Les Lumières: probing the cosmic Epoch of Reionization with high-redshift quasars

Kist, T.

1 Introduction

The lights on the night sky have fascinated us humans since time immemorial, and it might be hard to find civilisations that have not developed their own myths and legends trying to make sense of the phenomena that they would observe. Given the sky's intangibility, it is maybe not surprising that these explanations were often related to stories of creation, involving all sorts of deities, higher powers or other mythical creatures. Perhaps all the more remarkable is the fact that *science* has convinced us that looking at the night sky indeed means nothing less than *looking into the very past of our own Universe*. As if this was not enough, the last century came with the rapid consolidation of a theory – the ΛCDM model – that allows us to tell our Universe's *entire* evolution history – based on the variety of lights that we are observing in the sky.

Instrumental to that end was the unprecedented number of paradigm shifts that our understanding of physics has gone through in the past century, including the development of general relativity (Einstein 1916) and quantum mechanics (e.g. Heisenberg 1925; Born & Jordan 1925; Schrödinger 1926a,b,c,d), subsequently extended to quantum field theory (Tomonaga 1946; Feynman 1948; Schwinger 1948; Dyson 1949), as our most fundamental physical theories to date. Even more so in the field of cosmology, essentially our *entire* base of knowledge has been generated in that century. The rise of modern cosmology perhaps started looming in 1920 when Harlow Shapley and Heber Curtis in their *Great Debate* discussed the question if the then observed *spiral nebulae* would be of intra- or extragalactic origin, and therefore, more fundamentally, if the Universe would extend *beyond our own galaxy, the Milky Way*. The question ultimately got settled with Edwin Hubble's discovery that these nebulae's recessional velocities would linearly increase with their distance to Earth (Hubble 1929), confirming earlier observations by Vesto Slipher (Slipher 1913, 1917) and suggesting that they were indeed nothing else but separate *galaxies* other than the Milky Way.

Two conflicting theories arose in the subsequent decades as physical frameworks to explain the expansion of the Universe implied by these observations: first, the *steady-state theory* where the Universe would eternally expand, retaining its appearance over cosmic time by accommodating the spatial expansion via the concomitant creation of new matter that would help keep the matter density constant over time (Bondi & Gold 1948; Hoyle 1948). On the other hand, the *Friedmann equations*, derived as a homogeneous and isotropic solution to the Einstein field equations (Friedmann 1922; Lemaître 1927), got by without invoking the creation of new matter. This alternative idea, later dubbed the *Big Bang theory* by Fred Hoyle, instead postulated that the Universe started out in an initial hot dense state. Subsequently,

as the Universe spatially expanded, matter would dilute and the Universe would cool until it reached its present state. Irrefutable evidence for the latter theory ultimately arrived in 1964 with the discovery of the *cosmic microwave background* (CMB) radiation (Penzias & Wilson 1965), a nearly ideal blackbody spectrum, observed isotropically across the entire radio sky, that could convincingly be explained only as the result of a *Big Bang*.

While the discovery of the CMB had settled the fundamental debate between steady-state and Big Bang theory, many details were still yet to be explained. In the following years, the hypothesis of the existence of a *dark*, unseen type of matter gained new attention. Originally postulated significantly earlier (Thomson 1904; Poincare 1906; Kapteyn 1922; Oort 1932; Zwicky 1933), new observations of galaxy rotation curves (Rubin & Ford 1970) and other probes suggested (Ostriker & Peebles 1973; Einasto et al. 1974) that galaxies would have to be surrounded by *halos* of dark matter – an order of magnitude larger in extent than the galaxies themselves – and that this unknown type of matter would even make up the largest fraction of all matter in the Universe. At the present day, the existence of dark matter has become an integral assumption for our entire understanding of cosmology (e.g. Davis et al. 1985), and the Standard Model of Particle Physics has been extended in numerous directions to predict theoretical candidate particles that could account for this unseen type of matter. However, despite considerable efforts to detect dark matter on the particle-level, so far all searches have remained fruitless (e.g. Abdalla et al. 2022).

A further paradigm shift followed no more than 30 years ago, in 1998, when observations of Type Ia Supernovae (Riess et al. 1998; Perlmutter et al. 1999) suggested that our Universe is currently expanding at an *accelerated* rate. Explaining this behaviour requires the postulation of a further *dark* contributor to the Universe's total energy density: *dark energy* – whose nature is by no means better understood than that of dark matter. In the simplest interpretation, dark energy is nothing else but a *cosmological constant* Λ, originally introduced by Albert Einstein in his field equations of general relativity to keep the Universe static (Einstein 1917). Together with baryonic matter, dark matter, photons and neutrinos, dark energy forms the main ingredient of the ΛCDM *model* that has seen an unprecedented triumph in the recent decades, fitting observations at a remarkable level of precision (e.g. Planck Collaboration et al. 2020). The current cosmological best-fit value for Λ, however, is not less than 120 orders of magnitude away from the quantum field theory-based prediction from first principles, suggesting fundamental limitations in our understanding of the physics behind Λ (Weinberg 1989). At the same time, cosmologists keep challenging the simple picture of a cosmological constant from an observational perspective, conducting ever better surveys to search for evidence that the dark energy equation of state must be evolving over time (e.g. Abdul Karim et al. 2025), leaving vast space for new discoveries even after the revolutionary past century.

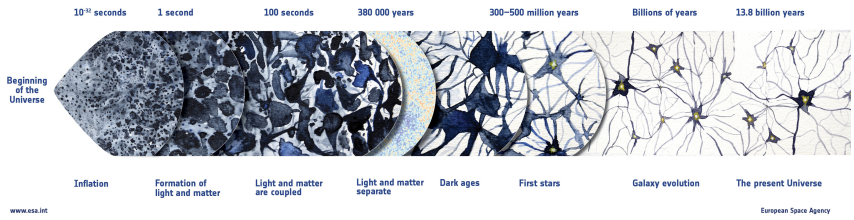

Figure 1.1: Schematic timeline of Universe's evolution history according to the ΛCDM model. Major events that took place over the past 13.8 billion years are labelled on the timeline, and the schematic drawings show the qualitative state of cosmic structure at the respective point in time. The Epoch of Reionisation is heralded by the formation of the first stars and galaxies, approximately within the first billion years. Credit: ESA.

1.1 The first billion years of our Universe

Despite the fact that its most prominent components are far from being understood, and that the ΛCDM model itself might not even be the final word in our understanding of cosmology, it provides an excellent fit to present observations and allows us to tell in remarkable detail the story of our Universe's past 13.8 billion years, following the timeline sketched in Figure 1.1. This thesis will contribute to telling the tale of its first billion years, and we start this chapter by telling the prologue.

1.1.1 The first seconds

The very beginnings of the Universe, or more specifically, its first $t_{\rm Pl} \sim 5 \times 10^{-44}\,\mathrm{s}$ will remain untold for the time being, at least for as long as quantum field theory and the theory of general relativity have not been merged into a unified theory of quantum gravity that would allow us to describe this regime. However, once the Planck time $t_{\rm Pl}$ has passed and general relativity enters its reign of validity, the evolution of spacetime – defined by the metric tensor $g_{\mu\nu}$ – is governed by the *Einstein field equations*

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{1.1}$$

relating the curvature of spacetime, quantified by the Einstein tensor $G_{\mu\nu}$, to the matter components of the Universe, summarised in the stress energy tensor $T_{\mu\nu}$ (Einstein 1916). Despite the fact that we can technically describe their *evolution* from first principles, we are still faced with an *initial conditions* problem: only a highly fine-tuned choice of parameters governing the Universe's initial curvature and matter content would result in a Universe as flat, homogeneous and isotropic as the one that we are living in today.

Cosmic Inflation, an early phase of *accelerated* spatial expansion of the Universe, has been proposed as a solution to these issues (Guth 1981; Linde 1982; Albrecht & Steinhardt 1982). Driven by one or more scalar fields, their contribution comes to dominate over all other stress energy components, smoothing and flattening the Universe. During such an inflationary epoch, physical distances grow exponentially faster than the Hubble radius whose size governs which scales are in causal contact at a given point in time. As a result, the previously smoothed modes exit the horizon until eventually entering again in the subsequent hot expansion phase, explaining the smoothness of the Universe at the present day, even on its largest scales that would otherwise never have been in causal contact before. At the same time, small inflationary quantum fluctuations provide the seeds for the subsequent formation of the cosmic large-scale structure that shapes our Universe today. The excellent agreement of these inflationary predictions with observations of the CMB (Planck Collaboration et al. 2020) are seen as strong evidence that our Universe indeed went through an inflationary phase in its first $\sim 10^{-33} - 10^{-32}$ s. Inflation has to end with a *reheating* process where the inflationary scalar field(s) decay and deposit their stress energy into a thermal bath of standard model particles.

A vast range of models can drive an inflationary epoch, sharing the above generic predictions. The fact that many of the simplest models, requiring the least amount of fine tuning, have already been ruled out has sparked new debate (Ijjas et al. 2013) and led to the development of alternative theories such as *bouncing models* (Ijjas & Steinhardt 2016) where a period of *slow contraction* would precede the standard phase of hot expansion and similarly solve the problems that inflationary scenarios are aiming to address (e.g. Cook et al. 2020; Kist & Ijjas 2022). B-mode polarisation patterns detected in future CMB experiments (e.g. Ade et al. 2019), predicted by many inflationary but not by bouncing models, are expected to provide new evidence on this matter in the coming years.

In either case, in what follows – with its initial conditions set – the Universe enters its hot Big Bang expansion phase, starting in a *radiation-dominated* epoch where relativistic components of the primordial plasma (i.e., photons and neutrinos) dominate the total energy density. The initially enormous temperature begins to decline as the Universe expands, inducing the successive breaking of Standard Model symmetries into the four known fundamental forces, and the emergence of massive elementary particles via the Higgs mechanism (Higgs 1964; Englert & Brout 1964; Guralnik et al. 1964) during the *electroweak phase transition* after 10^{-12} s, and later that of hadrons such as protons and neutrons during the *quantum chromodynamics (QCD) phase transition* (10^{-5} s after the Big Bang). After 1 second, the point is reached where neutrinos *decouple* from the rest of the primordial plasma, and freely stream through the Universe until the present day, forming the so-called *cosmic neutrino background (CνB)*. As the Universe cools further,

particles and their anti-particles annihilate once their respective mass (or, equivalently, energy) scale is reached, starting with hadronic particles and later continuing with leptons until at around 10 s after the Big Bang, electrons and positrons have annihilated as well. A major remaining conundrum about these (a priori symmetric) *baryogenesis* and *leptogenesis* processes is how the present-day matter-antimatter asymmetry could come into place.

More clarity exists about the subsequent nuclear fusion of light chemical elements (deuterium, helium, lithium and other light isotopes) during *Big Bang nucleosynthesis* (BBN, Alpher et al. 1948). This process began once the universe's temperature fell below the binding energies of the lightest elements and continued as long as a sufficient temperature for nuclear fusion was retained, resulting in a characteristic abundance of these primordial elements by a time of around 20 minutes after the onset of the hot Big Bang.

1.1.2 The first millennia

The evolution that follows upon these turbulent first minutes takes place on significantly longer – cosmological – timescales. After around 47 000 years, at a redshift of $z \sim 3400$, *matter-radiation equality* was reached where matter superseded radiation as dominant component of the Universe's total energy density. At that point, photons, protons and electrons were still in thermal equilibrium, forming an ionised plasma opaque to radiation due to frequent Thomson scattering events with the numerous free electrons. This changed around 370 000 years after the Big Bang, at a redshift of $z \sim 1100$, when the binding energy of neutral hydrogen was reached such that protons and electrons could *recombine* and form the first atoms of neutral hydrogen.

Photons, due to the sudden absence of free electrons to scatter off, now had a mean free path that exceeded the horizon distance such that they *decoupled* and started travelling freely for the first time in cosmic history. The Universe had become *transparent*. Since these photons had been in thermal equilibrium with the rest of the primordial plasma right until these stages, and they travelled freely ever since, they still form a nearly perfect blackbody spectrum, reflecting the state of the Universe at this *surface of last scattering*. Detected for the first time in 1965 by Penzias & Wilson (1965), this Cosmic Microwave Background (CMB) radiation – now redshifted to microwave frequencies – can be seen as an early echo of the Big Bang, reflecting in minute detail the state of the Universe at $z \sim 1100$ and acting as our single most important source of information not only about these early stages of our Universe but also many later epochs, based on the marks they leave on the CMB photons. Most prominently, besides its overall isotropy and smoothness, the CMB contains small temperature variations, reflecting the state of density perturbations in the early Universe that would keep growing in the matter-dominated epoch and serve as seeds for the formation of cosmic structure that we observe today.

1.1.3 The first billion years

Despite its newly found transparency, the Universe subsequently entered what we call its *dark ages*, owing to the fact that the recently released CMB photons quickly redshifted away from visible frequencies into the infrared. The only other source of radiation at that time were microwave photons emitted via the 21 cm spin flip transition of the recently recombined neutral hydrogen atoms. As such, the Universe remained devoid of visible light until eventually the first stars would form out of the gradually growing primordial density perturbations. Indirect evidence for the existence of these pristine, metal-poor *population III stars* exists in form of traces of otherwise inexistent heavy elements in quasar emission spectra, but direct observational proof of their existence is still a long time coming (see e.g. Bromm 2013).

The cosmic dark ages gradually came to an end when cosmic structure formation gathered pace, resulting in a sharp increase in the presence of luminous structure. This process is decisively supported by the presence of *dark matter*, outnumbering ordinary matter by a factor of $5-6$ and only interacting gravitationally with standard model particles. As such, it forms a hierarchical structure of filaments and halos that we call the *cosmic web*, enshrouding vast low-density cosmic voids. These *halos* of dark matter act as supporting structures where ordinary matter can collapse further into *stars and galaxies*. Indeed, the James Webb Space Telescope (JWST) has already enabled the discovery of galaxies out to redshifts of $z \sim 14$ (e.g. Carniani et al. 2024; Witstok et al. 2025), no more than 300 million years after the Big Bang. The light from these new luminous sources gradually ended the cosmic dark ages and initiated the cosmic era that is the subject of this thesis, the *Epoch of Reionisation*.

1.2 Reionisation – Les Lumières cosmologiques

Many cosmic epochs have been termed with distinctively pictorial names, some of them even in remarkable analogy to eras of the more recent *human* history. Coincidence or not, it is in any case worthwhile to fully appreciate the metaphor of the cosmic *dark ages* and the period that logically has to follow. In the previous millennium, the *Age of Enlightenment* ended the *medieval* dark ages when individuals emerged from their self-imposed immaturity (Kant 1784). In a similar vein, the Epoch of Reionisation – the 'Age of *Cosmic* Enlightenment' – is shaped by the formation of the first *astrophysical* individuals. These first stars, galaxies and quasars quite literally illuminated – or enlightened – our Universe, not only ending the cosmic dark ages but also causing a profound impact on the Universe's subsequent evolution by connecting the properties of galaxies and other astrophysical sources of ionising photons to the large-scale history of our

Universe and its fundamental physical properties. In other words, the Epoch of Reionisation links astrophysics to cosmology, and neither the former nor the latter can be truly understood without a firm grasp of what happened in this era.

The estates where the reionisation process unfolds are what we call the *intergalactic medium* (IGM), composed of the vast voids between galaxies – the space where most cosmic baryons actually reside. The evolution of the IGM's ionisation state is what informs us about the way reionisation proceeded and its interplay with the growth and physical properties of galaxies that drove it. The relevant physical field to that effect is the number density field of cosmic hydrogen $n_{\rm H}$, separated into the two components

$$n_{\rm H} = n_{\rm HI} + n_{\rm HII}, \tag{1.2}$$

where $n_{\rm HI}$ is the number density field of *neutral* hydrogen atoms and $n_{\rm HII}$ that of *ionised* ones. The ionisation state of the IGM is then commonly quantified in terms of the *fractional fields* $x_{\rm HI} \equiv n_{\rm HI}/n_{\rm H}$ and $x_{\rm HII} \equiv n_{\rm HII}/n_{\rm H}$. Of further relevance is the number density n_e of free electrons, given by counting the remaining ions, where – besides hydrogen – singly (He II) and doubly (He III) ionised helium are non-negligible contributors:

$$n_e = n_{\rm HII} + n_{\rm HeII} + 2\,n_{\rm HeIII}. \tag{1.3}$$

Before reviewing past and ongoing modelling efforts aiming to trace the full evolution of these fields, we start by describing how the reionisation process proceeded on a *global* level.

1.2.1 The integrated effect of reionisation

Most fundamentally, from a cosmologist's perspective, the impact of reionisation is summarised in a single free parameter, the reionisation optical depth τ_e – notably, one of *the* six free parameters of the ΛCDM model. This is because the Epoch of Reionisation leaves a characteristic imprint on the CMB due to Thomson scattering of CMB photons by the free electrons released during reionisation. Over the course of cosmic time, starting from the decoupling of the CMB at $z_{\rm CMB} \simeq 1100$, this results in an integrated optical depth

$$\tau_e = \sigma_{\rm T} \int_0^{z_{\rm CMB}} n_e \, \frac{c\,(1+z)^2}{H(z)} \, {\rm d}z, \tag{1.4}$$

where $\sigma_{\rm T} = 8\pi/3 \cdot (e^2/4\pi\epsilon_0 m_e c^2)^2$ is the Thomson scattering cross section of the electron. The result of this is a suppression of the amplitude of the CMB temperature power spectrum proportional to a factor of $e^{-2\tau_e}$, in nearly perfect degeneracy with a reduction of the primordial amplitude of scalar perturbations A_s as a further free parameter of the ΛCDM model.

However, scattering of the large-scale CMB quadrupole radiation field also introduces a linear polarization pattern that can be identified in the E-mode polarisation power spectrum of the CMB as the so-called *reionisation bump* at low multipoles l, scaling as τ_e^2 (see e.g. Reichardt 2016). Due to these effects, an understanding of reionisation is indispensable for precision constraints on other cosmological parameters based on observations of the CMB (Elbers 2025; Jhaveri et al. 2025; Sailer et al. 2026).

1.2.2 The global timing of reionisation

Moving beyond its overall integrated effect, the logical next step consists in understanding how reionisation proceeded as a function of cosmic time. Constraining this *global timing* of reionisation is the most fundamental goal of all astrophysical reionisation studies and also a main objective of this thesis. We measure its progression in terms of the global *volume-averaged* neutral and ionised fractions $\langle x_{\rm HI}\rangle$ and $\langle x_{\rm HII}\rangle$ in the IGM, conventionally defined as $\langle x_{\rm HI}\rangle \equiv \langle n_{\rm HI}\rangle/\langle n_{\rm H}\rangle$ and $\langle x_{\rm HII}\rangle \equiv \langle n_{\rm HII}\rangle/\langle n_{\rm H}\rangle$. An evolution equation for these quantities can then be derived via simple photon counting arguments in a comoving patch of the Universe with volume V (see e.g. Gnedin & Madau 2022). Let $N_{\rm ion}(t)$ be the number of ionising photons at cosmic time t in that given patch, then the number of ionised hydrogen atoms $N_{\rm HII}(t)$ is equal to $N_{\rm ion}(t)$ minus the number of those photons that recombine (on a given recombination time scale $t_{\rm rec}$):

$$N_{\rm HII}(t) = N_{\rm ion}(t) - \int_0^t N_{\rm HII}\,\frac{{\rm d}t'}{t_{\rm rec}}. \tag{1.5}$$

Averaging over a cosmologically representative volume V to take the limit $N_i/V \to \langle n_i\rangle$ for any of the aforementioned indices i, and differentiating with respect to t, we can derive the well-known Madau reionisation equation

$$\frac{{\rm d}\langle x_{\rm HII}\rangle}{{\rm d}t} = \frac{\dot{n}_{\rm ion}}{n_{\rm H,0}} - \frac{\langle x_{\rm HII}\rangle}{t_{\rm rec}} \tag{1.6}$$

that describes, on a global level, the transition of cosmic hydrogen from neutral to completely ionised (Madau et al. 1999).

It also emphasises the prominent distinction between *sources* and *sinks* of reionisation, embodied by the two distinct terms of the equation, whose nature is still subject to active investigation. Current consensus is that galaxies – as opposed to more powerful sources such as active galactic nuclei (AGN) and quasars – are the dominating sources of ionising photons, driving the reionisation process (Madau & Dickinson 2014; Bouwens et al. 2015; Robertson et al. 2015; Finkelstein et al. 2019) (although JWST observations of early galaxies and AGN have been sparking new debate; see e.g. Madau

et al. 2024; Muñoz et al. 2024; Simmonds et al. 2024). Under this assumption, the source term in Eq. (1.6) is commonly parameterised into the three factors

$$\dot{n}_{\rm ion} = f_{\rm esc}\,\xi_{\rm ion}\,\rho_{\rm SFR}, \tag{1.7}$$

where $\rho_{\rm UV}$ is the UV luminosity density, $\xi_{\rm ion}$ the ionising photon production efficiency, i.e., the number of ionising photons emitted per unit UV luminosity, and $f_{\rm esc}$ is the fraction of these photons escaping the interstellar medium (ISM) and circumgalactic medium (CGM) and thus actually ionising hydrogen atoms in the diffuse IGM. While UV luminosity functions are well-determined (Bouwens et al. 2021; Adams et al. 2024; Donnan et al. 2024; Robertson et al. 2024), $\xi_{\rm ion}$ and $f_{\rm esc}$ are significantly harder to constrain and still remain an active research subject (Endsley et al. 2023; Atek et al. 2024; Muñoz et al. 2024; Saxena et al. 2024; Simmonds et al. 2024).

Besides these sources of ionising photons, we also have to account for recombinations via the sink term in Eq. (1.6). Their timescale $t_{\rm rec}$ is impacted by the actual gas density distribution whose effect we can absorb in the *clumping factor* that is commonly defined as

$$C_{\rm HII} \equiv \frac{\langle n_{\rm HII}^2 \rangle}{\langle n_{\rm HII} \rangle^2}. \tag{1.8}$$

Its value has to be estimated externally via hydrodynamical simulations, commonly resulting in values of $C_{\rm HII} \lesssim 3$ at $z \gtrsim 6$ (Gnedin & Ostriker 1997; Madau et al. 1999; Pawlik et al. 2009; Kaurov & Gnedin 2014; So et al. 2014). With a recombination rate $\dot{n}_{\rm rec} = \alpha_{\rm HII}^{\rm A} \langle n_e\, n_{\rm HII} \rangle \sim \alpha_{\rm HII}^{\rm A} \langle n_{\rm HII}^2 \rangle = \alpha_{\rm HII}^{\rm A}\, C_{\rm HII} \langle n_{\rm HII} \rangle^2$, this then amounts to a recombination timescale

$$t_{\rm rec} = \frac{\langle n_{\rm HII} \rangle}{\dot{n}_{\rm rec}} = \frac{1}{\alpha_{\rm HII}^{\rm A} \langle n_{\rm HII} \rangle\, C_{\rm HII}}, \tag{1.9}$$

where $\alpha_{\rm HII}^{\rm A}$ is the Case A recombination coefficient, accounting for recombinations into all excited states *and* the ground state (as appropriate for reionisation modelling; see Kaurov & Gnedin 2014).

1.2.3 The topology of reionisation

While Eq. (1.6) provides us with a basic analytical prescription of the temporal progression of reionisation on a global level, it is far from explaining reionisation to a full degree. This is because reionisation proceeds in a highly inhomogeneous, *patchy* way. In other words, the $x_{\rm HI}$ field is largely binary, with either a neutral or an ionised value at a given point in space. The composition of such neutral and ionised patches determines the *topology* of reionisation which describes the way reionisation proceeded on a spatial level and is thus also highly informative about the nature of the sources and sinks of ionising radiation.

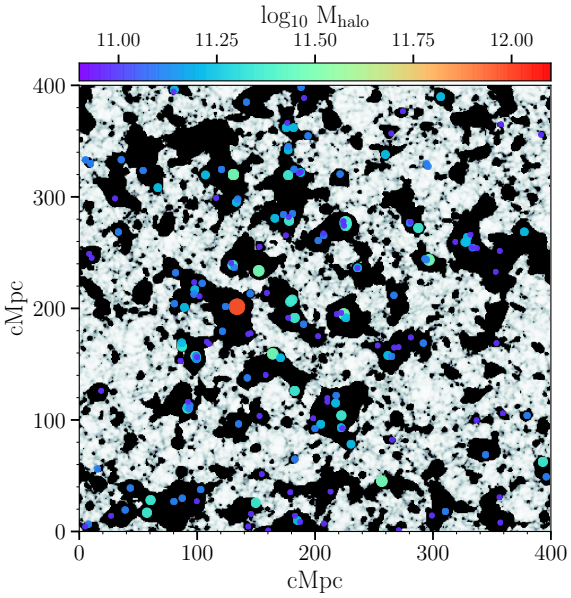

Figure 1.2: Two-dimensional slice of a simulated ionisation field $x_{\rm HI}$ at redshift $z = 7.5$ with a global volume-average neutral fraction of $\langle x_{\rm HI} \rangle = 0.5$, generated with the semi-numerical 21cmFAST code (Mesinger et al. 2011). White regions are neutral, black ones ionised. Dots denote the locations of the most massive halos within a 4 cMpc thick slice, color- and size-coded by their mass. Credit: Davies et al. (2018a).

Historically, two options have been considered for the overall progression of reionisation. First, *outside-in* scenarios, where photons – if they managed to escape from the dense host halos where the ionising sources are residing – would firstly ionise underdense regions of the IGM due to their modest need for ionising photons, whereas higher-density regions would first remain self-shielded from the ionising background until they would gradually get ionised as well (Miralda-Escudé et al. 2000).

On the other hand, motivated by the first numerical simulations of reionisation, Furlanetto et al. (2004) and Furlanetto & Oh (2005) developed a new model for the growth of ionised regions, instead favouring an *inside-out* scenario. Based on excursion-set theory (Bond et al. 1991) – in analogy to the Press-Schechter formalism for halo formation (Press & Schechter 1974) – they came to the conclusion that ionised bubbles would first form around clustered sources in overdense regions. These bubbles would then grow as time progresses, and eventually overlap and merge with other bubbles until the IGM has been ionised in its entirety. An example of a simulated topology at $\langle x_{\rm HI} \rangle = 0.5$, building upon this model, is depicted in Figure 1.2.

Improved numerical simulations – as well as observational hints that Lyman-α emitting galaxies (LAEs) at high redshift appear to be highly clustered in what would correspond to overdense, highly ionised regions (Hu et al. 2021; Trapp et al. 2023) – have led us to conclude that reionisation indeed proceeds largely in an inside-out fashion. Cosmic filaments, on the other hand, due to their lack of ionising sources at increased densities compared to cosmic voids, might be the last regions to reionise in accordance with the original outside-in argument (Kaurov 2016). Further details about the sizes of ionised bubbles and the topology of reionisation are yet to be determined, and as such, a large body of research is still concerned with these topics (e.g. Mason & Gronke 2020; Hayes & Scarlata 2023; Lu et al. 2024b; Witstok et al. 2024).

1.3 Modelling Reionisation

Modelling reionisation is a highly non-trivial task due to the plethora of different physical effects and scales involved. As such, *numerical simulations* of reionisation have been developed in many different flavours of varying complexity, owing to the difficulty of modelling the radiative processes occurring in this era jointly with the gravitational evolution of dark and baryonic matter, forming the large-scale structure of the Universe.

Efforts to simulate the cosmological evolution of our Universe started out with pure *N-body simulations* focussing exclusively on the dynamics of collisionless dark matter that would collapse into what we call the *cosmic web* (e.g. Davis et al. 1985; Springel et al. 2005). As computational resources grew, it became possible to additionally include the fluid-like dynamics of baryonic matter, as well as a growing number of *sub-grid prescriptions* for varying types of *feedback* effects from physical processes occurring below the resolution limit of these simulations, for example due to star formation or active galactic nuclei (AGN), matching the observed populations of galaxies at a remarkable level of detail and enabling unique new insights into the physical processes behind their formation and evolution (e.g. Vogelsberger et al. 2014; Schaye et al. 2015; Pillepich et al. 2018; Davé et al. 2019; Pakmor et al. 2023; Schaye et al. 2023).

Nonetheless, accounting for *all* physical effects that are at play – even if not modelled from first principles – would greatly exceed the computational resources available to us. As such, the selection of the physical effects included in the modelling has to be tailored to the cosmological and astrophysical observables to compare the simulations with. In the context of reionisation, this prominently requires accounting for the radiative effects of the ionising sources. The required *radiative transfer* calculations are expensive, especially when performing full *radiation-hydrodynamical simulations* where the radiative effects are self-consistently coupled to a conventional

hydrodynamical simulation tracing the evolution of dark and baryonic matter (e.g. Gnedin 2014; Ocvirk et al. 2016, 2020; Garaldi et al. 2022; Kannan et al. 2022; Lewis et al. 2022). As a result, many approximate, *semi-numerical* schemes have been developed with the aim of reducing the computational complexity while still modelling reionisation at a reasonable level of precision (e.g. Mesinger et al. 2011). Before providing a quick overview over these approaches, we build intuition by discussing the basic underlying evolution equations that need to be solved.

1.3.1 Reionisation equations

To model reionisation, we turn back to the neutral fraction field x_{HI} introduced in Section 1.2. Rather than describing the global evolution of its volume-averaged version $\langle x_{\mathrm{HI}} \rangle$, governed by Eq. (1.6), we now aim to construct an evolution equation for the field itself. We restrict this section to a basic set of simplified modelling equations, illustrating the most important contributions, and point the reader to Gnedin & Madau (2022) for a more comprehensive review of past and ongoing modelling approaches. We start by following the same reasoning as in Section 1.2, counting ionisation and recombination events to construct a simple rate equation with source and sink terms analogous to Eq. (1.6).

The atoms in a gas of neutral hydrogen can be ionised in two ways, via *photoionisation* and *collisional ionisation*, characterised by the ionisation rates Γ_{HI} and $\Gamma_{e\mathrm{HI}}$. *Photoionisation* occurs if a Lyman continuum photon, i.e., a photon with wavelength shorter than the Lyman limit $\lambda_{\mathrm{HI}} = 911.3\,\text{Å}$ (corresponding to an energy greater than $h\nu_{\mathrm{HI}} = 13.6\,\mathrm{eV}$), encounters a neutral hydrogen atom. We characterise the underlying radiation field by its (space- and angle-averaged) specific intensity J_ν whose evolution is described by the *radiative transfer equation*

$$(\partial_t - \nu\, H(t)\, \partial_\nu)\, J_\nu + 3H(t)\, J_\nu = -c\,\kappa_\nu\, J_\nu + \frac{c}{4\pi}\,\epsilon_\nu, \tag{1.10}$$

where ϵ_ν is the proper space-averaged volume emissivity and k_ν is the absorption coefficient (see e.g. Peebles 1993; Haardt & Madau 1996, 2012).

For a given radiation field, we can then obtain the *photoionisation rate* Γ_{HI} of neutral hydrogen by integrating the product of the number $J_\nu/h\nu$ of photons at frequency ν and the HI absorption cross section $\sigma_{\mathrm{HI}}(\nu)$ over all Lyman continuum frequencies:

$$\Gamma_{\mathrm{HI}} = 4\pi \int_{\nu_{\mathrm{HI}}}^{\infty} \frac{J_\nu}{h\nu}\, \sigma_{\mathrm{HI}}(\nu)\, \mathrm{d}\nu. \tag{1.11}$$

Collisional ionisation of neutral hydrogen atom is caused by thermal electrons, and hence its rate is given by

$$\Gamma_{e\mathrm{HI}} = n_e \int_{\nu_{\mathrm{HI}}}^{\infty} \sigma_{e\mathrm{HI}}(v)\, v\, f(v)\, \mathrm{d}v, \tag{1.12}$$

where $f(v)$ is the Maxwell–Boltzmann distribution of the electron velocity v, and $\sigma_{e\mathrm{HI}}(v)$ is the collisional ionization cross section.

Recombination, i.e., the capture of a free electron by a proton, takes place at a rate given by the product of the free electron density n_e and the Case A recombination coefficient $\alpha^{\mathrm{A}}_{\mathrm{HII}}$ as introduced in Section 1.2.2, given by the sum

$$\alpha^{\mathrm{A}}_{\mathrm{HII}} = \sum_{i=1}^{\infty} \sum_{l=0}^{n-1} \int_{\nu_{\mathrm{HI}}}^{\infty} \sigma_{nl}(v)\, v\, f(v)\, \mathrm{d}v, \tag{1.13}$$

where $\sigma_{nl}(v)$ is the recombination cross section into the state characterised by the quantum numbers n and l.

Putting together all ionisation and recombination contributions, the full rate equation for the evolution of the neutral (or ionised) fraction field reads

$$\frac{\mathrm{d}x_{\mathrm{HI}}}{\mathrm{d}t} = -x_{\mathrm{HI}}\,\Gamma_{\mathrm{HI}} - x_{\mathrm{HI}}\,\Gamma_{e\mathrm{HI}} + \alpha^{\mathrm{A}}_{\mathrm{HII}}\, n_e\, x_{\mathrm{HII}}. \tag{1.14}$$

Although we omit them here for brevity, a full set of modelling equations would include analogous evolution equations for the helium fractions x_{HeI}, x_{HeII} and x_{HeIII} (see e.g. Bolton & Haehnelt 2007a).

An additional important consequence of the photoionisation process is that the IGM gets *photoheated* to temperatures of several thousands of Kelvin at a photoheating rate of

$$\mathcal{H}_{\mathrm{HI}} = 4\pi\, n_{\mathrm{HI}} \int_{\nu_{\mathrm{HI}}}^{\infty} \frac{J_\nu}{h\nu}\, (h\nu - h\nu_{\mathrm{HI}})\, \sigma_{\mathrm{HI}}(\nu)\, \mathrm{d}\nu, \tag{1.15}$$

summarised along with all other sources of heating in the total heating rate $\mathcal{H}$. This means that we simultaneously have to trace the *thermal evolution* of the IGM based on all relevant heating and cooling sources, commonly modelled via the equation

$$\frac{\mathrm{d}T}{\mathrm{d}t} = -2H(t)\,T + \frac{2T}{3\Delta}\,\frac{\mathrm{d}\Delta}{\mathrm{d}t} + \frac{T}{\mu}\,\frac{\mathrm{d}\mu}{\mathrm{d}t} + \frac{(\gamma-1)\mu}{k_{\mathrm{B}} n}(\mathcal{H} - \Lambda), \tag{1.16}$$

where Δ is the cosmic overdensity field, μ the mean molecular weight and n the number density of all free baryonic particles (i.e., including ions *and* electrons), $\gamma = 5/3$ is the adiabatic index, and $\mathcal{H}$ and Λ are the total heating and cooling rates, respectively (see e.g. Hui & Gnedin 1997).

This completes the set of the most important modelling equations of the radiative transfer problem during the reionisation era. Note that we are faced with a highly non-trivial set of coupled partial differential equations. Specifically, we have to solve the radiative transfer transfer equation (1.10), governing the intensity of the ionising sources, jointly with the evolution equations of all ionised species (see Eq. (1.14) for hydrogen), coupled to each other via the photoionisation rate Γ_{HI} and the absorption coefficient

κ_ν. These equations, in return, are also coupled to the thermal evolution equation (1.16) via the heating and cooling rates $\mathcal{H}$ and Λ. Conversely, the temperature evolution back-reacts on the collisional ionisation rate $\Gamma_{e\mathrm{HI}}$ and the recombination coefficient $\alpha^{\mathrm{A}}_{\mathrm{HII}}$ by changing the thermal distribution of electrons $f(v)$. On top of this, everything is coupled to the evolution of dark and baryonic matter, governing the overdensity field Δ, as well as other astrophysical feedback effects that may have an impact on any of these fields. The solution of this coupled system of PDEs therefore poses a highly challenging numerical task.

1.3.2 Reionisation simulations

Self-consistent, fully coupled *radiation-hydrodynamical simulations* are considered the holy grail of these attempts, jointly modelling all of the aforementioned effects. Important examples of such simulation suites are CROC (Gnedin 2014), Thesan (Garaldi et al. 2022; Kannan et al. 2022) and CoDa (Ocvirk et al. 2016, 2020; Lewis et al. 2022), or the recently developed Nyx-RT code (Marshak et al. 2025) that is poised to extend the hydrodynamical Nyx simulations (Almgren et al. 2013; Lukić et al. 2015) to include radiative transfer. However, the computational cost of these simulations is considerable. As such, the underlying radiative transfer schemes often have to remain approximate, and the box sizes ($\sim 100\,\mathrm{cMpc}$) and resolutions ($\sim 0.1 - 1\,\mathrm{pMpc}$ at $z = 6$) of these simulations are not able to compete with those of their purely hydrodynamical counterparts. In addition, and most crucially, this means that most simulation suites are limited to at most a handful of high-resolution flagship runs, hindering a systematic exploration of the effects of variations in cosmological parameters and/or the timing and topology of reionisation.

A less costly alternative to such simulations are *partially coupled simulations* where the aforementioned modelling components are still included but not all of them are fully coupled to each other, for example by performing radiative transfer in post-processing on previously completed hydrodynamical simulations such as done in the C2-Ray simulations (Iliev et al. 2006; Mellema et al. 2006), or by Aubert & Teyssier (2010) with their Aton code, recently revised to also include helium (Asthana et al. 2024). A further example are the Sherwood-Relics simulations (Puchwein et al. 2023), building upon the Sherwood simulations (Bolton et al. 2017) and including a multi-step hybrid technique for treating radiative transfer effects to capture the effects of patchy reionisation on the IGM, allowing for a wider exploration of astrophysical and cosmological parameter variations.

Semi-numerical models are an even more efficient category of techniques that do not require cosmological (hydrodynamical or N-body) simulations to model the full non-linear distribution of cosmic matter but rather evolve the Gaussian field of cosmic initial conditions directly via perturbation

theory. The prime example of these techniques is the 21cmFAST code (Mesinger et al. 2011) which predicts the distribution of ionised gas at a given redshift via an excursion set-based formalism (Furlanetto et al. 2004; Furlanetto & Oh 2005) from the Zel'dovich approximation (Zel'dovich 1970). An alternative example is the Abundance Matching Box for the Epoch of Reionization (AMBER) that instead determines the reionisation-redshift field from a certain base field such as the angle-averaged specific intensity J_ν via an abundance matching approach (Trac et al. 2022). While being approximate, these methods come with the decisive advantage of allowing for cheap simulations of multiple large boxes with varying reionisation histories as needed for cosmological parameter inference.

This also motivates the hybrid approach for the simulation models adopted in this thesis: in order to simulate large cosmological volumes including the rarest overdensities where quasars are assumed to reside, and to explore variations with respect to the volume-averaged neutral fraction $\langle x_{\mathrm{HI}} \rangle$ of the IGM, we make use of the efficient semi-numerical 21cmFAST code to simulate an array of different ionisation topologies. However, for an optimal modelling of the diffuse IGM and its impact on the Lyman-α forest observed in high-redshift quasar spectra, we combine these semi-numerical topologies with density, velocity and temperature fields from the Nyx hydrodynamical simulations that were designed specifically for this purpose. Finally, as we only have to model single quasar sightlines, the radiative transfer problem can significantly be simplified to a single dimension. Hence, we perform radiative transfer in post-processing on the sightlines that have been extracted from the aforementioned hydrodynamical simulations and ionisation topologies. This illustrates how the complexity of modelling reionisation is often broken down via custom-built approaches tailored to the problem at hand.

1.4 Quasars – Les Lumières astrophysiques

Having obtained an overview over the cosmological epoch that we would like to study, as well as the state of our modelling efforts to that end, we now turn to the observational probes – *the astrophysical lights* – that provide a means to gain this understanding. Yet another time, the first relevant historical developments date back to the previous century, to the late 1950s, when radio astronomers detected unusually bright, point-like *quasi-stellar radio sources* (see e.g. Edge et al. 1959), dubbed *quasars* (Chiu 1964), whose nature remained unclear at first. A breakthrough followed in 1963 in the form of the first detection of a – still point-like – visual counterpart to one of these sources which suggested an unusually high redshift of $z = 0.158$, placing the object *at a cosmological distance* despite its star-like appearance (Hazard et al. 1963; Schmidt 1963). Within the next few decades, this interpretation

was actively debated, until eventually consensus was reached that quasars were indeed as distant and powerful sources as originally suggested, making them *the brightest (non-transient) astrophysical lights in the entire Universe*, appearing on the sky as *quasi-stellar objects* (QSOs) despite their distance of millions of parsecs from our own galaxy.

An explanation for their extraordinary power was provided one year later by Salpeter (1964) by invoking the (at that point by no means proven) existence of *black holes*. In that picture, the quasar radiation would be released by gas in an accretion disc of gas falling onto a supermassive black hole (SMBH) in the centre of a galaxy. While highly controversial in these years, especially the first images of *shadows* of black holes from the Event Horizon Telescope (Event Horizon Telescope Collaboration et al. 2019, 2022) have convinced us of their existence, and that such a supermassive black hole should reside in the centre of any larger galaxy, many of which are shining as active galactic nuclei (AGN) and the most powerful ones as quasars (e.g. Padovani et al. 2017).

1.4.1 The Lyman-α transition

To date, quasars can be considered as our main source of information about the Epoch of Reionisation, owing to the fact that these objects already existed at sufficiently early cosmic stages, and to the exquisite sensitivity of their spectra to imprints from reionisation.

Particularly sensitive in this regard is the highly resonant *Lyman-α transition*. A Lyman-α photon corresponds to the transition of a neutral hydrogen atom between its 1s ground state and the 2p excited state at a wavelength of $\lambda_{\mathrm{Ly}\alpha} = 1215.67\,\text{Å}$. Its cross section can be calculated from first principles (see e.g. Heitler 1954; Bethe & Salpeter 1957; Karzas & Latter 1961; Berestetskii et al. 1971), resulting in

$$\sigma_\alpha(\nu) = \frac{\pi e^2}{4\pi\varepsilon_0 m_e c} f_\alpha \, \phi_\alpha(\nu), \tag{1.17}$$

where the oscillator strength $f_\alpha = 2^{13}/3^9 \simeq 0.416$ follows from the corresponding quantum-mechanical matrix element, and $\phi_\alpha(\nu)$ is the normalised line profile function, accounting for *line broadening*.

Contributions from two distinct line broadening mechanisms are relevant in the case at hand, resulting in the common expression of ϕ_α as a *Voigt profile* ϕ_{Voigt} (Voigt 1912). First, the line is *broadened naturally* due to the Heisenberg uncertainty principle (Heisenberg 1927) since energy and lifetime of the excited state cannot both be sharply determined at the same time. The consequence of this is a Lorentzian profile

$$\phi_{\mathrm{Lor}}(\nu) = \frac{R_\alpha/(\pi\nu_\alpha)}{(\nu/\nu_\alpha - 1)^2 + R_\alpha^2} \tag{1.18}$$

with the factor $R_\alpha = \Gamma_\alpha / 4\pi\nu_\alpha$ determined by the decay constant $\Gamma_\alpha = 6.265 \times 10^8\,\mathrm{s}^{-1}$ of the Lyman-α transition that is inversely related to the lifetime of the 2p state.

Second, thermal motion of the gas described by a Maxwell-Boltzmann velocity distribution causes *Doppler broadening* according to a Gaussian profile

$$\phi_{\mathrm{Gauss}}(\nu) = \frac{1}{\sqrt{2\pi\sigma^2_{\nu_\alpha}}}\, e^{-(\nu-\nu_\alpha)^2/2\sigma^2_{\nu_\alpha}}, \tag{1.19}$$

where the gas temperature T sets the profile's width $\sigma_{\nu_\alpha} = \sqrt{\frac{k_\mathrm{B}T}{m_e c^2}}\,\nu_\alpha$. The full Voigt profile ϕ_{Voigt} is then obtained by convolving ϕ_{Lor} and ϕ_{Gauss}.

1.4.2 The Lyman-α forest

For quasars, the Lyman-α transition is relevant in both emission and absorption: a strong Lyman-α emission line is produced due to the highly energetic, ionising UV radiation in the accretion disk, and the resulting recombination and resonant scattering events. As the radiation from the quasar propagates into the IGM, a characteristic absorption signature is imprinted upon its spectrum, the *Lyman-α forest*. This is because any cloud of neutral hydrogen that intersects our line of sight to the quasar absorbs those photons that have been redshifted to the Lyman-α resonance by the time they reach that cloud. The result is a *forest* of absorption and transmission spikes tracing the locations where there were and where there were no such clouds of neutral hydrogen present. This makes the Lyman-α forest an exceptional tool of modern cosmology, probing the cosmic large-scale structure via baryon acoustic oscillations (BAO) (McDonald 2003; Busca et al. 2013; Slosar et al. 2013; Font-Ribera et al. 2014; Delubac et al. 2015; du Mas des Bourboux et al. 2017, 2020; Alam et al. 2021; Adame et al. 2025) and the flux power spectrum (Croft et al. 1998; McDonald et al. 2000; Croft et al. 2002; Seljak et al. 2005, 2006; McDonald et al. 2006; Palanque-Delabrouille et al. 2013, 2015; Chabanier et al. 2019; Ivanov et al. 2025; Karaçaylı et al. 2025; Ravoux et al. 2025) whose smaller scales are also sensitive to the thermal state of the IGM (Hui & Gnedin 1997; Lidz et al. 2010; Becker et al. 2011; Boera et al. 2019; Walther et al. 2019; Gaikwad et al. 2020, 2021; Villasenor et al. 2022) and the free-streaming of dark matter (Viel et al. 2004, 2005, 2013; Garzilli et al. 2017; Iršič et al. 2017a,b; Villasenor et al. 2023; Garcia-Gallego et al. 2025b).

1.4.3 Quasars during the Epoch of Reionisation

Towards higher redshifts, when the Epoch of Reionisation has not fully ended yet, the presence of neutral hydrogen starts attenuating the Lyman-α forest

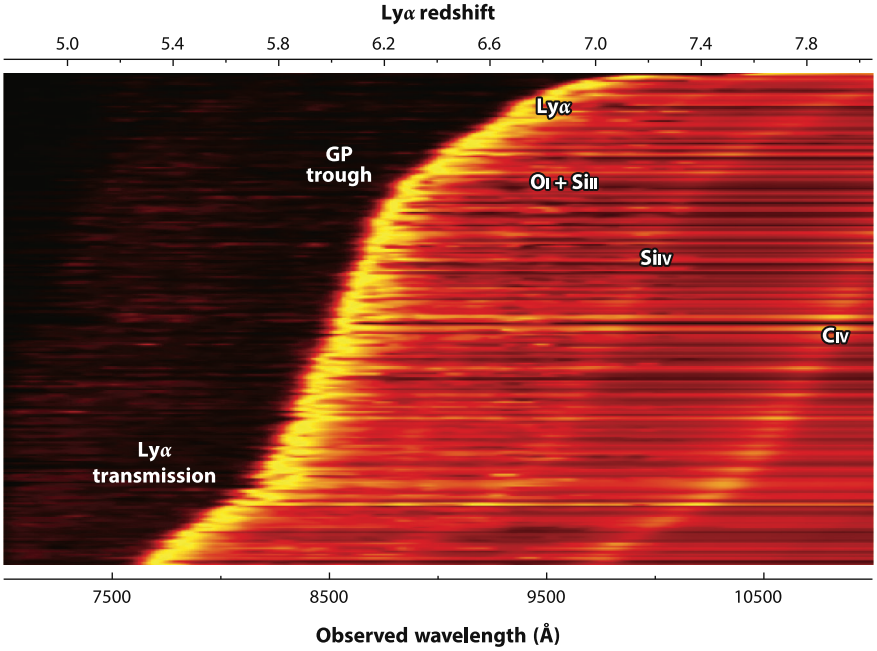

Figure 1.3: Two-dimensional representation of 527 quasar spectra at redshifts $z > 5.3$. The Lyman-α forest transmission seen blueward of the Lyman-α line gradually gets suppressed by increasingly large Gunn-Peterson troughs at higher redshifts. Credit: Fan et al. (2023).

as compared to how it appears at lower redshifts. To quantify this effect, we consider the Lyman-α transmission field

$$t(\lambda_{\rm obs}) = e^{-\tau(\lambda_{\rm obs})}, \tag{1.20}$$

where $\tau(\lambda_{\rm obs})$ is the Lyman-α optical depth at the observed-frame wavelength $\lambda_{\rm obs}$. The total optical depth $\tau(\lambda_{\rm obs})$ is obtained by integrating an infinitesimal optical depth element $\mathrm{d}\tau = \langle n_{\rm HI} \rangle \, \sigma_\alpha(\nu) \, \mathrm{d}l$ along the line-of-sight interval $\mathrm{d}l$ (e.g. Miralda-Escudé 1998; Mortlock 2016), resulting in

$$\tau_{\rm DW}(\lambda_{\rm obs}) = \int n_{\rm HI}(l) \, \sigma_\alpha(\nu) \, \mathrm{d}l. \tag{1.21}$$

1.4.3.1 The Gunn-Peterson effect

As Eq. (1.21) suggests, any enhancement of the density of neutral hydrogen along the line of sight causes increased Lyman-α attenuation. Remarkably, the Lyman-α resonance is strong enough that even the last remainders of neutral hydrogen during the final stages of reionisation are sufficient to

cause deep absorption troughs adjacent to the Lyman-α line where we would otherwise observe Lyman-α forest spikes, as first noted by Gunn & Peterson (1965). The strength of these *Gunn-Peterson (GP) troughs* can be derived via the lowest-order approximation of the Lyman-α line profile as a Dirac delta distribution

$$\phi_\alpha(\nu) = \delta(\nu - \nu_\alpha), \tag{1.22}$$

resulting in the Gunn-Peterson optical depth

$$\tau_{\mathrm{GP}}(z_{\mathrm{QSO}}) = \frac{\pi e^2 f_\alpha \lambda_\alpha}{4\pi\varepsilon_0 m_e c} \frac{\langle n_{\mathrm{HI}}\rangle(z_{\mathrm{QSO}})}{H(z_{\mathrm{QSO}})} \tag{1.23}$$

$$= 5.3 \times 10^5 \left(\frac{h}{0.68}\right)\left(\frac{1-Y}{0.76}\right)\left(\frac{\Omega_{\mathrm{b}}}{0.049}\right)\left(\frac{\Omega_{\mathrm{m}}}{0.31}\right)^{-\frac{1}{2}} \tag{1.24}$$

$$\times \left(\frac{1+z_{\mathrm{QSO}}}{1+7.54}\right)^{\frac{3}{2}} \left(\frac{H_0\sqrt{\Omega_{\mathrm{m}}(1+z_{\mathrm{QSO}})^3}}{H(z_{\mathrm{QSO}})}\right)\left(\frac{\langle x_{\mathrm{HI}}\rangle(z_{\mathrm{QSO}})}{1.0}\right)$$

that acts on all photons with $\lambda_{\mathrm{rest}} \leq \lambda_\alpha$ as they get redshifted into the Lyman-α resonance. This shows that even global volume-averaged neutral fractions of $\langle x_{\mathrm{HI}}\rangle(z_{\mathrm{QSO}}) = O(10^{-4})$ are sufficient to produce substantial GP absorption troughs. The emergence of these troughs at higher redshifts can be observed in Figure 1.3, representing 527 quasar spectra at redshifts $z > 5.3$. Measurements of τ_{GP} as well as length distributions of these absorption troughs have been among the earliest probes to place constraints on reionisation (Fan et al. 2000, 2002, 2006; Zhu et al. 2022).

1.4.3.2 Dark pixel fractions

The most straightforward – and in fact also the only model-independent – way of utilising this effect to constrain $\langle x_{\mathrm{HI}}\rangle$ as a function of redshift z is counting the fraction of *dark pixels* in high-redshift quasar spectra as proposed by Mesinger (2010). Given that any non-extinguished pixel must be ionised, whereas, given the high sensitivity of the Lyman-α transition, dark regions could still be neutral or already ionised as well, this places upper limits on $\langle x_{\mathrm{HI}}\rangle$ at a given redshift (McGreer et al. 2011, 2015; Jin et al. 2023) which can be complemented with their analogues from higher-order Lyman-series transitions (Davies et al. 2026). However, performing actual measurements of $\langle x_{\mathrm{HI}}\rangle$ rather than placing upper limits is not possible with this technique, and the derived limits quickly approach unity as soon as the absorption troughs start extending over large parts of the spectral range, limiting the constraining power of dark pixel fractions to the end stages of reionisation.

1.4.3.3 Effective optical depth distributions

More quantitative measurements are possible by not only estimating *if* – but *how much* – Lyman-α flux is effectively transmitted at a given redshift z. This, however, requires a model for the intrinsic spectrum $s(\lambda_{\rm obs})$ of the quasar, allowing us to determine the Lyman-α transmission fraction from the observed flux $f(\lambda_{\rm obs})$ as

$$t(\lambda_{\rm obs}) = \frac{f(\lambda_{\rm obs})}{s(\lambda_{\rm obs})}. \tag{1.25}$$

By evaluating this expression at a given position in the spectrum, we can obtain $t(z) = t(\lambda_{\rm obs} = (1+z)\,\lambda_\alpha))$ at the redshift corresponding to the respective observed-frame wavelength $\lambda_{\rm obs}$, conventionally averaged over length or redshift intervals of a certain width. A given spectrum thus provides us with an array of measurements of $t(z)$, extending along the redshift range that the respective line of sight is probing. We can then convert these values to an *effective Gunn-Peterson optical depth*

$$\tau_{\rm eff}(z) = -\ln t(z), \tag{1.26}$$

and the cumulative distribution function (CDF) of such $\tau_{\rm eff}(z)$ measurements along many lines of sight can then be compared to the CDF obtained from simulations (Becker et al. 2015; Bosman et al. 2018; Eilers et al. 2018; Yang et al. 2020b). This approach has recently led to the conclusion that reionisation must have ended *late*, not before $z \sim 5.3$ (Bosman et al. 2022) because the $\tau_{\rm eff}(z)$ distributions become incompatible with those obtained from homogeneous UV background (UVB) models starting at $z \gtrsim 5.3$. This suggests the presence of remaining neutral patches as late as $z \sim 5.3$ that could account for inhomogeneities in the UVB as favoured by the data.

1.4.3.4 The mean free path of ionising photons

As reionisation proceeded and ionised bubbles grew and merged, Lyman continuum photons could travel increasingly large distances before they would encounter an atom of neutral hydrogen that they would then ionise. In other words, the *mean free path* $\lambda_{\rm mfp}$ of ionising photons grew over time. Measurements of $\lambda_{\rm mfp}$ thus provide a further way of placing constraints on the timing of the reionisation era. In fact, the mean free path even forges a link between the IGM and the sources of ionising radiation since we can approximate the specific intensity of the ionising background as

$$J_\nu = \lambda_{\rm mfp}\,(1+z)^2\,f_{\rm esc}\,\frac{\epsilon_\nu}{4\pi}, \tag{1.27}$$

assuming that the comoving mean free path $\lambda_{\rm mfp}$, emitted by a population of Poisson distributed sources, is significantly smaller than the horizon

$\lambda_{\rm mfp} \ll c/H(z)$ such that cosmological expansion does not matter when solving Eq.(1.10) (see e.g. Bolton & Haehnelt 2007b; Haardt & Madau 2012). Hence the photoionisation rate follows $\Gamma_{\rm HI} \sim \epsilon\,\lambda_{\rm mfp}$ as per Eq. (1.11) for a given spectral shape of the radiation field with total ionising emissivity ϵ.

Various techniques have been deployed to measure the mean free path, such as inferring $\Gamma_{\rm HI}$ and $\lambda_{\rm mfp}$ based on models that relate these quantities to measurements of the Lyman-α optical depth (Miralda-Escude & Ostriker 1990; Meiksin & Madau 1993; Haardt & Madau 1996; Bolton & Haehnelt 2007b; Faucher-Giguère et al. 2008; Davies & Furlanetto 2016; D'Aloisio et al. 2018; Choudhury et al. 2021; Gaikwad et al. 2023). An alternative consists in directly measuring $\lambda_{\rm mfp}$ as the distance over which the transmitted ionising flux in mean quasar spectra has declined by a fraction $1/e$ (Prochaska et al. 2009; Fumagalli et al. 2013; O'Meara et al. 2013; Worseck et al. 2014; Becker et al. 2021). Lastly, the mean free path has also been constrained via the distribution of absorbers along lines of sight towards *individual* quasars (Songaila & Cowie 2010; Rudie et al. 2013; Romano et al. 2019; Bosman 2021). All these methods have lately converged on a rapid evolution of the mean free path between $5 \lesssim z \lesssim 6$ (Becker et al. 2021; Bosman 2021; Gaikwad et al. 2023), corroborating the picture of a rapid end to reionisation between these redshifts.

1.4.3.5 Damping wings

In order to use quasar spectra to obtain quantitative constraints on the Epoch of Reionisation even beyond its final stages, we can make use of the aforementioned natural broadening of the Lyman-α line which manifests as a so-called *damping wing* imprint in a quasar's spectrum when the global IGM neutral fraction $\langle x_{\rm HI} \rangle$ reaches order unity (Miralda-Escudé 1998) – and which is the subject of this thesis.

The signature is characterised by a suppression of the transmission *redward* of the Lyman-α line as depicted in Figure 1.4 for a simulated example sightline. As per Eq. (1.17), its shape and strength is determined by the Lyman-α cross section σ_α and the exact amount of neutral hydrogen along the line of sight. While technically the full Voigt profile would have to be considered, its shape is well-determined by the Lorentzian component $\phi_{\rm Lor}$. To build analytical intuition, we therefore restrict ourselves in this section to substituting Eq. (1.18) into Eq. (1.17), obtaining the damping wing optical depth

$$\tau_{\rm DW}(\lambda_{\rm obs}) = \frac{4R_\alpha \tau_{\rm GP}(z_{\rm ion})}{\pi} \left(\frac{\lambda_{\rm obs}}{(1+z_{\rm ion})\lambda_\alpha}\right)^{3/2} \times \left(I\left[\frac{(1+z_{\rm ion})\lambda_\alpha}{\lambda_{\rm obs}}\right] - I\left[\frac{(1+z_{\rm end})\lambda_\alpha}{\lambda_{\rm obs}}\right]\right), \tag{1.28}$$

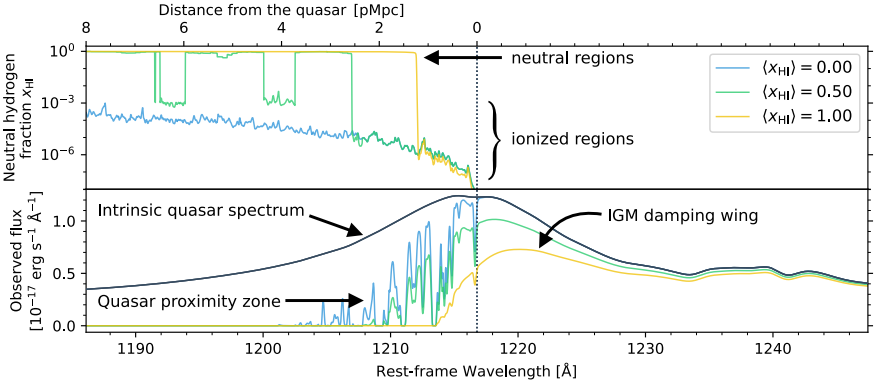

Figure 1.4: Quasar IGM Damping Wings. *Upper*: Neutral fraction field $x_{\rm HI}$ along the sightline from a quasar shining for 10^6 yr into three different environments with global neutral fractions of $\langle x_{\rm HI} \rangle = 0, 0.5$ and 1. *Lower*: The more neutral the IGM, the stronger the damping wing imprinted upon the quasar's intrinsic spectrum (black), and the smaller its proximity zone.

where we integrated from the redshift $z_{\rm ion}$ of the quasar's ionisation front to the end of reionisation at $z_{\rm end}$, and where $I(x)$ is a dimensionless factor given by

$$I(x) = \frac{x^{1/2}}{4(1-x)} + \frac{1}{8} \ln \left(\frac{1 - x^{1/2}}{1 + x^{1/2}} \right). \tag{1.29}$$

Mortlock et al. (2011) reported the first observational evidence for such an imprint in the spectrum of the quasar ULAS J1120+0641 at $z = 7.09$. This finding sparked a first wave of interest in probing reionisation with IGM damping wings, where broadly consistent constraints were derived from the aforementioned object and ULAS J1342+0928 at $z = 7.54$ (Bolton et al. 2011; Greig et al. 2017b; Bañados et al. 2018; Davies et al. 2018a; Greig et al. 2019; Ďurovčíková et al. 2020; Reiman et al. 2020), as well as DES J0252-0503 at $z = 7.00$ and J1007+2115 at $z = 7.51$ (Wang et al. 2020; Yang et al. 2020a; Greig et al. 2022), and a couple of other lower-redshift quasars (Schroeder et al. 2013). However, due to the the scarcity of quasars at sufficiently high redshifts, the analyses remained limited to this handful of objects until Ďurovčíková et al. (2024) and Greig et al. (2024a) presented the first studies of larger quasar samples, probing the end stages of reionisation between $5.8 \lesssim z \lesssim 7$ with quasar damping wings. To date, four largely independent analysis pipelines have been developed and applied to observational data (Greig et al. 2017a; Davies et al. 2018a; Ďurovčíková et al. 2020; Hennawi et al. 2025). The last of these four pipelines also forms the basis of the analyses presented in this thesis. Establishing such a new analysis framework is particularly timely in the light of the now ongoing Euclid mission whose wide-field survey is anticipated to identify up to hundreds of new quasars at

$z \gtrsim 7$ suitable for damping wing analysis (Euclid Collaboration et al. 2019), with the first objects already having been identified (Euclid Collaboration et al. 2026).

In addition to that, damping wing absorption may also provide evidence for the persistence of the last neutral islands at the end stages of reionisation. As pointed out by Malloy & Lidz (2015), if such a neutral island were to intersect a quasar sightline, it could cause a damping wing in the Lyman-α forest region of the quasar spectrum, adjacent to the island's Gunn-Peterson absorption trough. The first hints of such signatures have recently been identified in individual (Becker et al. 2024) and stacked (Spina et al. 2024; Zhu et al. 2024) quasar spectra at $5.5 \lesssim z \lesssim 6$, but their interpretation still requires careful simulation modelling (Sawyer et al. 2025; Gnedin & Zhu 2025).

1.4.3.6 Proximity zones

The ionisation state of the IGM in the foreground of a quasar is not only influenced by the topology of reionisation but also by the quasar's own ionising radiation. The consequence of this radiation is a Mpc-scale ionised bubble around the quasar whose size grows with its (UV-luminous) lifetime t_Q. To first approximation, assuming that each emitted photon ionises exactly one hydrogen atom in a homogeneous surrounding IGM, its radius R_ion can be approximated as

$$R_\mathrm{ion} = \left(\frac{3Q\, t_\mathrm{Q}}{4\pi \left\langle n_\mathrm{HI} \right\rangle} \right)^{1/3} , \tag{1.30}$$

where Q is the quasar's emission rate of ionising photons (Cen & Haiman 2000). The lifetime-dependence of R_ion establishes an important connection to the growth of the supermassive black holes (SMBHs) that are powering these quasars. SMBH growth mechanisms are actively debated (e.g. Bogdán et al. 2024; Maiolino et al. 2024a,b), trying to explain how these objects could grow so massive even at the earliest redshifts where quasars and AGN are observed. In combination with mass constraints, measurements of how long these objects have been shining as UV-luminous quasars are therefore essential in constraining their growth history.

Observationally, the ionised bubble around a quasar manifests as a *proximity zone* (or nearzone) R_prox in its spectrum blueward of the Lyman-α line. Due to the excess ionisation from the quasar, the transmission stays enhanced in this region of the spectrum, as seen in Figure 1.4. However, due to residual neutral gas or density clumps within an ionised bubble, as well as the observational measurement procedure[1], R_prox is not always identical

[1] R_prox is commonly defined as the location where the continuum-normalised flux, smoothed with a 20 Å boxcar filter, first drops below 10 %.

to the radius of the ionisation front $R_{\rm ion}$. As such, detailed proximity zone modelling is required (e.g. Davies et al. 2020; Satyavolu et al. 2023a; Zhou et al. 2024), but nonetheless, $R_{\rm prox}$ remains an important summary statistic, informative about the lifetime of a quasar, and measured extensively in the literature (e.g. Eilers et al. 2017, 2020; Ishimoto et al. 2020; Satyavolu et al. 2023b). However, due to the lifetime's degeneracy with the global ionisation state of the IGM, a joint measurement of the two quantities – as conducted in this thesis – is inevitable when probing quasars at redshifts where reionisation is ongoing.

1.5 Other observational probes of Reionisation

Quasars are only a single class of objects in a diverse collection of observational probes that are used to place constraints on the Epoch of Reionisation. A selection of recent constraints derived from this variety of probes is depicted in Figure 1.5. First, a largely model-independent integral constraint can be derived from the *Cosmic Microwave Background* as measured by the Planck satellite (Planck Collaboration et al. 2020). *Galaxies* provide a considerably more numerous alternative to quasars, probing reionisation through a number of distinct features such as the abundance and the clustering of Lyman-α emitters (LAEs) (e.g. Ouchi et al. 2010; Sobacchi & Mesinger 2015; Ouchi et al. 2018; Umeda et al. 2025), equivalent width distributions of Lyman-break galaxies (LBGs) (e.g. Stark et al. 2010, 2011; Mason et al. 2018, 2019; Tang et al. 2024), or damping wing features in their spectra (e.g. Curtis-Lake et al. 2023; Mason et al. 2026; Umeda et al. 2026), although at a significantly reduced constraining power per object. Damping wings can further be identified in the afterglow spectra of *gamma-ray bursts* (e.g. Lamb & Reichart 2000; Totani et al. 2006; Fausey et al. 2025a). Future radio observations of the *21 cm spin-flip transition* of neutral hydrogen (Furlanetto et al. 2006; Pritchard & Loeb 2012; Liu & Shaw 2020) and the *patchy kinetic Sunyaev-Zel'dovich (pkSZ) effect* (Sunyaev & Zeldovich 1980) targeted by current (Reichardt et al. 2021; Beringue et al. 2025) and next-generation (Matsumura et al. 2014; Abazajian et al. 2016; Ade et al. 2019) CMB surveys will set new standards in the coming decade. In this section, we provide a brief overview over these other probes.

1.5.1 The Cosmic Microwave Background

As discussed in Section 1.2.1, CMB photons carry information about reionisation as they get Thomson scattered off of the free electrons that get released during this epoch. As such, large CMB surveys such as the Wilkinson Microwave Anisotropy Probe (WMAP) and the Planck satellite have placed constraints on τ_e as defined in Eq. (1.4). While the first constraint from

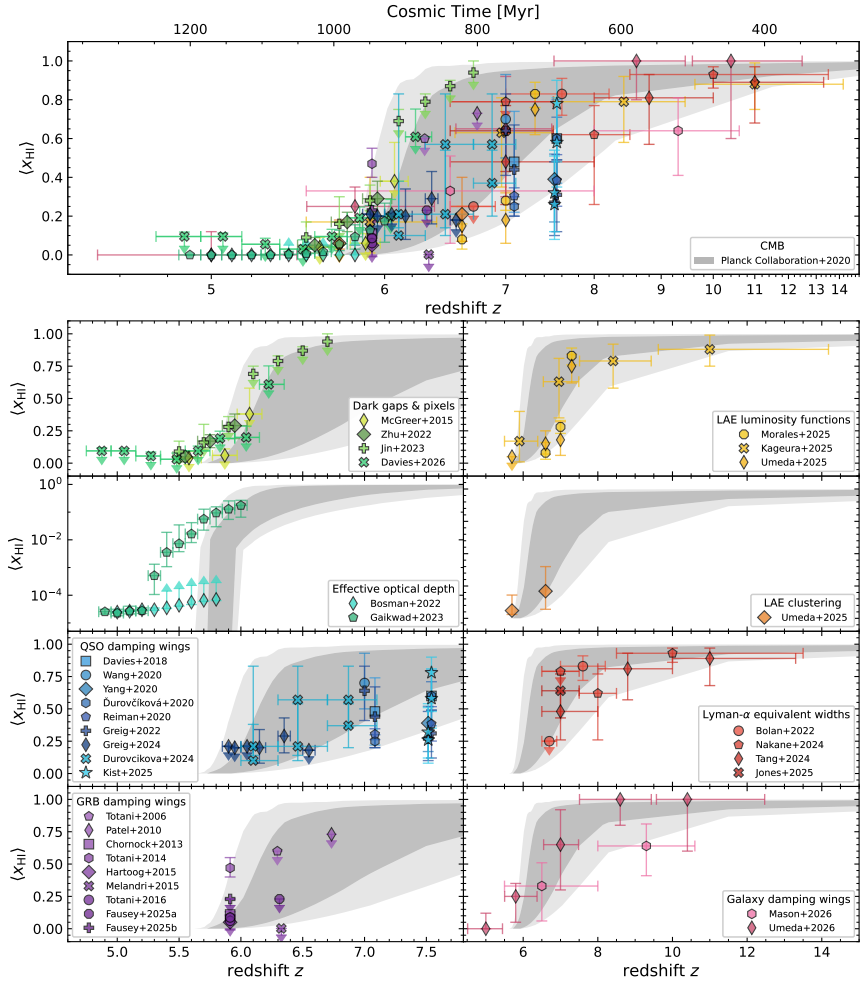

Figure 1.5: Selected constraints on the history of reionisation obtained from a variety of probes. Grey regions are based on CMB measurements by Planck Collaboration et al. (2020) combined with dark pixel fractions from McGreer et al. (2015). Quasar-based constraints include Lyman-α effective optical depth distributions (Bosman et al. 2022; Gaikwad et al. 2023), dark pixel fractions (McGreer et al. 2015; Jin et al. 2023; Davies et al. 2026), dark gap statistics (Zhu et al. 2022) and IGM damping wings (Davies et al. 2018a; Wang et al. 2020; Yang et al. 2020a; Ďurovčíková et al. 2020; Reiman et al. 2020; Greig et al. 2022, 2024a; Ďurovčíková et al. 2024; Kist et al. 2026). Galaxy-based probes are luminosity functions and clustering of Lyman-α emitters (Morales et al. 2021; Kageura et al. 2025; Umeda et al. 2025), equivalent width distributions of Lyman-break galaxies (Bolan et al. 2022; Nakane et al. 2024; Tang et al. 2024; Jones et al. 2025) and IGM damping wings (Mason et al. 2026; Umeda et al. 2026). Damping wing constraints from afterglow spectra of gamma ray bursts have been obtained by Totani et al. (2006); Patel et al. (2010); Chornock et al. (2013); Totani et al. (2014); Hartoog et al. (2015); Melandri et al. (2015); Totani et al. (2016); Fausey et al. (2025a,b). For better visibility, the upper summary panel scales linearly with cosmic time.

WMAP was still as high as $\tau_e = 0.17 \pm 0.04$ (Kogut et al. 2003), it reduced to $\tau_e = 0.089 \pm 0.014$ (Hinshaw et al. 2013) based on the final 9-year data release. Planck then successively reduced the inferred value further, starting with a value of $\tau_e = 0.097 \pm 0.038$ (Planck Collaboration et al. 2014) consistent with WMAP that then decreased via $\tau_e = 0.079 \pm 0.017$ (Planck Collaboration et al. 2016) down to $\tau_e = 0.0544 \pm 0.00755$ (Planck Collaboration et al. 2020) as the final legacy Plack result.

Recent baryon acoustic oscillation (BAO) measurements from the Dark Energy Spectroscopic Instrument (DESI) have led to claims of a $\sim 3\sigma$ tension with the basic ΛCDM model as supported by the Planck results, interpreted as a possible hint towards dynamical dark energy or neutrino masses smaller than the upper bound derived from neutrino oscillation experiments (Abdul Karim et al. 2025). This has sparked a new interest in the reionisation optical depth τ_e as a possible means to alleviate these tensions by absorbing them into a higher value for τ_e than preferred by Planck's legacy result, and more in line with the older Planck and WMAP constraints (Jhaveri et al. 2025; Sailer et al. 2026). This underpins the need for complementary constraints from other astrophysical probes of reionisation, independent of the Planck CMB data (Elbers 2025; Garcia-Gallego et al. 2025a; Kageura et al. 2026).

1.5.2 Galaxies

Galaxies – as the main drivers of reionisation (Madau & Dickinson 2014; Bouwens et al. 2015; Robertson et al. 2015; Finkelstein et al. 2019) – naturally carry a wealth of information about the progression of this epoch. The challenge in extracting this information consists in complex gas and star formation physics that simultaneously affect their evolution, rendering the derivable constraints highly model dependent. Their biggest advantage, on the other hand, is their sheer number density as compared to significantly rarer sources such as quasars. Moreover, JWST has allowed us to detect galaxies at unprecedented redshift frontiers, far beyond those of the most distant quasars known, reaching deep into the early stages of reionisation, with spectroscopically confirmed objects out to $z \sim 14$ (e.g. Bunker et al. 2023; Curtis-Lake et al. 2023; Carniani et al. 2024; Witstok et al. 2025).

Of particular interest for constraining reionisation is the class of *Lyman-α emitting galaxies* (LAEs). These typically young, low-mass and highly star-forming galaxies – with their characteristic Lyman-α emission line that gets attenuated with the increased presence of neutral hydrogen – therefore become harder to detect the more neutral hydrogen is present in the IGM. As a result, number counts of LAEs in the form of luminosity functions have extensively been used to constrain the reionisation state of the IGM as a function of redshift (e.g. Ouchi et al. 2010; Konno et al. 2014; Zheng et al. 2017; Inoue et al. 2018; Konno et al. 2018; Goto et al. 2021; Morales et al.

2021; Ning et al. 2022; Wold et al. 2022; Kageura et al. 2025; Umeda et al. 2025). In a similar manner, the LAEs that we do detect are more likely to reside in large ionised regions whereas those in neutral regions are more prone to be too attenuated. Measuring the clustering of LAEs therefore provides an alternative to simple number counts and can simultaneously yield information about bubble sizes and the topology of reionisation (Ouchi et al. 2010; Sobacchi & Mesinger 2015; Ouchi et al. 2018; Umeda et al. 2025).

Lyman-break galaxies, identified by the characteristic drop-off in their spectrum at the Lyman limit, are another class of galaxies often used to constrain reionisation by measuring their *Lyman-α equivalent widths*, i.e., the brightness of their Lyman-α line relative to their UV continuum. The distribution of equivalent width measurements then serves as a probe of $\langle x_{\rm HI} \rangle$ because the fraction of objects with high equivalent widths, showing strong Lyman-α emission, is expected to decrease in a more neutral IGM as already argued above (e.g. Stark et al. 2010, 2011; Ono et al. 2012; Schenker et al. 2012; Mesinger et al. 2015; Mason et al. 2018; Hoag et al. 2019; Mason et al. 2019; Jung et al. 2020; Bolan et al. 2022; Bruton et al. 2023; Nakane et al. 2024; Tang et al. 2024; Jones et al. 2025).

In addition to these probes, recent JWST observations have led to the first claims of *damping wing absorption* in individual (Curtis-Lake et al. 2023; Hsiao et al. 2024; Park et al. 2025; Mason et al. 2026) and stacked (Umeda et al. 2024, 2026) spectra of high-redshift galaxies. However, it has been pointed out that these signatures are in many cases dominated by very similar imprints from localised damped Lyman-α absorption systems (DLAs) in the foreground of these galaxies Heintz et al. (2024, 2025); Huberty et al. (2025). As such, the existing constraints need to be treated with caution and only marginalisation over the potential presence of DLAs can be expected to yield robust constraints (Mason et al. 2026).

1.5.3 Gamma-ray bursts

Damping wings have not only been reported towards galaxies and quasars, but also towards *gamma-ray bursts* (GRBs) – the brightest transient events seen on the sky, most likely to be associated to superluminous supernova events. Their brightness allows them to be observed out to extremely high redshifts, potentially out to $z \sim 20$ (Lamb & Reichart 2000). The fact that their afterglow spectra that can be modelled as a simple power law spectrum as well as the absence of significant own ionising radiation makes GRBs a conceptually very clean probe of the ionisation state of the IGM. The caveat is that such events are extremely rare and not more than a handful of such events have been detected at $z > 6$ to date (Totani et al. 2006; Patel et al. 2010; Chornock et al. 2013; Melandri et al. 2015; Fausey et al. 2025a), and, as a matter of fact, the derived constraints on $\langle x_{\rm HI} \rangle$ have not always been

consistent (Chornock et al. 2013; Totani et al. 2014; Hartoog et al. 2015; Totani et al. 2016; Fausey et al. 2025b).

1.5.4 The 21 cm line

The 21 cm spin flip transition of neutral hydrogen, lending its name from its rest-frame wavelength of $\lambda_{21\,\mathrm{cm}} = 21.106\,\mathrm{cm}$, is the transition between the two hyperfine levels of the 1s ground state of the neutral hydrogen atom. Due to the ubiquitous presence of neutral hydrogen in the IGM, the hope is that observations of this 21 cm signal – at present redshifted to radio frequencies – would allow us the create detailed *intensity maps* of neutral hydrogen at any position in the sky, probing the progression of reionisation at a level of spatial detail unachievable with any of the aforementioned astrophysical probes (see e.g. Furlanetto et al. 2006; Pritchard & Loeb 2012; Liu & Shaw 2020). Despite this conceptual simplicity and promise, the practical detection of the 21 cm signal has proven as a considerable challenge due to significant foregrounds and instrumental systematics, orders of magnitude larger than the signal itself (see e.g. Cheng et al. 2018).

Nonetheless, sensitivities are steadily increasing and ongoing experiments have already placed first upper limits on the strength of the signal (Mertens et al. 2020; Trott et al. 2020; Abdurashidova et al. 2022). A boost in sensitivity is expected in the coming decade with the advent of the low-frequency array of the Square Kilometre Array (SKA-low) (Koopmans et al. 2015).

1.5.5 The patchy kSZ effect

The Cosmic Microwave Background carries another signature from the reionisation era that could potentially be utilised by the next generation of CMB surveys. The *Sunyaev-Zel'dovich (SZ) effect* arises when CMB photons inverse-Compton scatter off of high-energy electrons, causing characteristic anisotropies in the CMB (Sunyaev & Zeldovich 1980). The effect exists in two flavours: the *thermal* SZ (tSZ) effect caused by energy transfer from hot electrons to colder CMB photons, and the *kinetic* SZ (kSZ) effect due to scattering off of electrons that are in bulk motion relative to the CMB, resulting in a Doppler shift that then produces small temperature anisotropies in the CMB. While the thermal SZ effect is dominated by the signal from galaxy clusters, the patchy nature of reionisation causes a non-negligible contribution to the kSZ effect. This is due to the free electrons within the ionised bubbles that are growing during that epoch. The consequence of the bulk motion of these bubbles is a *patchy kSZ* (pkSZ) imprint on the CMB (Gruzinov & Hu 1998; Knox et al. 1998).

However, this imprint has to be disentangled from other post-reionisation contributions to the total kSZ signal. First tentative detections have been

reported by the South Pole Telescope (SPT) (Reichardt et al. 2021) and the Atacama Cosmology Telescope (ACT) (Beringue et al. 2025), and more conclusive measurements are expected from the next stage of CMB surveys such as LiteBIRD (Matsumura et al. 2014), CMB-S4 (Abazajian et al. 2016) and the Simons Observatory (Ade et al. 2019).

1.6 Inference – Le bon sens réduit au calcul

Having reviewed the theory and modelling efforts behind the cosmic epoch that we are trying to constrain, as well as its observational signatures, the only missing ingredient is a framework that allows us to connect the former and the latter and draw conclusions from the observational data on our models of reionisation. A fundamental limitation inherent to cosmology is that we cannot set up and conduct experiments in the same vein as experimentalists in other fields of physics are able to. Instead, all we are provided with is *a single realisation of our Universe* – the one that we are living in – and we have to constrain our models based exclusively on observations of this single realisation, making the problem naturally suited for *Bayesian statistics*.

1.6.1 Bayes' theorem

Perhaps it is more than a coincidence that the development of the theory behind this brings us back from the *cosmic* to the *human* Age of Enlightenment. It was in 1763 when Richard Price published the manuscript of his friend Thomas Bayes who had passed away two years earlier, containing the first application of what we now know as *Bayes' theorem* (Bayes & Price 1763). Re-derived and extended by Pierre-Simon Laplace (Laplace 1812, 1814, 1820), probability in its Bayesian notion constitutes our subjective measure of belief of a given hypothesis or model. Bayesian logic is therefore sometimes interpreted as an extension of the Boolean logic of *true* and *false* to continuous probability values in between these two extremes, or, as Laplace famously put it himself, as *essentially just the extension of common sense to calculus* (Laplace 1814).

More formally speaking, we in the situation where we are provided with a theoretical model $\mathcal{M}$ whose free parameters $\boldsymbol{\theta}$ we are aiming to constrain based on a set of data $\boldsymbol{x}$ that we collected. In other words, we are trying to determine the *posterior* probability distribution $P(\boldsymbol{\theta}|\boldsymbol{x})$, quantifying our degree of belief in the parameters $\boldsymbol{\theta}$ after having seen the data $\boldsymbol{x}$. The law of conditional probabilities then allows us to write down Bayes' theorem, relating the *posterior* probability to three other distributions:

$$P(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{L(\boldsymbol{x}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})}{P(\boldsymbol{x})}. \tag{1.31}$$

Here $L(\boldsymbol{x}|\boldsymbol{\theta})$ is the *likelihood* of observing the data $\boldsymbol{x}$, given a fixed set of parameters $\boldsymbol{\theta}$, encoding the *predictions of our theory*. The *prior* probability $\pi(\boldsymbol{\theta})$ encodes our *prior beliefs* about what values of $\boldsymbol{\theta}$ we deem likely or unlikely. These beliefs could, for example, be based on physical arguments or previous experiments. Finally, the *evidence* $P(\boldsymbol{x})$ quantifies the probability of observing the data $\boldsymbol{x}$ *at all* (in the framework of our model $\mathcal{M}$). Computing $P(\boldsymbol{x})$ in the context of different models therefore also provides a way of performing *model comparison* by determining which model maximises the chance of collecting the observed data whatsoever. However, computing $P(\boldsymbol{x}) = \int \mathrm{d}^n\boldsymbol{\theta}\, L(\boldsymbol{x}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})$ can be highly challenging since it involves an integral over the parameter vector $\boldsymbol{\theta}$ whose dimensionality can be high. On the other hand, from the standpoint of evaluating the posterior $P(\boldsymbol{\theta}|\boldsymbol{x})$ as a function of the parameters $\boldsymbol{\theta}$, the evidence $P(\boldsymbol{x})$ does not constitute more than a normalisation constant and therefore does not warrant explicit computation when no comparison of different models is required.

1.6.2 Sampling

After setting up the posterior distribution $P(\boldsymbol{\theta}|\boldsymbol{x})$ by specifying likelihood and prior, we are still faced with the difficulty of practically evaluating Eq. (1.31). A naive evaluation on a parameter grid quickly becomes infeasible in high dimensions, and as such, the common approach is to resort to *sampling techniques*, amongst which *Markov Chain Monte Carlo* (MCMC) methods hold the most prominent position. The many different flavours of MCMC samplers that have been implemented to date all share the common idea of constructing a *Markov Chain* – a sequence of samples $\{\boldsymbol{\theta}_i\}_{i\in\mathbb{N}}$ where a given sample $\boldsymbol{\theta}_i$ only depends on its immediate precursor $\boldsymbol{\theta}_{i-1}$ – whose density approximates the distribution of interest, i.e., $P(\boldsymbol{\theta}|\boldsymbol{x})$. In essence, this is achieved via a random walk through the posterior distribution $P(\boldsymbol{\theta}|\boldsymbol{x})$. The first incarnation of this idea is the *Metropolis-Hastings algorithm* (Metropolis et al. 1953; Hastings 1970) which, at a given location in the chain, draws a new sample from a given proposal distribution, and either accepts or discards the proposed sample according to a specific acceptance criterion, ensuring that the principle of *detailed balance*, i.e., the reversibility of each transition, is satisfied.

The original Metropolis-Hastings algorithm has been extended in various ways in order to improve its efficiency and allow for a more favourable scaling to high dimensions. The specific technique that we will be applying in this thesis is the *Hamiltonian Monte Carlo* (HMC) algorithm (Duane et al. 1987; Neal 1996) that combines the Metropolis-Hastings algorithm with *Hamiltonian dynamics* (Hamilton 1834, 1835) to allow for a more efficient exploration of the probability landscape than a naive random walk

allows. Treating the parameter vector $\boldsymbol{\theta}$ as position variable, its dynamics is governed by the Hamiltonian

$$\mathcal{H}(\boldsymbol{\theta}, \boldsymbol{p}) = U(\boldsymbol{\theta}) + \frac{1}{2}\boldsymbol{p}^{\mathrm{T}} M^{-1} \boldsymbol{p}, \tag{1.32}$$

where the potential energy is chosen to be the negative logarithmic probability distribution of interest

$$U(\boldsymbol{\theta}) \equiv -\ln P(\boldsymbol{\theta}|\boldsymbol{x}), \tag{1.33}$$

and the kinetic term is defined by a given symmetric and positive definite *mass matrix* M for a set of artificially introduced conjugated momenta $\boldsymbol{p}$. After drawing a set of random momenta $\boldsymbol{p}_i$ at a given location in parameter space $\boldsymbol{\theta}_i$, *Hamilton's equations*

$$\frac{\mathrm{d}\boldsymbol{\theta}}{\mathrm{d}t} = \vec{\nabla}_{\boldsymbol{p}} \mathcal{H} = M^{-1}\boldsymbol{p} \tag{1.34}$$

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = -\vec{\nabla}_{\boldsymbol{\theta}} \mathcal{H} = \vec{\nabla}_{\boldsymbol{\theta}} \ln P(\boldsymbol{\theta}|\boldsymbol{x}) \tag{1.35}$$

can then be solved via a symplectic integrator such as the Leapfrog algorithm to propose a new sample $\boldsymbol{\theta}_{i+1}$ that is accepted or rejected according to the standard Metropolis-Hastings acceptance criterion. The advantage of the HMC algorithm is that it makes an informed guess for $\boldsymbol{\theta}_{i+1}$ by following a Hamiltonian trajectory through the potential landscape defined by the probability distribution, rather than drawing the new sample from a fixed proposal distribution, maximising the acceptance rate.

The number L of integration steps as well as the step size ϵ of the integrator are critical hyperparameters of the algorithm. Choosing these parameters too small would result in samples that are close to one another, effectively just recovering the inefficient random walk behaviour that the algorithm is designed to improve upon. A too large step size, on the other hand, would result in inaccurate integrations and hence a lower acceptance rate, while too many steps (of an adequate size ϵ) would eventually lead to a point where the trajectory starts to repeat itself in an oscillatory fashion, and hence the repeated computation of the same dynamics. The *No-U-Turn Sampler* (Hoffman & Gelman 2011) avoids these issues by successively constructing a binary tree of forward and backward leapfrog steps until the point where the trajectory would start to loop around, and then choosing a new sample from that tree.

As Eq. (1.35) suggests, the Hamiltonian Monte Carlo algorithm requires – and harnesses – gradient information. As such, it has only been able to gather pace in the past few years with the rise of *automatic differentiation* techniques as an alternative to symbolic and numerical differentiation. Unlike these approaches, automatic differentiation frameworks directly accumulate derivatives of composite functions via the repeated application of the chain

rule to all of their components. This has been a decisive factor that facilitated the current success of machine learning techniques by enabling efficient gradient-based decent and backpropagation. As such, the Hamiltonian Monte Carlo sampler takes advantage of the latest developments in computational infrastructure.

1.7 This Thesis – Les Lumières

Les Lumières – in the broader or the stricter sense, the title of this thesis unites all its main ingredients: the cosmological epoch we are aiming to constrain, the astrophysical lights we are using for that purpose, and the Bayesian way of reasoning that allows us to connect the former and the latter. Most of the chapters that will follow are based on or will extend the Bayesian inference pipeline for quasar damping wing analysis that has recently been introduced by Hennawi et al. (2025).

In **Chapter 2**, we leverage this pipeline to perform a detailed analysis of the precision of quasar IGM damping wing measurements. Several factors are contributing to the total error budget and need to be accounted for. 1) The (a priori unknown) *intrinsic continuum* of the quasar has to be reconstructed in order to disentangle it from the intergalactic absorption signature, introducing a continuum reconstruction error that will contribute to the overall uncertainty. 2) Since reionisation is a patchy process, sightline-to-sightline variations are a crucial factor. Especially during the mid-stages of reionisation, certain individual sightlines might still be mostly neutral while others are significantly ionised. A detailed modelling of the associated scatter is only possible via numerical simulations and introduces significant uncertainties. 3) The ionisation front of a given quasar – whose size is largely determined by the quasar's lifetime – has a significant impact on the damping wing strength. As such, the lifetime of the quasar has to be inferred simultaneously in order to be able to marginalise over its degeneracy with $\langle x_{\rm HI} \rangle$. 4) The observational setup determines the inference precision that we are ultimately able to achieve – and sets requirements on the sophistication of our likelihood prescription.

In this chapter, we perform a detailed study of all these contributions. The first part of the chapter is devoted to determining the optimal observational setup in terms of wavelength coverage, signal-to-noise ratio and spectral resolution. Crucially, we find that our approximate likelihood prescription prevents us from benefitting from higher-resolution spectra since the precision gains are evened out by the corrections we have to apply as to not quote overconfident (say: unjustly narrow) posterior distributions. We also determine the optimal dimensionality of our model for the quasar continuum and then proceed by breaking down the total error budget of IGM damping wing measurements as a function of astrophysical parameter

space. We find that we are able to constrain the IGM neutral fraction to $28.0^{+8.2}_{-8.8}$ % and the quasar lifetime to $0.80^{+0.22}_{-0.55}$ dex, where approximately half of the associated error budget is due to the continuum reconstruction task and the other half due to the stochastic nature of reionisation.

In **Chapter 3**, we further address that second contribution and introduce two new summary statistics that are robust against cosmic variance in the IGM and informative not only about the global timing of reionisation as quantified by $\langle x_{\rm HI} \rangle$ but also the *local* ionisation topology. We can extract this local line-of-sight information by defining 1) the Lorentzian-weighted HI column density $N_{\rm HI}^{\rm DW}$ whose weighting mimics the frequency dependence of the Lyman-α cross section, and 2) the quasar's distance $r_{\rm patch}$ to the first neutral patch. Both these statistics are defined on the original topology *before* it is altered by the quasar's own ionising radiation, and we propose to use them in combination with the lifetime $t_{\rm Q}$ of the quasar, summarising the effects of its ionising radiation, as an improved parameterisation of quasar IGM damping wings. We demonstrate that this parameterisation reduces the scatter of the IGM transmission in the damping wing region of the spectrum down to $\lesssim 1\,\%$.

In the final section of the chapter, we show that these statistics are independent of the chosen model of reionisation. In other words, when selecting sightlines that share the same values of our summary statistics, all scatter in the resulting IGM transmission profiles solely arises from density fluctuations in the IGM, independently of what ionisation topology the sightlines originated from. We demonstrate the latter by introducing a simple numerical procedure for generating neutral fraction skewers, showing that the statistical properties of the resulting IGM transmission profiles match those derived from realistic semi-numerical ionisation topologies.

In **Chapter 4**, we put this new IGM damping wing parameterisation into action by incorporating it into our previously introduced Bayesian inference pipeline. The aforementioned topology-independence allows us to constrain our new local summary statistics in a *model-independent* fashion. We show that we can tie these constraints to any given reionisation model in a separate subsequent step. In our framework, the assumption of a specific reionisation model amounts to imposing a specific prior on our local summary statistics, given by their distribution within that assumed topology. This then allows us to also quote a *model-dependent* constraint on the global IGM neutral fraction $\langle x_{\rm HI} \rangle$ – tied to the specific reionisation model that was assumed. By performing inference on a large set of mock spectra, we demonstrate that our local, model-independent measurements constrain the Lorentzian-weighted HI column density to $0.69^{+0.06}_{-0.30}$ dex and the quasar's original distance to the first neutral patch to $31.4^{+10.7}_{-28.1}$ cMpc if a noticeable damping wing is present in the spectrum.

Finally, in **Chapter 5**, we apply our new pipeline to JWST/NIRSpec spectra of two of the highest-redshift quasars known to date: J1007+2115 at

$z = 7.51$ and J1342+0928 at $z = 7.54$. After folding in the model-dependent topology information, we find that J1007+2115 is most likely to reside in a $\langle x_{\rm HI} \rangle = 0.32^{+0.22}_{-0.20}$ neutral IGM while shining for a remarkably short time of $\log_{10} t_{\rm Q} = 4.14^{+0.74}_{-0.18}$. The constraints on our local summary statistics turn out to be largely prior-dominated for this object, as is common for a sightline without a clear damping wing imprint in its spectrum, suggesting that the sightline is largely ionised. For J1342+0928, we infer a neutral fraction of $\langle x_{\rm HI} \rangle = 0.58^{+0.23}_{-0.23}$ along with an intermediately long lifetime of $\log_{10} t_{\rm Q} = 5.64^{+0.25}_{-0.43}$. This source does exhibit a non-negligible damping wing signature, and we are able to constrain its column density to $N_{\rm HI}^{\rm DW} = 20.24^{+0.25}_{-0.22}$, along with a distance $r_{\rm patch} = 10.9^{+5.6}_{-5.9}$ cMpc to the first neutral patch. We find our constraints on $\langle x_{\rm HI} \rangle$ and $t_{\rm Q}$ derived from both these objects to be consistent with those quoted in the literature.

In addition, we also demonstrate how our new parameterisation naturally allows us to quantify the impact of potential local absorption systems in the foreground of a source on the observed damping wing signal. The possibility of the presence of such absorbers in front of J1342+0928 has recently been pointed out by Davies et al. (2025). In our parameterisation, this can simply be understood as an additional contribution to the Lorentzian-weighted column density.

1.8 Outlook – Les nouvelles Lumières

More conclusive constraints on the Epoch of Reionisation are looming very near at the horizon – across all observational probes, and for quasar IGM damping wings in particular. The launch of JWST has allowed us to find galaxies at redshifts entirely unheard of, reaching deep into the core stages of reionisation (e.g. Bunker et al. 2023; Curtis-Lake et al. 2023; Carniani et al. 2024; Witstok et al. 2025). While significant efforts will be necessary to robustly extract information from the damping wing imprints towards these sources (Mason et al. 2026) or the spatial distributions of LAEs (Maitra et al. 2026), these probes will be complemented in the coming decade with new observing facilities that are poised to reveal first robust measurements of the cosmic 21 cm signal and the patchy kSZ effect. Both these signals are currently still hidden behind foregrounds, systematics and other nuisances, but the next generation of facilities – SKA-low (Koopmans et al. 2015), as well as LiteBIRD (Matsumura et al. 2014), CMB-S4 (Abazajian et al. 2016) and the Simons Observatory (Ade et al. 2019) – have the potential of constraining reionisation at an unprecedented level of precision.

The next leap forward in the context of quasar IGM damping wings is even closer ahead on the horizon. The Euclid wide-field survey has already drastically increased the number of known quasars at $z \gtrsim 7$ (Euclid Collaboration et al. 2026), and is anticipated to identify up to hundreds

of more objects in the coming years (Euclid Collaboration et al. 2019). Spectroscopic follow-up of these new discoveries with telescopes such as JWST will provide us with unprecedented samples for quasar damping wing analysis, with a first sample of 27 new JWST/NIRSpec spectra at $6.5 < z < 7.7$ already being taken at this writing.

This thesis aims to establish a solid statistical groundwork for their analysis. The new parameterisation introduced in Chapters 3 and 4 will make it possible for future analyses to separate out the model-independent extraction of constraints on the local ionisation topology from their model-dependent conversion to constraints on the global timing of reionisation. Larger sample sizes will eventually also open up possibilities for the comparison of different reionisation models by inferring the actual distribution of these local summaries. Further directions include their extension to the case of galaxies as background sources. Due to the higher prevalence of localised absorption systems in the foreground of galaxies as compared to quasars (Huberty et al. 2025), incorporating the absorber column density as an additional contribution to the inferred total Lorentzian-weighted column density as proposed in Chapter 5 will be particularly useful for jointly constraining the intergalactic damping wing signal and that due to localised absorption systems (Mason et al. 2026).

A further caveat identified in Chapter 2 remains unaddressed by this thesis. The approximate nature of our likelihood prescription introduces increasingly large biases when analysing higher-resolution data. In order to quote statistically faithful constraints, we are thus forced to retrospectively broaden our posteriors by an amount that makes them comparable in width to those inferred from low-resolution data. A natural approach to improve upon this issue are *simulation-based inference* methods (Alsing et al. 2019; Cranmer et al. 2020) that do not require any assumptions about the underlying likelihood by deploying a neural network model to learn the likelihood function directly from the simulated data. The application of simulation-based inference techniques therefore has the prospect of significant precision gains from high-resolution data. Addressing these points, and applying our techniques to larger samples of present (D'Odorico et al. 2023; Onorato et al. 2025) and upcoming (Euclid Collaboration et al. 2026) quasar spectra at ever higher redshifts (Euclid Collaboration et al. 2019) will soon shed bright new light on our understanding of the Epoch of Reionisation.