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

Kist, T.

3 A local, topology-independent parametrization of quasar IGM damping wings

Abstract

Lyman-α damping wings towards quasars provide a unique probe of the global volume-averaged neutral hydrogen (HI) fraction $\langle x_{\rm HI} \rangle$. Cosmic variance in the intergalactic medium (IGM), however, is a major source of stochasticity since the *local* neutral environment around a quasar varies significantly even when $\langle x_{\rm HI} \rangle$ is fixed. We show that the IGM damping wing carries additional information about this local ionization topology, unexploited by current analysis frameworks. We introduce two new physically motivated summary statistics encoding the local information about the HI distribution in the IGM *before* it is altered by the quasar's ionizing radiation, encompassing 1) the HI column density, weighted by a Lorentzian profile mimicking the frequency dependence of the Lyman-α cross section, and 2) the quasar's distance to the first neutral patch. This description, when combined with the quasar lifetime as a third parameter, reduces the IGM transmission scatter in the damping wing region of the spectrum to $\lesssim 1\,\%$ across physical parameter space. We introduce a simple procedure for generating synthetic HI sightlines around quasars and demonstrate that the resulting damping wing profiles are statistically indistinguishable from a realistic reionization topology. This opens the door for optimally extracting the damping wing's salient local information in a model-independent fashion. In the context of a specific reionization model, measurements of these local parameters can be translated into constraints on the global timing of reionization, but additionally provide information about the reionization topology, hitherto unused. A marginally modified version of our framework can also be employed for damping wings towards galaxies.

Work published in: **Timo Kist**, Joseph F. Hennawi and Frederick B. Davies, *Monthly Notices of the Royal Astronomical Society*, Volume 544, Issue 2, December 2025, Pages 2316–2339, doi.org/10.1093/mnras/staf1762. Reprinted here in its entirety.

3.1 Introduction

The epoch of reionization was a hallmark event in the evolution of our universe when the until then prevailing dark ages were ended by the light from the first galaxies and quasars. It is these very objects which we can also use as highly sensitive torchlights probing the presence of even the smallest amounts of neutral hydrogen in the intergalactic medium (IGM). This manifests in the form of extended Gunn-Peterson absorption troughs (Gunn & Peterson 1965) appearing at global volume-averaged IGM neutral fractions as low as $\langle x_{\rm HI} \rangle \gtrsim 10^{-4}$ where the Lyman-α transition saturates, and transitions into a smooth damping wing imprint redward of the Lyman-α line when $\langle x_{\rm HI} \rangle$ reaches order unity (Miralda-Escudé 1998). As such, every single high-redshift source offers a glimpse at the progress of reionization of the universe at its respective redshift. Existing (D'Odorico et al. 2023; Onorato et al. 2025) and future (Euclid Collaboration et al. 2019) statistical ensembles of their spectra provide us with a unique avenue to constrain the full timeline of the epoch of reionization, a notoriously hard problem due to the model-dependence of all probes going beyond the CMB electron scattering optical depth (Planck Collaboration et al. 2020) and Lyman-series dark pixel fractions (e.g. McGreer et al. 2015; Jin et al. 2023).

In the case of IGM damping wings, this undertaking is exacerbated by the fact that the IGM absorption imprint has to be disentangled from the intrinsic spectra of these sources, and, potentially, local damped Lyman-α (DLA) absorption systems in front of the source which can mimic the cosmological imprint from the IGM (Huberty et al. 2025). While both these tasks remain tractable when considering quasars as background sources, additional complications arise due to the patchy nature of the reionization process. A consequence of this is that at fixed IGM neutral fraction $\langle x_{\rm HI} \rangle$, the column density of neutral material *along a given sightline* can vary drastically, depending on how many neutral patches it actually contains. The distribution of neutral versus ionized sightlines at given $\langle x_{\rm HI} \rangle$ depends on the topology of reionization. The relation between damping wing strength and global IGM neutral fraction is thus inevitably *stochastic*. In essence, the Lyman-α damping wing only probes the *local* ionization state of the surrounding IGM as a single, statistical draw from the *global* ionization topology at the corresponding redshift.

To date, existing damping wing analysis frameworks (e.g. Greig et al. 2017b, 2024a; Davies et al. 2018a; Durovčíková et al. 2020, 2024; Hennawi et al. 2025; Umeda et al. 2024, 2025; Mason et al. 2026) have not made a clear methodological distinction between the two. The conventional approach has always consisted of directly inferring (among other parameters) the *global* IGM neutral fraction $\langle x_{\rm HI} \rangle$, building upon analytical expressions for the damping wing optical depth (Miralda-Escudé 1998), or realistic IGM

transmission profiles based on cosmological simulations. As demonstrated recently in the context of *quasar* damping wings (Kist et al. 2025b), not less than half the total error budget on the inferred IGM neutral fraction $\langle x_{\rm HI} \rangle$ arises due to density fluctuations and the stochasticity of reionization, comparable to the uncertainty due to the reconstruction of the unknown quasar continuum. An additional major source of uncertainty in the context of *galaxy* damping wings are local proximate Lyman-α absorption systems (Heintz et al. 2024, 2025) which can obfuscate the imprint arising from the IGM (Huberty et al. 2025), and would need to be marginalized out in order to obtain faithful constraints on the ionization state of the IGM (Mason et al. 2026).

In any case, eliminating the uncertainty sourced by the stochasticity of reionization would allow us to extract *all* the information that the damping wing imprint actually encodes—that is, physical line-of-sight information about the neutral hydrogen (HI) density field in front of the quasar which is informative not only about the timing but also the topology of reionization. Instead of directly processing this information into a constraint on the *global* IGM neutral fraction, leaving the latter unconstrained, we here propose a novel set of *physical* summary statistics which quantify the *local* HI content in front of a given object, and hence tightly parameterize the characteristic shape of the IGM damping wing.

Finding a better parameterization of the IGM damping wing has also been the primary motivation of Chen (2024) who proposed to re-center IGM transmission profiles based on the position in the spectrum where the transmission first approaches zero, and Keating et al. (2024b) who showed that the damping wing strength is tightly parameterized by the average HI density weighted with a Lorentzian profile, mimicking the Lorentzian decay of the Lyman-α cross section. Furthermore, Mason et al. (2026) for the first time inferred the distance between the source and the first neutral patch as a local measure in addition to the global IGM neutral fraction $\langle x_{\rm HI} \rangle$. Here we build upon the labels introduced by Keating et al. (2024b) and Mason et al. (2026), and establish two clearly defined physical summary statistics quantifying the local HI content around the source, readily applicable for astrophysical parameter inference. Leveraging these summary statistics, we show that the damping wing imprint does not only bear information about the *timing* of reionization in form of the evolution of the global IGM neutral fraction as a function of redshift, but even *topological* information about, e.g., the distribution of ionized patch sizes around the most massive halos *before quasars started shining*.

In addition to that, we show that our summary statistics are statistically insensitive to the underlying reionization model. Instead, all stochasticity in this parameterization is sourced by the well-understood distribution of density fluctuations in the IGM. This paves the way for *topology-independent*, local damping wing constraints which carry topological information, and can

be tied *subsequently* to a specific reionization topology, constraining, e.g., the global IGM neutral fraction $\langle x_{\rm HI}\rangle$ *in the context of that given topology.*

We start by introducing our formalism and defining our new local summary statistics in Section 3.2. We quantify the scatter of realistic IGM transmission profiles in the damping wing region within this parameterization in Section 3.3 and proceed in Section 3.4 by introducing a simplistic toy prescription which we then use to demonstrate the topology-independence of our parameterization. We conclude in Section 3.5.

3.2 Towards a three-parameter model of quasar IGM damping wings

Several works have recently been pushing towards a parameterization better capturing the characteristic shape of IGM damping wings (Chen 2024; Keating et al. 2024b). Specifically, Keating et al. (2024b) noted that (in the absence of a strong ionizing source) the Lorentzian-weighted average HI number density calculated along simulated sightlines from the Sherwood-Relics simulation suite (Bolton et al. 2017; Puchwein et al. 2023) coincides remarkably well with predictions from the analytical Miralda-Escudé (1998) damping wing model. Here, we build upon that finding and introduce novel two-parameter summary statistics of the local *pre*-quasar HI field that significantly reduce the scatter of the damping wing optical depth $\tau_{\rm DW}$ as compared to the global IGM neutral fraction $\langle x_{\rm HI}\rangle$.

3.2.1 Introducing a local HI column density label

We start by recalling that the damping wing optical depth $\tau_{\rm DW}$ of a strong ionizing source at redshift $z_{\rm QSO}$ as a function of rest-frame wavelength $\lambda_{\rm rest}$ is obtained by integrating the product of the HI density field $n_{\rm HI}^{\rm post}$ and the Lyman-α cross section of neutral hydrogen σ_α (evaluated at wavelength $\lambda = \lambda_{\rm rest}\,(1+z_{\rm QSO})/(1+z)$) along the (proper) line-of-sight interval $\mathrm{d}R$:

$$\tau_{\rm DW}(\lambda_{\rm rest}) = \int_0^{R(z_{\rm QSO})} n_{\rm HI}^{\rm post}(R)\cdot\sigma_\alpha\left(\frac{1+z_{\rm QSO}}{1+z(R)}\,\lambda_{\rm rest}\right)\mathrm{d}R, \qquad (3.1)$$

where $R(z_{\rm QSO})$ is the proper distance from the observer to the source, and $z(R)$ is the inverted distance-redshift relation, and we henceforth denote all associated comoving distances with a small r in place of a capital one.

Our goal is to construct summary statistics of the HI density field that best capture the characteristic shape of the IGM damping wing as given by $\tau_{\rm DW}(\lambda_{\rm rest})$. However, note that we are interested in maximizing the information content with respect to the *pre*-quasar HI density field which we denote as $n_{\rm HI}^{\rm pre}$ (in contrast to the post-quasar field $n_{\rm HI}^{\rm post}$ giving rise to

the actually observed IGM damping wing as per Eq. (3.1)). This is because $n_{\rm HI}^{\rm pre}$ directly relates to the *cosmological* reionization topology, removing the impact of the quasar as a local *astrophysical* source of ionizing radiation. In other words, we can decompose the pre-quasar HI field as

$$n_{\rm HI}^{\rm pre} = \langle n_{\rm H} \rangle (z_{\rm QSO}) \cdot x_{\rm HI} \cdot \Delta, \tag{3.2}$$

where $\langle n_{\rm H} \rangle (z_{\rm QSO})$ is the cosmic mean hydrogen density at the redshift of interest, Δ is the dimensionless matter overdensity field, and $x_{\rm HI}$ is the HI fraction field which informs us about the *global* volume-averaged IGM neutral faction $\langle x_{\rm HI} \rangle$.[1]

We now aim to condense the information contained in the *field* $n_{\rm HI}^{\rm pre}$ into a low-dimensional set of physically motivated summary statistics. The canonical choice for such a summary statistic is the HI column density $N_{\rm HI} = \int n_{\rm HI}^{\rm pre}(R)\, {\rm d}R$ along the line of sight. Noting the obvious similarity between this definition and Eq. (3.1), and following the weighting proposed in Keating et al. (2024b), we introduce a weighting function

$$w \equiv (R/R_{\rm T} + 1)^{-2} \sim (v - v_{\rm T})^{-2}, \tag{3.3}$$

mimicking the Lorentzian decay of the Lyman-α cross section σ_α in Eq. (3.1), evaluated at a fixed reference distance $R = R_{\rm T}$.[2] The equivalence with a Lorentzian profile becomes more apparent if we introduce the velocity grid $v = -H(z_{\rm QSO})\, R$, centered at the Lyman-α line (such that $\lambda_{\rm rest} = \lambda_\alpha \left(1 + \frac{v}{c}\right)$ for the Lyman-α wavelength $\lambda_\alpha \simeq 1215.67\,\text{Å}$), and express $R_{\rm T}$ as a velocity offset $v_{\rm T} \equiv +H(z_{\rm QSO})\, R_{\rm T}$.[3]

Given a field X, we now define the line-of-sight average $\langle\!\langle . \rangle\!\rangle_{\rm Lor}$ of this field with respect to the Lorentzian weighting kernel defined in Eq. (3.3) as

$$\langle\!\langle X \rangle\!\rangle_{\rm Lor} \equiv \frac{1}{\mathcal{N}(u_{\rm min}, u_{\rm max})} \int_{u_{\rm min}}^{u_{\rm max}} \frac{X(u)}{(u+1)^2}\, {\rm d}u, \tag{3.4}$$

where we introduced the dimensionless integration variable $u \equiv R/R_{\rm T}$ with integration limits $u_{\rm min}$ and $u_{\rm max}$, corresponding to the distances $R_{\rm min}$ and $R_{\rm max}$ whose values we will fix below. Further, $\mathcal{N}(u_{\rm min}, u_{\rm max})$ is a normalization factor ensuring $\langle\!\langle \mathbf{1} \rangle\!\rangle_{\rm Lor} = 1$ for the identity field $\mathbf{1}$. This factor can be explicitly determined, yielding

$$\mathcal{N}(u_{\rm min}, u_{\rm max}) = \frac{u_{\rm max} - u_{\rm min}}{(u_{\rm max}+1)(u_{\rm min}+1)}. \tag{3.5}$$

[1] For brevity, we write $\langle x_{\rm HI} \rangle \equiv \langle x_{\rm HI} \rangle (z_{\rm QSO})$ throughout this work.

[2] Note that away from the Lyman-α resonance, a Lorentzian profile is an excellent approximation of the full quantum-mechanical Lyman-α cross section σ_α.

[3] Note that our velocity grid follows the standard convention of $v > 0$ redward of the Lyman-α line, and $v < 0$ blueward where $R > 0$. However, to keep all distances positive, we define $R_{\rm T} > 0$ even though it corresponds to a *positive* velocity offset $v_{\rm T} > 0$ *redward* of Lyman-α.

Comparing Eq. (3.4) to Eq. (3.1) governing the damping wing optical depth, we now proceed by defining the *Lorentzian-weighted* HI column density as

$$N_{\rm HI}^{\rm DW} \equiv \mathcal{N}\left(\tfrac{R_{\rm min}}{R_{\rm T}}, \tfrac{R_{\rm max}}{R_{\rm T}}\right) \cdot R_{\rm T} \cdot \langle n_{\rm H}\rangle(z_{\rm QSO}) \cdot \langle\!\langle x_{\rm HI} \cdot \Delta \rangle\!\rangle_{\rm Lor}. \tag{3.6}$$

This definition involves three hyperparameters, i.e., the reference distance $R_{\rm T}$ as well as the integration limits $R_{\rm min}$ and $R_{\rm max}$. An informed choice of these parameters ensures that this summary statistic reflects the characteristic shape of the IGM damping wing not only for the pre-quasar topology (which we would like to maximize the information content about), but also the post-quasar one (which, due to the impact of the quasar ionizing radiation, inevitably carries a somewhat decreased amount of information about the pre-quasar IGM, but on the other hand gives rise to the actually observed damping wing imprint as per Eq. (3.1)).

Firstly, with regards to the *pre*-quasar topology, we show in Appendix 3.A that in the limit where $R_{\rm min} \to 0$ and $R_{\rm max} \to R(z_{\rm QSO})$, and where the Lyman-$\alpha$ cross section can be approximated as perfectly Lorentzian, the pre-quasar optical depth $\tau_{\rm DW}^{\rm pre}$ (as defined in Eq. (3.11)) evaluated at the velocity offset $v_{\rm T}$ becomes, up to a prefactor, approximately proportional to $N_{\rm HI}^{\rm DW}$:

$$\tau_{\rm DW}^{\rm pre}(v = v_{\rm T}) \simeq \frac{e^2}{m_e c^2} f_\alpha \, \gamma_\alpha \, \lambda_\alpha \, (c/v_{\rm T} - 1)^2 \times N_{\rm HI}^{\rm DW}, \tag{3.7}$$

where $f_\alpha \simeq 0.416$ is the Lyman-α oscillator strength, and $\gamma_\alpha \equiv \Gamma_\alpha \, \lambda_\alpha/(4\pi c)$ with the Lyman-α decay constant $\Gamma_\alpha = 6.265 \times 10^8\,{\rm s}^{-1}$. This motivates our choice of $N_{\rm HI}^{\rm DW}$ as the primary summary statistic that minimizes the scatter of the red-side damping wing transmission among different (pre-quasar) sightlines, and hence encodes the characteristic shape of the IGM damping wing in a single number.

Practically, however, we are concerned with profiles originating from the *post*-quasar topology rather than the pre-quasar one. As we will show in the consequent sections, $r_{\rm min} = 4\,{\rm cMpc}$, $r_{\rm max} = r_{\rm min} + 100\,{\rm cMpc}$, and $r_{\rm T} = 18\,{\rm cMpc}$ (corresponding to $v_{\rm T} \simeq 2000\,{\rm km/s}$ at $z_{\rm QSO} = 7.54$)[4] are adequate choices to ensure that Eq. (3.7) remains valid even for the post-quasar optical depth

[4] As our analysis is restricted to models at this fixed redshift, we henceforth interchangeably refer to reference distances $r_{\rm T}$ and reference velocity offsets $v_{\rm T}$, noting that only a fixed choice of $r_{\rm T}$ would account for redshift evolution in the density field.

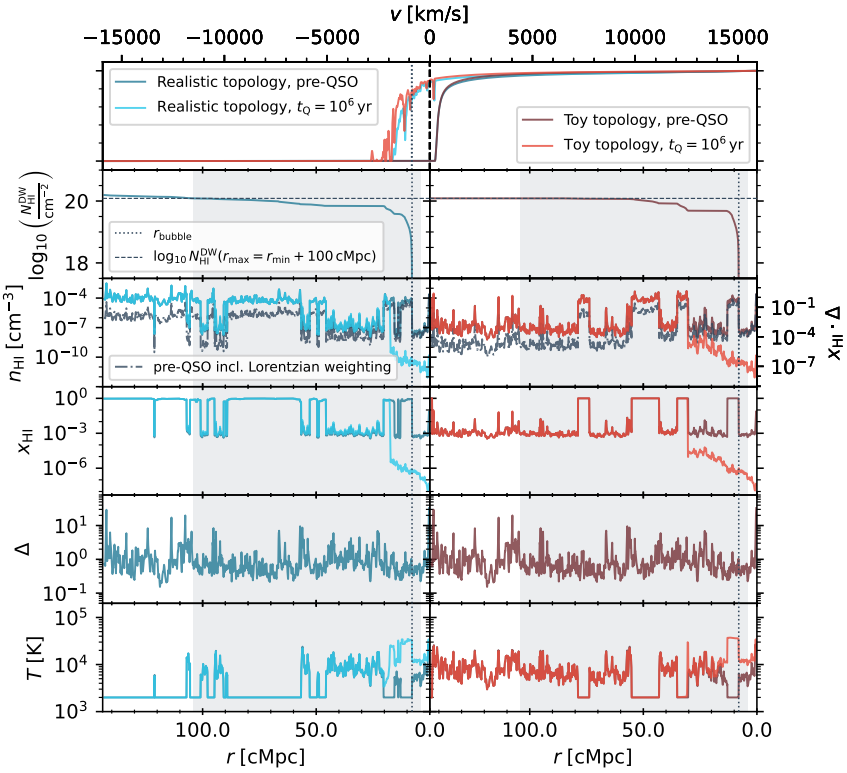

Figure 3.1: Lyman-α transmission and physical fields for an example Nyx sightline combined with a semi-numerical x_{HI} skewer extracted from a $\langle x_{HI} \rangle = 0.65$ topology (left panels/blue lines; see Section 3.3.1) and an analytical x_{HI} skewer generated according to our toy prescription (right panels/red lines; see Section 3.4.1), before (dark) and after (light colors) a quasar has been shining for $t_Q = 10^6$ yr. The panels show from top to bottom the Lyman-α transmission field, the Lorentzian-weighted HI column density N_{HI}^{DW} as a function of the upper integration limit r_{max}, the HI density field n_{HI} (also shown in units of $x_{HI} \cdot \Delta$), including its Lorentzian-weighted pre-quasar version (black dash-dotted line), the HI fraction x_{HI}, the overdensity field Δ, and the temperature field T. Our fiducial integration range for the Lorentzian-weighted HI column density N_{HI}^{DW} is highlighted by the grey-shaded region, and the horizontal dashed line marks the value of $\log_{10} N_{HI}^{DW}/cm^{-2} = 20.1$ corresponding to this integration range. The vertical dotted lines mark the location $r_{patch} = 8$ cMpc of the first neutral patch in the pre-quasar sightline.

$\tau_{\rm DW}$. Adopting these parameter choices and converting all quantities to comoving units, we can evaluate Eq. (3.6) as

$$N_{\rm HI}^{\rm DW} = 5.1 \times 10^{20}\,{\rm cm}^{-2} \times \left(\frac{\mathcal{N}\left(\frac{r_{\rm min}}{r_{\rm T}}, \frac{r_{\rm max}}{r_{\rm T}}\right)}{0.67}\right)\left(\frac{r_{\rm T}}{18\,{\rm cMpc}}\right) \times \left(\frac{1+z_{\rm QSO}}{1+7.54}\right)^2 \left(\frac{\langle\!\langle x_{\rm HI}\cdot\Delta\rangle\!\rangle_{\rm Lor}}{1}\right). \tag{3.8}$$

Here we used $\langle n_{\rm H}\rangle(z_{\rm QSO}) = \langle n_{\rm H}\rangle_0(1+z_{\rm QSO})^3$, where $\langle n_{\rm H}\rangle_0 = 3H_0^2(1-Y)\Omega_{\rm b}/(8\pi G m_{\rm p}) \simeq 1.9\times10^{-7}$ atoms/cm^3 with H_0 and $\Omega_{\rm b}$ chosen according to a Planck Collaboration et al. (2020) cosmology, and a big bang nucleosynthesis (BBN) hydrogen fraction of $1-Y=0.76$.

We show in the top and left panels of Figure 3.1 how we arrive at this summary statistic for a simulated example sightline of a $t_{\rm Q} = 10^6$ yr quasar in a globally 65 % neutral universe. A more detailed description of the underlying simulations will be provided in Section 3.3.1. In each panel, the post-quasar version of the corresponding field is depicted in light blue, the pre-quasar one in dark blue. The top panel shows the IGM transmission profiles,[5] and the second panel shows the Lorentzian-weighted HI column density (Eq. (3.8)) as a function of the upper integration limit $r_{\rm max}$, where the $N_{\rm HI}^{\rm DW}$ value resulting from our fiducial choice of $r_{\rm max} = r_{\rm min} + 100\,{\rm cMpc}$ is marked by the horizontal dotted line. The third panel depicts the HI density field $n_{\rm HI}$ (also given in units of $x_{\rm HI}\cdot\Delta$), and additionally the Lorentzian-weighted version of the pre-quasar field according to Eq. (3.3). The remaining panels show the neutral hydrogen fraction field $x_{\rm HI}$, the overdensity field Δ, and the temperature field T, respectively. The grey-shaded area in each panel depicts our fiducial integration range of $r_{\rm min} = 4\,{\rm cMpc}$ and $r_{\rm max} = r_{\rm min} + 100\,{\rm cMpc}$. We can see how distant neutral patches are downweighted by the Lorentzian weighting function (third panel), and therefore contribute less to $N_{\rm HI}^{\rm DW}$ (second panel) and hence do not notably affect the IGM damping wing (top panel) either.

3.2.2 Choice of the $N_{\rm HI}^{\rm DW}$ integration limits

As Eq. (3.7) shows, our $N_{\rm HI}^{\rm DW}$ statistic is an optimal prescription of the damping wing optical depth $\tau_{\rm DW}^{\rm pre}(v = v_{\rm T})$ of the pre-quasar density field $n_{\rm HI}^{\rm pre}$ evaluated at the velocity offset $v_{\rm T}$, provided that we integrate along the *entire* line of sight and assume a perfectly Lorentzian cross section. If we were to consider damping wings around galaxies, it would therefore be immediately clear that we would have to set our lower integration limit to $r_{\rm min} = 0$.

[5] Note that all profiles which are colored red will be discussed later in Section 3.4.

However, this simple picture changes when considering damping wings towards quasars instead. Even though we are still ultimately interested in the pre-quasar HI field, in practice we are inevitably probing the optical depth arising from the *post*-quasar field $n_{\rm HI}^{\rm post}$. Differences arise interior to the quasar ionization front where the pre-quasar neutral material is ionized away and hence the information about this part of the pre-quasar density field is lost; though we can account for this by also constraining the lifetime $t_{\rm Q}$ of the quasar. By tuning our integration limits $R_{\rm min}$ and $R_{\rm max}$ in Eq. (3.6), we can make sure that we still maximize the information content about $n_{\rm HI}^{\rm pre}$, and nevertheless obtain a statistic which describes the observed damping wing profile with a minimum amount of scatter. By nature of the required approximations, this scatter will be non-zero, but an informed choice of the integration limits ensures that it stays minimal and certainly remains smaller that the amount of scatter which $\langle x_{\rm HI}\rangle$ entails.

To do so, we firstly note that the optical depth integral in Eq. (3.1) effectively only starts at the radius of the ionization front $R_{\rm ion}$ that the quasar has carved out, since $n_{\rm HI}^{\rm post}$ is drastically reduced within the ionized bubble around the quasar (i.e., at $R < R_{\rm ion}$). On the other hand, at $R > R_{\rm ion}$, where the quasar ionization front has not arrived yet, we can safely approximate $n_{\rm HI}^{\rm post} \simeq n_{\rm HI}^{\rm pre}$. In a simple theoretical picture where every photon emitted by the quasar ionizes exactly one hydrogen atom in a (Strömgren) sphere around the quasar, the radius of the ionization front $R_{\rm ion}$ relates to the lifetime of the quasar $t_{\rm Q}$ and the volume averaged HI density $\langle n_{\rm HI}^{\rm pre}\rangle \equiv \langle n_{\rm HI}^{\rm pre}\rangle(z_{\rm QSO})$ as

$$R_{\rm ion} = \left(\frac{3Q\,t_{\rm Q}}{4\pi\langle n_{\rm HI}^{\rm pre}\rangle}\right)^{1/3} \tag{3.9}$$

(Cen & Haiman 2000), where Q is the quasar's emission rate of ionizing photons. Due to the approximate nature of this equation, defining $R_{\rm min} \equiv R_{\rm ion}$ is not guaranteed to give us the correct integration limit for an individual sightline. However, as we strive to define a sightline-independent integration limit, Eq. (3.9) constitutes a useful starting point for determining a universal value.

Since we aim to define our $N_{\rm HI}^{\rm DW}$ label agnostically to the reionization state of the universe, we first have to remove the dependence of the cosmic mean HI density $\langle n_{\rm HI}^{\rm pre}\rangle = \langle n_{\rm H}\rangle_0(1+z_{\rm QSO})^3\langle x_{\rm HI}\rangle$ on the IGM neutral fraction $\langle x_{\rm HI}\rangle$. We can achieve this by assuming $\langle x_{\rm HI}\rangle = 1$ which gives rise to the shortest $R_{\rm ion}$ possible. By doing so, we make our lower integration limit shorter, and hence conservatively also account for regions closer to the source where neutral material is only still present for sightlines from highly neutral topologies.

Secondly, as $N_{\rm HI}^{\rm DW}$ is supposed to be a summary statistic for the *pre*-quasar HI density $n_{\rm HI}^{\rm pre}$, the lifetime dependence of Eq. (3.9) is undesirable too, and we can eliminate it by fixing $t_{\rm Q}$ in Eq. (3.9) to a representative value

of $t_Q = 10^{4.5}\,\mathrm{yr}$. This specific value constitutes an empirical compromise between 1) quasars with longer lifetimes of $t_Q \gtrsim 10^{4.5}\,\mathrm{yr}$ and hence larger R_{ion}, where a fixed lower integration limit of $R_{ion}(t_Q = 10^{4.5}\,\mathrm{yr})$ includes pre-quasar IGM neutral material which has already been ionized away and therefore does *not* contribute to the damping wing, and 2) short-lived quasars with $t_Q \lesssim 10^{4.5}\,\mathrm{yr}$ where this integration limit excludes foreground neutral material that actually *will* impact the damping wing. As we will demonstrate in Section 3.3.3, our choice of $t_Q = 10^{4.5}\,\mathrm{yr}$ provides a reasonably accurate approximation for the full lifetime range expected for observed high-redshift quasars (see e.g. Morey et al. 2021; Khrykin et al. 2021). With these choices, we arrive at the comoving value of

$$r_{ion} = 4.0\,\mathrm{cMpc}\left(\frac{Q}{10^{57.14}\,\mathrm{s}^{-1}}\right)^{1/3}\left(\frac{t_Q}{10^{4.5}\,\mathrm{yr}}\right)^{1/3}\left(\frac{\langle x_{HI}\rangle}{1.0}\right)^{-1/3} \tag{3.10}$$

and fix our lower integration limit accordingly to $r_{min} = 4.0\,\mathrm{cMpc}$.[6]

The optical depth integral (Eq. (3.1)) formally runs all the way down to redshift zero; however, due to the Lorentzian decline in the outer parts of the HI cross section, we can safely stop the integration in Eq. (3.4) already significantly earlier than this, while still retaining a tight approximation of the full damping wing optical depth as per Eq. (3.7). Specifically, we only have to integrate until the Lorentzian line-of-sight average of neutral fraction and overdensity field $\langle\!\langle x_{HI}\cdot\Delta\rangle\!\rangle_{Lor}$ has converged. We practically find that this is the case after integrating over a range of $\sim 100\,\mathrm{cMpc}$ (as exemplarily seen in the second panel of Figure 3.1), where our weighting kernel (Eq. (3.3)) has decreased to $\sim 3\,\%$ of its value at r_{min}. Thus, we set the upper integration limit to $r_{max} = r_{min} + 100\,\mathrm{cMpc}$.

3.2.3 The damping wing as a one-parameter family

We argued in Section 3.2.1 that we identified a close-to optimal summary statistic N_{HI}^{DW} for the damping wing optical depth evaluated at a fixed reference distance r_T (or, equivalently, velocity offset v_T). We now demonstrate that by restricting ourselves to this fixed reference point, we do not give up a significant amount of additional information encoded in the imprint. In other words, we show that the Lyman-α transmission values at different velocity offsets are so correlated with each other that the damping wing profile essentially constitutes a one-parameter family, i.e., the entire spectral shape is determined by a single parameter, namely the transmission $\tau_{DW}(v = v_T)$ at our reference velocity offset.

To demonstrate this, we depict in Figure 3.2 the correlation between a large number of simulated IGM transmission profiles evaluated at our fixed

[6] Note that due to cosmological expansion manifesting in the density field, we arrived at the same comoving integration limit independently of redshift.

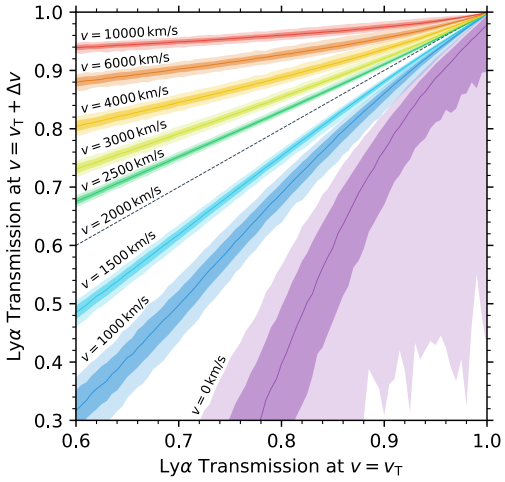

Figure 3.2: IGM transmission values at the fixed reference velocity offset $v_\mathrm{T} = 2000\,\mathrm{km/s}$ versus transmission values at a number of velocities offsets $v = v_\mathrm{T} + \Delta v$ away from this reference offset. Solid curves and shaded regions depict the median as well as the 68- and 95-percentile scatter of these transmission values, respectively, among $852 \times 21 \times 51$ realistic IGM transmission profiles simulated according to the description in Section 3.3.1. For reference, we are also showing the trivial one-to-one relation for $v = v_\mathrm{T}$ as a dashed black line.

reference velocity offset $v_\mathrm{T} = 2000\,\mathrm{km/s}$ versus their value at a number of other velocity offsets $v = v_\mathrm{T} + \Delta v$. Specifically, we considered $852 \times 21 \times 51$ transmission profiles consisting of 852 distinct sightlines at 21 global IGM neutral fraction values $\langle x_\mathrm{HI} \rangle$ covering a linear range between 0 and 1, and 51 lifetime values t_Q covering a logarithmic range between 10^3 and $10^8\,\mathrm{yr}$. The technical details of the simulations will be outlined in Section 3.3.1.

For each velocity offset, we show the median transmission values of all profiles as a solid curve, as well as the 68- and 95-percentile scatter as shaded regions. Our first observation is that in the largest parts of the damping wing region (at all velocity offsets $v \gtrsim 1500\,\mathrm{km/s}$), the width of the 68- (95-) percentile region never exceeds $\sim 1-2\,\%$ ($\sim 2-3\,\%$), even at the lowest IGM transmission values which are significantly impacted by damping wing absorption, decreasing down to $< 1\,\%$ ($< 2\,\%$) at higher transmission values. A number of effects conspire such that we observe a relatively similar degree of scatter for most of these curves: due to continuity reasons, the scatter naturally decreases for velocity offsets closer to our reference offset at $v_\mathrm{T} = 2000\,\mathrm{km/s}$ marked by the dashed black line.

However, the scatter does not change symmetrically around v_T, i.e., we find a more significant increase blueward of v_T compared to redward. As demonstrated by Keating et al. (2024b), this is due to the impact of infalling

gas as well as residual neutral gas in ionized patches which increasingly affects the shape of the transmission profiles in the vicinity of the Lyman-α line. These effects are also the reason why we cannot expect the damping wing to remain a one-parameter family at velocity offsets even closer to Lyman-α line center. Specifically, we see that the 68-percentile scatter has already increased to $\sim 6\,\%$ at $v = 1000\,\mathrm{km/s}$, and grows to $O(10\,\%)$ at Lyman-α line center. The scatter also grows in the other direction (i.e, further redward of v_T), albeit significantly more modestly than blueward. In this direction, the scatter is limited by the fact that transmission values are generally higher, approaching their upper bound of unity, and hence allowing for less scatter overall. As a result, the spread of the curves shows a mild peak around $v = 4000\,\mathrm{km/s}$, and then decreases again for even larger velocity offsets as the transmission values approach unity.

Overall, this demonstrates that the IGM transmission value at the reference velocity of $v_\mathrm{T} = 2000\,\mathrm{km/s}$ informs us about the *entire* shape of the IGM damping wing at *any* velocity offset $v \gtrsim 1500\,\mathrm{km/s}$ with remarkable precision, and hence a summary statistic that maximizes the information at this reference location also extracts the bulk of the information encoded in the entire damping wing imprint.

3.2.4 A second label: the distance to the first neutral patch

As discussed in the previous sections, the characteristic shape of the IGM damping wing is encoded to lowest order in the Lorentzian-weighted HI column density $N_\mathrm{HI}^\mathrm{DW}$ as defined in Eq. (3.6). By defining a second summary statistic, we can 1) further minimize the scatter among Lyman-α transmission skewers at a given set of parameter values, and 2) gain additional information about the reionization topology.

An obvious choice to achieve the first objective would be the location r_ion of the ionization front of the quasar, as this determines precisely where the optical depth integral in Eq. (3.1) effectively starts (c.f. the discussion in Section 3.2.2). However, at a fixed ionizing photon emission rate Q, and at a fixed column density $N_\mathrm{HI}^\mathrm{DW}$, the radius of the ionization front r_ion is, up to a minor contribution due to density fluctuations, largely degenerate with the quasar lifetime t_Q, as Eq. (3.9) suggests. The lifetime, however, is of direct physical interest and we therefore opt to rather keep t_Q as a post-quasar label.

By instead turning to the *pre*-quasar equivalent of r_ion, i.e., the distance r_patch between the source and the first neutral patch along the line of sight *in the pre-quasar topology*, we can simultaneously address both aforementioned objectives: while it is obvious that most long-lived quasars have carved out an ionized bubble extending significantly beyond the pre-quasar neutral patch location, the same is not true for objects whose ionization front r_ion

does not or only marginally exceed $r_{\rm patch}$. The latter can be the case primarily for young quasars, but also older ones where the first neutral patch is located at a large distance from the quasar itself. In any case, due to the fixed choice of the integration limit $r_{\rm min}$ for $N_{\rm HI}^{\rm DW}$, it is clear that $r_{\rm patch}$ encapsulates additional information about the pre-quasar ionization topology that is not fully captured by $N_{\rm HI}^{\rm DW}$. Besides this, $r_{\rm patch}$ is of direct physical interest as a key statistic of the topology of reionization, yielding complementary information to the values estimated from Lyman-α emission from galaxies (Mason & Gronke 2020; Hayes & Scarlata 2023; Umeda et al. 2024; Witstok et al. 2024; Torralba-Torregrosa et al. 2024; Lu et al. 2024a,b; Nikolić et al. 2025), and has already been introduced by Mason et al. (2026) as an additional summary statistic for IGM damping wings around galaxies.

Overall, this provides us with a three-parameter model, parameterizing quasar IGM damping wings as a function of $(N_{\rm HI}^{\rm DW}, r_{\rm patch}, t_{\rm Q})$. The two labels $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ are summary statistics of the *pre*-quasar topology, and could hence also be employed in the context of IGM damping wings towards galaxies, provided that the integration range of $N_{\rm HI}^{\rm DW}$ is adjusted accordingly. The third parameter $t_{\rm Q}$ is specific to quasars, encapsulating their impact as strong ionizing sources on the IGM transmission field.

3.2.5 Relation to previous definitions

Parameterizations aiming to better capture the characteristic shape of the IGM damping wing have recently been proposed in Chen (2024), Keating et al. (2024b), and Mason et al. (2026). Our $N_{\rm HI}^{\rm DW}$ statistic introduced in Section 3.2.1 particularly builds upon the study by Keating et al. (2024b), whereas the distance $r_{\rm patch}$ to the first neutral patch has already been employed by Mason et al. (2026) in the context of galaxy damping wings. In this section we investigate the similarities and differences among these definitions found in the literature and our own ones.

We chose to adopt as our first summary statistic the Lorentzian-weighted HI column density $N_{\rm HI}^{\rm DW}$ along the line of sight from the source, rather than the average velocity-weighted number density $\langle\!\langle n_{\rm HI} \rangle\!\rangle_{\rm Lor}$ as introduced in Keating et al. (2024b).[7] We deem this definition more appropriate for the following reasons: as an inherent line-of-sight quantity, $N_{\rm HI}^{\rm DW}$ emphasizes that the damping wing signature really only encodes one-dimensional information about the *local HI content in front of the quasar*. Only by combining the information from several such sightlines do we gain statistical information about the global average of the HI density and hence the global reionization topology. Secondly, the use of a line-of-sight quantity more clearly emphasizes the *physical scale* (i.e., $\sim 100\,$cMpc) impacting the IGM damping wing. In addition, the analogy to the column density of a local proximate damped

[7] Note that the symbol $\langle\!\langle n_{\rm HI} \rangle\!\rangle_{\rm Lor}$ is chosen in accordance with our own notation.

Lyman-α absorber (DLA) helps building intuition for the physical range of $N_{\mathrm{HI}}^{\mathrm{DW}}$, as well as the degree to which IGM damping wings and proximate DLAs can be disentangled.

We emphasize, however, that in practice, the notions of HI column densities and line-of-sight averaged HI number densities are equivalent, and we could easily convert between the two by realizing that $\langle\!\langle n_{\mathrm{HI}} \rangle\!\rangle_{\mathrm{Lor}} = \langle n_{\mathrm{H}} \rangle(z_{\mathrm{QSO}}) \cdot \langle\!\langle x_{\mathrm{HI}} \cdot \Delta \rangle\!\rangle_{\mathrm{Lor}}$ as follows from Eqs. (3.2) and (3.4). According to Eq. (3.6), this implies that the two statistics only differ by the geometrical factor $\mathcal{N}\left(\frac{r_{\mathrm{min}}}{r_{\mathrm{T}}}, \frac{r_{\mathrm{max}}}{r_{\mathrm{T}}}\right)$ as well as a factor of r_{T}. However, this equality demands that the average $\langle\!\langle n_{\mathrm{HI}} \rangle\!\rangle_{\mathrm{Lor}}$ be computed over the same integration range as $N_{\mathrm{HI}}^{\mathrm{DW}}$. The adequate lower integration limit for galaxy damping wings as considered in Keating et al. (2024b) would be $r_{\mathrm{min}} = 0$, in contrast to our choice of $r_{\mathrm{min}} = 4\,\mathrm{cMpc}$ due to the presence of a strong ionizing source which almost certainly ionizes away all neutral gas within the first few cMpc from the source. Further, with regards to quasar damping wings, Keating et al. (2024b) chose to operate on the post-quasar field, whereas we here opt to use pre-quasar labels with an adjusted integration limit r_{min}. This way, we directly constrain the pre-quasar field—which is of higher physical interest—at the price of a small (and, as we will show, negligible) additional amount of scatter.

In addition, despite the seeming agreement of our weighting function (Eq. (3.3)) with that adopted by Keating et al. (2024b), our summary statistic differs in that we do *not* realign the skewers in the way proposed in that work and in Chen (2024). We always keep the velocity grid centered at the Lyman-α line, whereas Chen (2024) and Keating et al. (2024b) re-center their velocity grid based on the individual realignment point of each transmission profile, defined as the point where the profile approaches zero for the first time. Note further that the meaning of the realignment point differs among the two works: Chen (2024) does not perform radiative transfer but attempts to model quasar lifetime effects through a simplistic model for the quasar ionization front. In return, their realignment point thus corresponds to what is conventionally measured as the *proximity zone size* in the spectrum of a quasar, often used as a summary statistic to gain information about its lifetime. Keating et al. (2024b), on the other hand, models lifetime effects through ionizing radiative transfer and performs the realignment at various but *fixed* lifetime values (as well as a set of pre-quasar skewers). Hence, the pre-quasar realignment point encodes similar information as our r_{patch} statistic, while the post-quasar one largely coincides with the proximity zone size *at a fixed lifetime value*, hence largely encoding the scatter due to density fluctuations in the IGM.

As a result, our three-parameter model makes both types of re-centering obsolete: while Chen (2024)'s realignment point is well characterized by the quasar lifetime t_{Q} which is a separate parameter of our model, Keating et al.

(2024b)'s pre-quasar realignment point is largely represented by our $r_{\rm patch}$ statistic. We find a negligible impact from realigning post-quasar skewers at a fixed lifetime which, while removing some excess scatter due to density fluctuations in the IGM, would be largely degenerate with $t_{\rm Q}$. The virtue of $t_{\rm Q}$ and $r_{\rm patch}$ is that both of these summaries are of direct physical interest, and we thus prefer to use them over abstract realignment points.

Furthermore, the distance $r_{\rm patch}$ between the source and the first neutral patch has been introduced in Mason et al. (2026) as an additional summary statistic in the context of galaxy IGM damping wings. Here we adopt the same *pre*-quasar statistic, and show that it remains a meaningful summary even for *post*-quasar IGM transmission profiles. Also note in this context that one of the earliest damping wing studies (Schroeder et al. 2013) already aimed to constrain the distance between the quasar and the first neutral patch, albeit in the *post*-quasar topology rather than the *pre*-quasar one as targeted in this work. Due to the absence of radiative transfer in their modeling, similar conclusions apply as drawn above regarding Chen (2024)'s realignment point, i.e., at fixed ionizing photon emission rate Q, the post-quasar distance to the first neutral patch encodes very similar information as the quasar lifetime $t_{\rm Q}$ which is already explicitly included as a parameter in our modeling framework.

3.3 Quantifying the IGM transmission scatter

With the definition of our new set of summary statistics at hand, we now describe how we simulate realistic IGM transmission profiles based on cosmological simulations combined with semi-numerical reionization topologies and 1d radiative transfer. We then use these simulated profiles to demonstrate the advantages of our labels over the global IGM neutral fraction $\langle x_{\rm HI} \rangle$ in parameterizing the characteristic shape of the IGM damping wing by comparing the IGM transmission scatter at fixed parameter values within the global and the local parameterization.

3.3.1 Simulating IGM transmission fields

We generate IGM transmission profiles $\boldsymbol{t}$ based on the hybrid approach introduced in Davies et al. (2018a), combining hydrodynamical skewers extracted from the Nyx cosmological simulations and $x_{\rm HI}$ skewers from semi-numerical reionization topologies with 1d ionizing radiative transfer.

We start by extracting 6×200 density, velocity, and temperature skewers originating at the 200 most massive halos ($M_{\rm halo} \geq 2 \times 10^{11}\,M_{\odot}$) from the $z = 7.0$ snapshot of the Nyx simulations (Almgren et al. 2013; Lukić et al. 2015), containing 4096^3 baryon and another 4096^3 dark matter particles in a 100 cMpc/h box. Our fiducial reionization topologies are simulated using an

adapted version of `21cmFast` (Mesinger et al. 2011; Davies & Furlanetto 2022) in a larger 400 cMpc box on a 2048^3 initial and a 512^3 output grid, providing us with sufficient statistics to probe the distribution of neutral/ionized regions around the 500 rarest, most massive halos ($M_{\rm halo} \geq 3 \times 10^{11}\,M_\odot$). We generate such topologies at 21 different global IGM neutral fractions $\langle x_{\rm HI} \rangle$, where we achieve $\langle x_{\rm HI} \rangle = 0.05, 0.1, ..., 0.95$ by tuning the ionizing efficiency ζ, and trivially add a completely ionized ($\langle x_{\rm HI} \rangle = 0.0$) and a completely neutral ($\langle x_{\rm HI} \rangle = 1.0$) model.

We subsequently combine each Nyx sightline with a random $x_{\rm HI}$ skewer pointing in a random direction originating at one of the aforementioned 500 most massive halos of the `21cmFast` box. Here we adopt the Nyx temperature field for all initially ionized regions, and assume an initially cold IGM for all neutral regions, fixed to $T = 2000\,{\rm K}$. To model the small-scale impact of the ionizing quasar radiation, we then perform one-dimensional radiative transfer along these sightlines (Davies et al. 2016), assuming a 'light bulb' lightcurve where the quasar has been shining at constant luminosity throughout its lifetime $t_{\rm Q}$ until the redshift $z_{\rm QSO}$ of interest. We here consider a model with $z_{\rm QSO} = 7.54$ and an ionizing photon emission rate of $Q = 10^{57.14}\,{\rm s}^{-1}$, resembling the quasar ULAS J1342+0928. We compute the time evolution up to a maximum lifetime of $t_{\rm Q} = 10^8\,{\rm yr}$, and store 51 intermediate outputs on a logarithmically spaced grid between $t_{\rm Q} = 10^3$ and $10^8\,{\rm yr}$. We then convolve the resulting output fields from the radiative transfer code with a Voigt profile to obtain the final IGM transmission profiles. This provides us with a set of 1200 IGM transmission profiles on a 21×51 grid of $(\langle x_{\rm HI} \rangle, t_{\rm Q})$ values. For the bulk of this work, we focus on the three representative lifetime values of $t_{\rm Q} = 10^4$, 10^6 and $10^8\,{\rm yr}$, representing a young, an intermediate, and a long-lived quasar.

We exclude sightlines that exhibit strong proximate optically thick absorption line systems. This is necessary due to the lack of a subgrid prescription for star formation in the Nyx simulations which implies that the properties and abundance of such systems are not guaranteed to match those found in the real universe. Observationally, such objects can be excluded based on the presence of associated metal absorption lines in the spectrum (Davies et al. 2025).

To identify such sightlines in our simulations, we consider all 1200 skewers from the completely ionized pre-quasar topology, divide each of them into chunks of size 0.1 pMpc, and compute the (unweighted) HI column density within each of these chunks. We then exclude all sightlines where at least one chunk (located within the first 5000 km/s from the source) exceeds a column density threshold of $10^{19}\,{\rm cm}^{-2}$. After applying this criterion, 852 skewers remain, and we end up with a set of $852 \times 21 \times 51$ transmission profiles covering the full $(\langle x_{\rm HI} \rangle, t_{\rm Q})$-parameter space.

3.3.2 Global parameterization

We start by investigating the IGM transmission scatter within the global parameterization, depicted in Figure 3.3 at select $(\langle x_{\rm HI}\rangle, t_{\rm Q})$ parameter values. Here, the solid lines mark the median transmission value of the 852 profiles, and the shaded regions denote the 68-percentile scatter (enclosed by the 16-th and the 84-th percentile). Besides the clear emergence of the damping wing imprint in more neutral environments (i.e., at higher global IGM neutral fractions $\langle x_{\rm HI}\rangle$ and shorter quasar lifetimes $t_{\rm Q}$), we can see that the bulk of the scatter emerges in the proximity zone due to density fluctuations in the quasar ionized IGM, as well as variation in the location of the quasar ionization front. However, even in the smooth damping wing region redward of the Lyman-α line we can identify a significant amount of scatter. This scatter is a direct consequence of the stochastic nature of reionization. In other words, at a given global value of the IGM neutral fraction $\langle x_{\rm HI}\rangle$, the *local* HI column density giving rise to the observed damping wing imprint can differ significantly from the global average.

To get a more quantitative idea of this scatter, we focus on the distribution of IGM transmission values at a specific location of the profile, at $v = v_{\rm T}$.[8] Figure 3.5 shows a heatmap of the 68-percentile widths $\sigma(t(v = v_{\rm T}))$ of these distributions as a function of the astrophysical parameters $\langle x_{\rm HI}\rangle$ and $t_{\rm Q}$, where $\sigma(t(v = v_{\rm T}))$ is defined as half the distance between 84-th and 16-th percentile.

For short quasar lifetimes $t_{\rm Q} \lesssim 10^4\,{\rm yr}$, we find similarly wide distributions as Keating et al. (2024b) do in profiles from the Sherwood-Relics simulations in the complete absence of a quasar. As expected, the scatter decreases with increasing quasar lifetime due to the larger quasar ionized regions which suppress the IGM damping wing such that the transmission is very close to unity redward of line center. Similarly, the scatter is smaller than 1 % in a fully ionized universe, even at the shortest quasar lifetimes, and grows with increasing IGM neutral fraction $\langle x_{\rm HI}\rangle$, peaking around $\langle x_{\rm HI}\rangle \sim 0.7$, and reaching values of up to $\sim 7\,\%$ at the shortest lifetimes of $10^3 - 10^4\,{\rm yr}$. Notably, the scatter is not exclusively sourced by the distribution of ionized patches of the underlying reionization topology but also by the mere presence of density fluctuations which are in fact the *only* source of scatter in an entirely neutral universe where $\langle x_{\rm HI}\rangle = 1$ and thus no topology variations are present.

In realistic settings where the intrinsic continuum of the quasar is unknown and has to be reconstructed too, the amount of scatter among different IGM transmission profiles seen in Figure 3.5 is comparable to the $\sim 5\,\%$ uncertainties in the reconstructed quasar continuum (c.f. Figure 6 in Hennawi

[8] As demonstrated in Section 3.2.3, due to the strong correlation across the entire IGM damping wing imprint, it suffices to evaluate $\tau_{\rm DW}$ at a single pixel to get a sufficient picture of the scatter across the entire IGM transmission profile.

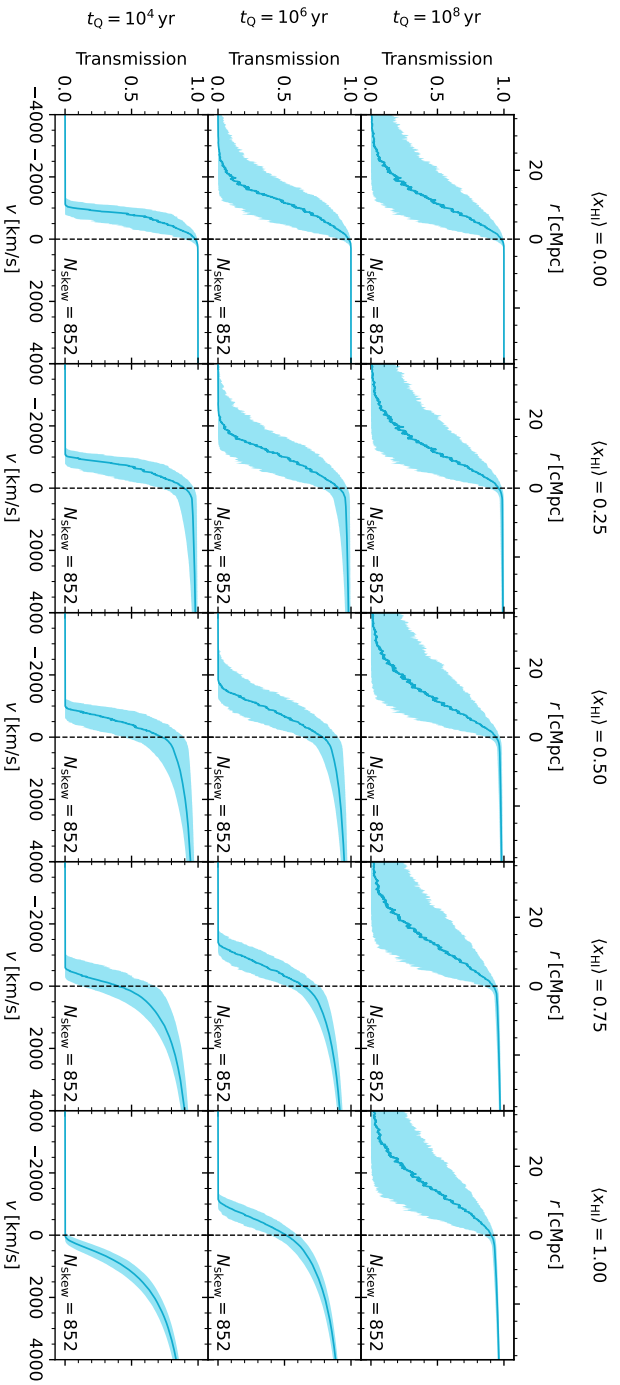

Figure 3.3: Median and 68-percentile scatter of the simulated IGM transmission profiles based on our semi-numerical reionization topology in the global $(\langle x_{\rm HI}\rangle, t_{\rm Q})$ parametenerization. The profiles are shown on a representative grid of $(\langle x_{\rm HI}\rangle, t_{\rm Q})$ parameter values.

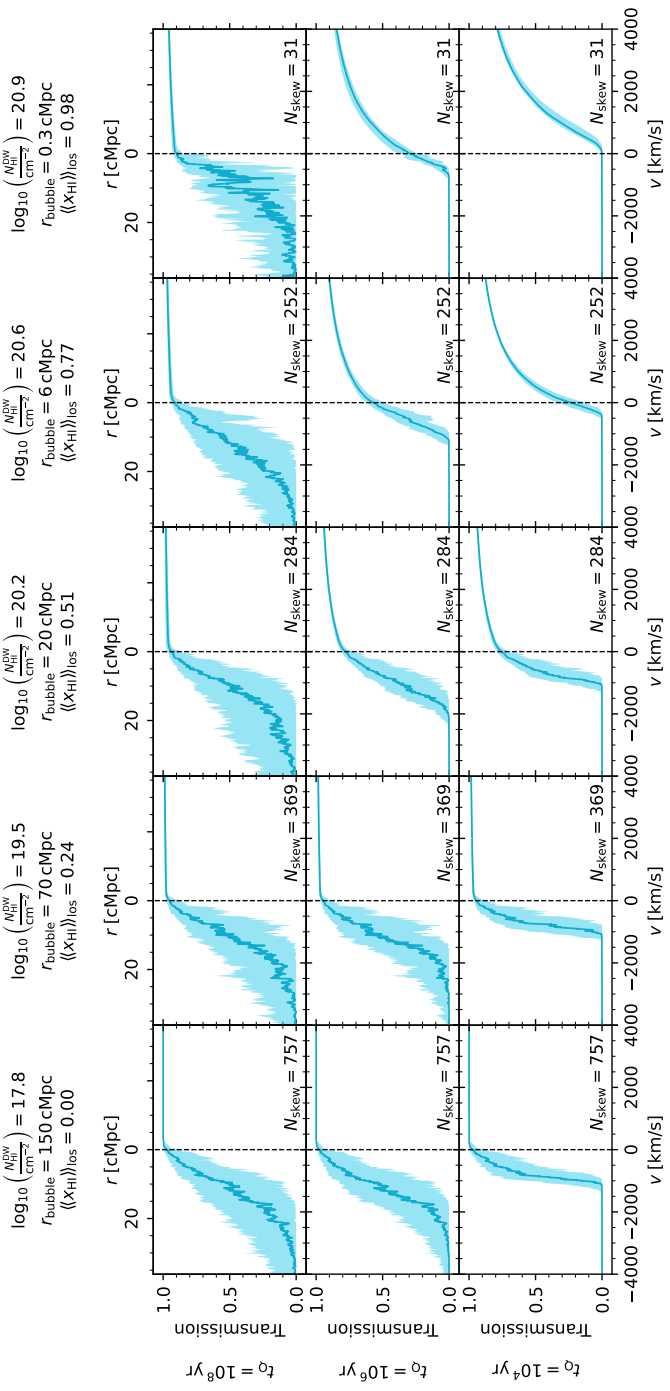

Figure 3.4: Median and 68-percentile scatter of the simulated IGM transmission profiles based on our semi-numerical reionization topology in the local $(N_{\rm HI}^{\rm DW}, r_{\rm patch}, t_{\rm Q})$ parameterization. The profiles are shown on a representative grid of $(N_{\rm HI}^{\rm DW}, r_{\rm patch}, t_{\rm Q})$ parameter values for the most typical combinations of $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$ across the full parameter space.

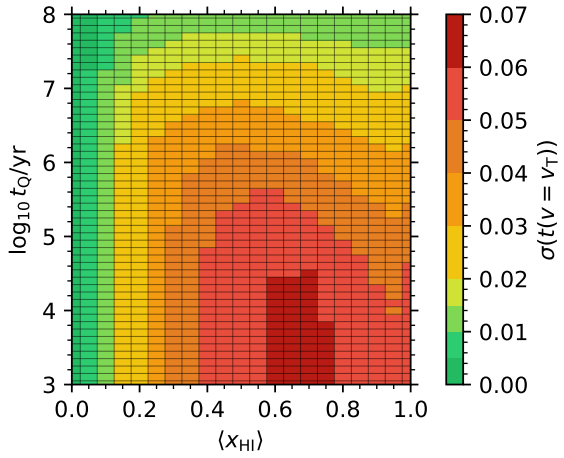

Figure 3.5: Sightline-to-sightline scatter of the IGM damping wing as a function of astrophysical parameter space in the global $(\langle x_{\rm HI}\rangle, t_{\rm Q})$-parameterization. The scatter is quantified as half the 68-percentile width $\sigma(t(v = v_{\rm T}))$ of the distribution of IGM transmission values at $v = v_{\rm T}$ among all sightlines at a given location in $(\langle x_{\rm HI}\rangle, t_{\rm Q})$-parameter space.

et al. 2025). As a result, this excess stochasticity in the IGM transmission model significantly adds to the total error budget on the reconstructed $\langle x_{\rm HI}\rangle$ and $t_{\rm Q}$ values (Kist et al. 2025b), whereas we could isolate it from the task of reconstructing the transmission profile if we were to work with a parameterization that more tightly captures the characteristic shape of IGM damping wings.

3.3.3 Local parameterization

We now demonstrate that the scatter among individual IGM transmission profiles decreases significantly when labelling the profiles with the local summary statistics introduced in Section 3.2 in place of the global IGM neutral fraction $\langle x_{\rm HI}\rangle$.

To that end, we compute the Lorentzian-weighted HI column density $N_{\rm HI}^{\rm DW}$ as defined in Eq. (3.6) and the distance $r_{\rm patch}$ to the first neutral patch for each (pre-quasar) sightline and subsequently aggregate all post-quasar profiles according to these new labels. We fix the velocity offset and the integration range for $N_{\rm HI}^{\rm DW}$ as discussed in Sections 3.2.1 and 3.2.2. To compute $r_{\rm patch}$, we smooth the $x_{\rm HI}$ field with a box-car filter of size 0.5 cMpc and determine $r_{\rm patch}$ as the distance from the source where the smoothed $x_{\rm HI}$ field first becomes more than 50 % neutral. This smoothing procedure makes sure that our $r_{\rm patch}$ measurement identifies the first *spatially extended*

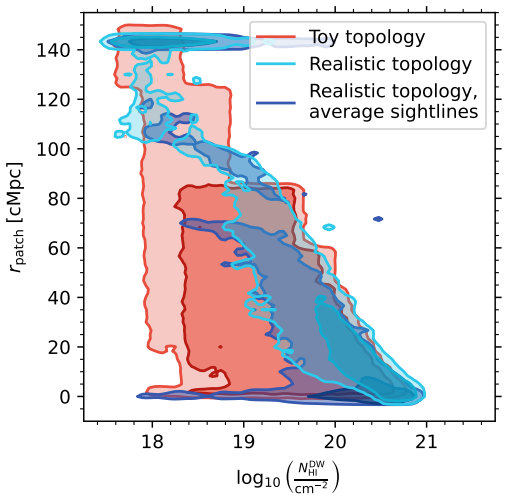

Figure 3.6: Distribution of our local summary statistics $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$ among all simulated IGM transmission profiles based on semi-numerical $x_{\rm HI}$ skewers originating from massive halos (cyan; see Section 3.3.1) and random locations (dark blue), and analytical $x_{\rm HI}$ skewers generated according to our toy prescription (red; see Section 3.4.1). Dark shaded regions denote the 68 % contours, lighter ones the 95 % ones. The distributions are smoothed via a kernel density estimation (KDE) with a Gaussian kernel of size $(0.05, 1.0)$ in $(\log_{10} N_{\rm HI}^{\rm DW}/{\rm cm}^{-2}, r_{\rm patch}/{\rm cMpc})$ parameter space. The sharp edges of the toy model contours result from the fact that these profiles are simulated on a fixed grid of parameter values.

neutral patch without being impacted by individual fluctuations in the $x_{\rm HI}$ field which do not have a notable impact on the damping wing shape. We find that we are not overly sensitive to the exact choices of these parameters.

The cyan contours in Figure 3.6 depict the 68 and 95 % regions of the distribution of $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ values among all 852×21 sightlines.[9] As our $\langle x_{\rm HI} \rangle$ parameter grid uniformly covers the range between zero and one, this is the $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ distribution that follows under the assumption that all neutral fraction values $0 \leq \langle x_{\rm HI} \rangle \leq 1$ are equally likely, i.e., in the absence of any knowledge about the reionization state of the IGM. We observe a clear degeneracy between $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$, where a higher column density $N_{\rm HI}^{\rm DW}$ implies a smaller $r_{\rm patch}$ value and vice versa. This degeneracy arises due to the fact that a closer distance to the first neutral patch clearly allows for more neutral material distributed within the range over which $N_{\rm HI}^{\rm DW}$ is computed. It is also apparent, however, that this is not a perfect degeneracy,

[9] Recall we are considering 852 different sightlines at 21 distinct global IGM neutral fraction values $\langle x_{\rm HI} \rangle$. Our 51 different quasar lifetime values do not play a role as $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$ are defined on the pre-quasar topology.

and therefore the $r_{\rm patch}$ statistic contains additional information not encoded in $N_{\rm HI}^{\rm DW}$.

To quantify this information, we define (unequally spaced) bins for the two labels and aggregate the transmission profiles accordingly. The binning is informed by the number of available skewers and the degree to which parameter variations impact the resulting IGM transmission profiles. Specifically, we choose a finer binning towards higher column densities and shorter distances to the first neutral patch which cause more variation in the profiles due to increasingly strong damping wing absorption.

The scatter among the aggregated IGM transmission profiles in select parameter bins representative of the variations across the full parameter space is shown in Figure 3.4. While a one-to-one comparison between individual bins in Figures 3.3 and 3.4 is not possible due to the redefined bin labels, we did compute the effective line-of-sight averaged neutral fractions $\langle\!\langle x_{\rm HI}\rangle\!\rangle_{\rm los}$ of the profiles in each parameter bin, and we depict in Figure 3.4 selected bins where the median $\langle\!\langle x_{\rm HI}\rangle\!\rangle_{\rm los}$ matches the cosmic average $\langle x_{\rm HI}\rangle$ values shown in Figure 3.3.[10] With this in mind, we can certainly note that the scatter redward of the Lyman-α line in Figure 3.4 is consistently lower than in Figure 3.3, implying that we identified a tighter parameterization of the characteristic shape of the IGM damping wing. Note that as expected, the scatter in the Lyman-α forest region is not impacted by the new parameterization.

For a more quantitative picture across the full range of parameter space, we summarize the 68-percentile widths $\sigma(t(v = v_{\rm T}))$ of the distributions of IGM transmission values at $v = v_{\rm T}$ in all $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ parameter bins in Figure 3.7.[11] In each given row, we are showing from left to right the scatter at the three representative lifetimes of $t_{\rm Q} = 10^4$, 10^6 and 10^8 yr, respectively. For reference, the three upper panels depict the scatter in the global $\langle x_{\rm HI}\rangle$ parameterization,[12] whereas the three middle panels depict the scatter of the same profiles aggregated according to the $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ parameterization. The three lower panels will be discussed in Section 3.4. The transparency of the colorbar is proportional to the logarithm of the number of profiles in each bin and thus represents the distribution of $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ values which was already explicitly depicted in Figure 3.6.

[10] Recall that we denote cosmic averages with single angle brackets $\langle . \rangle$, and their $z = 0$ values as $\langle . \rangle_0$ throughout. We always denote line-of-sight averages with double angle brackets, where $\langle\!\langle . \rangle\!\rangle_{\rm los}$ is the conventional line-of-sight average, and $\langle\!\langle . \rangle\!\rangle_{\rm Lor}$ as defined in Eq. (3.4) additionally involves a Lorentzian weighting function.

[11] Recall that we evaluate the IGM transmission scatter at $v = v_{\rm T}$, i.e., exactly at the reference velocity offset $v_{\rm T}$ used to determine $N_{\rm HI}^{\rm DW}$. Due to the strong pixel-by-pixel correlation across the entire IGM damping wing signature, this does *not* mean we are underestimating the scatter, since this scatter can maximally contribute another $\sim 0.5 - 1\,\%$ to $\sigma(t(v = v_{\rm T}))$ as demonstrated in Section 3.2.3.

[12] Note that these are exactly the $t_{\rm Q} = 10^4$, 10^6 and 10^8 yr slices of the heatmap in Figure 3.5.

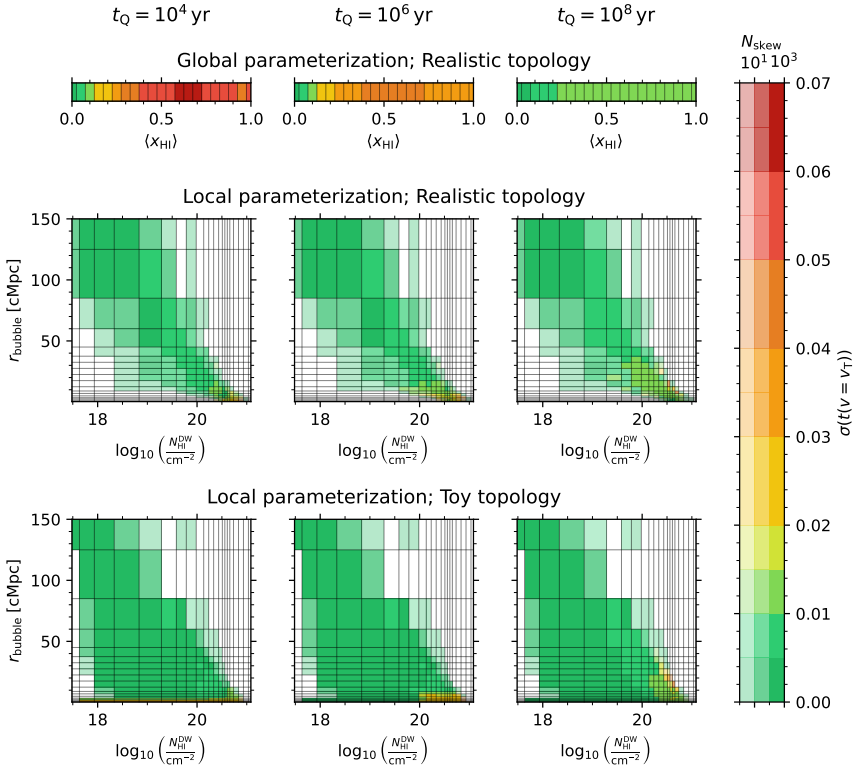

Figure 3.7: Sightline-to-sightline scatter of the IGM damping wing as a function of astrophysical parameter space for different parameterizations and topologies. The scatter is quantified as half the 68-percentile width $\sigma(t(v = v_T))$ of the distribution of IGM transmission values at $v = v_T$ among all sightlines at a given location in parameter space, and at quasar lifetimes of $t_Q = 10^4$ yr (left column), 10^6 yr (middle column) and 10^8 yr (right column). *Top row:* Profiles based on semi-numerical x_{HI} skewers (see Section 3.3.1) in the global $\langle x_{HI} \rangle$ parameterization. *Middle row:* Same profiles as above, but in the local (N_{HI}^{DW}, r_{patch}) parameterization. *Bottom row:* Profiles based on the analytical toy x_{HI} skewers (see Section 3.4.1) in the local (N_{HI}^{DW}, r_{patch}) parameterization. Bins which are occupied by a lower number of sightlines are more transparent, and white if not occupied at all.

Remarkably, the transmission in the vast majority of bins in the local parameterization seen in the three middle panels of Figure 3.7 varies by $\lesssim 1-2\,\%$, whereas it is consistently higher in the global parameterization depicted in the upper panels, reaching up to $\sim 7\,\%$ at $t_{\rm Q} = 10^4\,{\rm yr}$. This shows that our local summary statistics capture the characteristic shape of the IGM damping wing much more tightly than the global IGM neutral fraction $\langle x_{\rm HI}\rangle$.

At all lifetimes, however, we do observe a mild increase in scatter towards the highest column densities where the damping wing imprint gets more pronounced. This increased scatter is a result of the fact that we chose to define the two summaries on the *pre*-quasar $x_{\rm HI}$ field. By fixing the integration limit for $N_{\rm HI}^{\rm DW}$ independently of quasar lifetime, we did not fully pay tribute to the variable radius of the ionization front $r_{\rm ion}$ where the integration effectively starts. Hence, by fixing $r_{\rm min} = 4\,{\rm cMpc}$, we are theoretically including too much or too little neutral material in the $N_{\rm HI}^{\rm DW}$ statistic, depending on the size of the ionized bubble the quasar has carved out, which in turn depends on the quasar lifetime, as discussed in Section 3.2.2. Nevertheless, the fact that the resulting scatter does not exceed $\sim 2\,\%$ even at the highest HI column densities shows that our approximation was a valid one, still leading to a tight parameterization of the *post*-quasar IGM transmission field nearly independently of quasar lifetime.

3.4 Towards a topology independent local IGM damping wing parameterization

We introduced in the previous section a new three-parameter model for quasar IGM damping wings. We classified simulated IGM transmission profiles from a given reionization topology according to these labels and showed that the sightline-to-sightline scatter drastically decreases in this new parameterization as compared to the conventional (e.g. Greig et al. 2017b; Davies et al. 2018a; Ďurovčíková et al. 2024; Hennawi et al. 2025) one- or two-parameter model based on the global IGM neutral fraction $\langle x_{\rm HI}\rangle$ (and the quasar lifetime $t_{\rm Q}$). Specifically, by replacing $\langle x_{\rm HI}\rangle$ with the two *local* summary statistics $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$, we were able to reduce the scatter in the damping wing region of the profile down to $\lesssim 1\,\%$ across the entire range of parameter space. In this section, we go beyond these results and demonstrate that our conclusions hold true largely independently of the underlying reionization topology.

This is because once our local summaries are set, the detailed distribution of neutral hydrogen along the line of sight, set by the reionization topology, does not matter anymore for the shape of the profile in the damping wing region of the spectrum. Instead, all relevant information about the damping

wing shape is already contained in $N_{\rm HI}^{\rm DW}$ as Eq. (3.7) shows. The only remaining sources of scatter are the required approximations for this equality to hold, i.e., the finite integration limits of $N_{\rm HI}^{\rm DW}$, any differences between pre- and post-quasar density field that have not been mitigated by the choice of the lower integration limit and the inclusion of $r_{\rm patch}$ as additional label, as well as any deviations of the damping wing shape from being a one-parameter family determined at a fixed reference distance (or velocity offset). However, all these contributions remain negligible as we showed in the previous section, neither do our conclusions change if the profiles originate from a different reionization topology as we will demonstrate here.

The scatter in the proximity zone region remains unaffected either because here the dominant source of sightline-to-sightline scatter is density fluctuations, while the $x_{\rm HI}$ field—only varying discontinuously between ionized and neutral patches with values close to zero and one—effectively acts as a window function for the fluctuations in the density field (see Eq. (3.2) and the example realization of these fields in Fig. 3.1) which are already accounted for by considering a statistical sample of density skewers (independently of reionization topology). As a result, all effects of the ionization topology on the observed IGM transmission profile for a given sightline are already well captured in the two summary statistics $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$.

In Bayesian language, the mere effect of assuming a specific reionization topology is therefore imposing a corresponding prior on the parameters $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$ according to their distribution within that topology. For future damping wing analyses, this will allow one to perform a clear methodological distinction between 1) the topology-insensitive task of inferring the local damping wing statistics $(N_{\rm HI}^{\rm DW}, r_{\rm patch}, t_{\rm Q})$, and 2) tying these constraints to a particular reionization topology, constraining not only the timing of reionization via the global IGM neutral fraction $\langle x_{\rm HI}\rangle$, but also its topology which remains unconstrained by $\langle x_{\rm HI}\rangle$ alone. It is immediately clear that measurements of $N_{\rm HI}^{\rm DW}$ (as a statistical moment of the $x_{\rm HI}$ field) and $r_{\rm patch}$ for a statistical ensemble of objects carry additional topological information that we can extract and leverage to compare reionization models whose topologies are different.

Here we demonstrate the topology independence of our labels[13] in a simplified fashion by introducing a highly simplistic toy model to generate $x_{\rm HI}$ skewers analytically rather than extracting them from a realistic semi-numerical reionization topology. Specifically, we show that within our local parameter framework both models are statistically equivalent in terms of the resulting median IGM transmission profiles and scatter.

The specifics of the toy prescription which we introduce in the following section are set keeping in mind the practical task of performing inference with

[13]Again, the topology information is contained in the *distribution* of these summaries while the summaries themselves can be defined and measured topology-independently.

respect to these new local summary statistics. Specifically, we introduced in Hennawi et al. (2025) a fully Bayesian framework for inferring astrophysical model parameters from high-redshift quasar spectra. Although originally introduced to directly infer (among other parameters) the *global* IGM neutral fraction, the framework is equally applicable to other parameters governing the IGM transmission field such as our local summary statistics. All which is needed is an estimate of the IGM transmission likelihood given the model parameters. Under the assumption of Gaussianity, this practically means we need to be able to estimate means and covariances of the IGM transmission field smoothly as a function of the model parameters $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$.

As $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$ are derived quantities for IGM transmission profiles from the semi-numerical reionization simulation, the number of available sightlines can vary significantly from parameter bin to parameter bin. Especially for $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ values that are rare in a given reionization topology, our covariance estimates can easily become so noisy that the likelihood function is not guaranteed to vary smoothly as a function of $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$. While one could certainly beat down this noise with a massively increased number of simulations, building an analytical toy prescription where the continuity of the IGM transmission field with respect to $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$ is intrinsically built in drastically reduces the number of required sightlines and facilitates the inference task for future applications, even though establishing the full inference pipeline would exceed the scope of this work.

As far as this work is concerned, the toy prescription provides us with a simple way of testing the sensitivity of our summary statistics to the underlying reionization topology by comparing the scatter of the IGM transmission profiles within these two very different topologies—the latter topology not even physically motivated. The fact that this scatter agrees remarkably well between the two demonstrates the topology-insensitivity of our labels and legitimizes the future use of the analytical toy prescription for the purpose of astrophysical parameter inference.

3.4.1 An analytic toy prescription for generating neutral fraction profiles

To generate IGM transmission profiles according to the toy prescription which we will introduce now, we follow the same hybrid approach outlined in Section 3.3.1, only replacing the semi-numerical $x_{\rm HI}$ skewers with analytically generated ones. This implies in particular that we use the same hydrodynamical skewers as described in that section and the same radiative transfer code to model the impact of the quasar ionizing radiation.

3.4.1.1 Description of the prescription

We start by realizing that for a given density skewer, there exists a minimal and a maximal HI column density value $N_{\mathrm{HI,min}}^{\mathrm{DW}}$ and $N_{\mathrm{HI,max}}^{\mathrm{DW}}$ as computed according to Eq. (3.6). These values can only be achieved if the IGM is completely ionized or completely neutral (within the integration range), respectively. To determine $N_{\mathrm{HI,min}}^{\mathrm{DW}}$, we first have to compute the equilibrium ionization state for a fully ionized sightline based on the input UV background and temperature field, providing us with an x_{HI} field with values on the order of $O(10^{-3} - 10^{-4})$ which we can then use to compute $N_{\mathrm{HI,min}}^{\mathrm{DW}}$. To determine $N_{\mathrm{HI,max}}^{\mathrm{DW}}$, we subsequently set $x_{\mathrm{HI}} = 1$ at all distances greater than r_{patch}. Note that $N_{\mathrm{HI,min}}^{\mathrm{DW}}$ and $N_{\mathrm{HI,max}}^{\mathrm{DW}}$ differ from sightline to sightline, based on the specific realization of density fluctuations along the line of sight.

To achieve any column density value $\hat{N}_{\mathrm{HI}}^{\mathrm{DW}}$ between $N_{\mathrm{HI,min}}^{\mathrm{DW}}$ and $\leq N_{\mathrm{HI,max}}^{\mathrm{DW}}$, we start from the completely ionized sightline and add in neutral material until the desired $\hat{N}_{\mathrm{HI}}^{\mathrm{DW}}$ value is reached. We do so by subsequently adding in neutral patches i of a minimum size Δr_{min} and a maximum size Δr_{max} originating at locations r_i and growing them in the direction towards the observer.[14] Note that this procedure is by no means meant to resemble any physical processes occurring during reionization. It is a purely numerical procedure to ensure that small changes in the model parameters $N_{\mathrm{HI}}^{\mathrm{DW}}$ and r_{patch} result in small changes in the x_{HI} field and hence the IGM transmission profiles, allowing us to obtain smoothly varying means and covariances for future inference applications. To that end, we always follow the same sequence of locations $(r_i)_{i\in\mathbb{N}}$ when inserting new patches into a given sightline. In particular, we obtain the x_{HI} field for a given sightline and a desired parameter tuple $(\hat{N}_{\mathrm{HI}}^{\mathrm{DW}}, \hat{r}_{\mathrm{patch}})$ according to the following procedure:

1. Insert a neutral patch of size Δr_{min} at the location $r_0 = \hat{r}_{\mathrm{patch}}$, extending out to $\hat{r}_{\mathrm{patch}} + \Delta r_{\mathrm{min}}$.

2. Grow the current (i-th) patch (originating at r_i) in the direction of the observer by subsequently setting x_{HI} pixel values adjacent to the outer edge of the patch to unity. In case a given pixel is already neutral because it overlaps with one of the previous neutral patches, this step has no effect.

3. If the i-th patch has reached a size of Δr_{max}, insert a new patch of size Δr_{min} originating at the next location r_{i+1}.

4. If $r_{i+1} < \hat{r}_{\mathrm{patch}}$, do not add the $(i+1)$-th patch and instead skip to the $(i+2)$-th one because $\hat{r}_{\mathrm{patch}}$ is by definition the distance to the

[14] We refer to r_i as the 'origin' of this patch and stress that it can only 'grow' in the direction away from the source.

neutral patch *closest* to the source. If necessary, repeat this step until a location is found where a patch can be added.

5. While $N_{\rm HI}^{\rm DW} < \hat{N}_{\rm HI}^{\rm DW}$, repeat steps (ii) - (iv).

In this way, we are able to generate $x_{\rm HI}$ skewers on any desired $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ grid, varying smoothly with both parameters. Note, however, that the $N_{\rm HI}^{\rm DW}$ ranges are physically restricted by $N_{\rm HI,min}^{\rm DW}$ and $N_{\rm HI,max}^{\rm DW}$ for each individual skewer, so it is not necessarily the case that there exists a realization of every sightline at each point in $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ parameter space.

Moreover, for any given sightline, the maximal $N_{\rm HI}^{\rm DW}$ value changes as a function of $r_{\rm patch}$. This is because a skewer with a certain neutral patch distance $r_{\rm patch}$ is by definition guaranteed to be ionized at $r < r_{\rm patch}$, so the highest possible column density $N_{\rm HI,max}^{\rm DW}$ for a given sightline decreases with increasing $r_{\rm patch}$. A subspace of the full $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ parameter space is therefore *physically inaccessible*, and this is determined *exclusively* by the distribution of density fluctuations in the IGM, while entirely independent of any further assumptions about the reionization topology. Note in particular that our synthetic sightlines do not even necessarily respect the correlation between the locations of ionized patches and overdensity dictated by any inside-out reionization scenario.

The prescription introduced in this section comes with two hyperparameters $\Delta r_{\rm min}$ and $\Delta r_{\rm max}$ determining the minimum and the maximum size of neutral patches that we insert into the sightlines. We choose these as $\Delta r_{\rm min} = 0.5\,{\rm cMpc}$ and $\Delta r_{\rm max} = 5.0\,{\rm cMpc}$, loosely informed by the sizes we empirically find for skewers in our semi-numerical reionization topology (compare also Xu et al. 2017), and noting that we are not sensitive to these exact choices.

3.4.1.2 Generating IGM transmission profiles

Using this prescription, we now generate a grid of IGM transmission profiles based on the same 852 DLA-excluded density sightlines as in Section 3.3.2. We construct the grid in accordance with the 21 and 18 parameter bins used in Section 3.3.3 to classify the realistic profiles with respect to $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$, respectively. Accordingly, we perform ionizing radiative transfer for the same lifetime values as in Section 3.3.1, again focusing on the outputs after $t_{\rm Q} = 10^4$, 10^6 and $10^8\,{\rm yr}$.

An example sightline can be found on the right-hand side of Figure 3.1. The resulting IGM transmission profile shown in the top panel is overlaid in red with the one we obtained using a semi-numerical $x_{\rm HI}$ skewer (blue; left-hand panels) as considered in Section 3.3.1. For optimal comparability to the realistic reionization topology, we used the same underlying density skewer and matched the parameter values, i.e., both the semi-numerical sightline (blue) and the analytical toy-sightline (red) have $\log_{10} N_{\rm HI}^{\rm DW}/{\rm cm}^{-2} =$

20.1, $r_{\rm patch} = 8\,{\rm cMpc}$, and $t_{\rm Q} = 10^6\,{\rm yr}$, where the semi-numerical sightline originates from a topology with a global IGM neutral fraction of $\langle x_{\rm HI} \rangle = 0.65$. The pre-quasar versions of all physical fields are shown in dark, and those of the IGM transmission field in lighter colors.

Despite the locally very different $x_{\rm HI}$ fields, the resulting IGM transmission profiles look almost identical in the smooth damping wing region redward of the Lyman-α line, and even their proximity zones agree relatively well due to the matching density profiles giving rise to similar Lyman-α absorption signatures. The actual location of the quasar ionization front, however, differs notably among the two examples due to the larger amounts of neutral gas near the quasar for the semi-numerical sightline. Its proximity zone therefore does not extend quite as far out as that of the toy sightline. However, as we will show statistically in the next section, such differences in the proximity zone region due to the distribution of neutral material are entirely outweighed by the differences caused by density fluctuations alone. The damping wing region remains completely unaffected, as our $N_{\rm HI}^{\rm DW}$ label by construction accounts for neutral material along the *entire* line of sight with the correct weighting. This suggests that our parameterization well captures the characteristic shape of the IGM damping wing, even in a case where the sightlines originate from highly different reionization topologies. We proceed in the following section with a statistical confirmation of these sightline-based observations.

3.4.2 Comparison to a realistic reionization topology

We now investigate the scatter in the IGM transmission profiles based on our newly introduced toy prescription, and compare it to that found in Section 3.3.3 in the context of a more realistic semi-numerical reionization topology. We depict the median and the 68-percentile scatter of the IGM transmission profiles at representative locations in $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ parameter space in Figures 3.8, 3.9 and 3.10. The number of sightlines based on which the percentiles are computed are denoted in each panel. The three different figures show the same sets of profiles at quasar lifetimes of $t_{\rm Q} = 10^4$, 10^6 and $10^8\,{\rm yr}$, respectively. In addition to the profiles based on our toy topology (red), we also show the ones from the realistic reionization topology (blue; see also Figure 3.4), aggregated into the same parameter bins as introduced in Section 3.3.3.[15] For clarity, we omit profiles in noise-dominated parameter bins with no more than 30 sightlines.

[15] Note that all toy profiles are simulated at a unique set of parameter values for each given bin, whereas the semi-numerical ones are aggregated from the entire parameter range covered by this bin.

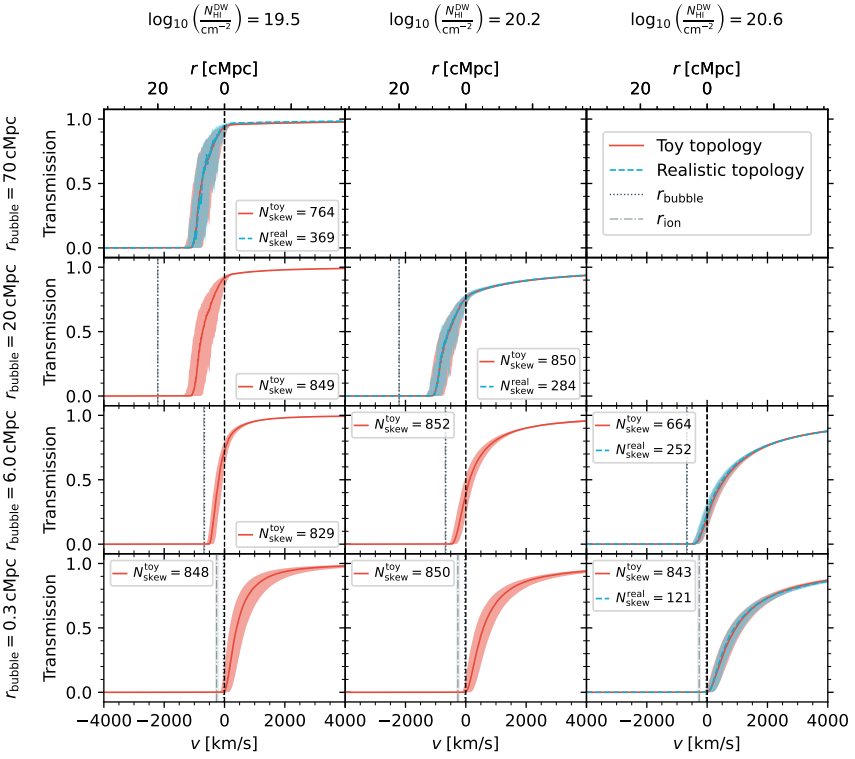

Figure 3.8: Median and 68-percentile scatter of simulated IGM transmission profiles in the local $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ parameterization at a quasar lifetime of $t_{\rm Q} = 10^4\,$yr, comparing profiles based on semi-numerical $x_{\rm HI}$ skewers (blue; see Section 3.3.1), and analytical ones generated according to our toy prescription (red; see Section 3.4.1). The profiles are shown on a representative grid of $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ parameter values. For clarity, bins which are occupied by no more than 30 sightlines are omitted.

3.4.2.1 Physically accessible regions of parameter space

The first thing of note is that our toy profiles extends over a significantly larger region in $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ parameter space as compared to the realistic ones, as we can immediately see by comparing the red and cyan contours depicted in Figure 3.6. Both these models show a non-trivial dependence on the parameter values due to the fact that not all regions of parameter space are accessible given the distribution of density fluctuations in the IGM. However, the toy prescription explores significantly larger regions in parameter space than the more realistic semi-numerical model. This shows that there are parameter configurations which are theoretically possible

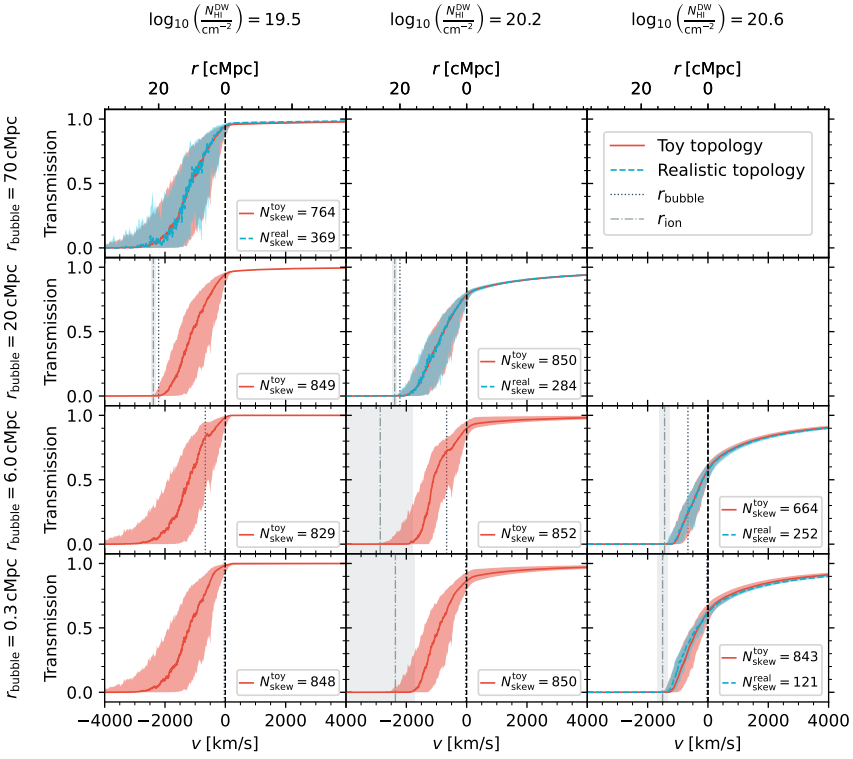

Figure 3.9: Like Figure 3.8, but for a quasar lifetime of $t_{\rm Q} = 10^6$ yr.

for a given sightline but which do not necessarily appear in our simulated reionization topology.

This mainly concerns combinations of small HI column densities $N_{\rm HI}^{\rm DW}$ with short neutral patch distances $r_{\rm patch}$. Such configurations are increasingly hard to achieve the smaller $r_{\rm patch}$, as neutral material closer to the source gets upweighted according to the Lorentzian weighting function. A very nearby neutral patch therefore has to be extremely small in extent as to not exceed a given HI column density $N_{\rm HI}^{\rm DW}$—a configuration that hardly occurs in the realistic reionization topology. The relatively tight correlation between $N_{\rm HI}^{\rm DW}$ and $r_{\rm patch}$ seen for the cyan contours in Figure 3.6 is therefore specific to this topology, and not necessarily a generic feature of other ones.

We demonstrate this fact by also depicting the parameter distribution that would follow if quasars were not to be placed in the most massive halos but in random environments (dark blue contours). For each of our 21

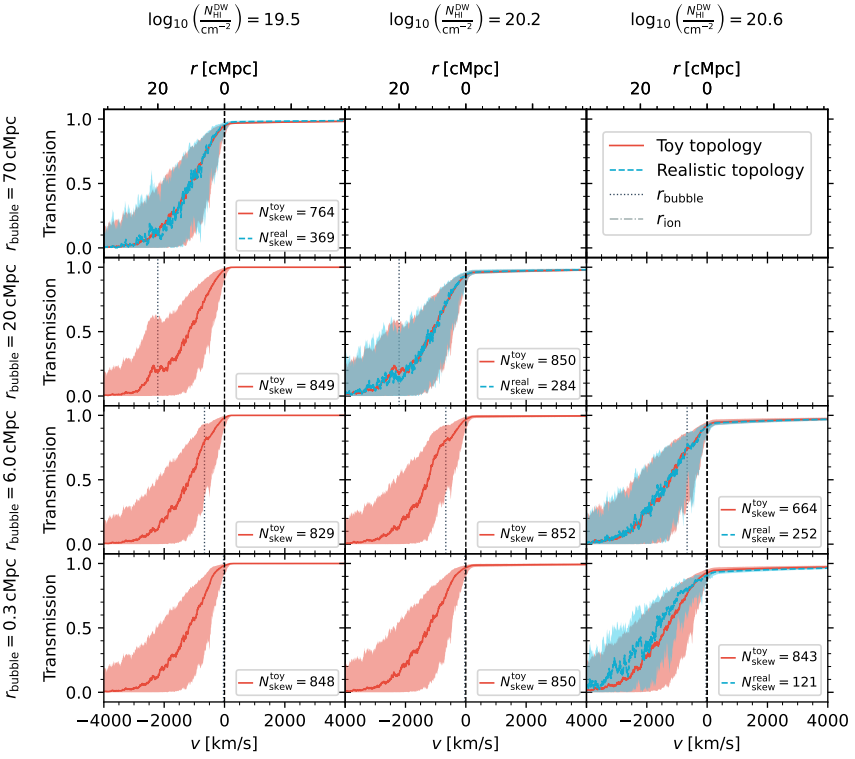

Figure 3.10: Like Figure 3.8, but for a quasar lifetime of $t_{\rm Q} = 10^8$ yr.

global topologies, we generate 1200 such sightlines[16] (770 of which remain after excluding sightlines with strong proximate absorbers according to the same criterion as adopted previously), and we analogously compute the values of our local summary statistics for each sightline. The resulting distribution shows clear differences to the one obtained under the assumption that quasars only populate massive halos (cyan). In particular, at short neutral patch distance $r_{\rm patch}$, the distribution shows an extended tail towards low column densities $N_{\rm HI}^{\rm DW}$. Such configurations are physically admitted (see red contours obtained with our toy prescription) but unlikely if quasars were to reside in the most massive halos which, assuming an inside-out reionization scenario, constitute the first regions to be ionized. As a result, largely ionized sightlines (i.e., sightlines where $N_{\rm HI}^{\rm DW}$ is low) preferentially go

[16] In practice, we take the same 1200 sightlines as before and instead place the source in the center rather than at the origin of the sightline. This suffices for a first qualitative impression as an overdensity ~ 70 cMpc away from the quasar does not significantly impact the damping wing signature anymore.

together with a large distance r_{patch} to the first neutral patch, whereas such a correlation is not necessarily required when quasars are placed in random locations. While this already qualitatively demonstrates the added value of our parameterization for distinguishing between different reionization scenarios and topologies, we leave a more thorough exploration of these application of our parameterization to future work.

In any case, as seen in the three lower panels of Figure 3.7, the IGM transmission scatter as quantified by the 68-percentile widths $\sigma(t(v = v_{\mathrm{T}}))$ remains similarly low even in those regions of parameter space which remain unexplored by the realistic topology with quasars living in the most massive halos. Our parameterization therefore tightly captures the shape of the IGM damping wing in *all* regions of parameter space which are physically allowed.

The unpopulated regions of parameter space at high column densities with large neutral patch distances, on the other hand, are a hard physical constraint, independent of reionization topology. As argued in Section 3.4.1, the maximum column density $N_{\mathrm{HI}}^{\mathrm{DW}}$ for a given sightline at a given neutral patch distance r_{patch} can only be achieved if the skewer is fully neutral starting at $r \geq r_{\mathrm{patch}}$ (and, by definition, fully ionized at $r < r_{\mathrm{patch}}$). This maximum value therefore decreases with increasing r_{patch} for a given sightline. Parameter combinations $(N_{\mathrm{HI}}^{\mathrm{DW}}, r_{\mathrm{patch}})$ where we cannot find *any* sightline that reaches the corresponding column density are therefore *physically excluded* since even the strongest density fluctuations do not suffice anymore to achieve this HI column density $N_{\mathrm{HI}}^{\mathrm{DW}}$. Again, this statement is solely based on the underlying distribution of density fluctuations and independent of the topology of reionization.

3.4.2.2 Agreement between realistic reionization model and toy prescription

Moving on to the comparison of the IGM transmission profiles between the realistic and the toy topology, we observe a remarkable agreement throughout nearly all bins that are populated by both models, across the entire lifetime range, as can be seen in the panels of Figures 3.8, 3.9 and 3.10 where profiles from both models are available. The only bins where differences become visible are the bottom right ones in each figure where $\log_{10} N_{\mathrm{HI}}^{\mathrm{DW}}/\mathrm{cm}^{-2} = 20.6$ and $r_{\mathrm{patch}} = 0.3\,\mathrm{cMpc}$.[17] While both models still perfectly agree at the lowest lifetime of $t_{\mathrm{Q}} = 10^4\,\mathrm{yr}$ (see Figure 3.8), the profiles increasingly differ at longer lifetimes of $t_{\mathrm{Q}} = 10^6$ and $10^8\,\mathrm{yr}$ as seen in Figures 3.9 and 3.10.

[17] Due to the lack of a star formation prescription, the Nyx simulations do not correctly capture the circumgalactic medium of massive halos, and therefore we exclude the first 0.3 cMpc of all sightlines. Hence $r_{\mathrm{patch}} = 0.3\,\mathrm{cMpc}$ is the shortest neutral patch distance possible in our simulations. Note that this does not imply our simulated transmission profiles are unphysical, as material this close to the quasar is certainly guaranteed to be ionized away, even for the shortest quasar lifetimes.

More specifically, we see that while the profiles still agree reasonably well in the damping wing region of the spectrum, the proximity zone transmission of the realistic profiles is systematically higher than that of the toy ones. The reason for this is related to the fact that in the realistic semi-numerical reionization topology, $r_{\rm patch}$ values of 0.3 cMpc are in practice only found for completely neutral sightlines, since a small $r_{\rm patch}$ value always implies a large column density $N_{\rm HI}^{\rm DW}$ as can be see in Figure 3.6. On the other hand, our toy model also allows for many sightlines with a first neutral patch at $r_{\rm patch} = 0.3$ cMpc and a number of ionized patches farther away from the source, leading to lower column densities, and hence the contours extending out to much smaller $N_{\rm HI}^{\rm DW}$. These configurations are extremely rare in realistic reionization topologies as the regions around the most massive halos where our quasars are placed are among the first ones to be reionized, whereas no such constraint applies to our toy toy prescription where we only place the first neutral patch at the desired location and subsequently add more neutral patches at *any* location behind it until a desired column density is reached.

Having available this broader range of $x_{\rm HI}$ skewers at $r_{\rm patch} = 0.3$ cMpc means that we can achieve a given HI column density $N_{\rm HI}^{\rm DW}$ for many additional density sightlines. These additional sightlines are generically not as underdense as the ones that achieve a given $N_{\rm HI}^{\rm DW}$ value in combination with a completely neutral sightline. This is because *both* $x_{\rm HI}$ and Δ contribute to $N_{\rm HI}^{\rm DW}$ (see Eq. (3.6)), and if $x_{\rm HI}$ is equal to unity, Δ must be comparatively low. Vice versa, the additional, globally more ionized, skewers from our toy topology can be combined with sightlines less underdense than those required for completely neutral skewers to achieve the same column density $N_{\rm HI}^{\rm DW}$. The more typical densities present in these additional sightlines in return lead to less pronounced transmission spikes in the proximity zone region which become increasingly apparent the larger the proximity zone sizes, i.e., the longer the quasar lifetimes, as the progression of the bottom right panels of Figures 3.8, 3.9 and 3.10 shows. The damping wing region on the other hand remains largely unaffected; if any, we observe a slight increase in damping wing absorption strength for the realistic sightlines as they are globally the most neutral ones. However, our $N_{\rm HI}^{\rm DW}$ statistic *by construction* already accounts for the competing effects between $x_{\rm HI}$ and Δ with the correct Lorentzian weighting, and hence the damping wing strength still agrees well among both models. The remaining small differences among the two models can be attributed to differences in the neutral hydrogen content in front of our fixed lower integration limit of $r_{\rm min} = 4$ cMpc.

While Figures 3.8, 3.9 and 3.10 only show the scatter of the IGM transmission profiles in select parameter bins, we verified that the effect described above only becomes relevant in parameter bins with $r_{\rm patch} = 0.3$ cMpc, while both models perfectly agree across the entire rest of $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ parameter space (modulo a somewhat increased statistical noise level for the realistic

topology in bins which are more scarcely populated). The prior volume where our toy prescription introduces a small bias therefore remains minimal, and this bias certainly remains subdominant to other major sources of uncertainty in realistic inference settings where e.g. the intrinsic continuum of the quasar has to be reconstructed as well (Kist et al. 2025b). Correspondingly, the 68-percentile widths $\sigma(t(v = v_{\mathrm{T}}))$ of the IGM transmission distributions coincide remarkably well between the two topologies in all regions of parameter space, as can be seen by comparing the middle to the lower panels of Figure 3.7.

Overall, this leads us to the striking conclusion that the statistical properties (median and 68-percentile scatter) of the IGM transmission profiles are largely insensitive to the peculiarities of *any* given reionization topology. Again, we compared transmission profiles from two entirely different models—the latter of which is not even physically motivated—and yet, no differences are recognizable across the largest parts of parameter space. Even the remaining differences at $r_{\mathrm{patch}} = 0.3\,\mathrm{cMpc}$ are a mere relic of the unphysical fact that our toy prescription does not account for regions around the most massive halos reionizing early. In future applications, we could alleviate this effect by modifying our toy model to respect other spatial correlations between x_{HI} and Δ such as their average radial profile in addition to $N_{\mathrm{HI}}^{\mathrm{DW}}$ and r_{patch} instead of randomly laying down the neutral patches. Our statement about topology independence of our new set summary statistics therefore remains unconditionally true when comparing profiles from any two realistic reionization topologies where this property is accounted for.

3.4.2.3 Parameter dependence of the IGM transmission profiles at short quasar lifetimes

After these general considerations, we now investigate how the shape of the IGM transmission profiles changes as a function of $N_{\mathrm{HI}}^{\mathrm{DW}}$ and r_{patch}. We start by considering the short lifetime ($t_{\mathrm{Q}} = 10^4\,\mathrm{yr}$) profiles depicted in Figure 3.8. As expected, higher HI column densities generally cause stronger IGM damping wing imprints. Also our second summary statistic r_{patch} has a clear impact on the shape of the IGM transmission profiles, even though it is defined as the distance to the first neutral patch in the *pre*-quasar topology. The first obvious reason for this is the fact that $N_{\mathrm{HI}}^{\mathrm{DW}}$ is defined to be insensitive to any kind of structure in front of our lower integration limit $r_{\mathrm{min}} = 4\,\mathrm{cMpc}$. For quasars whose first neutral patch is located at $r_{\mathrm{patch}} < r_{\mathrm{min}}$, and whose lifetime is not long enough to ionize away this neutral patch, r_{patch} is the only label that can carry information about the remaining nearby neutral material which $N_{\mathrm{HI}}^{\mathrm{DW}}$ by definition cannot capture.

However, we see in Figure 3.8 that the IGM transmission profiles still carry information about r_{patch} even if $r_{\mathrm{patch}} > r_{\mathrm{min}}$. This is because on average, a short-lived quasar which has only been shining for a few thousands of years

has not yet had sufficient time to carve out an ionized region extending (significantly) beyond the start of the first neutral patch (regardless how its original location $r_{\rm patch}$ relates to $r_{\rm min}$), and so $r_{\rm patch}$ constitutes a hard upper limit for the location of its ionization front $r_{\rm ion}$ (marked with a vertical dash-dotted line in each panel).[18] The transmission therefore certainly drops to zero at or prior to $r_{\rm patch}$ (dotted vertical lines), and hence $r_{\rm patch}$ carries additional information about the shape of the profile in this region, clearly supplementing the information about the red-side damping wing shape which is primarily encoded in the $N_{\rm HI}^{\rm DW}$ statistic. This can be seen by comparing the panels within a given row (i.e., at fixed $N_{\rm HI}^{\rm DW}$) in Figure 3.8. In these regions of parameter space where the transmission drops to zero close to $r_{\rm patch}$, this label therefore encodes similar information as the velocity realignment point proposed in Chen (2024) and Keating et al. (2024b) (when considered at a fixed quasar lifetime $t_{\rm Q}$).

On the other hand, no differences are recognizable between the profiles in the two upper rows of Figure 3.8 where all transmission is extinguished at a parameter-independent location significantly closer to the source than the position $r_{\rm patch}$ of the first neutral patch. As the IGM is fully ionized up to the latter point, the Gunn-Peterson optical depth $\tau_{\rm GP}$ in this region is inversely proportional to the photoionization rate $\Gamma_{\rm QSO}$ of the quasar, i.e., $\tau_{\rm GP} \sim 1/\Gamma_{\rm QSO}$. Since the quasar's photoionization rate decreases as the inverse square of the distance r from the quasar, $\Gamma_{\rm QSO} \sim 1/r^2$, all transmission eventually gets suppressed beyond a certain distance, even if the quasar ionization front has already passed through. As a result, profiles with $r_{\rm patch}$ values exceeding this distance are entirely degenerate. This degeneracy would be ameliorated if we were to consider a brighter quasar whose $\Gamma_{\rm QSO}$ would be higher and hence this distance would be located farther outwards from the source.

3.4.2.4 Parameter dependence of the IGM transmission profiles at intermediate to long quasar lifetimes

The sensitivity to $r_{\rm patch}$ decreases as we move to older objects with lifetimes of $t_{\rm Q} = 10^6$ or 10^8 yr such as depicted in Figures 3.9 and 3.10. Such quasars have had a sufficient amount of time to carve out a large ionized region around themselves, in most cases exceeding the first pre-quasar neutral patch location and hence, the transmission does not sharply drop to zero at $r_{\rm patch}$. Instead, $r_{\rm patch}$ is located somewhere within the proximity zone of most such quasars, and as such is hardly reconstructable from a given IGM transmission

[18] The location of the ionization front $r_{\rm ion}$ is measured empirically based on all toy sightlines which are available in a given parameter bin. To that end, we take the same algorithm used to determine $r_{\rm patch}$ for the pre-quasar topology, and apply it to the post-quasar topology instead. The dash-dotted line marks the median $r_{\rm ion}$ of all toy profiles in a given panel, and the grey-shaded region the 68-percentile scatter.

profile. Note that in the two panels of Figure 3.9 where $r_{\mathrm{patch}} = 20\,\mathrm{cMpc}$ (and $t_{\mathrm{Q}} = 10^6\,\mathrm{yr}$), we still do observe a similar drop-off close to r_{patch} as we did for shorter neutral patch distances at $t_{\mathrm{Q}} = 10^4\,\mathrm{yr}$. This shows that at a lifetime of $t_{\mathrm{Q}} = 10^6\,\mathrm{yr}$, a neutral patch located at $r_{\mathrm{patch}} = 20\,\mathrm{cMpc}$ still constitutes a comparably hard limit for the quasar ionization front which comes to a halt shortly thereafter, whereas closer neutral patches easily get ionized away such that r_{ion} significantly exceeds r_{patch}.

Remarkably, however, even in the case where the first pre-quasar neutral patch is ionized away, we statistically still do observe a remnant signature of r_{patch} in form of an excess transmission bump starting at exactly this location. This feature can most clearly be identified in the central panels of Figures 3.9 and 3.10 where both median and scatter are elevated starting precisely at r_{patch}. Most notably, a noisier version of this bump can even be identified for the semi-numerical IGM transmission profiles in the $(\log_{10} N_{\mathrm{HI}}^{\mathrm{DW}}/\mathrm{cm}^{-2}, r_{\mathrm{patch}}/\mathrm{cMpc}) = (20.2, 20)$ panel of Figure 3.10. We attribute this feature to photoelectric heating of the (neutral) IGM by the hard quasar spectrum.

The mechanism can be understood by returning to the example sightline depicted in Figure 3.1: comparing the temperature fields before and after the quasar has turned on, we see that the initially neutral regions get heated significantly (bottom row) after getting ionized away by the quasar. This is because at large optical depth, the soft photons all get absorbed, and the optical depth of the harder photons is also high so they get absorbed as well, causing photoelectric heating up to $40\,000\,\mathrm{K}$ (c.f. also Figure 6 of Davies & Hennawi (2023)). As a result, the transmission is enhanced in these regions. Since this significant amount of heating is restricted to initially neutral patches which recently got ionized by the quasar, the *pre*-quasar x_{HI} topology is still encoded in the *post*-quasar temperature field, and hence also leaves an imprint on the IGM transmission field. For an individual sightline, however, this signature is hard to disentangle from the stochasticity of the Lyman-α absorption features in the proximity zone of the quasar due to density fluctuations, even more so when the continuum is unknown, and in the presence of observational noise. Yet, our $(N_{\mathrm{HI}}^{\mathrm{DW}}, r_{\mathrm{patch}})$ parameterization allows us to statistically account for this feature when performing astrophysical parameter inference. We will explore this possibility in future work.

3.5 Conclusions

We introduced in this work a novel three-parameter model that tightly captures the characteristic shape of quasar IGM damping wings. As an alternative to the common parameterization based on the global IGM neutral fraction $\langle x_{\mathrm{HI}} \rangle$, we defined two new *local* summary statistics quantifying the

neutral hydrogen content along the sightline from the quasar *before* it started shining: 1) the Lorentzian-weighted HI column density $N_{\rm HI}^{\rm DW}$, and 2) the distance $r_{\rm patch}$ between the quasar and the first neutral patch. We supplemented these two local measures of the *pre*-quasar neutral topology with the quasar lifetime $t_{\rm Q}$ as a third parameter encapsulating the effects of the ionizing quasar radiation.

Since quasars are typically going to ionize away all surrounding neutral gas within the first few cMpc, we found that the damping wing is most sensitive to the Lorentzian-weighted HI column density $N_{\rm HI}^{\rm DW}$ starting at 4 cMpc from the source, and integrated over a range of 100 cMpc. By adding a Lorentzian weighting function (associated to a fixed reference distance $r_{\rm T} = 18\,$cMpc) to the column density integral, we account for the effect of the Lyman-α cross section σ_α in the optical depth integral, assuring a maximally tight parameterization of the IGM damping wing at the spectral pixel corresponding to $r_{\rm T}$. By demonstrating that the damping wing essentially constitutes a one-parameter family, we showed that $N_{\rm HI}^{\rm DW}$ tightly parameterizes the damping wing imprint not only at this specific location but across the *entire* spectral range.

We introduced as a second summary statistic the distance $r_{\rm patch}$ from the source to the first neutral patch in the *pre*-quasar topology, and showed that it still encodes information about *post*-quasar IGM transmission profiles, and that this information is complementary to $N_{\rm HI}^{\rm DW}$. We related the definitions of our summary statistics back to the labels recently proposed in Chen (2024), Keating et al. (2024b), and Mason et al. (2026), and argued that by starting the $N_{\rm HI}^{\rm DW}$ integration directly at the location of the source, our parameterization is also applicable in the context of IGM damping wings towards galaxies.

We simulated realistic IGM transmission profiles and compared their scatter in the damping wing region of the spectrum at fixed parameter values within both the global and the local parameterization. We found that due to the stochastic distribution of neutral patches during reionization and density fluctuations in the IGM, the 68-percentile scatter of the IGM transmission values at $v_{\rm T} = 2000\,$km/s in the global parameterization can be as large as 7 % (at a quasar lifetime of $t_{\rm Q} = 10^4\,$yr), and we demonstrated that this scatter decreases down to $\lesssim 1\,\%$ across the entire range of physical parameter space when aggregating the same sightlines according to our local parameterization.

We introduced a simple numerical prescription to generate synthetic HI density profiles at any desired location in $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$ parameter space, varying smoothly as a function of both these parameters. Even though the procedure does not attempt to describe any physical processes, it allowed us to demonstrate the robustness of our parameterization against the choice of reionization model since it results in a significantly different neutral topology. We observed an exceptional agreement between the realistic and

the toy profiles at almost every location in physical parameter space. Small differences between the models only became apparent for profiles with a neutral patch directly next to the quasar, which we could trace back to the fact that our toy prescription does not pay tribute to the fact that reionization takes place inside out. As this would be accounted for by any realistic reionization model, the remarkable overall agreement between the two models demonstrates the topology-insensitivity of our parameterization.

In addition, this agreement legitimizes the future use of our toy prescription for the purpose of astrophysical parameter inference, decoupling 1) the topology-insensitive task of inferring the local damping wing statistics introduced in this work, and 2) tying these constraints to a specific reionization model, resulting in near-optimal constraints not only on the global timing of reionization, but also the reionization topology, hitherto unconstrained with quasar IGM damping wings. In this context, all assumptions about the reionization model can be encoded in a prior on $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$, determined by the distribution of these parameters within the model of interest. This clear separation will facilitate future inference endeavors, and even the comparison of different reionization models, harnessing *all* the information encapsulated in quasar IGM damping wings.

Acknowledgements

We acknowledge helpful conversations with the ENIGMA group at UC Santa Barbara and Leiden University and would especially like to thank Shane Bechtel and Benjamin Snyder for comments on an early version of this manuscript. This work made use of `NumPy` (Harris et al. 2020), `SciPy` (Virtanen et al. 2020), `Astropy` (Astropy Collaboration et al. 2013, 2018, 2022), `h5py` (Collette 2013), `Matplotlib` (Hunter 2007), and `IPython` (Pérez & Granger 2007). TK and JFH acknowledge support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 885301). JFH acknowledges support from NSF grant No. 2307180.

Data Availability

The derived data generated in this research will be shared on reasonable requests to the corresponding author.

Appendices

3.A On the relation between τ_{DW} and $N_{\mathrm{HI}}^{\mathrm{DW}}$

We introduced in Section 3.2 the Lorentzian-weighted HI column density $N_{\mathrm{HI}}^{\mathrm{DW}}$ as a summary statistic of the local pre-quasar neutral hydrogen content along the line of sight from the source. By comparing $N_{\mathrm{HI}}^{\mathrm{DW}}$ as defined in Eq. (3.6) to the pre-quasar damping wing optical depth

$$\tau_{\mathrm{DW}}^{\mathrm{pre}}(\lambda_{\mathrm{rest}}) = \int_0^{R(z_{\mathrm{QSO}})} n_{\mathrm{HI}}^{\mathrm{pre}}(R) \cdot \sigma_\alpha \left(\frac{1+z_{\mathrm{QSO}}}{1+z(R)} \lambda_{\mathrm{rest}} \right) \mathrm{d}R, \tag{3.11}$$

we motivate in this section why this summary statistic is a near-optimal means to reduce the IGM transmission scatter redward of the Lyman-α line. Specifically, we show that in the limit where the Lorentzian-weighted average $\langle\!\langle x_{\mathrm{HI}} \cdot \Delta \rangle\!\rangle_{\mathrm{Lor}}$ as defined in Eq. (3.4) is computed along the *entire* line of sight, and where the Lyman-α cross section is approximated as perfectly Lorentzian, Eq. (3.7) holds, or, in other words, $N_{\mathrm{HI}}^{\mathrm{DW}}$ encodes the same information as the pre-quasar damping wing optical depth $\tau_{\mathrm{DW}}^{\mathrm{pre}}(v = v_{\mathrm{T}})$ evaluated at the reference velocity offset v_{T}.

We start by writing the integrand in Eq. (3.11) explicitly in terms of the integration variable R. This is required since we need to evaluate the Lyman-α cross section σ_α at wavelength

$$\lambda = \frac{1+z_{\mathrm{QSO}}}{1+z(R)} \lambda_{\mathrm{rest}}. \tag{3.12}$$

The relation $z(R)$ between redshift z and proper distance R from the quasar can be obtained by inverting the light-travel distance relation

$$R(z) = \int_z^{z_{\mathrm{QSO}}} \frac{c}{(1+z')H(z')} \mathrm{d}z'. \tag{3.13}$$

Certainly, the endpoint of reionization z_{end} puts an end to all contributions to the damping wing optical depth as per Eq. (3.11).[19] Since the integrand $c/(1+z')H(z')$ does not change significantly from z_{QSO} to z_{end}, we can expand Eq. (3.13) to linear order around $z = z_{\mathrm{QSO}}$, finding

$$R(z) \simeq \frac{c}{(1+z_{\mathrm{QSO}})\, H(z_{\mathrm{QSO}})} (z_{\mathrm{QSO}} - z). \tag{3.14}$$

This relation can now easily be inverted, and we obtain

$$1 + z(R) = (1+z_{\mathrm{QSO}}) \left(1 - \frac{H(z_{\mathrm{QSO}})}{c} R \right) \tag{3.15}$$

[19] In fact, Section 3.2.2 led us to the even stronger conclusion that the damping wing imprint is only sensitive to the structure within the first ~ 100 cMpc from the source.

If we now evaluate the Lyman-α rest-frame wavelength $\lambda_{\rm rest} = \lambda_\alpha\,(1 + \frac{v}{c})$ at the velocity offset $v = v_{\rm T}$, recalling that $R_{\rm T} \equiv +v_{\rm T}/H(z_{\rm QSO})$, and using Eq. (3.15), we immediately find that Eq. (3.12) turns into

$$\lambda = \lambda_\alpha \, \frac{c + H(z_{\rm QSO})\, R_{\rm T}}{c - H(z_{\rm QSO})\, R}. \tag{3.16}$$

With an expression at hand for $\lambda(R)$, we now proceed by substituting this relation into the following expression for the scattering cross section σ_α of the Lyman-α line:

$$\sigma_\alpha(\lambda) = \frac{\pi e^2}{m_e c}\, f_\alpha\, \phi_\alpha(\lambda). \tag{3.17}$$

Here, $f_\alpha \simeq 0.416$ is the Lyman-α oscillator strength, and $\phi_\alpha(\lambda)$ is the line profile function, commonly assumed as a Voigt profile. Away from the Lyman-α resonance itself, we can assume to good approximation that $\phi_\alpha(\lambda)$ is of Lorentzian shape (Draine 2011; Bach & Lee 2015), i.e.,

$$\phi_\alpha(\lambda) = \frac{\gamma_\alpha\, \lambda_\alpha/(\pi c)}{(\lambda_\alpha/\lambda - 1)^2 + \gamma_\alpha^2}, \tag{3.18}$$

where $\gamma_\alpha \equiv \Gamma_\alpha\, \lambda_\alpha/(4\pi c)$ with the decay constant $\Gamma_\alpha = 6.265 \times 10^8\,{\rm s}^{-1}$ of the Lyman-α transition. Evaluating $\lambda_{\rm rest}$ at the velocity offset $v_{\rm T}$ assures that λ (as given by Eq. (3.12)) is sufficiently far away from the Lyman-α resonance λ_α such that $\phi_\alpha(\lambda)$ is not only to good approximation of Lorentzian shape, but we can also safely neglect the second term in the denominator of Eq. (3.18). Utilizing Eq. (3.16), we can explicitly rewrite Eq. (3.17) for the Lyman-α scattering cross section as a function of R:

$$\sigma_\alpha(\lambda(R)) \simeq \frac{e^2}{m_e c^2}\, f_\alpha\, \gamma_\alpha\, \lambda_\alpha\, \frac{(c/H(z_{\rm QSO}) - R_{\rm T})^2}{(R + R_{\rm T})^2}. \tag{3.19}$$

Substituting this back into Eq. (3.11) shows that our weighting function (Eq. (3.3)) exactly mimics the scaling in the integrand of the optical depth integral. If we now let $R_{\rm min} \to 0$ and $R_{\rm max} \to R(z_{\rm QSO})$, we formally arrive at Eq. (3.7), showing that in these limits, $N_{\rm HI}^{\rm DW}$ is an optimal summary of the pre-quasar damping wing optical depth $\tau_{\rm DW}^{\rm pre}$. By instead choosing the integration limits in the way discussed in Section 3.2.2, we can ensure that an analogous version of Eq. (3.7) even holds for the *post*-quasar optical depth $\tau_{\rm DW}$.

3.B On the hyperparameter dependence of $N_{\rm HI}^{\rm DW}$

The Lorentzian-weighted HI column density $N_{\rm HI}^{\rm DW}$ defined in Eq. (3.6) comes with three hyperparameters, the integration limits $r_{\rm min}$ and $r_{\rm max}$ as well as

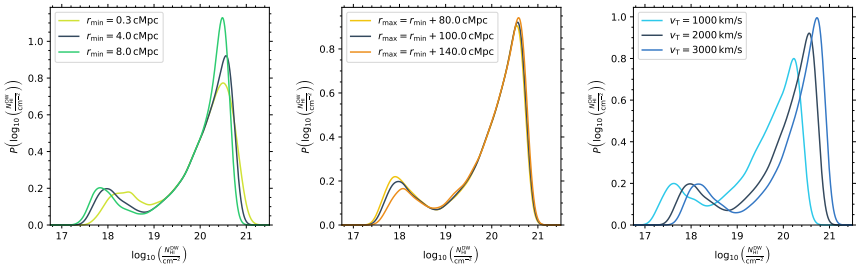

Figure 3.11: Distribution of the Lorentzian-weighted HI column density $N_{\mathrm{HI}}^{\mathrm{DW}}$ for different hyperparameter choices among all 21×852 simulated IGM transmission profiles based on semi-numerical x_{HI} skewers. Variations with respect to the lower integration limit r_{min} are shown in the left panel, the upper integration limit r_{max} in the middle, and the reference velocity offset v_{T} in the right panel. The distribution resulting from our reference choices of $r_{\mathrm{min}} = 4.0\,\mathrm{cMpc}$, $r_{\mathrm{max}} = r_{\mathrm{min}} + 100.0\,\mathrm{cMpc}$ and $v_{\mathrm{T}} = 2000\,\mathrm{km/s}$ is shown as black curve in each panel. All distributions are smoothed via a kernel density estimation (KDE) with a Gaussian kernel of size $\Delta(\log_{10} N_{\mathrm{HI}}^{\mathrm{DW}}/\mathrm{cm}^{-2}) = 0.1$.

the reference distance r_{T} (or, equivalently, velocity offset v_{T}) whose values we fix as discussed in Sections 3.2.2 and 3.2.3. We investigate in this appendix the effects of varying these three parameters.

Figure 3.11 depicts how the distribution of our $N_{\mathrm{HI}}^{\mathrm{DW}}$ statistic among all 21×852 simulated sightlines changes as a function of these three parameters. We start by studying the effects of the lower integration limit r_{min} depicted in the left-hand panel. As discussed in Section 3.2.2, this is the most sensitive hyperparameter as it determines how much nearby pre-quasar structure is included in the column density integral. This nearby structure carries the highest Lorentzian weight but also tends to have been ionized away by a quasar with a typical lifetime, and as such, our choice of $r_{\mathrm{min}} = 4\,\mathrm{cMpc}$ constitutes a compromise between maximizing the information about the pre-quasar topology that we can still gain while not including too much material that does not actually contribute to the observed damping wing signature anymore. In line with these considerations, we see in Figure 3.11 that the $N_{\mathrm{HI}}^{\mathrm{DW}}$-distribution gets more skewed towards higher values the shorter the integration limit is. In particular, we observe that the high-$N_{\mathrm{HI}}^{\mathrm{DW}}$ peak decreases in size and the distribution instead obtains a more extended tail towards the highest $N_{\mathrm{HI}}^{\mathrm{DW}}$ values.

Figure 3.12 depicts—analogously to Figure 3.7—half the 68-percentile scatter $\sigma(t(v = v_{\mathrm{T}}))$ among all 21×852 profiles when aggregated according to our $(N_{\mathrm{HI}}^{\mathrm{DW}}, r_{\mathrm{patch}})$ parameterization for three different choices of the lower integration limit $r_{\mathrm{min}} = 0.3\,\mathrm{cMpc}$, $4.0\,\mathrm{cMpc}$ and $8.0\,\mathrm{cMpc}$, and three different lifetimes of $t_{\mathrm{Q}} = 10^4\,\mathrm{yr}$, $10^6\,\mathrm{yr}$ and $10^8\,\mathrm{yr}$. Although the differences appear relatively weak, we see that the choice of $r_{\mathrm{min}} = 0.3\,\mathrm{cMpc}$ leads to a slight improvement at the shortest lifetime of $t_{\mathrm{Q}} = 10^4\,\mathrm{yr}$ where a shorter

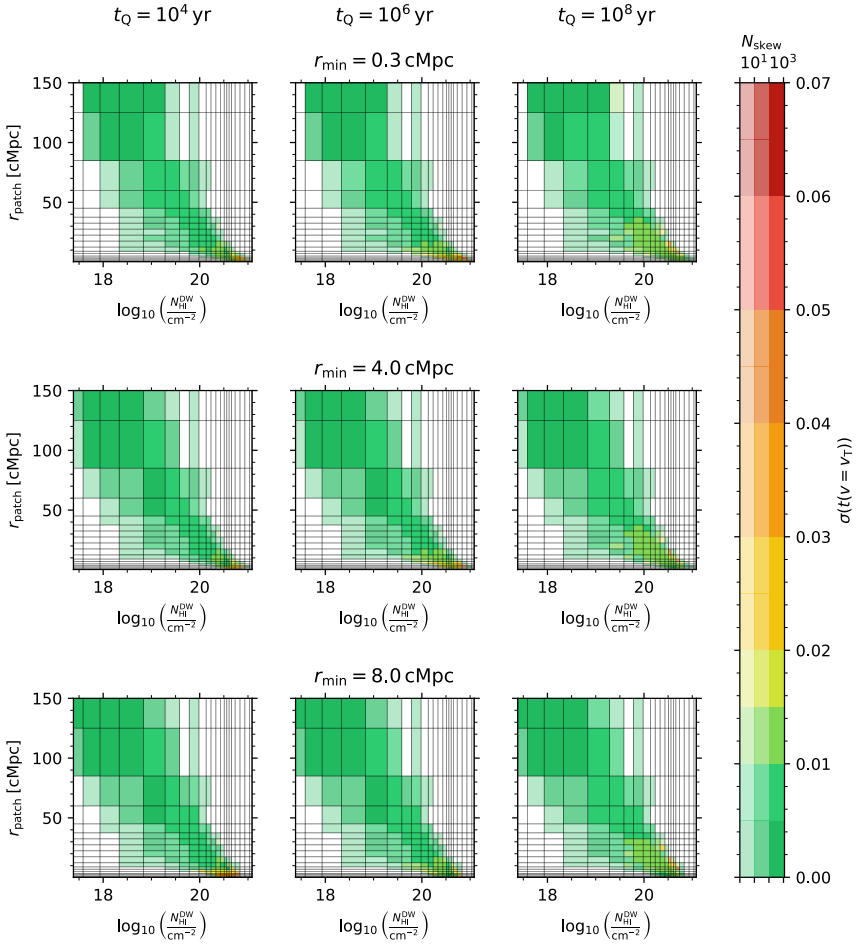

Figure 3.12: Sightline-to-sightline scatter of the IGM damping wing as a function of astrophysical parameter space for different lower integration limits $r_{\rm min}$ for the HI column density $N_{\rm HI}^{\rm DW}$. Analogously to the middle row of Figure 3.7, each panel shows half the 68-percentile width $\sigma(t(v = v_{\rm T}))$ of the distribution of IGM transmission values at $v = v_{\rm T}$ among all sightlines based on semi-numerical $x_{\rm HI}$ skewers as a function of $(N_{\rm HI}^{\rm DW}, r_{\rm patch})$-parameter space, for quasar lifetimes of $t_{\rm Q} = 10^4$ yr (left column), 10^6 yr (middle column) and 10^8 yr (right column), and for integration limits of $r_{\rm min} = 0.3$ cMpc (upper row), 4.0 cMpc (middle row) and 8.0 cMpc (lower row). Bins which are occupied by a lower number of sightlines are more transparent, and white if not occupied at all.

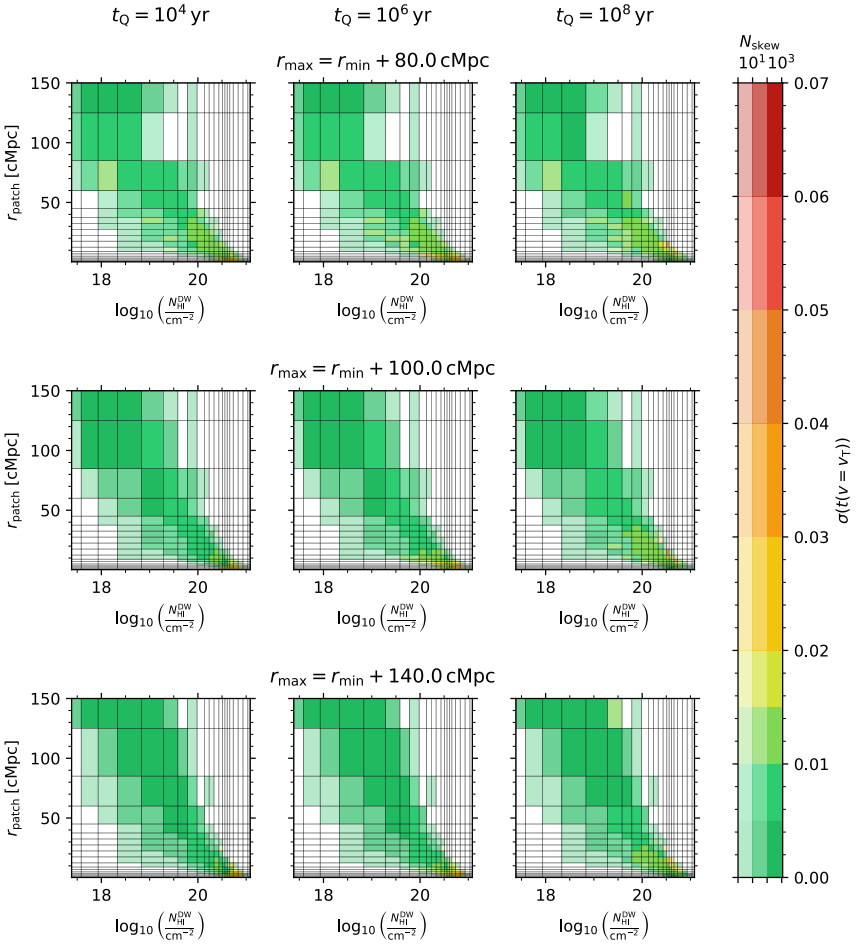

Figure 3.13: Like Figure 3.12 but for upper integration limits of $r_{\mathrm{max}} = r_{\mathrm{min}} + 80.0\,\mathrm{cMpc}$ (upper row), $r_{\mathrm{min}} + 100.0\,\mathrm{cMpc}$ (middle row) and $r_{\mathrm{min}} + 140.0\,\mathrm{cMpc}$ (lower row) for the HI column density $N_{\mathrm{HI}}^{\mathrm{DW}}$.

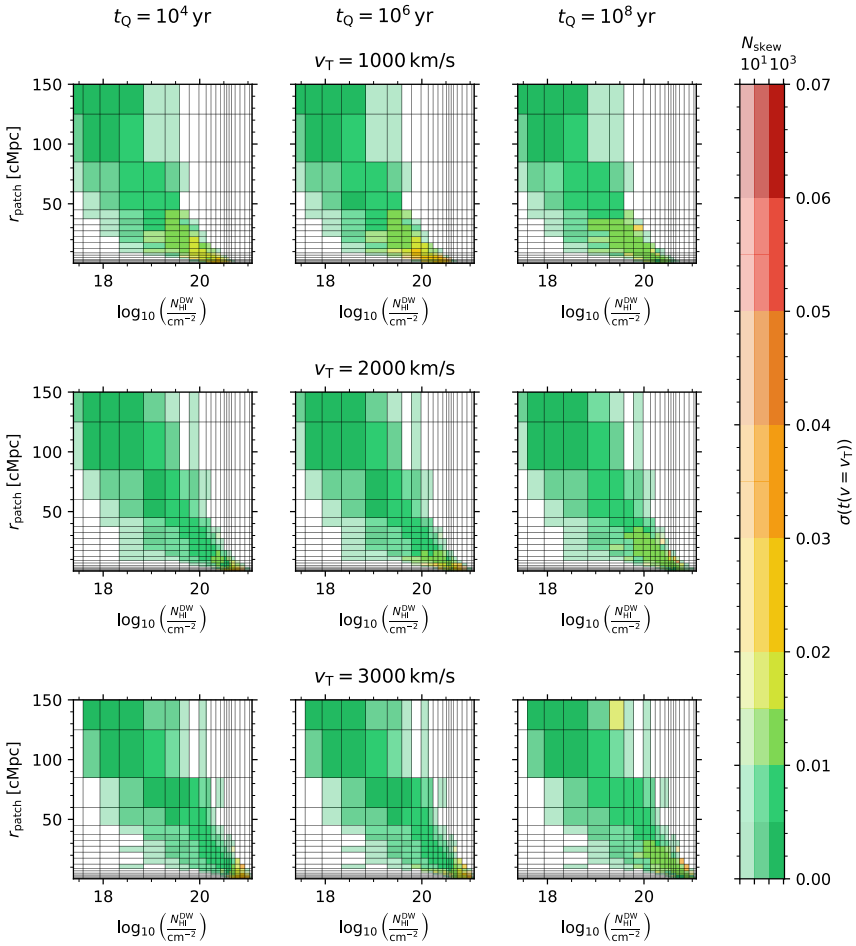

Figure 3.14: Like Figure 3.12 but for reference velocity offsets of v_T = 1000 km/s, 2000 km/s and 3000 km/s for the HI column density N_{HI}^{DW}. Regardless of the reference offset, the scatter is consistently evaluated at v = 2000 km/s.

integration limit would in fact be more appropriate, while we observe an opposite trend at longer quasar lifetimes. Vice versa, a large integration limit of $r_{\mathrm{min}} = 8.0\,\mathrm{cMpc}$ notably increases the scatter at the shorter lifetimes where this means we are clearly excluding too much neutral material from the integration range. These trends empirically confirm the aptness of our lower integration limit at 4.0 cMpc.

Going forward, we see that impact of the upper integration limit r_{max} is significantly weaker since all contributions from distant neutral material are anyways suppressed by the Lorentzian decline of our weighting function. As a result, the column density distribution in the middle panel of Figure 3.11 remains largely unaffected between $r_{\mathrm{max}} = r_{\mathrm{min}} + 80.0\,\mathrm{cMpc}$ and $r_{\mathrm{min}} + 140.0\,\mathrm{cMpc}$ corresponding to the maximum length of our sightlines. Note that we are not sensitive to the differences in the low-$N_{\mathrm{HI}}^{\mathrm{DW}}$ peak as these are the column densities that do not cause any observable damping wing imprint and therefore their exact value has little effect. Correspondingly, Figure 3.13 shows that the scatter $\sigma(t(v = v_{\mathrm{T}}))$ stays remarkably constant regardless of our choice of r_{max} and only starts to increase slightly towards our lowest value of $r_{\mathrm{max}} = 80.0\,\mathrm{cMpc}$.

Finally, the right-hand panel of Figure 3.11 suggests that a change of our reference velocity-offset v_{T} essentially just causes a linear shift in the column density distribution. This is in line with our considerations in Section 3.2.3 where we demonstrated that the damping wing imprint largely constitutes a one-parameter family whose value at any given velocity-offset is determined by the value at a fixed reference offset v_{T}. Correspondingly, the $N_{\mathrm{HI}}^{\mathrm{DW}}$-distribution only gets shifted by the corresponding proportionality constant. When considering the sightline-to-sightline scatter $\sigma(t(v = 2000\,\mathrm{km/s}))$ shown in Figure 3.14 for the three different reference offsets of $v_{\mathrm{T}} = 1000\,\mathrm{km/s}$, $2000\,\mathrm{km/s}$ and $3000\,\mathrm{km/s}$,[20] we see the scatter increases exactly in line with the amount one would deduce from Figure 3.2 based on the Lyman-α transmission itself.

[20] Note that for this analysis we always evaluate the scatter at $v = 2000\,\mathrm{km/s}$, regardless of the value of v_{T}.