



**HAL**  
open science

# Mesure, prédiction et modélisation de la topologie de l'évolution temporelle de la réionisation cosmologique

Emilie Thélie

► **To cite this version:**

Emilie Thélie. Mesure, prédiction et modélisation de la topologie de l'évolution temporelle de la réionisation cosmologique. Astrophysique [astro-ph]. Université de Strasbourg, 2023. Français. NNT : 2023STRAE013 . tel-04419717

**HAL Id: tel-04419717**

**<https://theses.hal.science/tel-04419717v1>**

Submitted on 26 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**ÉCOLE DOCTORALE de Physique et Chimie-Physique (ED182)**

**Observatoire astronomique de Strasbourg (UMR 7550)**

**THÈSE** présentée par :

**Emilie THELIE**

soutenue le : **22 septembre 2023**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Astrophysique

**Mesure, prédiction et modélisation de la  
topologie de l'évolution temporelle de la  
Réionisation cosmologique**

**THÈSE dirigée par :**

**M. AUBERT Dominique**

Professeur, Observatoire Astronomique de Strasbourg

**RAPPORTEURS :**

**M. DOUSPIS Marian**

Astronome, Institut d'Astrophysique Spatiale

**M. PICHON Christophe**

Directeur de recherche, Institut d'Astrophysique de Paris

---

**AUTRES MEMBRES DU JURY :**

**Mme FERRARI Chiara**

Astronome, Observatoire de la Côte d'Azur

**Mme LANÇON Ariane**

Professeur, Observatoire Astronomique de Strasbourg

---

## Liste des publications

### En tant que première autrice

- 2023** *Topology of Reionisation times : concepts, measurements and comparisons to gaussian random field predictions.*  
**E. Thélie**, D. Aubert, N. Gillet, J. Hiegel, P. Ocvirk.  
Astronomy & Astrophysics, Volume 672, id.A184, 21 pp.
- 2022** *First look at the topology of reionisation redshifts in models of the epoch of reionisation.*  
**E. Thélie**, D. Aubert, N. Gillet, P. Ocvirk.  
Astronomy & Astrophysics, 658, A139, 13pp.

### En tant que deuxième autrice

- 2023** *Reionisation time fields reconstruction from 21 cm signal map.*  
J. Hiegel, **E. Thélie**, D. Aubert, J. Chardin, N. Gillet, P. Galois, N. Mai, P. Ocvirk, R. Ibata.  
Accepté dans Astronomy & Astrophysics, arxiv:2307.00609.

### En tant que co-autrice

- 2023** *DUSTiER (DUST in the Epoch of Reionization) : dusty galaxies in cosmological radiation-hydrodynamical simulations of the Epoch of Reionization with RAMSES-CUDATON.*  
J. S. W. Lewis, P. Ocvirk, Y. Dubois, D. Aubert, J. Chardin, N. Gillet, **E. Thélie**.  
Monthly Notices of the Royal Astronomical Society, Volume 519, Issue 4, pp.5987-6007.
- 2022** *The short ionizing photon mean free path at  $z=6$  in Cosmic Dawn III, a new fully-coupled radiation-hydrodynamical simulation of the Epoch of Reionization.*  
J. S. W. Lewis, P. Ocvirk, J. G. Sorce, Y. Dubois, D. Aubert, L. Conaboy, P. R. Shapiro, T. Dawoodbhoy, R. Teyssier, G. Yepes, S. Gottlöber, Y. Rasera, K. Ahn, I. T. Iliev, H. Park, **E. Thélie**.  
Monthly Notices of the Royal Astronomical Society, Volume 516, Issue 3, pp.3389-3397.
- 2021** *Lyman-alpha opacities at  $z = 4-6$  require low mass, radiatively-suppressed galaxies to drive cosmic reionization.*  
P. Ocvirk, J. S. W. Lewis, N. Gillet, J. Chardin, D. Aubert, N. Deparis, **E. Thélie**.  
Monthly Notices of the Royal Astronomical Society, Volume 507, Issue 4, pp.6108-6117.

---

À toi, mon Papy, qui me disait toujours être très fier de moi.

---

## Remerciements

Je tiens tout d'abord à remercier Dominique, qui m'a donné l'opportunité de pouvoir travailler plus de trois ans en stage puis en doctorat avec lui à l'Observatoire de Strasbourg. Il a été très présent malgré un stage de Master et un début de thèse à distance en raison de la pandémie. Il m'a permis de comprendre et découvrir beaucoup de choses et a su me guider dans ce projet, tout en me laissant la liberté d'explorer ce que je souhaitais.

Je suis également reconnaissante envers Marian Douspis et Christophe Pichon pour avoir accepté de relire mon manuscrit de thèse. Je les remercie grandement, ainsi que Chiara Ferrari et Ariane Lançon, de faire partie du jury de cette thèse. Merci beaucoup aussi à Chiara et Benoît Famaey pour avoir participé à mon comité de suivi de mi-thèse et pour m'avoir donné des retours très positifs.

Je suis très heureuse d'avoir pu passer plus de trois ans au sein de l'Observatoire de Strasbourg, un endroit idéal pour faire ses premiers pas en astrophysique. J'ai ainsi pu développer mes connaissances et mon sens de la recherche dans un cadre accueillant et convivial, tout en faisant des rencontres inoubliables. J'ai apprécié de travailler et de discuter de science avec les personnes de notre équipe de cosmologie : Dominique, Joe, Jonathan, Julien, Katarina, Mei, Nicolas, et Pierre. Mes pensées se tournent ensuite vers tous les doctorants, et en particulier Amandine, Élisabeth, Ivan, Jakub, Julien, Lucie, PA, Thibaut et Wassim, pour tous les moments incroyables passés ensemble, que ce soit à l'Observatoire ou en dehors avec les Schluk, les karaokés, et autres. Je voudrais particulièrement vous remercier, Amandine et Élisabeth, pour ces trois années de discussions, d'entre-aide et de soutien, qui seront pour moi parmi les meilleurs souvenirs de ce doctorat.

En dehors de l'Observatoire, j'ai eu l'occasion de faire des rencontres tout aussi sympathiques qu'enrichissantes, et je pense en particuliers aux personnes de l'École Euclid, à Corentin Cadiou et à Adélie Gorce. Merci aussi à toi Chiara, pour avoir pris le temps de discuter avec moi et de me faire visiter l'Observatoire de la Côte d'Azur.

Je remercie chaleureusement mes parents, pour m'avoir donné la possibilité d'effectuer de si longues études, et pour m'avoir toujours soutenue dans ce long parcours, et dans tout ce qui va suivre. Je pense aussi plus largement à ma famille, et en particulier à ma soeur Manon et à mes petits neveux Maylan et Wyatt, qui ont apporté une note de bonheur dans nos vies. Merci à toi, Séverin, pour ta compagnie et pour m'avoir permis de passer de si bons moments tout au long de ces années. L'aventure nous attend dans les années à venir, et je ne pourrais être plus contente que tu aies décidé de me suivre là où ma carrière me portera.

Enfin, un grand merci à toutes les personnes qui ont été présentes dans ma vie pendant ces dernières années. Je fais référence à toutes les personnes mentionnées ci-dessus, à certaines qui ne sont plus parmi nous mais auxquelles je pense énormément, et à mes amis. Je pense à Coco, Ludo et Réhane, aux potes de Charly – Amandine, Emma, Margot, Océane, et Prettivy, et j'adresse aussi un clin d'oeil particulier au groupe de TPS – Émilie, Gaëtan, Nathou, PA, Séverin et Sylvain.

---

## Résumé

Pendant l'Époque de la Réionisation, les premières étoiles et galaxies de notre Univers sont apparues et ont réionisé le gaz, principalement de l'Hydrogène, en formant des bulles ionisées autour d'elles. Ces bulles ont grossi et fusionné jusqu'à ce que tout le gaz soit réionisé environ 1 milliard d'années après le Big Bang. Se pose alors la question de quand la Réionisation a eu lieu exactement, quelles ont été les sources de réionisation, ou encore comment ont évolué temporellement les bulles ionisées. Pour y répondre j'étudie la topologie du processus de réionisation au travers du champ de temps de réionisation ( $t_{\text{reion}}(\vec{r})$ ). Il contient les instants auxquels l'Hydrogène a été réionisé dans chaque cellule et peut être extrait des simulations cosmologiques faites avec **EMMA** (Aubert et al. 2015) ou des modèles semi-analytiques faits avec **21cmFAST** (Mesinger et al. 2011). Dans cette thèse, je vais montrer comment ce champ contient des informations spatiales et temporelles sur le processus de réionisation.

De plus en plus de télescopes ont aujourd'hui pour but de sonder la Réionisation. Par exemple, l'Hydrogen Epoch of Reionization Array (HERA), New Extension in Nançay Upgrading LOFAR (NenuFAR) ou le Square Kilometer Array (SKA) cherchent à détecter le signal 21 cm, émis par le gaz d'Hydrogène neutre. Avec SKA, on espère par exemple obtenir des cartes 2D de ce signal sur le plan du ciel et à plusieurs fréquences ou redshifts d'observation. Comme le champ  $t_{\text{reion}}(\vec{r})$  n'est pas une observable directe, nous avons utilisé dans Hiegel et al. (2023) un réseau de neurones pour le reconstruire à partir de cartes 2D du signal 21 cm à un redshift donné. Le réseau a de bonnes performances : les reconstructions sont fiables et nous permettent d'extraire l'histoire de réionisation par exemple, même si les petites échelles en sont lissées. Des reconstructions un peu moins performantes car d'autant plus lissées ont également été obtenues en ajoutant du bruit instrumental de SKA aux cartes de signal 21 cm.

J'ai ensuite étudié la topologie de  $t_{\text{reion}}(\vec{r})$  pour comprendre le processus évolutif de la Réionisation. J'ai donc effectué diverses mesures de statistiques topologiques dans ce champ en utilisant notamment le code de topologie **DisPerSE** (Sousbie 2011). Ce code utilise la théorie de Morse et l'homologie persistante pour extraire de multiples objets topologiques d'un champ, comme ses points critiques ou son squelette. Dans Thélie et al. (2022), j'ai mesuré l'influence radiative des sources avec les patches de réionisation (régions associées aux sources), et j'ai montré qu'ils sont majoritairement allongés et alignés aux filaments de matière. Cette configuration semble être dû à des sources perlées le long des filaments, empêchant ainsi les configurations en papillons, c'est-à-dire avec des patches de réionisation perpendiculaires aux filaments où le rayonnement suit les directions les moins denses. Dans Thélie et al. (2023), j'ai revisité l'interprétation physique de la topologie de  $t_{\text{reion}}(\vec{r})$  en associant l'histoire de réionisation globale, les sources de réionisation, les fronts d'ionisation qui s'en échappent, l'évolution des bulles ionisées et régions neutres, et la percolation des fronts à des concepts tels que les taux de remplissage, les gradients, les isocontours ou le squelette. J'ai enfin montré que ce champ peut être considéré en bonne approximation Gaussien aux grandes échelles, nous permettant ainsi d'accéder à un grand jeu de statistiques avec la seule connaissance du spectre de puissance de  $t_{\text{reion}}(\vec{r})$ .

---

## Abstract

During the Epoch of Reionisation (EoR), the first stars and galaxies of the Universe appeared and reionised the gas, mainly Hydrogen, forming ionised bubbles around them. These bubbles grow and merge until all the gas is reionised approximately 1 billion years after the Big Bang. This raises the question of when exactly the EoR takes place, what are the sources of reionisation, and how do the ionised bubbles evolve over time. To answer these questions, I am studying the topology of the reionisation process through the reionisation time field ( $t_{\text{reion}}(\vec{r})$ ). It contains the instants at which the Hydrogen is reionised in each cell and can be extracted from cosmological simulations made with EMMA (Aubert et al. 2015) or semi-analytical models made with 21cmFAST (Mesinger et al. 2011). In this thesis, I will show how this field contains spatial and temporal information about the reionisation process.

More and more telescopes are now being used to probe the EoR. For example, the Hydrogen Epoch of Reionization Array (HERA), the New Extension in Nançay Upgrading LOFAR (NenuFAR) and the Square Kilometer Array (SKA) are seeking to detect the 21 cm signal emitted by neutral hydrogen gas. With SKA, for example, we hope to obtain 2D maps of this signal on the plane of the sky and at several observation frequencies or redshifts. Since the reionisation time field is not a direct observable, in Hiegel et al. (2023) we use a neural network to reconstruct it from 2D maps of the 21 cm signal at a given redshift. The network performs well, with reconstructions being satisfactory and for example allowing us to extract the reionisation history even if the small scales are smoothed out. By adding SKA instrumental noise to the 21 cm signal maps, we obtain  $t_{\text{reion}}(\vec{r})$  maps that are slightly less well reconstructed, even more smoothed.

I then studied the topology of  $t_{\text{reion}}(\vec{r})$  to understand the evolution of Reionisation. I therefore measured various topological statistics in this field, in particular using the DisPerSE topology code (Sousbie 2011). This code uses Morse theory and persistent homology to extract multiple topological objects from a field, such as its critical points or its skeleton. In Th  lie et al. (2022), I measured the radiative influence of the sources with the reionisation patches (regions associated with the sources), and showed that they are predominantly elongated and aligned with the matter filaments. This configuration seems to be due to sources beaded along the filaments, thus preventing butterfly configurations, which corresponds to perpendicular reionisation patches to the filaments where the radiation follows the least dense directions. In Th  lie et al. (2023), I revisited the physical interpretation of the topology of  $t_{\text{reion}}(\vec{r})$  by associating the global reionisation history, the reionisation sources, the ionisation fronts escaping from them, the evolution of ionised bubbles and neutral regions, and the percolation of the fronts with concepts such as filling factors, gradients, isocontours or the skeleton. Finally, I have shown that this field can be considered Gaussian on large scales to a good approximation, allowing us to access a large set of statistics with knowledge of the  $t_{\text{reion}}(\vec{r})$  power spectrum alone.

# TABLE DES MATIÈRES

Liste des publications	i
Remerciements	ii
Résumé	iii
Abstract	iv
Table des matières	v
<b>1 De l’Univers jeune aux galaxies actuelles</b>	<b>1</b>
1.1 Introduction . . . . .	3
1.1.1 Qu’est-ce que j’étudie ? . . . . .	3
1.1.2 Le modèle standard de la cosmologie . . . . .	4
1.2 L’Univers pré-Réionisation . . . . .	6
1.2.1 Le Fond Diffus Cosmologique . . . . .	6
1.2.2 Les Ages Sombres . . . . .	9
1.3 La Réionisation cosmologique . . . . .	10
1.3.1 L’Aube Cosmique . . . . .	10
1.3.2 Les sources de réionisation . . . . .	13
1.3.3 La réionisation du milieu intergalactique . . . . .	17
1.4 L’Univers post-Réionisation . . . . .	18
<b>2 Comment étudier la Réionisation cosmologique ?</b>	<b>19</b>
2.1 Observer la Réionisation de manière indirecte . . . . .	21
2.1.1 Détection de galaxies lointaines avec HST et JWST . . . . .	21
2.1.2 Le rayonnement Lyman- $\alpha$ . . . . .	25
2.1.3 Le fond diffus cosmologique . . . . .	27
2.1.4 Contraintes sur l’histoire de réionisation . . . . .	28
2.2 Observer la Réionisation de manière directe avec le signal 21 cm . . . . .	28
2.2.1 Le signal 21 cm . . . . .	29
2.2.2 Signal global . . . . .	32
2.2.3 Spectre de puissance . . . . .	33

2.2.4	Tomographie avec le radio-télescope SKA . . . . .	35
2.3	Simulations de la Réionisation . . . . .	38
2.3.1	Simulations cosmologiques : l'exemple d'EMMA . . . . .	39
2.3.2	Modèles semi-analytiques : l'exemple de 21cmFAST . . . . .	41
2.3.3	Temps/redshifts de réionisation . . . . .	44
2.4	Défis à venir et objectifs de la thèse . . . . .	46
<b>3</b>	<b>Topologie du processus de réionisation</b>	<b>48</b>
3.1	Géométrie des régions ionisées et neutres, percolation et phases de la Réionisation	50
3.1.1	Les fonctionnelles de Minkowski . . . . .	50
3.1.2	Distribution de la taille des bulles neutres et ionisées . . . . .	55
3.1.3	Analyse des structures spatiales avec les spectres du signal 21 cm . . . . .	56
3.1.4	Décompte de structures 3D avec les nombres de Betti . . . . .	59
3.2	Topologie avec la théorie de Morse et l'homologie persistante . . . . .	62
3.2.1	La théorie de Morse . . . . .	62
3.2.2	Le code de topologie DisPerSE . . . . .	65
3.2.3	Le paramètre de la persistance . . . . .	67
3.3	Topologie à travers les champs gaussiens aléatoires . . . . .	67
3.3.1	Champs aléatoires gaussiens et isotropes . . . . .	69
3.3.2	Longueurs typiques et paramètre spectral pour un spectre de puissance donné . . . . .	73
3.3.3	Fonction de distribution de probabilité du champ . . . . .	74
3.3.4	Fonction de distribution de probabilité de la norme du gradient du champ	74
3.3.5	Longueur d'isocontours . . . . .	74
3.3.6	Fonction de distribution de probabilité des points critiques du champ	75
3.3.7	Longueur de squelette . . . . .	76
<b>4</b>	<b>Reconstruction des temps de réionisation à partir du signal 21 cm</b>	<b>78</b>
4.1	Introduction . . . . .	80
4.2	Simulations 21cmFAST . . . . .	81
4.3	Réseau de neurones convolutionnel . . . . .	83
4.3.1	Architecture du réseau . . . . .	83
4.3.2	Procédé d'apprentissage . . . . .	85
4.3.3	Métrique interne au réseau . . . . .	86
4.4	Reconstructions des temps de réionisation et tests des performances . . . . .	88
4.4.1	Exemples de carte reconstruite . . . . .	88
4.4.2	Évaluation pixel par pixel des reconstructions . . . . .	89
4.4.3	Spectres de puissances . . . . .	92
4.4.4	Fractions de volume neutre . . . . .	92
4.5	Premières reconstructions des temps de réionisation à partir de cartes du signal 21 cm bruitées . . . . .	93
4.5.1	Ajout de bruit dans les cartes de signal 21 cm . . . . .	93
4.5.2	Nouvelles reconstructions . . . . .	95
4.6	Conclusions & perspectives . . . . .	96
4.7	« Reionisation time fields reconstruction from 21 cm signal maps » . . . . .	98



<b>5</b>	<b>Topologie des « patches » de réionisation durant l'Époque de la Réionisation</b>	<b>115</b>
5.1	Introduction . . . . .	117
5.2	Simulations 21cmFAST . . . . .	118
5.3	Géométrie et orientation des patches de réionisation . . . . .	119
5.3.1	Nombre et taille des patches de réionisation . . . . .	120
5.3.2	Forme des patches de réionisation . . . . .	121
5.3.3	Orientation des patches de réionisation par rapport aux filaments de matière . . . . .	124
5.4	La persistance : un outil pour suivre la percolation des patches de réionisation	129
5.5	Comparaisons entre simulations cosmologiques et semi-analytiques . . . . .	131
5.5.1	Simulations EMMA . . . . .	131
5.5.2	Mesures topologiques des patches de réionisation . . . . .	132
5.6	Conclusions & perspectives . . . . .	133
5.7	« First look at the topology of reionisation redshifts in models of the epoch of reionisation » . . . . .	135
<b>6</b>	<b>Concepts et mesures de la topologie de l'évolution de la réionisation</b>	<b>149</b>
6.1	Introduction . . . . .	151
6.2	Extraction et traitement des temps de réionisation d'une simulation EMMA . .	152
6.3	Évolution du processus de réionisation au travers des temps de réionisation .	154
6.3.1	Histoire de réionisation . . . . .	154
6.3.2	Vitesse des fronts d'ionisation . . . . .	156
6.3.3	Évolution de la taille des bulles ionisées et régions neutres . . . . .	158
6.3.4	Percolation des fronts d'ionisation . . . . .	160
6.3.5	Décompte du nombre de sources de réionisation . . . . .	162
6.4	Gaussianité du champ de temps de réionisation . . . . .	163
6.4.1	Comment procédons-nous ? . . . . .	164
6.4.2	Étude de la gaussianité des temps de réionisation . . . . .	166
6.5	Comparaisons entre les temps et les redshifts de réionisation . . . . .	169
6.5.1	Temps et redshifts de réionisation . . . . .	169
6.5.2	Comparaisons des mesures topologiques . . . . .	169
6.6	Comparaisons entre simulations cosmologiques et semi-analytiques . . . . .	171
6.6.1	Simulation 21cmFAST . . . . .	171
6.6.2	Comparaisons des mesures topologiques . . . . .	172
6.7	Conclusions & perspectives . . . . .	172
6.8	« Topology of reionisation times : Concepts, measurements, and comparisons to Gaussian random field predictions » . . . . .	178
<b>7</b>	<b>Conclusions et perspectives</b>	<b>200</b>
	<b>Bibliographie</b>	<b>204</b>

# CHAPITRE 1

## DE L'UNIVERS JEUNE AUX GALAXIES ACTUELLES

*Ma thèse, et donc ce manuscrit, porte une sur époque spécifique de l'histoire de notre Univers, qui s'appelle l'Époque de la Réionisation. Pour l'étudier, je me place dans un contexte cosmologique bien défini que je vais rappeler dans une première partie. Dans un second temps, je vais expliquer comment l'Univers a évolué jusqu'à la Réionisation cosmologique. Ensuite, je détaille un peu plus précisément cette époque en apportant des précisions sur les sources de Réionisation et l'évolution de l'ionisation de l'Hydrogène. Enfin, je décris rapidement l'Univers tel qu'il est après cette dernière grande transition.*

## Sommaire

---

<b>1.1</b>	<b>Introduction</b>	<b>3</b>
1.1.1	Qu'est-ce que j'étudie?	3
1.1.2	Le modèle standard de la cosmologie	4
<b>1.2</b>	<b>L'Univers pré-Réionisation</b>	<b>6</b>
1.2.1	Le Fond Diffus Cosmologique	6
1.2.2	Les Ages Sombres	9
<b>1.3</b>	<b>La Réionisation cosmologique</b>	<b>10</b>
1.3.1	L'Aube Cosmique	10
1.3.2	Les sources de réionisation	13
1.3.3	La réionisation du milieu intergalactique	17
<b>1.4</b>	<b>L'Univers post-Réionisation</b>	<b>18</b>

---

## 1.1 Introduction

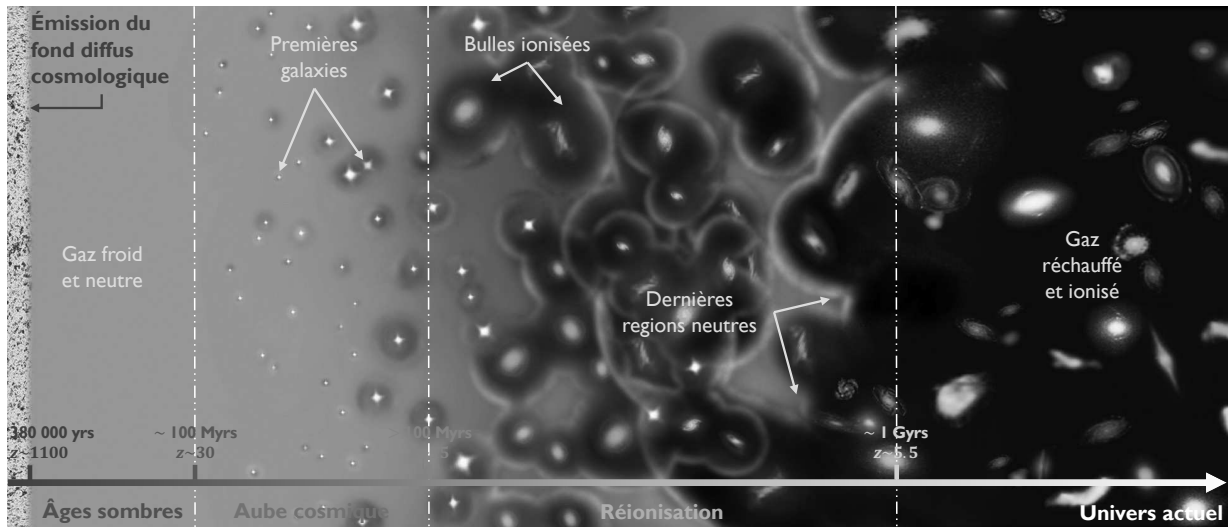
### 1.1.1 Qu'est-ce que j'étudie ?

Ma thèse porte sur une époque spécifique de l'histoire de notre Univers, qui s'appelle l'*Époque de la Réionisation*. Celle-ci marque la dernière grande transition que notre Univers a connu. Comme on le voit dans la frise de la Figure 1.1, cette époque est précédée de l'émission du Fond Diffus Cosmologique (Cosmic Microwave Background, CMB), après laquelle le gaz de l'Univers est neutre et froid. On parle alors de l'époque des Âges Sombres. Quand les premières étoiles et galaxies apparaissent, le gaz se réchauffe et l'Univers entre dans l'Aube Cosmique. Après cela, les galaxies émettent un rayonnement qui va ioniser le gaz et les premières bulles ionisées vont apparaître autour de ces galaxies. Ce processus commence quelques centaines de millions d'années après le Big Bang (soit à un redshift d'environ 15), et on entre dans l'Époque de la « Ré »-ionisation. Quand cette époque se termine, environ un milliard d'années après le Big Bang (ou vers un redshift de 5.5), le gaz est entièrement ionisé pour la deuxième fois dans l'histoire de l'Univers. Par ailleurs, il est chauffé à environ  $10^4$  K.

Le CMB a été découvert en 1965 par Penzias & Wilson (1965) et nous savons qu'après son émission, le gaz de l'Univers est sous forme d'atomes neutres pour la première fois (avec une prédominance de l'Hydrogène). Cependant, la même année, Gunn & Peterson (1965) étudient le spectre d'un quasar à un redshift de 2, et s'attendent à voir des raies d'absorption : même une faible fraction de gaz neutre aurait dû suffire à absorber le rayonnement UV du quasar. Or, ne détectant pas ou très peu de raies d'absorption, cela signifiait que le gaz devait être ionisé pour qu'il soit transparent aux rayonnements UV. Ils ont donc découvert qu'à ce redshift-là, le gaz de l'Univers était hautement ionisé, nous fournissant une première preuve observationnelle d'une réionisation du gaz de l'Univers. Aujourd'hui, nous disposons de plusieurs observations (dont je discuterai dans le Chapitre 2) qui contraignent le temps de fin de réionisation à environ un milliard d'années après le Big Bang (soit à un redshift  $z \sim 5.5$ , comme on le voit sur la frise).

Beaucoup de questions se posent sur cette époque, notamment parce qu'elle a longtemps été peu contrainte par les observations (comme nous le verrons dans le Chapitre suivant). On se demande par exemple comment l'Hydrogène s'est réchauffé et réionisé ? Quelles en sont les sources ? Quand est-ce que cette époque commence et se termine ? Quelle est la topologie des bulles ionisées ? Comment les galaxies impactent la Réionisation et en retour comment la Réionisation impacte les galaxies et la formation des structures ? Par ailleurs, comme nous le verrons dans les Sections suivantes, les premières étoiles et galaxies sont créées pendant l'Aube Cosmique par effondrement du gaz dans les halos de matière noire. Pendant la Réionisation, les galaxies ont continué à se former et les grandes structures de l'Univers, à l'intérieur desquelles la lumière est produite, sont apparues. Comprendre la Réionisation c'est donc aussi comprendre le phénomène de la formation des structures.

Pendant ma thèse, j'ai exploré ce processus de réionisation, en m'intéressant surtout au scénario global de réionisation plutôt qu'aux mécanismes individuels à l'oeuvre durant cette époque. Dans des modèles de Réionisation, je cherche à comprendre comment le rayonnement ionisant s'est propagé pendant tout ce processus, par quoi il a été émis et quel est le déroulé temporel du processus. Je me demande par exemple comment les bulles ionisées ont évolué



**FIGURE 1.1** – Frise chronologique des époques de l’Univers depuis l’émission du CMB jusqu’à aujourd’hui. On voit notamment après cette émission que le gaz est neutre et froid pendant les Âges Sombres. Puis, pendant l’Aube Cosmique, les premières étoiles et galaxies apparaissent, et le gaz se réchauffe. Les premières bulles ionisées apparaissent autour des galaxies pendant la Réionisation, et les dernières régions neutres disparaissent à la fin de cette époque. De nos jours, le gaz est réchauffé et complètement réionisé. Cette frise provient du site web de l’Agence Spatiale Européenne (ESA) et a été annotée par la suite.

dans le temps, comment elles grossissent, fusionnent et percolent. Est-ce que cela est corrélé avec la distribution de matière sous-jacente? En d’autres termes, j’étudie la topologie du processus de réionisation et de son évolution.

### 1.1.2 Le modèle standard de la cosmologie

La Réionisation est un processus qui a lieu à l’échelle entière de l’Univers. Mais étudier tout l’Univers est une tâche a priori impossible parce qu’il est très vaste, sauf à supposer que le « principe cosmologique » est valide. Il suppose que ses propriétés sont les mêmes pour tous les observateurs, peu importe où ils se placent dans l’espace et comment ils l’observent, tant que les échelles sondées sont assez grandes ( $> 200$  Mpc). À ces échelles, l’Univers est donc homogène et isotrope. Ce principe est fondamental ici parce qu’il nous permet d’analyser une portion de l’Univers à des échelles assez grandes en supposant que les propriétés que l’on en tire seront les mêmes peu importe où l’on se place.

Comme la Réionisation s’étale sur des périodes cosmologiques (près d’un milliard d’années), son évolution dépend fortement de l’expansion de l’Univers. Celle-ci a été découverte par Hubble et Lemaître dans les années 1920 en notant que les galaxies les plus distantes s’éloignaient plus rapidement de nous. Cette relation est linéaire :

$$v = H_0 d, \quad (1.1)$$

et a une pente donnée par le paramètre  $H_0 \sim 70$  km/s/Mpc aujourd’hui. On l’appelle le paramètre de Hubble et il évolue dans le temps ( $H = H(t)$ ). Notons ici que tous les temps  $t$  dont

il sera question dans ce manuscrit sont définis comme les temps écoulés depuis la naissance de l'Univers. On estime, qu'à l'heure actuelle, l'Univers a  $t_0 = 13.8$  milliards d'années.

Cette expansion peut être décrite par la métrique FLRW (Friedmann-Lemaître-Robertson-Walker), élaborée dans les années 1920-1930 grâce aux lois de la relativité générale et en supposant que l'Univers est homogène et isotrope. On peut en obtenir une solution (connue sous le nom de l'Équation de Friedmann) que l'on peut intégrer pour avoir une équation de la dépendance temporelle du paramètre de Hubble :

$$H(t) = H_0 \left( \frac{\Omega_m}{a^3} + \frac{\Omega_r}{a^4} + \Omega_\Lambda - \frac{\Omega_K}{a^2} \right)^{\frac{1}{2}} = \frac{\dot{a}}{a}. \quad (1.2)$$

Elle se paramétrise avec le facteur d'échelle  $a$ , qui vaut aujourd'hui  $a_0 = 1$ . Le facteur d'échelle sert à exprimer les grandeurs physiques en prenant l'expansion en compte. Par exemple, pour les distances, on définit

$$r(t) = a(t)r_0(t), \quad (1.3)$$

où  $r$  est la distance physique et  $r_0$  est une distance que l'on appelle « distance comobile ». On a la même relation en terme de longueurs d'onde :

$$\lambda(t) = a(t)\lambda_0(t). \quad (1.4)$$

Le facteur d'échelle peut être relié à une grandeur que l'on utilise en cosmologie pour décrire à la fois le temps et les distances et que l'on appelle le « redshift »  $z$  (ou « décalage vers le rouge ») :

$$a(z) = \frac{1}{1+z}. \quad (1.5)$$

Plus  $z$  est grand, plus on sonde un Univers lointain et précoce,  $z = 0$  correspondant à aujourd'hui. Le redshift correspond à un « rougissement » des longueurs d'onde quand on regarde des objets plus lointains.

L'Équation 1.2 fait intervenir les paramètres de densité  $\Omega_i = \rho_i/\rho_c$ , qui sont définis pour différents fluides :  $\rho_i$  est la densité d'un fluide et  $\rho_c = 3H_0^2/8\pi G$  est la densité critique. On considère trois fluides. Le premier est un fluide non-relativiste, qui est froid et qui décrit la matière. Sa contribution à  $H(t)$  en  $\Omega_m \propto a^{-3}$  indique qu'il est dilué géométriquement quand l'Univers s'expand. La matière est de deux sortes : il y a la matière comme définie par le modèle standard de la physique des particules et que l'on appelle matière baryonique (de densité  $\Omega_b$ ), et il y a la matière noire (CDM pour cold dark matter, de densité  $\Omega_c$ ). Cette dernière est donc supposée froide, non-baryonique, n'émettre et n'absorber aucun rayonnement et n'interagir avec aucune autre particule hormis à travers la gravité. Le second fluide est un fluide relativiste (incluant la lumière), dont la contribution à  $H(t)$  est en  $\Omega_r \propto a^{-4}$ , et il est également dilué avec le redshift en plus d'être dilué géométriquement. Le dernier est le fluide d'énergie noire, de contribution à  $H(t)$  en  $\Omega_\Lambda \propto a^0$ , et que l'on assimile à la constante cosmologique  $\Lambda$  parce qu'il n'est pas dilué et reste uniforme pendant que l'Univers évolue. L'Équation 1.2 étant vraie pour  $a = 1$ , nous obtenons donc :  $1 + \Omega_K = \Omega_m + \Omega_r + \Omega_\Lambda$ , où  $\Omega_K$  est lié à la courbure intrinsèque de l'Univers et quantifie le départ à un Univers critique.

De plus, on se rend compte que l'Univers passe par plusieurs ères dans lesquelles l'un ou l'autre des fluides domine le budget d'énergie. L'Univers est d'abord dominé par le fluide

$\Omega_\Lambda$	$\Omega_c$	$\Omega_b$	$\Omega_r$	$\Omega_K$
0.69	0.26	0.05	$9.3 \times 10^{-5}$	0

TABLE 1.1 – Densité des fluides de l’Univers contraints par (Planck Collaboration et al. 2020).

relativiste jusqu’à environ 50 000 ans après le Big Bang (soit  $z \sim 3500$ ), et  $a(t) \propto t^{\frac{1}{2}}$ . Ensuite, l’Univers est dominé par la matière entre 50 000 ans et 8 milliards d’années après le Big Bang environ (soit entre  $z \sim 3500$  et  $z \sim 0.5$ ), et  $a(t) \propto t^{\frac{2}{3}}$ . Cette époque de domination de la matière est aussi connue sous le nom de régime Einstein-de Sitter. Enfin, l’Univers est dominé par l’énergie noire après 8 milliards d’années après le Big Bang (soit  $z \sim 0.5$ ), et subit une expansion accélérée avec  $a(t) \propto \exp(H_0 t)$ . L’Époque de Réionisation, que j’étudie pendant cette thèse, se trouve donc dans l’ère où l’Univers est dominé par la matière, tandis que nous sommes actuellement dans la dernière ère, celle qui est dominée par l’énergie noire.

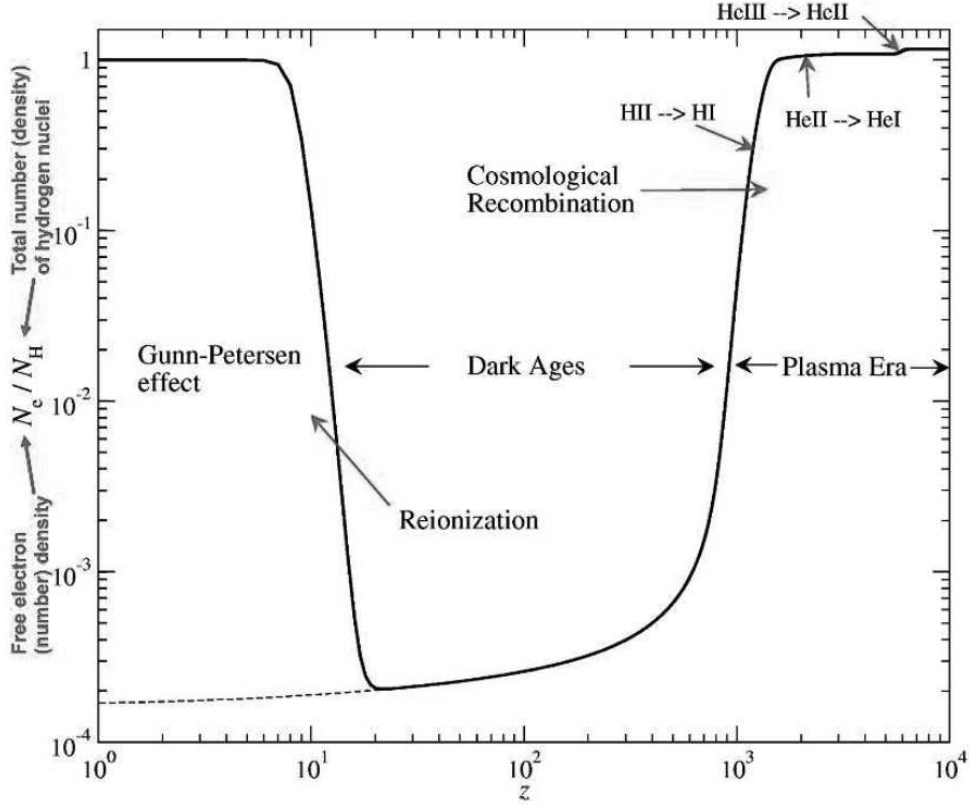
Le modèle cosmologique que je décris ici est le modèle standard de la cosmologie, que l’on appelle aussi modèle  $\Lambda$ CDM. Plusieurs observations – comme celles du CMB, du diagramme de Hubble des supernovae, du comptage des amas de galaxies, ou du spectre de puissance de la matière aux grandes échelles – ont contraint la densité des fluides de l’Univers dans le cadre de ce modèle (voir par exemple Eisenstein et al. (2005); Cole et al. (2005); Abbott et al. (2019)). Elles ont montré que l’Univers était effectivement plat et en expansion accélérée. Avec les dernières contraintes de Planck Collaboration et al. (2020) à partir du CMB, on estime qu’à l’heure actuelle  $H_0 \approx 68$  km/s/Mpc et que l’Univers est composé de 68% d’énergie noire, 26% de matière noire froide, 5% de matière baryonique (et donc 31% de matière), et  $9.3 \times 10^{-3}\%$  de radiation. On peut également exprimer ces pourcentages avec les paramètres de densité, pour lesquels je donne les valeurs dans la Table 1.1. Dans tout ce manuscrit, on suppose que l’Univers est décrit comme tel avec ces paramètres cosmologiques.

## 1.2 L’Univers pré-Réionisation

L’Époque de la Réionisation, étant la dernière grande transition que l’Univers a connu, est naturellement précédée par plusieurs époques. Le point le plus reculé de l’histoire de l’Univers est le *Big Bang*, et il est suivi d’une phase très rapide d’*inflation* (de  $10^{-35}$  s) pendant laquelle l’Univers s’expand rapidement. Ensuite, les premiers noyaux apparaissent environ 3 minutes après le Big Bang (*nucléosynthèse*). Après cela, l’Univers entre dans l’*Époque de la Recombinaison*, et c’est à partir de là que je commence ma description...

### 1.2.1 Le Fond Diffus Cosmologique

À la fin de la nucléosynthèse, les noyaux des atomes d’Hydrogène  $H$  (un simple proton), de Deutérium  $H_2$  (un proton et un neutron) et d’Hélium  $He$  (deux protons et deux neutrons) ont été créés. Comme les neutrons sont moins abondants (parce que légèrement plus massifs) que les protons, il existe moins de noyaux d’Hélium que de noyaux d’Hydrogène. On estime qu’environ 90% des noyaux sont des noyaux d’Hydrogène contre seulement 10% de noyaux



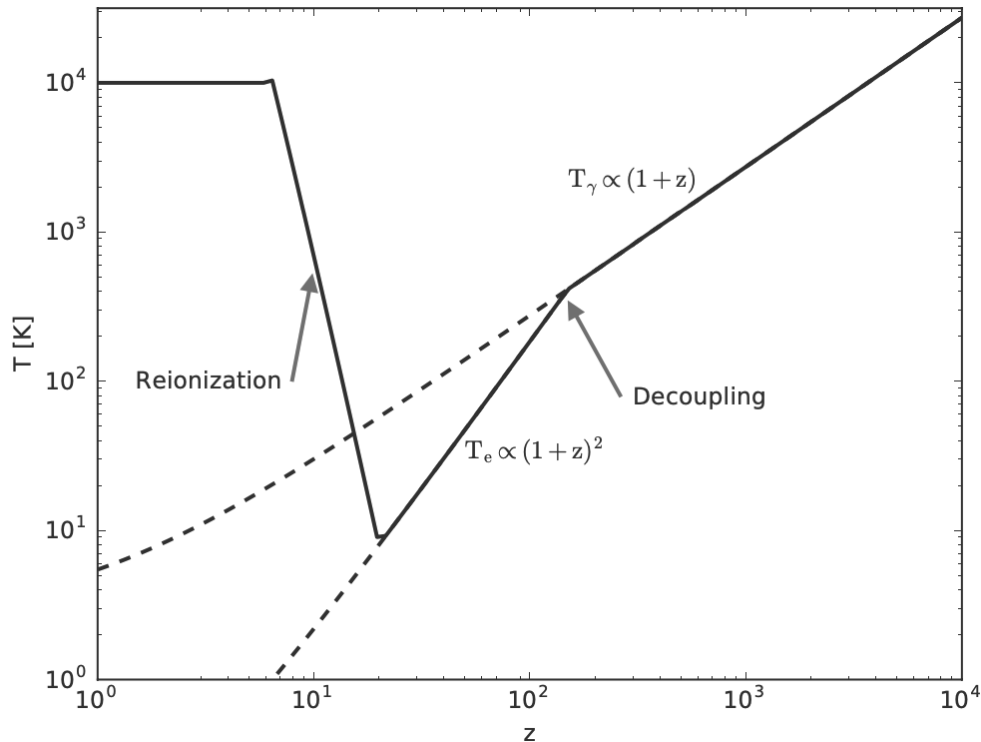
**FIGURE 1.2** – Schéma de l'histoire d'ionisation du gaz de l'Univers à haut redshift.  $x$  est le rapport entre la densité d'électrons libres et la densité d'atomes d'Hydrogène et représente la fraction de gaz ionisé dans l'Univers. On voit ici qu'à un redshift supérieur à 1100, le gaz de l'Univers est complètement réionisé. La recombinaison a ensuite lieu et pendant celle-ci, l'Helium puis l'Hydrogène recombine et le gaz devient complètement neutre. On est alors dans l'époque des Âges Sombres. Après cela, la Réionisation s'enclenche et le gaz se réionise, jusqu'à ce qu'il n'y ait plus de gaz neutre. Cette Figure est extraite de Sunyaev & Chluba (2009).

d'Hélium. Cette différence de proportion entre ces noyaux, et plus tard leurs atomes correspondants, justifie le fait que je m'intéresserai principalement à l'Hydrogène dans mes études sur la Réionisation, laissant de côté l'Hélium.

Ces noyaux, ainsi que des électrons, se déplacent alors librement dans l'Univers après la nucléosynthèse, et les photons sont fréquemment diffusés par les électrons libres qui n'ont pas encore recombinés : la matière et le rayonnement sont ainsi fortement couplés. Les noyaux vont être amenés à se combiner pour former les premiers atomes d'Hydrogène et d'Hélium de notre Univers, grâce aux réactions suivantes :





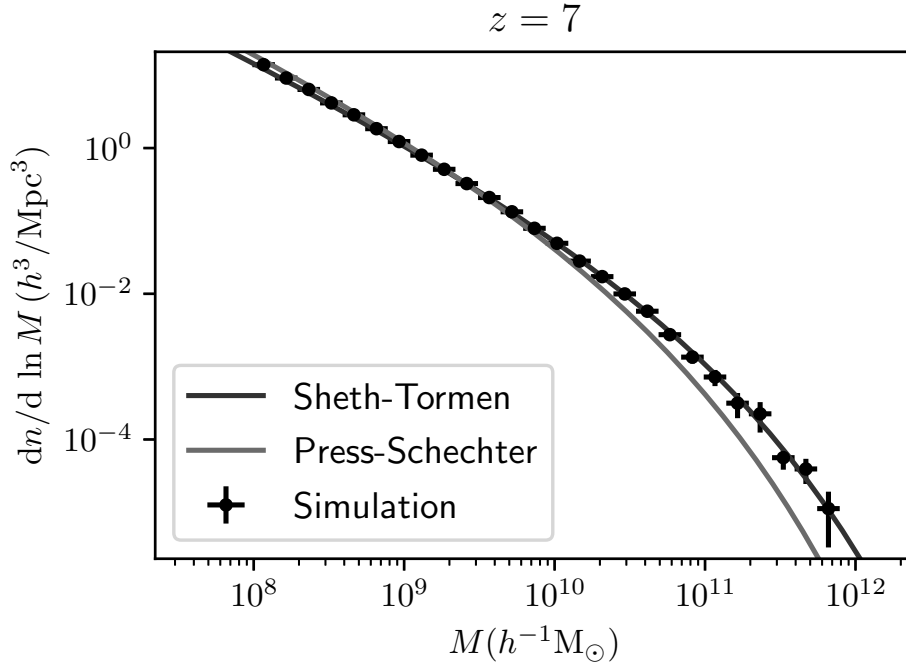


**FIGURE 1.3** – Schéma de l’histoire de température du gaz à haut redshift. On voit qu’à  $z > 150$ , la température du gaz est couplée à celle du CMB. À  $z \sim 150$ , elle s’en découple parce que la densité d’électrons libres devient trop faible pour qu’ils interagissent autant avec les photons du CMB. Le gaz suit alors l’expansion adiabatique et après les Âges Sombres (vers  $z > 30$ ), les premières étoiles réchauffent le gaz et le réionisent. Quand la Réionisation se termine, la température du gaz est de l’ordre de  $10^4$  K. Cette Figure est extraite de Gillet (2016) après avoir été adaptée de Loeb & Furlanetto (2013).

Ces réactions sont des réactions de recombinaison et on entre dans l’époque cosmologique du même nom.

Les schémas des Figures 1.2 et 1.3 montrent l’évolution de la fraction de gaz ionisé et de la température dans l’Univers depuis cette époque. On voit en particulier que la Recombinaison se termine à un redshift d’environ  $z \sim 1100$ , soit 380 000 ans après le Big Bang. C’est à ce moment, quand la température atteint environ  $T_\gamma \sim 3000$  K, que les atomes (d’Hydrogène et d’Hélium) deviennent abondants dans l’Univers, et que le plasma laisse sa place à un gaz neutre.

Tous les électrons sont « bloqués » dans les atomes, empêchant ainsi la diffusion Thomson des photons sur les électrons. On dit que cette diffusion est « gelée » et les photons primordiaux n’interagissent plus avec la matière à ce moment. L’Univers devient optiquement mince : on parle alors de *dernière diffusion* et les photons sont libres de se propager en ligne droite dans l’Univers. On mesure aujourd’hui ces photons avec ce qu’on appelle le *Fond Diffus Cosmologique* (Cosmic Microwave Background, CMB), et ils proviennent tous de la même époque et ont tous parcouru la même distance jusqu’à nous. Dans la frise chronologique de la Figure 1.1, cette surface de dernière diffusion des photons est représentée tout à gauche aux



**FIGURE 1.4** – Fonction de masse des halos modélisée par Press & Schechter (1974) en rouge, Sheth & Tormen (1999) en bleu et résultant d’une simulation avec les points noirs. On voit en particulier que les halos de masse  $10^8 M_{\odot}$  sont plus abondants que les halos les plus massifs. Cette Figure est extraite de Choudhury (2022).

couleurs du CMB mesuré par Planck Collaboration et al. (2020). On peut imaginer le CMB comme produit par une coquille, une surface 2D, qui nous entoure et qui est centrée sur nous : c’est la *surface de dernière diffusion*. La recombinaison n’est pas instantanée et cette coquille qui nous entoure a une petite extension spatiale le long de la ligne de visée, qui se traduit par un gradient important mais non infini de la courbe de la Figure 1.2 vers  $z \sim 1000$ . De plus, cette surface détoure l’Univers observable car au delà l’Univers est optiquement épais.

Les photons du CMB qui nous parviennent à l’heure actuelle proviennent de cette surface de dernière diffusion, mais ont été diffusés par des électrons libres pendant leur parcours. En particulier, quand la Réionisation a lieu, l’Univers contient à nouveau des électrons libres, sur lesquels les photons vont pouvoir être diffusés. Ainsi, à partir du CMB, on peut mesurer l’opacité des électrons libres aux photons du CMB le long de la ligne de visée, et c’est ce qu’on appelle la profondeur optique de réionisation  $\tau$ . Celle-ci permet alors de sonder l’état d’ionisation du milieu intergalactique (intergalactic medium, IGM). Je reviendrai plus en détail sur cela dans le Chapitre 2.

### 1.2.2 Les Ages Sombres

Après l’émission du CMB (380 000 ans après le Big Bang ou vers  $z \sim 1100$ ), l’Univers entre alors dans une période que l’on appelle les *Âges Sombres* (voir la frise de la Figure

1.1), qui porte ce nom parce que l'Univers est encore dépourvu d'étoiles (et de sources de lumière astrophysique en général). À cette époque, le gaz est refroidi, neutre, et découplé du rayonnement. Dans la Figure 1.3, on peut suivre l'évolution de la température du gaz avec le redshift, et on voit qu'au début de cette époque, elle est d'abord faiblement couplée à celle du CMB par l'effet de diffusion Compton, et décroît comme  $a^{-1}$  avec l'expansion. Par la suite, elle diminue de manière adiabatique en  $a^{-2}$  à mesure que l'Univers s'expand et tant qu'il n'y a pas de source de réchauffement.

C'est pendant cette époque que les « halos » de matière noire commencent à s'assembler. Ce sont des structures denses liées par la gravitation et ainsi capables d'attirer gravitationnellement les baryons pour former des nuages denses de gaz. Ils sont alors souvent désignés comme étant les « blocs fondateurs » des objets et structures que nous connaissons aujourd'hui parce que c'est en leur sein que les étoiles et galaxies se sont formées. Les fluctuations de densité de matière initiales ont permis aux premiers halos de se former par instabilité gravitationnelle. La formation de ces halos est généralement supposée suivre le modèle hiérarchique de formation des structures (White & Rees 1978). Ce modèle indique que des halos plus massifs se forment par fusion de halos moins massifs, que l'on appellera progéniteurs. L'abondance de halos présents dans l'Univers par intervalle de masse et de volume peut être estimée grâce à la fonction de masse des halos (Halo Mass Function, HMF), que je représente dans la Figure 1.4. Cette dernière est une distribution de masse des halos, elle est historiquement modélisée par Press & Schechter (1974) ou de Sheth & Tormen (1999), et elle indique que l'abondance de halos décroît avec leur masse.

## 1.3 La Réionisation cosmologique

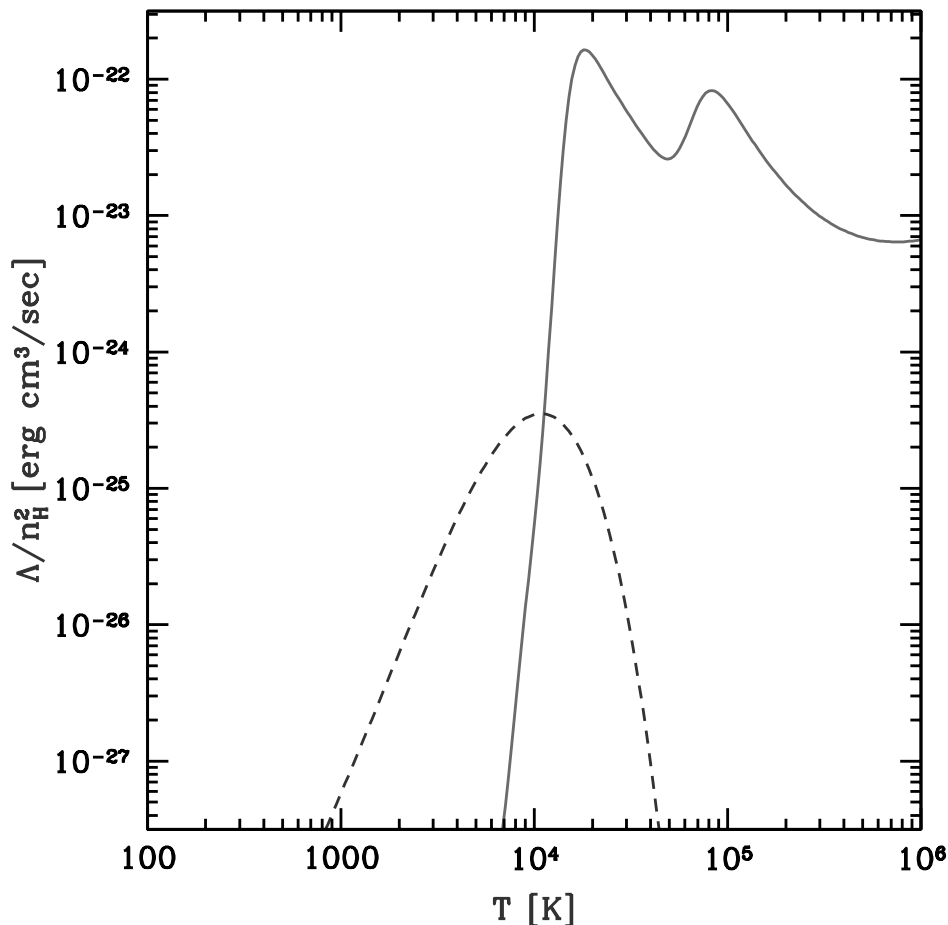
Après les Âges Sombres, et environ 100 millions d'années après le Big Bang (ou vers un redshift  $z \sim 30$ , voir la frise de la Figure 1.1), de nouveaux objets apparaissent dans l'Univers marquant le début de l'*Aube Cosmique* et les prémices de l'*Époque de la Réionisation*.

### 1.3.1 L'Aube Cosmique

Au sein des halos de matière noire, des sur-densités de gaz sont présentes et vont se contracter par instabilité gravitationnelle. Se faisant, la température du gaz augmente et sa pression va contrebalancer la gravité. On peut définir la masse de Jeans  $M_J$  comme suit :

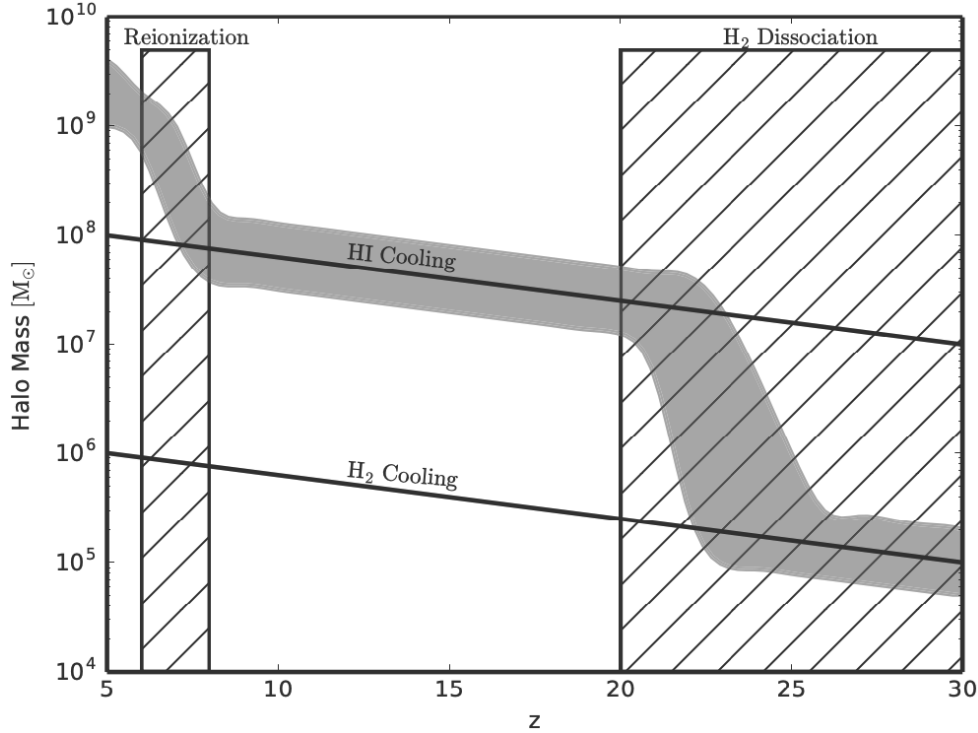
$$M_J \approx \frac{4}{3}\pi\lambda_J^3\rho, \quad (1.8)$$

où  $\rho$  est la densité du gaz dans le nuage, et  $\lambda_J \propto c_s/\sqrt{G\rho}$  est la longueur de Jeans. Cette dernière dépend de la vitesse du son  $c_s$ , qui croît avec la température. Quand la masse du nuage de gaz atteint la masse de Jeans, les instabilités gravitationnelles sont plus fortes que la pression thermique, et l'effondrement du nuage devient possible. Le gaz doit alors être assez froid pour que sa pression thermique diminue, que sa masse de Jeans baisse et qu'il s'effondre pour former des étoiles. Pour que cela soit possible, il doit évacuer de l'énergie et cela passe par le rayonnement émis par l'Hydrogène atomique  $H$  et moléculaire  $H_2$ . Dans la Figure 1.5, on voit en rouge le taux de refroidissement par l'Hydrogène atomique en



**FIGURE 1.5** – Taux de refroidissement en fonction de la température pour l’Hydrogène atomique HI et l’Hélium en traits pleins rouges et pour l’Hydrogène moléculaire  $H_2$  en tirets bleus. On remarque ici qu’une température seuil à  $10^4$  K apparaît comme limite à partir de laquelle le refroidissement atomique devient efficace. À plus basse température, le refroidissement est possible grâce à l’Hydrogène moléculaire. Cette Figure est extraite de Barkana (2016).

fonction de la température. Une température seuil de  $10^4$  K apparaît comme étant la limite à partir de laquelle le refroidissement atomique devient efficace. En bleu, on voit que le refroidissement d’un nuage de gaz est possible à plus basse température grâce à l’Hydrogène moléculaire. Pour que ce refroidissement ait lieu, il faut que le nuage de gaz ait atteint des températures suffisamment élevées. La température atteignable par le gaz est dictée par la masse du halo environnant, et quand le halo s’effondre, il se virielise et le gaz atteint la température de Viriel  $T_{\text{vir}}$ . Cette température est ainsi liée à la masse du halo qui s’effondre par la relation  $T_{\text{vir}} \propto M_{\text{min}}^{2/3}$  et on appelle cette dernière la masse de Viriel (Choudhury 2022). La température d’enclenchement du refroidissement peut alors être interprétée en terme de masse seuil à partir de laquelle un halo peut former des étoiles. Dans la Figure 1.6 extraite de Gillet (2016), on voit en rouge et en fonction du redshift les masses de halos auxquelles le



**FIGURE 1.6** – Schéma représentant le seuil de masse de halo minimum à différents redshifts pour que la formation stellaire s’enclenche en rouge. Ce schéma permet de voir approximativement les masses seuils auxquelles le refroidissement de l’Hydrogène atomique et moléculaire ont lieu. On voit notamment qu’au début de l’Aube Cosmique (entre  $z = 30$  et  $z = 20$  environ), le refroidissement moléculaire est en place dans les halos peu massifs. Ensuite, pendant la fin de l’Aube Cosmique et le début de la Réionisation (entre  $z = 20$  et  $z \sim 7.5$ ), c’est le refroidissement atomique qui permet majoritairement de former des étoiles dans les halos les plus massifs, suite à la suppression du refroidissement moléculaire après la dissociation de l’ $H_2$  par le rayonnement des premières étoiles. Cette Figure est extraite de Gillet (2016).

refroidissement du nuage de gaz se produit. Le refroidissement par l’Hydrogène atomique se produit à des températures plus élevées que le refroidissement par l’Hydrogène moléculaire, et cela se traduit dans le même sens en terme de masse : les halos de faibles masses ( $\sim 10^6 M_\odot$ ) vont pouvoir se refroidir par l’Hydrogène moléculaire, tandis que les halos de masse plus élevée ( $\sim 10^8 M_\odot$ ) se refroidissent par l’Hydrogène atomique. Les halos refroidis grâce à des réactions atomiques (de l’Hydrogène principalement) sont appelés « halos refroidis atomiquement », et ceux refroidis avec des réactions moléculaires sont appelés « halos refroidis moléculairement ». Par ailleurs, la présence de métaux dans le gaz par la suite pourra impacter le refroidissement et notamment le faciliter.

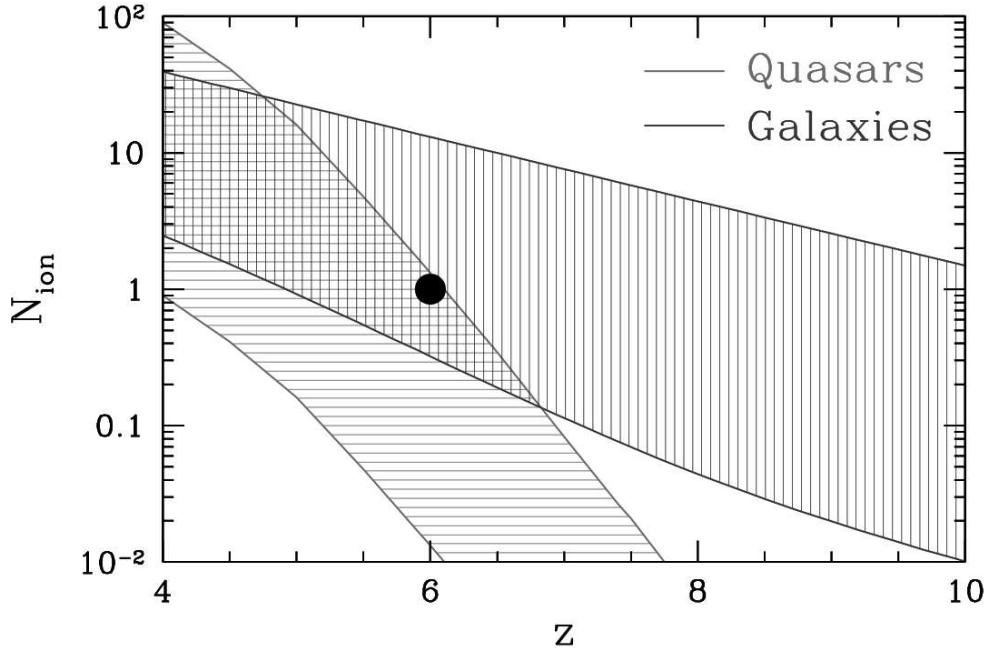
Les conditions de température et de masse permettant l’effondrement des nuages de gaz dans les halos sont atteintes environ 100 millions d’années après le Big Bang (soit vers  $z \sim 30$ ). À ce moment, les premiers effondrements gravitationnels se produisent, et cela marquent le début de l’Aube Cosmique. La naissance des premières étoiles et galaxies a ainsi lieu pendant cette époque, notamment dans les petits halos refroidis moléculairement. On

suppose que les premières étoiles à apparaître sont celles formant une population que l'on appelle « Population III » et qu'elles sont composées des éléments présents dans l'Univers à cette période-là, c'est-à-dire d'éléments légers (Hydrogène et Hélium). Ces étoiles sont issues d'un milieu sans métaux, entravant ainsi le refroidissement des nuages de gaz initiaux qui s'effondrent, et formant alors des étoiles relativement massives (Bromm et al. 2001; Abel et al. 2002). Ces premières sources vont également dissocier l'Hydrogène moléculaire  $H_2$  et couper ce canal de refroidissement, comme on le voit dans la Figure 1.5. Elles peuvent exploser en supernovae à la fin de leur vie et ainsi enrichir chimiquement leur environnement. Cet environnement enrichi va alors commencer à former des étoiles « métalliques ». Cela donnera naissance à la « Population I » et à la « Population II », qui contiennent des étoiles avec une richesse en métaux inférieure ou égale à celle du Soleil. Avec la présence de ces éléments plus massifs dans le gaz, le refroidissement du gaz est ainsi facilité et les premières galaxies peuvent se former dans les halos les plus massifs. Les étoiles très massives de la Population III vont terminer leur vie en trous noirs, qui émettront notamment des rayons X au sein de systèmes binaires (Wise 2019). Ce type de rayonnement, auquel s'ajoute également la contribution des noyaux actifs de galaxies (active galactic nuclei, AGNs), est supposé causer le réchauffement du gaz à mi-Aube Cosmique. Ce réchauffement se poursuit jusqu'à ce que l'on entre dans l'Époque de la Réionisation vers  $z \sim 15$  pendant laquelle le gaz change d'état d'ionisation à cause du rayonnement UV.

La répartition de masses des étoiles après l'effondrement d'un nuage de gaz dans un halo est en général décrite par la fonction de masse initiale (initial mass function, IMF). Il existe plusieurs modélisations de l'IMF (Salpeter 1955; Kroupa 2002) même si sa forme est encore débattue. On ne connaît pas cette répartition pour les premières étoiles formées, c'est-à-dire celles de la Population III, mais les simulations semblent montrer qu'elles étaient plus massives que celles d'aujourd'hui (Bromm & Larson 2004). L'IMF a donc une influence sur les différentes populations d'étoiles et en particulier sur la quantité de photons ionisants émis par les étoiles.

### 1.3.2 Les sources de réionisation

Une des questions importantes débattue quand on étudie la Réionisation concerne les sources qui produisent le rayonnement ionisant. Nous ne sommes pas sûrs actuellement du type d'objets dominant le budget de photons ionisants de l'Univers. Dans la Figure 1.7, des plages possibles de valeurs du nombre d'ionisations par atome d'Hydrogène sont représentées en considérant la contribution des quasars en rouge et celle des galaxies en bleu. On voit ici qu'à haut redshift (pendant la Réionisation), les quasars/AGNs sont moins efficaces que les galaxies pour réioniser l'IGM. Nous savons donc que ce ne sont pas des sources majoritaires de réionisation parce qu'ils ne sont pas présents à des densités suffisantes pour contribuer assez significativement au budget de photons de la Réionisation (Grissom et al. 2014; Kashikawa et al. 2015). À l'inverse, on peut voir que les galaxies contribuent majoritairement à ce budget au début de la Réionisation. On se demande notamment si les étoiles de la Population III, qui se forment pendant toute l'Aube Cosmique et la Réionisation, pourraient être les sources que l'on recherche. Elles ont une durée de vie assez courte (quelques millions d'années), mais comme elles sont assez massives, elles libèrent beaucoup (entre 20% et 90% selon leur masse) de photons ionisants (Schaerer 2002). Cependant, dès que le milieu s'enrichit en métaux, ces

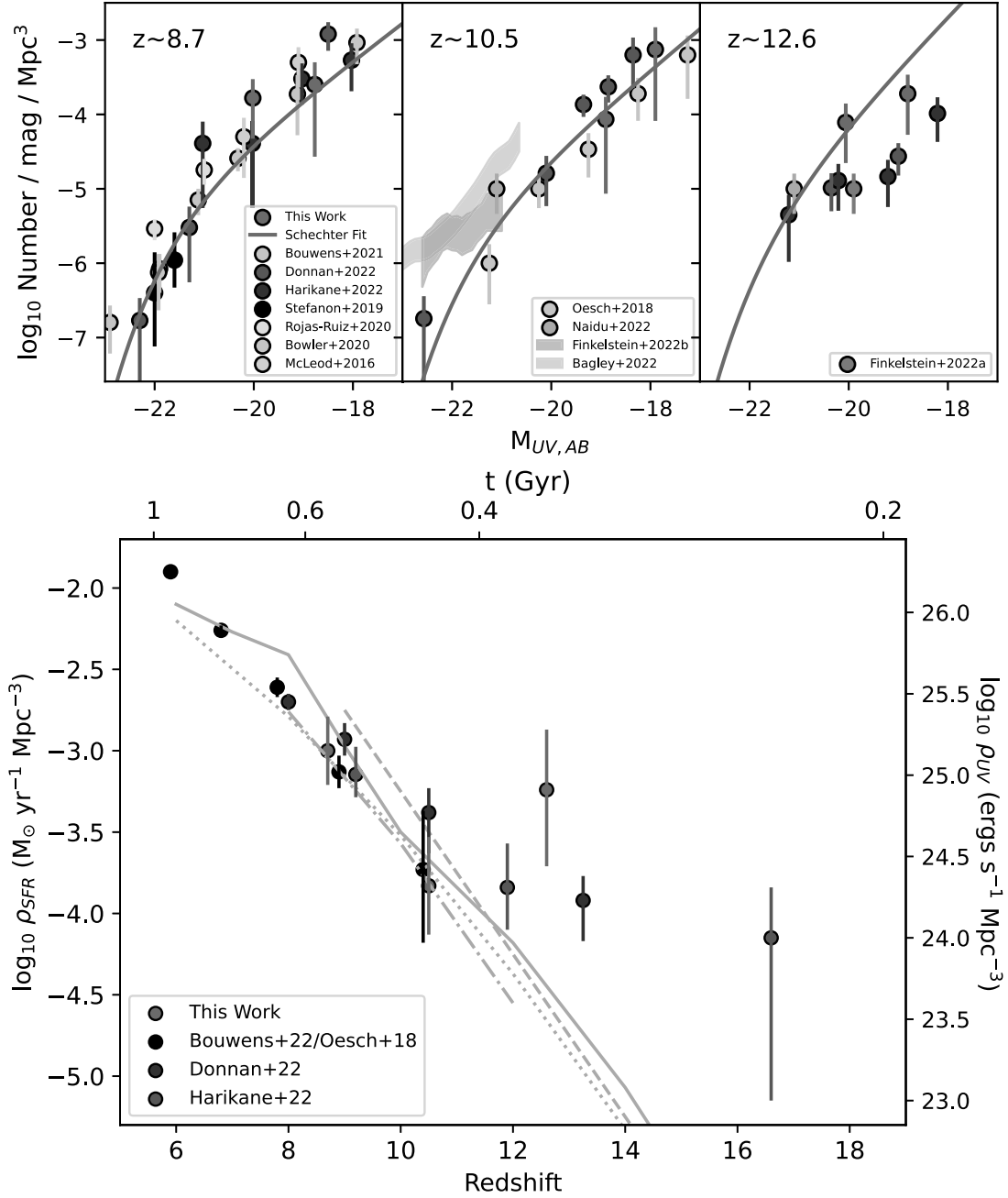


**FIGURE 1.7** – Plages possibles de valeurs pour le nombre d’ionisations par atome d’Hydrogène. En bleu, on voit la contribution des galaxies et en rouge, celle des quasars. Le point noir est une contrainte provenant de la réionisation de l’Hydrogène, indiquant que pour réioniser le gaz à  $z = 6$ , il faut au moins un photon produit par atome. On voit ici que ce sont les galaxies qui contribuent le plus à la Réionisation par rapport aux quasars. Cette Figure est extraite de Trac & Gnedin (2011).

étoiles sont de moins en moins nombreuses et leur contribution à la Réionisation est souvent débattue. D’autres études suggèrent que des sources plus exotiques, comme les étoiles binaires sources de rayons X, pourraient également contribuer à la Réionisation (Jeon et al. 2014).

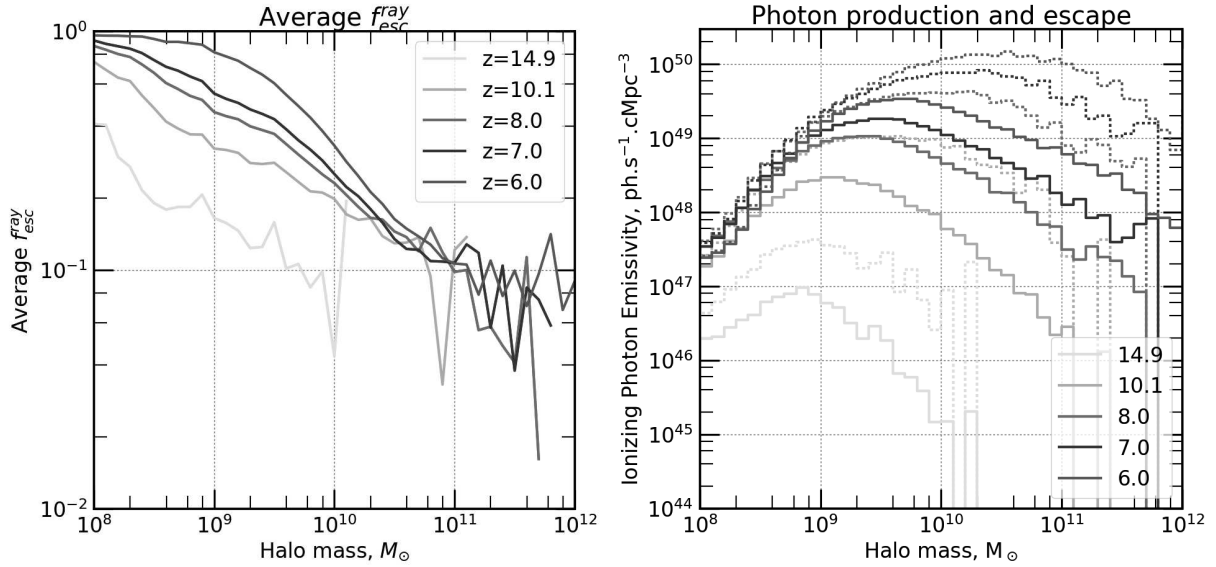
Ces incertitudes nous amènent à nous demander quelle est l’abondance des galaxies à haut redshift, et comment autant de photons ionisants ont pu s’en échapper pour réioniser l’IGM. Nous nous intéressons alors à la fonction de luminosité des galaxies, comme celles représentées en haut de la Figure 1.8 à différents redshifts. Ces fonctions indiquent le nombre de galaxies présentes à un certain redshift par tranche de la luminosité qu’elles émettent (voir par exemple les études récentes de Bouwens et al. (2023) et Harikane et al. (2023a,b)). L’HMF montre de plus qu’il y a plus de halos de faible masse qu’inversement. Naïvement, on imagine alors qu’un grand nombre de galaxies peu brillantes et peu massives ont dominé le budget de photons ionisants, tandis que les galaxies très massives et très brillantes seraient en minorité et donc en incapacité à dominer ce budget.

Ensuite, on se sait pas quelle quantité de photons ionisants s’échappent effectivement des galaxies. Une manière d’étudier ce problème peut se faire par le biais de la fraction d’échappement des galaxies  $f_{\text{esc}}$ . Les photons produits par les étoiles des galaxies peuvent être absorbés par le milieu interstellaire et  $f_{\text{esc}}$  permet d’indiquer le nombre de photons qui s’échappent des galaxies sources. Dans la Figure 1.9, on voit à gauche un exemple de fraction d’échappement en fonction de la masse des halos à différents redshifts modélisé pour



**FIGURE 1.8** – Fonctions de luminosité des galaxies à différents redshifts en haut. En bas, on peut voir les densités de taux de formation d'étoiles (axe de gauche) et de luminosité UV (axe de droite). Sur ces figures, les points représentent des contraintes observationnelles, et les lignes proviennent de diverses modélisations. Ces figures indiquent qu'il y a plus de galaxies ayant une faible luminosité et de halos de faibles masses pendant la Réionisation qu'inversement, ce qui pourrait indiquer naïvement que le budget de photons ionisants est dominé par un grand nombre de galaxies peu brillantes et peu massives. Ces Figures sont extraites de Bouwens et al. (2023).





**FIGURE 1.9** – Fraction d’échappement moyenne des halos formant des étoiles en fonction de leur masse et du redshift (voir les différentes couleurs) à gauche. À droite, on voit le nombre de photons s’échappant des galaxies par seconde et par unité de volume en fonction de la masse et du redshift. Ces figures indiquent que les halos de masses intermédiaires contribuent le plus à la Réionisation. En effet, les halos les moins massifs émettent peu de photons, même s’ils s’échappent plus efficacement dans l’IGM. Quant aux halos de grandes masses, même s’ils émettent beaucoup de photons ionisants, peu d’entre eux arrivent effectivement à s’échapper dans l’IGM. Ces Figures sont extraites de Lewis et al. (2020).

un scénario de réionisation donné (Lewis et al. 2020). On voit que les halos de plus faibles masses semblent laisser une plus grande fraction de photons s’échapper dans l’IGM que les galaxies les plus massives. À droite de la Figure 1.9, je montre la contribution des photons à l’ionisation de l’IGM en fonction de la masse de halo et du redshift. On voit ici que les halos de plus faibles masses sont nombreux, mais peu efficaces pour ioniser l’IGM, et les halos de plus grandes masses sont peu nombreux même si plus efficaces. Ce sont alors les halos de masses intermédiaires qui contribuent le plus à la Réionisation. Par ailleurs, en multipliant la fraction  $f_{\text{esc}}$  par la quantité de gaz galactique dans les étoiles et le nombre de photons ionisants qu’elles produisent et en prenant en compte le nombre de fois qu’un atome d’Hydrogène recombine, on se retrouve avec une efficacité d’ionisation des galaxies (ou émissivité de photons ionisants), que l’on note  $\zeta$ . Celle-ci contrôle le déroulement de la Réionisation et nous permet de quantifier et de connaître la contribution des galaxies au budget de photons ionisants, comme j’en discuterai à nouveau dans les Chapitre 4 et 5.

Enfin, la production de photons ionisants par les étoiles peut être affectée par diverses rétroactions, qui sont en compétition avec le processus de réionisation (Ciardi & Ferrara 2005; Dayal & Ferrara 2018; Choudhury 2022). Quand les étoiles explosent en supernovae, elles réchauffent le gaz autour d’elles, voire éjectent le gaz des plus petits halos, empêchant ainsi la formation d’étoiles (et donc de photons ionisants). Ceci correspond à la rétroaction mécanique. La formation d’étoiles peut également être réduite par la pollution en métaux

de l'environnement causée par les supernovae (et par exemple empêcher les étoiles de la Population III de se former), réduisant donc aussi l'émission de photons ionisants des étoiles. On parle de rétroaction chimique. Par ailleurs, la Réionisation elle-même, réchauffe le gaz et l'empêche de s'effondrer pour former des étoiles dans les plus petits halos (Dawoodbhoy et al. 2018). On appelle ceci la rétroaction radiative.

La Réionisation dépend donc fortement des propriétés des galaxies, avec entre autres leur masse, leur luminosité, leur fraction d'échappement, et leur efficacité d'ionisation. Plus ces quantités sont élevées, plus l'Hydrogène sera réionisé efficacement, et les bulles ionisées seront étendues. Dans ce cas-là, la réionisation sera rapide. La Réionisation dépend également du taux de formation d'étoiles des galaxies, des populations d'étoiles qu'elles contiennent, de l'IMF et de leurs métallicité. Ces propriétés peuvent être contraintes par exemple par les fonctions de luminosité et leur modélisation, comme je le montre dans la Figure 1.8 (Bouwens et al. 2017, 2023).

### 1.3.3 La réionisation du milieu intergalactique

Pour ioniser un atome d'Hydrogène neutre  $H$  (aussi souvent appelé « HI »), un photon  $\gamma$  d'énergie supérieure à 13.6 eV doit être absorbé par l'atome :



On appelle souvent « HII » l'Hydrogène ionisé produit par cette réaction. La probabilité d'absorption d'un photon traversant un milieu comportant de l'Hydrogène neutre est donnée par la section efficace  $\sigma_{\text{HI}} = 6.8 \times 10^{-18} (E/13.6 \text{ eV})^{-3} \text{ cm}^2$  (Wise 2019). On remarque que cette section efficace décroît avec l'énergie du photon : les photons à plus haute énergie vont pouvoir se propager sur de plus grandes distances avant de se faire absorber par un atome. Les rayons X (émis par les quasars et les étoiles binaires par exemple) ont en particulier une section efficace plus faible que les rayons UV (émis par les galaxies), et vont donc moins participer à la réionisation locale de l'Hydrogène. Les quasars et système binaires émettant des rayons X vont néanmoins pouvoir réioniser (et réchauffer) le gaz de l'IGM à plus grande distance, à l'inverse des galaxies.

Les étoiles des premières galaxies de l'Univers émettent un rayonnement UV, qui va réioniser l'IGM environnant. Au début de la Réionisation, seules les régions entourant les galaxies sont réionisées. On appelle ces régions des « régions HII » ou « bulles ionisées », et on peut en voir des exemples dans la frise de la Figure 1.1. Par définition, ces bulles contiennent du gaz hautement ionisé, c'est-à-dire avec une fraction d'ionisation proche de 1 et une fraction d'Hydrogène neutre autour de  $10^{-4}$ . En théorie, il existe un régime stationnaire (la sphère de Strömgen) dans lequel l'ionisation et la recombinaison du gaz sont à l'équilibre. Ces sphères sont décrites grâce au rayon de Strömgen et celui-ci stagne dans le temps dès que ces deux processus sont à l'équilibre (Iliev et al. 2006a). En pratique, comme les sources sont de plus en plus intenses et que la densité se dilue sous l'effet de l'expansion tout au long de la Réionisation, ce régime stationnaire n'est jamais atteint et les régions HII grossissent. De nouvelles sources apparaissent ailleurs en créant de nouvelles régions HII, et peu à peu les bulles fusionnent jusqu'à ce qu'elles percolent (comme on peut le suivre sur la frise). Après cela, les photons ionisants atteignent les régions les plus reculées et les moins denses pour

ioniser l'Hydrogène encore neutre et faire disparaître les dernières régions neutres environ 1 milliard d'années après le Big Bang (soit vers un redshift  $z \sim 5.5$ ).

## 1.4 L'Univers post-Réionisation

Depuis la fin de la Réionisation (un milliard d'années après le Big Bang, ou vers un redshift  $z \sim 5.5$ , voir sur la frise de la Figure 1.1), le gaz de l'Univers est réchauffé et réionisé. L'Univers n'a plus connu de grande transition et ses structures ont simplement continué à évoluer à partir de celles déjà formées pendant l'Aube Cosmique et la Réionisation. Il n'en demeure pas moins que la Réionisation peut être importante pour comprendre l'Univers actuel.

Le modèle  $\Lambda$ CDM, contraint par exemple par le CMB (Planck Collaboration et al. 2020) et par les oscillations acoustiques baryoniques (baryonic acoustic oscillations, BAO) mesurées dans les grands relevés de galaxies (Eisenstein et al. 2005), a aujourd'hui un grand succès aux grandes échelles. C'est donc le modèle naturel que j'ai utilisé pendant cette thèse pour étudier la Réionisation. Mais il faut savoir que ce modèle est parfois contesté. Par exemple, on ne sait pas ce qu'est exactement la matière noire, qui représente pourtant une grande majorité de la matière de notre Univers. On sait aussi que ce modèle a des difficultés à reproduire correctement les petites échelles. La structure des galaxies observées n'a pas la forme attendue par les modèles au centre des galaxies (problème « cusp-core », voir Walker & Peñarrubia (2011)). Dans le Groupe Local, que l'on définit comme l'ensemble de la Voie Lactée (Milky Way, MW), d'Andromède (M31) et de toutes leurs galaxies satellites, on repère également des problèmes que l'on n'explique pas encore avec le modèle  $\Lambda$ CDM, comme la distribution non-uniforme des galaxies naines satellites autour de nous et d'Andromède (Lynden-Bell 1976; Doliva-Dolinsky et al. 2023). Ces problèmes sont très étudiés et des modifications du modèle  $\Lambda$ CDM, ou bien de nouveaux modèles sont proposés. Par exemple, différents types de matière noire peuvent être étudiés, avec entre autres de la matière noire plutôt tiède (masse des particules de matière noire plus petite), ce qui pourrait impacter le processus de formation des galaxies et la Réionisation (Castellano et al. 2023).

Enfin, la Réionisation peut apporter des pistes de réflexion à ces problèmes. On sait qu'à cause du réchauffement du gaz pendant la Réionisation, la formation d'étoiles a été supprimée dans les plus petites galaxies (de masse  $< 10^9 M_\odot$ , voir Dawoodbhoy et al. (2018)). Étudier la Réionisation peut donc nous aider à comprendre la formation des galaxies. Dans le cas d'un processus de réionisation démarrant dans les endroits denses pour se terminer dans les vides (scénario « inside-out »), plus les galaxies satellites du Groupe Local sont proches des centres galactiques de MW et M31, plus elles ont été réionisées tôt (Ocvirk et al. 2014). En étudiant la Réionisation, on a aussi pu voir que la Voie Lactée et Andromède se sont réionisées indépendamment l'une de l'autre pendant une durée similaire (Ocvirk et al. 2013). Cela implique que le processus de réionisation a modelé les structures que l'on voit aujourd'hui, leur abondance et leur distribution spatiale, notamment dans l'Univers Local. La Réionisation, étant un processus ayant lieu aux échelles galactique et cosmologique, elle se présente ainsi comme une sonde unifiée pour la tester la cohérence de l'ensemble.

# CHAPITRE 2

## COMMENT ÉTUDIER LA RÉIONISATION COSMOLOGIQUE ?

*Dans ce Chapitre, je décris des moyens d'observations et d'études de l'Époque de la Réionisation. Je commence par présenter des manières d'observer indirectement la Réionisation, grâce au fond diffus cosmologique, au signal Lyman- $\alpha$  et aux nouvelles galaxies observées par le James Webb Space Telescope. Ensuite, je décris ce qu'est le signal 21 cm et comment on peut l'observer. Trois types d'observations seront à notre disposition pour détecter ce signal : le signal 21 cm moyen sur tout le ciel, le spectre de puissance du signal 21 cm ou des images 2D sur le plan du ciel observées à plusieurs redshifts (ou fréquences). Je continue en présentant les simulations cosmologique **EMMA** et semi-analytique **21cmFAST** de la Réionisation, qui nous permettent d'étudier cette époque avant d'en avoir des observations. Je détaille également ici le champ de temps de réionisation qui est central pour la suite. Enfin, je conclus ce Chapitre en mentionnant les défis auxquels nous faisons face pour comprendre la Réionisation, et les objectifs de cette thèse en regard à tous ces défis.*

## Sommaire

---

<b>2.1</b>	<b>Observer la Réionisation de manière indirecte . . . . .</b>	<b>21</b>
2.1.1	Détection de galaxies lointaines avec HST et JWST . . . . .	21
2.1.2	Le rayonnement Lyman- $\alpha$ . . . . .	25
2.1.3	Le fond diffus cosmologique . . . . .	27
2.1.4	Contraintes sur l’histoire de réionisation . . . . .	28
<b>2.2</b>	<b>Observer la Réionisation de manière directe avec le signal 21 cm</b>	<b>28</b>
2.2.1	Le signal 21 cm . . . . .	29
2.2.2	Signal global . . . . .	32
2.2.3	Spectre de puissance . . . . .	33
2.2.4	Tomographie avec le radio-télescope SKA . . . . .	35
<b>2.3</b>	<b>Simulations de la Réionisation . . . . .</b>	<b>38</b>
2.3.1	Simulations cosmologiques : l’exemple d’EMMA . . . . .	39
2.3.2	Modèles semi-analytiques : l’exemple de 21cmFAST . . . . .	41
2.3.3	Temps/redshifts de réionisation . . . . .	44
<b>2.4</b>	<b>Défis à venir et objectifs de la thèse . . . . .</b>	<b>46</b>

---

## 2.1 Observer la Réionisation de manière indirecte

Comme je l’ai présenté dans le Chapitre 1, l’Époque de la Réionisation a eu lieu quelques centaines de millions d’années après le Big Bang et s’est terminé un milliard d’année après le Big Bang, soit environ entre les redshifts  $z = 15$  et  $z = 5$ . Cette époque est lointaine et est donc difficile à observer avec les moyens actuels. Jusqu’à présent, plusieurs méthodes indirectes permettent de l’étudier.

### 2.1.1 Détection de galaxies lointaines avec HST et JWST

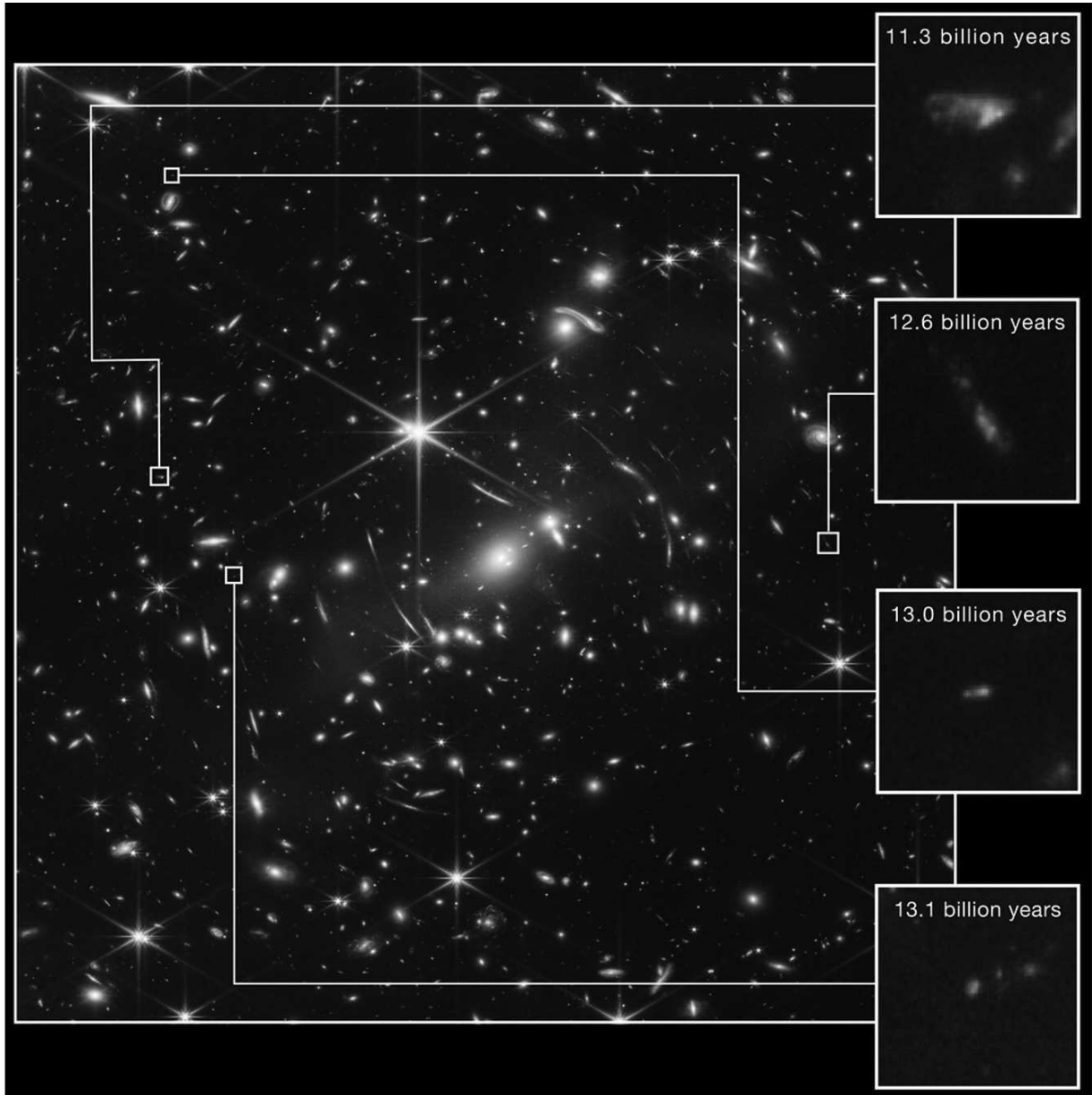
La dernière grande transition de l’état d’ionisation de l’hydrogène dans l’Univers qu’est la Réionisation résulte de l’émission du rayonnement UV des premières sources lumineuses ayant été créées. Ainsi, pour comprendre le processus de réionisation, il faut étudier les propriétés de ces sources, leur distribution spatiale et leur évolution temporelle.

Les sources de réionisation sont supposément les galaxies qui apparaissent tout au long de la Réionisation, et on pense généralement que les étoiles résidant dans ces galaxies sont les principales contributrices au budget de photons ionisants. Est-on alors capable de détecter ces galaxies, sources de la réionisation cosmologique de l’hydrogène ? Il y a une trentaine d’années, on a commencé à détecter de tels objets à haut redshift (jusqu’à  $z \sim 6$ ) avec les observations en champ profond dans les longueurs d’onde visibles menées par le télescope spatial Hubble<sup>1</sup> (Hubble Space Telescop, HST, Williams et al. (1996)). Ce type d’observations s’est poursuivi avec d’autres instruments pour détecter des galaxies toujours plus lointaines (jusqu’à  $z \sim 10$ ) et très récemment le télescope spatial James Webb<sup>2</sup> (James Webb Space Telescope, JWST) a produit un champ profond autour d’un amas de galaxies dans les longueurs d’onde infrarouges, et que je présente dans la Figure 2.1. Une majorité des galaxies colorées en rouge dans l’image sont probablement des galaxies très lointaines et qui datent de l’Époque de la Réionisation, comme celles qui sont encadrées sur la droite. Dans la Figure 2.2, je montre une Figure provenant de Harikane et al. (2023a) qui présente toutes les galaxies détectées avec ces instruments en intégrant les dernières observations du JWST. Elles sont classées en fonction leur luminosité et de leur redshift. On voit que le JWST (avec sa caméra NIRCam et son spectromètre NIRSpec) est capable de détecter des galaxies de plus en plus lointaines par rapport à celles que l’on voyait avec le HST. Jusqu’à présent, le JWST a ainsi permis d’identifier des galaxies candidates à des redshifts de l’ordre de  $9 < z < 17$  pour la première fois, ce qui va nous permettre d’accéder aux propriétés des objets menant la Réionisation.

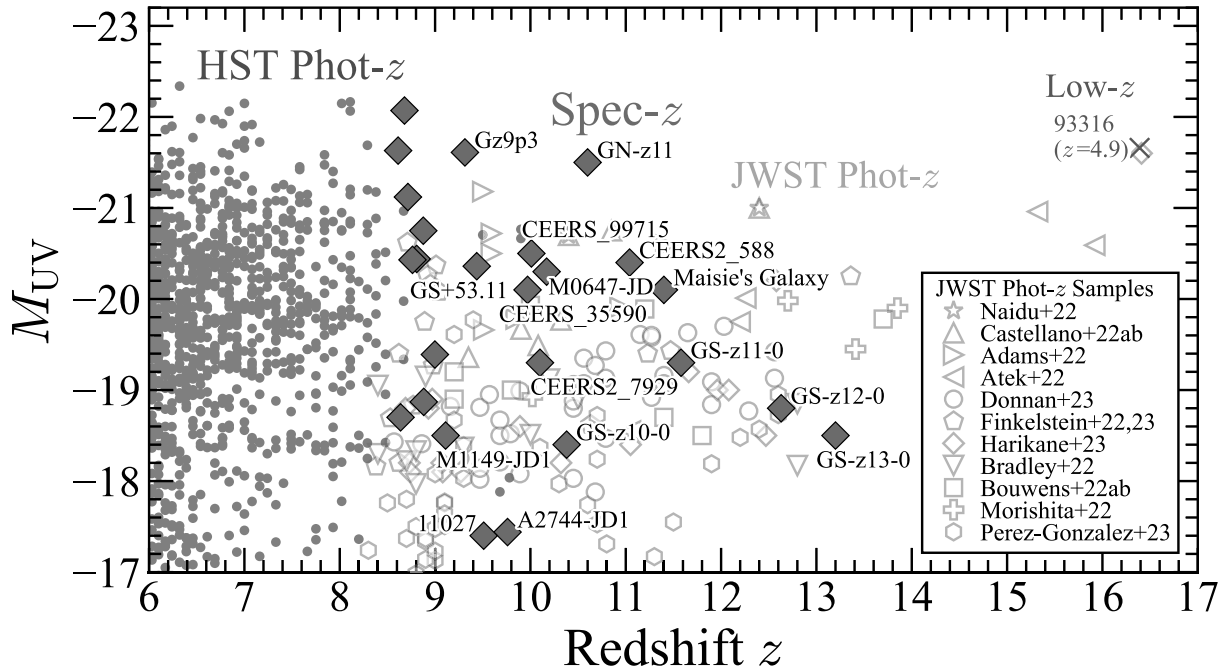
Depuis la sortie des données du JWST, un très grand nombre d’études se sont penchées sur les données photométriques et spectroscopiques récentes pour détecter des galaxies candidates à haut redshift et analyser leurs caractéristiques physiques (Bradley et al. 2022; Castellano et al. 2022; Naidu et al. 2022; Adams et al. 2023; Atek et al. 2023; Bouwens et al. 2023; Donnan et al. 2023; Finkelstein et al. 2023; Fujimoto et al. 2023; Harikane et al. 2023b,a; Labbé et al. 2023; Morishita & Stiavelli 2023; Yan et al. 2023). Ces études s’intéressent souvent à la fonction de luminosité dans l’ultraviolet de ces candidates et je montre un exemple de mesures faites à  $z \sim 9$ ,  $z \sim 12$  et  $z \sim 16$  par Harikane et al. (2023a) dans la

1. [https://www.nasa.gov/mission\\_pages/hubble/main/index.html](https://www.nasa.gov/mission_pages/hubble/main/index.html)

2. <https://webb.nasa.gov>



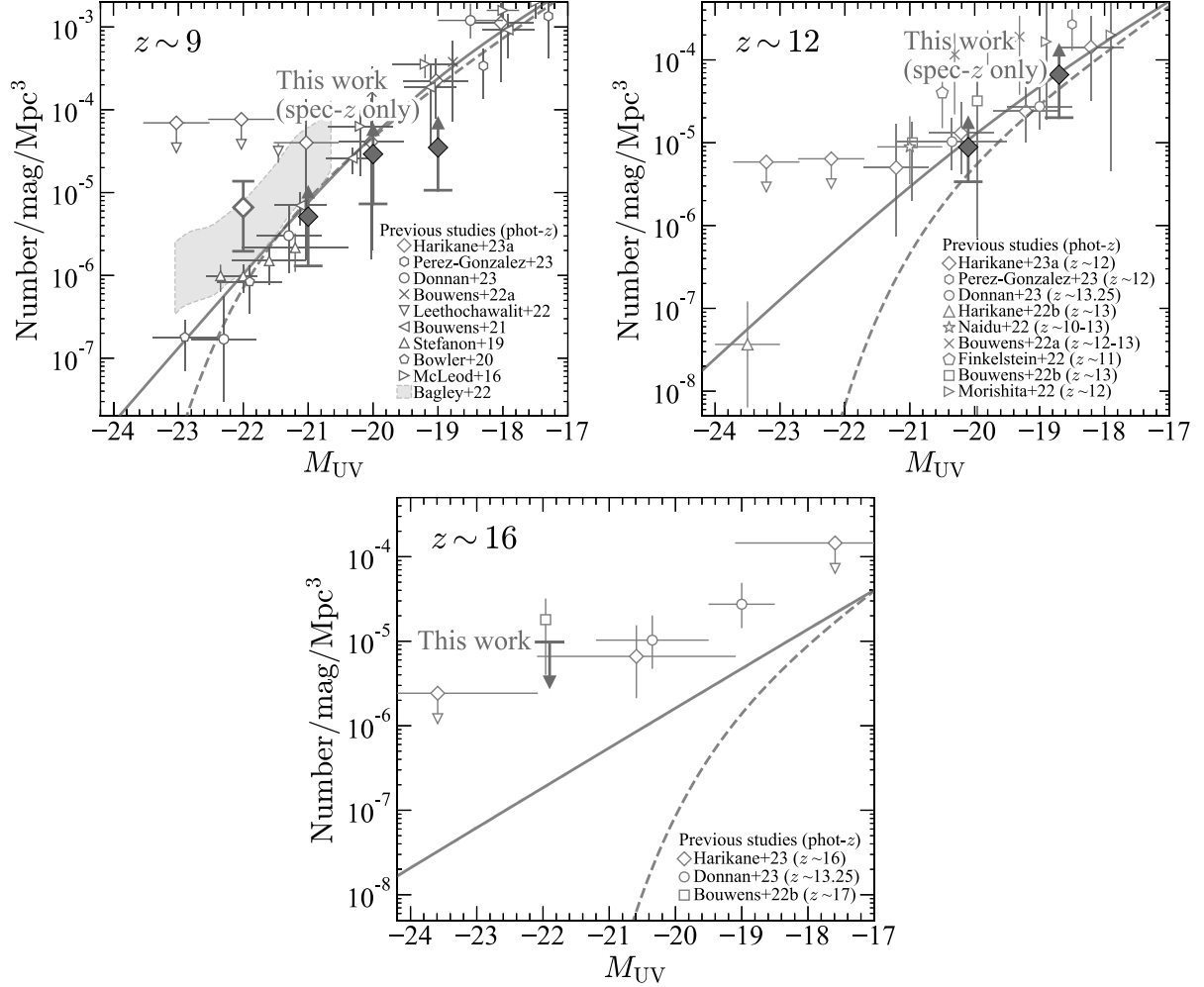
**FIGURE 2.1** – Champ profond autour de l’amas de galaxies SMACS 0723 observé par le JWST en infrarouge. Cette image a été faite avec l’instrument NIRCcam du JWST en juillet 2022 et provient du site web de la NASA (<https://www.nasa.gov/image-feature/goddard/2022/nasa-s-webb-delivers-deepest-infrared-image-of-universe-yet>). On voit en particulier des galaxies (en rouges et très ténues) de l’Époque de la Réionisation. Certaines d’entre elles sont agrandies dans les cadres sur la droite de l’image.



**FIGURE 2.2** – Magnitudes des galaxies détectées à hauts redshifts par le HST (en gris) et par le JWST (en rouge). Le redshift des galaxies présentées ici est majoritairement obtenu par photométrie. Cette Figure, extraite de Harikane et al. (2023a), présente aussi des galaxies dont le redshift est mesuré par spectroscopie (losanges rouges). Notamment, la galaxie, détectée à un redshift  $z > 16$  par photométrie, se trouve en fait à un redshift  $z < 5$  (voir en vert). On voit que le HST a permis de détecter des galaxies jusqu’à un redshift d’environ 8, tandis que le JWST repousse cette limite jusqu’à un redshift d’environ 15.

Figure 2.3. Cette fonction représente le nombre de galaxies par unité de volume détectées dans une tranche de luminosité donnée, et permet de contraindre les propriétés des sources de réionisation comme leur émissivité ou leur taux de formation d’étoiles (ou SFR pour star formation rate). Dans leur première étude (Harikane et al. 2023b) ainsi que dans plusieurs autres travaux menés récemment, un trop grand nombre de galaxies apparaissent comme étant trop lumineuses et/ou trop massives à haut redshift pour le volume dans lequel on les détecte. Dans le modèle actuel et favorisé de formation des structures (modèle hiérarchique), on pense que les galaxies se forment par fusions de galaxies moins massives au cours du temps et on ne s’attend en particulier pas à détecter autant de galaxies massives aussi tôt (Boylan-Kolchin 2023). Pendant ce temps, ces études proposent des explications reposant sur une absence de suppression de la formation d’étoiles dans les phases pré-Réionisation, ou bien sur une émission de rayonnement très efficace par les premières étoiles très lumineuses (population III). Keller et al. (2023) a aussi montré que ce type de galaxies peut être en fait simulé avec des modèles de formation des galaxies classiques et déjà existants. De nouvelles données spectroscopiques continuent à être mesurées par le JWST, et toutes ses données sont encore en train d’être analysées par les diverses équipes pour parfaire leurs études. Les résultats qui paraissent en ce moment évoluent donc assez rapidement. Les dernières





**FIGURE 2.3** – Fonctions de luminosité à  $z \sim 9$  (en haut à gauche), à  $z \sim 12$  (en haut à droite), et à  $z \sim 16$  (en bas). Les points rouges représentent les densité de galaxies avec des redshifts spectroscopiques détectées par Harikane et al. (2023a) et les flèches rouges sont leurs limites supérieures à  $1\sigma$ . Les points gris sont des estimations basées sur des galaxies détectées par photométrie dans d’autres études. Les courbes grises en traits pleins et en pointillés sont des ajustements obtenus à partir d’une double loi de puissance et d’une fonction de Schechter respectivement. Cette Figure est extraite de Harikane et al. (2023a). Les mesures sont en accord, ou tout du moins comparables, avec les résultats précédents obtenus avec le HST et le JWST.

données spectroscopiques montrent par exemple que la galaxie très lumineuse avec un redshift photométrique  $z > 16$  montrée dans la Figure 2.2 a en fait un redshift  $z < 5$  (Harikane et al. 2023a). Ces auteurs concluent que les fonctions de luminosité (montrées dans la Figure 2.3) à  $z \sim 9$  et  $z \sim 12$  sont en accord avec les mesures précédemment faites sur les données HST et JWST, tandis que celle à  $z \sim 16$  est une nouvelle contrainte sur la luminosité des galaxies. Elles permettent ainsi d’assurer que le modèle  $\Lambda$ CDM est toujours en accord avec les mesures.

### 2.1.2 Le rayonnement Lyman- $\alpha$

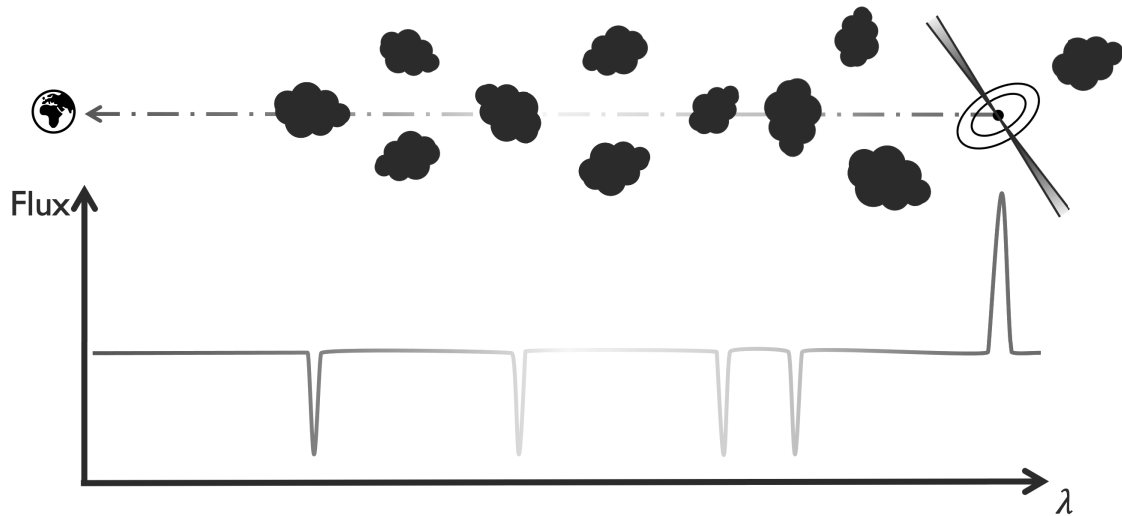
Les quasars sont des objets très lumineux et qui peuvent donc être détectés à haut redshifts. Dans la Figure 2.4, on voit un schéma de quasar à droite qui émet une raie Lyman-alpha ( $\text{Ly}\alpha$ ) visible par le pic rouge aux grandes longueurs d’onde dans le spectre du dessous. On voit également que son rayonnement est absorbé par des nuages d’hydrogène neutre le long de la ligne de visée du quasar, ce qui produit des pics d’absorption dans le spectre qui sont décalés vers le rouge pour chaque nuage de gaz. Ces raies d’absorption sont en fait dues à la transition de l’électron de l’hydrogène entre les premier et deuxième niveaux d’énergie. Dans ce schéma, elles ne sont pas nombreuses et indiquent alors un Univers plutôt globalement réionisé : on observe alors un quasar à une époque post-Réionisation. Si un plus grand nombre de nuages d’hydrogène sont présents le long de la ligne de visée d’un quasar, alors on observera bien plus de raies d’absorption. L’ensemble de ces raies est appelé la « forêt Lyman-alpha ».

Quand on étudie la Réionisation à des redshifts  $z > 5.5$ , beaucoup de gaz d’hydrogène neutre est présent. Au début de la Réionisation, en particulier, la présence d’une grande quantité de gaz neutre rend le milieu intergalactique (IGM) opaque au rayonnement  $\text{Ly}\alpha$  qui s’échappe des premières galaxies. Un exemple de spectre de quasar observé à cette époque ( $z = 6.0345$ ) par Bosman et al. (2022) est montré dans la Figure 2.5 en noir. Aux grandes longueurs d’onde à droite, on voit le pic de rayonnement UV émis par le quasar. En bleu, on voit le spectre continu que l’on devrait mesurer dans le cas où aucun nuage d’hydrogène neutre ne se trouverait le long de sa ligne de visée. Dans le spectre observé en noir, on voit que le flux s’annule parce que le rayonnement est quasiment totalement absorbé par le continu de gaz d’hydrogène neutre : c’est ce qu’on appelle l’effet Gunn-Peterson (Gunn & Peterson 1965). Grâce à cet effet, on est ainsi en mesure de détecter les régions neutres le long de la ligne de visée des quasars lointains, et de mesurer la profondeur optique du milieu, nous informant aussi sur sa densité. Une manière de sonder la fraction d’hydrogène neutre est ainsi de compter le nombre de pixels « noirs », c’est-à-dire de pixels ayant un flux nul, ce qui permet d’imposer une limite inférieure à la fraction de volume ionisé indépendamment du modèle de Réionisation.

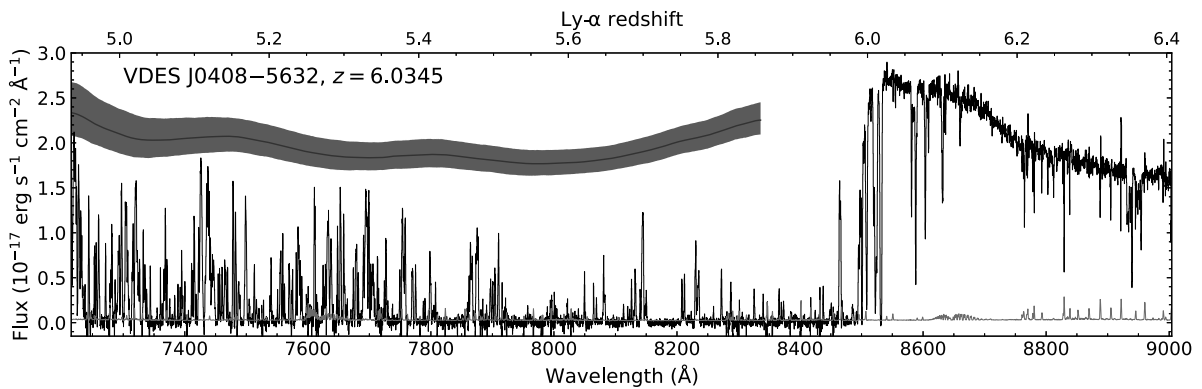
Un grand nombre d’études se sont intéressées au spectre du rayonnement émis par des quasars lointains observés par plusieurs relevés dans l’infrarouge proche (comme le Sloan Digital Sky Survey, SDSS<sup>3</sup>, ou plus récemment le programme XGR-30 mené au Very Large Telescope, VLT<sup>4</sup>). Ce type de spectre est notamment utilisé pour sonder les phases tardives de la Réionisation, et en particulier contraindre l’histoire de réionisation avec la fraction de

3. <https://classic.sdss.org>

4. <https://www.eso.org/public/france/teles-instr/paranal-observatory/vlt/>



**FIGURE 2.4** – Schéma représentant un cas post-Réionisation d’absorption du rayonnement UV émis par un quasar lointain. Le pic d’émission sur la droite correspond au rayonnement Lyman-alpha de ce quasar. Les nuages d’hydrogène neutre, qui sont distribués le long de la ligne de visée du quasar, absorbent le rayonnement.



**FIGURE 2.5** – Spectre (en noir) de la transition Lyman-alpha dans le quasar VDES J0408-5632 observé à un redshift  $z = 6.0345$ . L’incertitude de ce flux est représentée en rouge, et une reconstruction d’un flux continu sans absorption est montré en bleu. On remarque une absence de signal due à l’absorption de celui-ci par des nuages d’Hydrogène neutre le long de la ligne de visée du quasar. Cette Figure est extraite de Bosman et al. (2022).

volume neutre dans l’environnement proche de ces quasars (Fan et al. 2002; Mesinger & Haiman 2004; Fan et al. 2006; Malhotra & Rhoads 2006; Mortlock et al. 2011; Mesinger et al. 2015; Greig et al. 2017; Davies et al. 2018; Greig et al. 2019; Wang et al. 2020; Greig et al. 2022). La caractéristique de l’absorption en « aile amortie » (« damping wing ») dans les spectres de quasars à haut redshift est souvent utilisée dans ces études comme une preuve significative d’un Univers encore neutre et permet de calculer la fraction de gaz neutre autour des quasars. Greig et al. (2019) appuient de plus sur le fait que la Réionisation est un processus assez étendu dans le temps et Bosman et al. (2022) ajoutent que ce processus doit se finir assez tard ( $z < 5.5$ ). Les propriétés des galaxies sources de la réionisation, comme leur émissivité, l’opacité de l’IGM, ou encore les modèles de la Réionisation peuvent également être contraints avec ce genre d’observation (Fan et al. 2002; Mesinger & Haiman 2004; Dijkstra et al. 2011; Eilers et al. 2018; Mason & Gronke 2020). Enfin, la topologie du processus de réionisation, avec la morphologie des bulles ionisées et leur superposition, peuvent également être explorées avec l’émission Ly $\alpha$  (Fan et al. 2002, 2006; Mesinger & Haiman 2004; Mortlock et al. 2011; Mason & Gronke 2020). La taille des régions ionisées autour des quasars peut par exemple être déduite des spectres, comme le font Mortlock et al. (2011). Plus récemment, Mason & Gronke (2020) expliquent que l’on a accès aux propriétés des bulles ionisées avec ce type de spectre et que l’on peut par exemple contraindre leur taille.

### 2.1.3 Le fond diffus cosmologique

Depuis les années 1960, et notamment avec le lancement du satellite Planck<sup>5</sup> en 2009, la communauté astrophysique s’est beaucoup investie dans l’étude du fond diffus cosmologique (CMB pour cosmic microwave background). Par exemple, le spectre de puissance du CMB a été contraint et les paramètres cosmologiques estimés avec un très grand niveau de précision. La collaboration Planck a ainsi publié ces contraintes dans Planck Collaboration et al. (2020).

Le CMB est aussi un moyen de contraindre d’autres phénomènes astrophysiques parce que ce signal est grandement influencé par les structures qui se forment plus tard le long de la ligne de visée. Les photons émis par le CMB peuvent notamment être dispersés par les électrons libres (par diffusion Thomson), qui sont produits en grande quantité par le processus de réionisation entre les redshifts 15 et 5 environ. Ce phénomène de dispersion est une source d’anisotropies dans le CMB et résulte en un signal secondaire dans son spectre de puissance angulaire. À partir du CMB, on peut contraindre la profondeur optique  $\tau$  de cette diffusion, qui est une grandeur intégrée le long de la ligne de visée :

$$\tau = c \sigma_T \int_{t_0}^{t_{\text{CMB}}} \bar{n}_e(t) a^{-3}(t) dt \quad (2.1)$$

où  $c$  est la vitesse de la lumière,  $\sigma_T$  est la section efficace de la dispersion Thomson,  $t_0$  et  $t_{\text{CMB}}$  sont les temps aujourd’hui et à l’émission du CMB,  $\bar{n}_e$  est la densité d’électrons libres moyennée sur le volume et  $a$  est le facteur d’échelle. La collaboration Planck a contraint cette profondeur optique à  $\tau = 0.0544$  et je montre dans la Figure 2.6 une courbe de vraisemblance obtenue autour de cette valeur (Planck Collaboration et al. 2020). Cette dépendance du CMB

5. [https://www.esa.int/Enabling\\_Support/Operations/Planck](https://www.esa.int/Enabling_Support/Operations/Planck)

à la Réionisation a donc permis à la collaboration Planck de contraindre le redshift moyen de réionisation à  $z_{\text{mid\_reion}} = 7.68$  (Planck Collaboration et al. 2020).

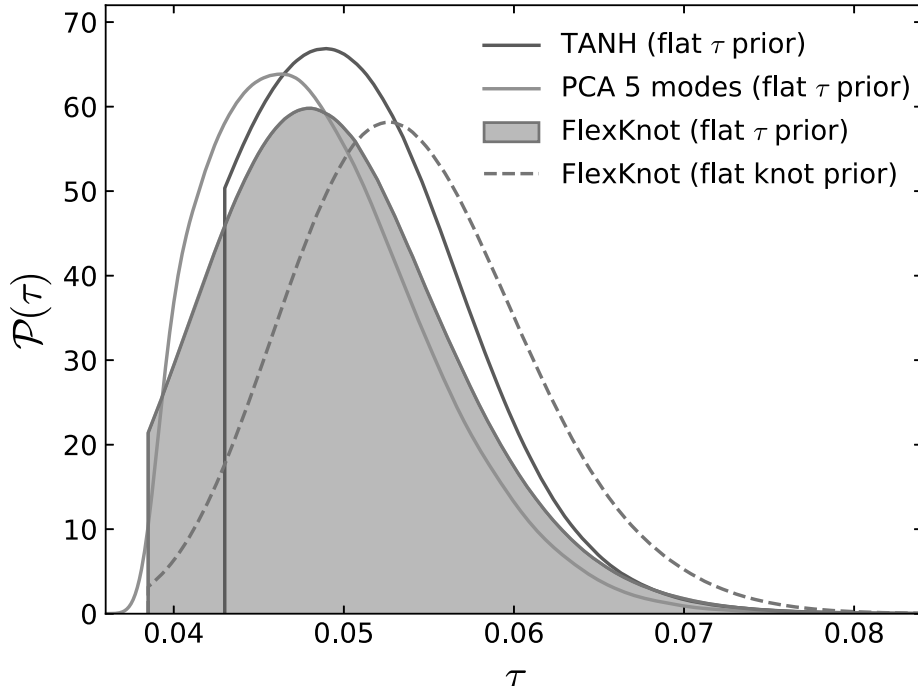
Par ailleurs, les photons du CMB peuvent aussi être dispersés par des électrons libres qui ont une énergie cinétique élevée. Cet effet est connu sous le nom de l'« effet cinétique Sunyaev-Zel'dovich » (kSZ pour kinetic Sunyaev-Zel'dovich), et ajoute de la puissance aux basses échelles angulaires du spectre du CMB. Il impacte ainsi le spectre de puissance du CMB de deux manières. La première provient du signal homogène kSZ qui est dû à une distribution homogène dans l'IGM d'électrons libres à la fin de la Réionisation. La seconde provient du signal « patchy » (ou sporadique) kSZ qui est dû à la dispersion des photons du CMB dans les bulles ionisées (pendant la Réionisation) le long d'une ligne de visée qui contient autrement du gaz neutre. Dans ce second cas, l'inhomogénéité de la Réionisation (par la présence de bulles ionisées distribuées dans l'IGM autour des sources) induit une empreinte à l'échelle de quelques arcminutes dans le CMB. Plusieurs études, comme McQuinn et al. (2005); Zahn et al. (2005); Iliev et al. (2007); Mesinger et al. (2012); Gorce et al. (2020), ont alors montré que le spectre de puissance angulaire de l'effet kSZ dépend du processus de réionisation, et en particulier de la morphologie du processus. Il semblerait que différents scénarios de réionisation pourraient aussi être contraints avec ce spectre de puissance. Zahn et al. (2005), Iliev et al. (2007) et Gorce et al. (2020) ajoutent que l'on a accès à la taille des bulles ionisées avec le pic de ce spectre de puissance.

#### 2.1.4 Contraintes sur l'histoire de réionisation

Comme expliqué précédemment, l'histoire de réionisation, autrement dit la fraction d'hydrogène neutre pendant la Réionisation, peut être contrainte avec les observations mentionnées ci-avant. Par exemple, Greig et al. (2022) utilisent toutes ces observations pour contraindre et reconstruire par modèle Bayésien cette fraction de volume neutre pendant la Réionisation. Dans la Figure 2.7, je montre leur résultat. On voit que les contraintes ponctuelles obtenues à partir du signal Ly $\alpha$  imposent à l'histoire de réionisation d'être comprise entre 0.3 et 0.9 environ dans la gamme de redshifts 6 et 8. La reconstruction par modèle Bayésien de la fraction de neutre en bleu nous donnent une idée globale de l'histoire de réionisation. On remarque en particulier que la Réionisation n'est pas un événement instantané, mais est au contraire un processus étalé dans le temps qui se termine bien à un redshift inférieur à 6.

## 2.2 Observer la Réionisation de manière directe avec le signal 21 cm

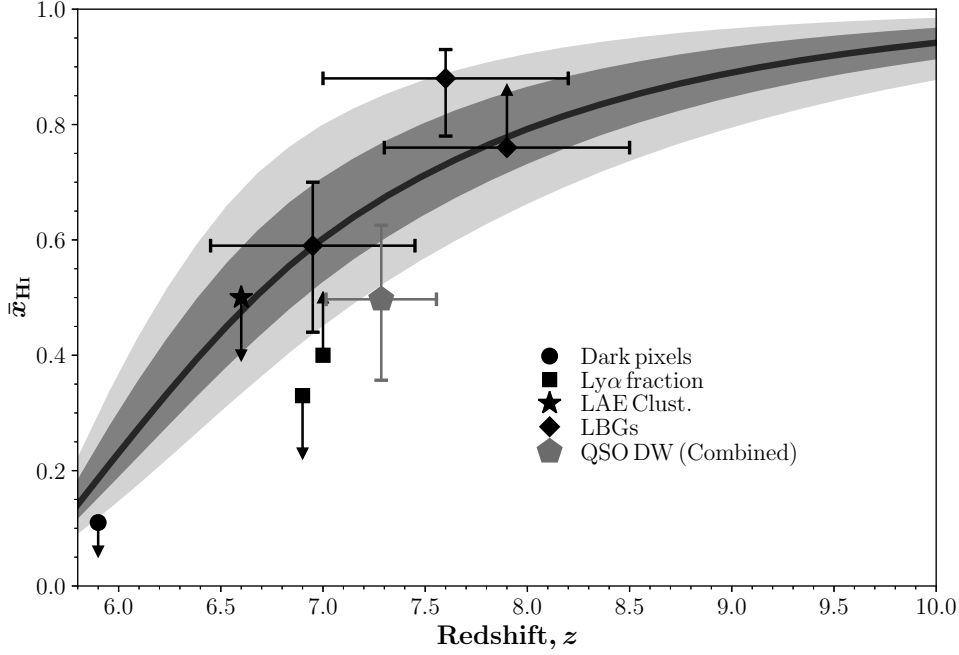
De nouveaux télescopes dans le domaine des ondes radio, ayant entre autres pour but d'observer l'Époque de la Réionisation directement, sont en train d'être construits. L'objectif de ces instruments est de détecter le signal 21 cm émis par les atomes d'hydrogène neutre. Je décris ici ce signal, ainsi que les observations possibles que l'on peut en obtenir.



**FIGURE 2.6** – Contraintes marginalisées obtenues par la collaboration Planck sur la valeur de la profondeur optique de la Réionisation  $\tau$ . Les différentes courbes représentent différents modèles de Réionisation testés. Ces courbes ne seront pas détaillées explicitement ici : le lecteur intéressé peut se référer à l’article Planck Collaboration et al. (2020), duquel cette Figure est extraite. On voit globalement que le maximum de vraisemblance obtenu par cette collaboration est autour de  $\tau = 0.05$ .

### 2.2.1 Le signal 21 cm

En astrophysique, nous pouvons observer des évènements dans l’espace qui seraient trop rares pour se produire sur Terre. Ce genre de phénomènes peut nous permettre d’étudier la physique de divers processus, et c’est en particulier le cas du signal 21 cm pour l’étude de la Réionisation (Furlanetto et al. 2006; Mesinger et al. 2011; Mellema et al. 2015). Dans la Figure 2.8, je montre une illustration qui permet d’expliquer l’origine du signal 21 cm. La transition entre les deux sous-niveaux hyperfins ( $n_0$  et  $n_1$ ) du niveau fondamental ( $n$ ) de l’atome d’hydrogène (neutre) est due au retournement du spin de l’électron de cet atome. Si cette transition est causée par un photon de longueur d’onde 21 cm qui est absorbé par l’atome d’hydrogène, alors il gagne de l’énergie. Il en perd en revanche quand un photon d’une même longueur d’onde est émis par désexcitation. Ce signal se produit à basse énergie, à une longueur d’onde de 21 cm, une fréquence de 1420 MHz, et sa température associée est  $T_{10} = 0.068$  K. On associe également à cette transition énergétique la température dite de spin,  $T_s$ , qui caractérise l’équilibre thermique entre les deux états de spin. Deux processus peuvent modifier cette température de spin. Le premier est simplement l’excitation directe ou la désexcitation de l’atome d’hydrogène entre les deux états de spin  $n_0$  et  $n_1$ , et le second est connu sous le nom de l’effet Wouthuysen-Field (WF ; Wouthuysen (1952); Field (1958)). Un

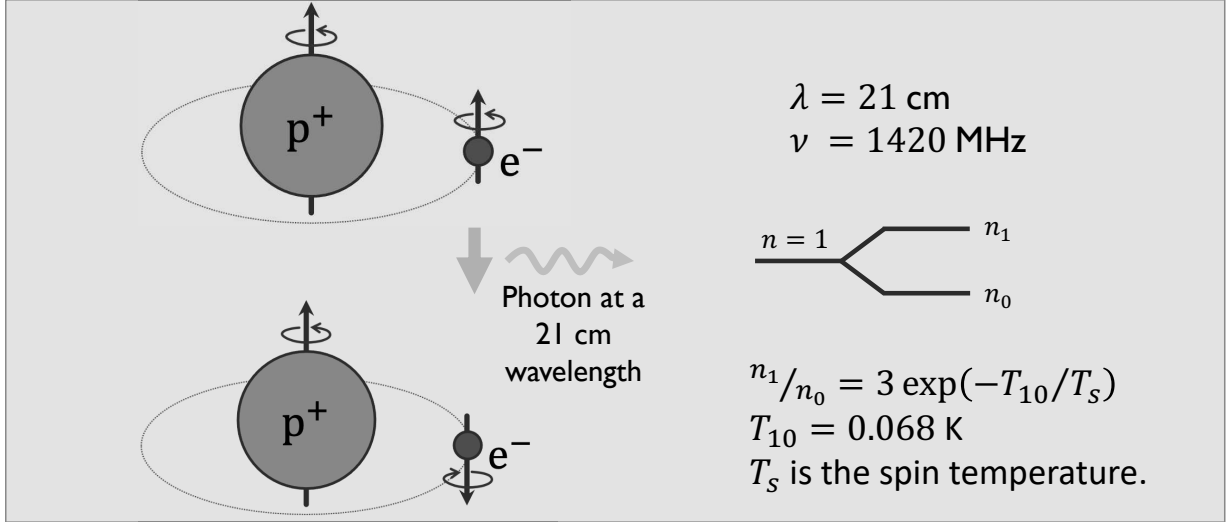


**FIGURE 2.7** – Compilation des contraintes existantes sur la fraction d’hydrogène neutre en fonction du redshift. Cette Figure est faite par Greig et al. (2022), et ils utilisent des contraintes obtenues avec le signal Ly $\alpha$ . Celles-ci sont représentées par les différents points. La courbe bleue et ses incertitudes sont la médiane et les erreurs à  $1\sigma$  et  $2\sigma$  qu’ils obtiennent par un modèle Bayésien à partir du signal Ly $\alpha$ , de la profondeur optique du CMB et de fonctions de luminosités. On voit que la Réionisation se termine à un redshift inférieur à 6 et la courbe nous donne une idée du déroulé de l’histoire de réionisation.

atome dans l’état fondamental  $n = 1$  peut alors absorber un photon Ly $\alpha$  et se retrouver dans l’état  $n = 2$  pour ensuite se désexciter spontanément et revenir dans l’état fondamental mais avec un état de spin différent. Cet effet signifie que la température de spin est couplée avec la température correspondant au rayonnement Ly $\alpha$  et les deux états de spin sont présents dans le gaz.

Les photons du CMB se propagent dans l’Univers et quand ils traversent une région contenant du gaz d’hydrogène neutre, ils peuvent être absorbés par ce gaz en excitant l’hydrogène de l’état  $n_0$  à l’état  $n_1$ . Les atomes d’hydrogène se trouvant dans l’état  $n_1$  peuvent également se désexciter spontanément pour revenir à l’état fondamental  $n_0$  en émettant un photon. Ces émissions ou absorptions ont alors un impact sur la température de brillance  $\delta T_b$ , par rapport à celle du CMB, que l’on note  $T_{\text{CMB}}$ . Ce changement de température de brillance est le signal qui nous intéresse et il s’exprime de la manière suivante (Mellema et al. 2015; Bianco et al. 2021) :

$$\begin{aligned} \delta T_b(z) \approx & 27 x_{\text{HI}}(z) (1 + \delta_b(z)) \left( \frac{1+z}{10} \right)^{\frac{1}{2}} \left( 1 - \frac{T_{\text{CMB}}(z)}{T_s(z)} \right) \\ & \times \left( 1 + \frac{1}{H(z)} \frac{dv_{\parallel}}{dr_{\parallel}} \right) \left( \frac{\Omega_b}{0.044} \frac{h}{0.7} \right) \left( \frac{\Omega_m}{0.27} \right)^{-\frac{1}{2}} \text{ mK}. \end{aligned} \quad (2.2)$$



**FIGURE 2.8** – Illustration du retournement du spin de l’électron dans l’atome d’hydrogène neutre. On remarque notamment l’émission d’un photon à une longueur d’onde de 21 cm ou une fréquence de 1 420 MHz.

On voit dans cette équation que ce signal dépend de plusieurs physiques :

- Il dépend tout d’abord de l’état d’ionisation du gaz  $x_{\text{HI}}$ . Si tout le gaz est réionisé, alors  $x_{\text{HI}} = 0$  et  $\delta T_b = 0$ .
- Ensuite, ce signal dépend de la fluctuation de densité des baryons  $\delta_b$  : plus une région est dense et plus le signal 21 cm sera intense.
- On retrouve également dans cette différence de température  $\delta T_b$ , le rapport entre la température du CMB  $T_{\text{CMB}}$  et la température de spin  $T_s$ , ou plutôt le contraste  $T_s - T_{\text{CMB}}$ . Celui-ci provient de l’état d’excitation du gaz d’hydrogène : si  $T_s > T_{\text{CMB}}$  et  $\delta T_b > 0$ , alors le gaz est en émission, et si  $T_s < T_{\text{CMB}}$  et  $\delta T_b < 0$ , alors il est en absorption. De plus, si  $T_s = T_{\text{CMB}}$  (et  $\delta T_b = 0$ ), alors les émissions sont compensées par les absorptions et on ne pourra pas détecter d’excitation particulière du gaz. Dans ce cas, l’état de spin est en équilibre thermique avec le CMB.
- $\delta T_b$  dépend aussi de la cosmologie par le biais des valeurs actuelles de la densité de baryons  $\Omega_b$ , de la densité de matière  $\Omega_m$  et de la constante de Hubble  $h = H_0/100$ .
- Ce signal dépend enfin des distorsions de l’espace des redshifts qui sont dues aux vitesses particulières  $dv_{\parallel}/dr_{\parallel}$  le long de la ligne de visée.

La température de brillance différentielle est de plus une observable qui dépend du redshift auquel on l’observe. Cette dépendance se retrouve dans les grandeurs physiques dont elle dépend. Cela signifie qu’en observant le signal  $\delta T_b$  à différents redshifts (ou fréquences le long de la ligne de visée), on peut suivre l’évolution temporelle des différents mécanismes physiques, dont le processus de réionisation. Cette observable est donc plus pratique que celles mentionnées dans la Section précédente si l’on souhaite étudier en détail la Réionisation parce qu’elle nous donne accès à l’évolution temporelle de la distribution spatiale de l’hydrogène neutre. En effet, on ne détecte qu’un nombre limité de galaxies ou de quasars (qui produisent les spectres Ly $\alpha$ ) par les télescopes, ce qui restreint notre observation temporelle du processus



de Réionisation. Il en va de même pour les observations basées sur le CMB à cause de leur nature intégrée du processus. Le signal 21 cm est donc une sonde de la Réionisation très prometteuse.

### 2.2.2 Signal global

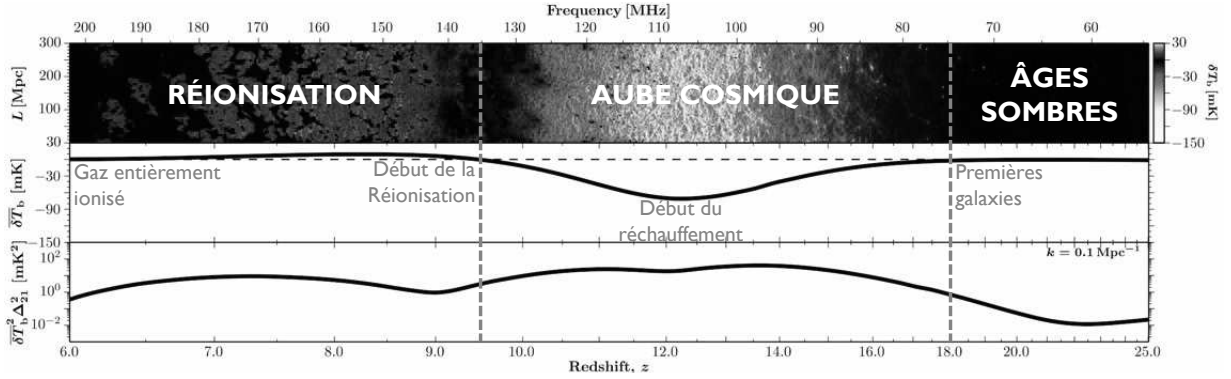
Il existe plusieurs moyens de mesurer le signal 21 cm. On peut tout d'abord mesurer le signal 21 cm moyen sur tout le ciel, que l'on appelle signal global. Ce signal correspond directement à la température de brillance différentielle moyenne  $\overline{\delta T_b}$  sur le ciel à un redshift donné, et j'en montre un exemple extrait d'une simulation faite par Park et al. (2019) dans la Figure 2.9. Dans cette Figure, on peut suivre l'évolution temporelle du signal 21 cm dans une frise sur la première ligne, ainsi que l'évolution temporelle du signal global dans la seconde ligne. Ces résultats proviennent d'une simulation 21cmFAST et ne représentent donc pas forcément la réalité, en particulier pour la forme du signal global et pour le déroulé temporel, mais ils vont nous servir ici de base pour expliquer les différentes phases auxquelles on peut s'attendre dans ce type de signal. Le temps (ou la fréquence d'observation) augmente de droite à gauche et inversement pour le redshift, et la physique du signal évolue comme suit :

- À un redshift  $z > 20$  environ, on est dans l'Époque des Âges Sombres, pendant laquelle  $T_s = T_{\text{CMB}}$ . À cette époque, le signal global est donc nul ( $\overline{\delta T_b} = 0$ ).
- À mesure que le temps avance, des instabilités gravitationnelles apparaissent et le gaz s'effondre pour former les premières galaxies (vers  $z \sim 18$  sur la frise). Les étoiles de ces galaxies émettent alors un rayonnement UV, notamment Ly $\alpha$ , et l'effet WF se met en place. Il y a donc un couplage Ly $\alpha$  et l'excitation du gaz commence à être régie par le rayonnement des premières étoiles et impose  $T_s = T_{\text{gaz}} < T_{\text{CMB}}$ . Le gaz passe alors en absorption et  $\overline{\delta T_b} < 0$ .
- Ensuite (vers  $z \sim 12$  sur la frise), le gaz commence à se réchauffer à cause du rayonnement émis par les premières étoiles et en particulier avec l'apparition d'objets sources de rayons X, comme des étoiles binaires ou les premiers noyaux actifs des galaxies (AGN). À partir de ce moment, la dynamique du processus change parce que la température de spin, qui est toujours couplée avec celle du gaz, suit l'élévation de cette dernière, même si le gaz reste toujours majoritairement en absorption ( $\overline{\delta T_b} < 0$ ). La température du gaz et donc de spin augmente jusqu'à atteindre de nouveau celle du CMB ( $T_s = T_{\text{CMB}}$  et  $\overline{\delta T_b} = 0$  vers  $z \sim 9.5$  sur la frise), et la dépasser quand le gaz passe en émission ( $T_s > T_{\text{CMB}}$  et  $\overline{\delta T_b} > 0$ ).
- C'est à ce moment que la Réionisation s'enclenche. Le rayonnement UV des premières galaxies commence à ioniser le gaz autour d'elles et le signal 21 cm est alors dominé par ce processus.  $\overline{\delta T_b}$  diminue en suivant la fraction de gaz neutre  $x_{\text{HI}}$  jusqu'à la fin de la Réionisation. Ainsi, vers  $z \sim 6$  sur la frise, on ne détecte plus de signal 21 cm parce qu'il n'y a plus d'atomes d'hydrogène neutre dans l'Univers et  $\overline{\delta T_b} = 0$ .

Si l'on peut mesurer ce signal global, on pourra alors comprendre comme a évolué le gaz d'hydrogène neutre en moyenne dans le temps, et en particulier suivre le processus de réionisation. Plusieurs instruments comme EDGES<sup>6</sup> (Experiment to Detect the Global EoR

---

6. <http://loco.lab.asu.edu/edges/>



**FIGURE 2.9** – Frise du signal 21 cm (en haut), signal moyen (au milieu) et spectre de puissance (en bas) représentés en fonction de la fréquence et du redshift d’observation. Ce signal 21 cm est modélisé avec une simulation semi-analytique 21cmFAST. Comme ce signal 21 cm est obtenu avec un modèle de Réionisation donné, l’évolution en redshift du signal sert seulement d’exemple ici et ne représente pas nécessairement la réalité. Cette Figure est extraite de Park et al. (2019) et a été annotée par la suite.

Signature), SARAS<sup>7</sup> (Shaped Antenna measurement of the background RAdio Spectrum) ou LEDA<sup>8</sup> (Large-aperture Experiment to detect the Dark Ages) ont essayé de mesurer ce signal. SARAS n’a pour le moment pas fait de mesure significative et a juste permis de rejeter certains scénarios de Réionisation, ayant un réchauffement par photons X-ray inefficace ou une réionisation trop rapide (Singh et al. 2018). EDGES a en revanche détecté un puits profond dans le signal global après l’apparition des premières galaxies, dont l’interprétation est incertaine (Bowman et al. 2018). Ce signal trop intense signifierait que le gaz est bien plus froid que ce à quoi on s’attendait. On ne l’explique pas avec les modèles standards et des modèles plus exotiques ont été proposés, comme avec de la matière noire qui refroidit. Néanmoins, le consensus actuel est que ce résultat n’est pas d’origine physique et proviendrait de la chaîne de traitement (Hills et al. 2018; Bradley et al. 2019; Sims & Pober 2020; Singh et al. 2022).

### 2.2.3 Spectre de puissance

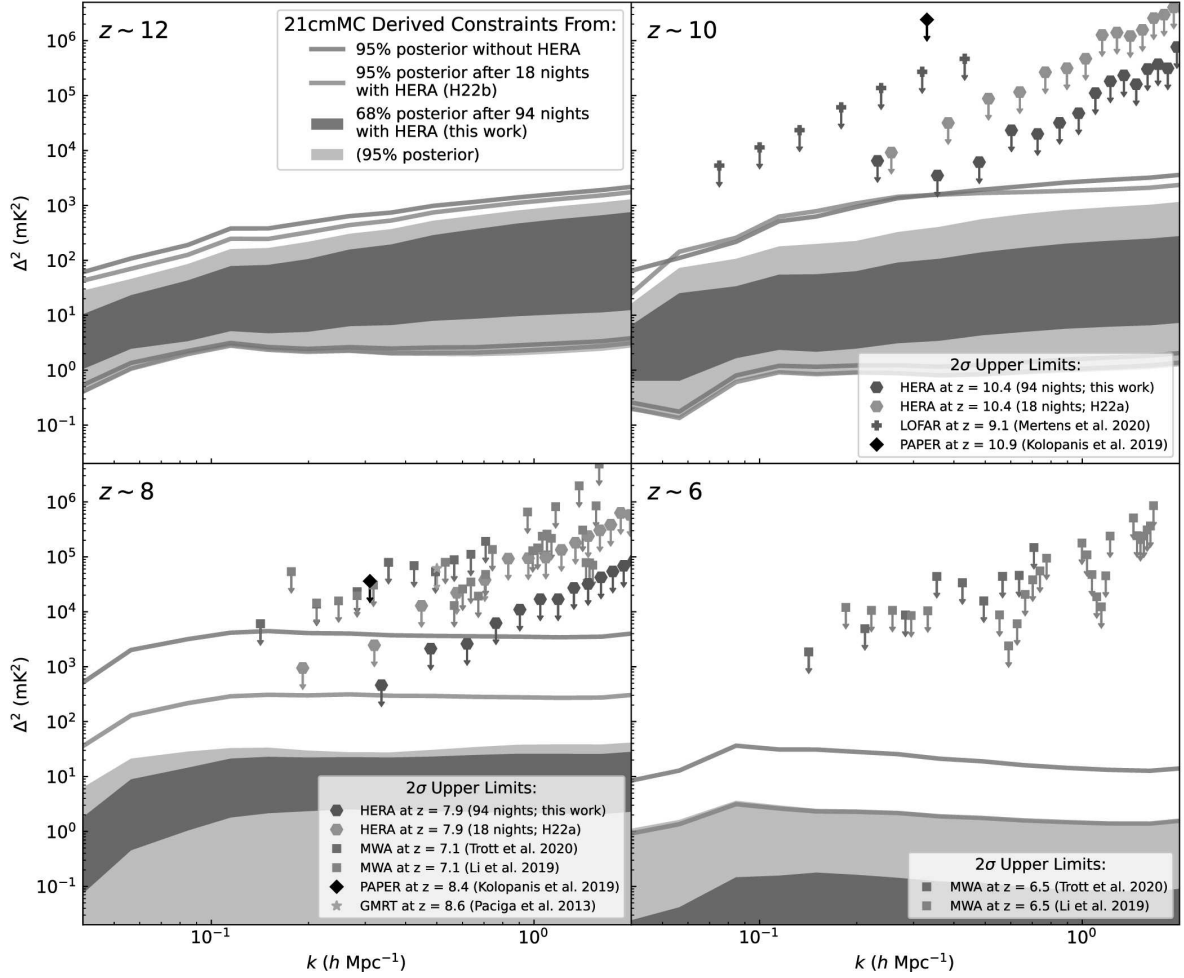
Une autre manière d’observer le signal 21 cm est de mesurer son spectre de puissance, que l’on peut définir comme suit (Furlanetto et al. 2006) :

$$\left\langle \tilde{\delta}_{21}(z, \vec{k}) \tilde{\delta}_{21}(z, \vec{k}') \right\rangle = (2\pi)^3 P_{21}(z, k) \delta(\vec{k} + \vec{k}'), \quad (2.3)$$

avec  $\delta$  la fonction de Dirac. On note ici  $\delta_{21}(z, \vec{r}) = (\delta T_b(z, \vec{r}) - \overline{\delta T_b}) / \overline{\delta T_b}$  et  $\tilde{\delta}_{21}$  la transformée de Fourier de  $\delta_{21}$ . Par définition, ce spectre contient alors des informations sur la structure spatiale du signal 21 cm. Tout comme le signal 21 cm  $\delta T_b$ , son spectre de puissance évolue dans le temps et il est différent selon le redshift auquel on l’étudie. Dans la Figure 2.9, je montre dans la troisième ligne le spectre de puissance pris à une échelle spatiale donnée par

7. <http://www.rri.res.in/DISTORTION/saras.html>

8. <http://www.tauceti.caltech.edu/leda/>



**FIGURE 2.10** – Dernières limites supérieures du spectre de puissance obtenues par la collaboration HERA à plusieurs redshifts (voir les points violets pour 94 nuits d’observations et les points oranges pour 18 nuits d’observations dans les différents panneaux). Les autres points représentent les résultats obtenus par d’autres télescopes. Les zones violettes représentent les spectres de puissance inférés par modèle Bayésien (avec 21cMMC) en utilisant 94 nuits d’observation avec HERA dans un intervalle de crédibilité à 68% en violet foncé et 95% en violet clair. Les lignes oranges et grises représentent l’intervalle à 95% de crédibilité du spectre de puissance inféré avec 18 nuits d’observation avec HERA et sans HERA respectivement. Cette Figure est extraite de HERA Collaboration et al. (2023).

$k = 0.1 \text{ Mpc}^{-1}$  en fonction du redshift, que l'on peut comparer à la frise du signal 21 cm et au signal global. On peut voir avec cette figure que l'on peut retrouver dans le spectre de puissance du signal 21 cm l'évolution de ce signal : on retrouve par exemple un minimum local au moment du réchauffement de l'hydrogène, ou bien au début de la Réionisation. Bien que le signal 21 cm est connu pour être non-Gaussien (Mellema et al. 2006; Dixon et al. 2016; Banet et al. 2021), le spectre de puissance de signal contient tout de même beaucoup d'information et notamment sur le processus de réionisation (Zaldarriaga et al. 2004; Furlanetto et al. 2004a; Bowman et al. 2006; Lidz et al. 2008). Avec ce spectre, on peut en effet contraindre les paramètres astrophysiques et la topologie de la Réionisation (Iliev et al. 2012; Mesinger et al. 2013; Pober et al. 2014; Greig & Mesinger 2015, 2017; Liu & Parsons 2016; Kern et al. 2017; Hutter et al. 2020; Pagano & Liu 2020; Shimabukuro et al. 2022). Je reviendrais sur ces études dans le Chapitre 3.

Plusieurs interféromètres dans le domaine radio sont en ce moment construits et utilisés pour détecter ce spectre de puissance : MWA<sup>9</sup> (Murchison Wide Field Array), LOFAR<sup>10</sup> (Low Frequency Array), PAPER<sup>11</sup> (Precision Array for Probing Epoch of Reionisation), HERA<sup>12</sup>, NenuFAR<sup>13</sup> (New Extension in Nançay Upgrading LOFAR), et SKA<sup>14</sup> (Square Kilometer Array). Les contraintes les plus récentes et les plus fortes du spectre de puissance ont été obtenues récemment avec le télescope HERA, et sont montrées dans la Figure 2.10 pour quatre redshifts différents avec les points violets. Pendant longtemps, ces contraintes ont surtout été des limites supérieures au spectre de puissance en raison du bruit d'observation, qui permettaient déjà d'exclure des modèles de Réionisation. Mais dans leurs résultats récents, HERA Collaboration et al. (2023) ont obtenu par modèle Bayésien des intervalles dans lesquels le spectre de puissance devrait se trouver. Ils commencent aussi à mettre de vraies contraintes sur le signal 21 cm, en nous fournissant la première preuve démontrant que l'IGM doit avoir été réchauffé a priori par des sources de rayons X pour repasser en émission au début de la Réionisation.

## 2.2.4 Tomographie avec le radio-télescope SKA

Enfin, la manière la plus directe d'observer l'évolution du processus de réionisation est de mesurer la tomographie du signal 21 cm, c'est-à-dire obtenir des cartes 2D sur le plan du ciel à différentes fréquences ou redshifts le long de la ligne de visée. C'est l'objectif du radio-télescope SKA, et plus particulièrement des antennes SKA-Low, que l'on peut voir dans la Figure 2.11, et qui sont en train d'être construites en Australie à l'heure actuelle. Ces 131 072 antennes mesurent 2 m de haut et sont réparties dans 512 stations pour obtenir une surface collectrice totale de  $0.4 \text{ km}^2$ . La distance maximale entre les stations d'antenne est de 74 km. On sait déjà que ce télescope va produire énormément de données (157 To/s, soit 4.9 Zo par an), qu'il faudra stocker dans de très grands centres de données et analyser. Cette quantité de données est l'équivalent du trafic internet mondial actuel. Par ailleurs,

9. <https://www.mwatelescope.org>

10. <https://www.astron.nl/telescopes/lofar/>

11. <http://eor.berkeley.edu>

12. <http://reionization.org>

13. <https://nenufar.obs-nancay.fr>

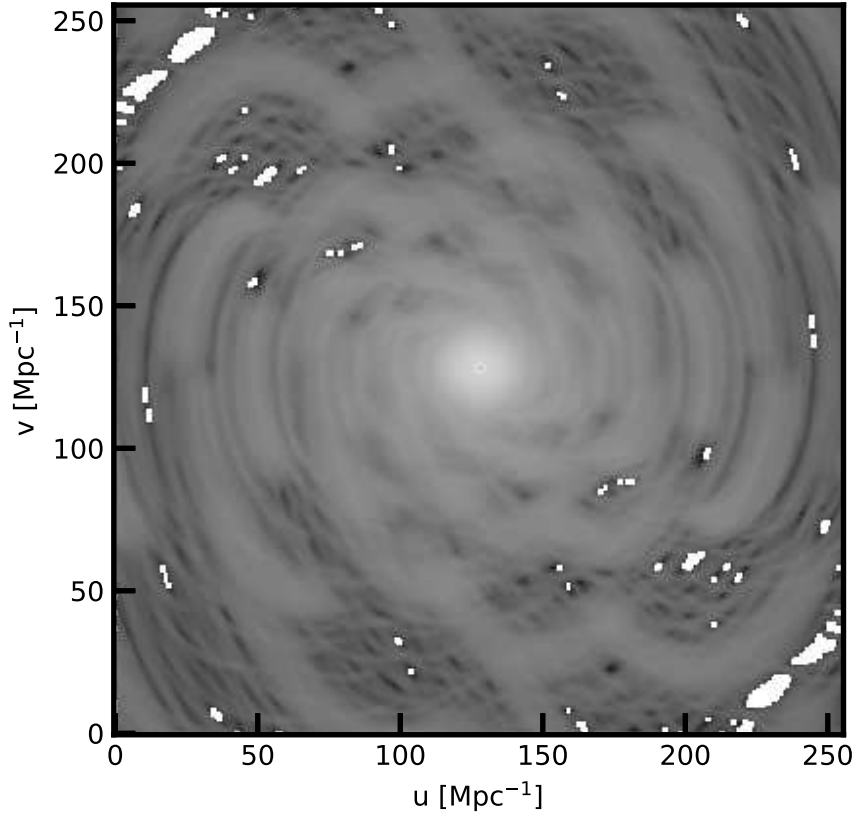
14. <https://www.skao.int>



**FIGURE 2.11** – Vue d’artiste représentant les futures antennes de SKA-Low qui seront construites en Australie. La disposition de ces antennes, ainsi que la bande de fréquences observée est spécifiée dans le cadre bleu en haut à gauche. Cette image est composée de données provenant du site internet de SKAO (<https://www.skao.int/en>).

pour comparer cet instrument à un instrument déjà construit, LOFAR, SKA-Low aura une résolution améliorée de 25% et il sera 8 fois plus sensible. Enfin, la bande de fréquences prévue pour les observations du signal 21 cm est comprise entre 50 et 350 MHz, soit entre les redshifts 6 et 25 (Aube Cosmique et Réionisation). Le champ de vue observé sera de  $200^\circ \times 200^\circ$ , soit plus de  $1 \text{ Gpc}^2$ . On peut espérer résoudre au moins les échelles de 7 Mpc pour des redshifts  $z > 6.5$  dans le plan du ciel et les échelles de 1 Mpc dans la direction des fréquences (Giri et al. 2018a). D’ici la fin de la décennie, l’observatoire SKA (SKAO pour SKA Observatory) pourra mesurer le signal 21 cm afin que la communauté essaie d’en extraire des informations sur le processus de réionisation en particulier.

Dans la Figure 2.9, la première ligne montre un exemple de tomographie du signal 21 cm : c’est un cône de lumière (lightcone) de  $\delta T_b$  simulé avec le code semi-analytique 21cmFAST par Park et al. (2019) pour un scénario de Réionisation donné. On retrouve les différentes époques de l’Univers à ces redshifts-là, à savoir les Âges sombres (à des redshifts  $z > 18$  sur la frise), l’Aube Cosmique (aux redshifts  $9.5 < z < 18$  sur la frise) et l’Époque de la Réionisation (aux redshifts  $6 < z < 9.5$  sur la frise). Si l’on se focalise sur la Réionisation, on peut par exemple voir les bulles ionisées qui apparaissent (en noir), qui grossissent, fusionnent et percolent à mesure que le redshift diminue. Avec ce genre d’observations tomographiques, on observe le ciel à différentes fréquences (ou redshifts), donc la région du ciel observée à un certain redshift sera différente de celle observée à un autre redshift. On ne peut donc pas étudier l’évolution d’une même portion de gaz à différent moment. En revanche, on peut s’intéresser



**FIGURE 2.12** – Couverture dans le plan  $uv$  par le radio-télescope SKA à un redshift  $z_{\text{obs}} = 8$  (correspondant à une fréquence  $\nu \sim 160$  MHz). On voit notamment ici que toutes les régions du ciel ne sont pas observées de manière équivalente, avec certaines zones qui ne sont pas du tout observées (en blanc). Cette couverture dans le plan  $uv$  est utilisée ultérieurement dans ce manuscrit (Chapitre 4), et elle est obtenue en reprenant les paramètres standards : nous supposons une durée d’observation par jour de 6 heures, 10 secondes de temps d’intégration, un total de 1000 heures d’observation, et une ligne de base de 2 km maximum (Giri et al. 2018b; Bianco et al. 2021; Prelogović et al. 2022).

à la manière globale dont l’hydrogène neutre évolue. Ces considérations, un peu triviales, auront une certaine importance pour nos travaux sur le champ de temps de réionisation mentionnés dans les Chapitre suivants.

L’exemple de cône de lumière donné dans la Figure 2.9 correspond à un signal parfait, c’est-à-dire sans bruit quelconque (comme du bruit instrumental) ou avant-plans (Voie Lactée, sources ponctuelles par exemple). Pour ce qui est du bruit instrumental, on sait déjà que la disposition des antennes de SKA-Low va impacter les observations. Les antennes sont réparties de manière à maximiser le signal en tout point de la région du ciel observée, mais le signal sera quand même détecté de manière non-uniforme. En interférométrie, on mesure le signal dans le domaine de Fourier. La couverture du plan  $uv$  (plan des fréquences spatiales dans le domaine de Fourier), que je montre dans la Figure 2.12, représente les différentes

échelles qui seront présentes dans l'image finale observée par SKA à un redshift d'observation de 8. Le centre de l'image correspond aux petites fréquences spatiales, c'est-à-dire aux grandes échelles spatiales, tandis que les bords de l'image correspondent aux grandes fréquences spatiales, c'est-à-dire aux petites échelles spatiales. On voit donc ici que SKA fournira beaucoup de signal aux grandes échelles et pourra en manquer aux plus petites échelles, avec des échelles (zones blanches du plan uv) qui seront totalement absentes des cartes produites. Maintenant, connaissant cette couverture, on pourra la prendre en compte dans l'analyse future des observations. Les avant-plans (qui ont une amplitude d'un ordre de grandeur 4 à 5 fois supérieur au signal cosmologique ; Gorce et al. (2023)) sont néanmoins une source de « nuisance » plus compliquée à prendre en compte, notamment parce qu'ils peuvent provenir de sources distantes et agir sur de grandes distances. À l'heure actuelle, SKAO organise des défis avec des données synthétiques pour amener la communauté à en retirer les avant-plans.<sup>15</sup>

Les observations tomographiques du signal 21 cm sont ainsi très prometteuses parce qu'elles pourront nous permettre pour la première fois de sonder directement l'Époque de la Réionisation. Le travail de ma thèse se place dans ce contexte et a notamment pour but de démontrer que l'on peut accéder à tout le déroulé du processus de réionisation à partir de ces cartes 2D (voir les Chapitres 4, 5 et 6). Par ailleurs, beaucoup d'équipes travaillent activement dans ce domaine et je vais revenir sur cela plus en détail dans le Chapitre 3.

## 2.3 Simulations de la Réionisation

Nous avons vu, qu'à l'heure actuelle, nous ne pouvons pas encore sonder directement l'Époque de la Réionisation et que nous sommes dans l'attente de futures observations directes de cette époque avec le signal 21 cm. Pour nous préparer à analyser ce signal, nous simulons la Réionisation avec différents types de modèles que je vais présenter dans cette Section. Ces simulations sont aussi très utiles pour comprendre la Réionisation au travers de différents modèles de Réionisation dont je vais présenter certains paramètres ici. Néanmoins, elles sont complexes à mettre en oeuvre parce que plusieurs échelles entrent en jeu : pour modéliser la formation d'étoiles au sein des galaxies, nous avons besoin d'une haute résolution, mais il faut aussi modéliser de grands volumes pour prendre en compte la variance cosmique. Par ailleurs, plusieurs physiques sont nécessaires pour une modélisation optimale. Il faut, en particulier, résoudre le transfert radiatif car la Réionisation est avant-tout un processus d'interaction entre lumière et matière (ionisation, chauffage, etc.).

Dans cette Section, je vais me focaliser sur deux modèles que j'ai utilisé pendant ma thèse, à savoir le code de simulation cosmologique **EMMA** (Aubert et al. 2015) et le modèle semi-analytique **21cmFAST**<sup>16</sup> (Mesinger et al. 2011; Murray et al. 2020). Je vais décrire la manière dont ces simulations fonctionnent, mais je ne suis pas experte sur ces aspect-là. En effet, pendant ma thèse, je n'ai pas développé de tels algorithmes de simulations. J'ai cependant généré plusieurs simulations sur différentes machines de calculs : sur **Krusty**, une machine locale à l'Observatoire de Strasbourg, sur le mésocentre de l'Université de Strasbourg<sup>17</sup> et sur

---

15. <https://sdc2.astronomers.skatelescope.org/sdc2-challenge/description>

16. <https://github.com/21cmfast/21cmFAST>

17. <https://hpc.pages.unistra.fr>

Occigen, un super-calculateur du CINES (Centre Informatique National de l’Enseignement Supérieur).

### 2.3.1 Simulations cosmologiques : l’exemple d’EMMA

Les simulations cosmologiques de la Réionisation sont des modèles multi-physiques qui résolvent des équations différentielles régissant les différentes physiques en jeu pendant la Réionisation. Il existe plusieurs codes de ce type, dont entre autres RAMSES<sup>18</sup> (Teyssier 2002), EMMA (Électromagnétique et Mécanique sur Maille Adaptative ; Aubert et al. (2015)), LICORICE (Semelin 2016), SCORCH (Simulations and Constructions of the Reionisation of Cosmic Hydrogen ; Doussot et al. (2019)), et le projet THESAN<sup>19</sup> (Kannan & Thesan Team 2022). À partir de ces codes de simulations, on peut générer des suites de simulations, comme CoDa I, II et III<sup>20</sup> qui est développée à Strasbourg (Ocvirk et al. 2016, 2020; Lewis et al. 2022), SPHINX-MHD<sup>21</sup> (Katz et al. 2021) et le projet Sherwood-Relics<sup>22</sup> (Puchwein et al. 2023). Dans cette Section, je vais me focaliser sur la simulation cosmologique EMMA et détailler son fonctionnement et les physiques qui y sont résolues.

EMMA est un code parallélisé et à raffinement adaptatif de mailles (Adaptive Mesh Refinement, AMR) qui décrit des quantités physiques sur une grille. L’AMR permet d’augmenter de manière dynamique la résolution spatiale tout en ayant un coût modéré en terme de consommation de mémoire. Il est dynamique dans le sens où les mailles vont être raffinées seulement à certains endroits, là où le gaz est très dense par exemple. EMMA implémente ce raffinement adaptatif avec la méthode FTT (Fully Threaded Tree) qui a été introduite par Khokhlov (1998). Une cellule raffinée va être divisée en huit cellules plus petites, qui peuvent aussi être chacune raffinée jusqu’à ce que le niveau de raffinement souhaité par l’utilisateur soit atteint. La parallélisation d’EMMA se fait à deux échelles. Il y a tout d’abord une parallélisation standard qui distribue les tâches sur plusieurs processeurs CPU communiquant entre eux grâce au protocole MPI. Ensuite, le second niveau de parallélisation se fait au sein d’un processus MPI dans lequel la charge de calcul est répartie entre plusieurs threads (fils) d’exécution. Cette parallélisation locale est faite par des processeurs graphiques (GPU) et implémentée grâce à CUDA.

Plusieurs physiques sont résolues dans EMMA et cela se fait par l’évolution de trois « fluides » qui sont couplés les uns aux autres :

- Le premier fluide est un fluide sans collisions qui modélise la dynamique des étoiles et de la matière noire non-baryonique. Cette dynamique est gérée avec un échantillonnage Monte-Carlo de l’espace des phases en utilisant des particules. Les particules évoluent dans l’espace grâce aux équations de Newton, qui nécessitent de connaître la densité et le potentiel sur chaque maille raffinée. La densité est obtenue avec le schéma d’interpolation standard CIC (Cloud-In-Cell). À partir de la densité, EMMA calcule le potentiel avec l’équation de Poisson dont la solution est obtenue avec des techniques de relaxation. Ces techniques sont pratiques dans le sens où elles peuvent s’appliquer dans des

18. <https://bitbucket.org/rteyssie/ramses/src/master/>

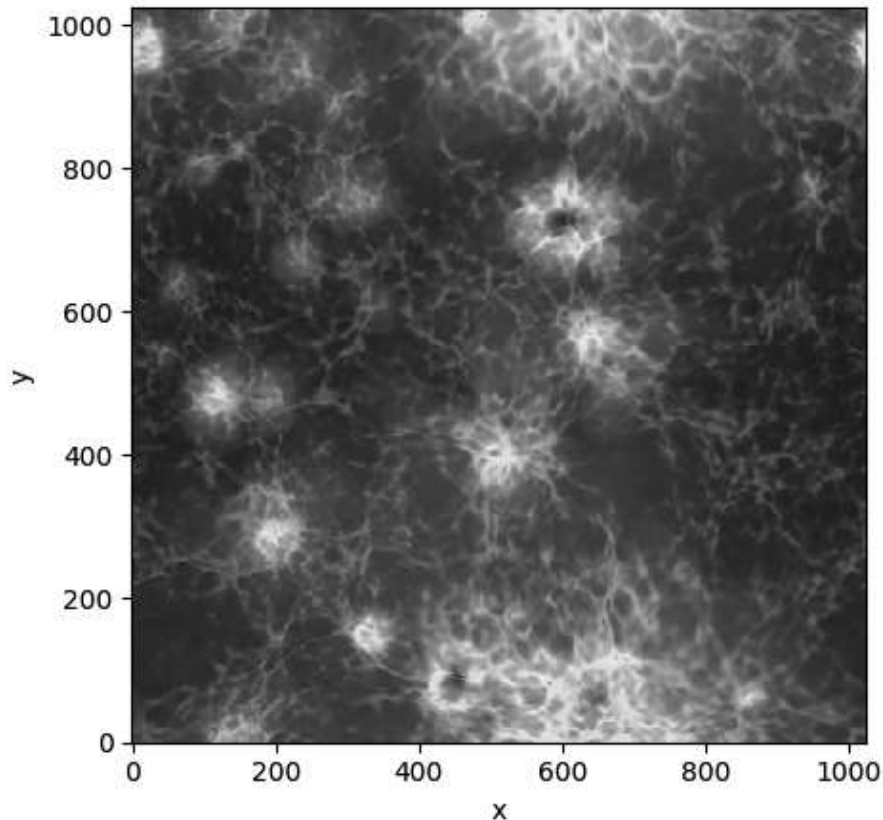
19. <https://www.thesan-project.com>

20. <https://coda-simulation.github.io>

21. <https://sphinx.univ-lyon1.fr>

22. <https://www.nottingham.ac.uk/astronomy/sherwood-relics/>





**FIGURE 2.13** – Carte extraite d’une simulation EMMA de taille  $1024^3 \text{ cMpc}^3/h^3$ . Cette carte représente la densité superposée au champ de temps de réionisation. On peut suivre la densité avec les filaments de matière. Plus une région est bleue, plus elle est dense et plus elle réionise tôt. Plus une région est rouge foncée, moins elle est dense et plus elle réionise tard.

géométries complexes, avec des conditions aux limites arbitraires et des grilles à résolutions multiples. Enfin, le champ d’accélération est obtenu en dérivant simplement le potentiel et est interpolé à la position des particules avec un CIC.

- Le deuxième fluide décrit l’hydrodynamique du gaz qui évolue selon les équations d’Euler. EMMA les résout grâce au schéma de Godunov, qui permet de résoudre des équations aux dérivées partielles en utilisant des volumes finis. La stabilité de la solution obtenue est assurée en imposant une limite supérieure au pas de temps grâce à la condition de Courant.
- Le troisième fluide, qui décrit la propagation du rayonnement, est modélisé par des équations de transfert radiatif, à savoir les deux premiers moments de l’équation de Liouville, qui décrivent la conservation de la densité de photons ionisants et de son flux. On ajoute à ces équations la relation de fermeture M1 (Levermore 1984; Aubert & Teyssier 2008). Cette relation indique que la pression radiative peut suivre un régime de diffusion et un régime de transport pur. Dans cette approximation, la lumière est décrite comme étant un fluide.

En plus de ces fluides, des processus thermiques et chimiques sont modélisés. EMMA se foca-

lise sur les processus atomiques de refroidissement et d’ionisation de l’hydrogène uniquement, en plus de l’évolution adiabatique de ce gaz. Ces processus affectent les processus hydrodynamiques et radiatifs de plusieurs façons. La densité et le flux de photons sont modifiés par la création de sources (les étoiles) et de puits (le gaz neutre). Dans les simulations que j’utilise pendant ma thèse, les sources sont créées via un processus de formation d’étoiles développé et calibré à partir de la simulation CoDa II pour modéliser cette formation d’étoiles sur des échelles galactiques (Gillet et al. 2021). Cela permet de modéliser la Réionisation dans de grands volumes tout en résolvant les (petites) structures responsables de ce processus. L’état d’ionisation du gaz est aussi pris en compte et est impacté par les effets de recombinaison et d’ionisation, qui sont en compétition. La densité d’énergie interne du gaz est également modifiée à cause du refroidissement atomique et du réchauffement radiatif. Enfin, comme il s’agit de simulations cosmologiques, **EMMA** utilise les coordonnées « super-comobiles », dictée par le facteur d’expansion, pour prendre en compte l’effet de l’expansion.

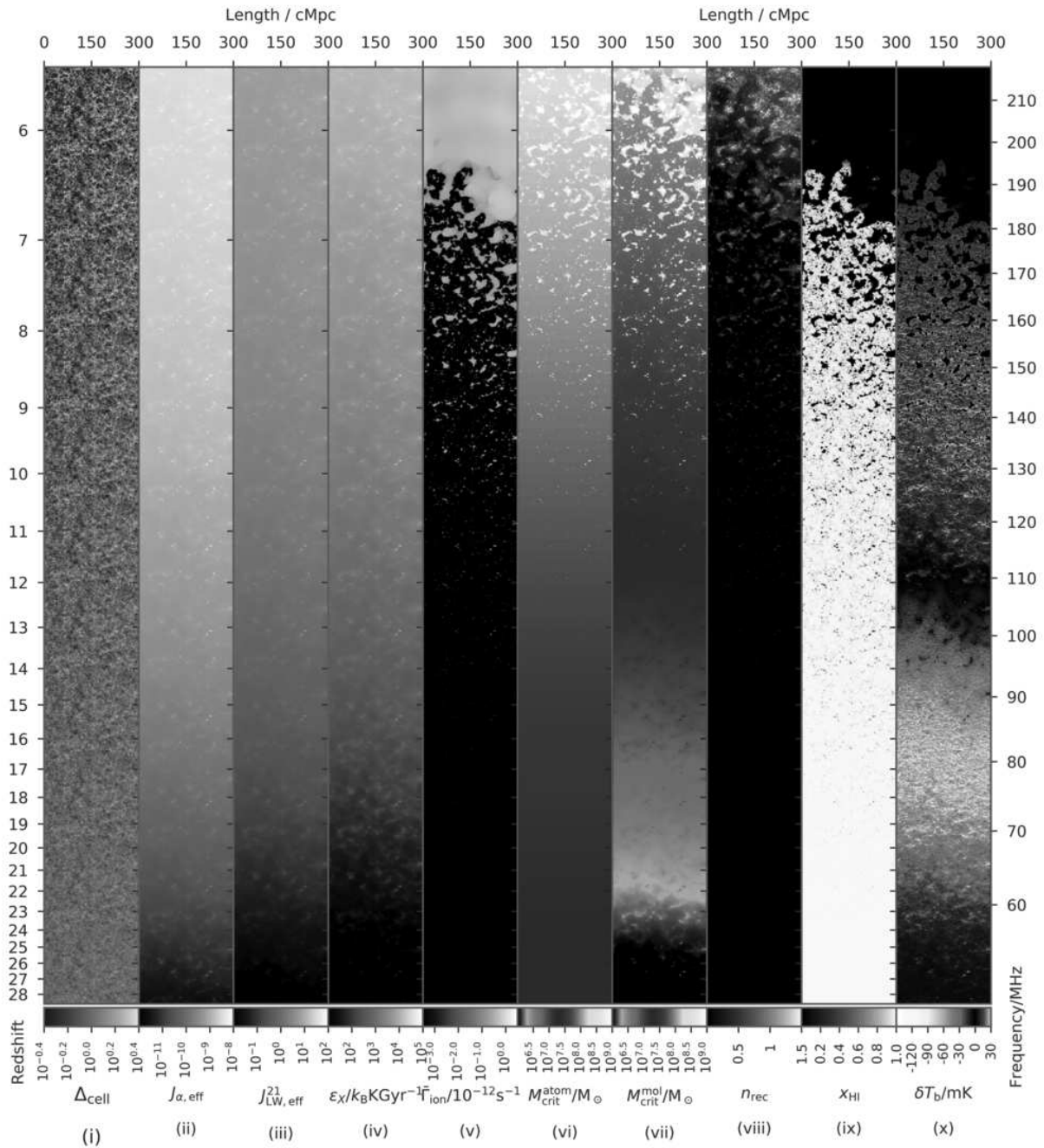
Par ailleurs, le pas de temps d’une itération d’une simulation **EMMA** est calibré en fonction des différentes physiques en jeu ici. Il va être limité par l’expansion cosmologique et par la taille d’une cellule. Il doit permettre de résoudre le temps de dynamique local, et doit satisfaire la condition hydrodynamique de Courant (qui permet la stabilité des solutions de l’hydrodynamique du gaz) et la condition radiative de Courant quand des sources sont présentes. La plus forte contrainte sur le pas de temps provient de cette dernière condition, qui impose alors des pas de temps courts liés à la grande vitesse de propagation du rayonnement, augmentant le temps de calcul de la simulation. Une manière de réduire ce temps de calcul est d’utiliser une vitesse de la lumière réduite, comme le propose la simulation **EMMA** de manière optionnelle. La vitesse de la lumière réduite peut avoir un impact sur le processus de réionisation modélisé et ceci est par exemple discuté par [DeParis et al. \(2019\)](#). Toutefois, grâce à l’utilisation de GPU (qui accélèrent le temps de calcul), simuler la Réionisation avec une vitesse de la lumière réduite n’a pas été nécessaire par la suite dans nos travaux.

Une simulation **EMMA** nous donne ainsi accès à plusieurs champs physiques permettant de décrire la Réionisation. On peut par exemple extraire le champ de densité et le champ de temps de réionisation, que je montre de manière superposée dans la Figure 2.13.

### 2.3.2 Modèles semi-analytiques : l’exemple de 21cmFAST

Comme nous l’avons vu précédemment, les simulations cosmologiques résolvent plusieurs physiques pour modéliser le processus de réionisation de la manière la plus précise possible. Prendre en compte toutes ces physiques, et notamment le transfert radiatif, demande ainsi un temps de calcul relativement grand. Les simulations semi-analytiques ont donc vu le jour afin de réduire ce temps de calcul en faisant des approximations de sorte à faire un compromis entre temps de calcul et auto-consistance de la modélisation. Ces simulations ont en général un temps de calcul bien inférieur aux simulations cosmologiques (quelques minutes pour **21cmFAST** pour un problème demandant plusieurs jours pour **EMMA** par exemple). Néanmoins, les deux types de simulations implémentent les physiques différemment, et de ce fait il n’existe pas de correspondance directe entre les paramètres des différents modèles. En pratique, on ne peut donc pas toujours faire des modèles identiques avec ces deux types de simulations, et l’on ne sait pas toujours comment se comparent les différents modèles entre eux.

Plusieurs simulations semi-analytiques ont été développées, notamment par [Mesinger &](#)



**FIGURE 2.14** – Cônes de lumière extrait d’une simulation 21cmFAST et présentés sur le site <https://github.com/21cmfast/21cmFAST>. On voit plusieurs champs physiques auxquels on a accès avec ce genre de simulation. En particulier, on peut extraire la densité de matière (première colonne), le champ d’ionisation (avant-dernière colonne) et le signal 21 cm (dernière colonne).

Furlanetto (2007), Alvarez et al. (2009), Choudhury et al. (2009), Santos et al. (2010), Mesinger et al. (2011), Battaglia et al. (2013), Hassan et al. (2016), Xu et al. (2017), Hutter (2018), et Trac et al. (2022). Le modèle de simulation semi-analytique de la Réionisation le plus connu est **21cmFAST** (Mesinger et al. 2011; Murray et al. 2020), et c’est celui que je vais détailler ici parce que j’ai travaillé avec pendant ma thèse. Xu et al. (2017) ont développé le modèle **IslandFAST** en reprenant de **21cmFAST** pour modéliser de manière plus précise la fin de la Réionisation et pouvoir en particulier suivre correctement les dernières régions neutres se faisant réioniser. Plus récemment, Trac et al. (2022) ont développé **AMBER** (Abundance Matching Box for the Epoch of Reionisation) en utilisant une approche différente par rapport à **21cmFAST** et en se basant sur le champ de redshift de réionisation.

**21cmFAST** a été développé dans le but de modéliser rapidement et efficacement le signal 21 cm cosmologique que nous commençons tout juste à observer. Le fonctionnement global de **21cmFAST** repose sur la génération des réalisations 3D des champs de densité, d’ionisation, de vitesse particulière et de température de spin qu’il combine pour construire le champ de signal 21 cm, à savoir la température de brillance différentielle.

**21cmFAST** génère tout d’abord des conditions initiales avec des champs de densité et de vitesses (de manière similaire à celles générées pour des simulations cosmologiques). Pour faire évoluer le champ de densité, l’effondrement gravitationnel est approximé en déplaçant les particules selon la théorie des perturbations au premier-ordre développée par Zel’dovich (1970). Cette approximation permet d’aller au delà du régime linéaire et de se placer dans un régime quasi-linéaire. Elle permet également de faire évoluer le champ de vitesse. Par ailleurs, **21cmFAST** choisit de prendre en compte les distorsions de l’espace des redshifts dans leur champ de vitesse parce qu’elles sont connues pour impacter le signal 21 cm avant la Réionisation et dans ses premières phases. Il faut de plus noter que les baryons et les particules de matière noire sont ici traités de la même manière.

Le champ d’ionisation (ou la distribution des régions ionisées) est généré grâce à une approche par ensemble d’excursion (plus connu sous le nom d’« excursion set formalism » en anglais), qui a été utilisée dans ce contexte pour la première fois par Furlanetto et al. (2004a). Ce formalisme permet de détecter des structures dans un champ en s’intéressant à l’ensemble des points qui sont supérieurs à un certain seuil. **21cmFAST** utilise la méthode semi-analytique FFRT (Fast Fourier Radiative Transfer) développée par Zahn et al. (2011) qui implémente ce formalisme en calculant des fractions d’effondrement  $f_{\text{coll}}$  de halos à partir du champ de densité. L’idée de cette technique repose sur le fait que les sources de lumière réionisent l’hydrogène environnant, impliquant que la distribution des sources structure directement la Réionisation. Les régions ionisées sont alors identifiées en sondant les différentes échelles du champ de densité évolué en commençant par les plus grandes pour finir par les plus petites. Elles sont définies comme les régions dans lesquelles il y a un plus grand nombre de photons ionisants que d’atomes d’hydrogène neutre. Ainsi, à partir du champ de densité évolué, **21cmFAST** sonde les halos effondrés dans une sphère d’un certain rayon  $R$ . La cellule centrale de cette sphère, qui est localisée au point de coordonnées  $\vec{x}$ , est signalée comme ionisée si le critère suivant est respecté :  $f_{\text{coll}}(\vec{x}, z, R) \geq \zeta^{-1}$ .  $\zeta$  est un paramètre d’efficacité d’ionisation et  $f_{\text{coll}}$  est la fraction d’effondrement dans la sphère de rayon  $R$ . En suivant le formalisme d’ensemble d’excursion, le seuil de filtrage  $R$  diminue de  $R_{\text{max}}$  (exprimé en Mpc) à la taille d’une cellule  $R_{\text{cell}}$ .  $R_{\text{max}}$  est un paramètre libre de **21cmFAST** et peut être assimilé au parcours libre moyen dans l’IGM d’un photon ionisant à  $z \sim 7 - 10$ . Ce formalisme implique

ainsi que 21cmFAST ne résout pas les sources de réionisation, mais modélise simplement leur impact sur leur environnement.

21cmFAST calcule aussi le champ de température de spin en prenant en compte l'évolution de la température du gaz, le réchauffement et l'état d'ionisation du gaz, ainsi que le rayonnement Ly $\alpha$ . Ensuite, avec tous ces champs calculés auparavant, 21cmFAST peut les combiner pour reconstruire l'évolution du signal 21 cm pendant l'Aube Cosmique et la Réionisation.

21cmFAST permet ainsi de générer divers champs 3D reliés à la physique de l'Aube Cosmique et de la Réionisation. De plus, il peut fournir en sortie des cônes de lumière pour imiter les observations que l'on espère avoir avec SKA (tomographie du signal 21 cm, comprenant deux directions d'espace sur le plan du ciel et une direction en fréquence le long de la ligne de visée). Dans la Figure 2.14, je montre différents cônes de lumière que l'on peut extraire de ces simulations. On voit en particulier le champ de densité (première colonne), le champ d'ionisation (avant-dernière colonne) et le champ de signal 21 cm (dernière colonne). On a également accès à d'autres champs, comme le champ de redshift de réionisation, qui m'a particulièrement intéressée pendant ma thèse.

### 2.3.3 Temps/redshifts de réionisation

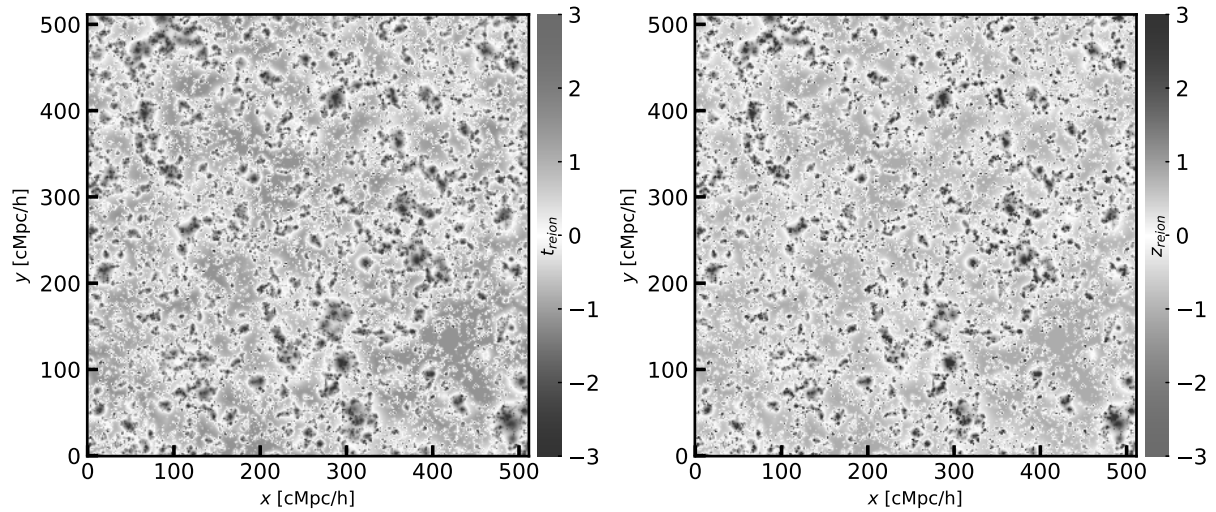
La réionisation de l'hydrogène durant le processus de réionisation est souvent étudiée au travers de boîtes de simulations montrant la distribution d'hydrogène ionisé à plusieurs redshift ou bien à travers le signal 21 cm. De manière complémentaire, j'ai choisi de m'intéresser pendant cette thèse à des boîtes indiquant directement le temps (ou le redshift) auquel l'hydrogène est réionisé dans chaque cellule d'une simulation. C'est ce qu'on appelle les temps (resp. redshifts) de réionisation  $t_{\text{reion}}(\vec{r})$  (resp.  $z_{\text{reion}}(\vec{r})$ ). Ce champ présente l'aspect pratique de résumer la même quantité d'information (que des champs ionisés pris à plusieurs redshifts tout au long d'une simulation) en un champ unique par simulation.

Dans la Figure 2.15, je montre un exemple de temps de réionisation à gauche et de redshifts de réionisation à droite extraits d'une simulation cosmologique EMMA. Dans ces cartes, je retire la moyenne à ces grandeurs et elles sont aussi normalisées par leur écart-type. La couleur représente le temps (resp. le redshift) auquel l'hydrogène est réionisé. Les régions bleues réionisent en premier autour des sources de réionisation, et les régions rouges réionisent en dernier dans les régions les moins denses en général. On voit ainsi que la Réionisation est effectivement un processus qui évolue dans le temps, et  $t_{\text{reion}}(\vec{r})$  (ou  $z_{\text{reion}}(\vec{r})$ ) permet d'étudier comment les fronts d'ionisation se propagent localement autour des sources de réionisation jusque dans les vides de matière.

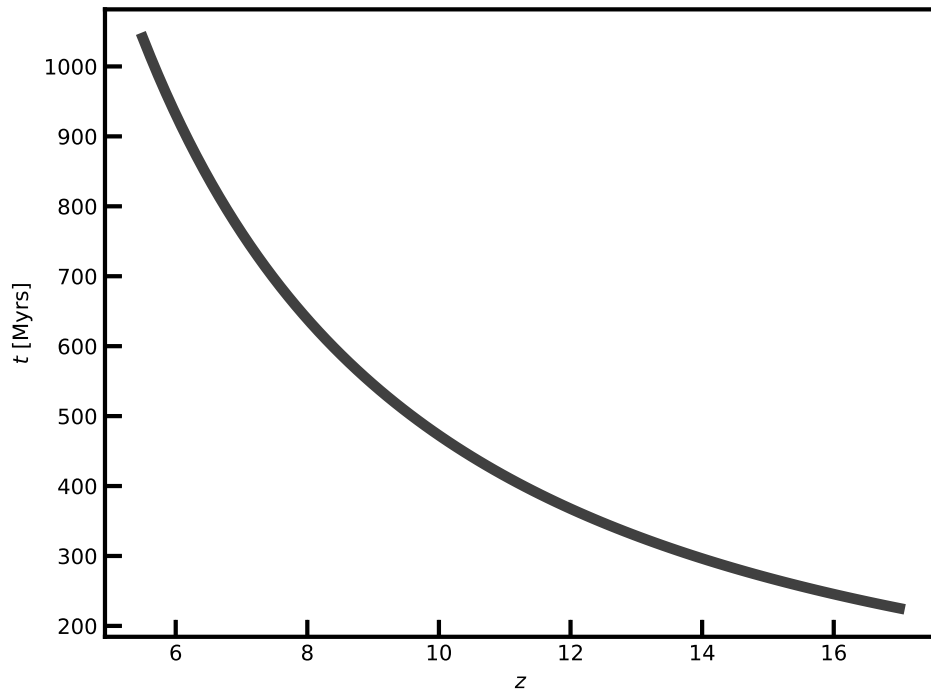
Les temps et redshifts de réionisation sont deux grandeurs différentes même si elles représentent la même information. Elles sont reliées par une relation non linéaire impliquant le facteur d'échelle  $a$  et l'expansion de l'Univers :

$$z(t) = \frac{1}{a(t)} - 1 \propto t^{-\frac{2}{3}}, \quad (2.4)$$

Pendant l'ère où la Réionisation a lieu, le facteur d'échelle  $a$  est proportionnel à  $t^{\frac{2}{3}}$ . Le temps et le redshift ont donc des monotonies opposées et le temps croît rapidement quand le redshift décroît (voir la Figure 2.16). Cette relation non-linéaire implique que la distribution spatiale des temps et redshifts de réionisation est différente et on voit par exemple dans les cartes



**FIGURE 2.15** – Cartes 2D de temps de réionisation (à gauche) et de redshifts de réionisation (à droite) extraits d’une simulation cosmologique 3D faite avec **EMMA**. Les temps et les redshifts ont une moyenne nulle et sont normalisés par leur écart-type. Les régions qui réionisent de manière précoce sont sondées en bleu et celles qui réionisent tardivement sont sondées en rouge.



**FIGURE 2.16** – Relation entre le temps et le redshift pendant l’Époque de la Réionisation. Le temps croît plus rapidement que le redshift décroît.

de la Figure 2.15 que  $t_{\text{reion}}(\vec{r})$  a des maxima plus intenses que  $z_{\text{reion}}(\vec{r})$  (voir la grande zone rouge en bas à droite des cartes). On verra notamment dans le Chapitre 6 que ces différences impactent les propriétés de ces champs, rendant le champ de temps de réionisation plus Gaussien que celui de redshift de réionisation.

## 2.4 Défis à venir et objectifs de la thèse

Comme le mentionne Choudhury (2022) dans sa revue sur l'Époque de la Réionisation, c'est une époque de l'Univers que l'on ne peut pas encore observer directement : on est à l'heure actuelle capable de sonder l'Univers jeune (avec le CMB à  $z \sim 1000$ ) et l'Univers post-Réionisation ( $z < 5$ ). Avec des observations d'objets lointains, comme les quasars (avec le signal Ly $\alpha$ ) et les galaxies (notamment grâce au JWST), on commence à repousser la limite de ce que l'on peut observer et comprendre de la Réionisation. On commence aussi tout juste à l'observer directement avec le spectre de puissance du signal 21 cm. Le défi des télescopes (notamment SKA) et des grands relevés à venir est de nous permettre de sonder la Réionisation de manière directe et plus précise.

Nous avons particulièrement beaucoup d'attentes à propos de la tomographie du signal 21 cm. Nous nous attendons à avoir énormément de données d'observation, qui seront stockées sur des serveurs dédiés et que nous allons devoir apprendre à manipuler. Nous savons aussi déjà que les observations de ce signal vont contenir du « bruit » que nous ne savons pas encore retirer. C'est une des raisons pour lesquelles SKAO organise des défis et demandent aux participants de développer des algorithmes retirant par exemple les avant-plans aux données (massives, de 1 To) qu'ils génèrent pour l'exercice.

Il est d'ailleurs nécessaire de créer des simulations et jeux de données sur lesquels nous pouvons travailler pour nous préparer à ces futures observations. Beaucoup de modèles, comme 21cmFAST, ont déjà pour but de générer des champs comme ceux que nous observerons, nous permettant ainsi de développer à l'avance des outils d'analyse pour comprendre les observations que nous aurons potentiellement à l'avenir, et ainsi en apprendre plus sur la Réionisation.

### *Objectifs de cette thèse...*

Ayant tout cela en tête, l'objectif de ma thèse est de fournir des outils pour comprendre la topologie du processus de réionisation avec les futures observations tomographiques du signal 21 cm. On sait que les bulles ionisées apparaissent autour des premières galaxies, puis grandissent, fusionnent et percolent, jusqu'à ce que tout le gaz soit réionisé à la fin de la Réionisation. Cette évolution peut être décrite par des outils mathématiques de topologie et c'est ce dont je discuterai dans le Chapitre 3. Ces outils sont en particulier très intéressants pour comparer différents scénarios de Réionisation et différents types de simulations (cosmologiques ou modèles semi-analytiques). Ce genre de comparaisons est ainsi évoqué plusieurs fois tout au long de ce manuscrit.

Comme je l'ai mentionné dans ce Chapitre, je m'intéresse à un champ purement simulé, à savoir le champ de temps de réionisation. Une étude très importante faite pendant ces trois années a été de développer un outil (avec Julien Hiegel, un autre doctorant de Dominique Aubert) pour reconstruire ce champ à partir des observations tomographiques du signal 21

cm. Je décris cet outil et les reconstructions que nous pouvons faire avec dans le Chapitre 4.

Ensuite, la raison pour laquelle le champ de temps de réionisation est l'objet de ma thèse est qu'il contient beaucoup d'informations sur le processus de réionisation et qu'avec les bons outils pour extraire ces informations, nous sommes capables de comprendre de manière détaillée le déroulé de ce processus. Dans le Chapitre 5, j'utilise ce champ pour expliquer comment les fronts d'ionisation se propage dans l'Univers par rapport aux filaments de matière, ainsi que le phénomène de percolation. J'étudie également le nombre, la forme, et taille de ce que j'appelle les « patchs de réionisation », qui contiennent l'hydrogène ionisé par la source de première réionisation originaire du patch.

Dans le Chapitre 6, je propose un formalisme permettant de sonder en grands détails le processus de réionisation au travers de  $t_{\text{reion}}(\vec{r})$ . Grâce à celui-ci, je suis capable d'étudier les sources de réionisation, les fronts d'ionisation, les bulles qui grossissent à partir des sources, la percolation, et les dernières régions neutres. J'utilise aussi la théorie des champs Gaussiens aléatoires pour comparer ces mesures à ce que l'on aurait dans le cas d'un champ Gaussien.

Le but de ces études est ainsi de me former à travailler sur l'Époque de la Réionisation au travers de simulations, tout en préparant des outils pour analyser les futures observations du signal 21 cm. En espérant que l'on obtienne des images tomographiques de ce signal, nous espérons pouvoir reconstruire le champ de temps de réionisation à partir de ces observations pour améliorer notre compréhension du processus de réionisation.



# CHAPITRE 3

## TOPOLOGIE DU PROCESSUS DE RÉIONISATION

*Dans ce Chapitre, je présente des outils pour étudier la topologie du processus de réionisation. On s'intéresse à la topologie de la Réionisation (étudiée dans des champs comme des cartes de signal 21 cm, des cartes d'ionisation ou bien des cartes de temps de réionisation) parce que cela nous permet de savoir comment elle se déroule spatialement et temporellement. Cela nous permet aussi d'accéder à des informations qui ne se trouveraient par exemple pas dans le spectre de puissance des champs en question. Dans un premier temps, je mentionne quelques études extraites de la littérature et qui portent sur la géométrie des bulles ionisées et des régions neutres, le décompte de structures 3D comme les pics, tunnels et vides, la percolation, ou encore les différentes phases de la Réionisation. Dans un second temps, je présente la théorie de Morse et l'homologie persistante, qui sont implémentées dans le code de topologie *DisPerSE* et souvent utilisées pour extraire des objets topologiques dans des champs d'intérêts. Enfin, j'explique la théorie des champs Gaussiens aléatoires et la manière dont elle peut nous être utile pour décrire la topologie de la Réionisation.*

---

## Sommaire

---

<b>3.1 Géométrie des régions ionisées et neutres, percolation et phases de la Réionisation . . . . .</b>	<b>50</b>
3.1.1 Les fonctionnelles de Minkowski . . . . .	50
3.1.2 Distribution de la taille des bulles neutres et ionisées . . . . .	55
3.1.3 Analyse des structures spatiales avec les spectres du signal 21 cm . . . . .	56
3.1.4 Décompte de structures 3D avec les nombres de Betti . . . . .	59
<b>3.2 Topologie avec la théorie de Morse et l'homologie persistante . . . . .</b>	<b>62</b>
3.2.1 La théorie de Morse . . . . .	62
3.2.2 Le code de topologie DisPerSE . . . . .	65
3.2.3 Le paramètre de la persistance . . . . .	67
<b>3.3 Topologie à travers les champs gaussiens aléatoires . . . . .</b>	<b>67</b>
3.3.1 Champs aléatoires gaussiens et isotropes . . . . .	69
3.3.2 Longueurs typiques et paramètre spectral pour un spectre de puissance donné . . . . .	73
3.3.3 Fonction de distribution de probabilité du champ . . . . .	74
3.3.4 Fonction de distribution de probabilité de la norme du gradient du champ . . . . .	74
3.3.5 Longueur d'isocontours . . . . .	74
3.3.6 Fonction de distribution de probabilité des points critiques du champ . . . . .	75
3.3.7 Longueur de squelette . . . . .	76

---

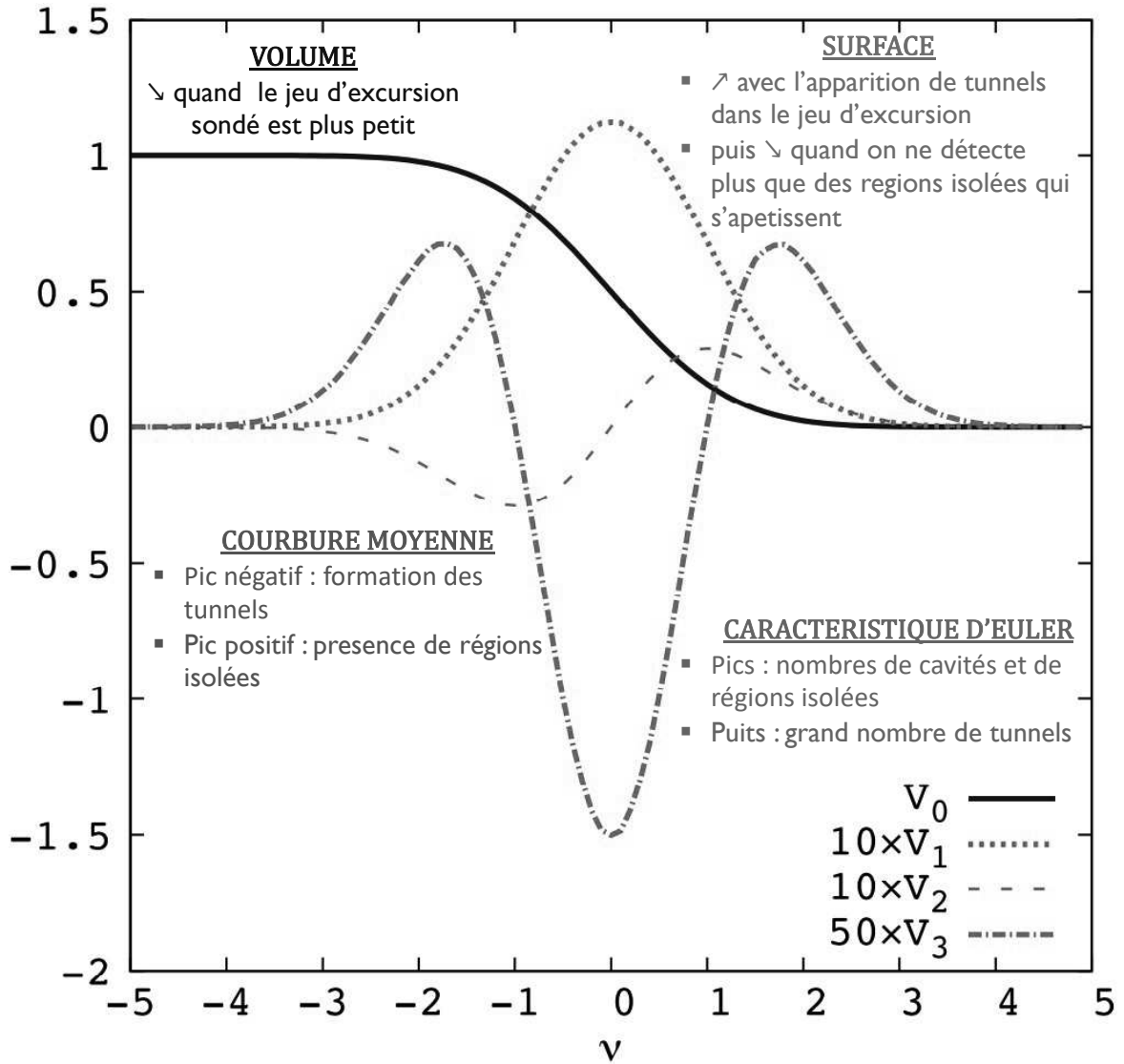
### 3.1 Géométrie des régions ionisées et neutres, percolation et phases de la Réionisation

Les bulles ionisées résultent de l’ionisation du gaz (de l’hydrogène notamment) par le rayonnement UV des sources de première réionisation autour de celles-ci, et sont donc directement corrélées avec les sources qui les ont créées, leur géométrie dépendant des propriétés des sources. Furlanetto et al. (2004a) étudient différents scénarios de réionisation (tels que des modèles inside-out et outside-in ou avec une double réionisation) en modifiant les propriétés des sources de réionisation et ils montrent que cela impactent la morphologie des bulles. McQuinn et al. (2007) montrent aussi que les bulles ionisées dépendent des propriétés de ces sources et en particulier que si peu de sources sont présentes alors les régions ionisées vont être plutôt grandes et sphériques. Friedrich et al. (2011) démontrent également que la suppression des sources d’ionisation dans les régions ionisées ralentit la croissance de ces régions et a un impact sur la distribution de leurs tailles. Comprendre les bulles ionisées et leur morphologie nous permet donc de contraindre les sources de réionisation.

#### 3.1.1 Les fonctionnelles de Minkowski

La morphologie de la réionisation peut tout d’abord être explorée par le biais de ce que l’on appelle les fonctionnelles de Minkowski (Minkowski functionals, MFs, Minkowski (1903)). Ces fonctions sont des outils mathématiques utiles pour décrire la topologie d’un ensemble d’excursions d’un champ (ensemble des points qui ont une valeur supérieure à un certain seuil). Elles ont l’avantage d’être invariantes par translations et rotations et d’avoir des expressions connues si l’on s’intéresse à des champs Gaussiens aléatoires (Gaussian random fields, GRFs), ce qui peut s’avérer utile si l’on souhaite comparer des mesures au cas Gaussien Schmalzing & Buchert (1997). En 3D, il en existe quatre indépendantes. La première,  $V_0$ , décrit le volume des régions dans lesquelles le champ dépassent un certain seuil (c’est en fait un taux de remplissage); la seconde,  $V_1$ , décrit l’aire de la surface de leur frontière; la troisième,  $V_2$ , décrit la courbure moyenne sur cette surface; et la dernière,  $V_3$ , décrit la courbure Gaussienne de ces régions. Un objet oblate (par exemple une sphère écrasée) a alors une large surface  $V_1$  et une faible courbure moyenne  $V_2$ . En prenant l’exemple d’un champ Gaussien, on obtient les courbes des MFs montrées dans la Figure 3.1 qui est extraite de Yoshiura et al. (2017) et dans laquelle j’ai détaillé leur comportement. La dernière MF  $V_3$  est également connue sous le nom de la caractéristique d’Euler  $\chi$ , qui est reliée au genre  $g$  par l’expression  $g = 1 - \chi/2$ . Celui-ci peut être vu comme le nombre de coupures d’un bord à l’autre que l’on peut faire dans un objet sans que celui-ci ne se retrouve coupé en deux (Gott et al. 1987), et il décrit la connectivité de l’objet ( $g = \#_{\text{trous}} - \#_{\text{structures isolées}}$ ). On peut par exemple couper un tore une fois seulement sans qu’il ne se retrouve en deux morceaux (il prend alors la forme d’un « C » en 3D). Son genre est donc  $g = 1$ , et sa caractéristique d’Euler est  $\chi = 0$ .

Dans le cadre d’études de la Réionisation, plusieurs travaux ont mesuré le genre ou la caractéristique d’Euler dans divers champs. Lee et al. (2008) ont par exemple quantifié la topologie de la distribution d’hydrogène neutre grâce au genre et ils ont montré que la Réionisation évolue de manière « inside-out » et en quatre phases (pré-réionisation, pré-

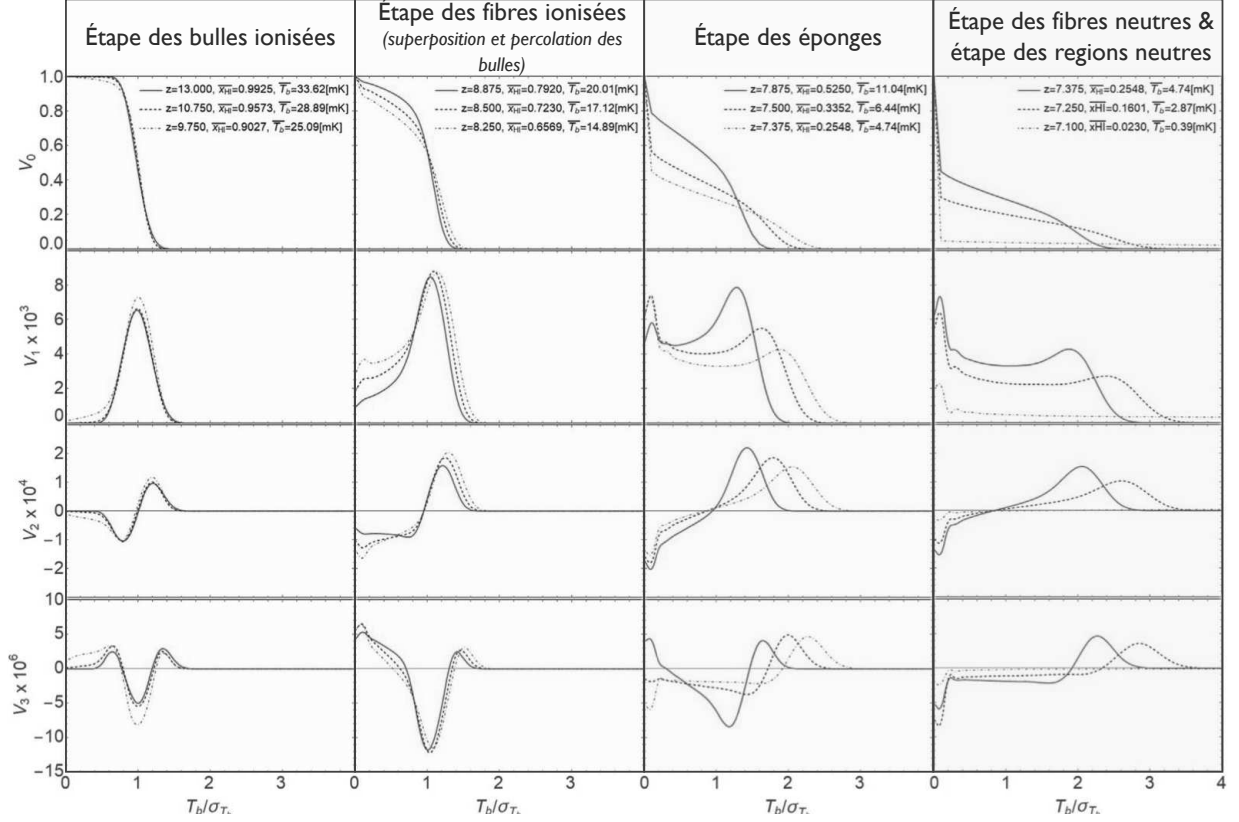


**FIGURE 3.1** – Fonctionnelles de Minkowski mesurées dans un champ Gaussien dont la valeur est donnée par  $p(\vec{x})$  avec  $\vec{x}$  les coordonnées 3D d'un point de ce champ.  $\nu$  représente ici un seuil définissant un ensemble d'excursion. Celui-ci contient tous les points de valeurs  $p(\vec{x})$  tels que  $p(\vec{x}) > p_{\text{thresh}}$  avec  $p_{\text{thresh}} = \nu\sigma$  où  $\sigma$  est l'écart-type du champ. Donc, quand  $\nu$  augmente, les valeurs des points présents dans l'ensemble d'excursion sont de plus en plus grandes et l'ensemble d'excursion contient de moins en moins de points. Cette Figure est extraite de Yoshiura et al. (2017) et a ensuite été annotée.

superposition des bulles ionisées, superposition et post-superposition). Friedrich et al. (2011) mesurent l'évolution de la caractéristique d'Euler pour étudier la topologie des bulles ionisées. Ils montrent qu'au tout début de la Réionisation, la topologie est dominée par un large nombre de bulles ionisées isolées ( $\chi > 0$ ), puis quand  $x_{\text{HII}} > 0.2$ , la topologie se complexifie avec un réseau de bulles ionisées connectées et formant des tunnels ( $\chi < 0$ ). Ils soulignent également que, pendant la première moitié de la Réionisation, leur topologie est cohérente avec un scénario de réionisation « inside-out » d'un champ Gaussien de densité. Hong et al. (2014) s'intéressent aussi au genre, mais cette fois-ci dans des cartes 2D du signal 21 cm, qu'ils comparent à celui du champ de densité sous-jacent. Ils montrent que l'on peut de nouveau suivre l'évolution de la Réionisation dans ce type de cartes, et que l'on peut aussi contraindre divers scénarios et propriétés des sources de réionisation avec le genre. Ils démontrent par contre que leurs mesures dévient rapidement du cas Gaussien.

D'autres études s'intéressent quant à elles à toutes les MFs, et Gleser et al. (2006) sont les premiers à les mesurer pour étudier le processus de réionisation et montrent ainsi qu'elles peuvent être utilisées pour décrire la topologie de la Réionisation. Selon eux, les MFs de surfaces d'iso-densité sont pratiques pour comprendre la morphologie des distributions de gaz et d'hydrogène neutre tout au long de la Réionisation et l'évolution en plusieurs phase de ce processus, ainsi que contraindre des modèles. Ils démontrent que les propriétés statistiques de l'hydrogène neutre sont impactées par la distribution d'hydrogène ionisé et la propagation des fronts d'ionisation dans le milieu intergalactique (IGM) pendant la Réionisation. Celle-ci cause également le départ à la Gaussianité de la distribution d'hydrogène neutre. Yoshiura et al. (2017) mesurent les MFs dans des cartes 3D de signal 21 cm et appuient sur le fait que cet outil est pratique pour extraire des informations non-Gaussiennes de ce champ, ce qui n'est par exemple pas possible avec le spectre de puissance. Ils montrent que la physique de ce champ et donc ses structures sont surtout reliées à la température de spin aux temps précoces de la Réionisation et à la fraction d'ionisation aux temps tardifs. Chen et al. (2019) extraient aussi les MFs du champ de signal 21 cm pour quantifier la topologie des bulles ionisées et régions neutres pendant la Réionisation. Je montre dans la Figure 3.2 leurs mesures des quatre MFs (dans les différentes lignes) à diverses fractions d'ionisation (voir les colonnes et différentes courbes) en fonction du signal 21 cm  $T_b$  normalisé par son écart-type. Quand  $T_b = 0$ , on sonde alors les régions ionisées et les contours à  $T_b = 0$  suivent les fronts d'ionisation. Chaque colonne nous permet de suivre des étapes différentes de la Réionisation, et je résume ici rapidement leur topologie respective, que Chen et al. (2019) décrivent de manière plus approfondie. Pour illustrer ce résumé, je montre dans la Figure 3.3 une carte binaire d'ionisation (hydrogène neutre en blanc et ionisé en noir) pour chaque étape du processus comme décrit par Chen et al. (2019). Il y a tout d'abord *l'étape des bulles ionisées* (première colonne) où les MFs sont quasiment toutes identiques au cas Gaussien et pendant laquelle les premières bulles ionisées apparaissent (augmentation de la surface des contours  $T_b \sim 0$  dans les MFs  $V_1$  et  $V_3$ ). Ensuite, de  $x_{\text{HII}} \sim 0.9$  à  $x_{\text{HII}} \sim 0.7$  (deuxième colonne), les bulles se connectent pour former des *fibres ionisées* (contours  $T_b = 0$  de plus en plus grands avec  $V_1$  et  $V_3$ ). À cette étape, la topologie des régions ionisées se complexifie, comme on peut le voir avec des pics qui apparaissent à bas  $T_b$  dans les MFs  $V_1$ ,  $V_2$  et  $V_3$ . Cette étape marque ainsi la superposition des bulles ionisées et leur percolation. Quand  $x_{\text{HII}} < 0.6$  (troisième colonne), on entre dans *l'étape des éponges* avec des fibres ionisées entremêlées aux fibres neutres. La caractéristique d'Euler  $V_3$  reflètent ceci car elle est positive pour  $T_b = 0$  à  $x_{\text{HII}} = 0.5$  montrant

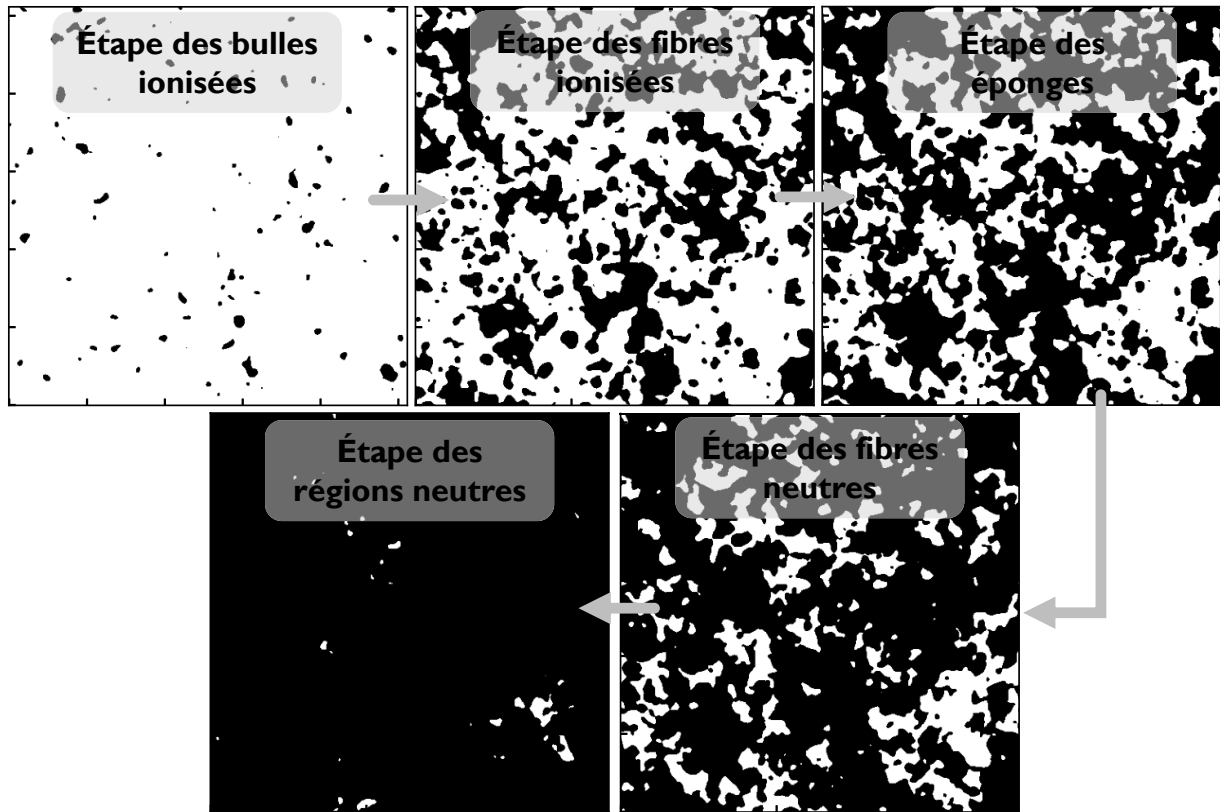
### 3.1. GÉOMÉTRIE DES RÉGIONS IONISÉES ET NEUTRES, PERCOLATION ET PHASES DE LA RÉIONISATION



**FIGURE 3.2** – Fonctionnelles de Minkowski mesurées dans des cartes 3D de signal 21 cm à différents redshifts (trois redshifts par panneau et qui augmentent de gauche à droite). Les quatre MFs sont montrées sur les différentes lignes ( $V_0$  à  $V_4$  de haut en bas). Les cartes dont elles sont extraites proviennent de simulations semi-analytiques faites avec 21cmFAST. Cette Figure est extraite de Chen et al. (2019) et a ensuite été annotée.

encore des bulles ionisées entourées de régions neutres, puis devient négative à  $T_b = 0$  pour  $0.3 < x_{\text{HI}} < 0.5$  montrant qu’autant de régions neutres que de bulles ionisées sont présentes. Enfin, quand  $x_{\text{HI}} < 0.3$  (deux premières courbes de la dernière colonne), on entre dans *l’étape des fibres neutres* (équivalente à l’étape des fibres ionisées), puis quand  $x_{\text{HI}} < 0.16$  (dernière courbe de la dernière colonne), c’est *l’étape des régions neutres* (équivalente à l’étape des bulles ionisées). Celles-ci sont marquées par  $V_0$  qui diminue fortement jusqu’à être presque nul à la fin de la Réionisation. Comme on peut le voir ici les MFs du signal 21 cm comportent beaucoup d’informations sur le processus de réionisation, mais ne sont pas triviales à interpréter.

Enfin, la forme et la morphologie évolutive des bulles ionisées peuvent aussi être étudiées dans différents modèles de réionisation par les shapefinders (SFs), comme en discutent Pathak et al. (2022). Les SFs sont en fait des rapports des MFs et décrivent l’épaisseur, la largeur et la longueur d’objets 3D. Dans cet article, ils expliquent notamment que la SF mesurant la longueur des structures permet d’étudier la percolation, tandis que les deux autres SFs ne varient pas. Ils montrent que pendant cette phase de percolation, les régions ionisées sont très filamenteuses et dépendent du modèle de sources utilisé, signifiant ainsi que l’on peut



**FIGURE 3.3** – Schéma représentant l'évolution d'un champ d'ionisation binaire pendant la Réionisation. L'hydrogène neutre est en blanc et l'hydrogène ionisé est en noir. Les cinq panneaux correspondent chacun à une étape de la Réionisation comme décrite par Chen et al. (2019). Au début de la Réionisation, on voit quelques bulles ionisées apparaître pendant *l'étape des bulles ionisées*. Ensuite, elles fusionnent pendant *l'étape des fibres ionisées*. À mesure que les fibres ionisées s'élargissent, elles s'entremêlent avec les fibres neutres. C'est *l'étape des éponges*. Ensuite, le gaz ionisé remplit de plus en plus l'Univers et le gaz neutre est sous forme de fibres neutres pendant *l'étape des fibres neutres*. Le gaz s'ionise de plus en plus et seules quelques régions neutres persistent pendant *l'étape des régions neutres*. À la fin de la Réionisation, l'entièreté du gaz d'hydrogène est réionisé.

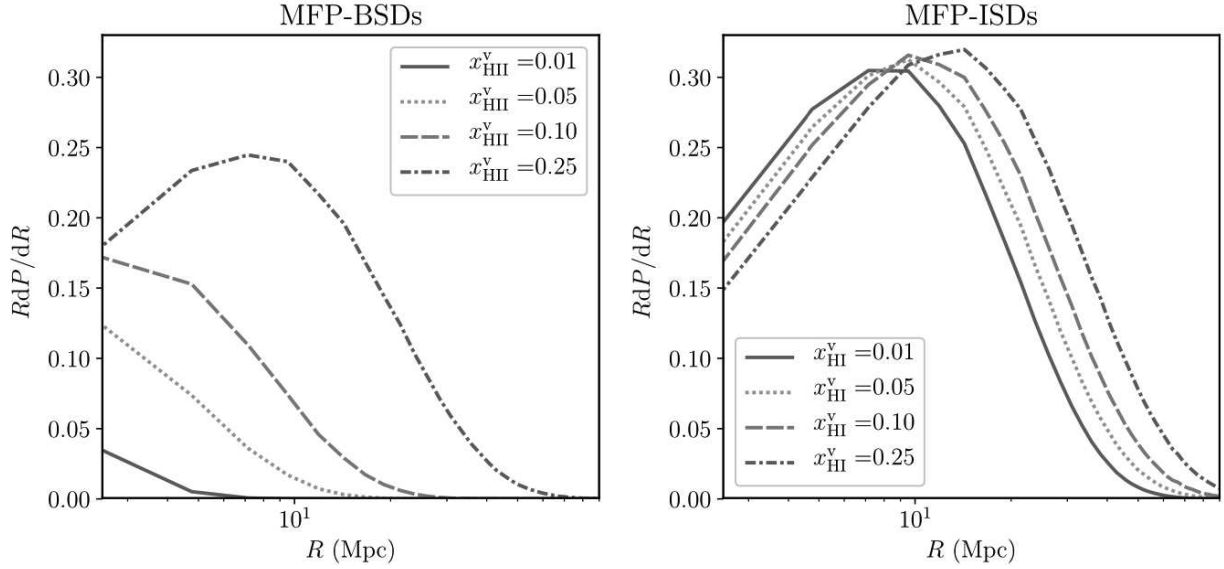
étudier les propriétés des sources de réionisation avec ce formalisme.

### 3.1.2 Distribution de la taille des bulles neutres et ionisées

On peut également étudier la physique du processus de réionisation avec un outil utilisé couramment dans ce domaine, qui est la distribution des tailles de bulles ionisées ou de régions neutres. On peut les mesurer dans divers champs, comme dans des cartes tomographiques du signal 21 cm, des cartes binaires de gaz ionisé ou des cartes de fraction d’ionisation. Il n’existe également pas une unique façon de définir la taille d’une bulle ionisée ou d’une région neutre et par conséquent, il n’existe pas une seule manière de les mesurer. Je mentionne donc ici de manière non-exhaustive plusieurs méthodes que l’on peut trouver dans la littérature. Les travaux de Furlanetto et al. (2004b) font parties des premiers à s’intéresser aux distributions de tailles de bulles ionisées. Ils construisent un modèle analytique simple de la distribution des tailles de bulles dépendant de l’efficacité d’ionisation des objets effondrés et de la masse minimale des halos contenant des objets lumineux. Ils estiment d’ailleurs que des statistiques au-delà du spectre de puissance sont plus puissantes pour sonder la Réionisation. Plus tard, plusieurs méthodes pour mesurer les distributions de tailles de bulles ionisées sont introduites. Zahn et al. (2007) introduit la méthode de la « moyenne sphérique » (spherical average, SPA), pour laquelle on construit des sphères autour de chaque cellule et on moyenne la fraction d’ionisation à l’intérieur dans le but de trouver la plus grande sphère telle que cette moyenne est supérieure à un certain seuil. Cette méthode est reprise par Friedrich et al. (2011), Lin et al. (2016) et Giri et al. (2018a). Ces derniers proposent également une méthode basée sur un algorithme Friend-Of-Friend (FOF), qui a été introduite par Iliev et al. (2006b). Friedrich et al. (2011) mentionnent aussi la méthode  $3V/A$ , qui consiste à mesurer le rapport entre le volume total des régions ionisées et leur surface totale (qui proviennent respectivement des MFs  $V_0$  et  $V_1$ ). Mesinger & Furlanetto (2007) construisent une méthode (mean free path, MFP) basée sur le ray-tracing pour mesurer la distribution des tailles d’objets d’intérêts, qui est par exemple reprise par Lin et al. (2016), Giri et al. (2018a) et Bianco et al. (2021). Giri et al. (2019) l’utilisent aussi dans le cas des bulles ionisées et régions neutres. Ces derniers sont ainsi les premiers à introduire les distributions de tailles des régions neutres, et ils soulignent en particulier que ces régions sont larges et que comme à bas redshift le rapport signal à bruit sera plus élevé, elles seront plus facilement détectables dans les observations tomographiques du 21 cm que les régions ionisées. Plus récemment, Kakiichi et al. (2017) ont créé une méthode appelée « granulométrie » pour trouver la taille moyenne et la distribution de tailles de bulles ionisées. Celle-ci se base sur le concept du tamisage défini par la morphologie mathématique et l’analyse d’image.

Avec toutes ces définitions différentes et indépendantes de la distribution de tailles de bulles ionisées ou régions neutres, on a en fait autant de points de vue différents sur les tailles de régions. Par exemple, les méthodes FOF déterminent plutôt les tailles de bulles ionisées aux petites échelles, et donc s’intéressent à celles qui ne contribuent que très peu à la fraction totale de volume ionisé, alors que les méthodes SPA sont plutôt des mesures concernant les bulles ionisées moyennes et donc celles qui contribuent le plus. Je présente des distributions de tailles de bulles ionisées et régions neutres mesurée avec la méthode MFP, et qui sont extraites de Giri et al. (2019). Cette méthode donne des résultats comparables à la méthode SPA, mais avec de plus grandes bulles selon Giri et al. (2018a). À gauche de la Figure 3.4, on



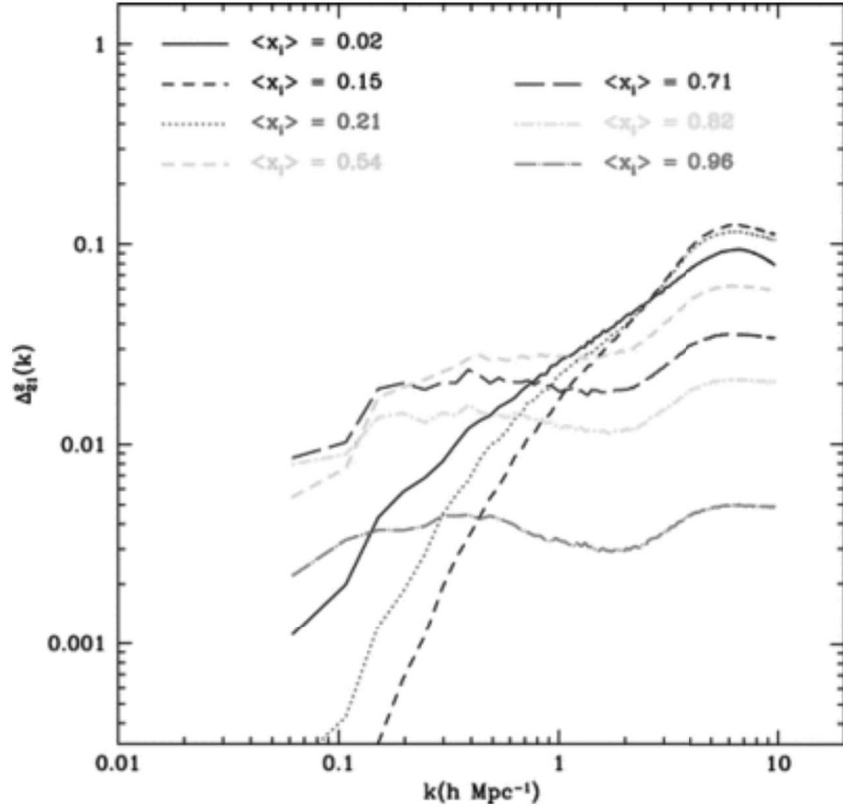


**FIGURE 3.4** – Distributions de tailles de bulles ionisées à gauche et de régions neutres à droite. Les distributions de tailles de bulles ionisées sont montrées pour des petites fractions d’hydrogène ionisé et donc à des temps précoces, et les distributions de tailles de régions neutres sont montrées pour de petites fractions d’hydrogène neutre et donc à des temps tardifs. Ces distributions sont mesurées avec la méthode MFP dans des cartes de signal 21 cm provenant de simulations cosmologiques. Cette Figure est extraite de Giri et al. (2019). On voit notamment que les BSDs croissent plus rapidement avec la fraction d’hydrogène ionisé que les ISDs avec la fraction d’hydrogène neutre.

voit les distributions de tailles de bulles ionisées (bubble size distribution, BSDs) en début de réionisation représentées à différentes fractions d’ionisation. À ces mêmes fractions de gaz neutre, on voit à droite les distributions de tailles de régions neutre (islands size distributions, ISDs). On peut déjà noter que quand la fraction d’ionisation augmente, les bulles ionisées grandissent, et de même quand la fraction de neutre diminue, les régions neutres rapetissent. Ces phénomènes sont évidemment attendus lors d’un processus de réionisation. De plus, aux mêmes fractions (d’ionisation pour les bulles ionisées et de neutre pour les régions neutres) les BSDs ont de plus petites tailles que les ISDs équivalentes. Les ISDs n’évoluent aussi que très peu avec la fraction de neutre, tandis que les BSDs évoluent bien plus rapidement.

### 3.1.3 Analyse des structures spatiales avec les spectres du signal 21 cm

Comme expliqué dans le Chapitre 2, nous disposerons peut-être avant la fin de la décennie d’observations du signal 21 cm. Des radio-télescopes comme HERA (Hydrogen Epoch of Reionization Array), LOFAR (Low Frequency Array), ou MWA (Murchison Widefield Array) prévoient de détecter le spectre de puissance du signal 21 cm avant que nous ayons potentiellement accès aux images 2D tomographiques du signal 21 cm faites par SKA (Square Kilometer Array). Il est donc utile de comprendre au préalable le spectre de puissance du



**FIGURE 3.5** – Spectres de puissance du signal 21 cm extraits d’une simulation cosmologique. Les différentes courbes montrent l’évolution selon la fraction moyenne de volume d’hydrogène ionisé  $\langle x_{\text{HI}} \rangle$  (ou selon le redshift de manière équivalente). Cette Figure est extraite de Lidz et al. (2008).

signal 21 cm et de savoir quelles informations il est possible d’en extraire. Plusieurs études, dont Zaldarriaga et al. (2004), Furlanetto et al. (2004a) et Bowman et al. (2006), ont conclu que le spectre de puissance du signal 21 cm contient des empreintes de l’évolution du processus de réionisation, mais aussi des bulles ionisées et de leurs tailles. Lidz et al. (2008) ont aussi montré que l’évolution temporelle de l’amplitude et la pente de ce spectre de puissance peuvent nous donner des informations sur la fraction de volume ionisé et les bulles ionisées. Je montre les spectre de puissance du signal 21 cm qu’ils extraient à partir d’une simulation cosmologique dans la Figure 3.5. On voit notamment un pic en amplitude apparaître aux grandes échelles quand 50% de l’IGM est réionisé. La pente du spectre de puissance s’aplatit ensuite quand la fraction d’hydrogène ionisé augmente et quand la taille des bulles ionisées augmentent.

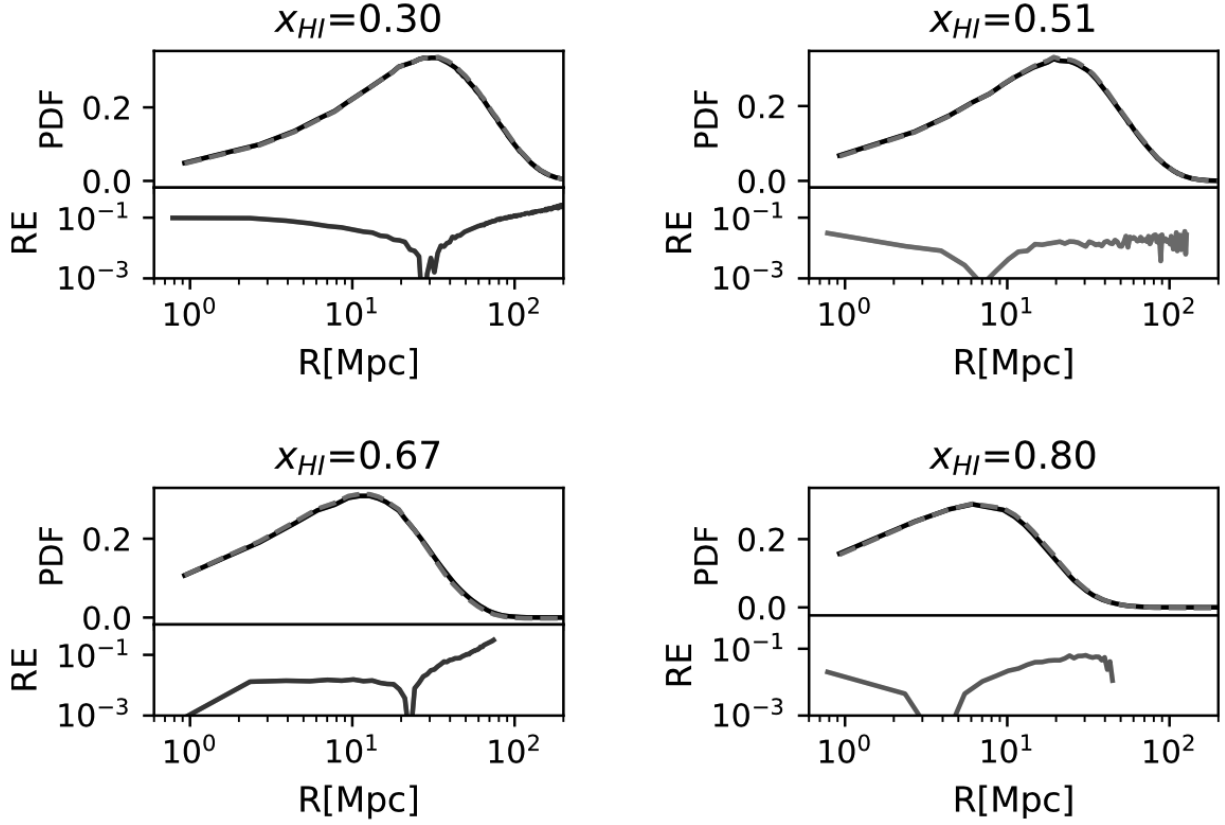
Ensuite, beaucoup d’études cherchent à contraindre les paramètres astrophysiques de la Réionisation à l’aide du spectre de puissance (Iliev et al. 2012; Mesinger et al. 2013; Pober et al. 2014; Greig & Mesinger 2015, 2017; Liu & Parsons 2016; Kern et al. 2017). Iliev et al. (2012) s’intéressent par exemple à la masse des sources de réionisation et montrent que le spectre de puissance du signal 21 cm permet de distinguer des cas où la Réionisation est menée par les halos de grandes masses ( $M > 10^9 M_{\odot}$ ) uniquement ou par ces halos et

ceux de plus faibles masses ( $10^8 M_\odot < M < 10^9 M_\odot$ ) conjointement. Mesinger et al. (2013) étudient l’impact des rayons X sur le spectre de puissance du signal 21 cm et en déduisent que dans des scénarios non-extrêmes leur impact est faible quand la Réionisation est déjà enclenchée ( $x_{\text{HII}} > 0.3$  environ). Cependant, ils démontrent que l’impact des rayons X dépend du moment et de la durée du réchauffement du gaz par les rayons X après l’apparition des premières galaxies. En effet, si ce réchauffement par rayons X finit de se produire après que la réionisation débute, alors le spectre de puissance du signal 21 cm aura une bien plus grande amplitude (d’un facteur 10 à 100 pour  $x_{\text{HII}} < 0.1$ ). Greig & Mesinger (2015) ont créé l’outil 21CMMC qui effectue des MCMC (Monte Carlo Markov Chain) et qui intègre le code de simulation semi-analytiques 21cmFAST dans le but de permettre à l’utilisateur de contraindre des paramètres astrophysiques (comme l’efficacité d’ionisation des galaxies  $\zeta$  et la température de viriel minimale pour qu’un puisse former des étoiles  $T_{\text{vir}}$ ) à partir d’observations du signal 21 cm. Dans Greig & Mesinger (2017), ils s’en servent par exemple pour estimer des paramètres astrophysiques à partir du spectre de puissance du signal 21 cm, ce qui leur permet de contraindre l’histoire de réionisation au pourcent près.

La morphologie de la Réionisation peut aussi être contrainte grâce au spectre de puissance du signal 21 cm, comme le démontrent Shimabukuro et al. (2022). En utilisant un réseau de neurones artificiels, ils reconstruisent à un redshift donné (ou à une fraction d’ionisation donnée) les distributions de tailles de bulles ionisées à partir de spectre de puissance du signal 21 cm. Ceci est illustré dans la Figure 3.6, où l’on voit des distributions de tailles de bulles ionisées (mesurée avec la méthode utilisée par 21cmFAST, à savoir la méthode SPA) provenant des simulations en ligne noir et les reconstructions à partir du réseau de neurones en pointillés rouges. Ces reconstructions sont très proches des distributions provenant de la simulation, montrant ainsi que les informations sur la taille des bulles ionisées se trouvent dans le spectre de puissance du signal 21 cm. Cependant, les auteurs précisent que ces reconstructions sont modèles-dépendantes et qu’elles ne doivent être utilisées que pour comprendre la morphologie des bulles ionisées et leur évolution. De plus, Pagano & Liu (2020) mesurent les corrélations entre le champ de densité et le champ d’ionisation et démontrent que celles-ci ont des conséquences sur le spectre de puissance du signal 21 cm. Ils s’en servent pour étudier également la morphologie de la Réionisation et exclure certains scénarios de réionisation.

Par ailleurs, Hutter et al. (2020) montrent également que la morphologie de la Réionisation (distributions de tailles de bulles ionisées et de régions neutres, morphologie à grande échelle) ont un impact sur le bispectre du signal 21 cm.<sup>1</sup> Pour également étudier les structures spatiales pendant la Réionisation avec les cartes 2D de signal 21 cm, Gorce & Pritchard (2019) proposent un nouvel outil, la fonction de corrélation triangulaire (triangle correlation function, TCF). Celle-ci permet d’extraire des informations supplémentaires par rapport au spectre de puissance car elle repose sur l’extraction des phases de la transformée de Fourier du champ d’ionisation, contrairement au spectre de puissance qui ne se concentre que sur l’amplitude. Dans cette étude, les auteurs montrent qu’ils ont accès à des limites supérieures de la taille des bulles ionisées pendant les phases précoces de la Réionisation. Ensuite, ils démontrent que l’absence de signal de la TCF implique que l’on est dans les

1. Le bispectre est la transformée de Fourier de la fonction de corrélation à trois points du champ. Cette fonction de corrélation représente l’excès de probabilité de présence dans le champ de structures spatiales à trois points. Le bispectre informe alors sur le taux de structures pouvant être décrites par un triangle fermé dans le champ.



**FIGURE 3.6** – Distributions de tailles de bulles ionisées mesurées dans le champ d’ionisation de simulations 21cmFAST par la méthode SPA en ligne noire et reconstructions de ces distributions à partir du spectre de puissance du signal 21 cm grâce à un réseau de neurones artificiel en pointillés rouges. En bas de chaque panneau, l’erreur relative (RE) des reconstructions par rapport aux vraies distribution est représentée. Chaque panneau montre un instant différent pendant le processus de réionisation et qui est dénoté par la fraction d’hydrogène neutre  $x_{\text{HI}}$ . Plus celle-ci est faible et plus l’hydrogène est ionisé et donc plus on sonde les phases tardives de la Réionisation. Cette Figure est extraite de Shimabukuro et al. (2022). Cette étude montre ainsi une très bonne capacité de reconstruction des distributions de tailles de bulles ionisées à partir de spectres de puissance du signal 21 cm, signifiant que l’information des tailles de bulles ionisées se trouve dans ces spectres de puissance.

phases intermédiaires de la Réionisation (superposition des bulles ionisées) et que l’on peut connaître la durée de ces phases. Enfin, vers la fin de la Réionisation, ils trouvent deux tailles caractéristiques des régions neutres, à l’inverse des distributions de tailles de régions ionisées qui n’en trouvent qu’une.

### 3.1.4 Décompte de structures 3D avec les nombres de Betti

Une autre manière de quantifier la topologie du processus de réionisation est de dénombrer les différentes structures d’un champ. Plusieurs études font cela grâce aux nombres de Betti (Kapahtia et al. 2019, 2021; Giri & Mellema 2021; Bianco et al. 2021; Elbers & van de

Weygaert 2023). Ces nombres sont des objets mathématiques bien définis et ont été introduit par E. Betti en 1870. De manière purement qualitative, le nombre de Betti  $\beta_k$  compte le nombre de surface à  $k$ -dimensions qui sont indépendantes. En 3D, il y a trois nombres de Betti :  $\beta_0$  compte le nombre de composantes connexes,  $\beta_1$  compte le nombre de courbes fermées indépendantes, et  $\beta_2$  compte le nombre de surfaces indépendantes. Autrement dit,  $\beta_0$  décompte le nombre d’objets isolés,  $\beta_1$  le nombre de tunnels et  $\beta_2$  compte le nombre cavités entourées par une surface. Si on imagine l’exemple 3D d’un tore, on compte un objet isolé ( $\beta_0 = 1$ ), on peut tracer deux courbes fermées indépendantes ( $\beta_1 = 2$ ) et il y a une surface entourant le tore ( $\beta_2 = 1$ ). Dans le cadre de la Réionisation et de cubes tomographiques du signal 21 cm, on peut associer les nombres de Betti aux bulles ionisées ( $\beta_0$ ), aux tunnels ( $\beta_1$ ) et aux régions neutres ( $\beta_2$ ) d’après Giri & Mellema (2021) ou Elbers & van de Weygaert (2023).

Originellement, les nombres de Betti ne sont pas explorés tels quels mais plutôt sous une forme combinée dans ce qu’on appelle le genre  $g$  ou la caractéristique d’Euler  $\chi$  ( $\chi = 2 - 2g$ ). Cette dernière est aussi l’équivalent de la troisième fonctionnelle de Minkowski et est une addition alternée de ces trois nombres de Betti :

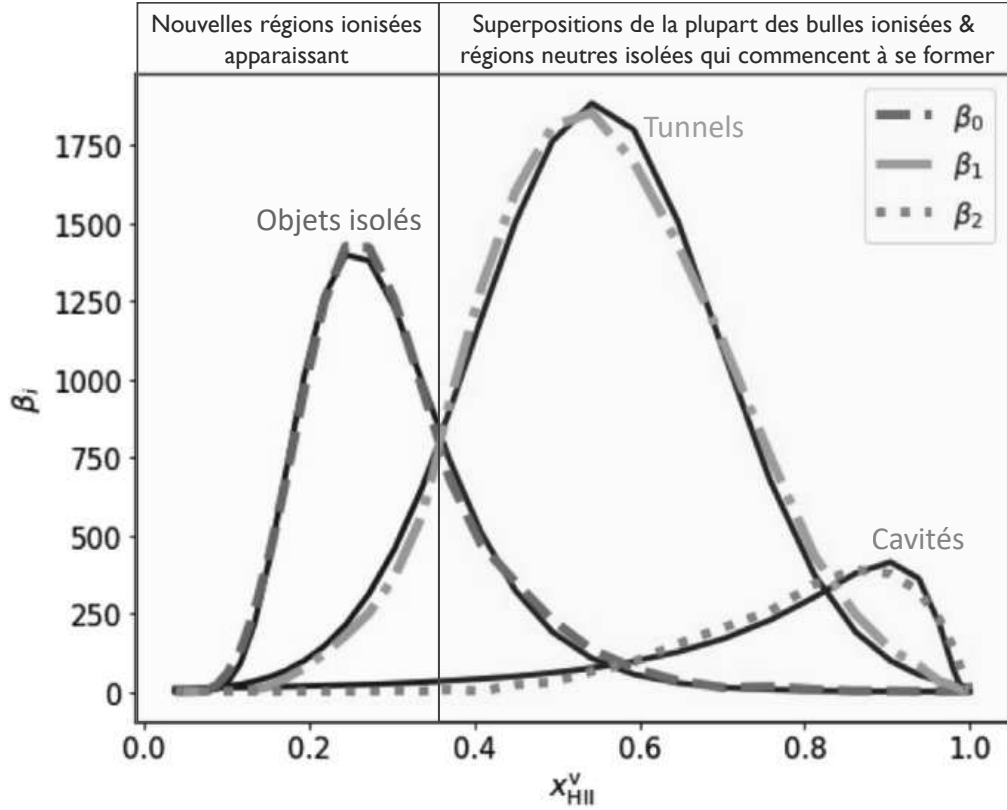
$$\begin{aligned}\chi &= \beta_0 - \beta_1 + \beta_2 \\ &= \#_{\text{structures connectées}} - \#_{\text{tunnels}} + \#_{\text{cavités}}.\end{aligned}\tag{3.1}$$

Avec cette expression, on retrouve  $\chi = 1 - 2 + 1 = 0$  pour un tore. De plus, mesurer cette caractéristique dans un champ peut aussi revenir à compter le nombre de points critiques de ce champ (Bardeen et al. 1986; Coles et al. 1996). Elle peut donc s’écrire de manière équivalente comme la somme alternée des points critiques d’un champ, comme par exemple à deux dimensions :

$$\chi = \#_{\text{maxima}} - \#_{\text{points selles}} + \#_{\text{minima}},\tag{3.2}$$

Le genre et la caractéristique d’Euler combinent l’information de trois nombres de Betti et contiennent alors nécessairement moins d’informations que ces nombres pris séparément, comme le soulignent aussi Giri & Mellema (2021). Plusieurs études ont mesuré  $g$  ou  $\chi$  dans le cadre de la Réionisation, comme Lee et al. (2008) dans la distribution d’hydrogène neutre, Friedrich et al. (2011) dans le champ de fraction d’ionisation, et Hong et al. (2014) dans les cartes 2D du signal 21 cm. Ces études ont déjà été discutées dans la Section 3.1.1 relatant des fonctionnelles de Minkowski.

Par la suite, les nombres de Betti sont plutôt utilisés pour décrire la topologie de la Réionisation. Par exemple, Kapahtia et al. (2019, 2021) et Elbers & van de Weygaert (2023) les mesurent dans des cartes 2D de signal 21 cm. Bianco et al. (2021) les mesurent dans leurs cartes 2D de régions ionisées (reconstruites à partir de cartes 2D du signal 21 cm). Giri & Mellema (2021) proposent de les extraire des cartes tomographiques 3D (deux directions sur le ciel et une direction de fréquences) du signal 21 cm similaires à celles que l’on observera avec SKA. Dans la Figure 3.7, je montre une Figure annotée que j’ai extraite de leur article et qui représente les trois nombres de Betti (en couleur) mesurés dans les cartes tomographiques du signal 21 cm. On peut voir une évolution de ces nombres en fonction de la fraction de volume ionisé marquée par un pic à des fractions d’ionisation différentes pour chacun de ces nombres. De plus, l’évolution des nombres de Betti est clairement connectée au processus de percolation. On voit d’abord le nombre d’objets isolés ( $\beta_0$ ) augmenter aux fractions d’ionisation précoces à mesure que de nouvelles régions ionisées apparaissent. Les bulles ionisées



**FIGURE 3.7** – Évolution des nombres de Betti en fonction de la fraction d’hydrogène ionisé. Le nombre de Betti 0 est montré en bleu et représente le nombre d’objets isolés. Le premier nombre de Betti est montré en orange et représente le nombre de tunnels. Le second nombre de Betti est montré en vert et représente le nombre de cavités. Les lignes noires sont des fonctions ajustés sur les mesures en couleur de ces nombres de Betti dans une carte tomographique 3D du signal 21 cm extraite d’une simulation cosmologique. Cette Figure est extraite de Giri et al. (2019) et a été ensuite annotée.

continuent à apparaître et commencent à fusionner. Mais elles apparaissent plus rapidement qu’elles ne fusionnent donc  $\beta_0$  augmente et  $\beta_1$  augmente.  $\beta_1$  compte le nombre de tunnels qui se forment quand les bulles fusionnent, et est donc un traceur de la fusion des bulles ionisées. Ensuite, les bulles ionisées fusionnent de plus en plus tandis que de moins de nouvelles bulles sont créées ( $\beta_0$  diminue et  $\beta_1$  augmente toujours). Puis, les régions neutres commencent à être décomptées quand les bulles ionisées percolent et que des cavités d’hydrogène neutre apparaissent entre les bulles ( $\beta_2$  augmente). Il y en a de plus en plus à mesure que les bulles ionisées commencent à remplir le volume, et celles ci commencent à disparaître comme tout l’hydrogène se fait réionisé ( $\beta_2$  diminue). On conclut donc que ces nombres permettent de suivre tout le processus de réionisation, et ils peuvent être ajustés à l’aide de fonctions analytiques simples à base d’exponentielles (voir les lignes noires sur la Figure 3.7).

Giri & Mellema (2021) mesurent également ces nombres dans des « fausses » observations du futur signal 21 cm produit par SKA, et ils ajoutent qu’ils contiennent de l’information sur le milieu et la fin de la Réionisation, et qu’ils pourront permettre de contraindre des

modèles de Réionisation. Elbers & van de Weygaert (2023) ont aussi montré que les nombres de Betti sont dépendants des propriétés des sources de réionisation. Ceci est attendu par le fait que ces nombres sont reflètent l'évolution des bulles ionisées (cf. les explications ci-dessus) qui sont directement impactées par les sources (Furlanetto et al. 2004b; McQuinn et al. 2007; Friedrich et al. 2011). Cette étude affirme aussi que les tunnels ( $\beta_1$ ) seront plus facilement observables parce qu'ils sont nombreux et qu'ils sont les plus sensibles aux différents modèles de Réionisation. Les auteurs soulignent de nouveau que les nombres de Betti peuvent contenir des informations additionnelles à celles contenues dans le spectre de puissance du même champ. Par ailleurs, ils expliquent qu'il y a une correspondance entre les nombres de Betti d'un réseau de bulles ionisées et ceux de la toile cosmique (amas, filaments et vides).

## 3.2 Topologie avec la théorie de Morse et l'homologie persistante

Dans cette thèse, je m'intéresse également à la topologie de la Réionisation, et dans cette Section, j'introduis un autre outil qui peut permettre de l'explorer. Cette outil est le code de topologie `DisPerSE`<sup>2</sup> (Discrete Persistent Structure Extractor; Sousbie (2011)) et il est basé sur la théorie de Morse et l'homologie persistante, que je décris plus en détail par la suite.

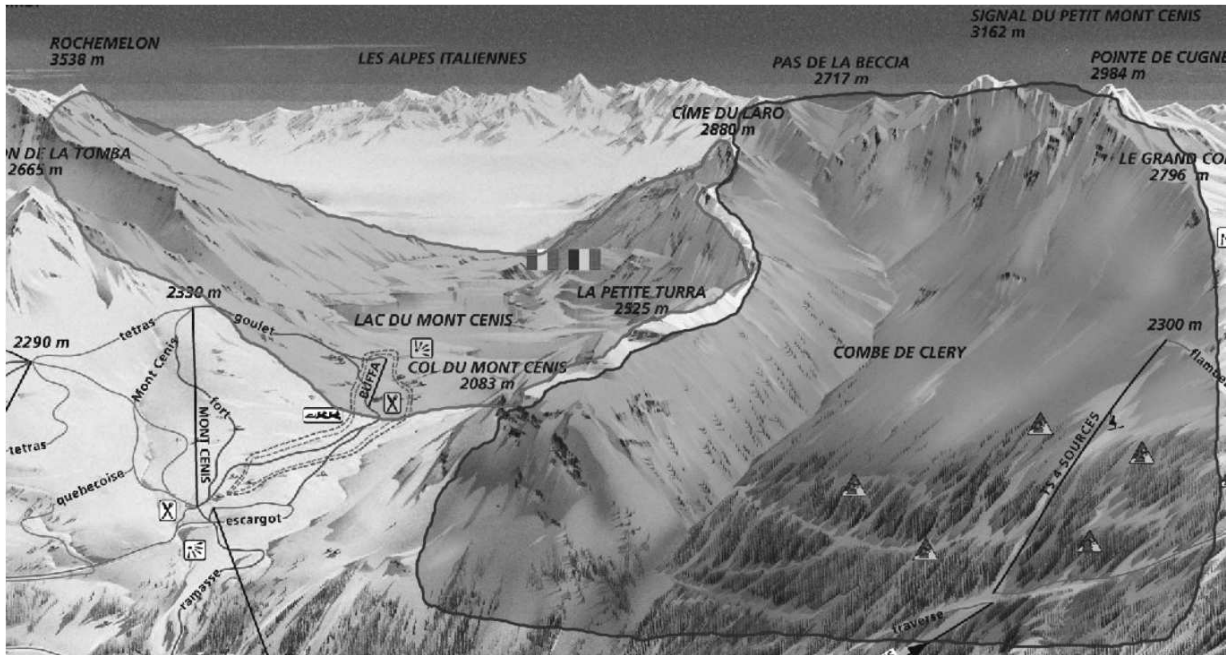
### 3.2.1 La théorie de Morse

La théorie de Morse permet d'accéder aux propriétés topologiques d'une fonction scalaire lisse définie dans un certain espace. Cette fonction peut par exemple être l'altitude des montagnes dans un paysage 2D montagneux, ou bien un champ 2D ou 3D de densité ou de temps de réionisation. Les propriétés topologiques du champ nous informent sur la manière dont les points de ce champ sont connectés et elles sont robustes et invariantes si l'on lisse le champ ou si on lui fait subir des transformations de rotation ou de translation par exemple. La théorie de Morse capture ainsi la relation entre les gradients d'un champ, la manière dont il s'«écoule» et sa topologie. Tout au long de la description de cette théorie, je vais faire une analogie de la topologie d'un champ 2D aux structures d'un paysage montagneux. Dans la Figure 3.8, je montre une illustration d'un tel paysage, et dans la Figure 3.9, je montre une carte 2D de temps de réionisation extraite d'une simulation cosmologique. Dans la suite de cette Section, je reviens sur l'analogie montagneuse et cette carte simulée pour illustrer les concepts mathématiques assez abstraits présentés ici.

La théorie que nous étudions ici nécessitent de connaître les points critiques du champ. Ces points sont localisés aux endroits de gradients nuls, et en 2D il en existe trois types (et 4 en 3D, etc.) et chacun est caractérisé par un indice critique  $k$ . L'indice critique  $k$ , ou l'ordre  $k$ , d'un point critique correspond par définition au nombre  $k$  de valeurs propres négatives dans la matrice Hessienne  $H$  du champ. Cette matrice est définie pour une fonction dérivable  $f$  et a la dimension  $n \times n$ , où  $n$  est le nombre de variables dont dépend la fonction. Chaque élément

---

2. <http://www2.iap.fr/users/sousbie/web/html/indexd41d.html>

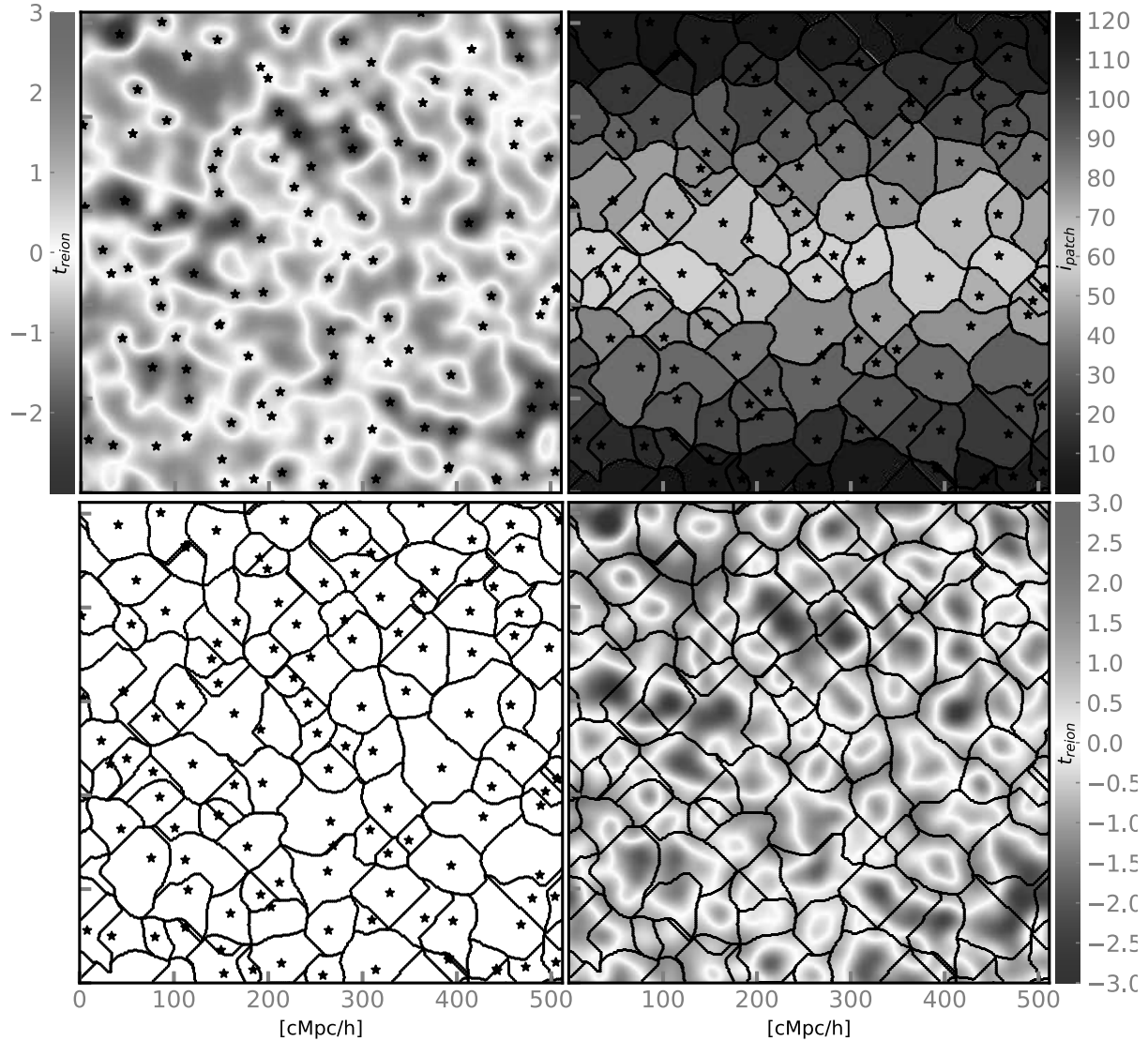


**FIGURE 3.8** – Représentation des variétés ascendantes de type 2 (régions rouge et bleu) par analogie avec les vallées d’un paysage 2D montagneux. Les pics des montagnes sont les maxima du champ, les cols sont ses points selles, et les creux des vallées sont ses minima. Les bordures des vallées, ou les lignes de crête, sont les filaments formant le squelette des montagnes. L’ensemble des vallées est une segmentation en variétés ascendantes de type 2. Cette Figure est extraite de Gay (2011).

$H_{ij}$  de la matrice contient la dérivée partielle par rapport aux variables  $i$  et  $j$  :  $H_{ij} = \frac{\partial f}{\partial i \partial j}$ . Les *minima* du champ (ou les creux des vallées dans notre analogie montagneuse), représentés par les étoiles dans la Figure 3.9, ont un indice critique de 0 et sont les points à partir desquels les valeurs du champ augmentent localement (dérivée seconde positive). Les *maxima* (ou les pics des montagnes) ont un indice critique de 2 et sont les points à partir desquels les valeurs du champ diminuent localement (dérivée seconde négative). Les *points selles* (ou les cols des montagnes) ont un indice critique de 1 et sont les points à partir desquels le champ augmente dans une direction et diminue dans l’autre (dérivée seconde positive dans une direction et négative dans l’autre). On appelle fonctions de Morse les fonctions qui n’ont aucun point critique dégénéré, c’est-à-dire aucune valeur propre de la Hessienne nulle aux points critiques.

Tous les points qui ne sont pas critiques ont un gradient qui indique une direction préférentielle de la fonction. Cette direction préférentielle suit ce qu’on appelle des lignes de champs ou lignes intégrales. Ce sont les courbes tangentes au gradient en chacun des points, et elles doivent avoir pour origine un point critique et se terminer par un point critique. Elles couvrent tout l’espace avec une unique ligne de champ passant par chaque point non-critique de l’espace. Avec ces lignes, on peut alors définir une segmentation de l’espace en régions que l’on appelle des variétés ascendantes (respectivement descendantes) de type  $k$ . Ces variétés contiennent alors toutes les lignes de champs qui ont pour origine (respectivement destina-





**FIGURE 3.9** – Champ de temps de réionisation (en haut à gauche en couleur) et ses minima (représentés par les étoiles noires). Les petites valeurs du champ sont en bleu, ses valeurs moyennes en blanc et ses grandes valeurs en rouges. Le champ a une moyenne nulle ici et est normalisé par son écart-type. En haut à droite, les variétés ascendantes de type 2 (ou les régions associées aux minima) délimitées par les lignes noires sont montrés avec les différentes couleurs qui représentent simplement une numérotation de ces régions. En bas à gauche, le squelette des temps de réionisation est représenté avec les lignes noires. En bas à droite, le champ de temps de réionisation est superposé à son squelette, et on voit que ce dernier suit bien les régions aux valeurs les plus élevées (il passe par les maxima).

tion) le même point critique.  $k$  dépend de l’indice du point critique associé à la variété : les variétés descendantes de type  $k$  sont celles originaires de points critiques d’indice  $k$  tandis que les variétés ascendantes de type  $k$  sont originaires des points critiques d’indice  $D - k$  avec  $D$  la dimension du champ. Par exemple, dans notre analogie avec un paysage montagneux, les variétés ascendantes de type 2 sont les vallées (voir sur l’illustration de la Figure 3.8) et les variétés descendantes de type 2 sont les montagnes. Autrement, en haut à droite sur la Figure 3.9, je représente les variétés ascendantes de type 2 du champ de temps de réionisation. Chaque région encadré en noir est une de ces variétés et la couleur représente simplement une numérotation de ces variétés. Celles-ci sont ainsi l’ensemble des points qui sont associés au minimum de la variété (représenté par une étoile) en descendant le long du gradient. Dans les Chapitres suivants, on appellera ces variétés des « patches », ou des « patches de réionisation » quand on les extraira du champ de temps de réionisation. L’ensemble des variétés ascendantes (ou descendantes) recouvrent ainsi tout l’espace et est appelé le complexe de Morse du champ.

Les bordures de ces régions s’appellent des arcs et sont les lignes de champ qui joignent deux points critiques. Pour un paysage montagneux, ces arcs sont les lignes de crêtes qui rejoignent les pics des montagnes aux cols. Dans le cas présenté dans la Figure 3.9, les bords des variétés ascendantes de type 2 (liées aux minima) sont des arcs joignant un maximum à un point selle. L’ensemble de ces arcs (maximum-point selle) forme ce que l’on appelle le *squelette* du champ, qui est représenté pour notre champ de temps de réionisation en bas à gauche de la Figure 3.9. Si l’on s’intéresse aux variétés descendantes de types 2 (associées aux maxima), les arcs qui les bordent relient des minima à des points selles et forment ce que l’on appelle ici l’*anti-squelette*.

### 3.2.2 Le code de topologie DisPerSE

DisPerSE a été développé par Sousbie (2011) et a d’abord été utilisé dans le contexte des grandes structures de l’Univers pour étudier les filaments de matière extraits de simulations cosmologiques ou observés dans des grands relevés, ainsi que pour explorer l’impact de ces filaments sur les galaxies et inversement. Par exemple, Sousbie et al. (2011) et Malavasi et al. (2020) ont créé des catalogues de filaments de la toile cosmique à partir de distributions de galaxies observées, et Tanimura et al. (2020) utilisent notamment le catalogue créé par Malavasi et al. (2020) pour étudier les propriétés physiques de ces filaments (densité et température). Des études comme Codis et al. (2018), Kraljic et al. (2020), Singh et al. (2020), Galárraga-Espinosa et al. (2020) et Song et al. (2021) explorent les propriétés des galaxies (comme leur masse ou leur taux de formation d’étoiles) et/ou leur distribution spatiale et orientation par rapport aux filaments de la toile cosmique qu’ils détectent avec DisPerSE dans des simulations cosmologiques. DisPerSE peut aussi être utilisé pour étudier les grandes structures de l’Univers pendant la Réionisation. Katz et al. (2020) s’intéressent à l’impact spatial et temporel de la Réionisation sur le contenu en baryon des filaments de la toile cosmique extraits de simulations cosmologiques grâce à DisPerSE.

DisPerSE a donc permis d’analyser ces grandes structures à partir du champ de densité ou de distributions de galaxies pour identifier les composants de la toile cosmique (les vides, les filaments, les murs et les amas denses à l’intersection de filaments). Pour étudier la Réionisation, j’ai surtout utilisé le champs de densité de gaz et de temps/redshift de réionisation

comme entrée de `DisPerSE`. Cet outil fonctionne ainsi sur tout type de champ : sur des champs définis sur une maille régulière (comme le champ de densité ou de temps de réionisation), non-régulière ou sur un ensemble de points (comme une distribution de galaxies). Ces données d'entrée sont généralement à 2 ou 3 dimensions, mais aucune limite n'est inscrite dans `DisPerSE`, même si le temps de calcul est en général une contrainte. Des données ayant une dimension plus élevée vont nécessairement mener à des calculs plus longs. `DisPerSE` prend également en compte des champs qui ont des conditions aux limites périodiques, ce qui est en général le cas en modélisation en cosmologie et qui sera le cas dans cette thèse.

La théorie de Morse se base sur des fonctions de Morse, qui sont définies comme étant deux fois différentiables et n'ayant aucun point critique dégénéré. Dans cette théorie, les fonctions sont ainsi définies sur des ensembles continus. Les données astrophysiques, étant discrétisées, ne satisfont cependant pas tout le temps ces conditions, c'est pourquoi `DisPerSE` repose sur la théorie de Morse discrète. Cette dernière a été développée par Forman (1998, 2001) et vient avec un nouveau jeu de concepts et de définitions que je vais brièvement décrire. J'invite le lecteur intéressé à se référer à Sousbie (2011) par exemple pour plus de détails.

Le principe de cette théorie discrète est qu'elle est applicable sur des espaces discrets, que l'on appelle « complexes simpliciaux ». On peut par exemple citer la triangulation de Delaunay comme exemple de complexe simplicial beaucoup utilisé en cosmologie (Schaap 2007). Les éléments discrets de ces complexes sont appelés « simplexes ». Une « face » d'un simplexe est un simplexe de dimension inférieure et dont les sommets sont contenus dans le simplexe. Une « coface » d'un simplexe est un simplexe de dimension supérieure, faisant de ce simplexe une face de la coface. On peut alors définir un nouveau type de fonctions : les fonctions de Morse discrètes, qui sont définies sur un complexe simplicial et qui associent ainsi une valeur à chaque simplexe. Ces fonctions portent une condition supplémentaire qui est équivalente dans le cas discret à celle imposant une non-dégénérescence des points critiques dans le cas des fonctions de Morse. Pour les fonctions de Morse discrètes, cette condition devient alors une condition sur les valeurs de la fonction. Celles-ci doivent alors être telles qu'il existe au maximum une face où la fonction prend une valeur supérieure à celle d'un simplexe donné et une coface où elle prend une valeur inférieure. Cette condition crée alors une direction privilégiée et donc un gradient discret (qui est l'équivalent du gradient de la théorie continue de Morse). Des lignes de gradient peuvent ensuite être définies à partir de là et les caractéristiques topologiques retrouvées dans ce cas de fonction discrète. Schaap (2007) montre avec la méthode DTFE (Delaunay Tessellation Field Estimator) que l'on peut en pratique toujours construire une fonction qui satisfait cette condition de Morse discrète. Il montre ainsi tout l'intérêt de cette théorie discrète et son outil, que `DisPerSE` reprend, permet d'obtenir une fonction qui satisfait cette condition par construction.

Après avoir donné une fonction (champ ou distribution de points) en entrée à `DisPerSE`, nous pouvons directement lire en sortie ses différentes caractéristiques topologiques, à savoir les points critiques, les variétés et les arcs. Les arcs permettent de reconstituer le squelette ou l'anti-squelette du champ. Ainsi, selon le champ donné en entrée, nous avons accès à différents types d'informations sur la topologie du champ. Par exemple, dans le cas d'un champ de densité, les maxima sont les régions les plus denses localement, le squelette représente directement les filaments de la toile cosmique, et les variétés sont par exemple les vides cosmiques et les murs à l'intersection de ceux-ci. Dans le cas du champ de temps de réionisation, ce sont ces objets que nous montrons dans la Figure 3.9, et nous allons voir dans la suite de

ce manuscrit comment ils peuvent être interprétés dans le contexte de la Réionisation.

### 3.2.3 Le paramètre de la persistance

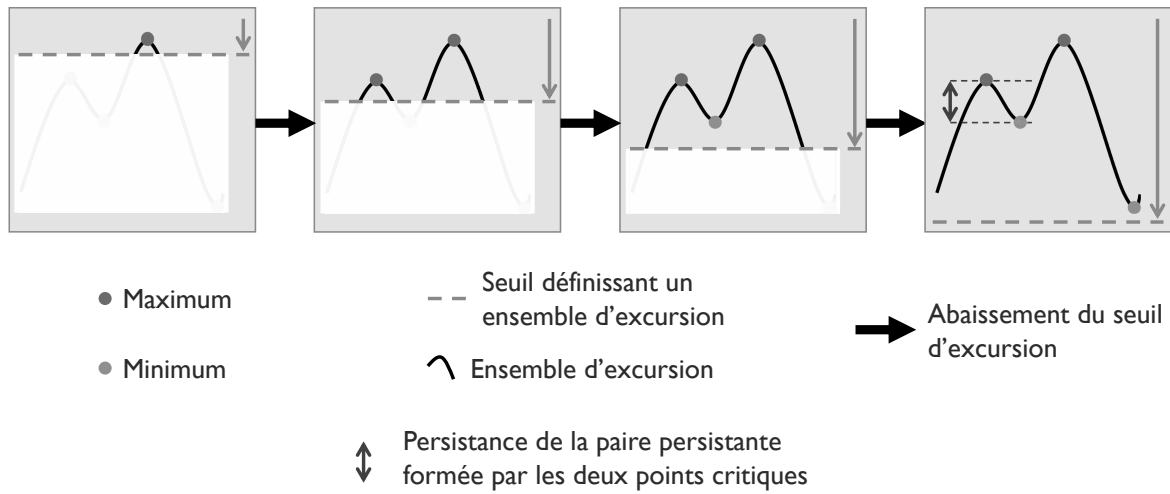
`DisPerSE` fait également appel au concept de la persistance pour permettre à l'utilisateur de se focaliser seulement sur les endroits les plus significatifs du champ. La topologie d'un champ peut être étudiée à travers ses ensembles d'excursion, qui correspondent à l'ensemble des points qui ont une valeur du champ supérieure à certain seuil et qui évoluent à mesure que l'on fait évoluer le seuil. Dans le schéma de la Figure 3.10, on peut par exemple voir des ensembles d'excursion en noir : quand le seuil passe par un point critique du champ, la topologie des ensembles d'excursion en est modifiée. Par exemple, si le seuil diminue et passe en dessous d'un maximum, alors de nouvelles structures apparaissent dans l'ensemble d'excursion (voir le deuxième panneau dans la Figure 3.10) ; et si le seuil passe en dessous d'un minimum, alors des structures fusionnent faisant disparaître des structures isolées (voir le troisième panneau dans la Figure 3.10). Les points critiques à l'origine d'apparition de nouvelles structures (les maxima dans l'exemple) sont appelés « positifs » et ceux à l'origine de disparitions de structures (les minima dans l'exemple) sont appelés « négatifs ». Les composants topologiques d'un champ peuvent alors être représentés par un paire de points critiques positif et négatif, que l'on appelle « paire persistante ». Ce que l'on appelle la « persistance » est la différence absolue entre les valeurs des deux points critiques de la paire (voir la flèche rose dans la Figure 3.10). La persistance représente ainsi « la durée de vie » de la structure topologique qui correspond à une paire persistante dans l'ensemble d'excursion.

`DisPerSE` intègre cette notion de persistance et permet à l'utilisateur de faire varier un seuil de persistance qui contrôle ainsi le nombre de structures topologiques extraites du champ. On exprime souvent ce paramètre de la persistance comme étant proportionnel à l'écart-type  $\sigma$  du champ. Un exemple d'application d'un seuil de persistance sur une fonction ou un champ est représenté avec le schéma de la Figure 3.11. On voit que seules les structures avec des paires persistantes ayant une persistance plus grande que le seuil de persistance se maintiennent. Le filtre de la persistance applique également un lissage au champ et permet ainsi de garder en sortie les structures les plus significatives à  $n - \sigma$  du champ. En d'autres termes, ce paramètre contrôle le niveau de précision que l'on souhaite avoir sur les produits de `DisPerSE` ou le niveau de bruit que l'on souhaite y soustraire. La persistance peut effectivement être très utile si les données en entrée du code sont bruitées parce qu'elle sert de seuil en-dessous duquel `DisPerSE` ne va pas chercher de structures topologiques.

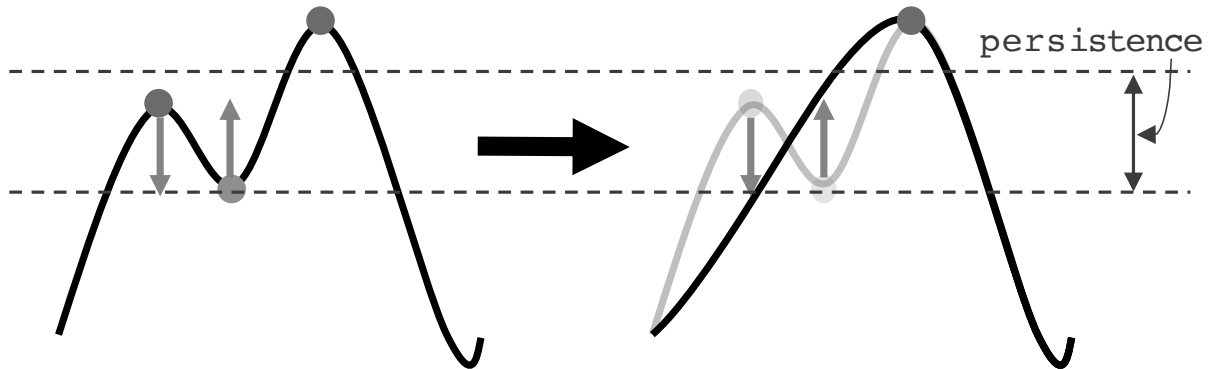
## 3.3 Topologie à travers les champs gaussiens aléatoires

Pour finir ce Chapitre, je vais maintenant m'intéresser à différentes statistiques topologiques que l'on peut extraire de nos champs d'intérêt (en particulier  $t_{\text{reion}}(\vec{r})$ ) et que l'on peut également calculer analytiquement avec la théorie des champs Gaussiens aléatoires (Gaussian random fields, GRFs). J'ai utilisé ces statistiques pendant ma thèse et je décris ces études dans le Chapitre 6.

La théorie des GRFs est un outil mathématique permettant d'accéder de manière analytique à diverses statistiques. Elle a été introduite par Rice (1944) pour extraire des statistiques



**FIGURE 3.10** – Schéma illustrant l'évolution d'un ensemble d'excursion (montré en noir) avec son seuil (en gris) qui diminue. Dans le premier panneau on remarque une structure isolée. On voit de nouvelles structures apparaître quand le seuil passe en dessous du maximum de gauche (en rouge) dans le deuxième panneau, et des structures isolées fusionner avec d'autres structures isolées qui disparaissent donc quand le seuil passent en dessous du minimum (en bleu) dans le troisième panneau. La persistence est illustrée dans le dernier panneau par la flèche rose.



**FIGURE 3.11** – Schéma illustrant le paramètre de la persistence intégré dans DisPerSE. On voit une fonction 1D représentée en noire, avec ses maxima en rouge et minima en bleu. Le seuil de persistence est ici représenté par la distance entre les traits violets et il est en particulier plus grand que la persistence de la paire persistante formée par les deux points critiques de gauche. À gauche, la fonction est représentée telle quelle, et à droite, on voit le résultat de l'application du filtre de la persistence à la fonction. Les flèches vertes représentent l'effet de ce seuil de persistence, qui « efface » les deux points critiques de gauche et donne un effet de lissage de la fonction.

de bruits Gaussiens 1D provenant d'appareils électriques. Ensuite, Longuet-Higgins (1957) l'a réutilisée dans le cas de vagues sur des surfaces à deux dimensions. Plus tard, elle est appliquée pour la première fois dans un contexte astrophysique par Bardeen et al. (1986) pour étudier la formation cosmologique des grandes structures en trois dimensions. Ces auteurs posent de nouveau les bases d'une théorie qui va amplement être utilisée en astrophysique par la suite pour étudier des champs Gaussiens (Pogosyan et al. 2009b,a; Pichon et al. 2010; Gay et al. 2012; Codis et al. 2018; Cadiou et al. 2020) ou légèrement non-Gaussiens (Matsubara 2003; Pogosyan et al. 2011; Gay et al. 2012; Cadiou et al. 2020; Matsubara & Kuriki 2021). Par exemple, Gay et al. (2012) s'intéressent à des statistiques 2D et 3D de champs cosmologiques, comme le décompte des pics dans des fond diffus cosmologiques (CMB) synthétiques pour étudier leurs non-Gaussianités.

Plusieurs statistiques souvent utilisées pour l'étude de la Réionisation cosmologique peuvent être obtenues dans le cadre de cette théorie. C'est par exemple le cas des MFs, ou du genre et de la caractéristique d'Euler dérivés des MFs, ainsi que des nombres de Betti dont je discute dans les Section 3.1.1 et 3.1.4 (Schmalzing & Gorski 1998; Matsubara 2003; Gleser et al. 2006; Lee et al. 2008; Gay et al. 2012; Kapahtia et al. 2019; Matsubara & Kuriki 2021). Comme discuté dans les Section précédentes, Gleser et al. (2006) utilisent les expressions Gaussiennes des MFs pour étudier la Gaussianité de la morphologie du processus de réionisation. Pendant cette thèse, j'utilise également la théorie des GRFs pour étudier la Gaussianité du processus de réionisation en comparant les expressions théoriques de diverses statistiques topologiques décrivant la topologie de ce processus à des mesures faites dans le champ de temps de réionisation (Thélie et al. 2023).

Dans cette Section, je vais commencer par décrire ce qu'est un GRF et montrer comment on obtient les matrices de covariance et les fonctions de distribution de probabilité (PDFs), qui sont utiles pour le calcul analytique des statistiques par la suite. Après cela, je détaille les calculs de longueurs typiques et d'un paramètre spectral pouvant être obtenus à partir du spectre de puissance d'un GRF. Enfin, je synthétise brièvement le calcul de certaines statistiques topologiques que l'on peut obtenir avec cette théorie. Parmi ces statistiques, je discute de la PDF d'un champ  $F$  et de la norme de son gradient, de la longueur de ses isocontours et de son squelette, et de la PDF des valeurs à ses points critiques. Je ne m'intéresse ici qu'à ces cinq statistiques parce qu'elles permettent de décrire une grande partie du processus de réionisation (histoire globale de réionisation, sources de réionisation, vitesse des fronts d'ionisation, évolution des bulles ionisées et des régions neutre et percolation) si on les extrait du champ de temps de réionisation. Ceci sera plus expliqué dans le Chapitre 6, et le détail des calculs donné ici n'est pas strictement nécessaire pour comprendre les études menées dans les Chapitres à venir.

### 3.3.1 Champs aléatoires gaussiens et isotropes

Dans toute cette Section, on s'intéresse à un GRF que l'on notera  $F$ . On suppose que ce champ est à deux dimensions et tous les calculs présentés ici sont menés en 2D. Ces calculs sont notamment extraits de Bardeen et al. (1986); Gay et al. (2012); Cadiou et al. (2020). Par définition, un GRF est un champ dont les valeurs sont distribuées selon une distribution Gaussienne. On peut ainsi écrire la PDF du champ Gaussien  $F$  de moyenne nulle de la

manière suivante :

$$P(F)dF = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{F^2}{2\sigma_0^2}\right) dF, \quad (3.3)$$

avec  $\sigma_0$  l'écart-type du champ  $F$ . On voit ici que le champ Gaussien est entièrement défini avec l'unique connaissance de son écart-type  $\sigma_0$ .

Pour accéder à des statistiques diverses de ce champ Gaussien  $F$ , de multiples intégrations de la PDF des valeurs de ce champ et de ses dérivées peuvent être nécessaires. Cette PDF à plusieurs variables est une probabilité jointe et elle est également Gaussienne (Bardeen et al. 1986). Elle s'écrit ainsi :

$$P(\vec{x})d^n\vec{x} = \frac{1}{(2\pi)^{n/2} \cdot \det(C)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\vec{x} \cdot C^{-1} \cdot \vec{x}\right) d^n\vec{x}. \quad (3.4)$$

Ici,  $\vec{x}$  est un vecteur  $n$ -D contenant la fonction  $F$  et potentiellement ses dérivées (selon la statistique que l'on souhaite calculer par la suite). Par exemple,  $\vec{x}$  pourrait s'écrire  $\vec{x} = (F, F_1, F_2, F_{11}, F_{22}, F_{12}, \dots)$  avec  $F_1$  et  $F_2$  les dérivées premières du champ  $F$  et  $F_{11}$ ,  $F_{22}$  et  $F_{12}$  ses dérivées secondes.  $C = \langle \vec{x} \otimes \vec{x} \rangle$  est la matrice de covariance de ce champ.

Pour un champ Gaussien, homogène et isotrope,  $C$  peut s'exprimer relativement simplement. En suivant les notations introduites par Pogosyan et al. (2009a) et Gay et al. (2012), je normalise le champ  $F$  et ses dérivées par leur moment :

$$x = \frac{F}{\sigma_0}, \quad x_i = \frac{\nabla_i F}{\sigma_1}, \quad \text{and} \quad x_{ij} = \frac{\nabla_i \nabla_j F}{\sigma_2}, \quad (3.5)$$

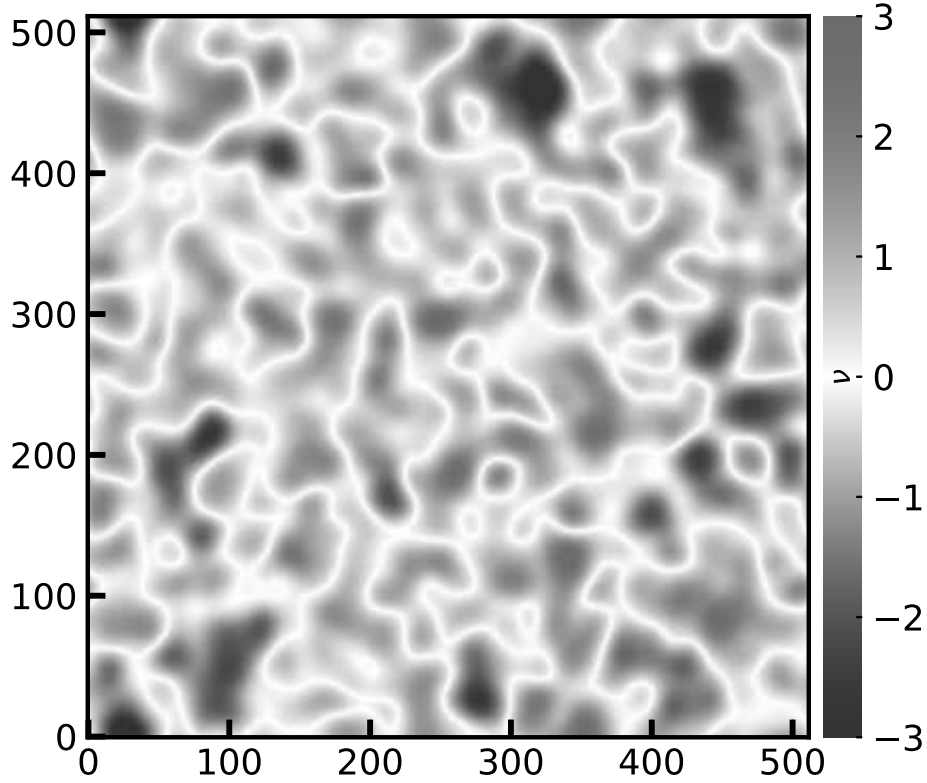
où  $i, j \in \{1, 2\}$  correspondent aux dimensions de l'espace Euclidien 2D. Les trois premiers moments de ce champ sont la variance du champ  $F$ , celle de son gradient, puis de son Laplacien, et s'expriment comme suit :

$$\sigma_0^2 = \langle F^2 \rangle, \quad \sigma_1^2 = \left\langle \left( \vec{\nabla} F \right)^2 \right\rangle, \quad \text{et} \quad \sigma_2^2 = \langle (\Delta F)^2 \rangle. \quad (3.6)$$

Quand on étudie un champ Gaussien, homogène et isotrope, les moments sont tout ce que l'on a besoin de connaître pour le définir entièrement. Un exemple de GRF 2D normalisé est montré dans la Figure 3.12, et les valeurs du champ sont exprimées par la notation  $\nu$ .

Maintenant, les statistiques qui nous intéressent par la suite dépendent du champ  $x$  en lui-même et potentiellement de plusieurs de ces dérivées. Dans certains cas, comme pour la PDF de la norme du gradient de  $F$ , on peut avoir une dépendance au champ et à ses premières dérivées seulement. On écrit alors  $\vec{x} = (x, x_1, x_2)^T$ . La matrice de covariance est relativement simplifiée dans le cas Gaussien parce que plusieurs covariances s'annulent par indépendance des dérivées paires et impaires (Gay 2011). Elle prend la forme suivante :

$$\begin{aligned} C &= \begin{pmatrix} \langle F^2 \rangle & \langle F \nabla_1 F \rangle & \langle F \nabla_2 F \rangle \\ \langle \nabla_1 F F \rangle & \langle (\nabla_1 F)^2 \rangle & \langle \nabla_1 F \nabla_2 F \rangle \\ \langle \nabla_2 F F \rangle & \langle \nabla_1 F \nabla_2 F \rangle & \langle (\nabla_2 F)^2 \rangle \end{pmatrix} \\ &= \begin{pmatrix} \sigma_0^2 & 0 & 0 \\ 0 & \frac{1}{2}\sigma_1^2 & 0 \\ 0 & 0 & \frac{1}{2}\sigma_1^2 \end{pmatrix}. \end{aligned} \quad (3.7)$$



**FIGURE 3.12** – Exemple de champ aléatoire Gaussien 2D normalisé par son écart-type  $\sigma_0$ , dont les valeurs du champ sont noté  $\nu$ , et qui a été généré avec un spectre de puissance comme celui présenté dans l'Équation 3.17.

En utilisant les variables normalisées  $x$ ,  $x_1$ , and  $x_2$ , la matrice de covariance devient :

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}. \quad (3.8)$$

Grâce à cette matrice, on peut alors accéder à la PDF jointe d'un champ Gaussien et de ses premières dérivées :

$$P(x, x_1, x_2) dx dx_1 dx_2 = \frac{2}{(2\pi)^{3/2}} e^{-\left(\frac{1}{2}x^2 + x_1^2 + x_2^2\right)} dx dx_1 dx_2. \quad (3.9)$$

Dans d'autres cas, comme pour la PDF des valeurs du champ en ses extrema, les statistiques peuvent dépendre du champ  $F$  et de ses dérivées premières et secondes. On écrit alors  $\vec{x} = (x, x_1, x_2, x_{11}, x_{22}, x_{12})^T$ , et on a besoin de connaître les covariances suivantes, qui sont les seules à être non nulles du fait de l'indépendance des dérivées paires et impaires (Gay 2011) :

$$\begin{cases} \langle \nabla_i F \nabla_j F \rangle = \frac{1}{2} \sigma_1^2 \delta_{ij} \\ \langle \nabla_i \nabla_j F \nabla_k \nabla_l F \rangle = \frac{1}{8} \sigma_2^2 (\delta_{ij} \delta_{kl} + \delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}) \\ \langle F \nabla_i \nabla_j F \rangle = -\langle \nabla_i F \nabla_j F \rangle = -\frac{1}{2} \sigma_1^2 \delta_{ij} \end{cases} \quad (3.10)$$



La matrice de covariance prend alors la forme :

$$C = \begin{pmatrix} \sigma_0^2 & 0 & 0 & -\frac{1}{2}\sigma_1^2 & -\frac{1}{2}\sigma_1^2 & 0 \\ 0 & \frac{1}{2}\sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2}\sigma_1^2 & 0 & 0 & 0 \\ -\frac{1}{2}\sigma_1^2 & 0 & 0 & \frac{3}{8}\sigma_2^2 & \frac{3}{8}\sigma_2^2 & 0 \\ -\frac{1}{2}\sigma_1^2 & 0 & 0 & \frac{3}{8}\sigma_2^2 & \frac{3}{8}\sigma_2^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{3}{8}\sigma_2^2 \end{pmatrix}. \quad (3.11)$$

En utilisant les variables normalisées  $x$ ,  $x_1$ ,  $x_2$ ,  $x_{11}$ ,  $x_{22}$  et  $x_{12}$ , ainsi que le paramètre spectral  $\gamma = \frac{\sigma_1^2}{\sigma_0\sigma_2}$ , la matrice de covariance devient :

$$C = \begin{pmatrix} 1 & 0 & 0 & -\frac{1}{2}\gamma & -\frac{1}{2}\gamma & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ -\frac{1}{2}\gamma & 0 & 0 & \frac{3}{8} & \frac{3}{8} & 0 \\ -\frac{1}{2}\gamma & 0 & 0 & \frac{3}{8} & \frac{3}{8} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{3}{8} \end{pmatrix}. \quad (3.12)$$

Cette matrice montre une dépendance de plusieurs variables entre elles, que l'on peut lever en la diagonalisant. Pour se faire, on définit les nouvelles variables suivantes (Bardeen et al. 1986; Gay 2011) :

$$u = -(x_{11} + x_{22}), \quad v = \frac{1}{2}(x_{11} - x_{22}), \quad \zeta = \frac{x - \gamma u}{\sqrt{1 - \gamma^2}}, \quad (3.13)$$

et la matrice de covariance s'exprime ainsi comme suit :

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{8} \end{pmatrix}. \quad (3.14)$$

Grâce à cette matrice, on peut alors accéder à la PDF jointe d'un champ Gaussien, de ses premières dérivées et de ses dérivées secondes :

$$P(\zeta, x_1, x_2, u, v, x_{12}) = \frac{16}{(2\pi)^3} e^{-Q(\zeta, x_1, x_2, u, v, x_{12})}, \quad (3.15)$$

avec  $Q$  définit comme suit :

$$Q(\zeta, x_1, x_2, u, v, x_{12}) = \frac{1}{2}\zeta^2 + x_1^2 + x_2^2 + \frac{1}{2}u^2 + 4v^2 + 4x_{12}^2. \quad (3.16)$$

Cette PDF donne donc accès par le biais de diverses intégration à la majorité des statistiques que je vais mentionner dans les Sections suivantes.

### 3.3.2 Longueurs typiques et paramètre spectral pour un spectre de puissance donné

Toute l'information statistique d'un GRF se trouve dans son spectre de puissance  $\mathcal{P}_k$ . Le spectre de puissance peut prendre n'importe quelle forme, mais dans les études présentées dans cette thèse, je m'intéresse surtout à des champs dont le spectre de puissance est défini par deux lois de puissance, comme suit :

$$\mathcal{P}_k = \begin{cases} A_1 k^{n_1} & \text{if } k \leq k_{\text{thresh}} \\ A_2 k^{n_2} & \text{if } k > k_{\text{thresh}} \end{cases}, \quad (3.17)$$

avec  $A_1$  et  $A_2$  les amplitudes de chacune des lois de puissance,  $n_1$  et  $n_2$  leur puissance et  $k_{\text{thresh}}$  le nombre d'onde séparant ces deux lois. De plus, je travaille aussi avec des champs lissés par un filtre Gaussien pour assurer la convergence de certaines statistiques. Ce lissage peut directement être inclus dans le spectre de puissance du champ :

$$\mathcal{P}_k^{\text{smoothed}} = \begin{cases} A_1 k^{n_1} e^{-2R_f^2 k^2} / 2\pi & k \leq k_{\text{thresh}} \\ A_2 k^{n_2} e^{-2R_f^2 k^2} / 2\pi & k > k_{\text{thresh}} \end{cases}. \quad (3.18)$$

$R_f$  est ici l'écart-type du filtre Gaussien (c'est-à-dire sa longueur), et s'exprime en nombre de cellules.

On peut maintenant remarquer que les moments du champ  $F$  peuvent aussi se définir à partir du spectre de puissance  $\mathcal{P}_k$  par la fonction suivante :

$$\sigma_i^2 = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \int_0^\infty k^{2i} \mathcal{P}_k k^{d-1} dk, \quad (3.19)$$

où  $i \in \mathbb{N}$  correspond au nombre de fois que l'on dérive le champ, et  $d$  est la dimension du champ (ici  $d = 2$ ). Si l'on prend la définition du spectre de puissance incluant le lissage donné dans l'Équation 3.18, les moments prennent alors la forme suivante :

$$\sigma_i^2 = \frac{A_2 \pi^{\frac{d}{2}-1}}{2\Gamma\left(\frac{d}{2}\right)} \left[ \frac{A_1}{A_2} \left( \frac{1}{\sqrt{2}R_f} \right)^{d+2i+n_1} \gamma\left( \frac{d+n_1}{2} + i, 2R_f^2 k_{\text{thresh}}^2 \right) + \left( \frac{1}{\sqrt{2}R_f} \right)^{d+2i+n_2} \Gamma\left( \frac{d+n_2}{2} + i, 2R_f^2 k_{\text{thresh}}^2 \right) \right]. \quad (3.20)$$

Dans cette expression, les fonctions gamma sont utilisées et elles se définissent comme suit :

$$\gamma(a, x) = \int_0^x t^{a-1} e^{-t} dt \quad \text{et} \quad \Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt. \quad (3.21)$$

Notons ainsi que les moments  $\sigma_i$  ne dépendent que des paramètres du spectre de puissance et de la dimension du champ  $F$  :  $\sigma_i^2 = \sigma_i^2(d, A_1, n_1, A_2, n_2, R_f)$ . À partir de ces moments (et donc du spectre de puissance), nous définissons les longueurs suivantes :

$$R_0 = \frac{\sigma_0}{\sigma_1}, \quad \text{et} \quad R_* = \frac{\sigma_1}{\sigma_2}. \quad (3.22)$$

$R_0$  et  $R_*$  peuvent se voir comme des longueurs caractéristiques du champ  $F$  (distances typiques entre deux zéros et deux extrema respectivement). On retrouve le paramètre spectral  $\gamma$  en faisant le rapport de ces deux longueurs :

$$\gamma = \frac{R_*}{R_0} = \frac{\sigma_1^2}{\sigma_0\sigma_2}. \quad (3.23)$$

Les moments  $\sigma_i^2$  ont la même unité que le spectre de puissance, mais le paramètre spectral  $\gamma$  est sans dimension.

### 3.3.3 Fonction de distribution de probabilité du champ

La PDF du champ Gaussien  $F$  est ce que l'on peut calculer de plus simple analytiquement puisque c'est directement la PDF de  $x$  :

$$P(x)dx = \frac{1}{2\pi} e^{-\frac{1}{2}x^2} dx. \quad (3.24)$$

À partir de cette PDF, on peut aussi calculer le facteur de remplissage du champ  $F$  :

$$FR(\nu) = \int_0^\nu P(x)dx = \frac{1}{2} \operatorname{erf}\left(\frac{\nu}{\sqrt{2}}\right), \quad (3.25)$$

avec  $\operatorname{erf}(\nu) = \frac{2}{\sqrt{\pi}} \int_0^\nu e^{-y^2} dy$ .

### 3.3.4 Fonction de distribution de probabilité de la norme du gradient du champ

On définit la norme du gradient du champ  $F$  comme suit :

$$\|\vec{\nabla}x\| = \frac{1}{\sigma_1} \sqrt{(\nabla_1 F)^2 + (\nabla_2 F)^2} = R_0 \sqrt{(\nabla_1 x)^2 + (\nabla_2 x)^2}, \quad (3.26)$$

avec  $\nabla_i x$  pour  $i \in \{1, 2\}$  les deux composantes du gradient de  $x = F/\sigma_0$ , et  $R_0$  défini dans l'Équation 3.22.

La PDF de cette norme dépend ainsi du champ  $F$  et de ses dérivées premières et elle est donnée dans l'Équation 3.9. Avec un changement de variable en  $w^2 = x_1^2 + x_2^2$ , la PDF de la norme du gradient du champ devient :

$$2\pi P(w)w dw = 2we^{-w^2} dw. \quad (3.27)$$

### 3.3.5 Longueur d'isocontours

La longueur moyenne des isocontours 2D d'un champ  $\mathcal{L}$  nous permet de caractériser les différents niveaux du champ en mesurant leur longueur. Cette longueur est en fait la MF  $V_1$  d'un champ 2D Schmalzing & Gorski (1998); Matsubara (2003). Pour n'importe quel champ, on peut donc définir la longueur d'isocontours à un niveau  $\nu$  donné comme étant :

$$\mathcal{L}(\nu) = \left\langle \frac{1}{R_0} \delta(x - \nu) (x_1^2 + x_2^2)^{1/2} \right\rangle = \left\langle \frac{1}{R_0} \delta(x - \nu) w \right\rangle, \quad (3.28)$$

avec  $\delta(x - \nu)$  la distribution de Dirac valant 1 si  $x = \nu$  et 0 sinon. Pour rappel,  $\nu$  est la valeur du champ normalisé en un point de l'espace, et on reprend la notation  $w^2 = x_1^2 + x_2^2$ . La longueur  $R_0$  est présente dans l'expression en raison de la normalisation du champ  $F$  et donne sa dimension à  $\mathcal{L}$ . Cette définition de la longueur d'isocontours peut être vu comme une somme de toutes les petites variations  $w$  du champ à un seuil  $\nu$  donné.

Pour calculer cette longueur, nous avons besoin que la PDF du champ dépende du champ  $x$  et de ses premières dérivées  $x_1$  et  $x_2$ . Dans le cas d'un champ Gaussien, on peut utiliser la même transformation de  $P(x, x_1, x_2)dx dx_1 dx_2$  en  $2\pi P(w)w dw$  que dans la Section précédente, on obtient :

$$2\pi P(x, w)w dx dw = \frac{2}{\sqrt{2\pi}} e^{-\left(\frac{1}{2}x^2 + w^2\right)} w dx dw. \quad (3.29)$$

En calculant la moyenne de l'Équation 3.28 avec cette distribution, on obtient directement la longueur moyenne d'isocontours à un niveau  $\nu$  :

$$\mathcal{L}(\nu) = \int_{-\infty}^{\infty} dx \int_0^{\infty} dw \frac{2\pi}{R_0} P(x, w) \delta(x - \nu) w^2 = \frac{1}{2\sqrt{2}R_0} e^{-\frac{1}{2}\nu^2}. \quad (3.30)$$

### 3.3.6 Fonction de distribution de probabilité des points critiques du champ

La densité moyenne d'extrema d'un champ quelconque est donnée par l'expression ci-dessous, comme l'explique Gay (2011) et que l'on retrouve aussi dans Bardeen et al. (1986) :

$$\frac{\partial n_{\text{ext}}}{\partial \nu} = \left\langle \frac{1}{R_*^2} |x_{11}x_{22} - x_{12}^2| \delta(x_1)\delta(x_2)\delta(x - \nu) \right\rangle. \quad (3.31)$$

Cette définition de la distribution repose sur la définition même d'un point critique, c'est-à-dire un point de gradient nul qui fait apparaître dans l'expression la partie  $\delta(x_1)\delta(x_2)$ . Le déterminant de la matrice Hessienne de la fonction  $F$ , qui est  $|x_{11}x_{22} - x_{12}^2|$ , permet de compter le nombre de points critiques présents dans le champ. La présence de  $R_*$  est due à la normalisation du champ  $F$ . Pour calculer cette distribution, on a donc besoin de connaître la PDF jointe du champ, de ses dérivées premières et secondes, que dans le cas d'un champ Gaussien, je la donne dans l'Équation 3.15.

Pour calculer  $\frac{\partial n_{\text{ext}}}{\partial \nu}$  avec sa définition dans l'Équation 3.31 et la PDF à 6 dimensions du champ  $F$ , on doit effectuer des intégrations non-triviales que je ne détaillerai pas ici par simplicité. Elles sont explicitées dans l'Appendice A de Bardeen et al. (1986). Ces intégrations nécessitent en particulier des contraintes supplémentaires quant au type du point critique (minimum ou maximum par exemple). Dans le cas d'une distribution des valeurs du champ

en ses minima, on a alors :

$$\begin{aligned}
 \frac{\partial n_{\min}}{\partial \nu} &= \frac{\exp\left(\frac{-\nu^2}{2}\right)}{\sqrt{2\pi}R_*^2} \left[ 1 - \operatorname{erf}\left(\frac{\gamma\nu}{\sqrt{2(1-\gamma^2)}}\right) \right] K_1(\nu, \gamma) \\
 &+ \frac{\exp\left(\frac{-3\nu^2}{6-4\gamma^2}\right)}{\sqrt{2\pi(1-\frac{2}{3}\gamma^2)}R_*^2} \left[ 1 - \operatorname{erf}\left(\frac{\gamma\nu}{\sqrt{2(1-\gamma^2)(3-2\gamma^2)}}\right) \right] K_2 \\
 &- \frac{\exp\left(\frac{-\nu^2}{2(1-\gamma^2)}\right)}{\sqrt{2\pi(1-\gamma^2)}R_*^2} \left[ 1 + \exp\left(\frac{\nu^2}{2(1-\gamma^2)}\right) \right] K_3(\nu, \gamma),
 \end{aligned} \tag{3.32}$$

avec

$$K_1(\nu, \gamma) = \frac{\gamma^2(\nu^2 - 1)}{8\pi}, \quad K_2 = \frac{1}{8\pi\sqrt{3}}, \quad \text{and} \quad K_3(\nu, \gamma) = \frac{\gamma(1-\gamma^2)\nu}{2(2\pi)^{3/2}}. \tag{3.33}$$

Si l'on s'intéresse plutôt à une distribution des valeurs du champ en ses maxima, on a :

$$\begin{aligned}
 \frac{\partial n_{\max}}{\partial \nu} &= \frac{\exp\left(\frac{-\nu^2}{2}\right)}{\sqrt{2\pi}R_*^2} \left[ 1 + \operatorname{erf}\left(\frac{\gamma\nu}{\sqrt{2(1-\gamma^2)}}\right) \right] K_1(\nu, \gamma) \\
 &+ \frac{\exp\left(\frac{-3\nu^2}{6-4\gamma^2}\right)}{\sqrt{2\pi(1-\frac{2}{3}\gamma^2)}R_*^2} \left[ 1 + \operatorname{erf}\left(\frac{\gamma\nu}{\sqrt{2(1-\gamma^2)(3-2\gamma^2)}}\right) \right] K_2 \\
 &+ \frac{\exp\left(\frac{-\nu^2}{2(1-\gamma^2)}\right)}{\sqrt{2\pi(1-\gamma^2)}R_*^2} \left[ 1 + \exp\left(\frac{\nu^2}{2(1-\gamma^2)}\right) \right] K_3(\nu, \gamma).
 \end{aligned} \tag{3.34}$$

avec seulement quelques signes différents par rapport à  $\frac{\partial n_{\min}}{\partial \nu}$ .

### 3.3.7 Longueur de squelette

Le calcul de la distribution des longueurs de squelette a été décrit dans plusieurs études, comme Pogosyan et al. (2009b), Gay (2011) et Gay et al. (2012). Je résume ici de manière rapide l'obtention de cette distribution. Le lecteur intéressé peut par exemple se référer à Pogosyan et al. (2009b) pour une explication détaillée. Pour rappel, le squelette (resp. l'anti-squelette) correspond aux lignes de champ qui relient les maxima (resp. les minima) aux points selles. Le squelette est ainsi l'ensemble pour lesquels le gradient est aligné avec le vecteur propre de la Hessienne  $\mathcal{H} = \vec{\nabla}\vec{\nabla}F$  du champ  $F$  et suit la direction de la valeur propre positive (resp. négative) de la Hessienne. Cette définition peut se traduire mathématiquement comme suit :  $\mathcal{H} \cdot \vec{\nabla}F = \lambda \vec{\nabla}F$ , avec  $\lambda$  la valeur propre positive (resp. négative) de la Hessienne. De manière équivalente, un point se trouve sur une ligne critique (c'est-à-dire sur le squelette) si  $s = \det(\mathcal{H} \cdot \vec{\nabla}x, \lambda \vec{\nabla}x) = 0$ , avec  $x = F/\sigma_0$  le champ normalisé. Ainsi, pour un champ quelconque, la distribution des longueurs d'un squelette ou anti-squelette prend la forme :

$$\frac{\partial L^{\text{skel}}}{\partial \nu} = \left\langle \frac{1}{R_*} \delta(s) |\vec{\nabla}s| \delta(x - \nu) \right\rangle, \tag{3.35}$$

avec  $|\vec{\nabla}s|$  la longueur des lignes de champ et  $R_*$  comme défini dans l'Équation 3.22. Pour calculer analytiquement cette expression, la PDF jointe du champ  $F$  et de ses premières, secondes et troisièmes dérivées est nécessaire. L'« approximation de squelette raide » est généralement utilisée ici parce qu'elle permet de négliger les dérivées supérieures à la troisième, ce qui simplifie grandement les calculs (Novikov et al. 2006; Pogosyan et al. 2009b; Gay 2011). Celle-ci signifie que le squelette est considéré comme relativement droit, réduisant probablement la longueur totale du squelette donnée par ce calcul. Dans le cas d'un champ Gaussien, tout ceci mène ainsi à la distribution suivante de longueur de squelette par unité de surface :

$$\begin{aligned} \frac{\partial L^{\text{skel}}}{\partial \nu} &= \frac{1}{\sqrt{2\pi}R_*} \exp^{-\frac{\nu^2}{2}} \left[ \frac{1}{8\sqrt{\pi}}(\sqrt{\pi} + 2\gamma\nu) \left( 1 + \operatorname{erf} \left( \frac{\gamma\nu}{\sqrt{2(1-\gamma^2)}} \right) \right) \right. \\ &\quad \left. + \frac{\sqrt{1-\gamma^2}}{2\sqrt{2\pi}} \exp \left( -\frac{\gamma^2\nu^2}{2(1-\gamma^2)} \right) \right], \end{aligned} \quad (3.36)$$

avec  $\gamma$  le paramètre spectral défini dans l'Équation 3.23. La distribution équivalente pour la longueur d'anti-squelette est la suivante :

$$\begin{aligned} \frac{\partial L^{\text{anti-skel}}}{\partial \nu} &= \frac{1}{\sqrt{2\pi}R_*} \exp^{-\frac{\nu^2}{2}} \left[ \frac{1}{8\sqrt{\pi}}(\sqrt{\pi} - 2\gamma\nu) \left( 1 + \operatorname{erf} \left( \frac{-\gamma\nu}{\sqrt{2(1-\gamma^2)}} \right) \right) \right. \\ &\quad \left. + \frac{\sqrt{1-\gamma^2}}{2\sqrt{2\pi}} \exp \left( -\frac{\gamma^2\nu^2}{2(1-\gamma^2)} \right) \right]. \end{aligned} \quad (3.37)$$

Enfin, en intégrant ces distributions sur tout l'espace des valeurs du champ, on obtient la longueur totale de squelette ou d'anti-squelette suivante (Pogosyan et al. 2009b; Gay 2011) :

$$L^{\text{tot}} = \left( \frac{1}{8} + \frac{\sqrt{2}}{4\pi} \right) \frac{1}{R_*}. \quad (3.38)$$

Il faut noter ici que nous utilisons l'« approximation de squelette raide » dans le calcul de cette distribution, et cette approximation ne nous donne accès ici qu'à un squelette local (en omettant des points qui devrait se trouver sur le squelette). L'extraction d'un squelette par **DisPerSE** nous donne en revanche un squelette global sous la forme d'un réseau connexe. La longueur totale d'un squelette **DisPerSE** est donc plus grande que  $L^{\text{tot}}$ . Par la suite, quand nous souhaiterons comparer ces deux résultats, nous multiplierons la mesure **DisPerSE** par un facteur de normalisation (rapport entre les longueurs totales de squelettes mesurés et prédites), comme en discute Gay (2011). Nous verrons dans le Chapitre 6 que cette re-normalisation n'impacte pas les formes des distributions, qui coïncident par ailleurs.

# CHAPITRE 4

## RECONSTRUCTION DES TEMPS DE RÉIONISATION À PARTIR DU SIGNAL 21 CM

*Les travaux présentés dans ce Chapitre n'ont pas été réalisés en premier pendant ma thèse, mais ils servent de point de départ à toutes les études présentées par la suite. Je présente en effet ici un des intérêts majeurs de travailler avec le champ de temps de réionisation, à savoir une possible reconstruction de ce champ à partir de futures observables de la Réionisation. Vers la fin de la décennie, SKA pourrait nous fournir des images 2D du signal 21 cm sur le plan du ciel et à plusieurs fréquences, et nous montrons dans ce Chapitre que nous pouvons nous en servir pour reconstruire ce qui nous intéresse, les temps de réionisation. Ces reconstructions, bien qu'elles dépendent du modèle de Réionisation, pourront nous permettre à l'avenir d'extraire des informations sur la physique du processus de réionisation (voir les Chapitre 5 et 6). Cette étude est ainsi fondamentale dans le déroulé de mon projet de thèse et c'est pour cela que nous commençons par celle-ci. Elle a été faite en collaboration avec Julien Hiegel, qui a développé le réseau de neurones et certains outils permettant de vérifier nos résultats. J'ai contribué à cette étude en générant les simulations sur lesquelles le réseau a été entraîné, en participant aux entraînements du réseau et au développement de certains outils de vérification des résultats. Les résultats détaillés dans ce Chapitre sont extraits de l'article « Reionisation time fields reconstruction from 21 cm signal maps », présenté en dernière Section et dont je suis deuxième autrice.*

---

## Sommaire

---

<b>4.1</b>	<b>Introduction</b>	<b>80</b>
<b>4.2</b>	<b>Simulations 21cmFAST</b>	<b>81</b>
<b>4.3</b>	<b>Réseau de neurones convolutionnel</b>	<b>83</b>
4.3.1	Architecture du réseau	83
4.3.2	Procédé d'apprentissage	85
4.3.3	Métrie interne au réseau	86
<b>4.4</b>	<b>Reconstructions des temps de réionisation et tests des performances</b>	<b>88</b>
4.4.1	Exemples de carte reconstruite	88
4.4.2	Évaluation pixel par pixel des reconstructions	89
4.4.3	Spectres de puissances	92
4.4.4	Fractions de volume neutre	92
<b>4.5</b>	<b>Premières reconstructions des temps de réionisation à partir de cartes du signal 21 cm bruitées</b>	<b>93</b>
4.5.1	Ajout de bruit dans les cartes de signal 21 cm	93
4.5.2	Nouvelles reconstructions	95
<b>4.6</b>	<b>Conclusions &amp; perspectives</b>	<b>96</b>
<b>4.7</b>	<b>« Reionisation time fields reconstruction from 21 cm signal maps »</b>	<b>98</b>

---

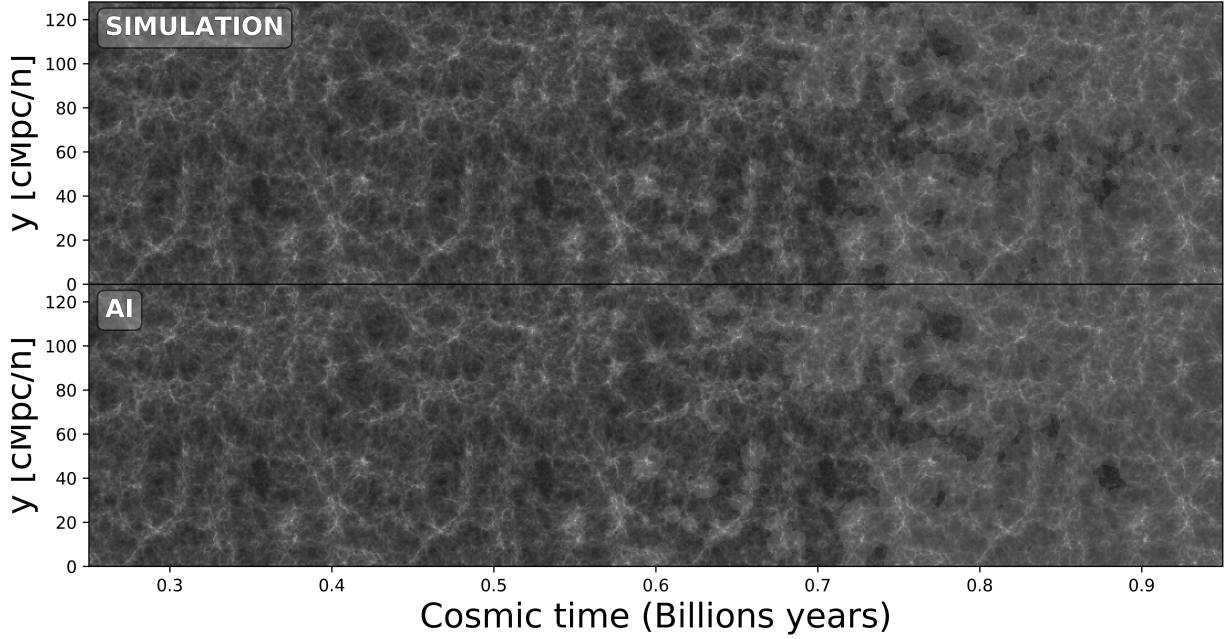


## 4.1 Introduction

Ce Chapitre a pour but de montrer que le champ de temps de réionisation, qui est surtout un outil lié à la modélisation de la Réionisation, peut également être accessible en principe par des observations. Avant la fin de la décennie, le radio-télescope SKA pourrait mesurer des cartes 2D du signal 21 cm sur le plan du ciel à différentes fréquences (ou redshifts) le long de la ligne de visée. Ces cartes sont le résultat de nombreux phénomènes physiques ayant lieu lors de la Réionisation (voir la Section 2.2), et en particulier de l'état d'ionisation du gaz d'hydrogène à cette époque. Elles contiennent ainsi de l'information sur le moment et la manière dont l'hydrogène a été réionisé. On retrouve également ces informations dans une carte de temps de réionisation, qui nous indique le moment d'ionisation de l'hydrogène dans chacun des pixels. Nous nous sommes ainsi demandé dans cette étude si l'on pouvait reconstituer des cartes 2D de temps de réionisation à partir de ces cartes 2D du signal 21 cm.

Dans ce Chapitre, nous nous servons d'un outil de plus en plus utilisé de nos jours, à savoir les réseaux de neurones. Pour étudier la Réionisation, beaucoup d'équipes ont développé ce genre d'outil. Par exemple, Shimabukuro et al. (2022) utilisent un réseau de neurones artificiel (ANN) pour prédire des distributions de taille de bulles ionisées à partir du spectre de puissance du signal 21 cm. Grâce à leurs distributions prédites, on peut ainsi avoir une idée de la morphologie et de l'évolution temporelle de ces bulles. D'autres études s'appuient sur des réseaux de neurones convolutionnels (CNN), comme nous le faisons dans ce Chapitre. Ce type de réseau est souvent utilisé pour faire des prédictions à partir de cartes en deux dimensions. Gillet et al. (2019) peuvent notamment contraindre avec un CNN et à partir de cartes 2D du signal 21 cm des propriétés des galaxies pendant la Réionisation, comme l'émissivité  $\zeta$  des galaxies ou encore la température de viriel  $T_{vir}$ , qui contrôle la température minimale des halos qui déclenchent le processus de formation d'étoiles. Bianco et al. (2021) prédisent par ailleurs des cartes binaires de régions ionisées ou neutres à partir d'images 2D du signal 21 cm. Toujours avec un CNN, mais cette fois-ci à partir de cartes 2D à  $z_{obs} = 6$  de densité d'étoiles et de gaz, Chardin et al. (2019) sont capables de construire des cartes de temps de réionisation. La frise présentée dans la Figure 4.1 est un exemple de leur résultat, et montre une tranche de densité qui évolue pendant la Réionisation, superposée au temps de réionisation (indiquant si l'hydrogène est ionisé ou non à chaque temps). Ces champs sont extraits d'une simulation EMMA en haut, et en bas on y voit le champ de densité d'EMMA et les temps de réionisation reconstruits par le CNN. Dans l'étude présentée ici, nous nous intéressons également à la reconstruction de cartes 2D de temps de réionisation, mais cette fois-ci à partir des futures observations d'images 2D du signal 21 cm avec un CNN. Nous souhaitons ainsi explorer la morphologie et l'évolution de la propagation du rayonnement ionisant par le biais des futures observations du signal 21 cm.

Ce Chapitre commence par décrire les simulations utilisées dans la Section 4.2, et le procédé de reconstruction des cartes  $t_{reion}(\vec{r})$  par réseau de neurones convolutionnel dans la Section 4.3. La Section 4.4 présente ensuite les reconstructions obtenues par le réseau et des métriques généralement utilisées pour tester les performances d'un réseau. Dans la Section 4.5, nous prenons enfin en compte le bruit instrumental généré par le radio-télescope SKA dans les images 21 cm. Nous concluons cette étude et offrons quelques perspectives dans la Section 4.6. La Section 4.7 est dédiée à l'article « Reionisation time fields reconstruction from



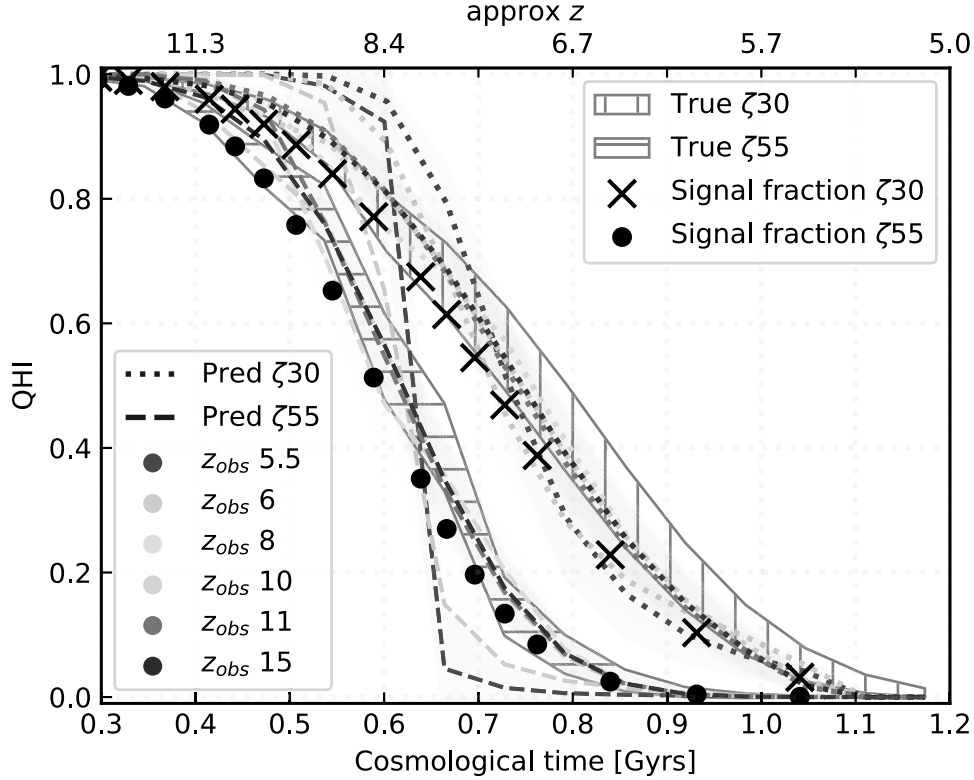
**FIGURE 4.1** – Frises de la densité d’hydrogène (couleurs bleues) superposée à la fraction binaire d’hydrogène ionisé (coloration rose). Le gaz réionisé est ainsi coloré en rose, tandis que les régions encore neutre ne le sont pas. La densité est extraite d’une simulation **EMMA**. En haut, l’information d’ionisation provient du champ de temps de réionisation correspondant produit par **EMMA**, et en bas elle provient de la reconstruction de ces temps faites par le réseau de neurones de Chardin et al. (2019). À l’oeil, les temps de réionisation reconstruits semblent relativement bien correspondre à ceux simulés.

21 cm signal maps » dont les résultats de ce Chapitre sont extraits.

## 4.2 Simulations 21cmFAST

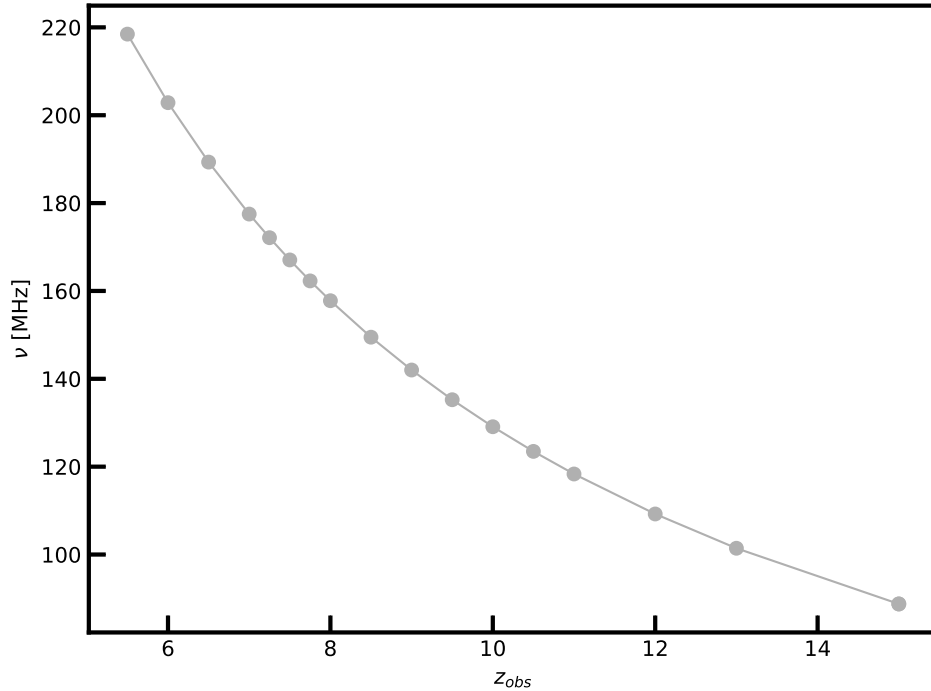
Pour reconstruire le champ de temps de réionisation, nous utilisons ici des modèles semi-analytiques. Ceux-ci sont pratiques lorsque l’on souhaite entraîner un réseau de neurones car un nombre important de cartes est nécessaire et leur temps de calcul est bien plus faible que celui de simulations cosmologiques.

Nous générons alors deux jeux de simulations **21cmFAST** de taille  $256^3 \text{ cMpc}^3/h^3$  et de résolution  $1 \text{ cMpc}^3/h^3$  avec des histoires de réionisation différentes. On fait seulement varier l’émissivité  $\zeta$  des galaxies en créant un jeu de simulations avec  $\zeta = 30$  et un jeu avec  $\zeta = 55$ , que l’on nomme respectivement  $\zeta_{30}$  et  $\zeta_{55}$ . Les histoires de réionisation de ces simulations sont montrées dans la Figure 4.2 en noir. Dans **21cmFAST**, l’émissivité  $\zeta$  contrôle le nombre de photons ionisants qui sont émis par les sources de réionisation : plus elle est grande et plus les sources émettent de photons ionisants, résultant en une réionisation plus précoce. C’est ce que l’on voit avec les courbes de fraction de volume neutre dans la Figure 4.2 : dans la simulation  $\zeta_{30}$ , l’hydrogène reste globalement neutre plus longtemps et réionise alors plus tardivement que dans la simulation  $\zeta_{55}$ .



**FIGURE 4.2** – Fraction de volume d’hydrogène neutre mesurée dans la carte  $t_{\text{reion}}(\vec{r})$  de la simulation  $\zeta_{30}$  en pointillés noir (avec les erreurs représentées par la zone avec les lignes verticales) et dans la carte  $t_{\text{reion}}(\vec{r})$  de la simulation  $\zeta_{55}$  (avec les erreurs représentées par la zone avec les lignes horizontales). Les croix et les points noirs représentent la fraction de signal nul dans les cartes du signal 21 cm aux différents redshift d’observation de la simulation  $\zeta_{30}$  et  $\zeta_{55}$  respectivement. Les pointillés colorés représentent les fractions de volume neutre mesurées dans les cartes de temps de réionisation reconstruites à partir du signal 21 cm observé à différents redshifts et extraits de la simulation  $\zeta_{30}$ . Les tirets colorés sont les fractions équivalentes aux pointillés mais pour la simulation  $\zeta_{55}$ . On voit ici que les histoires de réionisation provenant de cartes de temps de réionisation reconstruites sont proches des vraies histoires de réionisation pour une gamme de redshift comprise entre 8 et 15.

Chaque jeu de simulations contient 50 réalisations du modèle de réionisation correspondant. On extrait de chaque réalisation les cartes de signal 21 cm (ou de température de brillance différentielle  $\delta T_b$ ) à 18 redshifts différents, que l’on appelle par la suite des redshifts d’observation (en référence aux différentes fréquences ou redshifts auxquels on observera l’hydrogène neutre avec SKA). La Figure 4.3 montre les fréquences d’observation en fonction du redshift d’observation. Ces redshifts sont pris entre  $z_{\text{obs}}=5.5$  et  $z_{\text{obs}}=15$  (ou dans une gamme de fréquence comprise entre  $\nu \sim 220$  MHz et  $\nu \sim 90$  MHz) pour représenter une majeure partie de l’histoire de réionisation. On extrait également la carte de temps de réionisation  $t_{\text{reion}}(\vec{r})$  de chaque réalisation (qui est, pour rappel, unique pour chaque simulation). Par ailleurs, les cartes traitées dans le réseau de neurone (qui sera présenté ensuite) sont uniquement à deux dimensions. Nous extrayons alors de chaque cube de données ( $t_{\text{reion}}(\vec{r})$ ) ou



**FIGURE 4.3** – Fréquence en fonction du redshift d’observation. Les points montrent les redshifts auxquels on s’intéresse dans cette étude et auxquels on extrait le signal 21 cm de nos simulations. La gamme de redshifts étudiée ici correspond à une bande de fréquence qui sera observable en théorie par le radio-télescope SKA.

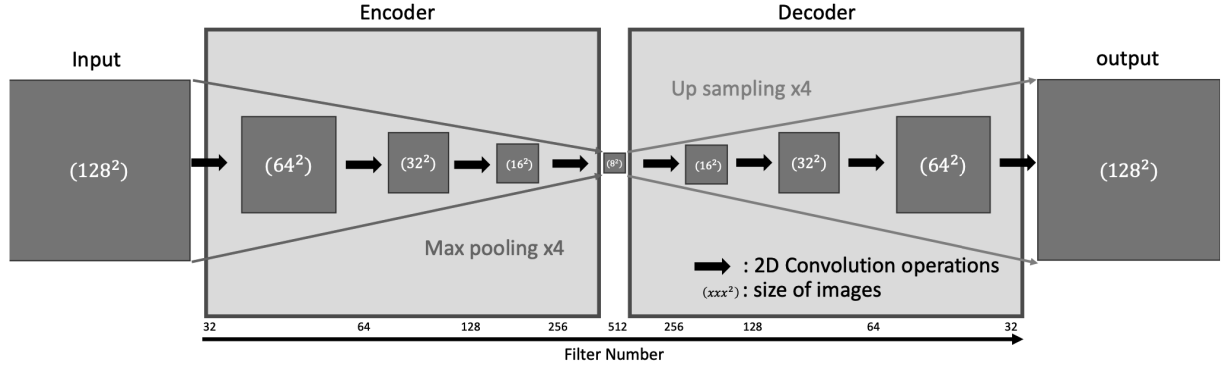
21 cm pour chaque réalisation et modèle) 768 cartes<sup>1</sup> de taille  $128^2 \text{ cMpc}^2/h^2$ , de résolution  $1 \text{ cMpc}^2/h^2$  sur la plan du ciel et d’épaisseur  $1 \text{ cMpc}/h$  le long de la ligne de visée. Pour obtenir la performance maximale des réseaux de neurones, ces cartes sont normalisées en leur soustrayant leur moyenne et en divisant le résultat par leur écart-type. Avec toutes les réalisations de chaque jeu de simulations, on a alors 38 400 cartes de signal 21 cm par redshift et de  $t_{\text{reion}}(\vec{r})$  correspondant.

## 4.3 Réseau de neurones convolutionnel

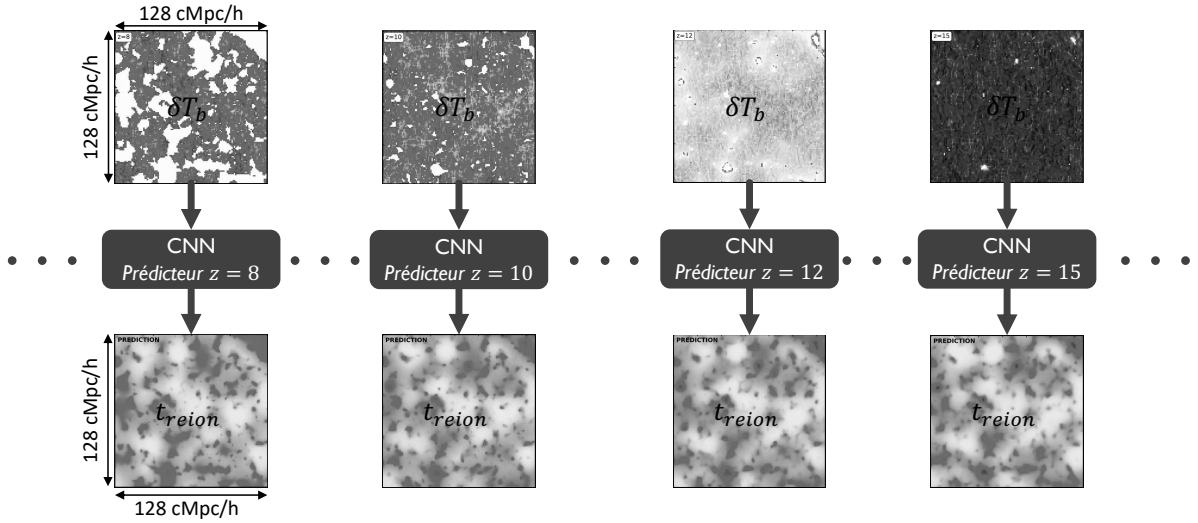
### 4.3.1 Architecture du réseau

Le but de ce Chapitre est de reconstruire les temps de réionisation à partir du signal 21 cm sur le plan du ciel à un redshift donné. Pour cela, nous utilisons un réseau de neurones convolutionnel (CNN), que nous développons grâce aux bibliothèques TensorFlow (Abadi et al. 2015) et Keras, et il prend la forme d’un U-net, représenté dans le schéma de la Figure

1. Pour extraire ces cartes 2D, on en prend 64 espacées de  $4 \text{ cMpc}/h$ , et ce dans chacune des 3 directions des cubes de données. Cela revient à en extraire 768. Ce choix est fait de manière à maximiser le nombre de cartes 2D extraites (plus on donne de cartes au réseau de neurones et mieux il peut s’entraîner), tout en évitant d’avoir des cartes trop ressemblantes parce que trop proches dans l’espace.



**FIGURE 4.4** – Schéma de l’architecture de notre réseau de neurones. Les images d’entrées sont en premier données à l’encodeur, qui les convolue en réduisant leur taille. Le décodeur se charge de « déconvoluer » les données en augmentant leur taille pour reconstruire une image proche de celle attendue en sortie.



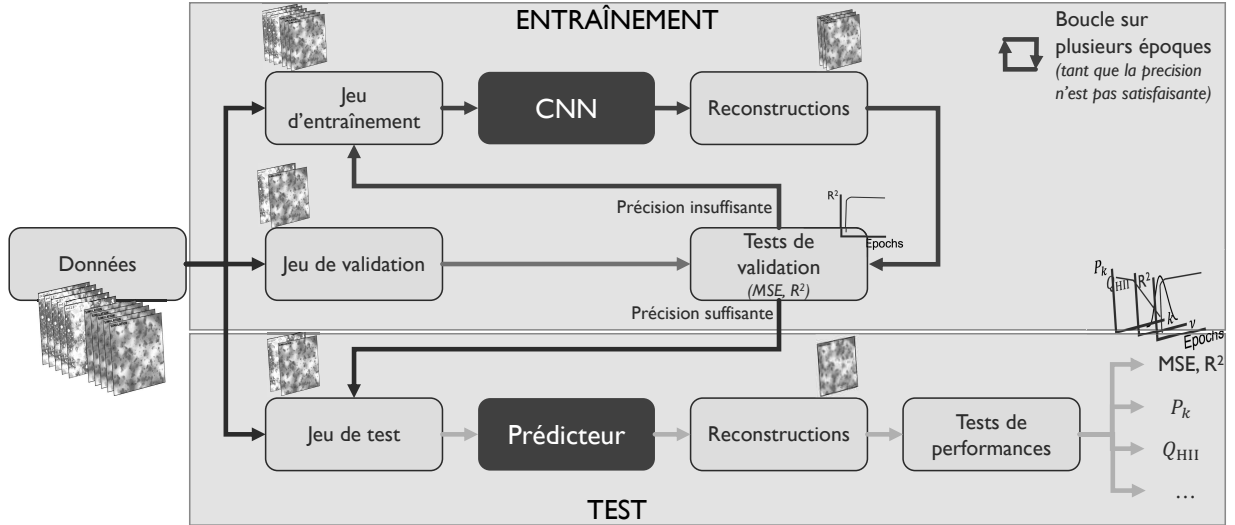
**FIGURE 4.5** – Schéma illustrant la reconstruction d’une carte de temps de réionisation (cartes du bas) à partir d’une carte de signal 21 cm ( $\delta T_b$ , carte du haut) à un redshift d’observation donné. On voit en particulier que nous avons ainsi autant de prédicteurs que de redshifts d’observations (dans les différentes colonnes). Ces cartes sont extraites des simulations 21cmFAST  $\zeta 30$ . Les reconstructions des cartes de temps de réionisation avec les prédicteurs aux différents redshifts devraient être identiques en tout point si les prédictions étaient parfaites.

4.4. Ce genre d’algorithme a été créé par Ronneberger et al. (2015) pour des applications du domaine biomédical. Un U-net est un auto-encodeur composé de deux parties, encodeur et décodeur (voir les deux blocs du schéma de l’architecture de notre U-net, Figure 4.4). À partir de l’image d’entrée (ici une carte  $\delta T_b$ , prise à un redshift donné) et à mesure qu’il descend en profondeur, l’encodeur effectue une série de convolutions 2D et de réductions de la taille de l’image (avec des couches de « max-pooling » qui sous-échantillonnent l’image sans en perdre les propriétés importantes). Les filtres de convolution du réseau sont initialisés pour avoir des poids aléatoires qui sont propres à chaque filtre, et qui sont mis à jour à chaque fois que les données passent dans le réseau, pendant ce qu’on appelle une époque de la phase d’apprentissage. La deuxième partie du réseau, le décodeur, fait le travail inverse : il applique le même nombre de convolutions et de couches d’augmentations de la taille des données (couches « up-sampling »), lui permettant de propager les informations obtenues avec le décodeur. Je laisse ici le lecteur intéressé se référer à l’article de Hiegel et al. (2023), qui est présenté en Section 4.7, pour plus de détails sur l’architecture du réseau. En sortie, l’algorithme est censé reconstruire la carte  $t_{\text{reion}}(\vec{r})$  du modèle correspondant. Il faut noter ici que l’on souhaite entraîner le réseau de neurones sur les cartes de signal 21 cm à un redshift donné afin de reconstruire la carte  $t_{\text{reion}}(\vec{r})$  correspondante, comme le schématise la Figure 4.5. On aura donc 18 réseaux de neurones à entraîner (un par redshift ou fréquence d’observation) pour chaque jeu de simulations, soit 36 en tout en prenant en compte les deux modèles de réionisation.

### 4.3.2 Procédé d’apprentissage

Nous disposons grâce à nos simulations 21cmFAST de 38 400 images (de signal 21 cm par redshift et de  $t_{\text{reion}}(\vec{r})$  correspondant) en tout, et celles-ci sont divisées en deux jeux, correspondant aux deux phases majeures de l’entraînement du réseau de neurones. Ces deux phases sont schématisées dans la Figure 4.6.

La première phase correspond à l’entraînement du réseau (voir le cadre du haut du schéma de la Figure 4.6), et nous réservons pour celle-ci 35 000 images. Pendant cette phase, et pour apprendre à reconstruire une carte  $t_{\text{reion}}(\vec{r})$  proche de celle attendue, l’algorithme doit effectuer un nombre d’époques suffisamment grand pendant lesquelles il va s’améliorer et se rapprocher des vrais images, jusqu’à un niveau de correspondance optimal. Cette phase se divise ainsi en deux étapes. La première est l’étape d’apprentissage, pendant laquelle 31 500 images sont utilisées pour entraîner le CNN et constituent le *jeu d’entraînement*. Pendant cette étape, le réseau s’entraîne avec les images de ce jeu et calcule des fonctions de coût (ici des erreurs quadratiques moyennes, MSE). Le but est de minimiser ces fonctions pour que les reconstructions se rapprochent des vrais cartes  $t_{\text{reion}}(\vec{r})$ , et cela se fait en modifiant les poids des filtres de convolutions. Ensuite, la seconde étape de la phase d’entraînement permet de contrôler les performances du réseau. Pour cela, on fournit au réseau 3 500 images dans un *jeu de validation*, qu’il utilise pour reconstruire des cartes  $t_{\text{reion}}(\vec{r})$ . À partir de ces reconstructions, il calcule des métriques internes, comme la fonction de coût (MSE, sur laquelle nous reviendrons dans la Section suivante) de nouveau ou le coefficient de détermination  $R^2$ . Ces métriques sont ainsi calculées au fil des époques et nous permettent de contrôler l’évolution de l’apprentissage au fil de l’eau, ainsi que de faire une synthèse de ses performances globales à la fin de l’apprentissage. Après ces deux étapes, l’apprentissage est terminé et nous disposons



**FIGURE 4.6** – Schéma expliquant le procédé d'apprentissage et de test du réseau de neurones. Pendant la phase d'entraînement, le CNN apprend sur les données du jeu d'entraînement à reconstruire de bonnes cartes de temps de réionisation et valide les reconstructions à la fin de chaque époque. Tant que les performances ne sont pas satisfaisantes, il effectue de nouvelles époques. Quand elles sont enfin satisfaisantes, on passe dans la phase de test, pendant laquelle on fait des reconstructions à partir du jeu de test (que le CNN n'a encore jamais vu) et du prédicteur (qui est le même CNN avec des poids optimaux des filtres de convolution). Dans cette phase, on effectue des tests de performances sur les reconstructions pour les évaluer.

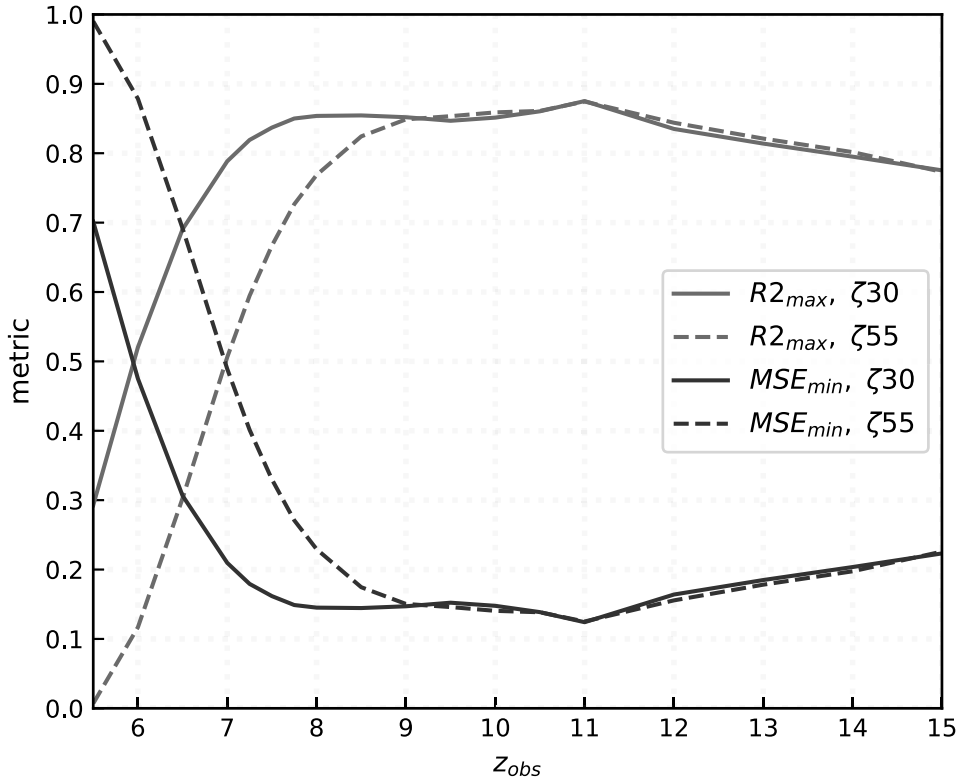
normalement d'un jeu de poids (des filtres de convolution) optimal pour la reconstruction de cartes de temps de réionisation à partir de cartes du signal 21 cm.

Enfin, nous arrivons dans la deuxième phase (voir le cadre du bas du schéma de la Figure 4.6), pour laquelle il nous reste 3 400 images, qui forment le *jeu de test*. Ce jeu n'a ainsi pas encore été vu par le réseau et il permet de tester la capacité du réseau à reconstruire de « bonnes » cartes de temps de réionisation. Nous reconstruisons alors avec les images du signal 21 cm de ce jeu des cartes de temps de réionisation, et ce sont ces cartes qui seront globalement analysées dans les Sections suivantes pour vérifier leur véracité (hormis pour les résultats concernant la fonction de coût ou le  $R^2$ , qui sont eux obtenus à partir du jeu de validation).

### 4.3.3 Métrique interne au réseau

Pendant la phase d'entraînement du réseau, nous avons calculé la fonction de coût MSE et le coefficient de détermination, comme mentionnés dans la Section précédente. On ne détaille par exemple que la MSE ici, dont l'expression est la suivante :

$$MSE = \frac{1}{N_{\text{images}}} \sum_{i=1}^{N_{\text{images}}} (X_i - Y_i)^2, \quad (4.1)$$



**FIGURE 4.7** – Métriques internes au réseau de neurones. La MSE est représentée en bleu et le coefficient de détermination  $R^2$  en rouge. Le modèle de Réionisation  $\zeta_{30}$  est représenté avec des tirets, et le modèle  $\zeta_{55}$  avec des traits pleins. On voit que les reconstructions faites avec un prédicteur à un redshift de 11 sont les meilleures pour les deux modèles.

où  $X$  et  $Y$  sont les images d'entrée (carte du signal 21 cm) et de sortie (carte de temps de réionisation) respectivement.  $N_{images}$  correspond au nombre d'images total à partir desquelles la MSE est calculée. Cette métrique est connue pour l'évaluation de prédicteurs tels que des réseaux de neurones et est d'ailleurs souvent utilisée dans les réseaux. On peut ainsi la calculer à chaque époque de la phase d'apprentissage pour savoir quand converge le réseau de neurones. Le minimum de la MSE nous indique aussi la performance maximale que l'on peut atteindre avec le réseau.

Dans la Figure 4.7 (courbes bleues), nous représentons le minimum de MSE atteint pour chacun des prédicteurs (indiqué par  $z_{obs}$ ) que nous avons entraîné avec des cartes de signal 21 cm prise à un redshift d'observation qui leur est propre. Les deux modèles de réionisation sont représentés en traits pleins pour le modèle  $\zeta_{30}$  et en pointillés pour le modèle  $\zeta_{55}$ . Idéalement, tous les prédicteurs devraient donner la même reconstruction pour un modèle de Réionisation donné parce qu'il n'y a qu'une carte de temps de réionisation pour plusieurs cartes de signal 21 cm à des redshifts d'observation différents. Cependant, on voit que la métrique n'est pas indépendante du redshift, ce qui signifie que les prédicteurs entraînés à différents redshifts d'observation ne sont pas capables de reconstruire les temps de réionisation de la même manière. Pour les plus bas redshifts, on remarque notamment que les performances sont vraiment mauvaises, avec des valeurs de MSE élevées. Ceci est certainement dû au fait qu'à



bas redshifts, tout l’hydrogène est réionisé, réduisant le signal 21 cm à zéro. Les meilleures performances (MSE les plus bas) sont obtenues avec des prédicteurs entraînés à des redshifts d’observation entre 10 et 12, le meilleur prédicteur étant celui à  $z_{\text{obs}} = 11$  pour les deux modèles. On suppose que cette gamme de redshifts d’observation est celle qui fournit le plus d’informations à la fois sur les bulles ionisées et sur les régions neutres. Par exemple, les cartes de signal 21 cm à  $z_{\text{obs}} = 11$  contiennent 95% de pixels avec un signal non-nul, comme le montre la Figure 4.2 avec les courbes de fraction de signal (croix et ronds), et on peut voir en blanc sur la carte de signal 21 cm à gauche de la Figure 4.8 les premières régions ionisées qui apparaissent. À ce redshift d’observation, les cartes de signal 21 cm contiennent alors des informations sur la position d’une grande partie des sources de réionisation, ainsi que sur toutes les régions encore neutre qui produisent le signal 21 cm. Après ces redshifts, la MSE augmente de nouveau, ce qui donne des performances moins bonnes. En effet, à ces redshifts, comme l’hydrogène n’est en moyenne pas très réionisé, le réseau manque d’informations sur les sources d’ionisation pour localiser correctement les minima de  $t_{\text{reion}}(\vec{r})$  et produire de bonnes reconstructions de ces cartes.

Par ailleurs, on remarque aussi que les performances du réseau de neurones varient selon les modèles de Réionisation. Ces variations semblent surtout présentes à bas redshifts d’observation, n’impactant pas la fenêtre de redshift dans laquelle nous obtenons les meilleures prédictions. On voit donc que les prédicteurs du modèle  $\zeta 55$  sont bien moins performants plus rapidement (dès que  $z_{\text{obs}} < 8$ ) alors que le modèle  $\zeta 30$  est moins performant à partir de  $z_{\text{obs}} < 7$ . Nous supposons que cela provient du fait que le modèle  $\zeta 55$  achève la Réionisation plus tôt que le modèle  $\zeta 30$ , avec une fraction d’ionisation de 50% de l’hydrogène vers  $z_{\text{obs}} \sim 8$  pour ce modèle contre  $z_{\text{obs}} \sim 7$  pour le modèle  $\zeta 30$  (voir la Figure 4.2), et ceci impacte les cartes de signal 21 cm en supprimant le signal plus tôt. On remarque donc ici une forte corrélation entre le déroulé de la Réionisation et les performances de notre CNN.

## 4.4 Reconstructions des temps de réionisation et tests des performances

Nous disposons à présent d’un réseau de neurones entraîné sur plusieurs jeux de données différents, un par redshift et par modèle. Dans cette Section, je détaille les résultats des tests de performances effectués sur les reconstructions de ces réseaux obtenues à partir du jeu de test sur lequel le réseau ne s’est pas entraîné et qui lui est donc inconnu. Ces tests sont les suivants : comparaison visuelle entre une reconstruction et une vraie carte de temps de réionisation, histogrammes des vrais pixels versus ceux reconstruits, fraction de pixels correctement reconstruits, spectre de puissance et fraction de volume neutre. D’autres tests sont étudiés dans Hiegel et al. (2023), comprenant la distribution normalisée de la différence entre les valeurs des cartes  $t_{\text{reion}}(\vec{r})$  et celles reconstruites, le paramètre de Dice et la distribution des valeurs de  $t_{\text{reion}}(\vec{r})$  à ses minima. Pour rester brève, je ne les détaillerai pas ici.

### 4.4.1 Exemples de carte reconstruite

Nous pouvons tout d’abord regarder l’aspect des cartes de temps de réionisation reconstruites avec le CNN et les comparer visuellement aux cartes originelles. La Figure 4.8 montre

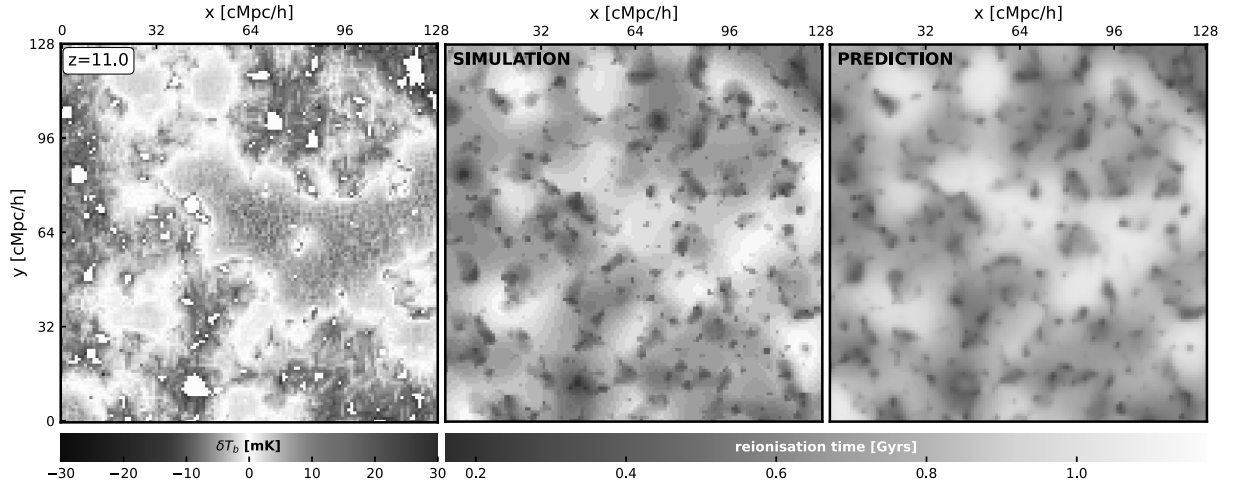
un exemple de cartes de signal 21 cm à gauche, et de temps de réionisation au centre extraites d'une même simulation (dont le modèle est le  $\zeta 30$ ), et accompagnées de la reconstruction de  $t_{\text{reion}}(\vec{r})$  à droite. Cette reconstruction est faite à partir d'une carte de 21 cm prise à un redshift d'observation  $z_{\text{obs}} = 11$ , c'est-à-dire faite avec le meilleur prédicteur que nous disposons. À l'oeil, nous pouvons déjà voir que la reconstruction semble proche de la vraie carte  $t_{\text{reion}}(\vec{r})$  à ce redshift d'observation : les mêmes structures semblent s'y trouver. Néanmoins, on peut aussi voir que les petites structures ne sont pas reconstruites, la reconstruction de  $t_{\text{reion}}(\vec{r})$  semblant être lissée.

#### 4.4.2 Évaluation pixel par pixel des reconstructions

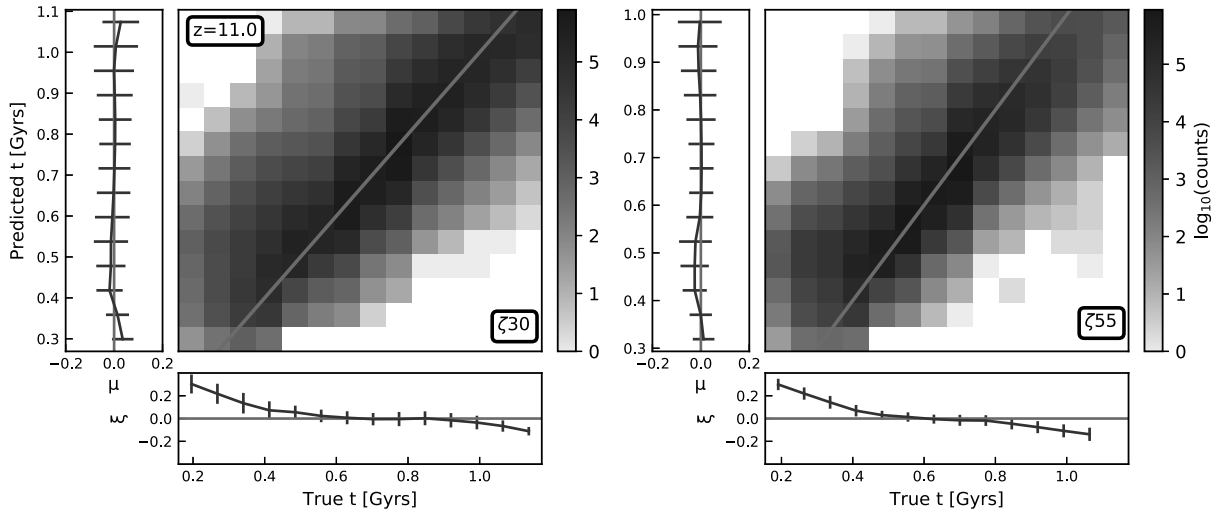
Une métrique que l'on retrouve souvent dans l'analyse des performances d'un réseau de neurones est ce que l'on appelle « true versus predicted » (« vrai versus prédit », que l'on note TvP). C'est un histogramme des valeurs des pixels des cartes reconstruites en fonction de ceux de la carte originelle et je les montre dans la Figure 4.9 pour le modèle  $\zeta 30$  à gauche et  $\zeta 55$  à droite. Les reconstructions proviennent du prédicteur  $z_{\text{obs}} = 11$  afin de connaître les performances du meilleur prédicteur. Dans ces histogrammes, on a pris les pixels de toutes les cartes reconstruites pour avoir la meilleure statistique possible. Les lignes rouges représentent la corrélation parfaite entre les vrais pixels et ceux reconstruits. On voit avec ces histogrammes que la majorité des pixels sont en effet très proches de cette corrélation parfaite (à 0.2 Gyrs).

Les histogrammes du bas et de gauche montrent les moyennes et écart-types des résidus  $r = p_{\text{prédit}} - p_{\text{vrai}}$  (où  $p$  est la valeur d'un pixel) dans les directions verticales et horizontales respectivement. L'historgramme du bas permet donc d'avoir accès à l'erreur d'apprentissage, tandis que l'historgramme de gauche permet de connaître l'incertitude de reconstruction. En pratique, seul le second estimateur est accessible parce que le premier dépend des vraies cartes dont nous ne disposerons généralement pas. Avec ces histogrammes des résidus, on remarque cependant que les valeurs extrêmes de  $t_{\text{reion}}(\vec{r})$  sont moins bien reconstruites (avec des résidus plus grands que pour les autres valeurs), et c'est en particulier le cas des valeurs à bas temps de réionisation. Les bas temps de réionisation correspondent aux endroits qui réionisent en premier, et on les assimile donc aux sources de première réionisation. Ces endroits à bas temps sont rares et correspondent aussi à un signal 21 cm  $\delta T_b$  nul, ce qui explique que le réseau échoue à les reconstruire correctement. Sur la reconstruction montrée dans la Figure 4.8, on peut aussi voir que ces minima sont lissés (voir les régions rouges foncées), ce qui est cohérent avec la mesure des résidus présentées ici.

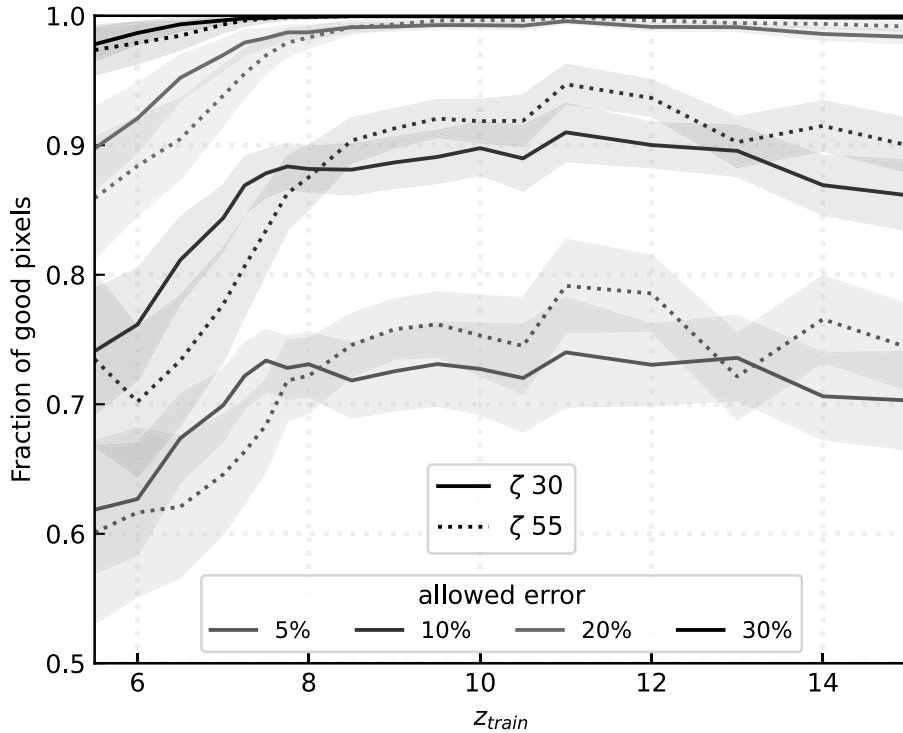
Une manière de synthétiser les informations présentes dans tous l'historgramme TvP obtenu avec les prédicteurs pour chaque redshift d'observations est de mesurer la fraction d'accord montrée dans la Figure 4.10. Cette fraction indique entre 0 et 1 si les pixels reconstruits le sont correctement à un certain pourcentage d'acceptation par rapport aux vrais pixels. Logiquement, plus le pourcentage d'erreurs autorisés est grand et plus on va valider des pixels reconstruits. Notons déjà que les prédicteurs des deux modèles présentent ici des performances similaires. De plus, les reconstructions avec les prédicteurs à bas  $z_{\text{obs}} (< 8)$  sont particulièrement mauvaises par rapport aux autres, avec par exemple 75% de pixels corrects quand on autorise une erreur de 10% à  $z_{\text{obs}} = 5.5$ , contre presque 90% à  $z_{\text{obs}} > 8$ . Ceci se comprend par le fait que les cartes  $\delta T_b$  à bas redshifts contiennent de moins en moins de si-



**FIGURE 4.8** – Exemple de reconstruction (à droite) faite par le CNN à partir d’une carte de signal 21 cm prise à un redshift d’observation  $z_{\text{obs}} = 11$  (à gauche). La carte au centre est la « vraie » carte de temps de réionisation extraite de la simulation, et le modèle de réionisation est ici le  $\zeta_{30}$ . La carte de temps de réionisation reconstruite ressemble à l’œil à la carte simulée, malgré un lissage apparent qui élimine les structures des petites échelles.

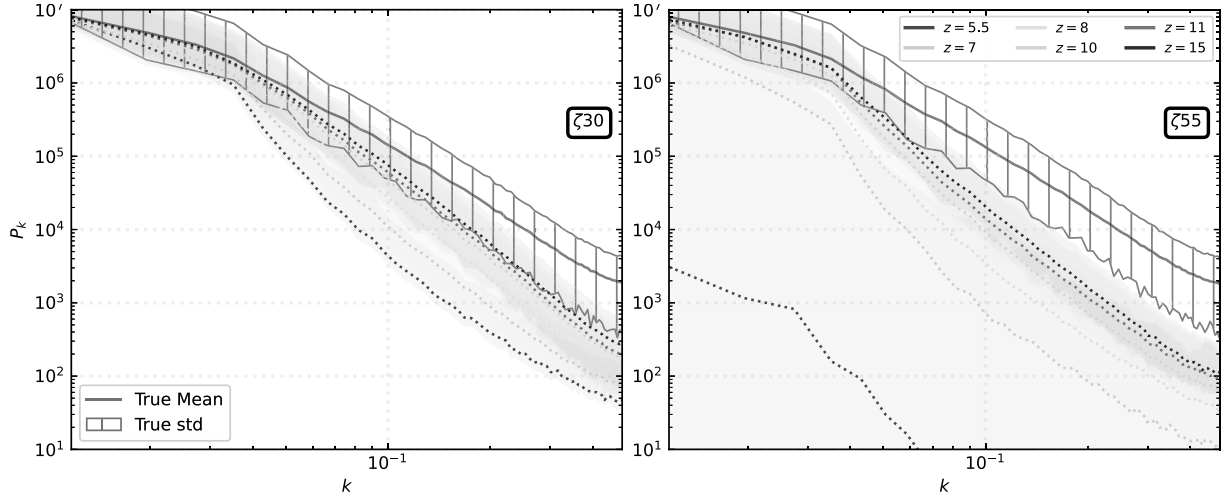


**FIGURE 4.9** – Histogrammes des vrais pixels versus ceux des cartes reconstruites avec le prédicteur  $z_{\text{obs}} = 11$  du modèle  $\zeta_{30}$  à gauche et  $\zeta_{55}$  à droite. Les lignes rouges représentent la corrélation parfaite entre les vrais pixels et ceux reconstruits. Les histogrammes du bas et de gauche montrent les moyennes et écart-types des résidus  $r = p_{\text{predict}} - p_{\text{vrai}}$  (où  $p$  est la valeur d’un pixel) dans les directions verticales et horizontales respectivement. L’histogramme du bas permet donc d’avoir accès à l’erreur d’apprentissage, tandis que l’histogramme de gauche permet de connaître l’incertitude de reconstruction. En pratique, seul le second estimateur est accessible parce que le premier dépend des vrais cartes dont nous ne disposerons généralement pas. On voit ici que les histogrammes suivent l’accord parfait donné par la ligne rouge, avec un écart notamment présent aux temps précoces et tardifs de réionisation.



**FIGURE 4.10** – Fraction d'accord en fonction du redshift d'observation. En couleur, différentes erreurs autorisées sont représentées. Plus ces erreurs autorisées sont grandes, plus on va compter de pixels des cartes reconstruites comme étant correctement reconstruits à ce pourcentage d'erreur près. Les traits pleins sont les mesures dans le modèle  $\zeta 30$ , et les pointillés sont celles du modèle  $\zeta 55$ . Par exemple, on en conclut que lorsque 20% d'erreur est autorisé, plus de 90% des pixels sont correctement reconstruits pour le modèle  $\zeta 30$  pour tous les redshifts d'observation, et plus de 95% des pixels sont correctement reconstruits pour ce même modèle avec des prédicteurs ayant un redshift supérieur à 8.

gnal ( $< 50\%$ ) parce que l'hydrogène est de plus en plus réionisé. À  $z_{\text{obs}} = 5.5$ , quasiment tout l'hydrogène est réionisé, justifiant ainsi la plus basse fraction d'accord. Les meilleures performances mesurées avec cet estimateur sont donc obtenues avec des prédicteurs à  $z_{\text{obs}} > 8$ . Avec cette métrique, on a également l'impression que tous les prédicteurs à  $z_{\text{obs}} > 9$  ont des performances similaires, avec une légère décroissance à mesure que  $z_{\text{obs}}$  augmente. On retrouve les mêmes conclusions que nous avons tirées précédemment : les prédicteurs aux redshifts d'observation extrêmes ont de moins bonnes performances. Plus précisément, les prédicteurs à bas redshifts n'ont pas assez de signal et ceux à hauts redshifts manquent de l'empreinte des sources de réionisation. Un compromis se trouve alors dans les redshifts intermédiaires, entre  $z_{\text{obs}} = 8$  et  $z_{\text{obs}} = 12$ , auxquels on peut faire les meilleures reconstructions (avec une fraction d'accord de près de 90% quand 10% d'erreur est acceptée).



**FIGURE 4.11** – Spectres de puissance des cartes  $t_{\text{reion}}(\vec{r})$ , avec le modèle  $\zeta_{30}$  à gauche et  $\zeta_{55}$  à droite. Les spectres de puissances des vraies cartes des simulations sont en noir, tandis que ceux des reconstructions sont en couleur pour différents redshifts d’observation. Les courbes sont des médianes faites sur toutes les cartes disponibles. Les zones colorées et rayées de noires représentent la dispersion autour de la médiane (1<sup>er</sup> et 99<sup>ème</sup> quantiles). Avec ces courbes, et pour la majorité des prédicteurs, on voit de nouveau que les structures à grande échelle sont bien reconstruites, et que celles à petite échelle ne le sont pas.

### 4.4.3 Spectres de puissances

Une statistique qui nous intéresse particulièrement concernant le champ de temps de réionisation, et que nous utiliserons davantage dans le Chapitre 6, est le spectre de puissance du champ. Il nous permet aussi ici d’analyser les performances de la reconstruction par le CNN aux différentes échelles. Nous comparons donc le spectre de puissance mesuré dans les vraies cartes de  $t_{\text{reion}}(\vec{r})$ , à celui extraits des reconstructions avec les CNN. Ceux-ci sont montrés dans la Figure 4.11 pour les modèles  $\zeta_{30}$  et  $\zeta_{55}$  à gauche et à droite respectivement. Pour les deux modèles, nous pouvons voir que les spectres de puissance des cartes reconstruites (en couleur) sont assez proches des vrais spectres de puissance (en noir) aux grandes échelles ( $k < 0.06$  h/cMpc environ), hormis pour les cartes reconstruites aux plus bas redshifts du modèle  $\zeta_{55}$ . Cependant, dès que  $k > 0.06$  h/cMpc, on voit que les spectres de puissances des cartes reconstruites décroissent plus rapidement que les spectres originels. Cela est attendu avec la perte des petites échelles que nous pouvions voir dans la carte reconstruite de la Figure 4.8. On remarque aussi que plus le redshift d’observation est grand et plus les spectres de puissance des cartes reconstruites se rapprochent des vrais spectres. Pour les redshifts  $z_{\text{obs}} > 11$  du modèle  $\zeta_{30}$ , les spectres de puissance sont encore compatibles avec le vrai spectre.

### 4.4.4 Fractions de volume neutre

À partir d’une carte de temps de réionisation, on a accès à l’histoire de réionisation, ou plus précisément ici la fraction de volume d’hydrogène neutre. En effet, le nombre de

cellules ayant un temps de réionisation plus grand qu'un temps seuil correspond directement au nombre de cellules dont l'hydrogène est encore neutre. On peut donc définir la fraction de volume neutre comme suit :

$$Q_{\text{HI}}(\nu) = \frac{N(> \nu)}{N_{\text{total}}}, \quad (4.2)$$

avec  $N_{\text{total}}$  le nombre total de cellules dans la carte considérée. On peut donc mesurer ces fractions dans les vraies cartes, ainsi que dans les cartes reconstruites avec les différents prédicteurs, comme le montre la Figure 4.2. Les zones noires correspondent aux mesures de  $Q_{\text{HI}}$  dans les vraies cartes de chaque modèle, et les tirets et les pointillés en couleur correspondent aux mesures dans les cartes reconstruites avec des prédicteurs à différents redshifts pour le modèle  $\zeta 55$  et  $\zeta 30$  respectivement. On voit que les mesures dans les cartes reconstruites sont relativement bien reconstruites pour les deux modèles, et sont d'autant plus proches des vraies histoires de réionisation que le redshift d'observation est élevé. De nouveau, on retrouve le fait qu'à bas redshifts, les reconstructions ne sont pas bien faites et que pour des redshifts  $z_{\text{obs}} > 8$ , les reconstructions sont proches des cartes attendues.

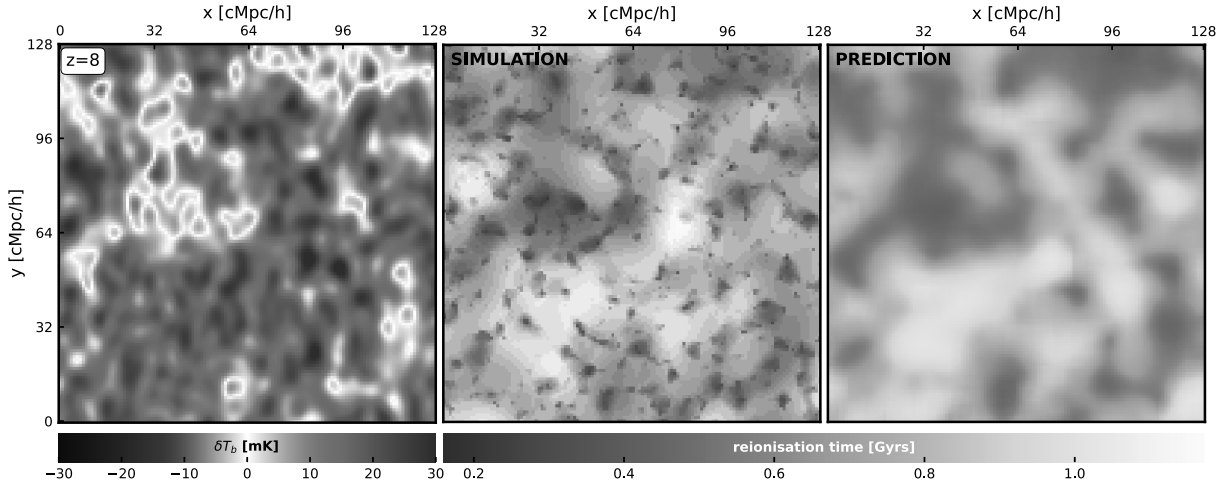
Par ailleurs, on rappelle ici qu'une reconstruction obtenue avec un prédicteur  $z_{\text{obs}} = 11$  par exemple est faite à partir de carte de signal 21 cm observée à un redshift de 11. On montre ici qu'à ce redshift, nous sommes capables d'accéder à une histoire de réionisation assez véridique grâce aux cartes  $t_{\text{reion}}(\vec{r})$  reconstruites. On en déduit ainsi qu'avec les observations futures sur le plan du ciel à un seul redshift (ou une seule fréquence, par exemple  $z_{\text{obs}} = 11$  correspondant à une fréquence  $\nu \sim 120$  MHz), on sera capable de reconstruire toute l'histoire de réionisation, et ainsi de reconstruire l'histoire de réionisation passée et d'extrapoler celle à venir dans la région du ciel observée.

## 4.5 Premières reconstructions des temps de réionisation à partir de cartes du signal 21 cm bruitées

Les observations du signal 21 cm sous forme de cartes 2D sur le plan du ciel à plusieurs fréquences (ou redshifts) produites par SKA ne seront pas dénuées de bruit. Plusieurs types de bruit peuvent être présents, comme le bruit instrumental ou les avant-plans (Koopmans et al. 2015; Giri et al. 2018b; Bianco et al. 2021). Dans cette Section, nous menons une étude préliminaire en incluant du bruit instrumental (uniquement pour le moment) à nos cartes de signal 21 cm dans le but de nous rapprocher des données SKA. Ces observations ont aussi une résolution limite que nous allons prendre en compte ici. Nous nous attendons à ce que ces limites observationnelles affectent les reconstructions de  $t_{\text{reion}}(\vec{r})$  et nous estimons dans cette première étude leur impact.

### 4.5.1 Ajout de bruit dans les cartes de signal 21 cm

Pour cette étude, nous créons un nouveau jeu de données avec 38 400 cartes de signal 21 cm (du modèle  $\zeta 30$  à  $z_{\text{obs}} = 8$ ) bruité et mises à la résolution que nous espérons avoir avec SKA, dont un exemple est donné à gauche dans la Figure 4.12. Nous implémentons ce bruit en prenant en compte la couverture dans le plan uv (qui est montrée dans la Figure 2.12) et la résolution de SKA grâce à la bibliothèque `Python Tools21cm` (Giri et al. 2020).

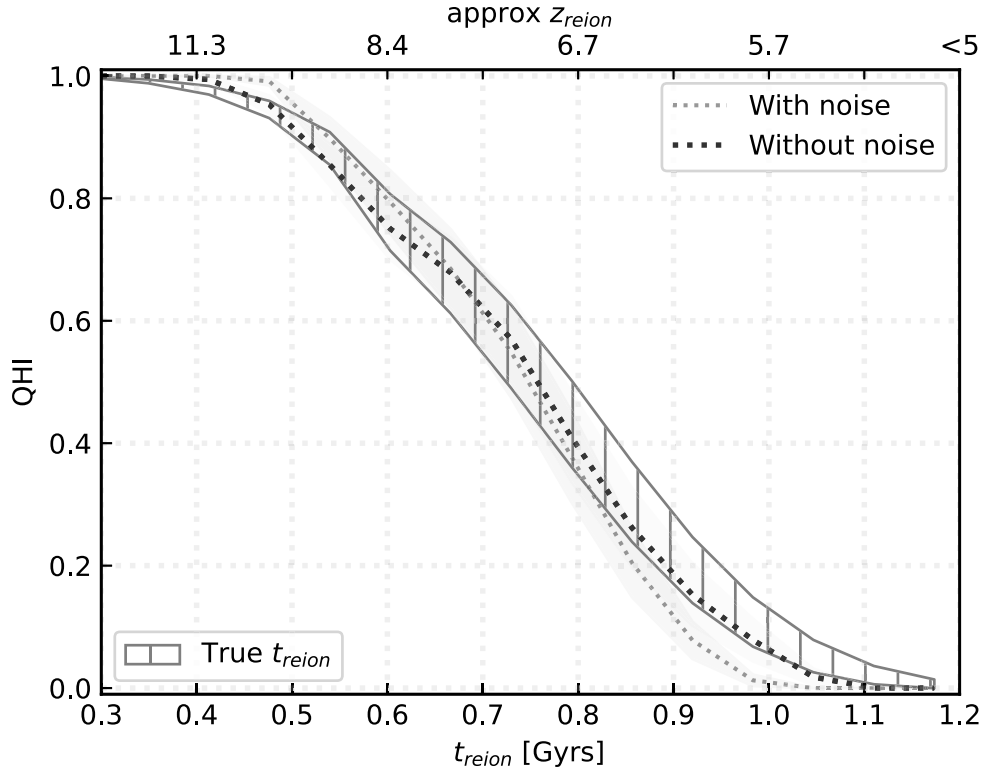


**FIGURE 4.12** – Exemple de reconstruction (à droite) faite par le CNN à partir d’une carte de signal 21 cm bruité prise à un redshift d’observation  $z_{\text{obs}} = 8$  (à gauche). La carte au centre est la « vraie » carte de temps de réionisation extraite de la simulation, et le modèle de réionisation est ici le  $\zeta 30$ . On voit qu’en présence de bruit dans les cartes de signal 21 cm, d’autant plus de structures aux petites échelles ne sont pas reconstruites, donnant un effet de lissage plus important. Malgré cela, les structures aux grandes échelles semblent bien conservées à l’œil.

En reprenant les chiffres usuels, nous supposons ici une durée d’observation par jours de 6 heures, 10 secondes de temps d’intégration, un total de 1000 heures d’observation, et une ligne de base de 2 km (Giri et al. 2018b; Bianco et al. 2021; Prelogović et al. 2022). À notre redshift d’observation de  $z_{\text{obs}} = 8$ , la résolution angulaire est d’environ  $\Delta\theta \sim 2.8$  arcmin, soit 7.35 cMpc sur le plan du ciel. Pour la résolution dans la direction des fréquences, `Tools21cm` nous permet d’appliquer un filtre tophat et d’avoir une résolution de  $\Delta\nu \sim 0.43$  MHz, soit 7.35 cMpc également.

Nous nous concentrons ici sur le redshift d’observation  $z_{\text{obs}} = 8$  comme mentionné ci-avant parce que plus le redshift d’observation est élevé et plus il y aura de bruit intégré dans les données d’observation. En effet, nous nous sommes rendus compte que les reconstructions à partir de cartes de signal 21 cm à  $z_{\text{obs}} = 11$  (soit au redshift qui donnait les meilleures reconstructions quand nous n’ajoutions pas de bruit) étaient bien moins bonnes. Nous préférons alors commencer à travailler avec des cartes à un redshift d’observation qui donne tout de même de bonnes reconstructions sans bruit tout en essayant de minimiser l’impact de l’ajout du bruit. Bien sûr, pour parfaire cette étude, il faudrait à l’avenir refaire le même travail à d’autres redshifts d’observations.

Nous utilisons donc le nouveau jeu de cartes de signal 21 cm bruité, ainsi que les cartes  $t_{\text{reion}}(\vec{r})$  correspondant (qui sont inchangées par rapport aux études précédentes), pour entraîner de nouveau un réseau de neurones qui a la même architecture que celle présentée dans la Figure 4.4. Ceci nous donne donc un prédicteur à  $z_{\text{obs}} = 8$  de cartes de temps de réionisation à partir d’observations bruitées, et grâce auxquels nous pouvons faire de nouvelles reconstructions de  $t_{\text{reion}}(\vec{r})$ .



**FIGURE 4.13** – Histoires de réionisation extraites de carte de temps de réionisation provenant des simulations 21cmFAST (en noir), ainsi que de reconstructions de  $t_{\text{reion}}(\vec{r})$  à partir du signal 21 cm (en bleu) et du signal 21 cm bruité (en rouge). Les zones colorées et rayées représentent la dispersion autour de la médiane (1<sup>er</sup> et 99<sup>ème</sup> quantiles). On voit que l’histoire de réionisation de la reconstruction est toujours proche de la vraie histoire de réionisation de la simulation, bien qu’elle soit plus rapide.

## 4.5.2 Nouvelles reconstructions

La vérification des performances avec les métriques internes du réseau est faite de la même manière que pour les autres réseaux mentionnés auparavant, et nous ne la détaillons pas ici. Nous nous concentrons au contraire sur deux résultats spécifiques, à savoir un exemple de cartes reconstruites à partir de ce nouveau réseau (montré à droite dans la Figure 4.12), et l’histoire de réionisation que l’on peut extraire des reconstructions (montrée en bleu dans la Figure 4.13). Nous pouvons tout d’abord voir sur la carte reconstruite de la Figure 4.12 que quand on regarde aux grandes échelles, on retrouve les mêmes structures que dans la vraie carte simulée. Cependant, nous remarquons que la carte reconstruite est encore plus dépourvue de petites structures que quand il n’y avait pas de bruit (voir par exemple la carte de droite de la Figure 4.8). Le bruit ajouté et la mise à la résolution de SKA semble ainsi lissée encore plus les reconstructions qu’auparavant sans bruit.

En regardant les histoires de réionisation reconstruites avec et sans bruit instrumental (en rouge et bleu dans la Figure 4.13), on voit néanmoins que les reconstructions à partir de données bruitées reconstruisent une histoire de réionisation correcte, même si elle est



légèrement plus abrupte en début et en fin de Réionisation. Nous pouvons donc en conclure que les reconstructions à partir de données bruitées contiennent toujours de l'information sur le déroulé global de la Réionisation.

Nous n'avons pas fait d'analyses plus poussées encore à ce stade, mais les résultats que nous obtenons avec les données bruitées sont prometteurs de mon point de vue. Ces résultats pourraient bien sûr être améliorés en changeant par exemple l'architecture du réseau de neurones (en ajustant les hyperparamètres, ou bien en modifiant le nombre de couches, de filtres ou la profondeur du réseau). Donner plus d'informations en entrée du réseau pourrait aussi lui permettre de faire des reconstructions plus précises. Par exemple, on pourrait ajouter en entrée du réseau la carte du plan uv montrée dans la Figure 2.12, ce qui lui permettrait notamment de donner un poids plus important aux régions du ciel qui seront effectivement observées pour qu'il se concentre sur celles-ci. Nous pourrions également utiliser d'autres types d'algorithmes tels que les réseaux adversariaux génératifs (GANs, voir Ullmo et al. (2021)), connus pour avoir de bonnes performances pour les petites structures. Par ailleurs, nous pourrions à l'inverse partir du principe que les données d'observation de SKA seront déjà traitées et que nous aurons accès à des données avec un minimum de bruit, ce qui nous permettrait d'avoir les meilleures reconstructions possibles indépendamment du réseau de neurones.

## 4.6 Conclusions & perspectives

Dans ce Chapitre, nous avons construit une méthode permettant de reconstruire des cartes 2D de temps de réionisation  $t_{\text{reion}}(\vec{r})$  à partir de cartes 2D de signal 21 cm  $\delta T_{\text{b}}$ , non-bruité et bruité, prises à un redshift d'observation  $z_{\text{obs}}$ . La méthode se base sur un réseau de neurones convolutionnel (CNN) qui a la forme d'un U-net. C'est un auto-encodeur, qui est donc composé d'une partie encodeur faisant des convolutions et des diminutions de la taille des données, et d'un décodeur qui fait le travail inverse pour reconstituer la carte  $t_{\text{reion}}(\vec{r})$ . Nous avons également vérifié les performances de ce réseau par le biais des métriques internes au réseau, comme les erreurs des moindres carrés (MSE), et d'autres statistiques. Nous en avons conclu que les reconstructions à partir de cartes de signal 21 cm non bruitées sont satisfaisantes bien que lissées par rapport aux cartes originelles. De plus, il existe une fenêtre de redshifts (entre  $z_{\text{obs}} = 8$  et  $z_{\text{obs}} = 12$ ) pour lesquelles les reconstructions sont correctes. Dans cette fenêtre, nous sommes capables de calculer l'histoire de réionisation du modèle, et celle-ci correspond bien à celle attendue. Cela signifie qu'à partir de cartes de signal 21 cm à *un redshift d'observation* donné, par exemple 8 ou 11, avec notre méthode, nous avons accès à toute l'histoire de réionisation. Nous pouvons ainsi reconstituer le passé par rapport à ce redshift et extrapoler le futur de la portion du ciel observée. Par ailleurs, les reconstructions à partir des cartes de signal 21 cm bruité observée à  $z_{\text{obs}} = 8$  correspondent relativement bien aux cartes attendues aux grandes échelles mais sont d'autant plus lissées que les précédentes. Cela dit, cela ne les empêche pas de contenir l'information nécessaire au calcul des histoires globales de réionisation, qui sont proches de celles attendues. Ainsi, même si ces reconstructions manquent d'autant plus de petites échelles, on peut quand même en extraire de l'information.

Ces reconstructions ne sont pas encore parfaites et il pourrait être intéressant de les

améliorer par la suite (notamment pour les petites échelles perdues). Nous pourrions changer l’architecture du réseau, ou bien changer de type de réseau et utiliser par exemple un GAN, qui est connu pour être efficace aux petites échelles. Une autre piste pour reconstruire également les petites échelles est d’utiliser des jeux de données avec des cartes zoomées (en 32 cMpc/h par exemple) pour lesquelles on augmente la résolution (jusqu’à obtenir des cartes  $128^2$ ).

Nous tenons à noter ici que parce que nous utilisons un CNN, nous sommes limités en terme de taille des cartes. Plus celles-ci sont grandes et plus l’entraînement sera long. D’autre part, nous avons utilisé des simulations semi-analytiques, faites avec 21cmFAST, pour construire nos jeux de données. Nous savons qu’il existe des différences entre les simulations cosmologiques et les simulations semi-analytiques. C’est pourquoi nous souhaitons par la suite mener le même genre d’étude sur des jeux de données extraits de simulations cosmologiques. Cela peut être compliqué à faire parce que ces dernières demandent beaucoup de temps de calcul. Il existe néanmoins des bases de données de simulations déjà faites, comme Semelin et al. (2017). Celle-ci ne contient cependant pas les cartes de temps de réionisation dont nous avons besoin pour entraîner notre réseau.

D’autre part, il pourrait être intéressant d’essayer de reconstruire des cartes  $t_{\text{reion}}(\vec{r})$  à partir de cartes de signal 21 cm obtenues avec un modèle donné avec des prédicteurs déjà entraînés sur un autre modèle. Cependant, cela n’est pas trivial parce que les prédicteurs sont modèles-dépendants a priori, ce qui peut affecter les reconstructions. Nous souhaitons par la suite analyser des reconstructions avec un prédicteur entraîné sur un modèle différent de celui des données d’entrée pour voir si nous sommes capables de contraindre des modèles de réionisation (ou de matière noire, voir les prochaines études de Julien Hiegel). On peut par exemple se demander s’il est possible d’écarter certains modèles en comparant deux reconstructions faites avec le « mauvais » prédicteur à différents redshifts d’observation. Nous pourrions aussi contraindre des modèles par le biais de l’histoire de réionisation. Si, à l’avenir, on peut d’une part accéder à l’histoire de réionisation grâce à des observations à différentes fréquences (avec le spectre de puissance ou les cartes de signal 21 cm par exemple), et que d’autre part on mesure l’histoire de réionisation dans nos reconstructions sur la plan du ciel (qui sont a priori modèle-dépendantes), on pourrait les comparer dans le but de déterminer le prédicteur qui donne la meilleure correspondance entre ces différentes mesures de l’histoire de réionisation. Ce prédicteur aurait alors le modèle le plus réaliste.

Enfin, le fait que nous parvenions à faire des reconstructions correctes des temps de réionisation à partir de cartes du signal 21 cm dans une gamme de redshift montre que le champ  $t_{\text{reion}}(\vec{r})$  est contenu dans le signal 21 cm dans cette gamme de redshift. Ceci n’est pas trivial et nous ne disposons pas à l’heure actuelle d’expression reliant ces deux grandeurs. De plus, nos reconstructions sont utiles en ce que le champ de temps de réionisation contient beaucoup d’informations sur le processus de réionisation. Nous allons montrer dans les Chapitres 5 et 6 des manières d’extraire de l’information sur le processus de réionisation à partir de cartes de temps de réionisation, et nous souhaitons à l’avenir les appliquer aux reconstructions de ce Chapitre, que nous pourrions faire à partir d’observations, pour mieux comprendre l’Époque de la Réionisation. Ces méthodes sont complémentaires aux études déjà présentes dans la littérature, qui se concentrent en général sur des données prises à différentes fréquences le long de la ligne de visée et relatent alors d’une réionisation moyenne à un redshift donné. Une reconstruction de  $t_{\text{reion}}(\vec{r})$  faite sur une région du ciel observée nous permet alors de savoir spécifiquement comment s’est réionisé localement l’hydrogène dans cette région du ciel.

## 4.7 « Reionisation time fields reconstruction from 21 cm signal maps »

Cette section présente l'article « Reionisation time fields reconstruction from 21 cm signal maps » (ou en français « Champs de temps de réionisation reconstruits à partir de cartes du signal 21 cm »). Cet article a été soumis le 24 mars 2023 dans la revue à comité de lecture *Astronomy & Astrophysics*, et accepté le 12 septembre 2023. Il a été publié sur arxiv en juillet 2023 et sera bientôt publié dans *A&A*.

# Reionisation time field reconstruction from 21 cm signal maps

Julien Hiegal, Emilie Th  lie, Dominique Aubert, Jonathan Chardin, Nicolas Gillet, Pierre Galois, Nicolas Mai, Pierre Ocvirk, and Rodrigo Ibata

Universit   de Strasbourg, CNRS UMR 7550, Observatoire Astronomique de Strasbourg, Strasbourg, France  
e-mail: julien.hiegal@astro.unistra.fr

Received ..., accepted ...

## ABSTRACT

**Context.** During the epoch of reionisation, the intergalactic medium is reionised by the UV radiation from the first generation of stars and galaxies. One tracer of the process is the 21 cm line of hydrogen that will be observed by the Square Kilometre Array (SKA) at low frequencies, thus imaging the distribution of ionised and neutral regions and their evolution.

**Aims.** To prepare for these upcoming observations, we investigate a deep learning method to predict from 21 cm maps the reionisation time field ( $t_{\text{reion}}(\mathbf{r})$ ), the time at which each location has been reionised. The  $t_{\text{reion}}(\mathbf{r})$  method encodes the propagation of ionisation fronts in a single field, and gives access to times of local reionisation or to the extent of the radiative reach of early sources. Moreover it gives access to the time evolution of ionisation on the plane of sky, when this evolution is usually probed along the line-of-sight direction.

**Methods.** We trained a convolutional neural network (CNN) using simulated 21 cm maps and reionisation time fields produced by the simulation code 21cmFAST. We also investigated the performance of the CNN when adding instrumental effects.

**Results.** Overall, we find that without instrumental effects the 21 cm maps can be used to reconstruct the associated reionisation times field in a satisfying manner. The quality of the reconstruction is dependent on the redshift at which the 21 cm observation is being made, and in general it is found that small-scale features ( $<10\text{cMpc}^{-1}$ ) are smoothed in the reconstructed field, while larger-scale features are recovered well. When instrumental effects are included, the scale dependence of reconstruction is even further pronounced, with significant smoothing on small and intermediate scales.

**Conclusions.** The reionisation time field can be reconstructed, at least partially, from 21 cm maps of IGM during the epoch of reionisation. This quantity can thus be derived in principle from observations, and should then provide a means to investigate the effect of local histories of reionisation on the first structures that appear in a given region.

**Key words.** Cosmology: large-scale structure of Universe, dark ages, reionisation, first stars – Methods: numerical, Galaxies: formation, high-redshift

## 1. Introduction

One of the most important transitions in the history of the Universe is the epoch of reionisation (EoR), a period driven by collapsed dark matter halos where the first galaxies and stars emerge (Loeb & Barkana (2001), Wise (2019), Dayal & Ferrara (2018), Mu  noz et al. (2020)). The light emitted by these sources started to reionise the intergalactic medium (IGM), mainly composed of hydrogen. This phenomenon is often pictured as a network of growing ionised bubbles, where the centre of the bubbles host the sources of light (Furlanetto et al. (2004), Th  lie et al. (2022)). Eventually, these growing regions percolate until the whole IGM gets reionised, ending the EoR near  $z=5.5\text{--}6$  (e.g. Kulkarni et al. (2019), Konno et al. (2014)).

This epoch can be probed using the 21 cm signal produced by a spin-flip transition (Furlanetto et al. (2006)). This process releases a photon with an initial frequency  $f_0 = 1420$  MHz that will be redshifted until it reaches us. Such low-frequency radio observations allow us to infer EoR properties, for example from the 21 cm power spectrum (e.g. Furlanetto et al. (2004), Zaldarriaga et al. (2004), Mesinger et al. (2013), Iliev et al. (2012), Greig & Mesinger (2017), Zhao et al. (2022), Nasirudin et al. (2020), Pagano & Liu (2020), Gazagnes et al. (2021), Liu & Parsons (2016), Gorce et al. (2023)) or the 21 cm bispectrum. (KaraGiannis et al. (2022), Hutter et al. (2019)) For example, the Low

Frequency Array<sup>1</sup> (LOFAR, van Haarlem, M. P. et al. (2013)) sets upper limits on the 21 cm signal power spectrum, putting the first constraints on the state of the IGM on the high emissivity of UV photons (Ghara et al. (2020) or on the radio background (Mondal et al. (2020)). Likewise, the Hydrogen Epoch of Reionisation Array<sup>2</sup> (HERA) is designed to study the 21 cm power spectrum to constrain several parameters such as the EoR timing (DeBoer et al. (2017)); for example, it was recently able to put actual boundaries on the X-ray heating produced by the first galaxies (Abdurashidova et al. (2022)).

The Square Kilometer Array<sup>3</sup> (SKA, see e.g. Mellema et al. (2013)), will soon be built, and will have enough sensitivity, resolution, and coverage at low frequencies to measure the 21 cm signal at high redshift and map the hydrogen distribution during the EoR. While SKA will also be able to investigate the EoR from the 21 cm power spectrum, SKA will give us the unique opportunity to get images of the HI state. Such observations at different frequencies, hence different redshifts, will not only track the HI in 2D on the sky, but also along the line of sight, providing the time evolution of the signal.

<sup>1</sup> <https://www.astron.nl/telescopes/lofar/>

<sup>2</sup> <http://reionisation.org/>

<sup>3</sup> <https://skatelescope.org>

This tomography is a great opportunity to explore the EoR (e.g. Giri (2019), Mellema et al. (2015)). SKA will allow us to study astrophysical parameters providing information on the IGM, size, and distribution of ionised bubbles or the properties of the first generation of galaxies (e.g. Mellema et al. (2013)). By extension, 21 cm observations from the EoR would help to improve our understanding of the early universe and to constrain many of its facets, such as the optical depth  $\tau$  of the last scattering surface (e.g. Billings et al. (2021)) or the properties of sources and propagation of ionising photons (e.g. Shaw et al. (2022)); these observations would also help us understand the properties of dark matter by studying the non-linear matter power spectrum (e.g. Mosbech et al. (2022)).

In this spirit, the aim of this paper is to investigate how these future 21 cm observations can help us to study how the reionising radiation propagated, and how it started and evolved. We focus on finding the seeds of the ionising photons that set off the reionisation and on monitoring the propagation and eventual percolation of reionisation fronts.

The 21 cm signal contains a significant amount of physical information, encoded by the temperature brightness  $\delta T_b$  (see Bianco et al. (2021), Prelogović et al. (2022a), Furlanetto et al. (2006), Mellema et al. (2006)):

$$\delta T_b(z) \approx 27 x_{\text{HI}}(z) (1 + \delta_b(z)) \left( \frac{1+z}{10} \right)^{\frac{1}{2}} \left( 1 - \frac{T_{\text{CMB}}(z)}{T_s(z)} \right) \left( \frac{\Omega_b}{0.044} \frac{h}{0.7} \right) \left( \frac{\Omega_m}{0.27} \right)^{-\frac{1}{2}} \text{ [mK]}, \quad (1)$$

which depends on the neutral fraction of hydrogen  $x_{\text{HI}}$ , the density contrast of baryons  $\delta_b$ , the cosmic microwave background temperature  $T_{\text{CMB}}$ , and the so-called spin temperature  $T_s$  driven by the thermal state of the gas or the local amount of Ly- $\alpha$  radiation (Liszt, H. (2001)).

A single 21 cm observation can therefore provide direct insight into the state of these quantities at the observed redshift  $z$ . Figure 1 shows examples of mock 21 cm observations, obtained thanks to 21cmFAST (Mesinger et al. (2011), Murray et al. (2020)), see Sect. 2). From  $z=15$  to  $z=5.5$  we can observe HII bubbles (in white), inside of which no signal can be observed, growing with time until only HII remains and the radio signal vanishes. Since each observation in this sequence is a snapshot of a propagation process, they are correlated. At the extreme, it can even be envisioned that a single 21 cm observation may be used as an anchor point to trace the sequence into the past (at higher  $z$ ) or be extrapolated into the future (to lower redshift) relative to the observed  $z$ . This is the assumption that we test in this work, and more specifically we aim to testing whether the chronology of the spatial distribution of ionised gas can be recovered from a 21 cm observation at a single redshift.

To obtain this chronology we can use the reionisation time field  $t_{\text{reion}}(\mathbf{r})$  (Chardin et al. (2019)). Mapped on 2D images (see Fig. 2),  $t_{\text{reion}}(\mathbf{r})$  returns the time of reionisation for each pixel of the map and encodes the complete history of ionisation propagation in a single field. In Thélie et al. (2022), it was shown how its topology contains a wealth of information on the reionisation process. For example,  $t_{\text{reion}}(\mathbf{r})$  minima are the seeds of the propagation fronts where presumably the first sources can be found,  $t_{\text{reion}}(\mathbf{r})$  isocontours track HII bubbles at a given time or its skeleton provides the sites of ionisation front encounters. It also gives information on the influence of radiation sources on each other (Thélie et al. 2022), opening the door to study distant radiative suppression by nearby objects in the environment.

More generally,  $t_{\text{reion}}(\mathbf{r})$  gives information on local reionisations rather than the global reionisation, putting an emphasis on the environmental modulation of the ionisation history. This local modulation of how light is produced and propagates can translate into local variations of star formation suppression (see e.g. Ocvirk et al. (2020)) or influence the spatial distribution of low-mass galaxies (see e.g. Ocvirk & Aubert (2011)). Galaxies experience a great diversity of reionisation from their point of view (e.g. (Aubert et al. (2018), Zhu et al. (2019), Sorce et al. (2022)), and the reionisation time distribution probes this diversity. Other examples of using a similar description include Trac et al. (2008) on the thermal imprint of local reionisations, Trac et al. (2022) for reionisation modelling, or Deparis et al. (2019) for ionisation front speed measurements. It should be noted that these specific examples use reionisation redshifts instead of reionisation times; while directly related, we found that times are more easily reconstructed than redshifts for our purposes (see Appendix), and we focus only on reionisation times in this paper.

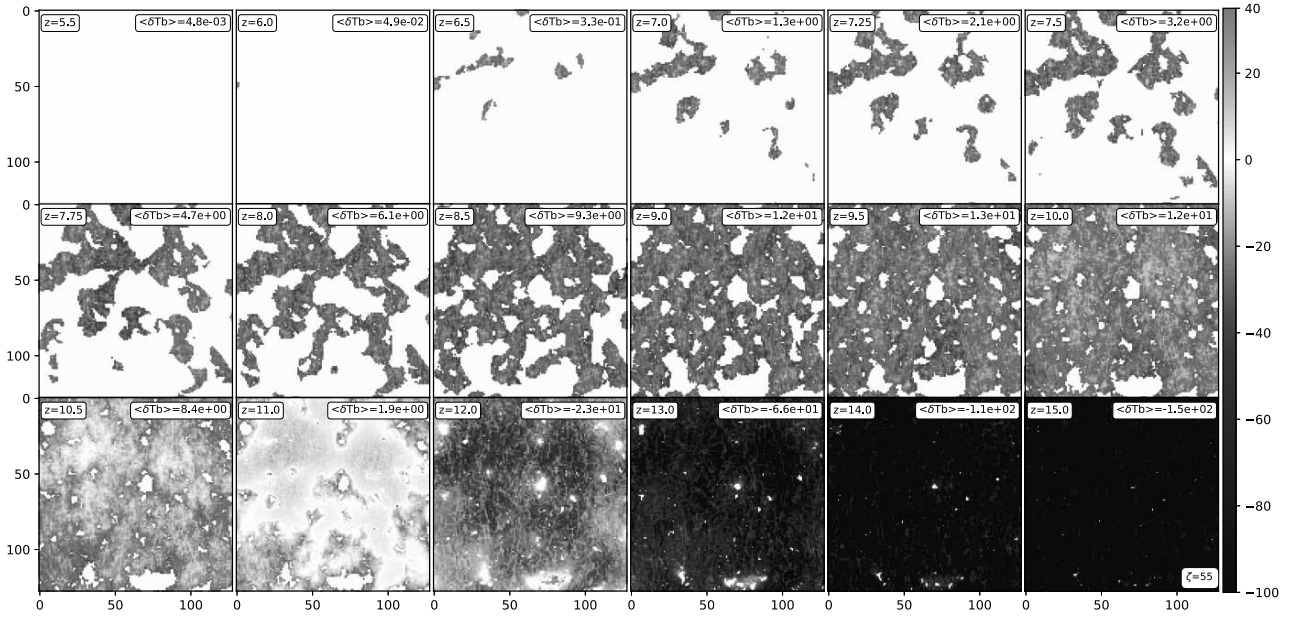
As a means to predict  $t_{\text{reion}}(\mathbf{r})$ , we use convolutional neural network (CNN) methods, which are capable of detecting and learning complex patterns in images. This tool has been widely used in different problems of astrophysics and cosmology (e.g. Bianco et al. (2021), Gillet et al. (2019), Chardin et al. (2019), Prelogović et al. (2022a), Ullmo et al. (2021)). In a recent study Korber et al. (2023) successfully retrieve the growth history of bubbles using mock physical fields. In this study, we extend the CNN applications to  $t_{\text{reion}}(\mathbf{r})$  field reconstructions from mock observations of the 21 cm signal using a U-shaped convolutional neural network (Ronneberger et al. (2015)), which allowed us to get the whole history of reionisation of a sky patch from a single observation.

This article is structured as follows. In Sect. 2 the CNN algorithm and the procedure to deal with the analysis are described. We also present the simulations used to obtain the data. In Sects. 3 and 4 we present the metrics used and the results obtained from monitoring the neural network performance. We discuss instrumental effects in Sect. 5, and conclude in Sect. 7.

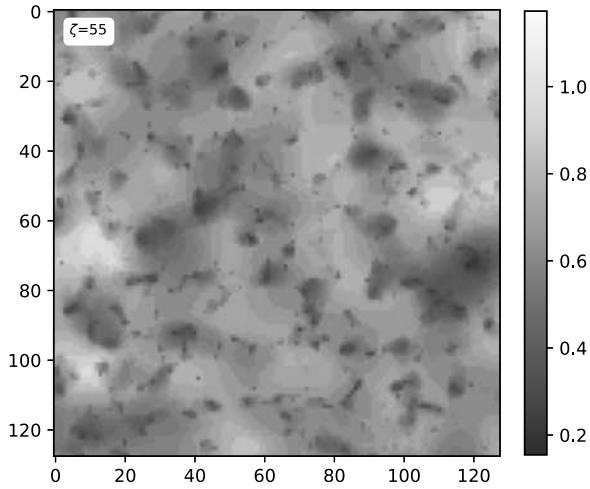
## 2. Convolutional neural network and simulation

The main purpose of this study is to reconstruct the spatial distribution of the reionisation times from 21 cm images using a convolutional neural network (CNN). CNNs are often used to process pixel data, and became widely used for image recognition (LeCun et al. (1999)). Our neural network is implemented thanks to the Tensorflow (Abadi et al. (2015)) and Keras (Chollet (2015)) Python libraries. It took root in the well-known U-net network first developed by Ronneberger et al. (2015). The particularity of this network architecture lies in two distinct parts (Fig. 3). The first part is a contracting path called the encoder, applying series of 2D convolutions and downsamplings to the input image (a 21 cm map here) where its size shrinks as it goes deeper through the neural network. Then the second part does the opposite; it consists of an expansive path (the decoder) applying the same number of convolutions with upsamplings to propagate the information obtained in the encoder. The resulting final output is then another image,  $t_{\text{reion}}(\mathbf{r})$  in our case. This special case of CNN is called an auto-encoder.

For the learning process, we generated a dataset of histories of reionisation, with their corresponding sequence of 21 cm maps. One CNN predictor was considered for each  $z_{\text{obs}}$  redshift at which we have mock 21 cm observations. In practice, we considered 18 predictors for each  $z_{\text{obs}}$  shown in Fig. 1. Ideally, all CNN predictors create the same  $t_{\text{reion}}(\mathbf{r})$  map from mock obser-



**Fig. 1.** Timeline of  $\delta T_b$ . The colour bars are in units of [mK];  $\langle \delta T_b \rangle$  represents the mean of the temperature brightness. These maps come from a 21cmFAST simulation, taken at a given depth from the model  $\zeta 55$  and for each redshift used in this study. White corresponds to the absence of signal, meaning that the hydrogen is ionised at this region. For this simulation, at  $z=5.5$ , there is no neutral hydrogen anymore. Starting at  $z=15$ ,  $\delta T_b$  is mainly seen in absorption (negative values in blue) until  $z=11$ , where  $\langle \delta T_b \rangle$  becomes positive and the signal is seen in emission (positive values in red). During the whole process, HII bubbles grow with time. Figure 2 corresponds to the  $t_{\text{reion}}(\mathbf{r})$  associated to these maps.

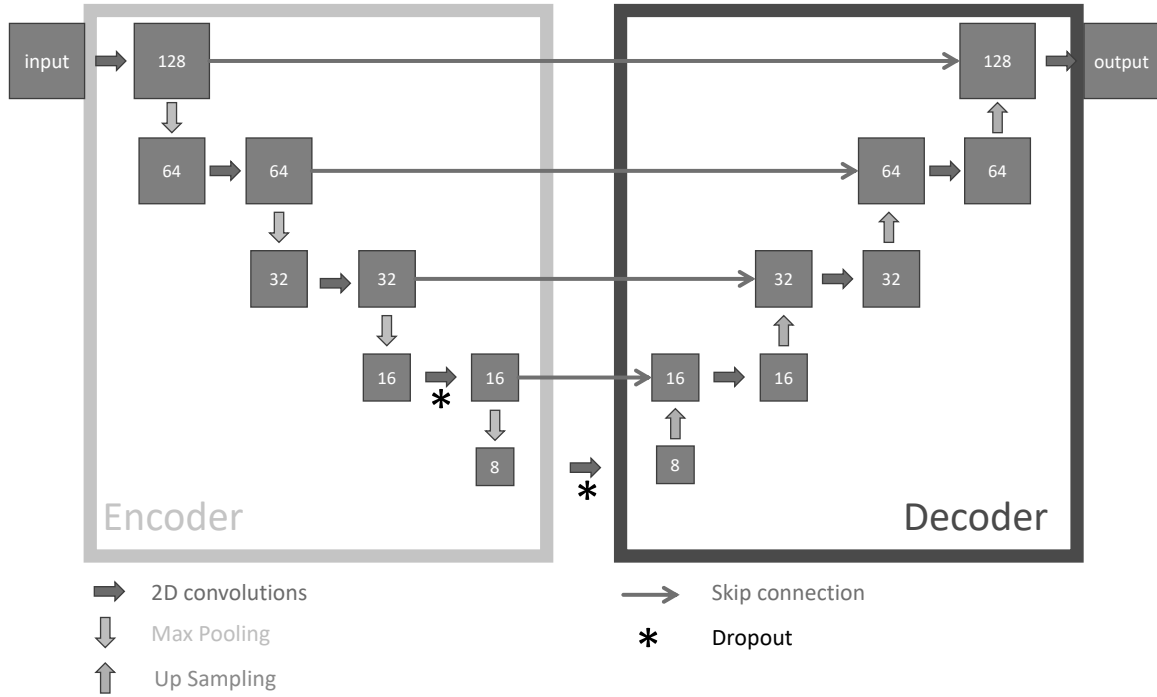


**Fig. 2.** Example of 2D  $t_{\text{reion}}(\mathbf{r})$  map from a  $\zeta=55$  21cmFAST model. The darker the region, the sooner it reionised. In this scenario (and for the whole  $\zeta 55$  dataset), the time of reionisation of the first HII regions is approximately 0.15 Gyr ( $z \approx 20$ , darkest spots), and the last HI regions are reionised at around 1.1 Gyr ( $z \approx 5.5$ , brightest regions). The mean value is 0.61 Gyr ( $z \approx 8.4$ ). This  $t_{\text{reion}}(\mathbf{r})$  map is associated with Fig. 1, taken at the same depth in the simulation box.

variations drawn from the same reionisation history. However, depending on the specific properties of a given 21 cm observation (e.g. the non-zero signal fraction) at a given  $z_{\text{obs}}$ , the predictions will not perform equally well.

The public 21cmFAST simulation code (Mesinger et al. (2011), Murray et al. (2020)) was chosen to obtain the dataset (i.e. 21 cm signal and  $t_{\text{reion}}(\mathbf{r})$  fields). Coeval simulation cubes of size  $256 \text{ cMpc}^{-1}$  with resolution  $1 \text{ cMpc}^{-1}/\text{pixel}$  were produced using a  $\Lambda\text{CDM}$  cosmology with  $(\Omega_m, \Omega_b, \Omega_\Lambda, h, \sigma_8, n_s) = (0.31, 0.05, 0.69, 0.68, 0.81, 0.97)$  consistent with the results from Planck Collaboration et al. (2020) and using standard  $(T_{\text{vir}}, \zeta)$  parameters. The parameter  $T_{\text{vir}}$  sets the minimal virial temperature for halos to enable star formation (see Gillet (2016), Muñoz et al. (2020), Barkana & Loeb (2001)) and Oh & Haiman (2002)) and was chosen such that  $\log_{10}(T_{\text{vir}}) = 4.69798$ . The parameter  $\zeta$  sets the ionising efficiency of high- $z$  galaxies, and allows us to modify the reionisation timing: the larger this value is, the faster the reionisation process will be (Greig et al. (2015)). We considered two ionising efficiencies  $\zeta = 30$  and  $\zeta = 55$  (referred as  $\zeta 30$  and  $\zeta 55$ ), leading to a total of 36 CNN models to be trained, 18 redshifts per  $\zeta$  value. For each  $\zeta$ , 50 different realisations with different seeds were run, giving us access to  $t_{\text{reion}}(\mathbf{r})$  and 21 cm 3D fields. As discussed in the introduction, an alternative approach is to consider the reionisation redshift  $z_{\text{reion}}(\mathbf{r})$  instead of  $t_{\text{reion}}(\mathbf{r})$ . However, we found that the times were better reconstructed, and a brief analysis using  $z_{\text{reion}}(\mathbf{r})$  is presented in Appendix B.

To produce 2D images of  $t_{\text{reion}}(\mathbf{r})$  and the 21 cm signal, we took 64 evenly spaced slices, one out of four (of  $1 \text{ cMpc}^{-1}$  thickness, corresponding to one cell), in the three directions of each cube. Each slice was cut into four  $128 \times 128$  images, finally leading to a total of 768 21 cm images per realisation and per  $z$ , giving 38,400 maps per redshifts. We standardised the 21 cm images to ensure that the range of pixel values was consistent across all images in the dataset and to help the model's training process. The mean value was subtracted and the result was divided by the standard deviation (std), both computed over the



**Fig. 3.** Convolutional Neural Network used in this work. The Encoder (left) refers to the first sequence of 2D convolutions and Max Pooling operations, while the Decoder (right) refers to the second sequence of 2D convolutions and upsampling operations. Each Max Pooling operation reduces the input map size by a factor of 2, keeping mainly large-scale information. The upsampling operator does the opposite, propagating the information at larger scales and increasing the input map size. Each convolution modifies the number of filters in such a way that we have the maximum number of filters at the deepest point of the network. Dropout and skip connection layers are depicted with black asterisks and purple arrows, respectively. Each square contains the size of the feature maps, which remain consistent along a given horizontal line (see Appendix A for additional details about the CNN model).

training set. The mean and standard deviation values are thus the parameters of our predictors.

As shown in Fig. 4, the neutral volume fraction  $Q_{HI}$  is shifted (on the time axis) according to the ionising efficiency  $\zeta$ . Since  $\zeta$  controls how many photons escape from galaxies,  $\zeta_{30}$  gives a delayed history of reionisation compared to  $\zeta_{55}$ . We discuss below the non-zero signal fraction (i.e. the fraction of pixels with non-zero 21 cm signal). Its time evolution is plotted as dots and crosses in Fig. 4, and is shown to follow  $Q_{HI}$ .

The entire dataset is split into three subsets, from which 35,000 images are used for the learning phase. The first subset, known as the training set, comprises 31,500 images. At each epoch during the learning phase this set is fed to the CNN, which computes the loss function (i.e. mean square error, MSE in our case) and modifies the weights to minimise it. Another separate subset of the entire dataset, called the validation set, consists of 3,500 images; it is exclusively used to evaluate the CNN’s performance during the learning phase after each epoch, without being used in the weight adjustment process. The final set is called the test set, and consists of the remaining 3,400 images, which are never processed by the CNN during the learning stages. All the results shown in this paper (except for the loss function and  $R^2$ ; see Sect. 3.1) were obtained via the test set.

### 3. Monitoring the algorithm performance

Once all the hyper-parameters were set and predictions made, we measured the training performance and the prediction accuracy

by comparing the predicted  $t_{\text{reion}}(\mathbf{r})$  maps with the ground truth given by the 21cmFAST simulation.

#### 3.1. Network internal metrics

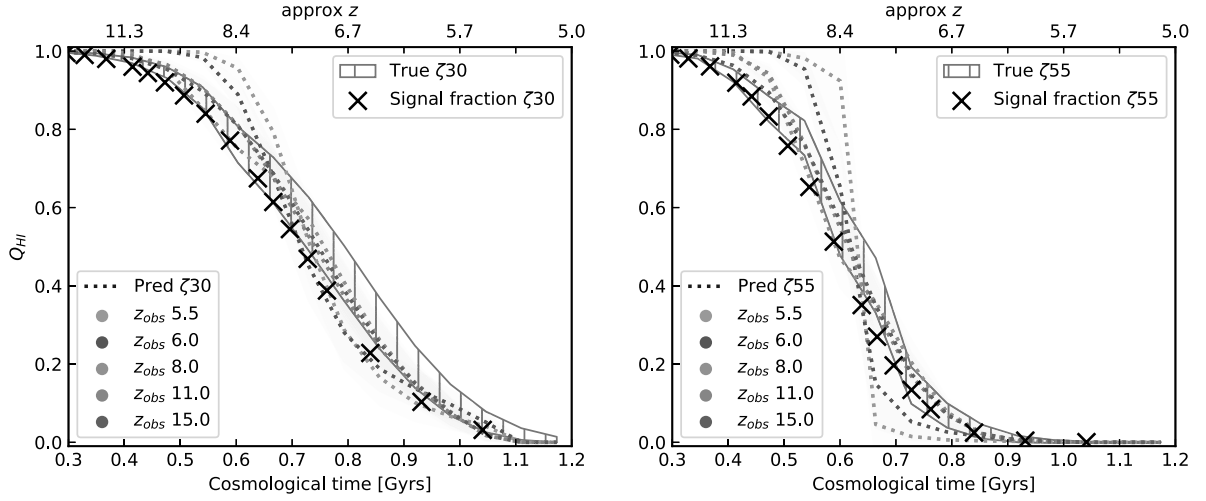
First, two internal metrics were used to monitor the training process. Starting with the loss function, the mean square error (MSE) was defined as the average of the squares of the errors. At each epoch the algorithm tries to minimise this loss function (MSE) by comparing the ground truth (given by the simulation) with the prediction (given by the CNN).

A second indicator was used, called the determination coefficient, defined as

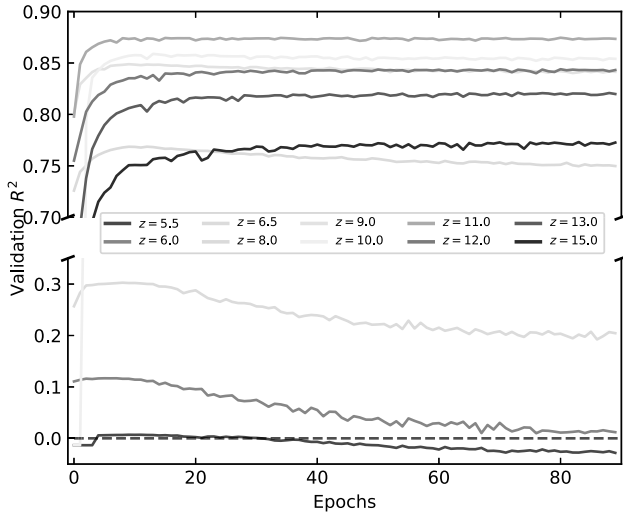
$$R^2 = 1 - \frac{\Sigma(\text{Pred} - \overline{\text{True}})^2}{\Sigma(\text{True} - \overline{\text{True}})^2} = 1 - \frac{\Sigma_{n=1}^{N_{\text{pix}}} (\text{Pred}_n - \text{True}_n)^2}{\Sigma_{n=1}^{N_{\text{pix}}} (\text{True}_n)^2}, \quad (2)$$

where Pred and True are the  $t_{\text{reion}}(\mathbf{r})$  maps of the prediction and ground truth, respectively;  $\text{Pred}_n$  and  $\text{True}_n$  correspond to the  $n$ -th pixel of the considered batch of images; and  $\overline{\text{True}}$  depicts the average of the true field and is equal to zero after normalisation. In our case the predicted values, Pred, and the ground truth values, True, are both (3500, 128, 128) cubes. The summations are performed over  $N_{\text{pix}}$  pixels to measure the network performance on a set it has already or never seen (training or validation set, respectively). Identical values of PRED and TRUE lead to  $R^2=1$ .

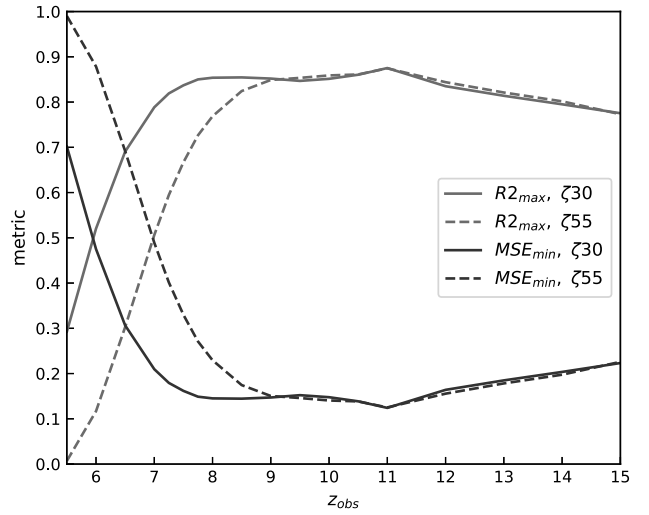
Figure 5 shows the  $R^2$  coefficient during the validation phase for several observation redshifts  $z_{\text{obs}}$  and for  $\zeta_{55}$ . For the vali-



**Fig. 4.** Neutral fraction of hydrogen with respect to time (bottom axis) and redshift (top axis) for  $\zeta_{30}$  (left panel) and  $\zeta_{55}$  (right panel). The crosses stand for the average non-zero signal fraction obtained from 21 cm images at different values of  $z$  (as if taken along the line of sight). The dashed areas stand for the true  $Q_{HI}$  computed from true maps of  $t_{reion}(\mathbf{r})$  (mean and std). The dotted lines stand for the average  $Q_{HI}$  obtained from maps of  $t_{reion}(\mathbf{r})$  predicted by the CNN.



**Fig. 5.**  $R^2$  coefficient with respect to epochs for a selection of training redshifts. These curves correspond to the validation phase and are for the  $\zeta_{55}$  model. The curves for  $\zeta_{30}$  are not shown here, but a comparison between the two models is shown in Fig. 6.



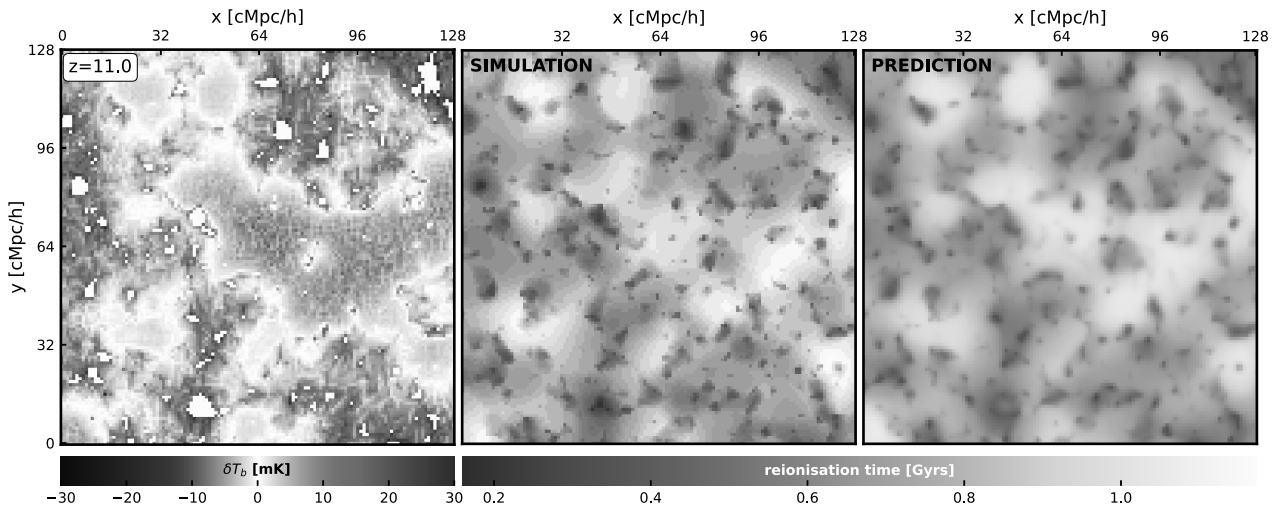
**Fig. 6.** Metric maximum and minimum with respect to the training redshift for the validation set. The red line is for the coefficient  $R2$  and the blue line is for the MSE. The solid lines are for  $\zeta_{55}$  and the dashed line for  $\zeta_{30}$  model.

validation set, this coefficient gives a first estimation of the similarities between true fields and predicted fields immediately at the end of each epoch, allowing us to follow the model's accuracy during the learning process. At low redshift the  $R^2$  coefficient is small and, even worse, negative for this model and for the lowest  $z$ . Nothing can be predicted from low redshift ranges since the non-zero signal fraction at small  $z_{obs} \in [5, 7]$  is low or even equal to zero. Then, above a value of 8 we observe that the prediction performance increases for increasing  $z_{obs}$  until 11, where it eventually degrades again until 15.

Figure 6 shows the maximum value across epochs for  $R^2$  for each  $z_{obs}$  and the two  $\zeta$  models, as well as the minimum value reached by the MSE loss. According to these metrics, the best reproduction of  $t_{reion}(\mathbf{r})$  is obtained from the CNN predictor us-

ing 21 cm maps at  $z_{obs}=11$ , corresponding to 95% of non-zero signal (see Fig. 4). Furthermore, the  $\zeta_{30}$  model returns better results at lower  $z_{obs}$ : at  $z_{obs}=7$ ,  $R^2_{max} \approx 0.79$  for  $\zeta_{30}$ , while  $R^2_{max} \approx 0.51$  for  $\zeta_{55}$ . This can be easily understood from Fig. 4, where the non-zero signal fraction of  $\zeta_{30}$  is considerably larger than for  $\zeta_{55}$  in the  $z_{obs}$  range [5.8] (top axis): at  $z_{obs}=7$ ,  $Q_{HI} > 40\%$  for  $\zeta_{30}$ , while  $< 17\%$  for  $\zeta_{55}$ . At this range of signal fraction values, the gain in terms of information is such that the network performance increases significantly. From Fig. 4 and Fig. 6, we can estimate that between [0.90, 0.96] of non-zero signal fraction the neural network achieves better performance. For a non-zero signal fraction greater than 0.96, the performance decreases again. At these levels of non-zero signal fraction there are only a few HII bubbles to be found, inducing a loss of information on the





**Fig. 7.** Example of prediction done by the CNN trained from images at  $z_{\text{obs}}=11$  and for  $\zeta 30$  model. The left panel shows the 21 cm signal  $\delta T_b$  at this redshift. The middle panel is the ground truth of the  $t_{\text{reion}}$  field. The right panel is the CNN prediction for  $t_{\text{reion}}$  obtained from  $\delta T_b$  (left panel).

location of the seeds of most reionisation regions. Without HII bubbles, the sources of reionisation cannot be located and the UV radiation propagation cannot be determined. Hence, in order to get the best performance, the CNN algorithm requires a compromise between a minimal set of HII bubbles and a significant non-zero signal fraction. The peak value of  $R^2=0.88$  at  $z=11$  also suggests that this redshift of observation is peculiar. The timeline in Fig. 1 shows that  $z=11$  seems to be the transition between a global negative temperature brightness and a positive one in our model as the long range influence of X-rays on the gas becomes effective. At this  $z_{\text{obs}}$  the 21 cm map contains small HII regions with no signal, easily interpretable for the CNN as the places where the first seeds of reionisation are found. Then there are regions that are hotter than average (shown in red) that will reionise sooner, and blue regions that are colder than average that will be the last regions to reionise. This  $z_{\text{obs}}$  thus contains information of the sequence of radiation propagation that seem to be more easily extracted compared to other observation redshifts.

### 3.2. Reionisation time prediction

Beyond the CNN internal metrics, the immediate result is the predicted map itself, as shown in Fig. 7. This  $z_{\text{obs}}=11$  map is one of the best reconstructions ( $R^2 = 0.84$ ) we could create for the  $\zeta 30$  model. The predicted map on the right seems quite close to the ground truth, but smoother. We note that the best predicted maps for  $\zeta 30$  and the best predicted maps for  $\zeta 55$  (not shown here) present a similar qualitative behaviour.

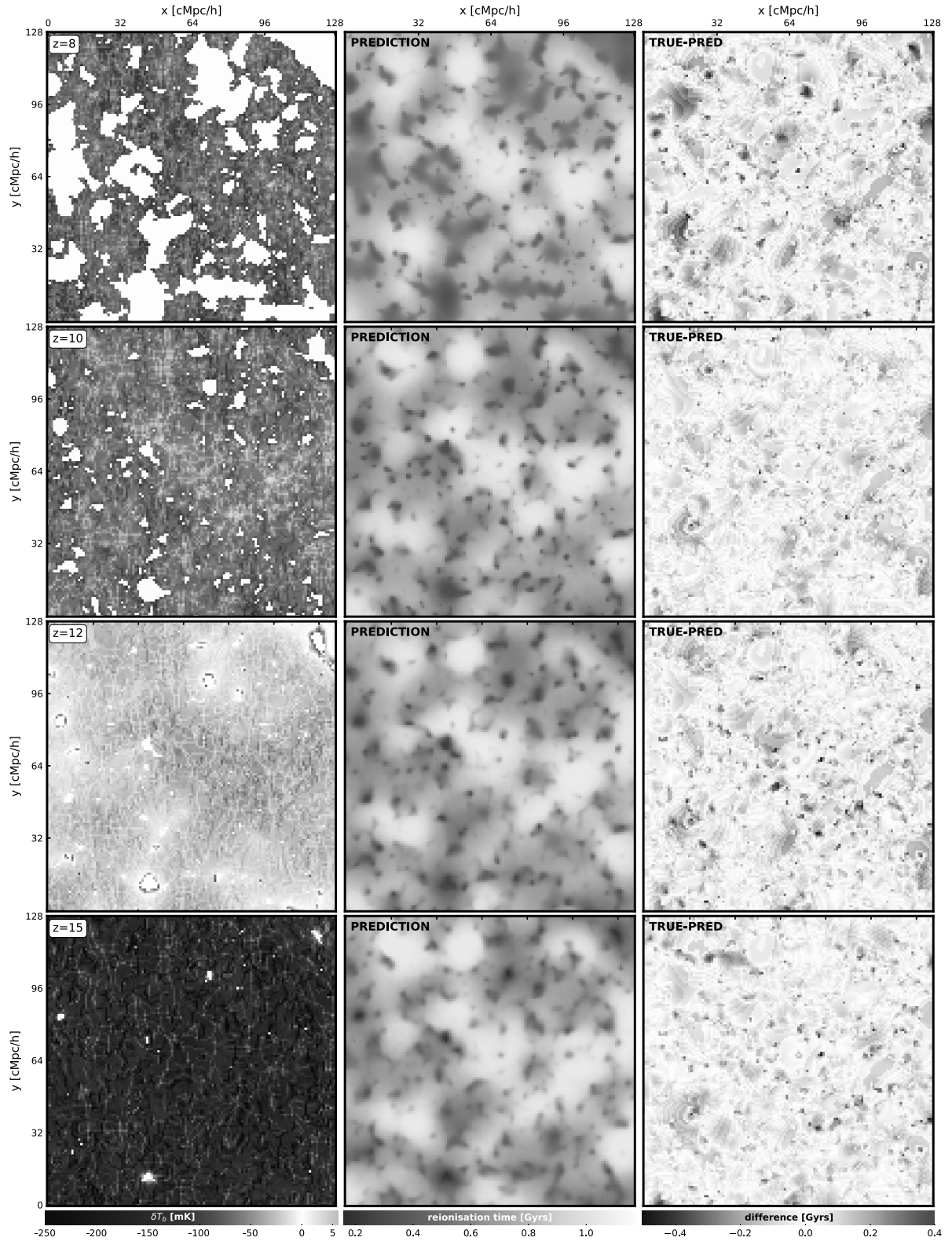
With the true map of  $t_{\text{reion}}(\mathbf{r})$  and its prediction, we can count the number of pixels with values larger than a given reionisation time to obtain  $Q_{\text{HI}}(t)$  on the sky (see Fig. 4). For  $z_{\text{obs}}=11$ , both TRUE and PRED measurements match within the dispersion of true values and are consistent with the signal fraction evolution computed from the actual evolution of the 21 cm signal with  $z$ . This implies that the information obtained across the sky at a single  $z_{\text{obs}}$  via the predicted  $t_{\text{reion}}(\mathbf{r})$  is consistent with (or can be cross-checked against) the evolution along the line of sight.

Figure 8 shows examples of predictions obtained for different models trained at different  $z_{\text{obs}}$ . The first column shows the mock observations (21 cm maps) at several redshifts, the middle

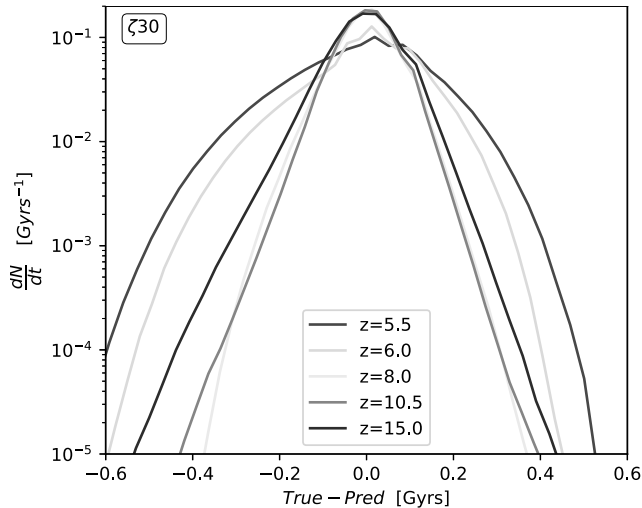
column shows the predictions obtained with the left panel and the right column shows the difference between the ground truth and the predicted  $t_{\text{reion}}(\mathbf{r})$  field. Looking at the two first columns, at low redshift, the predictions are suboptimal as the inferred field has been completely smoothed. For  $z_{\text{obs}} \geq 10$ , Our CNN becomes able to capture small-scale features ( $<10 \text{ cMpc}^{-1}$ ), such as extrema. However, in the right column the CNN seems to have more difficulties in predicting the local extrema of reionisation times, even though their locations are well predicted, especially at low redshift (e.g. for  $z_{\text{obs}}=8$ :  $(x,y) \approx (120,125) \text{ cMpc/h}$ ). These points correspond to the seeds of the propagation of fronts, presumably linked to the first sources of radiation, and seem to be subject to a smoothing intrinsic to our CNN implementation. Compared to  $z_{\text{obs}}=10$  or 11, the  $z_{\text{obs}}=15$  prediction appears to be slightly smoother, although the earliest reionisation times seem to be well reproduced. Finally, Fig. 9 depicts the normalised histogram of TRUE-PRED maps. Distributions are centred on zero, with an asymmetry towards negative values: our CNN predictions returns greater reionisation times than the ground truth (i.e. a delayed reionisation history). For example, taking  $z_{\text{obs}}=15$  there are more pixels at True-Pred=-0.4 Gyr ( $dN/dt > 2e-4$ ) than for True-Pred=0.4 Gyr ( $dN/dt < 3e-5$ ). This systematic effect is less severe for the best CNN predictors trained to process  $z_{\text{obs}}=8$  or 10.5 observations in this figure.

### 3.3. True versus Predicted histograms and fitting fraction

One of the most standard tests is the true versus predicted (TvP), where all the predicted pixels are compared one by one to their true value given by the simulation. Figure 10 shows the TvP corresponding to all the maps in the test set using the  $z_{\text{obs}}=11$  CNN. Most values follow the perfect correlation for typical  $t_{\text{reion}}(\mathbf{r})$  values (0.4-1 Gyr), while extreme values ( $<0.4$  Gyr and  $>1$  Gyr) are not as well recovered by the CNN (mean value up to +0.25 and down to -0.15 respectively). This is not surprising, given the predicted maps on Fig. 7. The extreme values of  $t_{\text{reion}}(\mathbf{r})$  coincide with small-scale features that are smoothed out, where the first sources with lowest  $t_{\text{reion}}$  are found. These values are also rare (4.2% of the total number of pixels), explaining why the algorithm fails at learning how to recover them.



**Fig. 8.** Examples of prediction made by the CNN trained with maps at several redshifts  $z_{\text{obs}}$  and  $z_{30}$ . The left panels show the 21 cm  $\delta T_b$  maps. The middle column shows the predicted  $t_{\text{reion}}$  fields. The right panels show the difference TRUE-PRED. Hence, the redder it appears on the map, the more the CNN overestimates the true value at the given pixel; instead, the darker it is, the more the CNN underestimates the real value. The true  $t_{\text{reion}}(\mathbf{r})$  is shown in Fig. 7.



**Fig. 9.** Normalised distribution of TRUE-PRED values (as shown in Fig. 8) for several  $z_{\text{obs}}$ . These curves were obtained from the whole test set of  $\zeta 30$  models.

Figure 11 presents a synthesis of the true versus predicted maps of all the predictors at different  $z_{\text{obs}}$ . The fitting fraction is a value between 0 and 1, which corresponds to the predicted pixel’s fraction whose value fits within an arbitrary error calculated as  $\epsilon\%$  of the true pixel’s value. The larger the allowed error is, the more ‘good’ pixels will be found. It is then clear how low  $z_{\text{obs}}$  values ( $<8$ ) gives less accurate results, especially when allowing a small error ( $<10\%$ ): the fitting fraction value is more than 10% lower than for  $z_{\text{obs}} > 8$ . It can be understood from the maps of temperature brightness (or from Fig. 4) where the maps contain less signal with decreasing redshifts from  $z=8$ : less than 50% of the map contains observable neutral hydrogen. At the extreme redshift  $z=5.5$  there is no signal left because all HI is reionised and no prediction is possible. A value of  $z_{\text{obs}} > 8$  seems to give the best results, and all predictions made from  $z_{\text{obs}}$  above 9 seem to have similar performance: between 70 and 75% of matching pixels for 5% error, and a slight decrease with growing  $z_{\text{obs}}$ . Overall, we recover the two trends identified previously: low  $z_{\text{obs}}$  simply lacks the signal for a good reconstruction, while high  $z_{\text{obs}}$  lacks the direct imprint of sources that appear later. The best compromise is found for  $z_{\text{obs}}$  values between 8 and 12, corresponding to signal fraction  $\approx 0.8$ , with an optimal value of 0.95.

### 3.4. $\zeta 30$ and $\zeta 55$

The first results comparing the  $\zeta 30$  and  $\zeta 55$  scenarios are very close. The prediction accuracy is similar except for lower  $z_{\text{obs}}$  ( $<8$ ) where the signal fraction tends to be quite different between both situations. In the most extreme case, at  $z_{\text{obs}}=5.5$ , the CNN trained with the  $\zeta 30$  maps is quite limited, but still returns a prediction, whereas the CNN trained with  $\zeta 55$  maps cannot predict anything, which was expected since there is no HI left at late times in the  $\zeta 55$  scenario. In the following sections, we only consider the  $\zeta 30$  scenario to discuss results in the whole range of  $z_{\text{obs}}$  used.

## 4. Structure of predicted maps

We now investigate the spatial structure of the reconstructions using three metrics: the power spectrum, the Dice coefficient, and the minima statistics.

### 4.1. Power spectrum $P_k$

We now compare the power spectrum  $P_k$  of the reionisation time field with that predicted by the neural network in order to have a statistical point of view on how well the network reconstructs the different scales present on the map. Figure 12 depicts the  $t_{\text{reion}}(\mathbf{r})$  power spectrum of model  $\zeta 30$ . A first look shows that the lack of 21 cm signal drastically erases the possibility to predict anything. Predictions for  $z=5.5$  and  $z=6$  are incompatible with the real  $P_k$  at mid-scales  $k \approx 7e-2 \text{ cMpc}^{-1}\text{h}$ ; less than 30% of the power remains for  $z=6$ , against up to 85% for  $z=8$ . At small scales ( $k > 2e-1 \text{ cMpc}^{-1}\text{h}$ ) less than 12% of the power remains for  $z=6$  against up to 57% for  $z=8$ . Now looking at large scales (low spatial frequencies such as  $k < 3e-2 \text{ cMpc}^{-1}\text{h}$ ), our model reproduces the power at more than 95% for  $z_{\text{obs}} > 8$ . However, at  $k=0.2 \text{ cMpc}^{-1}\text{h}$  and beyond, the prediction cannot produce enough power, meaning that the smaller scales are difficult to predict at all  $z_{\text{obs}}$ . Again, the predictor smoothes the  $t_{\text{reion}}(\mathbf{r})$  field, predicting a map generally blurrier than the ground truth. To improve results at the smallest scales, generative adversarial networks (GANs) could be a solution (see Ullmo et al. (2021)).

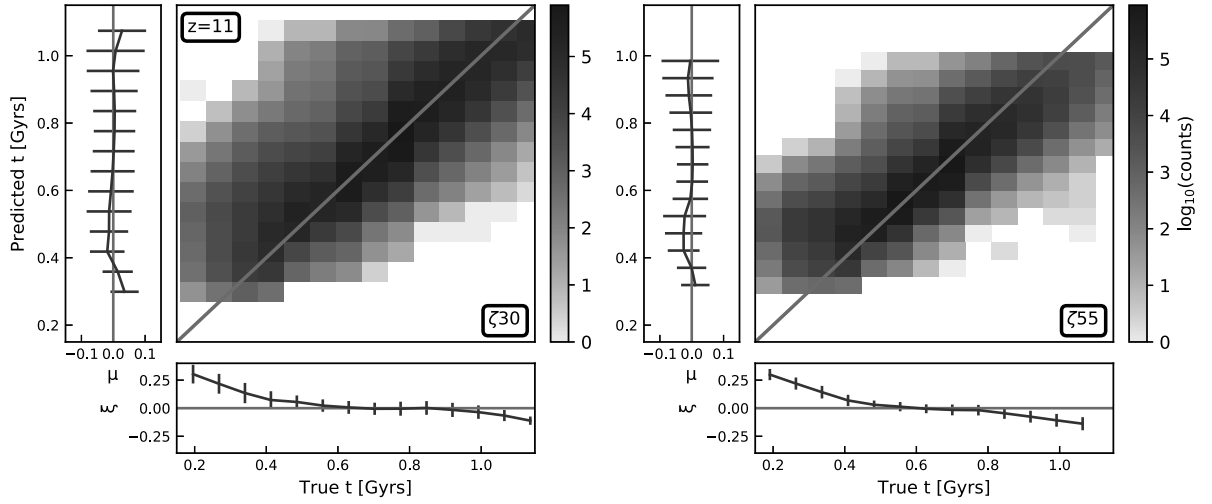
### 4.2. Dice coefficient

Another way to look at the predictor performance is the Dice coefficient (see Ullmo et al. (2021)). This method is useful to see what kind of regions the algorithm reconstructs best; for example, whether the first regions that become reionised are clearly predicted or if, conversely, late regions are reconstructed in a better way. This coefficient focuses on the map structure by looking at regions with given values. It determines how the CNN recovers structures instead of giving an accuracy according to the value of pixels or the considered scale.

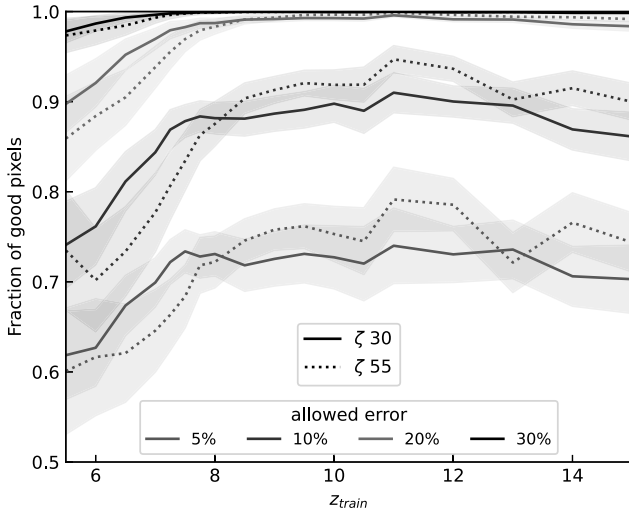
The Dice coefficient is used by taking a threshold  $t$  (0 to 100) and by considering only the pixels with the largest  $t_{\text{reion}}$  values in the true and predicted maps. We can estimate the regions of the map where the prediction overlaps with the ground truth, using a newly formed map with pixels in only three possible states:

- Predicted and true pixels both have a value above the threshold (both fit the criterion), denoted the yellow state.
- Both have a value below the threshold (both do not fit the criterion), denoted the blue state.
- There is a mismatch between the prediction and simulation, denoted the green state.

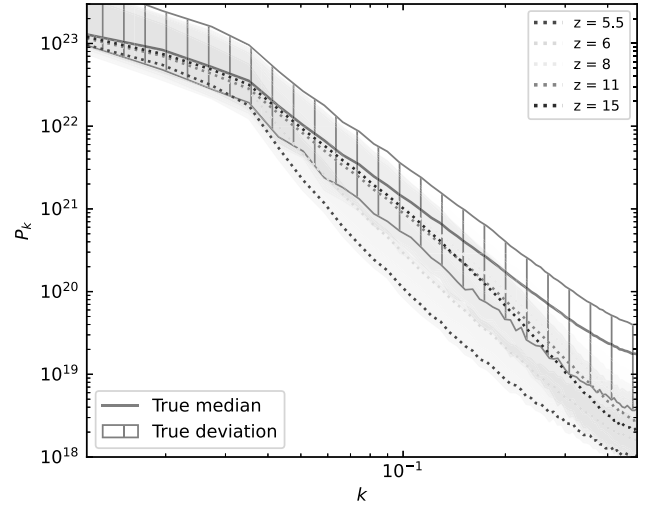
It is important to note that the value of the threshold corresponds to a given cosmological time (or redshift). Using 10% for the threshold, the constructed map will only contain the information for large values of cosmological time (low redshifts), typically the last regions to reionise. On the other hand, taking 100% as threshold, the whole map will be considered. An example of an overlap map is depicted in Fig. 13 using  $z_{\text{obs}}=10$ . The threshold example on the figure is 0.4, corresponding to 40% of the largest values. Only a few green regions (corresponding to 15% of the pixels) are present, and more than 85% of the pixels are in agreement between the prediction and the ground truth. This range of values is actually well reconstructed, and the remaining differences are located at the edge of these regions (in the green areas, e.g.  $(x,y) \approx (70,10) \text{ cMpc/h}$ ).



**Fig. 10.** True vs predicted maps obtained from the full dataset in the test set, for  $z_{\text{obs}} = 11$  and for both models. The red lines stand for the perfect correlation. The bottom and left histograms are the mean and the standard deviation of the residual:  $r = \text{Predicted} - \text{True}$ , in both vertical and horizontal directions. The bottom histogram stands for the learning error, while the side histogram stands for the recovery uncertainty. In practice, the recovery uncertainty is the only accessible estimator since the ground truth will not be accessible.



**Fig. 11.** Fitting fraction with respect to redshift. The greater the allowed error for a pixel (between prediction and ground truth) is, the larger the number of pixels that have a value close to the true one within this error (i.e. ‘good pixels’). The solid lines are for  $\zeta 55$  and the dotted lines for  $\zeta 30$ . The shaded areas depict the standard deviation around the mean for the whole predicted dataset.



**Fig. 12.** Power spectrum of the  $t_{\text{reion}}(r)$  field obtained for  $\zeta 30$ .  $k$  is the spatial frequency with dimension  $[\text{cMpc}^{-1}\text{h}]$ . The black line stands for the average power spectrum of the ground truth and the dotted coloured lines stand for  $P(k)$  predictions with different  $z_{\text{obs}}$ . The shaded areas depict the dispersion around the median (1st and 99th percentiles) for both predictions and ground truth.

The Dice coefficient, or association index, is calculated at a given threshold as (Dice 1945)

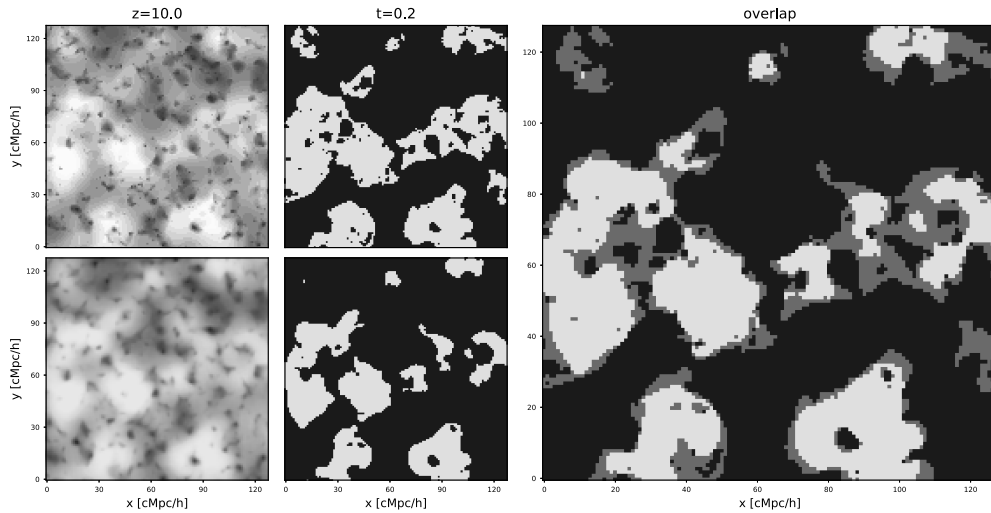
$$\text{Dice} = \frac{n_{\text{yellow}}}{n_{\text{yellow}} + n_{\text{green}}}, \quad (3)$$

with  $n_i$  the number of pixels with colour  $i$ . The Dice coefficient can only take values between 0 and 1 (0 for no correspondence between prediction and ground truth and 1 for a perfect reconstruction).

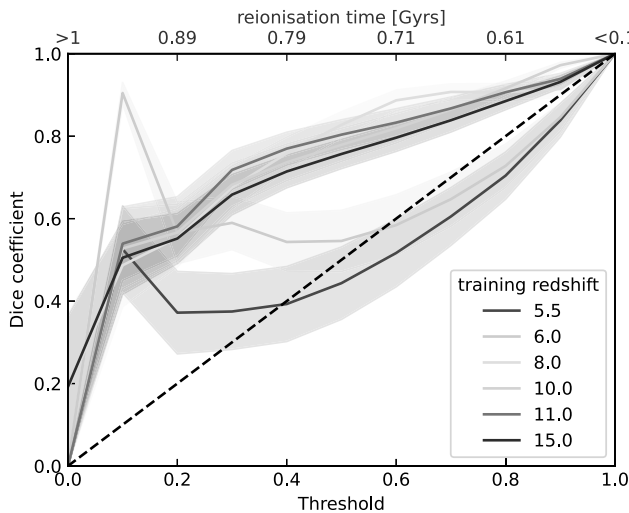
Figure 14 shows the Dice coefficient for the  $\zeta 30$  model. The dotted black line stands for the Dice coefficient obtained if we compare the true thresholded map with a map randomly filled

with zeros and ones. Globally, we recover that  $z_{\text{obs}} > 8$  provides better accuracy compared to the random situation (for threshold=0.4: Dice>0.7 against 0.4, respectively), with similar performance at all  $z_{\text{obs}}$ . Furthermore,  $z_{\text{obs}} = 11$  seems to dominate until median values for the threshold. Afterwards,  $z_{\text{obs}} = 8$  coefficients catch up, followed by  $z_{\text{obs}} = 10$ , meaning that these observation redshifts recover efficiently the first structures of reionisation.

Looking at low threshold values, the Dice coefficient provides additional insights into the performance at low  $z_{\text{obs}}$ , such as  $z_{\text{obs}} = 5.5$  or 6. At these redshifts, predictions are slightly better for low threshold values meaning that at  $z_{\text{obs}} = 6$ , the neural network predicts in a very efficient way the last regions reionised



**Fig. 13.** Example of overlap map for  $\zeta_{30}$  and a threshold value  $t = 0.2$ . The left panel is the true  $t_{reion}$  field (top) and the predicted  $t_{reion}(\mathbf{r})$  for a  $z_{obs}$  of 10 (bottom). The middle panel depicts a 1 or 0 map where 1 (yellow) stands for the pixels fitting 20% of the highest values and 0 (purple) for pixels below the threshold for the true field (top) and the predicted field (bottom). The right picture is the overlap of the middle panels, where yellow stands for coincidence, green for incorrect pixels in the prediction, and purple where the values are below the threshold in both fields.

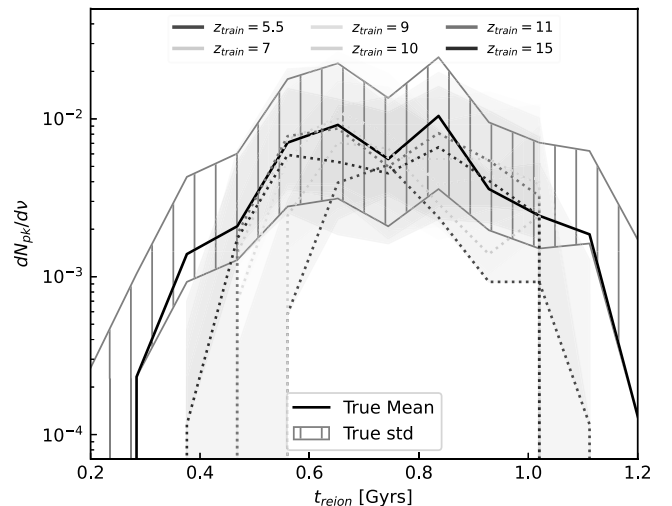


**Fig. 14.** Dice coefficient for several  $z_{obs}$  and for  $\zeta_{30}$ . The dashed black line stands for the Dice coefficient when matching the true map with a random mask. Pixels are considered starting with the largest pixel value (last pixel to reionise) down to a threshold. A 0.2 threshold means that we consider the last 20% to be reionised. The shaded areas depict the standard deviation around the mean DICE coefficient.

(for threshold=0.1, Dice>0.9): since they are the only regions where a non-zero signal can be found, the predictor can locate them accurately.

#### 4.3. Minima statistics

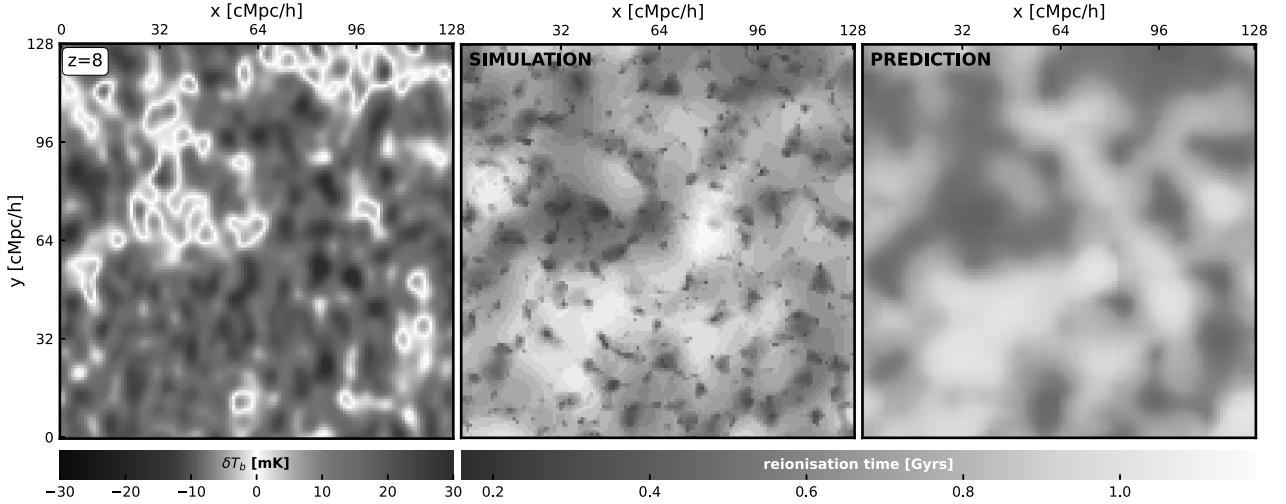
We now investigate at the minima of  $t_{reion}(\mathbf{r})$  (i.e. the regions that reionised at the earliest times) to probe how our CNN detects sources of reionisation. We used DisPerSE (Discrete Persistent Structures Extractor) (Sousbie (2011), Sousbie et al. (2011)) to identify the distribution of reionisation times minima; it uses discrete Morse theory to identify persistent topological features in



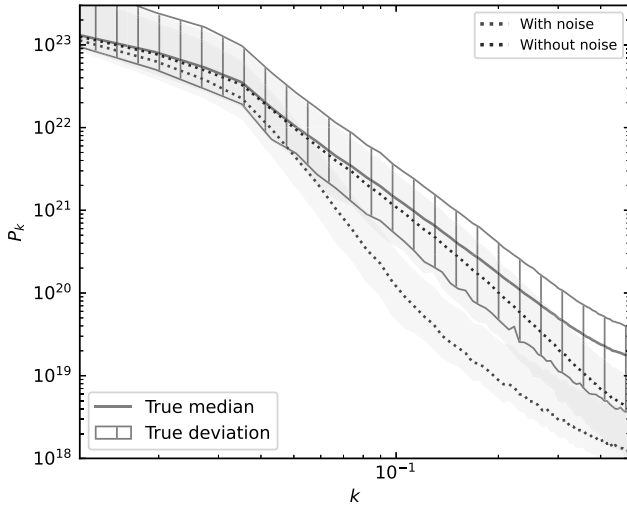
**Fig. 15.** Distribution of reionisation times for  $t_{reion}(\mathbf{r})$  minima. The solid line is obtained from true  $t_{reion}(\mathbf{r})$ , while the dotted lines are obtained from predictions at several  $z_{obs}$ . The shaded areas show the dispersion around the median (mean and std) for both predictions and ground truth.

two-dimensional maps, such as voids, walls, filaments, and clusters. While we focus here only on minima, valuable insights into the underlying topology of  $t_{reion}(\mathbf{r})$  can also be obtained from the persistent structures detected by DisPerSE, to see how they relate to the physical processes that shape the distribution of reionisation (see also Th  lie et al. (2022) a, b).

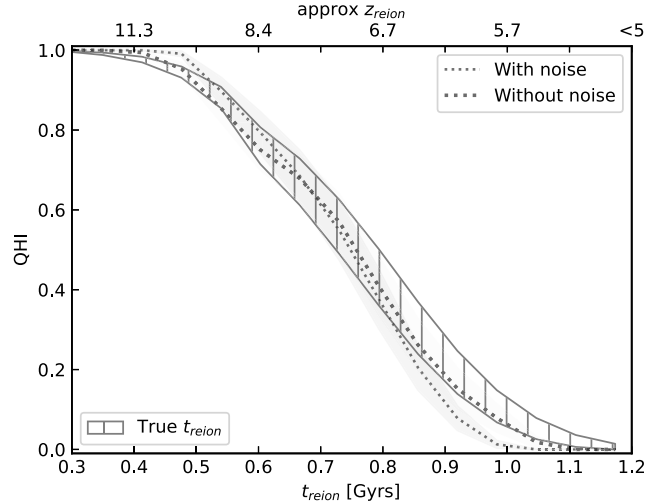
The results of this analysis are illustrated in Fig. 15. The black line stands for this analysis performed on the true  $t_{reion}(\mathbf{r})$  field: at low  $t_{reion}$  (high  $z$ ) sources are rare, and their number is highest at  $t=0.8$  Gyr (approx  $z=7$ ). Their number then drops at higher  $t_{reion}$  values, not because sources becomes rarer but because they appear in already reionised regions and cannot be traced by peaks in  $t_{reion}(\mathbf{r})$  maps. The statistics of our CNN predictions, depicted with dotted lines, clearly show that the



**Fig. 16.** Predictions for the model  $\zeta_{30}$  at  $z_{\text{obs}}=8$ , including instrumental effects for a typical SKA observation (left panel, see details in main text). The middle panel is the ground truth for  $t_{\text{reion}}(\mathbf{r})$  and the right panel is the CNN prediction.



**Fig. 17.** Power spectrum of the  $t_{\text{reion}}(\mathbf{r})$  field obtained with the model  $\zeta_{30}$  at  $z_{\text{obs}}=8$ , including instrumental effects for a typical SKA observation. The black line and hatched area stand for the mean and standard deviation of the true field  $P(k)$ . The dotted lines and shaded area depict the median and dispersion (1st and 99th percentiles) of CNN predictions  $P(k)$  with (blue) and without (red) instrumental effects.



**Fig. 18.** QHI obtained from predictions at  $z_{\text{obs}}=8$  with the model  $\zeta_{30}$ , including instrumental effects for a typical SKA observation. Hatched area stands for the true field QHI. Dotted lines and shaded area depict the mean and standard deviation of CNN predictions QHI with (blue) and without (red) instrumental effects.

CNN has some difficulties to detect the first sources of ionising photons ( $t_{\text{reion}}(\mathbf{r}) < 0.4$ ), resulting from the smoothing of  $t_{\text{reion}}(\mathbf{r})$  maps. However, the maximum of the distribution is well matched at least for  $z_{\text{obs}} > 8$ , whereas low  $z_{\text{obs}} = 5.5-7$  predictors fail unsurprisingly to recover the seeds of ionisation fronts from 21 cm maps with very low non-zero signal fractions. Interestingly, an observation made for example at  $z_{\text{obs}} = 10$  manages to predict in a satisfying manner, by being consistent with dispersion of time distribution, the population of peaks (and thus seeds and/or sources of reionisation) at later times ( $t_{\text{reion}} = 1$  Gyr), emphasising the ability of our CNN to extrapolate the future of a given observation. When compared with the previous power spectrum analysis, these results emphasise that the loss of accuracy on small scales mostly have an impact on high- $z$  (low  $t_{\text{reion}}$ ) peaks,

whereas the seeds of ionisation fronts at lower redshifts are much better predicted with the lowest  $z_{\text{obs}}$ .

## 5. Instrumental effects and prediction

The work discussed previously only takes into account a ‘perfect’ 21 cm signal, without any noise or instrumental effect. Studying the effect of foreground contamination and how to manage it is of primordial importance to being able to properly analyse the future observation. While this paper does not delve into these effects, this section aims to shed some light on them. Nonetheless, there are other studies in the field of deep learning that specifically address the impact of foreground contamination, such as the work by Bianco et al. (2023). These effects are expected to degrade the predictor’s capability to infer the

$t_{\text{reion}}(\mathbf{r})$  field. As a means to study the potential impact of these effects in our predictions, we created a new dataset of 21 cm maps with instrumental and noise characteristics corresponding to SKA. The UV coverage and instrumental effects are calculated using the tools21cm<sup>4</sup> library (Giri et al. (2020)), assuming a daily scan of 6 hours, 10s integration time, for a total observation of 1000 hours (Prelogović et al. (2022b), Ghara et al. (2016), Giri et al. (2018)). Our investigation is limited to  $z_{\text{obs}}=8$ , corresponding to the lowest redshift where the predictor accuracy in terms of the  $R^2$  coefficient remains satisfying ( $R^2=0.86$ ), while deeper observations are found to be significantly more degraded by noise. A maximum baseline of 2 km is assumed and the angular resolution is  $\Delta\theta \sim 2.8$  arcmin, corresponding to 7.35 cMpc at this redshift. The tools21cm library also convolves the coeval 21 cm cube in the frequency direction with a matching resolution  $\Delta\nu \sim 0.43$  MHz. Figure 16 shows the prediction from noisy observations with the input 21 cm observations shown in the left panel. The predicted  $t_{\text{reion}}(\mathbf{r})$  map is much blurrier at first sight; adding instrumental effects on input observations smoothes the prediction even more.

Figures 17 and 18 depict the power spectrum of the  $t_{\text{reion}}(\mathbf{r})$  field and  $Q_{HI}$  for  $\zeta 30$ , for both ground truth and prediction. The two predicted curves (dotted lines) have been implemented with observation at  $z=8$ , one with instrumental noise (in blue) and the other using a perfect 21 cm map (in red). In both predictions the power spectrum is successfully recovered at large scales ( $k < 3e-2 \text{ cMpc}^{-1}h$ ), with approximately 70% and 95% of the remaining power for the noisy case and perfect case, respectively. At smaller scales ( $k > 2e-1 \text{ cMpc}^{-1}h$ ) the power spectrum recovered from noisy maps has a sharper turn-off (17% of power remaining against 86% for the perfect case) and comes out of the error bars (hatched and shaded areas). It results in missing the small-scale structures, making it difficult and even impossible, to detect the first sources of reionisation accurately. On the other hand,  $Q_{HI}$  gives a fair history of reionisation, although more sudden from SKA maps than the ground truth and the perfect signal scenario: noisy reionisation ends around  $t_{\text{reion}}=1$  Gyr and the perfect scenario ends around 1.1 Gyr.

This outcome raises several questions. The first is linked to the architecture of the CNN, and whether it is possible to improve the CNN such that the accuracy of the prediction improves significantly, especially for noisy observations. For this purpose a solution could be to tune the hyper-parameters in order to find the best combination to recover  $t_{\text{reion}}(\mathbf{r})$ . In addition, keeping the U-shaped CNN and modifying the number of hidden layers or filters, or the deepness of the algorithm could change the prediction in the right direction. Another possibility could be to add an attention block to help our CNN to focus on small-scale features (Oktay et al. (2018)). Using GANs (Ullmo et al. (2021)) could also improve the output field in order to recover small scales ( $k > 2e-1 \text{ cMpc}^{-1}h$ ). Another solution would be to preprocess the noisy 21 cm observation in order to remove or reduce noise and instrumental effects. In such a situation we would hopefully recover a perfect observation scenario, drastically boosting the prediction accuracy.

## 6. Discussion and perspectives

The work presented in this paper has involved numerous decision-making processes that may have been influenced by factors such as default settings, initial ideas, and implementation challenges. When it comes to the architecture and hyper-

parameters of the CNN algorithm, the choices made are typically based on the fact that they yield improved performance (in terms of  $R^2$ ). However, some choices, such as the number of filters, the inclusion or location of dropout layers, or the choice of loss function for weight adjustment, can potentially be modified; it is conceivable that untested combinations of hyper-parameters might yield better results. Ongoing investigations are being conducted to further explore this matter.

Another decision was to use images of the 21 cm signal at a single redshift or frequency, leading to one CNN per redshift (and per model). A possibility would have been to train the predictors with multiple redshifts channels or even light cones. This could possibly help the predictors to infer maps even in the regime of low non-zero signal fraction at low  $z_{\text{obs}} (<8)$ ; the inclusion of higher  $z_{\text{obs}}$  information in the prediction process can provide additional constraints (e.g. on the global reionisation history) that cannot be inferred from single low  $z_{\text{obs}}$  21cm maps alone. Our choice is largely the result of the history of this work, where it was not obvious at first that any prediction would have been possible, even in the case of a perfect 21cm signal. Investigations are currently ongoing to see what can be gained from a multiple channel prediction.

However, we also believe that having multiple CNNs has some merit regarding the adequacy of the parameters (cosmological, astrophysical) of a predictor to the parameters that drive a given 21cm observation. In a real case scenario, the ‘real’ parameters of an observation are unknown, and we therefore face a situation where it is unclear which CNN should be used to reconstruct  $t_{\text{reion}}(\mathbf{r})$ . One possibility is to assume that the model parameters will be obtained from another analysis (e.g. using the 21cm power spectrum) and the role of a CNN predictor is therefore limited to reconstructing the spatial distribution for the reionisation times in a specific observation. However, preliminary investigations also show that when a set of 21cm maps at different  $z_{\text{obs}}$  are processed by the multiple predictors of a ‘wrong’ model (for example  $\zeta 55$  maps in  $\zeta 30$  predictors), they lead to a set of  $t_{\text{reion}}(\mathbf{r})$  maps that are inconsistent, for example with regard to their average reionisation history. Meanwhile, a CNN that can reconstruct multiple  $t_{\text{reion}}(\mathbf{r})$  at once from multiple 21m maps would always, by construction, ensure some consistency between its predictions, even for a wrong model. The optimal situation is likely to be an intermediate situation with CNNs dedicated to reconstruct a given  $z_{\text{obs}}$ , but that use multiple 21cm maps at different redshifts.

## 7. Conclusions

In this study, we implemented and tested a U-net architecture to infer the  $t_{\text{reion}}(\mathbf{r})$  field from 21 cm maps simulated by the 21cmFAST simulation code. These predictions are especially effective to recover the large-scale features ( $> 10 \text{ cMpc}^{-1}$ ) of reionisation times and can to some extent recover the past and extrapolate the future evolution of an observation made at a given  $z_{\text{obs}}$ .

For our models,  $z_{\text{obs}}$  values between 8 and 12 seem to provide the best results according to several metrics (e.g.  $R^2$ , the Dice coefficient, power spectrum, true vs predicted), corresponding to a signal fraction of 65% up to 96% for the  $\zeta 30$  model. For  $z_{\text{obs}} < 8$ , even though the last regions to be reionised can be reconstructed, the lack of 21 cm signal in general significantly degrades the network’s capability to predict  $t_{\text{reion}}(\mathbf{r})$ . For deep observations ( $z_{\text{obs}} > 12$ ), the CNN still manages to reproduce quite well the very first sources of reionisation due to the rare and narrow HII bubbles imprinted in the 21 cm signal, but has more difficulties

<sup>4</sup> <https://github.com/sambit-giri/tools21cm>

to predict the location of sources that appear later, which leads to smoother maps. It also might be interesting to keep the information with low signal fractions ( $z_{obs} < 8$ ) since it reconstructs quite well the last regions to become reionised.

In addition, our CNN model works well at recovering the largest scales, as seen for example in the power spectrum analysis. Nevertheless, there are still some limits to what our network can do; for example, it has difficulties to recover the smallest scales ( $< 10 \text{ cMpc}^{-1}$ ). That could indeed be a problem to constrain physics related to small-scale structures (such as the physics of low-mass objects or that related to the nature of dark matter). It might be still possible to improve results at small scales with the use of GANs to generate a more detailed  $t_{\text{reion}}(\mathbf{r})$  field. In addition, implementing an attention block to insert it in our CNN could help predictors to focus on small-scale features.

Two scenarios have been used with different histories of reionisation. No significant difference in the training phase or in the prediction phase has been detected. The main difference comes at the lowest redshifts: the  $\zeta 55$  scenario reionises sooner, lacks signal more rapidly, and is more difficult to predict for low  $z_{\text{obs}}$  ( $< 8$ ).

We believe that the method presented here can prove to be useful for the future interpretation of 21 cm data. First, it demonstrates that the information of reionisation times is encoded in the 21 cm signal. The field  $t_{\text{reion}}(\mathbf{r})$  gives access to chronology of light propagation in the transverse plane of the sky, that could be for example cross-checked with other estimates of the reionisation evolution obtained along the line of sight (21 cm light cones or *Lya*/21 cm forests for example), and presumably it can be related to the global history of structure build-up and star formation. Another application would be the cross-correlation of reionisation time maps with galaxy distributions or intensity maps other than 21 cm; having access to the propagation history around objects observed through other channels could provide insights into their own local history of the production of light (and therefore of stars and sources) (see also e.g. Aubert et al. (2018), Sorce et al. (2022)). There should also be an environmental modulation of star formation suppression by reionisation (e.g. Ocvirk et al. (2020), Ocvirk & Aubert (2011)) and a map of reionisation times could provide a way to test this by providing an insight into how local reionisation proceeded. As illustrated in Sect. 5, the reconstruction of reionisation times from actual 21 cm data will be certainly be challenging, but surely holds some potential that we have not fully investigated yet.

## Acknowledgement

The authors would like to acknowledge the High Performance Computing Centre of the University of Strasbourg for supporting this work by providing scientific support and access to computing resources. Part of the computing resources were funded by the Equipex Equip@Meso project (Programme Investissements d’Avenir) and the CPER Alsacalcul/Big Data. This work was granted access to the HPC resources of TGCC under the allocations 2023-A0130411049 “Simulation des signaux et processus de l’aube cosmique et Réionisation de l’Univers” made by GENCI. We also thank J. Freundlich for his help and advice. The authors acknowledge funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No. 834148).

## References

- Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, software available from tensorflow.org
- Abdurashidova, Z., Aguirre, J. E., Alexander, P., et al. 2022, *ApJ*, 925, 221
- Aubert, D., Deparis, N., Ocvirk, P., et al. 2018, *The Astrophysical Journal*, 856, L22
- Barkana, R. & Loeb, A. 2001, *Physics Reports*, 349, 125
- Bianco, M., Giri, S. K., Iliev, I. T., & Mellema, G. 2021, *MNRAS*, 505, 3982
- Bianco, M., Giri, S. K., Prelogović, D., et al. 2023, Deep learning approach for identification of HII regions during reionization in 21-cm observations – II. foreground contamination
- Billings, T. S., La Plante, P., & Aguirre, J. E. 2021, Extracting the Optical Depth to Reionization  $\tau$  from 21 cm Data Using Machine Learning Techniques
- Chardin, J., Uhlrich, G., Aubert, D., et al. 2019, *MNRAS*, 490, 1055
- Chollet, F. 2015, keras, <https://github.com/fchollet/keras>
- Dayal, P. & Ferrara, A. 2018, *Phys. Rep.*, 780, 1
- DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, *Publications of the Astronomical Society of the Pacific*, 129, 045001
- Deparis, N., Aubert, D., Ocvirk, P., Chardin, J., & Lewis, J. 2019, *Astronomy & Astrophysics*, 622, A142
- Dice, L. R. 1945, *Ecology*, 26, 297
- Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, *Phys. Rep.*, 433, 181
- Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2004, *The Astrophysical Journal*, 613, 1
- Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2004, *ApJ*, 613, 16
- Gazagnes, S., Koopmans, L. V. E., & Wilkinson, M. H. F. 2021, *Monthly Notices of the Royal Astronomical Society*, 502, 1816
- Ghara, R., Choudhury, T. R., Datta, K. K., & Choudhuri, S. 2016, *Monthly Notices of the Royal Astronomical Society*, 464, 2234
- Ghara, R., Giri, S. K., Mellema, G., et al. 2020, *Monthly Notices of the Royal Astronomical Society*, 493, 4728
- Gillet, N. 2016, PhD thesis, thèse de doctorat dirigée par Aubert, Dominique Astrophysique Strasbourg 2016
- Gillet, N., Mesinger, A., Greig, B., Liu, A., & Ucci, G. 2019, *MNRAS*, 484, 282
- Giri, S. K. 2019, PhD thesis, Stockholm University, Department of Astronomy, at the time of the doctoral defense, the following papers were unpublished and had a status as follows: Paper 4: Accepted. Paper 5: Manuscript.
- Giri, S. K., Mellema, G., & Ghara, R. 2018, *Monthly Notices of the Royal Astronomical Society*, 479, 5596
- Giri, S. K., Mellema, G., & Jensen, H. 2020, *Journal of Open Source Software*, 5, 2363
- Gorce, A., Ganjam, S., Liu, A., et al. 2023, *Monthly Notices of the Royal Astronomical Society*, 520, 375
- Greig, B. & Mesinger, A. 2017, *MNRAS*, 472, 2651
- Greig, B., Mesinger, A., & Koopmans, L. V. E. 2015, arXiv e-prints, arXiv:1509.03312
- Hutter, A., Watkinson, C. A., Seiler, J., et al. 2019, *Monthly Notices of the Royal Astronomical Society*, 492, 653
- Iliev, I. T., Mellema, G., Shapiro, P. R., et al. 2012, *MNRAS*, 423, 2222
- Karagiannis, D., Maartens, R., & Randrianjanahary, L. F. 2022, *Journal of Cosmology and Astroparticle Physics*, 2022, 003
- Konno, A., Ouchi, M., Ono, Y., et al. 2014, *ApJ*, 797, 16
- Korber, D., Bianco, M., Tolley, E., & Kneib, J.-P. 2023, *Monthly Notices of the Royal Astronomical Society*, 521, 902
- Kulkarni, G., Keating, L. C., Haehnelt, M. G., et al. 2019, *Monthly Notices of the Royal Astronomical Society: Letters*, 485, L24
- Labach, A. & Valaee, S. 2019, Regularizing Neural Networks by Stochastically Training Layer Ensembles
- LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. 1999, *Object Recognition with Gradient-Based Learning (Berlin, Heidelberg: Springer Berlin Heidelberg)*, 319–345
- Liszt, H. 2001, *A&A*, 371, 698
- Liu, A. & Parsons, A. R. 2016, *Monthly Notices of the Royal Astronomical Society*, 457, 1864
- Loeb, A. & Barkana, R. 2001, *Annual Review of Astronomy and Astrophysics*, 39, 19
- Mao, X.-J., Shen, C., & Yang, Y.-B. 2016, Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections
- Mellema, G., Iliev, I. T., Pen, U.-L., & Shapiro, P. R. 2006, *MNRAS*, 372, 679
- Mellema, G., Koopmans, L., Shukla, H., et al. 2015, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 10
- Mellema, G., Koopmans, L. V. E., Abdalla, F. A., et al. 2013, *Experimental Astronomy*, 36, 235
- Mesinger, A., Ferrara, A., & Spiegel, D. S. 2013, *MNRAS*, 431, 621
- Mesinger, A., Furlanetto, S., & Cen, R. 2011, *MNRAS*, 411, 955
- Mondal, R., Fialkov, A., Fling, C., et al. 2020, *Monthly Notices of the Royal Astronomical Society*, 498, 4178



- Mosbech, M. R., Boehm, C., & Wong, Y. Y. Y. 2022, Probing dark matter interactions with SKA
- Muñoz, J. B., Dvorkin, C., & Cyr-Racine, F.-Y. 2020, *Phys. Rev. D*, 101, 063526
- Murray, S. G., Greig, B., Mesinger, A., et al. 2020, *Journal of Open Source Software*, 5, 2582
- Nasirudin, A., Murray, S. G., Trott, C. M., et al. 2020, *The Astrophysical Journal*, 893, 118
- Ocvirk, P. & Aubert, D. 2011, *MNRAS*, 417, L93
- Ocvirk, P., Aubert, D., Sorce, J. G., et al. 2020, *Monthly Notices of the Royal Astronomical Society*, 496, 4087
- Oh, S. P. & Haiman, Z. 2002, *ApJ*, 569, 558
- Oktay, O., Schlemper, J., Folgoc, L. L., et al. 2018, *Attention U-Net: Learning Where to Look for the Pancreas*
- Pagano, M. & Liu, A. 2020, *Monthly Notices of the Royal Astronomical Society*, 498, 373
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *Astronomy & Astrophysics*, 641, A6
- Prelogović, D., Mesinger, A., Murray, S., Fiameni, G., & Gillet, N. 2022a, *MNRAS*, 509, 3852
- Prelogović, D., Mesinger, A., Murray, S., Fiameni, G., & Gillet, N. 2022b, *MNRAS*, 509, 3852
- Ronneberger, O., Fischer, P., & Brox, T. 2015, *arXiv e-prints*, arXiv:1505.04597
- Shaw, A. K., Chakraborty, A., Kamran, M., et al. 2022, *Probing early universe through redshifted 21-cm signal: Modelling and observational challenges*
- Sorce, J. G., Ocvirk, P., Aubert, D., et al. 2022, *MNRAS*, 515, 2970
- Sousbie, T. 2011, *MNRAS*, 414, 350
- Sousbie, T., Pichon, C., & Kawahara, H. 2011, *MNRAS*, 414, 384
- Thélie, E., Aubert, D., Gillet, N., & Ocvirk, P. 2022, *A&A*, 658, A139
- Thélie, E., Aubert, D., Gillet, N., & Ocvirk, P. 2022, *Topology of Reionisation times: concepts, measurements and comparisons to gaussian random field predictions*
- Trac, H., Cen, R., & Loeb, A. 2008, *ApJ*, 689, L81
- Trac, H., Chen, N., Holst, I., Alvarez, M. A., & Cen, R. 2022, *The Astrophysical Journal*, 927, 186
- Ullmo, M., Decelle, A., & Aghanim, N. 2021, *A&A*, 651, A46
- van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, *A&A*, 556, A2
- Wise, J. H. 2019, *arXiv e-prints*, arXiv:1907.06653
- Zaldarriaga, M., Furlanetto, S. R., & Hernquist, L. 2004, *ApJ*, 608, 622
- Zhao, X., Mao, Y., & Wandelt, B. D. 2022, *Implicit Likelihood Inference of Reionization Parameters from the 21 cm Power Spectrum*
- Zhu, H., Avestruz, C., & Gnedin, N. Y. 2019, *ApJ*, 882, 152

## Appendix A: CNN architecture details and hyper-parameters

In this section we describe the details of the CNN algorithm used in this study. We used the Python libraries Tensorflow (Abadi et al. (2015)) and Keras (Chollet (2015)) to implement our CNN. Table A.1 shows properties of the hidden layers. A convolution layer consists in applying a filter of a given size (3x3 in this study) to the whole input resulting in a featured map. Each convolution is performed with same padding, meaning that each convolution conserves the size of the input. For the encoder part, the first four convolutions are followed by a Max Pooling operation (of size 2x2) that is shrinking the size of the input by a factor of 2. When the 2x2 matrix passes through the input, it only conserves the pixel with the highest value. Dropout (Drop) layers are also used in order to help the CNN to prevent overfitting (Labach & Valaee (2019)). A dropout layer acts by shutting down a given fraction (0.5 for us) of neurons/filters in the corresponding hidden layer where it is applied. For the decoder part, each convolution is followed by an UpSampling layer that doubles the size of the input. In addition, a concatenated layer (Merge or skip connection) is applied to fuse features of a given hidden layer within the encoder with features of the same size in the decoder. In practice, skip connections tend to improve the accuracy of CNN and to make it converge faster (Mao et al. (2016)). Eventually, the activation function for each layer is Relu, except for the last one that is a linear activation function since we want to predict an output field with continuous values. During the learning phase of a CNN algorithm, the weights of each convolution filter are updated each time a batch of data is passed through. In our implementation, we set the hyper-parameter batch size to 16. This means that our model updates its weights after processing each batch of 16 images, and that one epoch is completed after N batches have been processed, where N is the number of batches needed to cover the entire dataset. In addition, to optimise the performance of the CNN algorithm, we needed to carefully choose the hyper-parameters. Some of them have already been discussed, such as batch size, dropout, and loss function. The ‘optimizer’ hyper-parameter was set to Adam. Another important factor is the initial weight of the model. The ‘kernel\_initializer’ hyper-parameter controls this and was set to ‘He Normal’. However, because the weights are randomly initialised, there is a possibility that the learning process may get stuck in a local minimum without learning any more. To prevent this, we added a feature to the code that restarts the weight initialisation if such a situation is detected.

## Appendix B: Using $z_{reion}(\mathbf{r})$ instead of $t_{reion}(\mathbf{r})$

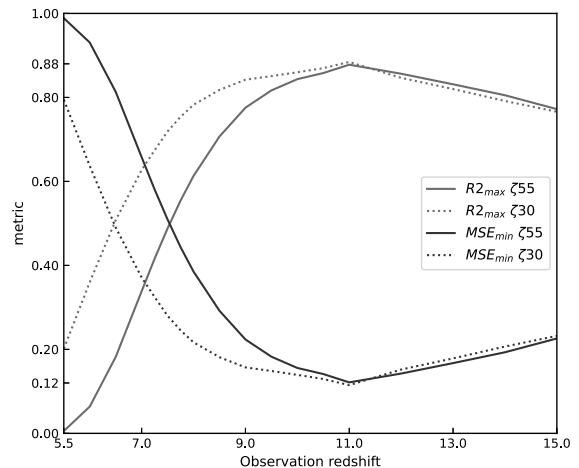
The redshift of reionisation  $z_{reion}(\mathbf{r})$  and  $t_{reion}(\mathbf{r})$  are two fields depicting the same quantity: the time of reionisation of regions in the sky, with  $z_{reion}(\mathbf{r}) \sim t_{reion}(\mathbf{r})^{-2/3}$  during the reionisation epoch. We can investigate how choosing redshift instead of time affects the capability of the CNN to predict  $z_{reion}(\mathbf{r})$  or  $t_{reion}(\mathbf{r})$  and which field gives the best results.

First, Fig. B.1 shows the same plot as Fig. 6, but for  $z_{reion}(\mathbf{r})$ . For  $z=11$  the values of the R2 and the MSE are quite similar. Nevertheless, going to lower values of  $z$  results in both metrics getting worse much faster than for  $t_{reion}(\mathbf{r})$ . Working with time and not redshift provides a greater range of redshifts for which the results are relevant.

Eventually, Fig. B.2 shows the power spectrum obtained with true and predicted  $z_{reion}(\mathbf{r})$  field. For  $\zeta_{30}$ , the true power spectrum is closely followed, especially at large scale. However, the

smaller scales turn off faster. This effect is even worse with  $\zeta_{55}$ . For  $t_{reion}(\mathbf{r})$  (Fig. 12), the smaller scales turn off more efficiently, showing that small scales are more represented when working with cosmic time instead of redshift.

The exact reason for this discrepancy is unclear. In Th  lie et al. (2022) we found that reionisation times are close to Gaussian random fields (GRFs) and can be analysed by means of GRF theory, unlike  $z_{reion}(\mathbf{r})$  which is a non-linear function of  $t_{reion}(\mathbf{r})$ . We suspect that GRFs are more easily reconstructed as they provide, for instance, a symmetric distribution of values around the mean, whereas  $z_{reion}(\mathbf{r})$ , for example, presents an asymmetric distribution of values that is affected the most from the smoothing of extrema inherent to our implementation of CNNs.

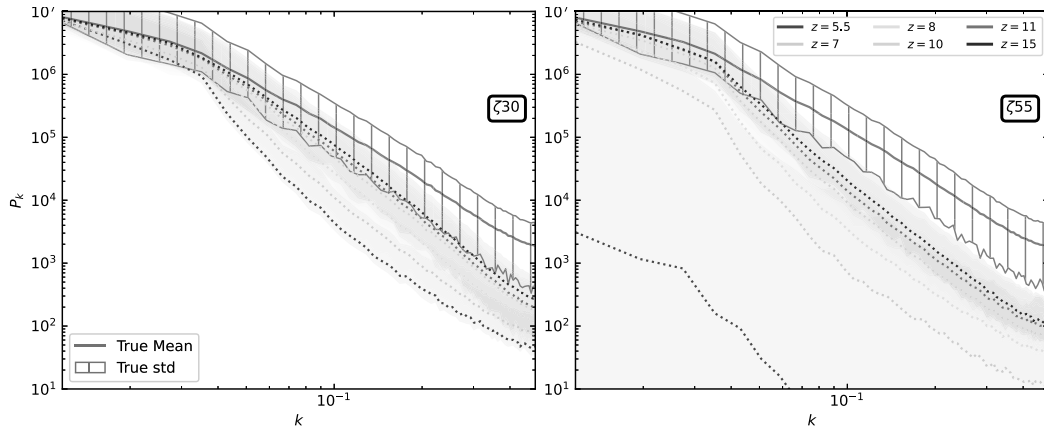


**Fig. B.1.** Metric maximum and minimum with respect to the training redshift for the validation set. The red line is for the coefficient R2 and the blue line is for the MSE. The solid lines are for  $\zeta_{55}$  and the dashed line for  $\zeta_{30}$  model.

**Table A.1.** Details of the architecture the CNN used to predict maps of  $t_{\text{reion}}(\mathbf{r})$ .

Network branch	Layer number	Layer type	Nbr of filters/data	Size of filter/data	Activation function
Encoder	1	Input	31 500	128x128	.
	2	Conv2D+Max Pooling	32	3x3	Relu
	3	Conv2D+Max Pooling	64	3x3	Relu
	4	Conv2D+Max Pooling	128	3x3	Relu
	5	Conv2D+Max Pooling+Dropout	256	3x3	Relu
	6	Conv2D+Dropout	512	3x3	Relu
Decoder	7	UpSampling+Merge+Conv2D	256	3x3	Relu
	8	UpSampling+Merge+Conv2D	128	3x3	Relu
	9	UpSampling+Merge+Conv2D	64	3x3	Relu
	10	UpSampling+Merge+Conv2D	32	3x3	Linear
	11	Output	31 500	128x128	.

Each convolution layer within the encoder part is followed by a Max Pooling layer, except the sixth. Instead, each convolution layer within the decoder part is followed by an up-sampling layer plus a Merge layer that concatenates layers of same dimension of the encoder part with the corresponding layer of the decoder.



**Fig. B.2.** Power spectrum of the  $z_{\text{reion}}(\mathbf{r})$  field obtained with the model  $\zeta_{30}$  on the left and  $\zeta_{55}$  on the right.  $k$  is the spatial frequency with dimension  $[\text{cMpc}^{-1}h]$ . The black line stands for the ground truth and the dotted coloured lines depict predictions with different training redshifts. The shaded areas depict the dispersion around the median (1st and 99th percentiles) for both prediction and ground truth.

# CHAPITRE 5

## TOPOLOGIE DES « PATCHS » DE RÉIONISATION DURANT L'ÉPOQUE DE LA RÉIONISATION

*Dans ce Chapitre, je présente une première analyse topologique des cartes 3D de redshift de réionisation. Cette étude a tout d'abord été une opportunité pour me familiariser avec les simulations cosmologiques EMMA, les simulations semi-analytiques 21cmFAST, ainsi que le code de topologie DisPerSE. J'y mesure le nombre, la taille, la forme et l'orientation des patchs de réionisation par rapport aux filaments de matière, qui sont tous extraits des simulations par DisPerSE. J'étudie ici ces patchs dans le but d'explorer des statistiques au delà du spectre de puissance par exemple, et qui comportent potentiellement des informations non présentes dans des statistiques plus simples. Ce travail est également important en ce qu'il montre l'utilité de sonder la topologie des cartes d'instant de réionisation pour la compréhension des modèles de l'Époque de la Réionisation. Ces cartes nous fournissent notamment un point de vue complémentaire aux études de la littérature faites sur des champs ionisés binaires ou de 21 cm à plusieurs redshifts. Les résultats dans ce Chapitre sont extraits de l'article « First look at the topology of reionisation redshifts in models of the epoch of reionisation » présenté en dernière Section.*

## Sommaire

---

<b>5.1</b>	<b>Introduction</b>	<b>117</b>
<b>5.2</b>	<b>Simulations 21cmFAST</b>	<b>118</b>
<b>5.3</b>	<b>Géométrie et orientation des patchs de réionisation</b>	<b>119</b>
5.3.1	Nombre et taille des patchs de réionisation	120
5.3.2	Forme des patchs de réionisation	121
5.3.3	Orientation des patchs de réionisation par rapport aux filaments de matière	124
<b>5.4</b>	<b>La persistance : un outil pour suivre la percolation des patchs de réionisation</b>	<b>129</b>
<b>5.5</b>	<b>Comparaisons entre simulations cosmologiques et semi-analytiques</b>	<b>131</b>
5.5.1	Simulations EMMA	131
5.5.2	Mesures topologiques des patchs de réionisation	132
<b>5.6</b>	<b>Conclusions &amp; perspectives</b>	<b>133</b>
<b>5.7</b>	<b>« First look at the topology of reionisation redshifts in models of the epoch of reionisation »</b>	<b>135</b>

---

## 5.1 Introduction

Cette étude résulte des premiers travaux menés pendant cette thèse et dont le but était d’explorer pour la première fois la topologie des cartes 3D de redshift de réionisation  $z_{\text{reion}}(\vec{r})$ . Ces cartes nous donnent l’instant de réionisation de l’hydrogène dans chaque cellule, nous permettant ainsi de suivre le rayonnement ionisant émis par les sources de réionisation.

Ce travail est dans la continuité de [Deparis et al. \(2019\)](#), qui ont utilisé ces cartes pour analyser la vitesse et la direction des fronts d’ionisation. Ici, nous nous intéressons également à cette propagation des fronts ionisants depuis les sources de réionisation. Les maxima de  $z_{\text{reion}}(\vec{r})$  correspondent à ces sources, et ce que nous appelons les « patches de réionisation »<sup>1</sup> sont les régions sur lesquelles ces sources ont une influence radiative. Ces patches sont porteurs d’informations : on peut par exemple s’attendre à ce que les sources qui émettent le plus de photons aient une influence à plus grande distance que leurs homologues les plus ténues. La forme des patches peut ainsi nous informer sur la manière dont se propagent les radiations ionisantes à partir des galaxies, et leur taille peut nous donner une idée du type de galaxies impliquées dans la réionisation (voir par exemple [Ocvirk et al. \(2014\)](#)). Les redshifts de réionisation sont aussi des outils pour étudier les variations locales du processus de réionisation, et sont par exemple étudiés par [Trac et al. \(2008\)](#), qui montrent des différences de redshifts de réionisation du gaz selon la densité de son environnement. [Battaglia et al. \(2013\)](#) démontre aussi que les champs de densité et de redshift de réionisation sont très corrélés et s’appuient sur cette corrélation pour développer un modèle semi-analytique de la Réionisation. Par ailleurs et en utilisant  $z_{\text{reion}}(\vec{r})$ , [Aubert et al. \(2018\)](#) et [Zhu et al. \(2019\)](#) attestent d’une réionisation à différents redshifts pour des galaxies de masses différentes, et [Sorice et al. \(2022\)](#) expliquent que des paires de halos similaires au Groupe Local peuvent avoir différentes histoires de réionisation. Les variations locales de la réionisation et cette non-homogénéité ont des conséquences sur les histoires de formation d’étoiles et leur distribution spatiale, comme en discutent par exemple [Ocvirk & Aubert \(2011\)](#) et [Ocvirk et al. \(2020\)](#). Ces études mettent ainsi en avant un processus non-homogène de réionisation, que nous proposons également d’explorer dans ce Chapitre.

Ce Chapitre commence par décrire dans la Section 5.2 les simulations 21cmFAST utilisées dans cette étude. Ensuite, la géométrie des patches de réionisation est analysée dans la Section 5.3 par des mesures de nombre, taille et forme des patches de réionisation. Leur orientation par rapport aux filaments de matière est aussi mesurée. Puis, nous explorons le phénomène de fusion et de percolation des patches dans la Section 5.4. Nous comparons également la géométrie des patches des simulations semi-analytiques 21cmFAST à celle des patches des simulations cosmologiques EMMA dans la Section 5.5. Nous concluons enfin notre travail dans la Section 5.6 tout en l’ouvrant sur quelques perspectives. La Section 5.7 présente l’article « First look at the topology of reionisation redshifts in models of the epoch of reionisation » dont sont extraits tous ces résultats.

---

1. En français, on peut traduire les « patches de réionisation » par « parcelles de réionisation ».

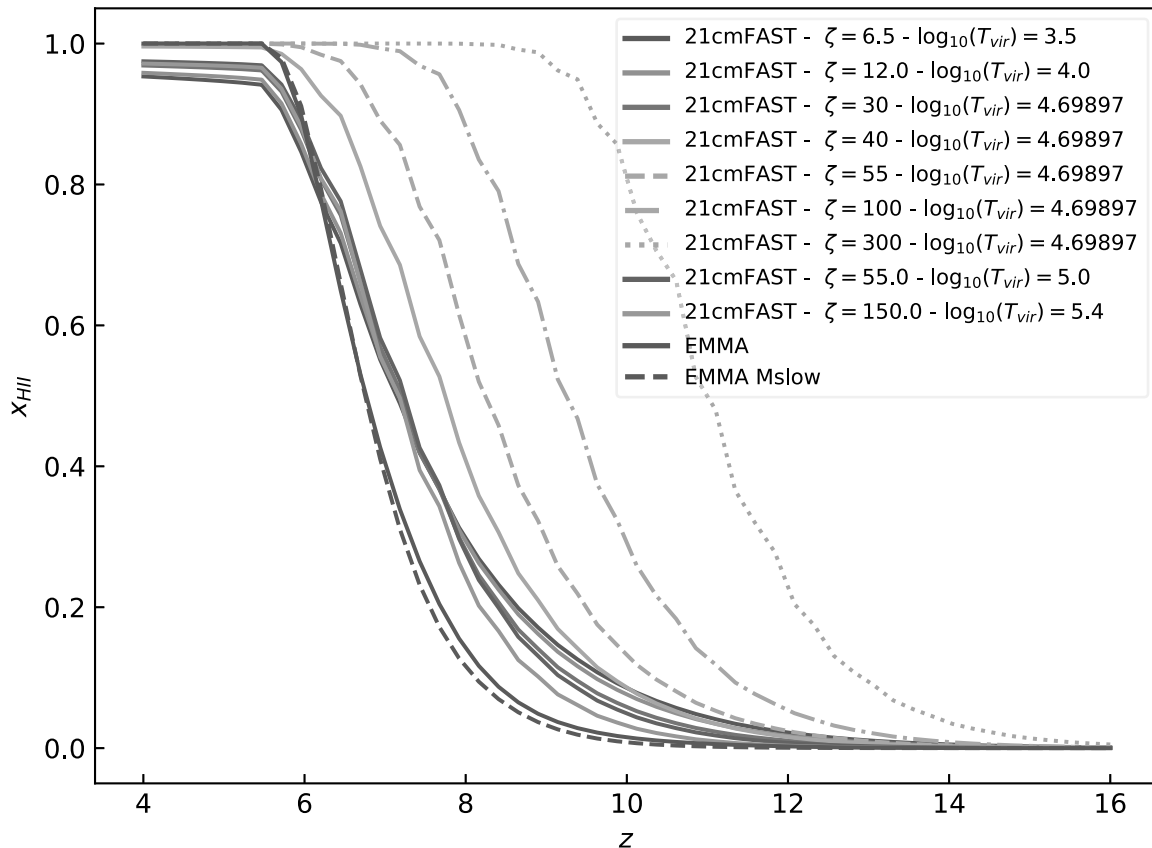
## 5.2 Simulations 21cmFAST

Notre étude se fait d'abord sur les simulations semi-analytiques 21cmFAST. Nous générons neuf jeux de 101 réalisations de simulations 21cmFAST avec des modèles de réionisation différents. Ces simulations sont de taille  $128^3 \text{ cMpc}^3/h^3$  avec une résolution de  $1 \text{ cMpc}^3/h^3$ . Dans un premier cas, et dans le contexte d'une même réionisation globale, on cherche à étudier l'effet sur la topologie des patchs de réionisation de différents seuils de masse dans la population de halos de matière noire dont les galaxies émettent des photons ionisants. Pour cela, on augmente la température de viriel  $T_{vir}$  dans les différents jeux de simulations<sup>2</sup>, ce qui implique que moins de petits halos enclenchent le processus de formation d'étoiles, signifiant aussi que moins de petits halos participent à la réionisation (ce qui la retarde). Afin que l'histoire de réionisation soit la même, nous compensons alors en augmentant l'émissivité  $\zeta$  des galaxies avec la température de viriel. Ces jeux de simulations sont ainsi générés avec les jeux de paramètres suivants :  $(\zeta, T_{vir}) \in \{(6.5, 3.2 \times 10^3 \text{ K}), (12, 1 \times 10^4 \text{ K}), (30, 5 \times 10^4 \text{ K}), (55, 1 \times 10^5 \text{ K}), (150, 2.5 \times 10^5 \text{ K})\}$ , et le modèle fiduciel est défini avec les paramètres  $(\zeta, T_{vir}) = (30, 5 \times 10^4 \text{ K})$ . On appelle ce groupe de simulations « SRH » pour « Same Reionisation History » (« Histoires de Réionisation Identiques »), et on voit en effet sur la Figure 5.1 que les histoires de réionisation de chaque modèle sont similaires (voir les courbes bleu, orange, vert foncée, violette et cyan). Ces modèles ont en effet le même redshift de milieu de réionisation, mais présentent tout de même de légères différences. Par exemple, le modèle avec la plus petite température de viriel ( $T_{vir} = 3.2 \times 10^3 \text{ K}$ , en vert) commence le processus de réionisation plus tôt mais a une histoire de réionisation plus douce que le modèle avec la plus grande température de viriel ( $T_{vir} = 2.5 \times 10^5 \text{ K}$ , en cyan). Ce dernier a une réionisation plus abrupte parce que les halos émetteurs de photons ionisants sont plus massifs et commencent à les émettre plus tardivement et en plus grande quantité.

Dans un second cas, on cherche à connaître l'influence de l'époque, précoce ou tardive, de réionisation sur la morphologie des patchs de réionisation en considérant des halos de même température de viriel, mais qui émettent plus ou moins de photons ionisants ( $\zeta$  différents). Ces jeux de simulations sont ainsi générés avec les jeux de paramètres suivants :  $(\zeta, T_{vir}) \in \{(30, 5 \times 10^4 \text{ K}), (40, 5 \times 10^4 \text{ K}), (55, 5 \times 10^4 \text{ K}), (100, 5 \times 10^4 \text{ K}), (300, 5 \times 10^4 \text{ K})\}$ . On appelle ce groupe de simulations « DRH » pour « Different Reionisation History » (« Histoires de Réionisation Différentes »). La Figure 5.1 montre bien que les histoires de réionisation de ces modèles sont différentes (voir les différentes courbes vertes) : moins les sources émettent de photons (plus  $\zeta$  est petit) et plus la Réionisation est tardive.

---

2. On rappelle que  $T_{vir}$  et la masse minimale des halos émetteurs sont reliées par  $M_{min} \propto T_{vir}^{3/2}$ . Greig & Mesinger (2015) indique que la température de viriel  $T_{vir} \sim 10^4 \text{ K}$  est généralement celle qui est choisie comme température minimale pour qu'un refroidissement atomique efficace commence. Cette température est équivalente à une masse minimale de halos émetteurs en photons ionisants de  $M_{min} \sim 10^8 M_\odot$  à un redshift d'environ 10.



**FIGURE 5.1** – Fractions d’ionisation en fonction du redshift  $z$  des simulations 21cmFAST et EMMA moyennées sur 64 réalisations de chaque modèle. On retrouve ici les différents jeux de données, avec les simulations EMMA en rouge, les simulations 21cmFAST du groupe SRH en trait plein, et les simulations 21cmFAST du groupe DRH en trait vert.

### 5.3 Géométrie et orientation des patchs de réionisation

Les études présentées dans ce Chapitre ont été menées avec le code de topologie `DisPerSE` (Discrete Persistent Structure Extractor, Sousbie (2011)). Les cartes 3D de redshifts de réionisation données en entrée de `DisPerSE` sont lissées par un filtrage gaussien sur une cellule afin qu’elles ne comportent pas de grandes zones à gradient nul qui sont difficilement prises en compte par `DisPerSE`. De plus, on fixe le seuil de persistance à  $0.5 - \sigma$  dans `DisPerSE` (sauf indication contraire).<sup>3</sup>

3. Ce seuil est exprimé en fonction de l’écart-type  $\sigma$  du champ et je l’ai choisi de manière arbitraire, étant dans le contexte d’une première analyse topologique de mes simulations. Cette valeur est toutefois du même ordre de grandeur que celles choisies dans la littérature (Sousbie et al. 2011; Codis et al. 2018; Galárraga-Espinosa et al. 2020; Cohn 2022). Une persistance très faible (proche de 0) nous permet de détecter tous les objets topologiques du champ en question, tandis qu’une persistance plus élevée effectue un filtrage pour ne garder que les objets les plus significatifs.



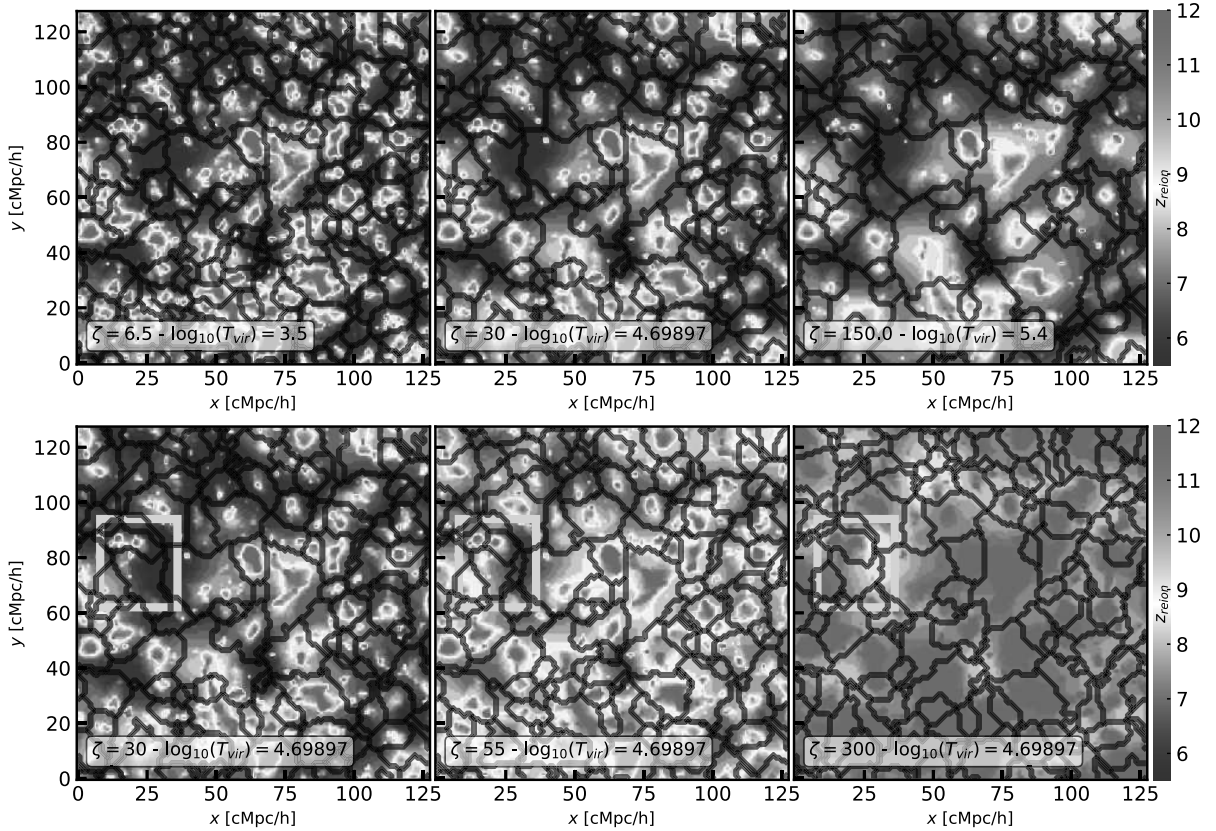
### 5.3.1 Nombre et taille des patchs de réionisation

Dans une carte de redshifts de réionisation, nous avons accès aux sources de réionisation en extrayant les maxima grâce à `DisPerSE`. Le nombre de sources de réionisation présentes dans le champ nous permet de connaître la densité des sources de réionisation et leur distribution spatiale. Chacune de ces sources de réionisation a une influence radiative sur son environnement, et cette influence est délimitée par les patchs de réionisation, montrés par exemple dans la Figure 5.2 (en noir) sur les cartes de  $z_{\text{reion}}(\vec{r})$  pour différents modèles. On voit en particulier que dans un modèle donné les patchs peuvent avoir différentes tailles et différentes formes, tout en étant en général centrés sur des endroits qui réionisent tôt (zones rouges). Le processus de réionisation n'est visiblement pas homogène ni isotrope dans ces cartes.

Topologiquement, ces patchs sont tous associés à un maximum différent. Tous les points qu'ils contiennent se trouvent sur des lignes de gradient qui ont pour origine le maximum du patch. Cela correspond mathématiquement à des variétés descendantes de type 3, et `DisPerSE` nous permet de les extraire des cartes de redshift de réionisation. On compte ainsi autant de patchs que de maxima ou de sources de réionisation dans le champ. Dans une simulation d'une taille donnée, le nombre de patchs est généralement inversement proportionnel à leur taille : plus il y a de patchs et plus ils sont petits. Par exemple, si très peu de patchs sont présents, très peu de sources de réionisation sont alors présentes et le rayonnement qu'elles produisent ont une grande zone d'influence, autrement dit de grands patchs de réionisation. C'est le comportement que l'on retrouve dans les cartes du haut de la Figure 5.2, où sur celle de gauche il y a plus de patchs que sur celle de droite, où ils y sont plus petits. Dans les deux groupes de simulations, même si la taille des patchs de réionisation varie selon le modèle, on mesure une taille typique de régions d'environ 10 cMpc/h.

La Figure 5.2 montre des cartes des simulations SRH sur la première ligne. Pour rappel, ces simulations ont la même histoire de réionisation, mais des populations de halos émetteurs différentes ( $\zeta$  et  $T_{\text{vir}}$  augmentant tous les deux). On remarque ici que plus  $T_{\text{vir}}$  est grand, plus le nombre de patchs de réionisation diminue (de gauche à droite). Les patchs sont donc plus grands quand  $T_{\text{vir}}$  augmentent, ce qui signifie que les sources de réionisation ont une influence à plus grande distance.

Dans le cas des simulations DRH avec des réionisations plus ou moins tardives ( $T_{\text{vir}}$  fixe et  $\zeta$  qui augmente), on s'attend à ce que le nombre de patchs de réionisation diminue avec l'efficacité  $\zeta$  des galaxies à émettre des photons. En effet, quand  $\zeta$  augmente, les sources produisent plus de photons ionisants, qui peuvent donc atteindre des régions plus éloignées. Les premiers patchs vont alors s'étendre rapidement empêchant les halos moins émetteurs de photons ou qui se forment plus tard de créer leur propre patch de réionisation. Ces plus petits halos sont donc « réionisés » par des sources extérieures et sont localisés dans les patchs formés par ces premières sources. La Figure 5.2 montre sur la ligne du bas trois modèles du groupe DRH. On y voit bien une diminution du nombre de patchs de réionisation quand l'efficacité  $\zeta$  des galaxies à émettre des photons augmente de 55 à 130 (avec des régions qui se sont regroupées, comme dans le cadre vert). Des effets numériques peuvent aussi se manifester : le nombre de patchs augmente quand  $\zeta$  augmente de 30 à 55 (par exemple des régions apparaissent dans le cadre vert). On attribue cette augmentation à notre processus d'extraction des patchs, en particulier au lissage gaussien appliqué aux boîtes et au filtre de

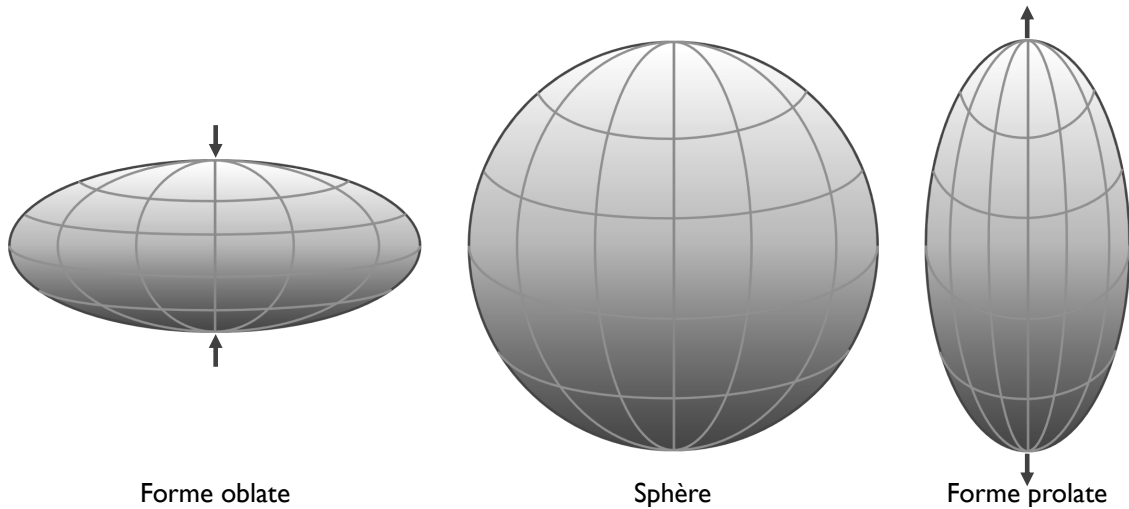


**FIGURE 5.2** – Cartes de redshifts de réionisation provenant de simulations 21cmFAST. Les cartes du haut font partie du groupe SRH (avec les jeux de paramètres respectifs  $(\zeta, T_{vir}) \in \{(6.5, 3.2 \times 10^3 \text{ K}), (30, 5 \times 10^4 \text{ K}), (150, 2.5 \times 10^5 \text{ K})\}$  de gauche à droite), et celles du bas, du groupe DRH (avec les jeux de paramètres respectifs  $(\zeta, T_{vir}) \in \{(30, 5 \times 10^4 \text{ K}), (55, 5 \times 10^4 \text{ K}), (300, 5 \times 10^4 \text{ K})\}$  de gauche à droite). Les contours noirs délimitent les patchs de réionisation extraites par DisPerSE. On voit ici que les patchs ont des tailles et des formes variées pour un même modèle. De plus, pour les cartes du haut (du groupe SRH) par exemple, on voit que plus les sources émettrices sont massives (de la gauche vers la droite), plus les patchs sont grands et moins ils sont nombreux.

la persistance. Le lissage filtre les plus petits patchs qui contiennent de faibles émetteurs et qui ont alors de faibles gradients de  $z_{reion}(\vec{r})$ . Ces gradients sont par contre amplifiés quand l'émissivité  $\zeta$  augmente et sont alors toujours présents dans les champ lissés et filtrés par la persistance avec les modèles à plus grand  $\zeta$ .

### 5.3.2 Forme des patchs de réionisation

Des exemples de patchs de formes très variées se manifestent dans les cartes de redshift de réionisation dans la Figure 5.2. La forme des patchs nous informe sur la direction de propagation des fronts ionisants et/ou sur la distribution des sources de réionisation. Le rayonnement ionisant ne semble ainsi pas se propager de manière isotrope et résulter d'un processus non-homogène, comme mentionné en introduction du Chapitre (Section 5.1).



**FIGURE 5.3** – Schéma des différents types de forme d'une ellipsoïde. Au milieu, une sphère est illustrée. Si on l'aplatit, elle prend alors une forme oblate (ellipsoïde de gauche), et si on l'étire, elle prend alors une forme prolata (ellipsoïde de droite).

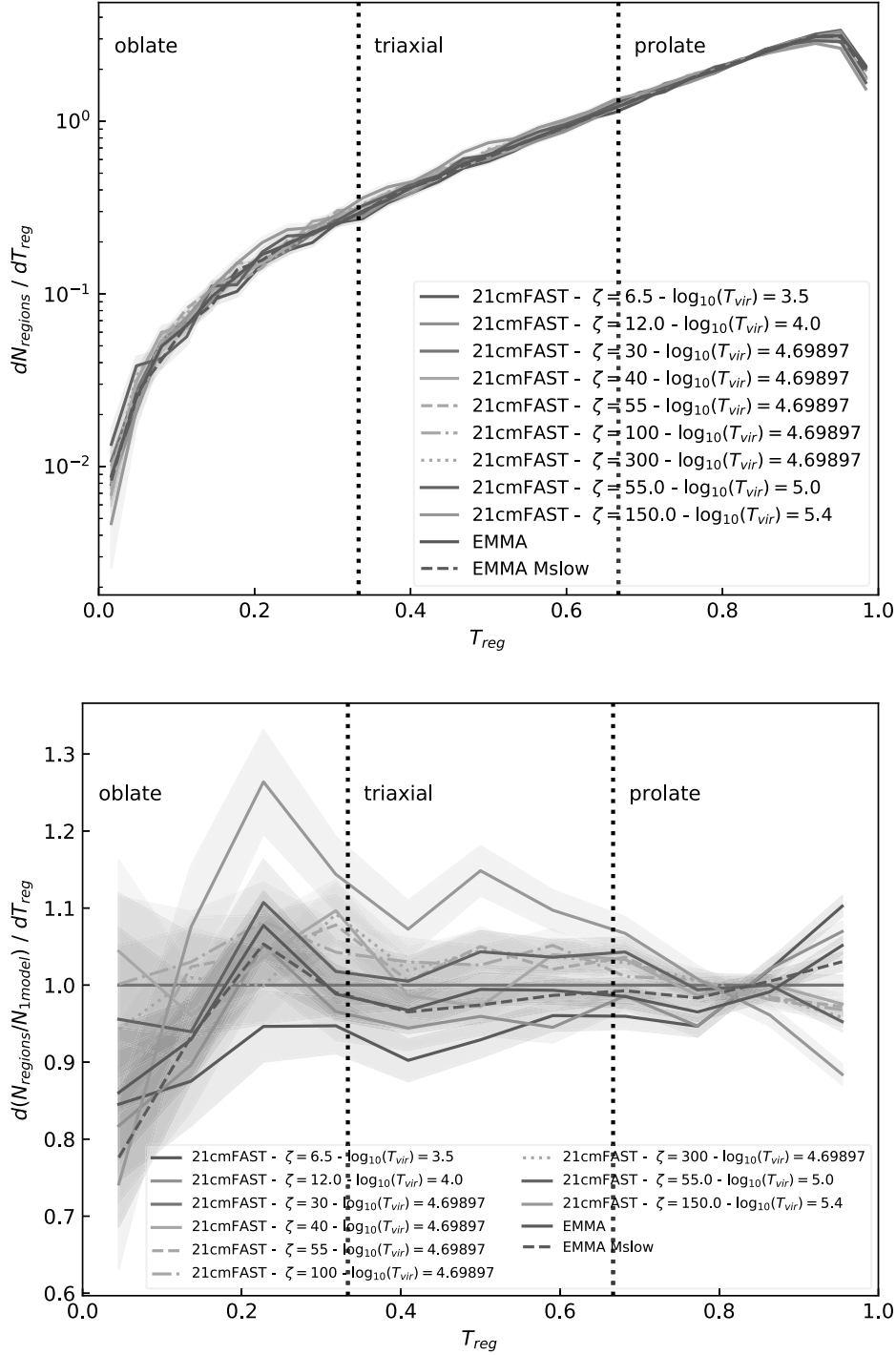
La mesure de la forme des patchs peut se faire par le biais du tenseur d'inertie. Pour un patch donné, on définit le tenseur comme suit :

$$I_{i,j} = \sum_k \omega_k \cdot (x_{k,i} - x_{ref,i}) \cdot (x_{k,j} - x_{ref,j}). \quad (5.1)$$

$k$  est un indice parcourant toutes les cellules du patch.  $i, j \in \{1, 2, 3\}$  sont des indices indiquant la composante cartésienne de la position de la cellule en 3D.  $x_{ref}$  est un point de référence, choisit ici comme étant le centre géométrique du patch. Ici, nous choisissons  $w_k = 1$  pour chaque cellule du patch car seule sa forme géométrique nous intéresse dans cette étude. À partir des valeurs propres  $\lambda_i$  (pour  $i \in \{1, 2, 3\}$ ) du tenseur d'inertie, on peut estimer la forme d'un patch en calculant une grandeur appelée le « paramètre de triaxialité ». Nous le définissons comme Tormen (1997) :

$$T = \frac{\lambda_3^2 - \lambda_2^2}{\lambda_3^2 - \lambda_1^2}, \quad (5.2)$$

avec l'ordre des valeurs propres suivant :  $\lambda_1 \leq \lambda_2 \leq \lambda_3$ . Dans le cas où  $\lambda_1, \lambda_2 \ll \lambda_3$ , alors une direction est très fortement privilégiée par rapport aux autres et  $T$  sera proche de 1. On définit alors un patch comme étant prolata (ou allongé, en forme de balle de rugby, voir Figure 5.3) si  $\frac{2}{3} < T < 1$ . Si  $\lambda_1 \ll \lambda_2, \lambda_3$ , alors une direction est sous-dominante et  $T$  sera proche de 0. On définit alors un patch comme étant oblate (comme une sphère aplatie, voir Figure 5.3) si  $0 < T < \frac{1}{3}$ . Enfin, si  $\lambda_1 \simeq \lambda_2 \simeq \lambda_3$ , alors toutes les directions ont des extensions similaires, on est alors dans le cas triaxial et  $\frac{1}{3} < T < \frac{2}{3}$ . Dans le cas des patchs de réionisation, les directions privilégiées ou sous-dominantes résultent de l'étendue de l'influence radiative des sources de réionisation et nous permettent de savoir comment se répartissent les sources et/ou comment se propage le rayonnement ionisant.



**FIGURE 5.4** – Fonctions de distribution de probabilité du paramètre de triaxialité (première ligne). La seconde ligne représente les mêmes PDFs mais divisées par la PDF du modèle fiduciel de 21cmFAST (avec  $\zeta = 30$  et  $T_{vir} = 5 \times 10^4$  K), et permet de voir les différences fines entre chaque modèle. Tous les modèles des simulations 21cmFAST et EMMA sont montrés ici en additionnant toutes les régions de 64 réalisations pour chaque modèle. On voit que la majorité des patchs de réionisation de tous les modèles ont une forme prolate.

Dans les mesures statistiques du paramètre de triaxialité montré en haut de la Figure 5.4, on compte une majorité de patchs prolates pour tous les modèles générés avec 21cmFAST, soit autour de 70% des patchs de chaque modèle. Le rayonnement suit donc une direction privilégiée depuis les sources de réionisation ou alors les sources sont distribuées selon cette direction particulière. De plus, dans le cas des simulations SRH, on remarque une légère tendance dans les rapports des PDFs en bas de la Figure 5.4 (courbes bleu et cyan par exemple) : plus  $T_{vir}$  augmente et moins les régions sont prolates, montrant ainsi que la masse minimale des halos de matière noire qui émettent des photons ionisants a une influence radiative sur l'environnement et/ou la manière dont sont distribuées les sources. Pour les simulations DRH (courbes vertes), on ne remarque pas de tendance particulière selon les modèles, indiquant une distribution des sources équivalentes et/ou que le rayonnement se propage globalement de la même manière pour chaque modèle, mais pas au même instant (voir le décalage des histoires de réionisation dans la Figure 5.1).

### 5.3.3 Orientation des patchs de réionisation par rapport aux filaments de matière

On sait maintenant que les patchs de réionisation sont plutôt de forme prolate, c'est-à-dire allongée. La question qui se pose alors est la suivante : pour quelle raison les patchs ont-ils cette forme ? Est-ce dû simplement au rayonnement qui se propage dans une direction particulière ? Ou est-ce que la distribution des sources de réionisation est anisotrope ? Pour répondre à ces questions, je me suis alors demandé si la distribution de matière à grande échelle jouait un rôle dans cette réionisation anisotrope. On sait en effet qu'aux échelles des grandes structures de l'Univers, la matière se distribue dans des filaments formant un squelette, connu sous le nom de la « toile cosmique » (Geller & Huchra 1989; Bond et al. 1996; Springel et al. 2006). La suite de cette étude porte donc sur une comparaison de l'orientation des patchs de réionisation par rapport aux filaments de matière. On peut les extraire des cartes de densité de matière de nos simulations grâce à DisPerSE (ils forment en fait le squelette topologique du champ de densité). On compare ici l'orientation de chaque patch par rapport au filament dominant du patch. Naïvement, on s'attendrait par exemple à ce que les photons ionisants s'échappent des sources de manière perpendiculaire aux filaments, en formant ce qu'on appelle des « papillons ». En effet, la matière est dense le long des filaments, donc on imagine plutôt que les photons vont suivre le parcours de moindre résistance dans une direction perpendiculaire.

La mesure de l'orientation des patchs par rapport aux filaments de matière peut se faire par le biais du tenseur d'inertie défini dans l'équation 5.1. Pour les tenseurs d'inertie des filaments,  $w_k$  correspond à la densité de la cellule  $k$ . Les vecteurs propres d'un tenseur d'inertie nous donnent les directions favorisées dans l'espace 3D du patch ou du filament, et les valeurs propres nous informent de l'étendue de leur extension le long de ces directions. Le vecteur propre associé à la plus grande valeur propre représente ainsi la direction principale du patch ou du filament. L'orientation des patchs de réionisation par rapport au filament de matière est donné par le cosinus de l'angle entre ces deux objets (qui est mesuré entre la direction

principale du patch de réionisation et celle du filament). On note alors :

$$\cos(\langle z_{reion}, \delta \rangle) = \cos(\langle \vec{\Psi}_{z_{reion},3}, \vec{\Psi}_{\delta,3} \rangle), \quad (5.3)$$

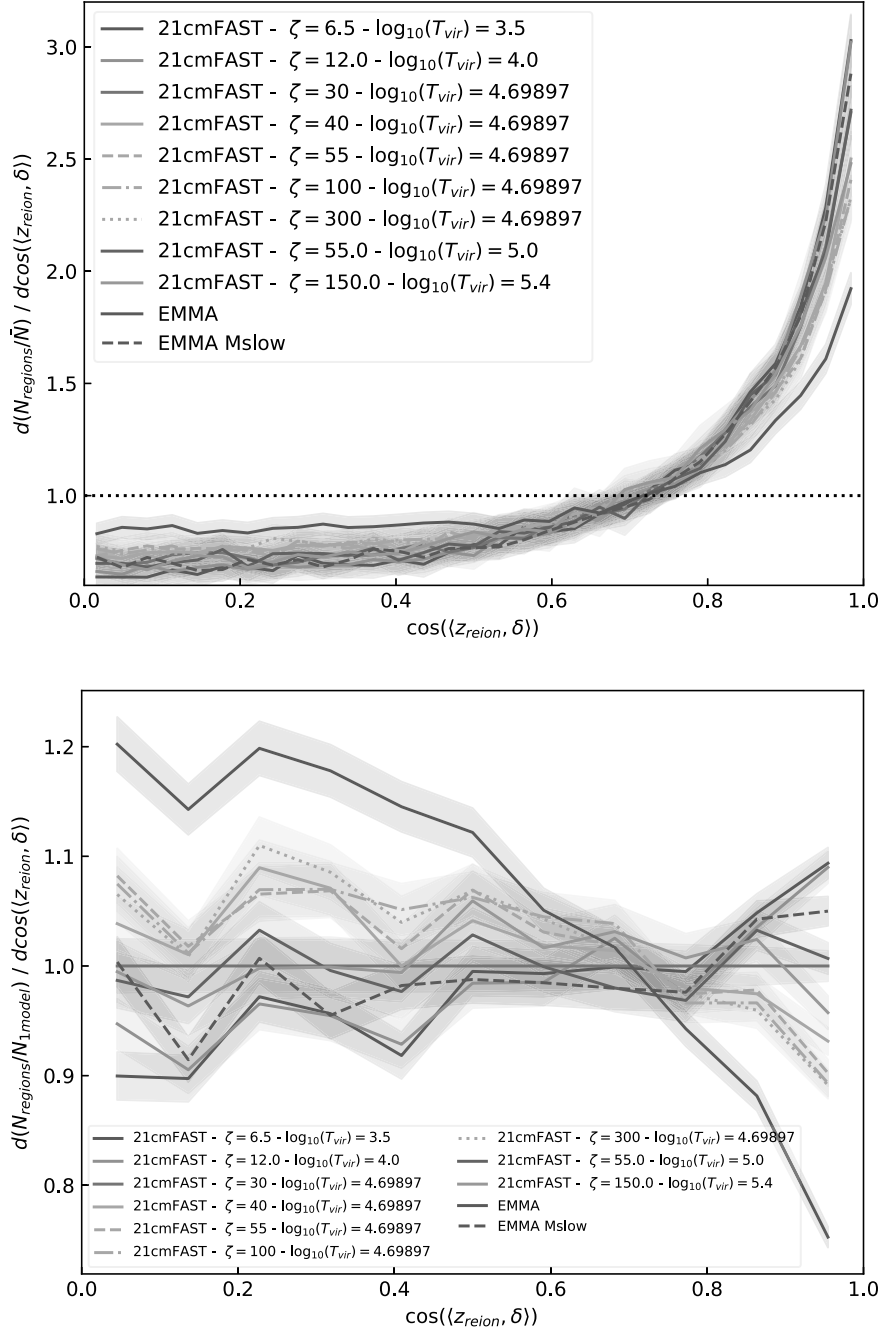
$$= \frac{\vec{\Psi}_{z_{reion},3} \cdot \vec{\Psi}_{\delta,3}}{\|\vec{\Psi}_{z_{reion},3}\| \|\vec{\Psi}_{\delta,3}\|}, \quad (5.4)$$

où  $\vec{\Psi}_{z_{reion},3}$  et  $\vec{\Psi}_{\delta,3}$  sont les vecteurs propres respectifs des patchs du champ de redshift de reionisation et des filaments du champ de densité, et ces vecteurs correspondent à leur valeur propre la plus grande ( $\lambda_3$ ). En cas de distribution aléatoire des orientations des patchs de réionisation par rapport aux filaments de matière, on s'attend à une distribution uniforme de  $\cos(\langle z_{reion}, \delta \rangle)$ .

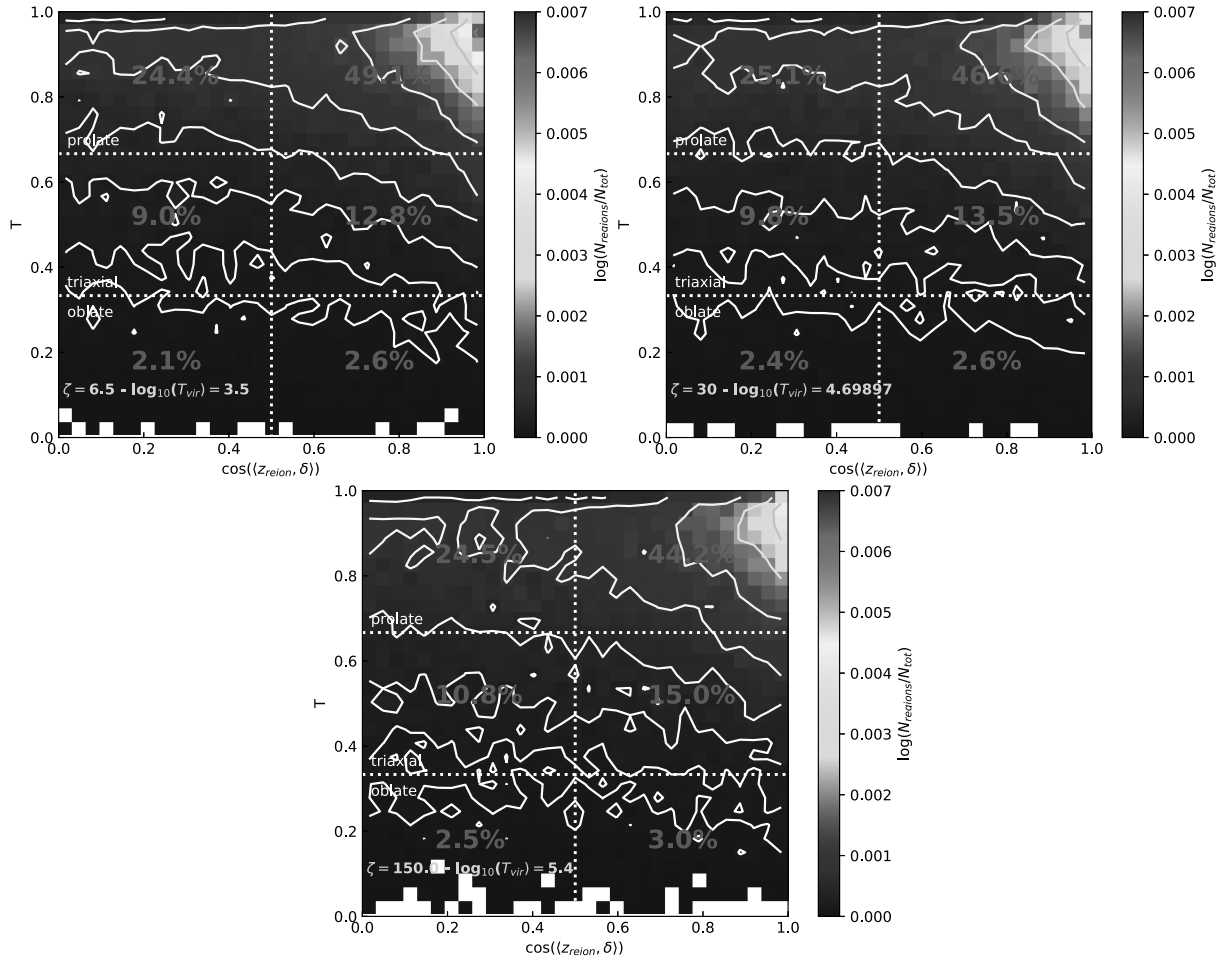
Pour toutes les simulations SRH et DRH, on compte une majorité de régions alignées aux filaments de matière, comme le montre la première ligne de la Figure 5.5. Le front de réionisation global de ces régions (qui est potentiellement un composite de plusieurs fronts qui proviennent d'une ou plusieurs sources de réionisation) se dirige donc le long du filament de matière. En regardant plus finement sur la deuxième ligne de la Figure 5.5, on remarque une légère tendance dans la groupe de simulations SRH (courbes bleu et cyan par exemple) : plus la masse minimale des halos de matière noire qui émettent des photons ionisants augmente et moins les régions sont alignées aux filaments et on voit même qu'elles sont légèrement plus perpendiculaires. Pour le groupe de simulations DRH, on n'observe pas de tendance particulière selon les modèles, montrant qu'une réionisation précoce ou tardive n'impacte pas la manière dont le gaz se fait réioniser.

De plus, la Figure 5.6 montre des histogrammes 2D de la forme des patchs de réionisation en fonction de leur orientation par rapport aux filaments de matière pour 3 modèles de simulations SRH. On voit ici que le cas des patchs prolates et alignés aux filaments de matière domine toutes les autres configurations. Dans notre modèle fiduciel ( $\zeta = 30$  et  $T_{vir} = 5 \times 10^4$  K), on compte 46.6% de ces patchs (que l'on définit par  $\frac{2}{3} < T < 1$  et  $0.5 \leq \cos(\langle z_{reion}, \delta \rangle) \leq 1$ ). À l'inverse, la configuration la moins probable est celle des patchs oblates et perpendiculaires aux filaments de matière, que l'on appelle des « papillons ». On en compte 2.4% dans le modèle fiduciel (que l'on définit par  $0 < T < \frac{1}{3}$  et  $0 \leq \cos(\langle z_{reion}, \delta \rangle) \leq 0.5$ ). Les autres 51% des configurations possibles sont répartis dans l'espace des paramètres  $T$  versus  $\cos(\langle z_{reion}, \delta \rangle)$ , avec une dominance notable des patchs de forme prolate, montrant ainsi qu'il existe une direction préférentielle des patchs de réionisation.

Comme nous nous attendions plutôt à avoir plus de « papillons » dans les simulations que ce que nous en mesurons, nous avons été un peu surpris par ce résultat. Nous avons donc émis des hypothèses quant aux possibles configurations qui mèneraient à ces formes et orientations de patchs de réionisation par rapport aux filaments de matière, et la Figure 5.7 illustre ces possibilités. On suppose ainsi qu'une majorité des sources sont perlées le long des filaments et certaines sont associées à un seul patch de réionisation par DisPerSE. Dans ce cas, on est dans le cas des patchs prolates-alignés, schématisés à droite de la Figure 5.7 et dont un exemple dans nos simulations est donné à droite de la Figure 5.8. On imagine que les sources le long du filament ont des propriétés relativement similaires : par exemple, elles apparaissent en même temps, elles ont la même émissivité ou la même masse. La première source à apparaître est un maximum local dans son environnement et se voit associée son propre patch de réionisation, qui contient alors les autres sources qui sont apparues légèrement



**FIGURE 5.5** – Fonctions de distribution de probabilité du cosinus de l’angle entre les patches de réionisation et leur filament de matière correspondant (première ligne). La seconde ligne représente les mêmes PDFs mais divisées par la PDF du modèle fiduciel de 21cmFAST (avec  $\zeta = 30$  et  $T_{\text{vir}} = 5 \times 10^4$  K), et permet de voir les fines différences entre chaque modèle. Tous les modèles des simulations 21cmFAST et EMMA sont montrés ici en additionnant toutes les régions de 64 réalisations pour chaque modèle. On voit que la majorité des patches de réionisation de tous les modèles sont alignés avec le filament de matière qui les traverse.

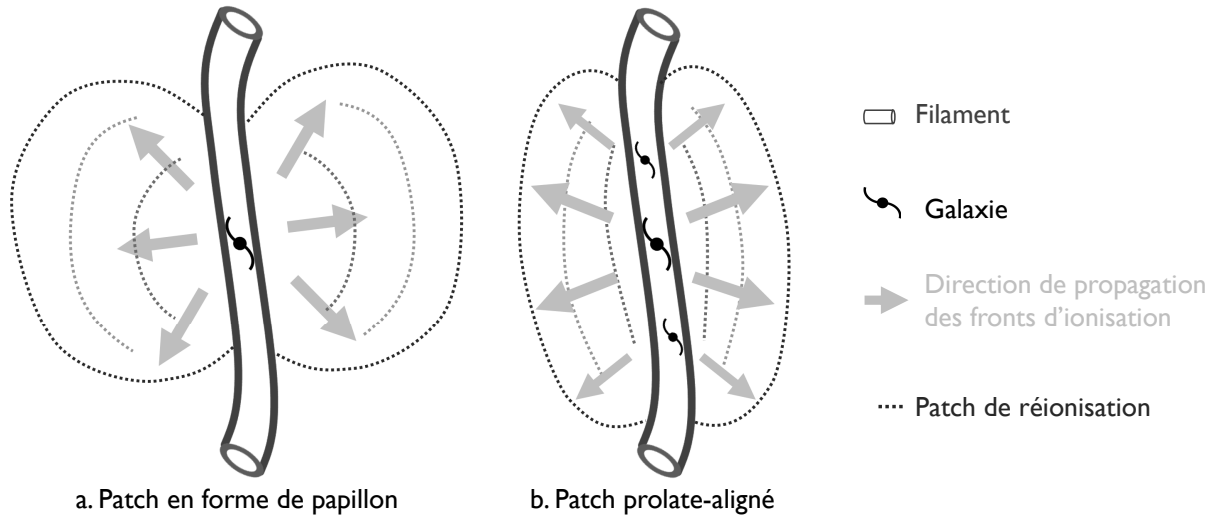


**FIGURE 5.6** – Histogrammes 2D représentant le logarithme du rapport du nombre de patchs de réionisation sur le nombre total de régions dans des tranches de valeur du paramètre de triaxialité et du cosinus de l’angle entre les patchs et les filaments de matière. Les lignes blanches sont les isocontours de l’histogramme. Les cellules colorées représentent les décomptes non nuls de patchs dans l’histogramme, et les cellules blanches indiquent des 0. Chaque histogramme représente un modèle différent parmi le jeu de modèle ayant la même histoire de réionisation (avec  $(\zeta, T_{vir}) \in \{(6.5, 3.2 \times 10^3 \text{ K}), (30, 5 \times 10^4 \text{ K}), (150, 2.5 \times 10^5 \text{ K})\}$ ). On voit que la majorité des patchs de réionisation de tous les modèles sont à la fois prolates et alignés au filament de matière qui les traverse.

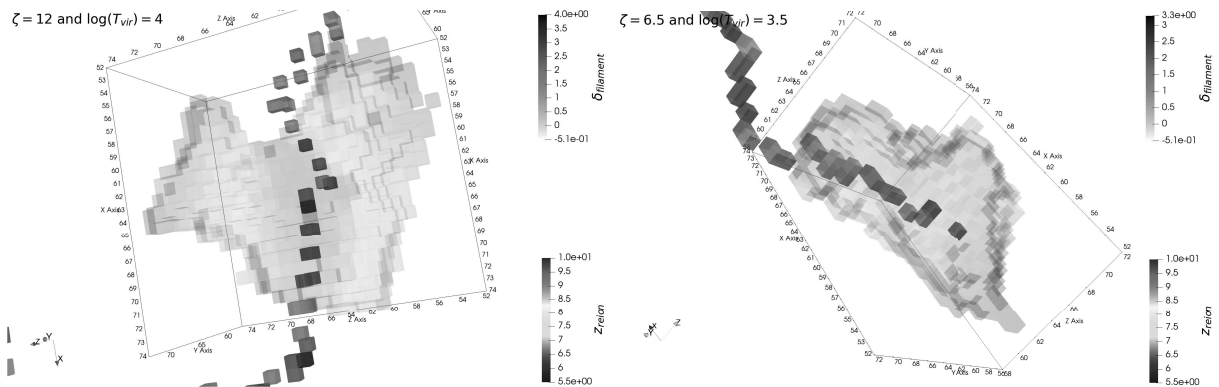
plus tard (mais pas assez de temps après pour être suffisamment significatives par rapport au seuil de persistance de DisPerSE pour avoir leur propre patch). Dans cette configuration, le rayonnement UV émis par toutes ces sources va faire partie du même patch et ainsi l’allonger le long du filament de matière. Ainsi, même si les fronts d’ionisation peuvent se propager un à un perpendiculairement par rapport à leur source respective, ils sont associés dans un seul patch de forme prolate et aligné au filament.

À l’inverse, dans le cas des papillons illustrés à gauche de la Figure 5.7, on suppose qu’ils correspondent aux rares sources de réionisation très brillantes et/ou isolées de nos simula-





**FIGURE 5.7** – Schéma de patchs de réionisation (traits pointillés bleus foncés) représentés dans deux configurations différentes. À gauche, le patch est orienté perpendiculairement au filament de matière (tube violet) : on parle de patch « papillon ». À droite, le patch est aligné le long du filament de matière : on parle de patch prolate-aligné. Les sources de réionisation sont représentées en noir et la direction de propagation des fronts d'ionisation est indiquée par les flèches jaunes.



**FIGURE 5.8** – Exemple de patchs de réionisation en forme de papillon sur la gauche. Sur la droite, on montre un exemple de patch prolate et aligné au filament de matière. Ces exemples sont extraits de simulations 21cmFAST. Les couleurs arc-en-ciel représentent les redshifts de réionisation des patchs, et les couleurs rose représentent la densité des filaments de matière.

tions et qui dominent leur environnement. Leur rayonnement se propage dans la direction perpendiculaire au filament car c'est la direction de plus faible résistance pour les photons, le filament étant une région dense de matière. Un papillon extrait des simulations est montré à gauche de la Figure 5.8, et on voit bien la propagation des fronts d'ionisation perpendiculaire au filament, en regardant les valeurs de  $z_{\text{reion}}(\vec{r})$  qui diminuent plus on s'éloigne du filament.

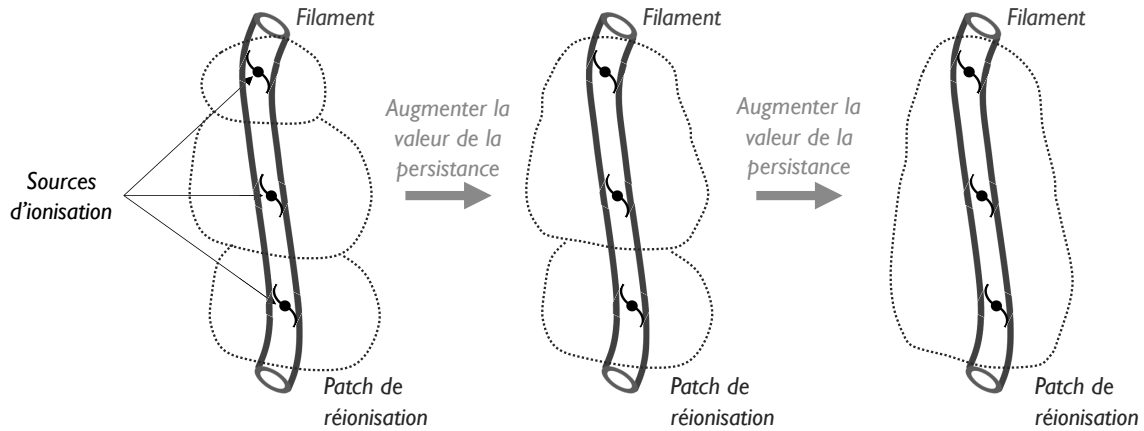
Finalement, pour les simulations SRH, on mesure une légère tendance en faveur d'un plus petit nombre de patchs prolates et alignés aux filaments et d'un plus grand nombre de papillons quand  $T_{\text{vir}}$  augmente (voire le panneau du bas de la Figure 5.5 ou la Figure 5.6). Cela indique que quand la masse minimale des halos qui émettent des photons ionisants est plus grande, les sources vont avoir légèrement plus tendance à dominer leur environnement radiatif. Dans le cas du groupe des DRH faisant varier les histoires de réionisation, on ne distingue de nouveau pas de tendances particulières selon les modèles.

## 5.4 La persistance : un outil pour suivre la percolation des patchs de réionisation

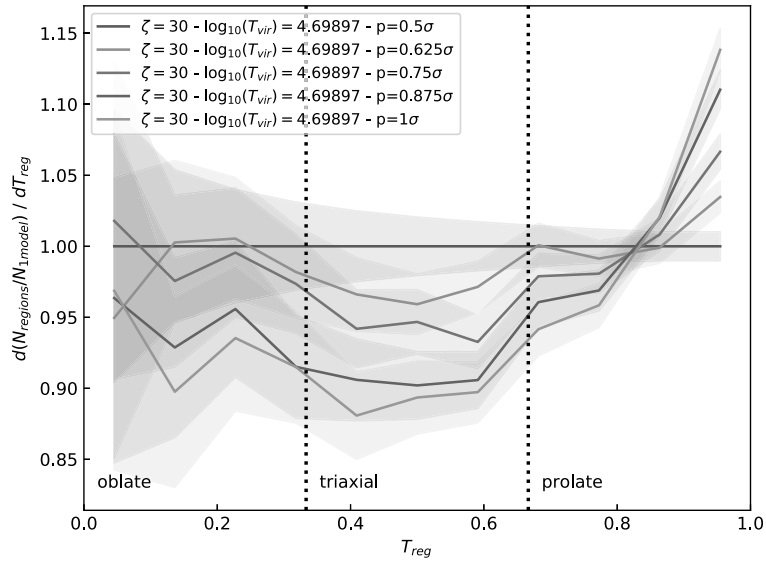
J'ai montré précédemment que pendant la Réionisation les patchs sont majoritairement prolates et alignés aux filaments de matière parce que les sources de réionisation sont perlées le long des filaments de matière. On peut alors se demander comment ces patchs évoluent au cours du temps et si l'on retrouve les mêmes comportements qu'en suivant l'évolution des bulles ionisées pendant la Réionisation. La croissance des bulles lors de la Réionisation est un phénomène très étudié dans la littérature (voir le Chapitre 3) parce qu'elle nous permet de suivre l'ionisation de l'hydrogène tout au long de la Réionisation. Les bulles ionisées sont également des traceurs des propriétés des sources qu'elles contiennent. Ces dernières ont directement une influence sur la taille et la morphologie des bulles par exemple (Furlanetto et al. 2004a; McQuinn et al. 2007; Friedrich et al. 2011; Giri et al. 2019).

Dans cette étude, nous présentons une méthode basée sur le seuil de persistance de `DisPerSE`, qui nous permet de faire une première étude de la fusion et de la percolation des patchs de réionisation, et ainsi d'explorer de nouveau le perlage des sources de réionisation le long des filaments de matière. Comme détaillé dans le Chapitre 2, ce seuil contrôle la différence entre les valeurs de deux points critiques (par exemple des maxima pour  $z_{\text{reion}}(\vec{r})$ ) détectés par `DisPerSE`, ignorant ainsi les points critiques qui ne sont pas assez significatifs (voir la Figure 3.11). Pour n'importe quel champ en entrée de `DisPerSE`, augmenter la persistance a alors l'effet d'extraire moins de points critiques, et donc moins de patchs associées à ces points critiques. Le schéma de la Figure 5.9 est une illustration de ce processus, appliqué au cas des patchs de réionisation extraits de  $z_{\text{reion}}(\vec{r})$ . On voit en effet qu'en augmentant la persistance (de gauche à droite sur le schéma), on extrait alors moins de maxima, et les patchs de réionisation sont en moins grand nombre et plus étendus. Cela signifie alors qu'en augmentant la persistance, on observe les patchs de réionisation qui fusionnent. Un seuil de persistance appliquée à  $z_{\text{reion}}(\vec{r})$  est alors une mesure relative de la différence entre les redshifts de création de deux sources. On peut ainsi effectivement suivre la fusion des patchs et leur percolation en suivant l'effet de la persistance.

En mesurant de nouveau la forme et l'orientation des patchs de réionisation par rapport aux filaments de matière dans le modèle fiduciel ( $\zeta = 30$  et  $T_{\text{vir}} = 5 \times 10^4$  K), on montre



**FIGURE 5.9** – Schéma illustrant le processus de percolation via des changements du seuil de persistance. Un filament de matière est représenté en violet contenant plusieurs sources de réionisation (en noir). Ces sources de réionisation produisent leur propre patch de réionisation (en pointillés bleus), et quand la persistance augmente (de gauche à droite), on voit que les patches ont fusionné pour donner de plus grandes patches qui sont associés à la source la plus significative.



**FIGURE 5.10** – Rapport entre les fonctions de distribution de probabilité du paramètre de triaxialité du modèle fiduciel 21cmFAST (avec  $\zeta = 30$  et  $T_{\text{vir}} = 5 \times 10^4$  K) pour différentes valeurs de persistance et celle obtenue avec la persistance  $0,5 - \sigma$ . On additionne dans ces distributions toutes les régions des 101 réalisations de chaque modèle. En regardant finement avec ce type de courbes, on voit que plus la persistance est grande, plus les patches de réionisation ont tendance à être prolates, nous permettant ainsi de suivre leur percolation.

encore qu'ils sont majoritairement prolates et alignés aux filaments pour tous les niveaux de persistance. On comprend aussi avec le schéma de la Figure 5.9 que plus la persistance est grande et plus les patchs vont avoir tendance à s'allonger et s'aligner avec le filament. On observe une légère tendance dans ce sens dans les mesures. Par exemple, la Figure 5.10 montre que les patchs sont de plus en plus prolates quand la persistance augmente. Cela signifie alors qu'ils ont tendance à fusionner sous forme de fibre ionisée le long des filaments de matière. On retrouve la même évolution que pour les bulles ionisées pendant la Réionisation mentionnée par Chen et al. (2019). Selon ces auteurs, la Réionisation évolue en plusieurs phases et pendant les deux premières phases, les bulles fusionnent et forment des fibres avant de percoler. Il semble que nous retrouvons le même effet.

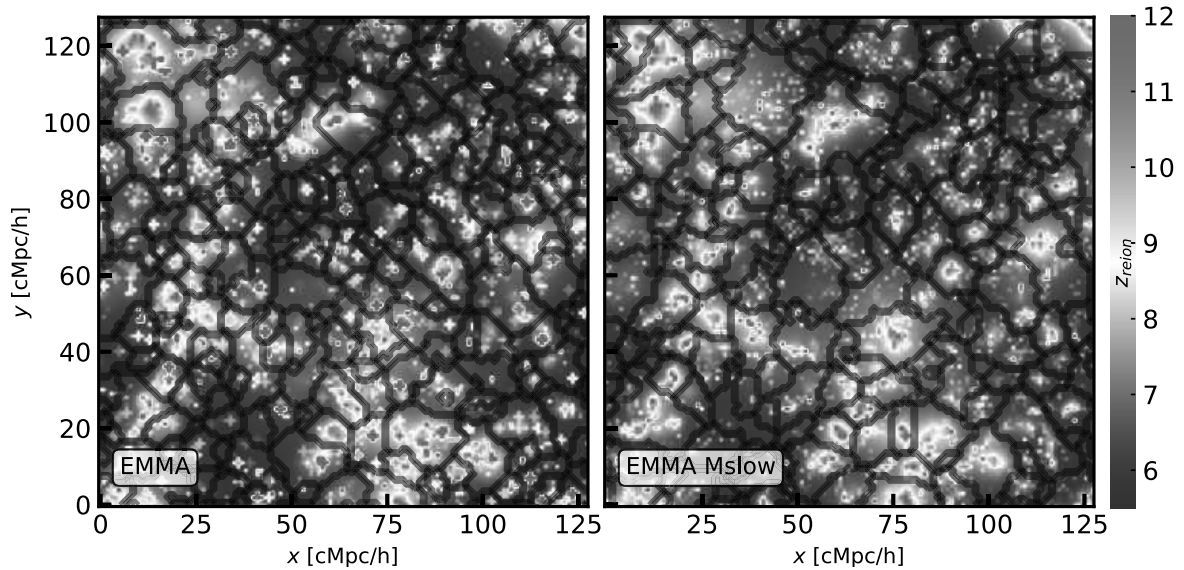
## 5.5 Comparaisons entre simulations cosmologiques et semi-analytiques

Dans cette section, je cherche à comparer les mesures de la topologie des simulations cosmologiques **EMMA** avec hydrodynamique radiative à celles des simulations semi-analytiques **21cmFAST**. Ce genre de comparaison est nécessaire pour savoir par exemple si l'on est capable de retrouver les mêmes comportements dans ces deux types de modélisations. Je propose ainsi ici de les distinguer par le biais des études sur les patchs de réionisation.

### 5.5.1 Simulations **EMMA**

On utilise deux simulations **EMMA** de taille  $512^3 \text{ cMpc}^3/h^3$  et résolution  $1 \text{ cMpc}^3/h^3$ . Comme elles ont des tailles de boîtes plus grandes que celles produites par **21cmFAST**, elles sont coupées en 64 sous-boîtes de  $128^3 \text{ cMpc}^3/h^3$  et seront ainsi comparées dans les études ci-dessous à 64 réalisations de simulations **21cmFAST**. Le seul paramètre qui différencie les deux simulations **EMMA** est la résolution de la masse des particules stellaires ( $10^7 M_\odot$  et  $10^8 M_\odot$ ). Gillet et al. (2021) indique que ce paramètre contrôle la manière dont le gaz est transformé en étoile dans une cellule de simulation : une faible valeur donnera une population d'étoiles diffuse, tandis qu'une valeur plus élevée donnera une population d'étoiles plus stochastique, focalisée sur les cellules les plus denses. La Figure 5.1 montre que les deux simulations **EMMA** (en rouge) ont été calibrées pour avoir une histoire de réionisation très similaire. La simulation simplement appelée « **EMMA** » a une résolution en masse des particules stellaires de  $10^8 M_\odot$ . Sa population d'étoiles est dite « biaisée », c'est-à-dire focalisée sur les cellules les plus denses de la simulation. La simulation appelée « **EMMA Mslow** », a une résolution en masse des particules stellaires de  $10^7 M_\odot$ , et donc une population d'étoiles plus diffuse.

Si l'on compare leur histoire de réionisation à celle des simulations **21cmFAST** du groupe SRH (Figure 5.1), on voit qu'elle commence après celle des simulations **21cmFAST**, tout en se terminant à peu près en même temps. L'histoire de réionisation des simulations **EMMA** est proche de celle des simulations SRH, même si elle est un peu plus abrupte.



**FIGURE 5.11** – Cartes de redshifts de réionisation provenant de la « simulation EMMA » à gauche (résolution de masse de particules stellaires de  $10^8 M_\odot$ ) et de la « simulation EMMA Mslow » à droite (résolution de masse de particules stellaires de  $10^7 M_\odot$ ). Les contours noirs délimitent les patches de réionisation extraites par DisPerSE.

## 5.5.2 Mesures topologiques des patches de réionisation

La Figure 5.11 montre une carte de  $z_{\text{reion}}(\vec{r})$  extraite de chaque simulation EMMA. On voit que dans le modèle avec la population d'étoiles biaisée (« simulation EMMA », à gauche), les sources de réionisation sont plus localisées (assez larges zones rouges), tandis que dans le modèle avec la population d'étoiles plus diffuse (« simulation EMMA Mslow », à droite), il y a plus de petites sources de réionisation (beaucoup de petits points rouges clairs et blancs). Même si ces petites sources sont en plus grand nombre, elles sont en fait moins significatives, apparaissent souvent au même moment (même redshift de réionisation sur la carte  $z_{\text{reion}}(\vec{r})$ ) et sont généralement associées par DisPerSE au patch de réionisation d'une source plus précoce et plus dominante. Dans la « simulation EMMA Mslow », les patches sont donc plus larges (et en moins grand nombre) que dans le cas de la « simulation EMMA ».

Le nombre de patches de réionisation détectés par DisPerSE dans les simulations EMMA est un peu plus grand que dans les simulations 21cmFAST, impactant directement la taille de ces régions qui va donc être légèrement plus faible que dans les simulations 21cmFAST. Par exemple, le rayon des patches de la simulation « EMMA » est d'environ 7 cMpc/h contre 10 cMpc/h pour les simulations 21cmFAST. Ensuite, les mesures de forme et d'orientation des patches de réionisation dans les simulations EMMA donnent globalement les mêmes résultats que pour 21cmFAST (voir les courbes rouges en haut des Figures 5.4 et 5.5). On remarque aussi en bas des Figures 5.4 et 5.5 que le modèle « Mslow » avec des particules stellaires moins massives ( $10^7 M_\odot$ ) et donc une distribution de sources plus diffuse a des caractéristiques topologiques assez proches de celles mesurées dans les modèles 21cmFAST. Cela paraît cohérent avec la manière dont 21cmFAST produit le rayonnement ionisant. En effet, dans ce modèle semi-analytique, chaque cellule des simulations peut produire du rayonnement ionisant et les halos

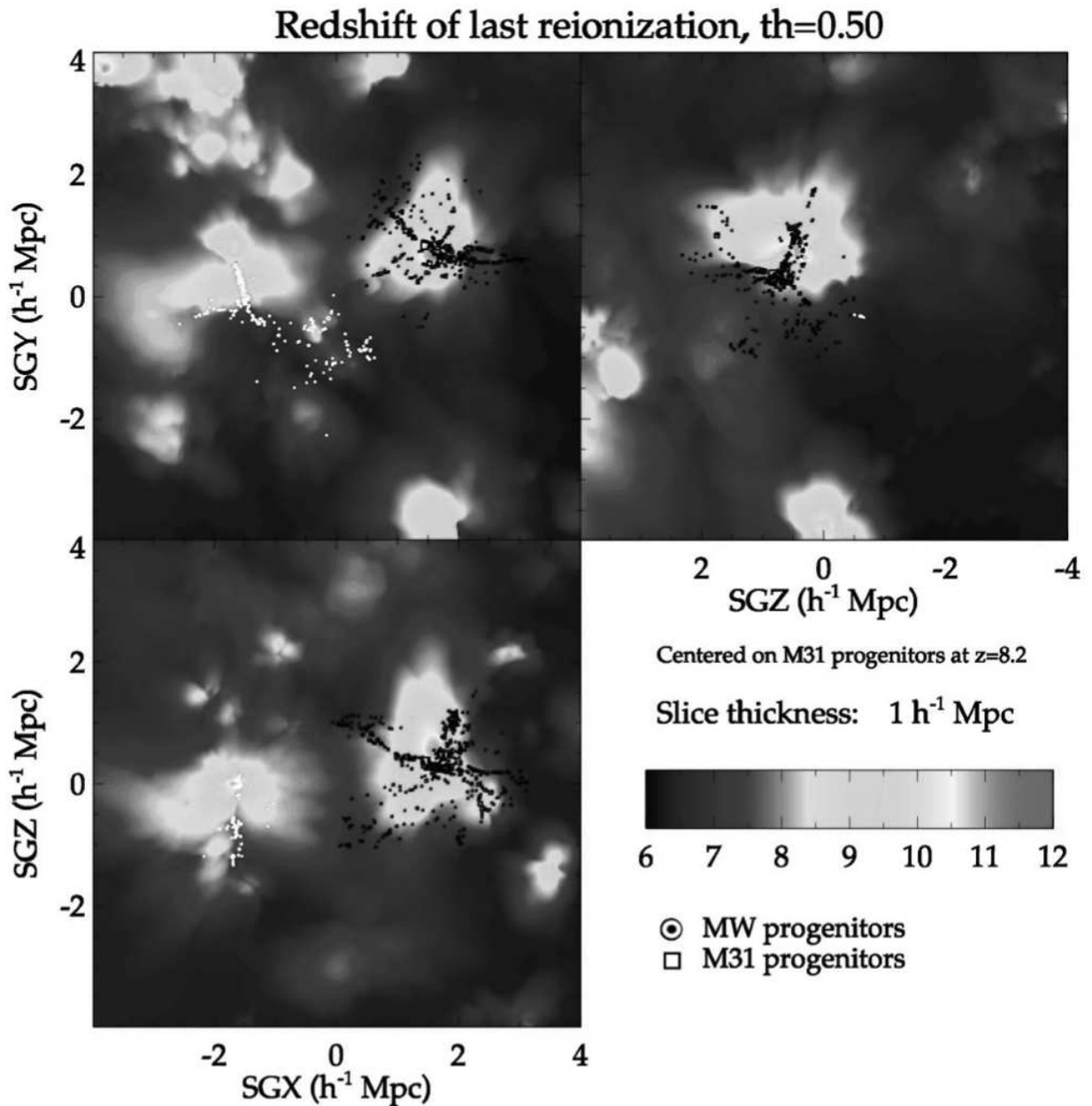
sources ne sont pas explicitement modélisés (Mesinger et al. 2011). Par ailleurs, la population biaisée d'étoiles de l'autre modèle **EMMA** implique un plus grand nombre de petits patches de réionisation (voir à gauche sur la Figure 5.11), et légèrement moins de patches alignés aux filaments de matière (voir en bas de la Figure 5.5).

Par conséquent, une façon de voir les choses est que l'on peut retrouver les mêmes comportements dans la topologie des patches de réionisation des simulations cosmologiques et semi-analytiques. Mais une autre façon de considérer ces résultats est de constater que des topologies différentes peuvent être tout de même simulés dans **EMMA** quand on biaise la formation d'étoile en la rendant plus stochastique et même si l'histoire de réionisation est similaire. Cela signifie que des modèles de simulations qui paraissent se ressembler de prime abord (de part leur histoire de réionisation par exemple) peuvent avoir des géométries différentes quand on les analyse avec ce type de méthodes.

## 5.6 Conclusions & perspectives

Dans cette étude, le champ de redshift de réionisation a été utilisé pour la première fois pour extraire des propriétés sur les patches de réionisation. On a montré que l'on avait accès au nombre de sources de réionisation, à la taille, la forme, et l'orientation des patches par rapport aux filaments de matière, ainsi qu'au processus de percolation. Dans des simulations semi-analytiques **21cmFAST**, la taille typique des patches de réionisation est d'environ  $10c\text{Mpc}/h$ . Généralement, ces patches sont prolates et alignés aux filaments de matière, ce qui signifie que les sources de réionisation sont perlées le long des filaments de matière et que le rayonnement UV s'en échappe de manière à former un patch allongé le long des filaments. Ces sources perlées ont a priori des propriétés similaires : elles apparaissent par exemple en même temps, émettent autant photons ou ont des fractions d'échappement proches. Le cas minoritaire dans ces simulations est celui des patches en forme de papillon avec un rayonnement se propageant perpendiculairement au filament dans un patch oblate, et dans laquelle la source domine totalement son environnement parce qu'elle est très brillante et/ou très massive. On a de plus montré que ces mesures sont dépendantes du modèle de réionisation en faisant varier l'efficacité des galaxies à produire des photons et la masse minimale des halos émetteurs de photons ionisants (reliée au paramètre  $T_{vir}$  de **21cmFAST**). En comparant les mêmes mesures topologiques faites dans des simulations semi-analytiques **21cmFAST** et cosmologiques **EMMA** avec des histoires de réionisation similaires, on montre que des comportements « topologiques » similaires peuvent être modélisés dans les deux types de simulations en appliquant un modèle de sources diffus dans **EMMA**. Cependant, on trouve aussi des comportements différents en appliquant cette fois-ci un modèle de sources biaisé. On dispose donc ici d'outils nous permettant de distinguer des modèles et différents types de simulations.

Nos conclusions de cette étude nous amènent à penser que la topologie des patches de réionisation dépend des sources modélisées et de leurs propriétés, ce qui signifie que la topologie peut aussi varier selon la résolution et/ou de la taille de la simulation. Il pourrait donc être intéressant de refaire ce genre d'analyses dans des simulations moins grandes mais plus résolues, comme la suite de simulations **CoDaII** (Ocvirk et al. 2020) (qui a une taille de  $64^3 c\text{Mpc}^3/h^3$  et contient  $4096^3$  cellules ; elle est donc 8 fois plus petite en volume que les simulations de cette étude, mais a une bien meilleure résolution). La Figure 5.12 montre le



**FIGURE 5.12** – Cartes de redshifts de réionisation vues dans plusieurs directions supergalactiques représentant le progéniteur du Groupe Local. Ces cartes proviennent de la simulation CoDaII (Ocvirk et al. 2020), dont la taille est de  $64^3 \text{ cMpc}^3/\text{h}^3$  comprenant  $4096^3$  cellules. Cette Figure est extraite de Sorce et al. (2022). Les zones les plus rouges correspondent à des zones qui réionisent le plus tôt, et les régions les plus foncées réionisent en dernier. Les cercles blancs sont les progéniteurs de la Voie Lactée (MW) et les carrés noirs sont les progéniteurs d'Andromède (M31). Les temps de réionisation autour des progéniteurs de MW ou de M31 prennent effectivement une forme de papillon à cette résolution.

champ de redshift de réionisation de cette simulation vu dans plusieurs directions. On y voit aussi les progéniteurs de la Voie Lactée (cercles blancs) et d'Andromède (carrés noirs). Dans ce genre de simulation, on s'attendrait à obtenir plus de régions en papillon parce qu'avec une meilleure résolution, le gaz est également plus finement décrit et donc les variations locales d'absorption et de création des sources sont aussi mieux décrites. C'est en effet ce que l'on voit dans la Figure 5.12 avec les fronts d'ionisation en forme des papillons autour des galaxies.

Par ailleurs, nous nous sommes focalisés dans cette étude sur les sources topologiques de réionisation (que sont les maxima du champ de redshift de réionisation). En extrayant les halos de matière noire des simulations, on pourrait également en déduire des propriétés de ces sources telles que leur masse, et étudier les mesures topologiques évoquées ici en fonction des propriétés des halos. Cela nous indiquerait entre autres quels types de sources sont impliqués dans la réionisation et comment elles sont distribuées dans l'espace. Par exemple, on pourrait étudier l'influence de la masse du « halo source » sur la taille, la forme ou l'orientation des patches de réionisation. On pourrait aussi étudier les patches de réionisation des progéniteurs de la Voie Lactée ou d'Andromède au travers des cartes de redshifts de réionisation, comme l'ont fait *Sorce et al. (2022)*. D'autre part, des simulations à plus petite échelle et plus résolues (de l'ordre du rayon de Viriel) pourraient nous permettre d'étudier la manière dont le rayonnement s'échappe des galaxies, sur quelle distance il affecte leur environnement ou comment l'environnement radiatif proche des galaxies évolue dans le temps. Ces analyses pourraient être par exemple menées dans des simulations de type *CoDa* (*Ocvirk et al. 2020*) ou de futurs modèles faits par exemple avec *Dyab1o* (*Aubert & Durocher 2021*).

## 5.7 « First look at the topology of reionisation redshifts in models of the epoch of reionisation »

Cette section présente l'article « First look at the topology of reionisation redshifts in models of the epoch of reionisation » (ou en français « Premier regard sur la topologie des redshifts de réionisation dans des modèles de l'époque de réionisation »). Cet article a été soumis le 8 juillet 2021 dans la revue à comité de lecture *Astronomy & Astrophysics*, et accepté le 22 novembre 2021. Il a été publié en février 2022 dans le volume 658.



# First look at the topology of reionisation redshifts in models of the epoch of reionisation

Emilie Th  lie<sup> </sup>, Dominique Aubert<sup> </sup>, Nicolas Gillet<sup> </sup>, and Pierre Ocvirk

Universit   de Strasbourg, CNRS UMR 7550, Observatoire Astronomique de Strasbourg, Strasbourg, France  
e-mail: emilie.thelie@astro.unistra.fr

Received 8 July 2021 / Accepted 22 November 2021

## ABSTRACT

*Context.* During the epoch of reionisation (EoR), the first stars and galaxies appeared while creating ionised bubbles that eventually percolated near  $z \sim 6$ . These ionised bubbles and percolation process are closely scrutinised today because observations of neutral hydrogen will be carried on in the next decade with the Square Kilometre Array radio telescope, for instance. In the meantime, EoR studies are performed in semi-analytical and fully numerical cosmological simulations to investigate the topology of the process, for instance.

*Aims.* We analyse the topology of EoR models by studying regions that are under the radiative influence of ionisation sources. These regions are associated with peak patches of the reionisation redshift field, for which we measure the general properties such as their number, size, shape, and orientation. We aim to gain insights into the geometry of the reionisation process and its relation to the matter distribution, for example. We also assess how these measurements can be used to quantify the effect of physical parameters on the EoR models or the differences between fully numerical simulations and semi-analytical models.

*Methods.* We used the framework of Morse theory and persistent homology in the context of the EoR, which was investigated via the DisPerSE algorithm on gas density and redshift of reionisation maps. We analysed different EoR scenarios with semi-analytical 21cmFAST and fully numerical EMMA simulations.

*Results.* We can distinguish between EoR models with different sources using simple analyses of the number, shape, and size distributions of the reionisation redshift patches. For every model (of the semi-analytical and fully numerical simulations), we statistically show that these bubbles are rather prolate and aligned with the underlying gas filaments. Moreover, we briefly highlight that the percolation process of HII bubbles during the EoR can be followed by studying the reionisation redshift fields with different persistence thresholds. Finally, we show that fully numerical EMMA simulations can be made consistent with 21cmFAST models in this topological framework as long as the source distribution is diffuse enough.

**Key words.** large-scale structure of Universe – dark ages, reionization, first stars – methods: numerical – galaxies: formation – galaxies: high-redshift

## 1. Introduction

The epoch of reionisation (EoR) marks the transition of a totally neutral to a fully ionised Universe, whose large-scale structures are composed of filaments, voids, and very dense regions. During the EoR, the first sources of radiation appear and release the photons that are required to ionise the cosmic gas. Light escaping from these first stars and galaxies created HII regions (also known as HII bubbles) that eventually percolated at the end of EoR between  $z = 5.3\text{--}6$  (Kulkarni et al. 2019).

This epoch is closely studied today to try to understand, for instance, which types of sources drove the EoR, or how and when the HII bubbles overlapped. Analysing the large-scale topology of the EoR is useful to discover the properties of the objects that drove it, such as the nature of the driving sources of the EoR, their spatial distribution, the timing of their appearance, or the spectrum of their radiation (see e.g. Chardin et al. 2017).

Complementary to studies using the power spectrum, for instance (e.g. Mellema et al. 2006; Dixon et al. 2016; Shaw et al. 2020), many works tried to answer these questions with a topological approach of the EoR. For example, Minkowski functionals (and the derived genus or Euler characteristic) are used to analyse the volume, surface area, and curvature of ionised

and neutral regions, giving information about the shape of these regions and the percolation process (as in e.g. Gleser et al. 2006; Lee et al. 2008; Friedrich et al. 2011; Hong et al. 2014; Yoshiura et al. 2017; Chen et al. 2019). Other studies focused on the size distributions of neutral or ionised regions (also called island size distribution, ISD; and bubble size distribution; BSD, respectively), which characterise the size of neutral islands or HII bubbles and their percolation during the EoR. Examples of methods used in these studies are the friends-of-friends algorithm (Iliev et al. 2006; Friedrich et al. 2011; Lin et al. 2016; Giri et al. 2018, 2019), the spherical average method (Zahn et al. 2007; Friedrich et al. 2011; Lin et al. 2016; Giri et al. 2018), the mean free path method (Mesinger & Furlanetto 2007; Lin et al. 2016; Giri et al. 2018, 2019), and the granulometry method (Kakiichi et al. 2017). The Betti numbers, related to 3D structures of a field (isolated objects, tunnels, and cavities), are also used to probe the topology of ionised bubbles and neutral islands during the EoR (Giri & Mellema 2021). The triangle correlation function (TCF) is another tool for extracting topological information, such as the radius and shape of ionised and neutral bubbles (Gorce & Pritchard 2019). It relies on a three-point correlation function based on the inverse Fourier transform of the bispectrum that describes the phase of the signal of interest.

Upcoming observations from radio interferometers such as the New Extension in Nançay Upgrading loFAR<sup>1</sup> (NenuFAR; Zarka et al. 2012) and the low-frequency component of the Square Kilometre Array<sup>2</sup> (SKA-Low; see e.g. Mellema et al. 2013) will allow us to map the sky with HI regions through the redshifted 21 cm signal that is emitted during the EoR (Mellema et al. 2015). The SKA will produce 2D tomographic images of the 21 cm emission of the EoR at many redshifts. Geometrical studies can therefore focus directly on the 21 cm fields, for example, using the 21 cm power spectrum (see Kim et al. 2013; Choudhury & Paranjape 2018; Seiler et al. 2019; Pagano & Liu 2020) or the 21 cm bispectrum (see Hutter et al. 2020) to extract topological information, such as the size distribution of ionised bubbles.

In this study, we also propose to explore the EoR through its topology, but using the discrete persistent structure extractor, also called DisPerSE<sup>3</sup> (Sousbie 2011). This code relies on discrete Morse theory and persistent homology. By computing gradients and critical points of fields, it extracts geometrical properties, including peaks, walls, filaments, and voids for a 3D field. The persistent homology is used in this algorithm only to focus on significant features, and it takes the noise in data sets into account, for instance, by providing a way to suppress the noise by tuning a persistence level. DisPerSE has been written in order to analyse astrophysical structures, in particular, the large-scale density structures of the Universe. Many studies have extracted and analysed the matter filaments with DisPerSE from cosmological simulations (e.g. Singh et al. 2020; Galárraga-Espinosa et al. 2020; Katz et al. 2020; Song et al. 2021), observations (e.g. Malavasi et al. 2020; Tanimura et al. 2020) or both (e.g. Sousbie et al. 2011). Persistent homology has also been used in phenomenological models of the EoR to study the topology of HII regions (Elbers & van de Weygaert 2019).

In this work, we extract the filaments of the matter density field, but also the peak patches of the redshift of reionisation field or  $z_{\text{reion}}$  (Battaglia et al. 2013; Deparis et al. 2019) from EoR simulations produced by the semi-analytical code 21cmFAST<sup>4</sup> (Murray et al. 2020; Mesinger et al. 2011) and the simulation code électromagnétisme et mécanique sur maille adaptative (EMMA; Aubert et al. 2015). The  $z_{\text{reion}}$  field is a measure of the redshift at which each point in space has been reionised. It therefore contains spatial and temporal information about the propagation of radiation from sources of reionisation in a single field per simulation. Using DisPerSE, we apply the Morse theory framework to the 3D field of our different models. As detailed below, maxima (peaks) in  $z_{\text{reion}}$  correspond to the seeds, the first sources, of local reionisations. Peak patches around these seeds represent the extent of their time-integrated radiative influence. Even though we did not investigate them in the current work, walls or filaments in  $z_{\text{reion}}$  define the regions in which percolation occurred. This approach can complement the approaches that are based on the study of HII regions at a single redshift or a sequence of redshifts. The shape of these  $z_{\text{reion}}$  patches might give us an indication about the way radiation escapes the galaxies. Their size could inform us about the extent of the radiative influence of the different types of sources, for example, in relation to the large absorption trough in the quasar spec-

tra (Becker et al. 2015) or to understanding the observed strong fluctuations of UV radiation (Chardin et al. 2017). In general, a different long-range influence might be expected from faint or locally sub-dominant sources compared to more dominant emitters. Overall, the Morse theory framework as implemented by DisPerSE provides a solid and reproducible mathematical description of the  $z_{\text{reion}}$  geometry that could be used to characterise and compare models of the EoR. In this initial study, we gain first insights into the extent of the  $z_{\text{reion}}$  patches, their shapes, and their relative orientation to the underlying gas distribution. We compare different EoR models from 21cmFAST and EMMA.

We first describe in Sect. 2 the method we employed by presenting our simulations, the reionisation redshift field, the DisPerSE algorithm, and the method we used to extract the orientation and shape of the patches. Afterwards, Sect. 3 shows our first topological analyses: the number of reionisation seeds, and the shape and size of the reionisation patches. Sections 4 and 5 highlight the differences between simulations with different parameters, as well as studies of the orientation of radiation with respect to the underlying density. Section 6 presents a comparison of the topology of the EoR in semi-analytical and cosmological simulations. We finally conclude in Sect. 7. The cosmology parameters we used throughout are  $(\Omega_m, \Omega_b, \Omega_\Lambda, h, \sigma_8, n_s) = (0.31, 0.05, 0.69, 0.68, 0.81, 0.97)$ , as given by Planck Collaboration VI (2020).

## 2. Method

### 2.1. Simulated data

First, we obtained 3D cosmological simulations from the semi-analytical code 21cmFAST (version 3.0.3; Murray et al. 2020; Mesinger et al. 2011). 21cmFAST evolves initial density and velocity fields based on the first- or second-order perturbation theory of Zel'Dovich (1970), and provides associated predictions such as temperature, ionisation, 21 cm signal, and radiation fields.

In our case, the ionised gas evolved in a  $189^3 \text{ cMpc}^3$  box at a resolution of  $1.48^3 \text{ cMpc}^3$  (or boxes of  $128^3 \text{ cMpc}^3 \text{ h}^{-3}$  with  $128^3$  cells). Different physical models were studied using two parameters of 21cmFAST: the virial temperature  $T_{\text{vir}}$  and the ionising efficiency  $\zeta$ , as defined by Greig & Mesinger (2015).  $T_{\text{vir}}$  is the minimum virial temperature for a halo to start the star formation process. It allows controlling the gas accretion, cooling, and retainment of supernova outflows, and it is related to the halo mass through  $M_{\text{min}} \propto T_{\text{vir}}^{3/2}$ . The higher the virial temperature, the higher the halo mass.  $\zeta$  is the ionising efficiency of high- $z$  galaxies. It defines the number of photons that escape from galaxies, and higher values tend to accelerate the reionisation.

The parameters of our different models are shown in Table 1. We separated the models into two sets of simulations: a set in which  $\zeta$  and  $T_{\text{vir}}$  were varied, which we called the same reionisation history (SRH) set, and a set in which  $\zeta$  alone was varied, which we called the different reionization history (DRH) set. For the SRH simulations set, virial temperatures were set first before we tuned the ionising efficiency in order to obtain the same reionisation history. Then, the higher  $T_{\text{vir}}$ , the larger  $\zeta$  is in order to maintain the same reionisation history because only haloes with the highest mass radiate, meaning that there are also fewer haloes that release photons. The DRH simulation sets were made in order to analyse the effect of different emissivities and thus reionisation histories. We ran each model 101 times using different seeds in 21cmFAST. The ionisation histories of Fig. 1

<sup>1</sup> <https://nenufar.obs-nancay.fr>

<sup>2</sup> <https://www.skatelescope.org>

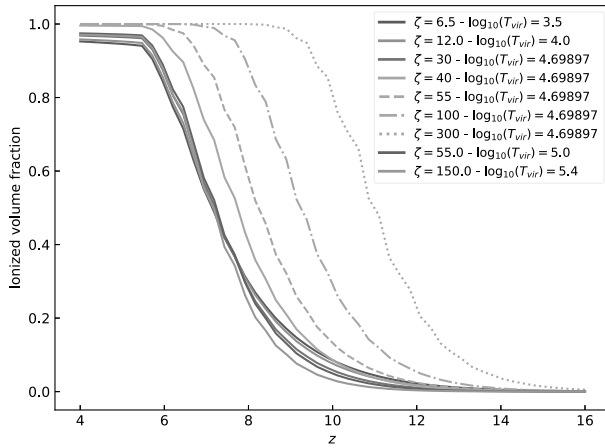
<sup>3</sup> <http://www2.iap.fr/users/sousbie/web/html/indexd41d.html>

<sup>4</sup> <https://github.com/andreimesinger/21cmFAST>

**Table 1.** Parameters of the 21cmFAST simulated models.

Model	$\zeta$	$T_{\text{vir}}$ [K]	$(\log_{10}(T_{\text{vir}}))$
1	6.5	$3.2 \times 10^3$	(3.5)
2	12	$1 \times 10^4$	(4)
3	30	$5 \times 10^4$	(4.69897)
4	55	$1 \times 10^5$	(5)
5	150	$2.5 \times 10^5$	(5.4)
3	30	$5 \times 10^4$	(4.69897)
3-1	40	$5 \times 10^4$	(4.69897)
3-2	55	$5 \times 10^4$	(4.69897)
3-3	100	$5 \times 10^4$	(4.69897)
3-4	300	$5 \times 10^4$	(4.69897)

**Notes.** Five models change  $\zeta$  and  $T_{\text{vir}}$ . They belong to the SRH set. Four other models modify the  $\zeta$  parameter from the fiducial model 3. They form the DRH set. The values of  $T_{\text{vir}}$  are also given in logarithm.


**Fig. 1.** Ionized volume fraction of sets of 21cmFAST simulations with different models, each averaged on their 101 runs.

are averaged over all the runs of each model. Figure 1 shows the ionised volume fraction of all models.

## 2.2. Reionisation redshift maps

As mentioned in the Introduction, the geometry of ionisation is usually investigated via the ionised fraction distribution  $x_{\text{HII}}(\mathbf{r}, z)$  at a given redshift  $z$  or with a sequence of redshifts, where  $\mathbf{r}$  stands for the 3D position. Here, we rather focus on the reionisation redshift field  $z_{\text{reion}}$ . It is obtained by saving the redshift at which the gas is considered to be ionised for each position, that is, when the ionisation fraction exceeds a given threshold, which was set at 50%<sup>5</sup>, as follows:

$$z_{\text{reion}}(\mathbf{r}) = z(\mathbf{r}, x_{\text{HII}} = 0.5). \quad (1)$$

It incorporates spatial and temporal information about the propagation of radiation during the EoR in a single field (Battaglia et al. 2013; Deparis et al. 2019; Trac et al. 2021) and can be seen as a reciprocal of or dual to the  $x_{\text{HII}}$  field.

<sup>5</sup> A few cells are not reionised at the end of the simulations ( $z = 5.5$ ). The reionisation redshift of these cells was manually set to 5.5 in the  $z_{\text{reion}}$  map.

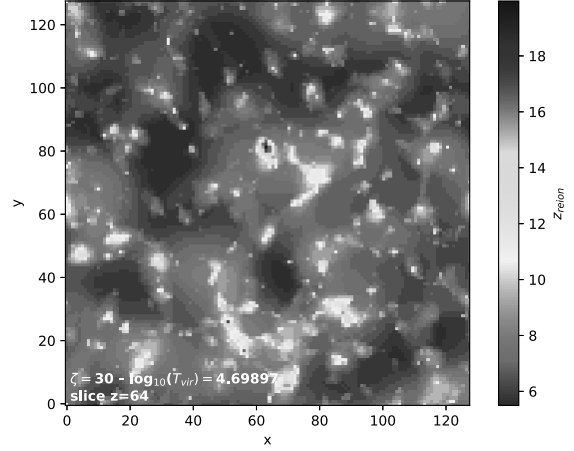
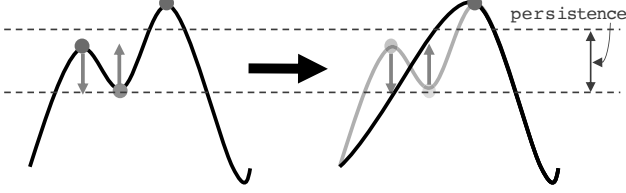

**Fig. 2.** 2D slice at depth  $z = 64$  of the reionisation redshift field extracted from the 21cmFAST simulation (model  $\zeta = 30$  and  $T_{\text{vir}} = 5 \times 10^4$  K).

Figure 2 shows an example of a reionisation redshift map extracted from a 21cmFAST simulation. The bluest regions correspond to the places at which the gas was first reionised and can be interpreted as the local seeds of the local reionisation process. Ionization fronts propagated around these  $z_{\text{reion}}$  peaks towards the reddest regions that reionised later on, forming connected regions of varying shapes and extent: these regions represent the local radiative influence of these local seeds, integrated over time.

## 2.3. DisPerSE

DisPerSE relies on Morse theory in order to analyse the topology of manifolds through the study of differential functions. This code searches for critical points in a field, such as maxima, minima, and saddle points. DisPerSE also looks for the integral lines (or field lines) of the field: they are the tangent curves to the gradient in every point, and always have critical points as origin and destination. As these lines cover the entire space, we can produce a tessellation of the space, creating regions called ascending or descending manifolds. Finally, the set of all ascending or descending manifolds (also known as peak or void patches) is called the Morse complex of a field.

Within DisPerSE, we can tune the so-called ‘‘persistence’’ to control the significance of the topological features found in fields (see Sousbie 2011). This can be considered a significance threshold that separates two critical points, and the local critical points of the field that are not significant enough are ignored. This allows controlling the smoothness of the resulting topological features and can be used to remove the noise in the input data. Figure 3 illustrates the impact of applying a persistence threshold on a 1D field (left column) on which a persistence threshold is applied (right column). We consider a pair of critical points: a minimum (in blue) and a maximum (in red). The distance between the value of the field at these points (green arrows) is compared to the value of the chosen persistence threshold (distance between the two purple dashed lines): if it is higher than the persistence, these extrema are considered when points are detected and assigned to patches. When it is lower, however, they are discarded, as if they were filtered out from the field. In this case, the field becomes topologically smoother.



**Fig. 3.** Illustration of the persistence parameter of the DisPerSE algorithm. It shows a 1D function, with red dots as maxima and blue dots as minima. The left column is the actual function to which the persistence threshold is applied, and the right column shows the result after this threshold is applied. The length between the dashed purple lines represents the persistence threshold in that case. The green arrows represent the smoothing of the curve when the corresponding critical points are removed. This illustration is inspired by Sousbie (2011).

Before we used DisPerSE, the gas density  $\delta(\mathbf{r})$  and the reionisation redshift  $z_{\text{reion}}(\mathbf{r})$  were first converted into dimensionless 3D fields by applying the following transformation:

$$\delta_a = \frac{\delta - \bar{\delta}}{\sigma_\delta}, \quad (2)$$

$$z_{\text{reion},a} = \frac{z_{\text{reion}} - \overline{z_{\text{reion}}}}{\sigma_{z_{\text{reion}}}}, \quad (3)$$

where  $\bar{\delta}$  and  $\overline{z_{\text{reion}}}$  are the mean of each field, and  $\sigma_\delta$  and  $\sigma_{z_{\text{reion}}}$  are their standard deviation. These transformations allowed us to express the persistence parameter in terms of the deviation of a field to its mean value. They show at how many  $\sigma$  the fields are smoothed from a topological point of view. Moreover, the fields were also filtered with a Gaussian kernel (with a standard deviation of one cell for each field) in order to avoid large patches of constant field values that prevent DisPerSE from running properly.

We first applied the DisPerSE algorithm with a persistence level of  $0.5\sigma$  (Sousbie et al. 2011; Codis et al. 2018; Galárraga-Espinosa et al. 2020; Cohn 2021) on the 3D reionisation redshift field of our simulations in order to determine the  $z_{\text{reion}}$  patches associated with the  $z_{\text{reion}}$  maxima<sup>6</sup>. They correspond to the regions that are topologically associated with a given  $z_{\text{reion}}$  maximum. These regions gather all the cells of the simulations that have a positive gradient towards their maximum. These maxima are the first places to reionise: they are the seeds of local reionisations. Again, these patches are different from the ionised bubbles that are usually mentioned in the literature (see e.g. Chen et al. 2019; Gorce & Pritchard 2019; Giri et al. 2019; Giri & Mellema 2021), which are regions in which hydrogen gas is ionised at a given time. The cells within the patches discussed here are connected through time via the topology of  $z_{\text{reion}}$ , providing another insight into the propagation of ionisation fronts and into the extent of the influence of the reionisation seeds. Figure 4 shows reionisation redshift maps superimposed on the contours of their segmentation obtained with DisPerSE. The patches are naturally centred around the maxima, are not necessarily spherical, and have different sizes and shapes (we recall that 3D effects can be hidden in these 2D maps) depending on the local specificities of the production and absorption of radiation.

<sup>6</sup> In practice, DisPerSE computed the ascending 3-manifolds of the field using the integral lines that are related to the maxima of the 3D field (see Sousbie 2011 for more details).

We are also interested in the filaments of the gas density structures in our models. With DisPerSE, they are detected as the set of points or cells in the density field belonging to arcs that originate in saddle points and end in extrema. We applied this selection to the gas density fields in our simulations using a  $0.5\sigma$  persistence level. Figure 5 is an example of a density map from a 21cmFAST simulation superimposed on the filaments detected with DisPerSE. The map is obtained from an average of several slices for better visualisation. The filaments clearly follow the distribution of the gas in the box.

#### 2.4. Inertial tensor: A tool for determining the orientation of an object

Our study assesses the shape and size of the reionisation patches, as well as their orientation with respect to the gas filaments. We used inertial tensors, as defined by Tormen (1997),

$$I_{i,j} = \sum_k \omega_k \cdot (x_{k,i} - x_{\text{ref},i}) \cdot (x_{k,j} - x_{\text{ref},j}). \quad (4)$$

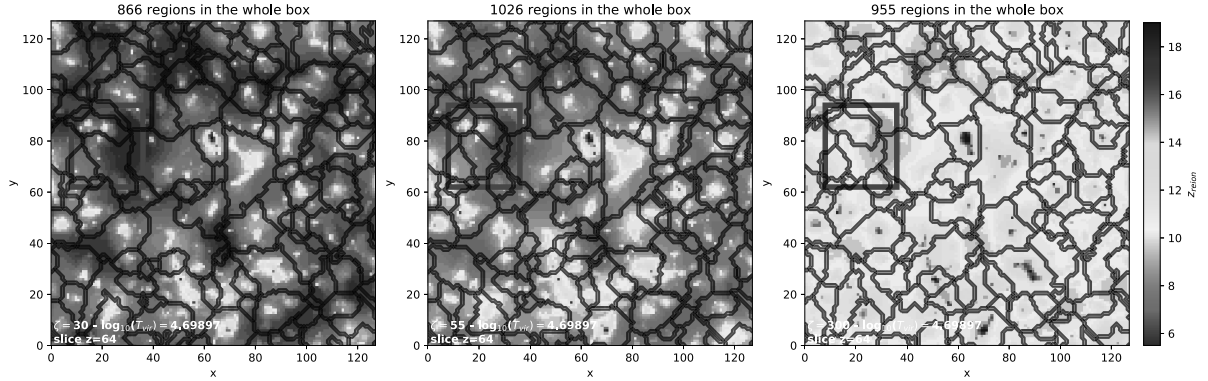
Here,  $k$  scans all of the cells within the patches,  $i, j \in \{1, 2, 3\}$  refer to the Cartesian components of the cell position  $x$ , and  $\omega$  is a weight.  $x_{\text{ref}}$  is a point of reference: in our case, it is the geometrical centre of each patch. The eigenvectors of the tensor provide the 3D directions of the extension of the patches, and the eigenvalues provide the typical extent of the cells along these directions. The eigenvector with the highest eigenvalue therefore represents the main axis of a given patch. The same procedure can be applied to the cells belonging to a given gas filament, in order to extract its main orientation.

For  $z_{\text{reion}}$  patches, we used  $\omega_k = 1$  because we are only interested in the geometrical shape. The main axis of these patches represents the direction favoured by radiation after it is emitted by the local seed. Presumably, this axis corresponds to the path of least resistance for radiation or is indicative of the geometry of light production as ionisation fronts propagated. Because we are interested in their orientation relative to the reionisation patch they are located in, we set  $\omega_k$  to the cell density for gas filaments and restricted ourselves to cells that are within this reionisation patch.

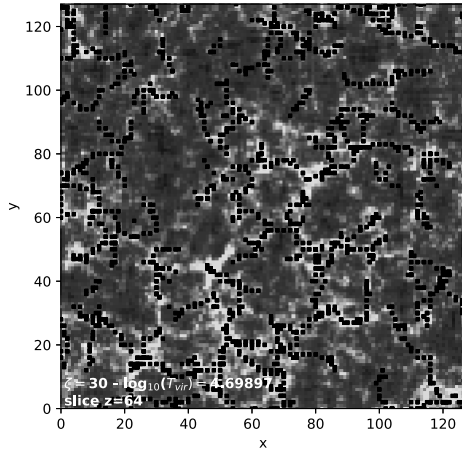
### 3. First topological insights

#### 3.1. Number of reionisation patches

Figure 6 shows the number of reionisation patches in every model (accumulating all runs). The parameters that evolve from one simulation to the next,  $T_{\text{vir}}$  and  $\zeta$ , only affect the radiation processes, whereas the density maps are not modified between the different  $(T_{\text{vir}}, \zeta)$  models, and the filaments are necessarily the same, for instance. The number of reionisation patches (or reionisation seeds) decreases with increasing  $T_{\text{vir}}$  (and then  $\zeta$ ; see the top panel of Fig. 6, which shows the SRH set): with higher  $T_{\text{vir}}$ , there are fewer emitting haloes because this temperature favours massive haloes (which also produce more photons to maintain a reionisation with the same history). For the DRH simulations set (i.e. with the same emitting haloes, but different ionising efficiencies), the number of reionisation patches as a function of the ionising efficiency parameter presents two regimes. We would have expected the number of reionisation patches to only decrease with increasing  $\zeta$ : when the sources emit more radiation, the first regions expand rapidly, preventing the smaller and later-formed haloes from creating their own

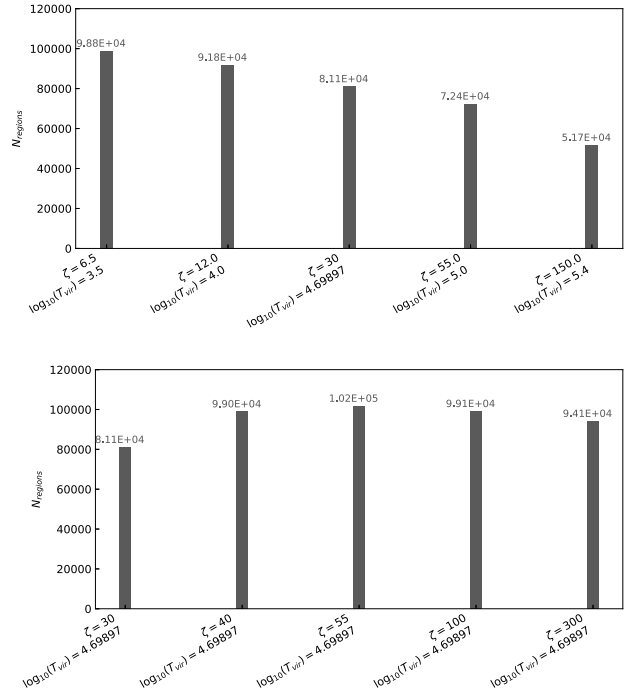


**Fig. 4.** 2D slices of  $z_{\text{reion}}$  fields extracted from a 21cmFAST simulation. The reionisation redshifts are superimposed on the contours (in black) of the patches from the segmentation obtained with DisPerSE (with a persistence level of  $0.5\text{-}\sigma$ ). The three panels represent three models (of the DRH simulations set) with different  $\zeta$  values: models 3 ( $\zeta = 30$ ), 3–2 ( $\zeta = 55$ ), and 3–4 ( $\zeta = 300$ ), from left to right.



**Fig. 5.** 2D slice of the density field extracted from a 21cmFAST simulation (model  $\zeta = 30$  and  $T_{\text{vir}} = 5 \times 10^4$  K). The density is superimposed on the filaments (in black) obtained with DisPerSE (with a persistence level of  $0.5\text{-}\sigma$ ). This map is made from an average of many slices to better visualise the filaments.

patches. The latter sources are then ionised by external emitters and belong to the patches of these earlier external reionisers. For instance, this behaviour can be seen in the purple boxed patch in the middle and right panels of Fig. 4. It shows reionisation redshift fields for  $\zeta$  values of 55 and 300 superimposed on their DisPerSE segmentations. As  $\zeta$  increases, four small patches merge as bright early emitters expand their influence more efficiently and late emitters are externally reionised. However, the number of patches increases as  $\zeta$  increases from 30 to 55. We verified that this behaviour is an artefact of the way the DisPerSE segmentations are produced in this study. The  $z_{\text{reion}}$  field was smoothed before being processed by DisPerSE, filtering out small patches of low emitters with weak gradients. As  $\zeta$  increases, the extent and the gradients are enhanced within these patches, and the gradients are not suppressed by the Gaussian smoothing and the persistence filter. This effect is illustrated in the left and middle panels of Fig. 4. Only two patches are detected within the purple box at low  $\zeta$ , while four patches can be seen as this parameter is increased to moderately higher values.



**Fig. 6.** Number of patches detected with DisPerSE (with a persistence level of  $0.5\text{-}\sigma$ ) for every model listed in Table 1. *Top panel:* models of the SRH set, and *bottom panel:* models of the DRH set. For each model, the patch numbers of every run are added.

The number of patches is related to the first Betti number  $\beta_0$  of an ionisation field as it represents the number of sources of ionisation (Giri & Mellema 2021). Giri & Mellema (2021) computed the  $\beta_0$  number of their  $x_{\text{HII}}$  ionisation field throughout the reionisation process for models that varied only the ionising efficiency (see their FN1 and FN2 models and the top left panel of their Fig. 6, where they accumulated the number of sources as  $x_{\text{HII}}$  grew). These models can be compared to our DRH set models for which we have histories of reionisation (for the lowest  $\zeta$  models) similar to theirs. They concluded that the  $\beta_0$  numbers of these two models are approximately the same, but with slightly fewer sources, for the model with the highest  $\zeta$ . We conclude

that reionisation patches appear sooner (see Fig. 1), and their total number throughout the reionisation history is lower for the highest  $\zeta$  models (see Fig. 6), which is consistent with the results of Giri & Mellema (2021).

### 3.2. Shapes of the reionisation patches

With the eigenvalues given by the inertial tensors, we can compute a number that quantifies the geometrical shape of the patches. Tormen (1997) called it the triaxiality parameter and defined it as follows:

$$T = \frac{\lambda_3^2 - \lambda_2^2}{\lambda_3^2 - \lambda_1^2}. \quad (5)$$

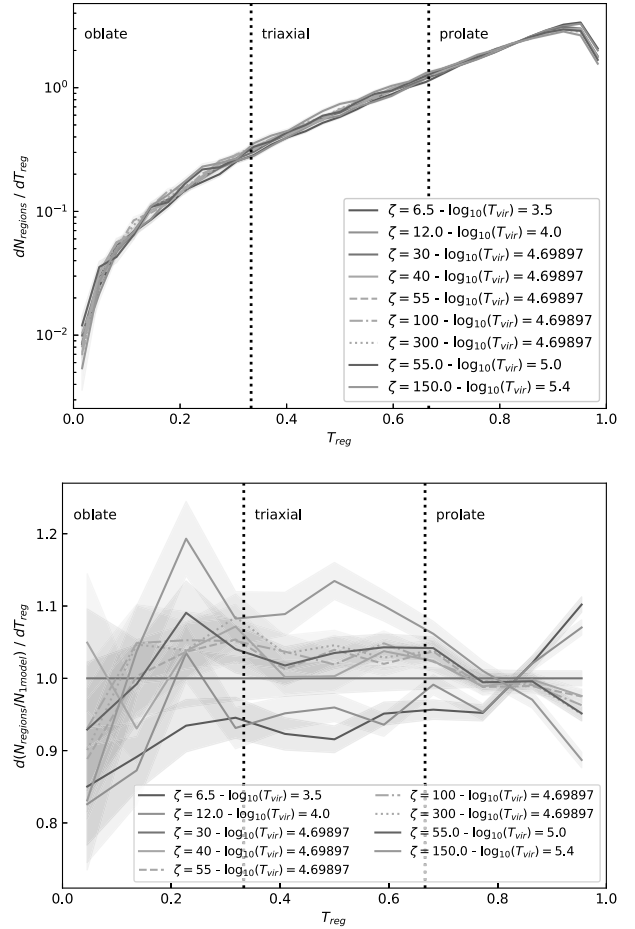
The  $\lambda_i$  (with  $i \in \{1, 2, 3\}$ ) are the eigenvalues of the inertial tensors with  $\lambda_1 \leq \lambda_2 \leq \lambda_3$ . When  $0 \leq T \leq \frac{1}{3}$ , the object is oblate: it has two large dimensions and one small dimension, like a flattened sphere. When  $\frac{1}{3} \leq T \leq \frac{2}{3}$ , the object is triaxial. When  $\frac{2}{3} \leq T \leq 1$ , the object is prolate: it has one large dimension and two small dimensions, like a rugby ball. The top panel of Fig. 7 shows the probability distribution functions of the triaxiality parameters computed for the reionisation patches of each run of every model. They all return a similar distribution, with a majority of prolate and a minority of oblate patches. The normalisation of these distributions with the fiducial model (bottom panel) reveals small differences between the SRH models, however: with increasing  $T_{\text{vir}}$  and  $\zeta$ , the prolate patches dwindle.

These prolate or cigar shapes of the reionisation patches are reminiscent of the ‘‘ionised fibre’’ stage, as it was called in Chen et al. (2019). Using Minkowski functionals, these authors were able to separate the process of reionisation into five stages: an ‘‘ionised bubble’’ stage during which the first isolated ionised bubbles appear, an ‘‘ionised fibre’’ stage during which bubbles start to connect to each other and form a large fibre structure, a ‘‘sponge’’ stage with intertwined ionised and neutral regions, and ‘‘neutral fibre’’ and ‘‘neutral island’’ stages, which are the neutral counterparts of the two first stages. In our segmentations, patches seem to have kept a memory of the fibre shape of the second stage of Chen et al. (2019).

### 3.3. Size distribution of the reionisation patches

We computed the probability distribution function of the volume of the reionisation patches. Figure 8 shows the PDFs for the two simulation sets. Patches with intermediate volumes dominate the simulation box. The largest volume is limited by the finite size of the box, and the smallest volumes are limited by the resolution. When we compare the different models of the SRH set, the highest virial temperatures favour the ionisation by the more massive haloes that are stronger emitters: it leads to larger but fewer reionisation patches. A similar trend is measured for the DRH set, when  $\zeta$  becomes large, with fewer and more prominent patches. From our size distributions, a typical radius of patches of  $\sim 10 \text{ cMpc } h^{-1}$  can be extracted, which is consistent for instance with the radii found by Gorce & Pritchard (2019) with their triangle correlation function.

The size distribution of the patches returns general information about the maximum radius that can be reached by HII regions before percolation during the whole reionisation process. It thus provides insights into the required size for cosmological simulations of the EoR to use the framework described in this work. For example, our PDFs of the patch volumes indicate that our boxes ( $128 \text{ cMpc } h^{-1}$ ) do not limit typical-sized patches



**Fig. 7.** Probability distribution functions of the triaxiality parameter of the reionisation patches of every model (top panel). Bottom panel: same PDFs, but each one of them is divided by one particular model (with  $\zeta = 30$  and  $\log_{10}(T_{\text{vir}}) = 4.69897$ ). All the patches of each run for one specific model are accumulated in the two panels.

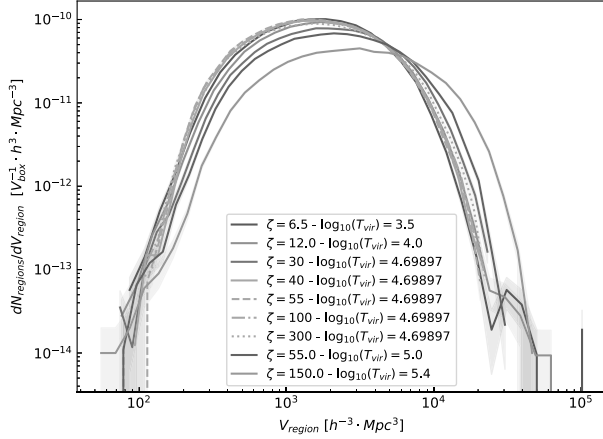
( $10 \text{ cMpc } h^{-1}$ ), but it is not clear yet what kind of cut-off is applied to these PDFs in the regime of large reionisation patches. It is for instance known that many features in the 21 cm signal only converge at larger scales ( $200 \text{ cMpc } h^{-1}$  at least; Iliev et al. 2014), and further investigations will be conducted in the future to show how this translates into our reionisation patches.

## 4. Orientation of the reionisation patches with respect to the gas filaments

This section examines the orientation of the reionisation patches with respect to matter. The directions of the main axis of the gas filaments and the patches they are located in are compared using the eigenvectors of their inertial tensors, computed as explained in Sect. 2.4. To do this, the cosine of the angle between these directions was computed as follows:

$$\cos(\langle z_{\text{reion}}, \delta \rangle) = \cos(\langle \Psi_{z_{\text{reion},3}}, \Psi_{\delta,3} \rangle), \quad (6)$$

$$= \frac{\Psi_{z_{\text{reion},3}} \cdot \Psi_{\delta,3}}{\|\Psi_{z_{\text{reion},3}}\| \|\Psi_{\delta,3}\|}, \quad (7)$$



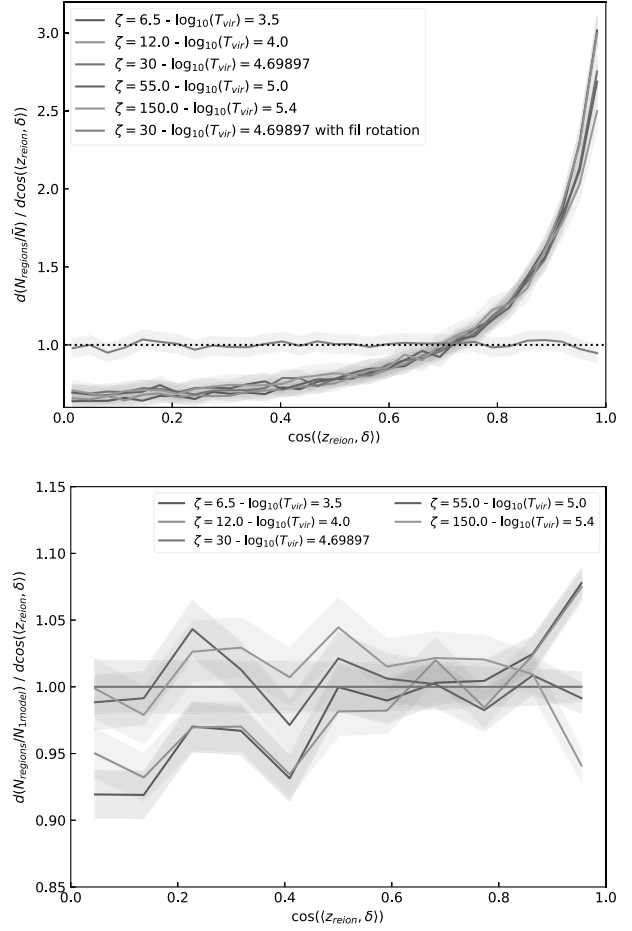
**Fig. 8.** Probability distribution functions of the volume of segmentation patches for each physical model listed in Table 1. Here, all the patches of each run for one specific model are accumulated.

$\Psi_{z_{\text{reion}},3}$  and  $\Psi_{\delta,3}$  are the eigenvectors corresponding to the highest eigenvalue of the  $z_{\text{reion}}$  patches and the filaments, respectively.

#### 4.1. Emitting haloes with different mass ranges and ionising efficiency (SRH set)

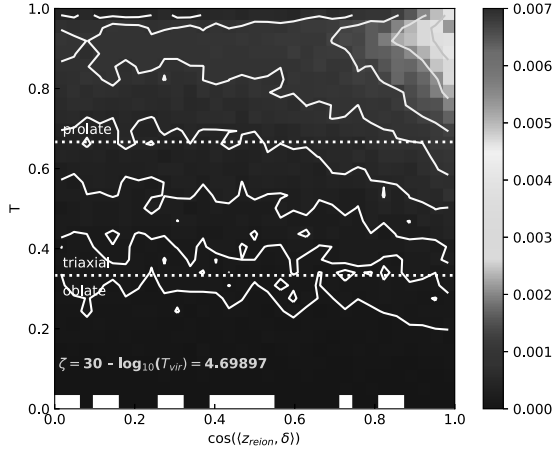
We first consider the probability distribution functions of the patch-filament alignment for every model of the SRH set. The PDFs are shown in Fig. 9. For each model, we accumulated the cosines of all of their patches in order to improve the statistics of the result. The top panel shows that most reionisation patches are rather aligned with their gas filament. In addition, there does not seem to be any great difference between the different models of this set. The bottom panel, showing PDFs normalised to the  $\zeta = 30$  and  $T_{\text{vir}} = 5 \times 10^4$  K fiducial model, shows the slight variations between the five models. The models are well ordered in the  $\cos(\langle z_{\text{reion}}, \delta \rangle) \sim 1$  regime: in simulations with the smaller emitting haloes, more patches are aligned with the density. Conversely, models with strong emitters more frequently show perpendicular configurations between matter and radiation, but they are still dominated by aligned situations. The grey curve in the top panel corresponds to the model with  $\zeta = 30$  and  $T_{\text{vir}} = 5 \times 10^4$  K, for which each filament was rotated with a randomly chosen angle for safety checks. In this case, filaments are oriented randomly, which means that their orientation compared to the radiation is also random. The PDF of this test is rather flat and shows the expected random behaviour.

We now know that most reionisation patches are aligned with their filament. In Sect. 2.4, we showed that most reionisation patches have a prolate shape. The question now is whether these behaviours are correlated. Figure 10 is a 2D distribution of the triaxiality parameter versus the cosine of angles between the radiations and filaments for the fiducial model ( $\zeta = 30$  and  $\log_{10}(T_{\text{vir}}) = 4.69897$ ). The shape and the orientation of  $z_{\text{reion}}$  patches are indeed correlated: they are generally prolate and aligned with the filaments. This distribution is very similar for every other model (not shown here). Our simulations therefore show a clear regime of prolate-aligned reionisation patches (see the top right region with  $T \geq \frac{2}{3}$  and  $\cos(\langle z_{\text{reion}}, \delta \rangle) \geq 0.5$  of the  $T$  versus  $\cos(\langle z_{\text{reion}}, \delta \rangle)$  space, which includes 46.6% of the total number of patches). Other types of patches are visible as well, even if there are rarer. For example, the opposite of



**Fig. 9.** Probability distribution functions of the cosine of the angle between the reionisation patches and their corresponding filaments (*top panel*). *Bottom panel*: same PDFs are shown, but divided by the one of the model with  $\zeta = 30$  and  $\log_{10}(T_{\text{vir}}) = 4.69897$ . Each model in which  $\zeta$  and  $T_{\text{vir}}$  varied at the same time is represented here, and all patches of their 101 runs are accumulated in both panels. *Top panel*: the model with  $\zeta = 30$  and  $\log_{10}(T_{\text{vir}}) = 4.69897$  is also represented (in grey) with a rotation of filaments for a comparison with total randomness.

the prolate-aligned patches are the oblate-perpendicular patches to the filaments (see the bottom left region with  $T \leq \frac{1}{3}$  and  $\cos(\langle z_{\text{reion}}, \delta \rangle) \leq 0.5$  of the  $T$  versus  $\cos(\langle z_{\text{reion}}, \delta \rangle)$  space): there are only a few of them (2.36%), and they are in fact patches with butterfly shapes. Figure 11 shows these two types of patches: the pink cells belong to the filament, and the rainbow-coloured cells belong to the reionisation patches. The top panel shows an example of a prolate-aligned region. It clearly shows its prolate shape, and the reionisation patch follows the direction of the filament well. In the bottom panel, the patch looks like a butterfly with two wings that surround the filament: it is easier for the radiation to escape perpendicularly to the filament, as these directions are less dense. Another interesting feature of this butterfly patch is that regions close to the filaments have clearly been reionised earlier, while the reionisation redshift decreases as the distance to the filaments increases. The same behaviour is observed in the prolate-aligned region of the top panel, but in a less impressive manner. These examples clearly show that the sources are



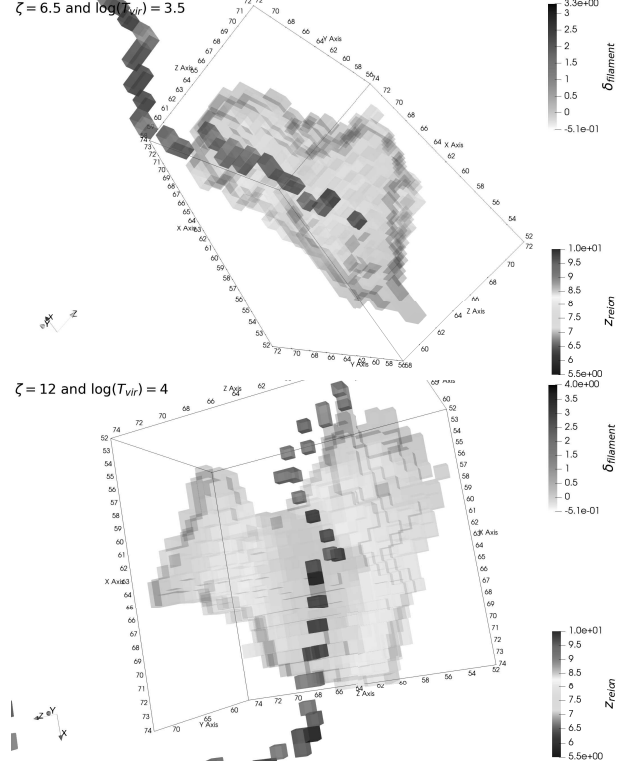
**Fig. 10.** 2D histogram representing the logarithm of the ratio of the number of patches to the total number of patches of the simulation in bins of triaxiality parameter and cosine of the angle between reionisation patches and filaments for the fiducial model (with  $\zeta = 30$  and  $\log_{10}(T_{\text{vir}}) = 4.69897$ ). The white full lines are the isocontours of the histogram. The coloured cells of this histogram correspond to non-zero values, and the white cells indicate zero values.

located within the filaments or very close to them. This leads to an inside-out scenario of the EoR.

The difference between fiber-like and butterfly shapes probably originates in the population of sources within a given patch. Fiber-like regions might include a collection of radiation sources that lie like pearls on a string along the filaments, with similar properties in terms of ignition time, emissivity, and environment: in these situations, the source, located at the reionisation peak of a given patch (i.e. the patch seed) is not different to close-by sources within the same region. The differences in reionisation times values and gradients created by these sources are not sufficient to cause DisPerSE to segment a region into small individual reionisation patches. It rather produces extended patches along filaments. On the other hand, butterfly patches might be the mark of a patch dominated by its seed source (because it is isolated and/or brighter than other local emitters), which alone drives the propagation of ionisation fronts from the very beginning, preferentially along the path of least resistance, perpendicularly to filaments. The latter geometry is clearly sub-dominant in our models, however. This trend can be accentuated by the late reionisation scenario of the SRH dataset (compared to e.g the DRH models): at late times, a greater number of locations are susceptible to host sources, and patches are rarely dominated by a single source. It should be noted that an increase in both  $T_{\text{vir}}$  and  $\zeta$  can boost the propagation of ionisation fronts, cancelling local small variations of reionisation times and gradients around sources along filaments. This would lead to larger coherent patches, and we therefore expect a decrease in the number of fibre-like regions in these cases, as measured indeed in the SRH set.

#### 4.2. Different reionisation histories

We now focus on the simulations of the DRH set (those that vary only  $\zeta$  for the fiducial  $T_{\text{vir}}$ ). Changing  $\zeta$  allows us to study the effect of different histories of reionisation and different halo-ionising efficiencies. Figure 12 shows the PDFs of the cosine of the angle between their reionisation patches and their filaments.



**Fig. 11.** Examples of reionisation patches and their corresponding filament extracted from the segmentations of our simulations: the *top panel* shows a prolate-aligned region, and the *bottom panel* shows an oblate-perpendicular or butterfly region. In shades of pink, we show filaments, and rainbow colors indicate the  $z_{\text{reion}}$  region. The redder the region, the earlier it reionised. In contrast, the bluer the region, the later it reionised.

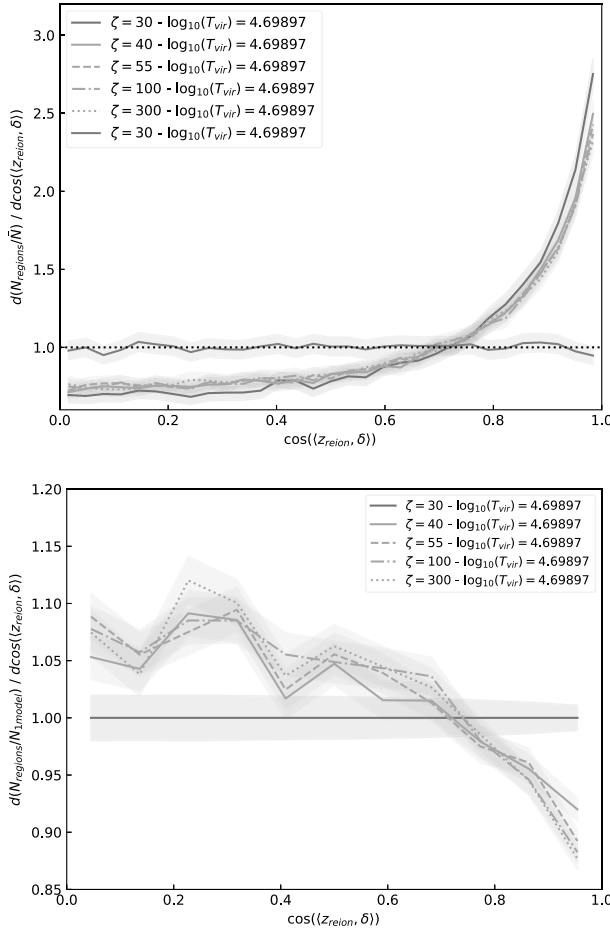
At first sight (top panel of this figure), there are no significant differences between the five models: most patches are aligned with their filaments. The PDF ratios of the fiducial model show a weak trend: the higher  $\zeta$ , the weaker the alignment of the patches with the filaments, and there is a slight increase in the number of perpendicular butterfly patches. This is consistent with the above hypothesis regarding the origins of the different geometries: DRH models with large  $\zeta$  (while keeping  $T_{\text{vir}}$  constant) attribute strong emissivities to the first and rare sources that seed the patches and let them extend their radiative influence perpendicularly to the filaments at early times.

### 5. Effect of the persistence parameter

In the results described above, all the segmentations of the density and the reionisation redshift fields were obtained using DisPerSE with a fixed value of the persistence threshold. In this section, we propose a brief comparison of the main results shown above, but with different persistence values. This comparison is only made for the fiducial 21cmFAST model ( $\zeta = 30$  and  $T_{\text{vir}} = 5 \times 10^4$  K) for the sake of brevity, but it is very likely that the conclusions remain the same for every other model.

The segmentations of the density and redshift of reionisation fields (of the fiducial 21cmFAST model) were therefore computed with DisPerSE using four other persistence values, which were  $0.625\text{-}\sigma$ ,  $0.75\text{-}\sigma$ ,  $0.875\text{-}\sigma$ , and  $1\text{-}\sigma$ . Figure 13 shows the accumulated numbers of reionisation patches found in these fields

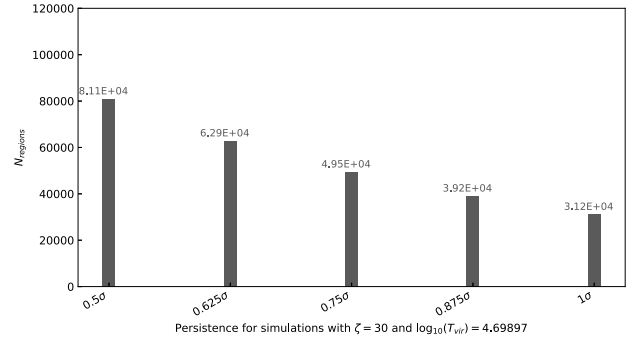




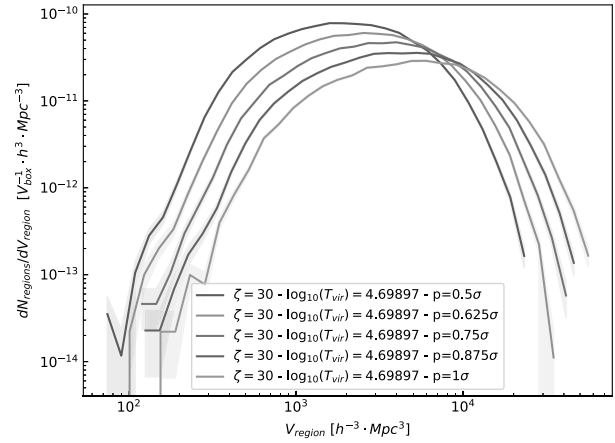
**Fig. 12.** Probability distribution functions of the cosine of the angle between the reionisation patches and their corresponding filaments (*top panel*). *Bottom panel*: same PDFs are shown, but they are divided by the one of the model with  $\zeta = 30$  and  $\log_{10}(T_{\text{vir}}) = 4.69897$ . Each of the models that only varied  $\zeta$  is represented here, and all patches of their 101 runs are accumulated in both panels. The model with  $\zeta = 30$  and  $\log_{10}(T_{\text{vir}}) = 4.69897$  is also represented (*top panel*, in grey) with a rotation of filaments for comparison with total randomness.

for every one of the 101 runs. The stronger the persistence, the lower the number of patches, as expected because critical points are more easily merged. This means that in the resulting segmentations with these stronger persistences, the number of critical points is smaller, which explains the smaller number of patches. This behaviour directly affects the volume of the reionisation patches, for which the probability distribution functions are shown in Fig. 14. The correlation of the number of patches and their volume is clearly visible here: the stronger the persistence, the larger the volume of patches and the smaller the number patches because the total volume is conserved.

Persistence is related to percolation. For instance, the 1D representation of the persistence in Fig. 3 shows that when a persistence threshold is applied, the leftmost maximum (reionisation seed in our case) disappears and is associated with the other maximum. A weak persistence applied to the reionisation redshift maps allows less prominent and more numerous patches to be taken into account in the final segmentations. Conversely, high persistence values favour the merging of patches



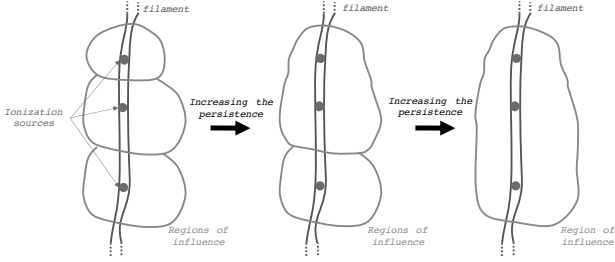
**Fig. 13.** Number of patches in the fiducial model ( $\zeta = 30$  and  $T_{\text{vir}} = 5 \times 10^4$  K) that is detected using DisPerSE with different persistence levels. For each one of them, the patch numbers of every run were added.



**Fig. 14.** Probability distribution functions of the volume of segmentations patches for the fiducial model ( $\zeta = 30$  and  $T_{\text{vir}} = 5 \times 10^4$  K) with the different persistences used in DisPerSE. Here, all the patches of each run for one specific model are accumulated.

by DisPerSE, as shown in Fig. 15. The persistence applied to the reionisation redshift maps can thus be seen as a relative measure of the different timings of creating two patches, and it can therefore follow the percolation process. We could in principle retrieve the relative progression of the fusion of ionised bubbles (regardless of the time), which would provide a complementary view of the time progression of the percolation, which is the subject of many studies in the literature (see e.g. Chen et al. 2019; Giri et al. 2019; Gorce & Pritchard 2019; Giri & Mellema 2021). Elbers & van de Weygaert (2019) pointed out that the persistence can follow the lifetime of a topological feature and then the creation and evolution of structures, which is the percolation process. They showed with simple models that they can follow the stages of reionisation and know in which stage the bubbles are by coupling the persistence homology with the Betti numbers.

By computing the triaxiality parameter for every new persistence, we can plot the PDFs shown in the top panel of Fig. 16. Generally, all the segmentations with the different persistences behave in the same way: they have a majority of prolate reionisation patches. In the detailed ratios with the  $0.5\text{-}\sigma$  curves (bottom panel of the same figure), however, a trend with increasing persistences is clear: the stronger the persistence, the larger the number of prolate patches. The percolation process



**Fig. 15.** 2D illustration of the percolation process using the persistences. A matter filament is represented in blue with many ionisation seeds (in red) within it. These seeds create their patch of radiative influence (shown in orange), and when the persistence increases (from left to right), smaller patches merge to form larger patches.

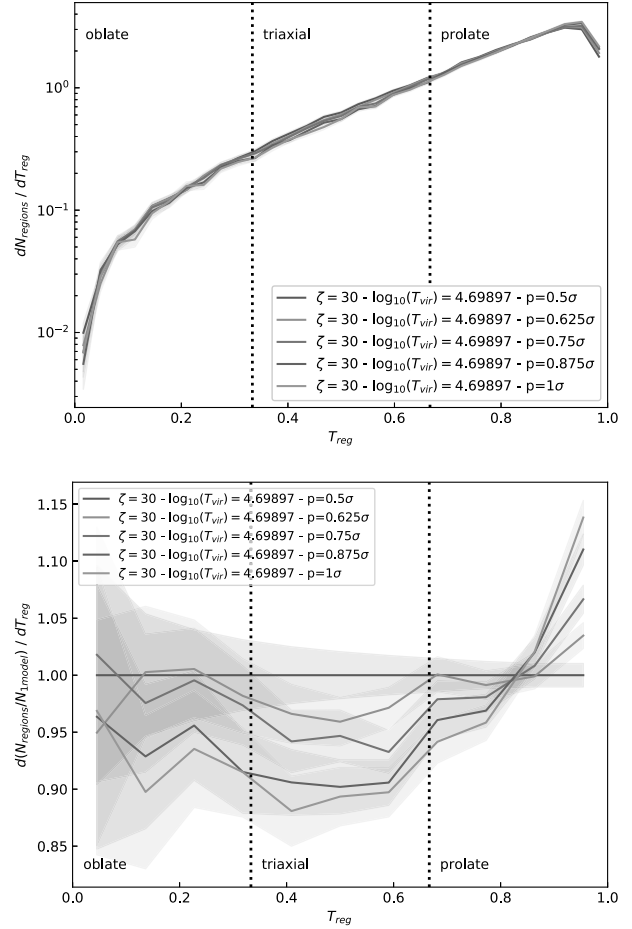
therefore favours the merging of patches along filaments and promotes prolate-aligned configurations, with patches like pearls on a string along filaments, as sketched again in Fig. 15. Again, our study of the persistence recalls the process detailed by Chen et al. (2019), who reported that small ionised bubbles grew during their “ionised bubble” stage and merged to form the larger and fibre-like regions during the “ionised fibre” stage. Changing the persistence parameter allows us to retrieve a similar behaviour in our segmentations.

Nevertheless, our segmentations appear to be robust to the changes in persistence, and the main conclusions described in previous sections are unchanged. The shape of the reionisation patches (see Fig. 16) remains very similar, and the orientation of the patches with respect to the filaments is almost unchanged as well, as we show in Fig. 17.

## 6. Comparison with EMMA simulations

We now compare the topology of EoR simulations obtained with the semi-numerical code 21cmFAST to that of the cosmological simulation code EMMA (Electromagnétisme et Mécanique sur Maille Adaptative, Aubert et al. 2015). EMMA is an adaptive mesh refinement (AMR) code that couples radiative transfer and hydrodynamics. Light is described as a fluid (resolved using the moment-based M1 approximation, Aubert & Teyssier 2008).

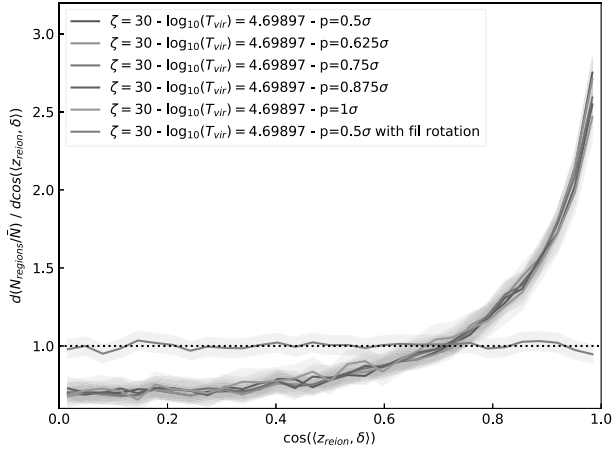
The EMMA simulations set we employed was composed of two simulations with a cosmology given by Planck Collaboration VI (2020), without AMR and no reduced speed of light (see Gillet et al. 2021 for more information about the simulations). The difference between the two simulations is the mass resolution of the stellar particles:  $10^7 M_{\odot}$  for the one labelled EMMA Mslow and  $10^8 M_{\odot}$  for the other simulation. As detailed in Gillet et al. (2021), stellar particles are converted from gas according to a Poissonian process. Cells that convert an amount of gas into stars that is only slightly larger than the stellar mass naturally lead to a stochastic star formation process. Conversely, if the same amount of gas is much larger than the stellar mass particle, it induces a steady stream of particles. As a consequence, the EMMA Mslow model presents a diffuse population of stars, whereas the other model exhibits a stellar biased population focused on the densest cells. For these two simulations, the boxes have  $512^3$  cells for  $512^3 \text{ cMpc}^3 h^{-3}$ . We then cut these boxes in 64 sub-boxes of  $128^3$  cells and  $128^3 \text{ cMpc}^3 h^{-3}$ , creating 64 runs of the two EMMA models. The resolution of these sub-boxes was the same as that of the 21cmFAST simulations. Comparisons with the 21cmFAST models were made with 64 runs of each 21cmFAST simulation for consistency.



**Fig. 16.** Probability distribution functions of the triaxiality parameter for the fiducial model ( $\zeta = 30$  and  $T_{\text{vir}} = 5 \times 10^4 \text{ K}$ ) with the different persistences used in DisPerSE (top panel). Bottom panel: same PDFs, but each one of them is divided by one particular model (the fiducial model with a persistence of  $0.5\text{-}\sigma$ ). All the patches of each run for one specific model are accumulated in both panels.

Figure 18 shows the ionised volume fractions of every simulation of this study. The EMMA simulations are fully ionised approximately at the same time as those of 21cmFAST. With EMMA, however, the reionisation starts later and is more abrupt, which might be explained by the fact that EMMA simulations have larger volumes than the 21cmFAST simulations, with greater voids and larger haloes.

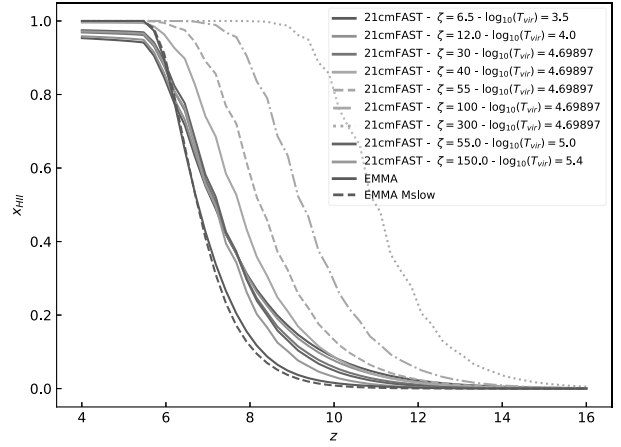
Again, before we used DisPerSE on the EMMA fields, the same transformations as described in Sect. 2.3 were applied to each sub-box, using the means and the standard deviations of the  $512^3 \text{ cMpc}^3 h^{-3}$  boxes instead of those of each sub-grid. We thus took the cosmic variance of the whole box into account. After obtaining the segmentations with DisPerSE with the same persistence as was used for all the 21cmFAST simulations ( $0.5\text{-}\sigma$ ), the distribution of patch volumes was computed. This is shown in Fig. 19 for every EMMA and 21cmFAST simulation, accumulating 64 runs for each one of them. The EMMA simulations have smaller patches and a greater number of them than the 21cmFAST simulations. The EMMA model with the massive,  $10^8 M_{\odot}$ , stellar particles is significantly different from the 21cmFAST simulations: the patch size distribution is clearly shifted to smaller



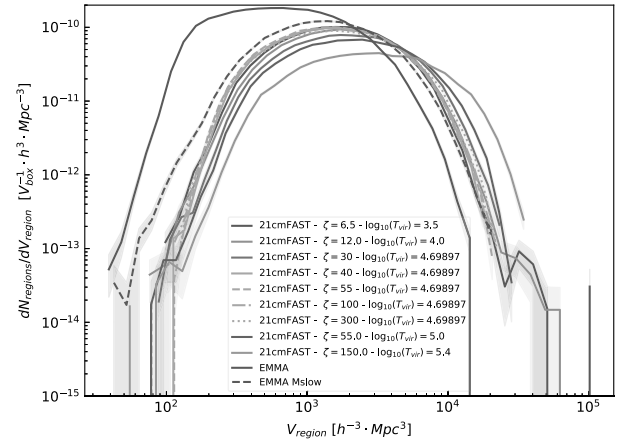
**Fig. 17.** Probability distribution functions of the cosine of the angle between the reionisation patches and their corresponding filaments (as explained in Sect. 2.4) of the fiducial model ( $\zeta = 30$  and  $T_{\text{vir}} = 5 \times 10^4$  K) with the different persistences used in DisPerSE, accumulating all patches of every run. The model with the  $0.5\text{-}\sigma$  persistence is also represented (in grey) with a rotation of filaments in order for comparison with total randomness.

volumes with a large number of small patches. The EMMA Mslow model with the less massive,  $10^7 M_{\odot}$ , stellar particles is more like the 21cmFAST models: the more diffuse source distribution allowed by this setting recovers topological features similar to those of the 21cmFAST models (in which all locations are a potential source of radiation with its own distribution of emitting haloes).

The tensors and main directions of the reionisation patches and filaments of the EMMA simulations were computed and allowed us to compare the direction of the propagation of radiation with respect to the filament. The top panel of Fig. 20 shows the distributions of the cosine of the angle between these directions for the EMMA and the 21cmFAST simulations. Patches and gas filaments in the EMMA simulation are aligned similarly to the 21cmFAST models, and slightly more aligned in the EMMA Mslow simulation. The bottom panel of Fig. 20 shows the ratio of each PDF of cosines with the fiducial 21cmFAST model. Again, the EMMA Mslow shows very similar PDFs to the 21cmFAST models, which implies that it is possible to recover the same type of behaviour as in semi-analytical simulations with a fully numerical code even if the reionisation occurs differently. Conversely, the EMMA model with massive,  $10^8 M_{\odot}$ , stellar particles is different in terms of alignments. The scarce, stochastic, and biased distribution of sources in this models leads to many more perpendicular patches than in all the other models. While this model is in reasonable agreement in terms of reionisation history, its topology is quite different: this emphasises that models that present some reasonable global agreement can be vastly different in terms of topology when the details of the source models are not equivalent, for example. The higher star formation stochasticity and the more biased distribution of sources lead to trends that are similar to the higher virial temperatures in the 21cmFAST simulations, with isolated reionisation seeds that dominate their environment more strongly. The shapes of the reionisation patches in the EMMA simulations set are also rather prolate in both simulations, as in the 21cmFAST simulations, and no significant difference between the models was found (not shown here).



**Fig. 18.** Ionized volume fraction of all the 21cmFAST simulations and the two EMMA simulations having different models, each one of them averaged over 64 runs.

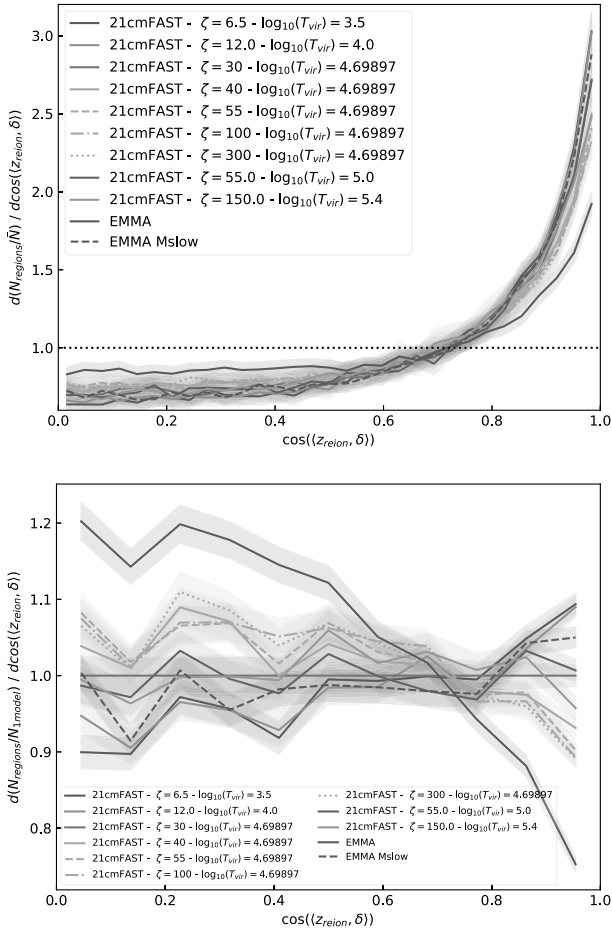


**Fig. 19.** Probability distribution functions of the volume of segmentation patches for every 21cmFAST and EMMA simulation. Here, all the patches of each 64 runs for one specific model are accumulated.

## 7. Conclusions

We have analysed the topological structures extracted from different EoR models made with the 21cmFAST semi-analytical code and the EMMA cosmological simulation code. For this purpose, we used the reionisation redshift and gas density fields. The reionisation redshift field corresponds to the time when the gas has reionised and contains spatial and temporal information about the physical processes during the EoR. To extract the topological information, we used DisPerSE, which applies concepts of Morse theory, to the density field to obtain the filaments of matter and to the reionisation redshift field to segment it into reionisation patches coming from the same seed of local reionisation.

In the 21cmFAST models, the size distributions of the reionisation patches have an average radius of  $10 \text{ cMpc } h^{-1}$ , quantifying the typical extent of the radiative influence of the seeds of reionisation. In general, an increase in the virial temperature and the ionising efficiency of haloes decreases the number of patches and expands them. This extends the reach of the radiative influence of the seeds of local reionisation.



**Fig. 20.** Probability distribution functions of the cosine of the angle between the reionisation patches and their corresponding filaments (*top panel*). *Bottom panel*: same PDFs, but divided by the one of the model with  $\zeta = 30$  and  $\log_{10}(T_{\text{vir}}) = 4.69897$ . Each 21cmFAST model is represented here, alongside with the two EMMA simulations, accumulating all patches of 64 runs in both panels.

Using the inertial tensors of the reionisation patches and density filaments, we were able to extract their shape and orientation. We found that the reionisation patches are statistically prolate and aligned with the local gas filaments. Radiation escapes from sources that lie like pearls on a string along the gas filament to reach the zones that reionise last (typically, the voids) in an inside-out manner. A minority of patches shows a butterfly shape. These are oblate patches with radiation that escapes perpendicularly to the local filaments. The difference probably arises from the properties of sources within the patches: if the initial source that defines a reionisation patch is locally dominant in terms of photon output, butterfly shapes can appear, whereas if the initial source is only the first of a collection of rather similar close-by emitters, patches become aligned with the local distribution of matter. We note that the dominant status of a source can depend on many factors (intrinsic emission, escape fraction, density, and configuration of local absorbers) and is therefore likely to be resolution and/or model dependent. Our results, obtained at rather low resolution ( $1.48^3 \text{ cMpc}^3$ ), should be revisited, for instance, at higher resolution.

Comparisons of the 21cmFAST segmentations with those obtained from EMMA simulations showed that EMMA simulations can provide results similar to the semi-analytical predictions in terms of patch properties (volume, number) and relative orientations to the filaments of matter. We also find that EMMA simulations with increased bias and stochasticity for star formation can also lead to quite different topologies on the same criteria, even though they present similar reionisation histories.

These results appear to be robust when the persistence parameter with DisPerSE was changed. As the persistence increases and patches are merged, the patches expand and become more prolate. This provides insight into the percolation process of our EoR models directly from a single reionisation map. We find that during the first stages of the reionisation, isolated ionised bubbles appear and merge with time in order to form larger and fibre-like ionised bubbles (Chen et al. 2019). This subject could be further investigated, as it is complementary to standard percolation studies based on sequences of snapshots of ionised fractions maps.

To summarise, the reionisation redshift field contains much spatial and temporal information and is found to be quite useful in the current topological framework. In addition, statistical analyses of this field are quite practical because only one  $z_{\text{reion}}$  field needs to be extracted from each simulation. We probably only skimmed the potential of the approach developed here. We already described the possibilities offered by the persistence parameter. We could also investigate the halo properties in relation to the host reionisation patches. Moreover, we could work on smaller scales to determine whether the study of the patch geometry (size, shape, orientation) might hold information about the ways in which radiation escapes from galaxies, how far they affect their environment, and even about the time evolution or duty cycles of the photons production (see e.g. Deparis et al. 2019). However, such studies of galaxies or haloes are not possible at the resolution discussed here. Ideally, they would be conducted with models with Cosmic Dawn (CoDa)-like characteristics (Ocvirk et al. 2016). Finally, the  $z_{\text{reion}}$  field is not a direct observable, and similar analyses might be directly conducted on the topology of 21 cm maps or light cones. These geometrical studies may reveal a way to relate these observations to the timing of the reionisation process.

*Acknowledgements.* We thank Adélie Gorce for her proofreading and her precious questions and advice. We also thank Christophe Pichon for his help with the topological code DisPerSE. Furthermore, we express our gratitude to the referee of this paper for the useful and constructive comments that allowed us to make a more comprehensive and complete paper. This work was granted access to the HPC resources of CINES under the allocations 2020-A0070411049 and 2021-A0090411049 “Simulation des signaux et processus de l’aube cosmique et Réionisation de l’Univers” made by GENCI. This research made use of ASTROPY, a community-developed core Python package for astronomy (Astropy Collaboration 2018); MATPLOTLIB, a Python library for publication quality graphics (Hunter 2007); SCIPY, a Python-based ecosystem of open-source software for mathematics, science, and engineering (Virtanen et al. 2020); NUMPY (Harris et al. 2020) and PYTHON (Perez & Granger 2007).

## References

- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, 156, 123
- Aubert, D., & Teyssier, R. 2008, *MNRAS*, 387, 295
- Aubert, D., Deparis, N., & Ocvirk, P. 2015, *MNRAS*, 454, 1012
- Battaglia, N., Trac, H., Cen, R., & Loeb, A. 2013, *ApJ*, 776, 81
- Becker, G. D., Bolton, J. S., Madau, P., et al. 2015, *MNRAS*, 447, 3402
- Chardin, J., Puchwein, E., & Haehnelt, M. G. 2017, *MNRAS*, 465, 3429
- Chen, Z., Xu, Y., Wang, Y., & Chen, X. 2019, *ApJ*, 885, 23
- Choudhury, T. R., & Paranjape, A. 2018, *MNRAS*, 481, 3821
- Codis, S., Jindal, A., Chisari, N. E., et al. 2018, *MNRAS*, 481, 4753

## E. Th  lie: First look at the topology of reionisation redshifts in models of the epoch of reionisation

- Cohn, J. D. 2021, ArXiv e-prints [arXiv:2108.02292]
- Deparis, N., Aubert, D., Ocvirk, P., Chardin, J., & Lewis, J. 2019, *A&A*, 622, A142
- Dixon, K. L., Iliev, I. T., Mellema, G., Ahn, K., & Shapiro, P. R. 2016, *MNRAS*, 456, 3011
- Elbers, W., & van de Weygaert, R. 2019, *MNRAS*, 486, 1523
- Friedrich, M. M., Mellema, G., Alvarez, M. A., Shapiro, P. R., & Iliev, I. T. 2011, *MNRAS*, 413, 1353
- Gal  rraga-Espinosa, D., Aghanim, N., Langer, M., Gouin, C., & Malavasi, N. 2020, *A&A*, 641, A173
- Gillet, N. J. F., Aubert, D., Mertens, F. G., & Ocvirk, P. 2021, ArXiv e-prints [arXiv:2103.03061]
- Giri, S. K., & Mellema, G. 2021, *MNRAS*, 505, 1863
- Giri, S. K., Mellema, G., Dixon, K. L., & Iliev, I. T. 2018, *MNRAS*, 473, 2949
- Giri, S. K., Mellema, G., Aldheimer, T., Dixon, K. L., & Iliev, I. T. 2019, *MNRAS*, 489, 1590
- Gleser, L., Nusser, A., Ciardi, B., & Desjacques, V. 2006, *MNRAS*, 370, 1329
- Gorce, A., & Pritchard, J. R. 2019, *MNRAS*, 489, 1321
- Greig, B., & Mesinger, A. 2015, *MNRAS*, 449, 4246
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, 585, 357
- Hong, S. E., Ahn, K., Park, C., et al. 2014, *J. Korean Astron. Soc.*, 47, 49
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, 9, 90
- Hutter, A., Watkinson, C. A., Seiler, J., et al. 2020, *MNRAS*, 492, 653
- Iliev, I. T., Mellema, G., Pen, U. L., et al. 2006, *MNRAS*, 369, 1625
- Iliev, I. T., Mellema, G., Ahn, K., et al. 2014, *MNRAS*, 439, 725
- Kakiichi, K., Majumdar, S., Mellema, G., et al. 2017, *MNRAS*, 471, 1936
- Katz, H., Ramsay, M., Rosdahl, J., et al. 2020, *MNRAS*, 494, 2200
- Kim, H.-S., Wyithe, J. S. B., Park, J., & Lacey, C. G. 2013, *MNRAS*, 433, 2476
- Kulkarni, G., Keating, L. C., Haehnelt, M. G., et al. 2019, *MNRAS*, 485, L24
- Lee, K.-G., Cen, R., Gott, J. R., III, & Trac, H. 2008, *ApJ*, 675, 8
- Lin, Y., Oh, S. P., Furlanetto, S. R., & Sutter, P. M. 2016, *MNRAS*, 461, 3361
- Malavasi, N., Aghanim, N., Douspis, M., Tanimura, H., & Bonjean, V. 2020, *A&A*, 642, A19
- Mellema, G., Iliev, I. T., Pen, U.-L., & Shapiro, P. R. 2006, *MNRAS*, 372, 679
- Mellema, G., Koopmans, L. V. E., Abdalla, F. A., et al. 2013, *Exp. Astron.*, 36, 235
- Mellema, G., Koopmans, L., Shukla, H., et al. 2015, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 10
- Mesinger, A., & Furlanetto, S. 2007, *ApJ*, 669, 663
- Mesinger, A., Furlanetto, S., & Cen, R. 2011, *MNRAS*, 411, 955
- Murray, S., Greig, B., Mesinger, A., et al. 2020, *J. Open Sour. Softw.*, 5, 2582
- Ocvirk, P., Gillet, N., Shapiro, P. R., et al. 2016, *MNRAS*, 463, 1462
- Pagano, M., & Liu, A. 2020, *MNRAS*, 498, 373
- Perez, F., & Granger, B. E. 2007, *Comput. Sci. Eng.*, 9, 21
- Planck Collaboration VI. 2020, *A&A*, 641, A6
- Seiler, J., Hutter, A., Sinha, M., & Croton, D. 2019, *MNRAS*, 487, 5739
- Shaw, A. K., Bharadwaj, S., & Mondal, R. 2020, *MNRAS*, 498, 1480
- Singh, A., Mahajan, S., & Bagla, J. S. 2020, *MNRAS*, 497, 2265
- Song, H., Laigle, C., Hwang, H. S., et al. 2021, *MNRAS*, 501, 4635
- Sousbie, T. 2011, *MNRAS*, 414, 350
- Sousbie, T., Pichon, C., & Kawahara, H. 2011, *MNRAS*, 414, 384
- Tanimura, H., Aghanim, N., Bonjean, V., Malavasi, N., & Douspis, M. 2020, *A&A*, 637, A41
- Tormen, G. 1997, *MNRAS*, 290, 411
- Trac, H., Chen, N., Holst, I., Alvarez, M. A., & Cen, R. 2021, *ApJ*, submitted [arXiv:2109.10375]
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Meth.*, 17, 261
- Yoshiura, S., Shimabukuro, H., Takahashi, K., & Matsubara, T. 2017, *MNRAS*, 465, 394
- Zahn, O., Lidz, A., McQuinn, M., et al. 2007, *ApJ*, 654, 12
- Zarka, P., Girard, J. N., Tagger, M., & Denis, L. 2012, in *SF2A-2012: Proceedings of the Annual Meeting of the French Society of Astronomy and Astrophysics*, eds. S. Boissier, P. de Laverny, N. Nardetto, et al., 687
- Zel'Dovich, Y. B. 1970, *A&A*, 500, 13

# CHAPITRE 6

## CONCEPTS ET MESURES DE LA TOPOLOGIE DE L'ÉVOLUTION DE LA RÉIONISATION

*Dans ce Chapitre, j'explore plus systématiquement la topologie du processus de réionisation au travers du champ 2D de temps de réionisation. J'y mesure de nouvelles statistiques : la distribution du gradient de la norme des temps de réionisation, la longueur de ses isocontours, la distribution du nombre de sources de réionisation et la distribution de la longueur de squelette. Ce large éventail de mesures montre d'autant plus l'utilité de travailler sur ce genre de champ. De plus, je les compare aux prédictions de la théorie des champs Gaussiens aléatoires (GRFs). Cette étude me permet ainsi de comprendre cette théorie, et de l'appliquer pour la première fois dans le contexte de la Réionisation. Je soulève en particulier un point fort du champ de temps de réionisation, qui est sa « quasi-Gaussianité » à des échelles de l'ordre de celles que nous allons observer à l'avenir avec SKA par exemple. Les résultats détaillés dans ce Chapitre sont extraits de l'article « Topology of reionisation times : Concepts, measurements, and comparisons to Gaussian random field predictions » présenté en fin de Chapitre.*

## Sommaire

---

<b>6.1</b>	<b>Introduction</b>	<b>151</b>
<b>6.2</b>	<b>Extraction et traitement des temps de réionisation d'une simulation EMMA</b>	<b>152</b>
<b>6.3</b>	<b>Évolution du processus de réionisation au travers des temps de réionisation</b>	<b>154</b>
6.3.1	Histoire de réionisation	154
6.3.2	Vitesse des fronts d'ionisation	156
6.3.3	Évolution de la taille des bulles ionisées et régions neutres	158
6.3.4	Percolation des fronts d'ionisation	160
6.3.5	Décompte du nombre de sources de réionisation	162
<b>6.4</b>	<b>Gaussianité du champ de temps de réionisation</b>	<b>163</b>
6.4.1	Comment procédons-nous ?	164
6.4.2	Étude de la gaussianité des temps de réionisation	166
<b>6.5</b>	<b>Comparaisons entre les temps et les redshifts de réionisation</b>	<b>169</b>
6.5.1	Temps et redshifts de réionisation	169
6.5.2	Comparaisons des mesures topologiques	169
<b>6.6</b>	<b>Comparaisons entre simulations cosmologiques et semi-analytiques</b>	<b>171</b>
6.6.1	Simulation 21cmFAST	171
6.6.2	Comparaisons des mesures topologiques	172
<b>6.7</b>	<b>Conclusions &amp; perspectives</b>	<b>172</b>
<b>6.8</b>	<b>« Topology of reionisation times : Concepts, measurements, and comparisons to Gaussian random field predictions »</b>	<b>178</b>

---

## 6.1 Introduction

Dans ce Chapitre, je continue d’analyser la topologie du processus de réionisation, mais en me focalisant cette fois-ci sur le champ de temps de réionisation (qui contient en chaque endroit le moment auquel l’hydrogène se fait réioniser). J’explore ici la Réionisation de manière plus approfondie, au travers de la vitesse des fronts d’ionisation, de l’évolution de la taille des bulles ionisées et régions neutres, du décompte des sources de réionisation et de la percolation. Par exemple, comprendre comment et à quelle vitesse les fronts d’ionisation se propagent nous permet de savoir comment le gaz a été réionisé, et [Deparis et al. \(2019\)](#) indiquent que deux phases ont eu lieu pendant la Réionisation. Pendant la première phase, les fronts d’ionisation sont ralentis par le gaz neutre et dense et leur vitesse est plus faible que celle de la lumière ; alors que pendant la seconde phase et vers la fin de la Réionisation, les fronts accélèrent en atteignant les régions les moins denses. Par ailleurs, étudier la percolation des bulles ionisées permet de comprendre les dernières phases de la Réionisation ([Chen et al. 2019](#)). Suivre les bulles ionisées qui percolent et les régions neutres pendant la Réionisation est utile pour contraindre leur morphologie et leur évolution temporelle ([Shimabukuro et al. 2022](#)), mais aussi les propriétés des galaxies sources de la Réionisation (par exemple la masse ou l’émissivité de ces objets), comme en discutent notamment [Giri et al. \(2018a, 2019\)](#). Les propriétés de ces sources de réionisation ont également un impact sur la topologie du processus de réionisation. [Friedrich et al. \(2011\)](#) ont montré que la topologie des bulles ionisées peut être plus ou moins complexe selon les caractéristiques des sources. Par exemple, les bulles ionisées individuelles grossissent bien plus avant de fusionner et de percoler quand seulement les halos les plus massifs (de masse supérieure à  $10^9 M_\odot$ ) contiennent des sources de photons ionisants (ce qui est cohérent avec notre étude sur la taille des patchs de réionisation mentionnée dans le Chapitre 5). Dénombrer les sources et savoir quand elles apparaissent est donc utile pour comprendre comment la Réionisation évolue ensuite. Dans ce Chapitre, les mesures que je vais présenter permettent ainsi d’étudier la topologie de l’évolution temporelle de la Réionisation d’une manière différente de ce qui a déjà été fait dans la littérature en utilisant les temps de réionisation qui nous permettent de résumer toutes ces informations avec un champ unique.

Par ailleurs, le processus de réionisation est connu pour être un phénomène hautement non-Gaussien à cause de la non-linéarité de la formation des structures et de la manière dont le rayonnement ionise la matière. Ces non-Gaussianités sont présentes à toutes les échelles et à tout moment pendant la Réionisation ([Iliev et al. 2006b](#)) et sont d’autant plus intenses à mesure que la Réionisation se poursuit ([Mellema et al. 2006](#); [Dixon et al. 2016](#)). [Iliev et al. \(2006b\)](#) montrent que les non-Gaussianités de la densité sont très fortes dans les bulles ionisées à cause des fortes densités qui s’y trouvent, et que les fonctions de distribution de probabilité (PDFs) de fraction de gaz ionisé et de masse de gaz ionisé sont encore plus non-Gaussiennes que la densité. Des études, faites par exemple par [Mellema et al. \(2006\)](#), [Dixon et al. \(2016\)](#) ou [Banet et al. \(2021\)](#), quantifient également le niveau de non-Gaussianité dans des PDFs reliées au signal 21 cm avec les paramètres d’asymétrie, d’aplatissement ou avec les quantiles. Ces travaux montrent donc que la Réionisation est elle-même également source de non-Gaussianité, et nous nous interrogeons aussi dans ce Chapitre sur le sujet de la Gaussianité du processus de réionisation, mais cette fois-ci au travers du champ de temps de réionisation. Étudier ces non-Gaussianités permet ainsi de comprendre comment la Réionisation évolue



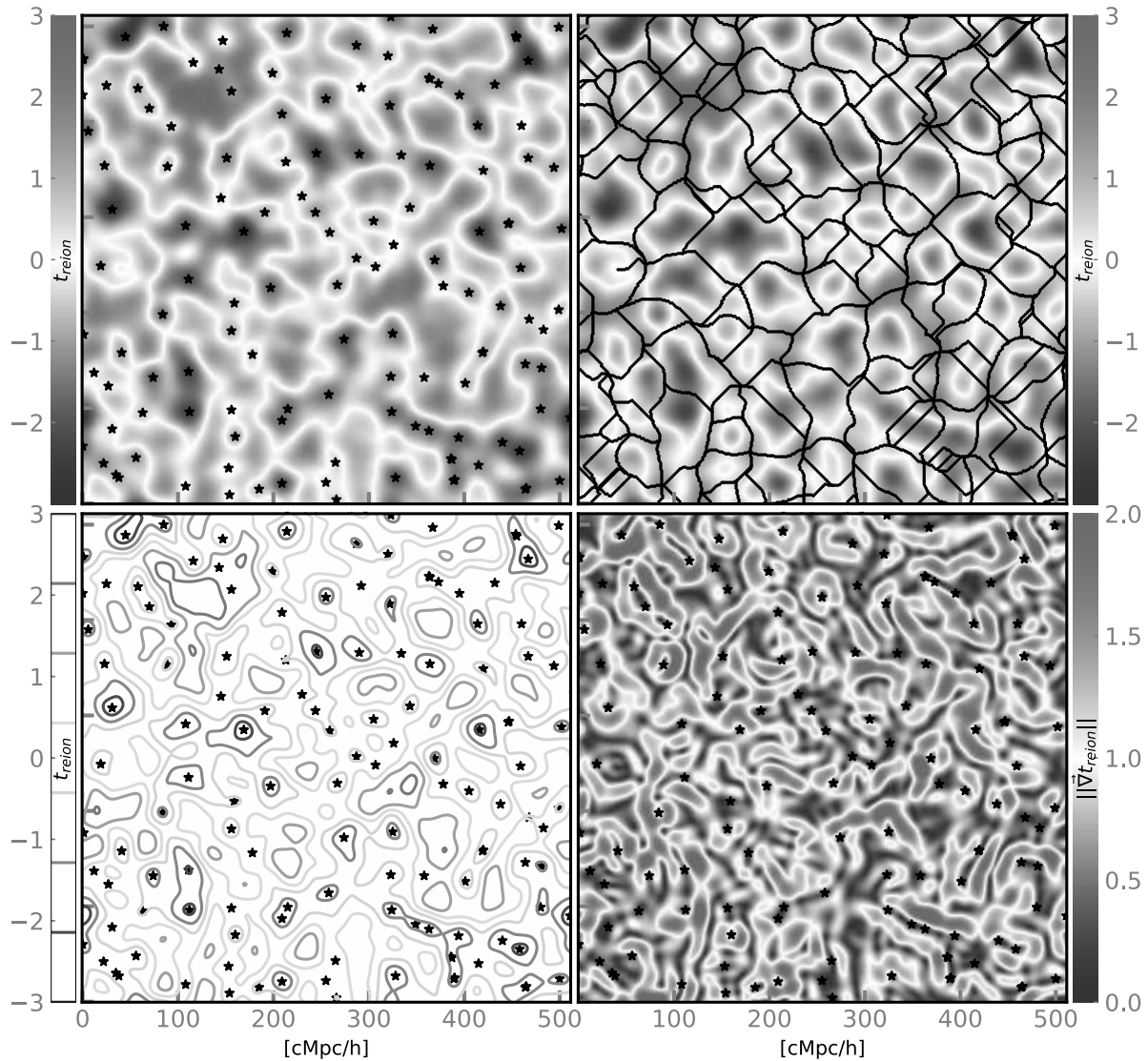
dans son ensemble. Elles compliquent cependant aussi la compréhension de ce processus parce que des statistiques usuelles, comme le spectre de puissance, ne sont plus suffisantes pour obtenir toute l'information présente dans les champs d'intérêt mentionnés dans ce paragraphe. Cette étude a donc pour but de fournir des outils pour étudier l'évolution temporelle et spatiale de la Réionisation de manière complémentaire aux études déjà présentes dans la littérature, tout en ayant l'avantage de pouvoir aussi accéder aux non-Gaussianité de la Réionisation. De plus, on sait avec le Chapitre 4 que l'on aura accès aux temps de réionisation à partir des futures observations du signal 21 cm par SKA, ce qui motive davantage ce travail.

Ce Chapitre commence par décrire la simulation **EMMA** utilisée pour cette étude dans la Section 6.2. Dans la Section 6.3, nous présentons des mesures statistiques de l'évolution du processus de réionisation dans le champ de temps de réionisation. Ensuite, la Gaussianité des temps de réionisation est étudiée dans la partie 6.4. Dans la Section 6.5, nous comparons les mesures faites dans les champs de temps et de redshift de réionisation. Puis, la Section 6.6 présente des comparaisons entre la simulation cosmologique **EMMA** et une simulation semi-analytique **21cmFAST**. Nous concluons enfin dans la Section 6.7. La Section 6.8 est dédiée à l'article « Topology of reionisation times : Concepts, measurements, and comparisons to Gaussian random field predictions » dont les résultats de ce Chapitre sont extraits.

## 6.2 Extraction et traitement des temps de réionisation d'une simulation **EMMA**

Dans ce Chapitre, nous commençons par étudier la topologie 2D des temps de réionisation extraits d'une simulation **EMMA**. Cette simulation a une taille de  $512^3 \text{ cMpc}^3/h^3$  et une résolution de  $1 \text{ cMpc}^3/h^3$  (Gillet et al. 2021). La résolution en masse des particules stellaires est ici de  $10^8 M_\odot$ . Dans ces travaux, je choisis d'étudier ces cartes en deux dimensions parce que nous sommes capables de reconstruire de telles cartes à partir d'observations 2D sur le plan du ciel du signal 21 cm par SKA. J'extrait alors de la simulation **EMMA** 100 cartes 2D de  $t_{\text{reion}}^*(\vec{r})$  (espacées de  $5 \text{ cMpc}/h$ ) de taille  $512^2 \text{ cMpc}^2/h^2$  et de résolution  $1 \text{ cMpc}/h$  sur le plan du ciel et  $1 \text{ cMpc}/h$  dans la direction des fréquences. On remarque ici que la résolution que nous avons choisie dans la direction des fréquences est en fait simplement la résolution des simulations dans la troisième dimension. En fréquence, cette résolution est inférieure à  $0.1 \text{ MHz}$  (elle varie selon le redshift). Les observations SKA auront cependant une résolution bien moins fine dans la direction des fréquences parce que l'on moyenne en général dans cette direction sur une bande de fréquence d'une largeur d'environ  $0.5 \text{ MHz}$  pour améliorer le ratio signal à bruit (Giri et al. 2018b). Les observations auront donc une résolution en fréquence au moins 5 fois moins bonne que celle que nous utilisons ici. Cela signifie que les données d'observations seront un peu plus lissées dans la direction des fréquences, filtrant ainsi les plus petites structures que nous pouvons avoir avec les simulations.

Les cartes de temps de réionisation sont également lissées par des filtrages Gaussiens de tailles  $R_f \in \{1, 2, 6\}$  (exprimés en  $\text{cMpc}/h$ ). Un exemple de carte de temps de réionisation normalisée et lissée par un filtrage Gaussien de taille  $R_f = 6$  est donnée dans les panneaux supérieurs de la Figure 6.1. Nous choisissons de filtrer ces cartes pour plusieurs raisons. On sait déjà que la résolution des images produites par SKA sera moins précise que ce que l'on modélise ici. En effet, elle sera supérieure à  $7 \text{ cMpc}$  pour des redshifts supérieurs à 6.5. On



**FIGURE 6.1** – Carte de temps de réionisation extraite de la simulation EMMA sur les panneaux supérieurs. Elle est superposée à ses minima (étoiles noires) en haut à gauche et aux patches de réionisation (traits noirs) en haut à droite, dont les bordures forment également le squelette de  $t_{\text{reion}}(\vec{r})$ . En bas à gauche, les isocontours de  $t_{\text{reion}}(\vec{r})$  sont montrés avec les minima. Le champ des temps de réionisation est ici normalisé et lissé par un filtrage Gaussien de taille  $R_f = 6$ . En bas à droite, on voit la norme du gradient de  $t_{\text{reion}}(\vec{r})$  avec les minima.

se rapproche donc de ces résolutions avec le lissage avec  $R_f = 6$ . Par ailleurs, lisser permet aussi de faire converger certaines intégrales sur le spectre de puissance de  $t_{\text{reion}}(\vec{r})$  aux hautes fréquences (voir la forme du spectre de puissance dans l'Équation 3.17, Section 3.3). De plus, comme dans le Chapitre 5, le lissage permet d'éviter les grandes zones à gradients nuls que **DisPerSE** ne prend pas en compte. Ce lissage Gaussien va ainsi avoir l'effet de « Gaussianiser »  $t_{\text{reion}}(\vec{r})$  à différentes échelles, nous permettant ainsi de trouver quelles échelles pourraient être à l'origine de non-Gaussianité dans le champ.

Par ailleurs, nous normalisons les cartes de temps de réionisation de la manière suivante :

$$x(\vec{r}) = \frac{t_{\text{reion}}(\vec{r}) - \overline{t_{\text{reion}}(\vec{r})}}{\sigma_0}, \quad (6.1)$$

avec  $\overline{t_{\text{reion}}(\vec{r})}$  la moyenne du champ, qui vaut ici 790 Myrs.  $\sigma_0$  est l'écart-type du champ et il est défini à partir du spectre de puissance du champ de temps de réionisation de la simulation (voir la Section 6.4). Sa valeur dépend du lissage du champ, et vaut 107 Myrs, 91.6 Myrs et 62.2 Myrs pour  $R_f = 1, 2$  et 6 respectivement. Dans ce Chapitre, on notera ainsi  $t_{\text{reion}}(\vec{r})$  le champ de temps de réionisation et  $x$  ce champ normalisé, et par la suite, on se réfère aux valeurs de ce champ normalisé de la manière suivante :

$$\nu = x(\vec{r}). \quad (6.2)$$

Quand  $\nu < 0$ , on sonde alors les temps inférieurs au temps de réionisation moyen et quand  $\nu > 0$ , on sonde les temps supérieurs au temps de réionisation moyen. Dit autrement, les faibles valeurs de  $\nu$  se réfèrent aux temps précoces de réionisation et les valeurs de  $\nu$  élevées se réfèrent aux temps tardifs de réionisation, en unité de la variance des valeurs du champ.

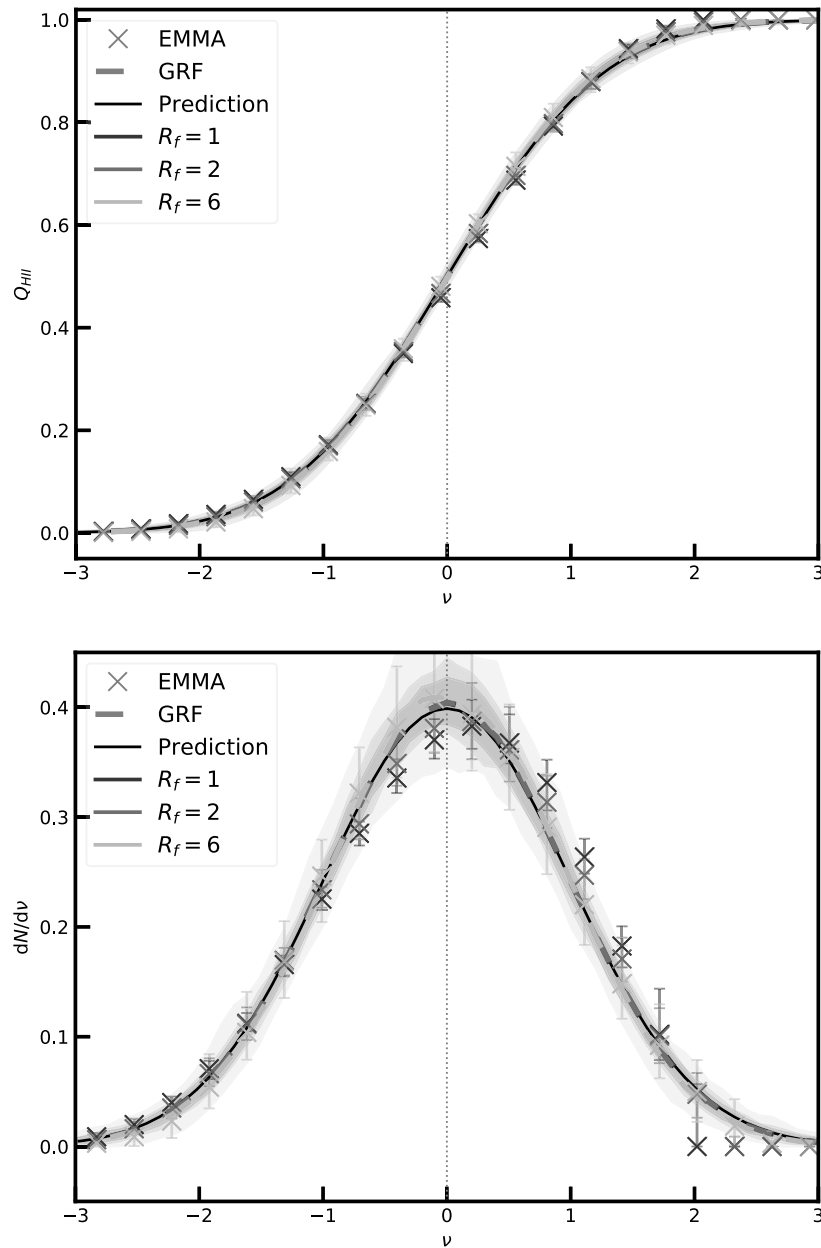
## 6.3 Évolution du processus de réionisation au travers des temps de réionisation

Dans la continuité des travaux menés dans le Chapitre 5, j'approfondis mes analyses topologiques des champs d'instant de réionisation. Diverses mesures statistiques sur le champ 2D des temps de réionisation sont discutées dans cette Section.

### 6.3.1 Histoire de réionisation

L'histoire de réionisation ou la fraction de volume ionisé est une statistique de base pour décrire l'évolution globale de l'état d'ionisation de l'hydrogène pendant la Réionisation. Le champ de temps de réionisation nous donne directement accès à cette statistique en mesurant son taux de remplissage (c'est-à-dire en comptant le nombre de cellules avec un temps de réionisation inférieur à un temps seuil) :

$$Q_{\text{HII}}(\nu) = \frac{N(< \nu)}{N_{\text{total}}}, \quad (6.3)$$



**FIGURE 6.2** – Fractions de volume ionisé dans le panneau supérieur et PDFs du champ de temps de réionisation dans le panneau inférieur. Ces courbes représentent la médiane de toutes les réalisations des champs pour différents lissages (avec différentes valeurs de  $R_f$ , en couleur). Les croix montrent les mesures faites dans la simulation EMMA, les tirets montrent les mêmes mesures faites dans les GRFs correspondants, et les traits noirs sont les prédictions de la théorie des GRFs.  $\nu$  représente les temps de réionisation normalisés. On voit ici que les mesures dans la simulation sont proches des prédictions Gaussiennes, malgré une légère asymétrie apparaissant à la fin de la Réionisation.

avec  $N(< \nu)$  le nombre de cellules qui ont une valeur du champ inférieure à  $\nu$  et  $N_{\text{total}}$  le nombre total de cellules. Il s'agit ainsi d'une PDF cumulée du champ de temps de réionisation. Le panneau supérieur de la Figure 6.2 montre les histoires de réionisation extraites des cartes  $t_{\text{reion}}(\vec{r})$  (avec les croix) lissées avec les trois filtres Gaussiens de largeurs différentes. Ces mesures sont faites sur des cartes provenant d'une seule et même simulation dont l'histoire de réionisation est unique. Cela nous paraît rassurant de voir qu'elles sont identiques pour tous les lissages car le processus de réionisation est ainsi globalement conservé pour tous les lissages. On voit de plus qu'elles semblent symétriques, signifiant que la Réionisation commence à la même vitesse qu'elle se termine. Sur cette distribution cumulée, il ne semble en particulier pas y avoir de début lent avec une fin accélérée ou inversement.

Cependant, le taux de remplissage mentionné ci-avant est une distribution cumulée qui peut alors cacher des comportements fins. On peut aussi mesurer la PDF du champ de temps de réionisation, qui nous informe ici sur la quantité de cellules réionisées à un temps donné. Cette PDF est montrée sur le panneau du bas de la Figure 6.2 avec les croix pour la simulation EMMA avec trois lissages différents. On voit de nouveau que les PDFs avec les différents lissages sont quasiment superposées, avec une légère asymétrie en fin de Réionisation qui apparaît pour le plus petit lissage. Cette asymétrie montre le pic de la PDF décalé vers des temps supérieurs au temps de réionisation moyen. Les petites échelles présentes dans la carte la moins lissée induisent ainsi une réionisation qui démarre lentement pour se terminer légèrement plus brutalement que dans les cartes les plus lissées.

### 6.3.2 Vitesse des fronts d'ionisation

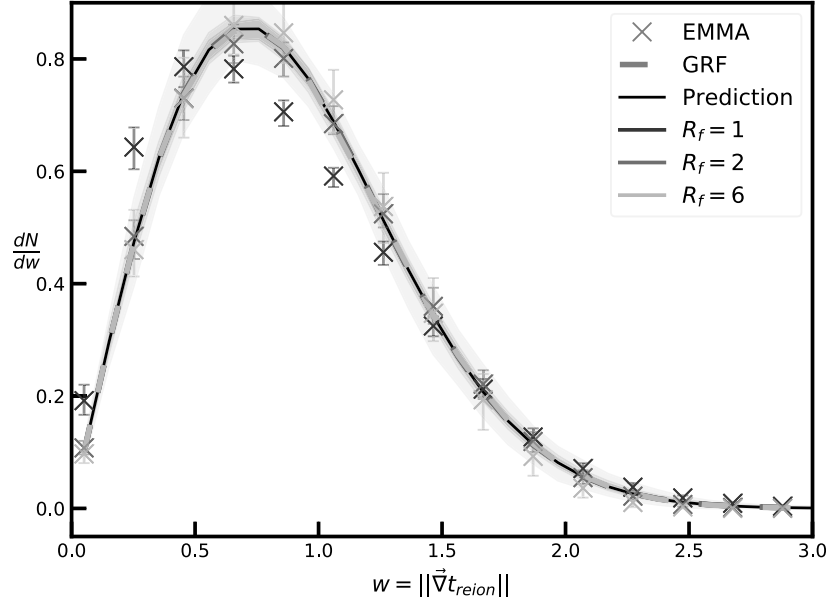
Maintenant que l'on peut estimer quand l'hydrogène se fait globalement réioniser, la question qui se pose ensuite est de savoir à quelle vitesse les fronts d'ionisation se propagent. [Deparis et al. \(2019\)](#) nous fournit une réponse en étudiant la vitesse des fronts d'ionisation. Ils calculent des cartes de vitesse de fronts d'ionisation à partir de cartes de redshift de réionisation. En effet, le gradient spatial de  $t_{\text{reion}}(\vec{r})$  nous permet de savoir le temps nécessaire pour ioniser le gaz sur une certaine distance. L'inverse de ce gradient représente donc une vitesse, qui est interprétée par [Deparis et al. \(2019\)](#) comme étant la vitesse du front d'ionisation qui traverse une cellule. Je peux calculer un gradient similaire dans mes cartes de temps de réionisation, que je définis comme suit :

$$\|\vec{\nabla}x\| = \frac{1}{\sigma_1} \sqrt{(\nabla_1 F)^2 + (\nabla_2 F)^2} = R_0 \sqrt{(\nabla_1 x)^2 + (\nabla_2 x)^2}, \quad (6.4)$$

où  $F$  représente le champ de temps de réionisation et  $x$ , le champ de temps de réionisation normalisé. Nous rappelons que  $\sigma_1$  est l'écart-type de la norme du gradient du champ  $F$  et  $R_0$  est une longueur caractéristique (distance typique entre deux zéros) définie dans l'équation 3.22 dans la Section 3.3.  $\nabla_i x$  pour  $i \in \{1, 2\}$  sont les deux composantes du gradient spatial 2D de  $t_{\text{reion}}(\vec{r})$ , et elles sont calculées par le biais de transformées de Fourier du champ :

$$\nabla_i x = \mathcal{F}^{-1}[\tilde{x} \times i k_i], \quad (6.5)$$

avec  $\tilde{x} = \mathcal{F}[x]$  la transformée de Fourier de  $x$ , et  $k^2 = k_1^2 + k_2^2$ .  $\|\vec{\nabla}x\|$  représente donc l'inverse de la vitesse des fronts d'ionisation. Comme l'étude montrée ici est faite exclusivement en 2D,



**FIGURE 6.3** – Fonctions de probabilité de distribution de la norme du gradient de  $t_{\text{reion}}(\vec{r})$ . Ces courbes représentent la médiane de toutes les réalisations des champs pour différents lissages (avec différentes valeurs de  $R_f$ , en couleur). Les croix montrent les mesures faites dans la simulation EMMA, les tirets montrent les mêmes mesures faites dans les GRFs correspondants, et les traits noirs sont les prédictions de la théorie des GRFs. On voit ici que les mesures dans les cartes lissées avec les deux plus grands lissages sont proches des prédictions Gaussiennes.

il faut noter que nous perdons ainsi la composante du gradient de la troisième direction et que la vitesse des fronts d’ionisation est probablement sous-estimée. La Figure 6.1 montre en bas à droite une carte de norme de gradient de  $t_{\text{reion}}(\vec{r})$  (avec  $t_{\text{reion}}(\vec{r})$  lissé avec un filtrage Gaussien de taille  $R_f = 6$ ). Les minima de  $t_{\text{reion}}(\vec{r})$ , autrement appelés les sources de réionisation, sont également représentés sur la carte par les étoiles noires. Au départ de ces sources de réionisation, on voit ainsi les fronts d’ionisation accélérer et former de plus grands fronts, parfois rectilignes, qui vont ioniser le gaz à plus grande distance. Ceci rappelle les résultats obtenus dans le Chapitre 5 avec des sources perlées le long des filaments dans des patchs de réionisation allongés et alignés aux filaments (les patchs étant le résultat de l’ensemble des fronts d’ionisation qui se propagent des sources localisées sur les filaments).

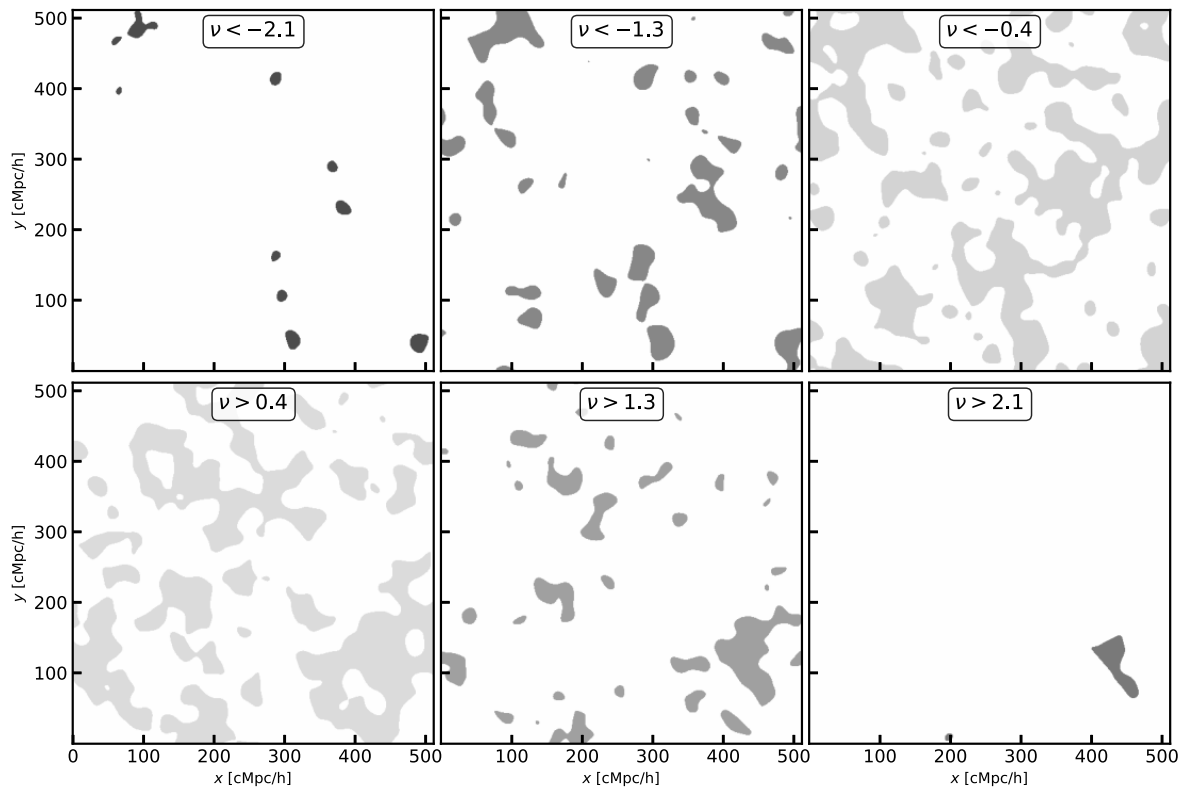
Une manière de quantifier la vitesse des fronts d’ionisation à tout temps de la Réionisation est de mesurer la PDF du champ de la norme de gradient de  $t_{\text{reion}}(\vec{r})$ . Elle est montrée dans la Figure 6.3 avec les croix pour la simulation EMMA et pour plusieurs lissages. On voit en particulier qu’avec le plus petit lissage, la distribution est décalée vers les gradients plus petits et donc les vitesses de fronts d’ionisation plus grande. On retrouve le même effet que [Deparis et al. \(2019\)](#) mentionnent, à savoir des fronts d’ionisations qui accélèrent en fin de réionisation. Cet effet va être expliqué plus en détail dans la Section 6.4.

### 6.3.3 Évolution de la taille des bulles ionisées et régions neutres

L'histoire de réionisation et la vitesse des fronts nous permettent de savoir quand et à quelle vitesse respectivement la Réionisation se propage. La prochaine étape est de savoir comment elle se produit : dans quelle direction les fronts d'ionisation se dirigent ? Comment sont créées et évoluent les bulles ionisées et les régions neutres au cours du temps ? Comment percolent les bulles ionisées ? Dans la littérature, ce type d'étude se fait notamment par l'analyse des bulles ionisées et régions neutres. Les distributions de taille des bulles ionisées ou régions neutres sont souvent utilisées pour les étudier. Giri et al. (2018b, 2019) expliquent par exemple que ces distributions nous permettent de contraindre les sources de réionisation (ce que sont ces objets, leur émissivité, etc.). Shimabukuro et al. (2022) se servent de spectres de puissance simulés du signal 21 cm et d'un réseau de neurones pour reconstruire ces distributions de tailles. Ils expliquent que leurs reconstructions dépendent du modèle de Réionisation utilisé mais que ce sont des outils pour comprendre la morphologie et l'évolution des bulles ionisées.

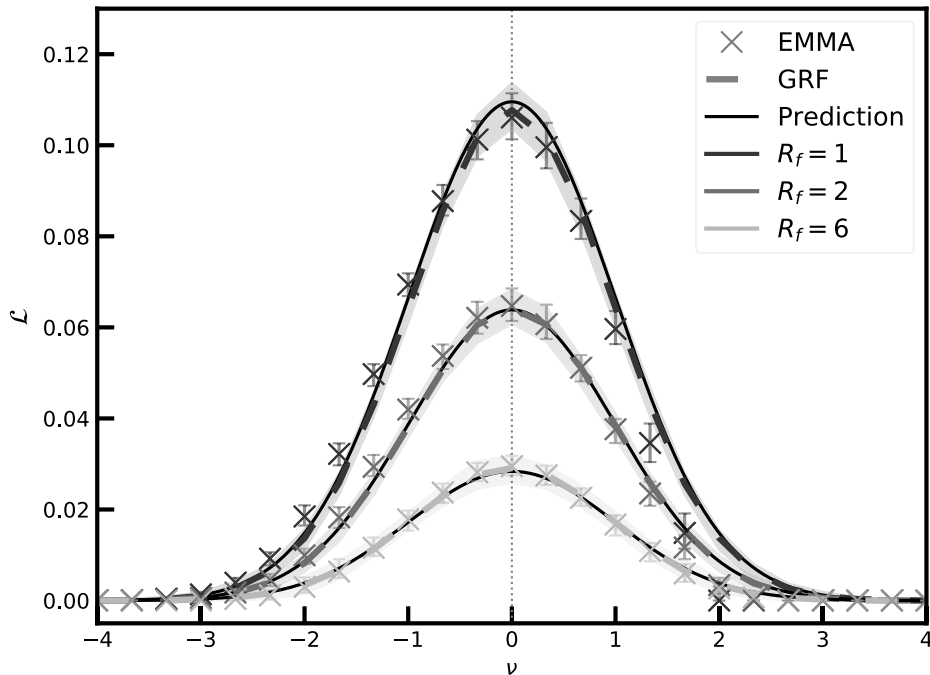
Nous proposons ici une manière alternative de suivre la croissance de ces bulles en extrayant les isocontours du champ de temps de réionisation. Ces contours sont montrés par exemple dans le panneau en bas à droite de la Figure 6.1 accompagnés des sources de réionisation (minima) de  $t_{\text{reion}}(\vec{r})$ , qui sont identifiées par les étoiles noires. Ils correspondent aux endroits de même temps de réionisation. La Figure 6.4 montre des cartes de bulles ionisées (première ligne) ou régions neutres (deuxième ligne) extraites d'une carte  $t_{\text{reion}}(\vec{r})$  de la simulation EMMA lissée par le filtre Gaussien d'écart-type  $R_f = 6$ . Les isocontours bleus foncés de  $t_{\text{reion}}(\vec{r})$  sont ceux qui ont des petits temps de réionisation et ils apparaissent donc autour des minima de  $t_{\text{reion}}(\vec{r})$ , c'est-à-dire des sources de premières réionisations. Ils contiennent alors toutes les régions qui ont des temps de réionisation plus petits et donc qui ont déjà réionisé : ils entourent ainsi les premières bulles ionisées montrées sur la carte de gauche de la Figure 6.4. Ces bulles (et donc les isocontours) grandissent et fusionnent, comme on peut le voir avec la seconde carte de bulles ionisées représentées en bleu un peu plus clair dans la Figure 6.4. On voit même qu'elles commencent à percoler sur la carte en haut à droite de la Figure 6.4. Après le temps de réionisation moyen ( $\nu = 0$ ), les isocontours rouges rétrécissent jusqu'à disparaître à mesure que le temps passe et qu'ils deviennent plus foncés. Dans cette phase, on suit en fait les dernières régions neutres qui se font toutes peu à peu réioniser. Ces régions neutres correspondent à l'ensemble des cellules dont le gaz n'est pas encore réionisé et qui sont à l'intérieur des isocontours rouges. Sur les cartes en bas de la Figure 6.4, ces régions contiennent ainsi les cellules qui ont un temps  $\nu$  supérieur à la valeur de l'isocontour correspondant. On peut voir que ces régions neutres en rouge rétrécissent puis disparaissent.

Une manière de quantifier la croissance des bulles avec les isocontours est de mesurer l'évolution de la longueur  $\mathcal{L}$  des isocontours en fonction du temps. Cette longueur est montrée dans la Figure 6.5 avec les croix pour la simulation EMMA et pour différents lissages (différentes valeurs de la taille  $R_f$  du filtre Gaussien). C'est une longueur moyenne normalisée par rapport à la surface de la carte. On peut y suivre les bulles ionisées qui grandissent quand  $\nu < 0$  avec  $\mathcal{L}$  qui augmente. Au temps de réionisation moyen  $\nu = 0$ , on imagine que les fibres ionisées sont entremêlées au fibres neutres, comme le décrit (Chen et al. 2019), donnant une longueur d'isocontours maximale, avant qu'elle diminue à mesure que les régions neutres réionisent et disparaissent à la fin de la Réionisation ( $\nu > 0$ ). On voit également que quand le lissage



**FIGURE 6.4** – Bulles ionisées (première ligne) et régions neutres (deuxième ligne) extraites d’une carte  $t_{\text{reion}}(\vec{r})$  de la simulation EMMA. Cette carte est ici lissée par un filtre Gaussien de taille  $R_f = 6$ . Les bulles ionisées sur les cartes de la première ligne sont l’ensemble des cellules qui ont un temps normalisé  $\nu$  inférieur au temps précisé sur les cartes. Les régions encore neutres des cartes de la deuxième ligne sont l’ensemble des cellules qui ont un temps normalisé  $\nu$  supérieur à celui précisé sur les cartes.



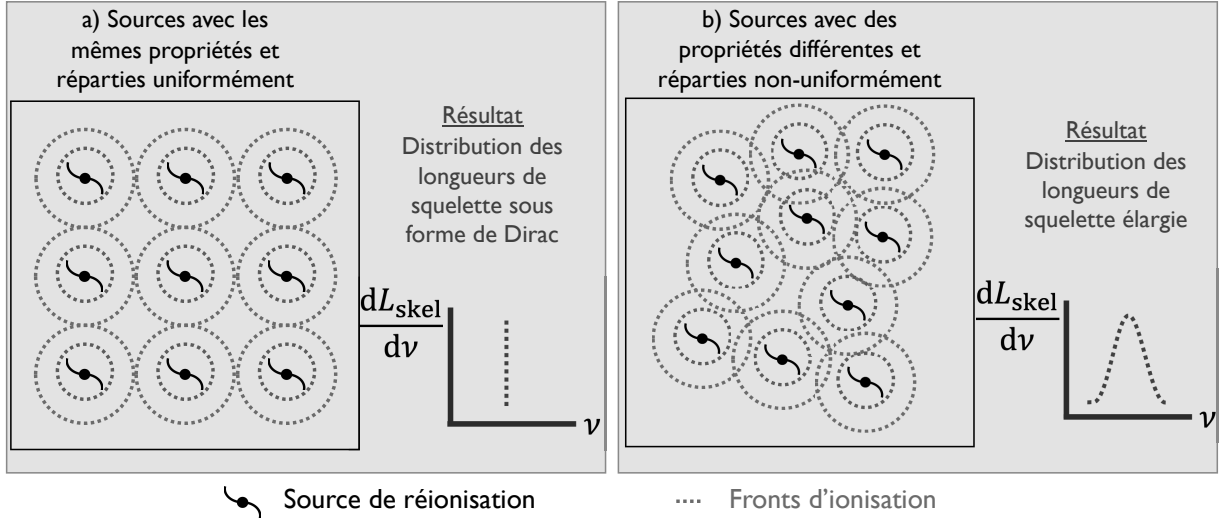


**FIGURE 6.5** – Longueur des isocontours de  $t_{\text{reion}}(\vec{r})$ . Ces courbes représentent la médiane de toutes les réalisations des champs pour différents lissages (avec différentes valeurs de  $R_f$ , en couleur). Les croix montrent les mesures faites dans la simulation EMMA, les tirets montrent les mêmes mesures faites dans les GRFs correspondants, et les traits noirs sont les prédictions de la théorie des GRFs.  $\nu$  représente les temps de réionisation normalisés. On voit ici que les mesures dans la simulation sont proches des prédictions Gaussiennes, malgré une légère asymétrie apparaissant à la fin de la Réionisation pour le plus petit lissage.

est plus fort, les isocontours sont plus petits, ce qui est attendu avec le filtrage des petites structures induites par le lissage.

### 6.3.4 Percolation des fronts d'ionisation

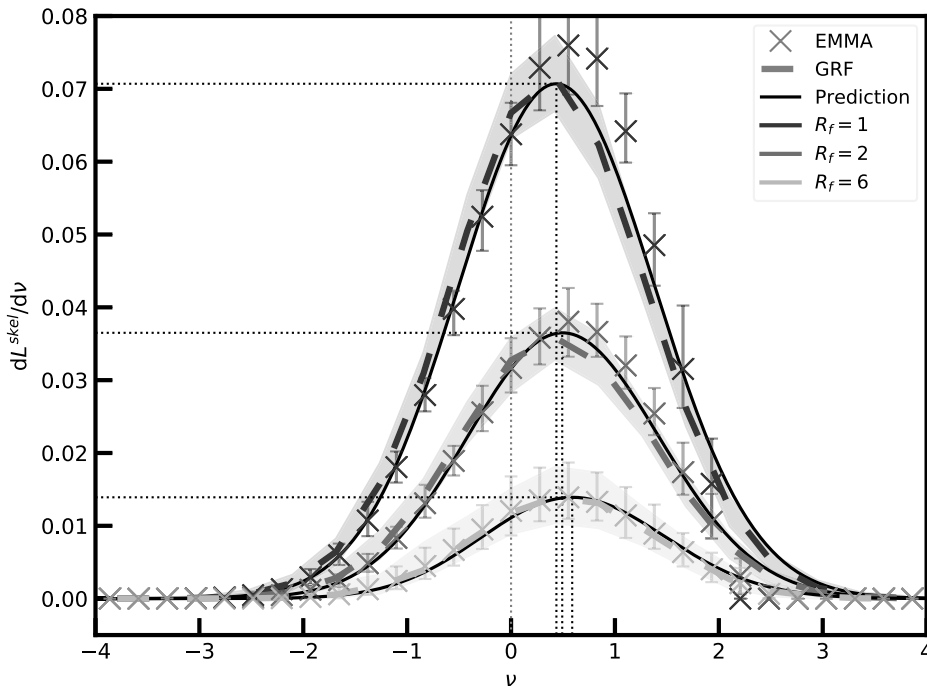
J'ai montré précédemment que l'on pouvait suivre les fronts d'ionisation et la croissance des bulles ionisées et régions neutres. Nous avons également un moyen d'étudier comment, à quel endroit et à quel moment les fronts d'ionisation et les bulles ionisées percolent. Cela peut se faire par le biais du squelette des temps de réionisation. DisPerSE nous permet de l'extraire des cartes de  $t_{\text{reion}}(\vec{r})$  (en prenant une valeur de la persistance de  $10^{-5} - \sigma \ll 1$  pour extraire le squelette dans son ensemble). Le squelette de  $t_{\text{reion}}(\vec{r})$  correspond en fait aux bords des patches de réionisation et donc à l'endroit où se rencontrent les fronts d'ionisation. Un exemple de squelette des temps de réionisation est montré en noir sur le panneau supérieur droit de la Figure 6.1. Sur cette carte, il semblerait que le squelette s'oriente souvent en forme de diamant ou losange. Dans Thélie et al. (2023), nous supposons que cet effet était numérique et dû aux formes en diamant autour des sources, qui sont créées par l'approximation M1 utilisée dans



**FIGURE 6.6** – Schéma illustrant deux cas de distributions des sources de réionisation menant à des distributions de longueurs de squelette différentes. Dans le cas a), les sources (représentées en noir) sont supposées avoir les mêmes propriétés et sont réparties uniformément dans une carte 2D. Leur rayonnement produit des fronts d’ionisation (traits pointillés bleus à bas temps et rouges à grand temps) qui percolent tous au même moment, ce qui produit une distribution des longueurs de squelette sous forme de Dirac. Dans le cas b), les sources n’ont plus nécessairement les mêmes propriétés et ne sont plus réparties uniformément. Les fronts d’ionisation ne percolent plus tous au même moment, produisant une distribution des longueurs de squelette plus large.

EMMA et par la représentation du rayonnement sous forme de fluide (Aubert & Teyssier 2008). Cependant, nous nous sommes depuis rendus compte que cet effet est également présent dans des cartes 21cmFAST : il peut donc être dû au fait que nous travaillons ici en 2D, à notre processus de transformation des cartes ou bien à DisPerSE. Par exemple, dans cette étude, nous avons aussi choisi une très petite valeur de la persistance, qui pourrait être à l’origine de cet effet en créant des connexions artificielles. Nous n’avons pour le moment pas de justification évidente à cet effet et cela mériterait de plus amples recherches.

À partir du squelette de  $t_{\text{reion}}(\vec{r})$ , on peut mesurer la distribution des longueurs de squelette en fonction du temps. Elle nous permet de connaître la durée pendant laquelle les fronts percolent (largeur de la distribution) et le moment auquel ils percolent (position du pic). Par exemple, avec un pic en  $\nu < 0$ , la percolation se produit tôt et avec un pic en  $\nu > 0$ , elle se produit tardivement. Le schéma de la Figure 6.6 illustre deux cas de distributions des sources de réionisation, produisant des distributions de longueurs de squelette différentes. Si cette distribution est un Dirac (voir le cas a) du schéma), alors les sources de réionisation sont distribuées uniformément et les fronts de réionisation rencontrent des fronts opposés tous au même moment. Dans le cas d’une distribution plus large (cas b) du schéma), ce n’est plus le cas et ils percolent pendant toute la Réionisation. La Figure 6.7 montre cette distribution mesurée dans la simulation EMMA avec les croix dans le cas de plusieurs lissages (voir les différentes tailles  $R_f$  du filtre Gaussien en couleurs). On voit que ces distributions ont un pic à un temps supérieur au temps moyen de réionisation ( $\nu > 0$ ) signifiant que la majorité

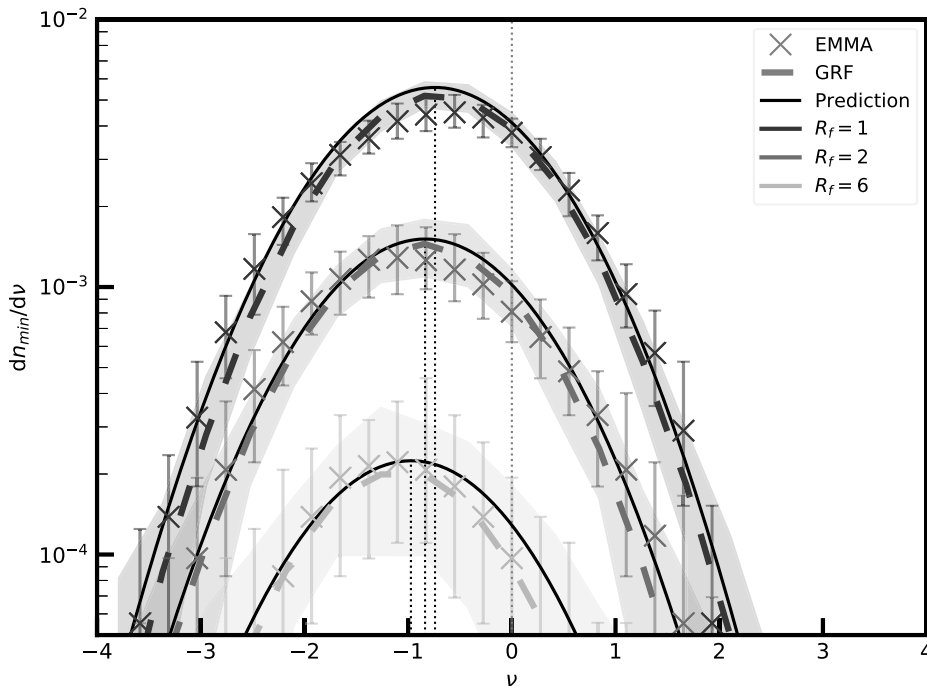


**FIGURE 6.7** – Longueurs de squelette de  $t_{\text{reion}}(\vec{r})$ . Ces courbes représentent la médiane de toutes les réalisations des champs pour différents lissages (avec différentes valeurs de  $R_f$ , en couleur). Les croix montrent les mesures faites dans la simulation EMMA, les tirets montrent les mêmes mesures faites dans les GRFs correspondants, et les traits noirs sont les prédictions de la théorie des GRFs.  $\nu$  représente les temps de réionisation normalisés. On voit ici que les mesures dans la simulation sont proches des prédictions Gaussiennes, malgré une légère asymétrie apparaissant à la fin de la Réionisation pour le plus petit lissage.

des fronts d'ionisation percolent après ce temps moyen et donc que la percolation se produit plutôt proche de la fin de la Réionisation. De plus, la croissance des distributions indique la fusion des patches de réionisation jusqu'à ce qu'ils percolent, et la décroissance indique que la majorité du gaz est réionisé et qu'il reste de moins en moins de régions neutres dans lesquels les fronts d'ionisation percolent. On voit aussi que plus le lissage est fort et moins le squelette est grand, ce qui est en accord avec le fait que les petites structures sont filtrées.

### 6.3.5 Décompte du nombre de sources de réionisation

Un aspect de la Réionisation encore non-abordé dans ce Chapitre sont les objets sources de la réionisation. Quand apparaissent-ils et combien sont nécessaires pour achever la Réionisation? On sait déjà que les minima du champ de temps de réionisation représentent ces sources de réionisation. On peut les extraire de nos cartes grâce à DisPerSE (on choisit la même persistance de  $10^{-5} - \sigma$ ), et ainsi mesurer la distribution du nombre de sources de réionisation en fonction du temps montrée dans la Figure 6.8 avec les croix pour la simulation EMMA en différentes couleurs selon le lissage. On remarque déjà que ces distributions ont un pic



**FIGURE 6.8** – Fonctions de distribution de probabilité des valeurs de  $t_{\text{reion}}(\vec{r})$  prises à ses minima. Ces courbes représentent la médiane de toutes les réalisations des champs pour différents lissages (avec différentes valeurs de  $R_f$ , en couleur). Les croix montrent les mesures faites dans la simulation EMMA, les tirets montrent les mêmes mesures faites dans les GRFs correspondants, et les traits noirs sont les prédictions de la théorie des GRFs.  $\nu$  représente les temps de réionisation normalisés. On voit ici que les mesures dans la simulation sont proches des prédictions Gaussiennes.

à des temps précoces de la réionisation ( $\nu < 0$ ), ce qui est attendu comme on s'intéresse aux minima de  $t_{\text{reion}}(\vec{r})$ . La partie croissante de ces distributions montre les premières sources de réionisation qui apparaissent et commencent à ioniser l'hydrogène. On voit aussi que la majorité des sources de réionisation sont apparues avant le temps moyen de réionisation  $\nu = 0$ . Ensuite, la partie décroissante de ces distributions reflètent le fait que, même si de nouvelles galaxies sont continuellement créées, la majorité d'entre elles sont localisées dans des régions déjà réionisées et ne seront alors pas comptées comme des sources de réionisation.

## 6.4 Gaussianité du champ de temps de réionisation

La Section précédente nous a ainsi permis de présenter des outils versatiles de mesures statistiques à partir d'un même champ. Ces outils peuvent être également prédits à partir de la théorie des GRFs, comme on va en discuter dans cette Section, et nous servir pour analyser également la Gaussianité des temps de réionisation et donc du processus de réionisation.

### 6.4.1 Comment procédons-nous ?

L'Époque de la Réionisation est connu pour être hautement non-Gaussienne, notamment au travers d'étude de la densité, de fractions d'ionisation ou du signal 21 cm (cf. l'introduction de ce Chapitre en Section 6.1 et voir par exemple Iliev et al. (2006b); Mellema et al. (2006); Dixon et al. (2016); Banet et al. (2021)). Dans cette étude, nous proposons d'explorer de nouveau la Gaussianité du processus de réionisation, mais cette fois-ci et pour la première fois au travers du champ de temps de réionisation. La Figure 6.9 est un schéma illustrant notre manière de procéder. Nous utilisons la théorie des champs Gaussiens aléatoires (GRFs) pour prédire analytiquement diverses statistiques d'un champ Gaussien (Rice 1944; Longuet-Higgins 1957; Bardeen et al. 1986; Gay et al. 2012; Cadiou et al. 2020) à partir de la PDF du champ normalisé et de ses dérivées qui proviennent de la connaissance du spectre de puissance (voir la ligne du haut du schéma de la Figure 6.9). Toutes les mesures mentionnées dans la Section précédente peuvent en particulier être prédites par cette théorie. Les expressions analytiques de ces prédictions et la manière de les obtenir sont détaillées dans la Section 3.3. Le but de cette Section est uniquement de comparer ces prédictions Gaussiennes aux mesures faites dans le champ de temps de réionisation extrait de la simulation EMMA, afin de tester sa Gaussianité (voir la ligne du milieu du schéma de la Figure 6.9). Pour un champ Gaussien (dont la définition est donnée dans la Section 3.3), nous pouvons ainsi prédire la PDF du champ (voir la Figure 6.2) et de la norme du gradient du champ (Figure 6.3), la longueur des isocontours en fonction de la valeur du champ (Figure 6.5), la PDF de la longueur de squelette (Figure 6.7)<sup>1</sup>, ainsi que la PDF des valeurs aux minima du champ (Figure 6.8).

Nous rappelons ici qu'un champ Gaussien se définit uniquement par son spectre de puissance et que si on le connaît, on détient alors toute l'information présente dans le champ. En extrayant le spectre de puissance du champ  $t_{\text{reion}}(\vec{r})$  de la simulation EMMA, qui est montré en trait plein bleu dans la Figure 6.10, on peut ainsi générer des GRFs et y mesurer les statistiques mentionnées ci-avant (voir les tirets colorés dans les Figures référencées précédemment). Ces champs Gaussiens nous servent ici à vérifier que les mesures statistiques présentées dans la Section précédente sont faites correctement (comme expliqué par la ligne du bas du schéma de la Figure 6.9).

Par ailleurs, les propriétés topologiques de ces GRFs seront alors les mêmes que celles de  $t_{\text{reion}}(\vec{r})$  si ce dernier est un champ Gaussien car son spectre de puissance contiendrait toute son information. Dans nos modèles, le spectre de puissance de  $t_{\text{reion}}(\vec{r})$  peut se définir par une double loi de puissance (autrement dit il a deux pentes dans l'espace logarithmique). Nous pouvons alors le paramétrer de la manière suivante :

$$\mathcal{P}_k = \begin{cases} A_1 k^{n_1} & \text{si } k \leq k_{\text{thresh}} \\ A_2 k^{n_2} & \text{si } k > k_{\text{thresh}} \end{cases} . \quad (6.6)$$

$A_1$  et  $A_2$  sont les amplitudes de chacune des lois de puissances et  $n_1$  et  $n_2$  leur puissance.  $k_{\text{thresh}}$  est le seuil qui sépare les deux parties du spectre de puissance et correspond donc

1. Il faut noter que DisPerSE nous donne un squelette global alors que le calcul de la théorie des GRFs repose sur l'approximation d'un squelette raide, qui a tendance à sous-estimer la longueur de squelette en ratant des filaments. Les mesures et la prédiction comportent ainsi des différences d'amplitude et il convient de multiplier les mesures par un facteur de normalisation (rapport entre les longueurs totales de squelettes mesurés et prédites, voir Gay et al. (2012)).

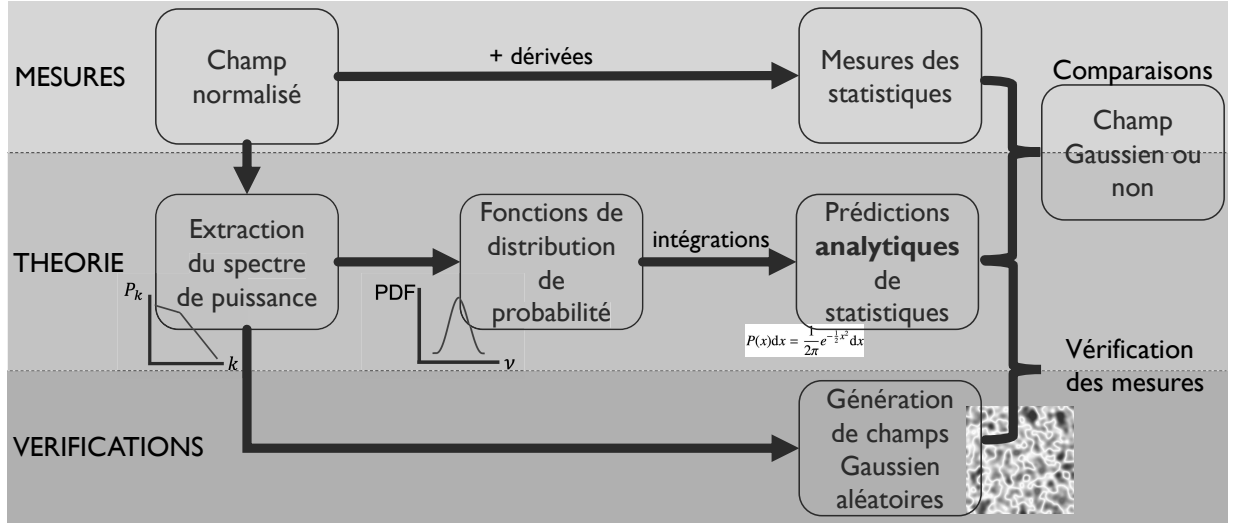


FIGURE 6.9 – Schéma expliquant le procédé de notre étude de Gaussianité.

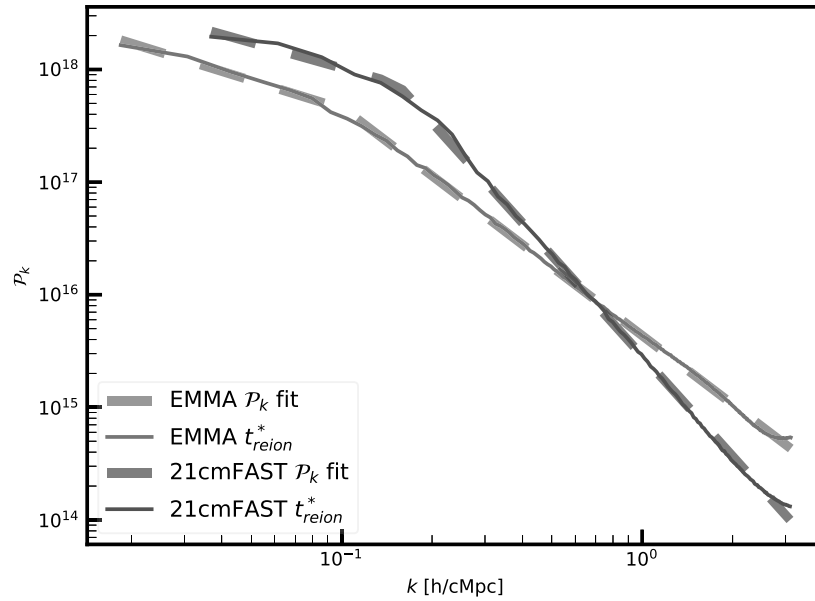


FIGURE 6.10 – Spectres de puissance des champ de temps de réionisation extraits de la simulation EMMA en trait plein bleu et de la simulation 21cmFAST en trait plein marron. Les pointillés bleus et marron représentent l’ajustement fait avec la forme analytique donnée dans l’Équation 6.6 des spectres de puissances des cartes EMMA et 21cmFAST respectivement. Pour la carte  $t_{\text{reion}}(\vec{r})$  d’EMMA, le jeu de paramètres ajustés est le suivant :  $\{A_1, n_1, A_2, n_2, k_{\text{thresh}}\} = \{6.70 \times 10^{16}, -0.83, 4.34 \times 10^{15}, -2.03, 0.10\}$ . Pour la carte  $t_{\text{reion}}(\vec{r})$  de 21cmFAST le jeu de paramètres ajustés est le suivant :  $\{A_1, n_1, A_2, n_2, k_{\text{thresh}}\} = \{1.83 \times 10^{17}, -0.75, 2.87 \times 10^{15}, -2.96, 0.15\}$ . La simulation 21cmFAST a légèrement plus de grandes structures et moins de petites structures que la simulation EMMA.

à une longueur caractéristique.<sup>2</sup> En ajustant le spectre de puissance de  $t_{\text{reion}}(\vec{r})$  avec cette expression, nous obtenons la forme montrée en pointillés bleus dans la Figure 6.10 avec le jeu de paramètres  $\{A_1, n_1, A_2, n_2, k_{\text{thresh}}\} = \{6.70 \times 10^{16}, -0.83, 4.34 \times 10^{15}, -2.03, 0.10\}$ . La longueur caractéristique  $k_{\text{thresh}}$  correspond à un longueur de l'ordre d'une dizaine de cMpc/h pour les temps de réionisation de la simulation EMMA. Cette expression du spectre de puissance est celle que nous avons originellement utilisé pour cette étude dans Thélie et al. (2023) et que nous utilisons aussi dans ce manuscrit (sauf contre-indication). D'autre part, nous lisons aussi les champs de temps de réionisation, et ce lissage peut être directement incorporé dans le spectre de puissance défini originellement :

$$\mathcal{P}_k^{\text{smoothed}} = \begin{cases} A_1 k^{n_1} e^{-2R_f^2 k^2} / 2\pi & \text{si } k \leq k_{\text{thresh}} \\ A_2 k^{n_2} e^{-2R_f^2 k^2} / 2\pi & \text{si } k > k_{\text{thresh}} \end{cases}. \quad (6.7)$$

On retrouve dans cette expression la taille  $R_f$  du filtre Gaussien appliqué au champ.

Enfin, un spectre de puissance d'un champ nous donne toujours accès aux moments de ce champ grâce à l'expression suivante :

$$\sigma_i^2 = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \int_0^\infty k^{2i} \mathcal{P}_k k^{d-1} dk, \quad (6.8)$$

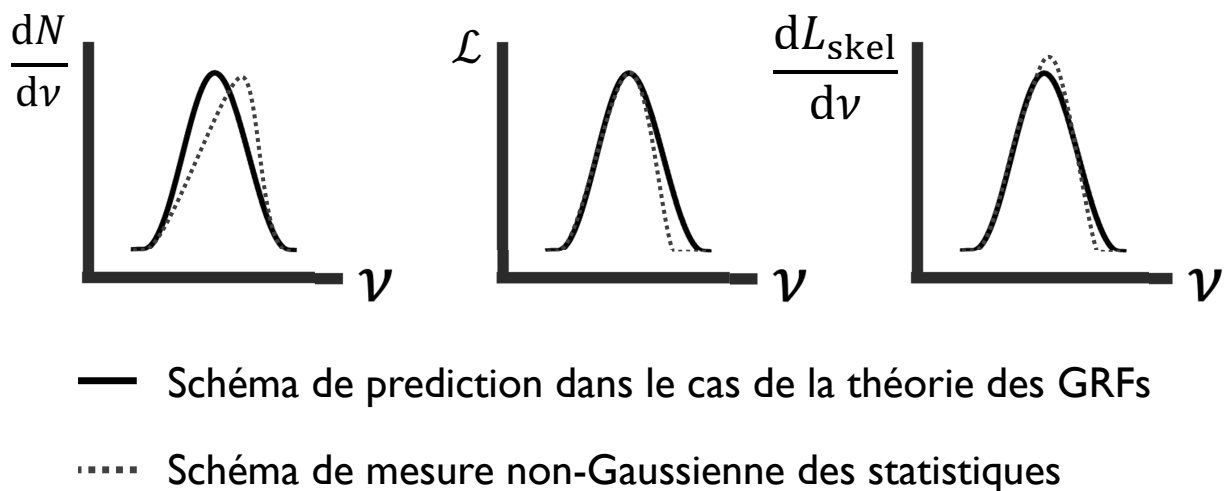
où  $i \in \mathbb{N}$  est le nombre de fois que le champ est dérivé et  $d$  la dimension du champ.  $\sigma_0$  est l'écart-type du champ, et  $\sigma_1$  est l'écart-type de la première dérivée du champ. Nous y avons ainsi accès avec la seule connaissance du spectre de puissance de  $t_{\text{reion}}(\vec{r})$ . Tout ceci est détaillé plus amplement dans la Section 3.3.

## 6.4.2 Étude de la gaussianité des temps de réionisation

Maintenant que nous disposons des prédictions de la théorie des GRFs (dont les expressions sont données dans le Chapitre 3) pour toutes les mesures mentionnées dans ce Chapitre, nous nous interrogeons sur la Gaussianité de  $t_{\text{reion}}(\vec{r})$  au travers de ces mesures. On compare donc ici ces mesures aux prédictions Gaussiennes. Pour rappel, la PDF du champ est montrée dans la Figure 6.2, la norme du gradient du champ dans la Figure 6.3, la longueur des isocontours en fonction de la valeur du champ dans la Figure 6.5, la PDF de la longueur de squelette dans la Figure 6.7, et la PDF des valeurs aux minima du champ dans la Figure 6.8. Dans ces Figures, les prédictions Gaussiennes sont les courbes en trait plein noir et les mesures EMMA sont montrées par les croix colorées pour les différents lissages. Avec ces cinq mesures, on voit que le champ de temps de réionisation est très proche d'un champ Gaussien. Cela implique que la simple connaissance du spectre de puissance de  $t_{\text{reion}}(\vec{r})$  nous donne accès à un large éventail de mesures statistiques de la topologie du processus de réionisation, qui nous informent sur tout le déroulé de la Réionisation (cf. la physique que l'on peut extraire de ces mesures mentionnée dans la Section précédente).

Cependant, nous détectons aussi des empreintes de non-Gaussianité dans ces mesures avec l'asymétrie présente dans la PDF des temps de réionisation, que l'on retrouve aussi dans les

2. Cette forme fonctionnelle du spectre de puissance peut en fait ne dépendre que de quatre paramètres. Cette dépendance est décrite plus loin dans l'Équation 6.10. Toute l'étude présentée dans ce Chapitre et extraite de Thélie et al. (2023) utilise cependant l'expression à cinq paramètres.

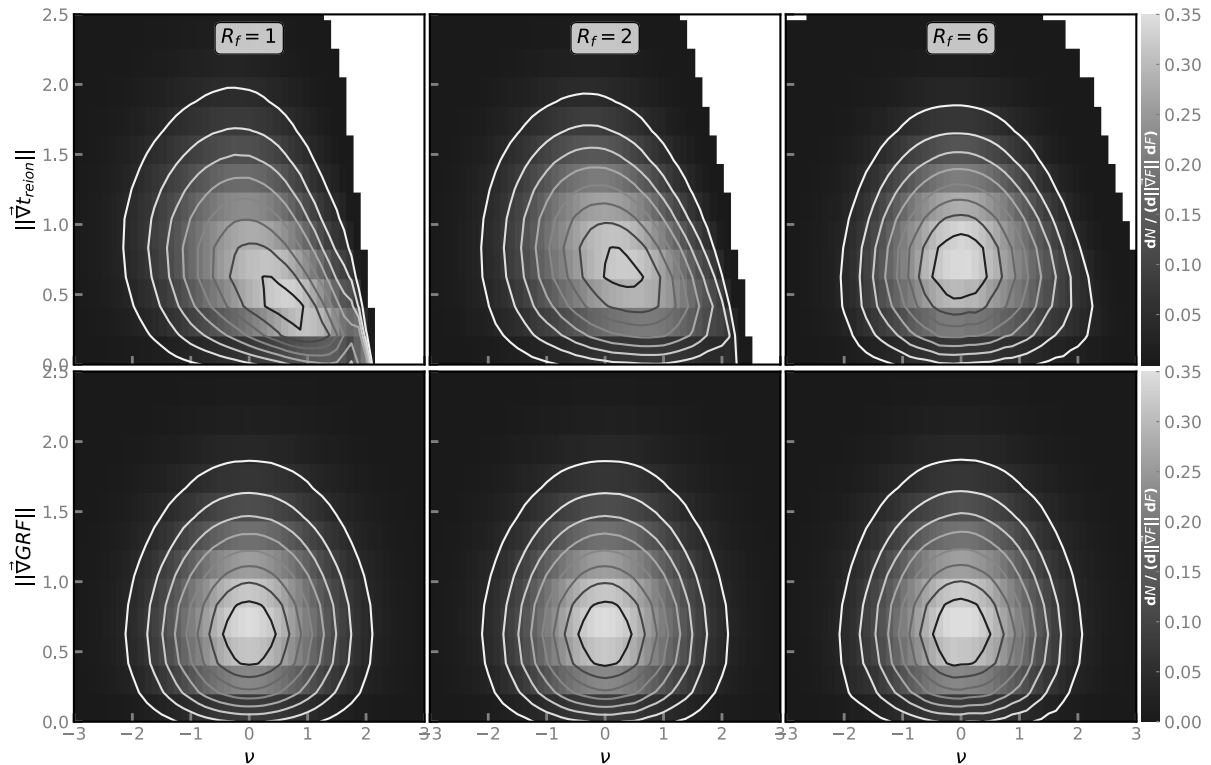


**FIGURE 6.11** – Schémas illustrant les non-Gaussianités présentes dans nos mesures (surtout au lissage avec  $R_f = 1$ ). Les traits pleins noirs représentent les prédictions Gaussiennes de la théorie des GRFs et les traits en pointillés rose schématisent les mesures des différentes statistiques. À gauche, la fraction de volume ionisé (ou l’histoire de réionisation) est représentée. Au centre, la longueur des isocontours est représentée. À droite, la longueur de squelette est représentée. On voit donc de nouveau que les mesures sont quasiment Gaussiennes, mais présentant tout de mêmes les asymétries schématisées ici et qui sont des empreintes de non-Gaussianité.

longueur des isocontours et du squelette. Ces trois types d’asymétries sont schématisées dans la Figure 6.11 de manière simplifiée. Elles sont surtout présentes dans les mesures avec le plus petit lissage ( $R_f = 1$ ), indiquant qu’elles sont dues en majorité aux structures des petites échelles. Elles montrent ainsi que le processus de réionisation de ces petites échelles ne se fait pas de manière symétrique et commence lentement avant de s’accélérer en fin de Réionisation (voir la PDF de  $t_{\text{reion}}(\vec{r})$  ou son schéma dans la Figure 6.11 avec un pic décalé vers les temps tardifs de réionisation pour les plus petits  $R_f$ ). Cela influe sur la réionisation des dernières régions neutres en fin de Réionisation, qui se fait de manière plus abrupte, et cela se répercute alors sur les longueurs d’isocontours (voir la chute des longueurs d’isocontours aux temps tardifs de réionisation pour  $R_f = 1$  schématisé dans la Figure 6.11). Comme la Réionisation commence plus lentement que pour un processus Gaussien, les bulles ionisées mettent plus de temps à fusionner et percolent aussi plus rapidement. La longueur du squelette de  $t_{\text{reion}}(\vec{r})$ , schématisé dans la Figure 6.11, a donc aussi pic décalé vers les temps tardifs pour les plus petits lissages (elle croît plus que pour un champ Gaussien et décroît ensuite plus abruptement).

Par ailleurs, en étudiant la vitesse des fronts d’ionisation en fonction du temps de réionisation, on peut fournir une raison probable de cette asymétrie du processus de réionisation. La Figure 6.12 montre des distributions 2D de la norme du gradient de  $t_{\text{reion}}(\vec{r})$  (qui est proportionnelle à l’inverse de la vitesse des fronts d’ionisation) en fonction du champ  $t_{\text{reion}}(\vec{r})$  normalisé pour différents lissages (différentes tailles  $R_f$  du filtre Gaussien dans les 3 colonnes). La ligne du bas montre les mesures faites dans les GRFs générés avec le même spectre de





**FIGURE 6.12** – Fonctions de distribution de probabilité 2D de la norme du gradient de  $t_{\text{reion}}(\vec{r})$  en fonction des valeurs  $\nu$  prises par le champ  $t_{\text{reion}}(\vec{r})$  normalisé pour différents lissages (voir les différentes valeurs de  $R_f$  dans chaque colonne). La première ligne montre ces mesures faites dans la simulation EMMA et la deuxième ligne montre celles faites dans les GRFs correspondants. Toutes les réalisations de chaque champ sont additionnées. Les lignes en échelle de gris sont les isocontours de ces histogrammes. On voit que les mesures dans la simulation sont proches des prédictions Gaussiennes, malgré une légère asymétrie apparaissant à la fin de la Réionisation pour les deux plus petits lissages. On comprend ici que l’asymétrie provient de l’accélération des fronts d’ionisation à la fin de la Réionisation.

puissance que les cartes  $t_{\text{reion}}(\vec{r})$ , et on voit qu’elles sont symétriques, ce qui est attendu par définition d’un GRF. En revanche, on retrouve l’asymétrie mentionnée précédemment dans les mesures EMMA et pour les plus petits lissages (première ligne et deux premières colonnes). On voit aussi qu’elle est en fait due à l’accélération des fronts d’ionisation vers la fin de la Réionisation quand les petites échelles sont présentes. *Deparis et al. (2019)* remarquent aussi, qu’à la fin de la Réionisation, les fronts d’ionisation s’accélèrent. Ils montrent que cette accélération est liée à la densité du gaz traversée par les fronts, ceux-ci atteignant en dernier les régions les moins denses, que l’on appelle les vides de matière. Notre étude va dans ce sens, et nous ajoutons même que les fronts d’ionisation percolent dans ces vides. Dans la distribution 2D avec  $R_f = 1$  de la Figure 6.12, on estime que l’accélération des fronts commence à partir de  $\nu = 1.5$  environ. Dans la distribution des longueurs de squelette de la Figure 6.7 avec  $R_f = 1$ , vers  $\nu = 1.5$ , on voit que l’asymétrie de la distribution implique une longueur

du squelette inférieure au cas Gaussien, avec de moins en moins d'endroits de percolation des fronts. On a donc bien un phénomène agissant aux petites échelles, c'est-à-dire dans les dernières petites régions encore neutres.

Comme les empreintes de non-Gaussianité se trouvent surtout dans les cartes  $t_{\text{reion}}(\vec{r})$  peu lissées, on conclue alors que quand  $t_{\text{reion}}(\vec{r})$  est lissé à une résolution d'environ 8 cMpc/h, le champ est quasi-Gaussien. Cette résolution est proche de celle que l'on aura dans les cartes 2D d'observations du signal 21 cm avec SKA. Si l'on est capable de reconstruire le spectre de puissance de  $t_{\text{reion}}(\vec{r})$  à partir de ces observations, on aura alors accès à toutes les statistiques mentionnées dans ce Chapitre.

## 6.5 Comparaisons entre les temps et les redshifts de réionisation

### 6.5.1 Temps et redshifts de réionisation

Pendant ma thèse, j'ai originellement travaillé sur le champ de redshifts de réionisation avec l'étude des patches de réionisation présentée dans le Chapitre 5. Nous nous sommes ensuite rendus compte que le champ de temps de réionisation présentait des caractéristiques plus intéressantes que celui de redshift de réionisation. Cette Section a pour but de donner quelques explications sur le choix de travailler avec les temps de réionisation dans la majorité des études présentées dans ce manuscrit.

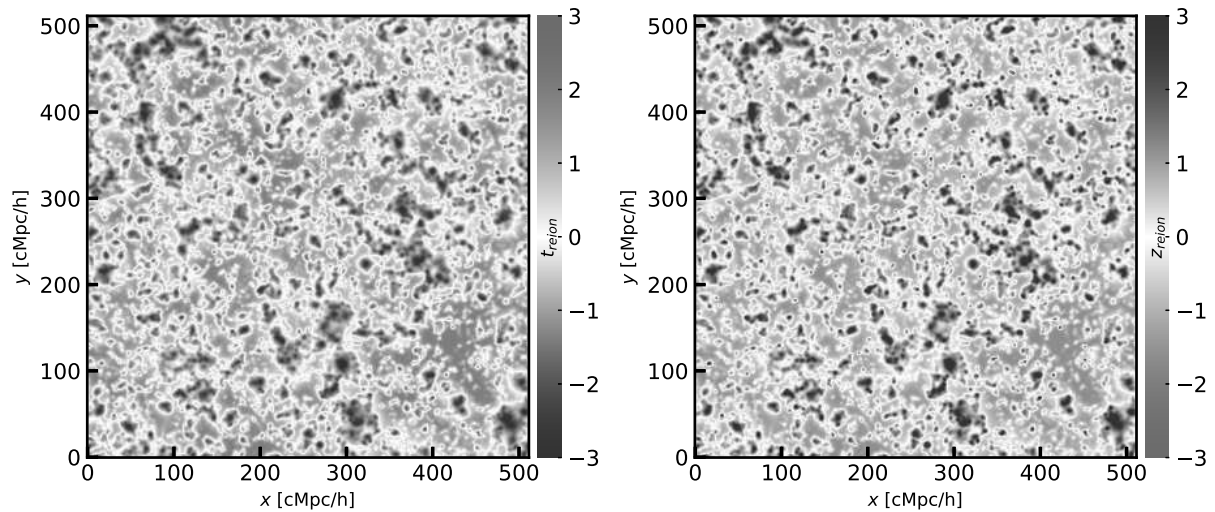
Les champs de temps et de redshift de réionisation ont une monotonie opposée et sont reliés par une relation non-linéaire (voir la Section 2.3.3). Cela provient de la relation entre le temps et le redshift. En effet, le temps croît rapidement quand le redshift décroît, comme le montre la relation suivante :

$$z(t) = \frac{1}{a(t)} - 1 \propto t^{-\frac{2}{3}}, \quad (6.9)$$

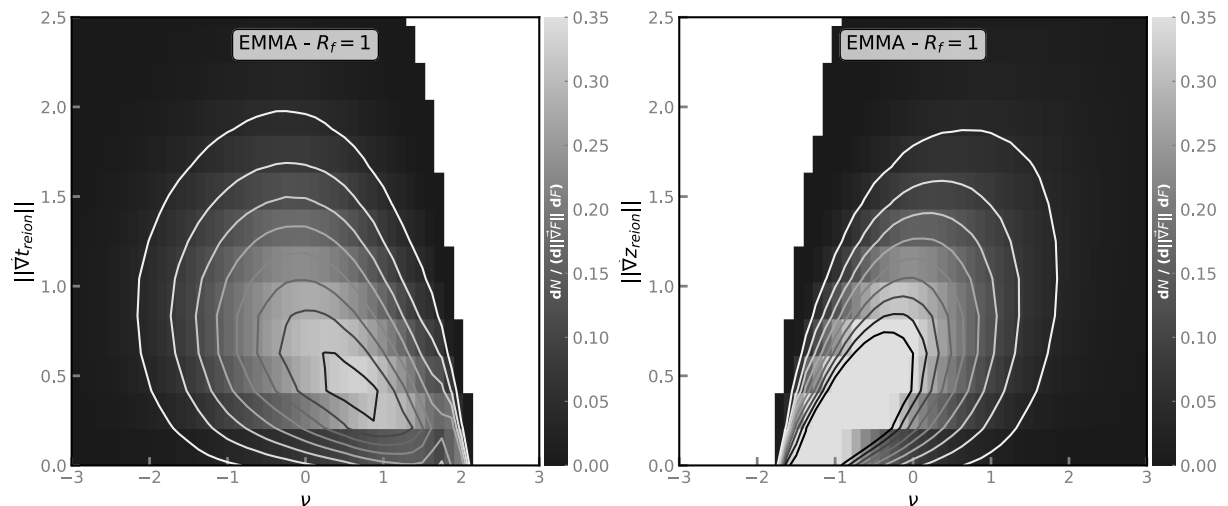
où  $a$  est le facteur d'échelle, qui est proportionnel à  $t^{\frac{2}{3}}$  pendant cette ère. Cette différence de monotonies peut être à l'origine de différences de propriétés entre les deux champs et en particulier pour la Gaussianité. La Figure 6.13 montre une carte de  $t_{\text{reion}}(\vec{r})$  à gauche et une carte de  $z_{\text{reion}}(\vec{r})$  à droite, qui sont extraites de la simulation EMMA et lissées avec un filtre Gaussien de taille  $R_f = 1$ . On peut déjà voir à l'oeil qu'elles présentent des différences : les temps de réionisation ont par exemple des zones rouges plus intenses et des zones bleues moins intenses que les redshifts de réionisation. Au début de la Réionisation par exemple, les temps de réionisation évoluent deux fois plus lentement que les redshifts de réionisation. Cela signifie que les temps de réionisation sont plus résolus que les redshifts au début de la Réionisation.

### 6.5.2 Comparaisons des mesures topologiques

On peut alors se demander dans cette Section si ces différences impactent la topologie et/ou la Gaussianité de ces cartes. Nous avons effectué des transformations similaires sur les cartes 2D de redshifts de réionisation extraites de la simulation EMMA à celles faites sur les



**FIGURE 6.13** – Cartes de temps de réionisation à gauche et de redshift de réionisation à droite. Elles sont extraites de la simulations EMMA et sont lissées avec un filtre Gaussien de taille  $R_f = 1$ . On remarque ici que les cartes ne sont pas exactement les mêmes, avec des valeurs plus intenses en fin de Réionisation notamment pour la carte de temps.



**FIGURE 6.14** – Fonctions de distribution de probabilité 2D de la norme du gradient du champ en fonction des valeurs  $\nu$  prises par le champ. Ces mesures sont faites dans les cartes  $t_{\text{reion}}(\vec{r})$  à gauche et  $z_{\text{reion}}(\vec{r})$  à droite, extraites de la simulation EMMA. Les cartes sont lissées par un filtrage Gaussien de taille  $R_f = 1$ . Toutes les réalisations de chaque champ sont additionnées. Les lignes en échelle de gris sont les isocontours de ces histogrammes. On voit ici que l'asymétrie semble plus marquée pour les redshifts de réionisation que pour les temps.

cartes de  $t_{\text{reion}}(\vec{r})$  (normalisation et lissage), et nous y avons mesuré les mêmes statistiques. La seule différence est qu'il faut alors chercher les maxima de  $z_{\text{reion}}(\vec{r})$  au lieu des minima pour localiser les sources de réionisation, ainsi que l'anti-squelette du champ au lieu du squelette. Un exemple de mesure est montré avec les distributions 2D de la norme du gradient du champ en fonction de la valeur normalisée du champ dans la Figure 6.14. À gauche, la mesure est faite dans les temps de réionisation (donc avec  $\nu = t_{\text{reion}}(\vec{r})$ ), et à droite, elle est faite dans les redshifts de réionisation (donc avec  $\nu = z_{\text{reion}}(\vec{r})$ ). Comme les monotonies sont opposées, les mesures sont renversées sur l'axe  $\nu$  : dans le cas  $t_{\text{reion}}(\vec{r})$  à gauche, la Réionisation commence à gauche (petit temps) et se finit à droite (grand temps), alors que dans le cas de  $z_{\text{reion}}(\vec{r})$ , la Réionisation commence à droite (grand redshift) et se finit à gauche (petit redshift). On voit déjà sur cette mesure qu'il y a une empreinte de non-Gaussianité plus marquée pour  $z_{\text{reion}}(\vec{r})$  que pour  $t_{\text{reion}}(\vec{r})$ , avec une asymétrie plus importante. C'est en fait le cas dans la majorité des mesures (que nous ne montrons pas ici pour rester brefs) : les redshifts de réionisation semblent présenter plus d'empreintes de non-Gaussianité que les temps de réionisation. Notre hypothèse est alors que ces différences proviennent de la relation non-linéaire entre le temps et le redshift, et que celle-ci renforce les non-Gaussianité de  $z_{\text{reion}}(\vec{r})$  par rapport à  $t_{\text{reion}}(\vec{r})$ .

## 6.6 Comparaisons entre simulations cosmologiques et semi-analytiques

Comme dans le Chapitre précédent, nous comparons ici les mesures statistiques de la topologie du processus de réionisation faites dans la simulation cosmologique EMMA à celles faites dans une simulation semi-analytique 21cmFAST. Le but est encore d'avoir des moyens de comparer le comportement de ces deux types de simulations avec des statistiques plus complexes que l'histoire de réionisation ou le spectre de puissance par exemple.

### 6.6.1 Simulation 21cmFAST

La simulation 21cmFAST a une taille de  $256^3 \text{ cMpc}^3/h^3$  et une résolution de  $1 \text{ cMpc}^3/h^3$ . Elle n'a pas la même taille que la simulation EMMA parce que nous ne disposons pas de calculateurs assez puissants pour obtenir une simulation de  $512^3 \text{ cMpc}^3/h^3$ . En revanche, elle a la même résolution. Les seuls paramètres différents de ceux par défaut dans le code 21cmFAST sont la température de viriel<sup>3</sup>  $T_{\text{vir}} = 10^5 \text{ K}$  et l'efficacité des galaxies à émettre des photons  $\zeta = 40$ . Ces paramètres ont été choisis de sorte à faire correspondre les histoires de réionisation des deux simulations. On extrait alors 51 cartes 2D de temps de réionisation de cette simulation 21cmFAST, qui ont une taille de  $256^3 \text{ cMpc}^2/h^2$ , une résolution de  $1 \text{ cMpc}^2/h^2$  sur le plan du ciel et de  $1 \text{ cMpc}/h$  dans l'axe des fréquences. Les mêmes transformations que pour les cartes EMMA sont appliquées à ces cartes (normalisation et lissage). Un exemple de carte 21cmFAST est montrée à droite de la Figure 6.16 à côté d'une carte EMMA à gauche. Ces

3. Nous rappelons que  $T_{\text{vir}}$  est relié à la masse minimale des halos émetteurs de photons ionisants par  $M_{\text{min}} \propto T_{\text{vir}}^{3/2}$ . Greig & Mesinger (2015) indique que la température de viriel  $T_{\text{vir}} \sim 10^4 \text{ K}$  est généralement celle qui est choisie comme température minimale pour qu'un refroidissement atomique efficace commence. Cette température est équivalente à une masse minimale de halos émetteurs en photons ionisants de  $M_{\text{min}} \sim 10^8 M_{\odot}$  à un redshift d'environ 10.

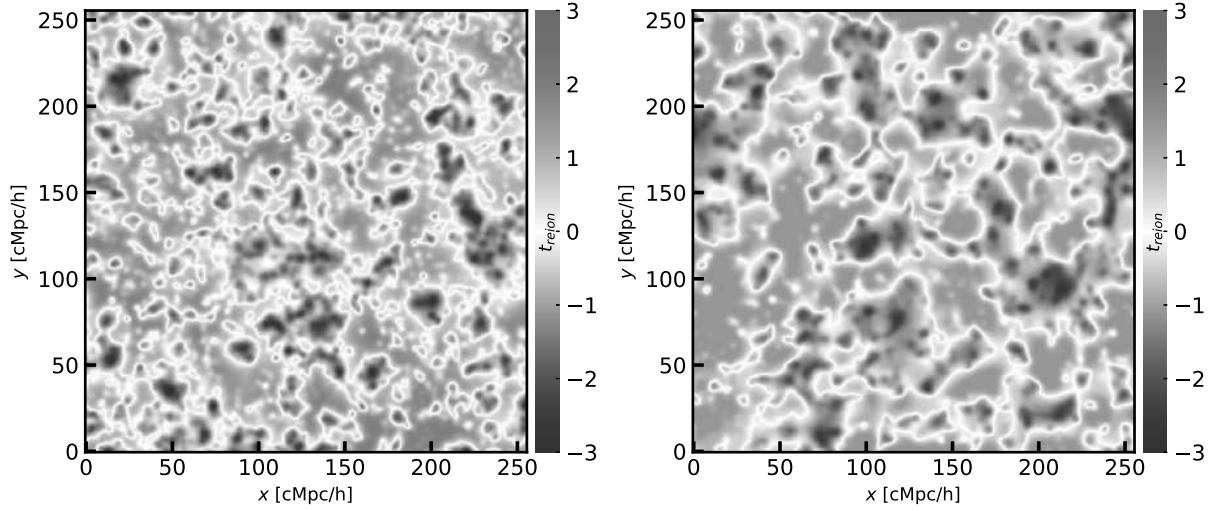
cartes sont lissées avec un filtre Gaussien de taille  $R_f = 1$ . On voit que la carte 21cmFAST a plus de grandes structures que la carte EMMA, ce qui correspond à ce que l'on voit dans leur spectre de puissance montrés dans la Figure 6.10. Je mesure dans les cartes 21cmFAST de la même manière les différentes statistiques, et j'en extrais le spectre de puissance afin de comparer ces mesures aux prédictions de la théorie des GRFs. Il faut remarquer que les mesures et prédictions prennent en compte la taille des cartes et en sont donc indépendantes.

### 6.6.2 Comparaisons des mesures topologiques

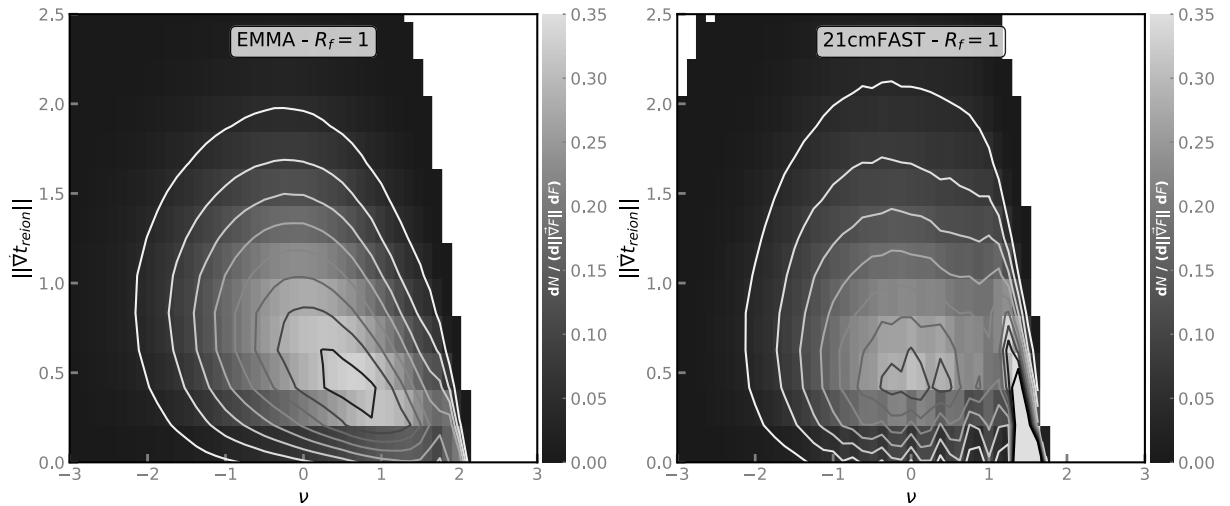
Globalement, les cartes  $t_{\text{reion}}(\vec{r})$  de la simulation 21cmFAST ont les mêmes caractéristiques que celles de la simulation EMMA, c'est-à-dire qu'elles sont aussi proches d'être Gaussiennes, que cette Gaussianité dépend de la même manière du lissage et que l'on y retrouve aussi des asymétries similaires. De même, on peut en extraire les mêmes informations sur la physique de la Réionisation. La différence majeure entre les mesures dans les deux simulations est en fait que les non-Gaussianités du champ 21cmFAST sont marquées différemment de celles du champ EMMA. On peut par exemple voir cet effet dans les distributions 2D de la norme du gradient de  $t_{\text{reion}}(\vec{r})$  en fonction des valeurs normalisées du champ, montrées dans la Figure 6.16. À gauche, la mesure est faite dans la simulation EMMA, et à droite, elle est faite dans la simulation 21cmFAST. On voit en effet que pour la simulation 21cmFAST l'asymétrie est très marquée due à une très forte hausse des vitesses en fin de Réionisation avec la norme du gradient de  $t_{\text{reion}}(\vec{r})$  qui tend vers 0. Cet effet provient en fait de l'absence de traitement du transfert radiatif dans 21cmFAST, qui n'impose alors pas de limite supérieure à la vitesse des photons et donc des fronts d'ionisation. Les fronts peuvent donc se propager plus rapidement que la vitesse de la lumière et c'est ce que l'on voit ici à partir de temps supérieur à  $\nu = 1$ , soit à un peu plus d'un sigma de la moyenne. On retrouve encore dans ce Chapitre que deux types de simulations différents peuvent avoir des topologies différentes malgré que leur histoire de réionisation soit similaires. On peut ainsi retenir avec cette étude que malgré la rapidité des simulations semi-analytiques, elles ont des approximations qui se remarquent particulièrement dans les statistiques présentées ici. Nous devons ainsi continuer à utiliser des simulations cosmologiques, en complément des modèles semi-analytiques.

## 6.7 Conclusions & perspectives

Dans cette étude, j'analyse pour la première fois la topologie du champ de temps de réionisation en 2D. Je montre que je peux en extraire de manière statistique un large jeu de mesures topologiques qui nous permettent de suivre l'évolution temporelle du processus de réionisation. Nous avons par exemple directement accès à l'histoire de réionisation qui nous dit *quand* l'hydrogène se fait globalement réioniser. Nous pouvons mesurer la distribution des vitesses des fronts d'ionisation, qui nous permettent de savoir *à quelle vitesse* l'hydrogène s'est fait réioniser. Nous pouvons aussi suivre l'évolution de la taille des bulles ionisées et régions neutres avec la longueur des isocontours pour comprendre *comment* l'hydrogène s'est fait réioniser. Ensuite, la longueur du squelette nous permet d'explorer *la fin* de la Réionisation et savoir à quel moment les fronts d'ionisation percolent. La distribution des temps de réionisation des sources de réionisation nous permet enfin de connaître *les instants*



**FIGURE 6.15** – Cartes de temps de réionisation extraites de la simulations EMMA à gauche et de la simulation 21cmFAST à droite. Elles sont lissées avec un filtre Gaussien de taille  $R_f = 1$ . La carte EMMA de gauche n'est qu'une portion de  $256^2$  cMpc<sup>2</sup>/h<sup>2</sup> de la carte originale pour pouvoir la comparer à la celle de 21cmFAST.



**FIGURE 6.16** – Fonctions de distribution de probabilité 2D de la norme du gradient de  $t_{\text{reion}}(\vec{r})$  en fonction des valeurs  $\nu$  prises par  $t_{\text{reion}}(\vec{r})$ . Ces mesures sont faites dans les cartes  $t_{\text{reion}}(\vec{r})$  de la simulation EMMA à gauche et de la simulation 21cmFAST à droite. Les cartes sont lissées par un filtrage Gaussien de taille  $R_f = 1$ . Toutes les réalisations de chaque champ sont additionnées. Les lignes en échelle de gris sont les isocontours de ces histogrammes. On voit ici que l'accélération des fronts ionisants est bien plus importante dans le cas de la simulation 21cmFAST .

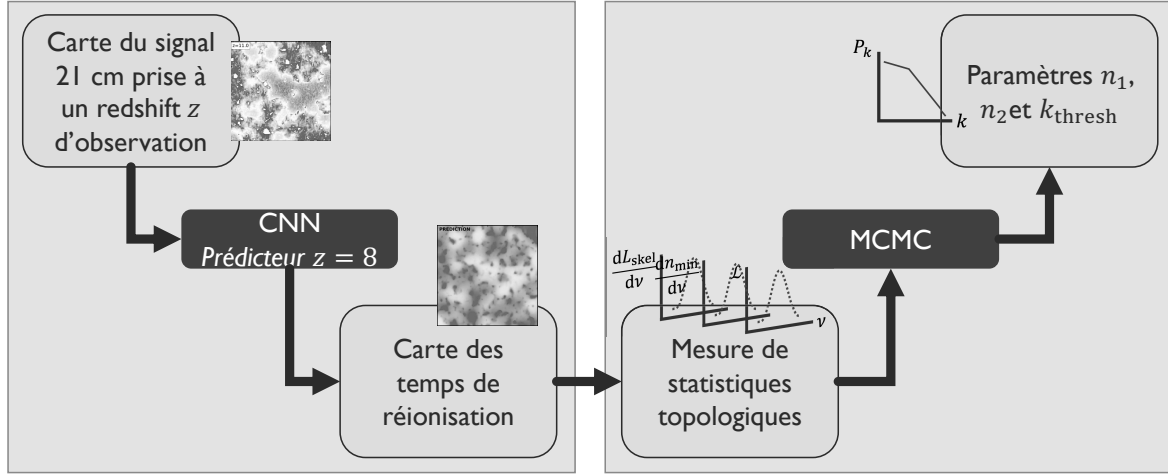
*d'apparition de ces sources.* Ces mesures sont également des outils qui nous permettent de distinguer des modèles de simulations et on voit ici qu'à cause de l'absence de transfert radiatif, la simulation 21cmFAST amplifie les asymétries. Toutes ces mesures peuvent aussi être comparées à des prédictions de la théorie des GRFs qui ne nécessitent que la connaissance du spectre de puissance du champ Gaussien. Ce travail montre que le champ  $t_{\text{reion}}(\vec{r})$  est Gaussien aux échelles supérieures à 8 cMpc/h et donc aux échelles que nous allons observer avec SKA. Cela signifie qu'à ces échelles et en première approximation, la seule connaissance du spectre de puissance de  $t_{\text{reion}}(\vec{r})$  nous permet d'accéder à la physique de la Réionisation. Par ailleurs, des empreintes de non-Gaussianités sont présentes dans le champ à cause des plus petites échelles. Elles prennent la forme d'asymétrie dans les distributions et sont dues à l'accélération des fronts d'ionisation dans les dernières régions neutres et peu denses vers la fin de la Réionisation.

Les travaux présentés dans ce Chapitre sont très importants pour mes projets à venir parce qu'ils posent les bases de ce que l'on peut faire avec le champ de temps de réionisation. Comme le champ est quasi-Gaussien aux grandes échelles, son spectre de puissance contient une grande partie de son information. On pourrait par exemple générer des modèles de temps de réionisation et de toutes ses propriétés aux grandes échelles (supérieures à environ 8 cMpc/h) à partir de son spectre de puissance uniquement. Parmi mes futurs projets, je souhaite ainsi reconstruire le spectre de puissance de  $t_{\text{reion}}(\vec{r})$  à partir de celui du signal 21 cm afin de pouvoir extraire des observations des informations sur la physique de la Réionisation d'une manière alternative à ce qui est déjà proposé dans la littérature. Cette reconstruction pourrait se faire par le biais de réseaux de neurones. D'autre part, j'ai aussi pour projet construire un modèle purement analytique grâce à la Gaussianité de  $t_{\text{reion}}(\vec{r})$  qui nous retournerait le signal moyen du 21 cm et son spectre de puissance pendant la Réionisation. C'est ce que Muñoz (2023) a commencé à faire pour l'Aube Cosmique, et le but est de pouvoir inférer les paramètres astrophysiques et contraindre le processus de la Réionisation.

Enfin, il serait intéressant de mener ce genre d'étude dans des simulations plus résolues et donc à plus petite échelle. Cela pourrait se faire par exemple avec la suite de simulation CoDa (Ocvirk et al. 2020) ou avec des futurs modèles faits par exemple par Dyablo (Aubert & Durocher 2021). Cependant, à ces résolutions, le champ de temps de réionisation sera d'autant plus non-Gaussien. Il conviendrait alors de prendre en compte cette non-Gaussianité en ajoutant des termes d'asymétrie par exemple dans les distributions avec la série Gram-Charlier, comme l'ont fait Gay et al. (2012) et Cadiou et al. (2020) notamment.

### ***Un exemple d'application...***

Pour illustrer un exemple possible d'application des études de ce Chapitre, je décris ici rapidement un projet en cours, dans lequel j'essaie d'inférer par une méthode de Monte-Carlo par chaîne de Markov (MCMC) les paramètres du spectre de puissance de  $t_{\text{reion}}(\vec{r})$  à partir des statistiques mesurées dans ce Chapitre (comme le représente la partie de droite du schéma de la Figure 6.17). L'objectif est d'être capable d'accéder au spectre de puissance des temps de réionisation (qui contient toute l'information de ce champ aux grandes échelles) même quand il est difficilement mesurable. Par exemple, nous avons vu dans le Chapitre 4 que le spectre de puissance des cartes de temps de réionisation reconstruites par réseau de neurones n'est pas exactement celui des cartes originales de la simulation. Dans ce genre de contexte, on peut imaginer que la combinaison de plusieurs statistiques mesurées dans le champ en



**FIGURE 6.17** – Schéma expliquant le procédé d’inférence des paramètres du spectre de puissance de  $t_{\text{reion}}(\vec{r})$  à partir de cartes de temps de réionisation. Les cartes dont sont extraits ces paramètres sont par exemple ici celles que nous pouvons reconstruire grâce au réseau de neurones présenté dans le Chapitre 4 à partir du signal 21 cm.

question contient toute son information. Je voudrais notamment essayer d’obtenir le « vrai » spectre de puissance des cartes  $t_{\text{reion}}(\vec{r})$  reconstruites avec les modèles de réionisation testés dans le réseau de neurones convolutionnel (CNN) présentées dans le Chapitre 4. J’illustre ce processus avec le schéma de la Figure 6.17.

Je tiens tout d’abord à remarquer que nous nous sommes rendus compte après la publication des travaux présentés dans ce Chapitre que le spectre de puissance de  $t_{\text{reion}}(\vec{r})$  ne nécessite en fait que 4 paramètres pour être pleinement défini. On a en effet une contrainte en  $k_{\text{thresh}}$  où le spectre de puissance des deux parties doit être identique, et cela impose  $A_2 = A_1 k_{\text{thresh}}^{n_1 - n_2}$ . On peut alors aussi donner la forme suivante à notre spectre de puissance :

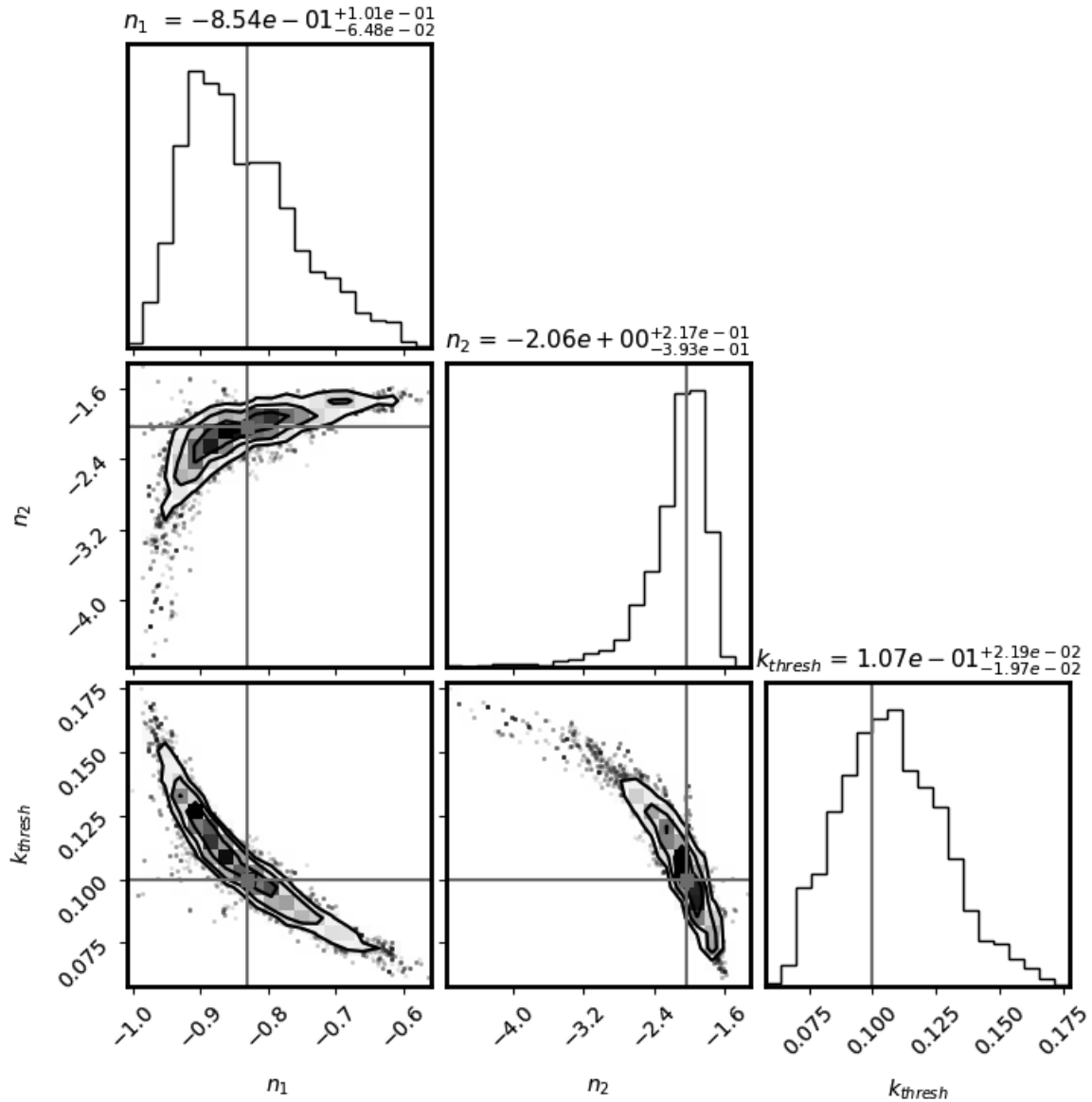
$$\mathcal{P}_k^{\text{alternatif}} = A_1 k_{\text{thresh}}^{n_1} \begin{cases} \left(\frac{k}{k_{\text{thresh}}}\right)^{n_1} & \text{si } k \leq k_{\text{thresh}} \\ \left(\frac{k}{k_{\text{thresh}}}\right)^{n_2} & \text{si } k > k_{\text{thresh}} \end{cases} . \quad (6.10)$$

Dans cette étude, nous utilisons cette forme du spectre de puissance pour permettre au MCMC de trouver les paramètres optimaux qui correspondent aux données d’entrée. De plus, seulement les puissances  $n_1$  et  $n_2$  et le seuil  $k_{\text{thresh}}$  du spectre de puissance sont recherchés ici parce que nous nous sommes rendus compte que les statistiques présentées dans ce Chapitre ne dépendent en fait pas de l’amplitude du spectre.

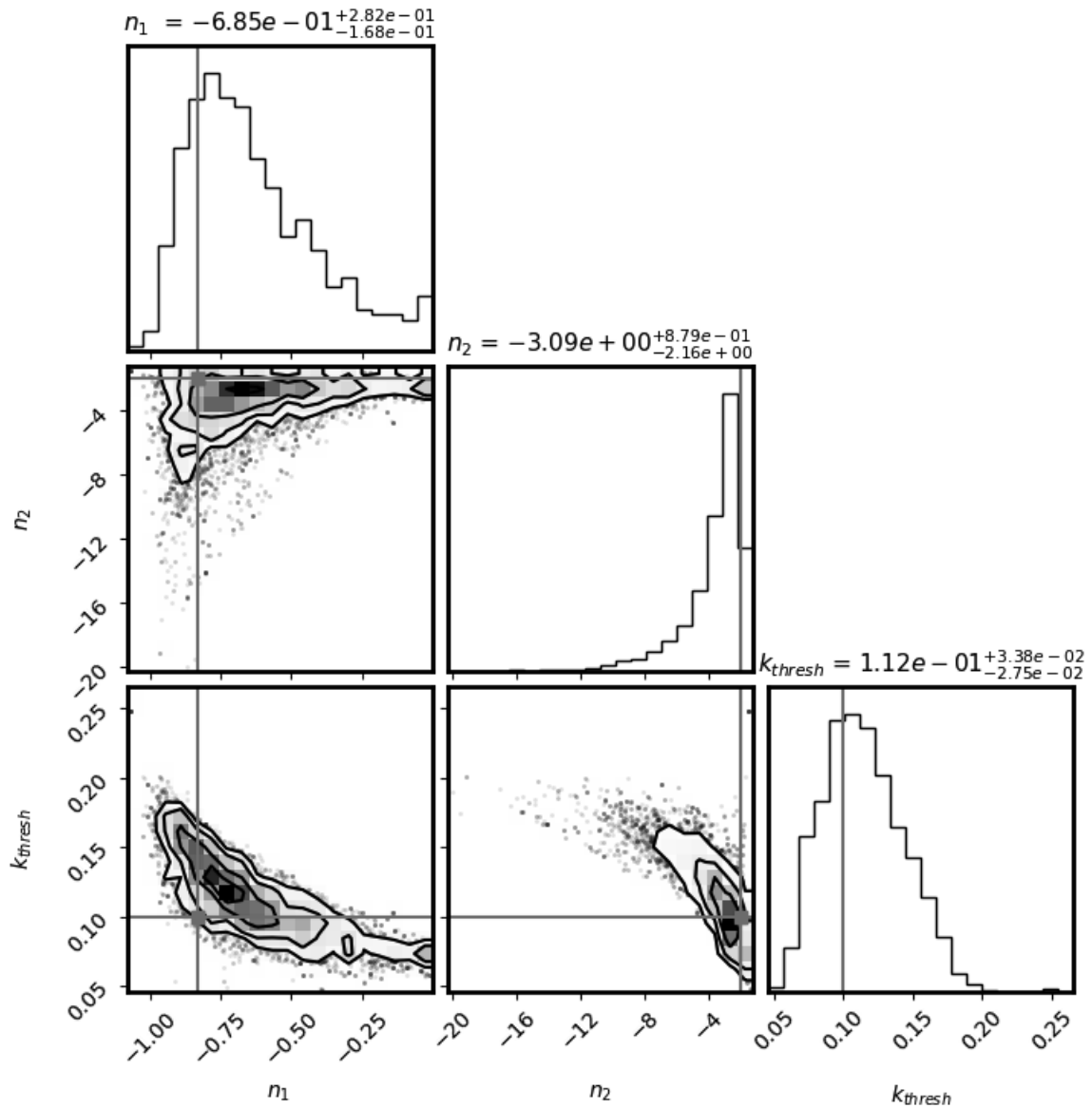
Je présente dans les Figures 6.18 et 6.19 des résultats préliminaires de ce projet. Dans la Figure 6.18, j’essaie de retrouver avec un MCMC (à 60 marcheurs qui effectuent 200 pas chacun<sup>4</sup>) les paramètres du spectre de puissance de  $t_{\text{reion}}(\vec{r})$  à partir des prédictions des longueurs d’isocontours, de la distribution des sources et de la longueur de squelette fournies par la théorie des GRFs en supposant un lissage avec  $R_f = 6$ . Dans un premier temps, j’utilise les prédictions Gaussiennes de  $\mathcal{L}$ ,  $\frac{\partial n_{\text{min}}}{\partial \nu}$  et  $\frac{\partial L^{\text{skel}}}{\partial \nu}$  pour être sûr qu’aucune non-Gaussianité

4. Ces valeurs sont choisies de sorte à amener le MCMC à converger.





**FIGURE 6.18** – Distributions des postérieurs marginalisés résultantes d'un MCMC ayant pour but de retrouver les valeurs des paramètres du spectre de puissance de  $t_{\text{reion}}(\vec{r})$  à partir des prédictions de la théorie des GRFs correspondantes. Les points rouges localisent les vrais paramètres du spectre de puissance de  $t_{\text{reion}}(\vec{r})$  que l'on cherche à retrouver. Ce MCMC contient 60 marcheurs qui effectuent chacun 200 pas.



**FIGURE 6.19** – Distributions des postérieurs marginalisés résultantes d’un MCMC ayant pour but de retrouver les valeurs des paramètres du spectre de puissance de  $t_{reion}(\vec{r})$  à partir des mesures faites sur la carte  $t_{reion}(\vec{r})$  de la simulation EMMA. Les points rouges localisent les vrais paramètres du spectre de puissance de  $t_{reion}(\vec{r})$  que l’on cherche à retrouver. Ce MCMC contient 60 marcheurs qui effectuent chacun 200 pas.

ne fausse les résultats. On voit ainsi que les paramètres inférés à partir de ces 3 prédictions Gaussiennes sont proches des 3 paramètres du spectre de puissance de  $t_{\text{reion}}(\vec{r})$  (représentés par les points rouges).

Dans un second temps, nous cherchons avec le même genre de MCMC dans la Figure 6.19 les paramètres du spectre de puissance à partir des mesures des trois statistiques faites dans les cartes de temps de réionisation de la simulation EMMA (lissées avec  $R_f = 6$ ). Les paramètres inférés sont de nouveau relativement proches des valeurs attendues pour le vrai spectre de puissance. Cela montre que dans un cas parfait, nous sommes capables de reconstituer la forme du spectre de puissance de  $t_{\text{reion}}(\vec{r})$  à partir de mesures topologiques.

La prochaine étape, que nous n'avons pas encore finalisée, est de retrouver avec le même genre de MCMC les « vrais » paramètres du spectre de puissance de  $t_{\text{reion}}(\vec{r})$  pour un modèle de réionisation donné à partir des mesures faites dans les reconstructions du CNN du Chapitre 4.

## 6.8 « Topology of reionisation times : Concepts, measurements, and comparisons to Gaussian random field predictions »

Cette section présente l'article « Topology of reionisation times : Concepts, measurements, and comparisons to Gaussian random field predictions » (ou en français « Topologie des temps de réionisation : concepts, mesures et comparaisons aux prédictions de champs Gaussiens aléatoires »). Cet article a été soumis le 15 septembre 2022 dans la revue à comité de lecture *Astronomy & Astrophysics*, et accepté le 20 février 2023. Il a été publié en avril 2023 dans le volume 672.

# Topology of reionisation times: Concepts, measurements, and comparisons to Gaussian random field predictions

Emilie Th  lie<sup> </sup>, Dominique Aubert<sup> </sup>, Nicolas Gillet, Julien Hiegel<sup> </sup>, and Pierre Ocvirk

Universit   de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, 11 rue de l'Universit  ,  
67000 Strasbourg, France  
e-mail: emilie.thelie@astro.unistra.fr

Received 15 September 2022 / Accepted 20 February 2023

## ABSTRACT

**Context.** In the next decade, radio telescopes, such as the Square Kilometer Array (SKA), will explore the Universe at high redshift, and particularly during the epoch of reionisation (EoR). The first structures emerged during this epoch, and their radiation reionised the previously cold and neutral gas of the Universe, creating ionised bubbles that percolate at the end of the EoR ( $z \sim 6$ ). SKA will produce 2D images of the distribution of the neutral gas at many redshifts, pushing us to develop tools and simulations to understand its properties.

**Aims.** With this paper, we aim to measure topological statistics of the EoR in the so-called reionisation time fields from both cosmological and semi-analytical simulations. This field informs us about the time of reionisation of the gas at each position; it is used to probe the inhomogeneities of reionisation histories and can be extracted from 21 cm maps. We also compare these measurements with analytical predictions obtained within Gaussian random field (GRF) theory.

**Methods.** The GRF theory allows us to compute many statistics of a field, namely the probability distribution functions (PDFs) of the field or its gradient, isocontour length, critical point distributions, and skeleton length. We compare these theoretical predictions to measurements made on reionisation time fields extracted from an EMMA simulation and a 21cmFAST simulation at 1 cMpc/h resolution. We also compared our results to GRFs generated from the fitted power spectra of the simulation maps.

**Results.** Both EMMA and 21cmFAST reionisation time fields ( $t_{\text{reion}}(\mathbf{r})$ ) are close to being Gaussian fields, in contrast with the 21 cm, density, or ionisation fraction, which have all been shown to be non-Gaussian. Only accelerating ionisation fronts at the end of the EoR seem to be the cause of small non-gaussianities in  $t_{\text{reion}}(\mathbf{r})$ . Overall, this topological description of reionisation times provides a new quantitative and reproducible way to characterise the EoR scenario. Under the assumption of GRFs, it enables the generation of reionisation models with their propagation, percolation, or seed statistics simply from the reionisation time power spectrum. Conversely, these topological statistics provide a means to constrain the properties of the power spectrum and by extension the physics that drive the propagation of radiation.

**Key words.** large-scale structure of Universe – dark ages, reionization, first stars – methods: numerical – methods: statistical – galaxies: formation – galaxies: high-redshift

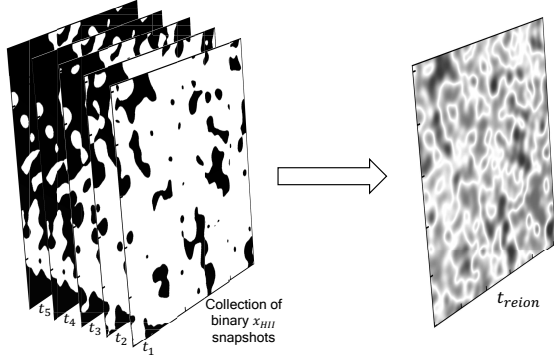
## 1. Introduction

The epoch of reionisation (EoR) saw the birth of stars and galaxies. Looking back at the history of the Universe, the first sources of radiation appear during the EoR, emitting photons that reionise the cosmic gas and create HII ‘bubbles’ around galaxies. These bubbles eventually percolate near the end of the EoR at  $z = 5.3\text{--}6$  (Barkana & Loeb 2001; Dayal & Ferrara 2018; Kulkarni et al. 2019; Wise 2019). This epoch marks the transition from a Universe with totally cold and neutral gas to the Universe we see today, where the gas is warmer and ionised.

The evolving geometry of the EoR has been widely investigated in the literature in order to understand physical processes, such as the growth of structures, the geometry of the ionised and neutral bubbles, and the percolation process. Many works focus on the geometry of the ionised and neutral bubbles and on percolation with Minkowski functionals (or derived statistics, such as the Euler characteristic, the genus, or the shapefinders; see Gleser et al. 2006; Lee et al. 2008; Friedrich et al. 2011; Hong et al. 2014; Yoshiura et al. 2017; Chen et al. 2019; Pathak et al. 2022), with the triangle correlation function (Gorce & Pritchard 2019), or with the Morse theory and persistent homology (Th  lie et al. 2022). Other studies

extract the size and shape of the ionised bubbles thanks to the contour Minkowski tensor (Kapahtia et al. 2018, 2019, 2021). Counting the numbers of 3D structures in a field (isolated objects, such as peaks, tunnels, and voids) can be done using the Betti numbers (Kapahtia et al. 2019, 2021; Giri & Mellema 2021; Bianco et al. 2021; Elbers & van de Weygaert 2023). The size of the ionised or neutral bubbles has also been investigated with methods, such as the friends-of-friends algorithm (Iliev et al. 2006; Friedrich et al. 2011; Lin et al. 2016; Giri et al. 2018a, 2019), the spherical average method (Zahn et al. 2007; Friedrich et al. 2011; Lin et al. 2016; Giri et al. 2018a), the mean free path method (Mesinger & Furlanetto 2007; Lin et al. 2016; Giri et al. 2018a, 2019; Bianco et al. 2021), and the granulometry method (Kakiichi et al. 2017; Busch et al. 2020). In addition, the low-frequency component of the Square Kilometre Array radio interferometer<sup>1</sup> (SKA-Low; see e.g. Mellema et al. 2013) will produce 2D tomographic images of the 21 cm HI emission at many redshifts during the EoR. There have therefore been many studies exploring the spatial structure of this signal, for example using the 21 cm power spectrum (Zaldarriaga et al. 2004; Furlanetto et al. 2004; McQuinn et al. 2006; Bowman et al. 2006;

<sup>1</sup> <https://www.skatelescope.org>



**Fig. 1.** Schematic summarising the use of the reionisation time field (on the right): it allows us to use only one field instead of a series of many snapshots of binary ionised fraction (on the left) for example.

Lidz et al. 2008; Iliev et al. 2012; Mesinger et al. 2013; Pober et al. 2014; Greig & Mesinger 2015, 2017; Liu & Parsons 2016; Kern et al. 2017; Park et al. 2019; Pagano & Liu 2020; Gazagnes et al. 2021), or the 21 cm bispectrum (Hutter et al. 2020).

**Reionisation time field: definition and motivation.** Our work focuses on the reionisation time (or redshift) map. This field is generated by cosmological simulations and models (e.g. EMMA and the 21cmFAST semi-analytical code), and corresponds to the time at which each position of the simulation box is considered to be reionised, which is when the ionisation fraction exceeds a 50% threshold, as follows:

$$t_{\text{reion}}^*(\mathbf{r}) = t(\mathbf{r}, x_{\text{HII}} = 0.5), \quad (1)$$

where  $\mathbf{r}$  is the position, and  $x_{\text{HII}}$  is the ionised fraction.  $t_{\text{reion}}^*$  is measured from the Big Bang, meaning that the cosmic gas is almost entirely reionised at  $t_{\text{reion}}^* \sim 1$  Gyr. As shown in Fig. 1 and thanks to  $t_{\text{reion}}(\mathbf{r})$ , we compress the information about the evolution of  $x_{\text{HII}}$  into a single field instead of using a collection of snapshots. In the reionisation time map, blue regions correspond to those where the gas reionises first, whereas the red ones are the last regions where the gas reionises. We focus on 2D  $t_{\text{reion}}(\mathbf{r})$  maps to study the EoR on the sky as it will be observed (with the upcoming SKA 2D images for instance).

The field  $t_{\text{reion}}(\mathbf{r})$  holds both spatial and temporal information on the reionisation scenario and is therefore often used to characterise or compare the evolving structure of the reionisation provided by models. For example, it can be used to measure the speed and direction of ionising radiation as it propagates from sources (Deparis et al. 2019; Thélie et al. 2022). Recently,  $t_{\text{reion}}(\mathbf{r})$  was also used to efficiently generate models of the reionisation (Trac et al. 2022). This parameter is also valuable in investigations of local variations of the reionisation scenario (Trac et al. 2008; Battaglia et al. 2013; Aubert et al. 2018; Zhu et al. 2019; Sorce et al. 2022) and the consequences of an inhomogeneous reionisation. These local modulations of reionisation histories could possibly manifest themselves in the star formation histories of low-mass galaxies (Ocvirk et al. 2020) or their spatial distribution (Ocvirk & Aubert 2011).  $t_{\text{reion}}$  is therefore a versatile descriptor of models and in this paper we propose to revisit its study in a more general manner. In particular, we show how the topological study of this field can unravel many properties of the summarised reionisation scenario in a physically meaningful, quantitative, and reproducible way.

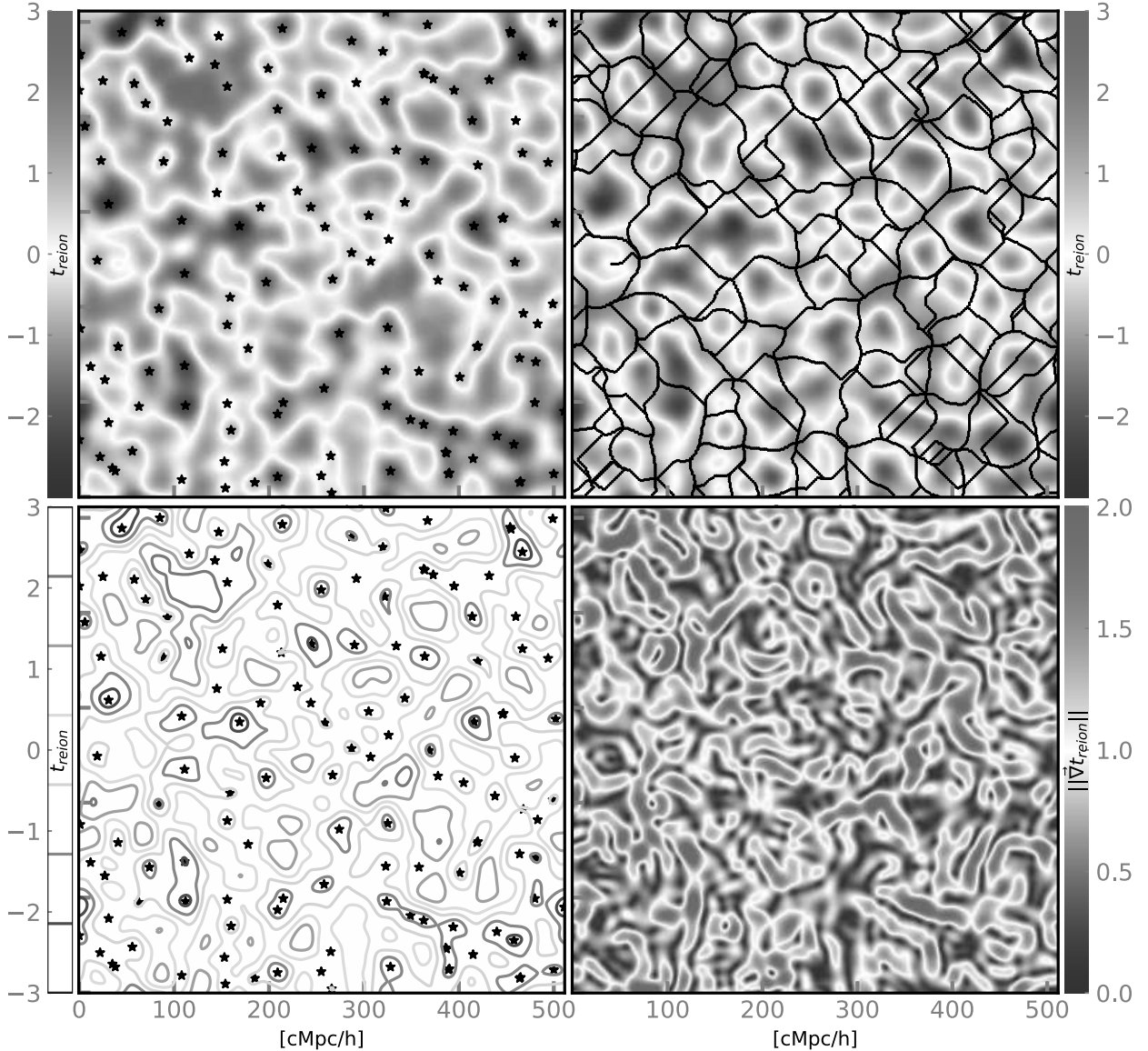
However, we also claim that this field, and the study of its topology, is not only useful in the strict and limited scope of reionisation models but also in the context of future observations. Indeed, in the next decade, radio telescopes, such as SKA, will map the intergalactic medium (IGM) during the EoR thanks to the 21 cm radiation coming from neutral hydrogen atoms (e.g. Koopmans et al. 2015). 21 cm lightcones along the line of sight will contain a wealth of information on the evolving reionisation state of the IGM and on the underlying matter density, the ionisation fraction of the gas, and its thermal state. In Hiegel et al. (in prep.), we show that it is also possible to reconstruct 2D reionisation time maps from 2D 21 cm images thanks to a convolutional neural network. Therefore,  $t_{\text{reion}}(\mathbf{r})$  could possibly be extracted from observations (even though small structures are smoothed out), thus granting access to the evolution of the reionisation in the transverse plane, which would complement line-of-sight studies. More details about these reconstructions are given in Appendix A. In the future, we intend to use the framework described in the present paper on the reionisation time field reconstructed from 21 cm observations.

**Topology of the reionisation time field.** The topological features of  $t_{\text{reion}}(\mathbf{r})$  can be described by analogy with a mountainous landscape (Gay 2011). Consider a mountainous landscape as a 2D field: for each 2D position of the space, the altitude is the value of the field. The mountain peaks are the maxima of the field, the bottoms of the valleys are the minima, and the mountain passes are the saddle points. The skeleton of this field corresponds to all the ridge lines, which join each of the passes (i.e. saddle points) to the peaks (i.e. maxima) and are the lines with the least slope. The skeleton forms a connected network throughout the space. With  $t_{\text{reion}}(\mathbf{r})$ , many geometrical quantities can be interpreted physically to describe the evolution of the ionised and neutral gas during the EoR. Recently, in Thélie et al. (2022), we studied the 3D topological properties of the EoR, such as the shape, size, and orientation of ‘peak patches of reionisation’. Here, we go further and work on a large set of geometrical properties of the 2D reionisation time fields  $t_{\text{reion}}(\mathbf{r})$ , as shown in Fig. 2.

– Probability distribution function (PDF) of the field values. A widely used statistic when studying the EoR is the fraction of ionised volume  $Q_{\text{HII}}$ . We can measure this with the reionisation time field as it is directly the cumulated distribution function of the  $t_{\text{reion}}(\mathbf{r})$  values (i.e. the number of cells that has a lower reionisation time than a given threshold). In other words, it is the reionisation history; it contains information about the timing of reionisation, as well as its global evolution.

– PDF of the gradient norm field. Moreover, we can compute the first derivative of the reionisation time field, and extract the norm of its gradients. This norm corresponds to the time interval within which the gas is reionised in each cell of the simulation boxes ( $\sim \Delta t / \Delta x$ ), which is equivalent to the inverse of a velocity field. An example of this gradient norm field is shown in the bottom-right panel of Fig. 2. The analysis of this field allows us to study the ‘velocity’ of the radiation fronts (Deparis et al. 2019).

– Isocontour length. The isocontours of  $t_{\text{reion}}(\mathbf{r})$  delineate the regions reached at a given time by the HII bubbles. Their length is interesting because it contains information about the growth of the ionised bubbles and the decrease in size of the last neutral bubbles. With the bottom left panel of Fig. 2, we see the first ionised bubbles with the darkest blue contours, their growth with the lighter blue contours, and their fusion when the blue contours merge. We also see the last neutral regions being ionised with the



**Fig. 2.** Two-dimensional slices related to the EMMA reionisation time field that are smoothed with a Gaussian kernel with a standard deviation of  $R_f = 6$ . The top left panel shows  $t_{\text{reion}}(\mathbf{r})$  with its minima (black stars). The top right panel shows  $t_{\text{reion}}(\mathbf{r})$  with its skeleton (black lines). The bottom left panel shows the isocontours of  $t_{\text{reion}}(\mathbf{r})$ , with its minima again in black stars. Here, the reionisation time field is normalised to put its mean at a value of 0, and its standard deviation at 1. The bottom right panel shows the norm of the gradient of  $t_{\text{reion}}(\mathbf{r})$ .

red contours. Here, we have an insight into the percolation process of the EoR.

– Reionisation seed count. We can extract the critical points of the reionisation time field, and in particular its minima. These points correspond to the sources of radiation (first zones that reionise) and are shown on the example field of Fig. 2 with the black stars on the left panels. As expected, the critical points are within the bluest zones, which reionise first, and are also in the middle of the first blue isocontours. Their distribution as a function of time allows us to know the time of appearance of these reionisation seeds. For instance, we can infer the moment when the maximum number of sources lights up. This distribution should also correlate with the star formation rate.

– Reionisation patches. We can extract the void patches from the reionisation time field (or the peak patches from the reionisation redshift field). These patches contain all of the cells that are linked to the  $t_{\text{reion}}(\mathbf{r})$  minima by a negative gradient. Thanks to them, we can study the extent of the radiative influence of a reionisation seed with size distributions. Their shape and orientation with respect to the density filaments informs us about the direction of propagation of the reionisation fronts. As we studied these patches in Thélie et al. (2022), we do not focus on them in this paper.

– Distribution of the skeleton lengths. We also calculate the distribution of the length of the skeleton of  $t_{\text{reion}}(\mathbf{r})$ . An example skeleton is shown with the black lines in the top right panel of Fig. 2. As explained before, they connect the maxima of a field

by passing through its saddle points as ridge lines. The skeleton of the reionisation time field  $t_{\text{reion}}(\mathbf{r})$  physically corresponds to the front lines between the propagating radiation that comes from the reionisation seeds. It therefore indicates the extent to which the photons can propagate from a source and ionise the medium before reaching an opposite ionising front coming from another source: it is the percolation lines between patches of reionisation. The skeleton length distribution with respect to the time tells us about the length of merging radiation fronts at a given time.

In this work, we study all of these properties (except the reionisation patches) through measurements in  $t_{\text{reion}}(\mathbf{r})$  maps that are extracted from simulations obtained with the EMMA cosmological code and the 21cmFAST semi-analytical code.

*Gaussianity of the reionisation time field.* We also analyse the gaussianity of the reionisation time field ( $t_{\text{reion}}(\mathbf{r})$ ), and therefore of the reionisation process. The EoR is known to be ruled by strongly non-Gaussian phenomena that are probed in different ways in the literature. Many studies look at this non-Gaussian nature of the EoR directly through the 21 cm PDF, sometimes using skewness, kurtosis, and quantile analyses (Mellema et al. 2006; Ichikawa et al. 2010; Dixon et al. 2016; Ross et al. 2019; Banet et al. 2021). Other studies use higher-order statistics, such as the bispectrum or trispectrum (Majumdar et al. 2018; Shaw et al. 2020). The density or ionisation fraction fields are also studied for their non-Gaussian features (Iliev et al. 2006). The non-gaussianities of the EoR are generally said to be due to the non-linear structure formation (Bernardeau et al. 2002), to be present at all scales, and to increase in importance as the reionisation process proceeds. Moreover, some studies show that non-gaussianities are increased within the HII bubbles due to high ionisation and high densities (Iliev et al. 2006; Dixon et al. 2016; Majumdar et al. 2018), which could be due to an inside-out process of ionisation (Iliev et al. 2006). Ross et al. (2019) also explains that the quasi-stellar objects (QSOs) and the X-ray heating can be another cause of non-gaussianity. Overall, understanding the source of non-gaussianities will help us to obtain a broader comprehension of the EoR physical processes, and to put constraints on the reionisation parameters (Shaw et al. 2020; Greig et al. 2022).

In this study, we compare the measured statistics on  $t_{\text{reion}}(\mathbf{r})$  mentioned above to Gaussian random field (GRF) theory predictions. This theory allows us to compute statistics of Gaussian-distributed field (Rice 1944; Longuet-Higgins 1957; Doroshkevich 1970; Bardeen et al. 1986; Hamilton et al. 1986; Pogosyan et al. 2009a,b; Pichon et al. 2010; Gay et al. 2012; Cadiou et al. 2020) or weakly non-Gaussian fields (Matsubara 2003; Pogosyan et al. 2011; Gay et al. 2012; Cadiou et al. 2020; Matsubara & Kuriki 2021). Rice (1944) first introduced the GRF theory to extract statistics from the 1D random noise of electronic devices. Longuet-Higgins (1957) later used the same theory, but this time on random waves on 2D surfaces. It is only later that Bardeen et al. (1986) used the GRF theory in astrophysics in order to study 3D structure formation in a cosmological context simply using information in a power spectrum. More recently, Gay et al. (2012) extracted many statistics of 2D and 3D cosmological fields thanks to this theory, for example counting the peaks on synthesised cosmic microwave background (CMB) maps to study their non-gaussianities. Some widely used statistics in the context of the EoR can be analytically calculated thanks to the GRF theory, such as Minkowski functionals, the derived genus or Euler characteristics, and Betti numbers (Schmalzing & Gorski 1998; Matsubara 2003; Lee et al.

2008; Gay et al. 2012; Kapahtia et al. 2019; Matsubara & Kuriki 2021). For example, Lee et al. (2008) computed the genus of the neutral hydrogen field  $x_{\text{HII}}$  to study the evolution of the EoR through many phases.

Here, we compare the statistics of the EoR that we measure on  $t_{\text{reion}}(\mathbf{r})$  with the analytic predictions of the GRF theory for the first time. The advantage of GRFs is that all of the field information is compressed within their power spectrum: we therefore envision the prospect of summarising the timing and the evolution of the EoR with  $t_{\text{reion}}(\mathbf{r})$  or its power spectrum. A Gaussian  $t_{\text{reion}}(\mathbf{r})$  has interesting applications. For example, associated evolving ionisation fields can easily be generated if we know the  $t_{\text{reion}}(\mathbf{r})$  power spectrum; one can imagine having a new class of fast-forward models of the reionisation process within which we could vary astrophysical parameters (that are hopefully encoded in the power spectrum of the reionisation time field). Conversely, the properties of the aforementioned topological statistics could be used to constrain the power spectrum under the GRF assumption, and by extension the physics that drives the propagation of radiation. Also, with a Gaussian  $t_{\text{reion}}(\mathbf{r})$ , its power spectrum could be directly retrieved from measurements of its topological statistics. This could be interesting in the case where the power spectrum has not been properly measured. For instance, with the reconstructed reionisation time maps from 21 cm observations by the CNN mentioned above and in Appendix A, the power spectrum would be suppressed on small scales due the CNN smoothing, and combining the diverse statistics measured in  $t_{\text{reion}}(\mathbf{r})$  could help to obtain the proper power spectrum. Finally, by comparing the measurements and predictions, we show in this paper the extent to which it is realistic to suppose that  $t_{\text{reion}}(\mathbf{r})$  is a Gaussian field, and we also infer a few causes of non-gaussianities.

*Organisation of the paper.* We start by describing the EMMA and 21cmFAST simulations we used in Sect. 2, as well as some generated GRFs. Subsequently, Sect. 3 presents every topological characteristic we studied with the GRF theory. We study the behaviour of each of these characteristics with the generated GRFs, and measure them using the EMMA reionisation time fields. The same analyses are also presented for a 21cmFAST simulation in Sect. 4. In Sect. 5 we present our conclusions about this work and introduce a few perspectives. Appendix A presents some details about how we can reconstruct reionisation time maps from observation-like data. Appendix B details the calculation of the spectral moments from a specific power spectrum. Appendix C is the full calculation to obtain the PDF of the norm of the gradient of a GRF. Some results are also shown for the reionisation redshift field in order to compare it to the reionisation time field in Appendix D. The cosmology parameters used are  $(\Omega_m, \Omega_b, \Omega_\Lambda, h, \sigma_8, n_s) = (0.31, 0.05, 0.69, 0.68, 0.81, 0.97)$ , as given by Planck Collaboration VI (2020).

*Notations.* Throughout this paper, we use the following notations, which were introduced in Pogosyan et al. (2009b), Gay et al. (2012). We use  $F$  to denote the studied field, which refers to the reionisation time fields. In this study, we work with normalised fields using the momenta of  $F$  and its derivatives:

$$\sigma_0^2 = \langle F^2 \rangle, \quad \sigma_1^2 = \langle (\nabla F)^2 \rangle, \quad \text{and} \quad \sigma_2^2 = \langle (\Delta F)^2 \rangle. \quad (2)$$

We introduce thereafter notations for the normalised fields and its derivatives:

$$x = \frac{F}{\sigma_0}, \quad x_i = \frac{\nabla_i F}{\sigma_1}, \quad \text{and} \quad x_{ij} = \frac{\nabla_i \nabla_j F}{\sigma_2}, \quad (3)$$

**Table 1.** Average and standard deviation of both EMMA and 21cmFAST reionisation time and redshift fields.

$R_f$	$\overline{t_{\text{reion}}^*}$	1	$\sigma_0^{t_{\text{reion}}^*}$	2	6
EMMA [Myrs]	790	107	91.6	62.2	
21cmFAST [Myrs]	801	165	149	97.5	
$R_f$	$\overline{z_{\text{reion}}^*}$	1	$\sigma_0^{z_{\text{reion}}^*}$	2	6
EMMA	6.99	0.82	0.67	0.43	
21cmFAST	6.40	1.32	1.17	0.74	

**Notes.** The standard deviations are computed thanks to the expression given in Appendix B, and depend on the power-spectrum parameters of the fields.

with  $i, j \in \{1, 2\}$  for 2D fields. Here, we mainly work on the reionisation time field  $t_{\text{reion}}^*(\mathbf{r})$ , which we normalise as follows:

$$x = t_{\text{reion}} = \frac{t_{\text{reion}}^* - \overline{t_{\text{reion}}^*}}{\sigma_0} \quad (4)$$

where  $\overline{t_{\text{reion}}^*}$  is the mean of the field. In addition, the normalised reionisation time fields are denoted  $x$  in the following work. However, when we refer to the values of the field, we use the following notation:

$$\nu = x(\mathbf{r}). \quad (5)$$

As we work on normalised reionisation time fields, when  $\nu < 0$ , we probe moments before the average reionisation times, and when  $\nu > 0$ , we probe those after the average reionisation time. Also, low values of  $\nu$  therefore refer to early reionisation times, and large values of  $\nu$  to late reionisation times.

We also introduce dimensionless spectral parameters:

$$R_0 = \frac{\sigma_0}{\sigma_1}, \quad R_* = \frac{\sigma_1}{\sigma_2}, \quad \text{and} \quad \gamma = \frac{R_*}{R_0} = \frac{\sigma_1^2}{\sigma_0 \sigma_2}. \quad (6)$$

These parameters can be analytically expressed if the power spectrum of the field is known. The calculation is shown in Appendix B for a specific type of power spectrum.

## 2. Simulated data

### 2.1. EMMA simulation

In this work, we used a  $512^3 \text{ cMpc}^3 \text{ h}^{-3}$  cosmological simulation with a resolution of  $1 \text{ cMpc}^3 \text{ h}^{-3}$ , as detailed in Gillet et al. (2021). This simulation was obtained with the cosmological code EMMA (Electromagnétisme et Mécanique sur Maille Adaptative, Aubert et al. 2015), which is an adaptive mesh refinement (AMR) code that couples hydrodynamics and radiative transfer, and in which light is described as a fluid (resolved using the moment-based M1 approximation; Aubert & Teyssier 2008). The EMMA simulation follows the cosmology given by Planck Collaboration VI (2020) and has no AMR, no reduced speed of light, and a stellar particle mass of  $10^8 M_\odot$ .

As this work is focused on 2D fields, 100 slices of  $512^2 \text{ cMpc}^3 \text{ h}^{-2}$  in size (spaced one from another by 5 slices) are extracted from the  $t_{\text{reion}}(\mathbf{r})$  field. Slices are smoothed with a Gaussian kernel of standard deviation  $R_f \in \{1, 2, 6\}$  (see Sect. 2.3.2), and normalised as described in Sect. 1. The average and standard deviation of the EMMA reionisation time field are given in Table 1.

**Table 2.** Angular resolutions corresponding to the size of the smoothing kernel applied to our simulation at a redshift  $z = 6.905$ .

$R_f$	Simulations			SKA
	1	2	6	–
$\Delta_x$ [cMpc]	1.48	2.96	8.88	8.3
$\Delta_\theta$ [arcmin]	0.57	1.14	3.42	3.11

**Notes.** The angular and spatial resolutions are also given for the radio-telescope SKA at the same redshift for a maximum baseline of 2 km (Giri et al. 2018b).

### 2.2. 21cmFAST simulation

We compare the statistical measurements of the EMMA simulation to those of a semi-analytical simulation generated with 21cmFAST<sup>2</sup> (version 3.0.3; Mesinger et al. 2011; Murray et al. 2020). The size of the simulation box generated is  $256^3 \text{ cMpc}^3 \text{ h}^{-3}$ , again with a resolution of  $1 \text{ cMpc}^3 \text{ h}^{-3}$ . The reionisation model used varies by only two parameters from the default ones: the ionising efficiency of high-redshift galaxies  $\zeta = 40$ , and the virial temperature  $T_{\text{vir}} = 10^5 \text{ K}$ . Here,  $\zeta$  controls the number of photons emitted by galaxies: the higher it is, the faster the reionisation.  $T_{\text{vir}}$  is the minimum virial temperature allowing a halo to start forming stars. Those parameters are chosen to approximately match the reionisation history of the 21cmFAST simulation with that of the EMMA simulation. 21cmFAST can provide us with the reionisation redshift maps that we can convert into reionisation time maps (with a given cosmology), similar to the EMMA ones. From this 3D simulation, 51 slices (of size  $256^2 \text{ cMpc}^2 \text{ h}^{-2}$ ) can be extracted (again spaced from each other by 5 slices), and they are also smoothed and normalised the same way the EMMA slices are. The average and standard deviation of the 21cmFAST reionisation time field are given in Table 1.

### 2.3. Choices for the simulation data sets

#### 2.3.1. Reionisation time and redshift fields

We also extracted the reionisation redshift fields  $z_{\text{reion}}(\mathbf{r})$  from both EMMA and 21cmFAST simulations. The following statistical analyses are performed on both reionisation time and redshift fields. In the main text, we only present the results of the  $t_{\text{reion}}(\mathbf{r})$  field, and we briefly present a similar analysis of  $z_{\text{reion}}(\mathbf{r})$  in Appendix D.

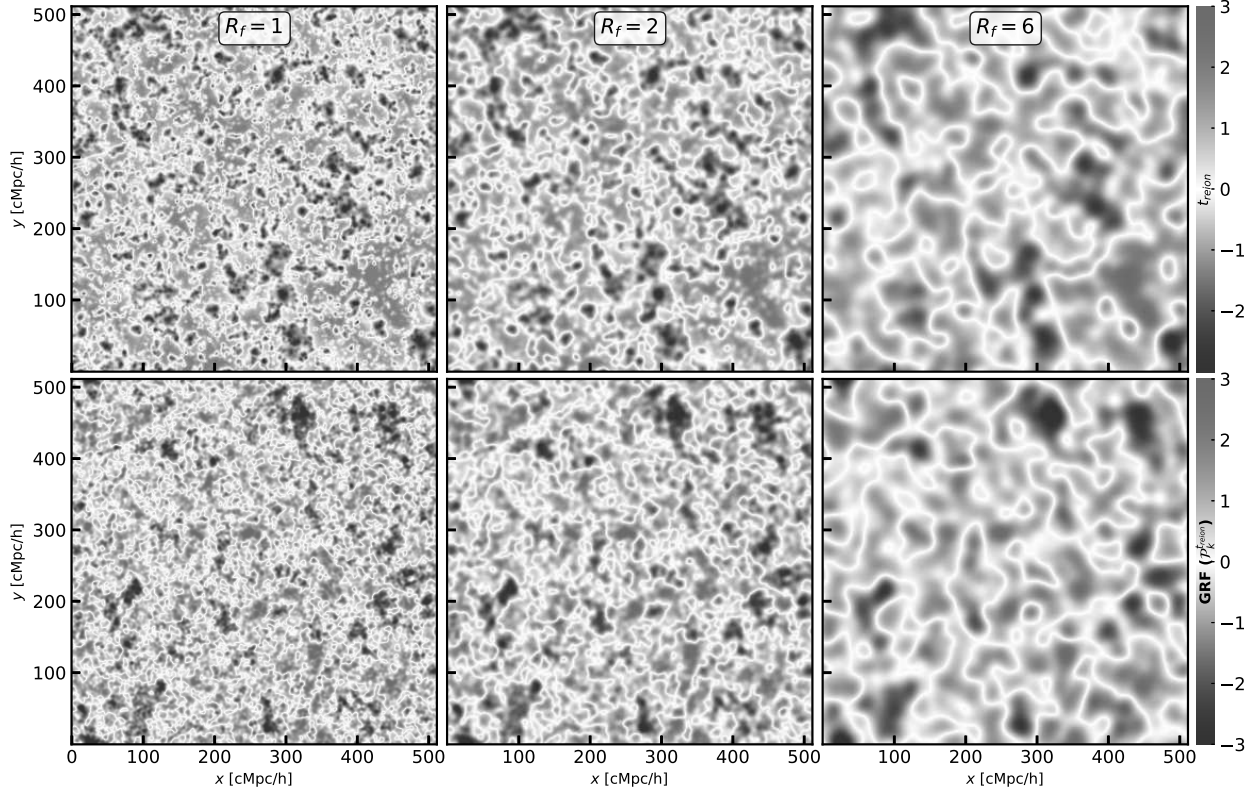
#### 2.3.2. Smoothing

As mentioned above, we apply different smoothings on the reionisation time fields. We use Gaussian kernels with different standard deviations. We chose to smooth the fields for the following reasons:

- The future observed images from the SKA for example will have a lower spatial resolution than what we simulate here with  $R_f \in \{1, 2\}$  (see the angular resolutions given for the different kernel sizes in Table 2 compared to the equivalent values for SKA), and so SKA will not be able to probe the smallest scales of our simulations.
- In order to compute the field momenta or spectral parameters, we need to integrate over power spectra, which leads to possible divergence (see the shape of the spectra in Sect. 2.4). Smoothing

<sup>2</sup> <https://github.com/andreimesinger/21cmFAST>





**Fig. 3.** Two-dimensional slices of the EMMA reionisation time field (first row) and the 2D GRFs obtained with the corresponding power spectrum (second row). Each column corresponds to different smoothings, with from left to right,  $R_f \in \{1, 2, 6\}$ . All the fields are normalised.

the fields circumvents this divergence problem, and is also well accounted for in the GRF theory.

– We use the discrete persistent structure extractor (DisPerSE<sup>3</sup>; Sousbie 2011), which assumes that the input field defines a Morse function, without large zero-gradient patches. As in Th  lie et al. (2022), we use it here to extract the critical points and the skeleton of  $t_{\text{reion}}(\mathbf{r})$ , and smoothing the field removes such patches.

– The EMMA reionisation time fields are represented in the first row of Fig. 3 for the different smoothings. The Gaussian smoothing filters out the smallest structures, while keeping the global shape of the larger scales: we ‘gaussianise’ the fields. Therefore, smoothing the reionisation time fields allows us to pinpoint the scales that are at the origin of non-Gaussian features in our measurements.

#### 2.4. Gaussian random fields

A GRF is a Gaussianly distributed field and if it has a null average, its PDF can be written as follows:

$$P(\mathbf{x})d^n \mathbf{x} = \frac{1}{(2\pi)^{n/2} \cdot \det(C)^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x} \cdot C^{-1} \cdot \mathbf{x}\right) d^n \mathbf{x}, \quad (7)$$

where  $\mathbf{x}$  is a n-D vector function of the position and  $C = \langle \mathbf{x} \otimes \mathbf{x} \rangle$  is the covariance matrix. For instance,  $\mathbf{x}$  could be expressed as follows:  $\mathbf{x} = (x, x_1, x_2, \dots)$  with the dimensionless field  $x = F/\sigma_0$ , and its first derivatives as defined in Eq. (3).

<sup>3</sup> <http://www2.iap.fr/users/sousbie/web/html/indexd41d.html>

To generate and analytically study this kind of field, we only need its power spectrum, as this defines a GRF entirely. From the power spectrum  $\mathcal{P}_k$ , we also have access to the momenta, as follows (Bardeen et al. 1986; Pogosyan et al. 2009b; Gay 2011):

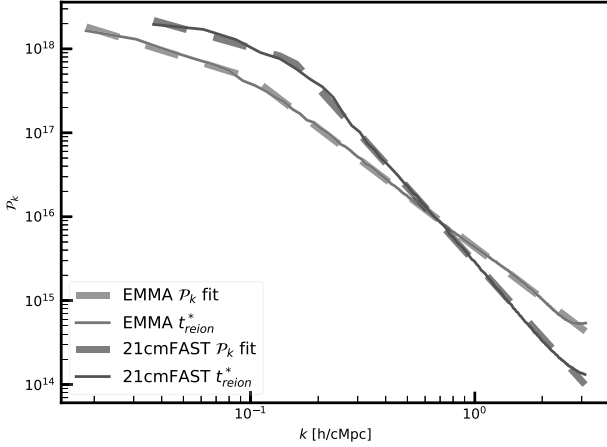
$$\sigma_i^2 = \frac{2\pi^{d/2}}{\Gamma(\frac{d}{2})} \int_0^\infty k^{2i} \mathcal{P}_k k^{d-1} dk, \quad (8)$$

where  $i \in \mathbb{N}$  corresponds to the number of times the field is derived, and  $d$  is the dimension of the field. The analytical derivation of the momenta is presented in Appendix B for a specific form of power spectrum detailed below.

We use the average power spectrum of the slices of our simulated fields: they are represented in logarithmic scales in Fig. 4 for the EMMA and 21cmFAST reionisation time fields. We use the following expression to fit these power spectra:

$$\mathcal{P}_k = \begin{cases} A_1 k^{n_1} & \text{if } k \leq k_{\text{thresh}} \\ A_2 k^{n_2} & \text{if } k > k_{\text{thresh}} \end{cases}, \quad (9)$$

where  $A_1$  and  $A_2$  are the amplitude of each part, and  $n_1$  and  $n_2$  are the power of each part.  $k_{\text{thresh}}$  is the threshold separating the two parts of the power spectrum. We obtain the parameters given in Table 3 after fitting the EMMA and 21cmFAST reionisation time power spectra. In Fig. 4, the dashed lines represent the average expected power spectra in logarithmic scales of  $t_{\text{reion}}(\mathbf{r})$ , which fit both simulations curves relatively well. This figure also demonstrates that there are more large structures in the 21cmFAST field than in the EMMA field, and also fewer small structures.



**Fig. 4.** Fitting (in dashed lines) of the reionisation time power spectra. The straight lines correspond to the power spectrum measured on the fields of each simulation (EMMA in blue and 21cmFAST in brown). The fittings are done on the average logarithmic power spectrum of every 2D slice for each field. The fields are not normalised when the fitting is carried out.

**Table 3.** Parameters defining the power spectra of the reionisation time and redshift fields ( $t_{\text{reion}}(\mathbf{r})$  and  $z_{\text{reion}}(\mathbf{r})$ ) for both EMMA and 21cmFAST simulations.

	$A_1$	$n_1$	$A_2$	$n_2$	$k_{\text{thresh}}$
<b>EMMA</b>					
$t_{\text{reion}}$	$6.70 \times 10^{16}$	-0.83	$4.34 \times 10^{15}$	-2.03	0.10
$z_{\text{reion}}$	4.03	-0.75	0.38	-1.80	0.10
<b>21cmFAST</b>					
$t_{\text{reion}}$	$1.83 \times 10^{17}$	-0.75	$2.87 \times 10^{15}$	-2.96	0.15
$z_{\text{reion}}$	6.17	-0.91	0.25	-2.86	0.20

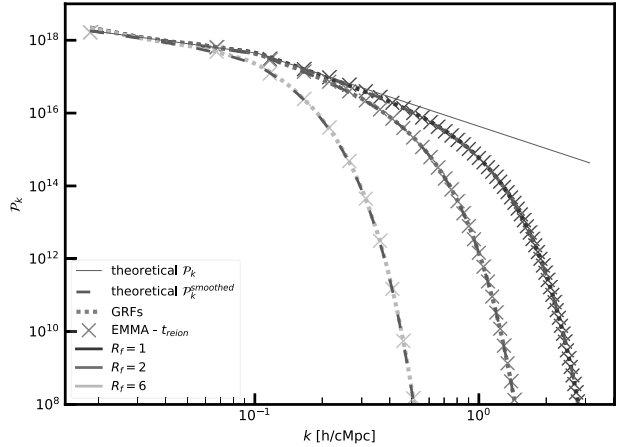
**Notes.** These are obtained from a fitting of the power spectrum of the fields in order to generate GRFs in the same units as the reionisation times (years) or redshift before they are normalised.

The smoothed power spectrum of our GRF is defined as follows:

$$\mathcal{P}_k^{\text{smoothed}} = \begin{cases} A_1 k^{n_1} e^{-2R_f^2 k^2} / 2\pi & k \leq k_{\text{thresh}} \\ A_2 k^{n_2} e^{-2R_f^2 k^2} / 2\pi & k > k_{\text{thresh}} \end{cases}, \quad (10)$$

where  $R_f$  is the standard deviation of the kernel (i.e. size of the kernel) expressed as the number of cells in the following analyses.

We also generate multiple sets of 100 runs (with different seeds) of GRFs with the power spectrum of both EMMA and 21cmFAST  $t_{\text{reion}}(\mathbf{r})$  (for which, the parameters are given in the Table 3). For each power spectrum, three sets of GRFs with different smoothing are created, with the following kernel sizes:  $R_f \in \{1, 2, 6\}$ . In addition, the GRFs are all normalised the same way as the simulations. Figure 3 shows example GRF maps (second row) with different smoothing (see each column): while differences can be spotted, such GRFs are close to the EMMA fields. Figure 5 shows the expected power spectra and those measured in the simulations: the GRF and  $t_{\text{reion}}(\mathbf{r})$  curves are very similar. When we increase the kernel size  $R_f$  (see the purple, blue, and green curves), larger and larger scales are smoothed, as expected.



**Fig. 5.** Power spectra of the GRFs (coloured dotted lines) and EMMA reionisation times (coloured crosses). All 100 runs have been averaged to show a mean power spectrum for each set of simulations. The theoretical power spectra are shown both with and without an incorporated smoothing by the dashed and straight red lines, respectively. Three different smoothings are represented with  $R_f \in \{1, 2, 6\}$ .

### 3. Topological measurements on EMMA simulations and comparisons to GRF theory predictions

In this section, we extract several topological statistics from the EMMA reionisation time field  $t_{\text{reion}}(\mathbf{r})$ . We also derive their expression and compare them to the different runs of GRFs in order to statistically check the behaviour of our theoretical curves and our measurements.

#### 3.1. Filling factor of the field: PDF and reionisation history

##### 3.1.1. Measurements on the reionisation time field

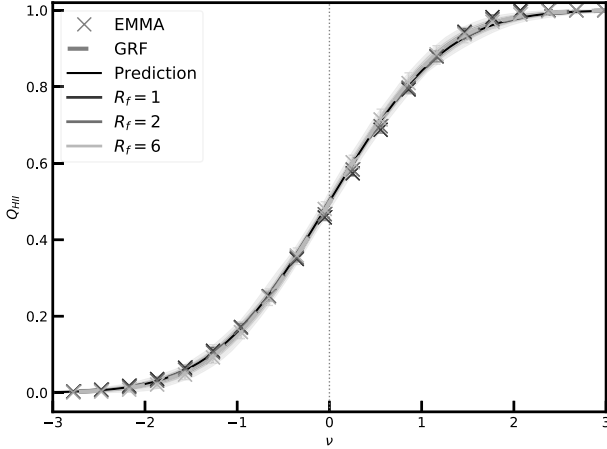
The filling factor of the reionisation time field  $t_{\text{reion}}(\mathbf{r})$  shows the reionisation history (or fraction of ionised volume  $Q_{\text{HII}}$ ) of the simulation box. This quantity allows us to study the global evolution of the ionisation of the gas during the EoR and can be directly extracted from the  $t_{\text{reion}}(\mathbf{r})$  map by counting the number of values lower than a time threshold, which gives us the cumulated PDF of the  $t_{\text{reion}}(\mathbf{r})$  values. The  $t_{\text{reion}}(\mathbf{r})$  PDF tells us about the distribution of reionisation times in the box: it also incorporates reionisation evolution information. If both not-cumulated and cumulated distributions are symmetric with respect to the average reionisation time, this means that the reionisation evolves in the same manner throughout the whole EoR. If they are asymmetric and if the PDF has its peak at a larger time, then reionisation begins slowly before accelerating. On the contrary, if the PDF has its peak at a smaller time, then reionisation starts rapidly before slowing down.

##### 3.1.2. GRF theoretical expression

We can almost directly compute the filling factor of a Gaussian field with the Gaussian field theory because it only requires the PDF of the value of the field (Gay et al. 2012):

$$P(x)dx = \frac{1}{2\pi} e^{-\frac{1}{2}x^2} dx. \quad (11)$$

To calculate the filling factor, the PDF only depend on the normalised field  $x = F/\sigma_0$ . Now, this statistic is the number of field



**Fig. 6.** Fraction of ionised volume for the different smoothings (in colours). The median of every run is computed for each field. The dashed lines correspond to the GRFs, and the crosses are for the EMMA reionisation time field. The black lines are the theoretical predictions. The shaded areas and the error bars represent the dispersion around the median (1st and 99th percentiles) of the GRFs and  $t_{\text{reion}}(\mathbf{r})$ , respectively. Here,  $\nu$  represents the value of the normalised reionisation times.

values exceeding a given threshold  $\nu$ . Applied to our reionisation time fields, the number of values that are higher than a given threshold is the same as the number of cells that are still neutral. This number therefore directly corresponds to the fraction of neutral gas volume  $Q_{\text{HI}}$ . However, in our case, we are interested in the fraction of ionised volume  $Q_{\text{HII}} = 1 - Q_{\text{HI}}$ . This latter corresponds to the number of values that are lower than a given threshold, as follows:

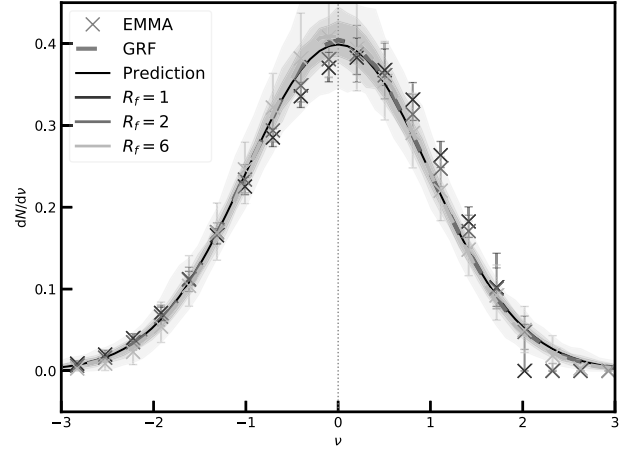
$$Q_{\text{HII}}(\nu) = \int_0^\nu P(x)dx = \frac{1}{2} \operatorname{erf}\left(\frac{\nu}{\sqrt{2}}\right), \quad (12)$$

$$\text{where } \operatorname{erf}(\nu) = \frac{2}{\sqrt{\pi}} \int_0^\nu e^{-y^2} dy.$$

### 3.1.3. Comparison of the measurements and the predictions

We show the filling factors in Fig. 6 and the PDFs in Fig. 7: the crosses (and error bars) are the EMMA measurements, the dashed lines (and the shaded areas) are the GRFs measurements, and the GRF predictions are shown in black. Firstly, in both figures, the GRF distributions closely follow the predictions by the GRF theory. The filling factor measurements on the EMMA reionisation time fields are relatively close to the predictions, depicting a rather symmetric reionisation process. However, in Fig. 7, the PDFs measured on the EMMA  $t_{\text{reion}}(\mathbf{r})$  maps are not symmetric, the peak being shifted towards later times; the filling factors or cumulated PDFs hide this imprint of non-gaussianity. At the same time, when we smooth the fields on larger areas (i.e.  $R_f$  increases), the distributions tend to become more symmetric around the mean reionisation time  $\nu = 0$  (i.e. close to the GRF predictions).

Therefore, the regions that ionise after the average time of the simulation ( $\nu = 0$ , or around 790 Myr or a redshift of 7; see Table 1) cause the asymmetry. This means that the reionisation process is slightly slower at early times and accelerates afterwards. Mellema et al. (2006) or Dixon et al. (2016) showed that asymmetry arises towards the end of the EoR using the brightness temperature field. This asymmetry is also a key parameter



**Fig. 7.** Probability distribution function of the median of every run of the fields for the different smoothings (in colours). The dashed lines correspond to the GRFs, and the crosses are for the EMMA fields. The black lines are the theoretical predictions. The shaded areas and the error bars represent the dispersion around the median (1st and 99th percentiles) of the GRFs and  $t_{\text{reion}}(\mathbf{r})$ , respectively. Here,  $\nu$  represents the value of the normalised reionisation times.

in the newly developed code AMBER (Trac et al. 2022), in which we can directly tune an asymmetry parameter of the reionisation history. We see here that the non-gaussianity of the reionisation fields are filtered out with the smoothing, and are therefore hidden in the time differences on small-scale structures, and at later times.

## 3.2. PDF of the gradient norm field: Ionising front velocities

### 3.2.1. Measurements on the reionisation time field

In this section, we analyse the norm of the spatial gradients of each field that we define as follows, for a GRF or a reionisation time field  $F$ :

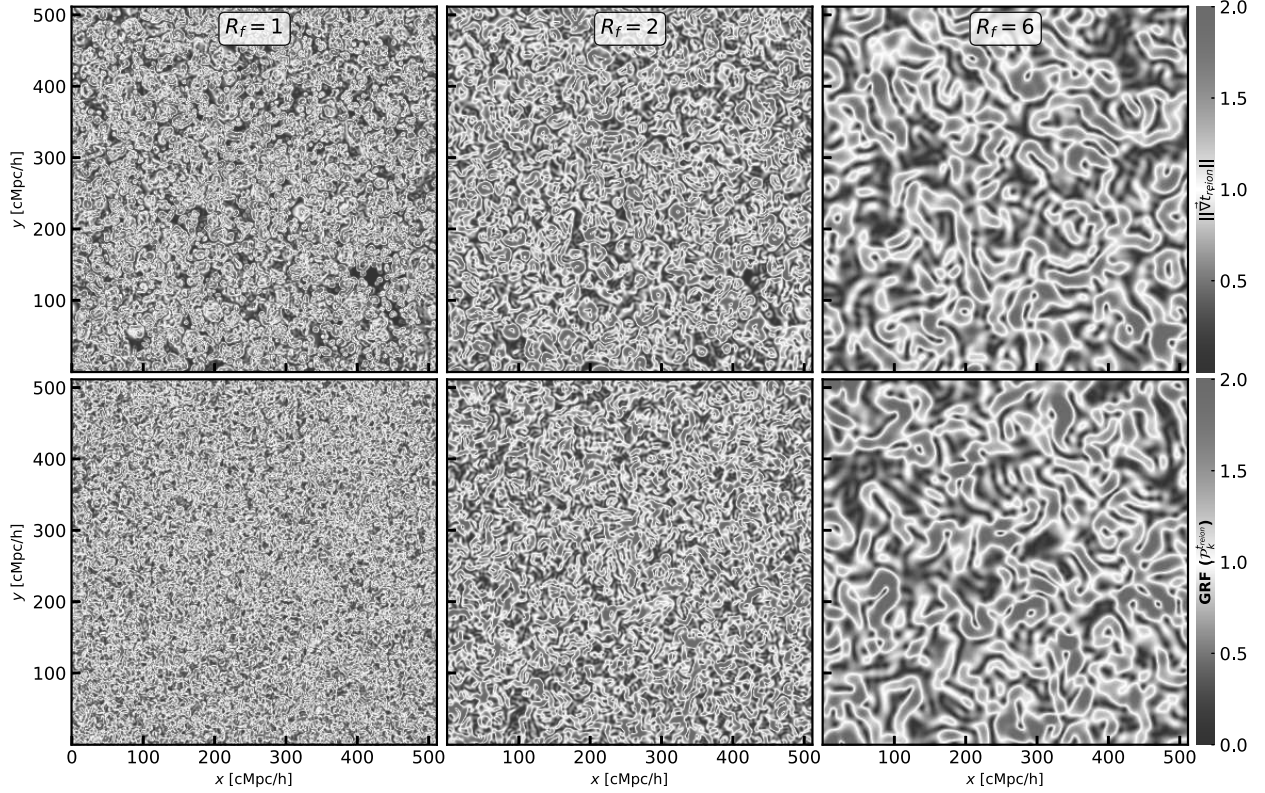
$$\|\nabla x\| = \frac{1}{\sigma_1} \sqrt{(\nabla_1 F)^2 + (\nabla_2 F)^2} = R_0 \sqrt{(\nabla_1 x)^2 + (\nabla_2 x)^2}, \quad (13)$$

where  $\nabla_i x$  for  $i \in \{1, 2\}$  are the two components of the gradient of the field  $x = F/\sigma_0$ , and  $R_0$  is given in Eq. (6). Numerically, the gradients of the fields are computed thanks to Fourier transforms. Each component of the gradient is obtained as follows:

$$\nabla_i x = \mathcal{F}^{-1}[\tilde{x} \times ik_i], \quad (14)$$

where  $\tilde{x} = \mathcal{F}[x]$  is the Fourier transform of the field, and  $k^2 = k_1^2 + k_2^2$ . We should note that here we observe a 3D phenomenon in 2D, which means that the gradient norm is probably underestimated as we miss the third direction component of the front velocities. Figure 8 shows gradient norm maps of the reionisation time fields in the first row, and gradient norm maps of the GRFs in the second row for each smoothing ( $R_f \in \{1, 2, 6\}$ , see the columns). The two maps are relatively similar at first glance for  $R_f \in \{2, 6\}$ . However, some disparities start to be visible for the smallest smoothing kernel  $R_f = 1$  again: the reionisation time field has larger structures than the corresponding GRF.

The gradient norm of  $t_{\text{reion}}(\mathbf{r})$  is linked to the reionisation velocity field defined by Deparis et al. (2019), which is the inverse of the spatial derivative of the reionisation time



**Fig. 8.** Two-dimensional slices of the norm of the EMMA reionisation time gradient (first row) and the norm of the GRF gradient obtained with the corresponding power spectrum (second row). Each column corresponds to different smoothings, with from left to right,  $R_f \in \{1, 2, 6\}$ .

field; it contains information about the ionising front velocity, and Deparis et al. (2019) showed that these fronts move forward in two stages: they are first slowed down by dense neutral gas and their speed is lower than the speed of light. However, when reaching the end of the EoR, the fronts accelerate because radiation reaches under-dense regions. This means that as time increases, the speed of the reionisation fronts increases, or conversely, the gradient norms of the reionisation time field decrease.

### 3.2.2. GRF theoretical expression

The PDF of the gradient norm of a Gaussian field  $\|\nabla F\|$  only depends on the field ( $x = F/\sigma_0$ ) and its first derivatives ( $x_1$  and  $x_2$  as defined in Eq. (3)), as

$$P(x, x_1, x_2) dx dx_1 dx_2 = \frac{2}{(2\pi)^{3/2}} e^{-\left(\frac{1}{2}x^2 + x_1^2 + x_2^2\right)} dx dx_1 dx_2. \quad (15)$$

This joint PDF is obtained relatively easily with Eq. (7) with a three-dimensional covariance matrix (as shown in Appendix C). From this expression, thanks to an integral over the field values and a change of variable, we can retrieve the PDF of the norm of the field gradient:

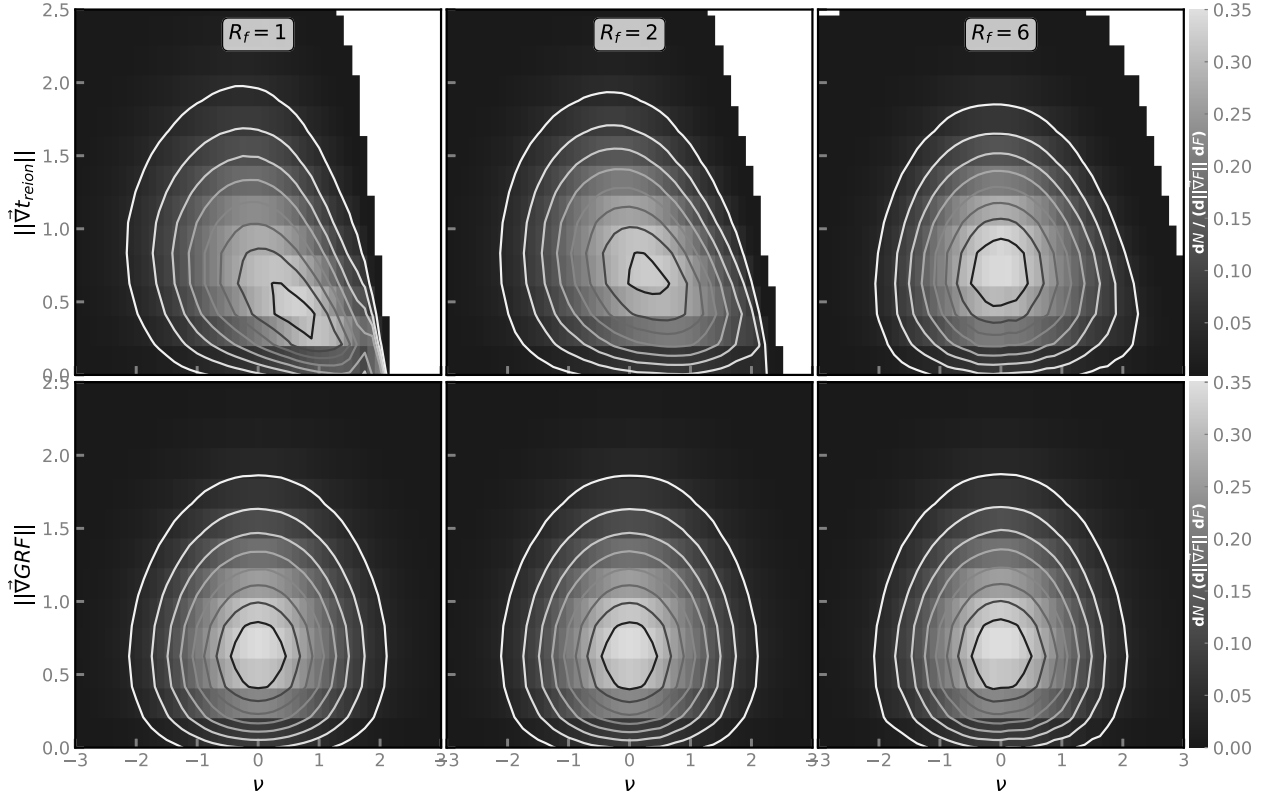
$$2\pi P(w) w dw = 2w e^{-w^2} dw \quad \text{with} \quad w^2 = x_1^2 + x_2^2. \quad (16)$$

### 3.2.3. Comparison of the measurements and the predictions

Firstly, the 2D distributions of the gradient norm of  $t_{\text{reion}}(\mathbf{r})$  with respect to the field values provide insight into the radi-

ation front velocities at each time of the EoR. The first row of Fig. 9 shows these 2D PDFs for the EMMA reionisation time field and their corresponding GRFs are shown in second row. Each smoothing kernel size is represented with  $R_f \in \{1, 2, 6\}$  in the columns. The GRF cases are symmetric around their mean reionisation time  $\nu = 0$ , as expected. We again see the asymmetry mentioned above for the EMMA reionisation time fields: the peak of the distributions is shifted towards the later times and lower velocities, especially for smaller smoothings. This asymmetry enables us to see the acceleration of the ionising fronts as the EoR progresses. If we integrate along the y-axis, we retrieve the PDF of the gradient norm, which is shown in Fig. 10. The dashed lines represent the GRF measurements, the GRF prediction is shown with the black line, and the crosses are for the EMMA reionisation time fields. The GRF measurements are superimposed on the predictions, and the EMMA measurements slightly underestimate the gradients norm because of the acceleration of the radiation fronts at the end of the EoR. On both 2D and 1D distributions, the GRF measurements are independent of the smoothing, as expected given that Eq. (15) does not depend on the kernel size. For the EMMA measurements, increasing the kernel size brings them closer to the predictions.

Our measurements on the EMMA reionisation time field therefore reflect the increase in the ionising front velocities as time increases, which is probably a strong cause of the asymmetry of our reionisation time field distributions, and therefore of the non-gaussianity of the process. In any case, this phenomenon mainly impacts the small scales of  $t_{\text{reion}}(\mathbf{r})$  as it tends to be filtered out with large smoothing.



**Fig. 9.** Two-dimensional PDFs of the gradient norms with respect to the values of the fields of every run for each field and different smoothings ( $R_f \in \{1, 2, 6\}$ , see each column). The first row shows the EMMA reionisation times, and the second row their corresponding GRFs. The grey-scale lines are the isocontours of the histograms. Here,  $\nu$  represents the value of the normalised reionisation times.

### 3.3. Isocontour length: Size evolution of ionised and neutral bubbles

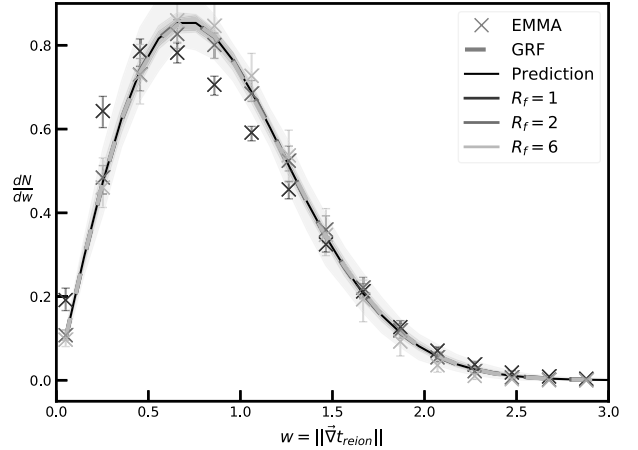
#### 3.3.1. Measurements on the reionisation time field

The isocontours of the reionisation time field tell us how far radiation propagates at a specific time. The first row of Fig. 11 shows the isocontours of the EMMA reionisation time field and the second row shows those of its corresponding GRFs. The three columns correspond to the three smoothings applied to the slices. The bluest contours represent the earliest times, and the reddest ones represent the latest times. The number of contours per level visually decreases as the size of the Gaussian kernel increases (because small structures disappear when increasing  $R_f$ ).

The isocontour length  $\mathcal{L}$  of  $t_{\text{reion}}(\mathbf{r})$  informs us about the extent of a reionisation time level. In Fig. 11, we can see that, as the reionisation time increases, the isocontours first encompass larger and larger regions (blue contours); when the mean reionisation time  $\nu = 0$ , they begin to encompass smaller and smaller regions (red contours). Their length therefore contains information on the size of the ionised and neutral bubbles, on the percolation of the ionised bubbles, and on the different reionisation stages.

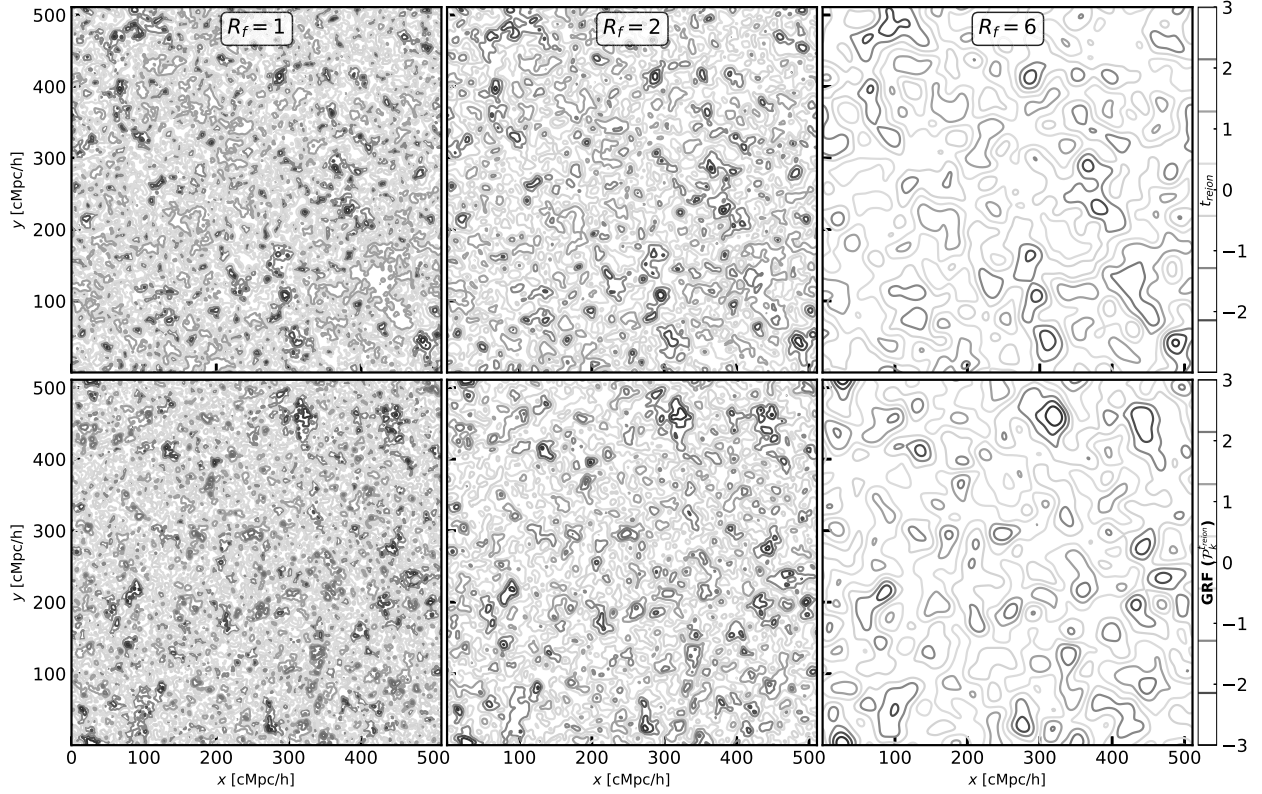
#### 3.3.2. GRF theoretical expression

The average 2D isocontour length allows us to characterise the levels of a field by a measurement of their length; moreover, it is one of the Minkowski functionals (Schmalzing & Gorski 1998;

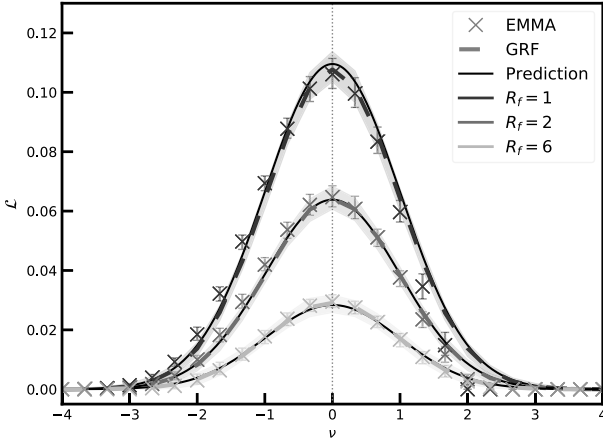


**Fig. 10.** Filling factors of the gradient norms of the fields for the different smoothings (in colours). The median of every run is computed for each field. The dashed lines correspond to the GRFs, and the crosses are for the EMMA reionisation time fields. The black lines are the theoretical predictions. The shaded areas and the error bars represent the dispersion around the median (1st and 99th percentiles) of the GRFs and  $t_{\text{reion}}(\mathbf{r})$ , respectively. Here,  $w$  represents the value of the gradient norm of the reionisation times.

Matsubara 2003), which have often been used to quantify the topology of the EoR (see e.g. Gleser et al. 2006; Lee et al. 2008;



**Fig. 11.** Isocontours of 2D slices of the EMMA reionisation time field (first row) and of the 2D GRF obtained with the corresponding power spectrum (second row). Each column corresponds to a different smoothing, with  $R_f \in \{1, 2, 6\}$  from left to right. All fields are normalised. Eight levels of contours are represented with the colours.



**Fig. 12.** Distribution of the isocontour length for the different smoothings (in colours). The median of every run is computed for each field. The dashed lines correspond to the GRFs, and the crosses show the EMMA reionisation time field. The black lines are the theoretical predictions. The shaded areas and the error bars represent the dispersion around the median (1st and 99th percentiles) of the GRFs and  $t_{\text{reion}}(r)$ , respectively. Here,  $\nu$  represents the value of the normalised reionisation times.

Friedrich et al. 2011; Hong et al. 2014; Yoshiura et al. 2017; Chen et al. 2019; Pathak et al. 2022). The isocontour length at

a level  $\nu$  can be defined as follows (Schmalzing & Gorski 1998; Matsubara 2003; Gay et al. 2012):

$$\mathcal{L}(\nu) = \left\langle \frac{1}{R_0} \delta(x - \nu)(x_1^2 + x_2^2)^{1/2} \right\rangle = \left\langle \frac{1}{R_0} \delta(x - \nu)w \right\rangle, \quad (17)$$

where  $\delta$  is the Dirac delta distribution;  $x$ ,  $x_1$ , and  $x_2$  are the normalised field and first derivatives as defined in Eq. (3);  $w^2 = x_1^2 + x_2^2$ ; and  $R_0$  is defined in Eq. (6) and corresponds to the ratio of the two first spectral moments. This latter appears due to the normalisation of the field and its derivatives.

As  $\mathcal{L}$  depends both on the field and its first derivative, we need to use  $P(x, x_1, x_2)$ , as defined in Eq. (15), in order to calculate it. This probability is expressed as follows:

$$2\pi P(x, w)w dx dw = \frac{2}{\sqrt{2\pi}} e^{-(\frac{1}{2}x^2 + w^2)} w dx dw. \quad (18)$$

Now, we compute the isocontour length by integrating the quantity of Eq. (17) over the field and gradient norm values:

$$\mathcal{L}(\nu) = \int_{-\infty}^{\infty} dx \int_0^{\infty} dw \frac{2\pi}{R_0} P(x, w) \delta(x - \nu) w^2 = \frac{1}{2\sqrt{2}R_0} e^{-\frac{1}{2}\nu^2}. \quad (19)$$

### 3.3.3. Comparison of the measurements and the predictions

The isocontours length  $\mathcal{L}(\nu)$  distributions are shown in Fig. 12 with dashed lines for the GRFs, black lines for the GRF prediction, and crosses for the EMMA reionisation time field  $t_{\text{reion}}(r)$

for the three smoothings. The measurements of the GRF distributions closely follow the GRF prediction curves, as well as the EMMA measurements. The contours grow in length until the mean reionisation time  $\nu = 0$  and then their length decreases. The measurements and predictions vary with the smoothing: the larger the Gaussian kernel, the smaller the lengths of the isocontours. Indeed, with small  $R_f$ , there are many more isocontours per reionisation level (i.e. more small structures), making a longer total length of isocontours than for large  $R_f$ .

As expected, the evolution of isocontour length with time traces the evolution of the ionised bubbles and neutral regions. At the beginning of the EoR, the first ionised bubbles appear because the gas starts to be ionised. These bubbles are seen as the dark blue contours in the bottom-left panel of Fig. 11. As reionisation progresses, the bubbles grow, traced by the larger contours represented by light blue levels in the same figure. The contours increase in length, and sometimes two small dark blue contours merge into one, providing insight into the percolation process. At  $\nu \sim 0$ , the isocontours have reached a maximum length, with the ionised regions intertwined with the neutral regions. Subsequently, the isocontour length  $\mathcal{L}$  decreases while the isocontours start to encompass the last neutral regions, until  $\mathcal{L}$  is close to zero at the moment when there is almost no more neutral gas. The process we mention here is very close to what Chen et al. (2019) found, namely several phases during the reionisation process, starting with an ionised bubble stage followed by an ionised fibre stage where the bubbles merge into long fibres throughout the box. Then, there is the sponge stage, which is the moment when the ionised fibres are intertwined with the neutral fibres. The process ends by a neutral fibre stage and a neutral bubble stage.

With these statistics, the EMMA reionisation times do not show significant imprints of non-gaussianities as their isocontour length follows the GRF predictions, and more so for the largest smoothing.

### 3.4. PDF of minima values: Reionisation seed counts

#### 3.4.1. Measurements on the reionisation time field

In order to compare our simulations to the predicted reionisation seed count, we use the topological code DisPerSE (Sousbie 2011), which allows us to extract the critical points of a field. DisPerSE relies on the Morse theory to get topological information from the fields by studying differentiable functions. It is run on every 2D slice of reionisation time fields, as well as on every GRF generated, with a  $10^{-5} - \sigma$  persistence threshold<sup>4</sup>.

We focus on the minima of the reionisation time field. We call these the ‘seeds’, or equivalently the ‘sources’ of reionisation because these are the first regions where the gas is locally reionised. We can measure the number of minima at a reionisation time (i.e. the PDF of the  $t_{\text{reion}}(\mathbf{r})$  values at its minima). Counting these reionisation seeds informs us about the evolution of the EoR: for example, if the PDF peaks at early times,

<sup>4</sup> The value of the persistence parameter chosen here is very low in order to apply no selection on the extracted critical points. It is a threshold controlling the maximal distance between the field values in a maximum–minimum pair within the extracted features by DisPerSE. The persistence allows us to control the significance of the extracted topological features, and therefore the smoothness of the features. It can be used to override the noise of the input field. More details are given by Sousbie (2011) or Thélie et al. (2022).

the majority of the reionisation seeds appear at the beginning of the EoR, whereas if the PDF is uniformly distributed, then sources steadily contribute to reionisation throughout the whole EoR.

#### 3.4.2. GRF theoretical expression

The PDF of the minima of the reionisation time field can also be derived with the GRF theory. To compute it, we need a joint PDF of the field that is dependent on the field, and on its first and second derivatives. Indeed, the critical points correspond to the zeros of the first derivative, and the sign of the second derivative gives us information on the type of critical point (it is positive for minima). The calculation is in six dimensions in 2D and requires changes of variables to make the covariance matrix diagonal:

$$u = -(x_{11} + x_{22}), \quad v = \frac{1}{2}(x_{11} - x_{22}), \quad \zeta = \frac{x - \gamma u}{\sqrt{1 - \gamma^2}}, \quad (20)$$

where  $\gamma = \sigma_1^2/(\sigma_0\sigma_2)$ , and  $x$ ,  $x_{11}$ , and  $x_{22}$  correspond to the dimensionless field and derivatives as defined in Eq. (3). The full calculation is presented in Gay (2011) and the resulting PDF is the following:

$$P(\zeta, x_1, x_2, u, v, x_{12}) = \frac{16}{(2\pi)^3} e^{-Q(\zeta, x_1, x_2, u, v, x_{12})}, \quad (21)$$

with

$$Q(\zeta, x_1, x_2, u, v, x_{12}) = \frac{1}{2}\zeta^2 + x_1^2 + x_2^2 + \frac{1}{2}u^2 + 4v^2 + 4x_{12}^2, \quad (22)$$

where  $x_1$ ,  $x_2$ , and  $x_{12}$  are dimensionless derivatives of the field as defined in Eq. (3).

The average extrema density is given by the following expression (which is explained in Gay 2011):

$$\frac{\partial n_{\text{ext}}}{\partial v} = \left\langle \frac{1}{R_*^2} |x_{11}x_{22} - x_{12}^2| \delta(x_1)\delta(x_2)\delta(x - v) \right\rangle. \quad (23)$$

The non-trivial part of the minima distribution calculation is the 6D integration involved in Eq. (23). A version of this integration for 3D fields is detailed in Appendix A of Bardeen et al. (1986). Additional constraints (on the eigenvalues of the hessian matrix of the field) are required to make Eq. (23) a distribution of minima. Here, we only give the resulting distribution of minima:

$$\begin{aligned} \frac{\partial n_{\text{min}}}{\partial v} = & \frac{\exp\left(\frac{-v^2}{2}\right)}{\sqrt{2\pi}R_*^2} \left[ 1 - \operatorname{erf}\left(\frac{\gamma v}{\sqrt{2(1-\gamma^2)}}\right) \right] K_1(v, \gamma) \\ & + \frac{\exp\left(\frac{-3v^2}{6-4\gamma^2}\right)}{\sqrt{2\pi(1-\frac{2}{3}\gamma^2)R_*^2} \left[ 1 - \operatorname{erf}\left(\frac{\gamma v}{\sqrt{2(1-\gamma^2)(3-2\gamma^2)}}\right) \right]} K_2 \\ & - \frac{\exp\left(\frac{-v^2}{2(1-\gamma^2)}\right)}{\sqrt{2\pi(1-\gamma^2)R_*^2} \left[ 1 + \exp\left(\frac{v^2}{2(1-\gamma^2)}\right) \right]} K_3(v, \gamma), \end{aligned} \quad (24)$$

where  $R_*$  is defined in Eq. (6), and

$$K_1(v, \gamma) = \frac{\gamma^2(v^2 - 1)}{8\pi}, \quad K_2 = \frac{1}{8\pi\sqrt{3}}, \quad \text{and} \quad K_3(v, \gamma) = \frac{\gamma(1-\gamma^2)v}{2(2\pi)^{3/2}}. \quad (25)$$

### 3.4.3. Comparison of the measurements and the predictions

The PDFs of the minima values are shown in Fig. 13, with dashed lines showing the GRFs, full black lines the predictions, and crosses the EMMA reionisation time field. The three smoothings are represented (see the three  $R_f$  values in different colours). The theoretical PDFs are centred around early reionisation times (i.e.  $\nu < 0$ ) because we look for minima. The GRFs and EMMA  $t_{\text{reion}}(\mathbf{r})$  fields closely follow the expected curves. When the fields are smoothed with a larger kernel (increasing  $R_f$ ), the least significant minima get smoothed out, decreasing the number of counted minima. We also note that smoothing on larger areas causes the distributions to be shifted towards smaller values of the field (i.e. towards the beginning of the EoR).

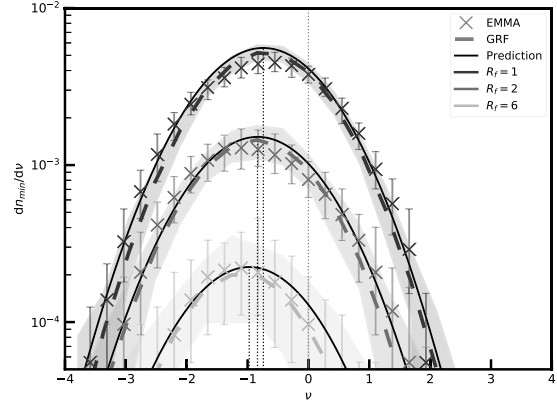
With these results, we can see that when the EoR begins (smallest  $\nu$ ), there are a few reionisation seeds that represent the first radiation sources of the Universe. The further the reionisation progresses, the more the number of reionisation seeds increases as more and more galaxies are created. This number continues to increase until some intermediate time, where the minima distributions peak, which happens before the average reionisation time at the mean reionisation time  $\nu = 0$ . The distributions then decrease because more and more intergalactic gas is reionised by the already existing radiation sources, until it is all reionised. At this time, no more new seeds of ionisation front propagation appear, although there can be new sources in already reionised places. The impact of the smoothing is that only the most exceptional and early reionisation seeds remain in the smoothed fields – because they are sufficiently significant –, while the other ones are filtered out. This means that these exceptional seeds reionise the other ones in an inside-out way. Globally, there is little to no imprint of non-gaussianities with this statistic within the error bars.

## 3.5. Skeleton length: Regions of ionising front percolation

### 3.5.1. Measurements on the reionisation time field

Thanks to DisPerSE, we can also extract the skeleton of  $t_{\text{reion}}(\mathbf{r})$  from every reionisation time field and GRF using the same persistence value as in Sect. 3.4. The skeleton is, as explained in Sect. 1, the network formed by all the segments joining the maxima and saddle points together along the ridge of the field. In 2D, the skeleton corresponds to the edges of the reionisation patches, which are regions under the radiative influence of a reionisation seed. These edges, or the place where the patches intersect, are also the front lines between the radiation fronts of reionisation. Looking at the skeleton (in black) in the top right panel of Fig. 2, we can see that it seems to be preferentially aligned diagonally on the map, which is due to diamond shapes produced around sources that are caused by the M1 approximation used in EMMA to model the radiative transfer (Aubert & Teyssier 2008).

We are interested in the length distribution of the skeleton as a function of time. As an example, if the distributions are as narrow as a Dirac distribution, then the reionisation seeds are uniformly distributed in space, which in turn means that all the radiation fronts encounter opposite fronts all at the same time. If the Dirac distribution peaks at  $\nu < 0$ , then the percolation happens at early times, and if it peaks at  $\nu > 0$ , then the percolation happens at late times. Otherwise, if the distribution is wider, the seeds are not uniformly distributed and the merger of the reionisation patches happens throughout the reionisation process. If there were many ionised bubbles at early times and then a single growing bubble, then the distribution would be asymmetric. As the skeleton is the place where radiation fronts meet, it is



**Fig. 13.** Distribution of the critical points of the fields for the different smoothings (in colours). The median of every run is computed for each field. The dashed lines correspond to the GRFs, and the crosses are for the EMMA reionisation time fields. The black lines are the theoretical predictions. The shaded areas and the error bars represent the dispersion around the median (1st and 99th percentiles) of the GRFs and  $t_{\text{reion}}(\mathbf{r})$ , respectively. The black dotted vertical lines represent the average of the predictions. Here,  $\nu$  represents the value of the normalised reionisation times.

impacted by the front velocities: if the fronts propagate faster, they can reach more-distant regions, meaning that the reionisation patches are larger. If one assumes a simple 2D lattice of circular patches of radius  $R$ , it is expected to have a total skeleton length per unit surface of  $L \sim 2\pi R/\pi R^2 \sim R^{-1}$ ; a scenario with large patches should lead to small  $L$  values. Also, the accelerated fronts can cause the percolation to happen more rapidly in the remaining neutral regions, causing the distribution to decrease more sharply at the end of the EoR, and to therefore be asymmetric with respect to time. Generally speaking, these distributions tell us whether the ionising fronts percolate in a longer or shorter period of time.

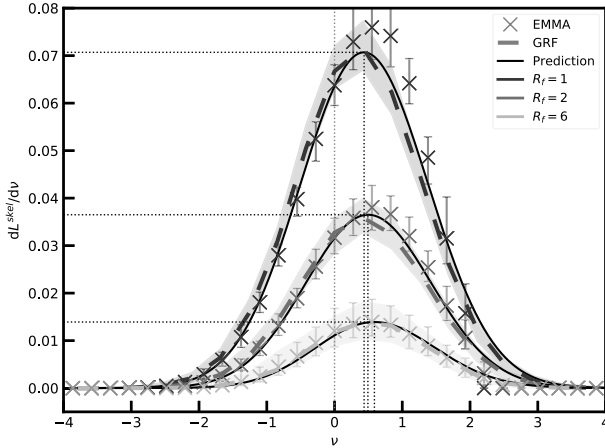
### 3.5.2. GRF theoretical expression

The analytical calculation of the skeleton-length distribution has been described by several authors, such as Pogosyan et al. (2009b), Gay (2011), Gay et al. (2012). The calculations are presented in detail by Pogosyan et al. (2009b) for example. In summary, the skeleton corresponds to all the points for which the gradient is aligned with an eigenvector of the hessian of the field,  $\mathcal{H} = \nabla\nabla F$ , following the gradient in the direction of the positive eigenvalue of the hessian<sup>5</sup>. This definition can be mathematically written as follows:  $\mathcal{H} \cdot \nabla F = \lambda \nabla F$ , where  $\lambda$  is the positive eigenvalue of the hessian. In an equivalent way, a point is on a critical line if  $s = \det(\mathcal{H} \cdot \nabla x, \lambda \nabla x) = 0$ , with the dimensionless field  $x (= F/\sigma_0)$ . The skeleton-length distribution can therefore be written as follows:

$$\frac{\partial L^{\text{skel}}}{\partial \nu} = \left\langle \frac{1}{R_*} \delta(s) |\nabla s| \delta(x - \nu) \right\rangle, \quad (26)$$

<sup>5</sup> Returning to our representation of the topology through a mountain landscape, the skeleton, and therefore the ridge lines, are referred to as critical lines in topology. Walking on a ridge line is coming from a pass and going in the direction that goes up to the peak: this means that we follow the gradient in the direction of the positive eigenvalue of the hessian. In addition, we note that, if we were to follow the gradient in the direction of the negative eigenvalue of the hessian, we would be on the anti-skeleton, aiming at reaching a minima.





**Fig. 14.** Distribution of the skeleton length of the fields for the different smoothings (in colours). The median of every run is computed for each field. The dashed lines correspond to the GRFs, and the crosses are for the EMMA reionisation time fields. The black lines are the theoretical predictions. The shaded areas and the error bars represent the dispersion around the median (1st and 99th percentiles) of the GRFs and  $t_{\text{reion}}(r)$ , respectively. The black dotted vertical lines represent the average of the predictions. Here,  $\nu$  represents the value of the normalised reionisation times.

with  $|\nabla s|$  giving the length of the critical lines, and  $R_*$  given in Eq. (6). The integration of Eq. (26) involves the PDF of the field, and its first, second, and third derivative. The ‘stiff approximation’ allows us to neglect the derivatives of higher than third order, which simplifies the calculations. The critical lines are therefore considered to be relatively straight: the total length of the skeleton may therefore probably be reduced with this approximation. The final expression of the skeleton length per unit of the total field surface is therefore

$$\frac{\partial L^{\text{skel}}}{\partial \nu} = \frac{1}{\sqrt{2\pi}R_*} \exp^{-\frac{\nu^2}{2}} \left[ \frac{1}{8\sqrt{\pi}} (\sqrt{\pi} + 2\gamma\nu) \left( 1 + \text{erf} \left( \frac{\gamma\nu}{\sqrt{2(1-\gamma^2)}} \right) \right) + \frac{\sqrt{1-\gamma^2}}{2\sqrt{2\pi}} \exp \left( -\frac{\gamma^2\nu^2}{2(1-\gamma^2)} \right) \right], \quad (27)$$

with  $\gamma$  defined in Eq. (6).

The total skeleton length of a field can be obtained by integrating the distribution of Eq. (27) over all values of  $\nu$ . The resulting expression is (Pogosyan et al. 2009b; Gay 2011)

$$L^{\text{tot}} = \left( \frac{1}{8} + \frac{\sqrt{2}}{4\pi} \right) \frac{1}{R_*}. \quad (28)$$

### 3.5.3. Comparison of the measurements and the predictions

The skeleton-length PDFs are shown in Fig. 14 with dashed lines for the GRFs, full black lines for the predictions, and crosses for the EMMA reionisation time field. The colours represent the three different smoothings (see the three  $R_f$  values in different colours). The ‘stiff approximation’ leads to discrepancies between the measurements and the predictions. In addition, the analytical calculation of the skeleton length is local and DisPerSE gives a global skeleton (Pogosyan et al. 2009b; Gay et al. 2012), meaning that the predictions tend to underestimate the skeleton length by probably missing filaments.

For these reasons, we multiply our measurements by a normalisation factor<sup>6</sup> to match the prediction amplitude following Gay et al. (2012). Still, the shape of the distributions seems to be conserved, as the renormalised measurements on the GRFs are almost perfectly superimposed on the predicted curves. The measurements on the reionisation time fields from EMMA are also relatively close to the predictions, especially when increasing  $R_f$ . The PDFs increase until reaching a maximum after the average reionisation time ( $\nu > 0$ ), before decreasing again. They are not centred on the average reionisation time  $\nu = 0$ ; instead they are shifted to later times when the smoothings increase, and are asymmetric, except for the highest smoothings ( $R_f = 6$ ), for which the measured EMMA distributions are ‘symmetrised’ thanks to the smoothing, and therefore closer to the predictions.

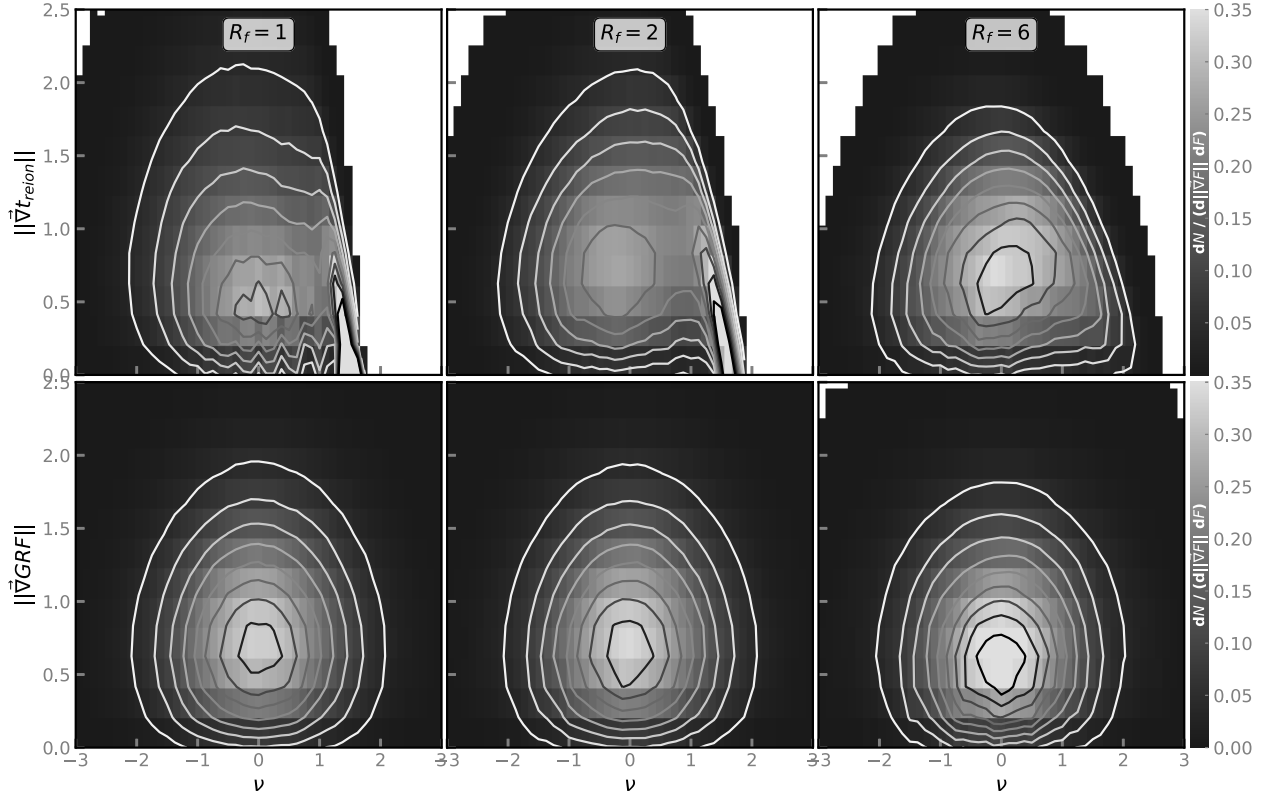
With this statistic, we can again follow the evolution of the EoR from the point of view of merging radiation fronts. During the first stages of the EoR, the ionised bubbles grow, and the radiation fronts start to reach farther out, increasing the skeleton length. At a reionisation time of  $\nu > 0$ , the distribution peaks when the ionising fronts percolate on the longer length; at this moment there are large ionised fibres. Subsequently, most gas is reionised, and there is less percolation of ionised bubbles, meaning that fewer ionised fronts encounter other fronts, until the gas is totally ionised at the end of the EoR ( $\nu \sim 3$  as in the reionisation history). With small smoothing kernel sizes ( $R_f \in \{1, 2\}$ ), we retrieve the asymmetry that results from the increase in the radiation front velocities at the end of the EoR, as expected.

## 4. Predictions from Gaussian random field theories compared to 21cmFAST simulations

In this section, we compare the theoretical statistics described in Sect. 3 to the 21cmFAST reionisation time field. For the sake of brevity, we only show the 2D histograms of the field values and its gradient norms, the fraction of the ionised volume of gas, and the skeleton length distribution for the 21cmFAST reionisation time fields and the GRFs generated with the corresponding power spectrum. Indeed, the 21cmFAST reionisation fields give similar results to the EMMA ones: the semi-analytical generated field is generally close to the GRF predictions. This latter field depicts an almost Gaussian behaviour in every statistic and smoothing studied in this paper, as for the EMMA simulation, albeit slightly less Gaussian for the minima and skeleton length distributions. Again, increasing the Gaussian kernel size (i.e. increasing  $R_f$ ) tends to ‘gaussianise’ the reionisation time field.

The first row of Fig. 15 shows the 2D histograms for  $t_{\text{reion}}(r)$  and the second row shows the GRFs. Each column represents a different Gaussian smoothing with  $R_f \in \{1, 2, 6\}$ , respectively. Figure 16 shows the fraction of ionised volume, and Fig. 17 shows the PDF of the skeleton length. Both figures show the GRF predictions in black, the 21cmFAST reionisation time fields in coloured crosses, and the GRF measurements in coloured dashed lines. In the 21cmFAST field, there are more imprints of non-gaussianities: the gradient norms decrease dramatically at higher times (i.e. the front velocities increase dramatically). This decrease was already the case in the EMMA simulation, but it is much more pronounced in the present case. For the two smallest smoothings, this non-gaussianity imprint also appears in the  $Q_{\text{HII}}$  statistic (see Fig. 16), where the gas seems to be totally reionised

<sup>6</sup> The normalisation factor is given by the ratio between the total length of the measured skeleton over the one of the prediction. It depends on the smoothing via the spectral parameter  $R_*$  (see Eq. (28)).



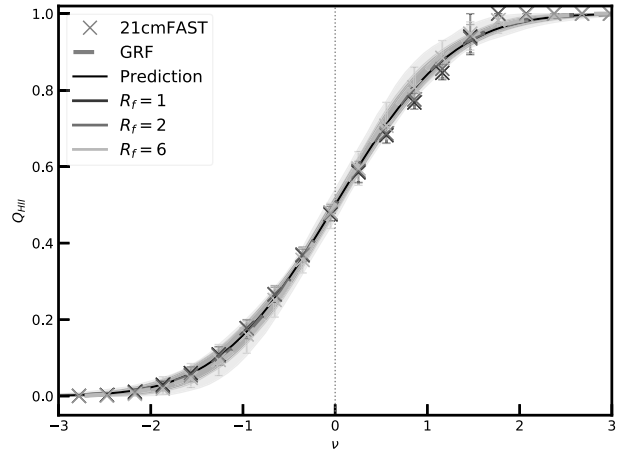
**Fig. 15.** Two-dimensional PDFs of the gradient norms with respect to the values of the fields of every run for each field and different smoothings ( $R_f \in \{1, 2, 6\}$ , see each column). The first row corresponds to the 21cmFAST reionisation times, and the second row to their corresponding GRFs. The grey-scale lines are the isocontours of the histograms. Here,  $\nu$  represents the value of the normalised reionisation times.

earlier than in the predictions or in the skeleton length PDFs (see Fig. 17), where at later times, the distributions depart from the predictions. Moreover, 21cmFAST does not explicitly model radiation propagation (Zahn et al. 2011), which means that the front velocities are not limited by the speed of light, for example. This is not the case in EMMA, and could explain the small gradient values at late times.

Nevertheless, even if the 21cmFAST measurements of Figs. 16 and 17 are not superimposed on the predictions, they remain within the measurement error bars. For the two smallest smoothing kernels, the 21cmFAST skeleton length distributions (Fig. 17) peak at a lower skeleton length value than the EMMA distributions (Fig. 14). This means that at a given time, the skeleton lengths are smaller in the 21cmFAST simulation than in the EMMA simulation: there are less percolation places in the 21cmFAST simulation than in the EMMA simulation at a given moment. As already mentioned, a shorter skeleton length implies larger patches for a simple lattice model; this would be consistent with the results obtained in Thélie et al. (2022), where 21cmFAST patches were usually found to be larger than the ones found in the EMMA simulation for similar models.

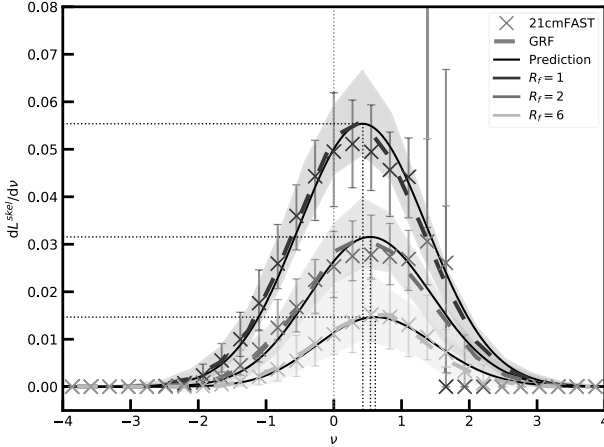
## 5. Conclusion and perspectives

In this work, we extract topological statistics from 2D reionisation time (and redshift in Appendix D) maps produced based on an EMMA cosmological simulation and on a 21cmFAST semi-analytical simulation (which have approximately the same reion-



**Fig. 16.** Fraction of ionised volume of the fields for the different smoothings (in colours). The median of every run is computed for each field. The dashed lines correspond to the GRFs, and the crosses are for the 21cmFAST reionisation time fields. The black lines are the theoretical predictions. The shaded areas and the error bars represent the dispersion around the median (1st and 99th percentiles) of the GRFs and  $t_{reion}(r)$ , respectively. Here,  $\nu$  represents the value of the normalised reionisation times.

isation history). Reionisation time maps contain a wealth of spatial and temporal information about the reionisation process. The



**Fig. 17.** Distribution of the skeleton length of the fields for the different smoothings (in colours). The median of every run is computed for each field. The dashed lines correspond to the GRFs, and the crosses are for the 21cmFAST reionisation time fields. The black lines are the theoretical predictions. The shaded areas and the error bars represent the dispersion around the median (1st and 99th percentiles) of the GRFs and  $t_{\text{reion}}(\mathbf{r})$ , respectively. The black dotted vertical lines represent the average of the predictions. Here,  $v$  represents the value of the normalised reionisation times.

fraction of ionised volume (i.e. the filling factor of the  $t_{\text{reion}}(\mathbf{r})$  map) contains information on the global timing and evolution of the reionisation process. The PDF of the gradient norm map informs us about the velocity of the radiation fronts. The average isocontour length of the field allows us to follow the percolation process. The critical point distributions inform us about the timing of the appearance of reionisation seeds. The skeleton lengths tell us about the moment, duration, and place of the percolation of ionisation fronts. We also apply the GRF theory (Rice 1944; Longuet-Higgins 1957; Doroshkevich 1970; Bardeen et al. 1986; Gay et al. 2012) in the context of the EoR to compare GRF predictions to measurements of these statistics in simulations. We generate GRFs from a fitted power spectrum of each simulation to check our simulation measurements.

We show that the topological statistics extracted from the EMMA and 21cmFAST reionisation time maps are relatively close to the GRF predictions, and are even closer when the maps are smoothed on larger areas. This means that  $t_{\text{reion}}(\mathbf{r})$  can be assumed to be Gaussian with a good level of approximation, and that we have therefore developed a simple tool that allows us to quickly generate fields related to reionisation. This result is surprising in a context where many other EoR fields have been shown to be highly non-Gaussian, such as the 21 cm or the density fields (see e.g. Mellema et al. 2006; Iliev et al. 2006; Majumdar et al. 2018; Ross et al. 2019). We find the major differences between the EMMA cosmological simulation and the 21cmFAST semi-analytical simulation reionisation time fields to be caused by the increase in the velocity of the fronts at the end of the EoR.

The topological statistics applied to the reionisation time field can therefore be used to characterise the evolution of the EoR. The reasonable agreement between GRF predictions and model measurements also suggests that it may be possible to generate histories of reionisation on the sky from the simple knowledge of the power spectrum of the reionisation time field. Such generated histories would automatically

come with a set of topological statistics (number of reionisation seeds, skeleton length, Minkowski functionals, etc.) fully determined by the power spectrum within the framework of GRF theory. In addition, we show here that the reionisation evolution can be inferred from the power-spectrum parameters (or the spectral parameters  $R_0$ ,  $R_*$ , and  $\gamma$ ) only, as long as the scales are sufficiently large for the reionisation time field to be close to a GRF. Finally, the topological statistics discussed here directly depend on the power-spectrum parameters (amplitudes, slopes, characteristic scales) in the GRF approximation. The physics of the propagation of reionisation, which is presumably encoded in the power spectrum, can be constrained even in situations where the power spectrum cannot be easily estimated, for example by fitting peaks, isocontours, or skeleton statistics with their Gaussian predictions. As such, these statistics can be used to constrain the power spectrum, even in situations where the reionisation time fields suffer from noise or poor resolution, for example.

However, our studies show that this similarity with GRF predictions operates on large scales of about 8 cMpc/h, which is similar to the SKA resolution at these redshifts. We still see small imprints of non-gaussianities on the smaller scales. Indeed, at the end of the EoR, the radiation fronts propagate increasingly quickly because of the remaining neutral voids. This velocity increase makes the process asymmetric with respect to the mean reionisation time, and it is poorly reconstructed with the symmetric theory that is the GRF theory. As the regions that remain to be ionised decrease in size as the EoR comes to an end, this phenomenon remains at small scales, and the velocity increase gets smoothed out with the largest smoothing. To take these asymmetries into account, we could add non-Gaussian terms in our expressions with the Gram-Charlier expansion (Gay et al. 2012; Cadiou et al. 2020). Also, it could be interesting to investigate how reduced-speed-of-light approximations (Deparis et al. 2019; Ocvirk et al. 2019) influence the statistics presented here and could lead to an even better agreement with GRF predictions. In any case, our results are probably resolution-dependent and we could verify this with models of higher resolution.

Finally, the reionisation time or redshift fields are not directly observable. In the next decade, the SKA observatory will provide 21 maps of the EoR that will be of similar resolution to our simulation smoothed with  $R_f = 6$ . For this reason, we are creating a method to be presented in a future paper (Hiegel et al., in prep.) to reconstruct 2D reionisation redshift maps from 2D 21 cm maps (that are taken at a given redshift), and from which we can compute the reionisation time maps. With these maps, we will be able to infer the topological characteristics of the reionisation process as we do here with simulations.

*Acknowledgements.* We thank Christophe Pichon and Corentin Cadiou for their help and advice on the understanding of the GRF theory. This work was granted access to the HPC resources of CINES and TGCC under the allocations 2020-A0070411049, 2021-A0090411049, and 2022-A0110411049 “Simulation des signaux et processus de l’aube cosmique et Réionisation de l’Univers” made by GENCI. This research made use of TOOLS21CM, a community-developed package for the analysis of the 21 cm signals from the EoR and Cosmic Dawn (Giri et al. 2020), ASTROPY, a community-developed core Python package for astronomy (Astropy Collaboration 2018); MATPLOTLIB, a Python library for publication quality graphics (Hunter 2007); SCIPY, a Pythonbased ecosystem of open-source software for mathematics, science, and engineering (Virtanen et al. 2020); NUMPY (Harris et al. 2020) and IPYTHON (Perez & Granger 2007).

## References

- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, AJ, 156, 123  
 Aubert, D., & Teyssier, R. 2008, MNRAS, 387, 295  
 Aubert, D., Deparis, N., & Ocvirk, P. 2015, MNRAS, 454, 1012

- Aubert, D., Deparis, N., Ocvirk, P., et al. 2018, *ApJ*, 856, L22
- Banet, A., Barkana, R., Fialkov, A., & Guttman, O. 2021, *MNRAS*, 503, 1221
- Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, *ApJ*, 304, 15
- Barkana, R., & Loeb, A. 2001, *Phys. Rep.*, 349, 125
- Battaglia, N., Trac, H., Cen, R., & Loeb, A. 2013, *ApJ*, 776, 81
- Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, *Phys. Rep.*, 367, 1
- Bianco, M., Giri, S. K., Iliev, I. T., & Mellema, G. 2021, *MNRAS*, 505, 3982
- Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2006, *ApJ*, 638, 20
- Busch, P., Eide, M. B., Ciardi, B., & Kakiichi, K. 2020, *MNRAS*, 498, 4533
- Cadiou, C., Pichon, C., Codis, S., et al. 2020, *MNRAS*, 496, 4787
- Chen, Z., Xu, Y., Wang, Y., & Chen, X. 2019, *ApJ*, 885, 23
- Dayal, P., & Ferrara, A. 2018, *Phys. Rep.*, 780, 1
- Deparis, N., Aubert, D., Ocvirk, P., Chardin, J., & Lewis, J. 2019, *A&A*, 622, A142
- Dixon, K. L., Iliev, I. T., Mellema, G., Ahn, K., & Shapiro, P. R. 2016, *MNRAS*, 456, 3011
- Doroshkevich, A. G. 1970, *Astrophysics*, 6, 320
- Elbers, W., & van de Weygaert, R. 2023, *MNRAS*, 520, 2709
- Friedrich, M. M., Mellema, G., Alvarez, M. A., Shapiro, P. R., & Iliev, I. T. 2011, *MNRAS*, 413, 1353
- Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2004, *ApJ*, 613, 16
- Gay, C. 2011, Ph.D. Thesis, thèse de doctorat en Physique dirigée par Christophe Pichon, Paris
- Gay, C., Pichon, C., & Pogosyan, D. 2012, *Phys. Rev. D*, 85, 023011
- Gazagnes, S., Koopmans, L. V. E., & Wilkinson, M. H. F. 2021, *MNRAS*, 502, 1816
- Gillet, N. J. F., Aubert, D., Mertens, F. G., & Ocvirk, P. 2021, *ArXiv e-prints* [arXiv:2103.03061]
- Giri, S. K., & Mellema, G. 2021, *MNRAS*, 505, 1863
- Giri, S. K., Mellema, G., Dixon, K. L., & Iliev, I. T. 2018a, *MNRAS*, 473, 2949
- Giri, S. K., Mellema, G., & Ghara, R. 2018b, *MNRAS*, 479, 5596
- Giri, S. K., Mellema, G., Aldheimer, T., Dixon, K. L., & Iliev, I. T. 2019, *MNRAS*, 489, 1590
- Giri, S., Mellema, G., & Jensen, H. 2020, *J. Open Source Softw.*, 5, 2363
- Gleser, L., Nusser, A., Ciardi, B., & Desjacques, V. 2006, *MNRAS*, 370, 1329
- Gorce, A., & Pritchard, J. R. 2019, *MNRAS*, 489, 1321
- Greig, B., & Mesinger, A. 2015, *MNRAS*, 449, 4246
- Greig, B., & Mesinger, A. 2017, *MNRAS*, 472, 2651
- Greig, B., Ting, Y.-S., & Kaurov, A. A. 2022, *MNRAS*, 513, 1719
- Hamilton, A. J. S., Gott, J., Richard, I., & Weinberg, D. 1986, *ApJ*, 309, 1
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, 585, 357
- Hong, S. E., Ahn, K., Park, C., et al. 2014, *J. Korean Astron. Soc.*, 47, 49
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, 9, 90
- Hutter, A., Watkinson, C. A., Seiler, J., et al. 2020, *MNRAS*, 492, 653
- Ichikawa, K., Barkana, R., Iliev, I. T., Mellema, G., & Shapiro, P. R. 2010, *MNRAS*, 406, 2521
- Iliev, I. T., Mellema, G., Pen, U. L., et al. 2006, *MNRAS*, 369, 1625
- Iliev, I. T., Mellema, G., Shapiro, P. R., et al. 2012, *MNRAS*, 423, 2222
- Kakiichi, K., Majumdar, S., Mellema, G., et al. 2017, *MNRAS*, 471, 1936
- Kapahtia, A., Chingangbam, P., Appleby, S., & Park, C. 2018, *J. Cosmol. Astropart. Phys.*, 2018, 011
- Kapahtia, A., Chingangbam, P., & Appleby, S. 2019, *J. Cosmol. Astropart. Phys.*, 2019, 053
- Kapahtia, A., Chingangbam, P., Ghara, R., Appleby, S., & Choudhury, T. R. 2021, *J. Cosmol. Astropart. Phys.*, 2021, 026
- Kern, N. S., Liu, A., Parsons, A. R., Mesinger, A., & Greig, B. 2017, *ApJ*, 848, 23
- Koopmans, L., Pritchard, J., Mellema, G., et al. 2015, *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 1
- Kulkarni, G., Keating, L. C., Haehnelt, M. G., et al. 2019, *MNRAS*, 485, L24
- Lee, K.-G., Cen, R., Gott, J., Richard, I., & Trac, H. 2008, *ApJ*, 675, 8
- Lidz, A., Zahn, O., McQuinn, M., Zaldarriaga, M., & Hernquist, L. 2008, *ApJ*, 680, 962
- Lin, Y., Oh, S. P., Furlanetto, S. R., & Sutter, P. M. 2016, *MNRAS*, 461, 3361
- Liu, A., & Parsons, A. R. 2016, *MNRAS*, 457, 1864
- Longuet-Higgins, M. S. 1957, *Philos. Trans. R. Soc. London Ser. A Math. Phys. Sci.*, 249, 321
- Majumdar, S., Pritchard, J. R., Mondal, R., et al. 2018, *MNRAS*, 476, 4007
- Matsubara, T. 2003, *ApJ*, 584, 1
- Matsubara, T., & Kuriki, S. 2021, *Phys. Rev. D*, 104, 103522
- McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., & Furlanetto, S. R. 2006, *ApJ*, 653, 815
- Mellema, G., Iliev, I. T., Pen, U.-L., & Shapiro, P. R. 2006, *MNRAS*, 372, 679
- Mellema, G., Koopmans, L. V. E., Abdalla, F. A., et al. 2013, *Exp. Astron.*, 36, 235
- Mesinger, A., & Furlanetto, S. 2007, *ApJ*, 669, 663
- Mesinger, A., Furlanetto, S., & Cen, R. 2011, *MNRAS*, 411, 955
- Mesinger, A., Ferrara, A., & Spiegel, D. S. 2013, *MNRAS*, 431, 621
- Murray, S., Greig, B., Mesinger, A., et al. 2020, *J. Open Source Softw.*, 5, 2582
- Ocvirk, P., & Aubert, D. 2011, *MNRAS*, 417, L93
- Ocvirk, P., Aubert, D., Chardin, J., Deparis, N., & Lewis, J. 2019, *A&A*, 626, A77
- Ocvirk, P., Aubert, D., Sorce, J. G., et al. 2020, *MNRAS*, 496, 4087
- Pagano, M., & Liu, A. 2020, *MNRAS*, 498, 373
- Park, J., Mesinger, A., Greig, B., & Gillet, N. 2019, *MNRAS*, 484, 933
- Pathak, A., Bag, S., Dasgupta, S., et al. 2022, *J. Cosmol. Astropart. Phys.*, 2022, 027
- Perez, F., & Granger, B. E. 2007, *Comput. Sci. Eng.*, 9, 21
- Pichon, C., Gay, C., Pogosyan, D., et al. 2010, in *AIP Conf. Ser.*, eds. J. M. Alimi, & A. Fuözfa, 1241, 1108
- Planck Collaboration VI. 2020, *A&A*, 641, A6
- Pober, J. C., Liu, A., Dillon, J. S., et al. 2014, *ApJ*, 782, 66
- Pogosyan, D., Gay, C., & Pichon, C. 2009a, *Phys. Rev. D*, 80, 081301
- Pogosyan, D., Pichon, C., Gay, C., et al. 2009b, *MNRAS*, 396, 635
- Pogosyan, D., Pichon, C., & Gay, C. 2011, *Phys. Rev. D*, 84, 083510
- Rice, S. O. 1944, *Bell Syst. Tech. J.*, 23, 282
- Ross, H. E., Dixon, K. L., Ghara, R., Iliev, I. T., & Mellema, G. 2019, *MNRAS*, 487, 1101
- Schmalzing, J., & Gorski, K. M. 1998, *MNRAS*, 297, 355
- Shaw, A. K., Bharadwaj, S., & Mondal, R. 2020, *MNRAS*, 498, 1480
- Sorce, J. G., Ocvirk, P., Aubert, D., et al. 2022, *MNRAS*, 515, 2970
- Sousbie, T. 2011, *MNRAS*, 414, 350
- Thélie, E., Aubert, D., Gillet, N., & Ocvirk, P. 2022, *A&A*, 658, A139
- Trac, H., Cen, R., & Loeb, A. 2008, *ApJ*, 689, L81
- Trac, H., Chen, N., Holst, I., Alvarez, M. A., & Cen, R. 2022, *ApJ*, 927, 186
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Methods*, 17, 261
- Wise, J. H. 2019, *Contemporary Physics*, 60, 145
- Yoshiura, S., Shimabukuro, H., Takahashi, K., & Matsubara, T. 2017, *MNRAS*, 465, 394
- Zahn, O., Lidz, A., McQuinn, M., et al. 2007, *ApJ*, 654, 12
- Zahn, O., Mesinger, A., McQuinn, M., et al. 2011, *MNRAS*, 414, 727
- Zaldarriaga, M., Furlanetto, S. R., & Hernquist, L. 2004, *ApJ*, 608, 622
- Zhu, H., Avestruz, C., & Gnedin, N. Y. 2019, *ApJ*, 882, 152

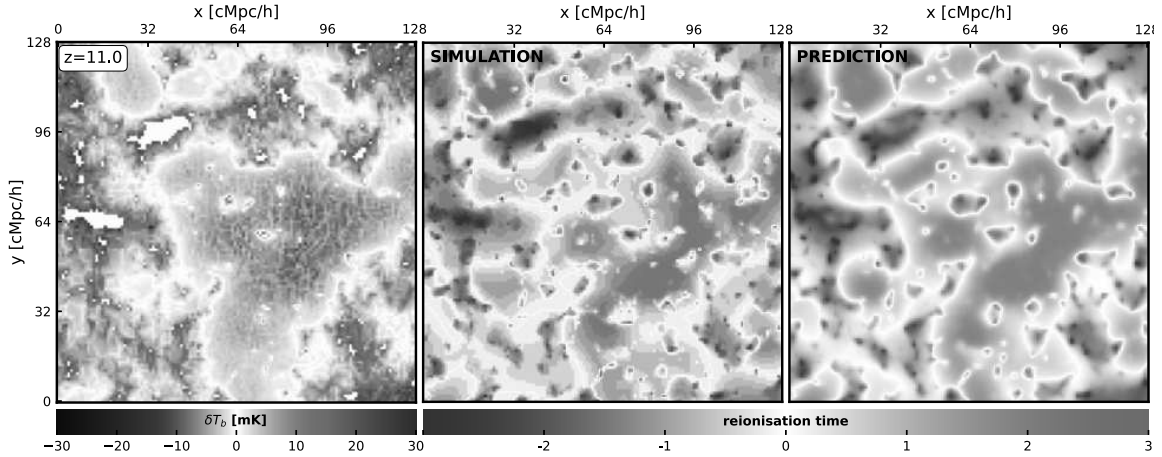
## Appendix A: Reconstruction of reionisation times from 21 cm

In this paper, we discuss the reionisation time field that informs us about the time at which the gas is reionised at each position. This field holds spatial and temporal information about the reionisation process. Even though the field is primarily available only via EoR models, our aim is to be able to also work on an ‘observed’ reionisation time field from 21 cm maps. For example, the SKA will collect the (redshifted) 21 cm signal to produce 2D images on the plane of the sky at many redshifts (or frequencies) along the line of sight. These images will therefore contain the differential brightness temperature  $\delta T_b$  relative to a background radio temperature, and map the distribution of neutral hydrogen at different redshifts. On the left panel of Fig. A.1, we show an example of a 21cmFAST 21 cm map at a redshift of 11. In Hiegel et al. (in preparation), we aim to reconstruct the reionisation times from this signal and this Appendix summarises what has been achieved so far towards this objective.

We used a convolutional neural network (CNN) algorithm that can learn and detect complex patterns within images. In particular, we developed a U-net that takes an image as input and reconstructs another image as output: in our case, the inputs are 2D 21 cm maps at a given redshift (such as the one in the left panel of Fig. A.1) and the CNN will learn to construct outputs

that will be as close as possible to the corresponding 2D reionisation time maps (such as the one in the middle panel of Fig. A.1). We therefore constructed data sets of 50 21cmFAST simulations that have a size of 256 cMpc/h with a resolution of 1 cMpc/h, from which we extract  $128 \times 128$  images. These images are split into a training set on which the CNN will learn to reconstruct  $t_{\text{reion}}(\mathbf{r})$  and a validation set to check its performance. In that paper, we do not smooth the reionisation time maps.

With this U-net, we are able to reconstruct reionisation time maps from observation-like maps with levels of correspondence to the true maps that vary with the observational redshift. For  $z \in [8 - 12]$ , 65% to 96% of  $t_{\text{reion}}(\mathbf{r})$  signal is well reconstructed. The reconstructed map shown in the right panel of Fig. A.1 was obtained with 21 cm maps taken at  $z = 11$ , which is one of the redshifts that returns the best results. We can see that in the process, and even for the best reconstructions, the small scales are smoothed out of the predicted  $t_{\text{reion}}(\mathbf{r})$  maps compared to the true ones, which is to be improved in future works. In Hiegel et al. (in preparation), we quantify the performance of the CNN via many diagnostics and show for example that we can extract an ionisation history that is consistent with the ones from the 21cmFAST simulations. This means that we can extract information about the evolution of the reionisation process and its topology simply from a 21 cm map obtained at a single redshift.



**Fig. A.1.** Example reconstruction of a 2D reionisation time map from a 2D map of the 21 cm signal taken at a redshift of  $z = 11$ . The left and middle panels are the brightness temperature and reionisation time fields generated by a 21cmFAST simulation. The right panel is the reconstruction of the reionisation times with the neural network designed to reproduce the true reionisation time map of the middle panel. Both  $t_{\text{reion}}(\mathbf{r})$  maps are dimensionless.

## Appendix B: Calculation of the moment of a field and its derivatives from a given power spectrum

In this work, we use the spectral moments  $\sigma_i$  (for  $i \in \mathbb{N}$ ) of the field of interest in order to normalise our fields, or because they appear in spectral parameters (defined in Eq. 6). These moments are only defined by the power spectrum of the field. The zeroth-order moment of a field is simply the standard deviation of the field, the first-order moment is the standard deviation of the first derivative of the field, and so on, as written in Eq. 2.

The moments can be expressed as follows (Bardeen et al. 1986; Pogosyan et al. 2009b; Gay 2011):

$$\sigma_i^2 = \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} \int_0^\infty k^{2i} \mathcal{P}_k k^{d-1} dk, \quad (\text{B.1})$$

where  $i \in \mathbb{N}$  corresponds to the number of derivation of the field, and  $d$  is the dimension of the field (in our case,  $d = 2$ ). In our case, we are interested in the power spectra of the reionisation field, which are defined in Eq. 9, and which have two slopes in logarithmic scales (in his thesis, Gay (2011) presented the calculation for a power spectrum with one slope in logarithmic scales). To do so, we need the gamma functions, defined below:

$$\gamma(a, x) = \int_0^x t^{a-1} e^{-t} dt \quad \text{and} \quad \Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt. \quad (\text{B.2})$$

The integral within the moments can then be separated into two integrals where the cut is at the threshold  $k_{\text{thresh}}$  separating the two parts of the power spectrum:

$$\begin{aligned} \sigma_i^2 &= \frac{2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} \left[ \int_0^{k_{\text{thresh}}} k^{2i} \mathcal{P}_k k^{d-1} dk + \int_{k_{\text{thresh}}}^\infty k^{2i} \mathcal{P}_k k^{d-1} dk \right] \\ &= \frac{A_2 2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} \left[ \frac{A_1}{A_2} \int_0^{k_{\text{thresh}}} k^{d-1+2i+n_1} e^{-2R_f^2 k^2} dk \right. \\ &\quad \left. + \int_{k_{\text{thresh}}}^\infty k^{d-1+2i+n_1} e^{-2R_f^2 k^2} dk \right] \\ &= \frac{A_2 2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} \left[ \frac{A_1}{A_2} \left( \frac{1}{\sqrt{2}R_f} \right)^{d+2i+n_1} \int_0^{2R_f^2 k_{\text{thresh}}^2} k^{\frac{d+n_1}{2}-1} e^k dk \right. \\ &\quad \left. + \left( \frac{1}{\sqrt{2}R_f} \right)^{d+2i+n_2} \int_{2R_f^2 k_{\text{thresh}}^2}^\infty k^{\frac{d+n_2}{2}-1} e^k dk \right]. \end{aligned} \quad (\text{B.3})$$

Now using the gamma functions defined above, we have the following expression for the spectral moments:

$$\begin{aligned} \sigma_i^2 &= \frac{A_2 2\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2})} \left[ \frac{A_1}{A_2} \left( \frac{1}{\sqrt{2}R_f} \right)^{d+2i+n_1} \gamma\left(\frac{d+n_1}{2} + i, 2R_f^2 k_{\text{thresh}}^2\right) \right. \\ &\quad \left. + \left( \frac{1}{\sqrt{2}R_f} \right)^{d+2i+n_2} \Gamma\left(\frac{d+n_2}{2} + i, 2R_f^2 k_{\text{thresh}}^2\right) \right]. \end{aligned} \quad (\text{B.4})$$

The moments  $\sigma_i^2$  have the same units as the power spectrum, but it is worth mentioning that the spectral parameters  $R_0$ ,  $R_*$ , and  $\gamma$  remain dimensionless. We also note that they are only dependent on the dimension and the power spectrum parameters:  $\sigma_i^2 = \sigma_i^2(d, A_1, n_1, A_2, n_2, R_f)$ .

## Appendix C: Calculation of the PDF of the gradient norm of a field

Our fields of interest being Gaussian, we reiterate that their PDF can be written as follows:

$$P(\mathbf{x}) d^n \mathbf{x} = \frac{1}{(2\pi)^{n/2} \cdot \det(C)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} \mathbf{x} \cdot C^{-1} \cdot \mathbf{x}\right) d^n \mathbf{x}, \quad (\text{C.1})$$

where  $\mathbf{x}$  is a n-D vector function of the position and  $C = \langle \mathbf{x} \otimes \mathbf{x} \rangle$  is the covariance matrix. To compute a PDF depending on the

field and its first derivative, we use  $\mathbf{x} = \begin{pmatrix} x \\ x_1 \\ x_2 \end{pmatrix}$ .

The covariance matrix of a 2D field  $F$  is the following:

$$\begin{aligned} C &= \begin{pmatrix} \langle F^2 \rangle & \langle F \nabla_1 F \rangle & \langle F \nabla_2 F \rangle \\ \langle \nabla_1 F F \rangle & \langle (\nabla_1 F)^2 \rangle & \langle \nabla_1 F \nabla_2 F \rangle \\ \langle \nabla_2 F F \rangle & \langle \nabla_1 F \nabla_2 F \rangle & \langle (\nabla_2 F)^2 \rangle \end{pmatrix} \\ &= \begin{pmatrix} \sigma_0^2 & 0 & 0 \\ 0 & \frac{1}{2}\sigma_1^2 & 0 \\ 0 & 0 & \frac{1}{2}\sigma_1^2 \end{pmatrix}. \end{aligned} \quad (\text{C.2})$$

Using the normalised variables  $x$ ,  $x_1$ , and  $x_2$ , the covariance matrix becomes:

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}. \quad (\text{C.3})$$

All the components of the PDF are now known, and after calculations, it is expressed as follows (and Eq. 15 is retrieved):

$$P(x, x_1, x_2) dx dx_1 dx_2 = \frac{2}{(2\pi)^{3/2}} e^{-\left(\frac{1}{2}x^2 + x_1^2 + x_2^2\right)} dx dx_1 dx_2. \quad (\text{C.4})$$

As we are interested in PDFs that only depend on the first derivative of  $F$ , an integration on the field values  $x$  is done (thanks to the Gaussian integral  $\int_{-\infty}^\infty e^{-\alpha y^2} dy = \sqrt{\frac{\pi}{\alpha}}$ ):

$$P(x_1, x_2) dx_1 dx_2 = \frac{1}{\pi} e^{-(x_1^2 + x_2^2)} dx_1 dx_2. \quad (\text{C.5})$$

Moreover, we are interested in the norm of the gradient of the field, which is why we make a change of variable and introduce  $w^2 = x_1^2 + x_2^2$ . This is, more precisely, a change of variables in 2D polar coordinates:

$$\forall \theta \in [0, 2\pi], \quad \begin{cases} x_1 = w \cos(\theta), \\ x_2 = w \sin(\theta). \end{cases} \quad (\text{C.6})$$

With this change of variable, and as the PDF is independent of the introduced angle  $\theta$ , we can write:

$$P(x_1, x_2) dx_1 dx_2 = P(w, \theta) w dw d\theta = 2\pi P(w) w dw = 2w e^{-w^2} dw, \quad (\text{C.7})$$

with a rewritten PDF depending only on the norm of the gradient of the field:

$$P(w) = \frac{1}{\pi} e^{-w^2}. \quad (\text{C.8})$$

## Appendix D: Reionisation redshift field analyses

The reionisation time and redshift fields are related to the same expression linking time and redshift:

$$z(t) = \frac{1}{a(t)} - 1, \quad (\text{D.1})$$

with  $a$  the scale factor. With this definition, they have opposite monotonicities, which has consequences for a topological study. Indeed, there are small differences between the two fields :  $t_{\text{reion}}(\mathbf{r})$  increases more rapidly as  $z_{\text{reion}}(\mathbf{r})$  decreases, and this causes some distinction in gaussianity analyses.

The aim of this Appendix is only to present a few results for the reionisation redshift field ( $z_{\text{reion}}(\mathbf{r})$ ), which presents some differences from the reionisation time field. To remain brief, we focus on the EMMA  $z_{\text{reion}}(\mathbf{r})$ . From the same simulation described in Sect. 2.1, we can extract 100 slices of the reionisation redshift field that are also smoothed with a Gaussian kernel of standard deviation  $R_f \in \{1, 2, 6\}$ , and normalised as below:

$$z_{\text{reion}} = \frac{z_{\text{reion}}^* - \overline{z_{\text{reion}}^*}}{\sigma_0^{z_{\text{reion}}^*}}, \quad (\text{D.2})$$

with  $\overline{z_{\text{reion}}^*}$  the mean of each field.  $\sigma_0^{z_{\text{reion}}^*}$  is the expected standard deviation of the field. The average and standard deviation of the reionisation redshift fields are given in Table 1. The power spectrum of the reionisation redshift field is also fitted as described in Sect. 2.1, and the resulting parameters are shown in Table 3.

We again generate 100 GRFs with the proper power spectrum to compare  $z_{\text{reion}}(\mathbf{r})$  to them; these GRFs are also smoothed and normalised as described in Sect. 2.4. The resulting predictions for the reionisation time field are also calculated for the reionisation redshift field and shown below. The expressions are slightly modified for  $z_{\text{reion}}(\mathbf{r})$ , changing some integral limits or signs.

### Filling factor

From Eq. 11, we can calculate the fraction of ionised volume of the reionisation redshifts in the same way as we did for the reionisation times. Here, the number of values that have a redshift higher than the threshold is equivalent to the number of cells that have already reionised. This therefore corresponds directly to the fraction of ionised volume  $Q_{\text{HII}}$ , as written below:

$$Q_{\text{HII}}(\nu) = \int_{\nu}^{\infty} P(x) dx = \frac{1}{2} \operatorname{erfc}\left(\frac{\nu}{\sqrt{2}}\right). \quad (\text{D.3})$$

### Filling factor of the gradient norm

As in the previous section, the filling factor of the gradient norm of redshift fields can be obtained from the joint PDF of Eq. 15, and is defined as follows:

$$f(\nu) = \int_{\nu}^{\infty} 2\pi P(w) w dw = e^{-\nu^2}. \quad (\text{D.4})$$

### Isocontour length

As it is symmetric, the isocontour length is the same for the reionisation time and redshift fields, for which the expression is reiterated below:

$$\mathcal{L}(\nu) = \frac{1}{2\sqrt{2}R_0} e^{-\frac{1}{2}\nu^2}. \quad (\text{D.5})$$

## Distribution of the maxima

The maxima of the reionisation redshift slices and the GRFs can be extracted with DisPerSE, as explained in Sect. 3.4. Their distribution can be theoretically calculated the same way as for the minima (as described in Sect. 3.4.2). We obtain the following distribution, where only some signs have changed:

$$\begin{aligned} \frac{\partial n_{\text{max}}}{\partial \nu} &= \frac{\exp\left(\frac{-\nu^2}{2}\right)}{\sqrt{2\pi}R_*^2} \left[ 1 + \operatorname{erf}\left(\frac{\gamma\nu}{\sqrt{2(1-\gamma^2)}}\right) \right] K_1(\nu, \gamma) \\ &+ \frac{\exp\left(\frac{-3\nu^2}{6-4\gamma^2}\right)}{\sqrt{2\pi(1-\frac{2}{3}\gamma^2)}R_*^2} \left[ 1 + \operatorname{erf}\left(\frac{\gamma\nu}{\sqrt{2(1-\gamma^2)(3-2\gamma^2)}}\right) \right] K_2 \\ &+ \frac{\exp\left(\frac{-\nu^2}{2(1-\gamma^2)}\right)}{\sqrt{2\pi(1-\gamma^2)}R_*^2} \left[ 1 + \exp\left(\frac{\nu^2}{2(1-\gamma^2)}\right) \right] K_3(\nu, \gamma). \end{aligned} \quad (\text{D.6})$$

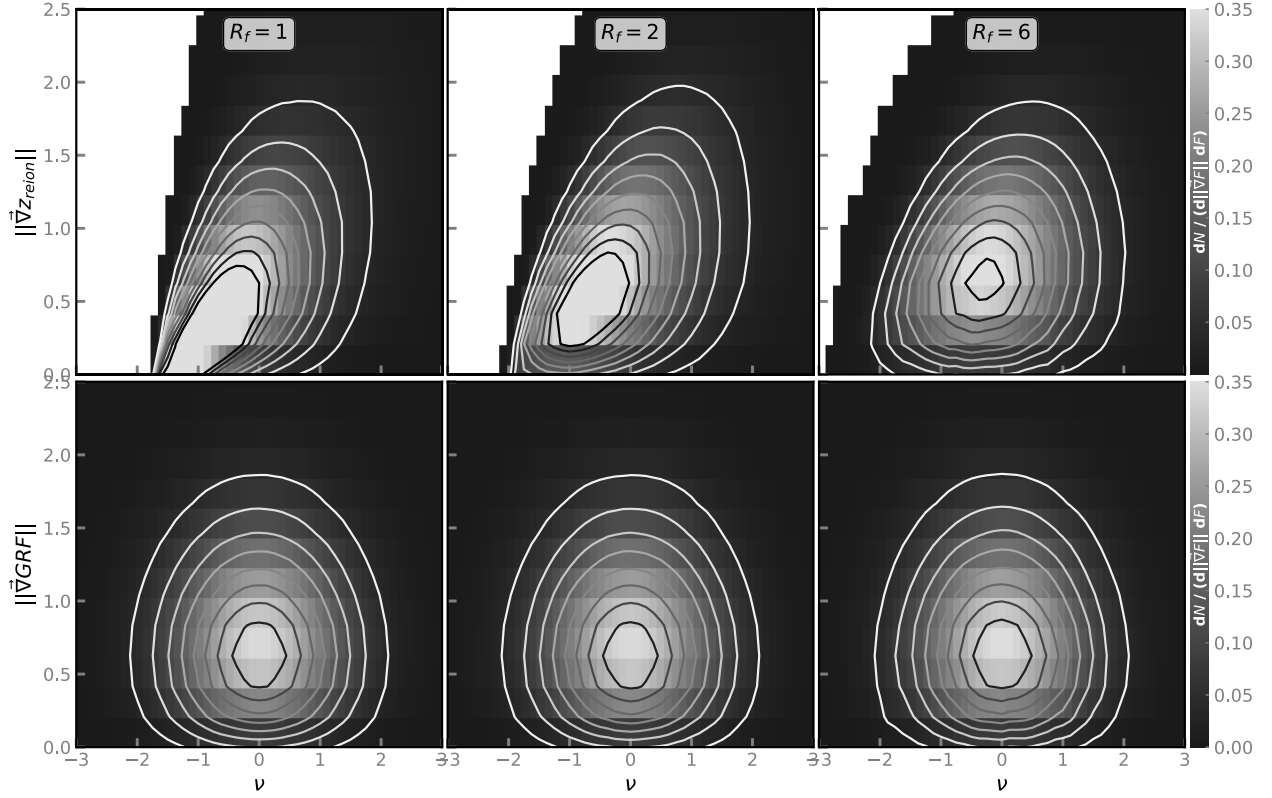
## Distribution of the anti-skeleton length

As the reionisation redshift field has an opposite monotony compared to the reionisation time field, the skeleton of  $t_{\text{reion}}(\mathbf{r})$  is equivalent to the anti-skeleton of  $z_{\text{reion}}(\mathbf{r})$ . The anti-skeleton joins minima together passing through saddle points, and can also be extracted from the fields thanks to DisPerSE. The distribution of the anti-skeleton lengths are calculated in the same way as the skeleton-length distribution. Gay et al. (2012) informs us that this results in the same expression as that for the skeleton length (see Eq. 27) but with  $\nu$  that becomes  $-\nu$ , as follows:

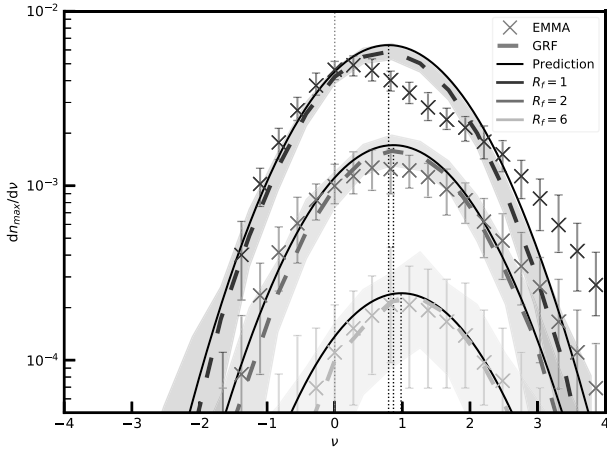
$$\begin{aligned} \frac{\partial L^{\text{skel}}}{\partial \nu} &= \frac{1}{\sqrt{2\pi}R_*} \exp^{-\frac{\nu^2}{2}} \left[ \frac{1}{8\sqrt{\pi}} (\sqrt{\pi} - 2\gamma\nu) \left( 1 + \operatorname{erf}\left(\frac{-\gamma\nu}{\sqrt{2(1-\gamma^2)}}\right) \right) \right] \\ &+ \frac{\sqrt{1-\gamma^2}}{2\sqrt{2\pi}} \exp\left(-\frac{\gamma^2\nu^2}{2(1-\gamma^2)}\right). \end{aligned} \quad (\text{D.7})$$

## Results

In this short result section, we only show figures that highlight the discrepancies between the reionisation time and redshift fields. We begin with the 2D histograms of the gradient norm of  $z_{\text{reion}}(\mathbf{r})$  versus  $t_{\text{reion}}(\mathbf{r})$  for the EMMA simulation in Fig. D.1. On the first row, there are the cosmological fields, and on the second row, there are the GRFs. There is again a ‘symmetrisation’ when the size of the Gaussian kernel increases (i.e.  $R_f$  increases), although it is less pronounced than for the reionisation time field (see Fig. 9). On Fig. D.2, we show the maxima PDF of the EMMA reionisation redshift fields (with the crosses) of the corresponding GRFs (with the dashed lines), and the prediction (in black). Comparing these to the minima PDF of the EMMA reionisation time field (see Fig. 13), we can see that the  $t_{\text{reion}}(\mathbf{r})$  minima PDFs are closer to the GRF ones than the  $z_{\text{reion}}(\mathbf{r})$  maxima PDFs, which again shows that  $t_{\text{reion}}(\mathbf{r})$  is more Gaussian than  $z_{\text{reion}}(\mathbf{r})$ . With these PDFs, we can again see the effect of the non-linear relation between  $z_{\text{reion}}(\mathbf{r})$  and  $t_{\text{reion}}(\mathbf{r})$ , which affects the x-axis. Indeed, for the smallest smoothing, the difference between the time and redshift of reionisation is evident:  $z_{\text{reion}}(\mathbf{r})$  underestimates the number of critical points (that are reionisation seeds) with respect to  $t_{\text{reion}}(\mathbf{r})$ . These figures globally show that the reionisation time field is more Gaussian than the reionisation redshift field, which is due to the non-linear relation between time and redshift, and which affects all the statistics studied in this paper.



**Fig. D.1.** Two-dimensional PDFs of the gradient norms with respect to the values of the fields of every run for each field and different smoothings ( $R_f \in \{1, 2, 6\}$ , see each column). The first row corresponds to the EMMA reionisation redshifts, and the second row to their corresponding GRFs. The grey-scale lines are the isocontours of the histograms. Here,  $\nu$  represents the value of the normalised reionisation redshifts.



**Fig. D.2.** Distribution of the critical points of the fields for the different smoothings (in colours). The median of every run is computed for each field. The dashed lines correspond to the GRFs, and the crosses are for the EMMA reionisation redshift fields. The black lines are the theoretical predictions. The shaded areas and the error bars represent the dispersion around the median (1<sup>st</sup> and 99<sup>th</sup> percentiles) of the GRFs and  $t_{\text{reion}}(P)$ , respectively. The black dotted vertical lines represent the average of the predictions. Here,  $\nu$  represents the value of the normalised reionisation redshifts.



# CHAPITRE 7

## CONCLUSIONS ET PERSPECTIVES

Les radio-télescopes à venir, comme le Square Kilometer Array (SKA), vont nous permettre de sonder (avec le signal 21 cm du gaz d'hydrogène neutre) l'époque de la formation des premières structures de l'Univers, que l'on appelle l'Époque de la Réionisation. L'un des défis des études actuelles portant sur cette époque est de fournir des outils qui nous permettront de comprendre et contraindre ce processus. Dans ce sens, j'ai exploré pendant ma thèse la topologie de l'évolution de la Réionisation à travers différents modèles de simulations numériques. Dans la foulée de Ocvirk et al. (2013, 2014); Aubert et al. (2018); Deparis et al. (2019); Chardin et al. (2019), je me suis en particulier intéressée au champ de temps de réionisation  $t_{\text{reion}}(\vec{r})$ , qui nous indique le moment auquel l'hydrogène a été réionisé, et qui contient une grande quantité d'informations à la fois temporelle et spatiale sur ce processus. Un de ses avantages est de résumer tout le déroulé du processus de réionisation en un seul champ.

Ce champ n'est pas une observable directe. Une partie importante de mes années de thèse a donc été orientée vers la reconstruction de ce champ à partir d'observations du signal 21 cm. Dans Hiegel et al. (2023), nous avons alors utilisé un algorithme d'apprentissage automatique basé sur un réseau de neurones convolutionnel, qui prend en entrée une carte de signal 21 cm à un redshift d'observation donné, et reconstruit en sortie la carte de  $t_{\text{reion}}(\vec{r})$  du modèle correspondant. Malgré le fait que les petites structures soient lissées dans les reconstructions, ces dernières sont tout de même satisfaisantes : par exemple, elles permettent d'accéder à des statistiques telles que le spectre de puissance, l'histoire de réionisation, ou la distribution des valeurs aux minima du champ. Les temps de réionisation peuvent également être reconstruits à partir de cartes de signal 21 cm bruitées avec néanmoins d'autant plus de structures aux petites échelles lissées : elles nous permettent encore tout de même de mesurer l'histoire de réionisation par exemple.

Ayant démontré que nous pouvons extraire  $t_{\text{reion}}(\vec{r})$  du signal 21 cm, je me suis notamment intéressée aux patches de réionisation de  $t_{\text{reion}}(\vec{r})$  dans Thélie et al. (2022), qui sont les régions associées aux minima de  $t_{\text{reion}}(\vec{r})$ , c'est-à-dire aux sources de première réionisation. J'ai en particulier montré qu'ils ont une taille moyenne de 10 cMpc/h et qu'ils sont majoritairement

---

allongés et alignés le long des filaments de matière. Par ailleurs, dans Th  lie et al. (2023), j’ai montr   comment un jeu de statistiques topologiques des cartes  $t_{\text{reion}}(\vec{r})$  nous informent sur la propagation des fronts d’ionisation, l’  volution de la taille des bulles ionis  es et r  gions neutres, les sources de r  ionisation et la percolation des bulles. J’ai pu   tudier les diff  rences entre les simulations cosmologiques EMMA et les mod  les semi-analytiques 21cmFAST et diff  rents sc  narios de R  ionisation. Par exemple, la vitesse des fronts et leur propagation ont des diff  rences remarquables aux temps tardifs de r  ionisation d      leur diff  rente mod  lisation.

Enfin, dans cette derni  re   tude, je me suis rendue compte que ces statistiques peuvent aussi   tre calcul  es de mani  re totalement analytique gr  ce    la th  orie des champs Gaussiens al  atoires (GRFs). Pour un tel champ, son spectre de puissance contient toute son information et permet ainsi    lui seul de calculer ces statistiques topologiques. J’ai ainsi compar   les mesures faites dans le champ    ces pr  dictions Gaussiennes et j’ai montr   que  $t_{\text{reion}}(\vec{r})$  peut   tre consid  r   comme un champ Gaussien dans une bonne approximation si sa r  solution s’approche de celle que nous auront dans les observations SKA. Dans ce cas, je peux ainsi caract  riser tout le processus de la r  ionisation avec l’unique connaissance du spectre de puissance des temps de r  ionisation.    l’inverse, les statistiques topologiques peuvent   tre aussi utilis  es pour contraindre ce spectre de puissance et ainsi le processus de r  ionisation.

### *Perspectives...*

J’ai plusieurs pistes pour aller au-del   des r  sultats obtenus pendant ma th  se. Nous avons vu que le champ de temps de r  ionisation peut   tre approxim   comme un champ Gaussien aux grandes   chelles. Nous n’avons pas encore explor   les causes de cette Gaussianit   et il pourrait   tre int  ressant d’y r  fl  chir    l’avenir. Nous savons par exemple que le champ de densit   est Gaussien quand il est   tudi   aux grandes   chelles, et nous pensons que  $t_{\text{reion}}(\vec{r})$  est tr  s corr  l   avec celui-ci, ce qui pourrait expliquer sa Gaussianit  . Aux petites   chelles, des empreintes de non-Gaussianit   apparaissent par exemple sous forme d’asym  trie de la r  ionisation. Celles-ci pourraient   tre prises en compte par la th  orie des GRFs en ajoutant des termes non-Gaussiens par le biais de d  veloppement Gram-Charlier (Gay et al. 2012; Cadiou et al. 2020) rendant la description analytique encore plus pr  cise. De plus, concernant les reconstructions de  $t_{\text{reion}}(\vec{r})$     partir du signal 21 cm faites avec le r  seau de neurones, nous avons vu que ces m  mes petites   chelles   taient mal reconstruites et nous pensons    plusieurs mani  res d’am  liorer nos r  sultats. Par exemple, nous pourrions am  liorer les performances du r  seau de neurones en ajustant diff  remment ses hyper-param  tres ou bien en d  veloppant une architecture diff  rente, notamment    l’aide des r  seaux adversaires g  n  ratifs (generative adversarial network, GAN, voir Ullmo et al. (2021)).

Ces petites   chelles sont li  es aux premi  res galaxies qui se forment au sein des halos de mati  re noire pendant la R  ionisation. Les   tudier est en particulier int  ressant dans une   poque o   le James Webb Space Telescope (JWST) commence    r  v  ler directement les propri  t  s de nombreuses galaxies durant l’  poque de la R  ionisation. Par le biais de simulations, je pourrais m’int  resser aux galaxies ou aux halos de mati  re noire afin de comprendre leur environnement radiatif, en   tudiant par exemple l’extension de la r  gion d’influence radiative des halos, ou en comparant leur position relative    celle des sources de r  ionisation (minima de  $t_{\text{reion}}(\vec{r})$ ) et en regardant comment cela influence ou d  pend de leurs propri  t  s. Je pourrais   galement   tudier des analogues du Groupe Local au travers des cartes de temps ou de redshifts de r  ionisation, comme l’ont fait Sorce et al. (2022). Ces auteurs,

ainsi que d'autres comme Ocvirk et al. (2013, 2014) et Aubert et al. (2018), ont montré que ce genre d'études peut permettre de comprendre la modulation par le rayonnement de la formation d'étoiles ou la distribution des galaxies.

À l'avenir, j'aimerais plus m'impliquer dans la communauté travaillant sur le signal 21 cm. Ce signal commence à être contraint par des radio-télescopes tels que l'Hydrogen Epoch of Reionisation Array (HERA ; HERA Collaboration et al. (2023)). Avec SKA, on espère cartographier l'évolution de la distribution de gaz neutre au cours du temps. Nous obtiendrons un grand flux de données (157 To/s) à partir desquelles il faudra extraire de l'information, et je souhaite participer à tout cela. Une meilleure connaissance des données me permettra de meilleures prédictions, et de mieux cerner ce qui doit être modélisé ou mesuré dans  $t_{\text{reion}}(\vec{r})$  par exemple. Nous savons dorénavant et déjà que les observations tomographiques du signal 21 cm par SKA contiendront un niveau de bruit non négligeable qui sera dû à l'instrument et aux avant-plans. Dans notre étude pour reconstruire les temps de réionisation à partir de ce signal, nous avons déjà essayé de travailler avec des cartes contenant du bruit instrumental et les performances en étaient diminuées. Nous souhaitons ainsi améliorer ces performances et éventuellement essayer de reconstruire  $t_{\text{reion}}(\vec{r})$  en ajoutant également les avant-plans. Alternativement, nous pourrions traiter et essayer de soustraire le bruit et les avant-plans des cartes de signal 21 cm : c'est d'ailleurs un *data challenge* proposé par l'Observatoire SKA (SKAO).<sup>1</sup> Par ailleurs, j'ai aussi pour projet d'étudier les corrélations entre le spectre de puissance du signal 21 cm et celui des cartes  $t_{\text{reion}}(\vec{r})$ , voire même d'essayer de reconstruire le second à partir du premier. Nous savons maintenant qu'aux grandes échelles  $t_{\text{reion}}(\vec{r})$  peut être considéré comme Gaussien avec une bonne approximation. Par conséquent, si je peux reconstruire son spectre de puissance à partir d'observations du spectre de puissance du signal 21 cm, je pourrais décrire la topologie du processus de réionisation avec ces futures observations. Cette reconstruction pourrait par exemple se faire à l'aide de réseau de neurones. Autrement, une relation analytique de transformation d'un spectre de puissance à l'autre pourrait être obtenue par régression symbolique, avec par exemple l'algorithme  $\Phi$ -SO développé par Tenachi et al. (2023).

Par la suite, je vais aussi m'intéresser à la modélisation du signal 21 cm en lui-même en plus de son interprétation. Par exemple, Muñoz (2023) a développé un modèle purement analytique de l'Aube Cosmique pour construire le signal global et le spectre de puissance du signal 21 cm. Nous souhaiterions étendre ce modèle à la Réionisation afin de contraindre le scénario de Réionisation avec les futures observations du signal 21 cm. Je vais d'ailleurs partir à l'Université du Texas à Austin, où je serais « Fellow » et travaillerai en collaboration avec Julian Muñoz sur l'Époque de la Réionisation et l'Aube Cosmique.

Enfin, de nouvelles infrastructures de calcul vont apparaître dans les années à venir, et les supercalculateurs exaflopiques vont en particulier voir le jour. Les simulations, qui sont un moyen fondamental d'étude de l'Univers jeune, vont devoir s'adapter à ces nouveaux calculateurs. De plus, étudier la Réionisation avec SKA va nécessiter la modélisation de grands volumes de données (de l'ordre du Gpc<sup>3</sup>) dans lesquels les sources de réionisation devront aussi être résolues. Les futures simulations devront donc être assez « massives » pour répondre à ces enjeux. De nouveaux codes, en particulier adaptés aux futures machines de calcul, sont donc en cours de développement. Pendant ma thèse, j'ai notamment eu l'occasion de suivre

---

1. <https://sdc2.astronomers.skatelescope.org/sdc2-challenge/description>

---

le projet *Dyablo* (DYnamics adaptive mesh refinement CFD applications with PABLO) et je souhaite m’y investir plus à l’avenir (Aubert & Durocher 2021). Il permet à l’utilisateur de générer des simulations avec raffinement de mailles adaptatives. *Dyablo* devrait ainsi pouvoir générer relativement « aisément » des simulations cosmologiques de la Réionisation qui simulent entre autres le signal 21 cm. D’autre part, avec l’essor de l’intelligence artificielle, de nouvelles méthodes de modélisations apparaissent, et celles-ci sont souvent construites pour avoir un bien moindre temps de calcul que les simulations classiques. Il existe par exemple des algorithmes basés sur de l’apprentissage profond permettant de générer des données cosmologiques, comme des cubes de densité ou des réalisations du fond diffus cosmologique (Han et al. 2021; Ullmo et al. 2021; Conceição et al. 2023). Un couplage entre les simulations et l’intelligence artificielle pourrait peut-être faire partie des outils standards de ma future recherche.

Pendant ma thèse, j’ai ainsi eu l’opportunité de commencer à explorer une époque de l’histoire de l’Univers qui va être au centre d’une attention particulière à l’avenir, notamment avec SKA. J’ai hâte de participer à tout cela au travers de ces divers projets et de bien d’autres encore à venir.

# BIBLIOGRAPHIE

- Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow : Large-Scale Machine Learning on Heterogeneous Systems, software available from tensorflow.org
- Abbott, T. M. C., Allam, S., Andersen, P., et al. 2019, ApJL, 872, L30
- Abel, T., Bryan, G. L., & Norman, M. L. 2002, Science, 295, 93
- Adams, N. J., Conselice, C. J., Ferreira, L., et al. 2023, MNRAS, 518, 4755
- Alvarez, M. A., Busha, M., Abel, T., & Wechsler, R. H. 2009, ApJL, 703, L167
- Atek, H., Shuntov, M., Furtak, L. J., et al. 2023, MNRAS, 519, 1201
- Aubert, D., Deparis, N., & Ocvirk, P. 2015, MNRAS, 454, 1012
- Aubert, D., Deparis, N., Ocvirk, P., et al. 2018, ApJL, 856, L22
- Aubert, D. & Durocher, A. 2021, in SF2A-2021 : Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics. Eds. : A. Siebert, ed. A. Siebert, K. Baillié, E. Lagadec, N. Lagarde, J. Malzac, J. B. Marquette, M. N'Diaye, J. Richard, & O. Venot, 473–475
- Aubert, D. & Teyssier, R. 2008, MNRAS, 387, 295
- Banet, A., Barkana, R., Fialkov, A., & Guttman, O. 2021, MNRAS, 503, 1221
- Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15
- Barkana, R. 2016, PhR, 645, 1
- Battaglia, N., Trac, H., Cen, R., & Loeb, A. 2013, ApJ, 776, 81
- Bianco, M., Giri, S. K., Iliev, I. T., & Mellema, G. 2021, MNRAS, 505, 3982
- Bond, J. R., Kofman, L., & Pogosyan, D. 1996, Nature, 380, 603
- Bosman, S. E. I., Davies, F. B., Becker, G. D., et al. 2022, MNRAS, 514, 55

- Bouwens, R. J., Oesch, P. A., Illingworth, G. D., Ellis, R. S., & Stefanon, M. 2017, *ApJ*, 843, 129
- Bouwens, R. J., Stefanon, M., Brammer, G., et al. 2023, *MNRAS*, 523, 1036
- Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2006, *ApJ*, 638, 20
- Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., & Mahesh, N. 2018, *Nature*, 555, 67
- Boylan-Kolchin, M. 2023, *Nature Astronomy*
- Bradley, L. D., Coe, D., Brammer, G., et al. 2022, arXiv e-prints, arXiv :2210.01777
- Bradley, R. F., Tauscher, K., Rapetti, D., & Burns, J. O. 2019, *ApJ*, 874, 153
- Bromm, V., Ferrara, A., Coppi, P. S., & Larson, R. B. 2001, *MNRAS*, 328, 969
- Bromm, V. & Larson, R. B. 2004, *ARA&A*, 42, 79
- Cadiou, C., Pichon, C., Codis, S., et al. 2020, *MNRAS*, 496, 4787
- Castellano, M., Fontana, A., Treu, T., et al. 2022, *ApJL*, 938, L15
- Castellano, M., Menci, N., & Romanello, M. 2023, arXiv e-prints, arXiv :2301.03854
- Chardin, J., Uhlrich, G., Aubert, D., et al. 2019, *MNRAS*, 490, 1055
- Chen, Z., Xu, Y., Wang, Y., & Chen, X. 2019, *ApJ*, 885, 23
- Choudhury, T. R. 2022, *General Relativity and Gravitation*, 54, 102
- Choudhury, T. R., Haehnelt, M. G., & Regan, J. 2009, *MNRAS*, 394, 960
- Ciardi, B. & Ferrara, A. 2005, *SSRv*, 116, 625
- Codis, S., Jindal, A., Chisari, N. E., et al. 2018, *MNRAS*, 481, 4753
- Cohn, J. D. 2022, *MNRAS*, 513, 624
- Cole, S., Percival, W. J., Peacock, J. A., et al. 2005, *MNRAS*, 362, 505
- Coles, P., Davies, A. G., & Pearson, R. C. 1996, *MNRAS*, 281, 1375
- Conceição, M., Krone-Martins, A., da Silva, A., & Moliné, Á. 2023, arXiv e-prints, arXiv :2304.06099
- Davies, F. B., Hennawi, J. F., Bañados, E., et al. 2018, *ApJ*, 864, 143
- Dawoodbhoy, T., Shapiro, P. R., Ocvirk, P., et al. 2018, *MNRAS*, 480, 1740
- Dayal, P. & Ferrara, A. 2018, *PhR*, 780, 1

- Deparis, N., Aubert, D., Ocvirk, P., Chardin, J., & Lewis, J. 2019, *A&A*, 622, A142
- Dijkstra, M., Mesinger, A., & Wyithe, J. S. B. 2011, *MNRAS*, 414, 2139
- Dixon, K. L., Iliev, I. T., Mellema, G., Ahn, K., & Shapiro, P. R. 2016, *MNRAS*, 456, 3011
- Doliva-Dolinsky, A., Martin, N. F., Yuan, Z., et al. 2023, arXiv e-prints, arXiv :2303.01528
- Donnan, C. T., McLeod, D. J., McLure, R. J., et al. 2023, *MNRAS*, 520, 4554
- Doussot, A., Trac, H., & Cen, R. 2019, *ApJ*, 870, 18
- Eilers, A.-C., Davies, F. B., & Hennawi, J. F. 2018, *ApJ*, 864, 53
- Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, *ApJ*, 633, 560
- Elbers, W. & van de Weygaert, R. 2023, *MNRAS*, 520, 2709
- Fan, X., Narayanan, V. K., Strauss, M. A., et al. 2002, *AJ*, 123, 1247
- Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, *AJ*, 132, 117
- Field, G. B. 1958, *Proceedings of the IRE*, 46, 240
- Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, *ApJL*, 946, L13
- Forman, R. 1998, *Advances in Mathematics*, 134, 90
- Forman, R. 2001, *Sém. Lothar. Combin.*, 48
- Friedrich, M. M., Mellema, G., Alvarez, M. A., Shapiro, P. R., & Iliev, I. T. 2011, *MNRAS*, 413, 1353
- Fujimoto, S., Arrabal Haro, P., Dickinson, M., et al. 2023, arXiv e-prints, arXiv :2301.09482
- Furlanetto, S. R., Oh, S. P., & Briggs, F. H. 2006, *PhR*, 433, 181
- Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2004a, *ApJ*, 613, 16
- Furlanetto, S. R., Zaldarriaga, M., & Hernquist, L. 2004b, *ApJ*, 613, 1
- Galárraga-Espinosa, D., Aghanim, N., Langer, M., Gouin, C., & Malavasi, N. 2020, *A&A*, 641, A173
- Gay, C. 2011, PhD thesis, thèse de doctorat dirigée par Pichon, Christophe Physique Paris 6 2011
- Gay, C., Pichon, C., & Pogosyan, D. 2012, *PhRvD*, 85, 023011
- Geller, M. J. & Huchra, J. P. 1989, *Science*, 246, 897
- Gillet, N. 2016, *Theses*, Université de Strasbourg

- Gillet, N., Mesinger, A., Greig, B., Liu, A., & Ucci, G. 2019, MNRAS, 484, 282
- Gillet, N. J. F., Aubert, D., Mertens, F. G., & Ocvirk, P. 2021, MNRAS, 507, 3179
- Giri, S., Mellema, G., & Jensen, H. 2020, The Journal of Open Source Software, 5, 2363
- Giri, S. K. & Mellema, G. 2021, MNRAS, 505, 1863
- Giri, S. K., Mellema, G., Aldheimer, T., Dixon, K. L., & Iliev, I. T. 2019, MNRAS, 489, 1590
- Giri, S. K., Mellema, G., Dixon, K. L., & Iliev, I. T. 2018a, MNRAS, 473, 2949
- Giri, S. K., Mellema, G., & Ghara, R. 2018b, MNRAS, 479, 5596
- Gleser, L., Nusser, A., Ciardi, B., & Desjacques, V. 2006, MNRAS, 370, 1329
- Gorce, A., Ganjam, S., Liu, A., et al. 2023, MNRAS, 520, 375
- Gorce, A., Ilić, S., Douspis, M., Aubert, D., & Langer, M. 2020, A&A, 640, A90
- Gorce, A. & Pritchard, J. R. 2019, MNRAS, 489, 1321
- Gott, J. Richard, I., Weinberg, D. H., & Melott, A. L. 1987, ApJ, 319, 1
- Greig, B. & Mesinger, A. 2015, MNRAS, 449, 4246
- Greig, B. & Mesinger, A. 2017, MNRAS, 472, 2651
- Greig, B., Mesinger, A., & Bañados, E. 2019, MNRAS, 484, 5094
- Greig, B., Mesinger, A., Davies, F. B., et al. 2022, MNRAS, 512, 5390
- Greig, B., Mesinger, A., Haiman, Z., & Simcoe, R. A. 2017, MNRAS, 466, 4239
- Grissom, R. L., Ballantyne, D. R., & Wise, J. H. 2014, A&A, 561, A90
- Gunn, J. E. & Peterson, B. A. 1965, ApJ, 142, 1633
- Han, D., Sehgal, N., & Villaescusa-Navarro, F. 2021, PhRvD, 104, 123521
- Harikane, Y., Nakajima, K., Ouchi, M., et al. 2023a, arXiv e-prints, arXiv :2304.06658
- Harikane, Y., Ouchi, M., Oguri, M., et al. 2023b, ApJS, 265, 5
- Hassan, S., Davé, R., Finlator, K., & Santos, M. G. 2016, MNRAS, 457, 1550
- HERA Collaboration, Abdurashidova, Z., Adams, T., et al. 2023, ApJ, 945, 124
- Hiegel, J., Thélie, E., Aubert, D., et al. 2023, Reionisation time fields reconstruction from 21 cm signal map
- Hills, R., Kulkarni, G., Meerburg, P. D., & Puchwein, E. 2018, Nature, 564, E32



- Hong, S. E., Ahn, K., Park, C., et al. 2014, *Journal of Korean Astronomical Society*, 47, 49
- Hutter, A. 2018, *MNRAS*, 477, 1549
- Hutter, A., Watkinson, C. A., Seiler, J., et al. 2020, *MNRAS*, 492, 653
- Iliev, I. T., Ciardi, B., Alvarez, M. A., et al. 2006a, *MNRAS*, 371, 1057
- Iliev, I. T., Mellema, G., Pen, U. L., et al. 2006b, *MNRAS*, 369, 1625
- Iliev, I. T., Mellema, G., Shapiro, P. R., et al. 2012, *MNRAS*, 423, 2222
- Iliev, I. T., Pen, U.-L., Bond, J. R., Mellema, G., & Shapiro, P. R. 2007, *ApJ*, 660, 933
- Jeon, M., Pawlik, A. H., Bromm, V., & Milosavljević, M. 2014, *MNRAS*, 440, 3778
- Kakiichi, K., Majumdar, S., Mellema, G., et al. 2017, *MNRAS*, 471, 1936
- Kannan, R. & Thesan Team. 2022, in *American Astronomical Society Meeting Abstracts*, Vol. 54, *American Astronomical Society Meeting Abstracts*, 420.01
- Kapahtia, A., Chingangbam, P., & Appleby, S. 2019, *JCAP*, 2019, 053
- Kapahtia, A., Chingangbam, P., Ghara, R., Appleby, S., & Choudhury, T. R. 2021, *JCAP*, 2021, 026
- Kashikawa, N., Ishizaki, Y., Willott, C. J., et al. 2015, *ApJ*, 798, 28
- Katz, H., Martin-Alvarez, S., Rosdahl, J., et al. 2021, *MNRAS*, 507, 1254
- Katz, H., Ramsøy, M., Rosdahl, J., et al. 2020, *MNRAS*, 494, 2200
- Keller, B. W., Munshi, F., Trebitsch, M., & Tremmel, M. 2023, *ApJL*, 943, L28
- Kern, N. S., Liu, A., Parsons, A. R., Mesinger, A., & Greig, B. 2017, *ApJ*, 848, 23
- Khokhlov, A. 1998, *Journal of Computational Physics*, 143, 519
- Koopmans, L., Pritchard, J., Mellema, G., et al. 2015, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 1
- Kraljic, K., Davé, R., & Pichon, C. 2020, *MNRAS*, 493, 362
- Kroupa, P. 2002, *Science*, 295, 82
- Labbé, I., van Dokkum, P., Nelson, E., et al. 2023, *Nature*, 616, 266
- Lee, K.-G., Cen, R., Gott, J. Richard, I., & Trac, H. 2008, *ApJ*, 675, 8
- Levermore, C. D. 1984, *JQSRT*, 31, 149
- Lewis, J. S. W., Ocvirk, P., Aubert, D., et al. 2020, *MNRAS*, 496, 4342

- Lewis, J. S. W., Ocvirk, P., Sorce, J. G., et al. 2022, MNRAS, 516, 3389
- Lidz, A., Zahn, O., McQuinn, M., Zaldarriaga, M., & Hernquist, L. 2008, ApJ, 680, 962
- Lin, Y., Oh, S. P., Furlanetto, S. R., & Sutter, P. M. 2016, MNRAS, 461, 3361
- Liu, A. & Parsons, A. R. 2016, MNRAS, 457, 1864
- Loeb, A. & Furlanetto, S. R. 2013, *The First Galaxies in the Universe*
- Longuet-Higgins, M. S. 1957, *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 249, 321
- Lynden-Bell, D. 1976, MNRAS, 174, 695
- Malavasi, N., Aghanim, N., Douspis, M., Tanimura, H., & Bonjean, V. 2020, A&A, 642, A19
- Malhotra, S. & Rhoads, J. E. 2006, ApJL, 647, L95
- Mason, C. A. & Gronke, M. 2020, MNRAS, 499, 1395
- Matsubara, T. 2003, ApJ, 584, 1
- Matsubara, T. & Kuriki, S. 2021, PhRvD, 104, 103522
- McQuinn, M., Furlanetto, S. R., Hernquist, L., Zahn, O., & Zaldarriaga, M. 2005, ApJ, 630, 643
- McQuinn, M., Lidz, A., Zahn, O., et al. 2007, MNRAS, 377, 1043
- Mellema, G., Iliev, I. T., Pen, U.-L., & Shapiro, P. R. 2006, MNRAS, 372, 679
- Mellema, G., Koopmans, L., Shukla, H., et al. 2015, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 10
- Mesinger, A., Aykutalp, A., Vanzella, E., et al. 2015, MNRAS, 446, 566
- Mesinger, A., Ferrara, A., & Spiegel, D. S. 2013, MNRAS, 431, 621
- Mesinger, A. & Furlanetto, S. 2007, ApJ, 669, 663
- Mesinger, A., Furlanetto, S., & Cen, R. 2011, MNRAS, 411, 955
- Mesinger, A. & Haiman, Z. 2004, ApJL, 611, L69
- Mesinger, A., McQuinn, M., & Spergel, D. N. 2012, MNRAS, 422, 1403
- Minkowski, H. 1903, MatAn, 57, 447
- Morishita, T. & Stiavelli, M. 2023, ApJL, 946, L35
- Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, Nature, 474, 616

- Muñoz, J. B. 2023, arXiv e-prints, arXiv :2302.08506
- Murray, S., Greig, B., Mesinger, A., et al. 2020, *The Journal of Open Source Software*, 5, 2582
- Naidu, R. P., Oesch, P. A., van Dokkum, P., et al. 2022, *ApJL*, 940, L14
- Novikov, D., Colombi, S., & Dore, O. 2006, *Monthly Notices of the Royal Astronomical Society*, 366, 1201
- Ocvirk, P. & Aubert, D. 2011, *MNRAS*, 417, L93
- Ocvirk, P., Aubert, D., Chardin, J., et al. 2013, *ApJ*, 777, 51
- Ocvirk, P., Aubert, D., Sorce, J. G., et al. 2020, *MNRAS*, 496, 4087
- Ocvirk, P., Gillet, N., Aubert, D., et al. 2014, *ApJ*, 794, 20
- Ocvirk, P., Gillet, N., Shapiro, P. R., et al. 2016, *MNRAS*, 463, 1462
- Pagano, M. & Liu, A. 2020, *MNRAS*, 498, 373
- Park, J., Mesinger, A., Greig, B., & Gillet, N. 2019, *MNRAS*, 484, 933
- Pathak, A., Bag, S., Dasgupta, S., et al. 2022, *JCAP*, 2022, 027
- Penzias, A. A. & Wilson, R. W. 1965, *ApJ*, 142, 419
- Pichon, C., Gay, C., Pogosyan, D., et al. 2010, in *American Institute of Physics Conference Series*, Vol. 1241, *Invisible Universe*, ed. J.-M. Alimi & A. Fuözfa, 1108–1117
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6
- Pober, J. C., Liu, A., Dillon, J. S., et al. 2014, *ApJ*, 782, 66
- Pogosyan, D., Gay, C., & Pichon, C. 2009a, *PhRvD*, 80, 081301
- Pogosyan, D., Pichon, C., & Gay, C. 2011, *PhRvD*, 84, 083510
- Pogosyan, D., Pichon, C., Gay, C., et al. 2009b, *MNRAS*, 396, 635
- Prelogović, D., Mesinger, A., Murray, S., Fiameni, G., & Gillet, N. 2022, *MNRAS*, 509, 3852
- Press, W. H. & Schechter, P. 1974, *ApJ*, 187, 425
- Puchwein, E., Bolton, J. S., Keating, L. C., et al. 2023, *MNRAS*, 519, 6162
- Rice, S. O. 1944, *The Bell System Technical Journal*, 23, 282
- Ronneberger, O., Fischer, P., & Brox, T. 2015, *CoRR*, abs/1505.04597
- Salpeter, E. E. 1955, *ApJ*, 121, 161

- Santos, M. G., Ferramacho, L., Silva, M. B., Amblard, A., & Cooray, A. 2010, MNRAS, 406, 2421
- Schaap, W. E. 2007, PhD thesis, University of Groningen, Netherlands
- Schaerer, D. 2002, A&A, 382, 28
- Schmalzing, J. & Buchert, T. 1997, ApJL, 482, L1
- Schmalzing, J. & Gorski, K. M. 1998, MNRAS, 297, 355
- Semelin, B. 2016, MNRAS, 455, 962
- Semelin, B., Eames, E., Bolgar, F., & Caillat, M. 2017, MNRAS, 472, 4508
- Sheth, R. K. & Tormen, G. 1999, MNRAS, 308, 119
- Shimabukuro, H., Mao, Y., & Tan, J. 2022, Research in Astronomy and Astrophysics, 22, 035027
- Sims, P. H. & Pober, J. C. 2020, MNRAS, 492, 22
- Singh, A., Mahajan, S., & Bagla, J. S. 2020, MNRAS, 497, 2265
- Singh, S., Jishnu, N. T., Subrahmanyan, R., et al. 2022, Nature Astronomy, 6, 607
- Singh, S., Subrahmanyan, R., Udaya Shankar, N., et al. 2018, ApJ, 858, 54
- Song, H., Laigle, C., Hwang, H. S., et al. 2021, MNRAS, 501, 4635
- Sorce, J. G., Ocvirk, P., Aubert, D., et al. 2022, MNRAS, 515, 2970
- Sousbie, T. 2011, MNRAS, 414, 350
- Sousbie, T., Pichon, C., & Kawahara, H. 2011, MNRAS, 414, 384
- Springel, V., Frenk, C. S., & White, S. D. M. 2006, Nature, 440, 1137
- Sunyaev, R. A. & Chluba, J. 2009, Astronomische Nachrichten, 330, 657
- Tanimura, H., Aghanim, N., Bonjean, V., Malavasi, N., & Douspis, M. 2020, A&A, 637, A41
- Tenachi, W., Ibata, R., & Diakogiannis, F. I. 2023, arXiv e-prints, arXiv :2303.03192
- Teyssier, R. 2002, A&A, 385, 337
- Thélie, E., Aubert, D., Gillet, N., Hiegel, J., & Ocvirk, P. 2023, A&A, 672, A184
- Thélie, E., Aubert, D., Gillet, N., & Ocvirk, P. 2022, A&A, 658, A139
- Tormen, G. 1997, MNRAS, 290, 411
- Trac, H., Cen, R., & Loeb, A. 2008, ApJL, 689, L81

## BIBLIOGRAPHIE

---

- Trac, H., Chen, N., Holst, I., Alvarez, M. A., & Cen, R. 2022, *ApJ*, 927, 186
- Trac, H. Y. & Gnedin, N. Y. 2011, *Advanced Science Letters*, 4, 228
- Ullmo, M., Decelle, A., & Aghanim, N. 2021, *A&A*, 651, A46
- Walker, M. G. & Peñarrubia, J. 2011, *ApJ*, 742, 20
- Wang, F., Davies, F. B., Yang, J., et al. 2020, *ApJ*, 896, 23
- White, S. D. M. & Rees, M. J. 1978, *MNRAS*, 183, 341
- Williams, R. E., Blacker, B., Dickinson, M., et al. 1996, *AJ*, 112, 1335
- Wise, J. H. 2019, *Contemporary Physics*, 60, 145
- Wouthuysen, S. A. 1952, *AJ*, 57, 31
- Xu, Y., Yue, B., & Chen, X. 2017, *ApJ*, 844, 117
- Yan, H., Ma, Z., Ling, C., Cheng, C., & Huang, J.-S. 2023, *ApJL*, 942, L9
- Yoshiura, S., Shimabukuro, H., Takahashi, K., & Matsubara, T. 2017, *MNRAS*, 465, 394
- Zahn, O., Lidz, A., McQuinn, M., et al. 2007, *ApJ*, 654, 12
- Zahn, O., Mesinger, A., McQuinn, M., et al. 2011, *MNRAS*, 414, 727
- Zahn, O., Zaldarriaga, M., Hernquist, L., & McQuinn, M. 2005, *ApJ*, 630, 657
- Zaldarriaga, M., Furlanetto, S. R., & Hernquist, L. 2004, *ApJ*, 608, 622
- Zel'dovich, Y. B. 1970, *A&A*, 5, 84
- Zhu, H., Avestruz, C., & Gnedin, N. Y. 2019, *ApJ*, 882, 152

# Mesure, prédiction et modélisation de la topologie de l'évolution temporelle de la Réionisation cosmologique

## Résumé

Les premières sources de lumière, formées pendant la Réionisation, ionisent le gaz autour d'elles dans des bulles qui grossissent jusqu'à ce qu'il soit entièrement réionisé un milliard d'années après le Big Bang. J'ai étudié ce processus au travers de modélisations du champ de temps de réionisation (*treion*) qui donne l'instant de réionisation du gaz en tout point.

Des radiotélescopes, comme le Square Kilometer Array, vont sonder la Réionisation avec le signal 21 cm émis par l'Hydrogène neutre. J'ai montré que *treion* peut être reconstruit à partir de cette observable via des réseaux de neurones profonds.

Je me suis intéressée à la topologie de *treion*. Elle indique que le rayonnement à grande échelle se propage globalement le long des filaments de matière. Les sources, fronts d'ionisation, bulles ionisées, régions neutres, et la percolation sont reliés à des statistiques mesurables dans *treion*. Ce champ peut enfin être considéré Gaussien aux grandes échelles, nous donnant ainsi accès à ces statistiques avec l'unique connaissance de son spectre de puissance.

**Mots-clés** : astrophysique, cosmologie, réionisation, grandes échelles de l'Univers, simulations

## Résumé en anglais

The first light sources, formed during the Epoch of Reionisation, ionise the gas around them in bubbles that grow until it is completely reionised one billion years after the Big Bang. I have studied this process by modelling the reionisation time field (*treion*), which gives us the instant of reionisation of the gas at any point.

Radio-telescopes such as the Square Kilometer Array will probe the Reionisation using the 21 cm signal emitted by neutral Hydrogen. I have shown that *treion* can be reconstructed from this observable using deep neural networks.

I put my interest on the topology of *treion*. It indicates that large-scale radiation propagates globally along the matter filaments. Reionisation sources, ionisation fronts, ionised bubbles, neutral regions and percolation are linked to measurable statistics in *treion*. This field can finally be considered Gaussian on large scales, giving us access to these statistics with the sole knowledge of its power spectrum.

**Keywords**: astrophysics, cosmology, reionisation, large scale structure of the Universe, simulations