exponent is linked to the gap between the first eigenvalue in the $q = 0$ and $q = 1$ sectors, cf. Eq. (25). The first reason is that, by construction, the $q = 1$ constraint acts as an excitation within the full state space (with all values of q included, as in Eq. (A3)), and hence should correspond to a subdominant eigenvalue within that space. Indeed, it is observed numerically that the second eigenvalue obtained from the transfer matrix of Ref. [15] coincides with the leading eigenvalue of the $q = 1$ sector, obtained by using the techniques outlined above.

As a second argument, note that in the language of the SFL model height mapping, encircling the first four sites of Fig. 12 yields a height dislocation of $\mathbf{A} - \mathbf{C} + \mathbf{B} - \mathbf{D}$. By the four-coloring rule, $\mathbf{A} + \mathbf{B} + \mathbf{C} + \mathbf{D} = \mathbf{0}$, this is the same as

$$\mathbf{m}_T = 2(\mathbf{A} + \mathbf{B}) = -2(\mathbf{C} + \mathbf{D}). \tag{A4}$$

But the latter is *also* the height defect associated with a defect vertex $(\mathbf{C}, \mathbf{D}, \mathbf{C}, \mathbf{D})$ that corresponds to excluding the real loops from that vertex, which is exactly a thermal-type excitation (and to wit the one that is used for computing the critical exponent x_T within the field theory).

In the field theory, one might compute the exponent corresponding to the insertion of magnetic defects $\pm q' \mathbf{m}_T$ at either end of the cylinder. In the transfer matrix, these should simply be linked to the gap between the sectors $q = 0$ and $q = q'$.

In Table II we show the sizes of the various transfer matrices used in this work. The columns labelled $(0,0)$ and 'Thermal' correspond to the expressions (A2) with $q = 0$ and $q = 1$, respectively. For the other columns, similar expressions may be worked out along the lines of [15].

Finally, let us remark that the computations for the generalized eight-vertex model introduced in Section VI B are produced from the same reference states, but slightly generalizing the transfer matrix to accommodate the contact-type vertices shown on Fig. 9.

ACKNOWLEDGMENTS

We are grateful to Ken Dill for introducing us to the Flory model of polymer melting. JK would further like to thank the KITP in Santa Barbara for hospitality, where this work was initiated. The research of JK is supported by

the NSF under grant number DMR-9984471. JK is a Cottrell Scholar of Research Corporation.

---

[1] P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
[2] B. Nienhuis, Phys. Rev. Lett. **49**, 1062 (1982).
[3] B. Miller and J.O. Rädler, Phys. Rev. Lett. **82**, 1911 (1999).
[4] H. W. J. Blöte and B. Nienhuis, J. Phys. A **22**, 1415 (1989).
[5] M. T. Batchelor, B. Nienhuis and S. O. Warnaar, Phys. Rev. Lett. **62**, 2425 (1989).
[6] K. A. Dill *et al.*, Protein Sci. **4**, 561 (1995).
[7] P.J. Flory, Proc. R. Soc. London, Ser. A **234**, 60 (1956).
[8] G. I. Menon and R. Pandit, Phys. Rev. E **59**, 787 (1999).
[9] J.F. Nagle, Proc. R. Soc. London, Ser. A **337**, 569 (1974).
[10] R.J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, New York, 1982).
[11] P.D. Gujrati and M.N. Goldstein, J. Chem. Phys. **74**, 2596 (1981).
[12] A. Baumgartner and D. Yoon, J. Chem. Phys. **79**, 521 (1983).
[13] H. Saleur, J. Phys. A **19**, 2409 (1986).
[14] J. Bascle, T. Garel and H. Orland, J. Phys. A **25**, L1323 (1992).
[15] J. L. Jacobsen and J. Kondev, Nucl. Phys. B **532** [FS], 635–688 (1998).
[16] J. Kondev and C. L. Henley, Phys. Rev. B **52**, 6628 (1995).
[17] J. Kondev, Phys. Rev. Lett. **78**, 4320 (1997).
[18] Vl. S. Dotsenko and V. A. Fateev, Nucl. Phys. B **240**, 312 (1984); *ibid.* **251**, 691 (1985).
[19] H. Orland, C. Itzykson, and C. de Dominicis, J. Phys. Lett. (Paris) **46**, 353 (1985).
[20] S. Higuchi, Nucl. Phys. B **540**, 731 (1999).
[21] J. Cardy, *Scaling and Renormalization in Statistical Physics*, (Cambridge University Press, 1996).
[22] B. Nienhuis, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and J. L. Lebowitz (Academic, London, 1987), Vol. 11.
[23] J. L. Jacobsen and J. Kondev, J. Stat. Phys **96**, 21–48 (1999).
[24] H. W. J. Blöte, J. L. Cardy and M. P. Nightingale, Phys. Rev. Lett. **56**, 742 (1986); I. Affleck, *ibid.* **56**, 746 (1986).
[25] J. L. Cardy, J. Phys. A **16**, L355 (1983).
[26] D. Kim and P. A. Pearce, J. Phys. A **20**, L451 (1987).
[27] H. Saleur, J. Phys. A **19**, L807 (1986).
[28] B. Duplantier and H. Saleur, Nucl. Phys. B **290**, 291 (1987).
[29] S. Lise, A. Maritan, and A. Pelizzola, cond-mat/9809210.
[30] B. Duplantier and H. Saleur, Phys. Rev. Lett. **59**, 539 (1987).
[31] R. Sedgewick, *Algorithms in C* (Addison-Wesley, 1990).
[32] We shall usually consider $n_{\rm b}$ and $n_{\rm g}$ real and non-negative, though we believe the model to have interesting properties for negative, or even complex, fugacities.
[33] The vertex weights of the six-vertex model are conventionally called a, b and c. We shall henceforth trade the variable c for $w_{\rm X}$ to avoid confusion with the central charge, which will play a major role in what follows.
[34] Note in particular that the field theory construction only applies to geometries that are isomorphic to the sphere with a point at infinity under a conformal transformation. Generalizations to higher genera have so far only been treated in mean-field theory [19, 20].

5.9 Article "Classification of conformal field theories"

Nous présentons une méthode de classification des théories conformes basées sur le gaz de Coulomb (construction aux champs bosoniques libres). Etant donné une configuration géométrique particulière des charges d'écran, nous produisons des conditions nécessaires pour l'existence des représentations dégénérées et pour la fermeture de l'algèbre des opérateurs de vertex. La classification résultante comprend la classification habituelle basée sur les algèbres de Lie classiques, mais elle est plus générale. Nous appliquons ensuite la méthode à la théorie de type gaz de Coulomb pour le modèle de deux types de boucles étudiée précédemment par Jacobsen et Kondev. Le but de l'étude est d'éclaircir la relation entre les modèles basés sur le gaz de Coulomb et les théories conformes avec des symétries étendues.

ELSEVIER

Nuclear Physics B 618 [FS] (2001) 523–550

NUCLEAR PHYSICS B

www.elsevier.com/locate/npe

Classification of conformal field theories based on Coulomb gases. Application to loop models

Vladimir S. Dotsenko [a], Jesper Lykke Jacobsen [b], Marco Picco [a]

[a] *LPTHE,* [1] *Université Pierre et Marie Curie, Paris VI, Université Denis Diderot, Paris VII, Boîte 126, Tour 16, 1er étage, 4 place Jussieu, F-75252 Paris Cedex 05, France*

[b] *Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Bâtiment 100, 91405 Orsay, France*

Received 14 June 2001; accepted 15 August 2001

Abstract

We present a method for classifying conformal field theories based on Coulomb gases (bosonic free-field construction). Given a particular geometric configuration of the screening charges, we give necessary conditions for the existence of degenerate representations and for the closure of the vertex-operator algebra. The resulting classification contains, but is more general than, the standard one based on classical Lie algebras. We then apply the method to the Coulomb gas theory for the two-flavoured loop model of Jacobsen and Kondev. The purpose of the study is to clarify the relation between Coulomb gas models and conformal field theories with extended symmetries.

PACS: 05.50.+q; 64.60.Fr; 75.10.Hk

1. Introduction

Conformal field theories (CFT) based on Coulomb gases arise in a variety of two-dimensional problems in statistical physics [1,2]. Furthermore, a great number of CFTs are known to possess a Coulomb gas formulation. Early examples, such as the Potts model or the critical and tricritical points in the O(n) model, were based on a single scalar field, and the physical operators could be interpreted as particles carrying scalar quantised electric and magnetic charges.

More recently, multicomponent Coulomb gases employing several bosonic free fields have appeared in the study of critical phases in the so-called fully-packed loop models [3].

E-mail addresses: dotsenko@lpthe.jussieu.fr (V.S. Dotsenko), jacobsen@ipno.in2p3.fr (J.L. Jacobsen), picco@lpthe.jussieu.fr (M. Picco).

[1] Unité Mixte de Recherche CNRS UMR 7589.

0550-3213/01/$ – see front matter
PII: S0550-3213(01)00409-6

A first step in their resolution consists in bijectively mapping configurations of oriented loops to those of a discretised surface. The basic idea is here to interpret the loops as contour lines of the surface height, but due to the fully-packing constraint it turns out that the height variables in general have to be vector valued. Based on symmetry and entropic considerations, an effective action of the Liouville type can then be written down for the continuum limit of this interfacial representation. However, this action contains a certain number of elastic constants whose exact values cannot be inferred directly from the discrete model.

Important technical progress was achieved with the use of the *loop ansatz*. It states that the most relevant vertex operators in a given model have to be exactly marginal and taken into the action as screenings, thus allowing all the elastic constants to be fixed [4]. This situation is truly remarkable: the discrete model precludes an a priori knowledge of the parameters defining its continuum limit, but it nevertheless fixes the geometry of the screening charges and thus permits an exact a posteriori determination of the very same parameters. As the end result one obtains a CFT in the form of a Coulomb gas (in general multicomponent, the number of components being the dimensionality of the height space) with a given background charge and screening operators. Physical operators are represented by vertex operators

$$V_{\vec{\beta}}(z,\bar{z}) = :\mathrm{e}^{i\vec{\beta}\vec{\varphi}(z,\bar{z})}:, \tag{1.1}$$

with $\vec{\beta}$ taking values on a particular lattice, specific to the given model. The vector $\vec{\varphi}(z,\bar{z})$ contains a set of bosonic free fields $\vec{\varphi}(z,\bar{z}) = \{\varphi_1(z,\bar{z}), \varphi_2(z,\bar{z}), \ldots, \varphi_D(z,\bar{z})\}$, which can be interpreted as the continuum limit of the components of the discrete interfacial height.

The question which appears naturally is whether there is a chiral algebra hidden behind this conformal theory, defined as a multicomponent Coulomb gas, and what it might be. In particular, given the fact that fully-packed loop models may possess a central charge $c > 1$ [3], one may wonder if the chiral algebra could be bigger than just the Virasoro one, ensured by the conformal invariance of the model. Put differently: might a particular model, represented in the continuum limit as a Coulomb gas, have additional, or extended symmetries? By this we mean not just global symmetries, which are usually explicit, but symmetries which are extended on the level of the chiral algebra, i.e., extended infinitely. In case of a positive answer this would imply the existence of extra chiral operators, like $W(z)$ operators, which form an extended chiral algebra together with the stress-energy tensor $T(z)$.

Presence of such symmetries in the Coulomb gas would not automatically ensure that they persist in the original model defined on the lattice, and that the model would be integrable. Still, in case of a positive answer for the continuum theory (existence of extended symmetries) but negative for the lattice model, one could try to find an integrable version of it on the lattice.

On the other hand, in case of a negative answer in the continuum (no extra chiral operators in the multicomponent Coulomb gas model, in addition to $T(z)$) the prediction for the lattice model is likely to be definitive: extended symmetries will not come about on the lattice, and the model will not be integrable. It is in this perspective, looking for

the existence of extra symmetries or their absence, that we shall try to clarify the relation between Coulomb gas models and CFTs which are based on extended chiral algebras.

The general method is presented in Section 2. We here study the necessary conditions for the existence of degenerate representations and for the closure of the vertex-operator algebra. These conditions are of course met by all W-type extended CFTs which are based on simply laced Lie algebras, but in fact they are less restrictive. In particular they are fulfilled by the two-flavour fully-packed loop model of Jacobsen and Kondev [3], whose continuum limit is reminiscent of, but not identical to, the classical WA_3 theory. To assess whether such theories actually contain extra chiral operators we work on a case-to-case basis, computing the exact form of such operators or proving that they do not exist. In Section 3 this method is applied to a variety of two- and three-component Coulomb gases, in particular to those which arise in the solution of loop models. We finally present our conclusions in Section 4.

2. General method

2.1. Background charge and screenings

The primary information defining a Coulomb gas model can be summarised as follows:

1. Dimensionality of the Coulomb gas (number of free-field components), D.
2. The background charge, $\vec{\alpha}_0$.
3. A set of D screening operators,

$$\left\{:\mathrm{e}^{i\vec{\alpha}_a\vec{\varphi}(z,\bar{z})}: \mid a = 1, 2, \ldots, D\right\}. \tag{2.1}$$

We shall assume that, for a generic model, the number of screenings equals the dimensionality of the Coulomb gas. Namely, if the number of screenings were less than D then a subset of free fields would decouple, and the model could be represented as a direct product of a submodel with the background charge and screenings and a submodel of actual free fields, no background charge, and no screenings. If, alternatively, due to a special symmetry of the lattice to which the screenings belong, the number of screenings exceeds D, a subset of exactly D screenings ought to be sufficient to define correlation functions of physical operators. Different choices of a subset of D screenings should give equivalent results. At least this should be the case for a Coulomb gas with an underlying chiral algebra, which classifies all the operators in the model, and with respect to which the physical operators are primaries. This last point will be made more precise in the following. For the moment we do not have any reasons for assuming such special symmetries of the lattice to which the screenings should belong.

The primary information on a Coulomb gas, as stated above, is sketched in Fig. 1. For various lattice models, like Potts or O(n) models [1,2], or more general loop models [3], the geometry of $(\{\vec{\alpha}_a\}, \vec{\alpha}_0)$ is known explicitly. But we shall here derive general constraints on the possible geometries, by using consistency conditions of the corresponding CFT.

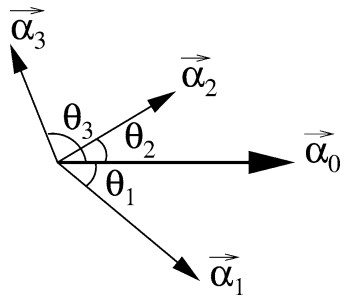


Fig. 1. Background charge vector $\vec{\alpha}_0$ and the vectors $\{\vec{\alpha}_a\}$ defining the screening operators $V_{\vec{\alpha}_a}(z,\bar{z}) = :\mathrm{e}^{i\vec{\alpha}_a\vec{\varphi}(z,\bar{z})}:$ for $a = 1, 2, \ldots, D$. On the figure, $D = 3$. $\{\Theta_a\}$ are the angles which $\{\vec{\alpha}_a\}$ make with the background charge $\vec{\alpha}_0$.

We take $\vec{\alpha}_0$ as given. The stress-energy tensor of D free fields $\{\varphi_a(z,\bar{z})\}$ will be taken in the form

$$T(z) = -\frac{1}{4}:\partial\vec{\varphi}(z)\partial\vec{\varphi}(z): + i\vec{\alpha}_0\partial^2\vec{\varphi}(z), \tag{2.2}$$

with the two-point correlation functions of the fields $\{\varphi_a(z,\bar{z})\}$ normalised as

$$\left\langle \varphi_a(z,\bar{z})\varphi_b(z',\bar{z}')\right\rangle = 2\delta_{a,b}\log\frac{1}{|z-z'|^2}. \tag{2.3}$$

With this normalisation, the conformal dimension (with respect to the stress-energy tensor (2.2)) of a vertex operator

$$V_{\vec{\alpha}}(z,\bar{z}) = :\mathrm{e}^{i\vec{\alpha}\vec{\varphi}(z,\bar{z})}: \tag{2.4}$$

will be equal to

$$\Delta_{\vec{\alpha}} = \vec{\alpha}^2 - 2\vec{\alpha}\vec{\alpha}_0. \tag{2.5}$$

The first condition on the screening operators (2.1) is that they have to be marginal, $\Delta_{\vec{\alpha}_a} = 1$ for $a = 1, 2, \ldots, D$. This condition ensures that contour integrals of the screenings

$$Q_a = \oint_C dz V_{\vec{\alpha}_a}(z), \quad a = 1, 2, \ldots, D, \tag{2.6}$$

commute with the Virasoro algebra generated by $T(z)$. The operator $V_{\vec{\alpha}_a}(z)$ in (2.6) is assumed to be the holomorphic part of the screening operator $V_{\vec{\alpha}_a}(z,\bar{z})$ in (2.1), in the sense of holomorphic–antiholomorphic factorisation of correlation functions of vertex operators, or, more generally, in the sense of holomorphic–antiholomorphic factorisation of two-dimensional integrals of correlation functions of vertex operators with respect to the contour integrals [5,6].

To make the screenings marginal, the vectors $\{\vec{\alpha}_a\}$ in (2.1) have to satisfy the condition

$$\Delta_{\vec{\alpha}_a} = \vec{\alpha}_a^2 - 2\vec{\alpha}_a\vec{\alpha}_0 = 1, \tag{2.7}$$

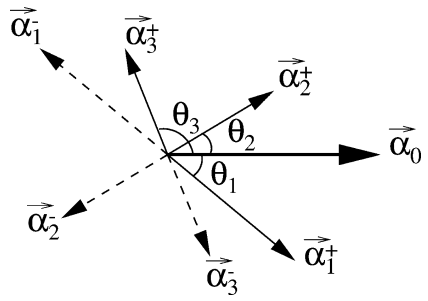


Fig. 2. The complete set of screenings $\{\vec{\alpha}_a^+, \vec{\alpha}_a^-\}$, and the background charge $\vec{\alpha}_0$.

or

$$\alpha_a^2 - 2\alpha_a\alpha_0\cos\Theta_a = 1, \tag{2.8}$$

$$\alpha_a^{\pm} = \alpha_0\cos\Theta_a \pm \sqrt{1+\alpha_0^2\cos^2\Theta_a}. \tag{2.9}$$

Here $\{\Theta_a\}$ are the angles in Fig. 1, and $\alpha_a^{\pm}$ are the "lengths" of the vectors $\vec{\alpha}_a$ (one positive, α_a^+, and one negative, α_a^-), which satisfy the condition (2.7).

In Fig. 1 we have indicated a particular geometry of the vectors $\vec{\alpha}_0$ and $\{\vec{\alpha}_a\}$. This choice of geometry is for the moment arbitrary. The only constraint, once the directions of $\vec{\alpha}_0$ and $\{\vec{\alpha}_a\}$ have been chosen, is on the lengths of the screenings, Eq. (2.9). There are two solutions, "+" and "−", for each direction, and we shall see in the following that the consistence of the corresponding CFT requires the use of them both. Since the "lengths" $\{\alpha_a^-\}$ are negative, the screening vectors $\{\vec{\alpha}_a^-\}$ are oriented in the opposite direction with respect to $\{\vec{\alpha}_a^+\}$. The set of screenings has thus been doubled, D screenings $\{\vec{\alpha}_a^+\}$ and D screenings $\{\vec{\alpha}_a^-\}$, as indicated in Fig. 2. We also note the relations between α_a^+ and α_a^-

$$\alpha_a^+ + \alpha_a^- = 2\alpha_0\cos\Theta_a, \tag{2.10}$$

$$\alpha_a^+\alpha_a^- = -1, \tag{2.11}$$

which follow from Eq. (2.8). These relations will be used in the following.

2.2. Necessary condition for degenerate representations. Kac formula

We shall be looking for a Coulomb gas theory based on a chiral algebra. The number of chiral operators forming this algebra has to be equal to the number of free fields, D. In addition to $T(z)$, this requires $D-1$ extra chiral operators to control all the degrees of freedom of the theory, i.e., to control D free fields.

We shall assume that these extra chiral operators belong to the module of the identity operator, just like $T(z)$, i.e., that they are made as linear combinations of products of derivatives of free fields. For a Coulomb gas theory based on D bosonic fields this last assumption appears to us to be natural. This is essentially because in this case the chiral operators will have zero Coulomb charge; their correlation functions will therefore not require the integrated screening operators, they will be "simple".

It should be noted that this argument is valid only in the case of $\vec{\alpha}_0 \neq 0$, where, in general, correlation functions of generic vertex operators *do* require screenings. Generic correlation

functions are therefore not simple, but those made exclusively of chiral symmetry operators are. This argument implicitly supposes that an acceptable Coulomb gas theory will allow for a deformation of the trivial free-field point $\vec{\alpha}_0 = 0$, which has $c = D$.

Assuming therefore that the extra chiral operators belong to the module of the identity, the possible chiral algebras for Coulomb gas theories should be made of operators with integer conformal dimensions, $\Delta \in \mathbb{N}$. Furthermore, since a dimension-one current is known to generate a continuous rather than a discrete symmetry, we require $\Delta \geqslant 2$. This means that the chiral algebras should be of the W type. If this algebra exists, the screening operators (integrated along the contours) have to commute with it, in order to respect the extended symmetry of the given theory.

For chiral operators in the form of linear combinations of products of derivatives of free fields, the operators which are primary with respect to them are the vertex operators

$$V_{\vec{\beta}}(z,\bar{z}) = :\mathrm{e}^{i\vec{\beta}\vec{\varphi}(z,\bar{z})}:, \tag{2.12}$$

where $\vec{\beta}$ is for the moment arbitrary. In fact, an operator $\Phi(z,\bar{z})$ being primary with respect to a particular chiral operator $W(z)$ means that the operator product expansion (OPE) of $W(z)$ with $\Phi(z,\bar{z})$ takes the form

$$W(z)\Phi(z',\bar{z}') = \frac{A}{(z-z')^{\Delta_W}}\Phi(z',\bar{z}') + \cdots, \tag{2.13}$$

where A is a (structure) constant. Loosely speaking, the most singular term in the expansion in powers of $(z - z')$ produces again the operator $\Phi(z,\bar{z})$. Evidently, when $W(z)$ is a linear combination of products of derivatives of free fields, each term involving a total of Δ_W derivatives, and when $\Phi(z,\bar{z}) = V_{\vec{\beta}}(z,\bar{z})$ is an exponential of free fields, cf. Eq. (2.12), the expansion $W(z)V_{\vec{\beta}}(z',\bar{z}')$ will be precisely of the form (2.13).

Until now $\vec{\beta}$ has been an arbitrary vector. For special values of $\vec{\beta}$ it may however happen that primary operators appear in the module of $V_{\vec{\beta}}(z,\bar{z})$. They are called singular states, and their presence means degeneracy of the module. Since the integrated screenings commute with the operators of the chiral algebra, the singular state in the module of $V_{\vec{\beta}}(z,\bar{z})$ may be obtained by mapping, with the use of screenings, another vertex operator $V_{\vec{\beta}'}(z,\bar{z})$ into the module of $V_{\vec{\beta}}(z,\bar{z})$. This is brought about in the following way:

$$S\big(V_{\vec{\beta}}(z,\bar{z})\big) = \prod_{a=1}^{D}\left(\oint_{C_a} du_a V_{\vec{\alpha}_a^+}(u_a)\right)^{n_a} V_{\vec{\beta}'}(z,\bar{z}), \tag{2.14}$$

where $n_a \in \mathbb{N}_0$ for $a = 1, 2, \ldots, D$. Here $S(V_{\vec{\beta}})$ is a singular state in the module of $V_{\vec{\beta}}$, and $\vec{\beta}'$ should be related to $\vec{\beta}$ by

$$\vec{\beta}' = \vec{\beta} - \sum_a n_a \vec{\alpha}_a^+. \tag{2.15}$$

The configuration of contours $\{C_a\}$ in Eq. (2.14) is shown in Fig. 3. The set of contours in the figure is only schematic, since there are in fact n_a distinct contours corresponding to the screenings of type $a = 1, 2, \ldots, D$.

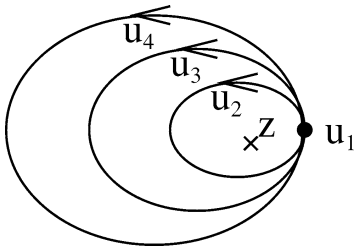


Fig. 3. Having integrated first all the screenings except for one type (whose choice is not important), one finally integrates the last screening, at the point u_1, around z.

This prescription for realising closed contour integrations of screenings around a fixed vertex operator, on which they act to induce the mapping (2.14), was first used in Ref. [7].

More precisely, the mapping of Eq. (2.14) is defined, and the singular state $S(V_{\vec{\beta}})$ exists, only if

$$\Delta_{\vec{\beta}'} - \Delta_{\vec{\beta}} = N, \tag{2.16}$$

where N is a positive integer. In fact, as the screenings commute with $T(z)$, the conformal dimension of $S(V_{\vec{\beta}})$ is equal to that of $V_{\vec{\beta}'}$. Being a descendent state, the conformal dimension of $S(V_{\vec{\beta}})$ should differ from $\Delta_{\vec{\beta}}$ by an integer. One thus obtains the condition (2.16).

Another way to get (2.16) is to require that the monodromy for the analytic continuation of the expression on the right-hand side of (2.14) with respect to the common variable u_1 (cf. Fig. 3) be trivial, so that the contour of integration over u_1 around z closes. Otherwise the integrated screenings would not actually commute with $T(z)$ and with other operators of the chiral algebra. It can be checked that this trivial monodromy requirement leads equivalently to the condition (2.16) [7].

This condition, with $\vec{\beta}'$ defined by (2.15) and $\Delta_{\vec{\beta}}$, $\Delta_{\vec{\beta}'}$ given by (2.5), fixes special values of $\vec{\beta}$ for which the module of $V_{\vec{\beta}}$ contains singular states, i.e., is degenerate.

Note that we have only applied the "+" screenings in Eqs. (2.14), (2.15). The reason is that once the solution for $\vec{\beta}$ is found, it will automatically satisfy the corresponding condition with negative screenings. For the analysis of singular states the "+" and "−" screenings are therefore complimentary. A similar remark holds true for a mixed mapping, i.e., one involving both "+" and "−" screenings: the corresponding conditions do not lead to new solutions. These statements should become more clear from the analysis that will follow.

We remark that this way of defining the degenerate representations, using the mapping by screenings between vertex operators and vertex operator modules, has probably been used first in Ref. [8] to reproduce the Kac formula for conformal dimensions for the degenerate representations of the Virasoro algebra. This method has also been used in [9,10] to define the degenerate representations of W-algebra CFTs based on classical Lie algebras.

Until now we have assumed a set of screenings, as depicted in Figs. 1, 2, with no restrictions yet, except for those on the lengths of screenings, cf. Eqs. (2.7)–(2.9). As we

are looking for Coulomb gas theories which should eventually be endowed with extended chiral algebras, we shall next identify the vertex operators $\{V_{\vec{\beta}}(z,\bar{z})\}$ which induce the degenerate representations of that chiral algebra.

Our approach can be outlined as follows: without knowing yet the chiral algebra, for a given set of screening operators we shall define, by using the mappings described above, the "would be degenerate representations". The screening operators are assumed to commute with the (presently unknown) chiral algebra. We shall define, for given screenings and background charge, the discrete set of values of $\vec{\beta}$ for which the operators $V_{\vec{\beta}}(z,\bar{z})$ would induce degenerate representations, if the chiral algebra really existed. In this way we shall arrive at a necessary condition for the existence of degenerate representations. The condition will also become sufficient when the corresponding extended chiral algebra is found.

The advantage of this approach will be that, having defined such special values of $\vec{\beta}$ corresponding to given screenings, it will allow us to derive next a set of constraints "acting backwards", i.e., that limits the allowed configurations of screenings and the background charge.

From the vast literature on W-algebras it is well-known that the proper way to look for singular states in the module of $V_{\vec{\beta}}(z,\bar{z})$ is to apply mappings of the type (2.14) by shifting $\vec{\beta}'$ away from $\vec{\beta}$ only along the "principal" directions defined by the individual screenings, and not as has been sketched schematically in Eqs. (2.14), (2.15), where several different screenings are involved simultaneously.

For the direction associated with the first screening, the mapping takes the form

$$S_1\big(V_{\vec{\beta}}(z,\bar{z})\big) = \left(\oint_C du\, V_{\vec{\alpha}_1^+}(u)\right)^n V_{\vec{\beta}-n\vec{\alpha}_1^+}(z,\bar{z}) = \prod_{i=1}^{n}\oint du_i\, V_{\vec{\alpha}_1^+}(u_i)\, V_{\vec{\beta}-n\vec{\alpha}_1^+}(z,\bar{z}), \tag{2.17}$$

cf. Figs. 3, 4. The condition (2.16) becomes

$$\Delta_{\vec{\beta}-n\vec{\alpha}_1^+} - \Delta_{\vec{\beta}} = N, \tag{2.18}$$

where N should be a positive integer. Substituting the formula (2.5) for $\Delta_{\vec{\beta}-n\vec{\alpha}_1^+}$ and $\Delta_{\vec{\beta}}$, one finds

$$\big(\vec{\beta}-n\vec{\alpha}_1^+\big)^2 - 2\vec{\alpha}_0\big(\vec{\beta}-n\vec{\alpha}_1^+\big) - \vec{\beta}^2 + 2\vec{\alpha}_0\vec{\beta} = N, \tag{2.19}$$

which gives

$$-2n\big(\vec{\beta}-\vec{\alpha}_0\big)\vec{\alpha}_1^+ + n^2\big(\alpha_1^+\big)^2 = N. \tag{2.20}$$

The general solution of this equation for the allowed values of $\vec{\beta}$ takes the form

$$2\big(\vec{\beta}-\vec{\alpha}_0\big) = \big(n\alpha_1^+ + n'\alpha_1^-\big)\vec{\omega}_1, \tag{2.21}$$

where the vector $\vec{\omega}_1$ should verify the equation $\vec{\omega}_1\vec{e}_1 = 1$ with $\vec{e}_1 = \vec{\alpha}_1^+/\alpha_1^+$, and n' in (2.21) is an another positive integer. By substituting (2.21) into (2.20) one finds $N = nn'$ so that, for $\vec{\beta}$ of the form (2.21), the constraint (2.18) is indeed verified.

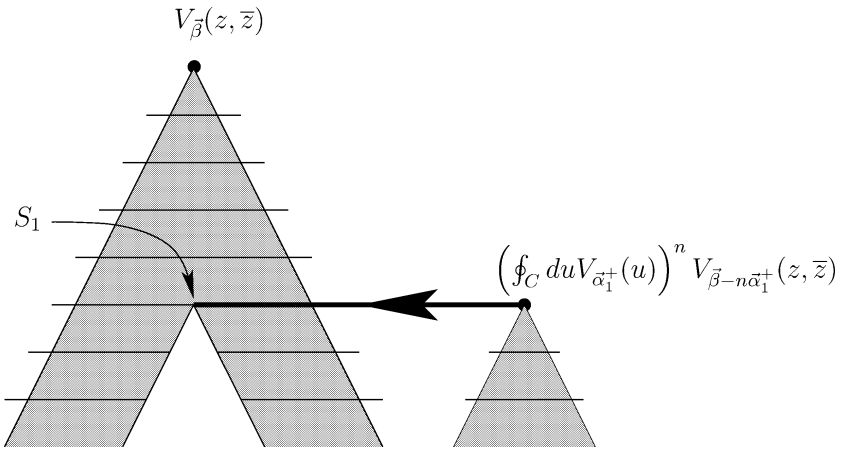


Fig. 4. Mapping of the module of $V_{\tilde{\vec{\beta}}} \equiv V_{\vec{\beta}-n\vec{\alpha}_1^+}$ into the module of $V_{\vec{\beta}}$, in the direction of the first screening $\vec{\alpha}_1^+$.

By similarly requiring that singular states be produced for each of the D screening directions, defined by the vectors $\{\vec{\alpha}_a^+\}$, one finds that $\vec{\beta}$ has to satisfy

$$2\big(\vec{\beta}-\vec{\alpha}_0\big)=\sum_{a=1}^{D}\big(n_a\alpha_a^+ + n'_a\alpha_a^-\big)\vec{\omega}_a. \tag{2.22}$$

Here the D vectors $\{\vec{\omega}_a\}$ are defined by

$$\vec{\omega}_a\vec{e}_b=\delta_{a,b}, \tag{2.23}$$

and $\{\vec{e}_a\}$ are the unit vectors which define the directions of the screenings:

$$\vec{\alpha}_a^{\pm}=\alpha_a^{\pm}\vec{e}_a \quad \text{for } a=1,2,\ldots,D. \tag{2.24}$$

Eq. (2.22) generalises (2.21). Alternatively it can be presented as

$$\vec{\beta}=\vec{\beta}_{(n'_1,n_1)(n'_2,n_2)\cdots(n'_D,n_D)}=\sum_{a=1}^{D}\left(\frac{n_a}{2}\alpha_a^+ + \frac{n'_a}{2}\alpha_a^-\right)\vec{\omega}_a+\vec{\alpha}_0. \tag{2.25}$$

This expression can be simplified by developing $\vec{\alpha}_0$ in terms of the vectors $\{\vec{\omega}_a/2\}$,

$$\vec{\alpha}_0=\sum_a X_a\frac{\vec{\omega}_a}{2}. \tag{2.26}$$

The expansion coefficients $\{X_a\}$ are then found by using the orthogonality property (2.23):

$$X_b=2\vec{\alpha}_0\vec{e}_b=2\alpha_0\cos\Theta_b=\alpha_b^+ + \alpha_b^-, \tag{2.27}$$

where in the last step we have used the relations (2.10) between α_a^+ and α_a^-. The angles $\{\Theta_a\}$ have been defined in Figs. 1, 2. Substituting (2.27) into (2.26), we arrive at

$$\vec{\alpha}_0=\sum_a\frac{\alpha_a^+ + \alpha_a^-}{2}\vec{\omega}_a, \tag{2.28}$$

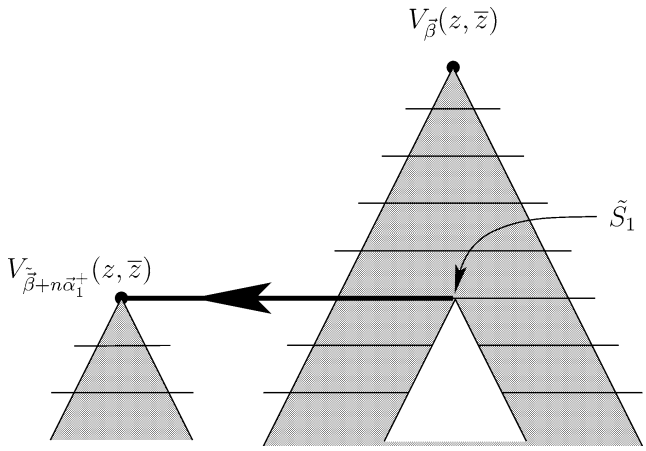


Fig. 5. The mapping is of the form $(\oint_C \mathrm{d}u\, V_{\vec{\alpha}_1^+}(u))^n \tilde{S}_1 = V_{\tilde{\vec{\beta}}+n\vec{\alpha}_1^+}$.

which can finally be inserted in Eq. (2.25) to yield the following set of values of $\vec{\beta}$:

$$\vec{\beta}_{(n'_1,n_1)(n'_2,n_2)\cdots(n'_D,n_D)} = \sum_{a=1}^{D}\left(\frac{1+n_a}{2}\alpha_a^+ + \frac{1+n'_a}{2}\alpha_a^-\right)\vec{\omega}_a. \tag{2.29}$$

For the "conjugate" values of $\vec{\beta}$, viz. $\tilde{\vec{\beta}} = 2\vec{\alpha}_0 - \vec{\beta}$, one finds:

$$\tilde{\vec{\beta}}_{(n'_1,n_1)(n'_2,n_2)\cdots(n'_D,n_D)} = \sum_{a=1}^{D}\left(\frac{1-n_a}{2}\alpha_a^+ + \frac{1-n'_a}{2}\alpha_a^-\right)\vec{\omega}_a. \tag{2.30}$$

The manifestation of singular states in the module of the vertex operator $V_{\tilde{\vec{\beta}}}(z,\bar{z})$ is so to say dual to that of $V_{\vec{\beta}}(z,\bar{z})$. If the module of $V_{\tilde{\vec{\beta}}}$ is degenerate, it contains particular primary states which produce, as images under appropriate mappings, the vertex operators $V_{\vec{\beta}'}$ several screenings away, as indicated in Fig. 5. More details on these properties can be found in the paper [7]. Physically, in theories with a chiral algebra, the set of dual operators $\{V_{\vec{\beta}}, V_{\tilde{\vec{\beta}}}\}$ can either represent the same (singlet) operator (such as the energy operator) or a doublet of conjugate operators (an example being the $\{\sigma, \sigma^\dagger\}$ spin operators in the three-state Potts model).

For the conformal dimensions of the vertex operators $V_{\vec{\beta}}(z,\bar{z})$, with the above values for $\vec{\beta}$, one obtains

$$\begin{aligned}\Delta_{\vec{\beta}_{(\ldots)}} &= \Delta_{2\vec{\alpha}_0-\vec{\beta}_{(\ldots)}} \equiv \Delta_{(n'_1,n_1)(n'_2,n_2)\cdots(n'_D,n_D)}\\ &= \sum_{a=1}^{D}\big(u^{(a)}_{n'_a,n_a}\big)^2(\vec{\omega}_a)^2 + 2\sum_{a<b}^{D} u^{(a)}_{n'_a,n_a}u^{(b)}_{n'_b,n_b}\,\vec{\omega}_a\vec{\omega}_b\\ &\quad - \sum_{a=1}^{D}\sum_{b=1}^{D}\big(\alpha_a^+ + \alpha_a^-\big)u^{(a)}_{n'_a,n_a}\vec{\omega}_a\vec{\omega}_b,\end{aligned} \tag{2.31}$$

where we have defined

$$u^{(a)}_{n'_a,n_a} = \frac{1-n_a}{2}\,\alpha_a^+ + \frac{1-n'_a}{2}\,\alpha_a^-. \tag{2.32}$$

The procedure that we followed above to define $\beta_{(n'_1n_1)(n'_2n_2)\cdots(n'_Dn_D)}$ in Eqs. (2.29), (2.30), is perfectly standard. Initially it has been used in Refs. [8–10].

Note that if the chiral algebra consisting of D chiral operators commuting with the screenings is found, Eq. (2.31) together with Eqs. (2.29), (2.30) would become the Kac formula for the degenerate representations of the Coulomb gas theory at hand. But in our approach it should rather be interpreted as a *necessary condition* for the existence of degenerate representations, for any given configuration of the screenings. The existence of the extended chiral algebra is not yet guaranteed.

It should be noticed that the degeneracy of the module of $V_{\vec{\beta}}$ in all D screening directions, with $\vec{\beta}$ taking its values in the discrete set (2.29), must be required in order for the operator algebra of the operators $\{V_{\vec{\beta}}(z,\bar{z})\}$ to be well defined. We shall consider the properties of this vertex-operator algebra shortly.

Another remark is on a detail which was actually implicit in the above arguments: for a theory with an extended chiral algebra to be acceptable we want it to contain one free parameter. Put differently, we want the theory to exist for general values of α_0, or, alternatively, of the central charge c. In the class of CFTs in which we are interested, c must remain a free parameter until we eventually impose the much stronger constraint of unitarity, in which case it will be forced to take its values in a discrete (but still denumerably infinite) set. In general we expect the structure of a CFT to be very rigid, in the sense that once we impose additional constraints that limit the generality of the theory, we will immediately either generate inconsistencies or, in the best case, fix c to take a finite number of values. We can illustrate this remark through the example of the two-flavoured loop model of Jacobsen and Kondev [3]. In the special cases where the fugacities of either loop flavour take the value 2 (resp. 1), the model is known to be equivalent to the $\mathrm{su}(4)_{k=1}$ Wess–Zumino–Witten model [11] (resp. to the equally-weighted six-vertex model), both of which are bona fide CFTs with fixed $c = 3$ (resp. $c = 1$). But it is not because of these two special cases that we shall accept the general model (i.e., with generic values of the loop fugacities) as a theory with an extended chiral algebra.

Returning to the general study, one could define the central charge from the two-point function of $T(z)$

$$\big\langle T(z)T(z')\big\rangle = \frac{c/2}{(z-z')^4}, \tag{2.33}$$

where $T(z)$ is given by (2.2), and the correlation functions of the fields $\{\varphi_a(z,\bar{z})\}$ are normalised as in Eq. (2.3). From (2.2), (2.3) and (2.33) one then finds

$$c = D - 24(\alpha_0)^2. \tag{2.34}$$

The Coulomb gas theory itself is defined for general values of α_0. Naturally, then, we are looking for conditions on the theory under which it would contain an extended chiral algebra for general values of α_0, or, equivalently, of the central charge.

2.3. Closure of the operator algebra of the operators $\{V_{\vec{\beta}_{(\cdots)}}\}$

Having defined the "would be degenerate" operators $\{V_{\vec{\beta}_{(\cdots)}}\} \equiv \{V_{(n'_1,n_1)(n'_2,n_2)\cdots(n'_D,n_D)}\}$, with $\{\vec{\beta}_{(\cdots)}\}$ defined by Eqs. (2.29), (2.30), we shall impose next the condition that their operator algebra closes. In a Coulomb gas theory with background charge and screenings the OPE for a product of two vertex operators $V_{\vec{\beta}_1}(z,\bar{z})$ and $V_{\vec{\beta}_2}(z',\bar{z}')$ produces by conservation of electric charge the vertex operator $V_{\vec{\beta}_1+\vec{\beta}_2}(z',\bar{z}')$, but also the vertex operators where the charge $\vec{\beta}_1+\vec{\beta}_2$ has been shifted by any number of screening vectors. Thus, schematically,

$$V_{\vec{\beta}_1}(z,\bar{z})V_{\vec{\beta}_2}(z',\bar{z}') \sim V_{\vec{\beta}_1+\vec{\beta}_2}(z',\bar{z}') + \text{descendents} + V_{\vec{\beta}_1+\vec{\beta}_2+\vec{\alpha}_1^+}(z',\bar{z}') + \text{descendents} + \cdots. \tag{2.35}$$

In general the OPE of $V_{\vec{\beta}_1}V_{\vec{\beta}_2}$ will produce the entire set of operators

$$\{V_{\vec{\beta}_1+\vec{\beta}_2+\sum_a (k_a\vec{\alpha}_a^+ + l_a\vec{\alpha}_a^-)}\} \tag{2.36}$$

as well as their descendents. If $\vec{\beta}_1$ and $\vec{\beta}_2$ belong to the Kac table, i.e., to the sets $\{\vec{\beta}_{(n'_1,n_1)(n'_2,n_2)\cdots(n'_D,n_D)}\}$, $\{\tilde{\vec{\beta}}_{(n'_1,n_1)(n'_2,n_2)\cdots(n'_D,n_D)}\}$ given by Eqs. (2.29), (2.30), then in order that the operators (2.36) also belong to the Kac table, the screening vectors $\{\vec{\alpha}_a^+\}$ must decompose as a integer linear combination of the basis vectors spanning (2.29), (2.30), i.e., the set $\{\alpha_a^+ \frac{\vec{\omega}_a}{2}\}$. It is sufficient to ensure this decomposition for the "+" screenings, since the corresponding decomposition of the "−" screenings over the basis $\{\alpha_a^- \frac{\vec{\omega}_a}{2}\}$ will then follow automatically. Writing

$$\vec{\alpha}_a^+ = \sum_b A_{ab}\alpha_b^+ \frac{\vec{\omega}_b}{2}, \tag{2.37}$$

the integer coefficients A_{ab} can be computed by projecting onto the normalised screening vectors, cf. Eq. (2.24):

$$\vec{\alpha}_a^+\vec{e}_c = \sum_b A_{ab}\alpha_b^+ \frac{1}{2}\vec{\omega}_b\vec{e}_c = \sum_b A_{ab}\alpha_b^+ \frac{1}{2}\delta_{bc} = \frac{1}{2}A_{ac}\alpha_c^+, \tag{2.38}$$

$$A_{ac} = \frac{2(\vec{\alpha}_a^+,\vec{e}_c)}{\alpha_c^+} = \frac{2\vec{\alpha}_a^+\vec{\alpha}_c^+}{(\vec{\alpha}_c^+)^2}. \tag{2.39}$$

We observe that when expressing the matrix of coefficients A_{ab} in terms of screening vectors, one recovers the same form as the Cartan matrix in the theory of classical Lie algebras, expressed in terms of the simple root vectors. The classification condition is the same: its components have to be integers.

It is well-known that this condition implies that the screening vectors $\{\vec{\alpha}_a^+\}$ must belong to the root lattice of one of the classical Lie algebras. This criterion appears here as a *necessary condition* for the closure of the operator algebra of the vertex operators $V_{(n'_1,n_1)(n'_2,n_2)\cdots(n'_D,n_D)}(z,\bar{z})$. Clearly, since we are interested in the most general Coulomb gas theory, and not just a closed sub-theory, we can further require the $\{\vec{\alpha}_a^+\}$ to be a set of *basis vectors* of the root lattice. However, there is for the moment no need that we should

constrain further to the *simple roots* of a classical Lie algebra, in which case we would immediately limit ourselves to the standard classifi cation of W-type extended CFTs.

Several remarks are in order.

In the Coulomb gas theory there are additional constraints on the lengths of the vectors $\{\vec{\alpha}_a^+\}$, in order to ensure the marginality of the screening operators. These constraints are expressed by Eq. (2.9).

For the case of simply laced algebras, all the lengths of $\{\vec{\alpha}_a^+\}$ have to be equal. In this case, to maintain the marginality constraint (2.9) upon varying α_0, all the angles $\{\Theta_a\}$ have to be equal.

For non-simply laced algebras, e.g., B_n, the ratio of the lengths of long and short roots is $\sqrt{2}$. This is incompatible with Eq. (2.9), at least for small values of α_0. We conclude that for Coulomb gas models defi ned in terms of bosonic fi elds only, the screening vectors $\{\vec{\alpha}_a^+\}$ have to correspond to the basic vectors of the root lattice of simply laced algebras only: A_n, D_n, E_6, E_7 or E_8. It is in fact well-known that the free-fi eld representation for W-theories based on non-simply laced algebras uses fermionic fi elds, in addition to the bosonic ones [10].

We emphasise once again that we are analysing the generic case of a Coulomb gas model of D bosonic fi elds $\{\varphi_a(z,\bar{z})\}$, with a non-degenerate set of D screening vectors $\{\vec{\alpha}_a^+\}$ which couple to *all* the fi elds $\{\varphi_a(z,\bar{z})\}$. On the contrary, if the set of the screening operators is reduced and some of the fi elds $\{\varphi_a(z,\bar{z})\}$ are decoupled from screenings, our arguments do not apply. For instance one could imagine having an extra bosonic fi eld which bosonises a couple of fermionic fi elds. Evidently, there is a wide range of possibilities of constructing non-generic Coulomb gas models.

As has been concluded above, in our Coulomb gas models all $\{\Theta_a\}$ have to be equal, i.e., the vector $\vec{\alpha}_0$ has to be "equidistant" from all the screening vectors, cf. Fig. 1. All the lengths of the screening vectors $\{\vec{\alpha}_a^+\}$ have to be identical, as have those of $\{\vec{\alpha}_a^-\}$. We shall henceforth denote them as α_+ and α_-. Eqs. (2.9)–(2.11) now take the simpler form

$$\alpha_\pm = \alpha_0 \cos\Theta \pm \sqrt{\alpha_0^2 \cos^2\Theta + 1}, \tag{2.40}$$

$$\alpha_+ + \alpha_- = 2\alpha_0 \cos\Theta, \tag{2.41}$$

$$\alpha_+\alpha_- = -1. \tag{2.42}$$

Similarly, Eq. (2.28) for the background charge vector becomes

$$2\vec{\alpha}_0 = \sum_a \bigl(\alpha_a^+ + \alpha_a^-\bigr)\vec{\omega}_a = (\alpha_+ + \alpha_-)\sum_a \vec{\omega}_a, \tag{2.43}$$

and Eq. (2.44) now reads

$$\vec{\alpha}_a^\pm = \sum_b A_{ab}\alpha_b^\pm \frac{\vec{\omega}_b}{2} = \frac{\alpha_\pm}{2}\sum_b A_{ab}\vec{\omega}_b. \tag{2.44}$$

From this latter equation one fi nds

$$\vec{\omega}_a = \sum_b A_{ab}^{-1}\frac{2\vec{\alpha}_b^\pm}{\alpha_\pm} = \sum_b 2A_{ab}^{-1}\vec{e}_b, \tag{2.45}$$

where A^{-1}_{ab} is a matrix inverse of A_{ab}. Substituting (2.45) into (2.43) one obtains

$$2\vec{\alpha}_0 = (\alpha_+ + \alpha_-) \sum_a \sum_b 2A^{-1}_{ab}\vec{e}_b \equiv (\alpha_+ + \alpha_-) \sum_b m_b \vec{e}_b, \tag{2.46}$$

with the notation $m_b = \sum_a 2A^{-1}_{ab}$. We finally arrive at the following relation between the background charge and the screenings:

$$2\vec{\alpha}_0 = \sum_a m_a \big(\vec{\alpha}^+_a + \vec{\alpha}^-_a\big). \tag{2.47}$$

Multiplying by $\vec{\alpha}_0$ we also deduce that

$$\frac{1}{\cos^2\theta} = \sum_a m_a. \tag{2.48}$$

Another remark can be made concerning the degeneracy assumption of the module of $V_{\vec{\beta}}(z,\bar{z})$, which has been used throughout the analysis. Its justification is that an extended chiral algebra must necessarily possess degenerate representations realised by the vertex operators, which are primaries as has been argued in the beginning of this section.

That the operator algebra of $\{V_{\vec{\beta}}\}$, corresponding to such degenerate representations, should close can be shown by using, for instance, the differential equations for correlation functions. These equations could be derived if the singular states were realised explicitly by the chiral algebra operators. In the analysis we have just used these assumptions.

One particular feature in the construction of "would be singular states" by mappings produced by the integrated screenings is that we have demanded the degeneracy in all D directions, corresponding to the D vectors $\{\vec{\alpha}^+_a\}$. This condition is necessary to ensure that the OPE of two operators $V_{\vec{\beta}_1}(z,\bar{z})$, $V_{\vec{\beta}_2}(z',\bar{z}')$ produces only a finite number of primary operators out of an, in principle, infinite set of operators:

$$\{V_{\vec{\beta}_1+\vec{\beta}_2+\sum_a (k_a\vec{\alpha}^+_a + l_a\vec{\alpha}^-_a)}\}, \tag{2.49}$$

cf. the arguments given above in connection with Eqs. (2.35), (2.36).

The support for this way of implementing the degeneracy assumption can be found, on one hand, in the fact that a finite OPE (in a sense of a finite number of primaries appearing in the OPE) is also a consequence of the above-mentioned differential equations. On the other hand, a subset of the primaries produced by the OPE (2.35) by adding an appropriate amount of screenings will consist of vertex operators $V_{\vec{\beta}'}(z,\bar{z})$ which are preimages of singular states, in the sense of the mapping (2.14). This subset, which constitutes the unphysical part of the Kac table, have to decouple from the rest in correlation functions as well as in the operator algebra of physical operators. In this way the infinite set of operators appearing in (2.49) will be restricted to a finite number of physical operators. The fact that we demand degeneracy in all D directions is thus tantamount to bordering the physical domain of the Kac table in all these directions.

It is interesting to apply the above result to the two-flavoured loop model introduced by Jacobsen and Kondev [3]. Its configurations are those of two colours of closed loops ($N_{\rm b}$ black loops and $N_{\rm g}$ grey ones) placed on the edges of a square lattice, in such a way

that every vertex of the lattice touches exactly one loop of either colour. Introducing loop fugacities $n_{\rm b}$ and $n_{\rm g}$ for the loop flavours, the partition function of the discrete model reads

$$Z = \sum_{\mathcal{G}} n_{\rm b}^{N_{\rm b}} n_{\rm g}^{N_{\rm g}}, \tag{2.50}$$

where $\mathcal{G}$ are the fully-packed loop configurations just defined. Interestingly, this model is critical for any $0 \leqslant |n_{\rm b}|, |n_{\rm g}| \leqslant 2$.

In the Coulomb gases obtained from loop models [3] the conditions defined above are not satisfied in general. In particular, the components of the matrix A_{ab}, cf. Eq. (2.39), are not integer-valued for generic values of $n_{\rm b}$ and $n_{\rm g}$ belonging to the critical manifold.[2] As a result, defining the "would be degenerate representations" and the Kac table would not make sense: the operator algebra of vertex operators $\{V_{\vec{\beta}_{(\dots)}}\}$ would not close.

In these models one defines physical operators differently, by using physical arguments and constructions that are directly linked to quantities in the discrete model of oriented loops. The operators coupling to the electric part[3] of the Coulomb gas again take the form of vertex operators $\{V_{\vec{\beta}}(z,\bar{z})\}$, with $\vec{\beta}$ belonging to a particular lattice, which, e.g., in case of the two-flavoured loop model will be a three-dimensional body-centered cubic lattice. The operator algebra of the operators $\{V_{\vec{\beta}}(z,\bar{z})\}$ is closed, by construction. But one will, in general, have to admit an infinite OPE for a couple of operators $V_{\vec{\beta}_1} V_{\vec{\beta}_2}$. As it stands, a model with such properties will not have an extended chiral algebra, will not have extended symmetries. It is just a Coulomb gas endowed with conformal invariance, nothing more.

On the other hand, one can check that in the special case $n_{\rm b} = n_{\rm g}$ of the above-mentioned model the matrix A_{ab} appearing in Eq. (2.39) actually *does* have integer valued components:

$$A_{ab} = \frac{2(\vec{\alpha}_a, \vec{\alpha}_b)}{(\vec{\alpha}_b)^2} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}. \tag{2.51}$$

The relation (2.47) between $2\vec{\alpha}_0$ and the screenings is also nicely satisfied, with $m_a = (1,0,1)$. The physical operators of this model (the electric ones) coincide with $\{V_{\vec{\beta}_{(\dots)}}\}$, and they satisfy the necessary condition for the degeneracy of the modules in the way

[2] It should be remarked that in the Coulomb gas models derived from the models of loops [1–3], the normalisation of free fields, of their action, and of the two-point function, is usually different from the one that we are using, defined by the two-point function (2.3). This implies a different definition for scalar products of vectors, like the screening vectors $\{\vec{\alpha}_a^{\pm}\}$, background charge $2\vec{\alpha}_0$, and the electric charges $\{\vec{\beta}\}$ of the physical operators $V_{\vec{\beta}}(z,\bar{z})$. This also results in a different formula for the conformal dimensions of the vertex operators. To compare the formulas in a general setting, independent of the details of a particular model, the matrix of elastic constants in the free field action of loop models has to be first diagonalised and then renormalised appropriately, so that the two-point functions of free fields take the form of Eq. (2.3), or, in any case, take a form which is symmetric with respect to the components of the free fields $\{\varphi_a(z,\bar{z})\}$.

[3] Operators with magnetic Coulomb charge are also physically relevant. They correspond to topological defects, vortices, in the interfacial representation, and they are reminiscent of disorder operators attached to a defect line.

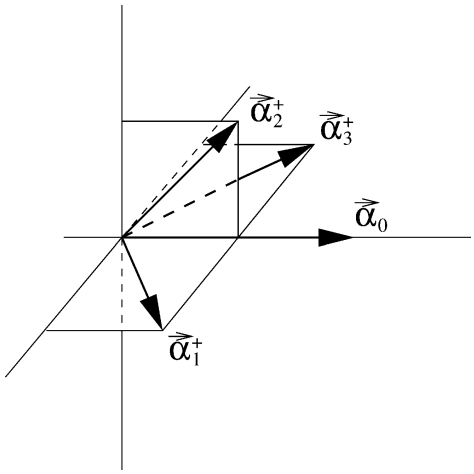


Fig. 6. Screening geometry in the two-flavoured loop model.

it has been described above. (Incidentally, in the loop model language, for $n_b = n_g$ the general model can be rewritten in a way so that all three elasticity constants are equal [4].)

So far, so good. The worrying point is that the matrix A_{ab} in Eq. (2.51) does not correspond to a Cartan matrix of any classical Lie algebra. The closest one would be the algebra A_3, with the Cartan matrix of the form

$$A_{ab} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}. \tag{2.52}$$

The fact that the off-diagonal elements of A_{ab} are now negative is a general property of the Cartan matrix associated with a classical Lie algebra, it being the matrix of scalar products of simple root vectors. On the other hand, the condition for the closure of the algebra of operators $\{V_{\vec{\beta}_{(\ldots)}}\}$ requires, in its general form presented above, the integer-valuedness only.

To make more clear the difference between the two sets of vectors $\{\vec{\alpha}_a^+\}$, corresponding to the matrices (2.51) and (2.52), we present them in Figs. 6 and 7. Evidently, the vectors in Fig. 6, cf. Eq. (2.51), belong to the same lattice as the vectors in Fig. 7, but they are not *simple* roots of A_3. The vector $\vec{\alpha}_2^+$ has been switched from one side to another, leading to a sign change in the off-diagonal elements of the matrix A_{ab}.

It might be possible that a more refined analysis could be given to the selection of screenings in the Coulomb gas models, taking into account the properties of representations: detailed properties of the operator algebra and of the correlation functions of physical operators. The purpose of the analysis would be to decide on the acceptance of the theories with the non-classical matrices A_{ab}, as in Eq. (2.51), which appear in the loop models.

In the absence of such a more detailed analysis, but also to test one more detail of our approach, we have directly calculated the chiral algebra operators, the W's, for the theories with the matrices A_{ab} in (2.51) and in (2.52).

This last part of our approach is in some sense akin to the inverse scattering problem: with the data of scattering one has to reconstruct the Hamiltonian of the problem. Here we

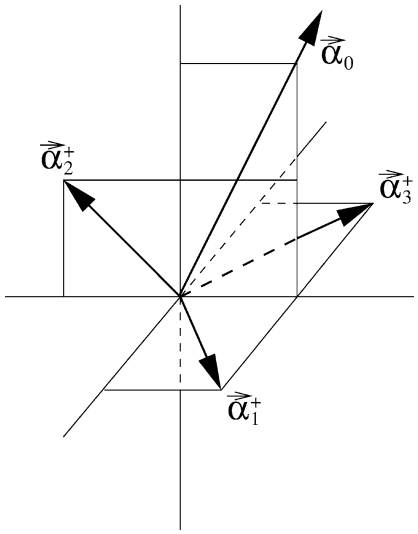


Fig. 7. Screening geometry in the WA_3 model.

have constructed the Coulomb gases starting from the geometry of the screening operators. But finally, having fixed them, it is then perfectly possible to calculate the extended chiral algebra operators. In particular, in this way it is possible to give a definitive answer whether an extended chiral algebra exists in the loop model with the matrix of screenings given by Eq. (2.51).

This last part of the analysis is described and applied to a variety of particular Coulomb gas problems in the section which follows. For the loop model with screening matrix (2.51) the answer will be negative. The model does not have an extended chiral algebra, and its symmetry is only a conformal one.

In the last section, which will be devoted to discussion and conclusions, we shall discuss the possibilities to define, in the loop models setting, models which might lead to the Coulomb gases related to the classical Lie algebras, with screenings realised by simple roots. This appears to be the only possibility to provide them with extended symmetries.

3. Applications

We have seen in the preceding section that the general consistency requirements means that the screening vectors must span the root lattice of a classical Lie algebra. In particular, we would like to consider situations where this is the case, without actually taking the screening vectors to be the *simple* roots of the concerned algebra. As we have seen above, one such situation is of interest to the two-flavoured loop model of Jacobsen and Kondev.

We shall consider various such examples with $D = 2$ and $D = 3$. It should be noticed that the fully-packed loop model on the honeycomb lattice (which has $D = 2$) possesses a unique screening vector, which is proportional to the background charge [3]. We have argued that such a situation is non-generic, and leads to the decoupling of one of the scalar fields from the other *and* from the background charge. Thus, from the point of view of

CFT this theory is trivial. Notwithstanding this "triviality", some of the critical exponents related to geometrical properties of the loops do take non-trivial values [3], different from those of the standard $D = 1$ loop model describing the dense phase of the O(n) model [1,2]. This is possible because these exponents do not belong to the physical part of the Kac table for minimal models. We also point out that despite of the triviality of its CFT, the model actually has an interesting integrable structure, due to its underlying $sl_q(2)$ quantum group symmetry [12].

Discarding such trivial possibilities, for both $D = 2$ and $D = 3$ there are two different choices of D screening vectors of equal length, spanning the root lattice of A_D. After defining the method of computations we turn to the detailed application to those two cases. One of the cases, of course, is the standard WA_D geometry, and it serves as a check of our computations that in this case we reproduce various known results.

3.1. Definitions

As already explained, we consider the case of a D-dimensional Coulomb gas with a fixed background charge $\vec{\alpha}_0$ and D screening charges. In addition to T, we must construct $D - 1$ extra chiral operators with integer dimensions. These operators are made as linear combinations of products of derivatives of free fields. We shall choose here the simplest construction, namely to search for chiral operators $W_3, \ldots, W_{D+1}$, where the subscript in W_i indicates its dimension. It is of course possible that the fields W_i should be found on higher levels, but for reason of simplicity we find this unlikely. After building the most general operators on the respective levels, we will impose on them that they commute with the integrated screening operators, as explained in Section 2.2, and that they be primary operators. The primarity condition is requested in order to ensure that $\{T(z), W_3(z), \ldots, W_{D+1}(z)\}$ forms a closed algebra.

The requirement that a field W commutes with the currents of Eq. (2.6) can be written

$$\big[W(z), Q\big] = \left[W(z), \oint_{C_0} \mathrm{d}z'\, V(z') \right] = \oint_{C_z} \mathrm{d}z'\, V(z') W(z) = 0, \tag{3.1}$$

where we have deformed the integration contour as indicated in Fig. 8. This implies that the residue must be zero:

$$\mathrm{Res}\big(V(z')W(z); z\big) = 0. \tag{3.2}$$

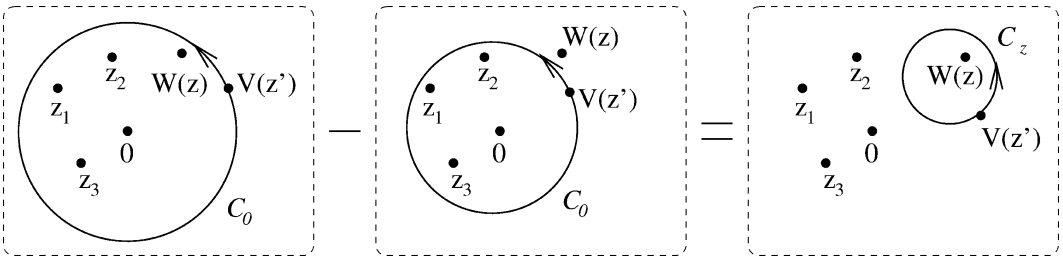


Fig. 8. Deformation of the integration contour in (3.1).

Second, W must be a primary operator of dimension Δ_W:

$$T(z)W(z') = \frac{\Delta_W}{(z-z')^2}W(z') + \frac{1}{(z-z')}\partial W(z') + \cdots. \tag{3.3}$$

We shall see below that for a specified choice of the screening charges, these two constraints, when properly expressed, suffice to uniquely determine the W operator, up to a global normalisation. To fix the latter, we impose a third constraint:

$$W(z)W(z') = \frac{c/\Delta_w}{(z-z')^{2\Delta_W}} + \cdots. \tag{3.4}$$

Keeping only the holomorphic part in (2.2), (2.3) and (2.4), we have

$$T(z) = -\frac{1}{4}\sum_{i=1}^{D}(\partial\varphi_i)^2 + i\vec{\alpha}_0\cdot\partial^2\vec{\varphi}, \tag{3.5}$$

$$\left\langle\varphi_i(z)\varphi_j(z')\right\rangle = -2\delta_{ij}\log(z-z') \tag{3.6}$$

with the corresponding central charge:

$$c = D - 24\sum_{i=1}^{D}\left(\alpha_0^i\right)^2 \tag{3.7}$$

and the screening operators:

$$V_k(z) = \exp\left(i\sum_i \alpha_k^i\varphi_i(z)\right) = \exp\left(\frac{1}{2}\sum_i g_k^i\varphi_i(z)\right) \tag{3.8}$$

with $g_k^i = 2i\alpha_k^i$. With this last definition, the Wick contraction with a derivative of a field reads simply

$$V_k(z')\partial^i\varphi_j(z) = \frac{a_i g_k^j}{(z'-z)^i}V(z') \tag{3.9}$$

with $a_1 = a_2 = 1$, $a_3 = 2$ and $a_4 = 6$. Forming suitable linear combinations, this formula will allow us to compute (3.2) for any operator $W(z)$ built out of the fields $\varphi_i(z)$ and their derivatives. However, we still need to develop the screening operator around z:

$$\begin{aligned} V_k(z') &= \exp\left(\sum_i \frac{g_k^i}{2}\varphi_i(z')\right) \\ &= \exp\left(\sum_i \frac{g_k^i}{2}\left(\varphi_i(z) + (z'-z)\varphi_i'(z) + \frac{(z'-z)^2}{2}\varphi_i''(z)\right.\right. \\ &\qquad \left.\left. + \frac{(z'-z)^3}{6}\varphi_i'''(z) + \cdots\right)\right) \end{aligned}$$

$$= V_k(z)\Bigg(1 + (z'-z)\sum_i \frac{g_k^i}{2}\varphi_i'(z) + (z'-z)^2\Bigg[\sum_i \frac{g_k^i}{4}\varphi_i''(z) + \sum_{i,j}\frac{g_k^i g_k^j}{8}\varphi_i'(z)\varphi_j'(z)\Bigg] + (z'-z)^3\Bigg[\sum_i \frac{g_k^i}{12}\varphi_i'''(z) + \sum_{i,j}\frac{g_k^i g_k^j}{8}\varphi_i'(z)\varphi_j''(z) + \sum_{i,j,k}\frac{g_k^i g_k^j g_k^l}{48}\varphi_i'(z)\varphi_j'(z)\varphi_l'(z)\Bigg] + \cdots\Bigg), \tag{3.10}$$

where we have used the notation $\varphi_i'(z') = \partial\varphi_i(z')$, $\varphi_i''(z') = \partial^2\varphi_i(z')$ and $\varphi_i'''(z') = \partial^3\varphi_i(z')$.

We consider an operator W_N of dimension N, built as a linear combination of the derivatives of the D scalar fi elds. Let us defi ne by $\mathcal{N}_D(N)$ the number of terms in the linear combination which forms the most general operator of dimension N. We then impose the constraint that this operator commutes with the D screening charges. As shown above, this corresponds to canceling the residue in z of the product $V(z')W_N(z)$, which for a general operator W_N amounts to canceling an operator of dimension $N-1$. We then expect $D\times\mathcal{N}_D(N-1)$ constraints. Next we have to impose the primarity. In the product $T(z)W_N(z')$, we have to cancel all the powers in $(z-z')^{-i}$ for $i = N+2, N+1, \ldots, 3$. Since the power $(z-z')^{-i}$ comes with an operator of dimension $N+2-i$, we will get an additional number of constraints equal to $\mathcal{N}_D(0)+\mathcal{N}_D(1)+\cdots+\mathcal{N}_D(N-1)$. Thus for a W_n operator with $\mathcal{N}_D(N)$ parameters, we have $(D+1)\mathcal{N}_D(N-1)+\mathcal{N}_D(N-2)+\cdots+\mathcal{N}_D(0)$ constraints. Of course, this is just the maximal number of constraints that we can expect, since in general not all of the corresponding linear equations will be independent.

3.2. Two scalar fields

The most general operator of dimension three built from two scalar fi elds is:

$$W_3 = \sum_i a_i\partial^3\varphi_i + \sum_i b_i\partial\varphi_i\partial^2\varphi_i + \sum_i c_i(\partial\varphi_i)^3 + \sum_{i\neq j} d_{ij}\big(\partial^2\varphi_i\big)\partial\varphi_j + \sum_{i\neq j} e_{ij}(\partial\varphi_i)^2\partial\varphi_j. \tag{3.11}$$

We shall choose the coordinate system so that the background electric charge is directed along the 1-direction. The stress-energy tensor then reads

$$T(z) = -\frac{1}{4}(\partial\varphi_1)^2 - \frac{1}{4}(\partial\varphi_2)^2 + i\alpha_0\partial^2\varphi_1. \tag{3.12}$$

For this case, we have $\mathcal{N}_2(0) = 1$, $\mathcal{N}_2(1) = 2$, and $\mathcal{N}_2(2) = 5$. Thus we will have 8 constraints from the primarity of W_3 and 10 constraints from the commutation of the W_3 operator with the two screening operators.

By applying first the constraints from the primarity, we obtain seven equations on the 10 constants entering in (3.11):

$$
\begin{aligned}
&a_1 = -4\alpha_0^2 c_1, \qquad b_1 = -6i\alpha_0 c_1, \qquad b_2 = 6i\alpha_0\big(1-8\alpha_0^2\big)c_1,\\
&e_{21} = -3\big(1-8\alpha_0^2\big)c_1, \qquad e_{12} = \frac{i}{4\alpha_0}(d_{12}+d_{21}), \qquad a_2 = -\frac{2}{3}i\alpha_0 d_{12},\\
&12i\alpha_0 c_2 = \big(1-24\alpha_0^2\big)d_{12} + \big(1+8\alpha_0^2\big)d_{21}.
\end{aligned}
\tag{3.13}
$$

3.2.1. Classical WA_2 *geometry*

Turning now to the commutation of W_3 with the screening operators, we begin by considering the geometry of the classical WA_2 theory:

$$
\vec{\alpha}_1^{+} = \alpha_+\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right), \qquad \vec{\alpha}_2^{+} = \alpha_+\left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right),
\tag{3.14}
$$

where α^+ is related to α_0 by the relation (2.7), i.e., $\alpha_0 = \frac{1}{2}(\alpha^+ - \frac{1}{\alpha^+})$.

The two screenings are related by a reflectional symmetry, and it is thus no surprise that considering just $\vec{\alpha}_1^{+}$ leads to a single-parameter solution

$$
\begin{aligned}
&a_1 = 0, \qquad a_2 = 24\alpha_0^2\tilde{c}, \qquad b_1 = 0, \qquad b_2 = 0, \qquad c_1 = 0,\\
&c_2 = 4\tilde{c}, \qquad q d_{12} = 12i\alpha_0\tilde{c}, \qquad d_{21} = 36i\alpha_0\tilde{c},\\
&e_{12} = -12\tilde{c}, \qquad e_{21} = 0.
\end{aligned}
\tag{3.15}
$$

If we also impose the normalisation condition (3.4), then we end up with the following expression of the parameters in terms of the central charge ($c = 2 - 24\alpha_0^2$):

$$
a_2 = \frac{i(2-c)}{12\sqrt{5c+22}}, \qquad c_2 = \frac{i}{3\sqrt{5c+22}}, \qquad d_{12} = \pm\sqrt{\frac{(2-c)}{24(5c+22)}},
\tag{3.16}
$$

$$
d_{21} = 3d_{12} = \pm\sqrt{\frac{3(2-c)}{8(5c+22)}}, \qquad e_{12} = -\frac{i}{\sqrt{5c+22}}.
\tag{3.17}
$$

Thus, our final result reads

$$
T = -\frac{1}{4}(\partial\varphi_1)^2 - \frac{1}{4}(\partial\varphi_2)^2 + i\sqrt{\frac{2-c}{24}}\,\partial^2\varphi_1,
\tag{3.18}
$$

$$
\begin{aligned}
W_3 = \frac{1}{\sqrt{5c+22}}\Bigg(&\frac{i(2-c)}{12}\partial^3\varphi_2 + \frac{i}{3}(\partial\varphi_2)^3 \pm \sqrt{\frac{2-c}{24}}\,\big(\partial^2\varphi_1\big)\partial\varphi_2\\
&\pm\sqrt{\frac{3(2-c)}{8}}\,\big(\partial^2\varphi_2\big)\partial\varphi_1 - i(\partial\varphi_1)^2\partial\varphi_2\Bigg).
\end{aligned}
\tag{3.19}
$$

Up to the normalization constant, this coincides with the solution of Fateev and Zamolodchikov [9].

3.2.2. Alternative geometry

Consider next the alternative geometry of the screening charges

$$\vec{\alpha}_1^+ = \alpha_+\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right), \qquad \vec{\alpha}_2^+ = \alpha_+\left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right). \tag{3.20}$$

These vectors still span a triangular lattice, i.e., the root lattice of the Lie algebra A_2. They also correspond to a "Cartan matrix" where the signs of the off-diagonal elements have been changed:

$$\frac{2(\vec{\alpha}_i, \vec{\alpha}_j)}{(\vec{\alpha}_j)^2} \equiv A_{ij} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}. \tag{3.21}$$

In this case we find only the trivial (null) solution for W.

3.3. Three scalar fields

In this section we have found it convenient to impose the following parametrisation for the screening operators

$$g_1^i = 2ia(1, 0, -1), \qquad g_2^i = 2ia(-1, 1, 0), \qquad g_3^i = 2ia(1, 0, 1) \tag{3.22}$$

and for the stress-energy tensor

$$\alpha_0^1 = X/2, \qquad \alpha_0^2 = X, \qquad \alpha_0^3 = 0. \tag{3.23}$$

Here the variable $X = (2a - 1/a)$ is related to the central charge by $c = 3 - 30X^2$.

We now have to consider the construction of two operators of dimensions 3 and 4. The most general such operators built from three scalar fields read

$$\begin{aligned} W_3 = {} & \sum_i a_i \partial^3 \varphi_i + \sum_i b_i \partial\varphi_i \partial^2 \varphi_i + \sum_i c_i (\partial\varphi_i)^3 + \sum_{i\neq j} d_{ij} \left(\partial^2 \varphi_i\right) \partial\varphi_j \\ & + \sum_{i\neq j} e_{ij} (\partial\varphi_i)^2 \partial\varphi_j + f \partial\varphi_1 \partial\varphi_2 \partial\varphi_3 \end{aligned} \tag{3.24}$$

and

$$\begin{aligned} W_4 = {} & \sum_i a_i \partial^4 \varphi_i + \sum_i b_i \partial\varphi_i \partial^3 \varphi_i + \sum_{i\neq j} c_{ij} \partial^3 \varphi_i \partial + \varphi_j \sum_i d_i \left(\partial^2 \varphi_i\right)^2 \\ & + \sum_{i<j} e_{ij} \partial^2 \varphi_i \partial^2 \varphi_j + \sum_i f_i \left(\partial^2 \varphi_i\right) (\partial\varphi_i)^2 + \sum_{i\neq j} g_{ij} \left(\partial^2 \varphi_i\right) (\partial\varphi_i)(\partial\varphi_j) \\ & + \sum_{i\neq j} h_{ij} \left(\partial^2 \varphi_i\right) (\partial\varphi_j)^2 + \sum_{i\neq j,k;\, j<k} i_{ijk} \left(\partial^2 \varphi_i\right) (\partial\varphi_j)(\partial\varphi_k) + \sum_i j_i (\partial\varphi_i)^4 \\ & + \sum_{i\neq j} k_{ij} (\partial\varphi_i)^3 (\partial\varphi_j) + \sum_{i<j} l_{ij} (\partial\varphi_i)^2 (\partial\varphi_j)^2 \\ & + \sum_{i\neq j,k;\, j<k} m_{ijk} (\partial\varphi_i)^2 (\partial\varphi_j)(\partial\varphi_k). \end{aligned} \tag{3.25}$$

Note that there is some overlap among the symbols used to designate the constants entering in the definitions of W_3 and W_4. These constants are of course independent in the two cases,

but in order to avoid complicating the notation we do not distinguish them by an extra index. This convention should lead to no confusion since we shall consider the operators W_3 and W_4 separately in the following.

3.3.1. Classical WA_3 *geometry*

In this case we have $\mathcal{N}_3(0)=1$, $\mathcal{N}_3(1)=3$, $\mathcal{N}_3(2)=9$, $\mathcal{N}_3(3)=22$ and $\mathcal{N}_3(4)=51$. Thus, Eq. (3.24) for W_3 contains 22 free parameters. Imposing the commutation of W_3 with the screening charges produces 3×9 constraints on these parameters.

Solving the corresponding equations gives the following relations, leaving just two free parameters, b_1 and d_{13}:

$$a_1=-iXb_1,\qquad a_2=-2iXb_1,\qquad a_3=-id_{13}X, \tag{3.26}$$

$$b_3=b_2=b_1,\qquad d_{32}=d_{31}=d_{13},\qquad f=i\frac{d13}{X}. \tag{3.27}$$

Next we impose the primarity of W_3, yielding $\mathcal{N}_3(0)+\mathcal{N}_3(1)+\mathcal{N}_3(2)=13$ additional constraints. It turns out that the only additional condition which result from these constraints is $b_1=0$. Thus we end with the following result:

$$\begin{aligned}
T(z)&=-\frac{1}{4}\sum_{i=1}^{3}(\partial\varphi_1)^2+i\sqrt{\frac{3-c}{120}}\left(\partial^2\varphi_1+2\partial^2\varphi_2\right),\\
W_3(z)&\propto\left(\frac{3-c}{30}\partial^3\varphi_3+i\sqrt{\frac{3-c}{30}}\left(\left(\partial^2\varphi_1\right)\partial\varphi_3+\left(\partial^2\varphi_3\right)\partial\varphi_1+\left(\partial^2\varphi_3\right)\partial\varphi_2\right)\right.\\
&\qquad\left.-\partial\varphi_1\partial\varphi_2\partial\varphi_3\right).
\end{aligned} \tag{3.28}$$

To normalise this properly, we demand that $W(z)W(z')=\frac{c/3}{(z-z')^6}+\cdots$. It follows that the above solution should be divided by $-\frac{4}{5}(7+c)$. The singularity at $c=-7$ will reappear in W_4 (see below).

The computation for the W_4 operator goes along the same line as for the W_3 case, but with much more parameters and constraints. In this case, we have $\mathcal{N}_3(4)=51$ parameters in the definition of W_4 (see (3.25)). There are $3\times\mathcal{N}_3(3)=66$ constraints produced by the commutation of the screening operators with W_4, and $\mathcal{N}_3(0)+\mathcal{N}_3(1)+\mathcal{N}_3(2)+\mathcal{N}_3(3)=35$ additional constraints coming from the requirement that W_4 be a primary operator. We thus have 101 constraints on the 51 parameters. Solving these constraints is rather straightforward since they are just linear equations. Our result (up to a multiplicative factor) can be expressed as a function of the central charge c (all the terms not present being zero):

$$a_1=i\frac{\sqrt{3-c}(-26+c+2c^2)}{\sqrt{30}},\qquad a_2=-i\frac{\sqrt{3-c}(c-2)(c+7)}{\sqrt{30}}, \tag{3.29}$$

$$b_1=b_3=-2(c-6)(c+2),\qquad b_2=3(c-2)(c+7), \tag{3.30}$$

$$c_{12}=-(c-3)(5c+22),\qquad d_1=\frac{1}{5}\left(-18-63c-2c^2\right), \tag{3.31}$$

$$d_2 = \frac{1}{10}(c+7)(19c-102), \qquad d_3 = -\frac{9}{2}(c-2)(c+7), \tag{3.32}$$

$$e_{12} = -\frac{2}{5}(c-3)(9c-2), \qquad f_1 = -i8\sqrt{\frac{6}{5}}\sqrt{3-c}\,(c+7), \tag{3.33}$$

$$f_2 = 2f_1, \qquad g_{12} = g_{31} = g_{32} = i\sqrt{30}\sqrt{3-c}\,(5c+22), \tag{3.34}$$

$$h_{12} = i\sqrt{\frac{6}{5}}\sqrt{3-c}\,(17c+54), \qquad h_{13} = -i8\sqrt{\frac{6}{5}}\sqrt{3-c}\,(c+7) = f_1, \tag{3.35}$$

$$h_{21} = h_{23} = i\sqrt{\frac{6}{5}}\sqrt{3-c}\,(9c-2), \qquad j_1 = j_2 = j_3 = 12(c+7), \tag{3.36}$$

$$l_{12} = l_{13} = l_{23} = -3(17c+54) \tag{3.37}$$

with the following solution:

$$\begin{aligned}
W_4 = {} & i\frac{\sqrt{3-c}(-26+c+2c^2)}{\sqrt{30}}\partial^4\varphi_1 + \frac{1}{5}\big(-18-63c-2c^2\big)\big(\partial^2\varphi_1\big)^2 \\
& + (c+7)\Bigg\{-i\frac{\sqrt{3-c}}{\sqrt{30}}\big[(c-2)\partial^4\varphi_2 \\
& \qquad + 48\big(\partial^2\varphi_1(\partial\varphi_1)^2 + 2\partial^2\varphi_2(\partial\varphi_2)^2 + \partial^2\varphi_1(\partial\varphi_3)^2\big)\big] \\
& \qquad + 3(c-2)\Big[\partial\varphi_2\partial^3\varphi_2 - \frac{3}{2}\big(\partial^2\varphi_3\big)^2\Big] + \frac{1}{10}(19c-102)\big(\partial^2\varphi_2\big)^2\Bigg\} \\
& + 12\big((\partial\varphi_1)^4 + (\partial\varphi_2)^4 + (\partial\varphi_3)^4\big) \\
& + (17c+54)\Bigg\{i\sqrt{\frac{6}{5}}\sqrt{3-c}\,\partial^2\varphi_1(\partial\varphi_2)^2 \\
& \qquad - 3\big((\partial\varphi_1)^2(\partial\varphi_2)^2 + (\partial\varphi_1)^2(\partial\varphi_3)^2 + (\partial\varphi_2)^2(\partial\varphi_3)^2\big)\Bigg\} \\
& + (9c-2)i\sqrt{3-c}\Bigg\{-i\frac{2}{5}\sqrt{3-c}\,\partial^2\varphi_1\partial^2\varphi_2 \\
& \qquad + \sqrt{\frac{6}{5}}\big(\partial^2\varphi_2(\partial\varphi_1)^2 + \partial^2\varphi_2(\partial\varphi_3)^2\big)\Bigg\} \\
& + i(5c+22)\sqrt{3-c}\,\big\{-i\sqrt{3-c}\,\partial^3\varphi_1\partial\varphi_2 \\
& \qquad + \sqrt{30}\big(\partial^2\varphi_1\partial\varphi_1\partial\varphi_2 + \partial^2\varphi_3\partial\varphi_3\partial\varphi_1 + \partial^2\varphi_3\partial\varphi_3\partial\varphi_2\big)\big\}.
\end{aligned} \tag{3.38}$$

Finally, we impose the standard normalisation $W(z)W(z') = \frac{c/4}{(z-z')^8} + \cdots$, which means that all the above should be divided by the factor

$$192(2+c)(7+c)(22+5c)(114+7c). \tag{3.39}$$

In particular, the W_4 operator becomes singular at a set of special values of the central charge:

$$c = -7, \qquad c = -2, \qquad c = -\frac{22}{5}, \qquad c = -\frac{114}{7}. \tag{3.40}$$

These singularities are exactly those found by Blumenhagen et al. [13]. Kausch and Watts [14] find in addition the singularities $c = 1/2$, $c = -68/27$ and $c = -24$ of which we see no sign.

3.3.2. Alternative geometry

Finally, we consider the geometry of the two-flavoured loop model (cf. Eq. (2.51)), for which we use the following parametrisation for the screening operators

$$g_1^i = 2ia(1, 0, -1), \qquad g_2^i = 2ia(1, 1, 0), \qquad g_3^i = 2ia(1, 0, 1) \tag{3.41}$$

and for the stress-energy tensor

$$\alpha_0^1 = X/2, \qquad \alpha_0^2 = 0, \qquad \alpha_0^3 = 0. \tag{3.42}$$

In this case, very tedious computations show that for W_3 as well as for W_4, all the parameters are zero under the application of the constraints.

4. Discussion

The classification of all possible two-dimensional CFTs has posed a major theoretical challenge ever since the introduction of the series of minimal models [15], and several fundamental questions remain open to this day. Among the approaches exploiting Coulomb gas techniques we should here mention the exact computation of partition functions on the torus [16,17], which has the merit of making direct correspondence between lattice models and modular invariants. In the context of the present paper, the generalisations to coset theories with $c > 1$ [18,19] and field theories based on Lie algebras [20] are particularly interesting. Also, an extension of the standard coset classification of CFTs using Coulomb gases [21] was used in Ref. [22] to obtain an alternative description of a particular W_3 invariant model.

In this paper we have discussed the general construction of CFTs based on Coulomb gases with several bosonic fields. The requirement that the CFT be consistent has dictated us two physical guiding principles: the existence of degenerate representations, and the closure condition on the vertex operator algebra. From these principles we have obtained the classification condition (2.39) stating that the screening vectors must belong to the root lattice of a classical Lie algebra. The screenings are however not required to be *simple* roots, and therefore the matrix A_{ac} in Eq. (2.39) is more general than the Cartan matrix in the theory of W-algebras.

One of the main motivations of our work has been the attempt to identify the CFT underlying two different model of fully-packed loops (FPL) [3]: the single-flavoured FPL model on the honeycomb lattice and the two-flavoured FPL model on the square lattice. Both these loop models are known to be critical for any value of the loop fugacities in the interval $[-2, 2]$. In both cases, the underlying Coulomb gas furnishes exact values of the central charge, the thermal scaling dimension, and the scaling dimensions of an infinite set of topological defects linked to the propagation of a set of strings between two

points ("watermelon dimensions") [1,2]. In the honeycomb case this information is further supported by a Bethe ansatz solution, the integrability of the model being assured by its $sl_q(2)$ quantum group symmetry [12], whereas in the square case no Bethe ansatz solution is known, except at the point $(n_b, n_g) = (2, 2)$ [23].

However, without access to the associated CFT, our knowledge of these models cannot be considered complete. To illustrate this point it is useful to compare with the finite-temperature O(n) model on the honeycomb lattice [1,2]. Its low-temperature critical phase is described by the dense phase of a (non-fully packed) loop model, the loops being defined in terms of the diagrammatic expansion of the associated spin model. This loop model is solvable, both as a Coulomb gas and by Bethe ansatz techniques, and it furnishes information on critical indices analogous to the FPL cases mentioned above. On the other hand, when $n = -2\cos(\pi g)$ with $m = g/(1-g)$ a positive integer ($m \geqslant 3$) the model is known to coincide with the series of minimal models of conventional CFT. For these unitary cases a wealth of further information is available, and many interesting applications become possible. To mention but one important example, the knowledge of exact operator product expansions makes it possible to study perturbatively the coupling of quenched randomness to the local energy density of such models. For the FPL cases one could imagine addressing the problem of a compact polymer in a random environment by similar techniques, but to do so knowing the corresponding CFT becomes indispensable. Another illustration of the complementarity of the two approaches (Coulomb gas versus CFT) is that the watermelon dimensions [1], which have a straightforward interpretation in terms of the loop model, coincide with CFT operators that are *outside* the physical part of the Kac table [2] (see also Ref. [24] for a review). Conversely, local physical operators in the CFT description do not in general seem to have an analogue in the loop approach.

In the present work we have argued that the honeycomb FPL model is in fact trivial from the CFT point of view, since one of the scalar fields decouples from the other, and from the background charge. On the other hand, we have showed that the non-trivial two-flavoured FPL model satisfies the classification condition (2.39). However, we have also found that the extended chiral operators needed to control the $D-1$ remaining degrees of freedom (with $D = 3$) do not exist, at least not on level 3 and 4 of the identity module. We can of course not rule out the eventuality that such operators exist at higher levels, but we find it rather unlikely.

To acquire a consistent CFT description of fully-packed loops, it thus seems to us that the most natural thing to do would be to somehow modify the definition of the loop models in question, so that the corresponding CFTs become the classical W-theories, WA_2 and WA_3, respectively. We recall that in the Coulomb gas formalism, the operator assigning the proper weights to the loops is a periodic function on the ideal state graph. The lattice to which screenings must belong is therefore fixed by standard Fourier analysis [3]. One generally assumes that all of the Fourier modes come with non-zero amplitudes, and the actual screenings are therefore singled out as being those closest to the background charge vector, since this ensures the lowest conformal dimension. For loop models with further adjustable parameters one may however imagine that the amplitude corresponding to the closest vectors can be made to vanish, in which case the screenings will have to be chosen

from the vectors second-closest to the background charge, and so on. This is actually what happens in the O(n) model, where the temperature T acts as an adjustable parameter. By making the obvious ("closest") choice for the screening vector, the one-dimensional Coulomb gas describes the dense (critical) phase, as mentioned above. But by fine-tuning the temperature it is possible to access another dilute (tricritical) phase, corresponding to the choice of the next-closest screening [24,25]. This situation is rather analogous to standard Ginzburg–Landau theory, where fine-tuning may serve to make the φ^4 term vanish, thus giving access to multi-critical behaviour governed by a more general φ^{2n} term. In a certain sense, successive fine-tuning of more and more parameters is tantamount to augmenting the symmetry of the corresponding critical theory. We would expect that the CFTs of the two FPL models under consideration may be turned into the classical WA_2 and WA_3 theories by fine-tuning suitable extra parameters, thus driving the systems to multi-criticality.

For the moment we do not have any concrete proposal for the construction of such tunable parameters. The introduction of temperature-like vacancies in FPL models is known [26] to introduce a flow towards the dense phase of the standard O(n) model [1, 2], so clearly the temperature is not a suitable parameter for endowing the model with a *higher* symmetry. A more promising possibility would be to progress in analogy with the ferromagnetic Ising model, which can be driven to tricriticality by introducing a staggered magnetic field and fine-tuning its strength (in addition to the critical temperature). For the fully-packed loop models, it is possible to impose several staggered fields that act so as to distinguish between the various ideal states [3].

Acknowledgements

We would like to thank D. Bernard, J. Kondev and A. Leclair for stimulating discussions. We are also grateful to the organizers of the programme "Integrable models in Condensed Matter and Non-Equilibrium Physics", May 14 – June 11, 2000, CRM, Université de Montréal. Part of the present work has been done and presented at this meeting.

References

[1] B. Nienhuis, Phys. Rev. Lett. 49 (1982) 1062;
B. Nienhuis, J. Stat. Phys. 34 (1984) 731;
B. Nienhuis, in: C. Domb, J.L. Lebowitz (Eds.), Phase Transitions and Critical Phenomena, Vol. 11, Academic Press, London, 1987.
[2] B. Duplantier, H. Saleur, Nucl. Phys. B 290 (1987) 291.
[3] M.T. Batchelor, J. Suzuki, C.M. Yung, Phys. Rev. Lett. 73 (1994) 2646;
J. Kondev, J. de Gier, B. Nienhuis, J. Phys. A 29 (1996) 6489;
J.L. Jacobsen, J. Kondev, Nucl. Phys. B 532 (1998) 635.
[4] J. Kondev, Phys. Rev. Lett. 78 (1997) 4320.
[5] Vl.S. Dotsenko, V.A. Fateev, Nucl. Phys. B 240 (1984) 312;
Vl.S. Dotsenko, V.A. Fateev, Nucl. Phys. B 251 (1985) 691.

[6] Vl.S. Dotsenko, Adv. Stud. Pure Math. 16 (1988) 123.
[7] G. Felder, Nucl. Phys. B 317 (1989) 215;
G. Felder, Nucl. Phys. B 324 (1989) 548, Erratum.
[8] C. Thorn, Nucl. Phys. B 248 (1984) 551.
[9] V.A. Fateev, A.B. Zamolodchikov, Nucl. Phys. B 280 (1987) 644.
[10] S.L. Lukyanov, V.A. Fateev, Sov. Sci. Rev. A Phys. 15 (1990) 1–117.
[11] J. Kondev, C.L. Henley, Nucl. Phys. B 464 (1996) 540.
[12] N.Yu. Reshetikhin, J. Phys. A 24 (1991) 2387.
[13] R. Blumenhagen, M. Flohr, A. Kliem, W. Nahm, A. Recknagel, R. Varnhagen, Nucl. Phys. B 361 (1991) 255.
[14] H.G. Kausch, G.M.T. Watts, Nucl. Phys. B 354 (1991) 740.
[15] A.A. Belavin, A.M. Polyakov, A.B. Zamolodchikov, Nucl. Phys. B 241 (1984) 333.
[16] P. Di Francesco, H. Saleur, J.-B. Zuber, J. Stat. Phys. 49 (1987) 57.
[17] V. Pasquier, J. Phys. A 20 (1987) L1229.
[18] P. Di Francesco, H. Saleur, J.-B. Zuber, Nucl. Phys. B 300 (1988) 393.
[19] I. Kostov, Nucl. Phys. B 300 (1988) 559.
[20] D. Gepner, Nucl. Phys. B 290 (1987) 10.
[21] M. Caselle, K. Narain, Nucl. Phys. B 323 (1989) 673.
[22] M. Caselle, Int. J. Mod. Phys. A 6 (1991) 5495.
[23] B. Nienhuis, cond-mat/0005274.
[24] B. Duplantier, Physica D 38 (1989) 71.
[25] J. Kondev, private communication.
[26] J.L. Jacobsen, J. Kondev, J. Stat. Phys. 96 (1999) 21.

Chapitre 6

Réseaux aléatoires : méandres et nœuds

Les différents modèles considérés jusqu'à présent, dans ce mémoire, sont tous définis sur des réseaux bidimensionnels dont les sommets et les arêtes forment un dessin régulier. Nous avons vu, notamment lors des chapitres 3–6, que les propriétés critiques des modèles peuvent dépendre de la topologie locale du réseau sous-jacent, en particulier dans le domaine antiferromagnétique. Une manière plus radicale d'étudier cette dépendance est de se demander ce qui se passe si, au lieu de définir le modèle sur un réseau planaire fixé, on somme sur un ensemble de réseaux au niveau de la fonction de partition.

La sommation sur des réseaux (cartes) planaires peut être obtenue par l'intégration sur certains ensembles de matrices aléatoires. En prenant des matrices de dimension $N \times N$, la limite planaire est celle de $N \to \infty$. Cette connexion avec les intégrales matricielles fut découverte en 1978 dans une publication classique de Brézin, Itzykson, Parisi et Zuber [29]. Depuis, le sujet a connu un grand essor et pourrait former à lui seul l'objet d'une revue comme celle-ci. Suffit-il de dire que le lecteur interessé peut consulter le livre de Mehta [128] pour un aperçu des développements des années 1980, ainsi que les textes plus modernes [61, 59].

L'intégration sur une métrique fluctuante est justifiée physiquement par l'étude de membranes fluides. La dynamique locale des molécules qui constituent la membrane peut être modélisée par celle d'un ensemble de triangulations du plan. Par dualité, ces triangulations équivalent à l'ensemble de réseaux planaires cubiques aléatoires : ceci est l'exemple le plus élémentaire auquel s'appliquent les intégrales matricielles.

6.1 Matière et gravité quantique

La sommation sur les triangulations du plan donne une théorie conforme avec charge centrale $c = 0$. L'action décrivant la métrique fluctuante de tels modèles est l'analogue, en deux dimensions, des équations de la théorie de relativité générale d'Einstein pour un univers sans matière. Pour cette raison, les modèles correspondants sont connus sous le nom de *gravité pure*.

Il est également possible de définir des modèles statistiques sur des réseaux aléatoires. La fonction de partition consiste alors en une trace sur les degrés de liberté statistiques ainsi que sur la métrique fluctuante. L'analogie avec la théorie de relativité devient alors celle d'un modèle de matière couplé à la gravité quantique. Le premier exemple d'un tel modèle fut le modèle d'Ising sur des réseaux aléatoires de degré trois ou quatre [105, 106]. Sa résolution exacte nous permet de faire plusieurs observations de validité générale :

1. Le couplage du modèle (ici, le modèle d'Ising ferromagnétique) à la gravité change les valeurs des exposants critiques par rapport à celles du modèle sur un réseau régulier.
2. Cependant, la charge centrale c est inchangée.
3. L'hypothèse d'universalité est valable dans les deux cas. De même que le comportement critique du modèle d'Ising ne dépend pas du choix du réseau régulier, le comportement de sa contrepartie gravitationnelle ne dépend pas de la classe de réseaux aléatoires (cubiques, tétravalents, ...) sur laquelle on somme.
4. L'inclusion d'interactions supplémentaires (telles qu'un champ magnétique) est parfois plus facile dans le modèle gravitationnelle que sur un réseau régulier.

La relation entre les comportements critiques sans ou avec couplage à la gravité fut expliquée par une équation, dite *relation de KPZ*, qui donne les dimensions conformes Δ_k du modèle gravitationnel en fonction de celles h_k du modèle sur un réseau régulier [110, 41, 42] :

$$\Delta_k = \frac{\sqrt{1-c+24h_k} - \sqrt{1-c}}{\sqrt{25-c} - \sqrt{1-c}}. \tag{6.1}$$

Ici, la charge centrale c est commune pour les deux réalisations du modèle.

Le comportement critique du modèle gravitationnel est caractérisé par la singularité de la fonction de partition (connexe) $Z(x)$ à une certaine valeur x_c du poids donné à un sommet (x est appellé la constante cosmologique) :

$$Z(x) \sim (x_c - x)^{2-\gamma(c)}. \tag{6.2}$$

La constante $\gamma(c)$, dite exposant de susceptibilité de corde, est reliée à la charge centrale par [110, 41, 42]

$$\gamma(c) = \frac{c - 1 - \sqrt{(1-c)(25-c)}}{12}. \tag{6.3}$$

Pour la gravité pure on trouve $\gamma(c=0) = -1/2$.

La relation de KPZ a été déduite pour la suite de modèles minimaux habillés par la gravité quantique. En tant que telle, elle est seulement valable pour $c \leq 1$ (ou $c \geq 25$), comme le montre aussi les singularités des relations données ci-dessus. Si un modèle avec $1 < c < 25$ est couplé à la gravité, on s'attend à ce que les théories correspondantes soient dominées par des surfaces avec des longs "doigts" (phase de polymères ramifiés). Dans cette phase, on a $\gamma = 1/2$ partout [55, 40].

La validité de la relation KPZ a été vérifiée pour un certain nombre de modèles dont la solution gravitationnelle peut être trouvée exactement par les techniques d'intégration matricielle. Parmi ces modèles on trouve notamment le modèle O(n) [117, 70, 118] et le modèle de Potts [39, 171, 58].

D'autres modèles discutés en ce mémoire ont également été considérés dans leur version gravitationnelle. A titre d'exemples, citons : les cycles hamiltoniens sur des réseaux cubiques ordinaires [57] et bicoloriables (eulériens) [73], la phase compacte du modèle O($n=1$) sur des réseaux cubiques eulériens [64] et des problèmes de tricoloriage [63, 119].

Dans les sections suivantes, nous allons discuter de deux problèmes de nature énumérative qui se prêtent à une formulation en termes de matrices aléatoires : les méandres et les nœuds. Pour ces problèmes, aucune solution exacte n'est connue. Cependant, il est possible de déduire le comportement critique des modèles correspondants sur un réseau régulier de manière exacte. La relation KPZ fournit alors des conjectures pour les exposants critiques associés. Une grande partie du travail consiste alors à vérifier ces conjectures numériquement. Pour ce faire, on procède souvent à une énumération par ordinateur.

6.2 Méandres

Le problème de méandres peut être considéré comme une version améliorée du fameux *problème des sept ponts à Kœnigsberg*, résolu par Euler en 1736 [56]. Son énoncé est le suivant : Etant donné un nombre positif entier n, en combien de manières peut-on traverser une rivière en empruntant chacun de $2n$ ponts exactement une fois sans couper sa propre trajectoire et en revenant au point de départ ? Ou bien, dans un langage plus moderne : Trouver le nombre de manières topologiquement distinctes de couper une courbe autoévitante fermée par une droite en $2n$ points.

L'intérêt physique du problème de méandres est qu'il décrit le pliage de polymères en une dimension. Il existe un certain nombre d'autres applications en topologie, algorithmique et analyse : voir [67] pour une discussion détaillée ainsi que des références.

Il est possible d'énumérer manuellement le nombre M_n de méandres d'ordre n pour n suffisamment petit. Une approche d'énumération par ordinateur, qui est décrite en détail dans l'article [66] reproduit ci-dessous, a permis d'obtenir les M_n jusqu'à $n = 24$ [100, 66] : voir le Tableau 6.1. En analogie avec d'autres problèmes de la gravité quantique, on s'attend à ce que ces nombres croissent comme

$$M_n \sim C x_c^{-2n} n^{-\alpha}, \tag{6.4}$$

n	M_n	n	M_n	n	M_n
1	1	9	933458	17	59923200729046
2	2	10	8152860	18	608188709574124
3	8	11	73424650	19	6234277838531806
4	42	12	678390116	20	64477712119584604
5	262	13	6405031050	21	672265814872772972
6	1828	14	61606881612	22	7060941974458061392
7	13820	15	602188541928	23	74661728661167809752
8	110954	16	5969806669034	24	794337831754564188184

TAB. 6.1 – Nombre de méandres à $2n$ intersections jusqu'à $n = 24$.

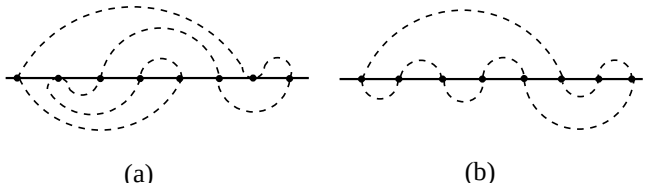


FIG. 6.1 – Méandres d'ordre $n = 4$ avec (a) ou sans (b) tangences.

où x_c est la valeur critique de la constante cosmologique et l'exposant critique $\alpha = 2 - \gamma(c)$ est relié à la charge centrale par (6.3). Ce qui est remarquable c'est que même si l'on ignore la valeur exacte de x_c (faute d'une solution de l'intégrale matricielle correspondante) on peut espérer calculer c en identifiant le modèle sur réseau régulier dont la version gravitationnelle est les méandres.

Sur la Figure 6.1.b nous montrons un des 42 méandres d'ordre $n = 4$, où les lignes en trait continu et pointillé représentent respectivement la rivière et le chemin. On voit que cela ressemble à la version gravitationnelle du modèle de boucles défini par la Figure 5.2, sauf que seuls les deux derniers sommets (où les deux couleurs de boucles se croisent) apparaissent. De l'autre côté, le modèle de boucles est en correspondance parfaite avec une version modifiée des méandres où l'on permet des croisements et des tangences. Un méandre de ce type, dit *méandre tangent*, est illustré sur la Figure 6.1.a. On pourrait paraphraser le nouveau type d'interaction en disant qu'il permet de visiter un pont tout en faisant demi-tour au milieu du pont.

Di Francesco, Golinelli et Guitter [65] ont avancé l'hypothèse que le modèle de boucles décrit bien la version régulière des méandres, à un détail important près. En effet, le modèle de boucles est défini sur le réseau carré qui est bicoloriable. Sur le réseau aléatoire, cette bicoloriabilité est perdue dès qu'on permet les tangences. Or, dans la théorie de champs [83] un manque de bicoloriabilité se traduit par le fait que la seconde composante de la hauteur devient un champ massif. Ce mécanisme est exactement celui responsable de la transition entre les phases compacte et dense du modèle de boucles [84]. En tant que tel, ses effets sur la charge centrale et les exposants critiques sont déjà connus. De l'autre côté, une fois la seconde composante de la hauteur disparue, la tangence devient une interaction non pertinente [65]. On s'attend alors à ce que pour n'importe quel poids (y compris zéro) de l'interaction de tangence, les méandres tangents soient décrits par la phase *dense* du modèle de boucles couplé à la gravité. Dans la limite où les fugacités de chaque type de boucles s'annulent, cette phase a $c = -4$ [84].

En particulier, la prédiction [65] pour l'exposant critique vaut

$$\alpha = 2 - \gamma(-4) = \frac{29 + \sqrt{145}}{12} \simeq 3.4201328\cdots. \tag{6.5}$$

Ceci est en bon accord avec les résultats numériques $\alpha = 3.4206 \pm 0.0004$ [100] et $\alpha \simeq 3.4207$ [66], même avec des tangences [66].

De nombreuses autres vérifications de cette identification sont possibles [65, 66]. En effet, il est possible de toucher le contenu en opérateurs du modèle de boucles en changeant les conditions au bord. Des résultats énumératifs dans différentes géométries sont rapportés dans la publication [66], jointe à ce chapitre, et ils supportent tous le scénario ébauché ci-dessus.

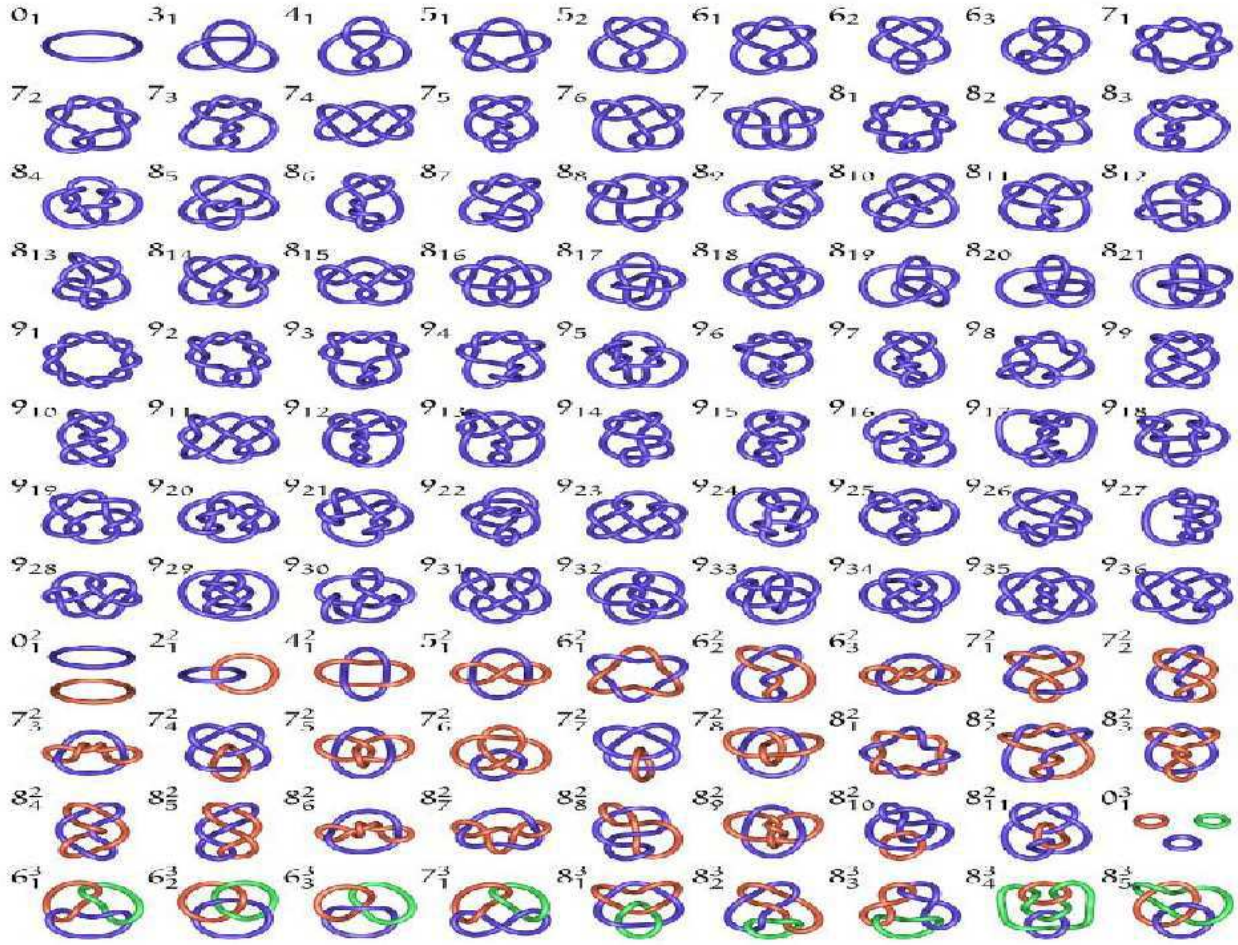


FIG. 6.2 – Tableau des premiers nœuds et entrelacs. Des nœuds non alternés apparaissent à partir de huit intersections. (Ce tableau ne montre pas tous les nœuds d'ordre 9.)

6.3 Nœuds et entrelacs

La théorie de nœuds est une autre branche des mathématiques issue de la *geometria situs* (géométrie de situation) [121, 159] et des récréations mathématiques [5, 132] du 17ème siècle. Pour une introduction moderne du sujet, voir [1, 104].

Ici, nous allons considérer le problème d'un point de vue énumératif : Etant donné un entier positif p, combien de nœuds topologiquement distincts existe-t-il dont le nombre minimal d'intersections (simples) dans une projection bidimensionnelle vaut p ?

Les quelques premières dizaines de nœuds sont montrées sur la Figure 6.2. On remarquera qu'il est possible de nouer plusieurs bouts de corde : dans ce cas, on parle d'*entrelacs*. Chacune de leurs composantes connexes porte une couleur différente sur la figure. Plus généralement, nous allons considérer la fonction génératrice (à constante cosmologique g)

$$F(n,g) = \sum_{k=1}^{\infty} \sum_{p=1}^{\infty} f_{k;p}\, n^k g^p, \tag{6.6}$$

où, en analogie avec le modèle O(n), chacune de k composantes porte un poids n. Les coefficients $f_{k;p}$ énumèrent les diagrammes du type montré en Figure 6.2. Ils sont liés à une intégrale sur des matrices hermitiennes $N \times N$ par

$$Z^{(N)}(n,g) \;=\; \int \prod_{a=1}^{n} \mathrm{d}M_a \, \mathrm{e}^{N \,\mathrm{tr}\left(-\frac{1}{2}\sum_{a=1}^{n} M_a^2 + \frac{g}{4}\sum_{a,b=1}^{n} M_a M_b M_a M_b\right)}, \tag{6.7}$$

FIG. 6.3 – L'équivalence topologique dite *flype*.

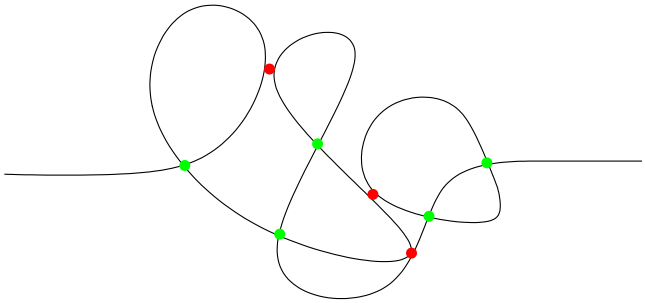


FIG. 6.4 – Diagrammes planaires avec des intersections et des tangences.

$$F(n,g) \quad = \quad \lim_{N\to\infty} \frac{\log Z^{(N)}(n,g)}{N^2}. \tag{6.8}$$

A ce point, plusieurs remarques techniques importantes s'imposent :

1. L'intégrale (6.7) pondère en fait chaque diagramme par l'inverse de son facteur de symétrie. Pour éliminer cette complication, nous allons désormais uniquement considérer des fonctions de corrélations, c'est-à-dire les entrelacs d'un ou plusieurs bouts de corde *ouverts*.

2. La fonction génératrice $F(n,g)$ énumère les diagrammes planaires sans prendre en compte les équivalences topologiques. Autrement dit, le comptage des diagrammes équivaut à un multiple comptage des entrelacs. Il est possible de résoudre ce problème pour une sous-classe d'entrelacs, dite les *entrelacs alternés*, dans laquelle la nature des croisements lorsqu'on suit un fil quelconque alterne entre "en dessous" et "au dessus". Nous considérons désormais uniquement cette sous-classe.

Avec ces contraintes, la question de surcomptage se réduit essentiellement à la factorisation d'une seule équivalence topologique dite *flype* [129]. (Ce mot en écossais fut introduit par Tait ; une traduction approximative est "chiquenaude".) Le *flype* est montré en Figure 6.3.

Par un travail sur la fonction génératrice $F(n,g)$ on peut alors montrer [91, 92, 97] que le comptage d'entrelacs équivaut à l'énumération de la classe de diagrammes planaires qui est illustrée en Figure 6.4. Ces diagrammes contiennent deux types de sommets : des intersections et des tangences. La sous-classe de tels diagrammes où seules les tangences sont permises définit le modèle O(n) gravitationnel. Ce modèle est exactement soluble [117, 70, 118]. Il parait pourtant plausible que les intersections agissent comme une perturbation pertinente qui introduit un flot vers une autre classe d'universalité.

Une manière de trancher est d'introduire les modèles correspondants sur un réseau régulier et d'utiliser la relation KPZ. Sur le réseau carré, un exemple d'un modèle ne comportant que des tangences est fourni par le modèle de Potts (dans sa représentation de modèle de boucles, voir le Chapitre 2). Comme nous en avons déjà discuté, ce modèle est dans la classe d'universalité du modèle O(n) dans la phase dense. Le lien avec le modèle O(n) gravitationnel est donc bien établi. Il semble alors logique d'étudier l'inclusion d'intersections dans le contexte du modèle de boucles sur le réseau carré.

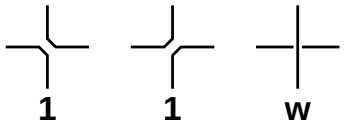


FIG. 6.5 – Sommets du modèle de boucles. Les intersections sont pondérées par w.

6.4 Modèle de boucles avec intersections

Motivés par la discussion ci-dessus, nous souhaitons étudier un modèle de boucles sur le réseau carré défini par les sommets de la Figure 6.5. Comme d'habitude, chaque boucle porte un poids n.

Sans intersections, ce modèle peut être traité par les méthodes du gaz de Coulomb [113]. Le résultat est un modèle d'interface avec une hauteur unidimensionnelle dont les exposants critiques confirment bien ceux qui sont connus de la solution exacte du modèle de Potts [13]. Or, une fois les intersections permises l'approche du gaz de Coulomb se heurte à un écueil : il n'est plus possible d'orienter les boucles de manière à faire une rédistribution locale de leurs fugacités. Cet obstacle peut également s'interpréter en termes de symétries : le défaut de pouvoir orienter les boucles diminue la symétrie U(n) à O(n).

Une autre possibilité d'attribuer un poids n à chaque boucle est fournie par la supersymétrie [140]. Contrairement au gaz de Coulomb, cette approche permet aux boucles de s'intersecter. Le poids n résulte d'une annulation partielle de degrés de libertés bosoniques et fermioniques dans la supertrace qui définit la fonction de partition. (Le modèle avec intersections peut cependant être étudié avec les techniques des modèles intégrables [127].)

Une étude récente [143] a souligné l'importance du choix d'espace cible pour le modèle sigma correspondant [80]. En effet, les deux choix possibles (une supersphère et une version supersymétrique d'un espace projectif) servent à distinguer entre les phases dense et diluée du modèle défini ci-dessus. Par rapport au gaz de Coulomb, les intersections représentent seulement une perturbation pertinente dans la phase dense [96] (pour $n < 2$), où elles introduisent un flot vers une phase de Goldstone caractérisée par une brisure spontanée de la supersymétrie. Le modèle de sigma correspondant flotte vers la limite de couplage zéro et est donc une théorie très simple de champs bosoniques et fermioniques libres (mais avec des corrections logarithmiques).

Les prédictions [96] pour les exposants critiques et la charge centrale

$$c = n - 1 \tag{6.9}$$

ont été vérifiées par des simulations numériques directes sur le modèle défini par les sommets de la Figure 6.5. La conclusion est que pour tout $w > 0$ le modèle est décrit par le point fixe supersymétrique.

6.5 Comportement asymptotique du nombre de nœuds

Comme dans le cas de méandres, on s'attend à ce que le nombre d'entrelacs (à fugacité n par composante connexe) croisse comme $\sim C x_{\rm c}^{-p} p^{-\alpha}$. L'exposant critique $\alpha = \alpha(n)$ dépend de n et est lié à la charge centrale c par la relation KPZ.

Or, il y a maintenant deux proposition pour la valeur de c. L'une est celle du modèle O(n),

$$c = 1 - 6(1-g)^2/g \tag{6.10}$$

avec $n = -2\cos(\pi g)$ et $0 < g < 1$ (phase dense), et l'autre est donnée par (6.9). Il est curieux de constater que dans les deux cas où le modèle d'entrelacs est exactement soluble, $n = 1$ [156] et $n = 2$ [170], ces valeurs coïncident. Dans la limite des nœuds, $n \to 0$, les valeurs correspondantes de l'exposant critique valent :

$$\alpha_{\rm CG}(c=-2) = 3, \qquad \alpha_{\rm SUSY}(c=-1) = \frac{13+\sqrt{13}}{6} \simeq 2.76759. \tag{6.11}$$

Il est physiquement raisonnable de supposer que l'élimination d'équivalences topologiques ne change pas la valeur de α. On peut alors trancher entre les deux conjectures (6.11) en énumérant les diagrammes planaires non renormalisés qui ont, à priori, des effets de taille finie plus réguliers.

Une énumération exacte par matrices de transfert jusqu'à $p = 22$ [91] ne mène pas à une précision suffisante pour distinguer entre les valeurs (6.11). Ce problème est en partie lié au fait qu'on ne connaît pas la forme des corrections (éventuellement logarithmiques) au comportement asymptotique dominant. Très récemment, une approche de simulation de Monte Carlo a permis d'étudier des systèmes beaucoup plus larges, de l'ordre $p \sim 10^5$ [152], et en particulier d'évaluer la dérivée $\alpha'(n)|_{n=1} = -0.30 \pm 0.01$. Les conjectures correspondantes sont $\alpha'_{\rm CG} = -\frac{3\sqrt{3}}{4\pi} \simeq -0.413$ et $\alpha'_{\rm SUSY} = -\frac{3}{10}$. Clairement, ce résultat est plutôt favorable à l'approche supersymétrique, qui paraît aussi la mieux justifiée d'un point de vue

théorique. Il serait intéressant de formuler des quantités liées aux nœuds qui permettraient de faire la liaison avec le contenu en opérateurs de la théorie supersymétrique.

6.6 Article "Exact meander asymptotics : a numerical check"

Cette note traite le problème d'énumération de méandres, c'est-à-dire le comptage des configurations topologiquement inéquivalentes dans lesquelles une courbe planaire fermée et autoévitante traverse une droite en un nombre de points prescrits. Nous résumons une description récemment introduite dans laquelle les méandres sont réalisés comme le couplage d'un modèle de deux types de boucles compactes à la gravité quantique. Cette description fournit des prédictions analytiques d'une multitude d'exposants décrivant les configurations des méandres. Ici, ces prédictions sont confrontées à des résultats d'énumération exacte obtenus par une méthode de matrice de transfert. L'accord est excellent.

ELSEVIER

Nuclear Physics B 580 [FS] (2000) 757–795

NUCLEAR PHYSICS B

www.elsevier.nl/locate/npe

Exact meander asymptotics: a numerical check

P. Di Francesco [a,1], E. Guitter [a,2], J.L. Jacobsen [b,3]

[a] *Service de Physique Théorique, C.E.A. Saclay, F-91191 Gif sur Yvette, France*
[b] *LPTMS, bâtiment 100, Université Paris-Sud, F-91405 Orsay, France*

Received 3 March 2000; accepted 1 May 2000

Abstract

This note addresses the meander enumeration problem: "count all topologically inequivalent configurations of a closed planar non self-intersecting curve crossing a line through a given number of points". We review a description of meanders introduced recently in terms of the coupling to gravity of a two-flavored fully-packed loop model. The subsequent analytic predictions for various meandric configuration exponents are checked against exact enumeration, using a transfer matrix method, with an excellent agreement.

PACS: 05.20.-y; 02.10.Eb; 04.60.Nc
Keywords: Meanders; Fully packed loop models; Coulomb gas; 2D quantum gravity; Transfer matrix

1. Introduction

The meander problem is one of those tantalizing questions that has resisted a definite solution for decades, although it is very easy to state: "given a positive integer n, in how many topologically distinct ways can a closed non-intersecting planar curve (*road*) cross a straight line (*river*) in exactly $2n$ points (*bridges*)?"

Originally an exercise of recreational mathematics [1], the meander problem turned out to have applications in the most various branches of science: sorting algorithms in computer science [2], enumeration of ovals of planar algebraic curves [3], classification of three-manifolds [4], and in connection with a particular type of self-avoiding walk describing the compact folding of a linear chain [5].

An obvious strategy would of course be to evaluate the first few meander numbers M_n, in the hope of finding an explicit formula, valid for arbitrary n. Such enumerative approaches exist on various levels of sophistication [6–9], and a recent transfer matrix method [10]

[1] philippe@spht.saclay.cea.fr
[2] guitter@spht.saclay.cea.fr
[3] jacobsen@ipno.in2p3.fr

0550-3213/00/$ – see front matter
PII: S0550-3213(00)00273-X

carried out this program up to $n = 24$. Although an explicit expression for M_n appears to be out of reach, it became clear that, in analogy with two-dimensional lattice polymers, the meander numbers scale asymptotically as $M_n \sim C R^{2n}/n^{\alpha}$, where R is a connectivity constant and α a configuration exponent.

A major achievement of random matrix theory has been to deal with precisely such asymptotic enumeration problems [11]. It is therefore natural to apply such techniques to the meander problem [12–14]. In particular it has emerged that a generalized multi-road multi-river meander problem, in which each closed segment of river (respectively, road) is given the statistical weight n_1 (respectively, n_2), can be cast as a Hermitian matrix model, known as the $\mathrm{O}(n_1, n_2)$ model [5]. In the special case of $n_2 = 1$ this model is soluble by a saddle-point method, leading to an exact evaluation of R and α [15,16]. Unfortunately these results do not pertain to the original meander problem, which is recovered in the limit $n_1, n_2 \to 0$.

In a recent publication [17] it was argued that the meander problem is a particular realization of the coupling to gravity of a certain two-flavored loop model [18], initially defined on the square lattice. The most general gravitational version of this loop model is a generalization of the meander problem in which river and road segments, counted with their respective weights of n_1 and n_2, are allowed to cross as well as to touch one another without crossing (tangency points) [17]. We shall refer to this model as *tangent meanders*. On the regular lattice, directed segments of river and road can be inserted by means of certain magnetic defect operators, for which the anomalous dimensions are known exactly. When dressed by quantum gravity, these dimensions transform according to the KPZ formula [19–21]. This transformation allows one to extract *exact* values for the configuration exponent α of tangent meanders, whereas R, being a non-universal quantity, is lost in the process. The connection to the original meander problem is then made by arguing that tangency is *irrelevant* from a renormalization point of view [17]. Thus, the result for α in fact pertains to the original meander problem, i.e., to the gravitational $\mathrm{O}(n_1, n_2)$ model. Moreover, the operator content of the theory gives access to other geometries, involving several rivers, possibly with marked points, as well as semi-meanders (river with a source).

Here we review and extend the arguments of [17]. In particular we establish the irrelevance of tangency rigorously in a number of special cases. We also add credibility to the theoretical predictions by performing extensive exact enumerations of various meander geometries, using a generalization of the transfer matrix method presented in [10]. We first address multi-component meanders, that allow for precisely checking the predicted value of the central charge of the underlying conformal theory. Next we explore two distinct river geometries, namely (i) two parallel rivers, and (ii) one semi-infinite river, that permit to validate their magnetic operator formulation within the corresponding conformal theory. We finally consider the case of tangent meanders and verify the irrelevance of tangency, thus confirming the cornerstone of the argument. In all cases we find an excellent agreement with theory, typically confirming the configuration exponents with 4–5 significant digits.

The paper is organized as follows. In Section 2 we review the square-lattice loop model and its solution, before presenting its gravitationally dressed version and the results for the asymptotics of a range of meander-related quantities. The transfer matrix algorithms are presented in Section 3, and in Section 4 we analyze the data and compare them to the theoretical predictions. Our conclusions and some perspectives can be found in Section 5.

2. Theory: from fully-packed loop gases to meanders

In this section, we review the arguments of Ref. [17] relating the meander problem to the gravitational version of a particular fully-packed loop model initially defined on the square lattice [18]. The effect of gravity is to replace the lattice with a random quadrangulation of the sphere. The lattice loop gas is described in Section 2.1 while its conformal structure is presented in Section 2.2. The connection to meanders via two-dimensional quantum gravity is explained in Section 2.3. This leads naturally to an effective field theory description detailed in Section 2.4 together with the subsequent predictions for various meandric configuration exponents.

2.1. Fully-packed loop gas on the square lattice

The configurations of the fully-packed loop model that we shall consider are defined by assigning to each edge of the two-dimensional square lattice either of two colors (say, black or white, represented as solid or dashed lines in Fig. 1, also referred to as 1 and 2 in the following) in such a way that each vertex has exactly two black and two white incident edges. Up to obvious rotations, this gives rise to the two vertex configurations depicted in Fig. 1 in which the black and white lines either avoid or cross each other. Note that with periodic boundary conditions the black and white lines form loops. It is interesting to remark that the fully-packed loop model's configurations defined here differ from those of the so-called densely packed loop model [18] in that each vertex is visited by a black *and* a white loop, whereas in the dense case, loops of a given (say black) color are not constrained to visit all vertices.

The partition function of the fully-packed loop model is then defined by assigning a weight n_1 per black loop and n_2 per white one,

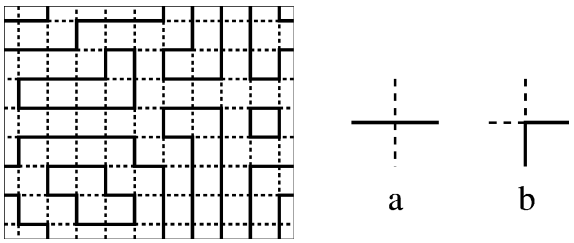


Fig. 1. A typical fully-packed loop configuration on the square lattice. Assuming doubly periodic boundary conditions, there are 6 black loops (solid lines) and 4 white ones (dashed lines). Up to rotations, the vertices of the model are of the two types (a) "crossing" or (b) "avoiding".

$$Z_{\mathrm{FPL}}(n_1,n_2)=\sum_{\substack{\text{fully-packed loop}\\ \text{configurations}}} n_1^{L_1}\,n_2^{L_2}, \tag{2.1}$$

where we have denoted by L_i the total numbers of loops of each color $i=1,2$. Following [18] we shall denote the model with partition function (2.1) as the $\mathrm{FPL}^2(n_1,n_2)$ model, while the densely packed version is referred to as the $\mathrm{DPL}^2(n_1,n_2)$ model.

The loop weights n_i may be recast as local Boltzmann weights as follows. This step is important in the field theoretic description of the model, since it leads to a *local* field theory. Let us assign to each black or white loop an arbitrary orientation, and attach to each vertex a local Boltzmann weight $e^{\mathrm{i}\pi(\varepsilon_1 e_1+\varepsilon_2 e_2)/4}$ where $\varepsilon_i=1$ if the oriented loop of color i makes a left turn, $\varepsilon_i=0$ if it goes straight, and $\varepsilon_i=-1$ if it makes a right turn. Summing over all possible orientations of all loops, we get a factor $2\cos\pi e_i$ per loop of color i, and therefore we reproduce the desired loop weights by setting

$$n_1=2\cos\pi e_1,\qquad n_2=2\cos\pi e_2. \tag{2.2}$$

2.2. Conformal field theory description

The $\mathrm{FPL}^2(n_1,n_2)$ model is known to be critical for $0\leqslant n_i\leqslant 2$ [18], and is described in the continuum limit by a simple conformal field theory based on free scalar fields. To identify its basic degrees of freedom, it is useful to rephrase the model as a (three-dimensional) height model as follows. Starting from an oriented fully-packed black and white loop configuration, we first bicolor the vertices of the square lattice, say with alternating filled (•) and empty (○) dots. Then we use the dictionary of Fig. 2 to assign one of the four labels A, B, C, D to each colored and oriented edge. With this convention, it is clear that edges of type $ABAB\ldots$ alternate along black loops, whereas edges of type $CDCD\ldots$ alternate along white loops, and that each vertex has one incident edge of each type A, B, C and D. It is seen that the four-labeling with A, B, C, D is in one-to-one correspondence with the coloring *and* orientation of edges of the FPL^2 model. In particular, the orientation of a given black or white loop is reversed if we interchange the A and B or C and D labels along the loop.

The above colors allow for the definition of a dual vector height variable on the center of each face of the lattice. Indeed, viewing as vectors the A, B, C, D labeling of the edges of

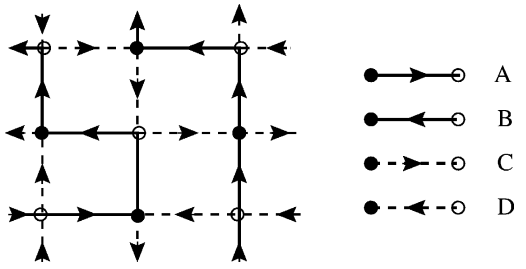


Fig. 2. A typical configuration of the FPL^2 model together with the bicoloration of its vertices (checkerboard of filled (•) and empty (○) dots). We have added the corresponding dictionary that allows to map the loop configurations onto A, B, C, D labelings of the edges.

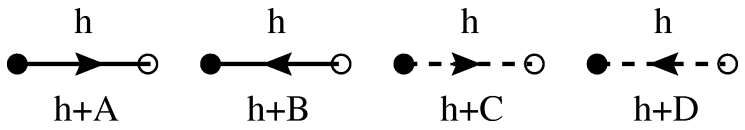


Fig. 3. Rules determining the change of the height variable across labeled edges. We adopt the Ampère convention that the height is increased (respectively, decreased) by the edge value if the arrow of the edge points to the left (respectively, right). The edge labels must be interpreted as three-dimensional vectors with the respective values $\mathbf{A}$, $-\mathbf{B}$, $\mathbf{C}$, $-\mathbf{D}$.

the lattice, let us arbitrarily fix the height to be zero on a given face of the lattice, and define it on all other faces by successive use the rules of Fig. 3 for the transition from a face to any of its neighbors. Note that it is necessary to impose the condition $\mathbf{A}+\mathbf{B}+\mathbf{C}+\mathbf{D}=0$ to ensure that the heights are consistently defined around each vertex. We may therefore assume in all generality that $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ are actually four vectors in $\mathbb{R}^3$ with vanishing sum. To get a more symmetric formulation, we may further fix $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ to be the four unit vectors pointing from the center of a tetrahedron towards its vertices. The heights are then clearly three-dimensional, as linear combinations of $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$. In the continuum limit, it was argued [18] that the three-dimensional height variable turns into a three-dimensional scalar field. Moreover the symmetries of the model completely fix the action for these fields and the corresponding field theory is conformal, with central charge

$$c_{\mathrm{FPL}}(n_1,n_2)=3-6\left(\frac{e_1^2}{1-e_1}+\frac{e_2^2}{1-e_2}\right), \tag{2.3}$$

where e_i have been defined in (2.2) and are constrained by $0\leqslant e_i\leqslant 1/2$. Actually the shift in the central charge away from 3 is due to the introduction of a background electric charge, ensuring that loops that have nontrivial winding with respect to the periodic boundary conditions still get correctly weighted, although for such loops the argument given before (2.2) no longer holds true.

2.3. Meanders: the coupling to gravity

To finally get to meanders, we must consider the coupling of the $\mathrm{FPL}^2(n_1,n_2)$ model to two-dimensional quantum gravity, by allowing the square lattice to fluctuate into arbitrary planar four-valent graphs. For each such graph the fully-packed loop model is still defined by coloring the edges black or white and allowing only the vertices shown in Fig. 1. As before, each colored loop is weighted by the appropriate n_i factor ($i=1,2$).

If we try to go through the steps of the previous section, namely by transforming the model into a height model, the issue of bicolorability of the vertices of the lattice becomes crucial on a random four-valent graph. Indeed, not all such graphs are vertex-bicolorable. So the coupling to gravity *stricto sensu* (sum over arbitrary planar four-valent graphs) will destroy this property.

We may now follow either of the two following paths. First, we can repair this and *impose* that the particular coupling to gravity preserve the bicolorability, namely that the gravitational model be defined on the set of vertex-bicolorable four-valent graphs only. These graphs are dual to the so-called eulerian quadrangulations. In genus zero (planar

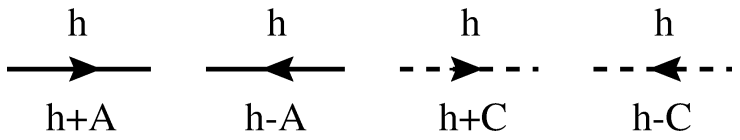


Fig. 4. Rules determining the change of the height variable across labeled edges in the non-bicolored case. These rules are identical to those of Fig. 3, with the further restriction that $\mathbf{B} = -\mathbf{A}$ and $\mathbf{D} = -\mathbf{C}$, allowing to ignore the bicoloration of vertices.

case), the latter are characterized by the fact that all their vertices have an even valency (Euler condition). If we couple the $\mathrm{FPL}^2(n_1, n_2)$ model to eulerian gravity, the A, B, C, D labeling is still well-defined, as well as the three-dimensional height, now defined on the centers of the faces of the graph. This preserves the degrees of freedom of the flat space model entirely in the gravitational formulation. This approach was initiated in [22,23] for the simpler case of the fully packed $\mathrm{O}(n)$ model with only one type of loops.

On the other hand, we can study how the degrees of freedom of the model are affected by the coupling to ordinary (non-eulerian) quantum gravity. Having lost the bicolorability of vertices, it is no longer possible to distinguish between A and B labels on one hand, and C and D on the other. We may still define an edge-labeling of the graph in one-to-one correspondence with oriented colored fully-packed loop configurations on the graph, but with vectors $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ satisfying the two constraints $\mathbf{A} + \mathbf{B} = 0$ and $\mathbf{C} + \mathbf{D} = 0$, and picking $\mathbf{A}$ and $\mathbf{C}$ to be two perpendicular unit vectors in $\mathbb{R}^2$. The correspondence between color/orientation and A, C labels reads as in Fig. 4. As in Eq. (2.1), the model is further completed by attaching weights n_i to each loop of color $i = 1, 2$. We may now define a height variable on the centers of the faces of the graph, by use of the previous rules. The main difference is that the height now lives in two dimensions (the plane generated by $\mathbf{A}$ and $\mathbf{C}$). Such a dimensional reduction is also observed on the square lattice when going from the $\mathrm{FPL}^2(n_1, n_2)$ to the $\mathrm{DPL}^2(n_1, n_2)$ model [18]; interestingly enough, the same dimensional reduction is also observed when the FPL^2 model is defined on the Manhattan square lattice, with oriented loops respecting the Manhattan orientation [24]. It results in a shift $c \to c - 1$ in the central charge of the underlying conformal theory, namely

$$c(n_1, n_2) = c_{\mathrm{DPL}}(n_1, n_2) \equiv 2 - 6\left(\frac{e_1^2}{1 - e_1} + \frac{e_2^2}{1 - e_2}\right). \tag{2.4}$$

In the following, we will concentrate on this formulation, eventually leading to the solution of the meander problem.

The partition function of the fully-packed model coupled to *ordinary* quantum gravity, hereafter referred to as the $\mathrm{GFPL}^2(n_1, n_2)$ model, reads in genus zero:

$$Z_{\mathrm{GFPL}}(n_1, n_2; x, y) = \sum_{\substack{\text{four-valent planar} \\ \text{graphs } \Gamma}} \sum_{\substack{\text{FPL configs C} \\ \text{on } \Gamma}} \frac{1}{|\mathrm{Aut}(\Gamma, \mathrm{C})|} n_1^{L_1} n_2^{L_2} x^{V_a(\Gamma)} y^{V_b(\Gamma)}, \tag{2.5}$$

where the sum extends over all the genus zero four-valent graphs Γ, and $|\mathrm{Aut}(\Gamma, \mathrm{C})|$ is the order of the symmetry group of Γ equipped with the loop configuration C. We have

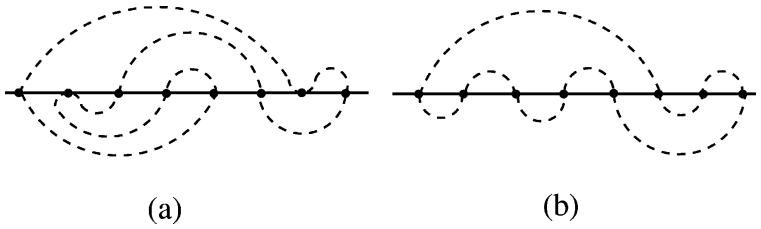


Fig. 5. Examples of (a) a tangent meander with 2 tangency points (b-vertices) and 6 bridges (a-vertices) and (b) a meander with 8 bridges.

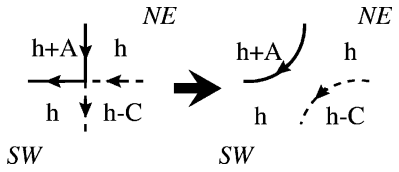


Fig. 6. A type b-vertex of the FPL^2 gravitational model, together with its dual height confi guration. We note that the NE and SW heights are identical. We may therefore undo the vertex as shown, which results in its irrelevance.

also denoted by V_a, V_b the total numbers of vertices of type a and b defi ned in Fig. 1 in the particular loop confi guration, namely we have weighted each crossing of a black and a white loop by x and each avoiding by y. When $x = y$, these are interpreted as the cosmological constant, as the total number of vertices $V_a + V_b$ of Γ is also the area of the corresponding dual random surface.

To generate meanders, we must now extract from (2.5) the confi gurations with only one black and one white loop, that will, respectively, play the role of the river and the road. This is done by taking the limit $n_1, n_2 \to 0$ in (2.5), resulting in

$$\begin{aligned} Z_{\mathrm{GFPL}}(x, y) &= \lim_{n_1, n_2 \to 0} \frac{1}{n_1 n_2} \big(Z_{\mathrm{GFPL}}(n_1, n_2; x, y) - 1 \big) \\ &= \sum_{\substack{n, p \geqslant 0, \\ n+p \geqslant 1}} \frac{x^{2n} y^p}{2(2n+p)} \mu_{2n,p}, \end{aligned} \tag{2.6}$$

where we have denoted by $\mu_{2n,p}$ the total number of *tangent meanders* with $2n$ crossings and p tangency points, i.e., confi gurations of a non-selfi ntersecting circuit (road) crossing a line (river) through $2n$ points (bridges) and touching the river p times (tangent contacts), as illustrated in Fig. 5. The usual meander numbers defi ned in [5] correspond to only crossings and no tangent points and read $M_n = \mu_{2n,0}$. In (2.6), the prefactor $1/(2(2n + p))$ stands for the symmetry factor attached to the tangent meanders: going from the graph to the representation where the river is a line, we may indeed cut the river loop in $2n + p$ places in between bridges and tangency points, and we still have 2 choices for the up/down position.

The meanders are therefore generated by the function (2.6) for $y = 0$. Let us now show that the universality class of the tangent meanders is the same as that of meanders. In the transformation into a (two-dimensional) height model, the "tangency" vertex b of Fig. 1 corresponds to the arrangements of heights on adjacent faces depicted in Fig. 6. We notice that the NE and SW heights are identical, irrespectively of the orientations of the two

loops. This means that as far as the height variable is concerned this vertex may be simply removed as shown. We conclude that the b-vertex of the model is irrelevant [17]. As a consequence, we expect the following asymptotics for $N = 2n + p$ large of the partition function $\mu_N(x, y)$ for tangent meanders with a total N of bridges and tangency points:

$$\mu_N(x, y) = \sum_{\substack{n,p\geqslant 0,\\ 2n+p=N}} x^{2n} y^p \mu_{2n,p} \sim C(x, y) \frac{R(x, y)^N}{N^\alpha}, \tag{2.7}$$

where the configuration exponent α is independent of x and y for $x > 0$ and $y \geqslant 0$. In particular, when $y = 0$, α is identified with the meander configuration exponent $M_n \sim C R^{2n}/n^\alpha$. The irrelevance of the vertex b will be checked numerically in Section 4.4 below.

More generally, we may consider tangent meanders with one single river, but arbitrarily many possibly interlocking closed roads. This is given by the $n_1 \to 0$ limit of (2.5) while $n_2 = q$ is kept finite, namely

$$\begin{aligned} Z_{\mathrm{GFPL}}(q; x, y) &= \lim_{n_1\to 0} \frac{1}{n_1}\big(Z_{\mathrm{GFPL}}(n_1, n_2 = q; x, y) - 1\big) \\ &= \sum_{n,p\geqslant 0; n+p\geqslant 1} \frac{x^{2n} y^p}{2(2n + p)} \mu_{2n,p}(q), \end{aligned} \tag{2.8}$$

where we have defined the tangent meander polynomial

$$\mu_{2n,p}(q) = \sum_{k=1}^{n+p} \mu_{2n,p}^{(k)} q^k \tag{2.9}$$

with coefficients $\mu_{2n,p}^{(k)}$ being the numbers of tangent meanders with k connected components of road, $2n$ bridges and p tangency points. The polynomial $\mu_{2n,0}(q) = m_n(q)$ coincides with the meander polynomial defined in [5]. We may also define the canonical partition function:

$$\mu_N(q; x, y) = \sum_{\substack{n,p\geqslant 0,\\ 2n+p=N}} x^{2n} y^p \mu_{2n,p}(q). \tag{2.10}$$

We expect a large N asymptotic behavior of the form

$$\mu_N(q; x, y) \sim C(q; x, y) \frac{R(q; x, y)^N}{N^{\alpha(q)}}, \tag{2.11}$$

where the configuration exponent $\alpha(q)$ only depends on q and not on y/x, due to the irrelevance of the b-vertex. In particular, it takes the same value at $y/x = 0$, where it coincides with the multi-component meander configuration exponent, i.e., $m_n(q) \sim C(q) R(q)^{2n}/n^{\alpha(q)}$.

In the special case $q = 1$ of arbitrarily many roads without extra weight, these numbers can be computed exactly. Indeed, we may decompose an arbitrary multi-component tangent meander with $2n$ bridges and p tangency points into its upper part (above the river) and lower part (below), and consider that these may be obtained first by picking p_1 points among the total $N = 2n + p$ to be the upper tangency points, and $p_2 = p - p_1$ to

be the lower ones. Let us then draw small semi-circles tangent to the p points, p_1 of them in the upper half, p_2 in the lower. With the $2n$ bridges, we have now a total of $2n+2p_1$ points in the upper half and $2n+2p_2$ in the lower one to be connected among themselves by pairs through non-intersecting arches. Such upper and lower arch configurations have already been extensively studied in [5]. In particular, there are $c_m = (2m)!/(m!(m+1)!)$ (c_m are the Catalan numbers) distinct upper arch configurations connecting $2m$ bridges by pairs in the upper half plane above the river. Hence we must choose among the c_{n+p_1} upper arch configurations and the c_{n+p_2} lower ones to form an arbitrary multi-component tangent meander. This results in the formula

$$\mu_{2n,p}(q=1) = \sum_{p_1=0}^{p} \frac{(2n+p)!}{p_1!(p-p_1)!(2n)!} c_{n+p_1} c_{n+p-p_1}, \tag{2.12}$$

where the combinatorial factor accounts for the choices of upper and lower tangent points among the total of $2n+p$. The corresponding partition function (2.10) reads

$$\mu_N(q=1;x,y) = N! \sum_{\substack{n,p,\\ 2n+p=N}} \frac{x^{2n}}{(2n)!} \sum_{\substack{p_1,p_2,\\ p_1+p_2=p}} \frac{y^{p_1}}{p_1!}\frac{y^{p_2}}{p_2!} c_{n+p_1} c_{n+p_2}. \tag{2.13}$$

When N is large, this is easily estimated by a saddle-point technique, making use of the Stirling formula. For $x, y \geqslant 0$, we find the large N behavior

$$\mu_N(q=1;x,y) \sim \frac{(4x+8y)^N}{N^3} \tag{2.14}$$

up to a multiplicative constant depending on x and y only. This shows explicitly that the exponent $\alpha(q=1)=3$ is robust and is not affected by the respective values of x and y. This confirms in particular the above-mentioned irrelevance of the b-vertex, as α keeps to the same value, irrespectively of y.

In conclusion, the meander numbers belong to the universality class of GFPL$^2(n_1,n_2)$ model at $n_1, n_2 \to 0$. The corresponding flat space theory has the central charge (2.4) with, according to (2.2), $e_1 = e_2 = 1/2$, hence $c=-4$. In the next subsection, we will concentrate on the partition function $Z_{\mathrm{GFPL}}(x, y=0)$ (2.6) that generates the meander numbers.

More generally, the multi-component meander polynomial belongs to the universality class of the GFPL$^2(n_1,n_2)$ model, with $n_1 \to 0$ and $n_2 = q$ and central charge (2.4). Finally, we may also consider meanders with arbitrarily many rivers and roads, with a weight n_1 per river and n_2 per road; these objects belong to the universality class of the GFPL$^2(n_1,n_2)$ model. It has been shown using matrix model techniques that the GFPL$^2(n_1=1,n_2=q)$ with $y=0$ belongs to the same universality class as the O(q) model when coupled to ordinary gravity (i.e., defined on arbitrary four-valent graphs) [15]. More precisely, the number of multi-river, multi-road meanders with a total of $2n$ intersections and with a weight q per road and 1 per river, also weighted by their inverse symmetry factor (and multiplied by $4n$ to make it comparable to M_n) behaves for large n as

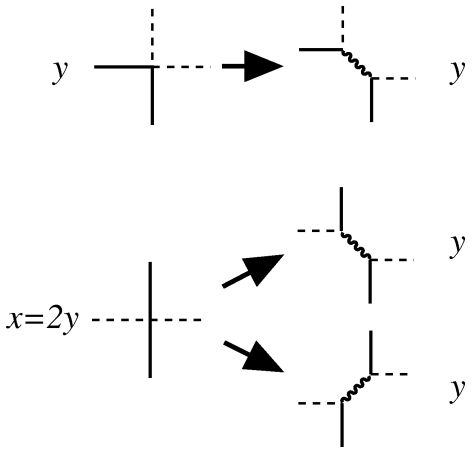


Fig. 7. The transformation of the four-valent a- and b-vertices of the FPL^2 model into pairs of trivalent ones connected by an extra (gray) edge, depicted as a wavy line. Due to the 1 to 2 correspondence in the case of the a-vertex, we must restrict ourselves to the case $x = 2y$, and then each trivalent vertex receives a weight $\sqrt{y}$.

$$C(1,q)R(1,q)^{2n}/n^{\alpha(1,q)}, \qquad R(1,q) = 2\frac{\sin^2(\pi e/2)}{e^2}, \\ \alpha(1,q) = \frac{2-e}{1-e}, \qquad q = 2\cos(\pi e). \tag{2.15}$$

The determination of $R(1,q)$ on which we have no prediction was made possible by an explicit mapping of the matrix model onto a particular version of the gravitational $\mathrm{O}(q)$ model, where the critical value of the cosmological constant x_c can be explicitly calculated. The value $\alpha(1,q)$ corresponds to the expected scaling behavior of a $c = 1 - 6e^2/(1-e)$ (cf. (2.4) with $n_1 = 1$ and $n_2 = q$) conformal theory coupled to two-dimensional quantum gravity. In particular, when $q = 1$, $c = 0$ we get $\alpha(1,1) = 5/2$ as expected in "pure gravity" without matter.

It is interesting to notice that if we restore the b-vertex (with a weight y per vertex), and choose specifically $x = 2y$, then the $\mathrm{GFPL}^2(1,q)$ model can be mapped onto that of tricoloring the edges of a random trivalent graph (with black, gray and white edge colors) with a weight q per loop of alternating say gray and white colors. Indeed, the quartic vertices a and b may be decomposed into pairs of connected trivalent ones as shown in Fig. 7 in which we have added a third type of edge, say with gray color. This transformation allows to map the configurations of the $\mathrm{GFPL}^2(1,q)$ model weighted by $x = 2y$ per a-vertex and y per b-vertex onto those of arbitrary edge-tricolored trivalent graphs, with a weight $\sqrt{y}$ per vertex and q per white/gray loop. The latter model was solved in [25] and identified with a particular version of the fully-packed $\mathrm{O}(2q)$ model on random trivalent graphs: indeed, the white and gray edges form naturally fully packed loops on the graph, and on each loop we may interchange the two colors to generate new tricolorings, hence we may simply draw fully-packed loops on trivalent graphs, and attach a weight $2q$ to each of these loops. But the loops have all even lengths and this turns out to dimensionally reduce the model to an effective $\mathrm{O}(2q/2 = q)$ one with central charge $c = 1 - 6e^2/(1-e)$ as before. This particular point therefore lies in the same universality

class as the model without b-vertex, which gives another explicit example of the irrelevance of the b-vertex.

2.4. Field theory description of meandric numbers

The coupling of a conformal field theory with central charge $c \leqslant 1$ to two-dimensional quantum gravity (i.e., its definition on random surfaces) has a simple field-theoretical formulation in terms of the Liouville field describing the conformal classes of metrics of the surfaces. This has led to a number of results, including the precise determination of various critical exponents. Indeed, the gravitational theory (say on genus zero surfaces) displays a critical behavior as a function of the cosmological constant x. In particular, there exists a finite value x_c of x at which the (connected) partition function behaves as

$$Z(x) \sim (x_c - x)^{2-\gamma(c)}, \tag{2.16}$$

where the string susceptibility exponent γ is related to the central charge c through [19–21]

$$\gamma(c) = \frac{c - 1 - \sqrt{(1-c)(25-c)}}{12}. \tag{2.17}$$

When applied to the $\mathrm{DPL}^2(0,0)$ model of the previous section (with $c = -4$), whose gravitational version was shown to describe meanders, we find that

$$\gamma \equiv \gamma(-4) = -\frac{5 + \sqrt{145}}{12}. \tag{2.18}$$

Comparing (2.16) with the expansion

$$Z_{\mathrm{GFPL}}(x) = \sum_{n \geqslant 1} \frac{x^{2n}}{4n} M_n \tag{2.19}$$

we deduce the asymptotic behavior [17]

$$M_n \sim C \frac{x_c^{-2n}}{n^{\alpha}}, \quad \alpha = 2 - \gamma = \frac{29 + \sqrt{145}}{12}. \tag{2.20}$$

Moreover, a number of the operators of the flat space conformal theory (in particular the spinless ones, with conformal dimensions $h = \bar{h}$) get dressed by gravity, in such a way that they acquire anomalous scaling dimensions. Any given operator ϕ_k with dimensions $h_k = \bar{h}_k$, is dressed into an operator $\tilde{\phi}_k$ with dressed dimension Δ_k such that the correlation functions behave as

$$\langle \tilde{\phi}_{k_1} \tilde{\phi}_{k_2} \ldots \tilde{\phi}_{k_p} \rangle \sim (x_c - x)^{2-\gamma+\Sigma(\Delta_{k_i}-1)}, \tag{2.21}$$

when x approaches the critical value x_c, and where the dressed dimension Δ_k is related to the flat space conformal dimension h_k through

$$\Delta_k = \frac{\sqrt{1 - c + 24h_k} - \sqrt{1-c}}{\sqrt{25-c} - \sqrt{1-c}}. \tag{2.22}$$

Let us now present the operator content of the $c = -4$ conformal theory describing the dense loop model $\mathrm{DPL}^2(0,0)$ [18]. For generic values of n_1, n_2, the $\mathrm{DPL}^2(n_1, n_2)$ has

a continuum description as a two-component scalar field with charges at infinity. More precisely, it is a Coulomb gas made of two decoupled scalar fields, with $c = c(n_i) = 1 - 6e_i^2/(1-e_i) = -2$ at $n_i = 0$ ($e_i = 1/2$), respectively, each viewed as the effective field theory of loops of one color. In particular, within each scalar field theory (indexed by the color $i = 1, 2$), there exist operators $\psi_k^{(i)}(z)$ that create k oriented defect lines (of color i) for the scalar field, with conformal dimensions [26,27]:

$$h_k^{(i)} = \frac{k^2 - 4}{32} \tag{2.23}$$

at $n_i = 0$. In the Coulomb gas formalism, these correspond to electromagnetic operators with electric charge e_i (spin-wave) and magnetic charge $\pm k/2$ (vortex), according to whether the defect line is oriented from or to the insertion point, and k is a strictly positive integer. The electric charge ensures that, if the defect lines wind around the insertion point, all extra curvature weights get cancelled [28]. For $k = 0$, (2.23) must be replaced by $h_0^{(i)} = 0$ corresponding to the identity operator. The correlation functions must have a vanishing total magnetic charge. Although these operators also carry an electric charge $1/2 = e_1 = e_2$, the electric neutrality of correlators imposes no extra condition. [4]

We may also combine operators for both colors, namely consider mixed operators $\psi_{k_1,k_2} = \psi_{k_1}^{(1)}\psi_{k_2}^{(2)}$ with conformal dimension $h_{k_1,k_2} = h_{k_1}^{(1)} + h_{k_2}^{(2)}$ for $k_1, k_2 \in \mathbb{Z}$.

Let us now study the dressing of these operators by gravity, and interpret them in meandric terms. The dressed operator $\tilde{\psi}_k^{(i)}$ again corresponds to the creation of a vertex with $|k|$ outcoming ($k > 0$) or incoming ($k < 0$) lines of color i.

As a first example, note that $h_2^{(i)} = 0$. From the formula (2.21), it is clear that the operators $\tilde{\psi}_{\pm 2}^{(i)}$ have the effect of *marking* a point on a loop of color i. However, due to the constraint of global magnetic neutrality of correlators, the marking is not arbitrary as the loop of color i must have an even number of marks (alternating $\tilde{\psi}_2^{(i)}$ and $\tilde{\psi}_{-2}^{(i)}$). In particular, the one-point function $\langle \tilde{\psi}_2^{(1)} \rangle$ that would naively count meanders with a marked river turns out to vanish, while for instance the two-point function $\langle \tilde{\psi}_2^{(1)} \tilde{\psi}_{-2}^{(1)} \rangle$ indeed counts meanders with two marked points on the river. We will meet more examples of this below.

The first application of the above concerns the two point function describing the insertion of a segment of river (color 1)

$$\left\langle \tilde{\psi}_1^{(1)} \tilde{\psi}_{-1}^{(1)} \right\rangle \sim (x_c - x)^{2\Delta_1 - \gamma}. \tag{2.24}$$

Recall that when we take the limit $n_1, n_2 \to 0$, only diagrams with one connected component of river and one of road are selected. In the case of (2.24), the river forms a segment, around which the road can freely wind (see Fig. 8). To fix ambiguities, let us send one end of the river to infinity (say to the left) and therefore represent the river as a half-line (this is allowed as we work on Riemann sphere). The number of configurations

[4] It can be shown [29] that the corresponding non-minimal conformal theory pertaining to the loops of color i (with central charge $c = 1 - 6e_i^2/(1-e_i)$) is equally well defined as a theory with central charge $c = 1$, in which the operator content has been reorganized, and in particular all of the above electromagnetic operators get a vanishing electric charge, while their magnetic one is unchanged. A supersymmetric version of this theory was also introduced in [30], in which arbitrary correlators could be calculated, irrespectively of the electric charge.

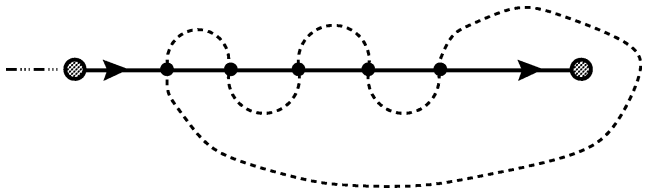


Fig. 8. A typical semi-meander configuration.

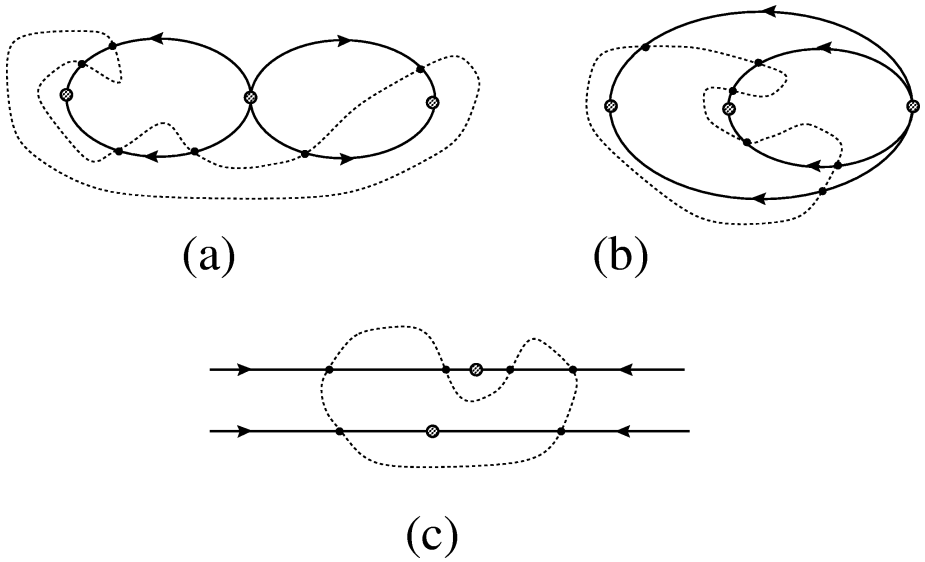


Fig. 9. A typical example of an eight figure river geometry (a). On the sphere, it is equivalent to the situation of (b), where the river crossing has been cleared of all winding pieces of road. This crossing may finally be sent to infinity (c) so as to form two parallel rivers. We have represented a particular configuration of road with $2p_1 = 4$ bridges on one loop and $2p_2 = 2$ on the other. Due to magnetic neutrality in the operator formulation, the two loops are marked as shown.

of a closed road crossing a half-line (river with a source) through n bridges is defined as the semi-meander number $\overline{M}_n$. We immediately identify the series expansion of (2.24) as a function of x to be

$$\langle \tilde{\psi}_1^{(1)} \tilde{\psi}_{-1}^{(1)} \rangle = \sum_{n \geqslant 1} \overline{M}_n x^n. \tag{2.25}$$

We therefore deduce the semi-meander asymptotics [17]

$$\overline{M}_n \sim \overline{C} \frac{{x_c}^{-n}}{n^{\bar{\alpha}}}, \qquad \bar{\alpha} = 1 + 2\Delta_1 - \gamma = 1 + \frac{\sqrt{11}}{24}\big(\sqrt{5} + \sqrt{29}\big). \tag{2.26}$$

Our second example is the "eight" meander, in which the river forms the figure "eight". To generate it, we need to insert an operator $\tilde{\psi}_4^{(1)}$, but magnetic neutrality forces us to insert two river-marking operators $\tilde{\psi}_{-2}^{(1)}$, one on each loop of the eight. The three-point correlation function $\langle \tilde{\psi}_4^{(1)} (\tilde{\psi}_{-2}^{(1)})^2 \rangle$ therefore generates the numbers of meanders whose river forms an eight figure, and with one marked point on each loop (see Fig. 9 (a)). To fix ambiguities, let us send the river-crossing to infinity. We may now represent the river as two parallel

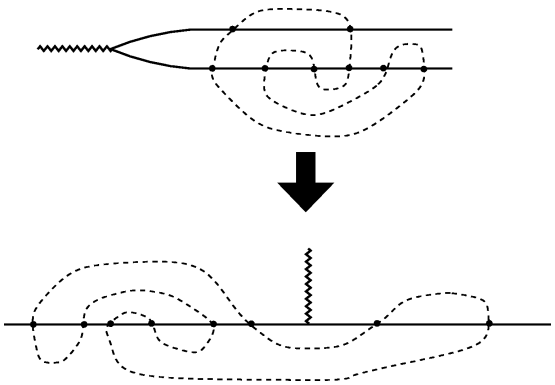


Fig. 10. A meander with two parallel rivers may be viewed as a meander "with a seam". The point where the seam is attached is a remnant of the point at infinity (say to the left), where the two rivers meet. The seam just prevents roads from encircling this point.

lines (all connected to the four-valent point at infinity), but the markings are still there (cf. Fig. 9 (c)). Hence we have

$$\begin{aligned}\langle \tilde{\psi}_4^{(1)} (\tilde{\psi}_{-2}^{(1)})^2 \rangle &= \frac{1}{2} \sum_{n \geqslant 1} x^{2n} M_n^{2\text{-mark}} \\ &= \frac{1}{2} \sum_{n \geqslant 1} x^{2n} \sum_{p_1+p_2=n} (2p_1+1)(2p_2+1) M_{p_1,p_2}, \end{aligned} \tag{2.27}$$

where $M_n^{2\text{-mark}}$ is the number of meanders with two rivers, each of which is marked, and M_{p_1,p_2} is the number of meanders with two parallel rivers, and with $2p_1$ bridges one the first and $2p_2$ on the second (there are indeed $2p_i+1$ possible markings on a river loop with $2p_i$ bridges). Note the prefactor of $1/2$ accounting for the fact that we distinguish the top and bottom of the figure when we represent a two-river meander. From (2.21), we find that [17]

$$\begin{aligned} M_n^{2\text{-mark}} &\sim \text{const}\, \frac{x_c^{-2n}}{n^{\alpha_{2\text{-mark}}}}, \\ \alpha_{2\text{-mark}} &= \Delta_4 + 2\Delta_2 - \gamma = \frac{1}{24}\big(\sqrt{5}+\sqrt{14}\big)\big(\sqrt{5}+\sqrt{29}\big). \end{aligned} \tag{2.28}$$

It is also easy to determine the large n behavior of the unmarked meanders with two parallel rivers. Let us draw the two rivers horizontally. We may view the two rivers as connected to each other at their far left (at infinity) and represent them as in Fig. 10 as one marked line, such that a seam originates from the marked point that prevents any road from encircling the point to the left. So we have mapped two-river meanders onto one-river meanders with a seam. We may now generate all meanders with a seam by considering all the ways of placing a seam on the upper-half of meanders. If $E(M)$ denotes the total number of upper exterior arches of a meander M (namely arches that have no other arch above them), then there are $E(M)+1$ distinct ways of decorating the meander with a seam. Hence the total number of two-river meanders reads

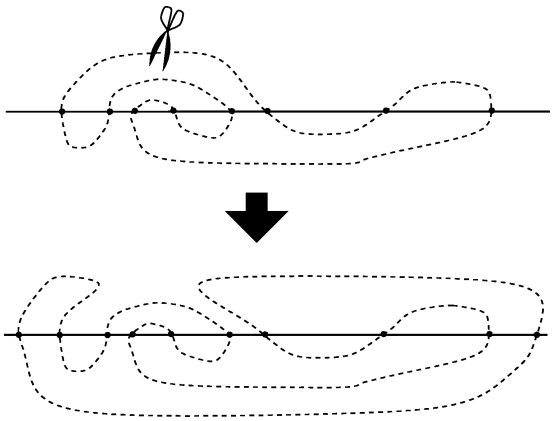


Fig. 11. How to construct a new meander with $2n+2$ bridges from a meander with $2n$ bridges and one distinguished exterior arch, here on the upper half of the meander. We cut the corresponding arch, and glue it back across the river, by encircling the whole lower part of the meander. This is easily inverted within the set of meanders with $2n+2$ bridges and one unique upper or lower exterior arch, showing that this construction gives only rise to distinct meanders.

$$M_n^{2\text{-riv}} = \sum_{\substack{\text{meanders } M \\ \text{with } 2n \text{ bridges}}} \big(E(M)+1\big) = \langle E+1\rangle_n M_n, \tag{2.29}$$

where $\langle\cdots\rangle_n$ stands for the average over all meanders with $2n$ bridges.

We now argue that $\langle E+1\rangle_n$ remains finite when $n\to\infty$. Indeed, starting from a meander with $2n$ bridges, we may construct one new meander with $2n+2$ bridges for each upper and lower exterior arch. As shown in Fig. 11, we simply cut the arch and close it on the other half, thus creating two new bridges and keeping a single connected road. Such meanders have $2n+2$ bridges and one single exterior arch on the side where we have closed the arch, and are all distinct. Hence we may write that their number is bounded by M_{n+1}, namely

$$2 \sum_{\substack{\text{meanders } M \\ \text{with } 2n \text{ bridges}}} E(M) \leqslant M_{n+1} \Rightarrow \langle E\rangle_n \leqslant \frac{1}{2}\frac{M_{n+1}}{M_n} \to \frac{R}{2}. \tag{2.30}$$

We therefore have the asymptotic bounds $1 \leqslant \langle E+1\rangle_n \leqslant 1+R/2$, which implies that $M_n^{2\text{-riv}} \propto R^{2n}/n^{\alpha}$ like ordinary meanders, hence $\alpha_{2\text{-riv}} = \alpha$. We also deduce that

$$\begin{aligned} \big\langle (2p_1+1)(2p_2+1)\big\rangle_n &\equiv \frac{\sum_{p_1+p_2=n}(2p_1+1)(2p_2+1)M_{p_1,p_2}}{\sum_{p_1+p_2=n} M_{p_1,p_2}} \sim n^{\beta}, \\ \beta = \alpha - \alpha_{2\text{-mark}} &= 2 - \Delta_4 = 2 - \frac{1}{24}\big(\sqrt{14}-\sqrt{5}\big)\big(\sqrt{5}+\sqrt{29}\big) \\ &= 1.521898\ldots. \end{aligned} \tag{2.31}$$

This shows the rather unexpected result that the two-river meanders tend to be very asymmetric, with a number of bridges of the order n on one river (since $p_1+p_2=n$) and of the order $n^{\beta-1}$ on the other, with $\beta-1 = 0.521898\ldots$. The above arguments can be generalized to the case of multi-component meanders and give rise to similar predictions.

In [17], a number of other results have been presented, all corresponding to more sophisticated river geometries, and making use of the magnetic defect operators $\tilde{\psi}_k^{(i)}$. It would also be possible in principle to use mixed operators to generate diagrams with both road and river geometries fixed. Another direction consists in going away from the point $n_1 = n_2 = 0$, for instance by considering the $\mathrm{GFPL}^2(n_1 = 0, n_2 = q)$ model in which meanders with arbitrary numbers of roads are considered, with a weight q per road. In that case, the corresponding conformal theory with central charge (2.4)

$$c = -1 - 6\frac{e^2}{1-e}, \qquad e = \frac{1}{\pi}\operatorname{Arccos}\left(\frac{q}{2}\right) \tag{2.32}$$

may be viewed as two decoupled bosonic field theories: a $c = -2$ theory (that of river loops, at $n_1 = 0$) and one with $c = c(q) = 1 - 6e^2/(1-e)$, where $q = 2\cos\pi e$ (that of road loops). The generating function $Z_{\mathrm{GFPL}}(q; x, y = 0)$ (2.8) for meander polynomials reads

$$Z_{\mathrm{GFPL}}(q; x, y = 0) = \sum_{n \geqslant 1} \frac{x^{2n}}{4n} m_n(q) \sim (x_c - x)^{2-\gamma(c)} \tag{2.33}$$

with $\gamma(c)$ as in (2.17), and c as in (2.32). This turns into the asymptotics

$$m_n(q) \sim C(q)\frac{R(q)^{2n}}{n^{\alpha(q)}}, \quad \alpha(q) = 2 - \gamma(c), \tag{2.34}$$

namely [17]

$$\alpha(q) = 2 + \frac{1 - e + 3e^2 + \sqrt{(1-e+3e^2)(13-13e+3e^2)}}{6(1-e)} \tag{2.35}$$

with e as in (2.32). An analogous formula holds for the (multi-component) semi-meander polynomials, namely

$$\bar{m}_n(q) \sim \overline{C}(q)\frac{\overline{R}(q)^n}{n^{\bar{\alpha}(q)}}, \quad \bar{\alpha}(q) = 1 + 2\Delta_1(c) - \gamma(c), \tag{2.36}$$

with $\Delta_1(c)$ as in (2.22), c as in (2.32), and $h_1 = -3/32$ as in (2.23). This yields [17]

$$\bar{\alpha}(q) = 1 + \frac{\sqrt{2(24e^2+e-1)}\left(\sqrt{1-e+3e^2} + \sqrt{13-13e+3e^2}\right)}{24(1-e)}, \tag{2.37}$$

with e as in (2.32).

An important remark is in order concerning the range of validity of (2.35) and (2.37). First, the $\mathrm{DPL}^2(0, q)$ model is critical only for $q \leqslant 2$, i.e., $e \geqslant 0$. We expect therefore a very different scaling behavior for the meandric numbers when $q > 2$. At large q, it was shown in Ref. [8] that

$$\alpha(q) = \frac{3}{2} \tag{2.38}$$

independently of q.

On the other hand, for semi-meanders, this transition at $q = 2$ is never reached. Indeed, another phenomenon appears at a value $q_c < 2$. It corresponds to a proliferation of

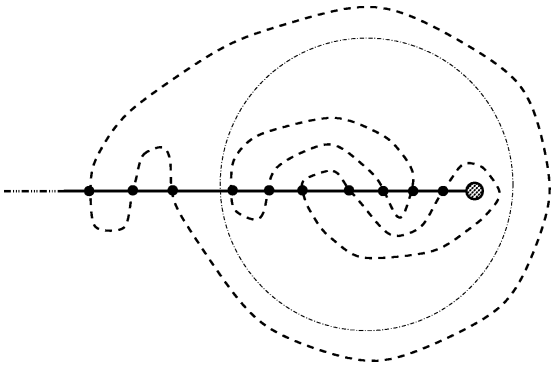


Fig. 12. A typical "branched" semi-meander.

connected components of road and also of the pieces of road that wind around the source of the river. More precisely, let us compare the two following contributions to the multi-component semi-meander numbers, according to whether:

(i) "branched" semi-meanders dominate: the semi-meander is typically separated into two parts, namely the interior and exterior of a "big circle" as shown in Fig. 12. The contribution of such semi-meanders reads asymptotically

$$\sum_{n_1+n_2=n} \frac{\overline{R}(q)^{n_1}}{n_1^{\bar{\alpha}(q)}} \times \frac{\overline{R}(q)^{n_2}}{n_2^{\bar{\alpha}(q)}} \sim \frac{\overline{R}(q)^n}{n^{2\bar{\alpha}(q)-1}}, \tag{2.39}$$

(ii) "connected" semi-meanders dominate: the semi-meanders cannot be cut as in (i), and typically behave as

$$\frac{\overline{R}(q)^n}{n^{\bar{\alpha}(q)}}. \tag{2.40}$$

The transition between the two regimes (i)–(ii) will take place when $2\bar{\alpha}(q) - 1 = \bar{\alpha}(q)$. We deduce that precisely at the transition, we must have $\bar{\alpha} = 1$, which according to (2.37) takes place when $24e^2 + e - 1 = 0$, namely at the critical value $e = e_c$, $q = q_c$ given by

$$e_c = \frac{\sqrt{97}-1}{48}, \qquad q_c = 2\cos\pi\frac{\sqrt{97}-1}{48}, \tag{2.41}$$

also corresponding to $c = 3/4$. Hence the formula (2.37) is only valid for $e_c \leqslant e$, namely $q \leqslant q_c = 1.673849\ldots < 2$. Beyond $q = q_c$, we expect the average number of connected components of road to be of the order of n, and we have [8]

$$\bar{\alpha}(q) = 0, \quad \text{for } q > q_c. \tag{2.42}$$

It is quite interesting to notice that the transition of semi-meanders is of a different nature than that of meanders. The latter simply encounters the $q = 2$ transition of the $O(q)$ model. The former undergoes a winding transition, in which semi-meanders become very different from meanders. Indeed, as long as the dominant semi-meanders have very little winding around the source of the river, they behave just like meanders (hence $R(q) = \overline{R}(q)$ in the regime $q < q_c$). But when the winding number of the semi-meanders becomes relevant (of the order of n) we expect many more semi-meanders than meanders with the same number of bridges, and $\overline{R}(q) > R(q)$ for $q > q_c$.

3. Meander enumeration algorithms

In this section, we will describe the algorithms that we used to check numerically our predictions of Section 2 for the different configuration exponents. All the results presented below concern the case $n_1 = 0$, i.e., a connected river configuration, and a varying weight $n_2 = q$ per road, i.e., multi-connected road configurations. We will also consider the several meander geometries discussed in Section 2.

All our algorithms are based on transfer matrix techniques. Transfer matrices have proven to be very powerful for studying a wide range of statistical mechanics systems, especially in two dimensions. Originally applied to systems with local interaction, such as the Ising model, the "row to row" transfer matrix describes the transition between two successive rows of spins, say at 'time' t and $t + 1$ along the transfer direction. More recently, the transfer matrix technique was also applied to systems with non-local degrees of freedom, but where, in a suitably designed basis, the statistical weights can still be evaluated between neighboring time slices. In this way it became possible to study self-avoiding polygons [31] and walks [32], and the random cluster model [33], to mention but a few important examples.

Finally, transfer matrices can also be applied to random lattice problems, as long as a definite transfer direction can be defined, as for instance in the case of lorentzian gravity [34]. This is also precisely the case in the meander problem, where we can simply choose to transfer along the river, adding one bridge in each time step. This was first recognized by Jensen in [10] where the method proved to be much better than previous enumeration algorithms, leading in particular to the largest accessible numbers of bridges.

When implemented on a computer, the transfer matrix algorithms will allow to enumerate *exactly* the various meandric objects that we are interested in for a fixed finite number N of bridges up to a certain maximal value of N (typically up to 48 bridges in the results presented below). From these exact finite N values, we can extract estimates for the large N asymptotic behaviors, and in particular for the configuration exponents described in Section 2. We can then compare these estimates to the corresponding expected theoretical values.

This section is organized as follows: the case of one infinite river (the original meander problem) is discussed in Section 3.1, while the other geometries, including the case of two infinite parallel rivers (connected at infinity), with possibly marked points on each of them, and the case of a semi-infinite river (semi-meanders), are discussed in Section 3.2. In the three cases above, we allow only crossing vertices of type a (cf. Fig. 1). The inclusion of tangency points for the infinite river case (tangent meanders) is discussed in Section 3.3.

3.1. One infinite river (meanders)

Fig. 13 shows a typical meander system with one infinite river and $N = 8$ bridges. The partition function is defined by giving a weight $n_2 = q$ to each road. The transfer matrix acts by transferring from the left to the right a vertical line intersecting the river between two consecutive bridges. A *state* characterizes that part of the system that is to the left of

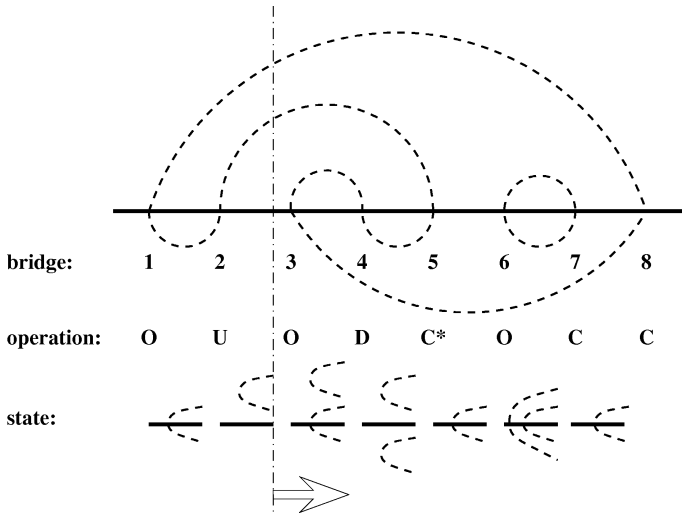


Fig. 13. A multi-road meander with 8 bridges and 2 connected components. Upon addition of a new bridge, four different operations (O, C, U and D) are allowed, as discussed in the text. We also show the state of that part of the meander that is to the left of a given vertical line perpendicular to the river (here in position after $t = 2$ steps). The transfer matrix acts by transferring this line from the left to the right.

the vertical line by listing the pairwise connectivities amongst the road segments as well as the position of the river.

Whenever the vertical line is transfered one step to the right, a new bridge is added and one out of four possible *operations* (O, C, U and D) can take place, as indicated on the figure. Each of these operations connects two consecutive states of the system, before and after the addition of the bridge. The operation O *opens* a new road segment on top of the river, thus connecting the two sides of the river. Similarly the operation C *closes* a road on top of the river. This operation comes in two variants, depending on whether the closed segment was already connected before the addition of the bridge (C) or not (C^*). In the former case, the road segment is erased and a non-trivial Boltzmann factor of q must be accounted for. In the latter case, the connectivity is transformed so as to connect the left-over partners of the two road segments that were eliminated. Finally, a road segment immediately below the river can move *up* (operation U), and a segment just above the river can move *down* (operation D).

In order to fully specify the transfer matrix of this problem we need to enumerate the possible connectivity states at a given time. The number of such states determines the size of the memory needed to store information at time $t = n$. We will therefore compute the number $F(n, N)$ of connectivity states after addition of the nth bridge for any $n = 1, 2, \ldots, 2N$ for the case of $2N$ bridges. Also, to implement the calculation of the meander partition function on a computer, one needs to *order* these states so that the entries of the transfer matrix can be accessed by means of a one-to-one mapping between the states and the set of integers $1, 2, \ldots, N_{\text{states}}$. Here N_{states} stands for the total number of states encountered in the whole transfer process from the left of the first bridge to the right of the last one. It is also the dimension of the explored state space and depends explicitly

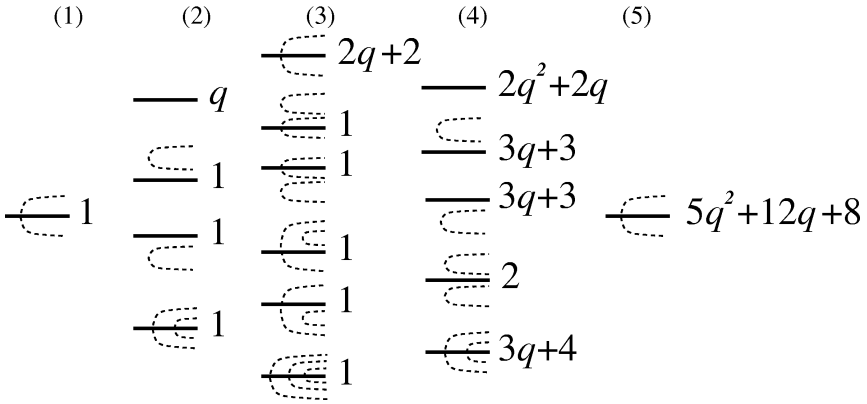


Fig. 14. Transfer matrix enumeration of all multi-road meanders with 6 bridges. The figure lists the complete set of intermediary states after the nth bridge ($n = 1, 2, \ldots, 5$) along with their respective weight.

on the number $2N$ of bridges added in the whole process. This number will be estimated below while the explicit ordering procedure will be presented at the end of this subsection.

Before we turn to the general case, let us illustrate the state counting for one infinite river with $2N = 6$ bridges. In Fig. 14, we detail the intermediate steps in the calculation of the corresponding meander polynomial $m_3(q)$. The figure depicts the complete set of states generated by the transfer process, when connecting the "empty" states (vacua) before bridge 1 and after bridge 6, along with their respective weights. Apart from the "empty" states $F(0, 6) = F(6, 6) = 1$, we read $F(1, 3) = 1$, $F(2, 3) = 4$, $F(3, 3) = 6$, $F(4, 3) = 5$, $F(5, 3) = 1$ and $N_{\text{states}} = 11$. The result of the calculation is the meander polynomial $m_3(q) = 5q^3 + 12q^2 + 8q$, indicating that with 6 bridges there are, respectively, 8, 12 and 5 meanders with 1, 2 and 3 connected components.

Dimension of the state space:

Let us now turn to the computation of the numbers $F(n, N)$. To this end we begin by relating these numbers to properties of a certain class of restricted Brownian walks. Consider a situation where there are p_1 (respectively, p_2) road segments above (respectively, below) the river, and where these segments are pairwise connected in such a way that exactly h arches cross the river. It is easy to check that the four operations O, C, D and U described above always shift p_1, p_2 and h by one unit, either $+1$ or -1, so that, by induction, n, p_1, p_2 and h have the same parity. Clearly, we also have $h \leqslant p_1$ and $h \leqslant p_2$ as an arch crossing the river connects a point above it to a point below it. There are then $(p_1 - h)/2$ (respectively, $(p_2 - h)/2$) non-crossing arches that stay above (respectively, below) the river. Now, while a crossing arch can be generated by means of a *single* bridge, by using the move O, the generation of a non-crossing arch necessitates at least *two* moves (O followed by either U or D). Therefore, twice the number of non-crossing arches plus the number of crossing arches cannot supersede the number of bridges added, i.e., $p_1 + p_2 - h \leqslant n$. In terms of h, the arch "height" above the river, this leads to the constraint:

$$\max(0, p_1 + p_2 - n) \leqslant h \leqslant \min(p_1, p_2). \tag{3.1}$$

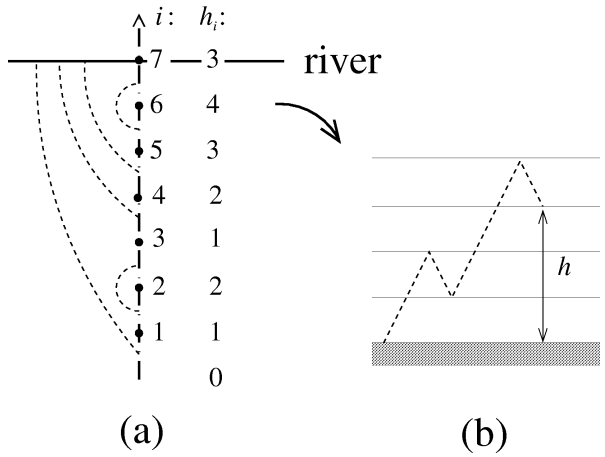


Fig. 15. A typical arch configuration (a) below the river and the corresponding walk (b). The walk has $p = 7$ steps and final height $h = 3$. At each step the height coordinate h_i on (b) changes by ± 1, according to whether an arch is opened or closed in (a).

Note that the above condition automatically implies that $p_i \leqslant n$, $i = 1, 2$.

The above constraint turns out to be the only one as long as $n \leqslant N$. For $n > N$ there are additional constraints, since we must always be able to annihilate any given state at level n in at most $2N - n$ moves so as to end up with the "empty" state at the right of the $2N$th bridge. Since p_1 and p_2 can decrease by at most one at each step, we thus have to impose for $n > N$ the two extra conditions:

$$p_1 \leqslant 2N - n, \qquad p_2 \leqslant 2N - n. \tag{3.2}$$

It is easy to estimate the number N_{states} of accessible states, i.e., those which satisfy the above constraints (3.1) and (3.2) for all $n = 1, \ldots, 2N$. An arch configuration, read from the bottom and upwards, can be mapped onto a $(p_1 + p_2)$-step Brownian walk, where the position h_i $(i = 0, \ldots, p_1 + p_2)$ of the walk starts from $h_0 = 0$ and is increased (respectively, decreased) by one unit for each opening (respectively, closing) of an arch, as illustrated on Fig. 15. Evidently these walks are constrained by $h_i \geqslant 0$ for all i. Decomposing the walk into its left part (describing the arches below the river) and its right part (describing the arches above the river) connected at height h, we are naturally lead to define $f(p, h)$ as the number of walks going from height 0 to height h (or equivalently that of walks going from height h to height 0) in p steps.

To evaluate $f(p, h)$ is a standard exercise. First note that without the constraint $h \geqslant 0$ this would simply read $\binom{p}{(p-h)/2}$ as me must choose the $(p - h)/2$ ascending steps. Each walk violating the constraint at some point hits $h_i = -1$. Let i be the first time this happens, and consider reversing each step from i and onwards. The result is a walk going from 0 to $(-1) - (h + 1)$, or equivalently from 0 to $h + 2$. Therefore:

$$f(p, h) = \binom{p}{(p - h)/2} - \binom{p}{(p - h)/2 - 1}. \tag{3.3}$$

In terms of $f(p, h)$, the number $F(n, N)$ of states accessible at level n can then be written as

$$F(n,N) = \sum_{p_1=0}^{\min(n,N-n)} \sum_{p_2=0}^{\min(n,N-n)} \sum_{h=\max(0,p_1+p_2-n)}^{\min(p_1,p_2)} f(p_1,h) f(p_2,h)$$
$$= \sum_{h=0}^{n} \sum_{p_1=h}^{\min(n,2N-n)} \sum_{p_2=h}^{\min(n+h-p_1,2N-n)} f(p_1,h) f(p_2,h), \tag{3.4}$$

where it is implicitly understood that h, p_1 and p_2 in the sums above all have the parity of n.

When $N \geqslant n$ it is possible to rewrite Eq. (3.4) in a simpler form, since the second constraint in the upper limit of the summations over p_1 and p_2 (namely $p_i \leqslant N-n$) does not come into play. Setting $p_1 + p_2 = 2p$, we simply have to enumerate all the set of p arches, cut by the river at a height h that has the same parity as n and satisfies the constraint $h \geqslant 2p - n$. Let us define

$$h_{\min}(n,p) \equiv \begin{cases} \max(0, 2p-n), & \text{for } n \text{ even}, \\ \max(1, 2p-n), & \text{for } n \text{ odd}. \end{cases} \tag{3.5}$$

In the walk language, we have to count all the walks of length $2p$ that stay non-negative with a *marking* at a point of height $h \geqslant h_{\min}$ and with a well defined parity, that of n. The total number of non-negative walks of length $2p$ going from 0 to 0 and with a marked point at position h can be easily calculated to be:

$$g(p,h) = f(2p,2h) + f(2p,2h+2) = \binom{2p}{p-h} - \binom{2p}{p-h-2}. \tag{3.6}$$

Summing over the heights $h \geqslant h_{\min}(n,p)$ having the parity of n (which is also that of $h_{\min}(n,p)$) we have

$$\sum_{\substack{h=h_{\min}(n,p),\\ h=h_{\min}(n,p) \bmod 2}}^{n} g(p,h) = \binom{2p}{p-h_{\min}(n,p)}. \tag{3.7}$$

The complete number of states used in the transfer matrix is now simply obtained by summing over the number of arches. Depending on the parity of n, we get:

$$F(n, N \geqslant n) = \sum_{p=0}^{n} \binom{2p}{\min(p, n-p)}, \quad \text{for } n \text{ even}, \tag{3.8}$$

and similarly:

$$F(n, N \geqslant n) = \sum_{p=1}^{n} \binom{2p}{\min(p-1, n-p)}, \quad \text{for } n \text{ odd}. \tag{3.9}$$

Note that the above expressions *do not depend* on N. This is because we assumed that $n \leqslant N$, in which case the second constraint (3.2) is ineffective. For large n, the above expressions for $F(n, N \geqslant n)$ can be evaluated by a simple saddle-point approximation. Setting $p = yn$, we get a saddle point at a value $y^\star$ solution of

$$(3y^\star - 1)^3 = (1 - y^\star)(2y^\star)^2, \tag{3.10}$$

Table 1
Number of intermediate states employed by the transfer matrix algorithm for multi-road meanders with $2N$ bridges. The number of states $F(N,N)$ after addition of the Nth bridge is less than the maximal number of states $F(n_{\max},N)$, which occurs at a value $n_{\max}$ slightly above N

N	$F(N,N)$	$F(n_{\max},N)$	$n_{\max}$	N	$F(N,N)$	$F(n_{\max},N)$	$n_{\max}$
1	1	1	1	13	12905	26770	15
2	4	4	2	14	27971	62959	16
3	6	6	3	15	59282	155153	18
4	16	16	4	16	128130	388695	19
5	29	29	5/6	17	272610	950128	20
6	68	68	6	18	588153	2279273	21
7	134	161	8	19	1254586	5733997	23
8	300	363	9	20	2703503	14523043	24
9	614	846	10	21	5777115	35946838	25
10	1349	1890	11	22	12438708	87192966	26
11	2813	4579	13	23	26613942	223196395	28
12	6126	11216	14	24	57268474	568622062	29

namely with the numerical value $y^\star = 0.611491992\ldots$. This value indicates that the state statistics is indeed governed by arch systems with a number $p = p_1 = p_2 = p^\star \equiv y^\star n$ of arches. The quantity $F(n, N \geqslant n)$ is then found to grow asymptotically as a^n, where

$$a = \frac{(2y^\star)^{2y^\star}}{(1-y^\star)^{1-y^\star}(3y^\star-1)^{3y^\star-1}}, \tag{3.11}$$

which, using (3.10), is also the solution of

$$a^3 = (1+a)^2. \tag{3.12}$$

This yields the numerical value $a = 2.147899036\ldots$.

In Table 1, we give some explicit values of the number of states needed when enumerating meanders with $2N$ bridges. It is seen that the maximum number of states occurs slightly after the addition of the middle bridge, $n = N$. For large N, one can easily estimate the value $n_{\max}$ of n where this maximum occurs. Indeed, for $n > N$, the second condition (3.2) starts to play a role. As we have seen, for large n, the state statistics without the second constraint is dominated by arch configurations with $p_1 = p_2 = p^\star = y^\star n$. Therefore, we expect this second constraint (3.2) to effectively affect the asymptotic behavior and start reducing the number of states whenever $p^\star = 2N - n$, i.e., for $n = 2N/(1+y^\star)$. Assuming that this precisely corresponds to the step having the maximum number of states, we have

$$n_{\max} = 2N/(1+y^\star) = 1.241085907\ldots N. \tag{3.13}$$

We can then estimate the asymptotic number of encountered states in the whole process to grow like:

$$N_{\text{states}} \sim F(n_{\max},N) \sim a^{n_{\max}} = (2.582603447\ldots)^N. \tag{3.14}$$

The estimates (3.13) and (3.14) agree with the values observed in Table 1.

Ordering the states:

According to Eq. (3.4), the full set of $F(n, N)$ states can be ordered if we know how to order the $f(p, h)$ restricted Brownian walks defined above. Namely, in that case one can order the complete set of states lexicographically after h, p_1, p_2, the value of the first Brownian walk $[1, 2, \ldots, f(p_1, h)]$, and finally the value of the second Brownian walk $[1, 2, \ldots, f(p_2, h)]$. By this we mean simply that state A precedes state B if $h_A < h_B$. If $h_A = h_B$ this criterion is inconclusive, and one compares the values of p_1, so that A precedes B if $p_{1,A} < p_{1,B}$. In case of further equality one proceeds to compare p_2, and so on.

To order the Brownian walks, consider first the example contributing to $f(7, 3)$ shown in Fig. 15. The idea is to obtain another formula for $f(p, h)$ which will in turn allow us to define a recursive ordering of the walks. We start by focusing on the first (lowermost) arch. Either this arch is open (as is the case on the figure), or it closes at some other point before p. In the first case the remaining arch configuration is a contribution to $f(p-1, h-1)$, and in the second the arches inside the first arch are independent of the arches above its termination point. We therefore have

$$f(p, h) = f(p-1, h-1) + \sum_{k=1,3,5,\ldots} f(k-1, 0) f(p-k-1, h), \tag{3.15}$$

with $f(0, h) = \delta_{0,h}$, and $f(p, h) = 0$ for $p < 0$. This is the required formula.

A walk contributing to $f(p, h)$ can now be recursively ordered, first by considering the termination point of the lowermost arch (which by definition is infinity if that arch is open), then by (recursively) considering the ordering of the smaller arch system inside the first arch, and finally by (recursively) considering the ordering of the arch system above the first arch (which only exists if that arch is closed). The procedure just described generalizes the ordering of the Catalan connectivities $c_{p/2} = f(p, 0)$ given in Ref. [33].

3.2. Other geometries

We now come to the case of the more involved meandric geometries encountered in Section 2.

Two infinite rivers:

It is possible to generalize the multi-road one-river transfer matrix to the case of several rivers, provided that the latter can be deformed into a system of parallel lines that are only connected among themselves at infinity. For simplicity we consider in the following the case of two such infinite rivers. As we have seen, this situation also corresponds to the deformation of the figure-eight configuration shown in Fig. 9, provided that a marked point is added on each river.

As before, the crucial point is to have a well-defined transfer direction, which we simply take to be parallel to both rivers. For any given configuration with $2N$ bridges we write $N = N_1 + N_2$, where $2N_1$ (respectively, $2N_2$) is the number of times the roads cross the first (respectively, the second) river. Note that contrary to the bridges on the same river which

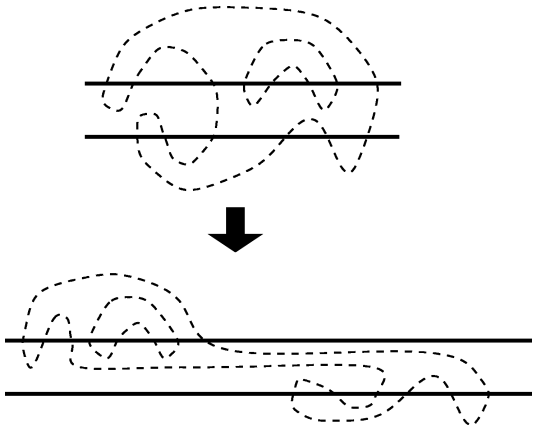


Fig. 16. The deformation of a two-river configuration by sending all the bridges crossing the upper river to the left and all the bridges crossing the lower river to the right. The transfer matrix acts by first adding the left bridges, then the right ones.

are naturally ordered, bridges on the first and on the second river are not naturally ordered with respect to one another. To avoid double counting in the transfer matrix approach, we can simply deform the roads as in Fig. 16 so as to send all the bridges of river 1 to the left and all those of river 2 to the right. In the transfer process, we thus add first the $2N_1$ bridges crossing the first river and then the $2N_2$ bridges crossing the second river.

Given a decomposition $N = N_1 + N_2$, the calculations therefore proceed exactly as in the one-river case, except that when both $N_1 > 0$ and $N_2 > 0$ not *all* of the connectivity states described in Section 3.1 come into use. Denoting by p_1 the number of road segments above the first river, by p_2 the road segments in between rivers, and by p_3 the segments below the second river, we now have the following constraints after the addition of the nth bridge:

$$\begin{aligned} p_1 &\leqslant \max\bigl(\min(n, 2N_1 - n), 0\bigr), \\ p_2 &\leqslant \min(n, 2N - n), \\ p_3 &\leqslant \max\bigl(\min(n - 2N_1, 2N - n), 0\bigr). \end{aligned} \tag{3.16}$$

The total two-river meander polynomial is obtained by summing over the possible decompositions. We can furthermore simply address the situation with a marked point on either river by weighing each term in the decomposition by a factor of $(2N_1+1)(2N_2+1)$.

One semi-infinite river (semi-meanders):

Finally, we have examined systems of semi-meanders, where the roads are allowed to wind around the source of a semi-infinite river. Each winding number w can be examined separately, by choosing an initial state of w arches nested inside one another, and simply applying the four operations discussed in Section 3.1. Unfortunately in this situation we have not found a simple and full description of the generated states (apart from $w = 0$ and $w = 1$). It is however still possible to carry out the transfer matrix calculations, by simply inserting the generated states in an unordered list. The price to be paid is that to find a given state one will have to sequentially search through the entire list, leading to a pitiful waste

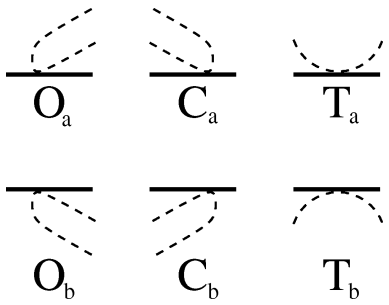


Fig. 17. The six new operations allowed for tangent meanders.

of computation time. This is nevertheless what we have done, and consequentially we have had to content ourselves with smaller system sizes.

3.3. Tangency points

We have also enumerated systems of *tangent* meanders, where apart from the four types of moves at each bridge described in Section 3.1, we allow for the six additional operations depicted in Fig. 17. These moves allow to *open* a road whilst staying *above* or *below* the river (O_a and O_b), to similarly *close* a road without crossing the river (C_a and C_b), and finally to *touch* the river whilst staying *above* it or *below* it (T_a and T_b). The meander polynomial is now defined by assigning a weight x to each of the crossing bridges and a weight y to each of the tangent vertices.

Here again, we have not been able to find an explicit way of enumerating and ordering the states obtained by applying this set of ten operations to the vacuum.

4. Numerical results

Let us now present our numerical results obtained by use of the algorithms presented in Section 3, together with an appropriate extrapolation procedure. Section 4.1 discusses multi-component meanders and introduces the extrapolation method. Section 4.2 is devoted to the case of two parallel rivers, while Section 4.3 addresses semi-meanders. Finally, we study tangent meanders in Section 4.4.

4.1. One infinite river (meanders)

We have enumerated multi-connected meanders with one infinite river and a *fixed fugacity* q up to $N = 24$, i.e., $2N = 48$ bridges for $q = 0$, $\sqrt{2}$, $\sqrt{3}$ and 2. In other words, we evaluated the quantities $M_N = \lim_{q\to 0}(m_N(q)/q)$, $m_N(\sqrt{2})$, $m_N(\sqrt{3})$ and $m_N(2)$ for $N = 1, \ldots, 24$. From these numbers, we can extract estimates for the large N "activity" per bridge $R(q)$ and configuration exponent $\alpha(q)$ defined as in Eq. (2.11) by

$$m_N(q) \sim C(q)\frac{R(q)^{2N}}{N^{\alpha(q)}}. \tag{4.1}$$

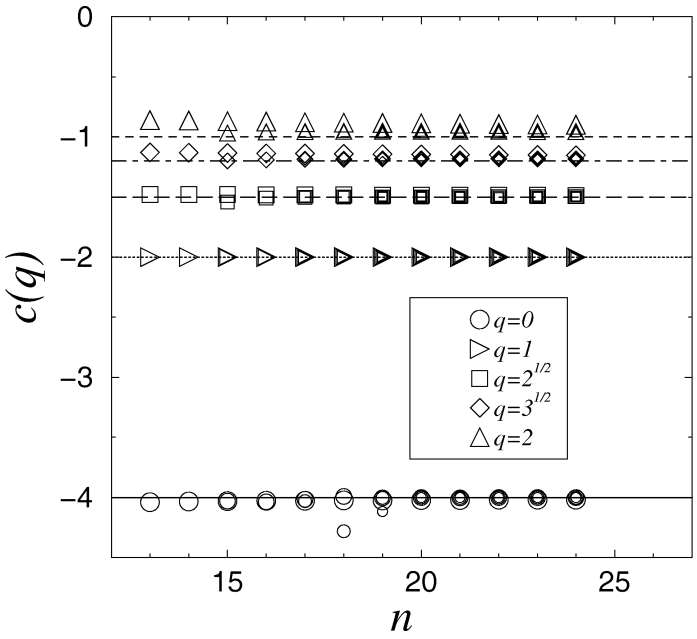


Fig. 18. The estimated central charge obtained from the values $\alpha_N^{(p)}(q)$ with $p = 1, 2, 3$ and 4 iterations. The size of the symbols decreases with the number of iterations. Each value is represented at an abscissa n corresponding to the largest index $N = n$ of $m_N(q)$ used in the estimate. We also draw horizontal lines at the predicted values of $c(q)$.

The estimates for $\alpha(q)$ can be transformed into estimates for the central charge $c(q)$ through the following relation, inverting (2.17) for $\gamma = 2 - \alpha(q)$:

$$c(q) = 1 - 6\frac{(2-\alpha(q))^2}{(\alpha(q)-1)}. \tag{4.2}$$

Starting from the zeroth order values:

$$\begin{aligned} R_N^{(0)}(q) &\equiv \sqrt{\frac{m_{N+1}(q)}{m_N(q)}}, \\ \alpha_N^{(0)}(q) &\equiv (N+1)^2\left(\frac{m_{N+2}(q)m_N(q)}{(m_{N+1}(q))^2} - 1\right), \end{aligned} \tag{4.3}$$

we can build better estimates $R_N^{(p)}$ and $\alpha_N^{(p)}$ by a recursive use of a standard convergence acceleration procedure:

$$R_N^{(p)} = R_N^{(p-1)} - \frac{p+1}{p}\,\frac{(R_{N+1}^{(p-1)} - R_N^{(p-1)})(R_N^{(p-1)} - R_{N-1}^{(p-1)})}{R_{N+1}^{(p-1)} + R_{N-1}^{(p-1)} - 2R_N^{(p-1)}} \tag{4.4}$$

and similarly for $\alpha_N^{(p)}$. The procedure implicitly assumes that the corrections to the asymptotic scaling (4.1) are regular, i.e., integer powers of $1/N$, in which case it guarantees that $R_N^{(p)}(q) = R(q) + \mathcal{O}(1/N^{p+1})$.

Fig. 18 shows the estimated central charge obtained through (4.2) using $\alpha_N^{(p)}(q)$ for $p = 1, 2, 3, 4$, together with the expected central charge according to (2.4) with $n_1 = 0$ and

Table 2
Numerical estimates for the meander configuration exponent $\alpha(q)$, the activity $R^2(q)$ per pair of bridges, and the central charge $c(q)$. The error is implicitly on the last digit. The corresponding theoretical values are also listed

q	$\alpha_{\text{theor}}(q)$	$\alpha_{\text{num}}(q)$	$R^2_{\text{num}}(q)$	$c_{\text{theor}}(q)$	$c_{\text{num}}(q)$
0	$\frac{29+\sqrt{145}}{12} = 3.4201328\ldots$	3.4207	12.26286	-4	-4.003
1	3	3.00000	16.000000	-2	-2.0000
$\sqrt{2}$	$\frac{53+\sqrt{265}}{24} = 2.8866175\ldots$	2.885	17.52468	$-3/2$	-1.496
$\sqrt{3}$	$\frac{131+\sqrt{1441}}{60} = 2.8160084\ldots$	2.812	18.68970	$-6/5$	-1.18
2	$\frac{13+\sqrt{13}}{6} = 2.7675918\ldots$	2.75	19.669	-1	-0.95

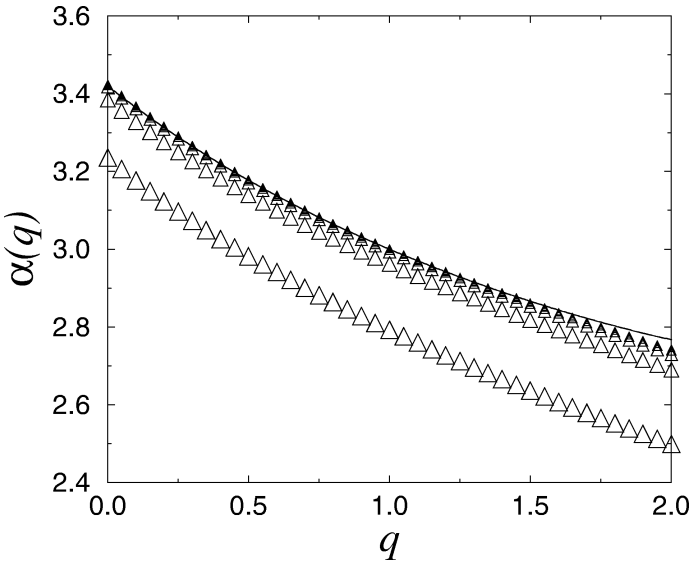


Fig. 19. Numerical estimates (triangles) for the meander configuration exponent $\alpha(q)$ for $0 \leqslant q \leqslant 2$, together with the predicted value (2.35)–(2.32) (solid line). We display the result of several iterations of our convergence acceleration procedure. The size of the symbols decreases with the number of iterations.

$n_2 = q$. For completeness, we also show the corresponding estimates obtained for $q = 1$ with $m_N(1) = (c_N)^2$ in terms of the Catalan numbers c_N. The corresponding quantitative values are displayed in Table 2.

We have also computed the numbers $M_N^{(k)} = \mu_{2N,0}^{(k)}$ of meanders with one infinite river, $2N$ bridges and a fixed number k of connected components ($k = 1, \ldots, N$) for $N = 1, \ldots, 20$ (40 bridges). As an illustration we display the results for $N = 20$ in Table 3.

From these values, we can extract the polynomials $m_N(q) = \sum_{k=1}^{N} M_N^{(k)} q^k$, hence the values $R(q)$ and $\alpha(q)$ for a *varying fugacity* q. These values are displayed in Fig. 19 together with the prediction (2.35)–(2.32).

Table 3
Multi-component meander numbers $M_N^{(k)}$ with $2N = 40$ bridges and k connected components of road, $k = 1, 2, \ldots, 20$

k	$M_{20}^{(k)}$	k	$M_{20}^{(k)}$
1	64477712119584604	11	706958959835806990
2	511563350415103008	12	235265604762448572
3	1901345329566422790	13	64713641205591820
4	4405839231880790648	14	14658557362753320
5	7145814923879522986	15	2709804590263296
6	8632733743310196256	16	402058856155712
7	8070705247685170684	17	46500885666900
8	5988061883039308848	18	3978168316200
9	3587066097601934530	19	226760523600
10	1755310029771295216	20	6564120420

All our estimates clearly validate the theoretical predictions with a very good accuracy. We note a small discrepancy for values of q close to $q = 2$. This corresponds to the regime which has the worst convergence of our acceleration procedure. This poor convergence might be due to either corrections which are not regular, as implicitly assumed, or to the vicinity of a transition at $q = 2$ where the DPL^2 model stops being critical. We anyway impute this small discrepancy to our estimation procedure.

4.2. Two infinite rivers

Beyond the central charge, we can also test the operator content of the theory. A first check concerns the dimension Δ_4 of the operator creating a vertex with four outgoing river segments, which, as we already explained, can be measured through the configuration exponent $\alpha_{2\text{-mark}}$ for two parallel infinite rivers with a marked point on each river. As we mentioned before, in the absence of marking, we expect the two-river configuration exponent $\alpha_{2\text{-riv}}(q)$ to be simply the same as that of meanders with a single infinite river.

Fig. 20 and 21 show our results based on enumerations with up to $2N = 42$ bridges for fixed $q = 0$, $\sqrt{2}$, $\sqrt{3}$ and 2. For completeness, we also added estimates for $q = 1$, as extracted from the exact expressions:

$$m_N^{2\text{-riv}}(q = 1) = \sum_{p=0}^{N} c_p c_{N-p} c_N = c_N c_{N+1} \tag{4.5}$$

for two un-marked rivers, and

$$\begin{aligned} m_N^{2\text{-mark}}(q = 1) &= \sum_{p=0}^{N} (2p + 1)\big(2(N - p) + 1\big) c_p c_{N-p} c_N \\ &= c_N \left(4^{N+1} - \binom{2n+3}{n+1} \right) \end{aligned} \tag{4.6}$$

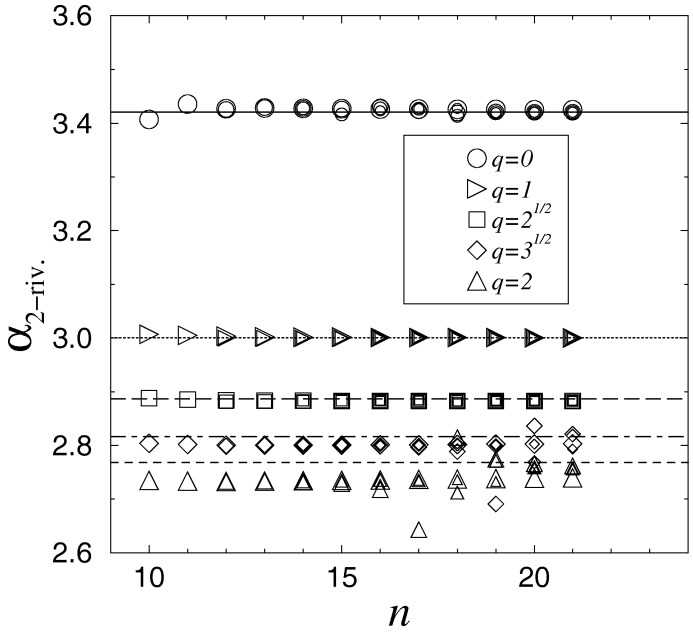


Fig. 20. The estimated configuration exponent for two *un-marked* rivers, obtained with $p = 1, 2, 3$ and 4 iterations of the acceleration procedure (4.4). The size of the symbols decreases with the number of iterations. Each value is represented at an abscissa n corresponding to the largest index $N = n$ of $m_N^{2\text{-riv}}(q)$ used in the estimate. We also draw horizontal lines at the predicted values of $\alpha_{2\text{-riv}}(q)$.

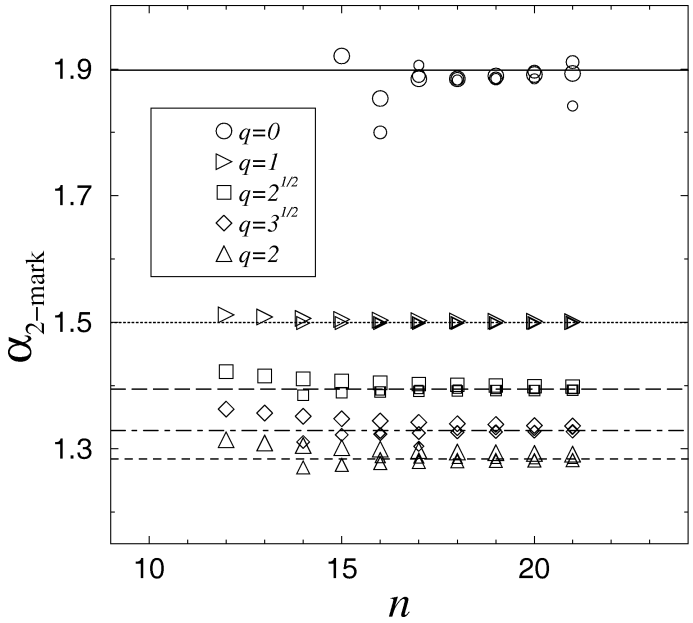


Fig. 21. The estimated configuration exponent for two *marked* rivers, obtained with $p = 2, 3$ and 4 iterations of the (modified, as explained in the text) acceleration procedure. The size of the symbols decreases with the number of iterations. Each value is represented at an abscissa n corresponding to the largest index $N = n$ of $m_N^{2\text{-mark}}(q)$ used in the estimate. We also draw horizontal lines at the predicted values of $\alpha_{2\text{-mark}}(q)$.

Table 4
Numerical estimates for the configuration exponent and the activity $R^2(q)$ per pair of bridges in the case of two un-marked rivers (upper half) and two marked rivers (lower half). The error is implicitly on the last digit. The corresponding theoretical values are also listed

q	$\alpha_{2\text{-riv}}(q)$	$\alpha_{\text{num}}(q)$	$R^2_{\text{num}}(q)$
0	$\frac{29+\sqrt{145}}{12} = 3.4201328\ldots$	3.4205	12.2627
1	3	3.00000	16.000000
$\sqrt{2}$	$\frac{53+\sqrt{265}}{24} = 2.8866175\cdots$	2.882	17.5246
$\sqrt{3}$	$\frac{131+\sqrt{1441}}{60} = 2.8160084\cdots$	2.80	18.688
2	$\frac{13+\sqrt{13}}{6} = 2.7675918\ldots$	2.75	19.663
q	$\alpha_{2\text{-mark}}(q)$	$\alpha_{\text{num}}(q)$	$R^2_{\text{num}}(q)$
0	$\frac{\sqrt{5}+\sqrt{14}}{\sqrt{29}-\sqrt{5}} = 1.8982348\ldots$	1.89	12.26
1	3/2	1.499	15.9999
$\sqrt{2}$	$\frac{\sqrt{5}+\sqrt{23}}{\sqrt{53}-\sqrt{5}} = 1.3941001\ldots$	1.393	17.527
$\sqrt{3}$	$\frac{\sqrt{11}+\sqrt{56}}{\sqrt{131}-\sqrt{11}} = 1.3285858\ldots$	1.33	18.69
2	$\frac{\sqrt{11}+\sqrt{2}}{\sqrt{26}-\sqrt{2}} = 1.2838772\ldots$	1.28	19.67

for two marked rivers. Note that in this latter case of two marked rivers, the leading contribution $m_N^{2\text{-mark}}(1) \sim (16)^N/N^{3/2}$ is corrected by a sub-leading contribution $\sim (16)^N/N^2$, with an exponent displaced by a half-integer. In this case, the corrections to the leading term thus involve generically half-integers instead of integers. We expect this phenomenon to be generic for all values of q with a double expansion involving both the exponent $\alpha_{2\text{-mark}}(q)$ and its "descendants" shifted by integers, and a sub-leading correction with the exponent $\alpha_{2\text{riv}}(q) - 1$ and its descendants. Since $\alpha_{2\text{-riv}} - 1 - \alpha_{2\text{-mark}}$ is always close to $1/2$ for the values of q of interest, we decided to use a modified acceleration procedure assuming half-integer corrections to scaling (the index p in (4.4) now takes half-integer values). The estimates of Table 4 and Fig. 21 have been obtained with this modified procedure.

Here again, the numerical values corroborate our predictions for the various configuration exponents. Note that we find that the activity per bridge is for these two cases the same as for a single infinite river. It should not come as a surprise that the activity $R(q)$ per bridge is independent of the particular meander geometry since it is a bulk quantity insensitive to the choice of boundary conditions (river shapes). It is also interpreted as the inverse of the convergence radius common to all the generating functions of the numbers at hand, also viewed as various correlators within the same theory.

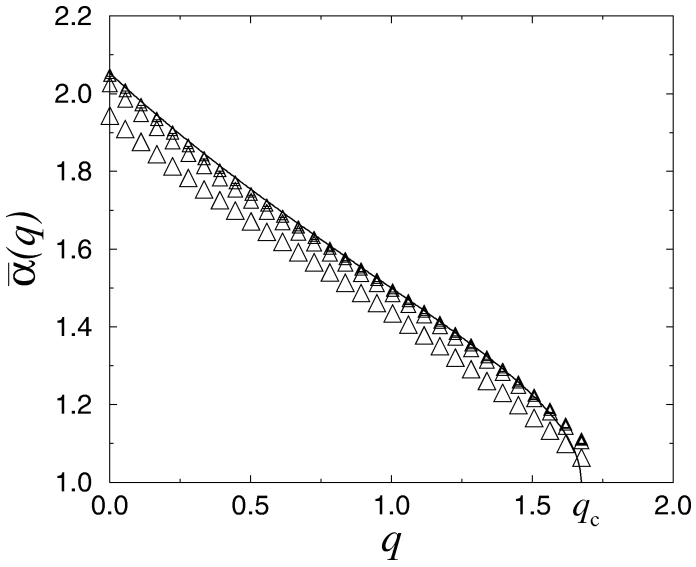


Fig. 22. Numerical estimates for the semi-meander configuration exponent $\bar{\alpha}(q)$ for $0 \leqslant q \leqslant q_c$, together with the predicted value (2.37)–(2.32). We display the result of several iterations of the acceleration procedure. The size of the symbols decreases with the number of iterations.

4.3. One semi-infinite river (semi-meanders)

As a second check of the operator dimensions of the theory, we also considered the case of semi-meanders with a single semi-infinite river ($n_1 = 0$) and a arbitrary number k of connected components of road. Our estimates rely on an exact enumeration of the numbers $\overline{M}_N^{(k)}(w)$ of semi-meanders with N bridges, k roads and winding number w for $1 \leqslant N \leqslant 33$, $1 \leqslant k \leqslant N$, $0 \leqslant w \leqslant N$ and $w = N \bmod 2$. For illustration, we list in Table 5, respectively, the numbers $\overline{M}_{33}^{(k)} = \sum_w \overline{M}_{33}^{(k)}(w)$ of semi-meanders with $N = 33$ bridges and k roads (irrespectively of the winding numbers) as well as the numbers $\overline{M}_{33}^{(1)}(w)$ of connected semi-meanders with winding number $w = 1, 3, 5, \ldots, 33$.

For this particular geometry, a clear parity effect occurs and it is thus important to make numerical evaluations from semi-meander numbers with either even or odd N to avoid large numerical errors. The results presented in Fig. 22 were extracted from even values of N. Table 6 lists quantitative values for the configuration exponent and the activity per pair of bridges. Here again, we note that the activity per bridge is identical to that of the other geometries. As shown in Fig. 22, the configuration exponent is in very good agreement for q less than 1.5. For larger values of q, a small discrepancy appears, again imputable to a poor convergence probably due to the vicinity of a transition point, now at $q = q_c$ where we expect a proliferation of connected components of road for the semi-meander geometry.

4.4. Tangent meanders

Finally, we also checked our assertion that the "tangency" vertices of type b (see Fig. 1) are irrelevant. We computed the canonical partition function $\mu_N(q; x, y)$ defined in (2.10)

Table 5
Multi-component semi-meander numbers $\overline{M}_N^{(k)}$ with $N = 33$ bridges and k connected components of road, $k = 1, 2, \ldots, 33$, and fixed winding connected semi-meander numbers $\overline{M}_N^{(1)}(w)$ with $N = 33$ and $w = 1, 3, 5, \ldots, 33$

k	$\overline{M}_{33}^{(k)}$	k	$\overline{M}_{33}^{(k)}$
1	455792943581400	18	3659252585228
2	3285874327160852	19	1040041120124
3	11119764476127424	20	279039302088
4	23598381333433844	21	70513532334
5	35436190513634790	22	16729859124
6	40334792072264540	23	3710923316
7	36464182713722576	24	765325248
8	27136413723456560	25	145710912
9	17127401092409102	26	25374900
10	9409371247346540	27	3992846
11	4602479751584184	28	558396
12	2042918178657320	29	67804
13	835326688674886	30	6904
14	318096906554664	31	557
15	113643690324368	32	32
16	38261586556480	33	1
17	12168938393766		

w	$\overline{M}_{33}^{(1)}(w)$	w	$\overline{M}_{33}^{(1)}(w)$
1	59923200729046	19	16277801502
3	121544501379440	21	1326698396
5	125267070807626	23	73827420
7	85716694410306	25	2638462
9	42336073574012	27	55052
11	15599486790514	29	568
13	4337132101822	31	2
15	908960663970	33	0
17	142142103262		

for the particular case of tangent meanders with a single connected component of road ($q = 0$), a fixed weight $x = 1$ for "crossing" vertices and with several values of the weight y per "tangency" vertex. We enumerated up to $N = 24$ vertices for $y = 0.5$, 1 and 2 while $y = 0$ simply corresponds to the meander configurations of Section 4.1 for which we have results up to 48 vertices.

Our results are represented in Fig. 23 and the corresponding quantitative values listed in Table 7. We find a clear evidence that the configuration exponent is independent of y, indicating that the universality class of tangent meanders and that of meanders are the same, as predicted in Section 2.3.

Table 6
Numerical estimates for the configuration exponent and the activity $R^2(q)$ per pair of bridges in the case of one semi-infinite river (semi-meanders). The error is implicitly on the last digit. The corresponding theoretical values are also listed

q	$\bar{\alpha}_{\text{theor}}(q)$	$\bar{\alpha}_{\text{even}}(q)$	$\bar{\alpha}_{\text{odd}}(q)$	$\overline{R}^2_{\text{even}}(q)$	$\overline{R}^2_{\text{odd}}(q)$
0	$1+\frac{\sqrt{11}}{24}(\sqrt{29}+\sqrt{5})$ $=2.0531987\ldots$	2.0532	2.051	12.26287	12.2626
1	3/2	1.500000	1.500000	16.000000	16.000000
$\sqrt{2}$	$1+\frac{\sqrt{2}}{48}(\sqrt{53}+\sqrt{5})$ $=1.2803730\ldots$	1.282	1.282	17.5247	17.5247

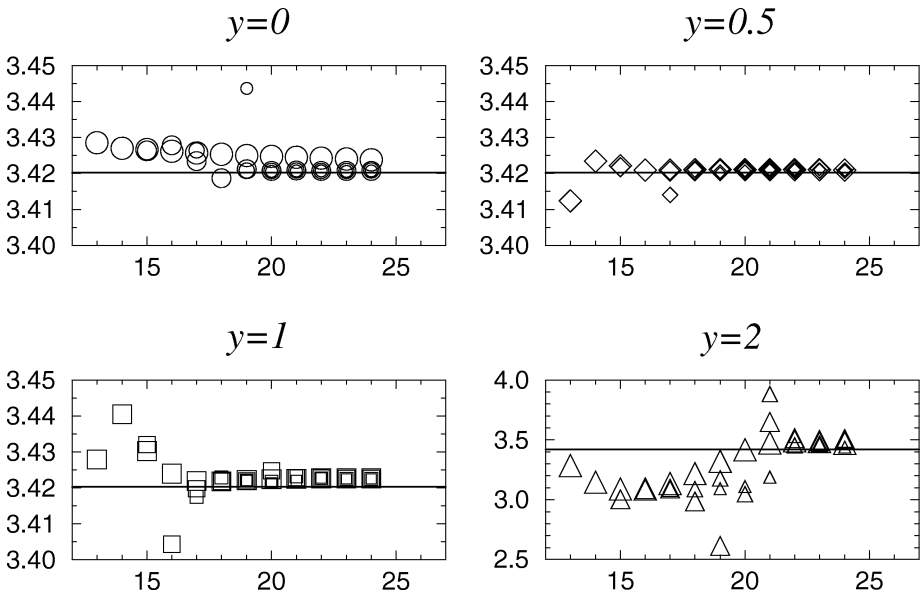


Fig. 23. The estimated configuration exponent α for tangent meanders with a weight $y = 0$, 0.5, 1 and 2 per tangency vertex. The size of the symbols decreases with the number p of iterations (here $p = 1, 2, 3, 4$). Each value is represented at an abscissa n corresponding to the largest index $N = n$ for the number of vertices (respectively, the number of bridge pairs in the case $y = 0$) used in the estimate. We also draw horizontal lines at the predicted value.

5. Discussion and conclusion

In this paper, we have presented theoretical and numerical evidence for a number of exact meandric configuration exponents. Although the sequence of physical arguments leading to the exact predictions is by no means a mathematical proof, the numerical evidence is compelling. Let us add a few comments on possible generalizations of our work.

Table 7
Numerical estimates for the configuration exponent and the activity $R(q)$ per bridges in the case of tangent meanders with $q = 0$, $x = 1$ and $y = 0$, 0.5, 1 and 2. The error is implicitly on the last digit. The corresponding theoretical value is also listed

y	α_{theor}	$\alpha_{\text{num}}(y)$	$R_{\text{num}}(y)$
0	$\frac{29+\sqrt{145}}{12} = 3.4201328\ldots$	3.4207	3.50184
1/2	$\frac{29+\sqrt{145}}{12} = 3.4201328\ldots$	3.4208	6.188
1	$\frac{29+\sqrt{145}}{12} = 3.4201328\ldots$	3.422	8.735
2	$\frac{29+\sqrt{145}}{12} = 3.4201328\ldots$	3.4	13.63

5.1. Ranges of validity and transitions

The physical arguments presented above allow to go much farther than just the case of one or two rivers. Indeed, we have seen that meanders are a particular case of the $\text{GFPL}^2(n_1, n_2)$ model at $n_1 = n_2 = 0$ with $y = 0$. Let us define multi-river and multi-road meander polynomials of order n, $m_n(n_1, n_2)$, by the following expansion of the partition function Z_{GFPL} (2.5) for $y = 0$

$$Z_{\text{GFPL}}(n_1, n_2; x) = \sum_{n \geqslant 1} \frac{x^{2n}}{4n} m_n(n_1, n_2). \tag{5.1}$$

Then we can as well predict the following large n asymptotic behavior

$$m_n(n_1, n_2) \sim C(n_1, n_2) \frac{R(n_1, n_2)^{2n}}{n^{\alpha(n_1, n_2)}}, \quad \alpha(n_1, n_2) = 2 - \gamma\big(c(n_1, n_2)\big), \tag{5.2}$$

where $c(n_1, n_2)$ is the central charge of the $\text{DPL}^2(n_1, n_2)$ model (2.4), and $\gamma(c)$ is as in (2.17). The general prediction (5.2) was actually proved in the case $n_1 = 1$ and arbitrary $0 \leqslant n_2 \leqslant 2$ in [15], by solving a particular matrix model.

Note however that the range of validity of (2.17) imposes that $c(n_1, n_2) \leqslant 1$. This gives the total range of validity of Fig. 24. Outside of this range, we must consider the two following cases:

(i) One of the $n_i > 2$: then the DPL^2 model is no longer critical, and we have no field theoretical prediction as to the value of the configuration exponent. It has been shown however, using matrix model techniques, that in the resembling case of the so-called $\text{O}(n)$ model, the critical exponent takes a constant value $\alpha = 3/2$ independently of $n > 2$ [16].

(ii) Both $n_1, n_2 \leqslant 2$ but $c(n_1, n_2) > 1$: one encounters the well-known "$c = 1$ barrier" phenomenon in two-dimensional quantum gravity [35,36]. The corresponding theories are dominated by configurations of surfaces with long fingers (branched polymer phase of quantum gravity). Remarkably, it was shown that $\gamma = 1/2$ throughout this phase, leading also to a constant exponent $\alpha = 3/2$.

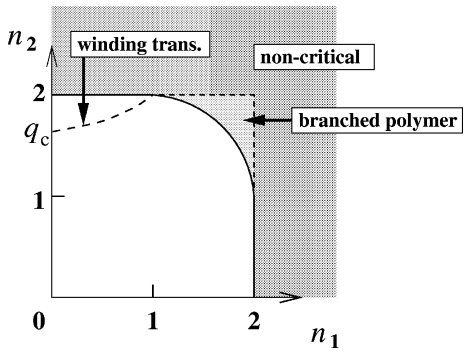


Fig. 24. The range of validity of the exact prediction for the meander configuration exponent in the general case of a weight n_1 per river and n_2 per road. In addition to the criticality constraint that $0 \leqslant n_i \leqslant 2$, we have represented the $c(n_1, n_2) = 1$ curve (joining the points $(n_1 = 2, n_2 = 1)$ and $(n_1 = 1, n_2 = 2)$, and passing by $(n_1 = \sqrt{2}, n_2 = \sqrt{2})$), beyond which the meanders are in a branched polymer phase. We have also represented by a dashed line the location of the winding transition joining the point $(0, q_c)$ to the point $(1, 2)$.

Hence we expect (for different reasons) that $\alpha = 3/2$ identically outside of the range of Fig. 24, for $n_1, n_2 \geqslant 0$. On the ordinate axis of Fig. 24, where $n_1 = 0$, we recover for meanders the announced transition at $n_2 = q = 2$, beyond which the DPL^2 model is no longer critical; numerical results however get poorer as we approach this point and we could not confidently analyze it numerically so far. In the case of semi-meanders, we have predicted an earlier "winding" transition point at $n_2 = q_c = 2\cos(\pi(\sqrt{97} - 1)/48)$. More generally, the configuration exponent for multi-component and multi-river semi-meanders (one half-line plus an arbitrary number of river loops) is predicted to be

$$\begin{aligned}\bar{\alpha}(n_1, n_2) &= \alpha(n_1, n_2) - 1 + 2\Delta_1(n_1, n_2) \\ &= 1 + \frac{1}{24}\big(\sqrt{25 - c(n_1, n_2)} + \sqrt{1 - c(n_1, n_2)}\big) \\ &\quad \times \sqrt{6(1 - e_1) - 4c(n_2)},\end{aligned} \tag{5.3}$$

where $c(n_2) = 1 - 6e_2^2/(1 - e_2)$ and $n_i = 2\cos\pi e_i$. The range of validity of (5.3) is smaller than the domain of Fig. 24, as it is delimited by the curve $3(1 - e_1)/2 + 6e_2^2/(1 - e_2) = 1$, with $n_i = 2\cos\pi e_i$ (it is represented in dashed line on Fig. 24). The latter corresponds to a winding transition as explained above, where the number of pieces of road winding around the source of the semi-infinite river becomes relevant. Above that curve, we expect the semi-meander configuration exponent to identically vanish.

5.2. Extensions

Extending the generalization of the previous subsection, we could consider *both* complicated river geometries and keeping n_1, n_2 finite. This requires some care when dealing with the electromagnetic operators creating river vertices, as several geometries might correspond to the same correlators. Also extra care should be exercised when imposing that the meandric objects be *connected*: in the multi-river and multi-road case, a meander must be globally connected, but can have disconnected rivers or roads. We may

also define higher-genus meandric numbers [5] which can also be enumerated by transfer matrix techniques, for instance on a torus.

As discussed before, we may also consider the coupling of the FPL^2 model to eulerian gravity, namely by summing over genus zero tetravalent graphs with only faces of even valency. Then all the above predictions are expected to still hold, except that we must take the formula (2.3) for the central charge, that remains one unit above that of the DPL^2 model as we have ensured the vertex-bicolorability of the graphs we sum over [22,23]. Note that, in this eulerian case, the b-vertex of Fig. 1 is now relevant and we *must* take $y \neq 0$ to get the above shifted central charge. The range of validity of Fig. 24 is reduced in this case to the zone of the square $[0,2] \times [0,2]$ delimited by the curve $c = 1$ joining the point $(2, q^\star)$ to the point $(q^\star, 2)$, with $q^\star = 2\cos(\pi(\sqrt{13}-1)/6) = 0.410135\ldots$. One should also be able to check numerically the predictions for eulerian gravity, using a suitable modification of our transfer matrix technique, in order to incorporate the vertex-bicolorability.

Another direction of generalization would be to consider more loop colors. For instance, one can define folding problems of two-dimensional lattices onto a d-dimensional target space, that in turn correspond to fully-packed loop models with a given number of loop colors, related to d. Attaching a different fugacity for loops of different colors, we may generate different universality classes, described by different conformal theories. Upon coupling to two-dimensional quantum gravity, these models would correspond to multi-colored meanders of some kind.

A final direction of generalization would be to add "matter" to the meandric graphs. Indeed, imagine we would like to consider a more involved model for a compactly folded polymer (protein), by attaching to its nodes a spin variable with inter and intra-chain Ising-like interactions. Then a simple way of describing it would be by first defining this matter spin model on the configurations of the square lattice fully-packed loop model, and then switching on gravity. If the matter model is still conformal, this should immediately lead to new configuration exponents for meanders in the presence of matter.

5.3. Algebraic exponents

It is both interesting and sad to notice that the only meandric numbers that can be calculated exactly with reasonably simple combinatorial formulas are all in one way or another related to Catalan numbers, and the corresponding exponents are always integers or half-integers (take for instance the case $n_1 = 0$, $n_2 = 1$, then $c = -2$, $\alpha = 3$, $\bar{\alpha} = 3/2$).

For general rational values of the e_i, we predict however that the exponents are algebraic numbers, roots of some quadratic equations with integer coefficients. Remarkably, such exponents are not commonplace in physics, but have emerged in some recent works on random walks [37,38]. If we could devise some relation between the meander problems and this other type of problems, we would probably be able to get a mathematically rigorous proof of our predictions.

Acknowledgements

We thank O. Golinelli for useful discussions.

References

[1] A. Sainte-Laguë, Avec des nombres et des lignes (Récréations mathématiques), Vuibert, Paris, 1937.
[2] K. Hoffman, K. Mehlhorn, P. Rosenstiehl, R. Tarjan, Sorting Jordan sequences in linear time using level-linked search trees, Information and Control 68 (1986) 170–184.
[3] V. Arnold, The branched covering of $CP_2 \to S_4$, hyperbolicity and projective topology, Siberian Math. J. 29 (1988) 717–726.
[4] K.H. Ko, L. Smolinsky, A combinatorial matrix in 3-manifold theory, Pacific J. Math. 149 (1991) 319–336.
[5] P. Di Francesco, O. Golinelli, E. Guitter, Meander, folding and arch statistics, Math. Comput. Modelling 26 (1997) 97–147.
[6] J. Touchard, Contributions à l'étude du problème des timbres poste, Canad. J. Math. 2 (1950) 385–398.
[7] W. Lunnon, A map-folding problem, Math. Comput. 22 (1968) 193–199.
[8] P. Di Francesco, O. Golinelli, E. Guitter, Meanders: a direct enumeration approach, Nucl. Phys. B 482 [FS] (1996) 497–535.
[9] O. Golinelli, A Monte-Carlo study of meanders, Eur. Phys. J. B 14 (2000) 145–155.
[10] I. Jensen, Enumerations of plane meanders, cond-mat/9910313.
[11] E. Brézin, C. Itzykson, G. Parisi, J.-B. Zuber, Planar diagrams, Commun. Math. Phys. 59 (1978) 35–51.
[12] S. Lando, A. Zvonkin, Plane and projective meanders, Theor. Comput. Science 117 (1993) 227–241; Meanders, Selecta Math. Sov. 11 (1992) 117–144.
[13] Y. Makeenko, Strings, matrix models, and meanders, Nucl. Phys. (Proc. Suppl.) 49 (1996) 226–237.
[14] G. Semenoff, R. Szabo, Fermionic matrix models, Int. J. Mod. Phys. A 12 (1997) 2135–2292.
[15] L. Chekhov, C. Kristjansen, Hermitian matrix model with plaquette interaction, Nucl. Phys. B 479 (1996) 683–696.
[16] B. Eynard, C. Kristjansen, More on the exact solution of the O(n) model on a random lattice and an investigation of the case $|n| > 2$, Nucl. Phys. B 466 [FS] (1996) 463–487.
[17] P. Di Francesco, O. Golinelli, E. Guitter, Meanders: exact asymptotics, Nucl. Phys. B 570 [FS] (2000) 699–712.
[18] J.L. Jacobsen, J. Kondev, Field theory of compact polymers on the square lattice, Nucl. Phys. B 532 [FS] (1998) 635–688; Transition from the compact to the dense phase of two-dimensional polymers, J. Stat. Phys. 96 (1999) 21–48.
[19] V.G. Knizhnik, A.M. Polyakov, A.B. Zamolodchikov, Fractal structure of 2D quantum gravity, Mod. Phys. Lett. A 3 (1988) 819–826.
[20] F. David, Conformal field theories coupled to 2D gravity in the conformal gauge, Mod. Phys. Lett. A 3 (1988) 1651–1656.
[21] J. Distler, H. Kawai, Conformal field theory and 2D quantum gravity, Nucl. Phys. B 321 (1989) 509.
[22] E. Guitter, C. Kristjansen, J. Nielsen, Hamiltonian cycles on random eulerian triangulations, Nucl. Phys. B 546 [FS] (1999) 731–750.
[23] P. Di Francesco, E. Guitter, C. Kristjansen, Fully packed O($n = 1$) model on random eulerian triangulations, Nucl. Phys. B 549 [FS] (1999) 657–667.

[24] P. Di Francesco, E. Guitter, J.L. Jacobsen, work in progress.
[25] B. Eynard, C. Kristjansen, An iterative solution of the three-colour problem on a random lattice, Nucl. Phys. B 516 (1998) 529–542.
[26] F. David, B. Duplantier, Exact partition functions and correlation functions of multiple Hamiltonian walks on the Manhattan lattice, J. Stat. Phys. 51 (1988) 327–434.
[27] H. Saleur, J. Phys. A 19 (1986) L807–L810.
[28] B. Nienhuis, in: C. Domb, J.L. Lebowitz (Eds.), Phase Transitions and Critical Phenomena, Vol. 11, Academic, London, 1987.
[29] P. Di Francesco, H. Saleur, J.-B. Zuber, Modular invariance in non-minimal two-dimensional conformal theories, Nucl. Phys. B 285 (1987) 454–480.
[30] H. Saleur, Polymers and percolation in two dimensions and twisted $N = 2$ supersymmetry, Nucl. Phys. B 382 (1992) 486–531.
[31] I.G. Enting, Generating functions for enumerating self-avoiding rings on the square lattice, J. Phys. A 13 (1980) 3713–3722.
[32] B. Derrida, Phenomenological renormalisation of the self avoiding walk in two dimensions, J. Phys. A 14 (1981) L5–L9.
[33] H.W.J. Blöte, M.P. Nightingale, Critical behaviour of the two-dimensional Potts model with a continuous number of states: A finite size scaling analysis, Physica A 112 (1982) 405–465.
[34] P. Di Francesco, E. Guitter, C. Kristjansen, Integrable 2D Lorentzian gravity and random walks, Nucl. Phys. B 567 [FS] (2000) 515–553.
[35] B. Durhuus, J. Frölich, T. Jónsson, Nucl. Phys. B 240 (1984) 453–480.
[36] F. David, Nucl. Phys. B 487 [FS] (1997) 633–649.
[37] G. Lawler, O. Schramm, W. Werner, Values of Brownian intersection exponents I: half plane exponents, math.PR/9911084.
[38] D.S. Fisher, P. Le Doussal, C. Monthus, Random walkers in 1-D random environments: exact renormalization group analysis, Phys. Rev. E 59 (1999) 4795–4840.

6.7 Article "Transfer matrix enumeration of colored links"

Nous proposons un algorithme de matrice de transfert pour l'énumération de diagrammes d'entrelacs alternés ayant des pattes extérieures. Chaque composante connexe a un poids n. Plus généralement, nous introduisons des diagrammes tétravalents avec des intersections et des tangences afin de pouvoir traiter des équivalences topologiques du type *flype*. Ceci entraîne une renormalisation finie d'un modèle de matrices associé. Nous donnons des résultats, sous forme de polynômes en n, pour les différentes fonctions génératrices jusqu'à l'ordre 19 (diagrammes à 2 pattes extérieures), 15 (4 pattes) et 11 (6 pattes) croisements. La limite $n \to \infty$ est résolue explicitement. Nous procédons à une analyse du comportement asymptotique des fonctions génératrices. Pour $0 \leq n \leq 2$, cette analyse est en bon accord avec une conjecture pour l'exposant critique basée sur la relation KPZ.

A Transfer Matrix approach to the Enumeration of Colored Links

J. L. Jacobsen *and* P. Zinn-Justin

Laboratoire de Physique Théorique et Modèles Statistiques
Université Paris-Sud, Bâtiment 100
91405 Orsay Cedex, France

We propose a transfer matrix algorithm for the enumeration of alternating link and tangle diagrams, giving a weight n to each connected component. Considering more general tetravalent diagrams with self-intersections and tangencies allows us to treat topological (flype) equivalences. This is done by means of a finite renormalization scheme for an associated matrix model. We give results, expressed as polynomials in n, for the various generating functions up to order 19 (2-legged tangle diagrams), 15 (4-legged tangles) and 11 (6-legged tangles) crossings. The limit $n \to \infty$ is solved explicitly. We then analyze the large-order asymptotics of the generating functions. For $0 \leq n \leq 2$ good agreement is found with a conjecture for the critical exponent, based on the KPZ relation.

04/2000

1. Introduction

It is well-known that a d-dimensional system in statistical mechanics can be conceived as a $(d-1)$-dimensional quantum field theory, by distinguishing one of the spatial coordinates as the direction of time. This correspondence lies at the heart of the *transfer matrix* formalism, where a linear operator is used to describe the discrete time evolution of the corresponding quantum system. More generally, transfer matrices have numerous applications for the combinatorial enumeration of discrete objects for which a definite direction (the transfer, or time, direction) can be singled out. Recently, this combinatorial aspect has come into focus through the enumeration of various objects pertaining to two-dimensional quantum gravity [1,2,3,4], such as plane meanders. Common to these examples is the existence of a preferred direction (e.g. the river, in the case of meanders), which can be straightforwardly promoted to the time direction.

In a previous paper [4] we have shown how this scheme also applies to the enumeration of alternating knot diagrams. Here the knot itself defines the transfer direction, since the algorithm essentially consists in reading the knot starting from one "ingoing" leg and ending at the other "outgoing" leg.

It is however far from obvious how this principle may generalize to the case of *link diagrams* with more than one connected component. This is the purpose of the present paper. More specifically, the final goal is to count alternating tangles *at fixed number of connected components*; it is therefore a generalization of the counting of alternating tangles with minimum number of components done in [4], but also of the counting of alternating tangles of [5,6] and of oriented alternating tangles of [7]. We also present some results for tangles with a higher number of outgoing strings ("external legs"), instead of just four as in the publications mentioned above. We shall in what follows present not just one, but two rather different transfer matrices addressing this enumeration problem.

After the definitions in section 2, which include various intermediate generating functions needed in the calculation, we shall present in section 3 the basic ideas behind the two proposed transfer matrices for alternating tangle diagrams. Section 4 is devoted to more technical details on the actual implementation of these ideas on a computer, and section 5 gives the numerical results and their analysis. Finally, in Section 6, we discuss how our algorithms may be adapted to various other problems of interest in graph theory and statistical physics.

2. Definitions of the generating functions

The objects we want to consider are tangles with $2k$ "external legs", that is roughly speaking the data of k intervals embedded in a ball B and whose endpoints are given distinct points on the boundary ∂B, plus an arbitrary number of (unoriented) circles embedded in B, all intertwined, and considered up to orientation preserving homeomorphisms of B that reduce to the identity on ∂B. Tangles with 4 external legs will be simply called tangles. The rest of the basic definitions is identical to those given in [4]. We represent these objects using diagrams, and restrict ourselves to alternating diagrams. This implies in particular that tangles can be considered as flype equivalence classes of diagrams [8].

Our goal is to count the number of prime tangles with a certain number of external legs and connected components. We shall relate in this section their generating functions to a simpler, more directly computable quantity, which is the following triple generating function

$$G(n, g_1, g_2) = \sum_{k,p_1,p_2=0}^{\infty} a_{k,p_1,p_2} n^k g_1^{p_1} g_2^{p_2} \tag{2.1}$$

where a_{k,p_1,p_2} is the number of topologically inequivalent open curves in the plane going from $(-\infty, 0)$ to $(+\infty, 0)$ together with k circles, connected together by p_1 regular intersections and p_2 tangencies, see Fig. 1. Here n, g_1 and g_2 are formal parameters, which can be evaluated at arbitrary complex values; however, it is natural to identify n with a number of *colors* one can assign to any of the closed loops of the diagram, so that the factor n^k correctly counts the total number of possible colorings of the diagram (assuming the external legs to carry a fixed color). The coefficient $a_{k,p,0}$ of the double generating function $G(n, g_1 = g, g_2 = 0)$ possesses the following interpretation: it is the number of alternating tangle diagrams with 2 external legs, k circles (i.e. $k+1$ connected components) and p crossings. The general coefficients do not possess such a clear knot-theoretic interpretation; however, they are needed to take into account the flyping equivalence (see [9,10]).

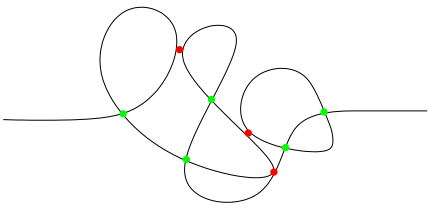


Fig. 1: Open curve and a circle with intersections (green dots) and tangencies (red dots).

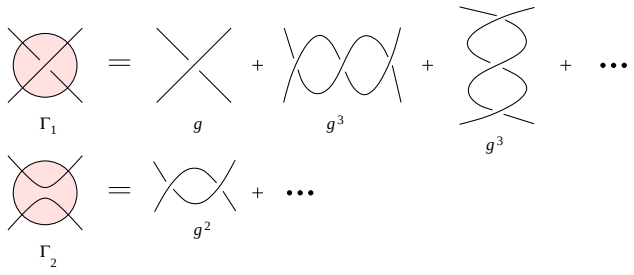


Fig. 2: Tangles of types 1 and 2 are distinguished by the two ways of connecting their external legs.

Next we define the generating functions of tangles Γ_1 and Γ_2 (see Fig. 2), which are necessary for the flyping equivalence. They are given by:

$$G = 1 + g_1 F_1 + g_2 F_2 \tag{2.2a}$$

$$\frac{\partial}{\partial g_2} F_1 = \frac{\partial}{\partial g_1} F_2 \tag{2.2b}$$

$$4F_1 = nG_1 + 2G_2 \tag{2.2c}$$

$$2F_2 = G_1 + (n+1)G_2 \tag{2.2d}$$

$$\Gamma_1 = G_1 \tag{2.2e}$$

$$\Gamma_2 = G_2 - G^2 \tag{2.2f}$$

via intermediate functions F_1, F_2, G_1, G_2. Note that inverting Eqs. $(2.2c,d)$ requires $n \neq 1, -2$. These special values of n will be investigated in detail in [10].

As in [4], we introduce an extra parameter to count edges of the diagram, according to the following definitions: $G(n, g_1, g_2, t) \equiv \frac{1}{t} G(n, g_1/t^2, g_2/t^2)$ and $\Gamma_i(n, g_1, g_2, t) \equiv \frac{1}{t^2}\Gamma_i(n, g_1/t^2, g_2/t^2)$.

The parameters t, g_1 and g_2 must then be chosen as a function of n and g according to the following *renormalization procedure* (see [10]):

$$1 = G(n, g_1(n,g), g_2(n,g), t(n,g)) \tag{2.3a}$$

$$g_1(n,g) = g(1 - 2H_2'(n,g)) \tag{2.3b}$$

$$g_2(n,g) = -g(H_1'(n,g) + V_2'(n,g)) \tag{2.3c}$$

where $H_1'(n,g)$, $H_2'(n,g)$ and $V_2'(n,g)$ are auxiliary quantities defined by:

$$H_2' \pm H_1' = \frac{(1 \mp g)(\Gamma_2 \pm \Gamma_1) \mp g}{1 + (1 \mp g)(\Gamma_2 \pm \Gamma_1) \mp g} \tag{2.4a}$$

$$H_2' + H_1' + nV_2 = \frac{(1-g)(\Gamma_2 + (n+1)\Gamma_1) - g}{1 + (1-g)(\Gamma_2 + (n+1)\Gamma_1) - g} \tag{2.4b}$$

These equations are independent only for $n \neq 0$, but have a smooth $n \to 0$ limit which is given in [4].

$\Gamma_1(n, g_1(n,g), g_2(n,g), t(n,g))$ and $\Gamma_2(n, g_1(n,g), g_2(n,g), t(n,g))$, once equations (2.3) are solved, are the desired generating functions for the number of prime alternating tangles of types 1 and 2 respectively [7] (see Fig. 2). The total number of tangles is given by $\Gamma_1 + 2\Gamma_2$. Similarly, one can define more general generating functions in the variables n, g_1 and g_2 which count tangles with more external legs; an explicit example will be given in Section 5.

3. The transfer matrices for alternating tangles

We now turn to the description of the two transfer matrices. The basic idea is common to both of them: starting from an initial state consisting of *all* external legs (i.e. two in the case under consideration), the system is time evolved through the addition of n intersections, until an empty final state is obtained.

Supposing the tangle diagram oriented from left to right, the first algorithm procedes by always evolving the uppermost vertex. We describe the details of this "single-step" algorithm in Section 3.1. The second algorithm, on the contrary, evolves all parts of the diagram simultaneously, adding one vertex to each of them in a given time step. In this way, the time can be defined as the geodesic distance from the pair of external legs. The details of this "geodesic" algorithm can be found in Section 3.2.

As in [4], we shall first concentrate on the enumeration of (prime, alternating) tangle *diagrams* with two external legs, which are related to the generating function $G(n, g = g_1, g_2 = 0)$ of Section 2. Adding tangencies, which is needed to take into account the flyping equivalence, will be discussed in Section 3.3, since it is an elementary extension of the algorithms.

Another minor modification of the algorithms will enable us to enumerate diagrams with more than two external legs; we shall develop this point in Section 3.4.

In both algorithms, the needs for CPU-time and memory increase exponentially with the system size p, though mercifully much more slowly than the number of knot diagrams actually being enumerated. As will become clear shortly, the single-step algorithm favors speed at the expense of memory consumption, while for the geodesic algorithm it is the other way around. However, since in practice both of these parameters are limiting factors for the maximally obtainable system size, it is a priori not clear which of the algorithms

performs best. We defer a detailed comparison to Section 4.3, and it turns out that the single-step algorithm comes out as the winner. Incidentally, even in the case of knots (one connected component), it performs slightly better than the algorithm described in [4].

3.1. The single-step algorithm

Let us briefly recall the working principle behind the knot enumeration algorithm presented in [4]. Reading the two-legged knot diagram from the first "ingoing" to the second "outgoing" leg, and calling at any instant the edge being read the "active line", there are two possibilities at each time step: 1) The active line is crossed by a line segment with edge labels that have not previously been encountered. We then add the new line segment to the current state. 2) The active line is identified with one of the endpoints of a line segment previously encountered. We then join the active line to that endpoint, wind around the line segment in question, and identify the new position of the active line with the opposite endpoint.

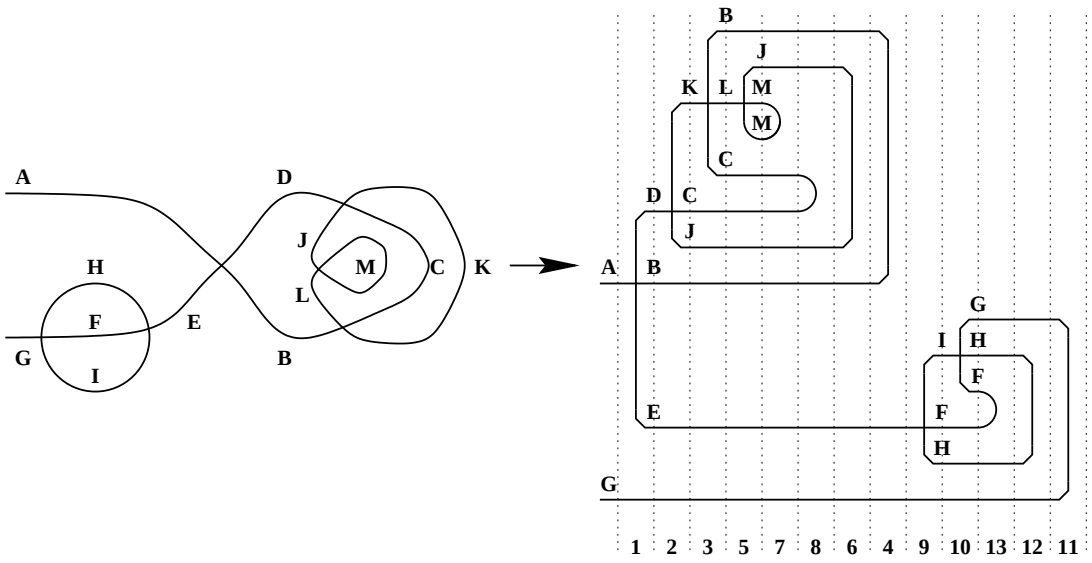


Fig. 3: Working principle of the single-step algorithm. a) A two-legged knot diagram with $p = 6$ intersections and $k = 3$ connected components. The edges are labelled from A to M. b) The same diagram in the time-slice representation. For reasons of clarity, the time slices are not drawn in chronological order.

The single-step algorithm can be viewed as a generalization of this principle. Let us, for the sake of illustration, consider the tangle diagram shown in Fig. 3. Since there is in general more than one connected component, clearly the concept of a unique "active line" no longer applies. Let us instead start from an initial state given by *both* external legs (edges A and G).

Moving along either of the edges A or G, a new line segment (DE resp. HI) is encountered. The question then arises which of these to process first. We resolve this ambiguity by stipulating that *in any given state, we evolve the line which at that instant is uppermost.*[1] At time $t = 1$, the edge A thus becomes B, and the new line segment DE is added. The edge D is now the new top line.

Analogously, at the instants $t = 2$ and $t = 3$, the top line (D resp. K) crosses a new line segment, which is then added to the current state. We can formalize this by stating the transformation rule shown in Fig. 4.1. This generalizes the corresponding rule of [4], except that the "active line" is now replaced by the uppermost line.

At time $t = 3$, the new top line carries the label B, which was however already produced by the transformation acting at $t = 1$. We therefore proceed, at $t = 4$, to the identification of the two "copies" of B, joining them through an arch. This is an example of the general transformation rule shown in Fig. 4.2. The addition of an arch means that the lines intermediate between the two instances of B (at positions $p = 1$ and q on Fig. 4.2) can henceforth not communicate with the lines at the exterior of the arch. These "trapped lines" must therefore eventually evolve to the empty state (vacuum), independently of the rest of the diagram. This observation has two implications: First, since both transformations conserve the parity (even/odd) of the number of lines, $p - q$ must be odd. Second, due to the above-mentioned rule that the current top line must always be treated first, the evolution of a possible set of "trapped lines" must take place at a later time. This means that when illustrating the sequence of moves on Fig. 3, we cannot *draw* the time-slices in chronological order. Note however, that the time ordering of the transformations is given by the time labels shown in the bottom of Fig. 3.

[1] With the optimization to be discussed in Sec. 4.2 we shall permit certain permutations of the lines. However, the line being evolved is in all cases the uppermost in the given *state*, though not necessarily in the corresponding time-slice representation.

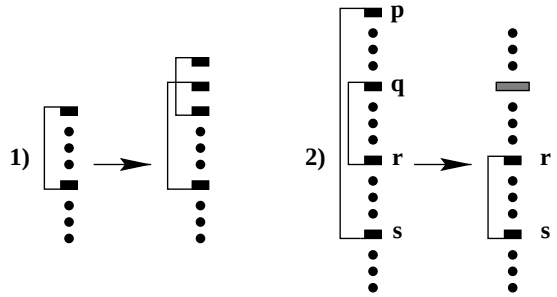


Fig. 4: The two types of transformations in the single-step algorithm. 1) Addition of a new line segment. 2) Identification of the top line (at position p) with another line (at position q), accompanied with the creation of a new block. Several remarks apply to the relative positions of p, q, r and s (see text).

The existence of a number of trapped lines is visualized on Fig. 4.2 by a *delimiter* (shown as a gray rectangle), which seperates the remaining lines into two *blocks*. Lines in different blocks cannot communicate, and must eventually evolve to the vacuum separately. In particular this means that the transformation 2) only applies when p and q belong to the same block. Conversely, q and r, and r and s may very well be separated by one or more delimiters (not shown). Also, although we have illustrated the case $r < s$, we may as well have $r > s$.

Of the thirteen transformations shown on Fig. 3, number 1, 2, 3, 5, 9, and 10 are of type 1, and the rest are of type 2. Clearly, the two types of transformations increase (resp. decrease) the number of lines by unity. Since the initial state consists of one line, the number of type 2 transformations must therefore exceed the number of type 1 transformations by one.

The purpose of the transfer matrix is not only to count the total number of tangle diagrams, but to do so for any fixed number of connected components. In particular, when performing a type 2 transformation, we need to know whether the points p and q were already connected though an arbitrary number of edges at an *earlier* time. On Fig. 4 we have represented this information by a number of lines on the left, connecting the points at a given instant into pairs. It may thus happen that on Fig. 4.2, $r = p$ and $s = q$. In this case, the type 2 transformation marks the completion of one connected component in the tangle diagram.

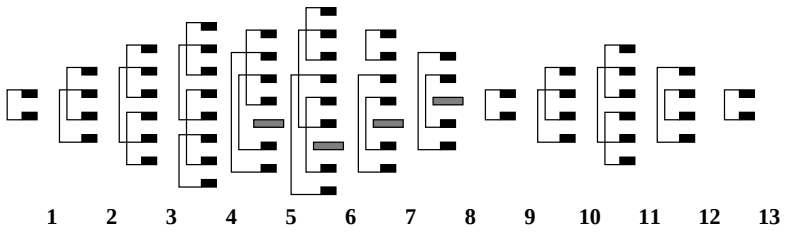


Fig. 5: Intermediate states produced by applying the single-step algorithm to the tangle diagram shown in Fig. 3.

We are now ready to define the set of *states* on which the transfer matrix acts. A state is defined by an even number of points (represented on Fig. 4 as black rectangles), connected into pairs by means of the edges encountered at previous times. In addition, the points are divided into $\ell + 1$ blocks by means of $\ell \geq 0$ delimiters. Note that points in different blocks can very well be connected, since any connection made beforehand persists after the addition of a delimiter.

On Fig. 5 we show the set of intermediate states corresponding to the time-slice representation of the tangle of Fig. 3. The initial state of a two-legged tangle is given by a pair of points (the exterior legs), implicitly connected at the point at infinity. It can be noted that the same state may occur at different instants of the transfer process. Also, any given state is not necessarily allowed at all instants.

After each type 2 transformation one may be able to simplify the set of delimiters. Namely, a delimiter may be eliminated if it is adjacent to another delimiter, or if it precedes the first point or succedes the last point of a state. On Fig. 5 we have implicitly assumed that such simplications have been carried out.

Finally, we must define the transfer matrix T which counts *all* tangle diagrams with $\tilde{p}$ vertices and k connected components. Its entries T_{ab}, where a and b are two basis states of the kind just defined, are 0 unless b is a descendant of a. An allowed state b is a descendant of a if it can be obtained via a transformation of one of the two types shown on Fig. 4 (for an arbitrary even $q \geq 2$ belonging to the same block as $p = 1$), followed by an arbitrary number of simplifications. T_{ab} is then the sum over all transformations from a to b of the corresponding weight: 1 or n depending on whether one closes a connected component or not (n can be either a given number, or a formal parameter, with a space of states enlarged by polynomials in n in the latter case). For the moment the simplications are just the elimination of superfluous delimiters. But we shall later (in Sec. 4.2) show that hitherto different states are equivalent by means of suitable transformations of the

blocks, and of the points within each block, such that a given state may be brought into a normal form. T then acts on the space of such normal forms. Apart from considerably reducing the dimension of the state space, these additional simplifications greatly enhance the efficiency of the algorithm.

As an example, we give in Appendix A the complete set of intermediate states with their corresponding weights for the counting of tangle diagrams up to four crossings.

3.2. The geodesic algorithm

We now turn to the description of our second algorithm. Apart from providing a highly non-trivial check of our results, our motivation for developing this alternative algorithm was to try to limit the number of intermediate states and thus lower the memory needs of the program. We still define the initial state as the set of external legs, but we redefine the chronological order of the tangle diagram by taking the time coordinate to be the *geodesic distance to the set of external legs.*

Roughly speaking, in each time step we apply one of the transformations shown in Fig. 4 to *each* of the lines present in the state at that instant. However, in order to introduce a valid time ordering of the diagram this rough idea needs to be refined.

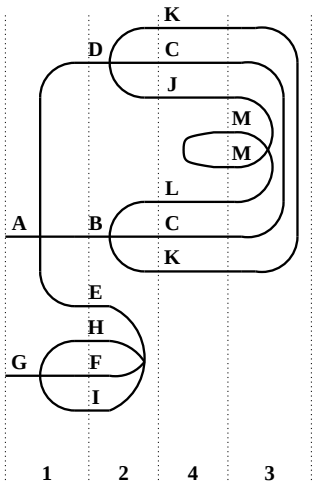


Fig. 6: The geodesic time-slice representation of the tangle diagram given in Fig. 3.

To progress, let us again consider the sample tangle diagram of Fig. 3. In Fig. 6 we show its new time-slice representation, this time using the above geodesic definition of time. At time $t = 1$ the edges A and G are both subject to the same transformation, in which three new edge labels (D, B, E resp. H, F, I) are encountered: as a short-hand notation we shall refer to this transformation as $1 \to 3$. It closely resembles the type 1 transformation in the single-step algorithm. When $t = 2$ the two upper edges (D and B) again undergo a $1 \to 3$ transformation, whereas the four lower edges (E, H, F, and I) annihilate at a common vertex: this is the $4 \to 0$ transformation. At the instant $t = 3$, it is recognized that two edge pairs (K and C) created at $t = 2$ carry the same label and thus must be identified. This transformation is reminiscent of the type 2 transformation in the single-step algorithm, and we shall here tag it $2 \to 0$. At the same time, the two edges J and L cross so as to become a new pair of edges, which are incidentally both labelled M. This is yet another transformation, the $2 \to 2$.

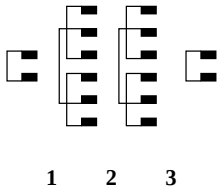


Fig. 7: Intermediate states used by the geodesic algorithm.

The intermediate states produced by this time-slice representation of this example are listed in Fig. 7. When comparing with the single-step algorithm (see Fig. 5) we note a considerable simplification.

Turning now to the general case, we see that apart from the $2 \to 0$ move, the transformations discussed above simply express the various ways of "time ordering" a tetravalent vertex, i.e. to assign the label t to at least one of its incident edges, and the label $t + 1$ to the remaining edges. The possibility $0 \to 4$ is excluded, as it would lead to the creation of disconnected diagrams. This leaves us with the transformations $4 \to 0$, $3 \to 1$, $2 \to 2$ and $1 \to 3$. But due to the planarity of the diagrams, one also needs to take into account that some of these transformations exist in several variants. For example, the "tadpole" labelled M on Fig. 6 is situated to the left of its adjacent vertex; however, a different diagram exists in which it is situated to the right. One must therefore accept that the transformation $2 \to 2$ comes in (at least) two guises: in the first, the outgoing edges bend backwards to the left, in the other they continue to the right.

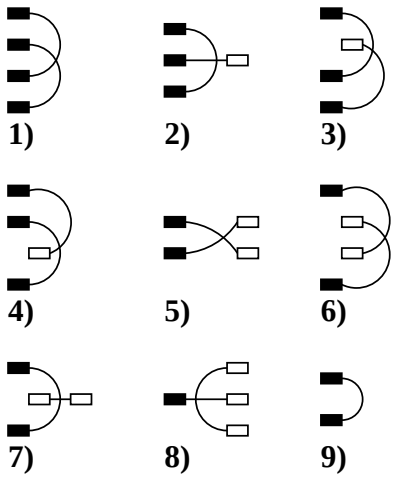


Fig. 8: Schematic transformation rules for the geodesic transfer matrix.

In Fig. 8 we show schematically the complete set of transformation rules. The black rectangles represent points at time t, and the white ones points at $t+1$. The solid lines indicate the action of the transfer matrix at time t. The first eight transformations are simply the topologically inequivalent ways of presenting a tetravalent vertex with at least one point labelled by t. In particular we note that the transformations of type $2 \to 2$ and $3 \to 1$ each occur in three different variants. The more exotic possibilities 3, 4 (resp. 5) start contributing to tangle diagrams with at least 4 (resp. 5) crossings. Finally, the ninth transformation is simply the $2 \to 0$.

We still need to transcribe these rules in terms of the states previously defined. In general, a given transformation leads to the creation of several new enclosed regions (blocks), and to represent the latter in terms of delimiters one needs to make use of the fact that cyclically permuting the points within a given block yields a topologically equivalent state (we shall come back to this point later, in Sec. 4.2). Also, since each block must evolve separately to the vacuum there are various parity constraints on the positions of the points entering a given transformation. By convention, we shall use capital letters to designate (possibly empty) blocks of points. Subscripts e and o indicate that the number of points must be even (resp. odd). Blocks with no subscript may have any parity: they are however subject to the global constraint that the total number of points must be even. Represented in this way, the exact transformation rules are given in Fig. 9.

Just like in the single-step algorithm, a transformation may be followed by an arbitrary number of simplifications (see Sec. 3.1).

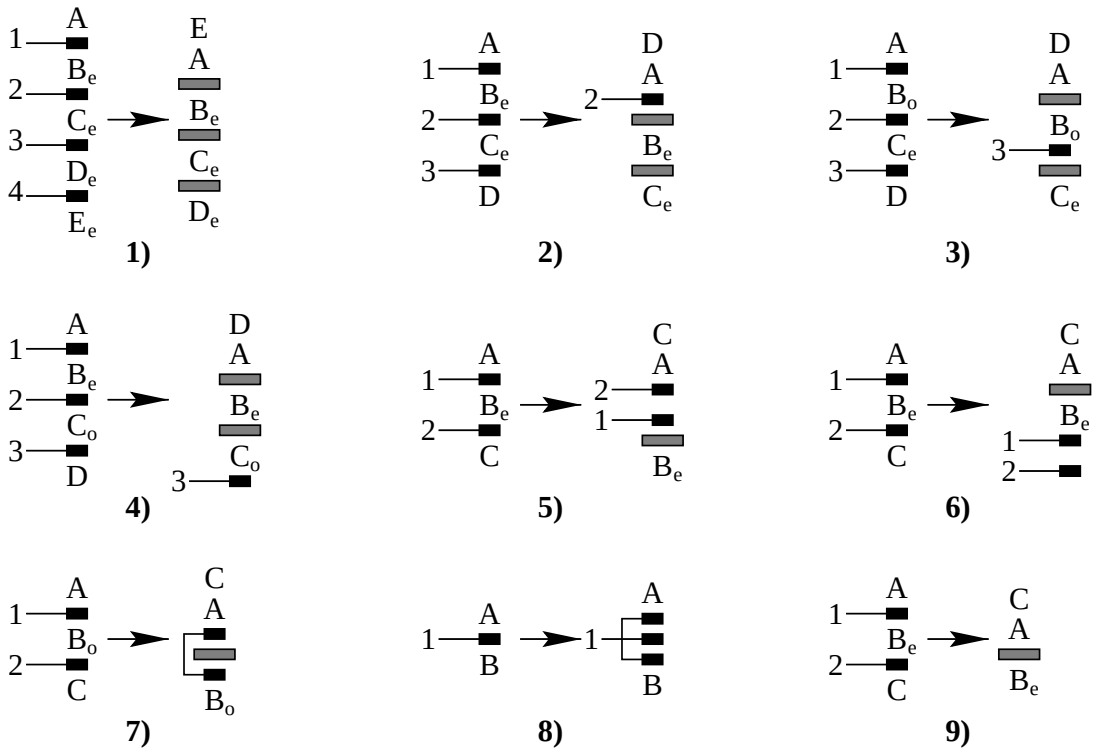


Fig. 9: Exact tranformation rules for the geodesic transfer matrix.

3.3. Tangencies

Until now we have been discussing the enumeration of tangle diagrams in which every vertex represents a crossing. However, to account for the flype equivalence we need to enumerate more general diagrams with p_1 intersections and p_2 tangencies, as discussed in Section 2. Fortunately, this is a very simple extension of either of our algorithms.

Let us for simplicity consider the case of the single-step algorithm. Adding a tangency rather than an intersection is obtained by modifying the transformation if Fig. 4.1, so that the two points added at time $t+1$ are both immediately above (resp. immediately below) the uppermost point at time t. Calling these variants respectively transformation 1a and 1b, a knot diagram with intersections and tangencies is then generated by acting on the initial state with a sequence of transformations 1, 1a, 1b and 2.

We also need to add to the characterization of each state a variable that, at any given time, specifies how many tangency transformations (type 1a or 1b) were used prior to that instant. The desired diagrams are then generated by sequences of p_1 transformations

of type 1, p_2 of type 1a or 1b, and $p_1 + p_2 + 1$ transformations of type 2, so that no intermediate state is empty.

Omitting the details, we notice that it is equally straightforward to include tangencies in the geodesic algorithm by obvious modifications of the eight first transformations of Fig. 8.

3.4. More external legs

Another extension of our algorithms consists in the enumeration of diagrams with 2ℓ external legs, $\ell > 1$. To this end we simply start from an appropriate initial state comprising ℓ line segments, instead of just one, and we demand that the total number of type 2 transformations exceed the total number of type 1 (i.e. 1, 1a, or 1b) transformations by ℓ.

For $\ell > 1$, such diagrams come in several types, corresponding to the number of ways of pairwise connecting the set of external legs at infinity. More precisely, given an ordered set of 2ℓ points $X_\ell \equiv \{x_1, x_2, \ldots, x_{2\ell}\}$, the number of types equals the number of ways to divide the set X_ℓ into pairs, considered up to the action of the dihedral group $D_{2\ell}$ on X_ℓ.[2] In particular, there are two types of (four-legged) tangles (see Fig. 2), and five types of tangles with $2\ell = 6$ external legs (see Section 5). The general integer sequence $1, 2, 5, 17, 79, \ldots$ thus defined is discussed in [11].

4. Implementational details

Although both of the algorithms described in the previous section are operational (as the reader may verify by studying Appendix A), we still need to give various details relative to their implementations on a computer. In particular, it is not clear how states of the type shown in Fig. 5 may be conveniently represented and manipulated. We shall address this question in Section 4.1. Another important observation is that states which until now have appeared to be different are in fact topologically equivalent. We shall discuss this point in Section 4.2 and demonstrate how it can be used to improve the efficiency of both algorithms. Finally, we compare the performances of the two different algorithms (single-step and geodesic) in Section 4.3.

[2] This also has a group-theoretic interpretation, in terms of number of $O(n) \times D_{2\ell}$-invariants in the tensor product of 2ℓ fundamental representations of $O(n)$, for generic n.

4.1. Representation of the states

In order to render the information contained in states of the type shown in Fig. 5 machine recognizable we shall represent each of them by an ordered list of non-negative integers. The length of the list representing a given state equals the number of points in the state plus the number of delimiters, and the order of its elements is given simply by reading the state from top to bottom. Each delimiter is represented by the digit zero. The other points each correspond to a positive integer, with the convention that two points are connected if and only if they are represented by the same integer. Clearly, this convention is not unique: for instance, $(1,2,0,1,2)$ and $(13,4,0,13,4)$ both describe the same state. To get rid of this ambiguity we shall stipulate that each consecutive digit, starting from the left, be chosen as small as possible, consistent with the above rules. Thus, $(1,2,0,1,2)$ is the unique normal form of our sample state.

The sequence of states shown in Fig. 5 can then be transcribed as follows:

$$\begin{aligned}(1,1) &\to (1,2,1,2) \to (1,2,1,3,2,3) \to (1,2,1,3,2,4,3,4) \to (1,2,3,1,0,3,2)\\ &\to (1,2,1,3,4,2,0,4,3) \to (1,1,2,3,0,3,2) \to (1,2,0,2,1) \to (1,1)\\ &\to (1,2,1,2) \to (1,2,1,3,2,3) \to (1,2,2,1) \to (1,1).\end{aligned}$$

At a given stage in the transfer process we need to run through the states present at time t, apply the transformation rules described in Section 3.1–3.2, and produce the set of descendant states (the states at time $t+1$) with their respective weight. The first time a given descendant state is produced, it must be inserted in a suitable data structure along with is weight. If subsequently the same state is produced again as a descendant of another parent state, rather than inserting it again we need to retrieve it in the data structure and update its weight. In order for the algorithm to be efficient, the operations of insertion and retrieval must be accomplished in constant time (i.e. in a time that does not depend on the number of states accomodated by the data structure).

These demands are fulfilled by a standard data structure known as a *hash table* [12]. It relies on the fact that to each state i we can assign a unique integer $k_i \in \mathbb{Z}_+$ (the hash key), and devise a function $f : \mathbb{Z}_+ \to \{0,1,2,\ldots,P-1\}$ (the hash function) that distributes the set of k_i's more-or-less uniformly on the set $\{0,1,2,\ldots,P-1\}$. By inserting the states i into an array of noded lists indexed by $f(k_i)$, we can retrieve any given state in a time proportional to the mean length of one of the pointer lists, $t \propto N/P$, where N is the total

number of entries. In practice we choose P to be a large prime such that $N/P \sim 10$, and we use the hash function $f(k) = k \bmod P$.

A convenient key k_i can be defined by concatenating the list of "digits" entering the normal form of the state i into one large integer. To find the minimum number of bits required to store one digit, we remark that for the counting of tangles with ℓ intersections the digits are all $\leq \ell + 1$. In the case at hand this means that we need to use at least five bits per digit; in practice we have however chosen to use eight bits, in order to profit from standard routines for handling character strings.

4.2. Equivalences between states

As has already been mentioned, it is true for either of the two tangle enumeration algorithms that some of the states generated at a given stage in the transfer process are topologically equivalent. Clearly, it is of the utmost interest to factor out as many topological equivalences as possible from the state space, since the memory demands as well as the time consumption of the algorithm are roughly proportional to the number of states being treated.

A first such equivalence is due to the fact that any two different blocks of points must evolve separately to the vacuum, without any mutual interaction. The relative position of the blocks is thus immaterial. The standard version of either algorithm (say, version 1) can thus be ameliorated by introducing a standard order among the blocks before inserting a given state in the hash table (version 2). We have done so by simply sorting the blocks according to their size. In the special case of the single-step algorithm it is advantageous to place the smallest blocks at the top of the state, since such blocks will then be evolved to the vacuum before touching any other block. The small block being eliminated, the remainder of the state will be smaller and can thus be processed more expeditiously. For the geodesic algorithm, the choice betwen ascending and descending ordering is irrelevant.

We take the convention of not changing the relative order of two equally sized blocks.[3] After the permutation of the blocks, the representation of the state in terms of a list of integers is brought back to its normal form (see Sec. 4.1).

[3] Inspecting by hand some modestly sized systems reveals that changing the order of equally sized blocks will only lead to a very small further gain. We have nevertheless made various attempts of imposing a more unique way of arranging the blocks, but since such transformations tend to break down a certain regularity in the connectivities which is imposed by the type 1 transformation, these attempts actually resulted in a slight *increase* in the number of states.

p	version 1	version 2	version 3
2	4	4	4
3	7	7	6
4	24	16	14
5	67	45	24
6	226	110	49
7	735	313	106
8	2573	804	209
9	9340	2160	479
10	32790	6345	1078
11	128794	17074	2382
12	468757	45858	5929
13	1933350	127751	13992

Tab. 1: Maximal number of intermediate states in the transfer process for three different versions (see text) of the single-step algorithm for tangle diagrams with p self-intersections and no tangencies.

In Table 1 we illustrate the resulting decrease in the number of states inserted in the hash table. It seems clear that not only is the number of states much smaller, but it even grows with a smaller exponent.

Second, the states are equivalent upon cyclic rotations (eventually combined with a reflection) of the points within any given block. Since, once again, the blocks are independent, these dihedral transformations can be performed independently within each block. The ultimate way of implementing this equivalence would be the following: before inserting a state in the hash table, subject it to all possible dihedral transformations, and check whether any of the transformed states is already present in the table. Unfortunately, the number of transformations increases faster than exponentially with the size of the state, and this exhaustive search would quickly end up usurping the majority of the CPU time.

We have therefore opted for a less perfect but much faster alternative (version 3).[4] Having sorted the blocks according to their size, we consider in turn all possible dihedral transformations on the points in the first block, keeping fixed the positions of the points in the other blocks. After each transformation we bring the integer representation of the state

[4] Comparing with the exhaustive method applied to some modestly sized systems shows that, once again, the number of topological equivalences not detected by the "approximate" method is negligible when compared to the number of states which are in fact topologically distinct.

into its normal form. We then identify (one of) the normal form(s) which lexicographically precedes all the others, and lock the points of the first block into their corresponding positions. Leaving the first block locked, we procede to apply the same procedure to the second block. We continue this way until all blocks have been locked, and only then the resulting representation of the state is inserted into the hash table.

The gain of version 3 over version 2 is comparable to the gain of version 2 over version 1, as witnessed by Table 1. In the following we shall therefore exclusively understand version 3 when referring to any one of the two algorithms (single-step or geodesic).

4.3. Comparing the two algorithms

A first striking difference between the single-step and the geodesic algorithm can be observed by comparing how their respective number of intermediate states (and thus the memory needs) evolve as a function of the "time" defined by the transfer process.

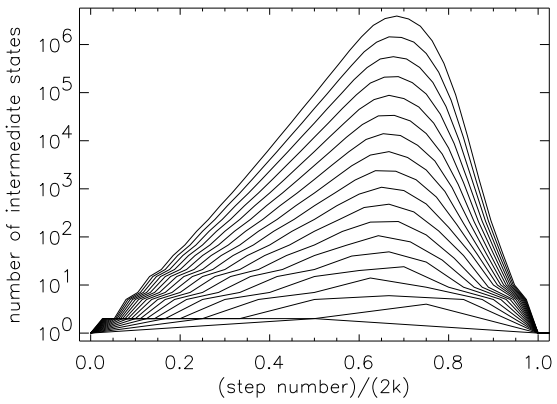


Fig. 10: Memory profile of the single-step algorithm (version 3, cf. Sec. 4.2). The curves represent knot diagrams with $p = 1, 2, \ldots, 19$ crossings and no tangencies.

In both cases, the number of states grows exponentially in the beginning, decreases exponentially towards the end, and reaches a maximum somewhere in between. However, for the single-step algorithm this maximum is reached at roughly $2T/3$ (where T is the total number of time steps), whereas for the geodesic algorithm the maximum is situated around $T/4$. The reason for this difference is that the geodesic algorithm will produce the majority of its states by applying the $1 \to 3$ rule as often as possible in the beginning of the process.

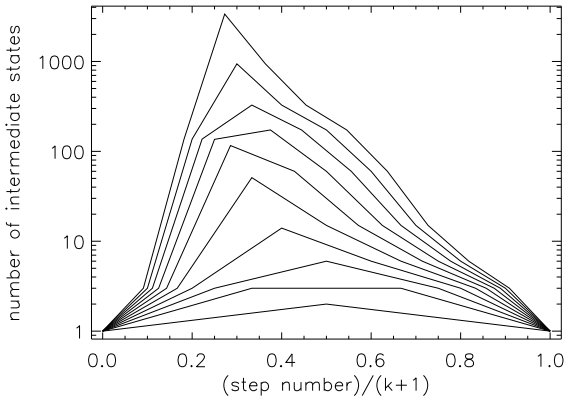


Fig. 11: Memory profile of the geodesic algorithm (version 2, cf. Sec. 4.2). The curves represent knot diagrams with $p = 1, 2, \ldots, 10$ crossings and no tangencies.

p	version 1	version 2	version 3
2	3	3	3
3	6	6	6
4	14	14	12
5	60	51	37
6	141	116	86
7	207	173	126
8	396	327	238
9	1308	941	544
10	5300	3367	1701

Tab. 2: Maximal number of intermediate states in the transfer process for three different versions (see Sec. 4.2) of the geodesic algorithm for tangle diagrams with p self-intersections and no tangencies.

To estimate the actual memory needs of the algorithms, it is instructive to compare the maximal number of intermediate states for the three different versions defined in Sec. 4.2. For the single-step algorithm the data were given in Tab. 1; we show the corresponding numbers for the geodesic algorithm in Tab. 2. As expected, version 1 of the geodesic algorithm employs considerably fewer states than version 1 of the single-step algorithm. However, quite surprisingly, the ameliorations implied by version 2 and version 3 lead to an enormous gain in the single-step case, but only a modest one in the geodesic case.

Thus, in version 3 the asymptotic growth of the number of states is significantly slower in the single-step algorithm than in the geodesic one, even though the latter was explicitly designed to use fewer states! Although a qualitative explanation of this phenomenon can be given by inquiring into the structure of a typical state we refrain from doing this here. Rather, let us simply accept the efficiency of the single-step algorithm as a remarkable fact.

Even when discarding the issue of memory, the geodesic algorithm has a serious drawback compared with its single-step counterpart as far as time consumption is concerned. Namely, the single-step algorithm processes each state in a time that grows roughly linearly with its size, whereas for the geodesic algorithm this time grows exponentially. To see this, consider the intermediate state which is obtained from the initial state by performing p transformations of type $1 \to 3$. To turn this state into tangle diagrams with exactly p crossings, one needs to complete it with $p+1$ transformation of type $2 \to 0$. Supposing p chosen so that the intermediate state has a complete number of time slices, a total of c_{p+1} diagrams will be recursively generated, where $c_k = \frac{(2k)!}{k!(k+1)!}$ are the Catalan numbers. The geodesic algorithm will therefore (asymptotically) spend the majority of the CPU time closing up this "maximally opened state".

We have therefore used the geodesic algorithm as a highly non-trivial check of our numerical results, but the data for large system sizes are generated exclusively by the single-step algorithm.

To conclude this section, let us briefly discuss the time complexity of our best algorithm (single-step, version 3). Based on the data in Tab. 1, we infer that both time and memory needs grow asymptotically as $\sim \kappa^p$, with $\kappa \approx 2.7 \pm 0.2$.

5. Numerical results

We now present the numerical results that we obtained using the single-step algorithm (version 3). Due to the enormous amount of data gathered we shall only give the main results.

The first data are obtained by running a program that implements the single-step algorithm without any tangencies. This corresponds to the generating function $G(n, g_1 = g, g_2 = 0)$ in the notation of Section 2. Its coefficients $a_{k,p,0}$ are given in Table 3 up to $p = 19$. The computation took a few hours on a 1 GHz single-processor work station with 1 GByte of memory.

p \ k	0	1	2	3	4	5	6	7	8	9
0	1									
1	2									
2	8	1								
3	42	12								
4	260	114	4							
5	1796	1030	90							
6	13396	9290	1349	22						
7	105706	84840	17220	728						
8	870772	787082	203568	14884	140					
9	7420836	7415814	2312094	244908	6120					
10	65004584	70867212	25691670	3575045	158354	969				
11	582521748	685839770	282000444	48517524	3185314	52668				
12	5320936416	6712285600	3074136464	628013796	55273668	1647728	7084			
13	49402687392	66349573368	33387698708	7871666088	871779428	39142116	460460			
14	465189744448	661680191832	361969672904	96451145091	12876308613	786444610	16890227	53820		
15	4434492302426	6651030871168	3921901043440	1162484964230	181430681094	14126467392	462455640	4071600		
16	42731740126228	67329662060890	42499598861832	13840075278704	2468480436152	234358127880	10552931952	171277860	420732	
17	415736458808868	685953949494774	460831546801414	163246693686684	32699872694298	3666111325052	212581611050	5308497112	36312408	
18	4079436831493480	7028941367108708	5001468564165262	1911737961254907	424232095742826	54835331971380	3912429396360	135564649071	1722788176	3362260
19	40338413922226212	72403769391718890	54341248085414380	22262254374655710	5413174461572394	791922013806504	67266181855770	3025712334552	59605106568	326023280

Tab. 3: Table of the number of alternating tangle diagrams with 2 external legs.

Let us first note that there are various quantities which can be extracted from this table and which are known exactly in an independent way. They provide a number of non-trivial checks. Let us define the number $a_p(n)$ of diagrams at fixed number of colors n

$$a_p(n) = \sum_{k=0}^{\infty} a_{k,p,0} n^k \tag{5.1}$$

The simplest choice is to set $n = 1$, that is to consider the sum of each row of the table. This series of numbers is known exactly (see [13] for a purely combinatorial argument, and [14] for a field theoretic one)

$$a_p(1) = 2\frac{(2p)!}{p!(p+2)!}3^p \tag{5.2}$$

and the corresponding generating function is

$$G(1,g,0) = \frac{1}{3}A(4-A) \tag{5.3a}$$

$$A = \frac{1-\sqrt{1-12g}}{6g} \tag{5.3b}$$

It is perhaps less well known that for $n = 2$, one also has an exact expression, in terms of elliptic integrals [15]:

$$G(2,g,0) = \frac{1}{g^2}\left(\frac{g}{2} - \frac{1}{8}\frac{k}{(1+k)^2}\right) \tag{5.4a}$$

$$g = \frac{E(k) - (1-k)K(k)}{2\pi(1+k)} \tag{5.4b}$$

where $K(k)$ and $E(k)$ are the complete elliptic integrals of the first and second kinds. We note that the generating function is non-algebraic, and it is not known how to find it by direct combinatorial arguments.

There exists a similar, although more complicated, formula for the case $n = -2$ which will be presented elsewhere [10].

One can also find expressions for the last non-zero element of each row, which formally corresponds to $n \to \infty$ (with $x \equiv ng^2$ fixed). However there is a parity effect which forces us to redefine separately odd and even generating functions:

$$H_{\rm even}(x) = \sum_{k=0}^{\infty} a_{k,2k,0}\, x^k \tag{5.5a}$$

$$H_{\rm odd}(x) = \sum_{k=0}^{\infty} a_{k,2k+1,0}\, x^k \tag{5.5b}$$

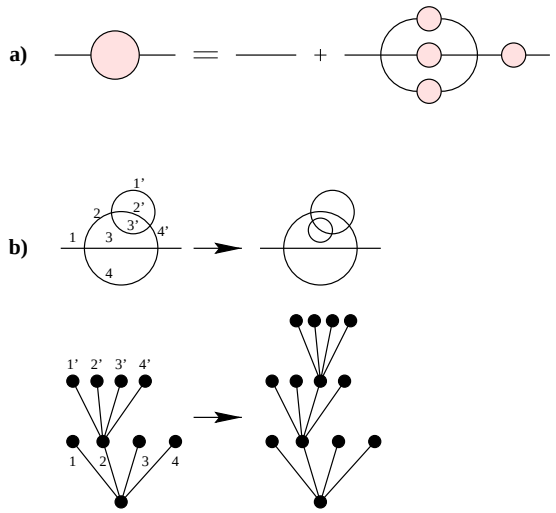


Fig. 12: a) Equation satisfied by $H_{\rm even}(x)$. b) Recursive definition of a circle diagram and its interpretation in terms of a rooted tree.

For the even case, a general diagram has the form of successive insertions of circles in the bare propagator, which leads to the equation, depicted on Fig. 12 a),

$$H_{\rm even} = 1 + xH_{\rm even}^4 \tag{5.6}$$

This can be described more explicitly in terms of rooted trees. Each insertion of a circle requires two additional intersections, and it leads to the creation of four new edges in which new circles can be inserted. On Fig. 12 b) we have labelled two successive generations of edges as $\{1,2,3,4\}$ and $\{1',2',3',4'\}$ respectively. Clearly, this reduces the problem to that of enumerating rooted trees in which each node (resp. the root) can have degree 1 or 5 (resp. 0 or 4). This is a simple example of a rather broad class of rooted trees discussed by Takács [16].

From Eq. (5.6) we infer that

$$a_{k,2k,0} = \frac{(4k)!}{(3k+1)!\,k!} \tag{5.7}$$

The proof of the formula for $H_{\rm odd}(x)$ is left as an exercize to the reader:

$$H_{\rm odd} = 2\frac{\mathrm{d}}{\mathrm{d}x}(xH_{\rm even}^3) \tag{5.8}$$

We infer that

$$a_{k,2k+1,0} = 2\frac{(4k+2)!}{(3k+2)!\,k!} \tag{5.9}$$

Finally, the first column in Table 3 reproduces the knot diagrams discussed in [4], of course.

Next we want to deduce some properties of the asymptotic behavior of these series from the numerical data. The first quantity one can extract is the "bulk entropy" of alternating tangles and links.[5] At fixed n it is defined by the leading exponential behavior of $a_p(n) \sim \mathrm{e}^{\hat{s}(n)p}$:

$$\hat{s}(n) = \lim_{p\to\infty} \frac{\log a_p(n)}{p} \tag{5.10}$$

It would however be more natural to consider alternating links/tangles at fixed number of connected components. This requires making an appropriate scaling ansatz for the coefficients $a_{k,p,0}$, which turns out to be

$$\log a_{k,p,0} \overset{k,p\to\infty}{\sim} p\, s(k/p) \tag{5.11}$$

where $s(x)$ is the bulk entropy at fixed ratio x of the number of connected components by the number of crossings. It is clear from Eq. (5.1) that the two entropies defined above are

[5] The bulk behavior should be independent of the number of external legs, and in particular be the same for links and tangles.

related to each other by a Legendre transform; namely, if one defines the *average* ratio x at fixed n:

$$x(n) = \frac{\mathrm{d}}{\mathrm{d}\log n}\hat{s}(n) = \left\langle \frac{k}{n} \right\rangle \tag{5.12}$$

then the following relation holds:

$$s(x(n)) = \hat{s}(n) - x(n)\log n \tag{5.13}$$

so that we also have the dual equation of (5.12)

$$-\log n = \frac{\mathrm{d}}{\mathrm{d}x}s(x(n)) \tag{5.14}$$

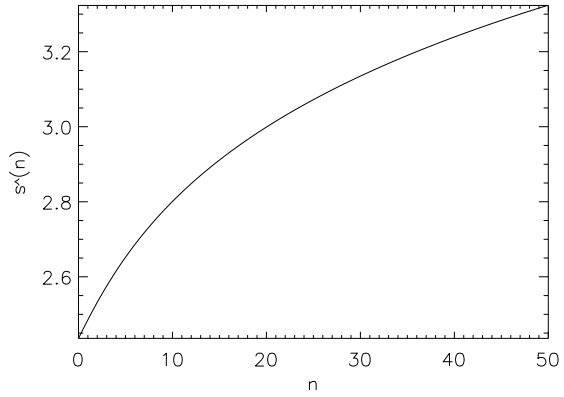


Fig. 13: The bulk entropy $\hat{s}(n)$.

On Fig. 13 the behavior of $\hat{s}(n)$ is shown for $n \in [0, \infty]$. The various exact solutions mentioned above correspond to the following known values of $\hat{s}(n)$:

$$\begin{aligned}
&\hat{s}(1) = \log 12 \\
&\hat{s}(2) = \log 4\pi \\
&\hat{s}(n) \overset{n\to\infty}{=} \frac{1}{2}\log n + \log\frac{16}{3\sqrt{3}} + o(1)
\end{aligned}$$

In [4], the following numerical value was given: $\exp\hat{s}(0) \approx 11.42$, which is confirmed here.

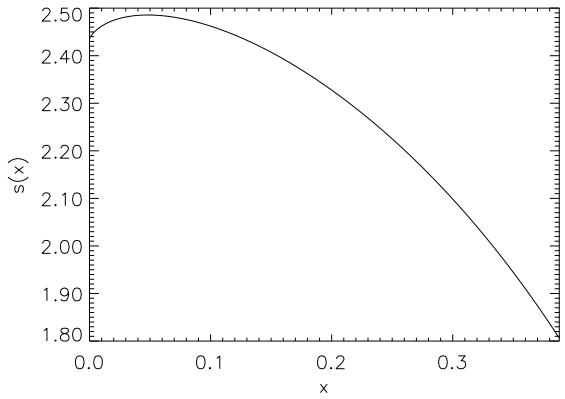


Fig. 14: The bulk entropy $s(x)$.

Even more interesting is the curve $s(x)$ shown on Fig. 14 for $x \in [0, 1/2]$. It displays a clear maximum at a value which is nothing but $x(1)$, cf Eq. (5.14). Numerically we find:

$$x(1) = 0.0481 \pm 0.0001 \tag{5.16}$$

This number has the following significance: at fixed number of crossings p, a typical tangle/link diagram will have $0.0481 \cdots \times p$ connected components when p goes to infinity. Note that this number is extremely small: any average made over (equally weighted) alternating link diagrams will be dominated by objects with few connected components.

Let us note that the bulk quantities *are affected* by the various renormalizations of Eq. (2.3), i.e. restriction to 2PI diagrams and inclusion of the flype equivalence. However, it is expected that the qualitative properties (and in particular the maximum of the entropy for a very small value of x) are unchanged. For example, if one considers 2PI diagrams (which corresponds to the counting of reduced prime alternating link/tangle diagrams), one finds a maximum at $x_2(1) \approx 0.033$ instead.

The discussion of the exponent associated to the subdominant power-law behavior of the series $a_p(n)$ (or, equivalently, $a_{k,p,0}$) is much more involved. We define the critical exponent

$$\alpha(n) = \lim_{p \to \infty} \frac{-\log a_p(n) + p\, \hat{s}(n)}{\log p} \tag{5.17}$$

Let us first recall the conjecture made in [9], which relies on several hypotheses: a) the asymptotic behavior of $a_p(n)$ is related to a singularity of the corresponding generating

function $G(n,g,0)$ which has the physical meaning of singularity of 2D quantum gravity, i.e. large link diagrams behave as continuum random surfaces for which conformal field theory techniques apply (KPZ formula [17]); b) the model describing link diagrams with n colors is in the same universality class as the usual $O(n)$ model of dense loops [18], which relies on the assumption that there is no phase transition in the generalized $O(n)$ matrix model. For $|n|<2$, $n=-2\cos(\pi\nu)$ $(0<\nu<1)$, this implies that

$$\alpha(n=-2\cos(\pi\nu)) = 1+1/\nu \tag{5.18}$$

Let us now discuss separately various regions of n and the corresponding numerical analysis.
• For $n<0$, a difficulty arises in that coefficients $a_p(n)$ do not have a fixed sign. This implies in particular that the dominant singularity of the generating function $G(n,g,0)$ is not necessarily on the real positive axis, as would be implied by hypothesis a) above. Numerically it seems that pairs of complex conjugated singularities do occur and become dominant in a large region of n which includes at least part of the interval $n\in[-2,0[$, thus invalidating conjecture (5.18) in this region. The analysis of such behavior is fairly involved and we leave it to future work.
• $0\le n<1$: at $n=0$ it was suggested in [4] that even though conjecture (5.18) is correct ($\alpha(0)=3$), there might be a logarithmic correction which spoils the asymptotic behavior of the coefficients. For n small, we expect several singularities extremely close to the dominant singularity making any analysis difficult. Estimates of the critical exponent do not contradict (5.18), but they have very low accuracy; for example,

$$n=\frac{\sqrt{5}-1}{2} \qquad \nu=3/5 \qquad \alpha_{\rm conj}=\frac{8}{3} \qquad \alpha_{\rm num}=2.6\pm0.1 \tag{5.19}$$

• $1\le n\le 2$: we can first extract from the exact solutions (Eqs. (5.2), (5.3) and (5.4)) the asymptotics

$$a_p(1) \sim 12^p p^{-5/2}\mathrm{cst} \tag{5.20.1}$$

$$a_p(2) \sim (4\pi)^p p^{-2}(\log p)^{-2}\mathrm{cst} \tag{5.20.2}$$

They are of course compatible with (5.18); however we note a logarithmic correction in (5.20.2) which comes from inverting the singularity of Eq. (5.4): $g-g_c\sim(k-k_c)\log(k-k_c)$. At this point it becomes clear that in order to remove the $(\log p)^{-2}$ factor, one just needs to perform an appropriate functional inversion on the generating series $G(2,g,0)$. Applying

the same procedure to the numerical data of $G(n, g, 0)$ for $1 \leq n \leq 2$, one can then use standard convergence acceleration methods and obtain precise estimates of the critical exponent. They are in good agreement with (5.18); for example,

$$n = \sqrt{2} \qquad \nu = 3/4 \qquad \alpha_{\rm conj} = \frac{7}{3} \qquad \alpha_{\rm num} = 2.34 \pm 0.01 \tag{5.21}$$

• $n > 2$: let us first recall that we have found exact expressions at $n \to \infty$ for odd and even coefficients separately (Eqs. (5.7) and (5.9)). Asymptotically,

$$a_{k,2k,0} \sim \left(\frac{256}{27}\right)^k k^{-3/2} \mathrm{cst} \tag{5.22a}$$

$$a_{k,2k+1,0} \sim \left(\frac{256}{27}\right)^k k^{-1/2} \mathrm{cst} \tag{5.22b}$$

i.e. the bulk terms are identical but the critical exponents are different. This can be understood easily since in the odd case there is one "defect" in the sequence of circles which corresponds to marking one connected component, that is multiplying by k (cf the differentiation in Eq. (5.8)).

Numerically, it is clear that for all $n > 2$ odd and even series behave differently. However once cannot perform any serious analysis on these series since they are too short. It may be that the exponents of Eq. (5.22) are preserved for any $n > 2$, or finite values of n might smooth the difference between odd and even series; the data we possess are unconclusive on this issue.

Let us end this analysis by noting that contrary to the bulk terms, critical exponents are expected to be *independent* of the various renormalizations of Eq. (2.3), due to universality arguments.

We now turn to the data obtained by inclusion of tangencies. For reasons of conciseness, we here refrain from displaying the three-dimensional array of coefficients a_{k,p_1,p_2}; these are electronically available from the authors upon request. Even the final results are fairly cumbersome to treat and display, so that we only show the results for the number of prime alternating tangles up to $p = 15$ (even though they can be easily obtained for p up to 18 or 19, as in [4], on a work station, and probably a bit further using larger computers).

These data satisfy once more various non-trivial checks, including the comparison with the table in the appendix of [5] (for $n = 1$), of the Tables 1 and 2 in [7] (for $n = 2$), and Table 3 of [4] (for $n = 0$).

p^k	Γ_1							Γ_2						
	0	1	2	3	4	5	6	0	1	2	3	4	5	6
1	1							0						
2	0							1						
3	2							1						
4	2							3	1					
5	6	3						9	1					
6	30	2						21	11	1				
7	62	40	2					101	32	1				
8	382	106	2					346	153	24	1			
9	1338	548	83	2				1576	747	68	1			
10	6216	2968	194	2				7040	3162	562	43	1		
11	29656	11966	2160	124	2			31556	17188	2671	121	1		
12	131316	71422	9554	316	2			153916	80490	15295	1484	69	1	
13	669138	328376	58985	5189	184	2		724758	425381	87865	6991	194	1	
14	3156172	1796974	347038	22454	478	2		3610768	2176099	471620	52231	3280	103	1
15	16032652	9298054	1864884	193658	10428	260	2	17853814	11376072	2768255	308697	15431	290	1

Tab. 4: Table of the number of prime alternating tangles.

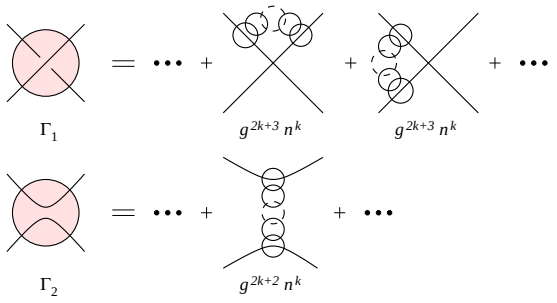


Fig. 15: Large n expansion of tangles.

The last term in each column can also be verified by means of the large n expansion. It is easy to convince oneself that in general the 2PI diagrams with the largest possible number of connected components is obtained by decorating the bare tangles by means of a festoon of chained circles, as shown on Fig. 15. In the case of Γ_1 (resp. Γ_2) there are two (resp. one) flype-inequivalent ways of doing so. As it stands, this argument holds true for odd p (resp. even p) in the case of Γ_1 (resp. Γ_2), but a similar reasoning holds true for the opposite parity. We conclude that the last term in each row of Tab. 4 should be 2 (for Γ_1

and $p \geq 6$) resp. 1 (for Γ_2 and $p \geq 2$), as is indeed observed. The next to leading terms should be obtainable in a similar fashion.

Finally, we demonstrate the power of our method by applying it to objects with more external legs. One may for example ask how many ways there are to intertwine *three* strings, and not just two as in the case of tangles. One must first establish the different ways the strings are coming out, which leads to Fig. 16.

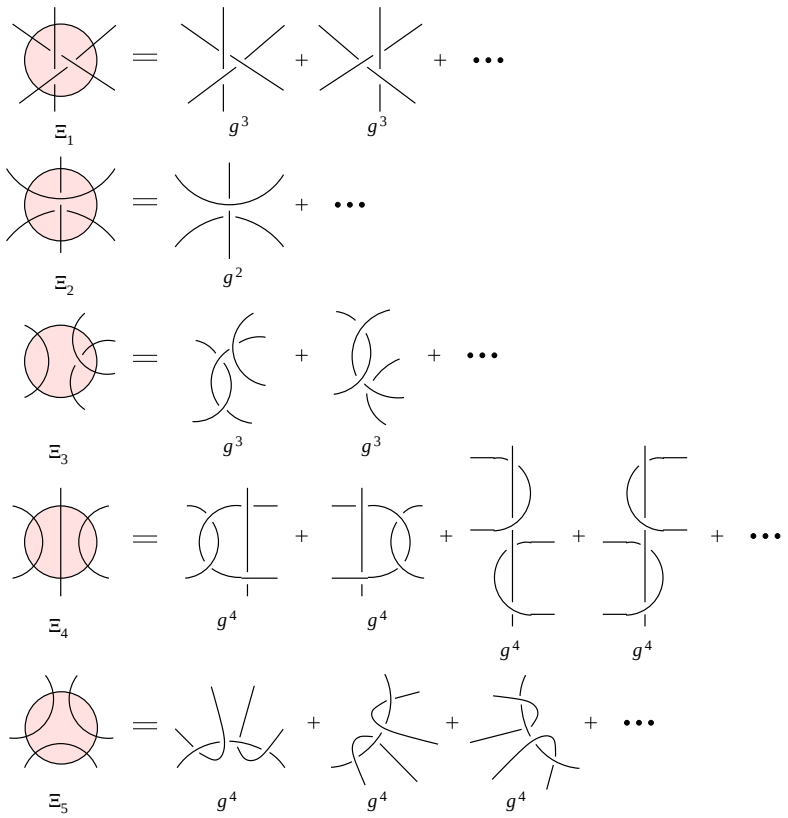


Fig. 16: The five types of tangles with 6 external legs.

We only consider configurations such that no strings can be pulled out altogether ("connected" correlation functions in the language of quantum field theory). Tab. 5 provides the first few orders of the series of the numbers of such objects.

The lowest order in p is explicited in Fig. 16. It is again relatively straightforward to check the correctness of the last entry in each row of Tab. 5 by considering the $n \to \infty$ limit. In the cases of [Ξ_1 with $p \geq 8$], [Ξ_2 with $p \geq 7$], and [Ξ_3 with $p \geq 3$] it is straightforward

	Ξ_1				Ξ_2				Ξ_3					Ξ_4				Ξ_5			
p^k	0	1	2	3	0	1	2	3	0	1	2	3	4	0	1	2	3	0	1	2	3
2	0				1				0					0				0			
3	2				0				2					0				0			
4	0				7				2					4				3			
5	18				6				16	2				8				9			
6	18				53	8			42	2				42	7			41	7		
7	156	24			154	6			171	44	2			156	14			168	21		
8	516	18			609	181	6		748	114	2			608	153	10		663	165	12	
9	2016	598	18		2956	422	6		2877	858	81	2		2850	586	20		3072	740	36	
10	10608	1428	18		11203	3498	318	6	14037	3752	213	2		11918	3445	364	13	13347	3966	438	18
11	40428	12318	1062	18	57664	15330	738	6	61028	19757	2511	131	2	57602	17558	1406	26	63393	20994	2040	54

Tab. 5: Table of the number of prime alternating tangles with 6 external legs.

to show that the diagrams having the highest power of g are just the trivial diagrams decorated by festoons (as in Fig. 15), meaning that the last entry in the corresponding rows should be respectively 18, 6 and 2. For the cases [Ξ_4 with $p \geq 4$] resp. [Ξ_5 with $p \geq 4$] we conjecture that the last entry of each row ξ_p should read, for p even,

$$\xi_p^{(4)} = \frac{1}{2}(3p-4)$$
$$\xi_p^{(5)} = \frac{1}{8}(p+8)(p-2).$$

For p odd we have $\xi_p^{(4)} = 2\xi_{p-1}^{(4)}$ resp. $\xi_p^{(5)} = 3\xi_{p-1}^{(5)}$.

6. Discussion and outlook

In this paper we have shown how to efficiently enumerate alternating tangle diagrams with a given number of connected components and external legs, and we have explained how the flypes can be easily incorporated into our algorithm to count only topologically inequivalent objects. We have illustrated our method with numerical data that are in agreement with the exact solutions at $n = 1$ [5,6,10], $n = 2$ [7], $n = -2$ [10], and the limit $n \to \infty$, and that surpass the general results of [7] by several orders. From a computational point of view, we note that the time required to compute order p grows exponentially with p, but much more slowly than the number of diagrams counted ($\sim 2.7^p$ compared to the number of diagrams $\sim 12^p$).

A remarkable advantage of our transfer matrix algorithms is that they allow to generate planar diagrams with external legs, even in the absence of a line or closed circuit defining an obvious transfer direction.

It should be noticed that our algorithms can be straightforwardly adapted to graphs of any coordination number $q \geq 3$. For the single-step algorithm, it suffices to modify the type 1 transformation so as to insert $q-2$ new points, instead of just two. For odd q, the parity constraints on transformation 2 no longer apply.

The geodesic algorithm can be similarly generalized. For example, on Fig. 8, changing to trivalent vertices would simply imply having five possible transformation rules instead of nine.

These generalizations open several interesting perspectives. One obvious possibility would be to numerically study matter theories defined on random graphs, by exact evaluation of correlation functions [19].

Appendix A. Tangle diagrams up to 4 crossings.

As an illustration we show on Fig. 17 the nine first iterations of the transfer matrix. We restrict ourselves to states which generate diagrams with at most four crossings. The states have all been simplified, as described in Sec. 4. The weight of the trivial state (represented by a cross on the figure) after step $2\ell - 1$ gives the number of two-legged tangle diagrams with exactly ℓ crossings, each connected component being weighed by a factor of n.[6]

[6] By convention, we also give a weight n to components connecting a pair of external legs, though from the point of view of diagrammatic perturbation theory this is, strictly speaking, not correct.

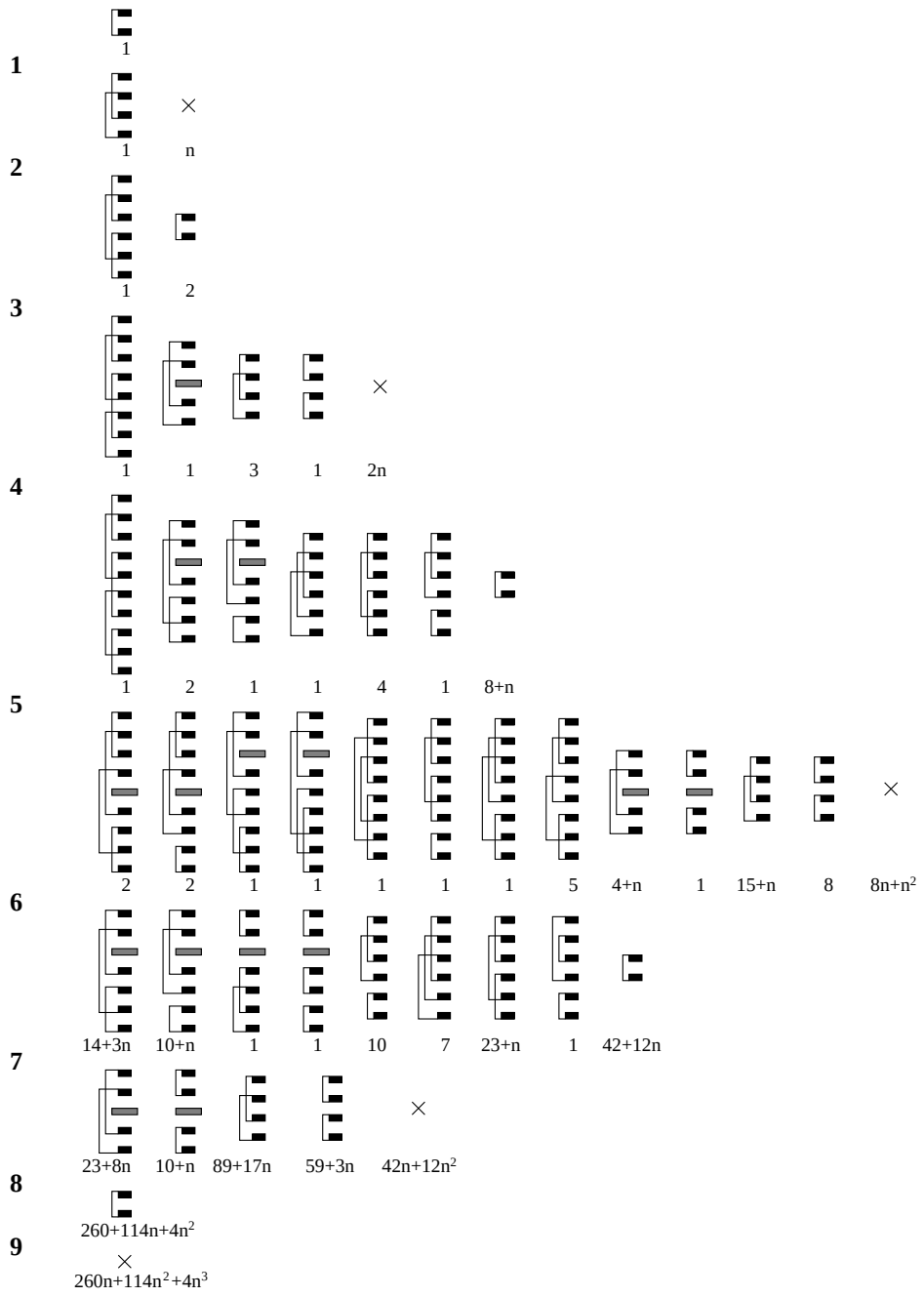


Fig. 17: The list of all intermediate states for tangle diagrams up to four crossings. Below each state we indicate its weight.

References

[1] P. Di Francesco, E. Guitter and C. Kristjansen, *Integrable 2D Lorentzian Gravity and Random Walks*, *Nucl. Phys.* **B 567** (2000) 515–553 (preprint `hep-th/9907084`).

[2] I. Jensen, *Enumerations of Plane Meanders* (preprint `cond-mat/9910313`); *A Transfer Matrix Approach to the Enumeration of Plane Meanders*, *J. Phys. A*, to appear (preprint `cond-mat/0008178`).

[3] P. Di Francesco, E. Guitter and J.L. Jacobsen, *Exact Meander Asymptotics: a Numerical Check*, *Nucl. Phys.* **B 580** (2000) 757–795 (preprint `cond-mat/0003008`).

[4] J. L. Jacobsen and P. Zinn-Justin, *A Transfer Matrix approach to the Enumeration of Knots* (preprint `math-ph/0102015`).

[5] C. Sundberg and M. Thistlethwaite, *The rate of Growth of the Number of Prime Alternating Links and Tangles*, *Pac. J. Math.* **182** (1998) 329–358.

[6] P. Zinn-Justin and J.-B. Zuber, *Matrix Integrals and the Counting of Tangles and Links*, to appear in the proceedings of the 11th International Conference on Formal Power Series and Algebraic Combinatorics, Barcelona June 1999 (preprint `math-ph/9904019`).

[7] P. Zinn-Justin and J.-B. Zuber, *On the Counting of Colored Tangles*, *Journal of Knot Theory and its Ramifications* 9 (2000) 1127–1141 (preprint `math-ph/0002020`).

[8] W.W. Menasco and M.B. Thistlethwaite, *The Tait Flyping Conjecture*, *Bull. Amer. Math. Soc.* **25** (1991) 403–412; *The Classification of Alternating Links*, *Ann. Math.* **138** (1993) 113–171.

[9] P. Zinn-Justin, *Some Matrix Integrals related to Knots and Links*, proceedings of the 1999 semester of the MSRI "Random Matrices and their Applications", MSRI Publications Vol. 40 (2001) (preprint `math-ph/9910010`).

[10] P. Zinn-Justin, *The General O(n) Quartic Matrix Model and its application to Counting Tangles and Links* (preprint `math-ph/0106005`).

[11] V.A. Liskovets, *Sequence A054499* in N.J.A. Sloane (red.), *The On-Line Encyclopedia of Integer Sequences*, published electronically at http://www.research.att.com/~njas/sequences/.

[12] R. Sedgewick, *Algorithms in C* (Addison-Wesley, 1990).

[13] W.T. Tutte, *A Census of Planar Maps*, *Can. J. Math.* **15** (1963) 249–271.

[14] E. Brézin, C. Itzykson, G. Parisi and J.-B. Zuber, *Planar Diagrams*, *Commun. Math. Phys.* **59** (1978) 35–51.

[15] V.A. Kazakov and P. Zinn-Justin, *Two-Matrix Model with ABAB Interaction*, *Nucl. Phys.* **B 546** (1999) 647 (preprint `hep-th/9808043`).

[16] L. Takács, *Enumeration of rooted trees and forests*, *Math. Scientist* **18**, 1–10 (1993).

[17] V. G. Knizhnik, A. M. Polyakov and A. B. Zamolodchikov, *Fractal structure of 2D quantum gravity*, *Mod. Phys. Lett. A* **3**, 819–826 (1988); F. David, *Conformal field*

theories coupled to 2D gravity in the conformal gauge, *Mod. Phys. Lett. A* **3**, 1651–1656 (1988); J. Distler and H. Kawai, *Conformal field theory and 2D quantum gravity*, *Nucl. Phys.* **B 321**, 509 (1989).

[18] I.K. Kostov, *Mod. Phys. Lett.* A4 (1989), 217;
M. Gaudin and I.K. Kostov, *Phys. Lett.* B220 (1989), 200;
I.K. Kostov and M. Staudacher, *Nucl. Phys.* **B 384** (1992), 459.

[19] J. L. Jacobsen and P. Zinn-Justin, work in progress.

6.8 Article "Dense loops, supersymmetry, and Goldstone phases"

En deux dimensions, il existe des relations entre des modèles de boucles et le modèle O(N). La phase de basse température (boucles denses) d'un tel modèle, ou de sa reformulation supersymétrique, peut posséder une phase de Goldstone avec une brisure de symétrie pour $N < 2$. Nous argumentons que cette phase est générique pour $-2 < N < 2$ dès qu'on admet des croisements de boucles. En outre, la phase de Goldstone est différente du modèle de boucles denses sans intersections qui fut étudié par Nienhuis [134]. Nos arguments sont confirmés par des résultats numériques et par un modèle sur réseau exactement résolu par Martins *et al.* [127].

Dense loops, supersymmetry, and Goldstone phases in two dimensions

J.L. Jacobsen,[1] N. Read,[2] and H. Saleur [3]

[1] *Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Bâtiment 100, F-91405 Orsay, France*

[2] *Department of Physics, Yale University, P.O. Box 208120, New Haven, CT 06520-8120, USA*

[3] *Department of Physics and Astronomy, University of Southern California, Los Angeles, CA 90089, USA*

(Date: September 23, 2002)

Loop models in two dimensions can be related to O(N) models. The low-temperature dense-loops phase of such a model, or of its reformulation using a supergroup as symmetry, can have a Goldstone broken-symmetry phase for $N < 2$. We argue that this phase is generic for $-2 < N < 2$ when crossings of loops are allowed, and distinct from the model of non-crossing dense loops first studied by Nienhuis [Phys. Rev. Lett. **49**, 1062 (1982)]. Our arguments are supported by our numerical results, and by a lattice model solved exactly by Martins *et al.* [Phys. Rev. Lett. **81**, 504 (1998)].

It has long been understood that theories of an N-component scalar field ϕ with O(N) symmetry in Euclidean space of dimension d can be related to statistical-mechanics models in which configurations of loops are given Boltzmann weights depending on their lengths and intersection properties [1]. In the $N \to 0$ limit, unwanted closed loops vanish [1], and the usual self-avoiding walks (polymers) are related to the generic critical point in the O($N \to 0$) theory. The $N \to 0$ limit can be avoided by using a model with a Lie superalgebra as symmetry ("supersymmetry") [2]. Other values of N are also of interest. In $d = 2$, exact results based on lattice models [3] show that there also exists a massless low-temperature (low-T) regime of the model, for $-2 < N \leq 2$. The properties of this phase—in which, geometrically, the loops become dense—are well-understood using Coulomb gas techniques [3].

In this Letter, we revisit the low-T phase of the O($N \leq 2$) models in $d = 2$. The supersymmetry approach, which can be generalized to all integer N, makes it clear that for $N < 2$ there should be a massless phase described by the Goldstone modes of the broken O(N) or supersymmetry, with very simple scaling dimensions. We argue that the usual dense-loops phase [3], in which loops never intersect, is not this Goldstone phase (in contrast to earlier expectations [4]): in fact, it is not generic, and an arbitrarily-weak perturbation that allows loop intersections causes a cross-over to the generic Goldstone phase for $-2 < N < 2$. We present numerical results that support our interpretation. There is also a soluble model that appears to be in the Goldstone phase [5].

The generic continuum O(N)-invariant action S for the scalar field ϕ includes the interaction term $-\lambda(\phi \cdot \phi)^2$ (by convention the Boltzmann weight is e^S). In terms of loops, this is the model introduced by de Gennes [1], in which the loops can cross, but for $\lambda > 0$ these crossings are disfavored. It has a second-order transition for $d > 2$, but for $d = 2$, a transition occurs only for $N \leq 2$; for $N > 2$, there is only the high-T or massive unbroken-symmetry phase. (For N not a positive integer, the theory is supposed to be defined by analytic continuation of its perturbation expansion. For N integer, this can be made rigorous by the use of supersymmetry, as we describe below.) When $-2 \leq N \leq 2$, the phase transition is second order, and the associated critical exponents have been conjectured by Nienhuis [3] and others [6]. For $N < -2$, the transition is expected to be first order.

The existence of the transition for $N \leq 2$ implies the existence of a low-T phase. From the point of view of the generic O(N) ϕ^4 theory, at low T the symmetry is broken to O($N-1$) according to Landau theory, and the low-T phase would be expected to be a Goldstone phase with massless excitations, described by a nonlinear sigma model with target manifold $\mathrm{O}(N)/\mathrm{O}(N-1) \cong S^{N-1}$, a sphere. For $N = 2$, the O(2) symmetry is not really broken, but there is a power-law phase with continuously-varying exponents, the Berezinskii-Kosterlitz-Thouless (BKT) phase [7]. For $N < 2$, use of the perturbative beta function in the S^{N-1} sigma model, defined by analytic continuation in N, implies that the coupling between the Goldstone modes in the sigma model becomes weak at large length scales [8], like Goldstone phases in higher d. This yields a low-T phase with simple forms for the scaling dimensions. This is possible in $d = 2$ for $N < 2$ because the Mermin-Wagner theorem no longer applies when the continuous SO(N) symmetry cannot be realized as unitary operations on a vector field. We discuss this theory more fully below, using supersymmetry.

In contrast, Refs. [9,3] defined a particular lattice model of strictly non-crossing loops. This was done by truncating the high-T expansion of a lattice version of the ϕ^4 theory as above [9], in order to simplify calculations of critical properties while hopefully remaining in the same universality class. The model has Boltzmann weight

$$e^S \equiv \prod_{\langle ij \rangle} \left(1 + K \vec{S}_i \cdot \vec{S}_j\right), \qquad (1)$$

where i, j are vertices on the honeycomb graph, $\langle ij \rangle$ denotes an edge of the graph, and the variables $\vec{S}_i$ are N-component fixed-length spins, $\vec{S}_i^2 = N$. The partition function can be evaluated as $Z = \sum N^L K^E$ where the sum is over graphs consisting of self-avoiding mutually-avoiding loops, L is the number of loops, E the number of edges they occupy, and $K \sim 1/T$. The critical K for

this model is known exactly, $K_c = \left(2+\sqrt{2-N}\right)^{-1/2}$ for $N \leq 2$ [3]. The transition is second-order for $-2 \leq N \leq 2$ only, and the critical exponents at this critical point, which is known as *dilute loops* (or as self-avoiding walks for $N \to 0$), are known by a variety of techniques [3,6]. The low-T region for each $-2 < N < 2$ is attracted to a unique massless phase (that varies with N) for all $K_c < K < \infty$. This phase has nontrivial scaling dimensions, which are again known [3,10]. It has become known as *dense loops* (or as dense polymers for $N \to 0$). In the limit $N \to 2$, this phase coincides with the critical endpoint of the BKT phase (which appears in the $N = 2$ model for $K \geq K_c$). For $N < -2$, we expect that the dense-loops phase is massive.

Both the critical and dense-loops phases are believed to be universal as continuum phases, in the sense that they are independent of the choice of lattice used in defining them. Recent work [11] on the $T = 0$ limits of certain lattice models identifies other massless phases of non-intersecting *fully-packed* loops, that depend on the lattice used. We are not concerned with these here, nor with the transition for $N > 2$ in the region $|K| > 1/N$, where the Boltzmann weight (1) can be negative [12].

From the critical exponents for dilute loops, it is natural to hope [3], that they are in the same universality class as the transition in the generic O(N) ϕ^4 theory. The idea is that any weak repulsion of the lines leads to a crossover to the infinitely-repulsive limit in which the loops never cross, so that the latter fixed point governs the transition for all $\lambda > 0$. We know of no reason to doubt this. As a check, we can reintroduce intersections of the lines in the dilute critical theory. The graphs of the expansion of the generic O(N) model that are not present in the truncated model would involve in particular crossings of the loops, or multiple occupancies of edges. In the continuum description, the most relevant (and most generic) of these is where two lines cross. This may be included as a perturbation of the dilute critical theory by adding to the action the integral of a multiple of the corresponding scalar operator, the so-called 4-leg operator, which has conformal weight [6] $h_4 = g - (g-1)^2/(4g)$, where we parameterized $N = -2\cos\pi g$, $g \in [1,2]$. We see that this operator is marginal at $N = 2$ and irrelevant (i.e. $h_4 > 1$) for $-2 \leq N < 2$. This strongly suggests that allowing more general interactions or configurations of loops in the lattice model would not change the universality class, at least when the perturbations are small, and hence that the *dilute* critical point [3] is indeed generic.

The model of non-intersecting loops was obtained by truncating the high-T expansion, and this would be expected not to change the behavior for sufficiently high T (small K); apparently this remains true down to $T = T_c$. However, this leaves open the possibility that the low-T behavior of the models is different, which we will now examine. The dense-loops phase is described by a theory similar to that of the dilute case, with the same parameterization of N and h_4 as above, but with $g \in [0,1]$. One sees that the 4-leg operator is now relevant in the whole region $-2 < N < 2$. As $N \searrow -2$, $h_4 \searrow -\infty$. Since the 4-leg perturbation is relevant in the low-T phase, it is *dangerously* irrelevant in the critical theory.

It is thus clear that the *dense*-loops phase is non-generic from an O(N)-model point of view. In $d > 2$, de Gennes' model appears generic, as the topological effect of strict non-intersection, present in $d = 2$, is absent (though in $d = 3$, there remains the topology of knots and linking instead). Even a model in $d > 2$ of strict non-intersection, when reduced to $d = 2$ by confining $d-2$ dimensions to finite intervals, gives an effective model in which intersections still occur. There is also a symmetry interpretation for the non-genericity: the dense phase of strictly non-intersecting loops has a larger symmetry, U(N) instead of O(N), which is based on the possibility of consistently orienting all the loops in the partition sum [8,3]. Though this symmetry may be broken by the boundary conditions, the massless dense-loops phase appears to be most easily understood in terms of a non-linear sigma model with this larger symmetry [8]. The 4-leg operator is a symmetry-breaking perturbation, that reduces the symmetry to the generic O(N). It would be surprising if the large-length-scale effect of this relevant symmetry-breaking perturbation led back to the same higher-symmetry phase. We note that the importance of the self-crossings has been observed in the related context of Lorentz lattice gases [13].

We now argue that the flow induced by the 4-leg perturbation leads to the generic Goldstone phase of the ϕ^4 theory, and begin by examining the structure of this phase, before giving numerical evidence that such a flow does in fact occur.

The loop (high-T) expansion of the O(N) model can be reproduced by considering instead a model with OSp($m|2n$) symmetry, with $N = m - 2n$; in particular, each closed loop incurs a factor of N [2,8]. The model then makes complete sense even for N non-positive, (but N must be an integer). The phase diagram, beta functions, and scaling dimensions are the same for all m at fixed N (though multiplicities of operators may vanish for m small [8]). The Goldstone phase should thus be described by a sigma model with target space $\mathrm{OSp}(m|2n)/\mathrm{OSp}(m-1|2n) \cong S^{m-1|2n}$, a supersphere, and a single coupling constant g_σ [8,14]. The perturbative beta function is, to leading order, $\beta \equiv dg_\sigma/d\ln L \propto (N-2)g_\sigma^2$, so for $N < 2$, the model flows to weak coupling for $g_\sigma \geq 0$ (which is the expected physical sign). Again, the Mermin-Wagner theorem does not apply here, and symmetry-breaking is allowed for $N < 2$. The weak-coupling fixed point is a theory of free massless scalars, with $m-1$ bosonic and $2n$ fermionic components, so is conformal with central charge $c = m-1-2n = N-1$, as it would be for S^{N-1}. There will be logarithmic corrections

due to the marginally-irrelevant coupling g_σ. We emphasize that in the sigma model, OSp$(m|2n)$ symmetry does not allow any other marginal or relevant couplings to be added to the action, so that this fixed point is robust, unlike the dense-loops theory above.

Explicitly, the target manifold can be parameterized by commuting coordinates x_i, $i = 1, \ldots, m$, and anticommuting coordinates η_j, $j = 1, \ldots, 2n$, subject to the constraint $\sum_{i=1}^m x_i^2 + \sum_{j=1}^n \eta_{2j-1}\eta_{2j} = 1$. The action of the sigma model is

$$S = -\frac{1}{g_\sigma} \int d^2r \left[\sum_{i=1}^m (\partial_\mu x_i)^2 + \sum_{j=1}^n \partial_\mu \eta_{2j-1} \partial_\mu \eta_{2j} \right] . \quad (2)$$

The simplest example is $m = n = 1$ ($N = -1$), which corresponds to the $S^{0|2}$ supersphere. The constraint can be solved as $x_1 = 1 - \frac{1}{2}\eta_1\eta_2$. After elimination of x_1 and some rescaling, one finds

$$S = -\int d^2r \left[\partial_\mu \eta_1 \partial_\mu \eta_2 - \frac{g_\sigma}{2} \eta_1\eta_2 \partial_\mu \eta_1 \partial_\mu \eta_2 \right] . \quad (3)$$

At long wavelengths the coupling g_σ renormalizes towards zero, and the theory becomes free symplectic fermions, with central charge $c = -2$, as claimed. In general, for $N < 1$, the partition function of the fixed-point theory vanishes, $Z = 0$. Similar behavior is well-known in the dense-polymer ($N = 0$) phase.

The supersymmetry analysis is valid for N integer. For general N, we are forced again to interpolate formally. The availability of the supersymmetry results greatly strengthens our confidence in this procedure and in the existence of the Goldstone phase for all $N < 2$.

Both at the dilute critical point and in the dense-loops phase, the central charge is $c = 1 - 6(g-1)^2/g$, in the parameterization above. For dense loops, $g \in [0, 1]$, this coincides with $c = N - 1$ for $N = 1, 2$. The $N = 2$ case is the BKT phase with $c = 1$ throughout, and not of interest here. We claim that, for all $-2 < N < 2$, the perturbed dense-loops phase flows to the Goldstone phase, which is a distinct massless phase even when it has the same c. For $1 < N < 2$, c decreases during the flow, but for $-2 < N < 1$ it increases. This latter behavior would not violate Zamolodchikov's c-theorem, as the theories involved are non-unitary. In particular, the $N = 0$ ($c = -2$) dense-polymer theory should flow to a $c = -1$ free theory. Note that for $N = 1$ the partition function is $Z = 1$ (after removing nonuniversal constants) in both the dense-loops and Goldstone phases, a consequence of $c = 0$.

Some particularly interesting operators in loop models in general are the k-leg operators (also known as fuseaux or watermelon operators), insertions of which give the probability that k lines terminate at the same point. In the general O(N) point of view, the leading such operator would be represented by k insertions of the field ϕ at the same point (with no derivatives), and must necessarily be in a totally-symmetric rank-k tensor representation of O(N), which can be assumed to be traceless on all pairs of indices. For the OSp$(m|2n)$ formulation, this becomes a supersymmetric tensor, that vanishes when contracted using the same invariant bilinear form as in the constraint in the sigma model. It is useful to choose the index values of the set of tensors in a correlation function to occur in distinct pairs, to force the lines in the loop model to connect the positions of the operators in some specified way. This is possible if m is sufficiently large, a typical situation in applications of supersymmetry.

At the dilute critical point and in the dense-loops phase, there are nontrivial scaling dimensions X_k (or conformal weights $h_k = X_k/2$) for the k-leg operators, which are known exactly [6,10] (for example, h_4 was given earlier). (In these phases, the symmetric-tensor k-leg operators are in general degenerate with others [8], and in the dense-loops phase, we assume there is a singlet 4-leg operator, which we used as the perturbation above; it is the most relevant invariant scalar operator.) In the Goldstone phases, on the other hand, the symmetric tensors correspond to functions like spherical harmonics on the target space, and their scaling dimensions tend to zero as $g_\sigma \to 0$. Since the symmetry is broken for $N < 2$, the precise form of each operator depends on which components are chosen. If the expectation of the field is $x_1 = 1$, all other x_i and $\eta_j = 0$, then any component of the k-leg operator can be rewritten in terms of the (rescaled, free) $m-1$ x_i's and the $2n$ η_j's, and the correlation functions will contain various logarithmic factors depending on which components are chosen. It is clear that the scaling dimensions of the leading k-leg operators are all zero, and there will be subleading operators with integer conformal weights (the precise form of correlation functions may require more attention to the limit $g_\sigma \to 0$). When the dense-loops phase is perturbed by the 4-leg operator, it should be possible to see the effective scaling dimensions of the k-leg operators (say, in two-point functions) cross over to zero, for all k.

Numerical transfer matrix calculations were done on a square-lattice cylinder of circumference L. The edges of the lattice are packed with one line per edge; two lines can meet at each vertex in three ways, the weight being 1 for a non-crossing and w for a crossing. The natural O(N)-invariant boundary condition for a periodic system is that the fields ϕ are periodic, both for Bose and Fermi components. This implies that a loop that winds the periodic direction is given the same factor N as a topologically trivial loop. In a conformally-invariant system, c can be extracted from the scaling of the free energy per vertex, $f(L) = f(\infty) - \pi c/(6L^2)$. We therefore use this formula to extract the effective value of c in a crossover, using $f(L)$ at pairs of values of L. Results for $N = 0$ (see Fig. 1) show a clear trend away from $c = -2$, and are consistent with a slow approach to $c = -1$ for $L \to \infty$.

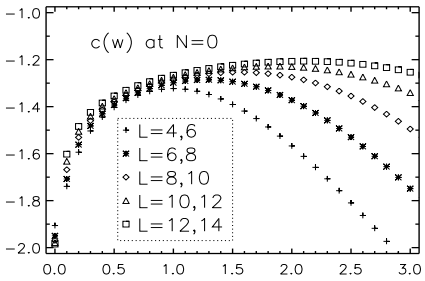


FIG. 1. Measures of the central charge vs. w for $N = 0$. The conjectured value for $L \to \infty$ is $c = -1$ for all $w > 0$.

For the k-leg scaling dimensions, we use the lowest transfer matrix eigenvalues E_k in each sector with k lines propagating along the cylinder. From conformal invariance [15], we define effective values of X_k at finite L from $E_k - E_0 = 2\pi X_k/L$; see Fig. 2 ($k = 2$ is not included as, for $N = 0$ in our model, $E_2 = E_0$ for all L and w). The results are consistent with $X_k = 0$ in the fixed-point theory. In a situation with a marginally-irrelevant operator, one expects [15] corrections to scaling to be $O(1/L \ln L)$, which presumably explains the slow convergence.

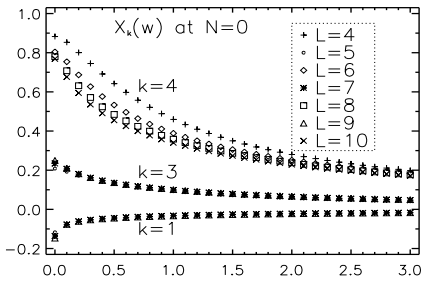


FIG. 2. Measures of the k-leg exponents $X_k = 2h_k$ vs. w for $N = 0$.

A useful check of our analysis is obtained by modifying the boundary conditions, in a similar way as in Ref. [8]. If we make the Fermi components of the ϕ fields obey antiperiodic boundary conditions, this will give a non-contractible (winding) loop the usual weight $w_e = m - 2n = N$ if its winding number is even, and $w_o = m + 2n = N + 4n$ if it is odd. In the Goldstone phase, this will give the n pairs of symplectic fermions the antiperiodic boundary condition, while leaving the bosonic scalar fields periodic. In the formula for $f(L)$, c will be replaced by $c - 24h_{\rm tw}$, where $h_{\rm tw}$ is the lowest conformal weight in the sector with boundary conditions twisted as described. We expect $c - 24h_{\rm tw} = m + n - 1$, which can be understood as the contribution of the twist operator of weight $h = -1/8$ for each symplectic fermion pair. Thus we expect

$$c - 24h_{\rm tw} = N - 1 + 3n = \frac{3w_o + w_e}{4} - 1, \qquad (4)$$

for general N, n. This appears to be confirmed numerically for $N = w_e = 0$ and several values of w_o in Fig. 3. Notice how this differs from the usual loop gas behavior ($w = 0$), where $c - 24h_{\rm tw} = 1 - 6\arccos^2(\frac{w_o}{2})/(\pi^2 g)$ [16].

For $N < -2$, we believe that the dense-loops phase is massive, and hence flows to the generic Goldstone phase only when w is greater than a positive critical value.

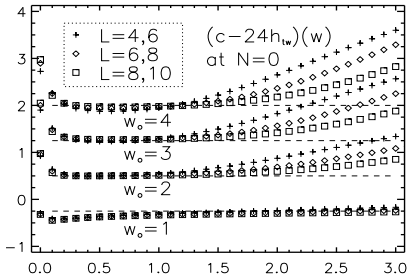


FIG. 3. Measures of $c - 24h_{\rm tw}$ vs. w for $N = 0$. Dashed lines show the prediction of Eq. (4).

Finally, it is important that a version of the OSp($m|2n$)-invariant model on the square lattice is integrable at $w = (2 - N)/4$ and can be solved by the Bethe ansatz [5]. The model is massless for $N < 2$, and there is strong evidence that the central charge is $c = N - 1$, and that the lowest scaling dimension, other than that of the identity operator, is zero; the authors conjecture that these are the exact values, but do not identify the phase with the Goldstone phase.

To conclude, we have argued that crossing of loops is a relevant perturbation in the dense-loops low-T phase for $-2 < N < 2$, and that the long-distance behavior is governed by the Goldstone phase instead, by analytical, numerical, and exact approaches.

This work was supported by the NSF under grant no. DMR-98-18259 (NR), and by the DOE (HS).

[1] P.G. de Gennes, Phys. Lett. A **38**, 339 (1972).
[2] G. Parisi and N. Sourlas, J. de Physique Lettres **41**, L403 (1980).
[3] B. Nienhuis, Phys. Rev. Lett. **49**, 1062 (1982).
[4] H. Saleur, Phys. Rev. B **35**, 3657 (1987).
[5] M.J. Martins, B. Nienhuis, and R. Rietman, Phys. Rev. Lett. **81**, 504 (1998).
[6] H. Saleur, J. Phys. A **19**, L807 (1986).
[7] V.L. Berezinskii, Sov. Phys.-JETP **34**, 610 (1971); J.M. Kosterlitz and D.J. Thouless, J. Phys. C **6**, 1181 (1973).
[8] N. Read and H. Saleur, Nucl. Phys. B **613**, 409 (2001).
[9] E. Domany, D. Mukamel, B. Nienhuis, and A. Schwimmer, Nucl. Phys. B **190**, 279 (1981).
[10] B. Duplantier and H. Saleur, Nucl. Phys. B **290**, 291 (1987).

[11] J.L. Jacobsen and J. Kondev, Nucl. Phys. B **515**, 701 (1998), and references therein.
[12] W. Guo, H.W.J. Blöte, and F.Y. Wu, Phys. Rev. Lett. **85**, 3874 (2000).
[13] M.S. Cao and E.G.D. Cohen, J. Stat. Phys. **87**, 147 (1997).
[14] H. Saleur, B. Wehefritz-Kaufmann, cond-mat/0112095, to appear in Nucl. Phys. B.
[15] J. Cardy, J. Phys. A **19**, L1093 (1986).
[16] J. Cardy, Phys. Rev. Lett. **84**, 3507 (2000).

Chapitre 7

Couplages aléatoires

7.1 Les différents types de désordre

Dans la nature, aucun matériau n'est parfaitement pur, même si l'on peut produire des cristaux avec une très faible concentration d'impuretés. Il est donc important d'étudier théoriquement comment une faible quantité d'impuretés peut influencer un matériau et en particulier changer le comportement critique d'un système magnétique.

L'effet d'impuretés peut être modélisé de plusieurs manières. Par exemple, la présence d'un *champ magnétique aléatoire* modifie fortement le comportement d'un modèle de spins : ce cas a été particulièrement étudié dans le contexte du modèle d'Ising. Ici nous allons uniquement discuter du cas de *couplages aléatoires*, c'est-à-dire $K \to K_{ij}$ dans (1.1). On distingue deux cas :

- Désordre *recuit* (les impuretés peuvent se déplacer). La moyenne statistique sur la distribution de désordre se fait alors au niveau de la fonction de partition.
- Désordre *gelé* (les impuretés sont figées). La moyenne sur le désordre est faite dans les quantités thermodynamiques extensives, telles que l'énergie libre.

C'est ce dernier cas, le plus difficile, qui nous intéresse ici.

Une autre distinction porte sur la distribution des couplages K_{ij}. Le cas d'un mélange de couplages positifs et négatifs mène à une situation intéressante où les effets du désordre sont amplifiés par de la frustration [93]. Ici, nous nous contraignons au cas plus simple des couplages purement ferromagnétiques, $K_{ij} > 0$. Ce cas est réalisé physiquement, par exemple, dans des expériences d'adsorption (d'un gaz noble sur un film mince) ou de déposition (du type epitaxie par faisceau).

7.2 Cas du modèle de Potts

Si le système sans désordre possède une transition de phase, l'effet des impuretés sur celle-ci dépend de l'ordre de la transition. Pour une transition de second ordre, un critère dû à Harris [76] prédit que les exposants critiques sont changés si, et seulement si, la chaleur spécifique C_V diverge à la transition du système pur. Inversement, les exposants critiques du système désordonné prennent alors des valeurs de manière à ce que C_V ne diverge pas [36, 37]. Si, par contre, le système sans désordre a une transition de premier ordre, une concentration infinitésimale de désordre agit de manière à produire une nouvelle transition de second ordre [2].

Dans ce contexte, le modèle de Potts bidimensionnel fournit un cadre particulièrement adapté à étudier les conséquences de ces prédictions théoriques, car il possède à la fois des transitions de phase du premier ($q > 4$) et du second ordre ($q \leq 4$). En outre, le critère de Harris prédit que le désordre est une perturbation pertinente pour $2 < q < 4$. Le cas $q = 2$ est marginal.

Très récemment, une très bonne revue [21] résume le travail fait par plusieurs groupe pendant les dernières deux décennies. Nous allons nous contenter d'en résumer quelques points principaux :

1. L'hypothèse d'universalité suggère que les propriétés critiques ne dépendent pas de la distribution de $K_{ij} > 0$. Ceci est confirmé par les simulations numériques qui étendent aussi l'universalité aux modèles de dilutions d'arêtes ou de sommets.

2. De manière générale, le désordre agit de manière à donner aux exposants critiques un spectre multifractal : les différents moments d'un corrélateur possèdent des exposants critiques indépendants. Par ailleurs, les corrélateurs sont non auto-moyennants.
3. Pour $2 < q < 4$, l'effet du désordre peut être traité perturbativement avec des techniques analogues à celles de la régularisation dimensionnelle de la théorie ϕ^4 en $4-\epsilon$ dimensions [44, 45, 123, 124, 48]. Actuellement, un grand nombre d'exposants multifractaux a été calculé par cette technique.
4. Des simulations numériques ont permis de valider les résultats du développement perturbatif. En particulier, elles ont établi que le point fixe aléatoire ne possède pas de brisure spontanée de la symétrie de répliques.
5. Pour $q > 4$, seuls des résultats numériques existent. Ils montrent que les exposants critiques sont des fonctions non triviales de q possédant aucune singularité à $q = 4$ [82].
6. La limite $q \to \infty$ a suscité un intérêt particulier [87, 101]. Elle possède des exposants critiques non triviaux.
7. Un diagramme de phase [30] prévoit une ligne de points tricritiques en dimension $d > 2$. En particulier, le point fixe aléatoire n'est atteint qu'au delà d'un seuil fini pour la concentration du désordre. Cette prédiction fut vérifiée en $d = 3$ très récemment [77, 35].

7.3 Article "Critical behavior of random-bond Potts models"

Dans le cadre du modèle de Potts à q états, nous étudions l'effet de désordre gelé sur un système ayant une transition du premier ordre. Pour $q \gg 1$, une relation avec le modèle d'Ising en champ aléatoire explique l'absence de chaleur latente en deux dimensions et suggère que pour $d > 2$ de tels systèmes possèdent un point tricritique avec un exposant ν relié à ceux du modèle en champ aléatoire par $\nu = \nu_{\mathrm{RF}}/(2 - \alpha_{\mathrm{RF}} - \beta_{\mathrm{RF}})$. En $d = 2$, nous analysons le modèle par des lois d'échelle et par l'invariance conforme : nous trouvons une transition continue avec un rapport β/ν qui varie continûment en fonction de q et avec un exposant $\nu \approx 1$ à faible variation. Nous trouvons une évidence forte pour l'existence des lois d'échelles multiples au niveau des fonctions de corrélations.

Critical Behavior of Random-Bond Potts Models

John Cardy[1,2] and Jesper Lykke Jacobsen[1,3]
[1]*University of Oxford, Department of Physics-Theoretical Physics, 1 Keble Road, Oxford OX1 3NP, United Kingdom*
[2]*All Souls College, Oxford OX1 4AL, United Kingdom*
[3]*Institute of Physics and Astronomy, University of Aarhus, DK-8000 Aarhus, Denmark*
(Received 6 May 1997)

The effect of quenched impurities on systems undergoing first-order phase transitions is studied within the framework of the q-state Potts model. For large q a mapping to the random-field Ising model explains the absence of any latent heat in 2D, and suggests that for $d > 2$ such systems exhibit a tricritical point with an exponent ν related to those of the random-field model by $\nu = \nu_{\mathrm{RF}}/(2 - \alpha_{\mathrm{RF}} - \beta_{\mathrm{RF}})$. In 2D we analyze the model using finite-size scaling and conformal invariance, and find a continuous transition with a ratio β/ν which varies continuously with q, and a weakly varying exponent $\nu \approx 1$. We find strong evidence for the multiscaling of the correlation functions. [S0031-9007(97)04550-X]

PACS numbers: 05.70.Jk, 64.60.Ak, 64.60.Fr

Although the effect on the critical behavior of adding quenched bond randomness to classical systems whose pure version undergoes a continuous phase transition is well understood in terms of the Harris criterion [1], the analogous situation when the pure transition is first order is less well studied. Following earlier work of Imry and Wortis [2], Aizenman and Wehr [3] and Hui and Berker [4] argued that in 2D any amount of randomness should lead to a vanishing of the latent heat. The arguments leading to this conclusion are analogous to those used by Imry and Ma [5] for the absence of any spontaneous magnetization in the random *field* Ising model (RFIM) for $d = 2$: the bond randomness couples to the local energy density, which is different for the coexisting phases of the pure model, in the same manner that the random field couples to the local magnetization of the RFIM. The vanishing of the latent heat should be accompanied by a divergent correlation length, and, if so, the question arises as to which universality class(es) the corresponding continuous transition belongs. A suitable model in which to study this is the q-state Potts model, whose pure version in 2D undergoes a first-order transition for $q > 4$, otherwise being continuous. Chen, Ferrenberg, and Landau [6] undertook an extensive Monte Carlo investigation of the case $q = 8$. In addition to confirming the continuous nature of the transition, these authors extracted numerical values of the critical exponents which appear to be consistent with those of the *pure* 2D Ising model. Similar values have also been claimed for the case $q = 4$, when the pure transition is continuous [7]. This disagrees with the predictions of Ludwig and Cardy [8], Ludwig [9], and Dotsenko *et al.* [10], who find a new random fixed point for $q > 2$, based on an expansion in powers of $q - 2$.

As will become clear, many of our results generalize, but let us for definiteness consider a Potts model on the square lattice with degrees of freedom s_i taking q values, and a reduced Hamiltonian $-\sum_{ij} K_{ij}\delta_{s_i s_j}$, where the sum is over nearest neighbor pairs. The ferromagnetic couplings K_{ij} are quenched random variables, taking the values K_1 and K_2, each with probability $\frac{1}{2}$. When $(e^{K_1} - 1)(e^{K_2} - 1) = q$ this model is, on average, self-dual, and, if the transition is unique, is therefore at its critical point [11]. It is useful to parametrize $e^{K_{ij}} - 1 = u_{ij} = q^{(1/2)+w_{ij}}$, where $w_{ij} = \pm w$, and $w > 0$ measures the strength of the randomness. The partition function of this model may be mapped onto that of the random cluster model [12], in which each bond of the lattice is either occupied, when it is counted with weight u_{ij}, or empty, in which case it is counted with weight 1. The partition sum is over all such configurations, in which each connected cluster of sites is weighted by a factor q. Let us first consider the pure model, with $w = 0$. In the limit $q \to \infty$, the sum over configurations is dominated by only two: the *empty* lattice, in which no bonds are occupied, which contributes a factor q^N, where N is the total number of sites, and the *full* lattice, with a weight $(\sqrt{q})^{2N}$, since the number of bonds is $O(2N)$. All other configurations are down by powers of $q^{1/2}$. At the self-dual point, there are therefore two coexisting states with identical bulk free energy and different internal energy densities, indicating, as expected, that the transition has a nonvanishing reduced latent heat per bond $\sim\frac{1}{2}\ln q$. For the pure model, this analysis may be extended to take into account higher order corrections in $q^{-1/2}$, with no essential change in the physical picture. Now consider an *interface* between these two phases. For large q, the lowest energy interface is parallel to a lattice direction, say the x axis, and is such that all the bonds with $y \le$ some integer are occupied, and those above this are empty (or vice versa). There will also be entropic fluctuations $y = h(x)$ of this interface, described by the usual solid-on-solid interfacial Hamiltonian, proportional to the length of the interface. The interfacial tension for large q is $\sigma \sim \frac{1}{4}\ln q$, independent of the local shape of the interface. This is to be compared with $\sigma \sim 2J$ between the *ordered* phases of a low temperature Ising model with reduced exchange coupling J.

Now consider the effect of adding bond randomness to the random cluster model. Each configuration of the interface will be weighted by an energy $\sum_x \sum_{y<h(x)} \times w(x,y) \ln q$, where (x,y) labels bond positions. This may be rewritten, up to a term independent of $h(x)$, as $\frac{1}{2}\sum_x (\sum_{y<h(x)} - \sum_{y>h(x)}) w(x,y) \ln q$. Compared with the energy of an RFIM interface with spins $s(x,y) = \pm 1$ coupled to a reduced random field $h(x,y) = \pm h_{\rm RF}$, the interfacial models are identical with the correspondence $J \leftrightarrow \frac{1}{8} \ln q$, and $h_{\rm RF} \leftrightarrow \frac{1}{2} w \ln q$. In addition, the imposition of a uniform reduced magnetic field h on the RFIM, which distinguishes between the two coexisting phases, is seen to be equivalent to a deviation $t \equiv (T - T_c)/T_c$ in the *temperature* variable away from the critical self-dual point. Since this couples to the energy density we find the correspondence $h \leftrightarrow \frac{1}{4} t \ln q$.

Of course, this is strictly valid only as $q \to \infty$. At finite q the q dependence of cluster configurations with more complicated topologies is not simply accounted for by the interfacial tension. For the same reason, the mapping is not between *bulk* configurations of the two models. However, it will be argued that certain universal properties are controlled by an a renormalization group (RG) fixed point at infinite q, and for these the mapping should be asymptotically exact. Although this has been described in terms of a 2D self-dual model, it should be clear that it is more general: lack of self-duality corresponds to a skewness in the distribution of the random fields $h(x,y)$, which may be compensated by adding a suitable uniform field (corresponding to a shift in the T_c of the Potts model), and, similarly, higher dimensions may be taken into account by appropriately replacing $\sqrt{q}$ by $q^{1/d}$.

The RG properties of the interface in the RFIM near $d = 2$ have been well studied [13]. When translated into the variables of the random-bond Potts model, the flow equations have the form

$$dw/dl = -(d/2 - 1)w + Aw^3 + \ldots , \qquad (1)$$

$$d(\ln q)^{-1}/dl = -(\ln q)^{-1}[(d-1) - Aw^2 + \ldots], \qquad (2)$$

$$dt/dl = t(1 + Aw^2 + \ldots) , \qquad (3)$$

where $A > 0$ is a nonuniversal constant. Corrections to these equations are supposed to be higher order in w and in $q^{-1/2}$. The RG flows for $d > 2$ and the consequent phase diagram are shown in Fig. 1. In the pure models, for $q >$ some $q_2(d)$ (low T in the RFIM), there is phase coexistence with a nonvanishing latent heat (spontaneous magnetization), controlled by a fixed point at infinite q ($T = 0$). For $d > 2$ this persists into the shaded region, bounded by a line of tricritical points where the latent heat vanishes. The universal behavior along this line is controlled by the fixed point R at $w = O((d-2)^{1/2})$ and infinite q. Using the correspondence $t \leftrightarrow \frac{1}{2} h \cdot T_{\rm RF}$, we conclude that the thermal eigenvalue ν^{-1} of the random-bond problem is related to the RG eigenvalues and exponents of the RFIM by the scaling relations $\nu^{-1} = y_h - \theta = (2 - \alpha_{\rm RF} - \beta_{\rm RF})/\nu_{\rm RF}$, where $-\theta$ is the eigenvalue

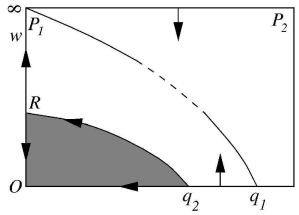


FIG. 1. Schematic phase diagram in the critical surface for $d > 2$. q increases to the left, and w is the disorder strength, with P_1P_2 being the percolation limit. RG flows are indicated. The latent heat is nonvanishing within the shaded region, and elsewhere the transition is continuous, controlled by the line of fixed points P_1q_1.

which controls the irrelevance of $T_{\rm RF}$ and the consequent violation of hyperscaling [13]. In the same manner, it may be shown that the latent heat vanishes as $(w_c - w)^{\beta_{\rm RF}}$ as the line Rq_2 is approached from below. Of course, these relations have been established only close to $d = 2$, but, if the topology of the RG flows does not change, they should hold also in three and higher dimensions.

Above the line Rq_2, the flows go to large w beyond the validity of (1)–(3). In addition, the renormalized interfacial tension flows to zero and the mapping between the models breaks down as domains of different topologies proliferate. However, for infinite q the mapping remains exact and the flows go to infinite w. This cannot happen for finite q since this is the percolation limit $K_1/K_2 = 0$, at which w^{-1} is *relevant* [14]. There must therefore exist another line of stable fixed points emerging from P_1, which control the universal continuous transition for large, but finite, values of w and q. It is tempting to conjecture, as indicated by the dashed line in Fig. 1, that this connects onto that found by expansion in powers of $q - q_1$ [8], where q_1 is the point where the exponent α of the pure model changes sign [1]. Our analysis indicates that, at least for $d = 2$, this is the case. In 2D (when $q_1 = 2$, $q_2 = 4$), the shaded region collapses, and for any nonzero w the renormalized interfacial tension, and thus the latent heat, vanish. The flows should be towards the line P_1q_1, with a crossover length which, from Eq. (1), has the form $\xi_X \sim e^{1/2Aw^2}$ and therefore may become very large for weak randomness.

We now turn to our numerical results for $d = 2$. These will be described in detail elsewhere [15]. As shown by Blöte and Nightingale [16] the transfer matrix for the pure q-state Potts model in a strip of width L may be constructed in a basis in which q enters only as a continuous parameter. We have generalized this to the case when the bond strengths u_{ij} are quenched random variables, and the transfer matrices $\mathcal{T}_i$ therefore depend on the row labels i. The size of the transfer matrices grows like 4^L, independent of q, making this a practicable method for larger q.

Starting with some suitable initial vector $\mathbf{v}_0$, the leading Lyapunov exponent is given by [17] $\Lambda_L^0 = \lim_{m\to\infty} \times \frac{1}{m} \ln \|(\prod_{i=1}^m \mathcal{T}_i)\mathbf{v}_0\|$. Higher exponents are found by

iterating a set of vectors $(\mathbf{v}_i)_{j=0}^{k}$, where a given $\mathbf{v}_j$ is orthogonalized to the set $(\mathbf{v}_i)_{i=0}^{j-1}$ after each multiplication by the transfer matrix. The average free energy per site is then $f_L = -\frac{1}{L}\Lambda_L^0$. For a system exhibiting a first-order transition with a bulk correlation length ξ we expect [16] $f_L \sim f_\infty + O(L^{-d}e^{-L/\xi})$, so that $\lambda(L) \equiv \ln[f_L - f_\infty] + d\ln L \sim \text{const} - L/\xi$. In Fig. 2 we show this for various values of q and the randomness strength $R \equiv K_2/K_1$. We see that the randomness changes the transition for $q = 8$ into one with an apparently diverging correlation length. In such a case the amplitude of the finite-size correction has the form [18] $f_L \sim f_\infty - \pi c'/6L^2$, where c' is the *effective* central charge [since f_L is the quenched free energy, in a replica formalism this is the derivative of the central charge $c(n)$ with respect to the number of replicas at $n = 0$]. The value of c' was determined by making parabolic least squares of f_L versus $1/L^2$ [19]. We found the optimum trade-off between statistical errors and a reasonable computation time by taking strips of length $m = 10^5$ and averaging f_L over 100 independent realizations of the randomness for $1 \le L \le 8$, and 3 realizations for $9 \le L \le 12$. Data collection was made for each 200 multiplications by $\mathcal{T}_i$, and the first 2000 iterations of each run were discarded in order to eliminate transients. The parabolic fits were made by including the data points for $L_0 \le L \le 12$, where L_0 must be chosen large enough to justify the finite-size scaling form, and small enough to minimize error bars. From the special cases of the Ising model and percolation it appeared that $L_0 = 3$ is optimal.

For the random-bond Ising model ($q = 2$) with $R = 2$ we found $c' = 0.495 \pm 0.006$, in agreement with the result of de Queiroz [19], $c' = 0.498 \pm 0.003$ using the spin basis, and with the expected value $c' = \frac{1}{2}$. For $q = 3$ and $R = 2$, our result $c' = 0.799 \pm 0.006$ is unable to distinguish between the pure value of $\frac{4}{5}$ and that of $c' \approx 0.8025$ obtained in Ref. [8] by an expansion in $q - 2$. For $q = 4$ the results $c' = 1.003 \pm 0.006$ for $R = 2$ and $c' = 1.010 \pm 0.022$ for $R = 10$ are consistent with each other and the pure value $c = 1$, but for larger q the $R = 2$ results appear to saturate while those for $R = 10$ show a gradual increase: $c' = 1.517 \pm 0.025$ for $q = 8$ and $c' = 3.003 \pm 0.031$ for $q = 64$. Similar values have recently been reported by Picco [20]. However, it should be pointed out that these are also very close to those expected at the percolation point $R \to \infty$. For then the replicated model is the Potts model with q^n states, so that $c' = (\partial/\partial n)c_{\text{Potts}}(q^n)|_{n=0} = (5\sqrt{3}/4\pi)\ln q \approx 0.689 \ln q$; this is confirmed by our transfer matrix calculations [15]. These are also remarkably close to the pure values for $2 \le q \le 4$. We conclude that measurement of the effective central charge does not distinguish well between pure, percolative, and nontrivial random behavior.

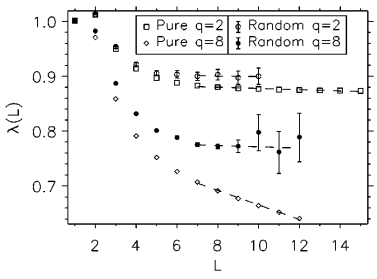


FIG. 2. Plots of $\lambda(L)$, normalized to $\lambda(1) = 1$, showing that bond randomness renders the phase transition second order.

We therefore turned to the magnetic exponent $x_1 = \beta/\nu$. Although this may be determined in principle by adding a ghost site [16], it proved more efficient to exploit duality and relate the spin-spin correlation function $G(m)$ on the strip to the free energy in the presence of a seam of frustrated bonds. Details of this relation and how to implement it in the transfer matrix will be given elsewhere [15]. In pure systems, according to conformal symmetry [21] $G(m)$ decays along the strip as $\exp(-2\pi x_1 m/L)$, so that the difference Δf_L in the free energy per site with and without the seam behaves as $2\pi x_1/L^2$. However, in the random system this difference typically has fluctuations which are $O(m^{-1/2})$. This has the consequence that, while $\Delta f_L \sim \frac{1}{mL}\ln G$ is self-averaging, G is not [22]. In fact, as shown by Ludwig [9], the moments $\overline{G(m)^N}$ (where the overline denotes the quenched average) exhibit *multiscaling;* that is, they scale with dimensions x_N which are not, in general, linear in N. Since conformal symmetry assumes translational invariance, it refers only to such averaged quantities. In practice, we can avoid the lack of self-averaging by performing a cumulant expansion

$$\ln \overline{G^N} = N\overline{\ln G} + \tfrac{1}{2}N^2\overline{(\ln G - \overline{\ln G})^2} + \ldots, \qquad (4)$$

where the higher cumulants may be directly extracted from the statistical fluctuations in Δf_L. For values of q and R which are not too large, this expansion appears to converge well, keeping the first 3 or 4 cumulants. The fact that the higher cumulants are nonzero implies multiscaling. Our values for x_1 are shown in Fig. 3. For $2 < q \le 3$ and $R = 2$ they are in perfect agreement with the predictions of the $(q - 2)$ expansion of Refs. [9,10], and for $q = 3$ and 4 they agree with recent Monte Carlo results of Picco [23]. For larger q the results appear to increase smoothly with q, with $x_1(q = 8) = 0.1415 \pm 0.0036$. Thereafter the cumulant expansion begins to break down.

Although the thermal exponent ν should be related in a similar manner to the first gap $\Lambda_L^1 - \Lambda_L^0$ in the Lyapunov spectrum, for reasons we do not understand this yields results which, if taken literally, appear to violate the bound $\nu \ge 1$ [24]. Instead we have measured ν directly by finite-size scaling of the magnetic correlation length away from T_c, using phenomenological RG methods [25]. For $q = 8$, as shown in Fig. 4, we find clear evidence for fixed points at $R = 1$, with $\nu = \frac{1}{2}$ ($x_T = 0$), and a random fixed point with $\nu_R = 1.01 \pm 0.02$. At $q = 3$ our results are consistent with the perturbative value $\nu_R \approx 1.02$, and for $q = 64$ we find $\nu_R = 1.02 \pm 0.03$.

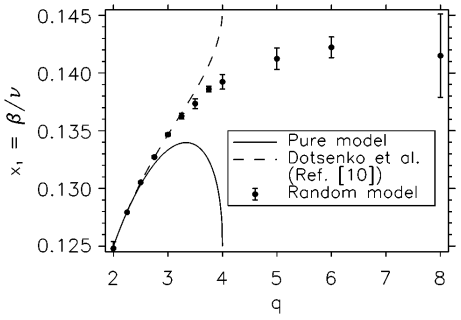


FIG. 3. Magnetic exponent as a function of q, with $R = 2$.

In summary, we have computed the exponents of the 2D random-bond Potts model and shown that while the thermal exponent ν_R is consistent, within error bars, with both the pure Ising value, and with the results of the $(q - 2)$ expansion, the magnetic exponent varies continuously with q in a manner which agrees with this expansion in the region it is expected to be valid, and with a value at $q = 8$ which is quite different from the Ising value of $\frac{1}{8}$. This is in sharp disagreement with the Monte Carlo results of Ref. [6]. One possible reason is that these authors use a nonstandard definition of the order parameter which, while it scales in the same way as the usual one in the pure case, may not when multiscaling is present. Another is that our very long strips are able to accommodate large regions in which all q values of the order parameter are realized, and this may not typically be the case in the square geometries of Ref. [6]. Our results should also be compared with those of Ref. [26], in which it is shown that the random *connectivity* of a Voronoi–Delauney lattice does not modify the first-order nature of the transition. However, in this case it may be shown that the typical random fluctuations in the energy of a region of size L^d are $O(L^{(d-1)/2})$ rather than $O(L^{d/2})$ as in our random-bond model, so they are always smaller than the domain wall energy. Another study [27], in which the transition was observed to be softened, refers to a lattice with random *curvature* which is fractal when embedded in the plane,

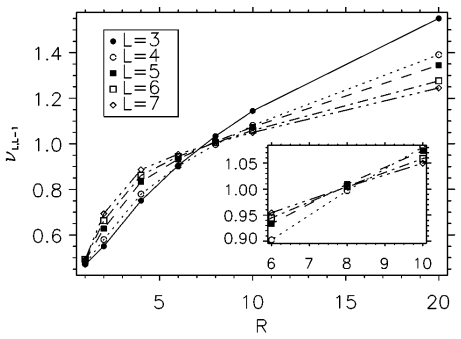


FIG. 4. Values of ν extracted from phenomenological RG for different strip widths and $q = 8$. There is a fixed point at $R \approx 8$ with $\nu \approx 1$ (see inset). Error bars are less than the symbol size.

and to which neither our arguments nor those of Refs. [3,4] directly apply.

We have also given a mapping of the random-bond problem to the RFIM which suggests that for $d > 2$ such systems should exhibit a tricritical point whose thermal exponent is related to those of the RFIM. This picture is quite generic, and, since there are many real 3D systems which undergo first-order transitions, it would be interesting to reexamine the effect of random impurities in such cases.

The authors acknowledge useful discussions with J. Chalker, A. Ludwig, and S. de Queiroz. This research was supported in part by the Engineering and Physical Sciences Research Council under Grant No. GR/J78327, and by the National Science Foundation under Grant No. PHY94-07194.

[1] A. B. Harris, J. Phys. C **7**, 1671 (1974).
[2] Y. Imry and M. Wortis, Phys. Rev. B **19**, 3580 (1979).
[3] M. Aizenman and J. Wehr, Phys. Rev. Lett. **62**, 2503 (1989).
[4] K. Hui and A. N. Berker, Phys. Rev. Lett. **62**, 2507 (1989); **63**, 2433(E) (1989).
[5] Y. Imry and S.-K. Ma, Phys. Rev. Lett. **35**, 1399 (1975).
[6] S. Chen, A. M. Ferrenberg, and D. P. Landau, Phys. Rev. E **52**, 1377 (1995).
[7] E. Domany and S. Wiseman, Phys. Rev. E **51**, 3074 (1995).
[8] A. W. W. Ludwig and J. L. Cardy, Nucl. Phys. **B285 [FS19]**, 687 (1987).
[9] A. W. W. Ludwig, Nucl. Phys. **B330**, 639 (1990).
[10] V. Dotsenko, M. Picco, and P. Pujol, Nucl. Phys. **B455**, 701 (1995).
[11] W. Kinzel and E. Domany, Phys. Rev. B **23**, 3421 (1981).
[12] P. W. Kasteleyn and C. M. Fortuin, J. Phys. Soc. Jpn. (Suppl.) **46**, 11 (1969).
[13] A. J. Bray and M. A. Moore, J. Phys. C **18**, L927 (1985).
[14] D. Stauffer, Z. Phys. B **22**, 161 (1975).
[15] J. L. Jacobsen and J. Cardy (to be published).
[16] H. W. J. Blöte and M. P. Nightingale, Physica (Amsterdam) **112A**, 405 (1982).
[17] G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn, Meccanica **15**, 9 (1980).
[18] H. W. Blöte, J. L. Cardy, and M. P. Nightingale, Phys. Rev. Lett. **56**, 742 (1986); I. Affleck, Phys. Rev. Lett. **56**, 746 (1986).
[19] S. L. A. de Queiroz, Phys. Rev. E **51**, 1030 (1995).
[20] M. Picco, Phys. Rev. Lett. **79**, 2998 (1997).
[21] J. L. Cardy, J. Phys. A **16**, L355 (1983).
[22] B. Derrida, Phys. Rep. **103**, 29 (1984).
[23] M. Picco, Phys. Rev. B **54**, 14 930 (1996); (private communication).
[24] J. T. Chayes, L. Chayes, D. S. Fisher, and T. Spencer, Phys. Rev. Lett. **57**, 2999 (1986).
[25] M. P. Nightingale, Physica (Amsterdam) **83A**, 561 (1976).
[26] W. Janke and R. Villanova, Phys. Lett. A **209**, 179 (1995); Nucl. Phys. B (Proc. Suppl.) **47**, 641 (1996).
[27] C. F. Baillie, W. Janke, and D. A. Johnston, Phys. Lett. B **388**, 14 (1996).

7.4 Article "Transfer matrix study of random-bond Potts models"

Nous étudions le modèle de Potts bidimensionnel sur le réseau carré en présence d'impuretés gelées (couplages aléatoires). Pour $q > 4$, la transition de premier ordre du modèle pur est adoucie à cause des impuretés. Nous déterminons les classes d'universalité résultantes en comparant des données de matrice de transfert avec des prédictions de l'invariance conforme. L'exposant magnétique β/ν varie continûment en fonction de q et prend des valeurs différentes de celles du modèle d'Ising pour $q > 4$. Par contre, l'exposant ν décrivant la divergence de la longueur de corrélation est numériquement consistant avec 1. Nous présentons des preuves de l'exactitude d'un diagramme de phase proposé précédemment : ce diagramme de phase unifie les comportements purs, percolatifs et aléatoires.

ELSEVIER

Nuclear Physics B 515 [FS] (1998) 701–742

NUCLEAR PHYSICS B

Critical behaviour of random-bond Potts models: a transfer matrix study

Jesper Lykke Jacobsen [a,b,c], John Cardy [a,d]

[a] *University of Oxford, Department of Physics – Theoretical Physics, 1 Keble Road, Oxford OX1 3NP, UK*
[b] *Institute of Physics and Astronomy, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark*
[c] *Somerville College, Oxford, UK*
[d] *All Souls College, Oxford, UK*

Received 28 November 1997; accepted 6 January 1998

Abstract

We study the two-dimensional Potts model on the square lattice in the presence of quenched random-bond impurities. For $q > 4$ the first-order transitions of the pure model are softened due to the impurities, and we determine the resulting universality classes by combining transfer matrix data with conformal invariance. The magnetic exponent β/ν varies continuously with q, assuming non-Ising values for $q > 4$, whereas the correlation length exponent ν is numerically consistent with unity. We present evidence for the correctness of a formerly proposed phase diagram, unifying pure, percolative and non-trivial random behaviour. © 1998 Elsevier Science B.V.

PACS: 05.70.Jk; 64.60.Ak; 64.60.Fr
Keywords: Potts models; Quenched-bond randomness; First-order phase transitions; Transfer matrices; Conformal field theory; Lyapunov spectra

1. Introduction

The effect of quenched bond randomness on a classical statistical mechanics system whose pure version undergoes a second-order phase transition is well understood. Namely, the so-called Harris criterion states that if the critical exponent α^{pure} governing the divergence of the specific heat at the transition point of the *pure* system is negative, weak bond randomness is irrelevant in the renormalisation group (RG) sense and the pure fixed point (FP) is stable [1]. On the other hand, if $\alpha^{\text{pure}} > 0$ the randomness is relevant and causes a cross-over to critical behaviour governed by a new random FP nearby, at least if the cross-over exponent α^{pure} is small.

PII S0550-3213(98)00024-8

This should be contrasted with the more dramatic effects of randomness in the field conjugate to the local magnetisation. Such randomness can eliminate low-dimensional phase transitions altogether, and at least it always changes the values of the critical exponents [2]. For this reason most early research was concentrated on field randomness. In this context a particularly popular model is the random-field Ising model (RFIM) for which a classical argument due to Imry and Ma [3] from a simple comparison of the field fluctuations with the stabilising effect caused by the formation of a domain wall concluded, that the lower critical dimension of the RFIM is $d_l = 2$.

The issue of quenched bond randomness imposed on a system that undergoes a thermal *first*-order phase transition is less studied. An adaptation of the Imry–Ma argument can be established by noting that the bond randomness couples to the local energy density, which differs for the two phases that co-exist at the critical point of the pure system, in exactly the same way that the random field couples to the local magnetisation in the RFIM. Consequently, the existence of a non-vanishing latent heat for $d < 2$ can be ruled out. Early work by Imry and Wortis [4] furnished a heuristic argument, reminiscent of that of the Harris criterion, that the bond randomness indeed softens any such phase transition in $d = 2$ to a continuous one. A subsequent phenomenological RG argument by Hui and Berker [5] confirmed that the lower critical dimension for random-bond tricriticality and end-point criticality is $d_l = 2$. As the dimensionality increases, tricritical points and critical end points emerge from $T = 0$. Finally, a mathematically rigorous theorem by Aizenman and Wehr [6] stated that quite generally for $d \leqslant 2$ an *arbitrarily* weak amount of quenched bond randomness leads to the elimination of any discontinuity in the density of the variable conjugate to the fluctuating parameter.

The question then emerges whether this softening of the phase transition can be verified for specific models and, if so, what are the universality classes of these novel second-order phase transitions. An investigation along these lines has recently been initiated by one of us [7], by considering a system of N two-dimensional Ising models coupled by their energy operators which, according to mean-field theory (MFT), is supposed to display a second-order phase transition. For $N > 2$, however, the RG flow of the model exhibits a runaway behaviour, which is characteristic of a fluctuation driven first-order transition [8]. In this sense the transition is only weakly first order and hence amenable to perturbative calculations. On adding weak bond randomness it was found that the RG trajectories curl back towards the pure decoupled Ising FP, and consequently Ising exponents are expected, up to possible logarithmic corrections. This study was extended by Pujol [9] to the case of N coupled q-state random-bond Potts models for $2 \leqslant q \leqslant 4$, but here the universality class of the impurity softened transition was found to depend on the coupling between the models.

A more interesting model for studying the effect of quenched bond impurities on a first-order transition is the q-state random-bond Potts model (RBPM). For $q > 4$ the phase transition of the pure system is first order with a latent heat that is an increasing function of q [10]. In fact, since the transition is first order already in MFT, on the RG level it is controlled by a zero-temperature discontinuity FP with the eigenvalue of the relevant scaling operator being $y = d$ [8]. Quenched randomness coupling to the

local energy density thus has the eigenvalue $d - 2(d - y) = d$ and is strongly relevant, whence an RG treatment appears to be problematic.

The work undertaken until now has therefore mainly been numerical. Extensive Monte Carlo (MC) simulations have been carried out for $q = 8$ by Chen, Ferrenberg and Landau [11] confirming the transition softening scenario outlined above, and finding critical exponents numerically consistent with those of the *pure* Ising model. Similar conclusions were reached by Domany and Wiseman [12] for $q = 4$ and also for the Ashkin–Teller model. It thus appears that in a variety of situations the universality class of the bond disordered models is that of the Ising model, *irrespective* of the symmetry underlying the original model.

To explain these findings Kardar et al. [13] have proposed an interface model for the RBPM which, after several approximations, is amenable to an RG treatment that is exact on the hierarchical lattice. In the pure model the interface exhibits a branching structure with fractal dimension at criticality, but when randomness is present the critical interface is asymptotically linear. Assuming that the vanishing of the interfacial free energy is governed by a zero-temperature FP, the Widom exponent μ turns out to be *independent* of q for all sufficiently large q, taking the Ising value $\mu = 1$.

This is in contrast to the perturbative expansion in powers of $(q - 2)$ investigated by Ludwig and Cardy [14], Ludwig [15,16], and Dotsenko et al. [17]. Using the RG approach for the perturbation series around the conformal field theories representing the pure models, these authors find the critical behaviour of the RBPM to be controlled by a new random FP which merges with the pure FP as $q \to 2$. Critical exponents are found to depend continuously on q, at least for $(q - 2)$ small, and in the case of the magnetic exponent x_H a calculation to three-loop order yields a prediction which is supposed to be very precise even up to $q = 3$ [17]. Unfortunately, extending these results beyond $q = 4$ is impossible, even in principle, since this is the limiting case in the range of minimal conformal theories around which the perturbative calculations take place. Another interesting implication of this line of research is that the local operators exhibit *multiscaling* [16], meaning that correlation functions of different moments of such operators decay with powers that are, in general, independent.

It has been suggested by Kardar et al. [13] and one of us [7] that these contrasting theories describe very different FPs. Indeed, it can be argued that the interface model pertains to the case of strong non-self-dual randomness, whilst the $(q - 2)$-expansion is relevant for weak self-dual randomness. Also, even though it may turn out that the critical exponents do not depend on q, the central charge c evidently must, since even when the critical behaviour is controlled by a decoupled Ising FP there is generally not just *one* Ising model but several.

To resolve this controversy we have undertaken an extensive study of the $d = 2$ RBPM where finite-size data obtained from transfer matrix (TM) calculations were combined with the powerful techniques of conformal invariance. We have extended the random-cluster model TMs of Blöte and Nightingale [18] to the case of bond randomness whilst taking into account that in the impure case such TMs do not commute and hence must be discussed in terms of their Lyapunov (rather than the eigenvalue) spectra.

Because of the lack of self-averaging the relation between the Lyapunov spectra and the critical exponents is inferred indirectly through a cumulant expansion, which has the advantage of illustrating the multiscaling properties of the correlation functions [16] explicitly. The number of Potts states q enters our TMs only as a continuous parameter, both facilitating the comparison with analytical results within the $(q-2)$-expansion and making the interesting regime $q > 4$ readily accessible.

Although the cumulant expansion yields very appealing results in the case of the magnetic exponent it works poorly for the thermal one. For reasons yet not fully understood such results, when taken at face value, seem to hint at a conformal field theory violating the bound $\nu \geqslant 2/d$ [19,20]. On the other hand, using phenomenological RG techniques [21] we find results consistent both with the bound and with the $(q-2)$-expansion [15,22].

Some of our results have been reported in a Letter [23], where we also described a mapping between the interfacial models of the RBPM for large q and the RFIM. This mapping, which is asymptotically exact in the limit $q \to \infty$, allowed us to establish a schematic phase diagram for the RBPM unifying pure, non-trivial random, and percolative behaviour. In the present paper the evidence for this phase diagram will be collected and discussed.

The outline of this paper is as follows. In Section 2 we define the model and discuss the principles of extracting physical information from the Lyapunov spectrum of the TMs. The proposed phase diagram is reviewed along with the translation of the renormalisation group equations from the RFIM to the problem at hand. Then, in Section 3, the TM formalism of Ref. [18] is generalised to the random case. This relies on the mapping of the RBPM to the random-cluster model and on two complementary representations of the connectivity of a row of spins in the latter model. By decomposing the TM into sparse single-bond TMs we arrive at a highly efficient algorithm, the implementation of which is considered in detail. The magnetic properties can be accessed by adding a ghost site, but we also describe an alternative route in which the two-point correlator is related to a disorder operator under duality. The corresponding implementation of the TMs has a seam spanning the length of the cylinder. In the percolation limit the TMs take on a particularly simple form, allowing us to obtain very accurate results.

Section 4 is dedicated to the presentation of our numerical results. From the scaling of the free energy we find evidence that the first-order phase transition is indeed softened due to the randomness. The effective central charge is determined both at the random FP and in the percolation limit. In the latter case we obtain excellent agreement with our analytical result. It is shown how a cumulant expansion leads to very accurate values of x_H. These depend continuously on q, and are in perfect agreement with the $(q-2)$-expansion at $q = 3$. For larger q the values stay far away from the Ising value, in sharp contrast to the results of Ref. [11]. For $q > 8$ the expansion begins to break down, and we give an argument why this must be so in terms of a model of coupled replicas. The problems encountered when trying to extract x_T in a similar fashion then lead us to discuss the method of extracting physical observables from the Lyapunov spectrum in more physical terms. We then consider the constraints put on the multiscaling exponents

by a conformal sum rule. A reliable determination of x_T is furnished by a variant of the phenomenological RG scheme, in which the shape of the self-dual surface is explicitly taken into account. The criticism of Ref. [19] recently raised by Pázmándi et al. [24] is shown not to apply to the RBPM. We conclude the section with a discussion of the higher Lyapunov spectrum and its possible relation to the (presently unknown) conformal field theory underlying the model.

Finally, Section 5 contains a discussion of our findings. We seek to explain the discrepancy with Ref. [11], and we discuss other types of randomness relevant to the question whether a first-order phase transition is softened due to impurities. A list of unsettled questions relevant for future research is also given.

2. The model and its phase diagram

2.1. The random-bond Potts model

The q-state Potts model [25] is defined by the reduced Hamiltonian

$$\mathcal{H} = -\sum_{\langle ij \rangle} K_{ij} \delta_{\sigma_i \sigma_j}, \tag{1}$$

where the spins, defined on the vertices of the square lattice, can take the values $\sigma_i = 1, 2, \ldots, q$, and the summation is over all nearest-neighbour bonds in the lattice. We shall specialise to the ferromagnetic case, where the reduced couplings $K_{ij} \geqslant 0$ measure the strength of the aligning tendency of nearest-neighbour spins.

Although the free energy of the pure model ($K_{ij} \equiv K$) is not known in closed form for general q, a wide range of exact results is nevertheless available [26]. In particular it is well known that the model exhibits a second-order phase transition for $q \leqslant 4$ and a first-order one for $q > 4$ [10].

However, in this paper we are mainly concerned with the random-bond Potts model (RBPM) for which much less is known. Here the couplings K_{ij} are quenched random variables, typically drawn from the symmetric binary distribution

$$P(K) = \tfrac{1}{2}[\delta(K - K_1) + \delta(K - K_2)], \tag{2}$$

where the ratio between strong and weak bonds $R = K_2/K_1$ measures the strength of the randomness. For the special choice

$$\left(\mathrm{e}^{K_1} - 1\right)\left(\mathrm{e}^{K_2} - 1\right) = q \tag{3}$$

the model is on average self-dual, as discussed in more detail in Section 3.4 below. Assuming that the phase transition is unique the model is therefore at its critical point [27].

Other self-dual distributions of the random bonds than that of Eq. (2) have also been investigated in order to check our results. In particular, we have found the trinary distribution introduced in Section 4.2 useful, since it gives us a clearer idea about the length scale associated with the random impurities.

2.2. Lyapunov spectrum of the transfer matrix

The construction of the transfer matrices (TMs) for the RBPM is described in detail in Section 3. It is well known that in the pure case ($R = 1$) the operator content of the conformal field theory (CFT) underlying the model is related to the eigenvalue spectrum $\{\lambda_i(L)\}$, $i = 0, 1, 2, \ldots$, of the TM for a strip of width L through [28]

$$f_i(L) - f_0(L) = \frac{2\pi x_i}{L^2} + \ldots, \tag{4}$$

where $f_i(L) = -\frac{1}{L} \ln \lambda_i(L)$ are the generalised free energies per site (in units of $k_B T$) and x_i the scaling dimensions of the corresponding operators. Similarly the central charge c, measuring the number of bosonic degrees of freedom of the CFT, is related to the finite-size corrections to the customary free energy through [29]

$$f_0(L) = f_0(\infty) - \frac{\pi c}{6L^2} + \ldots . \tag{5}$$

In the random case the TMs are no longer constant but depend on the particular realisation of the random bonds within each row of strip. Accordingly the concept of eigenvalues generalises to that of Lyapunov exponents. Starting with some suitable initial vector of unit norm $|v_0\rangle$, the leading Lyapunov exponent can be found by the Furstenberg method [30]

$$\Lambda_0(L) = \lim_{m\to\infty} \frac{1}{m} \ln \left\| \left(\prod_{j=1}^{m} \mathcal{T}_j \right) |v_0\rangle \right\|, \tag{6}$$

where $\mathcal{T}_j$ is the TM acting between rows $j-1$ and j. The average free energy per site is given as before by $f_0(L) = -\frac{1}{L}\Lambda_0(L)$. Higher exponents are found by iterating a set of n vectors $\{|v_k\rangle\}_{k=0}^{n-1}$, where a given $|v_k\rangle$ is orthogonalised to the set $\{|v_l\rangle\}_{l=0}^{k-1}$ after each multiplication by $\mathcal{T}_j$ [31]. Surprisingly, this method works even for a non-hermitian TM, and it is numerically shown to be independent of the choice of the initial vectors.

When some symmetry (e.g., spin reversal or duality) is manifest in $\mathcal{T}_j$ the orthogonalisation can be circumvented by iterating vectors which belong to definite irreducible components of that symmetry, but the Z_q permutational symmetry inherent in the Potts model has been lost through the mapping to the random-cluster model which forms the backbone of our TMs; see Section 3.1.

As to the extraction of physical information from the spectra, Eq. (5) is supposed to retain its validity provided that c is replaced by the effective central charge c', that in the standard replica formalism is the derivative of $c(n)$ with respect to the number of replicas at $n = 0$ [14]. The question to which extent Eq. (4) also remains valid is by no means trivial and we shall dedicate a fair part of the subsequent discussion to it.

2.3. Phase diagram

In the limit $q \to \infty$ the behaviour of the pure model is readily understood [23]. At the self-dual point the partition function is dominated by two contributions, namely those corresponding to the q completely ordered states and the completely disordered state respectively. All other configurations are down by powers of $1/\sqrt{q}$ and have recently been enumerated to 10th order in this small parameter [32]. The dominating states have identical free energy but different internal energy densities of K and 0 for the ordered and the disordered phase respectively, so the transition is, as expected, first order.

The bond randomness is then included through the parametrisation $e^{K_{ij}} - 1 = q^{\frac{1}{2}+w_{ij}}$, where $w_{ij} = \pm w$ and $w > 0$ measures the strength of the randomness. It can now be shown [23] that as $q \to \infty$ the model for an interface between these two phases of the RBPM is exactly the same as that of an interface between the spin-up and spin-down phases of the RFIM

$$\mathcal{H}^{\mathrm{RFIM}} = -J \sum_{\langle ij \rangle} s_i s_j - \sum_i h_i^{\mathrm{RF}} s_i - h \sum_i s_i \tag{7}$$

with $h_i^{\mathrm{RF}} = \pm h^{\mathrm{RF}}$, provided that one translates quantities between the two models using the "dictionary"

$$h^{\mathrm{RF}} \leftrightarrow \tfrac{1}{2} w \ln q, \qquad J \leftrightarrow \tfrac{1}{8} \ln q, \qquad h \leftrightarrow \tfrac{1}{4} t \ln q. \tag{8}$$

Here $t \equiv (T - T_c)/T_c$ is the reduced temperature.

The infinitesimal RG equations can now be inferred from the similar results for the RFIM [33]. Near $d = 2$ they read

$$dw/dl = -(d/2 - 1)w + Aw^3 + \ldots, \tag{9}$$

$$d(\ln q)^{-1}/dl = -(\ln q)^{-1}\big((d - 1) - Aw^2 + \ldots\big), \tag{10}$$

$$dt/dl = t(1 + Aw^2 + \ldots), \tag{11}$$

where $A > 0$ is a non-universal constant. The RG flows for $d > 2$ and the proposed phase diagram are shown in Fig. 1 (see Ref. [23] for details). The shaded region of non-vanishing latent heat is bounded by the line Rq_2 of tricritical points. This line, controlled by the fixed point R at infinite q, merges with the abscissa as $d \to 2$. At $q = \infty$ the interfacial mapping is exact so that the flows along RP_1 must extend all the way to $w = \infty$, and since w^{-1} is known to be a relevant scaling variable at the percolation limit [34], Ref. [23] concluded that Rq_2 must be separated from the percolative behaviour along P_1P_2 by another line of stable FPs emerging from P_1. It was then conjectured that this connects on to the line of random FPs found in the $(q - q_1)$-expansion [14]. In this paper we shall present the evidence for this conjecture for $d = 2$, when $q_1 = 2$ and $q_2 = 4$.

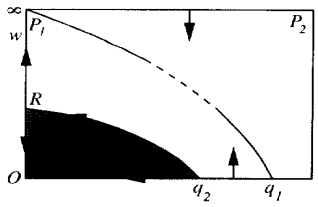


Fig. 1. Schematic phase diagram in the critical surface for $d > 2$. q increases to the left and w is the disorder strength, with P_1P_2 being the percolation limit. RG flows are indicated. The latent heat is non-vanishing within the shaded region, and elsewhere the transition is continuous, controlled by the line of fixed points P_1q_1. As $d \to 2$ the shaded region collapses to a line q_2O of first-order transitions in the pure system.

3. The transfer matrices

In spite of the large amount of high-precision results obtained by combining transfer matrix (TM) techniques with finite-size scaling for almost any conceivable type of pure statistical mechanics system (see, e.g., Ref. [35] for a review) the use of TMs in the study of disordered systems seems to have attracted rather little interest as compared with the complementary approach of Monte Carlo simulations.

A straightforward way of setting up the TMs for the q-state Potts model is to use the traditional spin basis where the state of a row of L spins is labelled by the q^L basis states $\{\sigma_1, \sigma_2, \ldots, \sigma_L\}$, $\sigma_i = 1, \ldots, q$. Whilst this approach is highly efficient for $q = 2, 3$ it has two major shortcomings in the general case. First, the dimension of the matrices becomes forbiddingly for large q, in particular making unaccessible the regime of $q > 4$ which is our main concern. Second, the restriction to integer values of q is unnecessary and in fact makes it difficult to compare numerical results with analytical calculations in the $(q-2)$-expansion [15–17].

Both these shortcomings can be remedied by writing the TMs in the connectivity basis introduced by Blöte and Nightingale [18]. In this representation the dimension of the TMs is independent of q which enters only as a *continuous* parameter. In fact, the number of basis states is asymptotically $\sim 4^L$ (or $\sim 5^L$ upon imposition of a magnetic field) with a rather small coefficient of proportionality, in practice making this basis the preferred choice for all but the Ising model ($q = 2$).

We have generalised these TMs to include quenched bond randomness, and also devised an alternative method of accessing the magnetic properties through the introduction of a seam along the strip. Furthermore, in the percolation limit the TMs are found to simplify in a manner that makes calculations for rather large strip widths feasible. For convenience these results are presented along with a review of the relevant parts of Ref. [18] thus making our description of the Potts model TMs self-contained.

3.1. Mapping to the random-cluster model

Introducing an imaginary 'ghost site' with fixed spin $\sigma_0 = 1$ the partition function for the Potts model can be written as

$$Z = \sum_{\{\sigma\}} \left(\prod_{\langle ij \rangle} \exp(K_{ij} \delta_{\sigma_i \sigma_j}) \right) \left(\prod_{\langle i0 \rangle} \exp(H_i \delta_{\sigma_i \sigma_0}) \right), \tag{12}$$

where $\prod_{\langle ij \rangle}$ is the usual product over pairs of nearest-neighbour sites and each site i has been connected to the ghost site 0 with a similar notation. The reduced magnetic field H_i, here taken to be site dependent, now enters at the same footing as the reduced exchange couplings K_{ij}. It should be pointed out, however, that a random coupling to the ghost site is not a true random field, since the latter would try to force different sites into *different* Potts states and not just into the particular state of the ghost site with a site-dependent probability. To avoid any confusion we shall therefore specialise to the case of a homogeneous field $H_i \equiv H$.

The site variables can now be traded for bond variables through the mapping to the random-cluster model introduced by Kasteleyn and Fortuin [36]. In terms of the variables $u_{ij} = e^{K_{ij}} - 1$ and $v = e^H - 1$ we arrive at

$$Z = q^N \sum_{G \subseteq \mathcal{L}} \sum_{G_0 \subseteq \mathcal{L}_0} \left(\prod_{\langle ij \rangle \in G} \frac{u_{ij}}{q} \right) \left(\prod_{\langle i0 \rangle \in G_0} \frac{v}{q} \right) q^{l(G \cup G_0)}, \tag{13}$$

where $\mathcal{L}$ denotes the set of all nearest-neighbour bonds, $\mathcal{L}_0$ the bonds from each of the N Potts spins to the ghost site, and $l(G \cup G_0)$ is the number of independent loops on the combined graph $G \cup G_0$.

The usual construction of the transfer matrix $\mathcal{T}$ for a strip of width L seems to be obstructed by the non-local factor $l(G \cup G_0)$, but this can be taken into account by choosing a basis containing information about which sites of a given row are interconnected through the part of the lattice below that row (including connections via the ghost site). This leads us to the concept of connectivity states, which we consider next.

3.2. The connectivity states

In order to determine the number of loop closures induced by appending a new row of L sites along with the corresponding L connections to the ghost site to the top of $G \cup G_0$, we need information about how the sites in the top row of $G \cup G_0$ were previously interconnected. This information is comprised in the *connectivity state* $(i_1 i_2 \ldots i_L)$, where $i_t = 0$ if site t is connected to 0 within the combined graph $G \cup G_0$ and, otherwise, $i_r = i_s$ is a (non-unique) positive integer if and only if sites r and s are connected within G.

Whilst this 'index representation' is useful for determining whether a newly appended bond does or does not close a loop, and thus will allow us to explicitly construct the single-bond TMs in the next subsection, a one-to-one mapping to the set of consecutive integers $\{1, 2, \ldots\}$ is clearly needed to define a 'number representation' which will enable us to label the entries of the TM and thus to perform actual computations. These representations and the mapping were supplied by Ref. [18] as were the determination of the number of connectivity states (d_L with and c_L without a magnetic field). We

shall review the necessary details and also give details on the construction of the inverse of the mapping just mentioned.

Consider first the case of $H = 0$ where all ghost bonds carry zero weight ($v = 0$). The connectivity states then have all $i_t > 0$ and can be recursively ordered by noting that the index representation is *well-nested*, i.e. for $r < s < t < u$,

$$(i_r = i_t) \wedge (i_s = i_u) \Rightarrow i_s = i_t. \tag{14}$$

It follows that if we define the *cut function* $\rho(i_1 i_2 \ldots i_L)$ to be the smallest $t > 1$ such that $i_1 = i_t$, if such a t exists, and $L + 1$ otherwise, the left $(i_2 i_3 \ldots i_{\rho-1})$ and right $(i_\rho i_{\rho+1} \ldots i_L)$ parts of the index representation are both well-nested. A complete ordering of the well-nested sequences is now induced by applying the cut function first to the whole sequence, then recursively to its right and finally to its left part.

More precisely, the mapping from the index to the number representation is effected by

$$\sigma(i_1 i_2 \ldots i_L) = \begin{cases} 1 & \text{if } L \leqslant 1, \\ c_{L,k-1} + [\sigma(i_k \ldots i_L) - 1] c_{k-2} & \\ \quad + \sigma(i_2 \ldots i_{k-1}) & \text{otherwise,} \end{cases} \tag{15}$$

where $k = \rho(i_1 \ldots i_L)$ and $c_{n,l} = \sum_{i=2}^{l} c_{i-2} c_{n-i+1}$ with

$$c_n \equiv c_{n,n+1} = \frac{(2n)!}{n!(n+1)!} \tag{16}$$

giving the number of well-nested n-point connectivities [18,37]. Explicit values are shown in Table 1.

To consider the general case of $v \neq 0$ we remark that the subsequence of non-zero indices is still well-nested. A complete ordering of an index representation $(i_1 i_2 \ldots i_L)$ with precisely s zero indices is then induced by first ordering according to the value of s, then lexicographically ordering the zeros, and finally using the ordering of the well-nested subsequence $(i_{p_1} i_{p_2} \ldots i_{p_{L-s}})$ given by Eq. (15). The lexicographic ordering is carried out by

$$\psi(i_1 i_2 \ldots i_L) = \begin{cases} 1 & \text{if } L = 1 \text{ or } s = L, \\ \psi(i_2 i_3 \ldots i_L) & \text{if } i_1 \neq 0, \\ \binom{L-1}{s} + \psi(i_2 i_3 \ldots i_L) & \text{if } i_1 = 0, \end{cases} \tag{17}$$

and the mapping to the number representation is finally

$$\tau(i_1 i_2 \ldots i_L) = d_{L,s-1} + [\psi(i_1 i_2 \ldots i_L) - 1] c_{L-s} + \sigma(i_{p_1} i_{p_2} \ldots i_{p_{L-s}}), \tag{18}$$

where $d_{n,l} = \sum_{i=0}^{l} \binom{n}{i} c_{n-i}$ with

$$d_n \equiv d_{n,n} = \sum_{i=0}^{n} \binom{n}{i} c_{n-i} \tag{19}$$

Table 1
The number of connectivity states for a Potts model transfer matrix of width L with (d_L) and without (c_L) an external magnetic field. Also shown is the size of the magnetic sector when using a ghost site ($d_L - c_L$) and a seam (Lc_L). For large strip widths the seam is advantageous. The number c_L of well-nested L-point connectivities should be compared to the total number of L-point connectivities b_L which increases *faster* than exponentially as a function of L.

L	c_L	d_L	$d_L - c_L$	Lc_L	b_L
1	1	2	1	1	1
2	2	5	3	4	2
3	5	15	10	15	5
4	14	51	37	56	15
5	42	188	146	210	52
6	132	731	599	792	203
7	429	2950	2521	3003	877
8	1430	12 235	10 805	11 440	4140
9	4862	51 822	46 960	43 758	21 147
10	16 796	223 191	206 395	167 960	115 975
11	58 786	974 427	915 641	646 646	678 570
12	208 012	4 302 645	4 094 633	2 496 144	4 213 597

giving the number of general n-point connectivities. Again, explicit values are presented in Table 1.

To construct the inverse mapping, i.e. the one taking us from the number to the index representation, we solve $\tau = \tau(i_1 i_2 \dots i_L)$ for the indices $(i_1 i_2 \dots i_L)$ by performing the following steps. First, the number of zero indices is found as $s = \max\{s | d_{L,s-1} < \tau\}$. Second, perform a slightly modified integer division by writing $\tau - d_{L,s-1} = Qc_{L-s} + R$, where the remainder R is restricted to take its values in the interval $[1, c_{L-s}]$. From Eq. (18) we infer that $\psi = Q + 1$ and $\sigma = R$. Third, the position of the first (left-most) zero index is given by $i_{L-k} = 0$, where $k = \max\{k | \binom{k}{s} < \psi\}$. This procedure of finding the zero indices is then iterated with $\psi \to \psi^{(1)} \equiv \psi - \binom{k}{s}$ until $\psi^{(s')} = 1$, and the remaining $s - s'$ zeros are filled in from the right: $i_{s'+1} = \dots = i_{s-1} = i_s = 0$.

It then remains to deduce the subsequence of non-zero indices by inverting $\sigma = \sigma(i_{p_1} i_{p_2} \dots i_{p_l})$ with $l = L - s$. After initialising $i_{p_1} = p_1$ we proceed by recursion as follows. First, choose $k = \min\{k | c_{l,k-1} + c_{k-2} c_{l-k+1} \geqslant \sigma\}$. If $k \leqslant l$ we have then found a connection: $i_{p_1} = i_{p_k}$. This procedure of finding the connections is now iterated on the left $(i_{p_2}, \dots, i_{p_{k-1}})$ and the right $(i_{p_k}, \dots, i_{p_l})$ parts of the remaining sequence. If $k \geqslant 2$ the assignment $i_{p_2} = p_2$ is performed. By (modified) integer division we then write $\sigma - c_{l,k-1} = Qc_{k-2} + R$ with $R \in [1, c_{l-k+1}]$, and pass over the left part of the sequence with $\sigma \to \sigma^{(1)}_{\text{left}} \equiv R$ and $l \to l^{(1)}_{\text{left}} \equiv k - 2$, and the right part with $\sigma \to \sigma^{(1)}_{\text{right}} \equiv Q + 1$ and $l \to l^{(1)}_{\text{right}} \equiv l - k + 1$. The recursion stops when for any sequence $l^{(m)} \leqslant 2$. If then $l^{(m)} = 2$ and the sequence is $(i_{p_a}, i_{p_{a+1}})$ we perform the assignment $i_{p_{a+1}} = i_{p_a}$ if $\sigma^{(m)} = 1$ and $i_{p_{a+1}} = p_{a+1}$ if $\sigma^{(m)} = 2$.

Any way of constructing the index representation $(i_1 i_2 \dots i_L)$ will of course reflect the above-mentioned arbitrariness as to the actual values of the non-zero indices, but the particular procedure just outlined is easily seen to ensure that all indices are $\leqslant L$.

This invariant is useful since then any given site t can be *disconnected* from the rest by assigning $i_t = L + 1$.

3.3. The single-bond transfer matrices

The amount of computer time necessary for building up a long strip by repeated application of the transfer matrix $\mathcal{T}$ can be enormously reduced by decomposing the latter as a product of sparse matrices, each corresponding to the addition of a single bond to $\mathcal{L}$.

Specifically we write $\mathcal{T} = \mathcal{T}^0\mathcal{T}^{\mathrm{h}}\mathcal{T}^{\mathrm{v}}$, where $\mathcal{T}^{\mathrm{v}} = \mathcal{T}^{\mathrm{v}}_L \ldots \mathcal{T}^{\mathrm{v}}_2\mathcal{T}^{\mathrm{v}}_1$ is connecting each of the L spin sites in the uppermost row of the strip to a new spin site situated *vertically* above it, and $\mathcal{T}^{\mathrm{h}} = \mathcal{T}^{\mathrm{h}}_{L,1} \ldots \mathcal{T}^{\mathrm{h}}_{2,3}\mathcal{T}^{\mathrm{h}}_{1,2}$ is finishing the new row of $\mathcal{L}$ by appending *horizontal* bonds between each of the nearest-neighbour dangling ends created by $\mathcal{T}^{\mathrm{v}}$. The matrix $\mathcal{T}^{\mathrm{h}}_{L,1}$ imposes periodic boundary conditions by interconnecting the newly added spins at sites L and 1. Finally $\mathcal{T}^0 = \mathcal{T}^0_L \ldots \mathcal{T}^0_2\mathcal{T}^0_1$ furnishes the bonds of $\mathcal{L}_0$ from each of the new spin sites to the ghost site. Each of these single-bond TMs is implicitly understood to depend on the particular realisation of the bond and, in the case of $\mathcal{T}^0_i$, the field randomness pertaining to the bond in question.

Upon addition of one single bond the summation over graphs in Eq. (13) is augmented by a sum over the two possible states of this new degree of freedom, viz. the bond added to $\mathcal{L}$ ($\mathcal{L}_0$) can be either *present* or *absent* in G (G_0). Correspondingly each column of the TM has at most two distinct non-zero entries.

Consider first adding a vertical bond by action of $\mathcal{T}^{\mathrm{v}}_l$, $l \in \{1, \ldots, L\}$. If the bond is 'present' any given connectivity state $(i_1 i_2 \ldots i_L)$ of the L uppermost spin sites will be left unchanged. In case of an 'absent' bond site l will be disconnected, and the number representation of the new connectivity state can be found by assigning $i_l = L+1$ and using Eq. (18). Interpreting the factor of q^N in Eq. (13) as an extra factor of q going with each vertical bond we see that the non-zero entries in $\mathcal{T}^{\mathrm{v}}_l$ corresponding to a column with a given connectivity number are a diagonal contribution of u_{ij} and a possibly off-diagonal contribution of q. In particular the vertical bonds do not induce any loop closures.

Similarly the TM of a horizontal bond $\mathcal{T}^{\mathrm{h}}_{l,l+1}$ has a diagonal entry of 1 for each column, corresponding to the bond being absent. The other non-zero entry corresponds to a present bond, and its value depends on whether a loop is being closed or not. Given the connectivity state $(i_1 \ldots i_l i_{l+1} \ldots i_L)$ of some column in the TM this is determined by comparing i_l and i_{l+1}: if they are equal we get an additional diagonal contribution of $u_{l,l+1}$ corresponding to a loop closure, whereas if they are different there is an off-diagonal entry with value $u_{l,l+1}/q$. In the latter case the connectivity number is found by assigning the value $\min\{i_l, i_{l+1}\}$ to all indices that were formerly equal to either i_l or i_{l+1} and applying Eq. (18). (The reason why we copy the *minimum* index is to ensure the proper handling of spins connected to the ghost site.)

Finally, the TM of a ghost bond $\mathcal{T}^{\mathrm{h}}_l$ has the same form as in the case of a horizontal bond if we make the substitutions $u_{l,l+1} \to v_l$ and $i_{l+1} \to 0$.

3.4. Magnetic properties

It is well known, at least in the case of a pure system, that physically interesting quantities like the central charge c as well as the thermal (x_T) and the magnetic (x_H) scaling dimensions can be extracted from the transfer matrix spectrum. Consider for the moment the case of vanishing magnetic field, $H = 0$. Since connections to the ghost site are then generated with zero weight ($v = 0$) such connections can only be present in any row if they were already there in the preceding row. In particular, noting that in the numbering of connectivities induced by Eq. (18) the non-ghost connectivities precede the others, we see that the TM assumes the following block form [18]

$$\mathcal{T} = \begin{bmatrix} \mathcal{T}^{11} & \mathcal{T}^{12} \\ 0 & \mathcal{T}^{22} \end{bmatrix}, \tag{20}$$

where superscript 2 (1) refers to the (non-)ghost connectivities.

The largest and the next-largest eigenvalues of $\mathcal{T}$ turn out to be the largest eigenvalue of block $\mathcal{T}^{11}$ and $\mathcal{T}^{22}$ respectively, and from the corresponding (reduced) free energies per site $f_0^{ii}(L) = -\frac{1}{L}\lambda_0^{ii}$ ($i = 1, 2$) for a strip of width L the magnetic scaling dimension can be found from the CFT formula [28]

$$f_0^{22}(L) - f_0^{11}(L) = \frac{2\pi x_H}{L^2} + \ldots . \tag{21}$$

Physically this relation to x_H can be understood by noting that by acting repeatedly with $\mathcal{T}^{22}$ on some initial (row) state $|v_0\rangle \neq 0$ one measures the decay of clusters extending back to row 0. This must have the same spatial dependence as the spin–spin correlation function and hence be related to x_H [18]. Analogously $\mathcal{T}^{11}$ measures the decay of two-point correlations between *pairs* of spins being interconnected within the random-cluster model. This is nothing but the energy–energy correlation in the strip geometry, and accordingly we expect that

$$f_1^{11}(L) - f_0^{11}(L) = \frac{2\pi x_T}{L^2} + \ldots . \tag{22}$$

We have checked the results for x_H by constructing a realisation of the TM in the presence of a *seam* spanning the length of the cylinder. Our algorithm also merits attention on its own right since it improves the asymptotic number of basis states necessary for finding $f_0^{22}(L)$ from $d_L - c_L \sim 5^L$ (the dimension of $\mathcal{T}^{22}$) to $Lc_L \sim L4^L$. In practice, however, with the strip widths L accessible using present-day computers the two algorithms perform more or less equally fast (see Table 1 for a comparison).

The well-known duality relation for the Ising model partition function without a magnetic field is easily extended to the case of the Potts model on a cylinder. For $v = 0$ the partition function of the random-cluster model, Eq. (13), can be rewritten as

$$Z = \sum_{G \subseteq \mathcal{L}} \left(\prod_{\langle ij \rangle \in G} u_{ij} \right) q^{C(G)}, \tag{23}$$

where $C(G)$ is the number of independent clusters on G. We first stipulate the duality between two very special graphs. Namely, the full graph $G = \mathcal{L}$ with partition function $Z_{\text{full}}(\{u_{ij}\}) = q\prod_{\langle ij\rangle} u_{ij}$ is taken to be dual to the empty graph $G^* = \emptyset$ with $Z^*_{\text{empty}}(\{u^*_{ij}\}) = q^{N^*}$, where the number of dual sites N^* is fixed by the Euler relation.

Establishing the duality then amounts to ascertaining that all other graphs have the same weight relative to this reference state as is the case in the dual model. In the terminology introduced above, duality states that a graph configuration G on the original lattice $\mathcal{L}$ is dual to a configuration G^* on the dual lattice $\mathcal{L}^*$ in which every bond of strength u_{ij} being 'present' in G corresponds to the dual bond of strength u^*_{ij} being 'absent' from G^*.

In particular, removing one bond from the full graph (relative weight: $1/u_{ij}$) must correspond to adding the corresponding dual bond to the empty dual graph (relative weight: u^*_{ij}/q), meaning that the bond strengths and their duals must obey the relation

$$u_{ij}u^*_{ij} = q. \tag{24}$$

When removing further bonds from G it may happen that a new cluster is separated from the rest of the graph, yielding an additional factor of q. But such a cluster formation corresponds precisely to a loop closure on the dual lattice, also giving an extra factor of q! Since all graph configurations G can be constructed by successive removals of bonds from the full reference state we have thus proven the fundamental duality relation [38]

$$Z(\{u_{ij}\}) = qCZ^*(\{u^*_{ij}\}), \tag{25}$$

where $C = q^{-N^*}\prod_{\langle ij\rangle} u_{ij}$ is some constant.

A similar duality relation can be established for the spin–spin correlation function. As usual we define the local order parameter as [8]

$$M_a(r) = \left(\delta_{\sigma(r),a} - \frac{1}{q}\right), \quad a = 1, \ldots, q. \tag{26}$$

In the high-temperature phase all components of the order parameter vanish, whilst in the ordered (low-temperature) phase the Z_q symmetry is spontaneously broken and one of the components, say $a = 1$, has a positive expectation value. A simple calculation now shows that the correlation function $G_{aa}(r_1, r_2) = \langle M_a(r_1)M_a(r_2)\rangle$ is proportional to the probability that the points r_1 and r_2 belong to the same cluster.

In a cylindrical geometry the graphs with r_1 and r_2, taken to be at opposite ends of the cylinder, connected correspond to dual graphs where clusters are forbidden to wrap around the cylinder. This is equivalent to computing the dual partition function with twisted boundary conditions $\sigma \to (\sigma + 1) \bmod q$ across a *seam* running from r_1 to r_2. By permuting the Potts spin states the shape of this seam can be deformed at will as long as it connects r_1 and r_2. Duality thus maps the correlation function onto a disorder operator

$$\langle M_a(r_1)M_a(r_2)\rangle = \left\langle \prod_{\text{seam}} \exp(-K^*\delta_{\sigma_i,\sigma_j}) \right\rangle_{Z^*}, \tag{27}$$

where $Z^* = Z^*(\{K^*\})$ is the dual partition function with periodic boundary conditions.

The construction of the TM in the presence of a seam is facilitated by the following observation: If no cluster is allowed to wrap the cylinder, each graph contributing to the partition function can be associated with a function $s(j)$ of the row number j, such that $s(j) = k \in \{1, \ldots, L\}$ means that in row j no horizontal bond connecting sites k and $k + 1 \pmod{L}$ is present. For obvious reasons we shall refer to s as the *virtual seam*. We can then write the TM in a basis which is the direct product of the L possible values of the virtual seam and the customary c_L non-ghost connectivities. The virtual seam is initialised by assigning to it a definite value in row 0, viz. $s(0) = L$ for all graph configurations of that row.

The single-bond TM of a vertical bond is diagonal in s, but a present horizontal bond *not* inducing a loop closure may alter the value of the virtual seam. Let us recall from Section 3.3 that to find the connectivity state $(i_1 \ldots i_l i_{l+1} \ldots i_L)$ giving the row label of $T^{\mathrm{h}}_{l,l+1}$ that corresponds to the off-diagonal entry with value $u_{l,l+1}/q$ we would join the two distinct clusters formerly labelled by either i_l or i_{l+1}. But such a merger would ruin the invariant stated above, unless we move the virtual seam at the same time. On the other hand, if $i_l = i_{l+1}$ and $s(j) = l$ we must explicitly prevent a cluster from wrapping the cylinder by leaving out that extra diagonal contribution which would otherwise by implied by the condition $i_l = i_{l+1}$. In this case the virtual seam is not moved.

To conclude this section we remark that in the case of a *planar* geometry any n-point Potts correlation function can be mapped to a generalised surface tension by duality [38–40].

3.5. The percolation limit

In the random-bond Potts model the couplings $u_{ij} \geqslant 0$ are quenched random variables, and the critical point can be accessed by drawing them from the symmetric binary distribution $P(u) = \frac{1}{2}[\delta(u - u_1) + \delta(u - u_2)]$, where $u_1 u_2 = q$. For details, see Section 4. Bond percolation can be studied in the limit $u_1 \to 0$, $u_2 \to \infty$ of infinitely weak and strong bonds, respectively. In this limit considerable simplifications occur in the TM, rendering computations with rather large strip widths feasible.

In the percolation limit all single-bond TMs have only one non-zero entry per column. Recall from Section 3.3 that in the general case there are two such entries of which one is diagonal and the other is 'non-trivial'. In the case of the strong vertical bonds and the weak horizontal bonds only the diagonal entries survive, so that the matrices $\mathcal{T}^{\mathrm{v}}_{\mathrm{strong}} = u_2 \mathbf{1}$ and $\mathcal{T}^{\mathrm{h}}_{\mathrm{weak}} = \mathbf{1}$ both become trivial. On the other hand, a weak vertical bond corresponds to a TM having one non-trivial entry of q per column, whilst a strong horizontal bond is represented by a TM that is u_2 times a non-trivial matrix with entries of 1's and $1/q$'s.

The factors of u_2 multiplying both $\mathcal{T}^{\mathrm{v}}_{\mathrm{strong}}$ and $\mathcal{T}^{\mathrm{h}}_{\mathrm{strong}}$ are innocuous albeit infinite, since of the $2L$ single-bond matrices constituting the entire $\mathcal{T}$ there will on average be L strong ones, hence L factors of u_2. On the level of the specific free energy this amounts to an infinite additive constant

$$f_0^{11}(L) = -\ln u_2 + \tilde{f}_0^{11}(L) \tag{28}$$

independent of the strip width L. In particular, the central charge c can be extracted from the *finite* quantity $\tilde{f}_0^{11}(L)$.

As we shall see in Section 4 this quantity can be found by measuring the asymptotic growth of the norm of $(\prod_{j=1}^{m} \mathcal{T}_j^{11})|v_0\rangle$, where $|v_0\rangle$ is some largely arbitrary initial vector. In the percolation limit the TM turn out to be so sparse that after a very few iterations the resulting vector has only one non-zero component. Computationally this means that it is sufficient to store the row index of that non-zero component as well as its norm. Both time and memory requirements are thus enormously reduced, allowing us to access larger system sizes.

The disadvantage of this projective quality of the percolation point TM is that neither the thermal nor the magnetic scaling dimensions can be found from the Lyapunov spectrum. In the case of x_H an initial vector in the $\mathcal{T}^{22}$ sector will rapidly decay to zero, thus invalidating the procedure for finding $\tilde{f}_0^{22}(L)$, and the alternative of using a seam is obstructed by the fact that disallowing the entry in the horizontal bond TM that corresponds to a cluster wrapping the cylinder is incompatible with the argument of pulling out an overall factor of u_2 from the TM.

4. Results

4.1. Softening of the transition

Before attempting to determine the universality classes of the RBPM it is essential to make sure that quenched bond randomness indeed renders the phase transitions second order. For $q > 4$ the pure system has a first-order transition for which the free energy per site is expected to scale like [18]

$$f_0(L) = f_0(\infty) + aL^{-d}\exp(-L/\xi), \tag{29}$$

where ξ is the bulk correlation length and a is an amplitude depending on q. In Fig. 2 we show plots of the function

$$\lambda(L) \equiv \ln[f_0(L) - f_0(\infty)] + d\ln L \sim \text{const} - L/\xi \tag{30}$$

for various values of q and the randomness strength R. These plots are rather sensitive to the value of $f_0(\infty)$, but although this is only known exactly for the pure model [10] it can nevertheless be determined with sufficient accuracy from the parabolic fits described in Section 4.2 below.

For $q = 8$ the finite correlation length of the pure system ($\xi \sim 70$) is seen to be rendered effectively infinite ($\xi \sim 10^3$) upon imposition of the randomness, whilst the transition of the Ising model ($q = 2$) simply stays second order. Despite the simplicity of these plots we also find a fair agreement with the recently found analytical values of ξ for the pure systems [41]; near $q = 4$ these assume the simple form

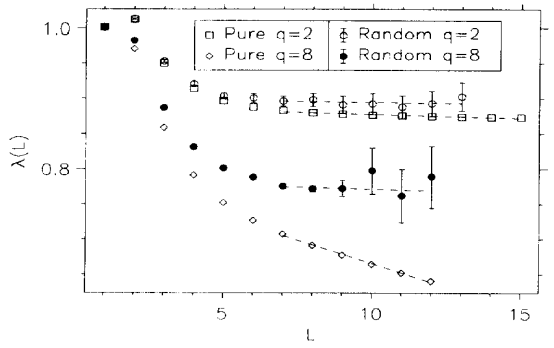


Fig. 2. Plots of $\lambda(L)$, normalised to $\lambda(1) = 1$, showing that bond randomness renders the phase transition second order. The random systems have $R \equiv K_2/K_1 = 2$.

$$\xi \simeq \frac{\sqrt{2}}{8} \exp\left(\frac{\pi^2}{\sqrt{q-4}}\right). \tag{31}$$

Another criterion for distinguishing between first and second-order phase transitions is the values of the (effective) exponents x_H and x_T as found from Eqs. (21) and (22) respectively. Generally speaking, for pure systems with $q > 4$ these equations give rise to rather poor fits which however have extrapolated values of the effective exponents that are in the vicinity of, and slightly below, zero, whereas when randomness is added the fits are much better and yield exponents in the interval $]0, 2[$. In view of the problems justifying such fits in the random case (see below) this evidence for a softening of the transition is however not to be taken too seriously.

4.2. Central charge at the random FP

The free energy per site $f_0^{11}(L)$ for the RBPM on long strips of width L is readily found from Eq. (6) applied to the $\mathcal{T}^{11}$ sector of the TM. We have performed extensive simulations for various values of q and the randomness strength R, though in most cases $R = 2$ was found to describe the random FP adequately.

Representative samples of our data are shown in Table 2. For each run a normalised initial vector $|v_0\rangle$ was prepared by choosing its components randomly, and after discarding the results of the first 2000 multiplications by $\mathcal{T}_i^{11}$ in order to eliminate transients, data collection was made for each 200 iterations until a strip of a total length of $m = 10^5$ had been built up. For $q > 2$ a total of 100 independent runs were made for $1 \leqslant L \leqslant 8$, and 3 runs for $9 \leqslant L \leqslant 12$, whilst for the Ising model ($q = 2$) we were able to make 100 runs for $1 \leqslant L \leqslant 13$ by using the conventional spin basis. Final results and error bars were extracted by computing the mean and the standard deviation for the totality of patches of length 200.

It is not a priori obvious that the Lyapunov exponents found from Eq. (6) are independent of the norm used. The standard norm in both the spin basis and the

Table 2
Critical free energies per site, $-f_0^{11}(L)$, for $R = 2$ and various values of q. The figures in parentheses indicate the error bar on the last quoted digits.

L	$q = 2$	$q = 3$	$q = 4$	$q = 8$	$q = 16$	$q = 64$
1	2.17460 (12)	2.62881 (13)	2.96193 (13)	3.80035 (16)	4.68198 (18)	6.54635 (24)
2	1.95329 (8)	2.26650 (9)	2.49558 (9)	3.06980 (10)	3.67393 (11)	4.95619 (14)
3	1.90971 (7)	2.19534 (7)	2.40405 (7)	2.92819 (8)	3.48241 (9)	4.67423 (11)
4	1.89550 (6)	2.17203 (6)	2.37431 (6)	2.88328 (7)	3.42387 (8)	4.59557 (10)
5	1.88895 (5)	2.16182 (6)	2.36126 (6)	2.86392 (6)	3.39934 (7)	4.56442 (9)
6	1.88568 (5)	2.15649 (5)	2.35449 (5)	2.85395 (6)	3.38683 (6)	4.54893 (8)
7	1.88377 (4)	2.15328 (5)	2.35040 (5)	2.84798 (5)	3.37948 (6)	4.53974 (7)
8	1.88250 (4)	2.15113 (4)	2.34782 (4)	2.84424 (5)	3.37479 (5)	4.53394 (7)
9	1.88164 (4)	2.14993 (24)	2.34624 (25)	2.84172 (11)	3.37186 (31)	4.53017 (39)
10	1.88098 (4)	2.14858 (23)	2.34504 (23)	2.84011 (26)	3.36918 (29)	4.52653 (25)
11	1.88048 (4)	2.14804 (22)	2.34386 (22)	2.83851 (24)	3.36768 (29)	4.52465 (34)
12	1.88017 (3)	2.14744 (20)	2.34314 (21)	2.83765 (24)	3.36639 (26)	4.52316 (34)
13	1.87991 (3)					

connectivity basis is given by the square root of the sum of the squared components, and these two are of course not identical. To impose the spin basis norm on the connectivity basis each term in the sum must be weighted by a factor q^C, where C is the number of clusters in the relevant connectivity state. We have checked the consistency of our results by comparing the first few Lyapunov exponents obtained from imposing the two different norms on the connectivity basis, and we find that not only are the results identical but there is even a complete agreement of the first three significant digits of the error bars. For $q = 2$ we found that the results using the spin basis and the connectivity basis were consistent, but that the error bars obtained using the spin basis were slightly smaller.

Our results for the free energies of the random-bond Ising model agree with, and are more precise than, those of de Queiroz[1] [42].

Values of the effective central charge c' can be extracted from Eq. (5) by employing various fitting procedures. In spite of the relatively slow convergence of both two-point fits $(L, L+1)$ and straight-line least-squares fits against $1/L^2$ [42], iterating such fits yields quite good results in the pure model. When randomness is added this is no longer so, since rather substantial error bars on the first estimates prevent us from efficiently iterating the fits.

A better scheme is to include the leading correction to the scaling of Eq. (5), which in the pure case has been shown numerically to take the form [18]

$$f_0^{11}(L) = f_0(\infty) - \frac{\pi c'}{6L^2} + \frac{A}{L^4} + \ldots \,. \tag{32}$$

[1] Actually they differ by a constant, since Ref. [42] defines the Hamiltonian as $-\sum_{\langle ij \rangle} K_{ij}\sigma_i\sigma_j$ as opposed to our Eq. (1). Since $\sigma_i\sigma_j = 2\delta_{\sigma_i\sigma_j} - 1$ there is a free energy difference of $2\overline{K_{ij}}$, which for $R = 2$ equals 0.91407.

Table 3
Effective central charge c' extracted from parabolic fits with $L_0 \leqslant L \leqslant L_{\max}$, as described in the text. Error bars on the last quoted digit are shown in parentheses. The choice $L_0 = 3$ appears to be optimal, provided that the strength of the randomness R is large enough (see text), and the corresponding values of c', shown in bold face, should be regarded as our results.

L_0	$q = 2$ $R = 2$	$q = 3$ $R = 2$	$q = 4$ $R = 2$	$q = 8$ $R = 2$	$q = 8$ $R = 10$	$q = 16$ $R = 2$	$q = 16$ $R = 10$	$q = 64$ $R = 2$	$q = 64$ $R = 10$
1	0.563(1)	0.927(1)	1.184(1)	1.787(1)	1.731(3)	2.330(1)	2.322(4)	3.120(1)	3.476(5)
2	0.508(2)	0.825(3)	1.042(2)	1.515(3)	1.586(10)	1.864(3)	2.101(10)	2.194(4)	3.150(13)
3	**0.499(3)**	**0.800(6)**	**1.003(6)**	1.441(7)	**1.521(23)**	1.752(8)	**2.052(25)**	2.065(10)	**3.034(30)**
4	0.500(6)	0.813(14)	0.994(14)	1.424(15)	1.548(52)	1.750(17)	2.089(57)	2.157(22)	3.079(68)
5	0.505(11)	0.842(30)	1.005(31)	1.426(33)	1.622(113)	1.785(38)	2.203(125)	2.305(47)	3.209(148)
6	0.500(20)	0.818(62)	0.963(63)	1.360(67)	1.587(228)	1.794(78)	2.196(251)	2.384 (93)	3.213(300)

One then performs either three-point fits $(L, L+1, L+2)$ or parabolic least-squares fits against $1/L^2$, and because of the much faster convergence no iteration is needed [42]. Although a correction proportional to $1/L^4$, due to the operator $T\overline{T}$, must necessarily be present in every system that is conformally invariant [43] it can of course not be guaranteed to be the dominant one in general.

In Table 3 the results of parabolic fits including the data points for $L_0 \leqslant L \leqslant L_{\max}$ have been shown as a function of L_0. It is seen that L_0 must be chosen large enough to justify the omission of higher order terms in the series (32), and small enough to minimise error bars. From the special cases of the Ising model and of the percolation point (see Section 4.3 below) we concluded that the choice $L_0 = 3$ is optimal.

Apart from the results shown in Table 3 we have also performed some runs for $q = 1.5$, finding, as expected from the Harris criterion [1], no difference between the results for the pure and the random model.

In the intermediate regime $2 \leqslant q \leqslant 4$ our results compare favourably to those of the $(q-2)$-expansion, at least up to $q = 3$. On the other hand, it is evident from Fig. 3 that the difference between c for the pure model and c' for the random one is of the same order of magnitude as our error bars, and only near $q = 4$, where the expansion is expected to break down anyway, are our results able to distinguish between the two different behaviours. Exactly at $q = 2$ the randomness is marginal and logarithmic corrections to the finite-size scaling forms, Eqs. (4) and (5), are expected. Whilst this issue has recently attracted considerable interest in the case of the critical exponents [44] the corrections to the central charge are much weaker [45] and accordingly our result is consistent with that of the pure Ising model.

In Fig. 4 we have displayed our results for c' as a function of $\log_{10} q$ for selected values of $q \in [1.5, 64]$. We have juxtaposed the results for two strengths of the randomness, namely weak randomness ($R = 2$, closed circles on Fig. 4) and strong randomness ($R = 10$, open circles). For small values of q both randomness strengths give rise to the same c', as witnessed by the overlap of the $q = 4$ data points. However, for larger q the $R = 2$ curve flattens out and grows slower than logarithmically. Sample runs show that the same is true for larger values of R, the difference being that the range of q-values

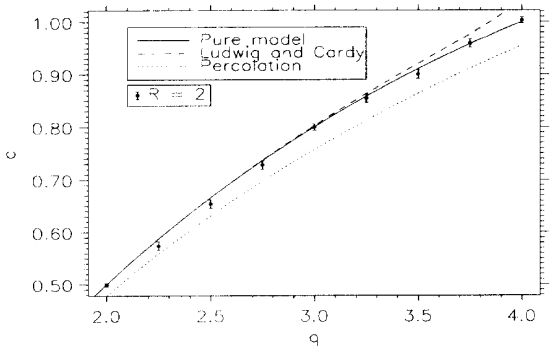


Fig. 3. The effective central charge c' as a function of q for $2 \leqslant q \leqslant 4$. The perturbative results by Ludwig and Cardy [14] do not differ appreciatively within the range of q-values where the expansion is supposed to be valid. Accordingly the numerical data are unable to distinguish between pure and non-trivial random behaviour. They are also quite close to, but distinguishable from, the percolation point values.

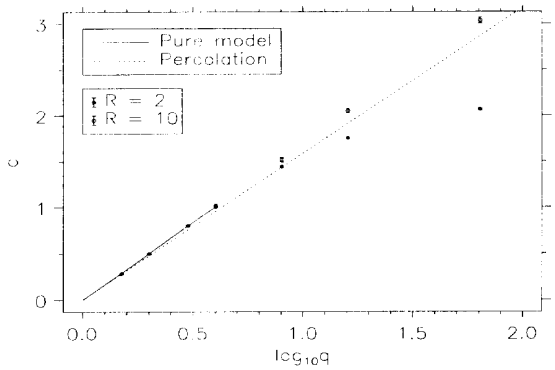


Fig. 4. Effective central charge as a function of $\log_{10} q$ for $1.5 \leqslant q \leqslant 64$. For large q the data for $R = 10$ are supposed to represent the true behaviour at the random fixed point, as argued in the text.

for which the growth is logarithmic is extended as R is increased. This is illustrated by the $R = 10$ curve's staying above, but very close to, the percolative result $\sim \log q$ (see Section 4.3 below) for the whole range of q-values shown on the plot. Another way to state this is that for fixed q and varying R, the quantity c' is an increasing function of R that eventually reaches a plateau as R becomes large enough. It then appears from Fig. 4 that for $q \leqslant 64$ the randomness strength $R = 10$ is sufficient to reach this plateau.

These findings are interpreted as follows. According to the $(q-2)$-expansion the randomness strength R^* corresponding to the random FP is an increasing function of q. Assuming this FP to persist as we enter the regime $q > 4$ (see Fig. 1) we now claim that the monotonicity of $R^*(q)$ also holds true when the $(q-2)$-expansion breaks down. From the RG flows given in Fig. 1 we see that any initial value of $R \in]1, \infty[$ will

Table 4
Effective central charge c' obtained using a trinary distribution of the random bonds. There is a fraction p of very weak and very strong bonds respectively, the remaining fraction $1-2p$ being critical.

p	$q=2$	$q=4$	$q=8$	$q=16$	$q=32$	$q=64$
0.05	0.522 (25)	1.030 (26)	1.477 (27)	1.780 (29)	1.920 (30)	1.974 (31)
0.10	0.519 (35)	1.032 (36)	1.510 (38)	1.915 (40)	2.251 (42)	2.552 (43)
0.15			1.539 (46)	1.996 (48)	2.416 (50)	2.817 (52)

eventually flow to the random FP as the system is viewed on larger and larger length scales. However, if we start out very far from R^* the onset of the asymptotic scaling given by Eq. (5) may be deferred to much larger length scales than the strip widths L numerically accessible for our TMs. We therefore expect poor scaling for strip widths $L \leqslant L_{\max}$. Conversely, if we choose the strength of the randomness as $R \sim R^*$ the resulting value of c' is expected to be more or less independent of the precise choice of R and equal to the true value of the central charge. But in our simulations we find that this is precisely accomplished by choosing R as an increasing function of q. Further justification for this interpretation is found from the phenomenological RG treatment in Section 4.6 below.

A heuristic argument explaining that the 'effective' $c'(R,q)$ obtained for small values of R is less than the 'correct' value $c'(R^*,q)$ associated with the random FP is readily furnished, at least for large q. Namely, from Zamolodchikov's c-theorem [46] we know that there exists a function $c(\{K\})$ of the couplings that decreases along the RG flow and equals the central charge at the fixed points. As a corollary the curves of constant c are orthogonal to the RG flow. In particular, for large q where the RG flow is known from the mapping to the RFIM (see Eqs. (9) and (10)), it is evident from Fig. 1 that $c'(R,q)$ is equal to $c'(R^*,q^*)$ for some $q^* < q$. Since our numerical results indicate that $c'(R^*,q)$ is an increasing function of q the proposition follows.

To check our results for c' we have also made 100 independent runs for each of the strip widths $1 \leqslant L \leqslant 8$ where the random bonds were drawn from the *trinary* distribution

$$P(K) = p[\delta(K-K_1) + \delta(K-K_2)] + (1-2p)\delta(K-K^*), \tag{33}$$

where K_1 and $K_2 = 1000K_1$ satisfy the criterion (3) and $(\exp K^* - 1)^2 = q$. Here $p \ll 1$ is the strength of the randomness. Of course this realisation of the randomness also preserves self-duality, and hence the model is again at its critical point [27].

Numerical results for c' using trinary randomness are shown in Table 4 and they are consistent with the binary results given above, again provided that p is increased as we go to larger and larger q. In particular it is reassuring to verify that we seem to probe the true random behaviour when $2/p$ (the length scale associated with this randomness) is comparable to the correlation length of the pure system (31).

An interesting question is whether the asymptotic value of c' is approached from above or below when the system is viewed on larger and larger length scales. For

Table 5
Effective central charge c' extracted from linear fits of $f_0^{11}(L) - f_0^{11}(\infty)$ versus $1/L^2$, with $L_0 \leqslant L \leqslant L_{\max}$. In all cases the approach towards the asymptotic values of Table 3 is from above. Error bars on the last quoted digit are shown in parentheses.

L_0	$q = 2$ $R = 2$	$q = 3$ $R = 2$	$q = 4$ $R = 2$	$q = 8$ $R = 10$	$q = 16$ $R = 10$	$q = 64$ $R = 10$
2	0.5662 (6)	0.9300 (7)	1.1903 (7)	1.723 (3)	2.313 (3)	3.469 (4)
3	0.535 (1)	0.878 (1)	1.117 (1)	1.656 (5)	2.207 (6)	3.311 (7)
4	0.521 (2)	0.848 (3)	1.075 (3)	1.605 (10)	2.148 (11)	3.206 (13)
5	0.514 (3)	0.836 (5)	1.051 (5)	1.585 (19)	2.125 (20)	3.163 (24)
6	0.512 (4)	0.839 (9)	1.043 (9)	1.595 (34)	2.144 (38)	3.174 (45)
7	0.510 (6)	0.840 (18)	1.022 (18)			
8	0.509 (9)	0.812 (33)	1.011 (34)			
9	0.503 (14)					
10	0.500 (23)					

models exhibiting reflection positivity Zamolodchikov's c-theorem [46] ensures that the convergence is from above. In particular the condition of positivity holds true for unitary models, whereas for a random model it may well fail to be fulfilled. Indeed, in the case of the RBPM a perturbative calculation [14] suggests that the convergence may be from below in some cases.

In order to discuss this point the parabolic fits versus $1/L^2$ employed above are no longer appropriate. Apart from speeding up the rate of convergence to a point where information about its direction becomes obliterated due to error bars the inclusion of higher-order corrections to the finite-size scaling form (5) may have the effect of reversing this direction. For example, in the case of the pure Ising model it is found [42] that the estimators obtained from parabolic fits converge from below, whereas the corresponding linear fits (i.e. without the $1/L^4$ correction) yield estimators that converge from above in accordance with the theoretical prediction.

In Table 5 we show the results of such linear least-squares fits for several values of q. The randomness strength R was chosen in accordance with the considerations given above. It appears that in all cases the finite-size estimators converge towards the asymptotic values of Table 3 from above.

We remark that values of c' similar to ours have recently been reported by Picco [47]. For $q = 8$ this author found $c' = 1.45 \pm 0.06$ which agrees with our result of respectively $c' = 1.52 \pm 0.02$ for binary randomness of strength $R = 10$, and $c' = 1.51 \pm 0.04$ for trinary randomness of strength $p = 0.10$. Our observation that c' appears to be an increasing function of R, eventually reaching a plateau as R becomes large enough, was confirmed by Ref. [47] that used binary randomness of strength $R = 10$ throughout. Strong evidence was also given that $c'(q)$ grows roughly logarithmically with q in the regime $q \in [5, 256]$, but a further discussion of what this implies will be deferred to Section 5 below.

It is worthwhile to compare the TM algorithm used in Ref. [47] to ours. It was found that the number of distinct entries in the pure model TM in the spin basis is

$$b_L = \sum_{i_2=1}^{2} \sum_{i_3=1}^{m_3} \sum_{i_4=1}^{m_4} \cdots \sum_{i_L=1}^{m_L} 1, \tag{34}$$

where $m_i = \max(i_2, i_3, \ldots, m_{i-1})$, and L designates the strip width as usually. Further taking into account the 2^L different realisations of the binary randomness in each strip, recursion relations between the different elements of the TM were found by computing a total of $(b_L)^2 2^L$ polynomials. Since this number of polynomials increases rapidly with L high-precision computations could only be performed up to $L_{\max} = 6$. The number of iterations used to determine $f_0(L = 6)$ was similar to ours, whereas more iterations were used for the smaller strip widths.

Evidently this algorithm also has the advantage that q enters only as a parameter, thus making accessible any value of q for the simulations. However, for large L it performs inefficiently, as we will now show. The numbers b_L of Eq. (34) are by no means unfamiliar. Indeed, they are nothing but the total number of L-point connectivities, including the non-well-nested ones [39]. Alternatively they can be viewed as the number of ways that L objects can be partitioned into indistinguishable parts [40]. With m_ν parts of ν objects each ($\nu = 1, 2, \ldots$) this can be rewritten as

$$b_L = {\sum_{m_\nu=0}^{\infty}}' \prod_{\nu=1}^{\infty} \frac{L!}{(\nu!)^{m_\nu} m_\nu!}, \tag{35}$$

where the primed summation is constrained by the condition $\sum_{\nu=1}^{\infty} \nu m_\nu = L$. From this representation the generating function can be immediately inferred

$$\exp(\mathrm{e}^t - 1) = \sum_{n=0}^{\infty} \frac{b_n t^n}{n!}. \tag{36}$$

Explicit values, found by Taylor expansion of the left-hand side, are shown in Table 1. Asymptotically the b_L are seen to grow faster than L^L whereas the well-nested connectivities only grow as $\sim 4^L$.

4.3. The percolation limit

In the case of the binary randomness (2) the percolation limit is reached by letting $(\mathrm{e}^{K_1} - 1) \to 0$ and $(\mathrm{e}^{K_2} - 1) \to \infty$ whilst maintaining the self-duality criterion (3). The partition function of the random-cluster model is then dominated by one graph only, viz. the one that covers all of the strong bonds and none of the weak ones. (Note in particular that the limits $R \to \infty$ and $q \to \infty$ do not commute.) Expressed in terms of the free energy per site this reads

$$f_0^{\mathrm{perc}} = -\frac{B}{N} \ln(\mathrm{e}^{K_2} - 1) - \frac{C}{N} \ln q, \tag{37}$$

where B is the number of strong bonds and C is the number of clusters in the dominant graph.

The quenched average over the randomness must be taken on the level of the free energy. Evidently, with the chosen distribution of the randomness, $\overline{B} = N$ whence the first term is simply a trivial, albeit infinite, constant. (Incidentally this is the same constant that was pulled out in Eq. (28).) On the other hand, the average number of percolation clusters is related to a derivative in the *pure* Q-state Potts model [8]

$$\overline{C} = \frac{\partial}{\partial Q} \ln Z(Q) \bigg|_{Q=1}, \tag{38}$$

thus determining the effective central charge $c'(q)$ at percolation as

$$c'(q) = \ln q \, \frac{\partial c(Q)}{\partial Q} \bigg|_{Q=1}. \tag{39}$$

An alternative argument for this relation is furnished by the observation that the replicated model is simply the Potts model with q^n states; differentiating $c(q^n)$ with respect to the number of replicas n and taking the limit $n \to 0$ one recovers the result (39). The central charge of the pure model is given by an expression due to Kadanoff [43,48]

$$c(Q) = \frac{(2-3y)(1+y)}{(2-y)}, \tag{40}$$

where $\sqrt{Q} = 2\cos(\pi y/2)$ and $0 \leqslant y \leqslant 1$, and taking the appropriate derivative of this we finally arrive at

$$c'(q) = \frac{5\sqrt{3}}{4\pi} \ln q. \tag{41}$$

As described in Section 3.5 the single-bond TMs in the percolation limit have only one non-zero entry per column, equal to either q, 1 or $1/q$. Taken together with their projective quality and Eq. (6) for the largest Lyapunov exponent it is clear that the free energy, and hence the central charge, must be explicitly proportional to $\ln q$. So it suffices to do the numerics for one value of $q \neq 1$. Because of the simple form of these TMs we were able to average $\tilde{f}_0^{11}(L)$ of Eq. (28) over 100 strips of length $m = 10^5$ for the range $1 \leqslant L \leqslant 19$. Consequently the factor of proportionality could be determined quite accurately as 0.688 ± 0.003, in excellent agreement with $5\sqrt{3}/4\pi \simeq 0.689$.

It is evident from the mapping between bond percolation and the pure $Q = 1$ Potts model that the critical exponents of the two models are identical: $x_T = \frac{5}{4}$ and $x_H = \frac{5}{48}$ [43]. Since all correlation functions at percolation can only take the values 0 and 1, it is also clear that different moments of a given correlation function all have the same scaling dimension. Thus, in the notation of Ludwig [16], $x_n = x_1$ for all $n > 1$. The pure model represents the other trivial extreme case of multiscaling behaviour: $x_n = nx_1$.

4.4. The cumulant expansion

The concept of multiscaling is best understood in terms of a simple example [49]: The random-bond Ising chain. From the reduced Hamiltonian $\mathcal{H} = -\sum_{i=1}^{N} K_i s_i s_{i+1}$ the

partition function is easily found as $Z = \prod_{i=1}^{N} c_i$, where $c_i = 2\cosh K_i$. In particular the quenched average $\overline{Z} = \exp[N \log \overline{c_i}]$ does *not* coincide with the most probable value $Z_{\text{m.p.}} = \exp[N\overline{\log c_i}]$. On the other hand, the reduced free energy is $F = -\sum_{i=1}^{N} \log c_i$ so that $\overline{F} = F_{\text{m.p.}} = -N\overline{\log c_i}$. The free energy is thus *self-averaging*, i.e. it takes on its sample averaged value with probability unity in the thermodynamic limit. Indeed, by the central limit theorem, F is normally distributed, it being a sum of random numbers, whereas Z is a product of random numbers and therefore *log*-normally distributed. Similarly the correlation function $\langle s_1 s_R \rangle = \prod_{i=1}^{R-1} \tanh K_i$ is non-self-averaging. In particular $\overline{\langle s_1 s_R \rangle^2} \neq \overline{\langle s_1 s_R \rangle}^2$.

In Section 3.4 we related the spin–spin correlation function $G(m)$ on a strip of the RBPM to the free energy in the presence of a seam of frustrated bonds (or with a ghost site). Taking the logarithm of Eq. (27) and exploiting the self-duality of the lattice we have

$$\Delta f(L) \equiv f_0^{22}(L) - f_0^{11}(L) = \frac{1}{mL} \ln G(m), \tag{42}$$

and in the pure system, according to conformal symmetry [28], this decays along the strip as $2\pi x_H / L^2$, cf. Eq. (21). When randomness is present $\Delta f(L)$ is a fluctuating quantity, and since free energies are supposed to be normally distributed these fluctuations are $\mathcal{O}(1/\sqrt{m})$. Consequently $\ln G$ is a self-averaging quantity and G is not [49], exactly as in the simple example given above.

In the *multiscaling* scenario of Ludwig [16] different moments $\overline{G(m)^n}$ scale with dimensions x_n which, as opposed to what is the case in the pure model, are not necessarily linear in n. (In this notation $x_H \equiv x_1$.) For $n_1 > n_2$ we have $x_{n_1} \geqslant x_{n_2}$ and $x_{n_1}/n_1 \leqslant x_{n_2}/n_2$ (convexity); pure and percolative behaviour are thus realisations of the two possible extremes of multiscaling.

Since translational invariance is one of the basic assumptions of conformal symmetry [43], the latter only refers to the averaged quantities $\overline{G(m)^n}$ and not to the $G(m)^n$ themselves. These averages cannot be computed directly in a numerical experiment because of the lack of self-averaging; this can however be circumvented by performing a cumulant expansion

$$\ln \overline{G^n} = n\overline{\ln G} + \tfrac{1}{2} n^2 \overline{(\ln G - \overline{\ln G})^2} + \ldots, \tag{43}$$

where each term on the right-hand side *is* self-averaging and can be directly extracted from the statistical fluctuations in $\Delta f(L)$ between the patches of length 200 into which we have divided our strip.

Quite generally for a stochastic variable X we have

$$\langle \exp X \rangle = \exp \sum_{j=1}^{\infty} \frac{1}{j!} k_j, \tag{44}$$

where explicit expressions for the six first cumulants k_i in terms of the moments m_i of X are given by [50]

Table 6
The first six cumulants of $-\Delta f(L)$ for $1 \leqslant L \leqslant 7$ and $R = 2$. The error bar on the first cumulant (shown in parentheses) is related to the second cumulant; error bars on the higher cumulants are not shown.

	L	1. cumulant	2. cumulant	3. cumulant	4. cumulant	5. cumulant	6. cumulant
$q = 3$	1	−1.039786 (242)	0.060716	−0.000791	0.000830	−0.004583	−0.000778
	2	−0.253209 (146)	0.012391	−0.000347	0.000279	0.000059	0.000386
	3	−0.106163 (113)	0.004963	−0.000246	0.000063	−0.000102	0.000046
	4	−0.057901 (95)	0.002784	−0.000143	0.000006	0.000034	0.000003
	5	−0.036521 (84)	0.001810	−0.000105	0.000001	−0.000002	−0.000003
	6	−0.025172 (76)	0.001289	−0.000075	0.000008	−0.000001	−0.000002
	7	−0.018426 (69)	0.000968	−0.000069	0.000002	0.000002	−0.000001
$q = 8$	1	−1.380171 (289)	0.104382	0.004069	0.014001	0.019452	−0.013889
	2	−0.326484 (177)	0.028366	−0.001683	−0.000432	0.000157	−0.003145
	3	−0.132560 (138)	0.014908	−0.001822	0.000356	0.000221	−0.000083
	4	−0.071296 (115)	0.010129	−0.001610	0.000319	−0.000959	0.002323
	5	−0.044886 (102)	0.007880	−0.001619	0.000252	−0.000082	−0.000456
	6	−0.031045 (92)	0.006450	−0.001538	0.000607	−0.000184	−0.001096
	7	−0.022851 (84)	0.005401	−0.001505	0.000237	0.000234	−0.000280

$$
\begin{aligned}
k_1 &= m_1, \\
k_2 &= m_2 - m_1^2, \\
k_3 &= m_3 - 3m_2m_1 + 2m_1^3, \\
k_4 &= m_4 - 4m_3m_1 - 3m_2^2 + 12m_2m_1^2 - 6m_1^4, \\
k_5 &= m_5 - 5m_4m_1 - 10m_3m_2 + 20m_3m_1^2 + 30m_2^2m_1 - 60m_2m_1^3 + 24m_1^5, \\
k_6 &= m_6 - 6m_5m_1 - 15m_4m_2 + 30m_4m_1^2 - 10m_3^2 \\
&\quad + 120m_3m_2m_1 - 120m_3m_1^3 + 30m_2^3 - 270m_2^2m_1^2 + 360m_2m_1^4 - 120m_1^6.
\end{aligned}
$$

We have computed these six cumulants of $\Delta f(L)$ for various values of R and q, based on 100 independent strips of length $m = 10^5$ and width $1 \leqslant L \leqslant 7$. Sample results for $R = 2$ and $q = 3, 8$ are shown in Table 6.

For $q = 3$ the cumulant expansion converges well. The magnitude of the higher cumulants decreases very rapidly, especially for $L \geqslant 3$, and reliable estimates for the left-hand side of Eq. (43) can be obtained simply by summing the first 3 or 4 cumulants, at least when n is not too large. Performing parabolic least-squares fits using Eq. (21) with an $1/L^4$ correction we thus expect to extract quite accurate values of x_n at the random FP.

As q increases the convergence is slower. This is witnessed by the $q = 8$ results of Table 6 decreasing noticeably slower, both for a definite cumulant as a function of L (vertically) and for a definite L as a function of the cumulant number (horizontally). The approximation of leaving out the higher cumulants in the sum (43) thus becomes increasingly difficult to justify, and eventually the cumulant expansion breaks down. This problem is enhanced by the fact that for $q > 8$ we expect a randomness strength of $R = 2$ to be insufficient in order to access the true behaviour at the random FP. We

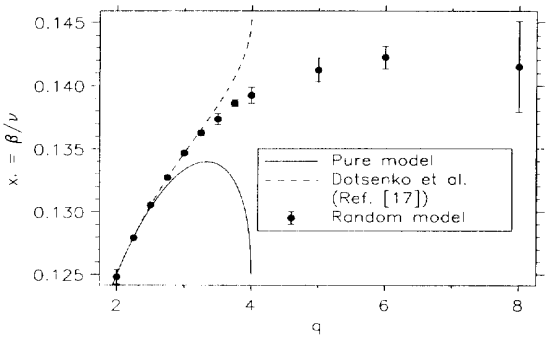


Fig. 5. Magnetic exponent $x_1 = \beta/\nu$ for $R = 2$ as obtained from the cumulant expansion. x_1 is an increasing function of q, continuously connecting onto the perturbative results near $q = 2$. For $q > 8$ the cumulant expansion begins to break down.

are thus forced to increase R, whence the fluctuations become even more violent and the cumulant expansion accordingly ill-behaved.

Our results for x_1 are shown in Fig. 5. Since error bars on the individual cumulants are related to the magnitude of the higher cumulants the question of how to assign a final error bar to x_1 becomes a delicate one. We have addressed this issue by averaging the estimates for x_1 obtained from various parabolic least-squares fits. More precisely, the average is calculated from 4 values, namely fits with $L_0 = 3$ or 4 and including the first 3 or 4 cumulants on the right-hand side of Eq. (43). The consistency of these 4 values is regarded as a check of the validity of the expansion.

In particular, for $q = 3$ we find $x_1(3) = 0.13467 \pm 0.00013$ which is 10 standard deviations above the value $x_1^{\text{pure}}(3) = \frac{2}{5} \simeq 0.13333$ of the pure three-state Potts model [43] and at the same time in perfect agreement with the result $x_1(3) = 0.13465 + \mathcal{O}(\epsilon^4)$ of the $(q-2)$-expansion [17]. The Monte Carlo result $x_1(3) = 0.1337 \pm 0.0007$ of Picco [51] was not able to distinguish convincingly between pure and random behaviour.

For $q = 4$ our result is $x_1(4) = 0.1396 \pm 0.0005$, in nice agreement with Picco's preliminary result $x_1(4) \sim 0.139$ [52] and decidedly different from the corresponding pure value of $x_1^{\text{pure}}(4) = \frac{1}{8}$.

As discussed at length in the Introduction a major motivation for this work was to determine whether the impurity softened transitions for $q > 4$ do or do not have the critical exponents of the pure Ising model. The data of Fig. 5 clearly show a smooth continuation of the perturbative results [16,17] exhibiting no singularity whatsoever at $q = 4$. Our result $x_1(8) = 0.1415 \pm 0.0036$ is comfortably away from the pure Ising value and provides a striking piece of evidence for both our phase diagram and the FP structure of the $(q-2)$-expansion.

All the results quoted for x_1 were computed using $R = 2$. We have checked that other values of R yield results consistent herewith, provided that R is chosen neither too small, in which case the cross-over length $\xi_X \sim \exp(1/2Aw^2)$ found from Eq. (9) becomes

too large for the random FP to be reached, nor too large, in which case the cumulant expansion breaks down. The same holds true when the random bonds are drawn from the trinary distribution (33) with various values for the dilution parameter p.

Because of the positive sign of the second cumulant the values of x_1 are invariably smaller than those one would have obtained without the cumulant expansion (i.e. using only the first cumulant). The latter, however, determine a universal exponent α_0 that describes the asymptotic decay of the spin–spin correlation function in a *fixed* sample at criticality. In terms of the multiscaling exponents this reads

$$\alpha_0 = \left.\frac{dx_n}{dn}\right|_{n=0}. \tag{45}$$

Near $q = 2$ Ludwig obtained the expansion [16]

$$\alpha_0 = x_1^{\text{pure}} + \tfrac{1}{16} y + \mathcal{O}(y^2), \tag{46}$$

where y is the RG eigenvalue of the energy operator coupling to the bond randomness. Our results for small fractional values of $(q - 2)$ are in good agreement with this expression, and if one takes into account the logarithmic corrections expected exactly at $q = 2$ it seems that the theoretical prediction and the numerical results have a common tangent at $q = 2$. For the physically interesting case of $q = 3$ the agreement is not so good. We believe that this apparent discrepancy would be resolved if the expansion (46) could be carried through to three-loop order as in the case of x_1.

Similar remarks can be made about the higher moments of the spin–spin correlation function for which we are unable to verify Ludwig's expansion [16]

$$x_n = n x_1^{\text{pure}} - \tfrac{1}{16} n(n-1) y + \mathcal{O}(y^2). \tag{47}$$

Nevertheless it should be remarked that the fact that the higher cumulants do not vanish in itself implies multiscaling.

Before concluding this section we should like to give a heuristic argument that the cumulant expansion breaks down for large q. In a replica formulation we can imagine the central charge $c(n)$ as a function of the number of replicas n. In this notation the central charge of the pure and the random systems are $c(1)$ and $c'(0)$ respectively, where the prime denotes a differentiation with respect to n. The partition function of the replicated strip is then

$$\overline{Z^n} = \int \exp(-nmLf) P(f)\, df = \exp\left(-mL\overline{f} + \frac{\pi m c(n)}{6L} - \ldots\right), \tag{48}$$

where $P(f)$ is the probability distribution of the free energy. Differentiating this expression twice with respect to n and taking the replica limit $n \to 0$ we infer that the second cumulant of f contains a term that is proportional to $c''(0)$. The cumulant expansion is thus expected to break down if $c(n)$ has a large curvature at $n = 0$.

For $2 \leqslant q \leqslant 4$ the replicas are weakly coupled, since $c(1) \simeq c'(0)$ [14]. Hence $c''(0) \ll 1$. But when $q = 4 + \epsilon$ the transition of the pure system goes first order

so that the function $c(n)$ starts out with slope $c'(0) = 1$ and somehow curves down to assume the value $c(1) = 0$. Consequently $c''(0) = \mathcal{O}(1)$ and the higher cumulants begin to contribute significantly to the sum (43). Finally, for $q \gg 4$ we are in the strong coupling regime. We still have $c(1) = 0$ and as our numerical data indicate that $c'(0) \sim \ln q$ it follows that $c''(0) \gg 1$. This means that the cumulant expansion must break down.

One may speculate whether the transition actually becomes first order whenever $q'' > 4$. Clearly this is the case for the pure Potts model [10], but a similar statement is true when N Ising models are coupled by their local energy density. Namely, in this case an RG analysis [7] implies a fluctuation-driven first-order transition whenever $N > 2$, that is to say for $2^N > 4$. If this conjecture is correct one would then suppose the function $c(n)$ to vanish for $n \geqslant n_0$, where $q^{n_0} = 4$. Evidently such a scenario is in accordance with our observation that $c''(0) \gg 1$ for $q \gg 4$.

4.5. *The thermal exponent*

Because of the rather striking success of the cumulant expansion for x_1 one would now expect the thermal exponent x_T to be similarly related to the fluctuations of $\Delta f_T(L) = f_1^{11}(L) - f_0^{11}(L)$. Surprisingly, this seems not to be the case. Computing the equivalent of α_0, i.e. using only the first cumulant, we find the following results for different values of q: $\alpha_0^T(2) = 1.028 \pm 0.001$, $\alpha_0^T(3) = 0.91 \pm 0.01$, $\alpha_0^T(4) = 0.81 \pm 0.02$ and $\alpha_0^T(8) = 0.65 \pm 0.01$. As remarked above the results using more cumulants can only be lower.

This is bad news since the quenched correlation length exponent ν can be shown quite rigorously to satisfy the bound [19,20]

$$\nu \geqslant \frac{2}{d}, \tag{49}$$

or, in our notation, $x_T \geqslant 1$. Though the proof of Ref. [19] refers to the divergence of the correlation length as the critical point is approached, and hence strictly speaking does not apply to the system under consideration since we work exactly *at* the critical point, the RBPM is among the simplest physical systems for which Eq. (49) is believed to be valid [20]. The point is strengthened by noting that the $(q-2)$-expansion yields $x_T = 1.02 + \mathcal{O}(\epsilon^3)$ at $q = 3$ [15]. It is therefore difficult to have confidence in the cumulant expansion for the thermal exponent, and independent methods of assessing x_T must be devised.

At this point we note that although the RG equation (11) seems to warrant an effective exponent of $x_T^{\text{eff}} = 1 - Aw^2$ for q large, this argument is only superficially true. Indeed, near $q = \infty$ the RG flows must extend to infinite w before reaching the random FP, and consequently an expansion valid for weak randomness is not to be trusted.

The alternative method for finding x_T that comes closest to the spirit of Refs. [19,20] is that of finite-size scaling off the critical point. This is discussed at length in the next subsection, and for the moment we concentrate on less 'obvious' possibilities.

One of the key points in the construction of the cumulant expansion was the realisation that the spin–spin correlation function was mapped onto a surface tension under duality, and hence could be expressed in terms of the *largest* Lyapunov exponent of a TM with twisted boundary conditions. Reinterpreting the latter as a free energy the self-averaging property was evident, and the cumulant expansion correspondingly behaved quite well if the fluctuations were not too large. It has recently been shown [39,40] that under duality four-point correlation functions are similarly mapped onto (generalised) surface tensions. Presently these duality relations have only been worked out for planar graphs, but there is some hope that they may be extended to the case of cylindrical boundary conditions as well. Taking two of the points as nearest neighbours on either end of the cylinder we would then recover the energy–energy correlator, and if the corresponding boundary conditions can be implemented in the TM x_T follows from a cumulant expansion.

Next, we discuss the method of iterating orthogonal vectors in order to extract the second Lyapunov exponent [31] in more physical terms. The energy–energy correlator (Green's function) can be written as

$$\langle E(r_1)E(r_2)\rangle = \frac{\operatorname{Tr} E(r_1)E(r_2)\exp(-\mathcal{H})}{\operatorname{Tr}\exp(-\mathcal{H})}. \tag{50}$$

Now imagine building up the strip by repeated action by the random TMs on some initial state situated at $r = -\infty$. When we reach r_1 the system is in a state $|a_0\rangle$ on which we act with the energy operator to define $|b_0\rangle = E(r_1)|a_0\rangle$. After n further iterations these states become

$$\begin{aligned} |a_n\rangle &= \mathcal{T}_n \ldots \mathcal{T}_2\mathcal{T}_1|a_0\rangle, \\ |b_n\rangle &= \mathcal{T}_n \ldots \mathcal{T}_2\mathcal{T}_1|b_0\rangle. \end{aligned} \tag{51}$$

Defining a new state $|\tilde{b}_n\rangle$ by orthogonalising $|b_n\rangle$ with respect to $|a_n\rangle$

$$|\tilde{b}_n\rangle = |b_n\rangle - \frac{\langle a_n|b_n\rangle}{\langle a_n|a_n\rangle}|a_n\rangle \tag{52}$$

we find that

$$\begin{aligned} \frac{\langle \tilde{b}_n|\tilde{b}_n\rangle}{\langle a_n|a_n\rangle} &= \frac{\langle b_n|b_n\rangle}{\langle a_n|a_n\rangle} - \frac{\langle a_n|b_n\rangle\langle b_n|a_n\rangle}{\langle a_n|a_n\rangle^2} \\ &= \langle E(r_1)E(r_2)\rangle - \langle E(r_1)\rangle\langle E(r_2)\rangle. \end{aligned} \tag{53}$$

Thus the process of orthogonalisation corresponds precisely to subtracting off the disconnected part of the correlation function.

When $n \gg 1$ the states $|b_n\rangle$ and $|a_n\rangle$ are almost identical due to contamination and have a huge norm $\sim \Lambda_0^n$. The idea of orthogonalising them is therefore numerically extremely unsound. Fortunately a simple calculation shows that orthogonalising after n_1 iterations and then again after $n - n_1$ further iterations is equivalent to orthogonalising only once, as above. Hence, by induction, we are allowed to orthogonalise after each

iteration, leaving us with the method of Benettin et al. [31]. Similar observations hold true for the higher Lyapunov spectrum.

At this point an objection may be raised. Since

$$\langle a_n|a_n\rangle = \langle a_0|\mathcal{T}_1^\dagger \mathcal{T}_2^\dagger \dots \mathcal{T}_n^\dagger \mathcal{T}_n \dots \mathcal{T}_2 \mathcal{T}_1|a_0\rangle, \tag{54}$$

where the dagger denotes transposition, the correlation function (53) corresponds to a realisation of the randomness that is always symmetric around the midpoint of r_1 and r_2. From the above physical picture leading to Eq. (53) it seems that what we really ought to compute is

$$\frac{\langle \tilde{b}'_n|\tilde{b}_n\rangle}{\langle a'_n|a_n\rangle}, \tag{55}$$

where the (transposed) TMs used to obtain the states on the left implement a different realisation of the randomness than that used to obtain the states on the right.

Numerically we are now facing the problem of computing the average of huge numbers that are no longer necessarily positive. As discussed in connection with Eq. (43) we do not expect the correlation function to be self-averaging, and because of possible negative values of Eq. (55) the subterfuge of averaging its logarithm will not help us out. Trial runs seem to indicate that for sufficiently small values of q and R (such as $q = 3$ and $R = 2$) the matrix elements appearing in Eq. (55) computed for the usual samples of length 200 may have either sign, but that their quotient is invariably positive. The corresponding result for x_T is roughly equal to that obtained from the cumulant expansion. Unfortunately, for larger values of q or R rare events of negative quotients begin to occur, and any attempt of averaging Eq. (55) without resorting to logarithms is hampered by such large fluctuations as to render the results insufficiently accurate at the very best. Computations along these lines, though physically appealing, must therefore be abandoned on numerical grounds.

Yet another possibility of determining at least an approximate value of x_T is through the conformal sum rule [53] that for an n-fold replicated system reads

$$\frac{c(n)}{12} = \frac{\sum_i d_i(n) x_i e^{-2\pi x_i}}{1 + \sum_i d_i(n) e^{-2\pi x_i}}, \tag{56}$$

where the sum runs over all operators in the theory, including the descendants of the Verma module with their appropriate multiplicities, and $d_i(n)$ are the multiplicities pertaining to the permutational symmetry of the n replicas and the q Potts states. For the magnetic operator $d_i(n) = n(q-1)$, since there are $(q-1)$ independent order parameters, and in the case of the energy operator $d_i(n) = n$. In the pure system this yields quite accurate estimates for x_T if the exact values of $c(1)$ and x_H are inserted along with the first descendant of the latter. Differentiating and going to the replica limit we find that for a random system

$$\frac{c'(0)}{12} = x_1(q-1)e^{-2\pi x_1} - \frac{x_2}{2}(q-1)^2 e^{-2\pi x_2} + \dots, \tag{57}$$

so that for values of x_1, x_2 and x_T near those of the Ising model the term with x_T enters only as a small correction. Consequently, at the very best only x_2 can be determined with some confidence from our previous results for $c'(0)$ and x_1. Its value appears to be consistent with that obtained from the cumulant expansion.

Finally we should like to mention that preliminary studies of exact partition function zeros for small $L \times L$ lattices with quenched bond randomness hints at an interesting new method of estimating x_T. Although the different realisations of the quenched bond randomness in general lead to a considerable scatter in the positions of such zeros, it turns out that the zeros that are closest to the real axis only exhibit a very weak dependence on the realisation. But these zeros are precisely those that fix x_T through their finite-size scaling. Results along these lines, both for zeros of the Lee–Yang and of the Fisher type, will be published elsewhere [54].

4.6. *Phenomenological renormalisation*

In view of the difficulties encountered in our attempts to extract x_T directly at the critical point we turn our attention to the method of *phenomenological renormalisation* [21], which is closer in spirit to the ideas that lead to the bound (49).

The magnetic correlation length can be found from the TM spectra through

$$\xi(L,T)^{-1} = \ln\left(\frac{\Lambda_0^{11}}{\Lambda_0^{22}}\right) = L(f_0^{22} - f_0^{11}), \tag{58}$$

and we note that this quantity would be self-averaging in the random case. Motivated by the form $\xi \sim (T - T_c)^{-\nu}$ of the divergence of the correlation length in the infinite system we make the finite-size scaling ansatz

$$\xi(L,t) = L\phi\big((T - T_c)L^{1/\nu}\big). \tag{59}$$

For pure systems, then, one traditionally scans through the vicinity of T_c to find an effective $T_c(L)$ as the solution of

$$\frac{\xi(L,T_c(L))}{L} = \frac{\xi(L-1,T_c(L))}{L-1}, \tag{60}$$

and computes an approximant $\nu(L)$ from

$$1 + \frac{1}{\nu(L)} = \left.\frac{\ln\big(\mu(L,T)/\mu(L-1,T)\big)}{\ln\big(L/(L-1)\big)}\right|_{T=T_c(L)}, \tag{61}$$

where the derivatives

$$\mu(L,T) \equiv \frac{\partial\xi(L,T)}{\partial T} = L^{1+\frac{1}{\nu}}\phi'\big((T - T_c)L^{1/\nu}\big) \tag{62}$$

are found by numerical differentiation. As $L \to \infty$ we have $T_c(L) \to T_c$ and $\nu(L) \to \nu$.

In the random case the extracted values of $\xi(L,T)$ are hampered by substantial error bars, and the method just outlined becomes by far too inefficient. Fortunately the very

Table 7
Phenomenological renormalisation for the thermal scaling dimension $x_T = 2 - 1/\nu$ at $q = 8$ and $R = 10$. For each strip width L the 100 independent strips of length $m = 10^5$ are divided into patches of length 200. Within each patch the *exact* $\mu(L, T_c)$ is computed, based on evaluations of $\xi(L, K_1', K_2')$ at $K_1' = K_1(1 \pm \epsilon_K)$ and $K_2' = RK_1'$, where K_1 is found from Eq. (3). Final results and error bars are obtained as the mean value and the standard deviation over the totality of patches.

L	$\mu(L)$	$\nu(L)$
1	1.087 (1)	–
2	4.229 (3)	1.041 (1)
3	10.426 (8)	0.816 (2)
4	19.682 (18)	0.827 (3)
5	31.867 (33)	0.863 (5)
6	46.994 (53)	0.885 (7)
7	65.020 (79)	0.904 (9)

costly idea of scanning for $T_c(L)$ can be discarded, since the exact T_c of the RBPM is known from Eq. (3). Consequently the derivatives (62) and the approximants (61) are evaluated at the exact T_c, whence the only remaining source of errors is that of statistical fluctuations over the different realisations of the randomness.

Naïvely one would now find the derivative (62) by subtraction of the free energies evaluated at $T = T_c(1 \pm \epsilon)$, where $\epsilon \ll 1$. Although this method yields reasonable results for $\epsilon \sim 10^{-2}$ it is way too inaccurate, since it involves the subtraction of almost identical quantities (with error bars). A superior strategy is to divide the strip into patches of length 200, calculate *exact*[2] values of $\mu(L)$ for each of those, and finally average over the totality of such patches. In this way one exploits the fact that $\xi\big(L, T_c(1-\epsilon)\big)$ and $\xi\big(L, T_c(1+\epsilon)\big)$ are strongly correlated when the realisation of the randomness is kept fixed. In practice we found that this trick leads to a reduction of the error bars with a factor ~ 120.

Sample results obtained by using these prescriptions are shown in Table 7. It is seen that although the convergence is still too slow for reliable extrapolations to the limit of an infinite system to be made, the conflict with the bound (49) appears to be resolved.

We have found that the convergence of the estimators $\nu(L)$ can be significantly sped up if one performs the numerical differentiation (62) by going perpendicularly off the self-duality criterion in (K_1, K_2) space instead of maintaining the condition $K_2 = RK_1$. Indeed, one may imagine that there is another exponent associated with motion *along* the critical surface, and maintaining $K_2 = RK_1$ one would then measure an admixture of this spurious exponent, in particular for large R. A simple calculation using Eq. (3) shows that one should then evaluate $\xi(L, K_1', K_2')$ at

$$K_1' = K_1\left(1 \pm \epsilon \frac{R\mathrm{e}^{K_1}}{q\mathrm{e}^{K_2}}(\mathrm{e}^{K_2} - 1)^2\right),$$

[2] Since we are now faced with differentiating a quantity that is known with full machine precision (10^{-16}) we can concentrate on minimising the rounding and truncation errors. This is accomplished by choosing $\epsilon = 10^{-5}$. See Section 5.7 of Ref. [58].

Table 8
Phenomenological renormalisation going perpendicularly off the critical surface for $q = 8$ and $R = 6$ and 10, respectively. The data collection was done as before.

	L	$\mu(L)$	$\nu(L)$
$R = 6$	2	1.898 (1)	–
	3	4.456 (2)	0.905 (1)
	4	8.172 (6)	0.902 (2)
	5	12.974 (11)	0.933 (4)
	6	18.883 (18)	0.945 (6)
	7	25.891 (27)	0.955 (8)
$R = 10$	2	1.832 (1)	–
	3	3.917 (2)	1.144 (2)
	4	6.815 (5)	1.081 (4)
	5	10.486 (9)	1.074 (6)
	6	14.948 (15)	1.059 (8)
	7	20.198 (22)	1.050 (11)

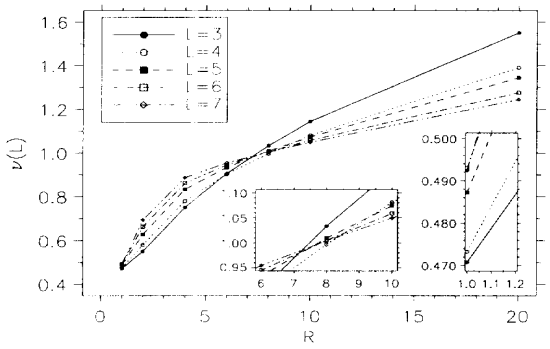


Fig. 6. Estimants $\nu(L)$ for the thermal exponent at $q = 8$ as obtained from phenomenological renormalisation applied to strips of width L and $L - 1$. In the pure system ($R \to 1$, see right-most inset) the estimants converge to $\nu(\infty) = \frac{1}{2}$. Curves for neighbouring system sizes intersect at values of ν and R that converge to those at the random fixed point as $L \to \infty$. In this case $\nu = 1.01 \pm 0.02$ and $R^* \approx 9$ (see left inset). Error bars are no larger than the size of the symbols.

$$K_2' = K_2(1 \pm \epsilon). \tag{63}$$

The sample results shown in Table 8 exhibit a conspicuous improvement over those of Table 7. Not only is the convergence faster, but it is even seen that the estimators $\nu(L)$ form a monotonically increasing sequence for low values of R and a monotonically decreasing one for high R. The extrapolated ν is pinched between those two sequences and consequently quite accurately determined.

Plots of the estimators $\nu(L)$ for $3 \leqslant L \leqslant 7$ and several values of R are shown in Figs. 6 and 7 for $q = 8$ and $q = 64$, respectively. It is seen that the curves for $\nu(L)$ and $\nu(L-1)$ intersect at a unique value of R, seemingly converging quite fast to a definite value R^* as L increases. We interpret R^* as the randomness strength at the critical

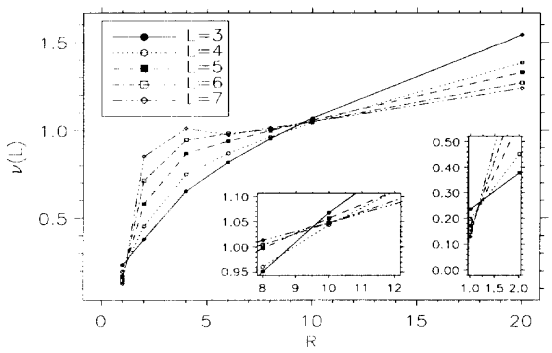


Fig. 7. Phenomenological renormalisation at $q = 64$. The curves intersect at larger angles than before, allowing for a rather accurate determination $\nu = 1.02 \pm 0.03$ in spite of the large fluctuations. Error bars are comparable to the size of the symbols. From the right-most inset it is seen that $\nu \to 0$ in the pure systems, as explained in the text. The left inset is a magnification of the region around $R^* \approx 10$.

Table 9
Values of the critical exponent ν and the randomness strength R^* at the random fixed point as obtained from phenomenological renormalisation.

q	ν	R^*
2	1.12 (3)	7 (1)
3	1.04 (4)	8 (1)
8	1.01 (2)	9 (1)
64	1.02 (3)	10 (1)

FP and the corresponding value of ν as the correct critical exponent. It is tempting to conjecture that the curves $\nu(L)$ approach ν on the entire interval $R \in]1, \infty[$ as $L \to \infty$. From the graphs it seems that the convergence is faster for large q.

The values of ν and R^* corresponding to this scenario are shown in Table 9. In accordance with the phase diagram (Fig. 1) R^* is a slowly, supposedly logarithmically,[3] increasing function of q. For $q = 2$ the deviation from $\nu = 1$ can be ascribed to logarithmic corrections [44], and for $q = 3$ our result for ν is in agreement with the $(q-2)$-expansion [15] though the possibility of replica symmetry breaking cannot be ruled out [22]. Also for $q > 4$ are our values for ν numerically consistent with unity, indicating that, unlike what is the case for the magnetic exponent, the thermal one displays only a weak q-dependence.

From Figs. 6 and 7 a remarkable feature about the pure system ($R = 1$) is apparent. For $q = 8$ the estimators $\nu(L)$ seem to converge to $\nu = \frac{1}{2}$ whilst for $q = 64$ the extrapolated value is $\nu \simeq 0$. The former value is hardly surprising since, as we also remarked above, a first-order transition is expected to exhibit scaling with trivial effective

[3] This supposition constitutes the simplest possibility allowed by the phase diagram of Fig. 1 in which $R^* \to \infty$ as $\ln q \to \infty$.

exponents (in this case: $x_T = 0$). On the other hand, $\nu \simeq 0$ for $q = 64$ has to do with the length scales of the system. Namely, from the asymptotic behaviour of the correlation length of the pure system [41]

$$\xi \sim \frac{2}{\ln q} \text{ as } q \to \infty, \tag{64}$$

we infer that $\xi \sim 1 \ll L$ at the transition point of the $q = 64$ system. But away from the transition point we also expect $\xi = \mathcal{O}(1)$, since the lattice spacing (unity) is the least length scale in the system. After all there is a ferromagnetic interaction between nearest-neighbour spins. We thus conclude that ξ is roughly temperature independent in this case. In order for this to be consistent with the asymptotic behaviour of the finite-size scaling ansatz (59)

$$\phi(x) \sim x^{-\nu} \quad \text{for } x \ll 1, \tag{65}$$

we must then have $\nu \simeq 0$. This is to be contrasted to the case of $q = 8$ where $\xi \gg L$ so that we clearly "see" the phase transitions in our strips of width L.

Very recently the bound (49) was challenged by Pázmándi et al. [24]. These authors claimed that the standard method of averaging over the disorder in finite-size (FS) systems introduces a new diverging length scale into the problem, whence the resulting ν_{FS} may be unrelated to the true exponent ν governing the divergence of the correlation length in the infinite system. In particular ν can be less than $2/d$, and if this is the case the standard method is liable to yield exactly $\nu_{\mathrm{FS}} = 2/d$. Ref. [24] then went on to suggest a noise reduction procedure that professedly would allow one to access the true ν. For each realisation of the binary randomness (2) used in the disorder average there will be a fraction p of weak bonds (K_1). The noise due to the fluctuations of p around its average value $\overline{p} = \frac{1}{2}$ can then be reduced by adjusting the couplings (K_1, K_2) for that particular realisation to the values they would assume at the critical point of an infinite system with a (mean) fraction of weak bonds equal to p.

To implement this we are faced with the task of finding the two-dimensional critical surface in the space (K_1, K_2, p) using our knowledge of its one-dimensional intersection with the plane $p = \frac{1}{2}$, viz. Eq. (3). Let the fraction of weak bonds in a particular realisation be

$$p = \tfrac{1}{2}(1 + \epsilon_p), \tag{66}$$

where $\epsilon_p \ll 1$. The symmetry $p \leftrightarrow 1 - p$ ensures that, to first order in ϵ_p, we must still go perpendicularly off the self-dual curve in the (K_1, K_2) subspace, as in Eq. (63). Since an increase in the number of weak bonds must be offset by an increase of the K's in order to keep the coupling to the energy density constant, the correct prescription is

$$\begin{aligned} K_1 &\to K_1 \left(1 + \epsilon_K \frac{R\mathrm{e}^{K_1}}{q\mathrm{e}^{K_2}} (\mathrm{e}^{K_2} - 1)^2\right), \\ K_2 &\to K_2 (1 + \epsilon_K) \end{aligned} \tag{67}$$

for some $\epsilon_K > 0$. Demanding that the combined change in p and in (K_1, K_2) must leave the coupling to the energy density invariant furnishes a relation between ϵ_p and ϵ_K

$$\epsilon_K = \epsilon_p \frac{K_2 - K_1}{K_1 (Re_1^K / q e^{K_2})(e^{K_2} - 1)^2 + K_2}, \tag{68}$$

and the derivatives (62) are now evaluated at these values of the parameters by going perpendicularly off the critical surface. To first order, of course, Eq. (63) still gives the correct way of doing so.

Our confidence in the results of Table 9 is increased by observing that the implementation of this novel averaging procedure does *not* alter our results. Indeed, trial runs for $q = 8$, where the discrepancy between the x_T extracted from the Lyapunov spectrum and phenomenological RG respectively is large, render the values of the estimators $\mu(L)$ unchanged within (small) error bars. It is thus concluded that even though our results for ν are conspicuously close to satisfying the bound (49) with equality, this is not due to an artifact in the averaging procedure.

4.7. *The higher Lyapunov spectrum*

Although the second Lyapunov exponent of $\mathcal{T}^{11}$ fails to yield the thermal scaling dimension x_T in the standard way it is hard to believe that the Lyapunov spectrum is not in some way related to the operator content of the CFT underlying the RBPM. In the case of the pure three-state Potts model, for example, it is well known [43] that the first five gaps of the Z_2-even sector are related to the energy density ϵ, its first descendants $L_{-1}\epsilon$ and $\overline{L}_{-1}\epsilon$, the stress tensor T and its conjugate $\overline{T}$. To wit, the scaling dimensions of these operators can be found from the gaps through Eq. (4), and we have verified this using our connectivity basis TMs.

In view of the bound (49) it is problematic to associate the first gap with the energy density in the random case, but it is nevertheless a beguiling question whether such concepts as descendants and the stress tensor are preserved by the randomness. To investigate this issue we have computed the first few Lyapunov exponents of $\mathcal{T}^{11}$ for $1 \leqslant L \leqslant 8$, averaging over 100 runs as usual. The scaling dimensions corresponding to the first five gaps for $q = 3, 4, 5$ and $R = 2$ are shown in Table 10. Self-averaging was ensured by utilising the cumulant expansion, and parabolic least-squares fits included the first three cumulants.

It is quite remarkable that even if the scaling dimension corresponding to the first gap may not be equal to x_T our data give strong reasons to believe that it has a descendant, and that this descendant has the expected degeneracy of two. And even though the scaling dimensions in general depend on q those corresponding to the fourth and the fifth gaps are constant within error bars and very close to 2, as is expected for the stress tensor of a conformally invariant system [43]. Preliminary data for even higher Lyapunov exponents seem to hint at descendants at level two, but since we have found that in the pure system higher and higher eigenvalues require larger and larger system

Table 10
Scaling dimensions corresponding to the first five gaps in the Lyapunov spectrum of $\mathcal{T}^{11}$ for $R = 2$. The parabolic least-squares fits included the first three cumulants of the probability distribution, and error bars were extracted based on the fits with $L_0 = 4, 5$ and 6.

q	1. gap	2. gap	3. gap	4. gap	5. gap
3	0.899 (4)	1.877 (13)	1.885 (12)	2.045 (24)	2.050 (23)
4	0.817 (5)	1.811 (9)	1.818 (8)	2.043 (23)	2.049 (24)
5	0.754 (6)	1.771 (6)	1.779 (6)	2.058 (24)	2.065 (25)

sizes before the asymptotic scaling form (4) is valid, massive computations are needed to establish reliable results for all but the first few scaling dimensions.

Another interesting feature of our data for the higher Lyapunov exponents is that the Harris criterion seems to be valid in a very complete sense. Namely, trial runs for $q = 1$ seem to indicate that although individual cumulants exhibit a pronounced dependence of R, their sum is virtually independent of the strength of the randomness in the whole range $R \in [1, 2]$. It thus appears that all exponents x_i that we can extract numerically from the Lyapunov spectrum, using Eq. (4) and the cumulant expansion, obey the Harris criterion. Since the connection between these exponents and the scaling dimensions of the underlying CFT is not completely known (witness x_T) this may well turn out to be a non-trivial observation.

5. Discussion and outlook

In a recent paper by Picco [47] it has been suggested that for $q = 2^p$ the effective central charge at the random FP is $c' = p/2$, and that this class of models thus behaves as p decoupled Ising models. Even without referring to our values of the magnetic exponent we should like to point out that all the data show is that $c' \propto \ln q$ with a constant of proportionality that is very close to $1/(2\ln 2) \simeq 0.721$. But this constant is *also* very close to that of the percolation point, viz. $5\sqrt{3}/4\pi \simeq 0.689$. Indeed, these two numbers differ by less than 5%, and since our error bars and those of Picco are in the 2% and the 4% range respectively, there is no irrefutable way of numerically distinguishing between percolative, Ising-like or indeed some other, presently unknown, behaviour of the central charge. A similar observation is valid for $2 \leqslant q \leqslant 4$ where our numerical data as displayed in Fig. 3 are compatible, within error bars, with both the values at the pure and the random FP (but not, in this case, with those at the percolation point).

On the other hand, our results for the magnetic exponent should leave no doubt that the correct CFT describing the RBPM cannot be that of a number of decoupled Ising models. In particular, the non-Ising value at $q = 8$ is in sharp disagreement with the Monte Carlo results of Ref. [11]. One possible explanation of this discrepancy is that these authors define a non-standard order parameter through

$$\rho = L'^{-d}\langle \max(N_1, N_2, \ldots, N_q)\rangle, \tag{69}$$

where N_i is the number of spins in state i, which is related to our local order parameter defined in Eq. (26) by $N_i = \sum_r (M_i(r) + q^{-1})$. The site label r runs over a hypercube of side L' with $24 \leqslant L' \leqslant 84$. Note that ρ may also be written as

$$\rho = L'^{-d} \lim_{k\to\infty} \left(\sum_{i=1}^{q} \langle N_i^k \rangle \right)^{1/k}. \tag{70}$$

Expressed in terms of the local order parameter, $\langle N_i^k \rangle$ gives a sum of terms each of the form

$$\sum_{r_j} \langle M_i(r_1)^{k_1} M_i(r_2)^{k_2} \ldots M_i(r_n)^{k_n} \rangle, \tag{71}$$

where $k_1 + k_2 + \ldots = k$. As $k \to \infty$ at fixed L', it is clear that at least some of the k_j must grow large. In the pure system, this should not matter, since each term will scale in the same manner. But when multiscaling is present, the scaling behaviour of the k_j power of the local order parameter may be different. Indeed, in the limit of $k \to \infty$ one would expect ρ to scale with dimension $\lim_{k\to\infty} x_k/k$, which is *less* than x_1 by convexity.

Another criticism of Ref. [11] is that the realisations of the binary randomness considered were confined to those for which the number of strong and weak bonds in each of the two lattice directions were equal. Though this restriction is clearly innocuous in the limit $L' \to \infty$ this may not be so as far as the finite-size scaling is concerned. From trial runs where similar restrictions were imposed to the bond distributions of the TMs we found that seemingly harmless noise reductions schemes can influence the output substantially.

Finally, the mapping to the RFIM [23] illustrated that for large q typical configurations consist of large clusters of spins in the same Potts state, separated by domain walls. Whilst our very long strips are guaranteed to accommodate large regions in which all q values of the order parameters are realised, it is not clear that this should be the case in the much smaller square geometries of Ref. [11]. Indeed it seems likely that one would find Ising exponents if the geometry under consideration typically can accommodate at most two different large clusters.

We now turn our attention to the thermal exponent. If the phenomenological RG scheme is to be trusted the values of x_T only exhibit a weak dependence on q, although the $(q-2)$-expansion gives us reason to believe that there is some variation [15]. It is interesting that x_T stays so close to unity even at very high q, but presently we do not have any arguments to explain this finding. Unfortunately the method employed was unable to resolve the slight deviations from unity, and it is indeed a challenge to future research to find more accurate ways of assessing x_T for disordered systems. Our results on the higher Lyapunov spectrum are nothing if not intriguing, and we believe that a great effort must be made to understand why the first gap in the spectrum fails to be

related to x_T in the standard way, even though the higher gaps show clear indications of a conformal field theory underlying the RBPM.

A very interesting issue to be addressed by future research is that of the dynamical universality class of the RBPM. In particular it would be interesting to see whether the dynamic critical exponent z does or does not agree with the Ising value of $z \approx 2$, or whether, in analogy with the RFIM, there is logarithmically slow dynamics due to the pinning of domain walls by impurities.

Other types of randomness are also of interest to the question whether a first-order phase transition is softened due to impurities. In this paper we have studied the effect of quenched bond randomness in a flat, regular lattice. A somewhat different scenario is obtained by investigating the pure q-state Potts model on lattices with quenched *connectivity* disorder. In Ref. [55] MC simulations of the $q = 8$ model on two-dimensional Poissonian random lattices (Voronoi tessellations) with toroidal topology unambiguously showed that the first-order nature of the transition was not modified.

An argument why this must be so, at least for large q, can readily be given. For simplicity consider the model on the dual Delaunay random lattice, which per construction is a triangulation of space [56]. As on the regular lattice, at large q there are only two important states in the equivalent random-cluster model: the empty lattice, which contributes a term $q^{N_{\text{vertices}}}$ to the partition function, and the full lattice, contributing a factor $u^{N_{\text{edges}}}$, where $u = e^K - 1$. Since for any triangulation $2N_{\text{edges}} = 3N_{\text{vertices}}$, the transition occurs when $u \sim q^{2/3}$ for *any* geometry. If we now consider that part of the lattice within a large hypercube of side L, the fluctuations in the difference of the energies of these two states inside this region will come solely from the edges which penetrate the boundary. On average, the difference in the energies of these two states will still be zero, but there will be fluctuations of the order of the square root of the number of bonds which penetrate the boundary, which will therefore be $\mathcal{O}(L^{(d-1)/2})$. These are very much smaller than the analogous fluctuations which are present when random bonds are added: these are $\mathcal{O}(L^{d/2})$, which leads to the Imry–Ma type of argument [3,5,6]. For $d = 2$ Voronoi tessellations the fluctuations are thus always smaller than the domain wall energy $\mathcal{O}(L^{d-1})$, and we conclude that such randomness is strongly irrelevant (at least for large q), in agreement with the results of Ref. [55].

Yet another kind of randomness is obtained by studying the Potts model on quenched random-gravity graphs, for which MC simulations for $q = 10$ have provided strong evidence for a softening scenario similar to ours [57]. However, in this case the *curvature* is random and hence when the lattice is embedded in the plane, it is fractal. Although the argument about compensation of the bulk energies when $u \sim q^{2/3}$ works for any triangulation, the number of boundary edges may well scale in a different manner for these lattices. Whilst it would be interesting to study this in detail, it is clear that this problem is quite different from ours, and neither our arguments nor those of Refs. [5,6] can be directly applied.

Acknowledgements

Stimulating discussions with J. Chalker, E. Domany, M. Kardar, D.P. Landau, A. Ludwig and S. de Queiroz are gratefully acknowledged. We also thank the staff at the Institute for Theoretical Physics at Santa Barbara, where part of this work was done, for its warm hospitality. This research was supported in part by the Engineering and Physical Sciences Research Council under Grant GR/J78327, and by the National Science Foundation under Grant PHY94-07194.

References

[1] A.B. Harris, J. Phys. C 7 (1974) 1671.
[2] A. Aharony, Phys. Rev. B 18 (1978) 3318.
[3] Y. Imry and S.-K. Ma, Phys. Rev. Lett. 35 (1975) 1399.
[4] Y. Imry and M. Wortis, Phys. Rev. B 19 (1979) 3581.
[5] K. Hui and A.N. Berker, Phys. Rev. Lett. 62 (1989) 2507; 63 (1989) 2433 (E).
[6] M. Aizenman and J. Wehr, Phys. Rev. Lett. 62 (1989) 2503.
[7] J. Cardy, J. Phys. A 29 (1996) 1897.
[8] J. Cardy, Scaling and Renormalization in Statistical Physics (Cambridge Univ. Press, Cambridge, 1996).
[9] P. Pujol, Europhys. Lett. 35 (1996) 283.
[10] R.J. Baxter, J. Phys. C 6 (1973) L445.
[11] S. Chen, A.M. Ferrenberg and D.P. Landau, Phys. Rev. E 52 (1995) 1377.
[12] S. Wiseman and E. Domany, Phys. Rev. E 51 (1995) 3074.
[13] M. Kardar, A.L. Stella, G. Sartoni and B. Derrida, Phys. Rev. E 52 (1995) R1269.
[14] A.W. W. Ludwig and J.L. Cardy, Nucl. Phys. B 285 [FS19] (1987) 687.
[15] A.W.W. Ludwig, Nucl. Phys. B 285 (1987) 97.
[16] A.W.W. Ludwig, Nucl. Phys. B 330 (1990) 639.
[17] Vl. Dotsenko, M. Picco and P. Pujol, Nucl. Phys. B 455 (1995) 701.
[18] H.W. J. Blöte and M.P. Nightingale, Physica A 112 (1982) 405.
[19] J.T. Chayes, L. Chayes, D.S. Fisher and T. Spencer, Phys. Rev. Lett. 57 (1986) 2999.
[20] J.T. Chayes, L. Chayes, D.S. Fisher and T. Spencer, Commun. Math. Phys. 120 (1989) 501.
[21] M.P. Nightingale, Physica A 83 (1976) 561.
[22] Vi. Dotsenko, Vl. Dotsenko, M. Picco and P. Pujol, Europhys. Lett. 32 (1996) 425.
[23] J. Cardy and J.L. Jacobsen, Phys. Rev. Lett. 79 (1997) 4063.
[24] F. Pázmándi, R.T. Scalettar and G.T. Zimányi, Phys. Rev. Lett. 79 (1997) 5130.
[25] R.B. Potts, Proc. Cambr. Philos. Soc. 48 (1952) 106.
[26] F.Y. Wu, Rev. Mod. Phys. 54 (1982) 235.
[27] W. Kinzel and E. Domany, Phys. Rev. B 23 (1981) 3421.
[28] J.L. Cardy, J. Phys. A 16 (1983) L355.
[29] H.W. Blöte, J.L. Cardy and M.P. Nightingale, Phys. Rev. Lett. 56 (1986) 742;
I. Affleck, Phys. Rev. Lett. 56 (1986) 746.
[30] H. Furstenberg, Trans. Am. Math. Soc. 68 (1963) 377.
[31] G. Benettin, L. Galgani, A. Giorgilli and J.-M. Strelcyn, Meccanica 15 (1980) 9.
[32] T. Bhattacharya, R. Lacaze and A. Morel, J. Phys. I France 7 (1997) 81.
[33] A.J. Bray and M.A. Moore, J. Phys. C 18 (1985) L927.
[34] D. Stauffer, Z. Phys. B 22 (1975) 161.
[35] J.L. Cardy (ed.), Finite-Size Scaling (North-Holland, Amsterdam, 1988).
[36] P.W. Kasteleyn and C.M. Fortuin, J. Phys. Soc. Jap. 46 (suppl.) (1969) 11.
[37] H.W. J. Blöte and B. Nienhuis, J. Phys. A 22 (1989) 1415.
[38] F.Y. Wu, Phys. Lett. A 228 (1997) 43.
[39] J.L. Jacobsen, Phys. Lett. A 233 (1997) 489.
[40] F.Y. Wu and H.Y. Huang, Phys. Rev Lett. 79 (1997) 4954.

[41] E. Buffenoir and S. Wallon, J. Phys. A 26 (1993) 3045.
[42] S.L. A. de Queiroz, Phys. Rev. E 51 (1995) 1030.
[43] J.L. Cardy, *in* Phase Transitions and Critical Phenomena Vol. 11, ed. C. Domb and J.L. Lebowitz (Academic Press, London, 1987).
[44] F.D. A. Aarão Reis, S.L. A. de Queiroz and R. R. dos Santos, Phys. Rev. B 56 (1997) 6013.
[45] J.L. Cardy, J. Phys. A 19 (1986) L1093.
[46] A.B. Zamolodchikov, Pis'ma Zh. Eksp. Teor. Fiz. 43 (1986) 565. [JETP Lett. 43 (1986) 730.]
[47] M. Picco, Phys. Rev. Lett. 79 (1997) 2998.
[48] VI.S. Dotsenko and V.A. Fateev, Nucl. Phys. B 240 (1984) 312.
[49] B. Derrida, Phys. Rep. 103 (1984) 29.
[50] Section 12.5.3 in Ledermann (ed.), Handbook of Applicable Mathematics, Vol. 2: Probability (Wiley, London, 1980).
[51] M. Picco, Phys. Rev. B 54 (1996) 14930.
[52] M. Picco, private communication.
[53] J.L. Cardy, Nucl. Phys. B 270 [FS16] (1986) 186; 366 (1991) 403.
[54] J.L. Jacobsen, in preparation.
[55] W. Janke and R. Villanova, Phys. Lett. A 209 (1995) 179; Nucl. Phys. B (Proc. Suppl.) 47 (1996) 641.
[56] A. Okabe, B. Boots and K. Sugihara, Spatial Tessellations: Concepts and Applications of Voronoi Diagrams (Wiley, London, 1992).
[57] C.F. Baillie, W. Janke and D.A. Johnston, Phys. Lett. B 388 (1996) 14.
[58] W.H. Press et al., Numerical Recipes in C, 2nd Ed. (Cambridge University Press, Cambridge, 1992).

7.5 Article "Large-q asymptotics of the random-bond Potts model"

Nous étudions numériquement le comportement asymptotique du modèle de Potts aléatoire, dans la limite d'un grand nombre d'états q. La paramétrisation de la ligne critique est soigneusement déterminée par une combinaison de la représentation de boucles de la matrice de transfert et le "théorème c" de Zamolodchikov. La charge centrale semble avoir le comportement asymptotique $c(q) = \frac{1}{2}\log_2(q) + O(1)$. Des valeurs très précises de l'exposant magnétique volumique x_1 sont fournies par des simulations du type Monte Carlo, directement au point critique. Quand $q \to \infty$, elles semblent atteindre une limite non triviale, $x_1 \to 0.192 \pm 0.002$.

Large-q asymptotics of the random-bond Potts model

Jesper Lykke Jacobsen[1,*] and Marco Picco[2]

[1]*Laboratoire de Physique Statistique,*[†]*Ecole Normale Supérieure, 24 rue Lhomond, F-75231 Paris Cedex 05, France*

[2]*LPTHE,* [‡]*Université Pierre et Marie Curie, Paris VI et Université Denis Diderot, Paris VII, Boîte 126, Tour 16, 4 place Jussieu, F-75252 Paris Cedex 05, France*

(Received 6 October 1999)

We numerically examine the large-q asymptotics of the q-state random bond Potts model. Special attention is paid to the parametrization of the critical line, which is determined by combining the loop representation of the transfer matrix with Zamolodchikov's c-theorem. Asymptotically the central charge seems to behave like $c(q)=\frac{1}{2}\log_2(q)+O(1)$. Very accurate values of the bulk magnetic exponent x_1 are then extracted by performing Monte Carlo simulations directly at the critical point. As $q\to\infty$, these seem to tend to a nontrivial limit, $x_1\to 0.192\pm 0.002$.

PACS number(s): 05.70.Jk, 64.60.Ak, 64.60.Fr

I. INTRODUCTION

Recently the two-dimensional q-state random bond Potts model with $q>4$ has attracted considerable interest, because it serves as a paradigm for examining the effect of quenched randomness [1] on a first-order phase transition [2]. Since in this case the randomness couples to the local energy density, a theorem by Aizenman and Wehr [3], along with related analytical work [4,5], suggests that the transition should become continuous, as has indeed been verified by subsequent numerical studies [6–13]. Unfortunately, analytical results have been scarce, except in the limit $q\to\infty$ where properties of a particular tricritical point were related to those of the zero-temperature fixed point of the random *field* Ising model in $d=2+\varepsilon$ dimensions [8]. From the conjectured phase diagram [8] it is, however, known that this fixed point is not the analytical continuation of the line of random fixed points found for finite $q>2$ [14,15]. Namely, the latter (henceforth referred to as the $q\to\infty$ limit of the model) is rather believed to be associated with a subtle percolationlike limit [8], the exact properties of which have not yet been fully elucidated.

In the present Rapid Communication we seek to gain further knowledge of this $q\to\infty$ limit by producing numerical results along the aforementioned line of critical fixed points for very large values of q. Since cross-over effects to the pure and percolative limits of the model have been shown to be important [9,10], special attention must be paid to the parametrization of the critical line. Generalizing a recently developed transfer matrix technique [16], in which the Potts model is treated through its loop representation [17], we were able to explicitly trace out this line, and as a by-product obtain very precise values of the central charge. Based on our numerical results for the $q=8^k$ state model with $k=1,2,\ldots,6$ we find compelling evidence that

$$c(q)=\tfrac{1}{2}\log_2(q)+O(1). \tag{1.1}$$

Although this behavior of the central charge is reminiscent of the Ising-like features of the tricritical fixed point discussed above, we shall soon see that from the point of view of the magnetic exponent the $q\to\infty$ limit is most definitely not in the Ising universality class. Note also that our precision allows us to convincingly distinguish the numerically computed central charge from its analytically known value in the percolation limit [9].

With the numerically obtained parametrization of the critical disorder strength at hand we then proceed to measure the corresponding magnetic bulk scaling dimension x_1 as a function of q. The most suitable technique here is that of conventional Monte Carlo simulations. Our results lend credibility to the belief [12] that $x_1(q)$ saturates as $q\to\infty$. Based on results for the $q=8^k$ state model with $k=1,2,3$ we propose the limiting value

$$x_1(q)\to 0.192\pm 0.002 \quad \text{for} \quad q\to\infty, \tag{1.2}$$

in agreement with the one reported in Ref. [12]. The fact that Eq. (1.2) does not coincide with any known scaling dimension of standard percolation is remarkable, and calls for further analytical investigations of the $q\to\infty$ limit.

After explaining the loop model transfer matrices in Sec. II, we state our results for the critical line and the central charge in Sec. III. The Monte Carlo method and the resulting values of the magnetic scaling dimension are presented in Sec. IV, and we conclude with a discussion.

II. LOOP MODEL TRANSFER MATRICES

The partition function of the random bond Potts model can be written as

$$Z=\sum_{\{\sigma\}}\prod_{\langle ij\rangle} e^{K_{ij}\delta_{\sigma_i,\sigma_j}}, \tag{2.1}$$

where the summation is over the q discrete values of each spin and the product runs over all nearest-neighbor bonds on the square lattice. The K_{ij} are the reduced coupling constants, which for the moment may be drawn from an arbi-

*Present address: LPTMS, bâtiment 100, Université Paris-Sud, F-91405 Orsay, France.

[†]Laboratoire associé aux universités Paris 6, Paris 7 et au CNRS.

[‡]Unité Mixte de Recherche CNRS UMR 7589.

trary distribution. By the standard Kasteleyn-Fortuin transformation [18], Eq. (2.1) can be recast as a random cluster model

$$Z=\sum_{\{\mathcal{G}\}} q^{C(\mathcal{G})} \prod_{\langle ij\rangle \in \mathcal{G}} (e^{K_{ij}}-1), \tag{2.2}$$

where $\mathcal{G}$ is a bond percolation graph with $C(\mathcal{G})$ independent clusters. Note that q now enters only as a (continuous) parameter, and since the nonlocality of the clusters does not obstruct the construction of a transfer matrix [19] the interesting regime of $q \gg 4$ becomes readily accessible, provided that one can take into account the randomness in the couplings [9].

In an analogous fashion we can adapt the even more efficient loop model representation [16] to the random case. Indeed, trading the clusters for their surrounding loops on the medial lattice [17], Eq. (2.2) is turned into

$$Z=q^{N/2}\sum_{\{\mathcal{G}\}} q^{L(\mathcal{G})/2} \prod_{\langle ij\rangle \in \mathcal{G}} \left(\frac{e^{K_{ij}}-1}{\sqrt{q}}\right), \tag{2.3}$$

where N is the total number of spins, and configuration $\mathcal{G}$ encompasses $L(\mathcal{G})$ loops. The strip width L is measured in terms of the number of ''dangling'' loop segments, and must be even by definition of the medial lattice [16].

A pleasant feature of the random bond Potts model is that the critical temperature is known exactly by self-duality [20]. Employing for simplicity the bimodal distribution

$$P(K_{ij})=\tfrac{1}{2}[\,\delta(K_{ij}-K_1)+\delta(K_{ij}-K_2)], \tag{2.4}$$

and choosing the parametrization $s_{ij}\equiv(e^{K_{ij}}-1)/\sqrt{q}$, the self-duality criterion takes the simple form

$$s_1 s_2 = 1. \tag{2.5}$$

To fully identify the critical point the only free parameter is then the strength of the disorder, which can be measured in terms of $R\equiv K_1/K_2>1$ or $s\equiv s_1>1$.

III. CENTRAL CHARGE

In Ref. [16] we showed that Zamolodchikov's c-theorem [21] is a powerful tool for numerically identifying the fixed points of a *pure* system. The idea is simple: From the leading eigenvalue of the transfer matrix, specific free energies $f_0(L)$ can be computed as a function of the strip width L. Effective central charges $c(L)$ are then obtained by fitting data for two consecutive strip widths according to [22]

$$f_0(L)=f_0(\infty)-\frac{\pi c}{6L^2}+\cdots. \tag{3.1}$$

By tuning the free parameter s of the system, local extrema $c(L,s_*(L))$ are sought for, and finally the fixed point is identified by extrapolation: $s_*=s_*(L\to\infty)$.

In principle this strategy can also be employed for a *disordered* system, provided that error bars are carefully kept under control. Now $f_0(L)$ is related to the largest Lyapunov exponent of a product of $M\to\infty$ random transfer matrices [23,8], and its statistical error vanishes as $M^{-1/2}$ by the central limit theorem. Thus, for large enough M any desired precision on $f_0(L)$ can be achieved.

TABLE I. Effective central charge of the $q=8$ state model, as a function of disorder strength s. Two- and three-point fits to Eq. (3.1) are labeled as $C(L,L+2)$ and $C(L,L+4)$, respectively.

s	$c(4,8)$	$c(6,10)$	$c(8,12)$	$c(4,6)$	$c(6,8)$	$c(8,10)$	$c(10,12)$
3	1.495	1.500	1.500	1.4101	1.4544	1.4731	1.4821
4	1.512	1.517	1.516	1.4157	1.4657	1.4868	1.4967
5	1.519	1.525	1.523	1.4152	1.4690	1.4918	1.5025
6	1.521	1.528	1.527	1.4116	1.4683	1.4927	1.5044
7	1.520	1.529	1.529	1.4067	1.4656	1.4915	1.5041
8	1.518	1.528	1.529	1.4013	1.4619	1.4890	1.5026
9	1.509	1.530	1.528	1.3972	1.4552	1.4860	1.5004
10	1.511	1.525	1.527	1.3908	1.4534	1.4826	1.4977
11	1.501	1.526	1.526	1.3873	1.4465	1.4791	1.4949
12	1.504	1.519	1.524	1.3816	1.4451	1.4756	1.4919

An important observation is that for larger and larger L, the $c(L)$ found from Eq. (3.1) become increasingly sensitive to errors in $f_0(L)$. Therefore, M must be chosen in accordance with the largest strip width $L_{\max}$ used in the simulations. For the system at hand we found that four significant digits in $c(L)$ were needed for a reasonable precise identification of $s_*(L)$, and with $L_{\max}=12$ this in turn implies that the $f_0(L)$ must be determined with six significant digits. We were thus led to choose $M=10^8$ for $q=8$, and $M=10^9$ for larger values of q. [Incidentally, improving our results to $L_{\max}=14$ would require augmenting M by at least a factor of 100 (apart from the increased size of the transfer matrices), and since several months of computations were spent on the present project this hardly seems possible in a foreseeable future.]

Data collection was done by dividing the strip into M/l patches of length $l=10^5$ lattice spacings, and for each patch the couplings were randomly generated from a *canonical* ensemble, i.e., the distribution (2.4) was restricted to produce an equal number of strong and weak bonds.

In the right part of Table I we show the resulting two-point fits (3.1) in the $q=8$ state model, as a function of s. The left part of the table provides analogous three-point fits, obtained by including a nonuniversal $1/L^4$ correction in Eq. (3.1). In all cases the error bars are believed to affect only the last digit reported. The two-point fits give clear evidence of a maximum in the central charge, and we estimate its location as $s_*=6.5\pm2.0$. The corresponding central charge is estimated from the three-point fits, as these are known to converge faster in the $L\to\infty$ limit [9], and we arrive at $c=1.530\pm0.001$. To appreciate the precision of this result, we mention that the numerical values of $c(q=8)$ first reported were 1.50 ± 0.05 [7] and 1.517 ± 0.025 [8].

Table II summarizes our results for other values of q. Two remarkable features are apparent. First, $s_*\propto q^w$ is well fitted by a power law with $w=0.31\pm0.02$. This gives valuable information on how the $q\to\infty$ limit of the model is approached, and implies that the ratio of the coupling constants $R\equiv K_1/K_2=\log(1+s\sqrt{q})/\log(1+\sqrt{q}/s)$ is a nonmonotonic function of q that tends to the *finite* limiting value $(1+2w)/(1-2w)=4.3\pm0.6$ as $q\to\infty$. We shall discuss this

TABLE II. Critical disorder strength s and central charge c, as functions of q.

q	s_*	c	$c/\log_2(q)$
8	6.5 (20)	1.530 (1)	0.5100 (3)
64	15.5 (20)	3.050 (3)	0.5083 (5)
512	32 (2)	4.545 (10)	0.5050 (11)
4096	65 (8)	6.038 (24)	0.5032 (20)
32768	135 (20)	7.54 (3)	0.5027 (20)
262144	250 (50)	9.04 (3)	0.502 (2)

finding further in Sec. V. Second, the central charge seems to fulfill the relation (1.1) as stated in the Introduction.

IV. MAGNETIC SCALING DIMENSION

In this section we explain the Monte Carlo method used for obtaining values of the magnetic scaling exponent. Simulations were performed on square lattices of size $L\times L$ with periodic boundary conditions, with L ranging from 4 to $L_{\max}=128$ for $q=8$, 64 and $L_{\max}=64$ for $q=512$.

We employed the Wolff cluster algorithm [24]. The first part of the simulations was to determine the autocorrelation times τ, which were found to increase with the lattice size and also with q. For the largest simulated lattices, we determined τ as follows: 88 ± 4 cluster updates for $q=8$ and $L=128$, 3000 ± 215 for $q=64$ and $L=128$, and $31\,000\pm3000$ for $q=512$ and $L=64$. This rapid increase of τ with q explains why we simulate only up to $L=64$ for the largest q.

Next, we measure the magnetization, defined for each disorder sample x by

$$m_x=\frac{q\langle\rho\rangle-1}{q-1}, \tag{4.1}$$

where $\rho=\max(N_1,N_2,\ldots,N_q)/L^2$ and N_σ is the number of Potts spins taking the value σ. Here $\langle\ldots\rangle$ denotes the thermal average. Then the magnetization $m(L)$ is obtained by averaging over 10^5 disorder configurations for $q=8$, and 10^4 configurations for $q=64$ and 512. For each disorder sample, 100τ updates were dedicated to the thermalization, and a further 100τ to the magnetization measurement. Error bars were computed from the disorder fluctuations (it can easily be checked [10] that the contribution from *thermal* fluctuations is negligible), and the strength of the disorder was chosen as indicated in Table II.

From a fit to $m(L)\simeq L^{-x_1}$, we obtain for the magnetic scaling dimension

$$x_1=\begin{cases}0.1535(10) & \text{for} \quad q=8\\ 0.172(2) & \text{for} \quad q=64\\ 0.180(3) & \text{for} \quad q=512.\end{cases} \tag{4.2}$$

We see that the magnetic exponent seems to saturate as we increase q. In view of the result (1.1) for the central charge we expect the asymptotic behavior should involve $\log(q)$ rather than q itself, and indeed the data are well fitted by

$$x_1(q)=a+b/\log(q), \tag{4.3}$$

with $a=0.192(2)$ and $b=-0.080(4)$. Thus, based on the form (4.3) we are led to propose the limiting value (1.2) of x_1 given in the Introduction.

V. DISCUSSION

It is useful to juxtapose our findings on the large-q behavior of the critical line with the phase diagram proposed in Ref. [8]. In that work the disorder strength was parametrized through $s=q^w$ with $w>0$, and the limit $w\to\infty$ was identified with classical percolation on top of the strong bonds. Actually it is easily seen from Eq. (2.2) that directly at $q=\infty$ this percolation scenario holds true whenever $w>\frac{1}{2}$, and assuming that the line of critical fixed points is described by a monotonic function $w_*(q)$ it can thus be confined to the region $w\leqslant\frac{1}{2}$. With this slight reinterpretation, Ref. [8] argues that at $q=\infty$ the critical point is located in the limit $w\to\frac{1}{2}$. Indeed, since for $q=\infty$ any initial $w\ll\frac{1}{2}$ will be driven to larger values due to mapping to the random field Ising model, this is nothing but the usual assumption of ''no intervening fixed points.''

However, this seems at odds with the results of Table II, where we found that for $q\gg4$ the critical line, when measured in terms of w, saturates at $w=0.31\pm0.02$. Unless our numerical method is flawed by some gross systematic error, it is thus *a priori* difficult to see how this can be reconciled with the above result of $w_*(q=\infty)=\frac{1}{2}$. A possible explanation is that the limits $q\to\infty$ and $w\to\frac{1}{2}$ are highly noncommuting. This is witnessed by the jump in the central charge, which in the percolation limit ($w=\infty$ and $q<\infty$) reads [8]

$$c_{\text{perc}}=\frac{5\sqrt{3}}{4\pi}\ln(q)\simeq0.477\,69\log_2(q), \tag{5.1}$$

to be contrasted with our numerical result (1.1).

ACKNOWLEDGMENT

We are grateful to J. Cardy for some very helpful comments.

[1] J. Cardy, Physica A **263**, 215 (1999).
[2] R. J. Baxter, J. Phys. C **6**, L445 (1973).
[3] M. Aizenman and J. Wehr, Phys. Rev. Lett. **62**, 2503 (1989).
[4] Y. Imry and M. Wortis, Phys. Rev. B **19**, 3580 (1979).
[5] K. Hui and A. N. Berker, Phys. Rev. Lett. **62**, 2507 (1989); **63**, E2433 (1989).
[6] S. Chen, A. M. Ferrenberg, and D. P. Landau, Phys. Rev. E **52**, 1377 (1995).
[7] M. Picco, Phys. Rev. Lett. **79**, 2998 (1997).
[8] J. Cardy and J. L. Jacobsen, Phys. Rev. Lett. **79**, 4063 (1997).
[9] J. L. Jacobsen and J. Cardy, Nucl. Phys. B **515** [FS], 701 (1998).
[10] M. Picco, Phys. Rev. E (to be published).
[11] C. Chatelain and B. Berche, Phys. Rev. Lett. **80**, 1670 (1998); Phys. Rev. E **58**, R6899 (1998); **60**, 3853 (1999).
[12] T. Olson and A. P. Young, Phys. Rev. B **60**, 3428 (1999).
[13] G. Palágyi, C. Chatelain, B. Berche, and F. Iglói, e-print cond-mat/9906067.

[14] A. W. W. Ludwig and J. L. Cardy, Nucl. Phys. B **285** [FS19], 687 (1987).
[15] A. W. W. Ludwig, Nucl. Phys. B **285**, 97 (1987); **330**, 639 (1990).
[16] Vl. S. Dotsenko, J. L. Jacobsen, M.-A. Lewis, and M. Picco, Nucl. Phys. B **546** [FS], 505 (1999).
[17] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, London, 1982).
[18] P. W. Kasteleyn and C. M. Fortuin, J. Phys. Soc. Jpn. **46** (suppl.), 11 (1969).
[19] H. W. J. Blöte and M. P. Nightingale, Physica A **112**, 405 (1982).
[20] W. Kinzel and E. Domany, Phys. Rev. B **23**, 3421 (1981).
[21] A. B. Zamolodchikov, Pis'ma Zh. Eksp. Teor. Fiz. **43**, 565 (1986) [JETP Lett. **43**, 730 (1986)].
[22] H. W. Blöte, J. L. Cardy, and M. P. Nightingale, Phys. Rev. Lett. **56**, 742 (1986); I. Affleck, *ibid.* **56**, 746 (1986).
[23] H. Furstenberg, Trans. Am. Math. Soc. **68**, 377 (1963).
[24] U. Wolff, Phys. Rev. Lett. **60**, 1461 (1988).

7.6 Article "Multiscaling of energy correlations"

Nous calculons numériquement les exposants qui gouvernent la décroissance du corrélateur énergie-énergie du modèle de Potts aléatoire à q états. Deux mesures différentes sont considérées : la moyenne sur le désordre et un échantillon fixe. Nous résultats sont en bon accord avec le développement à deux boucles autour de $q = 2$, qui a été trouvé récemment au moyen du groupe de renormalisation perturbatif. Ils satisfont la borne sur la longueur de corrélation, $\nu \geq 2/d$, et amplifient l'évidence contre une brisure de la symétrie de répliques dans cette classe de modèles.

PHYSICAL REVIEW E VOLUME 61, NUMBER 6 JUNE 2000

Multiscaling of energy correlations in the random-bond Potts model

Jesper Lykke Jacobsen
LPTMS, Bâtiment 100, Université Paris–Sud, F-91405 Orsay, France
(Received 16 December 1999)

We numerically calculate the exponents governing the disorder averaged and fixed-sample decay of the energy-energy correlator in the q-state random-bond Potts model. Our results are in good agreement with a two-loop expansion around $q=2$ recently found from perturbative renormalization group techniques, fulfill the correlation length bound $\nu \geqslant 2/d$, and give further evidence against replica symmetry breaking in this class of models.

PACS number(s): 05.70.Jk, 64.60.Ak, 64.60.Fr

I. INTRODUCTION

The q-state random-bond Potts model is an interesting framework for examining how a phase transition is modified by quenched disorder coupling to the local energy density. For $q>2$ such randomness acts as a relevant perturbation [1], and for $q>4$ it even changes the nature of the transition from first to second order (see Ref. [2] for a review). In the regime where $(q-2)$ is small, a score of analytical results have been obtained from the perturbative renormalization group, and the various expansions for the central charge and the multiscaling exponents for the moments of the spin-spin correlator compare convincingly to recent numerical work [3].

A particularly useful way of carrying out these simulations is to consider the finite-size scaling of the Lyapunov spectrum of the (random) transfer matrix, thus generalizing the method commonly applied to the eigenvalue spectrum in a pure system [4]. A definite advantage over the more traditional technique of Monte Carlo simulations [5] is that the transfer matrices allow for a representation in which q can be regarded as a continuously varying parameter [6,4], and in particular one can study small non-integer values of $(q-2)$.

The outcome of applying this method to the *energetic* sector of the transfer matrix, however, led to contradictory results [4]. Most notably, the exponent $\tilde{X}_1$ describing the asymptotic decay of the disorder-averaged first moment of the two-point function $\overline{\langle \varepsilon(x_1)\varepsilon(x_2)\rangle} \propto |x_1-x_2|^{-2\tilde{X}_1}$ seemed to be a rapidly decreasing function of q, in sharp disagreement with an exact bound on the correlation length exponent, $\nu \geqslant 2/d$ [7], which in our notation reads $\tilde{X}_1 \geqslant 1$.

More recent numerical work has emphasized the importance of crossover behavior [5] from the random fixed point to, on one side, the pure Potts model, and on the other, a percolationlike limit [4] in which the ratio $R=K_1/K_2$ between strong and weak couplings tends to infinity. It became clear that while the fixed ratio $R=2$ employed in Ref. [4] seems to have been adequate for studying the spin sector when $(q-2)$ is small, in general higher values of R are needed to measure the true random behavior in the regime $q>4$ [5].

These findings were put on a firmer ground when it was realized [8] that Zamolodchikov's c-theorem [9] can be used to explicitly trace out the critical disorder strength $R_*(q)$ as a function of q, by scanning for an extremum of the effective central charge. In conjunction with an improved transfer matrix algorithm in which the Potts model is treated through its representation as a loop model, this allowed Jacobsen and Picco [8] to produce very accurate results for the central charge and the magnetic scaling dimension in the regime $q \gg 4$.

On the analytical side, the perturbative expansions for the first three moments of the energetic two-point function have been known for quite some time [10]. It was, however, only very recently that Jeng and Ludwig [11] succeded in generalizing these computations to a general Nth moment of the energy operator $\overline{\langle \varepsilon(x_1)\varepsilon(x_2)\rangle^N} \propto |x_1-x_2|^{-2\tilde{X}_N}$, yielding

$$\tilde{X}_N = N\left(1-\frac{2}{9\pi^2}(3N-4)(q-2)^2+O(q-2)^3\right). \quad (1)$$

In particular, this makes available the experimentally relevant exponent $\tilde{X}_0'$ describing the typical decay of the energy-energy correlator in a fixed sample at criticality [10].

In this Rapid Communication we show that by combining the methods of Refs. [4,8] the exponents $\tilde{X}_1$ and $\tilde{X}_0'$ can be quite accurately determined numerically for small $(q-2)$. In particular, we find $\tilde{X}_1 \geqslant 1$ in full agreement with the correlation length bound [7], and our results lend strong support to the above two-loop results of the perturbative renormalization group.

II. SIMULATIONS

In order to compare our results with those of the $(q-2)$ expansion, while on the other hand staying comfortably away from $q=2$ where logarithmic corrections are expected [12], our main series of data has $q=2.5$. Iterating the transfer matrix for a strip of width L a large number $M \gg L$ of times, we examine the probability distribution of the ratio between the two largest Lyapunov exponents Λ_0, Λ_1 [13] in terms of the free energy gap $\Delta f(L)=(1/LM)\ln(\Lambda_0/\Lambda_1)$. We employ the loop representation of the transfer matrix where each loop on the surrounding lattice is given a weight $n=\sqrt{q}$ [14], and bond randomness is incorporated by weighing the two possible vertex configurations by w_i and $1/w_i$, where w_i is a quenched random variable that can assume two different values s and $1/s$, each one with probability 1/2 [8]. By construc-

TABLE I. Effective central charge $c(L,L+2)$ of the $q=2.5$ state model, as a function of disorder strength s and the strip width L.

s	$c(6,8)$	$c(8,10)$	$c(10,12)$
1.0	0.637 75	0.648 44	0.654 04
1.5	0.637 74	0.648 54	
2.0	0.637 70	0.648 74	
2.1	0.637 67	0.648 77	
2.2	0.637 64	0.648 79	0.654 15
2.3	0.637 59	0.648 80	0.654 15
2.4	0.637 54	0.648 79	0.654 13
2.5	0.637 48	0.648 79	0.654 11
2.6	0.637 40	0.648 77	0.654 09

tion, the system is then on average at its self-dual point [15]. The strength of the disorder is measured by $s>1$, which is related to the ratio between strong and weak bonds by $R=K_1/K_2=\ln(1+s\sqrt{q})/\ln(1+\sqrt{q}/s)$. The maximum strip width employed in the study is $L_{\max}=12$.

Following Ref. [8], we start by locating the critical disorder strength s_* by searching for a maximum of the effective central charge. To do so, we must be able to determine finite-size estimates $c(L,L+2)$ [16] with five significant digits, which means that the free energy $f_0(L)=(1/LM)\ln(\Lambda_0)$ must be known with seven digit precision. These considerations fix the necessary number of iterations to be $M=10^8$.

Our results for $c(L,L+2)$ as a function of L and s are shown in Table I. For a sufficiently large system size L, these data exhibit a maximum as a function of s, the position of which determines a finite-size estimate $s_*(L)$, which converges to s_* as $L\to\infty$. From the data of Table I, supplemented by improved three-point fits [4] (not shown), we extrapolate to $s_*(q=2.5)=2.5(1)$.

The fluctuations in $\Delta f(L)$ are examined by dividing the strip into M/m samples, each one of length $m=10^5$, from which the first few cumulants of $\Delta f(L)$ can be determined. As discussed in Ref. [4], the exponent $\tilde{X}_0'$ is related to the finite-size scaling [17] of the mean value (first cumulant) of $\Delta f(L)$, whereas $\tilde{X}_1$ is similarly determined from the sum of the entire cumulant expansion. In practice, the second cumulant is roughly two orders of magnetude smaller than the first, and higher cumulants are expected to be further suppressed, even though their determination is made difficult by numerical instabilities. We can therefore with confidence truncate the sum of the cumulants after the second one.

TABLE II. Effective exponent $\tilde{X}_0'(L)$ of the $q=2.5$ state model, as a function of disorder strength s and the strip width L.

s	$\tilde{X}_0'(6)$	$\tilde{X}_0'(8)$	$\tilde{X}_0'(10)$	$\tilde{X}_0'(12)$
1.0	0.9791	0.9513	0.9375	0.9293
1.5	1.0026	0.9754	0.9622	
2.0	1.0468	1.0191	1.0057	
2.2	1.0662	1.0377	1.0238	1.0158
2.3	1.0760	1.0470	1.0328	1.0246
2.4	1.0859	1.0563	1.0418	1.0333
2.5	1.0958	1.0656	1.0506	1.0419
2.6	1.1057	1.0749	1.0595	1.0504

TABLE III. Effective exponent $\tilde{X}_1(L)$ of the $q=2.5$ state model, as a function of disorder strength s and the strip width L.

s	$\tilde{X}_1(6)$	$\tilde{X}_1(8)$	$\tilde{X}_1(10)$	$\tilde{X}_1(12)$
1.0	0.9791	0.9513	0.9375	0.9293
1.5	1.0016	0.9746	0.9614	
2.0	1.0389	1.0124	0.9997	
2.2	1.0534	1.0269	1.0142	1.0065
2.3	1.0603	1.0338	1.0212	1.0133
2.4	1.0671	1.0405	1.0279	1.0198
2.5	1.0735	1.0469	1.0343	1.0260
2.6	1.0798	1.0531	1.0405	1.0320

Resulting finite-size estimates [17] of $\tilde{X}_0'$ and $\tilde{X}_1$ are shown in Tables II and III, respectively. Unlike what seemed to be the situation in the *magnetic* sector, these estimates exhibit a pronounced dependence on s. Jacobsen and Cardy [4] worked at fixed $R=2$, which for $q=2.5$ would correspond to $s\simeq 1.7$, and found $\tilde{X}_0'<1$ for all $q>2$. We see here that the correct way to extract these exponents is to extrapolate the $s=s_*$ data to the $L\to\infty$ limit. With the help of improved two-point estimates [4] (not shown), we thus obtain

$$\tilde{X}_0'=1.02(1), \quad \tilde{X}_1=1.00(1), \tag{2}$$

which verifies the bound of Ref. [7]. These exponents, as well as the result for their difference $\tilde{X}_0'-\tilde{X}_1=0.015(5)$, are in very good agreement with the $(q-2)$-expansion; see Eq. (1).

We have also performed simulations for higher values of q, where the discrepancy between Refs. [4] and [7] was even more pronounced, since s_* is an increasing function of q. For $q=2.75$ and $q=3$, we had to increase the number of iterations to $M=10^9$ in order to keep the error bars under control despite the higher disorder strength. In all cases we found good agreement with Ref. [7] and with the $(q-2)$ expansion, at least in the range where the latter can be assumed to be valid. A summary of our results is given in Table IV.

III. CONCLUSION

In summary, we have shown that the apparent violation of the correlation length bound [7] observed in Ref. [4] can be

TABLE IV. Numerical results for the critical disorder strength (s_* or R_*) and energetic scaling exponents ($\tilde{X}_0'$ and $\tilde{X}_1$) as functions of q. The agreement with the two-loop expansion, Eq. (1), is good.

q	s_*	R_*	$\tilde{X}_0'$		$\tilde{X}_1$	
			Numerics	Theory	Numerics	Theory
2.50	2.5(1)	3.3(2)	1.02(1)	1.023	1.00(1)	1.006
2.75	3.0(3)	4.1(5)	1.04(2)	1.051	1.01(1)	1.013
3.00	3.5(5)	4.7(10)	1.06(3)	1.09	1.02(2)	1.02

dismissed as a crossover effect due to the lack of tuning to the critical disorder strength. In conjunction with the results on degeneracy and descendents given in Ref. [4], we would thus claim that the transfer matrix method [4,8] can, at least in principle, be used to relate the entire Lyapunov spectrum to the operator content of the (as yet unknown) underlying conformal field theory.

In particular, we have supplied convincing numerical validation of the two-loop expansion (1) for the energetic multiscaling exponents [11]. Our results also provide further evidence in favor of the replica symmetric approach to the perturbative calculations, since the assumption of initial replica symmetry breaking leads to $\tilde{X}_1 = 1 + O(q-2)^3$ [18], which seems to be ruled out by the results given in Table IV.

[1] A.B. Harris, J. Phys. A **7**, 1671 (1974).
[2] J. Cardy, Physica A **263**, 215 (1999).
[3] C. Chatelain and B. Berche, Nucl. Phys. B **572**, 626 (2000).
[4] J.L. Jacobsen and J. Cardy, Nucl. Phys. B **515** [FS], 701 (1998).
[5] M. Picco, e-print cond-mat/9802092.
[6] H.W.J. Blöte and M.P. Nightingale, Physica A **112**, 405 (1982).
[7] J.T. Chayes, L. Chayes, D.S. Fisher, and T. Spencer, Phys. Rev. Lett. **57**, 2999 (1986); Commun. Math. Phys. **120**, 501 (1989).
[8] J.L. Jacobsen and M. Picco, Phys. Rev. E **61**, 13 (2000).
[9] A.B. Zamolodchikov, Pis'ma Zh. Éksp. Teor. Fiz. **43**, 565 (1986) [JETP Lett. **43**, 730 (1986)].
[10] A.W.W. Ludwig, Nucl. Phys. B **330**, 639 (1990).
[11] M. Jeng and A.W.W. Ludwig, e-print cond-mat/9910181.
[12] J.L. Cardy, J. Phys. A **19**, L1093 (1986).
[13] H. Furstenberg, Trans. Am. Math. Soc. **68**, 377 (1963).
[14] R.J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, London, 1982).
[15] W. Kinzel and E. Domany, Phys. Rev. B **23**, 3421 (1981).
[16] H.W. Blöte, J.L. Cardy, and M.P. Nightingale, Phys. Rev. Lett. **56**, 742 (1986); I. Affleck, *ibid.* **56**, 746 (1986).
[17] J.L. Cardy, J. Phys. A **16**, L355 (1983).
[18] Vik.S. Dotsenko, Vl.S. Dotsenko, M. Picco, and P. Pujol, Europhys. Lett. **32**, 425 (1995).

Chapitre 8

Modèles couplés

8.1 Désordre et limite de répliques

Dans l'approche du groupe de renormalisation perturbatif, le modèle de Potts aléatoire est intimement lié aux systèmes de plusieurs modèles de Potts purs couplés. Dans la limite continue, la fonction de partition pour une réalisation du désordre $m(x)$ s'écrit

$$Z(\beta) \propto \mathrm{Tr}\ \exp\left\{-\mathcal{H}_0 - \int d^2x\, m(x)\varepsilon(x)\right\}, \tag{8.1}$$

où $\mathcal{H}_0$ est le hamiltonien (action euclidéenne) du modèle sans désordre. Le désordre étant sous forme de couplages aléatoires entre les spins, il couple à l'opérateur de densité d'énergie $\varepsilon(x)$.

Nous avons déjà évoqué que, dans un modèle avec des impuretés gelées, la moyenne sur le désordre se fait au niveau de l'énergie libre. Formellement, cette moyenne se calcule en répliquant le système :

$$\overline{F(\beta)} \propto \overline{\log Z(\beta)} \propto \lim_{N\to 0} \frac{\overline{(Z(\beta))^N} - 1}{N}. \tag{8.2}$$

Il s'agit alors de résoudre un système de N modèles couplés deux à deux par leurs densités d'énergie :

$$\overline{(Z(\beta))^N} \propto \mathrm{Tr}\exp\left\{-\sum_{a=1}^{N}\mathcal{H}_0^{(a)} - m_0\int d^2x\sum_{a=1}^{N}\varepsilon_a(x) + g_0\int d^2x\sum_{a\neq b}\varepsilon_a(x)\varepsilon_b(x)\right\}. \tag{8.3}$$

Ici, g_0 est proportionnel à la largeur de la distribution $m(x)$, et sa valeur moyenne m_0 s'annule au point critique, $\beta = \beta_c$.

D'autre part, la fonction bêta calculée au moyen du groupe de renormalisation prend la forme [123, 124, 48]

$$\frac{\mathrm{d}g}{\mathrm{d}\xi} = \beta(g) \propto 3\epsilon g + (N-2)g^2 + \mathcal{O}(\epsilon^2, g^3), \tag{8.4}$$

où le paramètre $\epsilon \propto q - 2 \ll 1$. Il y a alors deux possibilités pour atteindre un point fixe à faible désordre :

- Soit avec $N < 2$ et un couplage initial positif,
- Soit avec $N > 2$ et un couplage initial négatif.

La limite de répliques, $N \to 0$, exploite la première possibilité, mais il est également possible de considérer $N > 2$ modèles couplés (sans désordre) afin d'examiner la seconde.

8.2 Unitarité des modèles couplés

La motivation pour étudier le cas de $N > 2$ modèles couplés (hormis l'intérêt intrinsèque de ce problème) est que les exposants critiques, en fonction de N, sont reliés à ceux du modèle désordonné par la limite $N \to 0$. Bien sûr, ceci est seulement vrai dans l'hypothèse où il n'y ait pas de brisure de la symétrie de répliques. Or, des comparaisons entre le développement perturbatif et des simulations numériques ont montré que cette hypothèse est bien satisfaite (voir le Chapitre 7).

Dans le contexte des théories conformes, la limite $N \to 0$ est très difficile à traiter directement, car elle correspond à une théorie non unitaire [123, 124]. Inversement, les modèles couplés (avec $N > 2$ entier) ont toutes chances d'être unitaires et l'on peut espérer que les techniques usuelles des théories conformes s'y appliquent.

Pour établir l'unitarité du point fixe des modèles couplés il est essentiel de vérifier qu'il se réalise bien comme la limite continue d'un modèle sur réseau avec des poids de Boltzmann non négatifs. Ce programme a été accompli dans l'article [49], attaché au présent chapitre. Ce travail a également permis de vérifier l'accord entre les exposants critiques obtenus par le développement perturbatif et des simulations numériques du modèle sur réseau. En particulier, le système $N = 3$, $q = 3$ possède un unique point critique non trivial avec une charge centrale déterminée numériquement comme

$$c = 2.377 \pm 0.003 \tag{8.5}$$

et perturbativement comme $c = 2.3808 + \mathcal{O}(\epsilon^5)$. La déterminaison numérique est facilitée par l'utilisation des relations de dualité discutées dans le Chapitre 2.

8.3 Théories parafermioniques

Lors du Chapitre 1, nous avons vu que les modèles minimaux unitaires ont tous $c < 1$. Plus précisement, ceci est vrai pour les représentations fortement dégénérées de l'algèbre de Virasoro qui encode l'invariance conforme [17]. Si l'on admet d'autres symétries que l'invariance conforme, l'algèbre correspondant devient plus large que celui de Virasoro et ses représentations irréducibles peuvent avoir des valeurs de c plus élevées.

Dans le cas de N modèles couplés il y a une symétrie discrète supplémentaire du type S_N qui correspond aux permutations des N indices dans (8.3). De bons candidats pour des théories conformes avec une symétrie étendue de ce type sont les théories parafermioniques introduites par Zamolodchikov et Fateev [168, 169]. Dans ces théories, seule la sous symétrie Z_N est réalisée de manière étendue, c'est-à-dire au niveau de l'algèbre chirale. La symétrie cyclique peut cependant être amplifiée jusqu'à une symétrie non abélienne à travers les champs de représentation.

Les parafermions Z_3 considérés dans [168, 169] ne peuvent pas expliquer le résultat (8.5) : leur nombre de degrés de liberté est encore trop petit. En effet, les théories [168, 71] et [169] décrivent plutôt les points critiques et tricritiques du modèle dit d'horloge. Cependant, il existe d'autres solutions à l'algèbre parafermionique [52]. On peut espérer que ces solutions s'appliquent au cas de modèles de Potts couplés.

Le besoin de théories conformes S_N est amplifié par l'observation que cette symétrie semble être restaurée dynamiquement même quand elle n'est pas présente dans l'action initiale. Par exemple, il a été observé numériquement que des systèmes de plusieurs modèles minimaux *différents* couplés par leurs densités d'énergie flottent vers un point fixe caractérisé par une symétrie S_N [51, 133].

q	$\Delta_{\varepsilon_1\varepsilon_2+\varepsilon_2\varepsilon_3+\varepsilon_3\varepsilon_1}$	$\Delta_{\varepsilon_1\varepsilon_2-\varepsilon_2\varepsilon_3}$	$\Delta_{\varepsilon_1\varepsilon_2\varepsilon_3}$	$\Delta_{\sigma_1\sigma_2}$	$\Delta_{\sigma_1\sigma_2\sigma_3}$
2.00	2.000	2.000	3.000	0.2500	0.3750
2.25	2.005	1.834	2.837	0.2775	0.4553
2.50	2.021	1.653	2.664	0.2949	0.5190
2.75	2.046	1.456	2.479	0.3030	0.5685
3.00	2.080	1.240	2.280	0.3023	0.6048
3.25	2.126	0.997	2.060	0.2921	0.6283
3.50	2.186	0.713	1.806	0.2703	0.6375
3.75	2.272	0.350	1.486	0.2303	0.6268
4.00	2.500	-0.500	0.750	0.0975	0.5301

TAB. 8.1 – Tableau 2 corrigé.

8.4 Article "Self-duality of coupled Potts models"

Nous considérons des modèles de Potts à q états couplés à travers leurs opérateurs d'énergie. Dans le cas des couplages autoduaux, des simulations numériques démontrent l'existence des points fixes non triviaux pour $2 \leq q \leq 4$. Ces points fixes était auparavant prédits par le groupe de renormalisation perturbatif. En utilisant de nouveaux algorithmes de matrices de transfert, algorithmes faisant intervenir des amas et des boucles, nous calculons des valeurs précises de la charge centrale et des exposants multifractales associés aux opérateurs d'énergie et de spin. Ces résultats se comparent favorablement avec ceux obtenus par la méthode du développement perturbatif dans la plage de paramètres où ce dernier est valide. La criticité des modèles au point fixe est vérifiée de manière indépendante en étudiant les valeurs propres plus élevées dans le secteur pair de la matrice de transfert ; nous mettons aussi en évidence l'existence des lois d'échelle par des simulations de type Monte Carlo. Ce programme pourrait marquer un premier pas vers l'identification des théories conformes décrivant le comportement critique de cette classe de modèles.

Erratum :
Il y a une erreur de signe dans l'équation (3.17). L'expression correcte est :

$$\alpha = 33 - \frac{29\sqrt{3}\pi}{3}. \tag{8.6}$$

En conséquence, le tableau 2 doit aussi être corrigé : voir le tableau 8.1 ci-dessus. L'accord avec les simulations numériques pour $\Delta_{\sigma_1\sigma_2\sigma_3}$ est maintenant très bon.

ELSEVIER

Nuclear Physics B 546 [FS] (1999) 505–557

NUCLEAR PHYSICS B

Coupled Potts models: Self-duality and fixed point structure

Vladimir Dotsenko [a], Jesper Lykke Jacobsen [b], Marc-André Lewis [a], Marco Picco [a]

[a] *LPTHE, Université Pierre et Marie Curie, Paris VI, Université Denis Diderot, Paris VII, Boîte 126, Tour 16, 1er étage, 4 place Jussieu, F-75252 Paris CEDEX 05, France*

[b] *Laboratoire de Physique Statistique, Ecole Normale Supérieure, 24 rue Lhomond, F-75231 Paris CEDEX 05, France*

Received 16 December 1998; accepted 11 February 1999

Abstract

We consider q-state Potts models coupled by their energy operators. Restricting our study to self-dual couplings, numerical simulations demonstrate the existence of non-trivial fixed points for $2 \leqslant q \leqslant 4$. These fixed points were first predicted by perturbative renormalisation group calculations. Accurate values for the central charge and the multiscaling exponents of the spin and energy operators are calculated using a series of novel transfer matrix algorithms employing clusters and loops. These results compare well with those of the perturbative expansion, in the range of parameter values where the latter is valid. The criticality of the fixed-point models is independently verified by examining higher eigenvalues in the even sector, and by demonstrating the existence of scaling laws from Monte Carlo simulations. This might be a first step towards the identification of the conformal field theories describing the critical behaviour of this class of models. © 1999 Elsevier Science B.V.

PACS: 05.50.+q; 64.60.Fr; 75.10.Hk; 75.40.Mg

Keywords: Potts model; Duality transformation; Perturbative renormalisation group; Multiscaling; Transfer matrix

[1] Unité Mixte de Recherche CNRS UMR 7589.

[2] Laboratoire associé aux universités Paris 6, Paris 7 et au CNRS.

PII S0550-3213(99)00097-8

1. Introduction

It is well-established that many spin models of statistical physics with random-valued nearest-neighbour couplings possess distinct critical points, their critical properties being different from those of the corresponding fixed-coupling models. To ensure the existence of a ferromagnetic-paramagnetic phase transition one usually restricts the study, within this class of models, to those where the random exchange couplings are exclusively ferromagnetic. This restriction may seem simplistic, but in fact it turns out that many models exhibiting site or bond dilution (systems where some couplings are zero), or even asymmetric distributions of positive and negative couplings, belong to the same critical universality class. Sometimes these statistical models are called "weakly disordered" in order to distinguish them from the strongly disordered models encountered in spin-glass theory, but in fact this nomenclature is somehow inappropriate, since the disorder (whose strength is here given by the spread of the distribution of random-valued couplings) could either be weak or strong. It is often observed that models with varying disorder strength display identical critical properties, a situation reminiscent of the well-known universality of the second-order phase transitions in non-disordered systems. Besides being of fundamental theoretical interest this universality is often of great practical importance: Whilst in calculations it is often simpler to consider weak randomness, in numerical studies stonger randomness is preferred, in order to reach more easily the new critical regime induced by the disorder.

An efficient laboratory for the study of disordered systems is provided by two-dimensional spin models. These models are non-trivial, critical, universal (unlike their one-dimensional counterparts) and easier to analyse, both analytically and numerically, than three-dimensional systems.

In this broad playground many results have been obtained over the years. Analytic results were mainly derived from perturbative renormalisation group calculations [1–3] whilst numerical results were obtained using Monte Carlo simulations and numerical diagonalisation of transfer matrices [4–7]. Much less has been found out concerning the conformal field theories (CFTs) which should describe exactly the random systems at their critical points. This is deceiving, considering the success of CFT in describing the critical behaviour of the corresponding pure models (without disorder). Our study of coupled Potts models is mainly motivated by our interest to progress towards the identification of these conformal theories.

The relation between the two problems, coupled and random, becomes clear when one considers the replica approach to random systems. For instance, the partition function of Ising-type models with random nearest-neighbour spin couplings can be cast in the form

$$Z(\beta) = \sum_{\{\sigma\}} \exp\left\{\beta \sum_{x,\alpha} J_{x,\alpha}\sigma_x\sigma_{x+\alpha}\right\} = \sum_{\{\sigma\}} \exp\left\{-\beta \sum_{x,\alpha} J_{x,\alpha}\varepsilon_{x,\alpha}\right\}. \tag{1.1}$$

We can consider, for simplicity, the model on a square lattice. Then x runs over the sites

of the lattice and the unit vector α over the neighbours, $\alpha = \mathbf{e}_1, \mathbf{e}_2$. The couplings $J_{x,\alpha}$ assigned to links of the lattice take random values, independently for each link, with some specified distribution. The product of neighbouring spins $-\sigma_x\sigma_{x+\alpha}$ in Eq. (1.1) is the energy operator $\varepsilon_{x,\alpha}$ assigned to a link, which becomes a local energy operator $\varepsilon(x)$ in the continuum limit near the critical point. In this limit the partition function (1.1) can symbolically be represented as

$$Z(\beta) \propto \mathrm{Tr}\ \exp\left\{-\mathcal{H}_0 - \int d^2x\, m(x)\varepsilon(x)\right\}, \tag{1.2}$$

$\mathcal{H}_0$ representing the Hamiltonian (or Euclidian action, in the field theoretic terminology) of the corresponding pure CFT if it is known.

$$m(x) \propto \frac{\beta J_{x,\alpha} - \beta_c J_0}{\beta_c J_0} \tag{1.3}$$

is a mass-type coupling, which replaces $\beta J_{x,\alpha}$ in the continuum limit: the randomness of $J_{x,\alpha}$, whose average value is J_0, translates into the randomness of $m(x)$. The trace Tr in Eq. (1.3) is assumed to represent, in the continuum limit theory, the summation over the spin configurations in the lattice model (1.1).

In quenched disordered systems, averaging over the randomness should be done at the level of the free energy $F(\beta) \propto \log Z(\beta)$. To this end one introduces replicas, that is, N copies of the same system

$$(Z(\beta))^N \propto \mathrm{Tr}\ \exp\left\{-\sum_{a=1}^{N}\mathcal{H}_0^{(a)} - \int d^2x\, m(x)\sum_{a=1}^{N}\varepsilon_a(x)\right\} \tag{1.4}$$

and averages over the randomness:

$$\overline{(Z(\beta))^N}. \tag{1.5}$$

By taking the limit $N \to 0$ one recovers the average of the free energy for a single system:

$$\overline{F(\beta)} \propto \overline{\log Z(\beta)} \propto \lim_{N\to 0}\frac{\overline{(Z(\beta))^N} - 1}{N}. \tag{1.6}$$

If the distribution of the couplings $\{J_{x,\alpha}\}$ in Eq. (1.1) is gaussian, and consequently that of $m(x)$ in Eq. (1.4), the disorder-averaged $(Z(\beta))^N$ in (1.5) is equivalent to a system of N models coupled by their energy operators:

$$\overline{(Z(\beta))^N} \propto \mathrm{Tr}\exp\left\{-\sum_{a=1}^{N}\mathcal{H}_0^{(a)} - m_0\int d^2x\sum_{a=1}^{N}\varepsilon_a(x) + g_0\int d^2x\sum_{a\neq b}\varepsilon_a(x)\varepsilon_b(x)\right\}. \tag{1.7}$$

Here g_0 is the width of the distribution of $m(x)$, and m_0 is its average value. At the critical point m_0 should vanish.

In cases where the distribution of $\{J_{x,a}\}$, or of $m(x)$, is not gaussian, the resulting theory in Eq. (1.7) contains higher-order coupling terms, like

$$\sum_{a,b,c} \varepsilon_a(x)\varepsilon_b(x)\varepsilon_c(x) \quad \text{and} \quad \sum_{a,b,c,d} \varepsilon_a(x)\varepsilon_b(x)\varepsilon_c(x)\varepsilon_d(x), \tag{1.8}$$

but these are either irrelevant (in the renormalisation group sense) or do not modify the fixed point structure and its stability. We can thus limit our study to second order energy-energy couplings.

On the lattice, before taking the continuum limit, a similar procedure applied to $Z(\beta)$ in Eq. (1.1) produces

$$\overline{(Z(\beta))^N} = \sum_{\{\sigma_a\}} \exp\left\{ -\beta J_0 \sum_{x,\alpha} \sum_{a=1}^{N} \varepsilon_{x,\alpha}^{(a)} + A g_0 \sum_{x,\alpha} \sum_{a \neq b} \varepsilon_{x,\alpha}^{(a)} \varepsilon_{x,a}^{(b)} \right\}, \tag{1.9}$$

where $\varepsilon_{x,\alpha}^{(a)} = -\sigma_x^{(a)} \sigma_{x+\alpha}^{(a)}$, and A is some constant coefficient.

Thus, in order to solve the random coupling problem, one first has to solve the theory of N coupled homogeneous models, Eqs. (1.7) and (1.9), ultimately taking the limit $N \to 0$. A complete solution to the problem would be the identification of the exact conformal field theory associated with the critical point.

When looking for possible candidates to this conformal theory, an important issue arising is the non-unitarity of the $N \to 0$ limit theory, as can be seen using perturbative renormalisation group (RG). For instance, in a non-unitary theory the central charge *increases* along the RG flow, in contradistinction to what is the case for a unitary (i.e. with reflection positivity) theory [8]. As a first step, in order to avoid the problems of non-unitarity and work with a well-defined problem, we suggest to consider the problem of N (positive, integer) coupled models, Eqs. (1.7) and (1.9), and to examine the critical properties of such systems. In the conformal theory language this amounts to studying N minimal models, coupled two by two by the operators $\Phi_{1,2}$ which are the energy operators in the case of Ising or Potts models. This problem is unitary. The β-function of the renormalisation group, to second order in perturbation, is given by

$$\frac{dg}{d\xi} = \beta(g) \propto (N-2)g^2 + \mathcal{O}(g^3) \tag{1.10}$$

in the case of Ising [1], and

$$\frac{dg}{d\xi} = \beta(g) \propto 3\epsilon g + (N-2)g^2 + \mathcal{O}(\epsilon^2, g^3) \tag{1.11}$$

in the case of q-state Potts models [2,3], where $\epsilon \propto q-2$. By the usual analysis, for $N < 2$ (and eventually for $N \to 0$) the theory (1.10) runs into zero coupling along the RG flow, and the theory (1.11) runs into a new non-trivial fixed point with

$$g_{\mathrm{FP}} \equiv g_* = \frac{3\epsilon}{2-N} + \mathcal{O}(\epsilon^2). \tag{1.12}$$

Thus, the random Potts model possesses a new non-trivial fixed point, and it is therefore of interest to look for the associated conformal theory.

These considerations hold true when g is initially positive, that is $g_0 > 0$, which is the case for the random model. If now, in order to attain unitarity, we replace $N < 2$ by $N > 2$ (N integer), the coefficients in Eqs. (1.10) and (1.11) change sign and the theory runs into a strong coupling regime, which is not controlled by pertubative RG and believed to be massive. This means a finite correlation length, and thus a non-critical theory.

To avoid this problem, when passing from $N < 2$ to $N > 2$ one should simultaneously change the sign of the initial coupling g_0. We expect some sort of similarity, or duality, of the domains $N < 2, g_0 > 0$ and $N > 2, g_0 < 0$. In fact, if multiplied by $(2 - N)$, Eq. (1.11) takes the form

$$\frac{d\lambda}{d\xi} \propto 3\epsilon\lambda - \lambda^2 + \mathcal{O}(\lambda^3), \tag{1.13}$$

where we have defined a new coupling $\lambda = (2 - N)g$. In other words, the RG flows depend on g in the combination $(2 - N)g$, so that the regions $N < 2, g_0 > 0$ and $N > 2, g_0 < 0$ are similar.

In this way one can gain unitarity whilst still staying critical. Note that the critical coupling g_*, Eq. (1.12), depends only on N and $\epsilon(q)$. It is not obvious, but we can hope that the unitary $N > 2$ problem being solved, that is the exact conformal field theory being found for general N, one could analytically continue into the domain $N < 2$ and extract exact information about the random model, $N \to 0$.

Following this approach, we first have to find the exact conformal theory of the non-trivial fixed point g_* of N coupled Potts models, $N = 3, 4, 5, \ldots$. The first step is to study three coupled models, expecting to be able to extend the result to general N later. It should be observed at this point that the model of N coupled Potts lattices (minimal conformal theories) is interesting on its own right, independently of its relation to the random problem. This makes the project doubly interesting.

The system composed of three coupled q-state Potts models can be studied by the usual methods, in particular numerically, either by Monte Carlo simulations or by diagonalising the transfer matrix in a strip geometry. Our first objective is to get some confidence in the existence of the fixed points predicted by perturbative RG, and second, to get fairly accurate numerical values for the central charge and dimensions of operators like spin, energy and their moments $\sigma_1(x)\sigma_2(x)$, $\sigma_1(x)\sigma_2(x)\sigma_3(x)$, etc. This should be useful when searching for the corresponding conformal theory.

To put the numerical study on a firm basis we need a properly defined model on the lattice, or rather on three lattices which are coupled, link to link, by their energy operators, see Eq. (1.9). For the Potts model the energy operator is of the form

$$\varepsilon_{x,\alpha} = -\delta_{\sigma_x,\sigma_{x+\alpha}}, \tag{1.14}$$

where $\delta_{\sigma,\sigma'} = 1$ for $\sigma = \sigma'$ and 0 otherwise, and the spin operator σ_x at the lattice site x takes q different values. The partition function of three coupled models is of the form

$$\sum_{\{\sigma,\tau,\eta\}} \exp\left(\tilde{\beta}\sum_{x,\alpha}(\delta_{\sigma,\sigma'}+\delta_{\tau,\tau'}+\delta_{\eta,\eta'})\right.$$
$$\left.+\tilde{g}_0\sum_{x,\alpha}(\delta_{\sigma,\sigma'}\delta_{\tau,\tau'}+\delta_{\tau,\tau'}\delta_{\eta,\eta'}+\delta_{\sigma,\sigma'}\delta_{\eta,\eta'})\right) \tag{1.15}$$

with $\sigma = \sigma_x$ and $\sigma' = \sigma_{x+\alpha}$, etc. In this way we have simply copied Eq. (1.9) for the particular case of $N = 3$ and $a, b = 1, 2, 3$, absorbing the coupling constant J_0 into a redefinition of $\tilde{\beta}$.

Next we have to look for critical points of Eq. (1.15), for some $\tilde{\beta}$ and negative $\tilde{g}_0$. If the position of the critical point is not known analytically and has to be determined numerically for a system as complicated as three coupled Potts lattices, with q^3 degrees of freedom at each site, accuracy is likely to be very poor, leading to unprecise or absurd measurements of critical properties. Thus, our numerical studies would greatly benefit from an exact determination of the position of the critical point. This can be done if the model is self-dual, like the single Ising or Potts models. For three coupled models, the existence of duality relations requires the inclusion of a three-energy interaction term

$$\sum_{x,\alpha}\delta_{\sigma,\sigma'}\delta_{\tau,\tau'}\delta_{\eta,\eta'}. \tag{1.16}$$

In the continuum limit this becomes

$$\int d^2x\,\varepsilon_1(x)\varepsilon_2(x)\varepsilon_3(x). \tag{1.17}$$

It is straightforwardly shown that such a term does not modify the fixed point structure nor its stability. This means that adding (1.16) to the lattice Hamiltonian should not modify the critical properies of the model, the continuum limit theory being the same. In this way we finally arrive at a lattice model with partition function

$$\sum_{\{\sigma,\tau,\eta\}} \exp\left(\sum_{x,\alpha}[a(\delta_{\sigma,\sigma'}+\delta_{\tau,\tau'}+\delta_{\eta,\eta'})+b(\delta_{\sigma,\sigma'}\delta_{\tau,\tau'}+\delta_{\tau,\tau'}\delta_{\eta,\eta'}+\delta_{\sigma,\sigma'}\delta_{\eta,\eta'})\right.$$
$$\left.+c\,\delta_{\sigma,\sigma'}\delta_{\tau,\tau'}\delta_{\eta,\eta'}]\right). \tag{1.18}$$

The coupling constants, a, b and c, can be chosen so as to render this model self-dual. In Section 2 we shall establish the corresponding duality transformations. It will turn out that the model possesses not a point, but a line in parameter space (a, b, c) on which it is self-dual. On symmetry grounds we should expect that its critical points belong to this line. In fact, it will be shown that generically there exists three self-dual fixed points:

(1) That of a single Potts model with q^3 states of the spin variable. At this point $a = b = 0$ and $c \neq 0$. Whenever the energy-energy coupling between the models is relevant (i.e. for $q > 2$), we have $q^3 \gg 4$. The phase transition is thus first-order, and the model is not critical.

(2) That of three decoupled q-state Potts models. At this point $b = c = 0$ and $a \neq 0$. The phase transition is second-order if $q \leqslant 4$.

(3) The non-trivial fixed point of three coupled models. The transition is here second-order, and the model is critical. The study of this point is the principal subject of this paper.

To make a better connection with the continuum limit theory, the energy operators of the lattice model will be redefined in the next section to take the form

$$E_{\sigma,\sigma'} = 1 - \delta_{\sigma,\sigma'}. \tag{1.19}$$

The principal coupling term then becomes

$$-g_0 \sum_{x,\alpha} \left(E_{\sigma,\sigma'} E_{\tau,\tau'} + E_{\tau,\tau'} E_{\eta,\eta'} + E_{\sigma,\sigma'} E_{\eta,\eta'} \right), \tag{1.20}$$

where the parameter g_0 is a linear combination of the parameters b, c in Eq. (1.18):

$$g_0 = b + c. \tag{1.21}$$

With respect to g_0 the critical point (1), corresponding to a first-order transition, has positive g_0, as it should have been expected. The decoupling point (2) has $g_0 = 0$, whilst the fixed point (3) is found for finite negative g_0.

As will be evident from the subsequent sections, to locate this last fixed point we have relied on the c-theorem of Zamolodchikov [8], which states that the effective central charge of the theory decreases along the RG flow. The effective central charge has been measured using the strip geometry, as will be explained in detail in Section 4. We assumed on symmetry grounds, and verified numerically, that the RG flow from the decoupling point to the non-trivial coupling point goes along the line of self-duality that we have found. To locate the point (3) we stayed on the self-dual line, on the negative g_0 side, and followed, using transfer matrices on a strip, the evolution of the effective central charge along the line. This led us to a particular limiting point on the line of self-duality, actually its end-point, at which the exact values of the couplings are known. We then used transfer matrices and Monte Carlo simulations, with the couplings tuned exactly to their end-point values, to check scaling laws and measure critical dimensions of various operators, comparing the result with those obtained by perturbative CFT.

The paper is laid out as follows. In Section 2, we present duality transformations and identify the self-dual lines for two and three coupled models. Section 3 is devoted to the computation of the central charge and the critical exponents of physical operators. To this end we employ perturbative CFT techniques. Section 4 introduces the various transfer matrix algorithms we used to numerically compute the critical properties. Since these algorithms are interesting on their own right, some of them being more efficient than those previously described in the Potts model literature, they are presented extensively. Section 5 presents the numerical results obtained using these newly introduced algorithms. A Monte Carlo study of scaling laws is also undertaken. Finally, Section 6 is devoted to concluding remarks and to a brief summary of the obtained results.

2. Self-duality and criticality

Duality relations, i.e. maps between sets of coupling constants that lead to the same partition functions, are central objects in the study of critical systems. These relations map one part of the coupling phase space to another one, a self-dual manifold separating the two. If the fixed points we are looking for are unique, they should be self-dual points in the phase space. If this were not the case, then fixed points would arise in dual pairs, which implies dual RG flows. Either these flows would never cross the self-dual line or they would both cross it at a given point. We are not able to picture in which way our system could behave like that. Numerical results presented later in this paper strongly support the conjecture that the fixed points of interest, for our models, are self-dual.

2.1. Duality relations

Let us first outline the construction of a global duality transformation from which the self-dual lines will be extracted. Following Ref. [10], we shall restrict our attention to Hamiltonians of the form

$$\mathcal{H} = -\sum_{\langle i,j\rangle} J(\xi_{ij}^{(1)}, \ldots, \xi_{ij}^{(N)}), \tag{2.1}$$

where $\xi_i^{(k)} = 1, \ldots, q$ are the spin variables of the kth model and $\xi_{ij}^{(k)} \equiv |\xi_i^{(k)} - \xi_j^{(k)}|$. Energy-energy coupled Potts models naturally have Hamiltonians of this form. For generic J, the partition function can be written as

$$Z = \sum_{\xi_i} \prod_{\langle i,j\rangle} u(\boldsymbol{\xi}_{ij}), \tag{2.2}$$

where $\boldsymbol{\xi}_{ij} = (\xi_{ij}^{(1)}, \ldots \xi_{ij}^{(N)})$ and the local Boltzmann weights $u(\boldsymbol{\xi})$ are

$$u(\boldsymbol{\xi}) = \exp(\beta J(\boldsymbol{\xi})). \tag{2.3}$$

The matrix U, whose elements are $u^{(k)}(\xi_{ij})$, is a $q \times q$ cyclic matrix. It is rather obvious that the partition function can be rewritten as

$$Z = q^N \sum_{\xi_{ij}} \prod_{\langle i,j\rangle} u(\boldsymbol{\xi}_{ij}), \tag{2.4}$$

where a factor of q^N has been pulled out in front, since after setting all the ξ_{ij}s, one still needs to fix the absolute value of one spin in each model in order to completely specify the configuration. Using Fourier transform the partition function can be defined on the dual lattice. Since it is translationally invariant, the eigenvalues of the matrix U are given by

$$\lambda(\boldsymbol{\eta}) = \sum_{\xi} \exp\left(\frac{2\pi i \boldsymbol{\xi} \cdot \boldsymbol{\eta}}{q}\right) u(\boldsymbol{\xi}), \tag{2.5}$$

which implies, using inverse Fourier transform, that

$$u(\boldsymbol{\xi}_{ij}) = \sum_{\boldsymbol{\eta}} T(\boldsymbol{\xi}_i, \boldsymbol{\eta}) \lambda(\boldsymbol{\eta}) T^*(\boldsymbol{\xi}_j, \boldsymbol{\eta}), \tag{2.6}$$

where

$$T(\boldsymbol{\xi}, \boldsymbol{\eta}) = q^{-\frac{1}{2}} \exp\left(\frac{2\pi \mathrm{i} \boldsymbol{\xi} \cdot \boldsymbol{\eta}}{q}\right). \tag{2.7}$$

This can be inserted in Eq. (2.2), and leads directly to

$$Z = q^{1+N-N_{\mathrm{D}}} \sum_{\boldsymbol{\eta}_{ij}=1}^{q} \prod_{\langle i,j \rangle} \lambda(\boldsymbol{\eta}_{ij}), \tag{2.8}$$

where N_{D} is the number of sites on the dual lattice. The additional factor of q comes from the fact that once all the $\boldsymbol{\eta}_{ij}$ are set, one still has to define the absolute value on one site. On a self-dual lattice, such as the square, the number of spin sites on the direct lattice and on its dual are equal (neglecting boundary effects), and one thus gets

$$Z(u) = qZ(\lambda). \tag{2.9}$$

The transformations (2.5) are well-defined global duality transformations. From these, self-dual solutions can be extracted. Let us treat the cases of two and three coupled models separately.

2.1.1. *Two models*

Consider the case of two coupled models, with Hamiltonian

$$\mathcal{H} = \sum_{\langle i,j \rangle} \mathcal{H}_{ij}, \tag{2.10}$$

$$\mathcal{H}_{ij} = -a\left(\delta_{\sigma_i,\sigma_j} + \delta_{\tau_i,\tau_j}\right) - b\,\delta_{\sigma_i,\sigma_j}\delta_{\tau_i,\tau_j}. \tag{2.11}$$

The choice of sign for the coupling b conforms with existing computations for the random model.

The duality relations, mapping the couplings (a, b) to (a^*, b^*), are given by

$$\mathrm{e}^{2a^*+b^*} = \frac{\mathrm{e}^{2a+b} + 2(q-1)\mathrm{e}^a + (q-1)^2}{\mathrm{e}^{2a+b} - 2\mathrm{e}^a + 1}, \tag{2.12}$$

$$\mathrm{e}^{a^*} = \frac{\mathrm{e}^{2a+b} + (q-2)\mathrm{e}^a - (q-1)}{\mathrm{e}^{2a+b} - 2\mathrm{e}^a + 1}. \tag{2.13}$$

The denominator is introduced to ensure that configurations of zero energy remain as such under duality. The self-duality condition constrains couplings a and b to satisfy

$$\mathrm{e}^b = \frac{2\mathrm{e}^a + (q-1)}{\mathrm{e}^{2a}}. \tag{2.14}$$

Two points of direct physical interpretation can be found on this line. First, the decoupled point $b = 0$ for which $a = \ln(1 + \sqrt{q})$, which is the usual critical temperature

$\beta_c(q)$ of the decoupled models. The second, at $a = 0$ and $b = \ln(1+q)$, corresponds to a q^2-state Potts model, for which $\beta_c(q^2) = \ln(1+q)$.

We remark that the case of two coupled models was previously considered by Domany and Riedel [11].

2.1.2. Three models

The Hamiltonian associated with the case of three coupled models is

$$\mathcal{H} = \sum_{\langle i,j \rangle} \mathcal{H}_{ij}, \tag{2.15}$$

$$\mathcal{H}_{ij} = -a\,(\delta_{\sigma_i,\sigma_j} + \delta_{\tau_i,\tau_j} + \delta_{\eta_i,\eta_j}) - b\,(\delta_{\sigma_i,\sigma_j}\delta_{\tau_i,\tau_j} + \delta_{\sigma_i,\sigma_j}\delta_{\eta_i,\eta_j} + \delta_{\tau_i,\tau_j}\delta_{\eta_i,\eta_j}) - c\,\delta_{\sigma_i,\sigma_j}\delta_{\tau_i,\tau_j}\delta_{\eta_i,\eta_j}. \tag{2.16}$$

The introduction of a three-coupling term is necessary to produce self-dual solutions, since it is generated by applying the duality relations to our original Hamiltonian. The duality transformations are given by

$$e^{3a^*+3b^*+c^*} = \frac{e^{3a+3b+c} + 3(q-1)e^{2a+b} + 3(q-1)^2 e^a + (q-1)^3}{e^{3a+3b+c} - 3e^{2a+b} + 3e^a - 1}, \tag{2.17}$$

$$e^{2a^*+b^*} = \frac{e^{3a+3b+c} + (2q-3)e^{2a+b} + (q^2-4q+3)e^a - (q-1)^2}{e^{3a+3b+c} - 3e^{2a+b} + 3e^a - 1}, \tag{2.18}$$

$$e^{a^*} = \frac{e^{3a+3b+c} + (q-3)e^{2a+b} + (3-2q)e^a + (q-1)}{e^{3a+3b+c} - 3e^{2a+b} + 3e^a - 1}. \tag{2.19}$$

Self-duality solutions are found to be

$$e^b = \frac{(2+\sqrt{q})e^a - (\sqrt{q}+1)}{e^{2a}},$$

$$e^c = e^{3a}\frac{3(e^a-1)(\sqrt{q}+1) + q\sqrt{q} + 1}{(e^a(2+\sqrt{q}) - (1+\sqrt{q}))^3}. \tag{2.20}$$

The two trivial points found in the two-models case also arise here; the decoupling point, for which $b = c = 0$ and $a = \ln(1+\sqrt{q})$, and the q^3-state Potts model, with $a = b = 0$ and $c = \beta_c(q^3) = \ln(1+q^{3/2})$.

2.2. A convenient reparametrisation

There is a certain arbitrariness in the way we choose to parametrise a point on the self-dual line. Ideally, the couplings entering the lattice Hamiltonian would in some way be comparable to those of the field theory describing its continuum limit. Evidently, decoupling points should be associated with null couplings in both schemes, but this still leaves plenty of room for a reparametrisation. We propose such a reparametrisation which makes closer contact with the physical considerations put forward in the Introduction, and, at the same time, facilitates the implementation of numerical methods.

2.2.1. General case

Introducing the parameter x ($x \in [1, +\infty[$) in Eq. (2.16), the self-duality relations can be rewritten as

$$e^a = \frac{\sqrt{q}+1}{\sqrt{q}+2} x, \tag{2.21}$$

$$e^b = \frac{(\sqrt{q}+2)^2}{\sqrt{q}+1} \frac{x-1}{x^2}, \tag{2.22}$$

$$e^c = \frac{3(\sqrt{q}+1)^2}{(\sqrt{q}+2)^4} \frac{x^3(x-\gamma)}{(x-1)^3}, \tag{2.23}$$

where $\gamma \equiv \frac{4-q}{3}$. Since the decoupled models we study have ferromagnetic ground states, it is more convenient to trade the $\delta_{\sigma_i,\sigma_j}$ for the operators

$$E_{\sigma_i,\sigma_j} = 1 - \delta_{\sigma_i,\sigma_j}. \tag{2.24}$$

Doing so, the Hamiltonian becomes

$$\begin{aligned}\mathcal{H}_{ij} = &-(3a+3b+c) + (a+2b+c)\,(E_{\sigma_i,\sigma_j} + E_{\tau_i,\tau_j} + E_{\eta_i,\eta_j}) \\ &-(b+c)\,(E_{\sigma_i,\sigma_j}E_{\tau_i,\tau_j} + E_{\sigma_i,\sigma_j}E_{\eta_i,\eta_j} + E_{\tau_i,\tau_j}E_{\eta_i,\eta_j}) \\ &+c\; E_{\sigma_i,\sigma_j}E_{\tau_i,\tau_j}E_{\eta_i,\eta_j}.\end{aligned} \tag{2.25}$$

The constant term, being the ground state energy, can be gauged out. Now define the new one- and two-energy coupling constants

$$\begin{aligned}\beta &= a + 2b + c, \\ g &= b + c.\end{aligned} \tag{2.26}$$

With this set of couplings the Hamiltonian is turned into

$$\begin{aligned}\mathcal{H}_{ij} = &\beta\,(E_{\sigma_i,\sigma_j} + E_{\tau_i,\tau_j} + E_{\eta_i,\eta_j}) \\ &-g\,(E_{\sigma_i,\sigma_j}E_{\tau_i,\tau_j} + E_{\sigma_i,\sigma_j}E_{\eta_i,\eta_j} + E_{\tau_i,\tau_j}E_{\eta_i,\eta_j}) \\ &+c\; E_{\sigma_i,\sigma_j}E_{\tau_i,\tau_j}E_{\eta_i,\eta_j}.\end{aligned} \tag{2.27}$$

On the self-dual line the couplings β and g are parametrised by

$$e^{-\beta} = \frac{\sqrt{q}+2}{3(\sqrt{q}+1)} \frac{x-1}{x-\gamma}, \tag{2.28}$$

$$e^{g} = \frac{3(\sqrt{q}+1)}{(\sqrt{q}+2)^2} \frac{x(x-\gamma)}{(x-1)^2}. \tag{2.29}$$

A similar reparametrisation for the case of two models leads to the Hamiltonian

$$\mathcal{H}_{ij} = \beta(E_{\sigma_i,\sigma_j} + E_{\tau_i,\tau_j}) - g(E_{\sigma_i,\sigma_j}E_{\tau_i,\tau_j}), \tag{2.30}$$

with

$$e^{-\beta} = \frac{x^2}{2x^2 + x(q-1)},$$
$$e^{g} = \frac{2x + (q-1)}{x^2}. \tag{2.31}$$

This particular parametrisation has some definite advantages over the original one. First, all Boltzmann weights remain finite along the self-dual line, something useful both in numerical studies and in a comparison with perturbative CFT. Second, as we shall see in Section 3, along the self-dual line, the coupling constant g has a sign in agreeement with perturbative computations. Finally, in the three-models case, the three-coupling term becomes infinite when $x \to \infty$, leading to a null Boltzmann weight. This implies some simplifications of numerical studies, as we now show.

2.2.2. Limits on the self-dual lines

As we just said, the limit $x \to \infty$ will be of special interest in the following sections. There we can simplify numerical computations (both using transfer matrices and Monte Carlo simulations) by suppressing some of the Boltzmann weights attached to the links. The local Boltzmann weight $W(L_{ij})$ for the degrees of freedom coupling spin sites i and j is completely determined by specifying the number L_{ij} of layers having differnt spin values at the two ends of the bond (ij). The total Boltzmann weight of a spin configuration is then given by

$$W = \prod_{\langle i,j \rangle} W(L_{ij}). \tag{2.32}$$

If we consider the case of two coupled models, Eq. (2.30), three possible weights $W(L)$ arise:

$$W(0) = 1, \qquad W(1) = e^{-\beta}, \qquad W(2) = e^{-2\beta+g}. \tag{2.33}$$

Using the self-dual parametrised solutions, we see that in the limit $x \to \infty$ this becomes

$$W(0) = 1, \qquad W(1) = 1/2, \qquad W(2) = 0. \tag{2.34}$$

For the case of three models, Eq. (2.25), there are four possible weights

$$W(0) = 1, \qquad W(1) = e^{-\beta}, \qquad W(2) = e^{-2\beta+g}, \qquad W(3) = e^{-3\beta+3g-c}. \tag{2.35}$$

At $x \to \infty$ this becomes

$$W(0) = 1, \qquad W(1) = \frac{\sqrt{q}+2}{3(\sqrt{q}+1)}, \qquad W(2) = \frac{1}{3(\sqrt{q}+1)}, \qquad W(3) = 0. \tag{2.36}$$

We shall use this limit extensively later on. The fact that configurations with one or more L_{ij} equal to three are forbidden greatly simplifies the problem. It even gives hope to solve exactly the lattice model at the critical point. In Section 4.5 below we shall see that it is possible to reformulate the Hamiltonians (2.30) and (2.25) so as to obtain even more null Boltzmann weights, by passing on to a random cluster picture.

3. Renormalisation group study of coupled Potts models

Coupled Potts models have already been studied in details using renormalisation group techniques. Details can be found in Refs. [2,3,12–15]. We shall only present here a summary of these results, which are exposed extensively in the given references.

The continuum limit of the models under consideration is defined by

$$\mathcal{H} = \sum_{i=1}^{N} \mathcal{H}_{0,i} - g \int d^2x \sum_{i \neq j}^{N} \varepsilon_i(x)\varepsilon_j(x), \tag{3.1}$$

where $\varepsilon_i(x)$ is the continuum limit of E_{σ_i,σ_j} [see Eq. (2.24)] and corresponds to the energy operator, and $\mathcal{H}_{0,i}$ are the Hamiltonians of the decoupled Potts models in the continuum limit. We shall restrict the discussion to q-state Potts models with $2 \leqslant q \leqslant 4$, for which the dimension of the energy operator varies between 1 (for $q = 2$) and $1/2$ (for $q = 4$). For such values of q, the continuum limit of the models are conformal field theories. The term $g \int \varepsilon\varepsilon$ is considered as a perturbation, which is relevant by dimensional analysis (except for $q = 2$, where it is marginal). The computation of the β-function to two loops was done in Refs. [2,3] and is given by

$$\begin{aligned}\beta(g) &\equiv \frac{dg}{d\log(r)} \\ &= 3\epsilon g(r) + 4\pi(N-2)g^2(r) - 16\pi^2(N-2)g^3(r) + \mathcal{O}(g^4(r)).\end{aligned} \tag{3.2}$$

Here ϵ is the perturbation parameter, and it corresponds to the dimension of the perturbation: $\epsilon = \frac{2}{3}(1 - \Delta_\varepsilon)$ where Δ_ε is the dimension of the energy operator of the decoupled models. For $q = 2$ we have $\epsilon = 0$, whilst for $q = 3$, $\epsilon = 2/15$ and for $q = 4$, $\epsilon = 1/3$. The parameter r is an infrared cut-off.

In order to compute critical exponents we need to compute the matrices $(Z_\varepsilon)_{ij}$ and $(Z_\sigma)_{ij}$ which are respectively the renormalisation constants of the energy and spin operators. These have to be understood in the following way:

$$\mathcal{O}_R = (Z_\mathcal{O})\mathcal{O}, \tag{3.3}$$

where the ith element of vector $\mathcal{O}$ is the operator $\mathcal{O}$ of model i, and $\mathcal{O}_R$ is the corresponding renormalised operator. These matrices are, again to second order in the perturbation, given by [3]

$$\frac{d\log(Z_\varepsilon(r))_{ij}}{dr} = \left(-4\pi(N-1)g - 8\pi^2(N-1)g^2 + \mathcal{O}(g^3)\right)(1-\delta_{ij}), \tag{3.4}$$

$$\frac{d\log(Z_\sigma(r))_{ij}}{dr} = \left(3(N-1)g^2\pi^2\epsilon\,\mathcal{F} + 4(N-1)(N-2)\pi^3g^3 + \mathcal{O}(g^4)\right)\delta_{ij}, \tag{3.5}$$

with

$$\mathcal{F} = 2\frac{\Gamma^2(-\frac{2}{3})\Gamma^2(\frac{1}{6})}{\Gamma^2(-\frac{1}{3})\Gamma^2(-\frac{1}{6})}. \tag{3.6}$$

The fact that the energy matrix is not diagonal was observed in Refs. [2,15]. This implies that the energy operators for each individual layer are no longer the proper ones to study critical behaviour. Instead, the eigenvectors of the matrix turn out to be the ones we observe in numerical studies. We shall come back to this point when we compute critical exponents in the next sub-section.

The possible behaviours of our coupled systems can be summarised as follows [13].

(i) For $q = 2$ the model corresponds to the N-colour Ashkin–Teller model. For $N = 2$ it is integrable [16]. For $N > 2$ the sign of g is determinant for the large-scale behaviour: When $g > 0$ a second-order phase transition of the Ising type is observed, whilst for $g < 0$ the scenario is that of a fluctuation-driven first order phase transition [17,18].

(ii) For $q > 2$ and $N = 2$ the model is still integrable, but now presents a mass generation [19]. This again indicates a first-order phase transition.

(iii) For the case $N > 2$ and $g > 0$ the coupling constants flow far from our perturbative region. Even if a definite proof is not given, a comparison with the case $q = 2$ seems to tell us that a mass gap is dynamically generated, indicating a first-order phase transition.

The situation is completely different for $g < 0$. In that case there is a non-trivial infrared fixed point for

$$g_* = -\frac{3\epsilon}{4\pi(N-2)} + \frac{9\epsilon^2}{4\pi(N-2)^2} + \mathcal{O}(\epsilon^3). \tag{3.7}$$

The critical exponents associated with the energy and spin operators for this fixed point, along with the values of the central charge, are computed below.

3.1. Critical exponents for $q > 2, N > 2$

We now compute the critical exponents of our coupled models at the fixed points of the renormalisation group. For the decoupling point, the critical exponents are evidently the ones of the pure models. We shall therefore concentrate on the non-trivial fixed point identified above. Since the renormalisation matrix for the spin operator is diagonal and the coupling is invariant under permutation, any linear combination of the spin operators has a well-defined critical exponent which is found to be [3]

$$\Delta'_\sigma = \Delta_\sigma - \tfrac{27}{16}\frac{N-1}{(N-2)^2}\mathcal{F}\epsilon^3 + \mathcal{O}(\epsilon^4), \tag{3.8}$$

where $\mathcal{F}$ was defined in Eq. (3.6).

Defining critical exponents for the energy operators is somehow more tricky. Correlation functions between renormalised operators (at a given cut-off R) are of the following form:

$$\langle \varepsilon_i(0)\varepsilon_j(R)\rangle \sim \sum_{k\neq i}\sum_{l\neq j}(Z_\varepsilon)_{ik}(Z_\varepsilon)_{jl}\frac{1}{R^{2\Delta_\varepsilon}}\langle \varepsilon_k(0)\varepsilon_l(1)\rangle, \tag{3.9}$$

mixing correlation functions of the N different layers. One way to extract unambiguous critical exponents is to diagonalise the renormalisation matrix, thus introducing a new basis of operators. This diagonalisation is exact, in the sense that eigenvectors for the one-loop renormalisation matrix remain eigenvectors to all order in the perturbation. For three models the eigenvectors are

$$\varepsilon_1+\varepsilon_2+\varepsilon_3, \qquad \varepsilon_1-\varepsilon_2, \qquad \varepsilon_2-\varepsilon_3. \tag{3.10}$$

When using this basis, the computation of the different exponents is straightforward. We have [15]

$$\Delta_{\varepsilon_1+\varepsilon_2+\varepsilon_3}=\Delta_\varepsilon+6\epsilon-9\epsilon^2+\mathcal{O}(\epsilon^3) \tag{3.11}$$

and

$$\Delta_{\varepsilon_1-\varepsilon_2}=\Delta_\varepsilon-3\epsilon+\tfrac{9}{2}\epsilon^2+\mathcal{O}(\epsilon^3). \tag{3.12}$$

Some explicit values of the different critical exponents for the case of three coupled model are provided by Table 1. The appearance of negative magnetic exponents for sufficiently large q shows the limitations of the perturbative expansion. It is also possible to compute the critical exponents for higher moments of the spin and energy operators. The physically significant operators are found to be

$$\sigma_1\sigma_2, \qquad \sigma_1\sigma_3, \qquad \sigma_2\sigma_3, \qquad \sigma_1\sigma_2\sigma_3; \tag{3.13}$$

$$\varepsilon_1\varepsilon_2+\varepsilon_2\varepsilon_3+\varepsilon_3\varepsilon_1, \qquad \varepsilon_1\varepsilon_2-\varepsilon_2\varepsilon_3, \qquad \varepsilon_1\varepsilon_2-\varepsilon_3\varepsilon_1, \qquad \varepsilon_1\varepsilon_2\varepsilon_3. \tag{3.14}$$

As was shown for the random Potts model [20], perturbative computations for higher moments of the spin and energy operators are much less precise than for the operators themselves, and one should keep in mind that they eventually become absurd for sufficiently high moments (for example, the third magnetic moment has a negative exponent for $q>3.7$). For the spin operators, the second-order perturbative computations lead to the following exponents [21,20]:

$$\Delta_{\sigma_1\sigma_2}=2\Delta_\sigma(\epsilon)+\frac{3\epsilon}{4(N-2)}\left[1-\frac{3\epsilon}{N-2}\left((N-2)\log 2+\tfrac{11}{12}\right)\right]+\mathcal{O}(\epsilon^3), \tag{3.15}$$

$$\Delta_{\sigma_1\sigma_2\sigma_3}=3\Delta_\sigma(\epsilon)+\frac{9\epsilon}{4(N-2)}\left[1-\frac{3\epsilon}{N-2}\left((N-2)\log 2+\tfrac{11}{12}+\frac{\alpha}{24}\right)\right]+\mathcal{O}(\epsilon^3), \tag{3.16}$$

with

$$\alpha=-33+\frac{29\sqrt{3}\pi}{3}. \tag{3.17}$$

For the energy operators, critical exponents for the diagonal operators are given by ($\varepsilon_S^2\equiv\varepsilon_1\varepsilon_2+\varepsilon_2\varepsilon_3+\varepsilon_3\varepsilon_1$ and $\varepsilon_A^2\equiv\varepsilon_1\varepsilon_2-\varepsilon_2\varepsilon_3$)

Table 1
Perturbative critical exponents of physically significant energy and spin operators for three coupled q-state Potts models

q	$\Delta_{\varepsilon_1+\varepsilon_2+\varepsilon_3}$	$\Delta_{\varepsilon_1-\varepsilon_2}$	Δ_{σ_1}
2.00	1.0000	1.0000	0.12500
2.25	1.1447	0.8499	0.12789
2.50	1.2639	0.7154	0.12964
2.75	1.3615	0.5930	0.12985
3.00	1.4400	0.4800	0.12805
3.25	1.5006	0.3737	0.12353
3.50	1.5429	0.2710	0.11501
3.75	1.5624	0.1656	0.09926
4.00	1.5000	0.0000	0.04238

Table 2
Perturbative critical exponents of physically significant energy and spin operator moments for three coupled q-state Potts models

q	$\Delta_{\varepsilon_1\varepsilon_2+\varepsilon_2\varepsilon_3+\varepsilon_3\varepsilon_1}$	$\Delta_{\varepsilon_1\varepsilon_2-\varepsilon_2\varepsilon_3}$	$\Delta_{\varepsilon_1\varepsilon_2\varepsilon_3}$	$\Delta_{\sigma_1\sigma_2}$	$\Delta_{\sigma_1\sigma_2\sigma_3}$
2.00	2.000	2.000	3.000	0.2500	0.3750
2.25	2.005	1.834	2.837	0.2775	0.4421
2.50	2.021	1.653	2.664	0.2949	0.4683
2.75	2.046	1.456	2.479	0.3030	0.4570
3.00	2.080	1.240	2.280	0.3023	0.4088
3.25	2.126	0.997	2.060	0.2921	0.3205
3.50	2.186	0.713	1.806	0.2703	0.1815
3.75	2.272	0.350	1.486	0.2303	−0.0397
4.00	2.500	−0.500	0.750	0.0975	−0.6949

$$\Delta_{\varepsilon_S^2} = 2\Delta_\varepsilon(\epsilon) + 3\epsilon + \tfrac{9}{2}\epsilon^2 + \mathcal{O}(\epsilon^3) \tag{3.18}$$

$$\Delta_{\varepsilon_A^2} = 2\Delta_\varepsilon(\epsilon) - \tfrac{3}{2}\epsilon - 9\epsilon^2 + \mathcal{O}(\epsilon^3) \tag{3.19}$$

$$\Delta_{\varepsilon_1\varepsilon_2\varepsilon_3} = 3\Delta_\varepsilon(\epsilon) - \tfrac{27}{4}\epsilon^2 + \mathcal{O}(\epsilon^3) \tag{3.20}$$

The numerical values for these operators are given in Table 2.

3.2. Central charge and the c-theorem for $q > 2, N > 2$

One of the most important tools of CFT, Zamolodchikov's c-theorem [8], provides a simple way of computing the central charge for perturbed theories. Moreover, it gives us crucial information:

(i) The c-function, to be defined below, is decreasing along the renormalisation flow.
(ii) If the flow has a fixed point, the field theory at the fixed point is conformally invariant and its central charge is the value of the c-function at that point.

With our conventions for the β-function taken into account, the c-function is defined as

Table 3
Comparison of the central charge values at the pure and the non-trivial fixed points

q	c_{pure}	c_{FP}
2.00	1.5000	1.5000
2.25	1.7627	1.7620
2.50	1.9975	1.9931
2.75	2.2089	2.1976
3.00	2.4000	2.3808
3.25	2.5734	2.5500
3.50	2.7309	2.7164
3.75	2.8734	2.9054
4.00	3.0000	3.3750

$$c(g') = c_{\text{pure}} - 24 \int_0^{g'} \beta(g)\mathrm{d}g, \tag{3.21}$$

with c_{pure} being the total central charge of the decoupled models. The central charge deviation from the decoupling point value is thus easily computed. Substituting $g' = g_*$, the fixed point value of the couplings, we get the following correction:

$$\Delta c = -24 \int_0^{g_*} \beta(g)\mathrm{d}g \tag{3.22}$$

$$= -\frac{27}{8}\frac{N(N-1)}{(N-2)^2}\left(\epsilon^3 - \frac{9}{2(N-2)}\epsilon^4\right) + \mathcal{O}(\epsilon^5). \tag{3.23}$$

Table 3 presents the pure and perturbative values of the central charge. Clearly, for large enough q the latter are not to be trusted, since the perturbation theory eventually breaks down. This is witnessed by the change of sign in Eq. (3.23) when $\epsilon > \frac{2}{9}$.

4. The transfer matrices

Systems of several coupled q-state Potts models possess an enormous number of states, and one should think that accurate numerical results cannot be obtained by the transfer matrix technique, since only very narrow strips can be accessed. However, there are several possibilities for dramatically reducing the size of the state space, as we show in the present section. In particular we shall see that it is possible to numerically study the q-state Potts model for *any* real q with less computational effort than is required in the traditional transfer matrix approach to the Ising model.

We have optimised the algorithm described in Ref. [5], and adapted it to the case of three coupled models, by taking into account all known symmetries of the Boltzmann weights in the spin basis. Further progress can be made by trading the spin variables for clusters or loops. The cluster algorithm of Ref. [22] has been adapted to the problem

of coupled models, and we also describe, for the first time in the literature, the practical implementation of the associated loop algorithm. The latter algorithm allows us to address the case of three coupled models on strips of width $L_{\max} = 12$.

4.1. The four algorithms

Consider a system of coupled Potts models defined on a cylinder of circumference L and length M, measured in units of the lattice constant. The imposition of periodic boundary conditions in the transversal direction is understood throughout. The transfer matrix can be viewed as a linear operator $\mathbf{T}$ that acts on the partition function of the M row system, where fixed boundary conditions have been imposed on the degrees of freedom of the Mth row, so as to generate the corresponding quantity for a system with $M+1$ rows. To fix terminology, we shall refer to the specification of the boundary condition on the last row as the *state* of the system. Thus, symbolically,

$$Z_{\alpha}^{(M+1)} = \sum_{\beta} T_{\alpha\beta} Z_{\beta}^{(M)}, \tag{4.1}$$

where the Greek subscripts label the possible states. The matrix elements $T_{\alpha\beta}$ are nothing but the Boltzmann factors $\exp[-\mathcal{H}(\alpha,\beta)]$ arising from the interaction between the degrees of freedom in the Mth and the $(M+1)$th row, when these are in the fixed states β and α respectively.

It is well known that crucial information about the free energy, the central charge, and the scaling dimensions of various physical operators can be extracted from the asymptotic scaling (with strip width L) of the eigenvalue spectrum of the transfer matrix, and so the question arises what is the most efficient way of diagonalising it. A quite common trick is to decompose it as a product of sparse matrices which each add a single degree of freedom to the Mth row, rather than an entire new row at once. But even more important is the realisation that the total number of states, and thus the dimensionality of the transfer matrix, can be reduced by taking into account the various symmetries of the interactions and possibly by rewriting the model in terms of new degrees of freedom other than simply the Potts spins. We shall refer to any such collection of states as a *basis* of the transfer matrix, and the cardinality of the basis as its *size*. The accuracy of the finite-size data, of course, improves significantly with increasing L, and since the size is a very rapidly (typically exponentially) increasing function of L we must face the highly non-trivial algorithmic problem of identifying and implementing the most efficient basis.

These considerations have led to the development of four different algorithms for the coupled Potts model problem. We list them here in order of increasing efficiency and, roughly, increasing algorithmic ingenuity. Details on the implementations will be given in the following sub-sections. As a rough measure of their respective efficiency we indicate how large L can be attained in practice[3] for the three-layered model.

[3] The largest sparse matrices that we could numerically diagonalise using a reasonable amount of computation

(i) `alg1` uses the trivial basis of Potts spins and can access $L = 5$ for $q = 3$ and $L = 4$ for $q = 4$. Although hopelessly inefficient in the general case this algorithm is still of some use for small, integer q since it allows for a straightforward access to magnetic properties.

(ii) `alg2` still works in the spin basis, but the $\mathbb{Z}_q$ symmetry of the Potts spins has been factored out. In this process the magnetic properties are lost. The size is now independent of q, but for $q < L$ a further reduction of the size occurs due to a truncation of the admissible states. For $q = 3$ this algorithm can access $L = 7$, but for general q only $L = 6$.

(iii) `alg3` utilises a mapping to the random cluster model, and the basis consists of the topologically allowed ('well-nested') connectivities with respect to clusters of spins that are in the same Potts state. Magnetic properties are lost, but can be restored through the inclusion of a ghost site at the expense of increasing the number of basis states. Since q enters only as a (continuous) parameter, this algorithm is particularly convenient for comparing with analytic results obtained by perturbatively expanding in powers of $(q-2)$. In the non-magnetic sector $L = 7$ can be accessed.

(iv) `alg4` works in a mixed representation of random clusters and their surrounding loops on the medial lattice. Magnetic properties can be addressed through the imposition of twisted boundary conditions, which by duality are equivalent to the topological constraint that clusters and loops do not wrap around the cylinder. Again q is a continuous parameter. Due to the definition of the medial lattice, only even L are allowed. This algorithm can access $L = 12$ in the non-magnetic sector and $L = 10$ in the magnetic one.

Before turning our attention to a detailed description of these algorithms we briefly recall how physically interesting quantities may be extracted from the transfer matrix spectra.

4.2. *Extraction of physical quantities*

The reduced free energy per spin in the limit $M \to \infty$ of an infinitely long cylinder is given by

$$f_0(L) = -\lim_{M\to\infty} \frac{1}{LM} \ln \mathrm{Tr} Z^{(M)} = -\frac{1}{L} \ln \lambda_0(L), \tag{4.2}$$

where λ_0 is the largest eigenvalue of $\mathbf{T}$. Starting from some arbitrary inititial vector of unit norm $|\mathbf{v}_0\rangle$, this can be found by simply iterating the transfer matrix [23]

$$\lambda_0(L) = \lim_{M\to\infty} \frac{1}{M} \ln \left\| \mathbf{T}^M |\mathbf{v}_0\rangle \right\|. \tag{4.3}$$

Higher eigenvalues $\lambda_k(L)$ are found by iterating a set of n vectors $\{|\mathbf{v}_k\rangle\}_{k=0}^{n-1}$, where a given $|\mathbf{v}_k\rangle$ is orthogonalised to the set $\{|\mathbf{v}_l\rangle\}_{l=0}^{k-1}$ after each multiplication by $\mathbf{T}$ [24].

---

time had size $\sim 10^7$.

In general, of course, there exists more expedient methods for diagonalising a square matrix, but since each of our four algorithms allows for factorising the transfer matrix as a product of *sparse* matrices this simple iteration method is superior. The advantage of using sparse-matrix factorisation is that it reduces the number of elementary operations from $(C_L)^2$ to LC_L per iterated row [25], where C_L is the size of the basis.

It is well known from conformal field theory how to relate the central charge to the finite-size scaling of the free energy [26,27]

$$f_0(L) = f_0(\infty) - \frac{\pi c}{6L^2} + \ldots \tag{4.4}$$

Similarly, the gaps of the eigenvalue spectrum fix the scaling dimensions x_i of physical operators[4] [28]

$$f_i(L) - f_0(L) = \frac{2\pi x_i}{L^2} + \ldots \tag{4.5}$$

In general one may construct several *sectors* of the transfer matrix by identifying the various irreducible representations of the symmetry group of the microscopic (spin) degrees of freedom. As a familiar example consider just one single Ising model ($q = 2$). Here there are two sectors (even and odd) corresponding to the possible transformations of a state vector $|\mathbf{v}\rangle \to \pm|\mathbf{v}\rangle$ under a global spin flip

$$\sigma \to (\sigma + 1) \bmod q\,. \tag{4.6}$$

This generalises straightforwardly to the q-state Potts model for which we must consider transformations $|\mathbf{v}\rangle \to \mathrm{e}^{2i\pi j/q}|\mathbf{v}\rangle$, $j = 0, 1, \ldots, q-1$. The thermal and magnetic scaling dimensions, x_t and x_h, are then extracted by applying Eq. (4.5) to $f_1^{\text{even}} - f_0^{\text{even}}$ and $f_0^{\text{odd}} - f_0^{\text{even}}$ respectively. In the ordinary spin basis (as used by `alg1`) the choice between the even and odd sector can be very easily accomplished by iterating initial vectors (cf. Eq. (4.3)) that are either even or odd under the transformation (4.6). When the $\mathbb{Z}_q$ symmetry has been factored out (as in `alg3` and `alg4`) the odd sector has to be constructed explicitly by means of a ghost site or a seam (see below for details).

When considering several coupled Potts models more than one magnetic exponent may be defined. Namely, for each individual model one may independently choose between the even (e) and odd (o) sector. At the decoupling point, of course, one has

$$\frac{1}{3}x_h^{\text{ooo}} = \frac{1}{2}x_h^{\text{eoo}} = x_h^{\text{eeo}}, \tag{4.7}$$

but at a general critical fixed point the magnetic exponents thus defined are independent.

When using Eqs. (4.4) and (4.5) to obtain finite-size estimates for c and x_i the convergence properties can be considerably improved by explicitly including higher-order terms in $1/L$. In an extensive study of the q-state Potts model Blöte and Nightingale [29] numerically showed that the sub-leading correction to the free energy takes

[4] The notational discrepancy with Section 3 is intentional. Indeed, the identification between the scaling dimensions of physical operators and numerically measured quantities is the object of Section 5.

the form of an additional $1/L^4$ term on the right-hand side of Eq. (4.4). Not surprisingly an analogous statement can be made about Eq. (4.5) for the scaling dimensions. From the viewpoint of conformal field theory such a non-universal correction to scaling can be rationalised as arising from the operator $T\overline{T}$, where T denotes the stress-energy tensor [6]. This operator of dimension 4 is present in any theory that is conformally invariant [30].

Finite-size estimates $c(L, L')$ can then be extracted either from two-point fits for $f_0(L)$ and $f_0(L' \equiv L+1)$ using Eq. (4.4) or from three-point fits for $f_0(L)$, $f_0(L+1)$ and $f_0(L' \equiv L+2)$ using

$$f_0(L) = f_0(\infty) - \frac{\pi c}{6L^2} + \frac{A}{L^4} + \ldots \tag{4.8}$$

Similarly, for the scaling dimensions the one-point estimates $x_i(L)$ obtained from Eq. (4.5) can be improved by extracting two-point estimates $x_i(L, L+1)$ from fits of the form

$$f_i(L) - f_0(L) = \frac{2\pi x_i}{L^2} + \frac{B}{L^4} + \ldots \tag{4.9}$$

4.3. Algorithm `alg1`

In the trivial spin basis each state is specified by the values of the L Potts spins in row M. The size of the transfer matrix for N coupled models is therefore

$$N_{q,N}^{\mathtt{alg1}}(L) = q^{LN}. \tag{4.10}$$

These numbers increase very rapidly as a function of strip width L, and they do not have a q-independent upper bound. To highlight the merits of the more sophisticated algorithms that we are about to develop we present some explicit values of the $N_{q,N}^{\mathtt{alg1}}$ in Table 4.

The reason that the trivial spin basis is still of some use is that it does not explicitly break the $\mathbb{Z}_q$ symmetry of the Potts spins. Therefore, $\mathbf{T}$ contains both the even and the odd sectors, and the various magnetic scaling exponents can easily be extracted. Furthermore, `alg1` has served as a check of the results obtained from the optimised algorithms presented below.

In the case of the Ising model ($q = 2$), states which are even and odd under a global spin flip [cf. Eq. (4.6)] are given by

$$|\mathrm{e}\rangle = \frac{1}{\sqrt{2}}(|00\ldots0\rangle + |11\ldots1\rangle), \qquad |\mathrm{o}\rangle = \frac{1}{\sqrt{2}}(|00\ldots0\rangle - |11\ldots1\rangle). \tag{4.11}$$

This is easily generalised to several layers of spins by simply forming direct product states. For example, for two coupled models we can define the states

$$|\mathrm{eo}\rangle = |\mathrm{e}\rangle \otimes |\mathrm{o}\rangle, \qquad |\mathrm{oo}\rangle = |\mathrm{o}\rangle \otimes |\mathrm{o}\rangle, \tag{4.12}$$

Table 4
Number $N_{q,N}^{\mathtt{alg1}}(L)$ of basis states in `alg1` as a function of strip width L and the number N of coupled Potts models. We indicate here these numbers up to the largest strip width for which we were able to diagonalise the transfer matrix

$N_{q,N}^{\mathtt{alg1}}(L)$	$q=2$		$q=3$		$q=4$	
N	2	3	2	3	2	3
$L=2$	16	64	81	729	256	4,096
$L=3$	64	512	729	19,683	4,096	262,144
$L=4$	256	4,096	6,561	531,441	65,536	16,777,216
$L=5$	1,024	32,768	59,049	14,348,907	1,048,576	
$L=6$	4,096	262,144	531,441		16,777,216	
$L=7$	16,384	2,097,152	4,782,969			
$L=8$	65,536	16,777,216				
$L=9$	262,144					

so that the gap between the largest eigenvalue in each of these sectors and the largest eigenvalue in the totally symmetric ($|\mathrm{ee}\rangle = |\mathrm{e}\rangle \otimes |\mathrm{e}\rangle$) sector defines two magnetic exponents, x_h^{eo} and x_h^{oo}, which are in general independent.

For the q-state Potts model the appropriate prescription reads

$$|\mathrm{e}\rangle = \frac{1}{\sqrt{q}}\left(|00\ldots0\rangle + |11\ldots1\rangle + \ldots + |(q-1)(q-1)\ldots(q-1)\rangle\right),$$
$$|\mathrm{o}\rangle = \frac{1}{\sqrt{q}}\left(|00\ldots0\rangle + \mathrm{e}^{\frac{2i\pi}{q}}|11\ldots1\rangle + \ldots + \mathrm{e}^{2i\pi\frac{q-1}{q}}|(q-1)(q-1)\ldots(q-1)\rangle\right), \tag{4.13}$$

where the physical state is obtained from Eq. (4.12) by taking the real part.

4.4. Algorithm `alg2`

When the $\mathbb{Z}_q$ symmetry of the Potts spins is factored out the size of the basis can be dramatically reduced. The drawback of this approach, however, is that the magnetic properties are lost. In other words, the resulting transfer matrix does only have an even sector, but this is still enough to extract finite-size estimates for the central charge and the thermal exponent. (On the other hand, we shall see in the following sub-sections that in the case of the algorithms `alg3` and `alg4` it is possible explicitly to reconstruct the odd sector from topological considerations.)

The algorithm `alg2` was already used in Ref. [5] to study the random-bond Potts model, and before adapting it to the case of several coupled Potts models we briefly recall its application to the single-layered model.

The basic idea is that the $\delta_{\sigma_i,\sigma_j}$-type interactions do not depend on the explicit values of the spins σ_i and σ_j. What matters is whether the two spins take identical or different values. Therefore, the possible number of states for a row of L spins is equal

to the number of ways b_L in which L objects can be partitioned into indistinguishable parts [31]. With m_ν parts of ν objects each ($\nu = 1, 2, \ldots$) this can be rewritten as

$$b_L = \sum_{m_\nu=0}^{\infty}{}' \prod_{\nu=1}^{\infty} \frac{L!}{(\nu!)^{m_\nu} m_\nu!}, \tag{4.14}$$

where the primed summation is constrained by the condition $\sum_{\nu=1}^{\infty} \nu m_\nu = L$. Alternatively one can write [5]

$$b_L = \sum_{i_2=1}^{2} \sum_{i_3=1}^{m_3} \sum_{i_4=1}^{m_4} \cdots \sum_{i_L=1}^{m_L} 1, \tag{4.15}$$

with $m_k = 1 + \max(i_2, i_3, \ldots, i_{k-1})$. From the former representation the generating function is found as

$$\exp(\mathrm{e}^t - 1) = \sum_{n=0}^{\infty} \frac{b_n t^n}{n!}, \tag{4.16}$$

whence

$$b_L = \frac{1}{\mathrm{e}} \sum_{k=0}^{\infty} \frac{k^L}{k!}. \tag{4.17}$$

Yet another way of interpreting the b_L is to notice [6] that they are the total number of possible L-point connectivities, including the non-well nested ones [29]. We shall come back to the notion of well-nestedness when we discuss `alg3`.

A major advantage of the basis used in `alg2` as compared to the trivial spin basis is that the b_L do not depend on q. Thus, any integer value of q can be accessed with this algorithm. However, for any fixed q the size of the transfer matrices that have $L < q$ can be further reduced.[5] Namely, in this case the number of permissible states is truncated due to the fact that L objects cannot be partitioned into more than L indistinguishable parts! In this way we identify the number of states $N_{q,N}^{\mathtt{alg2}}(L)$ for $N = 1$ Potts layers as a sum over Stirling numbers of the second kind [32]

$$N_{q,1}^{\mathrm{alg2}}(L) = \sum_{m=1}^{q} S_L^{(m)} = \sum_{m=1}^{q} \frac{1}{m!} \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} k^L. \tag{4.18}$$

Explicit values for $q = 2, 3, 4$ are shown in Table 5, and we conjecture that asymptotically $N_{q,1}^{\mathtt{alg2}}(L) \sim q^L$. We recall that the upper limit for general q is b_L, and this increases as $L^{\alpha L}$, where α is a constant of order unity [6]. However, by comparing Tables 4 and 5 we see that even though the number of states in algorithms `alg1` and `alg2` exhibit the same asymptotic growth, `alg2` is much more efficient than `alg1`.

[5] This possibility was not discussed in Ref. [5].

Table 5
Number $N_{q,N}^{\mathtt{alg2}}(L)$ of basis states in alg2. For $N = 2, 3$ coupled models these numbers are shown up to the largest strip width L for which we were able to diagonalise the transfer matrix

$N_{q,N}^{\mathtt{alg2}}(L)$	$q=2$			$q=3$			$q=4$		
N	1	2	3	1	2	3	1	2	3
$L=2$	2	4	4	2	3	4	2	3	4
$L=3$	4	10	20	5	15	35	5	15	35
$L=4$	8	36	120	14	105	560	15	120	680
$L=5$	16	136	816	41	861	12,341	51	1,326	23,426
$L=6$	32	528	5,984	122	7,503	310,124	187	17,578	1,107,414
$L=7$	64	2,080	45,760	365	66,795	8,171,255	715	255,970	
$L=8$	128	8,256	357,760	1,094	598,965		2,795	3,907,410	
$L=9$	256	32,896	2,829,056						

4.4.1. Layer indistinguishability

In the general case of N coupled Potts models further progress can be made by observing that since all layers interact symmetrically there is an additional S_N permutational symmetry. If we imagine numbering the one-layer states by an integer $i = 1, 2, \ldots, N_{q,1}(L)$ a general N-layer state can be represented by $(i_1, i_2, \ldots, i_N)$ where $i_1 \geqslant i_2 \geqslant \ldots \geqslant i_N$. The total number of states for N coupled models is then

$$N_{q,N} = \sum_{i_1=1}^{N_{q,1}} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} \cdots \sum_{i_N=1}^{i_{N-1}} 1. \tag{4.19}$$

For $N = 2, 3$ this is easily found to be

$$N_{q,2} = \frac{1}{2}\left(N_{q,1}^2 + N_{q,1}\right),$$
$$N_{q,3} = \frac{1}{6}\left(N_{q,1}^3 + 3N_{q,1}^2 + 2N_{q,1}\right). \tag{4.20}$$

Again explicit values pertaining to alg2 are shown in Table 5.

4.5. Mapping to random cluster model

An altogether different approach for setting up the Potts model transfer matrices consists in rewriting the partition function in terms of extended objects (clusters and loops) rather than the local spins. The resulting random cluster models [33] and loop gases [34] have the advantage that the specification of their states no longer depends on the value of q. Instead q enters only through the fugacity of the non-local objects, and hence it can be taken as a continuously varying parameter. Such reformulations of the problem are especially convenient for making contact with the predictions of conformal field theory, and in particular with the perturbative expansion in powers of $(q-2)$ which has been studied in Section 3.

The practical implementation of transfer matrices for such non-local degrees of freedom was pioneered by Blöte and collaborators [22,35]. The basic idea is here that in two dimensions the allowed connectivities of clusters and loops are strongly constrained by topological considerations. Accordingly the size of the corresponding transfer matrix is drastically reduced.

In the following two sub-sections we generalise these algorithms to the case of N coupled Potts models, and we give explicit details on the implementation for $N = 1, 2, 3$. The cluster algorithm `alg3` has previously been described for $N = 1$ by Blöte and Nightingale [29]. It was discussed in more detail in Ref. [6], where it was also shown how to adapt it to the case of bond randomness. The loop algorithm `alg4` has to our knowledge not previously been used to study the Potts model.[6] It is however the most efficient algorithm that we know of, and in fact it performs much faster than the spin basis algorithm for the Ising model!

Before focusing on the concrete implementation of `alg3` and `alg4`, which is the object of the following two sub-sections, we dedicate this sub-section to developing the appropriate mappings of the partition function. We need to consider the cases of two and three coupled Potts models in turn, but it should be clear that the results generalise straightforwardly to any number of models N.

4.5.1. Two models

The partition function for two coupled Potts models can be written as

$$Z = \sum_{\{\sigma,\tau\}} \prod_{\langle ij \rangle} \exp(\mathcal{H}_{ij}), \tag{4.21}$$

where $\mathcal{H}_{ij} = a(\delta_{\sigma_i,\sigma_j} + \delta_{\tau_i,\tau_j}) + b\delta_{\sigma_i,\sigma_j}\delta_{\tau_i,\tau_j}$. Using the standard Fortuin–Kasteleyn trick [33] the exponential of $\mathcal{H}_{ij}$ is turned into

$$\exp(\mathcal{H}_{ij}) = (1 + u_a\delta_{\sigma_i,\sigma_j})(1 + u_a\delta_{\tau_i,\tau_j})(1 + u_b\delta_{\sigma_i,\sigma_j}\delta_{\tau_i,\tau_j}), \tag{4.22}$$

where we have defined $u_a = \mathrm{e}^a - 1$ and $u_b = \mathrm{e}^b - 1$. After some straightforward algebra we obtain

$$\exp(\mathcal{H}_{ij}) = k_0 + k_1(\delta_{\sigma_i,\sigma_j} + \delta_{\tau_i,\tau_j}) + k_2\delta_{\sigma_i,\sigma_j}\delta_{\tau_i,\tau_j}, \tag{4.23}$$

where

$$k_0 = 1, \qquad k_1 = u_a, \qquad k_2 = u_a^2 + u_b(1 + u_a)^2. \tag{4.24}$$

Now imagine expanding the $\prod_{\langle ij \rangle}$ product in Eq. (4.21), using Eq. (4.23). In this way we obtain a total of 3^E terms ($E = |\langle ij \rangle|$ being the number of lattice edges), each consisting of a product of E factors of k_i ($i = 0, 1, 2$) multiplying a product of Kronecker deltas. Define $\mathcal{L}$ to be the graph consisting of two copies ('layers') of the lattice on which the Potts model is defined, one copy placed on top of the other. To

[6] Details on the implementation of related but more complicated loop models can be found in Refs. [35,36].

each of the terms in the expansion we associate a graphical representation $\mathcal{G}$ on $\mathcal{L}$ by colouring the lattice edges for which the corresponding Kronecker deltas occur in the product. In other words, $\mathcal{G}$ takes the form of a two-layered bond percolation graph.

To reproduce the partition function (4.21) we need to perform the sum $\sum_{\{\sigma,\tau\}}$ over the spin degrees of freedom. Because of the Kronecker deltas this is trivially done for each term $\mathcal{G}$, yielding simply a factor of q for each of the C connected components ('clusters') in $\mathcal{G}$.[7] The factors of k_i are easily accounted for by writing $E = E_0 + E_1 + E_2$, where E_j is the number of occurences in $\mathcal{G}$ of a situation where precisely $j = 0, 1, 2$ edges placed on top of one another have been simultaneously coloured. The partition function then takes the form

$$Z = \sum_{\mathcal{G}} q^C k_0^{E_0} k_1^{E_1} k_2^{E_2}. \tag{4.25}$$

On the self-dual line (2.14) the edge weights (4.24) assume the simple form

$$k_0 = 1, \qquad k_1 = \mathrm{e}^a - 1, \qquad k_2 = q. \tag{4.26}$$

As has already been mentioned, the limit $a \to \infty$ is of special interest. In this limit we can rewrite the partition function as $Z = \exp(Ea)\tilde{Z}$, where $\tilde{Z}$ has the same form as (4.25) but with the *finite* edge weights

$$\tilde{k}_0 = 0, \qquad \tilde{k}_1 = 1, \qquad \tilde{k}_2 = 0. \tag{4.27}$$

It is worthwhile noticing that in this limit it suffices to specify the colouring configuration of one of the layers to deduce that of the other layer. Namely, whenever an edge in the first layer is coloured its counterpart in the second layer has to be uncoloured, and *vice versa*. Therefore, the configuration sum $\sum_{\mathcal{G}}$ appearing in Eq. (4.25) in fact runs over one layer only.

4.5.2. Three models

In the case of three coupled Potts models the nearest-neighbour interactions take the form $\mathcal{H}_{ij} = a\delta_1 + b\delta_2 + c\delta_3$, where we have introduced the short-hand notation

$$\delta_1 = \delta_{\sigma_i,\sigma_j} + \delta_{\tau_i,\tau_j} + \delta_{\eta_i,\eta_j},$$

$$\delta_2 = \delta_{\sigma_i,\sigma_j}\delta_{\tau_i,\tau_j} + \delta_{\sigma_i,\sigma_j}\delta_{\eta_i,\eta_j} + \delta_{\tau_i,\tau_j}\delta_{\eta_i,\eta_j}, \tag{4.28}$$

$$\delta_3 = \delta_{\sigma_i,\sigma_j}\delta_{\tau_i,\tau_j}\delta_{\eta_i,\eta_j}. \tag{4.29}$$

The objects δ_1, δ_2, δ_3 are easily shown to obey the following relations:

$$\begin{aligned} &\delta_1\delta_1 = \delta_1 + 2\delta_2, \quad \delta_1\delta_2 = 2\delta_2 + 3\delta_3, \quad \delta_1\delta_3 = 3\delta_3, \\ &\delta_2\delta_2 = \delta_2 + 6\delta_3, \quad \delta_2\delta_3 = 3\delta_3, \quad \delta_3\delta_3 = \delta_3. \end{aligned} \tag{4.30}$$

Simple algebraic manipulations then lead to the following result, generalising Eq. (4.23),

[7] Note that a single isolated site is to be counted as a cluster on its own.

$$\exp(\mathcal{H}_{ij}) = k_0 + k_1\delta_1 + k_2\delta_2 + k_3\delta_3 \tag{4.31}$$

with edge weights

$$\begin{aligned} k_0 &= 1, \\ k_1 &= u_a, \\ k_2 &= u_a^2 + u_b(1+u_a)^2, \\ k_3 &= u_a^3 + 3u_a u_b(1+u_a)^2 + (1+u_a)^3\left[u_c(1+u_b)^3 + 3u_b^2 + u_b^3\right]. \end{aligned} \tag{4.32}$$

We recall that the Fortuin–Kasteleyn parameters are $u_{a,b,c} = \mathrm{e}^{a,b,c} - 1$. After the original spin degrees of freedom have been summed over, the partition function takes on the form

$$Z = \sum_{\mathcal{G}} q^C k_0^{E_0} k_1^{E_1} k_2^{E_2} k_3^{E_3} \tag{4.33}$$

with a notation analogous to that employed for the case of two coupled models. The graph configurations $\mathcal{G}$ now consist of three bond percolation graphs stacked on top of one another.

Along the self-dual line (2.20) the edge weights (4.32) simplify to

$$k_0 = 1, \qquad k_1 = \mathrm{e}^a - 1, \qquad k_2 = q^{1/2}(\mathrm{e}^a - 1), \qquad k_3 = q^{3/2}. \tag{4.34}$$

Finally, in the limit $a \to \infty$ the modified partition function $\tilde{Z} = \exp(-Ea)Z$ has the *finite* edge weights

$$k_0 = 0, \qquad k_1 = 1, \qquad k_2 = q^{1/2}, \qquad k_3 = 0. \tag{4.35}$$

4.6. Algorithm `alg3`

With the mapping to a random cluster model in hand we are now ready to discuss the implementation of `alg3`. The point of depart of this algorithm is the form (4.33) [or (4.25)] of the partition function, in which the spin degrees of freedom have been turned into non-local clusters.

For simplicity we consider first the case of a single Potts model, with partition function $Z = \sum_{\mathcal{G}} q^C u_a^{E_1}$ [33]. To construct the transfer matrix we need to specify a basis, so that the knowledge of the state before and after the addition of a new degree of freedom gives us enough information to compute the appropriate Boltzmann weights, cf. Eq. (4.1). As shown by Blöte and Nightingale [22] this is achieved by specifying the *connectivity* (with respect to the clusters in $\mathcal{G}$) of the L points in the last row of the strip. Since connections can only be mediated by the lattice edges which have previously been added there is a very powerful topological constraint (known as *well-nestedness*) on the connectivity states. Namely, given any four consecutive points, if the first is connected to the third and the second to the fourth, then all four points must be connected. Consequently, the number c_L of allowed (well-nested) L-point connectivities is less than the total number of connectivities b_L considered in Section 4.4.

Table 6
Number $N_N^{\mathtt{alg3},4}(L)$ of basis states in the even sector of algorithm alg3 and alg4. For $N = 2, 3$ coupled models these numbers are shown up to the largest strip width L for which we were able to diagonalise the transfer matrix. With alg4 it is possible to access larger strip widths $L^{\mathtt{alg4}} = 2(L^{\mathtt{alg3}} - 1)$ than with alg3, using the *same* number of states. Since for a strip of width L, alg4 needs to employ intermediary states of $(L+2)$-point connectivities, the size of the basis given here pertains to these intermediary states

$L^{\mathtt{alg3}}$	$L^{\mathtt{alg4}}$	$N = 1$	$N = 2$	$N = 3$
2	2	2	3	4
3	4	5	15	35
4	6	14	105	560
5	8	42	903	13,244
6	10	132	8,778	392,084
7	12	429	92,235	13,251,095
8	14	1,430	1,023,165	
9	16	4,862	11,821,953	

Using a recursive principle the well-nested connectivities can be enumerated [22], and they turn out to be nothing but the Catalan numbers

$$c_L = \frac{(2L)!}{L!(L+1)!} \sim 4^L \quad \text{for } L \gg 1. \tag{4.36}$$

Since the number of Potts states enters only as a (continuous) parameter in the partition function this is *independent* of q. A mapping from the connectivities to the set of consecutive integers $1, 2, \ldots, c_L$ can also be established. More details on the practical implementation of the transfer matrices can be found in Ref. [6].

For $N > 1$ layers of coupled Potts models we need to simultaneously keep track of the connectivity state of each layer in order to compute the factors of q occurring in Eq. (4.25) or (4.33). To specify the state of the layered system we can however take advantage of the indistinguishability of the layers, cf. Eq. (4.20). The resulting number of connectivity states $N_N^{\mathtt{alg3}}(L)$ used in alg3 can be found in Table 6.

4.6.1. Magnetic properties

At the expense of increasing the size of the basis it is possible to generalise alg3 to treat the case of a Potts model in a uniform magnetic field H [22]. To this end, note that the interaction with the field can be accounted for through the inclusion of a term $H\sum_i \delta_{\sigma_i,\sigma_0}$ in the Hamiltonian, where a so-called ghost spin 0 of fixed value $\sigma_0 \equiv 1$ has been introduced. Taking each site of the lattice to be connected to 0 through a 'ghost edge', this has the usual form of a nearest-neighbour interaction. The mapping to the random cluster model therefore goes through in exact analogy with Section 4.5. In specifying the connectivities one now needs both to keep track of which sites are connected (directly or indirectly) to the ghost site and, at the same time, to specify how the remaining sites are interconnected.

The extended L-point connectivity states thus defined can be ordered and enumerated [22], and their number d_L is found to grow asymptotically as 5^L. For $H \to 0$ the transfer matrix has the following block form:

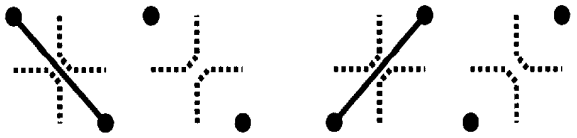


Fig. 1. The relation between the random cluster model on the square lattice and the loop model on the corresponding medial graph. The clusters consist of connected components of coloured edges (thick lines) or isolated sites (filled circles). Loops on the medial graph (dashed lines) are defined by the convention that they wrap around the cluster boundaries.

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}^{00} & \mathbf{T}^{01} \\ \mathbf{0} & \mathbf{T}^{11} \end{bmatrix}, \tag{4.37}$$

where superscript 1 (0) refers to the (non-)ghost connectivities, and the magnetic scaling dimension x_H is obtained via Eq. (4.5) from the largest eigenvalues of the $\mathbf{T}^{00}$ and $\mathbf{T}^{11}$ blocks. The physical content of this relation is that by acting repeatedly with $\mathbf{T}^{11}$ on some initial (row) state $|\mathbf{v}_0\rangle \neq 0$ one measures the decay of clusters extending back to row 0. This must have the same spatial dependence as the spin-spin correlation function and hence be related to x_H [22].

For $L = 1, 2, 3, \ldots$ the size of the magnetic block $\mathbf{T}^{11}$ is $d_L - c_L = 1, 3, 10, 37, 146, 599, \ldots$ For three coupled models this means that computations with one, two or all three layers in the magnetic sector are feasible for strip widths up to $L = 6$. Since our most refined algorithm `alg4` can access $L = 10$ we turn our attention to this next.

4.7. Algorithm `alg4`

The configurations of the random cluster model on some graph are in a one-to-one correspondence with configurations of a fully packed loop model on the *medial graph* [34]. By definition the nodes of the medial graph are situated at the mid-points of the edges of the original graph. In our case the graph on which the Potts model is defined is simply the square lattice, and the medial graph is then nothing but another square lattice that has been rotated through $\pi/4$ with respect to the original one, and rescaled by a factor of $1/\sqrt{2}$. Bipartitioning the medial graph into even and odd sublattices the precise correspondence is as shown in Fig. 1.

The partition function for N coupled Potts models can now be rewritten in the loop picture by using Euler's relation [34]. Namely, for each layer the number of clusters (C), loops (l), coloured edges (e) and sites (s) are related by $2C = l + s - e$. Thus, in the case of two models, Eq. (4.25) is turned into

$$Z = q^s \sum_{\mathcal{G}} q^{l/2} k_0^{E_0} \left(\frac{k_1}{q^{1/2}}\right)^{E_1} \left(\frac{k_2}{q}\right)^{E_2}. \tag{4.38}$$

Note that the factor of q^{-e} has been redistributed using $e = E_1 + 2E_2$. Similarly, for the three-layered model Eq. (4.33) is transformed into

$$Z = q^{3s/2} \sum_{\mathcal{G}} q^{l/2} k_0^{E_0} \left(\frac{k_1}{q^{1/2}}\right)^{E_1} \left(\frac{k_2}{q}\right)^{E_2} \left(\frac{k_3}{q^{3/2}}\right)^{E_3}. \tag{4.39}$$

When constructing the transfer matrix the advantage of this representation is that less information is needed to keep track of the loops than was the case in the cluster representation. Roughly speaking this is because the loop model is *fully packed*, so that one does not need to waste information to specify where is 'the empty space' in between the clusters.

The strip width L is now defined as the number of 'dangling ends' resulting from cutting the loops in between two rows of sites on the medial graph. A sparse-matrix decomposition can then be made by adding one vertex at a time, rather than one edge as was the case in the cluster representation. This has been illustrated in Fig. 2 for a strip of width $L = 6$. For each added vertex there are two possible configurations of the loop segments that must be summed over, cf. Fig. 1. When adding the first vertex of a new row the number of dangling ends increases from L to $L+2$, and it only goes back to L once the full row of L vertices has been completed. Therefore, we need to work interchangingly with bases specifying the loop connectivities of L and $L+2$ dangling ends [35,36].

By definition of the loops, each of the dangling ends is connected to exactly one other dangling end. In particular the strip width L must be *even*. Employing basically the same recursive argument as for the clusters [22] it is found that the possible number of connectivities among L dangling ends is now only $c_{L/2}$ [35]. For $L \gg 1$ this increases as 2^L, and in fact the loop representation of the Potts model is even more efficient than the spin representation of the Ising model![8] This is witnessed by Table 6, where we show some explicit values of $N_N^{\mathtt{alg4}}(L) = N_N^{\mathtt{alg3}}\big((L+2)/2\big)$.

At this point a brief comment on the boundary conditions is in order. Since the medial lattice is rotated through $\pi/4$ with respect to the original one, the imposition of periodic boundary conditions on the loops is not a priori equivalent to the boundary conditions hitherto used for the clusters. Indeed, these two possibilities are connected by a modular transformation, as should be clear from Fig. 2. The consistency between the results obtained from `alg3` and `alg4` will thus serve as a useful check of the modular invariance of the critical system under investigation.

To implement Eqs. (4.38) and (4.39) for several coupled Potts models we need to keep track of the edge weights on the original lattice as well as the number of closed loops on the medial graph. As shown in Fig. 1 the loop configuration suffices to determine the positions of the coloured edges on the original lattice, and so the edge weights can easily be determined from each vertex appended to the medial graph. By the same token the single-layer algorithm furnishes an even more efficient way of

[8] Using Stirling's formula a more accurate estimation of the asymptotic behaviour of $c_{L/2}$ is found as $2^L(L+2)^{-3/2}\sqrt{8/\pi}$. This approximation is asymptotically exact in a strict sense, and its relative precision is better than 10% for $L \geqslant 6$.

performing transfer matrix calculations for the random-bond Potts model than the one presented in Ref. [6].

4.7.1. Magnetic properties

In `alg4` the magnetic sector of the transfer matrix is constructed by using the fact that the spin-spin correlator maps onto a disorder operator under duality [6]. For a single Potts model at the self-dual (critical) point, or for several coupled Potts models along the self-dual lines described in Section 2, this leads to a relation between the magnetic scaling exponent and the largest eigenvalue of a modified transfer matrix on which *twisted boundary conditions* have been imposed.

Defining the local order parameter as

$$M_a(r) = \left(\delta_{\sigma(r),a} - \frac{1}{q}\right), \qquad a = 1, \ldots, q, \tag{4.40}$$

it is easily seen that the magnetic two-point correlator $G_{aa}(r_1, r_2) = \langle M_a(r_1) M_a(r_2) \rangle$ is proportional to the probability that the points r_1 and r_2 belong to the same cluster. Let us briefly recall that any given configuration of the random clusters is dual to one where each coloured edge in the direct lattice is intersected by an uncoloured edge in the dual lattice, and vice versa. Taking the two points r_1 and r_2 to reside at opposite ends of the cylinder, the graphs contributing to $G_{aa}(r_1, r_2)$ thus correspond to dual graphs where clusters are forbidden to wrap around the cylinder. This is equivalent to computing the dual partition function with twisted boundary conditions

$$\sigma \to (\sigma + 1) \bmod q \tag{4.41}$$

across a *seam* running from r_1 to r_2. By permuting the Potts spin states the shape of this seam can be deformed at will as long as it connects r_1 and r_2.

A realisation of the Potts model transfer matrix in the presence of a seam was first described in Ref. [6] within the context of `alg3`. Since there is a one-to-one correspondence between clusters and loops that wrap around the cylinder these ideas can also be applied to `alg4`.

In Fig. 2 we show a configuration contributing to the partition function with twisted boundary conditions. As the square lattice is self-dual we shall take the clusters and loops to live on the direct lattice and its medial lattice respectively. The extended connectivity state of the L $(L+2)$ dangling ends is the direct product of the $c_{L/2}$ possible connectivities mediated by the loops, and an integer specifying the position of the seam. The seam is a path inside the infinite *dual* cluster spanning the length of the cylinder, and after the addition of each vertex it must be updated according to the invariant that no loop closure take place across the seam.

Now consider the situation shown in Fig. 2 where we are about to add the fifth vertex of the top row. Since this is an even site the two possibilities for the loop configuration are those shown in the left part of Fig. 1. However, the first of these would lead to a forbidden loop closure, as witnessed by the fact that the seam is 'trapped' and cannot

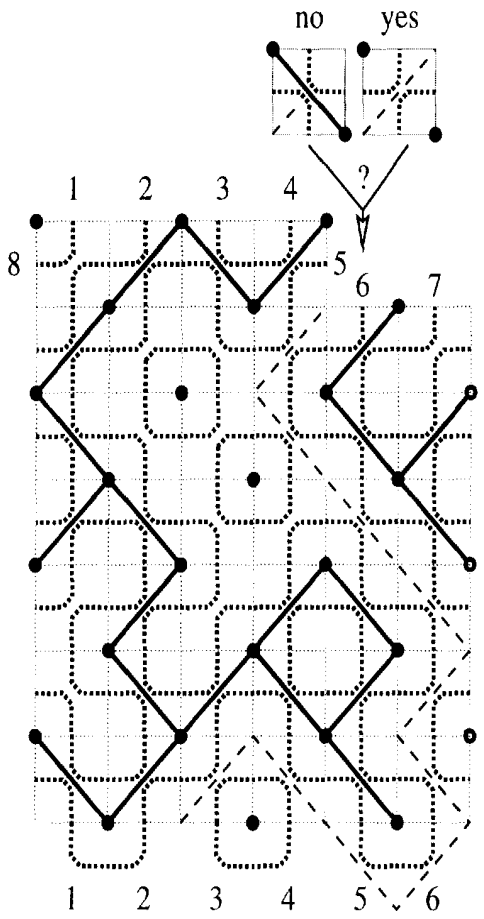


Fig. 2. Computation of the partition function with twisted boundary conditions. Coloured edges (solid linestyle) connect the sites of the direct lattice (filled circles) so as to form clusters. The loops on the medial lattice (short boldface dashes) surround the clusters, and both cluster and loops are subject to periodic boundary conditions (indicated by open circles) across the strip of width $L = 6$. A seam (long dashes) connecting sites of the dual lattice prevents the clusters and loops from wrapping around the cylinder. By adding a single vertex at a time the transfer matrix can be decomposed as a product of sparse matrices. Each multiplication by a single-vertex matrix corresponds graphically to augmenting the 'jigsaw puzzle' by one of the two 'pieces' shown in the top. In the situation at hand, one of these is forbidden by the twisted boundary conditions.

be updated. The corresponding entry in the transfer matrix is therefore zero, and only the graph realising the second possibility contributes.

Using the numbering of the L $(L+2)$ dangling ends of an (in)complete row shown at the bottom (top) part of Fig. 2 it is easily seen that the seam position is always immediately to the right of an *even* end in every other row, and to the right of an *odd* end in the remaining rows. Therefore, in all cases the number of permissible seam positions is equal to *half* the number of dangling ends. The total number of extended L-point connectivity states is therefore

$$\frac{Lc_{L/2}}{2} = \frac{L!}{\left(\frac{L}{2}-1\right)!\left(\frac{L}{2}+1\right)!} \sim 2^L. \tag{4.42}$$

Taking into account that for a strip of width L we need intermediate states of $(L+2)$-

Table 7
Number $N^{\mathtt{alg4}}_{N_{\rm mag}}(L)$ of basis states needed by `alg4` to access the scaling exponents of the three-layered system that correspond to the insertion of a magnetisation operator in $N_{\rm mag} = 1, 2, 3$ of the layers

L	$N_{\rm mag} = 1$	$N_{\rm mag} = 2$	$N_{\rm mag} = 3$
2	12	20	20
4	225	600	680
6	5,880	22,344	30,856
8	189,630	930,510	1,565,620
10	6,952,176	41,451,696	83,112,744

point connectivities, the size $N^{\rm alg4}_{N_{\rm mag}}$ of the transfer matrix for three coupled models with $N_{\rm mag} = 1, 2, 3$ layers in the odd sector is as shown in Table 7. Note that for $N_{\rm mag} = 1, 2$ two of the layers are indistinguishable in the sense of Section 4.4.1, whilst for $N_{\rm mag} = 3$ all three layers are indistinguishable.

A very important remark pertains to the proper choice of the initial vector $|\mathbf{v}_0\rangle$ used for finding the largest eigenvalue, cf. Eq. (4.3). Namely, when more than one layer is in the odd sector the seam positions on all layers must coincide at points r_1 and r_2. The reason why this is so can readily be illustrated for the case of two coupled models. With the initial vector (symbolically) chosen as $|\mathbf{v}_0\rangle = |1,1,1,\ldots\rangle \otimes |1,1,1,\ldots\rangle$ one would observe the asymptotic scaling of the correlator

$$\left\langle \sum_{r_1,r_1',r_2,r_2'} M_a^{(1)}(r_1) M_a^{(2)}(r_1') M_a^{(1)}(r_2) M_a^{(2)}(r_2') \right\rangle \sim \left\langle \sum_{r_1,r_2} M_a(r_1) M_a(r_2) \right\rangle^2, \tag{4.43}$$

where the summations run over the L points at each extremity of the cylinder. At large distances $|r_2 - r_1| \gg 1$ one would expect this to scale with the exponent $2x_H^{(1)}$, where $x_H^{(1)}$ is just the scaling dimension of the usual magnetisation operator. However, with $|\mathbf{v}_0\rangle = |1,0,0,\ldots\rangle \otimes |1,0,0,\ldots\rangle$ one would instead observe the scaling of

$$\left\langle \sum_{r_1,r_2} M_a^{(1)}(r_1) M_a^{(2)}(r_1) M_a^{(1)}(r_2) M_a^{(2)}(r_2) \right\rangle, \tag{4.44}$$

and this should decay as $|r_2 - r_1|^{-2x_H^{(2)}}$, where $x_H^{(2)}$ is the sought scaling dimension of the *local* operator $M_a^{(1)}(r) M_a^{(2)}(r)$. Indeed this is the simplest example of a *multiscaling exponent* as discussed by Ludwig [2] in the context of the random-bond Potts model $(N \to 0)$.

The proper choice of $|\mathbf{v}_0\rangle$ is tantamount to anchoring all the seams at the point r_1. When applying Eq. (4.3) one should then theoretically project $\mathbf{T}^{|r_2 - r_1|}|\mathbf{v}_0\rangle$ out on a state with a definite and identical seam position for all the layers before taking the norm. However, since we expect all the non-zero entries of the iterated vector to grow as $\exp\left(|r_2 - r_1|\lambda(L)\right)$ this would just correspond to multiplying by a constant before taking the logarithm. Clearly such a constant would not contribute to the computed value of $\lambda(L)$, and so the projection can be omitted.

5. Numerical results

Using the transfer matrix algorithms just described we are able to find the effective values of the central charge and of the various thermal and magnetic scaling dimensions, all as functions of the coupling constant a, parametrising the position on the self-dual lines identified in Section 2. We are furthermore able to monitor the scale dependence of these quantities by changing the strip width L.

Let us briefly recapitulate the physical information that we hope to extract from these data. First, we should like to provide compelling evidence that, for $N > 2$ coupled Potts models, a novel, non-trivial critical fixed point is located on the self-dual line, in the limit $a \to \infty$. We shall presently describe how such a conclusion may be attained from our numerical data. Second, we aim at fixing the values of the critical exponents at this fixed point. This is done by taking the limit $a \to \infty$ explicitly in the transfer matrices, in order to obtain high-precision data for rather large strip widths (up to $L = 12$ for $N = 3$). Extrapolating these data to the infinite system limit $L \to \infty$ we shall be able to identify the $a \to \infty$ fixed point with the one obtained from perturbative CFT in Section 3, and to associate the numerically obtained scaling dimensions with those of physical operators in the continuum limit. Third, we provide an independent check of the criticality of the fixed point under consideration. This is done by verifying the existence of scaling laws, using Monte Carlo simulations directly in the limit $a \to \infty$. As a by-product we extract values of the scaling dimensions that agree with the more precise estimates obtained from the transfer matrices. And fourth, we use our transfer matrices to inquire further into the structure of the (presently unknown) CFT governing the critical behaviour of the $a \to \infty$ model. To this end we take a closer look at the higher eigenvalues in the even sector. These data determine the scaling dimensions of less relevant operators in the Verma module, and they give crucial information of the operator content and descendence structure of the sought CFT. Finally, a comparison of the results obtained from algorithms `alg3` and `alg4` serves as a check of the modular invariance of the theory.

Although our main interest is $N = 3$ coupled models we have also produced numerical results for the case of two coupled q-state Potts models, for several values of $q \in [2, 4]$. For $q = 2$ this is the Ashkin–Teller model. This model is critical on the entire half line $a \in [0, \infty]$, and it provides a useful check of our algorithms since the critical exponents are known exactly as a function of a [37,34]. For $q > 2$, on the other hand, Vaysburd [19] has predicted a dynamical mass generation and thus a non-critical model. Since this prediction was made under several assumptions it is reassuring to see that our numerics confirms it. At the same time, our $N = 2$ data provide a clear illustration of the difference between a first-order phase transition scenario and a second order one.

5.1. *Finding critical points from effective exponents*

The key to the search for critical fixed points is, of course, Zamolodchikov's c-theorem [8]. This theorem states that the effective central charge (the c-function)

decreases along the RG flow and is stationary at its fixed points. In view of our definition (3.21) of the c-function, this is equivalent to the familiar statement that the β-function vanishes at a fixed point.

A local minimum (maximum) of the effective central charge thus corresponds to a (un)stable fixed point. When, as is the case here, the RG flow takes place in a multi-dimensional space of coupling constants one can also imagine the existence of saddle points, corresponding to a partially stable fixed point. But generically we expect the flow to take place *along* the self-dual line, and accordingly we can limit the search to this one-dimensional self-dual submanifold. This assumption is nicely corraborated by our numerical results. Indeed we find that in general any motion perpendicular to the self-dual line leads to a *decrease* of the central charge (see, e.g., Fig. 3b and Fig. 10 below). Hence, invoking duality, this line must serve as a mountain ridge for the central charge.

A notable exception to this scenario happens for the Ashkin–Teller model, where the self-dual half line $a \in [0, \infty]$ bifurcates into two mutually dual lines of critical points at $a = 0$. However, this anomaly is very clearly signalled by the vanishing of the c-function of the other self-dual half line, $a \in]-\infty, 0[$. We have not observed a similar situation for any other value of q, or of N.

Once a candidate for a fixed point has been identified the question arises whether it is critical or not. To address this point it is useful to examine the dependence of c_{eff} on the system size L. If a critical point is involved, the finite-size estimates $c_{\text{eff}}(L)$ should converge very rapidly, with increasing L, to the true value of the central charge at this point. On the other hand, if the point is not critical a finite correlation length ξ must be involved. As has been shown by Blöte and Nightingale [29] for the particular case of a first-order phase transition, Eq. (4.4) for the finite-size scaling of the free energy must then replaced by

$$f_0(L) = f_0(\infty) - \text{const.}\frac{e^{-L/\xi}}{L^2} + \ldots \tag{5.1}$$

This in turn implies that the effective central charge, i.e. the one measured assuming the scaling form (4.4), will vanish exponentially as a function of the strip width L. We shall soon see that our numerical data distinguish very easily between these two situations.

Complementary information about the location of possible fixed points is provided by the effective values of the thermal and magnetic scaling dimensions $x_{\text{eff}}(L)$ along the self-dual line. In many situations a fixed point is signalled by the fact that the curves $x_{\text{eff}}(L)$ and $x_{\text{eff}}(L+1)$ intersect at some coupling $a(L)$. The finite-size estimates $a(L)$ then converge very rapidly towards the true value of the fixed-point coupling, as $L \to \infty$. Indeed, this is nothing but the well-known technique of phenomenological renormalisation [38]. Evidently a fixed point can also be characterised as a point of high symmetry. It is thus hardly surprising to see from our numerics, that in many cases local extrema of x_{eff} as a function of a serve equally well to locate the position of the fixed point.

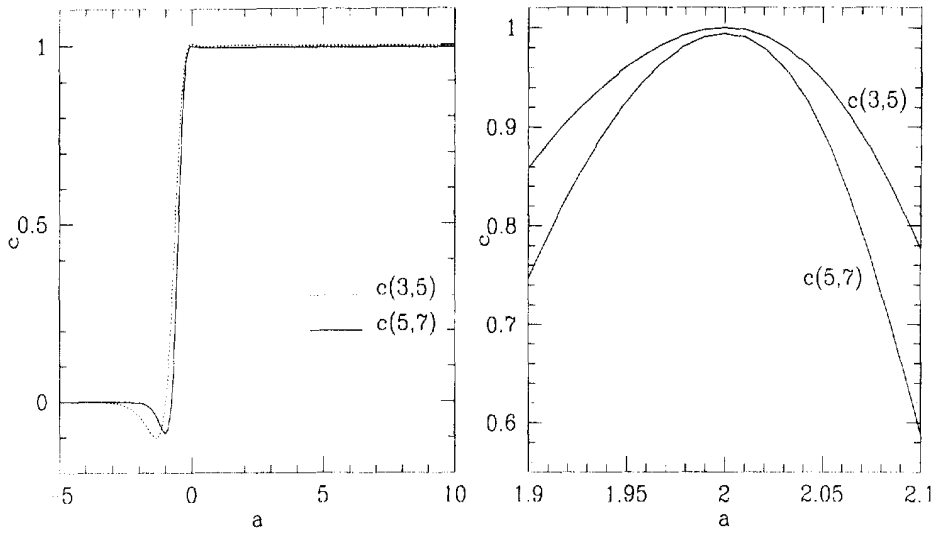


Fig. 3. Central charge for two coupled Ising models. (a) shows the effective central charge as a function of the parameter a. (b) corresponds to moving perpendicularly off the self-dual line.

As was the case for the central charge, once a fixed point has been located its precise nature (critical or not) can be inferred by observing the size-dependence of $x_{\text{eff}}(L)$. Since a finite correlation length implies an exponential, rather than a power law, decay of the various correlators, effective values of the scaling dimensions extracted from Eq. (4.5) will be observed to vanish exponentially with L, if $\xi < \infty$.

5.2. *Two coupled models*

5.2.1. *Ashkin–Teller model*

The case of two coupled Ising models is nothing but the well-known Ashkin–Teller model [39]. We shall nevertheless begin by investigating this case in order to demonstrate the method outlined above, and to see how we can reproduce known results.

As has been shown by Baxter [34] the isotropic Ashkin–Teller model can be mapped to a staggered eight-vertex model. This fact that this model is staggered seems to impede its solvability, but on the self-dual line given by Eq. (2.14), $a \in [-\infty, \infty]$, the eight-vertex model becomes *homogeneous* and hence soluble. For $a \geqslant 0$ the model is critical, with the following exact values of the critical exponents [37]:

$$c = 1, \qquad x_t = \frac{1}{2-y}, \qquad x_H^{(1)} = \tfrac{1}{8}, \qquad x_H^{(2)} = \frac{1}{8-4y}, \tag{5.2}$$

where

$$y = \frac{2}{\pi}\arccos\left(\mathrm{e}^{-a} + \frac{1}{2}\mathrm{e}^{-2a} - \frac{1}{2}\right). \tag{5.3}$$

In Fig. 3a we show the numerical values for the effective central charge, in the form of the three-point fits, cf. Eq. (4.8). For every $a < 0$ the central charge tends to zero in the large-system limit, indicating non-critical behaviour, and for $a \geqslant 0$ it approaches unity,

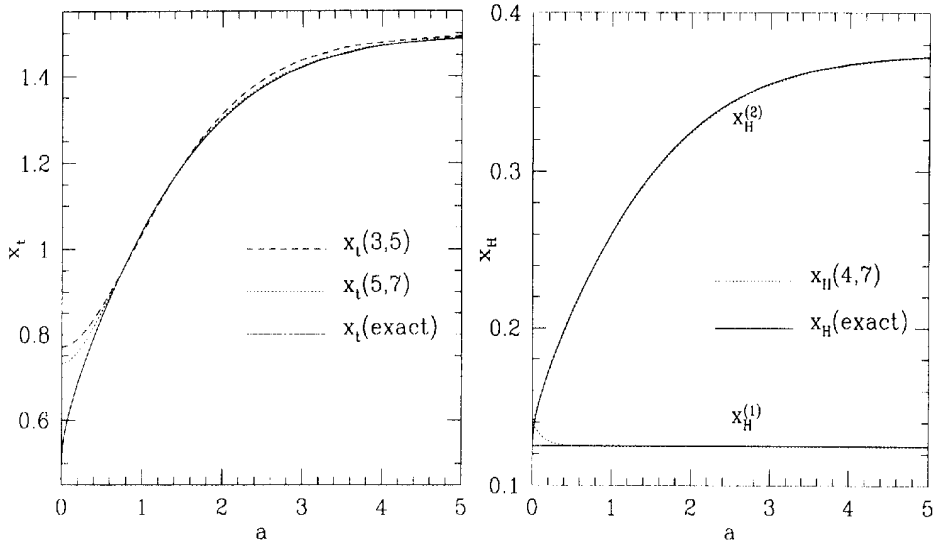


Fig. 4. Thermal dimension x_t and magnetic dimensions $x_H^{(1)}$, $x_H^{(2)}$ for two coupled Ising models.

in complete agreement with Eq. (5.2). The transfer matrix has also been diagonalised directly at $a = \infty$, again with the result $c \simeq 1$.

Fig. 3b depicts the behaviour of the central charge along a line perpendicular to the self-dual line at the point $a = 2$. The value $a = 2$ is chosen arbitrary, and we observe similar curves for other values of $a > 0$. As we move into the non-self-dual regime the central charge decreases, and the larger the strip width the faster the decrease. In the $L \to \infty$ limit we should have $c_{\text{eff}} = 0$ exactly. We can therefore conclude that the self-dual line indeed acts as a ridge of the central charge, thus confirming the hypothesis that the RG flow is along that line. We also observed that for $a < 0$ the critical line bifurcates into two mutually dual Ising-like lines [34].

Next, in Fig. 4, we present measured values of the thermal scaling dimension, x_t, as well as the two magnetic ones, $x_H^{(1)}$ and $x_H^{(2)}$. As the β-function is exactly zero we can expect very accurate values, and as the comparison with the analytical results (5.2) shows this is indeed the case. The only exception is the point $a = 0$, which corresponds to the 4-state Potts model. Due to the presence of a marginal operator there are strong logarithmic corrections [40], and the analytical results $x_t = 1/2$ and $x_H^{(1)} = x_H^{(2)} = 1/8$ are not accurately reproduced. Finally, by taking the explicit $a \to \infty$ limit in the transfer matrices we have verified the results $x_t = 3/2$, $x_H^{(1)} = 1/8$ and $x_H^{(2)} = 3/8$.

5.2.2. *Two coupled 3-state Potts models*

In the case $N = 2$, $q = 3$ there is an $a \leftrightarrow -a$ symmetry at the level of the free energy, as is reflected by the values of the effective central charge shown in Fig. 5. The origin of this symmetry becomes clear if one considers the possible local Boltzmann weigths $W(L_{ij}, a)$ [cf. Eq. (2.32)] associated with the interactions along edge (ij). For two coupled q-state Potts models these are given by

$$W(0,a) = 1, \qquad W(1,a) = \mathrm{e}^{a}, \qquad W(2,a) = \mathrm{e}^{2a+b} = 2\mathrm{e}^{a} + q - 1. \tag{5.4}$$

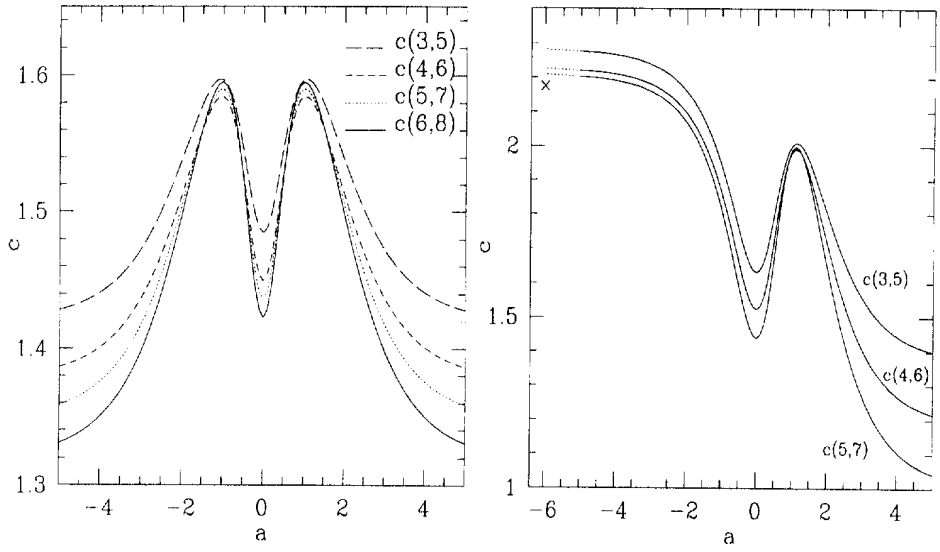


Fig. 5. Central charge for two coupled 3-state Potts models (a) and two coupled 4-state Potts models (b). The dots in (b) correspond to an interpolaton from the measurements made directly at $a = -\infty$ and $a = -5$. We also indicate by a cross a fit at $a = -\infty$ for strip widths L=10–14.

In the transfer matrix, starting from an arbitrary state, the first weight will appear $(q-1)^2$ times, the second one $2(q-1)$ times, and the last one just once. Specialising now to $q = 3$, we see that $W(0,a)$ and $W(1,a)$ have the same degeneracy, and that

$$W(0,-a) = \mathrm{e}^{-a}W(1,a), \quad W(1,-a) = \mathrm{e}^{-a}W(0,a), \quad W(2,-a) = \mathrm{e}^{-a}W(2,a). \tag{5.5}$$

Thus we have established an exact symmetry for the specific free energy on the self-dual line

$$f(-a, q=3) = f(a, q=3) - 2a. \tag{5.6}$$

Returning to the central charge values shown in Fig. 5a we see a local maximum at $a = \log(1+\sqrt{3})$, the fixed point corresponding to two decoupled 3-state Potts models. At this point the finite-size estimates converge very rapidly towards $c = 2 \times \frac{4}{5}$. The other trivial fixed point is located at $a = 0$ and corresponds to a single 9-state Potts model. Here the estimates $c_{\mathrm{eff}}(L)$ decrease steadily with increasing strip width, as expected for a first order phase transition. Any model with a bare coupling $a_0 > \log(1+\sqrt{3})$ renormalises towards the $a \to \infty$ fixed point, but this again corresponds to a first-order transition [19] as witnessed by the steady decrease of the estimates for c.

5.2.3. *Two coupled 4-state Potts models*

The effective central charge for two 4-state models can be inferred from Fig. 5b, and apart from the disappearence of the $a \leftrightarrow -a$ symmetry conclusions are as in the case of two 3-state models. The range of a-values shown on the figure cannot entirely exclude the possibility of a novel fixed point at $a = -\infty$, but the data of Table 8, obtained by

Table 8
Effective exponents for two coupled 4-state Potts models at $a = -\infty$

L	$c(L, L+4)$	$x_t(L, L+2)$
4	2.251	1.032
6	2.231	0.856
8	2.203	0.808
10	2.175	0.762
12		0.720

Table 9
Effective central charge for two coupled q-state Potts models at $a \to \infty$, obtained from alg4

q	$c(4, 8)$	$c(6, 10)$	$c(8, 12)$	$c(10, 14)$	$c(12, 16)$
2.00	0.988504	0.995838	0.998256	0.999142	0.999528
2.25	1.142687	1.149660	1.151260	1.151271	1.150843
2.50	1.263404	1.264818	1.260899	1.255641	1.250201
2.75	1.351006	1.338262	1.319944	1.300188	1.280109
3.00	1.406234	1.367834	1.322338	1.274092	1.224246
3.25	1.430583	1.353605	1.266530	1.174515	1.079463
3.50	1.426433	1.298639	1.157785	1.011642	0.864787
3.75	1.396976	1.208954	1.008126	0.807868	
4.00	1.346005	1.092727	0.833815	0.590962	

using alg4 directly at this point, tell us that if $a = -\infty$ indeed is a fixed point it must be unstable and non-critical.

5.2.4. The point $a = \infty$

As has already been stressed in Section 4 the topological algorithms alg3 and alg4 enable us to treat q, the number of Potts spin states, as a continuously varying parameter. We finish the discussion of the two-models case by presenting some results for the effective exponents for several fractional values of q, obtained directly in the $a \to \infty$ limit.

First, in Table 9 we display the central charge values. For $q = 2$ these converge very nicely towards $c = 1$, the exact Ashkin–Teller value (5.2). For higher q the increasing spacing between succesive estimates signals non-critical behaviour, or in other words, the free energies are not well fitted by Eq. (4.4). It is convenient to have such data at one's disposal, since they facilitate the distinguishing between first and second-order phase transitions in the sequel. Evidently, for $q = 2.25$ it not a priori easy to clearly point out the first-order nature of the transition, since the correlation length is very large. However, it is interesting to notice that whilst for $q = 2$ the estimates *increase* monotonically as a function of L, for all $q > 2$ they eventually begin to *decrease* for large enough L.

Table 10 provides the analogous estimates for the thermal scaling dimension. For $q = 2$ they converge towards $x_t = 3/2$, in accordance with Eq. (5.2). Otherwise their

Table 10
Effective values of the thermal scaling dimension for two coupled q-state Potts models at $a \to \infty$, obtained from alg4

q	$x_t(6,8)$	$x_t(8,10)$	$x_t(10,12)$	$x_t(12,14)$	$x_t(14,16)$
2.00	1.467238	1.475693	1.490527	1.495490	1.497566
2.25	1.745413	1.554317	1.588008	1.605994	1.618971
2.50	1.863461	1.768649	1.702778	1.739814	1.769780
2.75	1.780405	1.717780	1.672429	1.635165	1.603300
3.00	1.707436	1.630297	1.575072	1.532502	1.499100
3.25	1.648514	1.565197	1.510808	1.474971	1.452970
3.50	1.605054	1.524930	1.482309	1.464823	1.466321
3.75	1.576886	1.508815	1.487488	1.497813	
4.00	1.562949	1.514505	1.521838	1.566459	

variation is not monotonic in L, presumably signalling the presence of a finite correlation length.

5.3. Three coupled models

We now turn to our primary interest, namely the case of $N = 3$ coupled Potts models. The self-dual line here starts at the point where $b = -\infty$, that is $a \in [a_{\min}, \infty]$ with

$$a_{\min} = \log\left(\frac{\sqrt{q}+1}{\sqrt{q}+2}\right). \tag{5.7}$$

Unlike the case of two models just discussed we shall find that the $a \to \infty$ limit corresponds to a *critical* fixed point for *all* values of $q \in [2,4]$. The determination of the associated critical exponents thus becomes of paramount interest, and we have dedicated considerable computational effort to this purpose. Let us discuss the various cases in turn.

5.3.1. Three coupled Ising models

This is the three-colour Ashkin–Teller model. The effective central charge is displayed in Fig. 6a. At the trivial fixed point at $a = 0$ we find the usual dip in c, reflecting the first-order nature of the transition in the 8-state Potts model.

For all $a > 0$ the central charge is approximately constant, $c \simeq 3/2$, and we have also verified this value in the limit $a \to \infty$. This situation is very reminiscent of that of the (standard) *two*-colour Ashkin–Teller model, and seems to allow for a situation where the entire half line $a > 0$ consists of critical fixed points along which the critical exponents vary continuously. Whether this is indeed the case will be the subject of a separate publication [41].

Fig. 6b presents a closer look at the antiferromagnetic region $a < 0$. We find here a surprising candidate for an unstable critical fixed point at $a \simeq -0.22$ with a central charge of $c \simeq 1.5$. Further support for this suspicion is found in the maximum for the

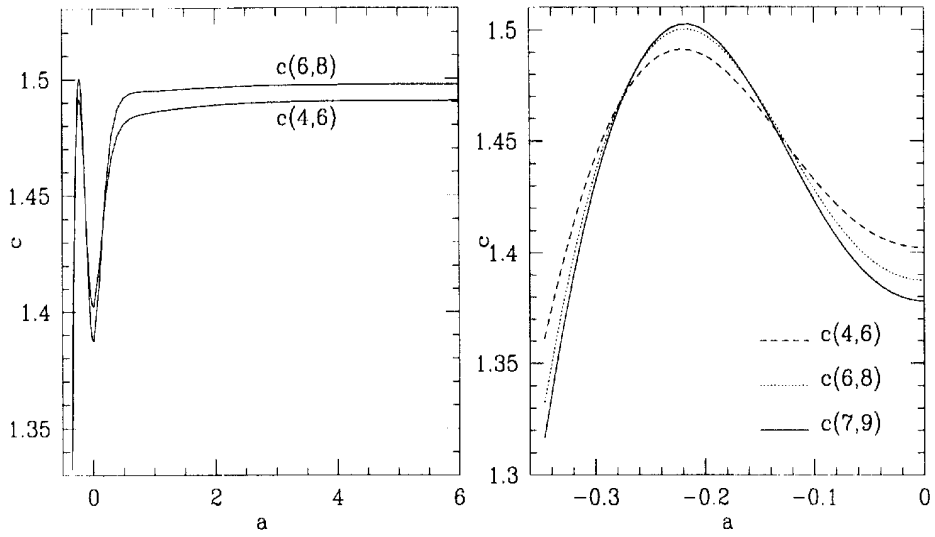


Fig. 6. Central charge for three coupled Ising models.

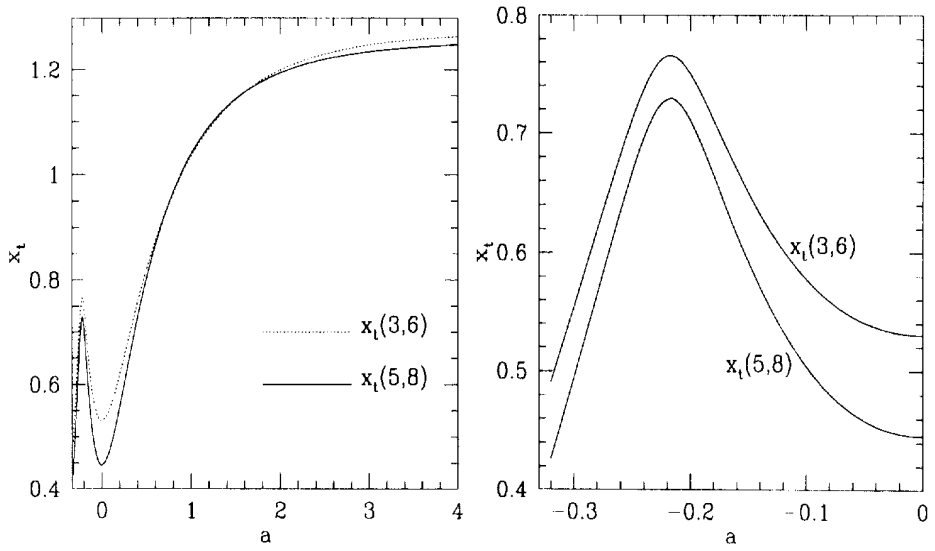


Fig. 7. Thermal exponent for three coupled Ising models.

effective thermal exponent in Fig. 7b, and in the crossing of the curves for the magnetic exponents in Fig. 8b. More results along these lines will be published elsewhere [41].

The results for the thermal exponent on the half line $a > 0$ are shown in Fig. 7a. Everywhere the convergence seems to be very rapid, except at $a = 0$ where we expect an extrapolated effective exponent of $x_t^{\text{eff}} = 0$ due to the first-order transition. In the $a \to \infty$ limit we estimate $x_t = 1.24 \pm 0.01$; more detailed results are provided by Table 12 below.

Finally, in Fig. 8 we present results for the three different magnetic exponents, $x_H^{(1)}$, $x_H^{(2)}$ and $x_H^{(3)}$. The first exponent is constant, $x_H^{(1)} \simeq 1/8$ for $a > 0$, as in the case of two coupled Ising models, whilst the other two depend on a. In the $a \to \infty$ limit we find $x_H^{(2)} \simeq 0.27$ and $x_H^{(3)} \simeq 0.46$ as witnessed by Tables 15 and 16.

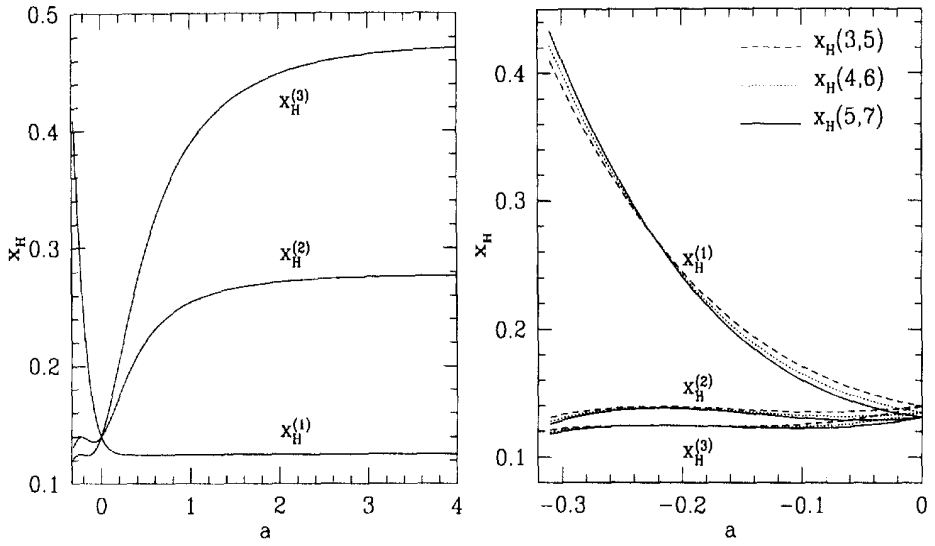


Fig. 8. Magnetic exponents for three coupled Ising models.

5.3.2. Three coupled 3-state Potts models

Until now, the models for which we have presented numerical data are believed to be somewhat atypical representatives for the general class of N coupled q-state Potts models. On one hand, for $N = 2$ coupled models the $\mathcal{O}(g^2)$-term in the β-function (1.11) vanishes, and neighter perturbative RG nor numerics predict any non-trivial critical fixed points for $q \neq 2$. On the other, for $q = 2$ the energy-energy coupling is exactly marginal, leading either to the Ashkin–Teller scenario with an entire line of critical fixed points along which the exponents vary continuously, or simply to a flow back towards the decoupled fixed point. For the case of $N = 3$ coupled 3-state Potts models, which we consider now, the situation can thus be believed to be generic: Perturbation theory predicts a novel critical fixed point, Eq. (1.12), with unique critical exponents, and so does numerics, as we shall presently see.

Consider first the effective central charge along the self-dual line $a \in [a_{\min}, \infty]$, where $a_{\min} = \log[(\sqrt{3}+1)/(\sqrt{3}+2)]$. In Fig. 9a we recognise the familiar structure of a local minimum and a local maximum, signaling the usual two trivial fixed points. However, for larger values of a a novel feature emerges, as witnessed by Fig. 9b. The central charge is here a very slightly decreasing function of a, indicating a flow towards the fixed point at $a = \infty$. Furthermore, the finite-size dependence of the estimates leads us to the conclusion that this point is now *critical*.

The more accurate data of Table 11 corroborate this scenario. First, we see that the finite-size estimates converge very rapidly to the value

$$c = 2.377 \pm 0.003, \tag{5.8}$$

in very good agreement with the perturbative prediction $c_{\mathrm{FP}} = 2.3808 + \mathcal{O}(\epsilon^5)$ from Eq. (3.23).[9] Second, the convergence of the estimates in Table 11 is now from *below*,

[9] In fact, the latter prediction would be slightly *smaller* if we could go to even higher order in perturbation theory, since the series (3.23) is known to be alternating.

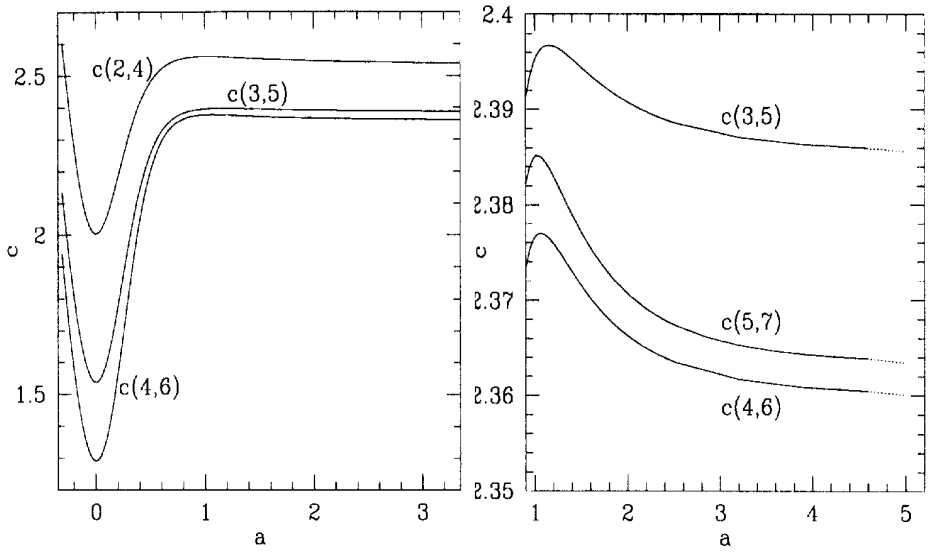


Fig. 9. Central charge for three coupled 3-state Potts models. The dots designate an interpolation from $a = 4.5$ to $a = +\infty$.

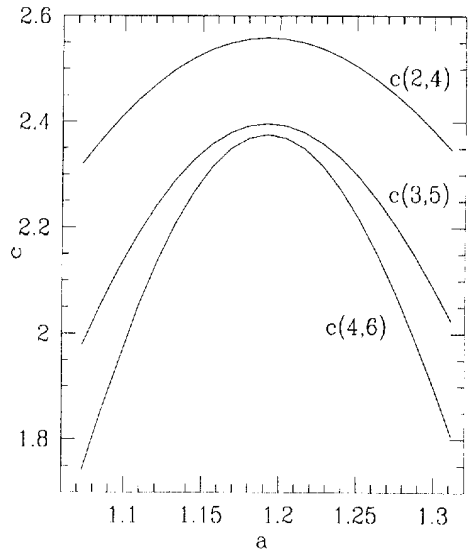


Fig. 10. Central charge for three coupled 3-state Potts models when moving perpendicularly off the self-dual line.

convincingly excluding the possibility that c_{eff} could eventually tend to zero, as would be the case for a non-critical fixed point. Third, the numerical value for c is comfortably away from that of the decoupled fixed point, $c_{\text{pure}} = 3 \times \frac{4}{5} = 2.4$.

In Fig. 10 we show the behaviour of the central charge along a line perpendicular to the self-dual line at the point $a = 1.19217$. Once again, this value of a was chosen arbitrarily; similar curves are observed for other values of $a > 0$.

Plots of the effective thermal and magnetic exponents are shown in Figs. 11a and 12a, respectively. More accurate values obtained directly at the $a = \infty$ critical point are given in Table 12–16.

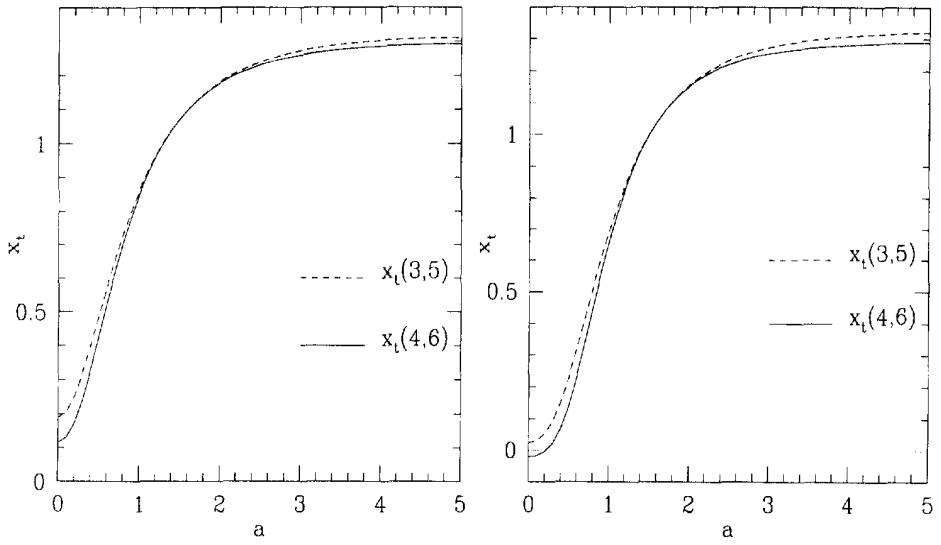


Fig. 11. Thermal exponent for x_t for three coupled q-state Potts models with $q = 3$ and $q = 4$ respectively.

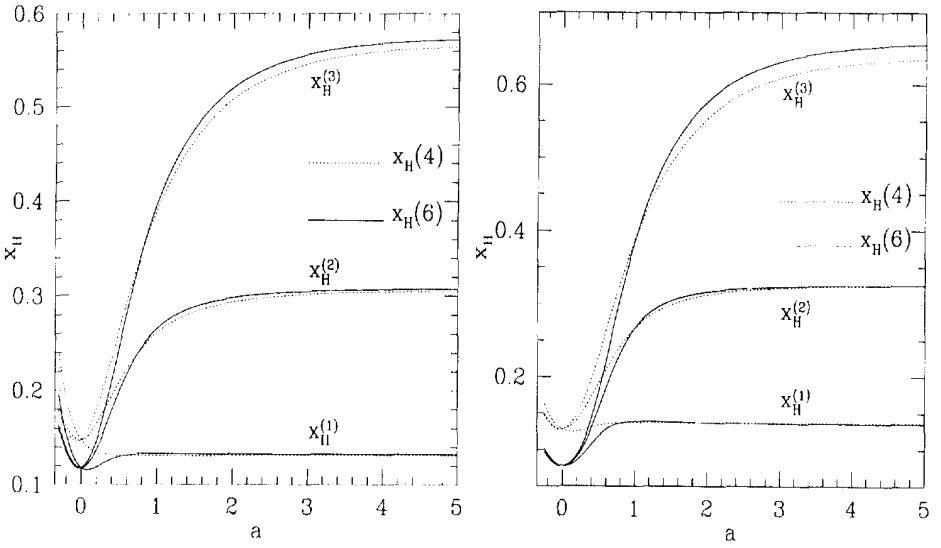


Fig. 12. Magnetic exponents $x_H^{(1)}$, $x_H^{(2)}$ and $x_H^{(3)}$ for three coupled q-state Potts models with $q = 3$ and $q = 4$ respectively.

5.3.3. *Three coupled 4-state Potts models*

The situation for three coupled Potts models is very similar to the case of $q = 3$ treated above. The plots of the effective central charge given in Fig. 13 again give us strong reasons to believe that the $a \to \infty$ limit represents the non-trivial critical fixed point predicted by perturbation theory. Also the effective thermal and magnetic exponents of Figs. 11b and 12b respectively exhibit a structure similar to the $q = 3$ case, but the critical exponents at $a \to \infty$ are of course distinct.

5.3.4. *The $a \to \infty$ limit for general $q \in [2,4]$*

After having numerically identified the non-trivial critical fixed point for three coupled models, that was predicted by the perturbative RG, we would like to accurately determine

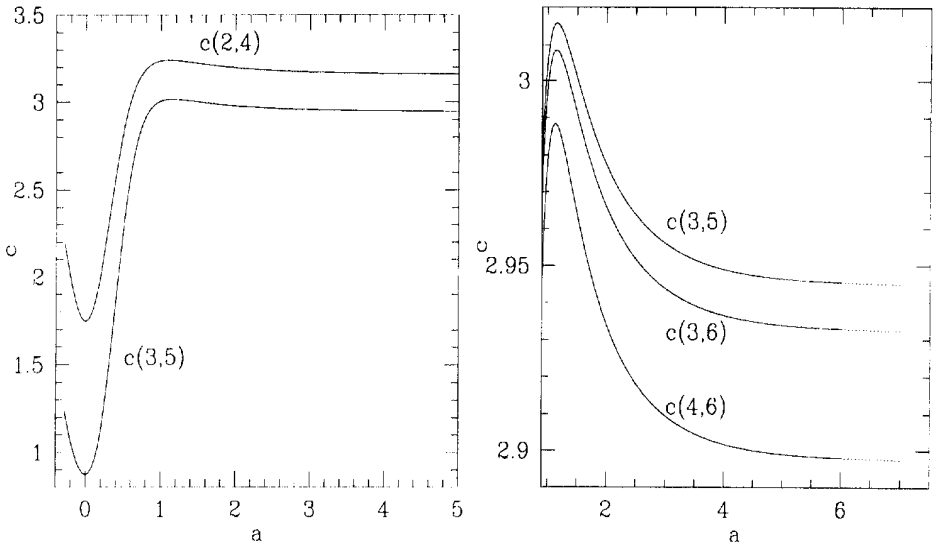


Fig. 13. Central charge for three coupled 4-state Potts models.

Table 11
Results for the central charge of three coupled models at $a = \infty$ and for varying values of q

q	$c(4,8)$	$c(6,10)$	$c(8,12)$
2.00	1.489114	1.497443	1.500008
2.25	1.747538	1.757162	1.760185
2.50	1.976384	1.986719	1.989905
2.75	2.179933	2.190166	2.193005
3.00	2.361502	2.370594	2.372345
3.25	2.523740	2.530453	2.530147
3.50	2.668818	2.671752	2.668217
3.75	2.798557	2.796195	2.788092
4.00	2.914511	2.905265	2.891139

the associated critical exponents as a function of q. This is done by diagonalising the transfer matrix directly at $a = \infty$, using `alg4`.

In Table 11 we show finite-size estimates of the central charge, obtained from three-point fits of the form (4.8). For all $q \in [2,4]$ these estimates seems to converge as $L \to \infty$. Of course, an explicit extrapolation to the infinite system limit from just three data points would easily turn out to be rather subjective. It should however be noted that the convergence is in most cases monotonic, and the last estimate therefore provides a quite accurate upper or lower bound for the extrapolated quantity.

The results for c are in very good agreement with the perturbative RG in the range $q \in [2,3]$. For larger values of q the perturbation theory is expected to break down, and so the numerical values are likely to be more reliable.

The corresponding values of the thermal scaling dimension are given in Table 12. In this case the comparison with the perturbative results is more tricky. We recall that x_t is obtained by fitting the first gap in the even sector of the transfer matrix to Eq. (4.9). This is expected to correspond to the scaling dimension of the most relevant symmetric

Table 12
Results for the thermal scaling dimension of three coupled models at $a = \infty$ using `alg4`

q	$x_t(6,8)$	$x_t(8,10)$	$x_t(10,12)$
2.00	1.257900	1.248706	1.242186
2.25	1.263919	1.255942	1.250623
2.50	1.268895	1.262901	1.259341
2.75	1.272294	1.268558	1.266905
3.00	1.273618	1.271938	1.271901
3.25	1.272423	1.272143	1.272966
3.50	1.286448	1.268344	1.268400
3.75	1.261129	1.260138	1.258753
4.00	1.250663	1.247052	1.242082

Table 13
Results for three models at $a \to \infty$ using `alg3`

q	$c(3,5)$	$c(4,6)$	$x_t(4,5)$	$x_t(5,6)$
2.00	1.5040	1.4906	1.2692	1.2600
2.25	1.7653	1.7490	1.2798	1.2691
2.50	1.9965	1.9775	1.2890	1.2774
2.75	2.2022	2.1801	1.2964	1.2842
3.00	2.3856	2.3601	1.3018	1.2890
3.25	2.5495	2.5198	1.3051	1.2913
3.50	2.6962	2.6615	1.3062	1.2907
3.75	2.8274	2.7869	1.3049	1.2868
4.00	2.9449	2.8975	1.3011	1.2796

energy operator, $\varepsilon_1 + \varepsilon_2 + \varepsilon_3$. However, $\Delta_{\varepsilon_1+\varepsilon_2+\varepsilon_3}$ has only been determined to order $\mathcal{O}(\epsilon^3)$ [see Eq. (3.11)], and so the discrepancy between the values of Table 12 and of Table 1 for large q should hardly come as a surprise. On the other hand, for q slightly larger than two numerics is supposed to be hampered by large logarithmic corrections due to the near-marginality of the energy operator [40]. In this range we thus expect the perturbative results to be the more reliable.

In Table 13 we show some more results for c and x_t, but this time obtained from `alg3`. The usefulness of these data is twofold. First, they clearly demonstrate the superiority of `alg4` as regards the accuracy and the convergence properties (monotonicity) of the finite-size estimates. Second, the agreement with Tables 11 and 12 serves as a check of the modular invariance of the $a = \infty$ critical point, as explained in Section 4.7.

We now turn our attention to the magnetic exponents $x_H^{(N_{\rm mag})}$ obtained by imposing twisted boundary conditions on $N_{\rm mag} = 1, 2$ or 3 of the $N = 3$ coupled Potts models. The numerical results are presented in Table 14–16. Comparing with the analytic results of Table 1–2 we find, as expected, that these scaling exponents must be those of the first three moments of the magnetisation. Excluding the case of $q = 2$ where the remarks made above hold true, we find a very good agreement between $x_H^{(1)}$ and Δ_{σ_1}, and between $x_H^{(2)}$ and $\Delta_{\sigma_1\sigma_2}$ in the range $q \leqslant 3$. For larger q the perturbative RG breaks

Table 14
$x_H^{(1)}$ for three coupled Potts models at $a = \infty$

q	$x_H^{(1)}(4,6)$	$x_H^{(1)}(6,8)$	$x_H^{(1)}(8,10)$
2.00	0.124815	0.125054	0.125112
2.25	0.127712	0.127847	0.127851
2.50	0.129955	0.129915	0.129810
2.75	0.131635	0.131334	0.131050
3.00	0.132825	0.132170	0.131623
3.25	0.133588	0.132480	0.131577
3.50	0.133977	0.132318	0.130962
3.75	0.134040	0.131734	0.129831
4.00	0.133816	0.130778	0.128237

Table 15
$x_H^{(2)}$ for three coupled Potts models at $a = \infty$

q	$x_H^{(2)}(4,6)$	$x_H^{(2)}(6,8)$	$x_H^{(2)}(8,10)$
2.00	0.277424	0.276866	
2.25	0.287411	0.287143	0.286780
2.50	0.296026	0.295985	0.295839
2.75	0.303408	0.303470	0.303445
3.00	0.309671	0.309661	0.309600
3.25	0.314915	0.314615	0.314310
3.50	0.319226	0.318393	0.317599
3.75	0.322685	0.321061	0.319512
4.00	0.325367	0.322693	0.320119

Table 16
$x_H^{(3)}$ for three coupled Potts models at $a = \infty$

q	$x_H^{(3)}(4,6)$	$x_H^{(3)}(6,8)$	$x_H^{(3)}(8,10)$
2.00	0.475424	0.471563	
2.25	0.503049	0.500930	
2.50	0.529688	0.529874	
2.75	0.555511	0.558541	
3.00	0.580638	0.587025	0.592223
3.25	0.605162	0.615388	
3.50	0.629150	0.643668	
3.75	0.652656	0.671885	
4.00	0.675719	0.700047	0.721371

down with a vengeance, as witnessed by its predictions of negative scaling dimension.

The agreement between $x_H^{(3)}$ and $\Delta_{\sigma_1\sigma_2\sigma_3}$ is less good, but this could easily have been anticipated, since we know that the perturbative scheme performs increasingly poorly for the higher moments.

Table 17
Scaling dimensions extracted from the higher gaps in the even sector as well as their identification with physical operators

Gap	$x(4)$	$x(6)$	$x(8)$	$x(10)$	$x(12)$	Extrapolation	Operator
1	1.694	1.471	1.385	1.344	1.322	1.27	$\varepsilon_1+\varepsilon_2+\varepsilon_3$
2	2.603	2.380	2.272	2.219	2.148	≈ 2.0	$T=L_{-2}I$
3	2.922	2.775	2.398	2.225	2.148	≈ 2.0	$\overline{T}=\overline{L}_{-2}I$
4	—	3.104	2.398	2.225	2.174	≈ 2.1	$\varepsilon_1\varepsilon_2+\varepsilon_2\varepsilon_3+\varepsilon_3\varepsilon_1$
5	—	3.104	2.685	2.633	2.599	≈ 2.3	$L_{-1}(\varepsilon_1+\varepsilon_2+\varepsilon_3)$
6	—	3.626	3.279	2.785	2.598	≈ 2.3	$\overline{L}_{-1}(\varepsilon_1+\varepsilon_2+\varepsilon_3)$
7	—	3.991	3.279	2.785	2.598	≈ 2.4	$\varepsilon_1\varepsilon_2\varepsilon_3$
8	—	3.991	3.289	3.146	3.066	≈ 3.0	$T'=L_{-3}I$

5.3.5. Higher exponents in the even sector for $N=3$, $q=3$

In an attempt to numerically identify more of the operators predicted by perturbation theory, we have looked at the scaling dimensions extracted from the finite-size scaling of higher gaps in the even sector of the transfer matrix. These computations are extremely time-consuming, since, in order to study the first k gaps, we need to iterate and orthogonalise $k+1$ vectors, cf. Eq. (4.3). We have therefore focused exclusively on the case of three coupled 3-state Potts models.

The primary operators we expect to find in the even sector must all be energetic, since the $\mathbb{Z}_q$ symmetry associated with the permutation of the Potts spin labels has been factored out. Furthermore, since `alg4` by construction treats all three layers symmetrically, we only expect to find such operators that are symmetric under any permutation of the layer indices. According to Section 3 the permissible operators are thus

$$I, \qquad \varepsilon_1+\varepsilon_2+\varepsilon_3, \qquad \varepsilon_1\varepsilon_2+\varepsilon_2\varepsilon_3+\varepsilon_3\varepsilon_1, \qquad \varepsilon_1\varepsilon_2\varepsilon_3. \tag{5.9}$$

If the (presently unknown) conformal field theory of the system can be assumed to have the usual structure [42] we can, in addition to these primaries, in general expect to find descendent operators with their appropriate degeneracies, reflecting the null vector structure of the Verma module. The identity operator I can be thought of as the primary corresponding to the zeroth gap, thus having scaling dimension $\Delta_I=0$ by definition.

In Table 17 we show the results for the first eight gaps, obtained by employing `alg4`. Since we do not wish to make any a priori assumptions on the existence of the stress tensor we report only the unadorned one-point fits (4.5), for strip widths up to $L=12$. For $L=4$ we were only able to iterate four orthogonal vectors, and consequently there are some empty entries in the table.

The extrapolation to the $L\to\infty$ limit becomes increasingly important for the higher gaps, but unfortunately also increasingly difficult, since the convergence is slower and there are fewer data points. The reported values of the extrapolated scaling dimensions are therefore only indicative.

Let us recall from Section 3 that perturbation theory predicts

$$\Delta_{\varepsilon_1+\varepsilon_2+\varepsilon_3}\simeq 1.44, \qquad \Delta_{\varepsilon_1\varepsilon_2+\varepsilon_2\varepsilon_3+\varepsilon_3\varepsilon_1}\simeq 2.08, \qquad \Delta_{\varepsilon_1\varepsilon_2\varepsilon_3}\simeq 2.28. \tag{5.10}$$

These values cannot be believed to be very precise, since at $q = 3$ we expect the perturbative expansion to be at the verge of breaking down. Nevertheless, they are supposed to be of the right order of magnitude, and we thus venture to make the identification shown in Table 17.

Note that for $L \geqslant 8$ the third and the fourth gap, and also the sixth and the seventh, are exactly degenerate with full 16-digit machine precision. One would therefore expect them to be connected by a simple symmetry, as reflected by our conjectured identification. However, for $L = 12$ a crossover takes place so that henceforth it is the second and third gap that are degenerate rather than the third and the fourth. This could have been anticipated by noticing that for $L < 12$ gaps 2–4 have quite differnet finite-size dependence. A similar crossover is expected to occur amongst gaps 5–7 for $L = 14$. Accordingly, our operator identifications pertain to the expected arrangement of the gaps in the $L \to \infty$ limit.

Apart from the primaries (5.9) we are able to identify the holomorphic and antiholomorphic components of the stress tensor, both with scaling dimension two. This provides further evidence for the criticality of the system, and is completely independent of the methods used this far. In addition we observe the level-one descendents of the operator $\varepsilon_1 + \varepsilon_2 + \varepsilon_3$ with the expected scaling dimension.

Unfortunately we have not been able to iterate enough vectors to see descendents at level two. This is a pity, since the degeneracy of these scaling dimensions would enable us to make predictions about the null vector structure of the Verma module.

5.3.6. Monte Carlo Results

Showing the existence of scaling operators, whose correlators obey scaling laws, would confirm that the $a \to \infty$ limit point on the self-dual line for three coupled models can be identified with a second-order phase transition. To do so, we performed Monte Carlo (MC) simulations with the simplified Boltzmann weights (2.36) pertaining to this point. A direct measure of correlators for the physically significant operators (i.e. those which are diagonal in the perturbation scheme) clearly shows that the model is indeed critical, exhibiting scaling laws with associated exponents sufficiently close to those obtained either by perturbative CFT or transfer matrix techniques. Since the critical exponents can be measured with much more accuracy using transfer matrices techniques, MC was used solely to establish the criticality of the fixed point model.

We chose to measure explicitly the correlation functions for the spin and energy operators. Whilst the definition of energy correlators is straightforward, the $\mathbb{Z}_q$ symmetry of spin variables has to be taken into account, and one should rather consider the operators

$$\Sigma_i \equiv \sigma_i - \overline{\sigma_i} \qquad \mod q, \tag{5.11}$$

where $\overline{\sigma_i}$ is the instantaneous average value of the spin σ on lattice i. The physically meaningful operators, for which we computed correlators, are thus the following:

$$\varepsilon_S \equiv \varepsilon_1 + \varepsilon_2 + \varepsilon_3 \tag{5.12}$$

$$\varepsilon_A \equiv \varepsilon_1 - \varepsilon_2 \tag{5.13}$$

$$\Sigma_S \equiv \Sigma_1 + \Sigma_2 + \Sigma_3 \tag{5.14}$$

$$\Sigma_A \equiv \Sigma_1 - \Sigma_2. \tag{5.15}$$

These operators have well-behaved distributions and are the ones obeying scaling laws. We performed simulations on a 160×160 lattice with periodic boundary conditions in both directions. Sixteen runs of 10,000 sweeps each were made, with a preliminary thermalisation stage of 500 times the autocorrelation time (which is around 25 updates for our lattice size). We also checked that the values of critical exponents were rather close to those predicted using this lattice size. A larger lattice would evidently provide a better description of critical behaviour, but computation time grows rapidly with the number of spins and, as we said, a precise measurement of critical exponents is not what was sought here.

Simulations were performed using a modified version of the Wolff one-cluster algorithm [43]. Let us consider the model given by Eq. (2.16), which we recall:

$$\begin{aligned}\mathcal{H}_{ij} = a\,(\delta_{\sigma_i,\sigma_j} + \delta_{\tau_i,\tau_j} + \delta_{\eta_i,\eta_j}) + b\,(\delta_{\sigma_i,\sigma_j}\delta_{\tau_i,\tau_j} + \delta_{\sigma_i,\sigma_j}\delta_{\eta_i,\eta_j} + \delta_{\tau_i,\tau_j}\delta_{\eta_i,\eta_j}) \\ + c\,\delta_{\sigma_i,\sigma_j}\delta_{\tau_i,\tau_j}\delta_{\eta_i,\eta_j}.\end{aligned} \tag{5.16}$$

We first choose randomly one of the three models, say the one with the indices σ (the model is invariant under permutation of any pair of layers). Then, we consider the model defined by

$$\mathcal{H}_{ij} = (a + b\,(\delta_{\tau_i,\tau_j} + \delta_{\eta_i,\eta_j}) + c\,\delta_{\tau_i,\tau_j}\delta_{\eta_i,\eta_j})\delta_{\sigma_i,\sigma_j} + \text{const.} \tag{5.17}$$

$$= a'\delta_{\sigma_i,\sigma_j} + \text{const.} \tag{5.18}$$

Next we update the lattice with the usual Wolff algorithm but with the local $a' = a + b\,(\delta_{\tau_i,\tau_j} + \delta_{\eta_i,\eta_j}) + c\,\delta_{\tau_i,\tau_j}\delta_{\eta_i,\eta_j}$. Repeated three times, this operation defines one update of the system.

The measurement procedure is quite straightforward: For a given operator $\mathcal{O}$, all correlation functions $G(|x-y|) \equiv \langle \mathcal{O}(x)\mathcal{O}(y)\rangle$, which are manifestly translationally invariant due to the choice of boundary conditions, are computed for $0 < |x-y| < 80$. Compiling results, one finds that the region where scaling laws exist is limited to $|x-y| < 20$ for all operators, except for the symmetric energy, whose scaling law behaviour is observed only for $|x-y| < 8$, this being due to the fact that the critical exponent for this operator is much larger than for the others. This is quite deceiving, but not unexpected. It is clear that for critical correlations to have some sense, they must be between points x and y such that $|x-y| \ll L$. This, combined with the fact that our statistics were rather low, explains why we had to restrict our attention to such narrow regions. Doing so, one observes nice scaling properties with associated exponents close to the ones computed by other analytical and numerical methods. Figs. 14 and 15 present the Log-Log plot of correlation functions along with statistical error bars. Linear fits of these graphs give $2\Delta_{\mathcal{O}}$, and thus leads to the following values for the critical exponents:

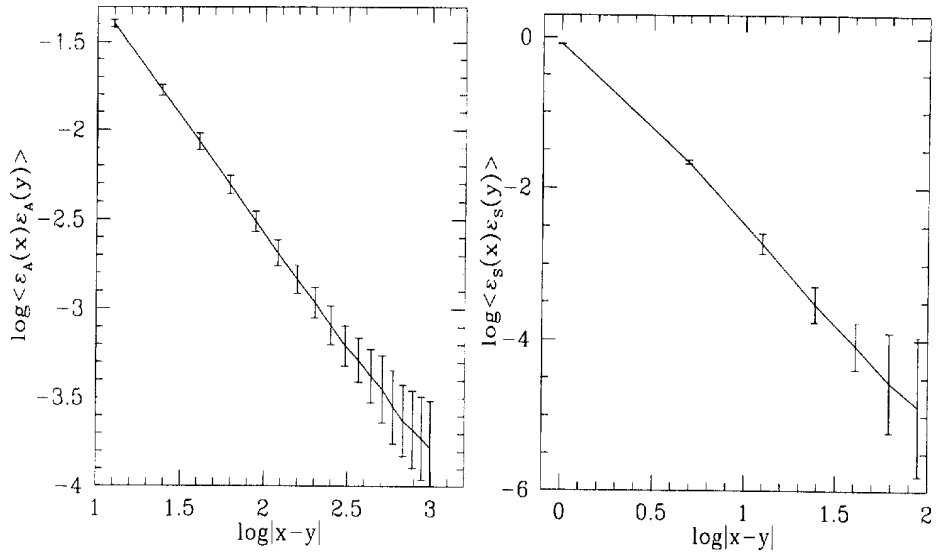


Fig. 14. (a) Monte Carlo results for $\log\langle\varepsilon_A(x)\varepsilon_A(y)\rangle$ as a function of $\log|x-y|$. (b) $\log\langle\varepsilon_S(x)\varepsilon_S(y)\rangle$ as a function of $\log|x-y|$.

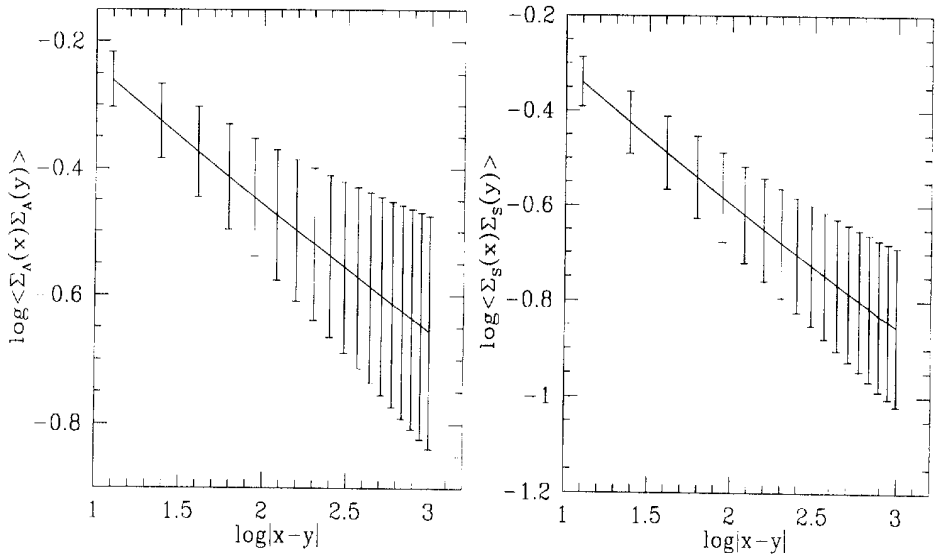


Fig. 15. (a) Monte Carlo results for $\log\langle\Sigma_A(x)\Sigma_A(y)\rangle$ as a function of $\log|x-y|$. (b) $\log\langle\Sigma_S(x)\Sigma_S(y)\rangle$ as a function of $\log|x-y|$.

$$\Delta_{\varepsilon_S} \equiv 1.27 \pm 0.13, \tag{5.19}$$

$$\Delta_{\varepsilon_A} \equiv 0.63 \pm 0.04, \tag{5.20}$$

$$\Delta_{\Sigma_S} \equiv 0.13 \pm 0.03, \tag{5.21}$$

$$\Delta_{\Sigma_A} \equiv 0.10 \pm 0.03. \tag{5.22}$$

Although not precise, the MC data clearly establish the existence of scaling operators. This confirms that the model under study is critical.

6. Concluding remarks

Besides the long-reaching goal of describing exactly disordered systems, the extensive study presented here is, we hope, interesting in many aspects. First, it introduces a new variant of the Potts model transfer matrix method which proves to be the most accurate and efficient up to now. Second, we have identified non-trivial fixed points both numerically and analytically and have presented evidence that these two types of fixed points are indeed the same and correspond to non-trivial critical models. The universality classes of these critical models are new, and several critical exponents were computed for the first time.

The next step in our long-term project is to solve the critical model, which was proven to be the end-point of the self-dual line. At this point, as we have shown, the Hamiltonian simplifies greatly, and it might be possible to map the system to a loop model, whose continuum limit is a Liouville field theory [36]. We believe that the results presented here will be of the outmost importance when searching for such a loop formulation.

Acknowledgements

We would like to thank P. Pujol for valuable discussions. M.-A.L. acknowledges financial support from NSERC Canada.

References

[1] Vik. S. Dotsenko anf Vl. S. Dotsenko, Sov. Phys. JETP Lett. 33 (1981) 37; J. Phys. A 17 (1984) L301.
[2] A.W.W. Ludwig, Nucl. Phys. B 285 (1987) 97; B 330 (1990) 639.
[3] Vl.S. Dotsenko, M. Picco and P. Pujol, Phys. Lett. B 347 (1995) 113, hep-th/9405003; Nucl. Phys. B 455 (1995) 701, hep-th/9501017.
[4] S. Chen, A.M. Ferrenberg and D.P. Landau, Phys. Rev. Lett. 69 (1992) 1213; Phys. Rev. E 52 (1995) 1377.
[5] M. Picco, Phys. Rev. Lett. 79 (1997) 2998, cond-mat/9704221; cond-mat/9802092.
[6] J. Cardy and J.L. Jacobsen, Phys. Rev. Lett. 79 (1997) 4063, cond-mat/9705038;
J.L. Jacobsen and J. Cardy, Nucl. Phys. B 515 [FS] (1998) 701, cond-mat/9711279.
[7] C. Chatelain and B. Berche, Phys. Rev. Lett. 80 (1998) 1670, cond-mat/9801028; Phys. Rev. E (1998) R6899.
[8] A.B. Zamolodchikov, Pis'ma Zh. Eksp. Teor. Fiz. 43 (1986) 565 [English translation in JETP Lett. 43 (1986) 730]; Sov. J. Nucl. Phys. 46 (1987) 1090.
[9] R.J. Baxter, J. Phys. C 6 (1973) L445.
[10] F.Y. Wu and Y.K. Wang, J. Math. Phys. 17 (1976) 439.
[11] E. Domany and E.K. Riedel, Phys. Rev. B 19 (1979) 5817.
[12] J.L. Cardy, J. Phys. A 26 (1996) 1897, cond-mat/9511112.
[13] P. Pujol, Europhys. Lett. 35 (1996) 283, cond-mat/9602116.
[14] P. Simon, Nucl. Phys. B 515 (1998) 624, cond-mat/9710024.
[15] M.-A. Lewis and P. Simon, Phys. Lett. B 435 (1998) 159, cond-mat/9805026.
[16] R.J. Baxter, Phys. Rev. Lett. 26 (1971) 832;
L.P. Kadanoff and F.J. Wegner, Phys. Rev. B 4 (1971) 3989;
F.Y. Wu, Phys. Rev. B 4 (1971) 3212.

[17] E. Fradkin, Phys. Rev. Lett. 53 (1984) 1967.
[18] R. Shankar, Phys. Rev. Lett. 55 (1985) 453.
[19] I. Vaysburd, Nucl. Phys. B 446 (1995) 387.
[20] M.-A. Lewis, Europhys. Lett. 43 (1998) 189, cond-mat/9710312.
[21] Vik.S. Dotsenko, Vl.S. Dotsenko and M. Picco, Nucl. Phys. B 520 (1998) 633, hep-th/9709136.
[22] H.W.J. Blöte and M.P. Nightingale, Physica A 112 (1982) 405.
[23] H. Furstenberg, Trans. Am. Math. Soc. 108 (1963) 377.
[24] G. Benettin, L. Galgani, A. Giorgilli and J.-M. Strelcyn, Meccanica 15 (1980) 9.
[25] M.P. Nightingale, *in* Finite-size scaling and numerical simulations of statistical systems, ed. V. Privman (World Scientific, Singapore, 1990).
[26] H. Blöte, J.L. Cardy and M.P. Nightingale, Phys. Rev. Lett. 56 (1986) 742.
[27] I. Affleck, Phys. Rev. Lett. 56 (1986) 746.
[28] J.L. Cardy, J. Phys. A 16, L355 (1983).
[29] H. Blöte and M.P. Nightingale, Physica A 112 (1982) 405.
[30] J.L. Cardy, *in* Phase Transitions and Critical Phenomena, Vol. 11, ed. C. Domb and J.L. Lebowitz (Academic Press, London, 1987).
[31] F.Y. Wu and H.Y. Huang, Phys. Rev. Lett. 79 (1997) 4954.
[32] I.S. Gradshteyn and I.M. Ryzhik, Table of Integrals, Series and Products, fourth ed. (Academic Press, New York, 1965).
[33] P.W. Kasteleyn, Physica 29 (1963) 1329.
[34] R.J. Baxter, Exactly Solved Models in Statistical Mechanics (Academic Press, New York, 1982).
[35] H.W.J. Blöte and B. Nienhuis, J. Phys. A 22 (1989) 1415.
[36] J.L. Jacobsen and J. Kondev, Nucl. Phys. B 532 [FS] (1998) 635, cond-mat/9804048.
[37] L.P. Kadanoff, Phys. Rev. Lett. 39 (1977) 903; Ann. Phys. (N.Y.) 120 (1979) 39;
L.P. Kadanoff and A.C. Brown, Ann. Phys. (N.Y.) 121 (1979) 318.
[38] M.P. Nightingale, Physica A 83 (1976) 561.
[39] J. Ashkin and E. Teller, Phys. Rev. 64 (1943) 178.
[40] J.L. Cardy, J. Phys. A 19 (1986) L1093.
[41] Vl. Dotsenko, J.L. Jacobsen, M.-A. Lewis and M. Picco, in preparation.
[42] A.A. Belavin, A.M. Polyakov and A.B. Zamolodchikov, Nucl. Phys. B 241 (1984) 333.
[43] U. Wolff, Phys. Rev. Lett. 60 (1988) 1461.

8.5 Article "Universality of coupled Potts models"

Nous étudions des systèmes de M modèles de Potts couplés à travers leur densité locale d'énergie. A chaque modèle est attribué un nombre d'états différent afin de briser la symétrie permutationnelle S_M présente dans le cas des modèles identiques couplés. Nous trouvons les transformations de dualité dans l'espace de $2^M - 1$ couplages entre des multiples opérateurs d'énergie : elles possèdent une forme particulièrement simple. La variété autoduale a la dimension $D_M = 2^{M-1} - 1$. Dans le cas particulier $M = 3$, nous identifions un unique point critique non trivial dans l'espace autodual tridimensionnel. Ses exposants critiques calculés à partir du groupe de renormalisation perturbatif sont comparés à des résultats numériques du type matrice de transfert. Notre but principal est d'apporter des preuves que la symétrie S_3 est restaurée au point critique de trois modèles couplés.

ELSEVIER

Nuclear Physics B 631 [FS] (2002) 426–446

NUCLEAR PHYSICS B

www.elsevier.com/locate/npe

Universality of coupled Potts models

Vladimir S. Dotsenko [a], Jesper Lykke Jacobsen [b], Xuan Son Nguyen [a], Raoul Santachiara [a]

[a] *LPTHE* [1], *Université Pierre et Marie Curie, Paris VI, Université Denis Diderot, Paris VII, Boîte 126, Tour 16, 1er étage, 4 place Jussieu, F-75252 Paris Cedex 05, France*

[b] *Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, Bâtiment 100, F-91405 Orsay, France*

Received 10 December 2001; accepted 11 March 2002

Abstract

We study systems of M Potts models coupled by their local energy density. Each model is taken to have a distinct number of states, and the permutational symmetry S_M present in the case of identical coupled models is thus broken initially. The duality transformations within the space of $2^M - 1$ multi-energy couplings are shown to have a particularly simple form. The selfdual manifold has dimension $D_M = 2^{M-1} - 1$. Specialising to the case $M = 3$, we identify a unique non-trivial critical point in the three-dimensional selfdual space. We compare its critical exponents as computed from the perturbative renormalisation group with numerical transfer matrix results. Our main objective is to provide evidence that at the critical point of three different coupled models the symmetry S_3 is restored.

PACS: 64.60.Fr

1. Introduction

In the study of coupled models, and also of disordered models in their replica formulation, the permutation group symmetry S_M is supposed to play an essential role [1–3]. Namely, the interaction part of the action for a set of M identical coupled models

$$A_{\text{int}} \propto \int \mathrm{d}^2x\, g \sum_{a\neq b} \varepsilon_a(x)\varepsilon_b(x) \tag{1.1}$$

E-mail address: jacobsen@ipno.in2p3.fr (J.L. Jacobsen).

[1] Unité Mixte de Recherche CNRS UMR 7589.

0550-3213/02/$ – see front matter
PII: S0550-3213(02)00213-4

is explicitly invariant with respect to any permutation of the models. Here a and b are replica indices, and $\{\varepsilon_a(x)\}$ designates the set of local energy operators. In the lattice definition of such coupled models the interaction part of the Hamiltonian takes a similar form, with only $\int \mathrm{d}^2x$ replaced by a summation and $\{\varepsilon_a(x)\}$ by an appropriate lattice expression.[2]

When one introduces asymmetric couplings, by generalising the common coupling constant g to a matrix g_{ab},

$$A_{\mathrm{int}} \propto \int \mathrm{d}^2x \sum_{a\neq b} g_{ab}\varepsilon_a(x)\varepsilon_b(x) \tag{1.2}$$

a perturbative renormalisation group (RG) analysis reveals that the S_M symmetry is restored at the fixed point. A detailed study of this scenario, within the ϵ-type perturbative RG for coupled Potts models, was carried out in [4]. Supposing all components of g_{ab} to stay of order ϵ, their initial values being all positive,[3] it was shown in [4] that the only non-trivial fixed point, having one attractive direction, all other directions being repulsive, is that with $g_{ab} \equiv g$.[4] This type of restoration of the symmetry S_M could be called "soft universality".

In this paper we are going to argue for a "strong universality" in the criticality of coupled models. We shall mainly be interested in coupling $M = 3$ different models,[5] via (1.2), with $\{\varepsilon_a\}$ belonging to Potts models with different number of states $\{q_1, q_2, q_3\}$. This breaks the permutational symmetry in a "strong" sense. Still, the RG calculations show the existence of a single fixed point with all $\{g_{ab}\}$ being positive, like it is the case for identical models.

To our knowledge, Simon [5] was the first to apply the RG analysis to a set of different coupled Potts models. The most general model studied by this author was that of M_1 Potts models with q_1 states and M_2 Potts models with q_2 states ($q_1 \neq q_2$), all of them being coupled. After determining the fixed point structure, he computed the dimensions of the spin operators, as well as the RG equations for the energy operators to two-loop order. The effect of disorder on these coupled systems was also analysed.

Here we generalise the RG calculations of [5] to the case of three different coupled Potts models ($q_1 \neq q_2 \neq q_3$). We shall compute the dimensions of energy operators, with a special focus on the symmetry of the theory, at the non-trivial fixed point that generalises the one found in [2] for three identical models. Within the space of couplings $\{g_{ab}\}$, the new fixed point is stable in one direction and unstable in the others, the topology of the RG flows

[2] In the case of the Potts model with nearest-neighbour spins $\sigma_i^{(a)}$ and $\sigma_j^{(a)}$ in the replica a, one has $\varepsilon_a(x) \sim 1 - \delta\big(\sigma_i^{(a)}, \sigma_j^{(b)}\big)$.

[3] In [4] the interaction part of the action A_{int} (1.2) was actually defined with an extra minus sign, so that initially all the components of g_{ab} were taken to be negative.

[4] Ref. [4] also identified other fixed points, which are related to the basic one (with all g_{ab} equal and of the same sign) by changing the sign of some of its components. This is equivalent to switching the sign of a certain number of energy operators. This last operation is a symmetry of the individual Potts models, corresponding to their selfduality (note however such duality transformations on the individual models are not directly related to the global duality transformations on the entire coupled system to be discussed in Section 3). This argument implies that the critical properties of the various critical points classified in [4] should be equivalent.

[5] The duality transformations of Section 3 are however valid for general M.

being similar to those of [2]. But there is also one apparent difference: the permutational symmetry has disappeared, the coupled models being different.

The purpose of this paper will be to provide evidence, by using various methods, that at the fixed point of three different coupled models the apparently lost symmetry S_3 is restored, implying a "strong universality".

This symmetry restoration cannot be observed on the level of the initial action (1.2), as discussed above, nor is it visible in the perturbative RG treatment, or in the Hamiltonian of the explicit lattice realisation. This is because $\{\varepsilon_a\}$ are the energy operators of different models, with different scaling dimensions in particular.

The way we shall check for the restoration of the symmetry is by looking at the spectrum of scaling dimensions at the new fixed point, within the sector of energy operators. Like in the case of identical models [1–3], the RG analysis implies that the three energy operators of the decoupled models $\{\varepsilon_1(x), \varepsilon_2(x), \varepsilon_3(x)\}$ will rearrange as three particular linear combinations so as to form the new primary operators at the fixed point of the coupled models.

In the case of identical models the corresponding linear combinations are easy to guess on symmetry grounds. The irreducible representations (irreps) of the group S_3 in the basis $\{\varepsilon_1(x), \varepsilon_2(x), \varepsilon_3(x)\}$ consist of a (symmetric) singlet

$$\varepsilon_{\mathrm{S}}(x) = \varepsilon_1(x) + \varepsilon_2(x) + \varepsilon_3(x) \tag{1.3}$$

and an (antisymmetric) doublet

$$\begin{cases} \varepsilon_{\mathrm{A}_1} = \varepsilon_1(x) - \varepsilon_2(x), \\ \varepsilon_{\mathrm{A}_2} = \varepsilon_1(x) - \varepsilon_3(x) \end{cases} \tag{1.4}$$

that act as the new primary operators at the fixed point [1–3]. The fact that the operators $\varepsilon_{\mathrm{A}_1}$ and $\varepsilon_{\mathrm{A}_2}$ belong to the same two-dimensional irrep means that their dimensions must coincide: $\Delta(\varepsilon_{\mathrm{A}_1}) = \Delta(\varepsilon_{\mathrm{A}_2})$. These are however in general different from the dimension $\Delta(\varepsilon_{\mathrm{S}})$ of the one-dimensional irrep.

When coupling different models, the corresponding linear combinations of $\varepsilon_1, \varepsilon_2, \varepsilon_3$ will have more complicated coefficients which have to be calculated by the RG technique. One will find something of the form

$$\begin{cases} \varepsilon_1^* = K_{11}\varepsilon_1 + K_{12}\varepsilon_2 + K_{13}\varepsilon_3, \\ \varepsilon_2^* = K_{21}\varepsilon_1 + K_{22}\varepsilon_2 + K_{23}\varepsilon_3, \\ \varepsilon_3^* = K_{31}\varepsilon_1 + K_{32}\varepsilon_2 + K_{33}\varepsilon_3. \end{cases} \tag{1.5}$$

Since the initial dimensions $\Delta(\varepsilon_1)$, $\Delta(\varepsilon_2)$, $\Delta(\varepsilon_3)$ of the decoupled models differ, one may expect that the critical dimensions (RG eigenvalues) of the newly formed primary operators ε_1^*, ε_2^*, ε_3^* might all be different.

Our argument is that, in the case of coupling different models, it is not the linear combinations (1.5) which have to be examined to analyse the symmetry, but rather the spectrum of their critical dimensions. Permuting ε_1, ε_2, ε_3 in the combinations (1.5) does not make much sense because they are different. To permute ε_1^*, ε_2^*, ε_3^*, one first has to know their properties, their scaling dimensions, to decide if it makes sense or not.

The conclusion will be that one has to study the spectrum of dimensions at the new fixed point. This provides a representation independent information, independent of the way one has defined the theory initially, as by its action (Hamiltonian) in Eq. (1.2).

If the symmetry S_3 is restored then the dimensions

$$\Delta(\varepsilon_1^*),\ \Delta(\varepsilon_2^*),\ \Delta(\varepsilon_3^*) \tag{1.6}$$

should form a singlet and a doublet, as is the case when one couples initially identical models.

The rest of the paper is organised as follows.

In Section 2 we analyse the case $M = 3$ by perturbative RG calculations. As will be explained towards the end of that section, in this case the RG results alone are not sufficient to decide whether the S_3 symmetry is restored or not. We therefore propose to study the coupled models through a particular lattice realisation, following [2]. To that end, in Section 3, we extend the duality transformations derived in [6] for symmetrically coupled models to the asymmetric case. The resulting relations are valid for any number M of coupled models, and for the most general asymmetric multi-energy couplings. In particular, we determine the selfdual solutions, thus simplifying dramatically the subsequent numerical simulations.

After recalling briefly, in Section 4, the most efficient transfer matrix algorithm constructed in [2], we show how it may be generalised to the case of asymmetrically coupled models. We then turn to the numerical analysis in Section 5. We locate the non-trivial critical point on the selfdual manifold for various systems of $M = 3$ coupled models, and we provide accurate values of the central charge and the energetic scaling dimensions. We shall find strong evidence for a singlet/doublet spectrum at the new fixed point. This we interpret as a signal of the restoration of the S_3 symmetry.

Finally, Section 6 is devoted to remarks and conclusions.

2. Renormalization group analysis

In the continuum limit, the three coupled Potts models can be represented by the action

$$A = \sum_{a=1}^{3} A_0^{(a)} + A_{\text{int}}, \tag{2.1}$$

$$A_{\text{int}} = \int \mathrm{d}^2x \sum_{a\neq b}^{3} g_{ab}^0 \varepsilon_a(x)\varepsilon_b(x). \tag{2.2}$$

Here $\{A_0^{(a)}\}$ represents three decoupled models, or more precisely the three decoupled conformal field theories for these models at their respective critical points. The interaction term A_{int} makes them coupled. In general, the initial couplings $\{g_{ab}^0\}$ are taken to be different. Since $\{\varepsilon_a\}$ have different dimensions when the models are different, even if we start from identical couplings $g_{ab}^0 = g_0$, they will become different in the course of renormalisation.

We shall parametrise the dimensions of the energy operators $\{\varepsilon_a\}$ as in the papers [2,4, 5,7]

$$\left(\Delta_{\varepsilon}^{(a)}\right)_{\text{phys}} = 1 - \frac{3}{2}\epsilon_a, \tag{2.3}$$

where the physical dimension $(\Delta_{\varepsilon}^{(a)})_{\text{phys}}$ corresponds to twice the conformal dimension, as usual. The quantity ϵ_a, which measures the deviation from the Ising model, appears also in the Coulomb gas parameter

$$\left(\alpha_{+}^{(a)}\right)^2 = \frac{4}{3} - \epsilon_a. \tag{2.4}$$

For the Ising model one has $\epsilon = 0$, whence $\alpha_{+}^2 = 4/3$ and $(\Delta_{\varepsilon})_{\text{phys}} = 1$. The parameter α_{+}^2 is useful in analytic calculations because it is simpler to use than the central charge of the corresponding conformal theory.

The details of the RG calculations are the same as in Refs. [5,7]. In particular, all the necessary integrals have been calculated in these papers. The generalisation to the case of three different models is straightforward, and we shall therefore only give the final results.

It turns out to be convenient to redefine the expansion parameters $\{\epsilon_a\}$ and the coupling constants $\{g_{ab}\}$ as follows:

$$\frac{3}{2}\epsilon_a = \tilde{\epsilon}_a, \qquad g_{ab} = \frac{1}{4\pi}\tilde{g}_{ab}. \tag{2.5}$$

From now on we shall adopt this convention, omitting the tildes throughout for simplicity of notation.

The β-functions for the interaction (2.2) are found in the form (to two-loop order)

$$\begin{cases} \frac{\mathrm{d}g_{12}}{\mathrm{d}\xi} \equiv \beta_{12}\left(\{g_{ab}\}\right) = \epsilon_{12}g_{12} - g_{13}g_{23} - \frac{1}{2}g_{12}\left(g_{13}^2 + g_{23}^2\right), \\ \frac{\mathrm{d}g_{13}}{\mathrm{d}\xi} \equiv \beta_{13}\left(\{g_{ab}\}\right) = \epsilon_{13}g_{13} - g_{12}g_{32} - \frac{1}{2}g_{13}\left(g_{12}^2 + g_{32}^2\right), \\ \frac{\mathrm{d}g_{23}}{\mathrm{d}\xi} \equiv \beta_{23}\left(\{g_{ab}\}\right) = \epsilon_{23}g_{23} - g_{21}g_{31} - \frac{1}{2}g_{23}\left(g_{12}^2 + g_{31}^2\right). \end{cases} \tag{2.6}$$

Here ξ is the RG parameter, $g_{ab} = g_{ba}$ by symmetry of the action (2.2), and we have defined $\epsilon_{ab} \equiv \epsilon_a + \epsilon_b$.

The renormalisation of the energy operators, to second order, is found to be given by the equations

$$\frac{\mathrm{d}\varepsilon_a}{\mathrm{d}\xi} = -(1-\epsilon_a)\varepsilon_a - \sum_{b\neq a} g_{ab}\varepsilon_b - \frac{1}{2}\left(\sum_{d\neq a}(g_{ad})^2\right)\varepsilon_a. \tag{2.7}$$

In matrix form this reads

$$\frac{\mathrm{d}\varepsilon_a}{\mathrm{d}\xi} = -\sum_{b=1}^{3} \Delta_{ab}\varepsilon_b, \tag{2.8}$$

$$\Delta_{ab} = (1-\epsilon_a)\delta_{ab} - \gamma_{ab}\left(\{g_{cd}\}\right), \tag{2.9}$$

$$\gamma_{ab} = \begin{pmatrix} -\frac{1}{2}\left(g_{12}^2 + g_{13}^2\right) & -g_{12} & -g_{13} \\ -g_{12} & -\frac{1}{2}\left(g_{21}^2 + g_{23}^2\right) & -g_{23} \\ -g_{13} & -g_{23} & -\frac{1}{2}\left(g_{31}^2 + g_{32}^2\right) \end{pmatrix}. \tag{2.10}$$

To define the new primary operators and their scaling dimensions at the new fixed point we shall have to diagonalise the matrix Δ_{ab}. It is however more convenient to regroup the terms $-\epsilon_a\delta_{ab} - \gamma_{ab}$ so that

$$\Delta_{ab} = \delta_{ab} - \Lambda_{ab}, \tag{2.11}$$

$$\Lambda_{ab} = \begin{pmatrix} \epsilon_1 - \frac{1}{2}\big(g_{12}^2 + g_{13}^2\big) & -g_{12} & -g_{13} \\ -g_{12} & \epsilon_2 - \frac{1}{2}\big(g_{21}^2 + g_{23}^2\big) & -g_{23} \\ -g_{13} & -g_{23} & \epsilon_3 - \frac{1}{2}\big(g_{31}^2 + g_{32}^2\big) \end{pmatrix} \tag{2.12}$$

and we need only diagonalise the matrix Λ_{ab}.

The non-trivial zeros of the β-functions (2.6) are found as

$$\begin{cases} g_{12}^* = \sqrt{\epsilon_{13}\epsilon_{23}}\big\{1 - \frac{1}{2}\epsilon_{12} - \frac{1}{4}(\epsilon_{13} + \epsilon_{23})\big\}, \\ g_{13}^* = \sqrt{\epsilon_{12}\epsilon_{23}}\big\{1 - \frac{1}{2}\epsilon_{13} - \frac{1}{4}(\epsilon_{12} + \epsilon_{23})\big\}, \\ g_{23}^* = \sqrt{\epsilon_{12}\epsilon_{13}}\big\{1 - \frac{1}{2}\epsilon_{23} - \frac{1}{4}(\epsilon_{12} + \epsilon_{13})\big\}. \end{cases} \tag{2.13}$$

They correspond to the non-trivial fixed point of the coupled models which we are interested in. It is readily checked that the fixed point (2.13) is stable in one direction and unstable in the two others, as is the case when one couples identical models. Furthermore, the topology of the RG flows is similar.

Substituting the fixed point values of the couplings in the matrix Λ_{ab} and diagonalising it, one obtains, after some algebra, the following expressions for the eigenvalues:

$$\lambda_1 = -\frac{a+b+c}{2} + \frac{3abc}{a+b+c}, \tag{2.14}$$

$$\begin{aligned} \lambda_{2,3} = {} & \frac{a+b+c}{2} - \frac{6abc + a^2b + ab^2 + a^2c + ac^2 + b^2c + bc^2}{2(a+b+c)} \\ & \pm \frac{1}{2(a+b+c)}\big\{6a^2b^2c^2 \\ & - 2abc\big(a^2b + ab^2 + a^2c + ac^2 + b^2c + bc^2 + a^3 + b^3 + c^3\big) \\ & + a^2b^2(a+b)^2 + a^2c^2(a+c)^2 + b^2c^2(b+c)^2\big\}^{1/2}, \end{aligned} \tag{2.15}$$

where we have simplified the notation by means of the abbreviations $a = \epsilon_{12}$, $b = \epsilon_{13}$ and $c = \epsilon_{23}$. According to the definition (2.11) of Λ_{ab}, the dimensions of the new primary operators $\{\varepsilon_a^*\}$, cf. Eq. (1.5), are related to the above eigenvalues through

$$\Delta\big(\varepsilon_a^*\big) = 1 - \lambda_a \tag{2.16}$$

for $a = 1, 2, 3$. We recall that all these calculations have been done to second order in ϵ.

It should be remarked that in the special case of two of the parameters being equal, $b = c$ and $a \neq b$, the expressions for the dimensions simplify considerably. One finds

$$\Delta\big(\varepsilon_1^*\big) = 1 + \frac{a+2b}{2} - \frac{3ab^2}{a+2b}, \tag{2.17}$$

$$\Delta\big(\varepsilon_2^*, \varepsilon_3^*\big) = 1 - \frac{a+2b}{2} + \frac{b(4ab + a^2 + b^2)}{a+2b} \pm \frac{b^2(a-b)}{a+2b}. \tag{2.18}$$

At first order in ϵ, the RG results for the dimensions

$$\Delta(\varepsilon_1^*) = 1 + \frac{a+b+c}{2}, \tag{2.19}$$

$$\Delta(\varepsilon_2^*) = \Delta(\varepsilon_3^*) = 1 - \frac{a+b+c}{2} \tag{2.20}$$

form a singlet and a doublet, in accordance with the scenario for the restoration of the S_3 symmetry discussed in the introduction. The degeneracy of the doublet is, however, lifted at order ϵ^2, by the last term in (2.15). This is true even when two of the models are identical, and only for $a = b = c$ does one recover the degeneracy $\Delta(\varepsilon_2^*) = \Delta(\varepsilon_3^*)$ [2].

In Section 5, where the results of our numerical work are presented, we shall show that this last result of the perturbative RG is wrong.[6] We provide evidence that the splitting between the dimensions of ε_2^* and ε_3^* is actually zero, and that the symmetry restoration scenario therefore holds true.

We could suggest the following argument for the failure of the perturbative RG. The ϵ-expansion calculations are valid for perturbed conformal theories because they respect the conformal symmetry, just as dimensional regularisation is valid in the context of pertubative calculations in a gauge theory because the method respects the gauge symmetry (otherwise the results would be dependent on the regularisation technique). In the present problem, the ϵ-regularisation should be correct as far as the conformal symmetry alone is concerned. But in case of extra symmetries, such as S_3, the method might well give wrong results.

The RG formulae (2.14)–(2.16) for the scaling dimensions, and for the central charge which will be given below, are still quite useful. Apart from the degeneracy issue just discussed, they compare well with the numerical results of Section 5 when $\{\epsilon_a\}$ are small enough.

As far as the dimensions $\Delta(\varepsilon_2^*)$, $\Delta(\varepsilon_3^*)$, defined by Eqs. (2.15) and (2.16), are concerned, it is their mean value which compares well with numerical results. It is worth noticing that the splitting, namely, the third term in (2.15), is numerically smaller than the principal ϵ^2 term, the second term in (2.15). This is because of the compensation of negative and positive terms in the third term of (2.15).[7]

The central charge can be obtained in a simple way using Zamolodchikov's c-theorem [8]. This theorem provides us with a function of the couplings $c(g_{12}, g_{13}, g_{23})$ which decreases along the renormalisation flow and takes a value $c(g_{12}^*, g_{13}^*, g_{23}^*)$ at the fixed point of the flow which equals the central charge of the associated conformal field theory.

With the conventions (2.5) taken into account, the c-function is uniquely determined by

$$\frac{\partial c(g_{12}, g_{13}, g_{23})}{\partial g_{ab}} = -\frac{3}{2}\beta_{ab}(g_{12}, g_{13}, g_{23}), \tag{2.21}$$

$$c(0,0,0) = c_0 \equiv c_1 + c_2 + c_3, \tag{2.22}$$

where c_0 is the total central charge of three decoupled models.

[6] We would still claim, of course, that our RG calculations are technically correct!

[7] It is easy to check that the expression under the square-root sign of (2.15) is always non-negative, so that $\lambda_{2,3}$ are well defined.

From Eq. (2.6) and Eqs. (2.21), (2.22) the c-function turns out to be given by

$$c(g_{12}, g_{13}, g_{23}) = c_0 - \frac{3}{4}\left(ag_{12}^2 + bg_{13}^2 + cg_{23}^2\right) + \frac{3}{2}g_{12}g_{13}g_{23} + \frac{3}{8}\left(g_{12}^2 g_{13}^2 + g_{12}^2 g_{23}^2 + g_{13}^2 g_{23}^2\right). \tag{2.23}$$

At the fi xed point we insert Eq. (2.13) into Eq. (2.23); up to the order ϵ^4 the correction Δc of the central charge is

$$\Delta c = -\frac{3}{4}abc + \frac{3}{8}\left(abc^2 + bca^2 + cab^2\right). \tag{2.24}$$

3. Duality transformations

The possibility of endowing the Potts model with a duality transformation was one of the main motivations for introducing it [9]. The study of duality for several Potts models coupled by their local energy density was initiated by Domany and Riedel, who worked out the case of two models with q_1 and q_2 states [10]. Technically, these authors used a particular version of the method of lattice Fourier transforms [11], due to Wu and Wang [12]. Dotsenko et al. generalised the computations to three symmetrically coupled q-state models [2]. However, it became clear that the complexity of the Fourier method grew rapidly as the number of models to be coupled was increased. Recently, it was pointed out by Jacobsen that trading the Potts spin variables for a formulation in terms of random clusters [13], or loops [14], the duality relations simplifi ed dramatically. This observation made it possible to work out the case of M coupled q-state Potts models, with the most general coupling by local energy densities consistent with an S_M symmetry [6].

We now show how to treat the even more general case, where each model does not necessarily have the same number of states, and the coupling by local energy densities is the most general one.

3.1. General case

Consider a set $\mathcal{L}^M$ of M identical planar lattices $\mathcal{L}$, which we imagine to be stacked on top of one another. On each lattice site $i \in \mathcal{L}$, and for each layer $\mu = 1, 2, \ldots, M$, we defi ne a Potts spin $\sigma_i^{(\mu)}$ taking the values $\sigma_i^{(\mu)} = 1, 2, 3, \ldots, q_\mu$. The layers interact by means of the reduced Hamiltonian

$$\mathcal{H} = \sum_{\langle ij \rangle} \mathcal{H}_{ij}, \tag{3.1}$$

where $\langle ij \rangle$ denotes the set of lattice edges, and the nearest-neighbour interaction is defi ned as

$$\mathcal{H}_{ij} = -\sum_{\ell \subset \mathcal{E}} K_\ell \prod_{\mu \in \ell} \delta\left(\sigma_i^{(\mu)}, \sigma_j^{(\mu)}\right). \tag{3.2}$$

By definition, the Kronecker delta function $\delta(x, y) = 1$ if $x = y$, and zero otherwise. We have here defined $\mathcal{E}$ as the product set $\prod_{\mu=1}^{M}\{\varnothing, \mathcal{L}^{(\mu)}\}$, so any one of the 2^M subsets in $\mathcal{E}$ can be interpreted as a certain subset of layer indices. One can also think of $\ell \subset \mathcal{E}$ as specifying the *state* of a given edge $\langle ij\rangle$, meaning that in the interaction term (3.2), ℓ determines which of the layers contribute to the product $\prod_{\mu\in\ell}$ of the corresponding delta functions. The layers specified by some $\ell \subset \mathcal{E}$ then interact by means of the product of their local energy densities, through a coupling constant K_ℓ.

For later convenience, we shall represent a subset $\ell \subset \mathcal{E}$ as a list of M open ($\circ$) or filled ($\bullet$) circles, the μth circle indicating, respectively, the absence (or presence) of the factor $\delta\big(\sigma_i^{(\mu)}, \sigma_j^{(\mu)}\big)$. Furthermore, we sometimes use a single open (respectively, filled) circle as an abbreviation of a list of M open (respectively, filled) circles.

When $M = 1$, the model defined by Eq. (3.2) reduces to the conventional Potts model, whilst for $M = 2$ it is identical to the Ashkin–Teller like model considered in Ref. [10]. For $M = 3$, it is the asymmetric version of the model discussed in Ref. [2], which had $q \equiv q_1 = q_2 = q_3$, and whose couplings possessed an S_3 symmetry upon permutation of the layers. Using the above symbolic notation, this symmetry can be expressed by the identities

$$\begin{aligned} &K_1 \equiv K_{\bullet\circ\circ} = K_{\circ\bullet\circ} = K_{\circ\circ\bullet},\\ &K_2 \equiv K_{\circ\bullet\bullet} = K_{\bullet\circ\bullet} = K_{\bullet\bullet\circ},\\ &K_3 \equiv K_{\bullet\bullet\bullet}. \end{aligned} \tag{3.3}$$

By means of a generalised Kasteleyn–Fortuin transformation [13] the local Boltzmann weights can be recast as

$$\exp(-\mathcal{H}_{ij}) = \prod_{\ell\subset\mathcal{E}}\left[1 + \big(\mathrm{e}^{K_\ell} - 1\big)\prod_{\mu\in\ell}\delta\big(\sigma_i^{(\mu)}, \sigma_j^{(\mu)}\big)\right], \tag{3.4}$$

where we have simply used that $\exp(K\delta) = 1 + [\exp(K) - 1]\delta$, if δ can only take the values 0 and 1. Since furthermore $\delta^2 = \delta$, expanding the product over ℓ will lead to the equivalent form

$$\exp(-\mathcal{H}_{ij}) = b_\circ + \sum_{\ell\subset\mathcal{E}} b_\ell \prod_{\mu\in\ell}\delta\big(\sigma_i^{(\mu)}, \sigma_j^{(\mu)}\big), \tag{3.5}$$

defining the coefficients b_ℓ. The normalisation of Eq. (3.4) is expressed by the fact that $b_\circ = 1$.

To relate the b_ℓ to the physical coupling constants K_ℓ, we evaluate Eqs. (3.4) and (3.5) in the situation where $\delta(\sigma_i^{(\mu)}, \sigma_j^{(\mu)}) = 1$ for $\mu \in \ell'$, and zero otherwise. This resulting equations

$$\exp\left(\sum_{\rho\subset\ell'} K_\rho\right) = \sum_{\rho\subset\ell'} b_\rho \tag{3.6}$$

can readily be solved by applying the principle of inclusion–exclusion [15], yielding

$$b_\ell = \sum_{\ell'\subset\ell}(-1)^{|\ell|-|\ell'|}\exp\left(\sum_{\rho\in\ell'} K_\rho\right). \tag{3.7}$$

The partition function in the spin representation

$$Z = \sum_{\{\sigma\}} \prod_{\langle ij \rangle} \exp(-\mathcal{H}_{ij}) \tag{3.8}$$

can now be transformed into the random cluster representation as follows. First, insert Eq. (3.5) on the right-hand side of the above equation, and imagine expanding the product over the lattice edges $\langle ij \rangle$. To each term in the resulting sum we associate an edge colouring $\mathcal{G}$ of $\mathcal{L}^M$, where an edge $\langle ij \rangle$ in layer μ is considered to be coloured (occupied) if the term contains the factor $\delta(\sigma_i^{(\mu)}, \sigma_j^{(\mu)})$, and uncoloured (empty) if it does not. In other words, an edge colouring $\mathcal{G}$ is determined by specifying an edge state $\ell_{ij} \subset \mathcal{E}$ for every edge $\langle ij \rangle \in \mathcal{L}$.

The summation over the spin variables $\{\sigma\}$ is now trivially performed, yielding a factor of q_μ for each connected component (cluster) in layer μ of the colouring graph. Keeping track of the prefactors multiplying the δ-functions, using Eq. (3.5), we conclude that

$$Z = \sum_{\mathcal{G}} \prod_{\ell \subset \mathcal{E}} b_\ell^{B_\ell} \prod_{\mu \in \ell} q_\mu^{C_\mu}, \tag{3.9}$$

where C_μ is the number of clusters in layer μ, and B_ℓ is the number of edges $\langle ij \rangle \in \mathcal{L}^M$ having the state ℓ.

It is worth noticing that the random cluster description of the model has the advantage that the q_μ only enter as parameters. By analytic continuation one can thus give meaning to a non-integer number of states. The price to be paid is that the C_μ are, a priori, non-local quantities.

In terms of the edge variables b_ℓ the duality transformation of the partition function is easily worked out. For simplicity we shall assume that the couplings constants K_ℓ are identical between all nearest-neighbour pairs of spins. The generalisation to an arbitrary inhomogeneous distribution of couplings is trivial; it suffices to let K_ℓ depend on $\langle ij \rangle$ in Eq. (3.4).

By definition, we take the colouring state $\ell \subset \mathcal{E}$ of the edge $\langle ij \rangle \in \mathcal{L}$ to be dual to the complementary colouring ℓ^* of its intersecting dual edge $\widetilde{\langle ij \rangle} \in \widetilde{\mathcal{L}}$. In the symbolic notation introduced above, the complementarity operation $*$ simply means replacing every $\bullet$ by a $\circ$, and vice versa. Also, note that we refer to dual quantities by a tilde throughout.

To establish the duality transformations, we begin by postulating that the configuration $\mathcal{G}_\bullet$ with all lattice edges coloured, be dual to the configuration $\mathcal{G}_\bullet^* \equiv \mathcal{G}_\circ$ with no coloured (dual) edge.

This requirement fixes the constant entering the duality transformation. Indeed, from Eq. (3.9), we find that $\mathcal{G}_\bullet$ has weight $b_\bullet^E \prod_{\mu=1}^M q_\mu$, where E is the total number of lattice edges, and $\mathcal{G}_\circ$ is weighted by $\tilde{b}_\circ^E \prod_{\mu=1}^M q_\mu^F$, where F is the number of faces, including the exterior one. We thus seek for a duality transformation of the form

$$\frac{Z(\{b_\ell\})}{\tilde{Z}(\{\tilde{b}_\ell\})} = \left(\frac{\tilde{b}_\circ}{b_\bullet}\right)^E \prod_{\mu=1}^M q_\mu^{F-1}, \tag{3.10}$$

where for any configuration $\mathcal{G}$ the edge weights must transform so as to keep the same *relative* weight between $\mathcal{G}$ and $\mathcal{G}_\bullet$ as between $\mathcal{G}^*$ and $\mathcal{G}_\circ$.

An arbitrary colouring configuration $\mathcal{G}$ entering Eq. (3.9) can be generated by applying a finite number of changes to $\mathcal{G}_\bullet$, in which an edge of weight $b_\bullet$ is changed into an edge of weight b_ℓ for some $\ell \subset \mathcal{E}$. By such a change, in general, a pivotal bond is removed from the colouring graph in each of some subset $\ell' \subset \ell^*$ of layers, thus creating $|\ell'|$ new clusters in the corresponding layers. The weight relative to that of $\mathcal{G}_\bullet$ will therefore change by $(b_\ell/b_\bullet)\prod_{\mu\in\ell'} q_\mu$. On the other hand, in the dual configuration $\widetilde{\mathcal{G}}$ a cluster will be lost in each of the layers $(\ell')^* \cap \ell^*$, since each of the $|\ell'|$ new clusters mentioned above will be accompanied by the formation of a *loop* in $\widetilde{\mathcal{G}}$. The weight change relative to $\mathcal{G}_\circ$ therefore amounts to $(\tilde{b}_{\ell^*}/\tilde{b}_0)\prod_{\mu\in(\ell')^*\cap\ell^*} q_\mu^{-1}$. Comparing these two changes we see that the factor $\prod_{\mu\in\ell'} q_\mu$ cancels nicely, and the duality transformation takes the simple form

$$\tilde{b}_\ell = \frac{b_{\ell^*}}{b_\bullet}\prod_{\mu\in\ell} q_\mu, \tag{3.11}$$

the relation with $\ell = \circ$ being trivial.

Note in particular that Eq. (3.11) with $\ell = \bullet$ implies that $b_\bullet \tilde{b}_\bullet = \prod_{\mu=1}^M q_\mu$, since $b_\circ = 1$ by definition. Then, dualising Eq. (3.11) once again yields

$$\tilde{\tilde{b}}_\ell = \frac{1}{\tilde{b}_\bullet}\prod_{\mu\in\ell} q_\mu \frac{b_\ell}{b_\bullet}\prod_{\mu\in\ell^*} q_\mu = b_\ell, \tag{3.12}$$

so that the duality transformation is indeed involutive, as required.

The duality relations (3.11) can be recast in an even simpler form by trading the random clusters for the loops surrounding them (and their duals) on the lattice $\mathcal{L}_{\mathrm{m}}$ medial to $\mathcal{L}$ [14]. Using the Euler relation, we find that Eq. (3.9) must be replaced by

$$Z = \left(\prod_{\mu=1}^M q_\mu\right)^{N/2} \sum_{\mathcal{G}} \prod_{\ell\subset\mathcal{E}} x_\ell^{B_\ell} \prod_{\mu\in\ell} q_\mu^{L_\mu/2}, \tag{3.13}$$

where, by a slight abuse of notation, we use the same notation $\mathcal{G}$ for the loop and the cluster configurations, since they are in bijective correspondence. L_μ are now the number of closed loops in layer μ, and N is the total number of vertices in $\mathcal{L}$. Note that the bond weights b_ℓ have now been replaced by

$$x_\ell = b_\ell \prod_{\mu\in\ell} q_\mu^{-1/2}. \tag{3.14}$$

It is easily verified that the duality relations (3.11) now simply read

$$\tilde{x}_\ell = x_{\ell^*}. \tag{3.15}$$

This is our main result.

In the case of a lattice $\mathcal{L} = \widetilde{\mathcal{L}}$ which is unchanged by the duality, such as the infinite square lattice, we can now search for selfdual solutions. These are obtained by imposing $\tilde{b}_\ell = b_\ell$, and read explicitly

$$x_\ell^{(\mathrm{s.d.})} = x_{\ell^*}^{(\mathrm{s.d.})}. \tag{3.16}$$

Since $x_\circ = 1$ by normalisation, the selfdual manifold has dimension $D_M = 2^{M-1} - 1$.

Two special points always belong to the selfdual manifold. The first one is

$$x_\ell = 1, \quad \text{for any } \ell \subset \mathcal{E}. \tag{3.17}$$

It is straightforward to verify that in terms of the original couplings this means $K_{\circ\cdots\circ\bullet\circ\cdots\circ} = 1 + q_\mu^{1/2}$ (with the unique $\bullet$ at position μ) for $\mu = 1, 2, \ldots, M$, all other K_ℓ being zero. In other words, this is just M non-interacting selfdual (critical) Potts models.

The other special point is

$$x_\circ = x_\bullet = 1 \quad (\text{all other } x_\ell = 0). \tag{3.18}$$

In terms of the original couplings this means $K_\bullet = 1 + \prod_{\mu=1}^{M} q_\mu^{1/2}$, all other K_ℓ being zero. In this case, the M models effectively couple so as to form a *single* critical Potts model with $\prod_{\mu=1}^{M} q_\mu$ states.

3.2. Two models

Let us briefly show how to recover the result for $M = 2$ [10] from the compact formulation of Eq. (3.15).

Introducing the shorthand notation $\delta_\mu = \delta\big(\sigma_i^{(\mu)}, \sigma_j^{(\mu)}\big)$ for $\mu = 1, 2$, the Hamiltonian (3.2) reads in this case

$$-\mathcal{H}_{ij} = K_{\bullet\circ}\delta_1 + K_{\circ\bullet}\delta_2 + K_{\bullet\bullet}\delta_1\delta_2. \tag{3.19}$$

From Eq. (3.7) we have

$$\begin{aligned} &\sqrt{q_1}\, x_{\bullet\circ} \equiv b_{\bullet\circ} = \mathrm{e}^{K_{\bullet\circ}} - 1, \\ &\sqrt{q_2}\, x_{\circ\bullet} \equiv b_{\circ\bullet} = \mathrm{e}^{K_{\circ\bullet}} - 1, \end{aligned} \tag{3.20}$$

$$\sqrt{q_1 q_2}\, x_{\bullet\bullet} \equiv b_{\bullet\bullet} = \mathrm{e}^{K_{\bullet\bullet}+K_{\bullet\circ}+K_{\circ\bullet}} - \big(\mathrm{e}^{K_{\bullet\circ}} + \mathrm{e}^{K_{\circ\bullet}}\big) + 1, \tag{3.21}$$

and using the selfduality criteria $x_{\bullet\circ} = x_{\circ\bullet}$ and $x_{\bullet\bullet} = x_{\circ\circ} \equiv 1$ we readily find the solutions

$$\begin{aligned} &\mathrm{e}^{K_{\circ\bullet}} = 1 + \sqrt{\frac{q_2}{q_1}}\big(\mathrm{e}^{K_{\bullet\circ}} - 1\big), \\ &\mathrm{e}^{K_{\bullet\bullet}} = \frac{\sqrt{q_1} + \sqrt{q_2}[1 + (q_1 - 1)\mathrm{e}^{-K_{\bullet\circ}}]}{\sqrt{q_1} + \sqrt{q_2}(\mathrm{e}^{K_{\bullet\circ}} - 1)}. \end{aligned} \tag{3.22}$$

3.3. Three models

For the case $M = 3$, the Hamiltonian (3.2) reads

$$\begin{aligned} -\mathcal{H}_{ij} = {} & K_{\bullet\circ\circ}\delta_1 + K_{\circ\bullet\circ}\delta_2 + K_{\circ\circ\bullet}\delta_3 \\ & + K_{\circ\bullet\bullet}\delta_2\delta_3 + K_{\bullet\circ\bullet}\delta_1\delta_3 + K_{\bullet\bullet\circ}\delta_1\delta_2 + K_{\bullet\bullet\bullet}\delta_1\delta_2\delta_3. \end{aligned} \tag{3.23}$$

From Eq. (3.7) we have

$$
\begin{aligned}
b_{\bullet\circ\circ} &= \mathrm{e}^{K_{\bullet\circ\circ}} - 1, \\
b_{\circ\bullet\bullet} &= \mathrm{e}^{K_{\circ\bullet\bullet}+K_{\circ\bullet\circ}+K_{\circ\circ\bullet}} - \left(\mathrm{e}^{K_{\circ\bullet\circ}} + \mathrm{e}^{K_{\circ\circ\bullet}}\right) + 1, \\
b_{\bullet\bullet\bullet} &= \mathrm{e}^{K_{\bullet\bullet\bullet}+K_{\circ\bullet\bullet}+K_{\bullet\circ\bullet}+K_{\bullet\bullet\circ}+K_{\bullet\circ\circ}+K_{\circ\bullet\circ}+K_{\circ\circ\bullet}} \\
&\quad - \left(\mathrm{e}^{K_{\circ\bullet\bullet}+K_{\circ\bullet\circ}+K_{\circ\circ\bullet}} + \mathrm{e}^{K_{\bullet\circ\bullet}+K_{\bullet\circ\circ}+K_{\circ\circ\bullet}} + \mathrm{e}^{K_{\bullet\bullet\circ}+K_{\bullet\circ\circ}+K_{\circ\bullet\circ}}\right) \\
&\quad + \left(\mathrm{e}^{K_{\bullet\circ\circ}} + \mathrm{e}^{K_{\circ\bullet\circ}} + \mathrm{e}^{K_{\circ\circ\bullet}}\right) - 1,
\end{aligned}
\tag{3.24}
$$

where the first and second line each represent three equations that can be obtained by cyclically rotating the layer indices.

Imposing the selfduality criteria $\sqrt{q_1}\, b_{\circ\bullet\bullet} = \sqrt{q_2 q_3}\, b_{\bullet\circ\circ}$ and $b_{\bullet\bullet\bullet} = \sqrt{q_1 q_2 q_3}$ we obtain after a little algebra

$$
\begin{aligned}
\mathrm{e}^{K_{\circ\bullet\bullet}} &= \frac{\sqrt{q_2 q_3/q_1}(\mathrm{e}^{K_{\bullet\circ\circ}} - 1) + \mathrm{e}^{K_{\circ\bullet\circ}} + \mathrm{e}^{K_{\circ\circ\bullet}} - 1}{\mathrm{e}^{K_{\circ\bullet\circ}+K_{\circ\circ\bullet}}}, \\
\mathrm{e}^{K_{\bullet\bullet\bullet}} &= \frac{A}{B}\mathrm{e}^{K_{\bullet\circ\circ}+K_{\circ\bullet\circ}+K_{\circ\circ\bullet}}, \\
A &= \left(\mathrm{e}^{K_{\bullet\circ\circ}} - 1\right)q_2 q_3 + \left(\mathrm{e}^{K_{\circ\bullet\circ}} - 1\right)q_1 q_3 + \left(\mathrm{e}^{K_{\circ\circ\bullet}} - 1\right)q_1 q_2 \\
&\quad + \left(\mathrm{e}^{K_{\bullet\circ\circ}} + \mathrm{e}^{K_{\circ\bullet\circ}} + \mathrm{e}^{K_{\circ\circ\bullet}} - 2\right)\sqrt{q_1 q_2 q_3} + q_1 q_2 q_3, \\
B &= \left[\left(\mathrm{e}^{K_{\circ\bullet\circ}} + \mathrm{e}^{K_{\circ\circ\bullet}} - 1\right)\sqrt{q_1} + \left(\mathrm{e}^{K_{\bullet\circ\circ}} - 1\right)\sqrt{q_2 q_3}\right] \\
&\quad \times \left[\left(\mathrm{e}^{K_{\bullet\circ\circ}} + \mathrm{e}^{K_{\circ\circ\bullet}} - 1\right)\sqrt{q_2} + \left(\mathrm{e}^{K_{\circ\bullet\circ}} - 1\right)\sqrt{q_1 q_3}\right] \\
&\quad \times \left[\left(\mathrm{e}^{K_{\bullet\circ\circ}} + \mathrm{e}^{K_{\circ\bullet\circ}} - 1\right)\sqrt{q_3} + \left(\mathrm{e}^{K_{\circ\circ\bullet}} - 1\right)\sqrt{q_1 q_2}\right].
\end{aligned}
\tag{3.25}
$$

These expressions generalise those given in Ref. [2].

It should be clear that for higher M, manipulating such expressions by direct use of the Fourier method [12] becomes extremely cumbersome.

4. Transfer matrix algorithm

In Ref. [2] it was shown that the transfer matrix $T(L)$ for coupled Potts models on semi-infinite strips of width L may be written in a variety of ways. For a given L, the smaller the dimension of $T(L)$, the more efficient will be the computations, since both time and memory consumption increase roughly linearly with $\dim T(L)$.

The best choice turns out to be to write the transfer matrix for the loop model on the medial lattice $\mathcal{L}_{\mathrm{m}}$ [14], which was referred to as algorithm `alg4` in [2]. This also gives us the advantage of having simple duality relations, cf. Eq. (3.15), and to be able to treat the numbers of states $\{q_1, q_2, q_3\}$ as continuous parameters.

Let us recall from Eqs. (3.13) and (3.16) that the partition function for $M = 3$ coupled models on the selfdual manifold can be written, up to a trivial multiplicative constant, as

$$
Z = \sum_{\mathcal{G}} x_1^{B_{\bullet\circ\circ}+B_{\circ\bullet\bullet}} x_2^{B_{\circ\bullet\circ}+B_{\bullet\circ\bullet}} x_3^{B_{\circ\circ\bullet}+B_{\bullet\bullet\circ}} q_1^{L_1/2} q_2^{L_2/2} q_3^{L_3/2},
\tag{4.1}
$$

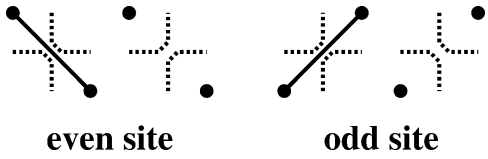


Fig. 1. The relation between the random cluster model on the square lattice and the loop model on the corresponding medial graph. The clusters consist of connected components of coloured edges (thick lines) or isolated sites (filled circles). Loops on the medial graph (dashed lines) are defined by the convention that they wrap around the cluster boundaries.

where we have put $x_1 = x_{\bullet\circ\circ}$, $x_2 = x_{\circ\bullet\circ}$ and $x_3 = x_{\circ\circ\bullet}$ for brevity. For simplicity, we shall take the lattice $\mathcal{L}$ to be the square one, so that $\mathcal{L}_{\rm m}$ is once again a square lattice.

In the loop picture, the symbols $\bullet$ and $\circ$ refer to the two different ways of splitting the vertex, with a definition that alternates between the even and the odd sublattice; see Fig. 1. But thanks to the selfduality, Eq. (4.1) is invariant with respect to a global colour conjugation $\bullet \leftrightarrow \circ$, and so one might as well forget about the distinction between the sublattices.[8] The number of occurrences of a given vertex splitting (and hence the B's) can be counted locally, and are thus easily realised as local Boltzmann factors in the transfer matrix.

The L_μ in (4.1) count the number of closed loops in each layer $\mu = 1, 2, 3$. Despite of the non-local nature of the loops, these quantities do not obviate the construction of the transfer matrix. Rather, they can be counted locally by writing the transfer matrix in the basis of Catalan-like connectivities, as described in [2]. Formally, if $\mathcal{C}_k$ designates the space of pairwise Catalan connectivities among a set of $L = 2k$ points, $T(L)$ acts on the product space $\mathcal{S}_k = \mathcal{C}_k \otimes \mathcal{C}_k \otimes \mathcal{C}_k$. One has $\dim \mathcal{C}_k = \frac{(2k)!}{k!(k+1)!}$.

Weighing the loops with a Boltzmann factor $\sqrt{q_\mu}$ that depends on the layer μ is a trivial modification of [2]. An important difference, however, is that the layers are now distinguishable. Accordingly, one has simply $\dim \mathcal{S}_k = (\dim \mathcal{C}_k)^3$.

As in Ref. [2] we transfer along one of the main directions of the square lattice $\mathcal{L}_{\rm m}$, with periodic boundary conditions in the transverse direction. To ensure that $\mathcal{C}_k$ be well defined, we must take the strip width to be even, $L = 2k$.

We shall also need to consider the situation where all vertex weights x tend to infinity in fixed ratios. Defining $x'_2 = x_2/x_1$ and $x'_3 = x_3/x_1$ as the relevant ratios, (4.1) can be rewritten, once again up to an irrelevant multiplicative constant, as

$$Z' = \sum_{\mathcal{G}'} x_2'^{B_{\circ\bullet\circ}+B_{\bullet\circ\bullet}} x_3'^{B_{\circ\circ\bullet}+B_{\bullet\bullet\circ}} q_1^{L_1/2} q_2^{L_2/2} q_3^{L_3/2}, \tag{4.2}$$

where the symbol $\mathcal{G}'$ means all colouring figurations in which the local colourings $\circ\circ\circ$ and $\bullet\bullet\bullet$ do not occur. In Ref. [2], the special case $(x'_2, x'_3) = (1, 1)$ with $q_1 = q_2 = q_3$ was identified as the non-trivial critical fixed point for three identical coupled models.

[8] Using the language of Ref. [14], coupled Potts models constitute a staggered vertex model, which however becomes homogeneous exactly on the selfdual manifold.

5. Numerical results

Using sparse matrix factorisation techniques, we have been able to numerically compute the first few eigenvalues of the transfer matrices $T(L)$ for Z (4.1) and $T'(L)$ for Z' (4.2) for even strip widths up to $L_{\max} = 12$. The largest matrices had dimension $\dim T(L_{\max}) = (132)^3$, but the sparse matrix factorisation necessitates the use of intermediate states with $L + 2$ dangling loop segments, and so the largest sparse matrices involved were of dimension $(429)^3$.

5.1. Phase diagram and central charge

As a first test of our algorithm, we checked that it gave the correct eigenvalues at the special points $(x_1, x_2, x_3) = (1, 1, 1)$ and $(0, 0, 0)$, cf. Eqs. (3.17) and (3.18).

We then extracted the (effective) central charge from the leading eigenvalue λ_0 in the standard way [16,17]

$$f_0(L) = f_0(\infty) - \frac{\pi c}{6L^2} + \cdots \tag{5.1}$$

with $f_0(L) = -\frac{1}{L}\log\lambda_0$. Using the c-theorem [8], we could readily establish the topology of the RG flows.

- In the space (x_1, x_2, x_3), there is a one-parameter curve along which the partial derivatives of the effective central charge with respect to the two perpendicular directions vanish identically. Moreover, the corresponding second derivatives are strictly negative, so that this curve acts as a "mountain ridge" for c_{eff}.
- The curve passes through the points $(0, 0, 0)$ and $(1, 1, 1)$, and then goes to infinity with some fixed ratios $(x_2', x_3') \equiv (x_2'^*, x_3'^*)$ that depend on (q_1, q_2, q_3).
- For various values $(q_1, q_2, q_3) \in [2, 4]^3$, with at most one $q_a = 2$, we observe that c_{eff} is a monotonically decreasing function when going along the curve from the point $(1, 1, 1)$ towards either $(0, 0, 0)$ or $x_1(1, x_2'^*, x_3'^*)$, $x_1 \to \infty$. In the former case, the decrease is rapid and gets more pronounced with increasing system size L, signalling the first-order nature ($c_{\text{eff}} = 0$) of the phase transition in the $(q_1 q_2 q_3)$-state Potts model. In the latter case, the decrease is very slight, distinguishing the point $(x_2'^*, x_3'^*)$ as a candidate for the non-trivial fixed point of three coupled Potts models in the lattice realisation.

Based on this evidence, we switched to the matrix $T'(L)$ in order to accurately locate the point $(x_2'^*, x_3'^*)$ and study its critical properties. Invoking again the c-theorem [8], this was done by scanning the space (x_2', x_3') for a maximum in c_{eff} for different system sizes. As usually [2], we obtained the best precision by including a non-universal $1/L^4$ term in (5.1), and so our fits are based on three consecutive strip widths L. The obtained maximum can be interpreted as a finite-size pseudo-critical point, which tends to $(x_2'^*, x_3'^*)$ as $L \to \infty$.

We have concentrated the main part of our computation time on two important special cases: a tricritical Ising model and a three-state Potts model coupled with either an Ising model or with a tricritical three-state Potts model. In the language of minimal models, we

Table 1
Maxima of the effective central charge for the system $\mathcal{M}_{345}$, and the deviation from the value at the decoupling point

L	$x_2'^*$	$x_3'^*$	$(c_{\text{eff}})_{\max}$	Δc
2, 4, 6	0.8523(1)	0.7133(1)	1.93540138	−0.00058
4, 6, 8	0.8476(1)	0.7136(1)	1.98040147	−0.00352
6, 8, 10	0.8481(1)	0.7076(1)	1.99047341	−0.00416
8, 10, 12	0.8483(1)	0.7061(1)	1.99348414	−0.00443

Table 2
Maxima of the effective central charge for the system $\mathcal{M}_{456}$, and the deviation from the value at the decoupling point

L	$x_2'^*$	$x_3'^*$	$(c_{\text{eff}})_{\max}$	Δc
2, 4, 6	0.9230(1)	0.8679(1)	2.27445933	−0.0115
4, 6, 8	0.9218(1)	0.8661(1)	2.32198052	−0.0185
6, 8, 10	0.9213(1)	0.8645(1)	2.33129236	−0.0190
8, 10, 12	0.9211(1)	0.8639(1)	2.33327361	−0.0225

refer to these two situations as $\mathcal{M}_{345} \equiv \mathcal{M}_3 \times \mathcal{M}_4 \times \mathcal{M}_5$ and $\mathcal{M}_{456} \equiv \mathcal{M}_4 \times \mathcal{M}_5 \times \mathcal{M}_6$, respectively.

The corresponding finite-size estimates for $(x_2'^*, x_3'^*)$ and for $(c_{\text{eff}})_{\max}$ are given in Tables 1 and 2. Rather than directly extrapolating $(c_{\text{eff}})_{\max}$ to the $L \to \infty$ limit, we consider instead for any L its deviation $\Delta c(L)$ with respect to the corresponding value found at the point where the three Potts models decouple. Since this deviation is numerically small, it is reasonable to expect that the finite-size corrections to c at $(x_2'^*, x_3'^*)$ are similar to those at the decoupling point. Tables 1 and 2 confirm that this is indeed the case: the estimates for Δc only depend very weakly on L. As $L \to \infty$, we obtain finally the following extrapolated values for Δc and for the critical couplings:

$$\begin{aligned} &\text{Model } \mathcal{M}_{345}: \quad \bigl(x_2'^*, x_3'^*\bigr) = \bigl(0.8485(1), 0.7053(5)\bigr), \qquad \Delta c = -0.0050(3), \\ &\text{Model } \mathcal{M}_{456}: \quad \bigl(x_2'^*, x_3'^*\bigr) = \bigl(0.9209(1), 0.8633(3)\bigr), \qquad \Delta c = -0.025(3). \end{aligned} \tag{5.2}$$

The values of Δc are seen to compare quite favourably with the two-loop result of the perturbative RG (2.24), whose numerical values read, respectively,

$$\Delta c(\mathcal{M}_{345})_{\text{RG}} = -0.0041, \qquad \Delta c(\mathcal{M}_{456})_{\text{RG}} = -0.0175. \tag{5.3}$$

Actually, assuming that the RG series is alternating, the fifth-order term not present in (2.24) would supposedly lead to even better agreement with the numerical results.

5.2. Higher exponents in the even sector

Examining the higher eigenvalues of the transfer matrices, one has also access to the scaling dimensions by using the formula [18]

$$f_i(L) - f_0(L) = \frac{2\pi (\Delta_i)_{\text{phys}}}{L^2} + \cdots. \tag{5.4}$$

Table 3
Dimension of the antisymmetric energy operator at $(x_2', x_3') = (1, 1)$ for three coupled three-state Potts models

L	$\Delta(\varepsilon_A)$
4, 6	0.6768
6, 8	0.6648
8, 10	0.6525
$L \to \infty$	0.63(3)
RG result	0.72

We here work in the even sector of the transfer matrix, and besides the identity operator (the free energy) we expect to find the various energy-like primaries. In particular, we are interested in the primary operators ε_1^*, ε_2^*, ε_3^* discussed in Section 2; cf. Eq. (1.5). According to Eqs. (2.14)–(2.16) we have $0 < \Delta(\varepsilon_2^*), \Delta(\varepsilon_3^*) < 1$ and $1 < \Delta(\varepsilon_1^*) < 2$ for small enough $\{\epsilon_a\}$, and so we expect the first three gaps of the transfer matrix to be associated with these operators.

As in Ref. [2], the scaling dimensions of the operators have been extracted by adding to Eq. (5.4) a non-universal $1/L^4$ term, and so the fits are based on two consecutive strip widths L.

We consider first the case of coupling three identical models.

At the non-trivial fixed point $(x_2', x_3') = (1, 1)$ the second and third eigenvalues of the transfer matrix are degenerate. These eigenvalues are associated with the operators ε_{A_1}, ε_{A_2} of (1.4) which generate the two-dimensional irrep of S_3, as discussed in the introduction.

It is important to notice that in the transfer matrix algorithm of [2] the three layers were treated symmetrically and only the operators which are symmetric under permutation of the layer indices were accessible. In that case, the degenerate eigenvalues associated with the energy doublet were not present in the spectrum of the transfer matrix. The fact that we find them here is a strong confirmation of the antisymmetric nature of the corresponding operators.

In Table 3 we report numerical values of the dimension $\Delta(\varepsilon_A) \equiv \Delta(\varepsilon_{A_1}) = \Delta(\varepsilon_{A_2})$ in the case of three coupled Potts models with $q = 3$. The final result $\Delta(\varepsilon_A) = 0.63(3)$ is in reasonable agreement with the RG value and also agrees well with the Monte Carlo result $\Delta(\varepsilon_A) = 0.63 \pm 0.04$ given in [2].

The fourth eigenvalue of the transfer matrix was already found in [2] and it corresponds to the symmetric energy operator ε_S of Eq. (1.3). It is a non-trivial check of our transfer matrices that the corresponding eigenvalue is identical to that of [2].

We now turn our attention to the case of three different coupled models.

In the space (x_2', x_3') of the couplings we have found that for any given system size L there is a unique point $(x_2', x_3')_{\text{deg}}$, at which the second and third eigenvalues are degenerate. We have determined the location of this point for the models $\mathcal{M}_{345}$ and $\mathcal{M}_{456}$ with $L = 4, 6, 8, 10$ and it is shown on Figs. 2 and 3.

In contradistinction to the case of three coupled identical models, the points $(x_2'^*, x_3'^*)$ and $(x_2', x_3')_{\text{deg}}$ do not coincide for finite systems. It is however evident from Figs. 2 and 3 that these two points become closer as the system size increases. This suggests that in

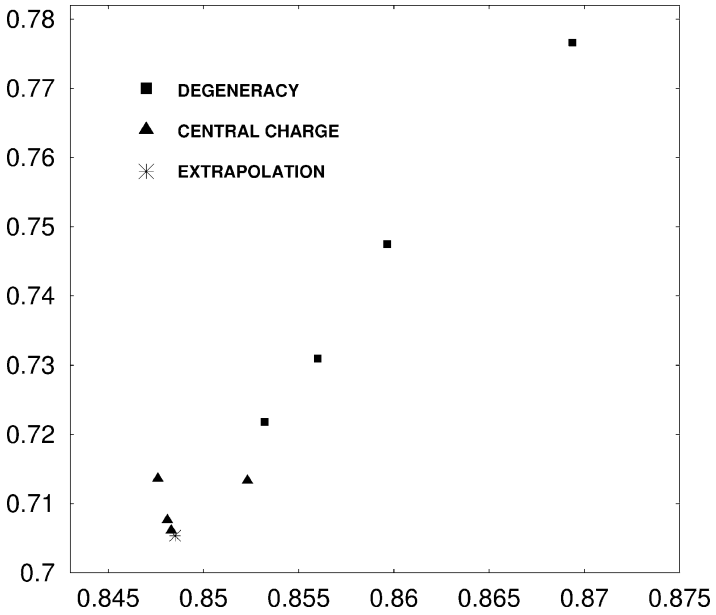


Fig. 2. The location of the critical point for the model $\mathcal{M}_{345}$, as determined from the maximum of the effective central charge. Filled triangles show the positions of the maxima $(x_2'^*, x_3'^*)$ as found from three-point fits with system sizes ranging from $L = 2, 4, 6$ to $L = 8, 10, 12$. The asterisk indicates the extrapolation of these points to the thermodynamic limit, $L \to \infty$. Filled boxes give the positions $(x_2', x_3')_{\text{deg}}$ of the degeneracy between the second and third eigenvalues of the transfer matrix, with system sizes ranging from $L = 4$ to $L = 10$. This latter sequence appears to converge to the same limiting point as the central charge data, in support of the symmetry restoration scenario.

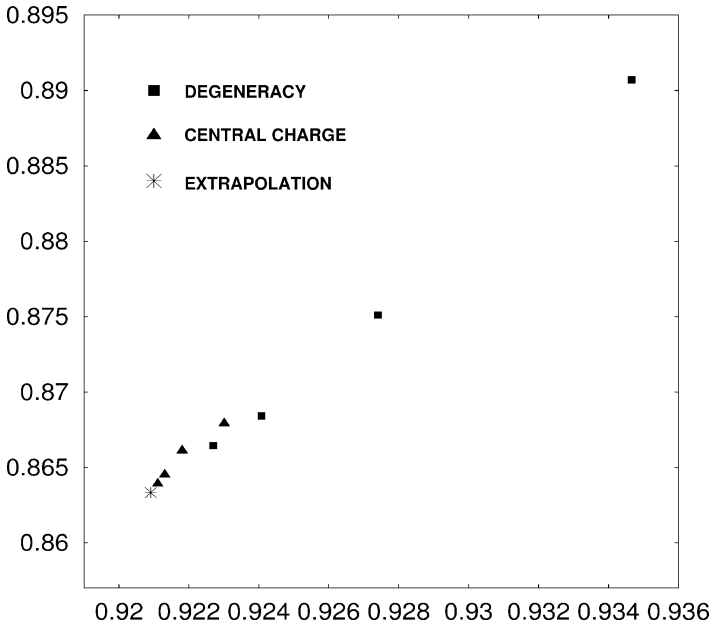


Fig. 3. Same data as in Fig. 2, but here for the model $\mathcal{M}_{456}$.

Table 4
Dimensions of the energy operators at $(x_2'^*, x_3'^*)$ for the $\mathcal{M}_{345}$ model

L	$\Delta(\varepsilon_2^*)$	$\Delta(\varepsilon_3^*)$	$\Delta(\varepsilon_1^*)$
4, 6	0.75734	0.7677	1.2991
6, 8	0.759	0.763	1.2803
8, 10	0.7564	0.75647	1.2761
RG result	0.746	0.771	1.287

Table 5
Dimensions of the energy operators at $(x_2'^*, x_3'^*)$ for the $\mathcal{M}_{456}$ model

L	$\Delta(\varepsilon_2^*)$	$\Delta(\varepsilon_3^*)$	$\Delta(\varepsilon_1^*)$
4, 6	0.6849	0.6857	1.2958
6, 8	0.6731	0.6756	1.2765
8, 10	0.6621	0.6637	1.2745
RG result	0.701	0.728	1.43

the thermodynamic limit they actually coincide. If this scenario holds true, we have at the critical point that $\Delta(\varepsilon_3^*) = \Delta(\varepsilon_2^*)$.

Since the finite-size effects in the sequence of points $(x_2', x_3')_{\text{deg}}$ are much stronger than is the case for $(x_2'^*, x_3'^*)$, it is reasonable to measure numerical values of the dimensions at the best available estimate for the latter point. We display the corresponding results for $\Delta(\varepsilon_1^*)$, $\Delta(\varepsilon_2^*)$, $\Delta(\varepsilon_3^*)$ in Tables 4 and 5, once again for the models $\mathcal{M}_{345}$ and $\mathcal{M}_{456}$, respectively. The numerical results are in quite good agreement with the two-loop RG calculation of Section 2. Notice that in the case of $\Delta(\varepsilon_2^*)$, $\Delta(\varepsilon_3^*)$ we compare with the RG values with the splitting ignored.

In Tables 4 and 5 we have given the results for $\Delta(\varepsilon_2^*)$ and $\Delta(\varepsilon_3^*)$ separately, although we do in fact believe that the two dimensions coincide. This point of view is corroborated by the fact that the difference between the dimensions of ε_2^* and ε_3^* decreases rapidly with the lattice size. For $L = 8, 10$ the numerical results give $\Delta_{\text{splitting}} \equiv |\Delta(\varepsilon_3^*) - \Delta(\varepsilon_2^*)| < 0.0003$ for the $\mathcal{M}_{345}$ model and $\Delta_{\text{splitting}} < 0.002$ for the $\mathcal{M}_{456}$ model.

In order to be sure that the very small values for $\Delta_{\text{splitting}}$ obtained at finite size are not somehow accidental, we have determined an upper bound for $\Delta_{\text{splitting}}$ by directly extrapolating the difference between the second and third eigenvalues of the transfer matrix at the point $(x_2'^*, x_3'^*)$. The result is shown below together with the extrapolated values for the dimensions of the three energy operators:

$$\begin{aligned}
&\text{Model } \mathcal{M}_{345}: \quad \Delta\big(\varepsilon_1^*\big) = 1.269 \pm 0.002, \\
&\qquad\qquad\quad \Delta\big(\varepsilon_2^*\big),\, \Delta\big(\varepsilon_3^*\big) = 0.750 \pm 0.005, \qquad \Delta_{\text{splitting}} < 0.003, \\
&\text{Model } \mathcal{M}_{456}: \quad \Delta\big(\varepsilon_1^*\big) = 1.272 \pm 0.002, \\
&\qquad\qquad\quad \Delta\big(\varepsilon_2^*\big),\, \Delta\big(\varepsilon_3^*\big) = 0.645 \pm 0.003, \qquad \Delta_{\text{splitting}} < 0.002. \qquad (5.5)
\end{aligned}$$

The difference $\Delta_{\text{splitting}}$ is well below the one predicted by the RG calculation, which reads, respectively, 0.025 and 0.027 for the models $\mathcal{M}_{345}$ and $\mathcal{M}_{456}$. The numerical work thus

provides clear evidence that at the non-trivial fixed point the splitting of the dimensions of ε_3^* and ε_2^* is actually zero.

6. Discussion

In this paper we have shown that coupling $M = 3$ different Potts models (with $q_1, q_2, q_3 > 2$ and not too large) one obtains a unique non-trivial critical point. The critical properties of this point, in particular its central charge and the values of various energetic scaling dimensions, have been determined quite accurately by a perturbative RG treatment and found to be consistent with large-scale numerical simulations. An exception is however the RG prediction that at two-loop order the degeneracy between the two antisymmetric energy operators should be lifted: this prediction has here been discarded on the basis of numerical evidence.

An extension of our investigation to the case of spin-like operators will be published elsewhere [19].

This work forms part of a larger project [2], in which we examine the possible universality classes of coupled Potts models, and eventually their relation to the random-bond Potts model. In particular, substantial evidence has been accumulated that in the random-bond case replica symmetry is not broken, and one can thus hope to make analytical progress by studying the *unitary* models that result from coupling a certain number of minimal models.

We believe that the symmetry properties of the coupled models play an essential role. It is in the light of this belief that the present work appears to be interesting: even in the absence of an explicit S_M symmetry in the initial action, this symmetry appears to be restored at the non-trivial critical point. As far as a putative CFT classification of S_M symmetrical critical points goes, we therefore see that the number of models to be classified is potentially very large. Indeed, assuming the conclusions of the $M = 3$ case to carry over to a general number M of coupled models, one may expect a distinct S_M symmetric universality class to arise from coupling any different set of M minimal models. If this is true, it would call for a substantial number of new CFTs endowed with extended symmetries. Further research along these lines is currently in progress.

References

[1] A.W.W. Ludwig, Nucl. Phys. B 330 (1990) 639.
[2] Vl.S. Dotsenko, J.L. Jacobsen, M.-A. Lewis, M. Picco, Nucl. Phys. B 546 (1999) 505.
[3] Vl.S. Dotsenko, X.S. Nguyen, R. Santachiara, hep-th/0104197, Nucl. Phys. B, in press.
[4] M.-A. Lewis, P. Simon, Phys. Lett. B 435 (1998) 159.
[5] P. Simon, Nucl. Phys. B 515 (1998) 624.
[6] J.L. Jacobsen, Phys. Rev. E 62 (2000) R1.
[7] Vl.S. Dotsenko, M. Picco, P. Pujol, Nucl. Phys. B 455 (1995) 701.
[8] A.B. Zamolodchikov, Pis' ma Zh. Eksp. Teor. Fiz. 43 (1986) 565, English translation in JETP Lett. 43 (1986) 730;
A.B. Zamolodchikov, Sov. J. Nucl. Phys. 46 (1987) 1090.
[9] R.B. Potts, Proc. Cambridge Philos. Soc. 48 (1952) 106.

[10] E. Domany, E.K. Riedel, Phys. Rev. B 19 (1979) 5817.
[11] R. Savit, Rev. Mod. Phys. 52 (1980) 453.
[12] F.Y. Wu, Y.K. Wang, J. Math. Phys. 17 (1976) 439.
[13] P.W. Kasteleyn, C.M. Fortuin, J. Phys. Soc. Jpn. (Suppl.) 26 (1969) 11.
[14] R.J. Baxter, Exactly Solved Models in Statistical Mechanics, Academic Press, New York, 1982.
[15] G.D. Birkhoff, Ann. Math. 14 (1912) 42.
[16] H. Blöte, J.L. Cardy, M.P. Nightingale, Phys. Rev. Lett. 56 (1986) 742.
[17] I. Affleck, Phys. Rev. Lett. 56 (1986) 746.
[18] J.L. Cardy, J. Phys. A 16 (1983) L355.
[19] Vl.S. Dotsenko, J.L. Jacobsen, X.S. Nguyen, R. Santachiara, work in progress.

Bibliographie

[1] C. C. Adams, *The knot book : An elementary introduction to the mathematical theory of knots* (W. H. Freeman, 1994).

[2] M. Aizenman et J. Wehr, Phys. Rev. Lett. **62**, 2503 (1989).

[3] K. Appel et W. Haken, Illinois J. Math. **21**, 429 (1977).

[4] K. Appel, W. Haken et J. Koch, Illinois J. Math. **21**, 491 (1977).

[5] C. G. Bachet de Méziriac, deux livres sur les récréations mathématiques (Paris ?, 1612 et 1624).

[6] M. T. Batchelor et H. W. J. Blöte, Phys. Rev. Lett. **61**, 138 (1988).

[7] M. T. Batchelor, J. Suzuki et C. M. Yung, Phys. Rev. Lett. **73**, 2646 (1994).

[8] M. T. Batchelor, H. W. J. Blöte, B. Nienhuis et C. M. Yang, J. Phys. A **29**, L399 (1996).

[9] R. J. Baxter, J. Math. Phys. **11**, 784 (1970).

[10] R. J. Baxter, J. Phys. C **6**, L445 (1973).

[11] R. J. Baxter, S. B. Kelland et F. Y. Wu, J. Phys. A **9**, 397 (1976).

[12] R. J. Baxter, H. N. V. Temperley et S. E. Ashley, Proc. Roy. Soc. London A **358**, 535 (1978).

[13] R. J. Baxter, *Exactly solved models in statistical mechanics* (Academic Press, London, 1982).

[14] R. J. Baxter, Proc. Roy. Soc. London A **383**, 43 (1982).

[15] R. J. Baxter, J. Phys. A **19**, 2821 (1986).

[16] R. J. Baxter, J. Phys. A **20**, 5241 (1987).

[17] A. A. Belavin, A. M. Polyakov et A. B. Zamolodchikov, Nucl. Phys. B **241**, 333 (1984).

[18] S. Beraha, non publié (1974).

[19] S. Beraha et J. Kahane, J. Combin. Theory B **27**, 1 (1979).

[20] S. Beraha, J. Kahane et N. J. Weiss, J. Combin. Theory B **28**, 52 (1980).

[21] B. Berche et C. Chatelain, cond-mat/0207421.

[22] V. L. Berezinskii, Sov. Phys. JETP **34**, 610 (1971).

[23] R. N. Berker et L. Kadanoff, J. Phys. A **13**, L259 (1980).

[24] N. L. Biggs, R. M. Damerell et D. A. Sands, J. Combin. Theory B **12**, 123 (1972).

[25] G. D. Birkhoff, Ann. Math. **14**, 42 (1912).

[26] H. W. J. Blöte et H. J. Hilhorst, J. Phys. A **15**, L631 (1982).

[27] H. W. J. Blöte et M. P. Nightingale, Physica A **112**, 405 (1982).

[28] H. W. J. Blöte et B. Nienhuis, Phys. Rev. Lett. **72**, 1372 (1994).

[29] E. Brézin, C. Itzykson, G. Parisi et J.-B. Zuber, Comm. Math. Phys. **59**, 35 (1978).

[30] J. L. Cardy et J. L. Jacobsen, Phys. Rev. Lett. **79**, 4063 (1997).

[31] J. L. Cardy, Phys. Rev. Lett. **84**, 3507 (2000).

[32] J. L. Cardy, J. L. Jacobsen et A. Sokal, J. Stat. Phys. **105**, 25 (2001).

[33] S.-C. Chang, J. Salas et R. Shrock, J. Stat. Phys. **107**, 1207 (2002).

[34] S.-C. Chang, J. L. Jacobsen, J. Salas et R. Shrock, cond-mat/0211623, J. Stat. Phys (soumis).

[35] C. Chatelain, B. Berche, W. Janke et P. E. Berche, Phys. Rev. E **64**, 036120 (2001).

[36] J. T. Chayes, L. Chayes, D. S. Fisher et T. Spencer, Phys. Rev. Lett. **57**, 2999 (1986).

[37] J. T. Chayes, L. Chayes, D. S. Fisher et T. Spencer, Comm. Math. Phys. **120**, 501 (1989).

[38] G. L. Chia, Discrete Math. **172**, 175 (1997).

[39] J.-M. Daul, hep-th/9502014.

[40] F. David, Nucl. Phys. B **487**, 633 (1997).

[41] F. David, Mod. Phys. Lett. A **3**, 1651 (1988).

[42] J. Distler et H. Kawai, Nucl. Phys. B **321**, 509 (1989).

[43] E. Domany et E. K. Riedel, Phys. Rev. B **19**, 5817 (1979).

[44] Vik. S. Dotsenko et Vl. S. Dotsenko, Sov. Phys. JETP Lett. **33**, 37 (1981).

[45] Vik. S. Dotsenko et Vl. S. Dotsenko, J. Phys. A **17**, L301 (1984).

[46] Vl. S. Dotsenko et V. A. Fateev, Nucl. Phys. B **240**, 312 (1984).

[47] Vl. S. Dotsenko et V. A. Fateev, Nucl. Phys. B **251**, 691 (1985).

[48] Vl. S. Dotsenko, M. Picco et P. Pujol, Nucl. Phys. B **455**, 701 (1995).

[49] Vl. S. Dotsenko, J. L. Jacobsen, M.-A. Lewis et M. Picco, Nucl. Phys. B **546** [FS], 505 (1999).

[50] Vl. S. Dotsenko, J. L. Jacobsen et M. Picco, Nucl. Phys. B **618**, 523 (2001).

[51] Vl. S. Dotsenko, J. L. Jacobsen, X. S. Nguyen et R. Santachiara, Nucl. Phys. B **631**, 426 (2002).

[52] Vl. S. Dotsenko, J. L. Jacobsen et R. Santachiara, *Representation theory of parafermionic current algebras*, en préparation (2002).

[53] B. Duplantier, J. Stat. Phys. **49**, 411 (1987).

[54] B. Duplantier et H. Saleur, Nucl. Phys. B **290**, 291 (1987).

[55] B. Durhuus, J. Frölich et T. Jónsson, Nucl. Phys. B **240**, 453 (1984).

[56] L. Euler, *Solutio problematis ad geometriam situs pertinentis*, Commetarii Academiae Scientiarum Imperialis Petropolitanae **8**, 128–140 (1736).

[57] B. Eynard, E. Guitter et C. Kristjansen, Nucl. Phys. B **528**, 523 (1998).

[58] B. Eynard et G. Bonnet, Phys. Lett. B **463** 273 (1999).

[59] B. Eynard, *Random matrices*, `http ://www-spht.cea.fr/lectures_notes.shtml`; Saclay Lecture Notes (2000).

[60] J. P. Flory, Proc. R. Soc. A **234**, 60 (1956).

[61] P. Di Francesco, P. Ginsparg et J. Zinn-Justin, Phys. Rap. **254**, 1 (1995).

[62] P. Di Francesco, P. Mathieu et D. Sénéchal, *Conformal field theory* (Springer, New York, 1996).

[63] P. Di Francesco, B. Eynard et E. Guitter, Nucl. Phys. B **516**, 543 (1998).

[64] P. Di Francesco, E. Guitter et C. Kristjansen, Nucl. Phys. B **549**, 657 (1999).

[65] P. Di Francesco, O. Golinelli et E. Guitter, Nucl. Phys. B **570**, 699 (2000).

[66] P. Di Francesco, E. Guitter et J. L. Jacobsen, Nucl. Phys. B **580** [FS], 757 (2000).

[67] P. Di Francesco, Bull. Am. Math. Soc. **37**, 251 (2000).

[68] P. Di Francesco, E. Guitter et J. L. Jacobsen, *Loop models and the folding of the square-diagonal lattice*, non publié (2000).

[69] D. Friedan, Z. Qiu et S. Shenker, Phys. Rev. Lett. **52**, 1575 (1984).

[70] M. Gaudin and I. K. Kostov, Phys. Lett. B **220**, 200 (1989).

[71] D. Gepner et Z. Qui, Nucl. Phys. B **285**, 123 (1987).

[72] B. Grünbaum et G. C. Shephard, *Tilings and patterns* (Freeman, New York, 1987).

[73] E. Guitter, C. Kristjansen et J. L. Nielsen, Nucl. Phys. B **546**, 731 (1999).

[74] W. Guo, H. W. J. Blöte et F. Y. Wu, Phys. Rev. Lett. **85**, 3874 (2000).

[75] F. Guthrie, non publié (1852). Voir `http ://www.math.gatech.edu/~thomas/FC/fourcolor.html` pour une discussion historique.

[76] A. B. Harris, J. Phys. C **7**, 1671 (1974).

[77] M. Hellmund et W. Janke, cond-mat/0206400.

[78] C. L. Henley, non publié (1996 ?).

[79] S. Higuchi, Nucl. Phys. B **540** 731 (1999).

[80] C. Itzykson et J. M. Drouffe, *Théorie statistique des champs* (InterEditions, 1989).

[81] J. L. Jacobsen, Phys. Lett. A **233**, 489 (1997).

[82] J. L. Jacobsen et J. L. Cardy, Nucl. Phys. B **515** [FS], 701 (1998).

[83] J. L. Jacobsen et J. Kondev, Nucl. Phys. B **532** [FS], 635 (1998).

[84] J. L. Jacobsen et J. Kondev, J. Stat. Phys. **96**, 21 (1999).

[85] J. L. Jacobsen, J. Phys. A **32**, 5445 (1999).

[86] J. L. Jacobsen et J. Vannimenus, J. Phys. A **32**, 5455 (1999).

[87] J. L. Jacobsen et M. Picco, Phys. Rev. E **61**, R13 (2000).

[88] J. L. Jacobsen, Phys. Rev. E **61**, R6060 (2000).

[89] J. L. Jacobsen, Phys. Rev. E **62**, R1 (2000).

[90] J. L. Jacobsen et J. Salas, J. Stat. Phys. **104**, 701 (2001).

[91] J. L. Jacobsen et P. Zinn-Justin, J. Knot Th. Ramif. **11**, 739 (2002).

[92] J. L. Jacobsen et P. Zinn-Justin, J. Knot Th. Ramif. **10**, 1233 (2001).

[93] J. L. Jacobsen et M. Picco, Phys. Rev. E **65**, 026113 (2002).

[94] J. L. Jacobsen et P. Zinn-Justin, J. Phys. A **35**, 2131 (2002).

[95] J. L. Jacobsen, J. Salas et A. D. Sokal, cond-mat/0204587, J. Stat. Phys (soumis).

[96] J. L. Jacobsen, N. Read et H. Saleur, cond-mat/0205033, Phys. Rev. Lett. (soumis).

[97] J. L. Jacobsen et P. Zinn-Justin, *The combinatorics of alternating tangles : From theory to computerized enumeration*, dans V. A. Malyshev et A. M. Vershik (eds.), *Asymptotic combinatorics with applications to mathematical physics*, p. 97–112 (Kluwer Academic Publishers, 2002).

[98] J. L. Jacobsen, J. Salas et A. D. Sokal, *Phase diagrams of planar Potts models*, en préparation (2002).

[99] J. L. Jacobsen et J. Kondev, cond-mat/0209247, Phys. Rev. E (soumis).

[100] I. Jensen, cond-mat/9910313.

[101] R. Juhasz, H. Rieger et F. Igloi, Phys. Rev. E **64**, 056122 (2001).

[102] P. W. Kasteleyn, Physica **29**, 1329 (1963).

[103] P. W. Kasteleyn et C. M. Fortuin, J. Phys. Soc. Jpn. **26** (suppl.), 11 (1969).

[104] L. H. Kauffman, *Knots and physics*, 3éme édition (World Scientific, 2001).

[105] V. A. Kazakov, Phys. Lett. A **119**, 140 (1986).

[106] D. V. Boulatov et V. A. Kazakov, Phys. Lett. B **186**, 379 (1987).

[107] C. King, J. Stat. Phys. **96**, 1071 (1999).

[108] C. King et F. Y. Wu, J. Stat. Phys. **107**, 919 (2002).

[109] W. Kinzel et E. Domany, Phys. Rev. B **23**, 3421 (1981).

[110] V. G. Knizhnik, A. M. Polyakov et A. B. Zamolodchikov, Mod. Phys. Lett. A **3**, 819 (1988).

[111] J. Kondev et C. L. Henley, Phys. Rev. B **52**, 6628 (1995).

[112] J. Kondev, J. de Gier et B. Nienhuis, J. Phys. A **29**, 6489 (1996).

[113] J. Kondev, cond-mat/9607181.

[114] J. Kondev, Phys. Rev. Lett. **78**, 4320 (1997).

[115] J. Kondev et J. L. Jacobsen, Phys. Rev. Lett. **81**, 2922 (1998).

[116] J. M. Kosterlitz et D. J. Thouless, J. Phys. C **6**, 1181 (1973).

[117] I. K. Kostov, Mod. Phys. Lett. A **4**, 217 (1989).

[118] I. K. Kostov and M. Staudacher, Nucl. Phys. B **384**, 459 (1992).

[119] I. K. Kostov, hep-th/0005190.

[120] T. D. Lee et C. N. Yang, Phys. Rev. **87**, 410 (1952).

[121] G. W. von Leibniz, cité dans *Histoire de l'Académie royale des Sciences de Paris* pour l'année 1771, p. 55.

[122] A. Lenard, cité pp. 169–170 dans E. H. Lieb, Phys. Rev. **162**, 162 (1967).

[123] A. W. W. Ludwig et J. L. Cardy, Nucl. Phys. B **285**, 687 (1987).

[124] A. W. W. Ludwig, Nucl. Phys. B **330**, 639 (1990).

[125] W. T. Lu et F. Y. Wu, J. Phys. A **31**, 2823 (1998).

[126] P. Martin, *Potts models and related problems in statistical mechanics* (World Scientific, Singapore, 1991).

[127] M. J. Martins, B. Nienhuis et R. Rietman, Phys. Rev. Lett. **81**, 504 (1998).

[128] M. L. Mehta, Matrix theory : Selected topics and useful results (Les Editions de Physique, Les Ulis, 1988).

[129] W. W. Menasco et M. B. Thistlethwaite, Bull. Amer. Math. Soc. **25**, 403 (1991).

[130] B. Miller et J. O. Rädler, Phys. Rev. Lett. **82**, 1911 (1999).

[131] C. Moore et M. E. J. Newman, J. Stat. Phys. **99**, 629 (2000).

[132] C. Mydorge, *Examen du livre des récréations mathématiques* (Paris ?, 1630).

[133] X. S. Nguyen, en préparation (2002).

[134] B. Nienhuis, Phys. Rev. Lett. **49**, 1062 (1982).

[135] B. Nienhuis, H. J. Hilhorst et H. W. J. Blöte, J. Phys. A **17**, 3559 (1984).

[136] B. Nienhuis, dans *Phase transitions and critical phenomena*, ed. C. Domb et J. L. Lebowitz (Academic Press, London, 1987), vol. 11.

[137] L. Onsager, Phys. Rev. **65**, 117 (1944).

[138] H. Orland, C. Itzykson et C. de Dominicis, J. Physique Lett. **46**, L353 (1985).

[139] J. Ozanam, *Récréations mathématiques et physiques* (Paris, 1694, 2 vols.). [Révision par Montucla (Paris, 1778, 4 vols.) ; traduction en anglais par Hutton (London, 1803, 4 vols.)]. Ici, nous faisons référence à p. 222, vol. 4 de l'édition 1725.

[140] G. Parisi et N. Sourlas, J. de Physique Lettres **41**, L403 (1980).

[141] A. M. Polyakov, Pis'ma Zh. Eksp. Teor. Fiz. **12**, 538 (1970) ; Zh. Eksp. Teor. Fiz. **66**, 23 (1974).

[142] R. B. Potts, Proc. Camb. Phil. Soc. **48**, 106 (1952).

[143] N. Read et H. Saleur, Nucl. Phys. B **613**, 409 (2001).

[144] R. C. Read, J. Combin. Theory **4**, 52 (1968).

[145] R. C. Read et W. T. Tutte, dans *Selected topics in graph theory 3*, ed. L. W. Beineke et R. J. Wilson (Academic Press, London, 1988).

[146] T. Saaty et P. Kainen, *The four color problem* (McGraw-Hill, New York, 1977).

[147] J. Salas et A. Sokal, J. Stat. Phys **104**, 609 (2001).

[148] H. Saleur, J. Phys. A **19**, 2409 (1986).

[149] H. Saleur, Comm. Math. Phys. **132**, 657 (1990).

[150] H. Saleur, Nucl. Phys. B **360**, 219 (1991).

[151] R. Savit, Rev. Mod. Phys. **52**, 453 (1980).

[152] G. Schaeffer et P. Zinn-Justin, en préparation (2002).

[153] T. D. Schultz, D. Mattis et E. H. Lieb, Rev. Mod. Phys. **36**, 856 (1964).

[154] A. Sokal, cond-mat/0012369.

[155] J. Stephenson, J. Math. Phys. **5**, 1009 (1964).

[156] C. Sundberg et M. Thistlethwaite, Pac. J. Math. **182**, 329 (1998).

[157] H. N. V. Temperley et E. H. Lieb, Proc. Roy. Soc. London A **322**, 251 (1971).

[158] W. T. Tutte, *Graph Theory* (Cambridge University Press, Cambridge, 1984).

[159] A.-T. Vandermonde, *Remarques sur les problèmes de situation*, Mém. de l'Ac. des Sc. de Paris pour l'année 1771, p. 566 (Paris, 1774).

[160] G. H. Wannier, Phys. Rev. **79**, 357 (1950).

[161] H. Whitney, Ann. Math. **33**, 688 (1932).

[162] F. Y. Wu et Y. K. Wang, J. Math. Phys. **17**, 439 (1976).

[163] F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).

[164] F. Y. Wu, Phys. Lett. A **228**, 43 (1997).

[165] F. Y. Wu et H. Y. Huang, Phys. Rev. Lett. **79**, 4954 (1997).

[166] F. Y. Wu, C. King et W. T. Lu, Ann. Inst. Fourier **49**, 1103 (1999).

[167] C. N. Yang et T. D. Lee, Phys. Rev. **87**, 404 (1952).

[168] A. B. Zamolodchikov et V. A. Fateev, Sov. Phys. JETP **62**, 215 (1985).

[169] A. B. Zamolodchikov et V. A. Fateev, Theor. Math. Phys. **71**, 451 (1987).

[170] P. Zinn-Justin et J.-B. Zuber, J. Knot Th. Ramif. **9**, 1127 (2000).

[171] P. Zinn-Justin, J. Stat. Phys. **98** 245 (2001).