The Dark Ages’ 21-cm Trispectrum

Thomas Flöss, a,b Tim de Wild, a P. Daniel Meerburg, a Léon V.E. Koopmans b

a Van Swinderen Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands
b Kapteyn Astronomical Institute, University of Groningen, P.O.Box 800, 9700 AV Groningen, The Netherlands
E-mail: t.s.floss@rug.nl

Abstract. We investigate tomography of 21-cm brightness temperature fluctuations during the Dark Ages as a probe for constraining primordial non-Gaussianity. We expand the 21-cm brightness temperature up to cubic order in perturbation theory and improve previous models of the signal by including the effect of the free electron fraction. Using modified standard perturbation theory methods that include baryonic pressure effects we derive an improved secondary bispectrum and for the first time derive the secondary trispectrum of 21-cm brightness temperature fluctuations. We then forecast the amount of information available from the Dark Ages to constrain primordial non-Gaussianity, including the imprints of massive particle exchange during inflation and we determine how much signal is lost due to secondary non-Gaussianity. We find that although secondary non-Gaussianity swamps the primordial signal, primordial non-Gaussianity can still be extracted with signal-to-noise ratios that surpass current and future CMB experiments by several orders of magnitude, depending on the experimental setup. Furthermore, we conclude that for the bi- and trispectra of massive particle exchange marginalizing over other primordial shapes affects signal-to-noise ratios more severely than secondary shapes. Baryonic pressure effects turn out to have a negligible impact on our forecasts, even at scales close to the Jeans scale. The results of this work reinforce the prospects of 21-cm brightness temperature fluctuations from the Dark Ages as the ultimate probe for primordial non-Gaussianity.
Contents

1 Introduction 2

2 21-cm fluctuations during the Dark Ages 4
   2.1 Global 21-cm signal 4
   2.2 Fluctuations in the 21-cm signal 6
      2.2.1 21-cm brightness temperature 6
      2.2.2 Gas temperature 7
      2.2.3 Free electron fraction 8
   2.3 Perturbative analysis of fluctuations 9
   2.4 Coupling 21-cm fluctuations to density and velocity fluctuations 12

3 Non-Gaussianity of 21-cm brightness temperature anisotropies 15
   3.1 Primordial non-Gaussianity 15
      3.1.1 Primordial bispectra 16
      3.1.2 Primordial trispectra 18
   3.2 Secondary non-Gaussianity 20
      3.2.1 Perturbation theory including baryonic pressure 20
      3.2.2 Secondary bispectrum 22
      3.2.3 Secondary trispectrum 23

4 The information content of the Dark Ages 23
   4.1 Total primordial information content 23
   4.2 Including secondary non-Gaussianity 26
   4.3 Forecasting experiments 31

5 Conclusions and Outlook 35

A Evolution of perturbations 38

B Fisher information matrix 39

C 21-cm secondary trispectrum 43
1 Introduction

Over the last few decades, inflation has become the leading paradigm for describing the early universe. Even the simplest models of inflation accurately predict the nearly Gaussian initial conditions of the universe and match the precise observations of the Cosmic Microwave Background (CMB) by the Planck satellite [1, 2]. Slight deviations from these Gaussian initial conditions of the early universe, colloquially known as primordial non-Gaussianity (pnG), may be used to further constrain the vast space of inflationary theories [3]. Currently, bounds on the size of pnG by CMB and Large Scale Structure (LSS) observations are consistent with a purely Gaussian distribution of fluctuations, and therefore unable to favour a particular theory of inflation. However, upcoming CMB and LSS experiments are forecast to improve on these bounds and might find a statistically significant deviation from Gaussianity.

The size of pnG can be probed through the higher statistical moments of the distribution of initial conditions. For a purely Gaussian distribution all information is then contained in the two-point correlation function of the primordial density field. Generally, the size of pnG is determined by the interactions of the inflaton, the scalar field driving the supposed inflationary expansion. However, even in the absence of such direct interactions, one expects indirect interaction through gravity (i.e. through coupling to the metric). This gravitational interaction sets the minimum amount of pnG that is guaranteed to be present in the initial conditions sourced by inflation and is commonly referred to as the gravitational floor. In the simplest theories such as single-field slow-roll inflation, this is the only source of pnG [4, 5], whereas more complicated models (i.e. inflaton self-interactions or multiple fields) could generate considerably larger amounts of non-Gaussianity. Another powerful probe of inflation is the famous Maldacena consistency condition [4, 6], which predicts the vanishing of the squeezed bispectrum for all single-field models of inflation. Hence a non-zero measurement of the squeezed bispectrum would rule out all single-field inflationary models and point toward multi-field dynamics.

Besides being a powerful probe of the physics driving the inflationary expansion, pnG can also contain signatures of additional physics at play during the inflationary era. Massive (spinning) extra fields could have been present, leaving their imprint on the higher order statistics of the initial conditions [7–9]. More specifically, heavy fields induce a characteristic oscillatory shape in the squeezed and collapsed limit of the three- and four-point correlation function of fluctuations respectively. The amplitude and frequency of this oscillation is directly related to the mass, in principle allowing one to probe the particle spectrum during inflation. Such extra fields are naturally present in string theory in which one might ultimately aim to embed the theory of inflation as well. Hence, we are motivated to study the existence of these extra fields and use inflation as a cosmological particle collider [7]. Although such oscillatory behaviour is a clean probe of the particle spectrum, its amplitude is unfortunately severely suppressed proportional to the mass, making it challenging for CMB and LSS experiments to ever measure such an effect.

Clearly, in order to advance precision cosmology to measure pnG, new probes and experiments are necessary. In this work we will study the use of 21-cm fluctuations during the cosmic Dark Ages as the ultimate probe of pnG. The Dark Ages refer to the epoch between recombination ($z \approx 1100$), when neutral hydrogen is formed and the CMB photons are re-
leased to free stream, and the formation of the first luminous objects (i.e stars) at $z < 30$. Soon after the CMB photons have been released, they redshift out of the visible wavelengths and the universe becomes truly dark. During this time, a neutral hydrogen gas permeates the universe while occasionally scattering with CMB photons, sometimes exciting the hyper-fine state of the electrons in the neutral hydrogen atoms. When the electron relaxes to its ground state, a photon with a wavelength of 21 centimeter is released. Once the hydrogen gas has cooled sufficiently and its temperature is decoupled from the CMB temperature, this 21-cm signal can be observed in absorption or emission to the background of CMB photons. Higher density regions containing more hydrogen have a brighter 21-cm signal, thereby tracing the matter density field which in turn traces the primordial fluctuations seeded by inflation. In this way, the 21-cm signal during $30 \leq z \leq 100$ can be used as a probe of $\text{pnG}$ \cite{10, 11}, containing an amount of information that is estimated to be several magnitudes more than that of the CMB and LSS, making it the ultimate probe of the early universe and possibly opening up the cosmological collider \cite{12}. However, several comments are in order. First, the small amplitude of the 21-cm signal makes it hard to measure even the global (mean) brightness temperature, let alone tiny fluctuations around it, making it a serious challenge from a technological and experimental point of view. Secondly, Earth’s ionosphere is opaque to the red-shifted 21-cm signal emitted at $z > 30$. Hence, probing this era will require an observatory in space or even on the far side of the moon, where also radio frequency interference (RFI) is minimized \cite{13}. Thirdly, although the 21-cm signal is expected to be a rather pristine tracer of the primordial initial conditions compared to the galaxy distributions that are the target of LSS experiments, it is nevertheless affected by non-linearity (i.e. gravity and peculiar velocity). Both the non-linear dependence of the 21-cm brightness temperature on the initial conditions as well as gravitational effects therefore leave their imprint on the distribution of fluctuations, inducing secondary non-Gaussianities that obscure the sought-after primordial non-Gaussianity by several orders of magnitude \cite{11, 14}. A detailed understanding of the 21-cm signal as well as these non-linear effects is warranted in order to reliably and accurately extract the primordial contribution and learn about inflation.

Secondary non-Gaussianity of the 21-cm brightness temperature during the Dark Ages was addressed previously in \cite{14} for the case of the bispectrum, where it was found to contribute significantly to the observed bispectrum of temperature fluctuations, introducing it as a nuisance that should be marginalized over. In this work we will improve on some of the assumptions and simplifications made in \cite{14}, in order to more accurately model the 21-cm signal during the Dark Ages. Furthermore, it was recently found that observational sensitivity to the primordial trispectrum can receive an enhanced scaling with respect to the smallest observable scale $k_{\text{max}}$ (or $\ell_{\text{max}}$ for CMB surveys) \cite{15}. We will see that this also affects primordial trispectra sourced by massive extra fields, possibly making it of prime observational interest to the cosmological collider. Extracting the primordial trispectrum will require an accurate modelling of the secondary trispectrum, which we derive here for the first time. In an effort to motivate experiments targeted at measuring the 21-cm signal from the Dark Ages we will determine the total information content of the Dark Ages that can be used to constrain primordial non-Gaussianity. Finally, we forecast the sensitivity of a simple experimental setup, that might be realized in the future.

This paper is organized as follows. In Section 2 we review the physics of the 21-cm signal during the Dark Ages. Section 3 will cover the non-Gaussian contributions to the statistics of
21-cm fluctuations. Then, in Section 4, we present Fisher forecasts on the amount of information available to constrain non-Gaussianity from the Dark Ages, as well as the sensitivity of more realistic experimental scenarios. We summarize our conclusions and outlook for future research in Section 5.

Conventions
In this work we denote spatial vectors in boldface, \( \mathbf{k} \). The magnitude of a vector is denoted as \( k \equiv |\mathbf{k}| \). Hats denote unit vectors \( \hat{x} \equiv \mathbf{x}/x \). We denote \( k_1 + k_j \equiv k_{ij} \) and \( k_i + k_j \equiv k_{ij} \).

Note that \( k_{ij} \neq |k_{ij}| \). We also define Mandelstam-like variables \( s = k_1 + k_2 \), \( t = k_1 + k_4 \) and \( u = k_1 + k_3 \). For correlation functions we often adopt the primed notation to hide usual prefactors and Dirac delta functions:

\[
\langle \delta(k_1)\delta(k_2)\cdots\delta(k_n) \rangle = (2\pi)^3 \delta_D(k_1 + k_2 + \cdots + k_n) \langle \delta(k_1)\delta(k_2)\cdots\delta(k_n) \rangle'
\]

Momentum integrals are sometimes written in the following condensed notation for readability:

\[
\int_{\mathbf{k}_i} = \int \frac{d^3k_i}{(2\pi)^3}
\]

For numerical computations we consider flat ΛCDM cosmology with parameters from Planck [16], see Table 1.

\[
\begin{align*}
H_0 &= 67.66 & \Omega_b h^2 &= 0.02242 & \sum m_\nu &= 0.06 \\
\Omega_k &= 0 & \Omega_c h^2 &= 0.11993 & \tau &= 0.0561 \\
n_s &= 0.9665 & A_s &= 2.1056 \times 10^{-9} & r &= 0
\end{align*}
\]

Table 1. Best-fit Planck parameters (specifically, Table 2 of Ref. [16] with TT, TE, EE+lowE+lensing+BAO) used in our numerical computations.

2 21-cm fluctuations during the Dark Ages

In this section we review the physics that gives rise to the 21-cm brightness fluctuations during the Dark Ages. First we will restrict the analysis to the background (or global) temperature. Subsequently we will discuss how 21-cm brightness temperature fluctuations arise due to fluctuations in the underlying density fields. In particular, the goal will be to derive how 21-cm fluctuations trace those of the baryonic density and velocity field.

2.1 Global 21-cm signal

We define the ratio of the abundance of Hydrogen in the hyperfine singlet \( (F = 0) \) and triplet state \( (F = 1) \) via the spin temperature \( T_s \) as follows [17]:

\[
\frac{n_1}{n_0} \equiv \frac{g_1}{g_0} e^{-T_s/T_s},
\]

where \( T_s \equiv E_{10}/k_B \) is the energy gap between the two hyperfine states and the degeneracies of the levels are \( g_{0,1} = 1, 3 \). During the Dark Ages, the spin temperature is determined by two processes: collision and radiative transitions between the two states. These processes are
described by the rates $C_{ij}$ and $R_{ij}$, respectively. Specifically, $R_{10}$ encodes spontaneous and stimulated emission whereas $R_{01}$ encodes absorption. The rates $C_{01}$ and $C_{10}$ describe upward and downward transition between the hyperfine states due to collisions between Hydrogen atoms. We are interested in the Dark Ages, before the first luminous objects are formed, and therefore do not take transitions induced by Lyman-$\alpha$ photons into account (known as the Wouthuysen-Field effect) [17].

During the Dark Ages, the steady state approximation is very accurate, since $C_{ij} + R_{ij} \gg H$ at all times [18], and the abundances are related via:

$$n_0(C_{01} + R_{01}) = n_1(C_{10} + R_{10}).$$

Then, in the limit $T_* \ll T_{\text{gas}}, T_\gamma$, also valid at all times of interest [14], we can write the spin temperature as:

$$T_s = T_\gamma + Y_c T_{\text{gas}},$$

where $Y_c \equiv T_0 C_{10} / T_{\text{gas}} A_{10}$ and $A_{10}$ the spontaneous decay rate. From the definition of $Y_c$, we find that atomic collisions drive $T_s \rightarrow T_{\text{gas}}$, while radiative interactions drive $T_s \rightarrow T_\gamma$.

In accordance with the standard convention in the literature, we define the brightness temperature of 21-cm radiation as [14, 18, 19]:

$$T_{21} = T_s - T_\gamma, \quad \tau_{21} = \frac{3 k_B T_s n_{\text{HI}} \lambda_{21}^3}{4 \pi T_s^2} \frac{A_{10}}{H + \partial_{v_\parallel} v_\parallel},$$

with wavelength $\lambda_{21} = h c / E_{10}$. Note that $\tau_{21}$ depends implicitly on the electron fraction via $n_{\text{HI}} \equiv n_\text{H}(1 - x_e)$. During the Dark Ages, however, $x_e \sim 10^{-4}$, so that we may approximate $n_\text{H} = n_{\text{HI}}$ and the optical depth $\tau_{21}$ becomes independent of the free electron fraction. We denote by $\partial_{v_\parallel} v_\parallel$ the gradient of the component of the peculiar velocity along the line-of-sight ($v_\parallel$). We define the dimensionless velocity gradient:

$$\delta_v \equiv - \frac{\partial_{v_\parallel} v_\parallel}{H}.$$

Then, we may write the 21-cm brightness temperature in the following way:

$$T_{21}(T_s, n_\text{H}, \delta_v) = \frac{3 k_B T_s}{32 \pi n_\text{H} \lambda_{21}^3} \frac{A_{10}}{H(1 - \delta_v)} \frac{T_s - T_\gamma}{1 + z}.$$

Notice that the brightness temperature depends explicitly on $n_\text{H}$ and $\delta_v$ and implicitly on $T_{\text{gas}}$ via $T_s$, which in turn depends on $n_\text{H}$, making the temperature a non-linear tracer of the underlying density field. For later convenience, we define $\tilde{T}_{21}$ as the brightness temperature in which the dependence on the velocity gradient term $\delta_v$ is factored out:

$$\tilde{T}_{21}(T_s, n_\text{H}) \equiv T_{21}(T_s, n_\text{H}, \delta_v) \times (1 - \delta_v).$$

\footnote{In literature, this is sometimes defined without a minus sign (e.g. [19]).}
Figure 1. Coefficients $\mathcal{T}$ coupling brightness temperature fluctuations to the gas temperature and baryon density fluctuations. Left, middle and right panel show the coefficients and linear, quadratic and cubic order, respectively.

2.2 Fluctuations in the 21-cm signal

The final expression for the 21-cm brightness temperature in equation (2.7) shows that it depends on the hydrogen density $n_H$, velocity gradient $\delta v$ and gas temperature $T_{\text{gas}}$ and the photon temperature $T_{\gamma}$. Therefore, fluctuations in the brightness ($\delta T_{21}$) are in principle sourced by those in the hydrogen density, velocity gradient, gas and photon temperature. However, following [14, 19], we will restrict the analysis to sub-horizon scales. On these scales, we assume the photon temperature to be uniform, since their high sound speed ($c_s = c/\sqrt{3}$) results in photon fluctuations being suppressed relative to those in the other fields. In this section, we will find out how 21-cm brightness fluctuations trace those in the hydrogen density, gas temperature and velocity gradient.

2.2.1 21-cm brightness temperature

We start by considering fluctuations in the density and gas temperature, and only include the velocity gradient fluctuations $\delta v$ once we transform to momentum space. That is, we effectively expand $\tilde{T}_{21}$ in terms of fluctuations in $n_H$ and $T_{\text{gas}}$. We define fluctuations in the latter by:

$$n_H(t, x) = \bar{n}_H(t) \left[1 + \delta_b(t, x)\right], \quad T_{\text{gas}}(t, x) \equiv \bar{T}_{\text{gas}}(t) \left[1 + \delta_T(t, x)\right],$$ 

(2.9)

where we have used that $\delta_H = \delta_b$ up to negligible corrections of order $\mathcal{O}(m_e/m_p)$ [19].

Taylor expanding $\delta T_{21}$ around $\delta_T = \delta_b = 0$ up to cubic order in fluctuations, we find:

$$\delta \tilde{T}_{21}(t, x) = \mathcal{T}_T \delta_T + \mathcal{T}_b \delta_b + \mathcal{T}_{TT} \delta_T^2 + \mathcal{T}_{bT} \delta_T \delta_b + \mathcal{T}_{bb} \delta_b^2 + \mathcal{T}_{TTT} \delta_T^3 + \mathcal{T}_{bTT} \delta_b \delta_T^2 + \mathcal{T}_{bTb} \delta_T \delta_b^2 + \mathcal{T}_{bhb} \delta_b^3,$$

(2.10)

where the coefficients $\mathcal{T}_{\nu\nu\nu}(t)$ depend only on time. We have numerically solved for the coefficients $\mathcal{T}$ following [19] and show them in Figure 1.

We can qualitatively understand the behavior of the coefficients $\mathcal{T}$ in the low and high-redshift limits as follows.

---

---

---
• **Efficient Collisional Coupling.**—At \( z \gtrsim 100 \), collisions between Hydrogen atoms efficiently couple \( T_s = T_{\text{gas}} \). The brightness temperature, in the absence of the velocity gradient term, then depends on \( n_{\text{H}} \) and \( T_{\text{gas}} \) as:

\[
T_{21} \propto n_{\text{H}} \left( 1 - \frac{T_{\gamma}}{T_{\text{gas}}} \right).
\]  

(2.11)

Since \( T_{21} \propto n_{\text{H}} \), we find that \( T_b \rightarrow \bar{T}_{21} \) and:

\[
T_{bb} \approx T_{bbb} \approx T_{bb} \rightarrow 0.
\]  

(2.12)

Note also that \( T_{21} \) is proportional to the difference \( T_{\text{gas}} - T_{\gamma} \) in this regime, which asymptotes to zero in the high redshift limit (\( z \gtrsim 200 \)) where Compton heating efficiently coupling the gas temperature to the photons, such that \( T_b \rightarrow 0 \). For the coefficients \( T_{TT} \), \( T_{bT} \) and \( T_{bb} \) we have:

\[
T_{TT}, T_{bT}, T_{TTT} \propto \bar{n}_{\text{H}} \frac{T_{\gamma}}{T_{\text{gas}}}, \quad T_{TTT}, T_{bTT} \propto -\bar{n}_{\text{H}} \frac{T_{\gamma}}{T_{\text{gas}}},
\]  

(2.13)

which yields for \( z \gtrsim 100 \) the following relations between the coefficients:

\[
T_{TT} \approx T_{bTT} \approx T_{TTT} \approx -T_{TTT}.
\]  

(2.14)

Note that these coefficients do not tend to zero, since they are not suppressed by the factor \( T_{\text{gas}} - T_{\gamma} \). Instead they grow as:

\[
T_{21} \propto \frac{n_{\text{H}}}{H} \propto (1 + z)^{3/2},
\]  

(2.15)

during matter domination via the dependence of the optical depth on the ratio \( n_{\text{H}}/H \). The above behavior is indeed verified by the numerical solutions in Figure 1.

• **Ineffective Collisional Coupling.**—For \( z \lesssim 50 \), collisions become very inefficient and \( T_s \) approaches \( T_{\gamma} \) again, which a small difference:

\[
T_s - T_{\gamma} \propto n_{\text{H}} \kappa_{10}^{HE}(T_{\text{gas}}).
\]  

(2.16)

The brightness temperature then scales approximately as \( T_{21} \propto n_{\text{H}}^2 \), implying:

\[
T_b \approx 2T_{bb} \approx 2\bar{T}_{21}, \quad T_{bbb} \approx 0.
\]  

(2.17)

As \( z \rightarrow 0 \), the optical depth decreases rapidly and all coefficients tend to zero quickly.

### 2.2.2 Gas temperature

The evolution equation for the gas temperature \( T_{\text{gas}} \) can be obtained from the first law of thermodynamics. Neglecting anything other than Compton heating, it reads [19]:

\[
\dot{T}_{\text{gas}} - \frac{2}{3} \frac{n_{\text{H}}}{n_{\text{H}}} T_{\text{gas}} = \Gamma_C \frac{x_e}{\bar{x}_e} (T_{\gamma} - T_{\text{gas}}),
\]  

(2.18)

where we have defined the Compton heating rate \( \Gamma_C \) as:

\[
\Gamma_C = \frac{8 \sigma_T a_r T_{\gamma}^4}{3(1 + x_{\text{He}} + x_e) m_e} \bar{x}_e.
\]  

(2.19)
In the above equation, \( \sigma_T \) is the Thomson cross section, \( a_c \) is the radiation constant, \( m_e \) is the electron mass, \( n_H \equiv n_{HI} + n_p \) is the total density of Hydrogen (both in neutral and ionized form), \( x_{He} \equiv n_{He}/n_H \) is the Helium fraction, \( x_e \equiv n_e/n_H \) is the free electron fraction (and \( \bar{x}_e \) its background value). Perturbing the above equation allows us to find a direct relationship between gas temperature fluctuations \( \delta_T \) and baryon density fluctuations \( \delta_b \).

To obtain the evolution equation for \( \delta_T \), one should in principle consistently include the coupling to fluctuations in \( T_\gamma, n_H, x_e \) and \( x_{He} \). To simplify the analysis, we will make the same assumptions as in [19]. As mentioned above, we consider small scales deep inside the horizon (\( k_{\text{min}} > 0.01 \text{ Mpc}^{-1} \)), so that photon temperature fluctuations are negligible and we set \( T_\gamma = \bar{T}_\gamma \). Secondly, we assume the Helium fraction to be uniform, i.e. we take \( x_{He} = \bar{x}_{He} \) and neglect any fluctuations. Finally, in the Compton heating rate \( \Gamma_C \), the free electron fraction only enters via the term \( 1 + x_e + x_{He} \). During the Dark Ages, \( x_e \ll 1 \), so that any electron fraction perturbation of \( \Gamma_C \) enters through \( x_e^2 \delta_{xe} \ll \delta_{xe} \ll 1 \). Hence, we can neglect electron fraction perturbations in \( \Gamma_C \) such that it only depends on background values \( x_e, \bar{x}_{He} \) and \( \bar{T}_\gamma \).

The full non-linear evolution of the gas temperature fluctuation \( \delta_T(t, \mathbf{x}) \) can be obtained directly from equation (2.18) and reads [19]:

\[
\delta_T - \frac{2}{3} \frac{1 + \delta_T}{1 + \delta_b} \frac{\bar{T}_\gamma}{T_{\text{gas}}} \Gamma_C \delta_T = \Gamma_C [(\bar{T}_\gamma/T_{\text{gas}} - 1) \delta_{xe} + \delta_x \delta_T].
\]

We obtain the background evolution of \( \bar{T}_{\text{gas}} \) as well as the recombination history \( \bar{x}_e \) from CAMB. Note that, as written in the form above, the right-hand-side couples the evolution of \( \delta_T \) to the electron fraction fluctuation \( \delta_{xe} \). In [14], the coupling to \( \delta_{xe} \) is neglected (i.e. the right-hand-side is taken to be zero), resulting in errors of order 10% at linear order during the Dark Ages. In this work, we will include the effect of electron fraction perturbations up to third order. To do so, we require an additional equation describing the evolution of \( \delta_{xe} \).

However, before deriving this additional equation, let us find out when the effect of \( \delta_{xe} \) on gas temperature fluctuations is substantial by examining the r.h.s. of equation (2.20). At high redshifts (\( z \gtrsim 500 \)), Compton scattering efficiently couples the gas and CMB temperature, rendering the term proportional to \( (\bar{T}_\gamma/T_{\text{gas}} - 1) \) vanishingly small. In addition, the strong coupling combined with the fact that we consider scales on which \( T_e = \bar{T}_\gamma \) implies \( \delta_T \equiv 0 \), and hence the second term (\( \propto \delta_{xe} \delta_T \)) is negligibly as well. In the low redshift regime \( (z \ll 200) \), the gas cools adiabatically as \( \Gamma_C \ll \dot{H} \) and fluctuations in the electron fraction also have no effect. However, at intermediate stages the effect is expected to be non-negligible.

### 2.2.3 Free electron fraction

Based on the discussion above, we only require an evolution equation for the free electron fraction that is accurate at late times \( (z \lesssim 500) \), and need not worry about the detailed recombination history at early times [14, 20]. We adapt the effective 3-level model for recombination [21]:

\[
\dot{x}_e = -C_P(T_{\text{gas}}, n_H, x_e) \left[ n_H x_e^2 \alpha_B(T_{\text{gas}}) - \beta_B(T_{\text{gas}}) (1 - x_e) e^{-E_{21}/k_B T_{\text{gas}}} \right]\]

where \( E_{21} = 10.2 \text{ eV} \) is the energy difference of the Ly-\( \alpha \) transition, \( \alpha_B \) is the case-B effective recombination coefficient, \( \beta_B \) the corresponding photo-ionization rate. Finally, the Peebles factor \( C_P \) gives the ratio of the effective downward transition rate from the \( n = 2 \) states to
their effective lifetime:

\[ C_P(T_{\text{gas}}, n_H, x_e) = \frac{1 + K \Lambda_{2s-1s} n_H (1 - x_e)}{1 + K \Lambda_{2s-1s} n_H (1 - x_e) + K \beta_B n_H (1 - x_e)}, \]  

where \( K \equiv \lambda_\alpha^3 / 8\pi H \) and \( \lambda_\alpha = 121.5 \text{ nm} \) is the Ly-\( \alpha \) rest wavelength.

Now, we simplify the evolution equation at late times. For \( z \lesssim 500 \), the second term in equation (2.21) describing the effect of photo-ionization, is completely negligible compared to the recombination term proportional to \( \alpha_B \). In addition, \( C_P \to 1 \) for \( z \lesssim 900 \) [19], so that we obtain:

\[ \dot{x}_e = -\alpha_B n_H x_e^2, \]  

(2.23)

to excellent precision. For the recombination coefficient \( \alpha_B \), we use the fit [22]:

\[ \alpha_B(T_{\text{gas}}) = F \frac{a_\alpha T_4^3}{1 + c T_4^4}, \]  

(2.24)

where \( T_4 \equiv T_{\text{gas}} / 10^4 \text{ K} \), \( a_\alpha = 4.309 \times 10^{-19} \text{ m}^3/\text{s} \), \( b = -0.6166 \), \( c = 0.6703 \) and \( d = 0.5300 \). The Fudge factor \( F = 1.14 \) is used to calibrate the effective 3-level result to a multi-level atom calculation [20].

To obtain the evolution equation for \( \delta x_e \), we perturb equation (2.23). Notice that evolution of fluctuations in the electron fraction are coupled to those in the gas temperature and the baryon density via \( \alpha_B(T_{\text{gas}}) \) and \( n_H \), respectively. We expand \( \alpha_B \) in terms of \( \delta T \) up to cubic order as follows:

\[ \alpha_B(T_{\text{gas}}) = \bar{\alpha}_B \left[ 1 + \sum_{n=1}^{3} A_{\delta T}^{n} \delta T^{n} \right], \quad A_{\delta T}^{n} \equiv \frac{1}{n!} \frac{T_{\text{gas}}}{\partial T_{\text{gas}}} \frac{\partial^n \bar{\alpha}_B}{\partial T_{\text{gas}}^{n}}, \]  

(2.25)

where \( \bar{\alpha}_B \equiv \alpha_B(T_{\text{gas}}) \). Upon defining \( \Gamma_R \equiv \bar{\alpha}_B \bar{n}_H \bar{x}_e \) as the background recombination rate, we obtain the evolution of electron fraction perturbations up to terms of cubic order in baryon, gas temperature and electron fraction fluctuations:

\[ \dot{\delta x}_e = -\Gamma_R \left[ \delta x_e + A_1 \delta_T + \delta_b \right. \]
\[ + \left. \delta x_e^2 + A_2 \delta_T^2 + 2A_1 \delta_x e \delta_T + A_1 \delta_T \delta_b + 2 \delta x e \delta_b \right. \]
\[ + A_3 \delta_T^3 + 2A_2 \delta x e \delta_T^2 + A_1 \delta_T^2 \delta_b + A_2 \delta_T^2 \delta_b + \delta_T^2 \delta_b + 2A_1 \delta_T \delta_b \delta_T \delta_b \right]. \]  

(2.26)

2.3 Perturbative analysis of fluctuations

Recall that the goal of this section is to find the relation between \( \delta T_{21} \) and the underlying baryonic density and velocity field. At this point \( \delta T_{21} \) is still a function of fluctuations in the gas temperature as well. However, \( \delta T \) can be traded effectively for \( \delta_b \) by recognizing that gas temperature fluctuations trace fluctuations in the density field. The physical mechanism behind this tracing relationship is simple and may be explained as follows. Consider an overdense region in the gas (\( \delta_b > 0 \)). Due to the higher density, the thermal motion of the particles in the gas in increased and hence the temperature will increase as well: \( \delta T \) traces \( \delta_b \).

In order to find the tracing relationship between \( \delta_T \) and \( \delta_b \), we start by writing the baryonic density contrast perturbatively, up to cubic order:

\[ \delta_b = \sum_{n=1}^{3} \delta_b^{(n)} \equiv \sum_{n=1}^{3} \delta_n, \]  

(2.27)
where we assume $\delta_b^{(n)} = \mathcal{O}\left[\left(\delta_b^{(1)}\right)^n\right]$. In the second equality, we defined $\delta_n \equiv \delta_b^{(n)}$ for notational brevity. We assume that the time-dependence of the baryons is identical to that of CDM, i.e. $\delta_b^{(n)} \propto a^{(n)}(t)$, which implies $\delta_b^{(n)} = nH(\delta_b^{(n)})^3$. Similarly, we expand $\delta_T$ and $\delta_{sx}$ to cubic order as well: \[ \delta_{sx}(t, x) = \sum_{n=1}^{3} \delta_{n}(t, x), \quad \delta_T(t, x) = \sum_{n=1}^{3} \delta_{n}(t, x). \] (2.28)

At each order, the evolution equations for the perturbations in the temperature and electron fraction can be obtained by inserting the above expansions into equation (2.20) and equation (2.26), respectively. In App. A we provide the resulting evolution equations and show how they can be used to solve for the coupling coefficients, treating $\delta_T$ and $\delta_{sx}$ to $\delta_b$ up to third order:

\[ \delta_T(z, x) = C_{T, 1, 1}^{(z)}(z) \delta_1 + C_{T, 2, 1}^{(z)}(z) \delta_1^2 + C_{T, 2, 2}^{(z)}(z) \delta_2 + C_{T, 3, 1}^{(z)}(z) \delta_1 \delta_2 + C_{T, 3, 2}^{(z)}(z) \delta_3, \] (2.29)

\[ \delta_x(z, x) = C_{x, 1, 1}^{(z)}(z) \delta_1 + C_{x, 2, 1}^{(z)}(z) \delta_1^2 + C_{x, 2, 2}^{(z)}(z) \delta_2 + C_{x, 3, 1}^{(z)}(z) \delta_1 \delta_2 + C_{x, 3, 2}^{(z)}(z) \delta_3, \] (2.30)

where the coupling coefficients are functions of redshift only. In the coefficient $C_{n,m}$, $n$ denotes the total order of the combination of density perturbations it multiplies, and $m = 1, \ldots, n$ labels the different coefficients at each order $n$. Note that $C_{n,m}$ is the coefficient coupling to density perturbation $\delta_n$, whereas $C_{n,m \neq n}$ couple to the product of lower order density fields with combined order $n$. At first order, we will often write $C_1 \equiv C_{1,1}$ for notational simplicity.

**Evolution of Coupling Coefficients**

In Figure 2, we show the numerical solutions for the gas coupling coefficients $C_{n,m}^{T}$. The solid lines show the solutions including the effect of electron fraction perturbations. The dotted lines exclude their effect by setting $C_{n,m}^{x} \equiv 0$ in all evolution equations. It is insightful to examine the numerical results at early (high-$z$) and late times (low-$z$), where the numerical analysis can be compared with analytic results:

\[ \text{Note that we do not assume the spatial dependence of baryons and CDM to be identical. In fact, we will include the spatial dependence due to baryonic pressure, in contrast to [14].} \]

\[ \text{In the notation below, the superscripts } T \text{ and } x \text{ are labels, not powers.} \]
- **High redshift limit**—At high redshifts, \( z \gtrsim 500 \), Compton heating is efficient \( (\Gamma_C/H \gg 1) \) and keeps the matter and photons in equilibrium so that \( T_\gamma = T_{\text{gas}} \). Since we consider small scales on which we neglect photon temperature fluctuations, we therefore expect no gas temperature fluctuations. This is indeed verified by the solutions for \( C_{n,m}^T \), which all tend to zero at high redshifts. Only when the gas temperature starts to decouple from the photon temperature at \( z \lesssim 500 \) due to adiabatic cooling, fluctuations in the former start growing.

- **Low redshift limit**—At low redshifts, Compton heating becomes completely inefficient, \( \Gamma_C/H \ll 1 \), and adiabatic cooling completely determines the gas temperature:

\[
\dot{T}_{\text{gas}} = \frac{2}{3} n_H T_{\text{gas}} ,
\]

which is just equation (2.18) in the absence of the Compton heating term (proportional to \( \Gamma_C \)). From the above equation, we easily obtain:

\[
T_{\text{gas}} \propto n_H^{2/3} ,
\]

characterizing a gas that is cooling adiabatically. The above relationship between the gas temperature and Hydrogen density yields the following relationship between \( \delta_T \) and \( \delta_b \):

\[
\delta_T = \frac{2}{3} \delta_b - \frac{1}{9} \delta_b^2 + \frac{4}{81} \delta_b^3 + \mathcal{O}(\delta_b^4)
\]

\[
= \frac{2}{3} (\delta_1 + \delta_2 + \delta_3) - \frac{1}{9} (\delta_2^2 + 2\delta_1 \delta_2) + \frac{4}{81} \delta_3^3 + \mathcal{O}(\delta_1^4),
\]

where we have expanded \( \delta_b \) up the third order in the second line. Comparing to equation (A.9), we find that the coefficients asymptote to constant values in the adiabatic or low redshift regime:

\[
C_{1}^T = C_{2,2}^T = C_{3,3}^T \rightarrow 2/3, \quad C_{2,1}^T \rightarrow -1/9, \quad C_{3,1}^T \rightarrow 4/81, \quad C_{3,2}^T \rightarrow -2/9,
\]

which is indeed verified by the numerical solution.

Below, we will first discuss the case in which electron fraction perturbations are neglected, and make contact with results obtained in literature. Then, we discuss the effect of electron fluctuations and provide a physical interpretation of their effect at first order.

As mentioned above, excluding the effect of electron fraction fluctuations leads to the dotted curves for the gas temperature coefficients. Our results agree with those of [14], in which the gas temperature coefficients where computed up to second order and in the absence of electron fluctuations. We find excellent agreement for the coefficients \( C_{1}^T \), \( C_{2,1}^T \) and \( C_{2,2}^T \), which are called \( C_1 \), \( C_2 \) and \( C_2' \) respectively in [14] and shown in their Figure 1. In addition, the first-order coupling coefficient agrees with the result obtained neglecting electron fraction fluctuations in [11], where it is called \( g_1 \) (see their Figure 4).

Comparing the solid and dotted curves in Figure 2, we find that the effect of electron fraction perturbations is most pronounced at linear order, i.e. for \( C_1^T \). At \( z = 30, 50 \) and 100, the solution excluding electron fraction perturbations is larger by \( 8\% \), \( 14\% \) and \( 34\% \), respectively. Our results confirm previous claims that neglecting \( \delta_x \) leads to errors of order \( 10\% \) for the linear evolution during the Dark Ages [14]. However, we also note that the effect
of \( \delta x_e \) rapidly becomes more significant at higher redshift during the Dark Ages, and should therefore be accounted for in a detailed analysis. At second and third order, the effect of electron fraction perturbations is smaller, but still starts to exceed the 10% level at \( z \geq 100 \), before decreasing again as \( z \to 1100 \).

In the low redshift limit, the solutions including and excluding \( \delta x_e \) agree again, which can be understood as follows. Due to the expansion of the universe, the Hydrogen gas becomes diluted at low redshift, rendering the recombination rate small: \( \Gamma_R / H \ll 1 \). The evolution equations for \( C_{n,m}^T \) then reduce to the form:

\[
\frac{dC_{n,m}^T}{da} \propto -\frac{C_{n,m}^T}{a},
\]

implying that the electron fraction coefficients decay as \( C_{n,m}^T \propto a^{-1} \) at low redshift. In turn, the effect of electron fraction perturbations on the evolution of \( C_{n,m}^T \) becomes small at low redshifts, and the dotted and solid curves become identical. Below, we will discuss the effect of \( \delta x_e \) in more detail at first order.

At linear order, the effect of electron fraction fluctuations on the evolution of gas temperature fluctuations can be understood in an intuitive way from the physical principles at play [20]. Consider a region with a density that is higher than average, so that \( \delta T_1 > 0 \). In such a region, Hydrogen density is higher than average, so that the local recombination rate is slightly higher and hence more recombinations occur. As a consequence, the electron fraction will be lower than average, resulting in \( \delta T_1 < 0 \) and reflected by \( C_{n,m}^T \) < 0. In turn, this reduces the coupling of the gas temperature to the photon temperature via Compton scattering, since there are less free electrons available to sustain the coupling. At background level, \( T_{\text{gas}} \) is below \( T_\gamma \) during the Dark Ages due to adiabatic cooling. In an overdense region, the reduced coupling between \( T_\gamma \) and \( T_{\text{gas}} \) makes the latter even smaller relative to the background difference. This results in \( \delta T_1 \) (or equivalently \( C_1^T \)) being smaller than it would have been if electron fraction fluctuations were neglected.

### 2.4 Coupling 21-cm fluctuations to density and velocity fluctuations

At this point, we are in the position to expand 21-cm fluctuations into fluctuations \( \delta b \) and \( \delta v \), as we have uniquely related temperature fluctuations to density fluctuations. We will make this expansion explicit in real space first. Subsequently, we transform to momentum space, where we can relate \( \delta v \) to the velocity divergence \( \theta_b \).

**Real Space**

Now we have all the ingredients to compute the coupling between the 21-cm brightness temperature and density fluctuations up to third order. We will denote these couplings by \( \alpha_{n,m} \), and they are defined in relation to \( \delta T_{21} \) as:

\[
\delta T_{21}(z, x) = \alpha_{1,1}(z)\delta_1 + \alpha_{2,1}(z)\delta_2 + \alpha_{3,1}(z)\delta_3 + \alpha_{3,2}(z)\delta_2 + \alpha_{3,3}(z)\delta_3. \tag{2.36}
\]

Using equation (2.29) and equation (2.30), we can write the couplings \( \alpha_{n,m} \) explicitly in terms of the \( T \) coefficients and the couplings \( C_{n,m}^T \), yielding:

\[
\alpha_{n,m} = T_T C_{n,m}^T + T_b, \tag{2.37}
\]

\[
\alpha_{2,1} = T_T C_{2,1}^T + T_{TT}[C_{1}^T]^2 + T_{bT}C_{1}^T + T_{bb}, \tag{2.38}
\]

\[
\alpha_{3,1} = T_T C_{3,1}^T + 2T_T T_{TT}[C_{1}^T]^2 + T_{bT} T_{TT}[C_{1}^T]^3 + T_{bT} T_{TT}[C_{1}^T]^2 + T_{bT} T_{TT} C_{1}^T + T_{bb}, \tag{2.39}
\]

\[
\alpha_{3,2} = T_T C_{3,2}^T + 2T_T T_{TT} C_{2,2} T_{TT} + T_{bT} C_{2,2} + T_{bb} (C_{1}^T + C_{1}^T) + 2T_{bb}. \tag{2.40}
\]
We plot the couplings in Figure 3 over the redshift range $10 - 1000$. Recall that in the adiabatic limit $C_{\alpha,\alpha}^T \rightarrow 2/3$ so that the couplings $\alpha_{n,n}$ become identical, which is indeed the case for $z \lesssim 50$. In addition, note that $C_{\alpha,\alpha}^T \approx 2C_{\alpha,\alpha}^T$ and $C_{\alpha,\alpha}^T \approx C_{\alpha,\alpha}^T$ in the adiabatic limit, such that $\alpha_{3,2} \approx 2\alpha_{2,1}$, which is indeed the case for $z \lesssim 30$. Note that in the approach above, we have effectively traded the dependence on $\delta_T$ for dependence on $\delta_\nu$. Ultimately, we are able to do so by using the first law of thermodynamics.

So far, we have worked with $\delta \tilde{T}_{21}$, i.e. we ignored the fluctuations $\delta_\nu$. To include the perturbations in $\delta_\nu$, we use the relation $T_{21} = \tilde{T}_{21}/(1 - \delta_\nu)$ and expand $\delta_\nu$ to third order, i.e. $\delta_\nu = \delta_\nu^{(1)} + \delta_\nu^{(2)} + \delta_\nu^{(3)}$ 5. Assuming $|\delta_\nu| \ll 1$ and using a geometric series expansion, we obtain:

$$
\delta T_{21}(z, x) = T_{21}(z)\left[\delta_\nu^{(1)}(1 + \delta_\nu^{(1)}) + \delta_\nu^{(2)}(1 + \delta_\nu^{(1)})^2 + \delta_\nu^{(3)}(1 + \delta_\nu^{(1)})^3\right](z, x)
$$

$$
+ \alpha_{1,1}(z)\left[1 + \delta_\nu^{(1)} + \delta_\nu^{(2)} + \delta_\nu^{(3)}\right]\delta_\nu^{(1)}(z, x)
$$

$$
+ \alpha_{2,1}(z)\left[1 + \delta_\nu^{(1)}\right]\delta_\nu^{(1)}(z, x) + \alpha_{1,2}(x)\left[1 + \delta_\nu^{(1)}\right]\delta_\nu^{(2)}(z, x)
$$

$$
+ \alpha_{3,1}(z)\left[\delta_\nu^{(1)}(z, x)\right]^3 + \alpha_{3,2}(z)\left[\delta_\nu^{(1)}\delta_\nu^{(2)}\right](z, x) + \alpha_{3,3}(z)\delta_\nu^{(3)}(z, x).
$$

(2.41)

**Momentum Space**

In the end, we wish to obtain correlation functions of 21-cm fluctuations in momentum space. Therefore, we transform the above expression for $\delta T_{21}(z, x)$ to momentum space. To do so we have to transform all different components separately. Products of perturbations in real space, e.g. the square of $\delta_\nu^{(1)}$, become convolutions in momentum space. In addition, the transform of $\delta_\nu$ can be related directly to the transform of the velocity divergence $\theta_b \equiv \nabla \cdot \nu_b$.

---

5 Although the fluctuations $\delta_\nu$ appear similar to redshift space distortions (RSD), they come from the velocity dependence of the optical depth and hence are of a qualitative different nature. In principle one should also include RSDs separately when fully modelling the 21-cm signal, introducing the linear growth rate $f = d\ln D_s/d\ln a$ as an additional parameter to marginalize over. We have chosen to ignore RSDs in our forecast as they do not introduce additional shape dependence and therefore overlap with the primordial shapes.
The Fourier transform of $\delta_v$ is written in terms of that of $\theta_b$ as:

$$\delta_v(z, k) = -\mu^2(k)\theta_b(z, k)\frac{\mathcal{H}}{k}, \quad (2.42)$$

where $\mathcal{H} = aH$ and $\mu(k) \equiv \hat{n} \cdot \hat{k}/k = k_\parallel/k$ is the cosine of the angle between the mode $k$ and the line-of-sight $\hat{n}$. At first order, this expression can be simplified even further by invoking the continuity equation, to obtain

$$\delta(v^{(1)}(z, k) = \mu^2(k)\delta_b^{(1)}(z, k). \quad (2.43)$$

Note that in transforming to momentum space, the angle between the line-of-sight and the Fourier mode is encoded in $\mu^2(k)$. This leads to a distinct difference between coupling coefficients in momentum space and those in real space ($T_{21}$ and $\alpha_{i,j}$); the former depend on the mode $k$ and time, whereas the latter depend only on time and not on $x$.

Now, we have all the relevant ingredients to compute the Fourier transform of the brightness fluctuation $\delta T_{21}$. At each order, we couple the brightness perturbations to those in the baryon density and velocity divergence via the coefficients $c_j^{(i)}(z, k, q)$, of which some explicitly depend on the principal mode $k$ and internal modes $q$ resulting from convolution integrals via the angle $\mu$ as explained above. The label $i$ denotes the total order of the quantity the coefficient couples to, and $j$ labels all different couplings at order $i$. Up to third order, the expansion of $\delta T_{21}(z, k)$ is given by:

$$\delta T_{21}(z, k) = \delta T_{21}^{(1)}(z, k) + \delta T_{21}^{(2)}(z, k) + \delta T_{21}^{(3)}(z, k), \quad (2.44)$$

with $\delta T_{21}^{(i)}(z, k)$ given in terms of the coefficients $c_j^{(i)}$ as:

$$\delta T_{21}^{(1)}(z, k) = c_1^{(1)}(z, k)\delta_b^{(1)}(k)$$

$$\delta T_{21}^{(2)}(z, k) = c_1^{(2)}(z)\delta_b^{(2)}(k) - \frac{1}{\mathcal{H}}c_2^{(2)}(z, k)\delta_b^{(2)}(k) + \int_{q_1} c_3^{(2)}(z, k, q_1)\delta_b^{(1)}(q_1)\delta_b^{(1)}(k - q_1)$$

$$\delta T_{21}^{(3)}(z, k) = c_1^{(3)}(z)\delta_b^{(3)}(k) - \frac{1}{\mathcal{H}}c_2^{(3)}(z, k)\delta_b^{(3)}(k)$$

$$+ \int_{q_1, q_2} c_3^{(3)}(z, k, q_1, q_2)\delta_b^{(1)}(q_1)\delta_b^{(1)}(q_2)\delta_b^{(1)}(k - q_{12})$$

$$- \frac{1}{\mathcal{H}}\int_{q_1} c_4^{(3)}(k, q_1)\delta_b^{(2)}(q_1)\delta_b^{(1)}(k - q_1)$$

$$+ \int_{q_1} c_5^{(3)}(k, q_1)\delta_b^{(2)}(q_1)\delta_b^{(1)}(k - q_1), \quad (2.45)$$

where we have suppressed the time dependence in $\delta_b^{(a)}$ and $\theta_b^{(a)}$ for brevity. We explicitly excluded the factors of $1/\mathcal{H}$ from the definition of the couplings $c_j^{(i)}$, so that the latter all have the dimension of temperature.
The couplings are a function of $\alpha_{i,j}$, $T_{21}$ and $\mu(k)$ and read:

\[
\begin{align*}
    c^{(1)}_1 &= \alpha_{1,1}(z) + T_{21}(z)\mu^2(k), \\
    c^{(2)}_1 &= \alpha_{2,2}(z), \\
    c^{(2)}_2 &= T_{21}\mu^2(k), \\
    c^{(2)}_3 &= T_{21}(z)\mu^2(q)\mu^2(k-q) + \alpha_{1,1}\mu^2(q) + \alpha_{2,1}(z), \\
    c^{(3)}_1 &= \alpha_{3,3}(z), \\
    c^{(3)}_2 &= T_{21}(z)\mu^2(k), \\
    c^{(3)}_3 &= T_{21}(z)\mu^2(q_1)\mu^2(q_2)\mu^2(k-q_12) + \alpha_{1,1}(z)\mu^2(q_1)\mu^2(q_2) + \alpha_{2,1}(z)\mu^2(q_1) + \alpha_{3,1}(z), \\
    c^{(4)}_4 &= 2T_{21}(z)\mu^2(q)\mu^2(k-q) + \alpha_{1,1}(z)\mu^2(q), \\
    c^{(3)}_5 &= \alpha_{2,2}(z)\mu^2(k-q) + \alpha_{3,2}(z).
\end{align*}
\]

(2.46)

Note that since $\mu(k) = \mu(-k)$, all coefficient are symmetric under inverting all momentum arguments, e.g. $c^{(3)}_3(k, q_1, q_2) = c^{(3)}_3(-k, -q_1, -q_2)$. The first order coupling, $c^{(1)}_1$, actually comprises two couplings, to $\delta^{(1)}_b$ and $\theta^{(1)}_b$, respectively. However, using equation (2.43) allows us to write down one overall coefficient coupling to $\delta^{(1)}_b$. In summary, we have found a direct relationship between $\delta T_{21}$ and $\delta_b$ to third order in perturbations.

We are now able to express the statistics of 21-cm temperature fluctuations in terms of the underlying tracer field $\delta_b$, which in turn traces primordial fluctuations, e.g. the power spectrum:

\[
P_{\delta T_{21}}(k_1) = \langle \delta T_{21}(k_1)\delta T_{21}(-k_1) \rangle' = \left( c^{(1)}_1(k_1) \right)^2 \langle \delta_b(k_1)\delta_b(-k_1) \rangle' = \left( c^{(1)}_1(k_1)M_b(k_1) \right)^2 P_{\zeta}(k_1)
\]

(2.47)

where $M_b$ is the linear transfer function of baryon fluctuations (obtainable through e.g. CAMB) and the time (redshift) dependence of the prefactor is implicit.

3 Non-Gaussianity of 21-cm brightness temperature anisotropies

In this section we review the non-Gaussian contributions to the statistics of 21-cm brightness temperature fluctuations during the Dark Ages. First we will go over the details of primordial non-Gaussianity and present templates that can be used to extract imprints of inflation from 21-cm data. Subsequently, we discuss the generation of secondary non-Gaussianity in the 21-cm signal due to non-linear evolution of fluctuations, such as gravitational collapse.

3.1 Primordial non-Gaussianity

As was mentioned in the introduction, inflation naturally seeds structure formation through the generation of quantum fluctuations. Over time, higher density regions collapse under gravitational attraction, forming stars and eventually galaxies. The statistics of the initial conditions set by inflation contain a wealth of information about the physics driving the accelerated expansion, as well as any extra particles present during this period. Primordial
(scalar) fluctuations are captured by the gauge invariant quantity $\zeta$ with mean zero. Generally, the $N$-point statistical correlation function of primordial fluctuations is given by:

$$\langle \zeta_{k_1}\zeta_{k_2}...\zeta_{k_N} \rangle = (2\pi)^3\delta_D(k_1 + k_2 + ... + k_N)F^{(N)}_\zeta(k_1, k_2, ..., k_N)$$  \hspace{1cm} (3.1)$$

where $F^{(N)}_\zeta$ encodes the size and shape of the $N$-point correlation as a function of the momentum configuration. The lowest order statistic is known as the power spectrum:

$$\langle \zeta_{k_1}\zeta_{k_2} \rangle' = P_\zeta(k_1)$$  \hspace{1cm} (3.2)$$

Moving beyond the power spectrum, in this work we will be concerned with the first two non-Gaussian contributions $N = 3$ (bispectrum) and $N = 4$ (trispectrum):

$$\langle \zeta_{k_1}\zeta_{k_2}\zeta_{k_3} \rangle' = B_\zeta(k_1, k_2, k_3)$$

$$\langle \zeta_{k_1}\zeta_{k_2}\zeta_{k_3}\zeta_{k_4} \rangle' = T_\zeta(k_1, k_2, k_3, s, t)$$  \hspace{1cm} (3.3)$$

where $s$ and $t$ are the Mandelstam-like variables as presented at the end of the introduction section. We will now discuss some common shapes of non-Gaussianity that arise in inflationary model building, first for the bispectrum, then the trispectrum.

### 3.1.1 Primordial bispectra

The simplest way to generate non-Gaussianity is by expanding fluctuations locally in terms of Gaussian fields as follows:

$$\zeta = \zeta_g + \frac{3}{5} f_{NL}^{\text{loc}} \zeta_g^2$$  \hspace{1cm} (3.4)$$

where the subscript $g$ denotes Gaussian fields, that have vanishing $(N > 2)$-point functions. Such an expansion gives rise to the so called local shape of the primordial bispectrum:

$$B^{\text{loc}}(k_1, k_2, k_3) = \frac{6}{5} f_{NL}^{\text{loc}} (P_\zeta(k_1)P_\zeta(k_2) + P_\zeta(k_2)P_\zeta(k_3) + P_\zeta(k_1)P_\zeta(k_3))$$  \hspace{1cm} (3.5)$$

which peaks in the squeezed triangle configuration (e.g. $k_1 \ll k_2 \approx k_3$). Local non-Gaussianity naturally arises in multi-field models of inflation, where extra light fields are present that modulate the dynamics. When the inflaton is able to interact with itself, bispectra of the equilateral type are generated. Their shape is well described by the template:

$$B^{\text{equil}}(k_1, k_2, k_3) = \frac{18}{5} f_{NL}^{\text{equil}} \left[ - (P_\zeta(k_1)P_\zeta(k_2) + 2 \text{ perms.}) - 2P_\zeta^{2/3}(k_1)P_\zeta^{2/3}(k_2)P_\zeta^{2/3}(k_3) 
\quad + (P_\zeta(k_1)P_\zeta(k_2))^{1/3}P_\zeta(k_3)^{2/3} + 5 \text{ perms.} \right]$$  \hspace{1cm} (3.6)$$

and peaks in the equilateral momentum configuration ($k_1 \approx k_2 \approx k_3$). All shapes that are due to self interactions can be systematically classified using the Effective Field Theory of

---

6 The number of degrees of freedom of the $N$-point correlation function is reduced by homogeneity and isotropy to be $3(N - 2)$.

7 From here on out we will use the primed correlator notation to remove the usual factor $(2\pi)^3\delta_D(\ldots)$. 

---
Inflation [23]. Finally, a third shape that is often encountered is the orthogonal shape
\[ B^{\text{ortho}}(k_1, k_2, k_3) = \frac{18}{5} f_{\text{NL}}^{\text{ortho}} \left[ -3(P_\zeta(k_1)P_\zeta(k_2) + 2 \text{ perms.}) - 8P_\zeta^{2/3}(k_1)P_\zeta^{2/3}(k_2)P_\zeta^{2/3}(k_3) 
+ 3(P_\zeta(k_1)P_\zeta(k_2)P_\zeta(k_3)^{2/3} + 5 \text{ perms.}) \right] \]
(3.7)
which is generated by more exotic models such as Galileon inflation. These three primordial shapes are the most commonly encountered bispectra in inflationary model building, hence in this work we will often refer to them together as the common primordial bispectra.

One of the most exciting prospects of measuring non-Gaussianity is the possibility of probing the particle spectrum during inflation. Heavy particles can naturally be present during this high energy phase, leaving a characteristic oscillatory shape in the squeezed limit of the primordial bispectrum, with frequency set by its mass [7, 9]. More concretely, under exchange of a scalar particle with mass \( m^2 > 3H/2 \) this oscillation is given by [7]:
\[
\lim_{k_3 \to 0} \langle \zeta(k_1)\zeta(k_2)\zeta(k_3) \rangle' \propto \frac{\pi^2}{k_1^2 k_3^2 \cosh^2 \pi \mu} \left( \frac{k_3}{k_1} \right)^{\frac{3}{2}} \left[ \left( \frac{k_3}{4k_1} \right)^{\frac{1}{2}i\mu} (1 - i \sinh \pi \mu)(\frac{5}{2} - i\mu)(\frac{3}{2} - i\mu)\Gamma(i\mu) \right] \frac{\Gamma(\frac{1}{2} + i\mu)}{\Gamma(\frac{1}{2} + i\mu)} + c.c.
\]
(3.8)
where \( \mu = \sqrt{m^2/H^2 - 9/4} \), clearly showing an oscillation in \( \log(k_3/k_1) \) with amplitude, frequency and phase set by the mass of the scalar particle. The oscillations induced by these heavy fields, could be used to follow the evolution of the scale factor, possibly providing additional information about inflation. For this reason these fields are often referred to as primordial standard clocks [24–27]. In [12] a template is proposed for capturing the oscillatory behaviour of the bispectrum:
\[
B^{\text{clock}}(k_1, k_2, k_3) = \frac{3^{3/2}}{10} f_{\text{NL}}^{\text{clock}} \frac{A_\zeta^2}{(k_1 k_2 k_3)^2} \alpha_{123}^{-1/2} \sin \left( \mu \log \left( \frac{\alpha_{123}}{2} \right) - \delta \right) \Theta(\alpha_{123} - \alpha_0) + 2 \text{ perms.}
\]
(3.9)
where \( \alpha_{123} = (k_1 + k_2)/k_3 \) and the Heaviside step function \( \Theta \) is included since the oscillation is only present for \( \alpha_{123} > 2 \) (or permutations). The authors consider a value of \( \alpha_0 = 10 \) in order to cut off near-equilateral configurations such that overlap with the equilateral shape is reduced. Such a cutoff also drastically reduces the amount of triangle configurations available to measure the signal, thereby lowering its signal-to-noise ratio (SNR). As we will see later in this work, the reduced overlap with the equilateral shape does not make up for the loss of SNR and a cutoff \( \alpha_0 > 2 \) is found to be optimal.

If the exchange particle has a mass \( 0 \leq m^2 \leq 3H/2 \) the oscillatory behaviour turns into a power-law [28], called the intermediate shape. Although there exists no analytical expression, the shape is captured well by the following template [12, 28]:
\[
B^{\text{int}}(k_1, k_2, k_3) = \frac{6}{5} A_\zeta^2 f_{\text{NL}}^{\text{int}} 3^2 k_1^2 + k_2^2 + k_3^2 (k_1 k_2 k_3)^{-\frac{1}{2}} \Theta(k_1, k_2, k_3, \alpha_0)
\]
(3.10)
\[
8\text{In order to stick to conventions in literature we use } \mu \text{ for both the line-of-sight angle } \mu(k) \text{ and the mass parameter } \mu(m). \text{ This is unfortunate but their meaning should be sufficiently clear from the context.}
where $\nu = \sqrt{9/4 - (m/H)^2} = -i\mu$ and $0 < \nu < 3/2$. This template interpolates between the behaviour of the local ($\nu > 3/2$) and equilateral ($\nu < 1/2$) shapes in the squeezed limit, hence coining the name intermediate. In order to reduce overlap with equilateral shapes, we introduce a cutoff similar to the clock template:

$$
\tilde{\Theta}(k_1, k_2, k_3, \alpha_0) = \Theta(\alpha_{123} - \alpha_0) + \Theta(\alpha_{132} - \alpha_0) + \Theta(\alpha_{231} - \alpha_0) \quad (3.11)
$$

In [12] the cutoff was again set to $\alpha_0 = 10$ whereas in this work we will find $\alpha_0 = 2$ to maximize the SNR when marginalising over the other primordial shapes.

Finally, a word about the normalization of the primordial bispectra. The common primordial shapes and intermediate shape are normalized such that $B(k, k, k) = \frac{18}{5} P_\zeta(k)^2$ in the equilateral configuration. The clock template is normalized such that its amplitude matches that of the intermediate shape for the transition mass $\nu = \mu = 0$ in the squeezed limit.

### 3.1.2 Primordial trispectra

The next order of correlation functions known as the trispectrum, correlate four fluctuations. As for the bispectrum, the simplest way of generating such a non-Gaussianity is by expanding the primordial field in terms of Gaussian fields to third order:

$$
\zeta = \zeta_g + \frac{3}{5} f_{NL} \zeta_g^2 + \frac{9}{25} g_{NL}^3 \zeta_g^3 \quad (3.12)
$$

in Fourier space, this expansion leads to two distinct local trispectra:

$$
T_{\zeta}^{NL} = \tau_{NL} \left[ P_\zeta(k_1) P_\zeta(k_3) P_\zeta(s) + P_\zeta(k_1) P_\zeta(k_3) P_\zeta(t) \right] + 6 \text{ perms.} \quad (3.13)
$$

$$
T_{\zeta}^{NL} = \frac{54}{24} g_{NL}^3 P_\zeta(k_1) P_\zeta(k_2) P_\zeta(k_3) + 3 \text{ perms.} \quad (3.14)
$$

where $s = |s| = |k_1 + k_2|$, $t = |t| = |k_1 + k_4|$ and $u = |u| = |k_1 + k_3|$ are diagonal momenta.

For single-field models, the $\tau_{NL}^{loc}$ is always generated in case of a nonzero local bispectrum, such that $\tau_{NL}^{loc} \geq (\frac{6}{5} f_{NL}^{loc})^2$, known as the Suyama-Yamaguchi bound [29]. Furthermore, this shape peaks in the collapsed limit, where one of the diagonals (e.g. $s$) becomes much smaller than the external momenta. The $g_{NL}^{loc}$ amplitude is an independent variable and is characterized by a peak in the double squeezed limit, in which two external momenta become small.

Quartic self-interactions of the inflaton field give rise to equilateral shapes, similar to the equilateral bispectrum. Such interactions can be studied systematically using the EFT of inflation [23]. This effective field theory approach allows one to study the interactions of the Goldstone of time translations $\pi = -\zeta/H$ in the decoupling limit (i.e. without gravity). Typically, quartic interactions for $\pi$ will imply the existence of cubic interactions that give rise to bispectra that have a much stronger signal than the corresponding trispectrum, unless some contrived mechanism is being considered [30]. Hence, such interactions will likely be detected already by observing the bispectrum. In the presence of an additional light scalar $\sigma$, we instead study the EFT of multifield inflation [31]. In this case, it is possible to write down quartic interactions for $\sigma$ that generate a trispectrum for $\zeta$ without generating a larger bispectrum, making the trispectrum the leading order observable non-Gaussianity in such theories [32]. At lowest order in derivatives one finds three interactions $\dot{\sigma}^4$, $\dot{\sigma}^2(\partial \sigma)^2$ and $(\partial \sigma)^4$, which give rise to three distinct trispectra.
we propose the template:

\[
T^\Delta_\zeta = \frac{221184}{25} g^4_{\text{NL}} \frac{1}{k_1 k_2 k_3 k_4 k_5^4},
\]

\[
T^2 (\partial \sigma)^2 = -\frac{27648}{325} g^{2}_\text{NL} \frac{\partial^2 (\partial \sigma)^2}{k_1} + 3 (k_3 + k_4) k_1 + 12 k_3 k_4 (\mathbf{k}_3 \cdot \mathbf{k}_4) + 5 \text{ perms},
\]

\[
T^4 (\partial \sigma)^4 = \frac{165888}{2575} g^4_{\text{NL}} \frac{2 k_1^2 - 2 k_1^2 \sum_i k_i^2 + k_i \sum_i k_i^2 + 12 k_1 k_2 k_3 k_4 ((\mathbf{k}_1 \cdot \mathbf{k}_2)(\mathbf{k}_3 \cdot \mathbf{k}_4) + 2 \text{ perms}),}
\]

where \( k_i = k_1 + k_2 + k_3 + k_4 \). The amplitudes are normalized such that they match the amplitude of the local shape for \( g_{\text{NL}} = 1 \) in the tetrahedral configuration where \( k_i = k \) and \( \mathbf{k}_i \cdot \mathbf{k}_j = -k^2/3 \): \( \langle \zeta^4 \rangle' = \frac{216}{25} \mathcal{P} \langle \zeta \rangle^3 \) [30]. As we will see, these shapes have a significant overlap. Therefore, when searching for an equilateral trispectrum, it suffices to start of with just one shape.

In [7] it was noted that the oscillatory signature of additional heavy particles is also present in the collapsed limit of the trispectrum. For example, the exchange of a heavy scalar gives rise to a signal:

\[
\lim_{s \to 0} \langle \zeta k_1 \zeta k_2 \zeta k_3 \zeta k_4 \rangle' = \frac{1}{4 k_1^2 k_2^2 k_3^2 k_4^2} \frac{1}{128 \pi} (k_{12} k_{34})^{3/2} \times \left[ \left( \frac{s^2}{4 k_{12} k_{34}} \right)^{\mu} (1 + i \sin \pi \mu) \left( \frac{3}{2} + i \mu \right) \left( \frac{5}{2} + i \mu \right) \Gamma(-i \mu) \Gamma(\frac{1}{2} + i \mu)^2 + \text{c.c} \right]
\]

(3.18)

where \( k_{12} = k_1 + k_2 \) (note that \( s \neq k_{12} \)). Again the frequency, amplitude and phase of the oscillation, are set by the mass through \( \mu \). In order to extract such an oscillation from data, we propose the template:

\[
T^\text{clock} = \frac{3^2}{16} g^2_{\text{NL}} \frac{1}{k_1^3 k_2^3 k_3^3 k_4^3} (k_{12} k_{34}) \frac{3}{2} \Theta(\alpha_{12} - \alpha_0) \Theta(\alpha_{34} - \alpha_0) \sin(\mu \ln(\alpha_{12, 34}) + \delta) + 2 \text{ perms}
\]

(3.19)

where \( \alpha_{12, 34} = \frac{s^2}{k_{12} k_{34}} \) and \( \alpha_{12} = \frac{k_{12}}{s} \). For scalars with an intermediate mass, the oscillatory behaviour in the collapsed limit again turns into a power law as found in [33]:

\[
\lim_{s \to 0} \langle \zeta k_1 \zeta k_2 \zeta k_3 \zeta k_4 \rangle' = 4 g^4_{\text{NL}} \mathcal{P}_\zeta (k_1) P \mathcal{P}_\zeta (k_2) P \mathcal{P}_\zeta (s) \left( \frac{s}{\sqrt{k_1 k_3}} \right)^{3-2\nu}
\]

(3.20)

To constrain this type of non-Gaussianity we propose the use of a template:

\[
T^\text{int} = 4 \frac{3^2}{2^\nu} g^4_{\text{NL}} \frac{1}{(k_1 k_2 k_3 k_4)^2} \frac{1}{s^8} \alpha_{12, 34}^{\frac{3}{2} - \nu} \Theta(\alpha_{12} - \alpha_0) \Theta(\alpha_{34} - \alpha_0) + 2 \text{ perms}
\]

(3.21)

As for the bispectrum, we included the Heaviside step functions in order to restrict to the collapsed limit. We will find \( \alpha_0 = 2 \) to be the cutoff value that optimises the signal-to-noise ratio for these templates. Furthermore, the intermediate template is normalized as to match the \( \mathcal{P} \) shape in the tetrahedral configuration and the amplitude of the oscillatory shape in the collapsed limit is normalized to match the intermediate shape for the mass limit \( \nu = \mu = 0 \).
3.2 Secondary non-Gaussianity

As mentioned before, by the time we enter the Dark Ages non-linear evolution of the density field will generate non-Gaussianity even if we start out with a perfectly Gaussian field. Furthermore, the fact that 21-cm temperature fluctuations depend non-linearly on the underlying fields \( \delta_b, \delta_v \) and \( \delta_T \) causes additional non-Gaussian imprints in the 21-cm signal. Since we are primarily interested in extracting (possible very weak) primordial non-Gaussianity from this signal, such secondary non-Gaussianity can be considered to be a confusion signal that should be accurately modeled and subtracted in order to adequately determine the primordial contribution. The magnitude and impact of these secondary non-Gaussianities were previously studied for the bispectrum in [11, 14]. There, it was found that the secondary contribution is several orders of magnitude stronger than the primordial signal. Including the amplitude of the secondary bispectrum as nuisance parameters and marginalising over them reduces the signal-to-noise ratio of the primordial amplitude. In this work we improve upon some of the assumptions made in those previous works, by including the effects of electron fraction fluctuations and baryonic pressure. The former has been studied in section 2.2 and its effect is captured in the coefficients \( C_{ij}^T \). The latter will be studied next. After having set up the framework for higher order perturbation theory with baryonic pressure, we will be able to derive both the secondary bispectrum as well as for the first time the secondary trispectrum of 21-cm temperature fluctuations during the Dark Ages.

3.2.1 Perturbation theory including baryonic pressure

By means of equation (2.45), we have direct perturbative relation between 21-cm fluctuations and baryon density contrast and velocity divergence up to third order. However, we have so far been agnostic about the functional form of \( \delta_b^{(n)} \) and \( \theta_b^{(n)} \). In earlier work [11, 14], two assumptions are made about the functional form. First, it assumed that the time-dependence of the baryonic fluctuations is identical to that of CDM fluctuations, so that the \( n \)-th order perturbations can be written as [34]:

\[
\delta^{(n)}_b(z, k) = a^n(z) \delta_{n,b}(k), \quad \theta^{(n)}_b(z, k) = \mathcal{H}(z) a^n(z) \theta_{n,b}(k).
\] (3.22)

Note that the functional form is product separable in temporal and spatial (momentum) dependence. The latter is contained in \( \delta_n \) and \( \theta_n \), which in standard perturbation theory read:

\[
\delta_{n,b}(k) = \int_{q_1 \cdots q_n} (2\pi)^3 \delta_D(k - q_1 \cdots q_n) F_{n,b}(q_1, \ldots, q_n) \delta_1(b(q_1) \cdots \delta_{1,b}(q_n)) \] (3.23)

\[
\theta_{n,b}(k) = \int_{q_1 \cdots q_n} (2\pi)^3 \delta_D(k - q_1 \cdots q_n) G_{n,b}(q_1, \ldots, q_n) \delta_{1,b}(q_1) \cdots \delta_{1,b}(q_n), \] (3.24)

where \( q_1 \cdots n \equiv q_1 + \cdots + q_n \) and the baryonic momentum kernels \( F_{n,b} \) and \( G_{n,b} \) are symmetric under interchanging the momenta. The second assumption made in previous works, is that the baryonic kernels are equivalent to that of CDM so that they can be replaced by the latter: \( F_{n,b} \rightarrow F_{n,c} \) and \( G_{n,b} \rightarrow G_{n,c} \), which can trivially be obtained from a simple recursion relation [34].

However, in reality baryons are very different from CDM in their physical nature. Where CDM is pressureless, baryons constitute pressure. In particular, on scales comparable to
the so-called baryonic Jeans scale $\lambda_J$, the pressure becomes considerable and suppresses the growth of fluctuations due to gravity. On scales smaller than the Jeans scale, pressure competes with gravity to generate acoustic sound waves and the notion of growing density fluctuations becomes ill-defined.

To take the first step in the direction of accounting for pressure effects, we use the formalism for baryonic fluctuations constructed in [35].\(^9\) Within this formalism, the Jeans scale is assumed to be constant, i.e. $\lambda_J \neq \lambda_J(z)$. Although strictly speaking this assumption becomes incorrect once the gas temperature decouples from the photon temperature, it has the convenient consequence that the time-dependence of baryonic fluctuations is still equivalent to that of CDM, $\delta_b^{(n)} \propto a^n$, and the results of section 2.3 are still applicable. Then, the baryonic fields $\delta_b$ and $\theta_b$ can be expanded as:

\[
\delta_b(z, k) = \sum_{n=1}^{\infty} a^n(z) g_n(k) \delta_{n,c}(k), \quad \theta_b(z, k) = \sum_{n=1}^{\infty} \mathcal{H}(z) a^n(z) h_n(k) \theta_{n,c}(k),
\]

(3.25)

where we have defined the filtering functions:\(^{10}\)

\[
g_n(k) = \frac{\delta_{n,b}(k)}{\delta_{n,c}(k)}, \quad h_n(k) = \frac{\theta_{n,b}(k)}{\theta_{n,c}(k)}.
\]

(3.26)

Note that while we solely require $g_1$ at first order (as the linear density contrast and velocity divergence are trivially related as $\theta_b = -\mathcal{H}\delta_b$), at higher order we require different filtering functions for $\delta_b$ (i.e. $g_n$) and $\theta_b$ ($h_n$).

We expect the filtering functions to be strongly decreasing for scales around and beyond the Jeans scale ($k/k_J \gg 1$), as pressure suppresses the evolution of baryonic density fluctuations relative to those in CDM on these scales. In addition, far away from the Jeans scale ($k/k_J \ll 1$), we assume the behaviour of baryons and CDM to be equivalent, i.e. $g_n, h_n \to 1$.

At first order, the filtering function is given by:

\[
g_1(k) = \frac{1}{1 + k^2/k_J^2},
\]

(3.27)

which indeed contains the limiting behavior described above. At second order, the function $g_2$ can be written in terms of $g_1$ as:

\[
g_2(k) = \sigma(k) \left[ 1 + \frac{7}{3} \frac{\delta_{2,c}(k)}{\delta_{2,c}(k)} \right], \quad \sigma(k) = \frac{1}{10/3 + (k/k_J)^2},
\]

(3.28)

where $\delta_{2,c}$ is given by (the prime does not denote a time derivative):

\[
\delta_{2,c}(k) \equiv \int_{q_1,q_2} (2\pi)^3 \delta_{10}(k - q_{12}) 
\cdot \mathcal{F}_2^{(a)}(q_1,q_2) \delta_{1,c}(q_1) \delta_{1,c}(q_2).
\]

(3.29)

---

\(^9\)We will only highlight the required essentials of the formalism, for details, we refer to [35].

\(^{10}\)In principle, these filtering functions depend on time. Therefore, they are often defined as: $g_n(z, k) = \delta_b^{(n)}(z, k)/\delta_b^{(n)}(z, k)$, including time dependence. However, as shown in [35], when decaying modes are omitted and the Jeans scale is taken to be constant, the filtering functions become time-independent as well, justifying this definition.
The modified kernel $\tilde{J}_2^{(s)}(q_1, q_2)$ (where the superscript $(s)$ denote the symmetrization of the kernel) reads:

$$\tilde{J}_2^{(s)}(q_1, q_2) \equiv F_2^{(s)}(q_1, q_2) + \frac{3}{14} \frac{k^2}{k_1^2} g_1(q_1)g_1(q_2), \quad (3.30)$$

which reduces to the CDM kernel $F_2^{(s)}(q_1, q_2)$ in the limit $k/k_1 \ll 1$. The function $h_2(k)$ can be written as:

$$h_2(k) = \frac{1}{\theta_{2, c}(k)} \int_{q_1, q_2} (2\pi)^3 \delta_D(k - q_{12}) \left[ 1 + \vartheta(q_1, q_2) \right] \delta_1, b(q_1)\delta_1, b(q_2) - 2\frac{\delta_2, b(k)}{\theta_{2, c}(k)}, \quad (3.31)$$

where we have defined $\vartheta(q_1, q_2) \equiv (q_1 \cdot q_2)(q_1^2 + q_2^2)/2q_1^2q_2^2$. In the limit $k/k_1 \to 0$, we obtain $h_2 \to 1$ as expected.

The higher order perturbations $\delta_b^{(3)}, \theta_b^{(3)}$ that show up in the secondary trispectrum will also be modified as compared to the CDM expressions. However, as we will see later, when comparing forecasts of the bispectrum with and without baryonic pressure effects, the differences are small. Hence, for computational simplicity we will not include baryonic pressure in our forecasts for the trispectrum. With the above perturbation expansions, we are now able to write down the secondary contributions to the non-Gaussianity of the 21-cm signal.

### 3.2.2 Secondary bispectrum

The lowest order non-Gaussian statistic of 21-cm brightness temperature fluctuations, i.e. the 21-cm bispectrum $B_{ST}$, now consists of a primary, primordial contribution:

$$\langle \delta T_{21}^{(1)}(k_1)\delta T_{21}^{(1)}(k_2)\delta T_{21}^{(1)}(k_3) \rangle' = \left( \prod_{i=1}^{3} c_i^{(1)}(k_i)M_b(k_i) \right) \times B_c(k_1, k_2, k_3) \quad (3.32)$$

where $M_b(k)$ is the linear transfer function for baryonic fluctuations. Moreover, there is also a higher order, secondary contribution:

$$\langle \delta T_{21}(k_1)\delta T_{21}(k_2)\delta T_{21}(k_3) \rangle_{sec.}' = \langle \delta T_{21}^{(1)}(k_1)\delta T_{21}^{(1)}(k_2)\delta T_{21}^{(2)}(k_3) \rangle' + 2\text{ perms.}$$

$$= c_1^{(1)}(k_1)c_1^{(1)}(k_2)c_2^{(2)}(k_3) \times 2\tilde{J}_2^{(s)}(k_1, k_2)P_c(k_1)P_c(k_2)$$

$$+ c_1^{(1)}(k_1)c_3^{(2)}(k_3, -k_1) + c_3^{(2)}(k_3, -k_2) \left[ \frac{g_1^2(k_1)g_1^2(k_2)P_c(k_1)P_c(k_2)}{3} \right]$$

$$+ 2\text{ perms.} \quad (3.33)$$

We have defined the modified kernels $\tilde{J}_2^{(s)}$ and $\tilde{\gamma}_2^{(s)}$ as:

$$\tilde{J}_2^{(s)}(k_1, k_2) \equiv g_1(k_1)g_1(k_2)\sigma_2(|k_{12}|) \left[ F_2^{(s)}(k_1, k_2) + \frac{7}{3} F_2^{(s)}(k_1, k_2) \right],$$

$$\tilde{\gamma}_2^{(s)}(k_1, k_2) \equiv 2g_1(k_1)g_1(k_2)\sigma_2(|k_{12}|) \left[ F_2^{(s)}(k_1, k_2) + \frac{7}{3} F_2^{(s)}(k_1, k_2) \right]$$

$$- g_1^2(k_1)g_1^2(k_2)[1 + \vartheta(k_1, k_2)], \quad (3.34)$$

In the limit $\lambda_J \to \infty$ and $P_c \to P_b$, this restores to the previously derived secondary bispectrum of 21-cm brightness temperature fluctuations [11, 14].

---

11See Appendix B of [35], Eqs. B23 and B25.
3.2.3 Secondary trispectrum

The secondary trispectrum is composed of 11 different contributions, which can be summarized schematically as:

\[
\langle \delta T_{21}(k_1)\delta T_{21}(k_2)\delta T_{21}(k_3)\rangle_{\text{sec}} = T_{\delta_1\delta_1\delta_2\theta_2} + T_{\delta_1\delta_1\delta_2\theta_2} + T_{\delta_1\delta_1\delta_2\theta_2} \\
+ T_{\delta_1\delta_1\delta_2\delta_1^2} + T_{\delta_1\delta_1\delta_2\theta_2^2} + T_{\delta_1\delta_1\delta_2\theta_2} \\
+ T_{\delta_1\delta_1\delta_2\delta_1^3} + T_{\delta_1\delta_1\delta_2\delta_1^2} + T_{\delta_1\delta_1\delta_2\theta_2} \\
+ T_{\delta_1\delta_1\delta_2\delta_3} + T_{\delta_1\delta_1\delta_2\delta_3},
\]

where for example

\[
T_{\delta_1\delta_1\delta_2\theta_2} = f(k_1, k_2, k_3, k_4)\langle \delta_b(1)(k_1)\delta_b(1)(k_2)\delta_b(2)(k_3)\delta_b(2)(k_4)\rangle + 11 \text{ perms}.
\]

Here the function \( f \) is composed of the coefficients \( c_i^{(j)} \) introduced above. The explicit expressions for all 11 contributions can be found in Appendix C.

4 The information content of the Dark Ages

In this section we will determine the amount of information in 21-cm brightness temperature fluctuations during the Dark Ages available to constrain primordial non-Gaussianity.\(^{12}\) First we will determine the total amount of information available when neglecting any form of signal confusion (either instrumental, systematic or secondary), giving an ultimate upper bound for the sensitivities that can be achieved. Subsequently, using our improved model of the 21-cm signal, we determine the amount of information when including signal confusion due to secondary non-Gaussianities. Finally, we will determine the sensitivity of different experimental scenarios.

4.1 Total primordial information content

In order to determine the total information content, we will assume a cosmic variance limited experiment of the Dark Ages between \( 30 \leq z \leq 100 \) and take the largest scale to be \( k_{\text{min}} = 0.01 \) Mpc\(^{-1} \), since on larger scales one has to include fluctuations in the CMB temperature and non-relativistic perturbation theory breaks down.\(^{13}\) When including the secondary bispectrum and trispectrum, following [36] we consider 14 redshift bins (i.e. \( \Delta z = 5 \)), assuming that they are sufficiently uncorrelated such that the covariance matrix can be taken to be diagonal in redshift space. More details about the Fisher analysis and how to evaluate the Fisher matrix for both the bispectrum and trispectrum, we refer to Appendix B.

The results for the common primordial bispectra are summarized in Table 2. The sensitivity to the amplitude of the primordial bispectrum is given on the diagonal. The increased sensitivity to the local shape as compared to the other two common shapes, can be understood as due to a slightly enhanced scaling with \( k_{\text{max}} \) [15]. Furthermore, on the off-diagonal we summarize the overlap coefficient between the common primordial bispectra.

---

\(^{12}\)The code used to perform the numerical calculations is openly available on GitHub via this link.

\(^{13}\)The survey volume sets the largest scale to be only slightly bigger at \( k = 0.005 \) Mpc\(^{-1} \). Hence, from an observational perspective one gains very little from going beyond our choice of \( k_{\text{min}} \).
Table 2. Diagonal: sensitivity to the amplitude of the common primordial shapes. Off-diagonal: cosine (B.18) between common primordial shapes. We assume a cosmic variance limited experiment between $30 \leq z \leq 100$ and $0.01 \leq k \leq 300 \text{ Mpc}^{-1}$.

|         | $f_{\text{loc}}^\text{NL}$ | $f_{\text{equil}}^\text{NL}$ | $f_{\text{ortho}}^\text{NL}$ |
|---------|-----------------------------|-------------------------------|-------------------------------|
| $f_{\text{loc}}^\text{NL}$ | $4.22 \times 10^{-6}$       | 0.29                          | -0.33                         |
| $f_{\text{equil}}^\text{NL}$ | $2.88 \times 10^{-5}$       | 0.19                          | 0.33                          |
| $f_{\text{ortho}}^\text{NL}$ |                               | 2.10 $\times 10^{-5}$        |

We now consider the massive scalar exchange template for $m \geq 3H/2$ (i.e. equation (3.9)). As mentioned, the authors of [12] suggested the use of a cutoff $\alpha_0 = 10$ in order to remove overlap with the equilateral shape equation (3.6). Such a cutoff also drastically reduces the amount of triangles available to observe, hence reducing the signal-to-noise ratio. Using the Fisher analysis we are able to determine the amount of overlap between the clock template and the common primordial shapes that we aim to look for in future experiments. In Figure 5a we present this overlap as a function of $\mu$ for both the cutoff $\alpha_0 = 10$ and $\alpha_0 = 2$. We find that although the overlap with the equilateral shape is almost completely removed for the former case, the overlap with the other shapes is increased for some mass values compared to the latter. In 4a we show the sensitivity to the amplitude of the clock template, which clearly shows that a larger cutoff is favorable only for masses $\mu \lesssim 1$. A similar observation can be made for the intermediate template (Figure 5b and 4b) where we again find $\alpha_0 = 2$ to optimise the sensitivity except for some small improvement with $\alpha_0 = 10$ when $\nu \gtrsim 1$. Hence, in what follows we will take $\alpha_0 = 2$.

Figure 4. Sensitivity to the amplitude of the massive exchange templates in a cosmic variance limited experiment between $30 \leq z \leq 100$ and $0.01 \leq k \leq 300 \text{ Mpc}^{-1}$, before (dashed) and after (solid) marginalising over the common primordial bispectra, for different cutoff values $\alpha_0$. Left: Clock template, equation (3.9). Right: Intermediate template, equation (3.10)
Figure 5. Cosine (B.18) between the massive exchange templates and common primordial shapes for different mass in a cosmic variance limited experiment between $30 \leq z \leq 100$ and $0.01 \leq k \leq 300 \text{ Mpc}^{-1}$. Solid lines correspond to the template with $a_0 = 2$. Dashed lines correspond to the template with $a_0 = 10$. Left: Clock template equation (3.9). Right: Intermediate template equation (3.10).

Proceeding to the trispectrum, in Table 3 (on the diagonal) we present the ultimate sensitivity of 21-cm temperature fluctuations from the Dark Ages to the amplitude of the local and equilateral trispectra. We find the $\tau_{\text{loc}}$ shape to be much better constrained than the other shapes. This can again be understood as due to an enhanced scaling [15], which we will comment on later in this subsection. The overlap between these primordial trispectra are given by the off diagonals, where we see that the enhanced scaling of the $\tau_{\text{loc}}$ shape makes it completely orthogonal to the other shapes, making it a very clean signal to look for.

| $\tau_{\text{loc}}$ | $\sigma_{\text{loc}}$ | $\sigma_{\text{eq}}^4$ | $\sigma_{\text{loc}}^4 (\sigma R)^4$ | $\sigma_{\text{loc}}^6 (\sigma R)^6$ |
|-----------------|-----------------|-----------------|-------------------------------|-----------------|
| 1.79 $\times$ 10$^{-7}$ | 0.00 | 0.00 | 0.00 | 0.00 |
| $g_{\text{loc}}$ | 0.02 | 0.13 | 0.11 | 0.06 |
| $g_{\text{eq}}^4$ | 0.229 | 0.95 | 0.65 | 0.254 |
| $(\sigma R)^4$ | 0.254 | 0.82 | 0.65 |
| $g_{\text{loc}}^6$ | 0.254 | 0.82 | 0.65 |

Table 3. Diagonal: sensitivity $\sigma$ to the amplitude of primordial trispectra. Off-diagonal: cosine (B.18) between common primordial trispectra. We assume a cosmic variance limited experiment between $30 \leq z \leq 100$ and $0.01 \leq k \leq 300 \text{ Mpc}^{-1}$.

The ultimate sensitivities for the massive exchange trispectra are given in Figure 6. Although it’s clear that the clock trispectrum does not outperform the clock bispectrum, we see that the intermediate template enjoys an enhanced scaling for masses $\nu > 3/4$. Such a scaling can improve the sensitivity to the trispectrum past the sensitivity to the bispectrum. However, for $\nu$ close to $3/2$, the shape acquires significant overlap with the $\tau_{\text{loc}}$ shape, which drastically reduces the sensitivity after marginalization making the sensitivity on par to that of the bispectrum. The overlap of the exchange trispectra with the local and equilateral trispectra is summarized in Figure 7, which shows how the intermediate template interpolates between the local and equilateral shapes depending on the mass $\nu$. 
The results presented in this subsection give an indication of how well 21-cm brightness temperature fluctuations from the Dark Ages are able to constrain primordial cosmology if we were able to harvest all of its information down to the Jeans scale $k_{\text{max}} = 300 \text{ Mpc}^{-1}$, demonstrating the ability of 21-cm cosmology to achieve the exquisite sensitivity necessary to advance our understanding of the primordial universe, possibly even paving a way towards the Cosmological Collider as noticed previously by [12].

4.2 Including secondary non-Gaussianity

Although 21-cm brightness temperature fluctuations are still a relatively pristine tracer of the underlying density field sourced by inflation, non-linear evolution such as due to gravitational collapse or velocity effects, will make even perfectly Gaussian fluctuations non-Gaussian over time, and between the surface of last scattering at $z \approx 1100$ and the Dark Ages $z \approx 200$ there is a considerable amount of time for such secondary non-Gaussianities to be generated. When trying to extract primordial non-Gaussianity from 21-cm observations, such secondary effects enter as confusion signals, introducing extra nuisance parameters that should be modeled.
and accounted for when analyzing data. Secondary non-Gaussianity of the 21-cm signal has already been studied in the context of the bispectrum before [11, 14]. In the previous sections we have improved upon some of the assumptions that went into those works in order to more accurately model the secondary bispectrum, as well as to extend this to the secondary trispectrum. In this subsection we will study the impact of secondary non-Gaussianity on the constraining power of the 21-cm signal from the Dark Ages.

The amount of additional nuisance parameters depends on how accurately we are able to model the secondary non-Gaussianity. Most optimally, we might be able to model the internal parameters of secondary non-Gaussianity ($\bar{T}_{21}$ and $\alpha_{i,j}$ in 2.4) to such precision that the only free parameter of the secondary $N$-point function is its amplitude. In that case the secondary $N$-point function can be treated as a single shape. More realistically, we obtain a best fit for the internal parameters such that their residuals become the free parameters of the secondary correlation function [14]:

$$F_{\delta T}(k_1, ..., k_N) = F_{\delta T}^{sec,0}(k_1, ..., k_N) + \sum_i \Delta A_i \frac{\partial F_{\delta T}^{sec}}{\partial A_i}$$

such that we end up with an individual shape $\frac{\partial F_{\delta T}^{sec}}{\partial A_i}$ for every internal parameter. Most pessimistically, one could consider every independent shape present in the secondary $N$-point function. Due to the large amount of shapes (already 21 for the bispectrum), calculating the full Fisher matrix quickly becomes intractable for higher $N$-point functions. In the following we will adopt the best fit approach of [14], giving rise to a four parameter secondary bispectrum and seven parameter secondary trispectrum. We have improved upon previous work by including free electron fraction perturbations to model the internal parameters of secondary non-Gaussianity, and we include baryonic pressure (in the bispectrum).

We quantify the impact of including secondary non-Gaussianity on the signal-to-noise ratio using the signal-to-noise degradation factor (SND) defined in equation (B.17). It’s interesting to note that although the signal-to-noise degradation at each redshift might be relatively large, when co-adding the redshifts the degradation is significantly reduced. This can be understood due to the difference in how the secondary shapes scale with $z$ as compared to the primordial shapes: going from equation (B.1) to equation (B.3), we see that the amount of signal from the primordial shape does not depend on $z$ whereas the secondary shapes (including off-diagonal contributions to the Fisher matrix) do. Hence, when adding information from redshift slices, one can also start distinguishing the primordial and secondary shapes by their redshift dependence.

In Figure 8 we plot the amount of degradation as a function of $k_{\text{max}}$ for the primordial bispectra. We find that including baryonic pressure slightly worsens the signal-to-noise degradation as compared to the simplified approach of [14], but not by much. We expect the same to be true for the trispectrum, so in order to keep the trispectrum calculations more tractable and efficient, we neglect baryonic pressure effects in the upcoming trispectrum analysis. Nevertheless, in an actual data analysis one should model the bi- and trispectrum as accurately as possible, in order to reliably extract the primordial signal. Furthermore, we find the four parameters of the secondary bispectrum to contribute evenly to the signal-to-noise degradation (meaning their cosine (B.18) with $f_{\text{NL}}$ is similar).
Figure 8. The signal-to-noise degradation factor after marginalising over the 4 parameter secondary bispectrum as a function of $k_{\text{max}}$ for different primordial bispectra. Note the different scales of the vertical axes.

We summarize the sensitivity to the amplitude of the common primordial bispectra as a function of $k_{\text{max}}$ in Figure 9, clearly showing the small loss of signal-to-noise due to the secondary bispectrum. Similar results are presented for the clock and intermediate template in Figure 10 and Figure 11 respectively. On the left side of figures we see that once overlap with the common primordial shapes is taken into account, the additional effect of the secondary bispectrum (difference between dotted and solid line in the figures) becomes negligible. We thus conclude that for the exchange templates, the overlap with other primordial shapes is more severe than any overlap with secondary shapes.

Figure 9. The sensitivity to the amplitude of the common primordial bispectra as a function of $k_{\text{max}}$ before (dashed) and after (solid) marginalising over the 4 parameter secondary bispectrum including baryonic pressure effects. The solid green line coincides with the dashed orange line.
Figure 10. *Left:* sensitivity to the amplitude of the bispectrum clock template with $\mu = 0.1$ as a function of $k_{\text{max}}$ without marginalising (dashed), after marginalising over the 4 parameter secondary bispectrum (dashed-dotted), after marginalising over the common primordial bispectra (dotted) and after marginalising over both secondary and common primordial bispectra (solid). *Right:* sensitivity to the amplitude of the clock template for different $\mu$ as function of $k_{\text{max}}$ after marginalising over both secondary and common primordial bispectra.

Figure 11. *Left:* sensitivity to the amplitude of the bispectrum intermediate template with $\nu = 1.4$ as a function of $k_{\text{max}}$ without marginalising (dashed), after marginalising over the 4 parameter secondary bispectrum (dashed-dotted), after marginalising over the common primordial bispectra (dotted) and after marginalising over both secondary and common primordial bispectra (solid). *Middle:* sensitivity to the amplitude of the int template for different $\nu$ as function of $k_{\text{max}}$ after marginalising over secondary and common primordial bispectra. *Right:* middle plot zoomed in on $100 \leq k_{\text{max}} \leq 300$ to distinguish between the various masses.

We now investigate the impact of the secondary trispectrum on the previously obtained sensitivities to primordial trispectra. In Figure 12 we summarize the signal-to-noise degradation after marginalising over the seven residual parameters of the secondary trispectrum. Due to the more complex shapes of the trispectra in $k$-space as compared to the bispectra, we see little overlap with the secondary trispectra, especially for higher values of $k_{\text{max}}$. The small overlap is reflected in the small loss of sensitivity to the amplitude of the common shapes in Figure 13. As for the bispectrum, we find similar overlap for all parameters. Results for
Figure 13. The sensitivity to the amplitude of the common primordial trispectra as a function of $k_{\text{max}}$ before (dashed) and after (solid) marginalising over the 7 parameter secondary trispectrum. Green and red lines coincide. The enhanced scaling of the $\tau_{NL}^\text{loc}$ shape is clearly visible.

Figure 12. The signal-to-noise degradation factor after marginalising over the 7 parameter secondary trispectrum as a function of $k_{\text{max}}$ for different primordial trispectra. Note the different scales of the vertical axes.

the clock and intermediate template are summarized in Figure 14 and Figure 15 respectively. Similar to the case of the bispectrum, in the left figures we see that when including the overlap of the common primordial trispectra, the additional effect of the secondary trispectra is negligible. What is different from the case of the bispectrum is the enhanced scaling for the intermediate template at mass values $\nu > 0.75$. 
4.3 Forecasting experiments

Until now we have studied the signal-to-noise with which we can measure primordial non-Gaussianity as a function of the maximum accessible momentum $k_{\text{max}}$. In reality however, there is a difference between the radial (or redshift) component of the momenta along the line-of-sight $k_{\parallel}$ and the perpendicular (or angular) component $k_{\perp}$. The resolution along these components is set by different properties of the experimental setup. The line-of-sight resolution is determined by our ability to distinguish redshift slices, which in practise corresponds the size of the frequency bins set by the window size $\delta \nu$ [36]:

$$k_{\text{max}} \approx \sqrt{\frac{17}{3}} \frac{1}{20 \delta \nu \sqrt{1 + z}}$$  \hspace{1cm} (4.2)
with $\delta \nu$ in units of MHz. Furthermore, the angular resolution $R$ is set by the baseline $b$ (the distance between two receivers in an interferometer) of the experiment through:

$$R = \frac{\lambda_{\text{obs}}}{b}$$

(4.3)

which yields the maximum perpendicular mode

$$k_{\text{max}}^\perp \approx \frac{2\pi \nu_0 b}{d(z)(1 + z)c}$$

(4.4)

with $b$ in units of km and where $\nu_0 \approx 1420 \times 10^6$ Hz is the frequency of the 21-cm signal, $c$ the speed of light in km/s and $d(z)$ the comoving distance to redshift $z$ in units of Mpc. In practice it is much easier to improve on the window $\delta \nu$ than it is to increase the baseline. In Figure 16 we show $k_{\text{max}}$ as a function of baseline and window size at several redshifts.

![Figure 16](image)

**Figure 16.** The maximal wavenumber $k_{\text{max}}$ that can be probed in a certain direction for several redshifts. **Left:** perpendicular $k_{\text{max}}^\perp$ as a function of baseline in kilometers. **Right:** parallel (line-of-sight) $k_{\text{max}}^\parallel$ as a function of window size $\delta \nu$.

Figure 17 shows the forecast sensitivity to the amplitude of the three common primordial bispectra. We find that already for a reasonably sized array with a baseline of several kilometers 21-cm observations from the Dark Ages can improve current as well as future CMB constraints by orders of magnitude. Notably, $\sigma_{f_{\text{loc}}^2} \sim O(10^{-2})$ can in principle be achieved with a baseline smaller than 10 kilometers, allowing for a test of the famous Maldacena consistency condition.
Figure 17. Forecast sensitivity to the amplitude of the common primordial bispectra for different baseline and window size, before (dashed) and after (solid) marginalising over the 4 parameter secondary bispectrum.

Figure 18 summarizes the ability to detect primordial features due to the exchange of massive particles during inflation for different baseline and window size. Depending on the nature of the exchange coupling the size of the amplitude of non-Gaussian signal can be anywhere between $f_{NL} \ll 10^{-2}$ and $f_{NL} > O(1)$ (see section E. of [12] for a clear discussion on the different possibilities). We see that a baseline of $100 - 1000$ km could already provide a clear picture of the particle content of the early universe, opening up an entirely new window into the physics of inflation.

Figure 18. Forecast sensitivity to the amplitude of some massive exchange bispectra for different baseline and window, before (dashed) and after (solid) marginalising over the 4 parameter secondary bispectrum and common primordial bispectra. Top row is the Clock template, bottom row is the intermediate template.

In Figure 19 we present for the first time forecasts of the sensitivity of different experimental setups to the amplitude of primordial trispectra, taking into account the secondary trispectrum of 21-cm brightness temperature fluctuations. We find that for most trispectra it proves hard to do better than $\sigma_g \sim O(1)$, a clear exception being the $\ell_{\text{loc}}$ local shape, that exhibits enhanced scaling with $k_{\text{max}}$ resulting in sensitivities on par with that of the
bispectrum.

Figure 19. Forecast sensitivity to the amplitude of the primordial trispectra for different baseline and window size, before (dashed) and after (solid) marginalising over the 7 parameter secondary trispectrum.

Moving on to the trispectrum of massive exchange, we show the results in Figure 20. We conclude that the oscillatory clock template can realistically only be constrained up to a sensitivity of $\sigma_g \sim O(10)$, still many orders of magnitude better than CMB observations (likely $\sim O(10^4)$ in the future). However, the intermediate trispectrum of massive exchange can reach sensitivities that are much more similar to the bispectrum sensitivities, allowing one to probe the same masses through both the bispectrum and trispectrum. Although we have not explicitly included spinning particles in this work, having two independent probes of the same mass could be used to break degeneracy between the mass and spin of the particle, that both contribute to the amplitude of the signal [7].
Figure 20. Forecast sensitivity to the amplitude of some massive exchange trispectra for different baseline and window size, before (dashed) and after (solid) marginalising over the 7 parameter secondary trispectrum and local and equilateral trispectra.

5 Conclusions and Outlook

Measurements of primordial non-Gaussianity play a key role in unraveling the mysteries of the early universe. Until now, the most stringent constraints on the amount of pnG come from observations of the cosmic microwave background radiation, which are still consistent with zero. Future CMB experiments (e.g. SO and CMBS4) could improve constraints by an order of magnitude, possibly zeroing in on non-zero pnG. However, in order to accurately distinguish different non-Gaussian signatures such as the imprint of massive particles present during inflation, sensitivities are required that can simply not be achieved with CMB measurements, due to the limited amount of modes available. To advance our understanding of the primordial universe, we ought to employ additional probes. Arguably the most promising avenue to this end is the use of 21-cm brightness temperature fluctuations. During a part of the cosmological Dark Ages (between $30 \leq z \leq 200$), these tiny fluctuation can in principle be observed in absorption to the CMB radiation. The result is a large observable 3D volume of linear modes that trace the primordial initial conditions, which could constrain primordial non-Gaussianity with unprecedented precision, opening up an exciting new window into the physics of the early universe.

In order to extract such primordial information from the tracer field, it’s required to accurately model the physics of the hydrogen gas. Furthermore, the non-linear relation between the initial conditions and the 21-cm brightness temperature fluctuations as well as gravitational collapse of the matter field induce secondary non-Gaussianity, swamping to primordial signal that we aim to observe. In this paper, we (re)address some of these aspects and made several advances. Specifically, we have:

- improved analytical modelling of the hydrogen gas by including the effect of free electron fraction perturbations, which turns out to be sizable (Figure 2) and hence should be
included in any detailed analysis of 21-cm observations of the Dark Ages

- derived the perturbation expansion of the 21-cm tracer field up to third order in the underlying matter field, while including baryonic pressure effects on scales close to the Jeans scale
- thereby obtained an improved expression for the secondary bispectrum of 21-cm brightness temperature fluctuations during the Dark Ages
- derived, for the first time, the secondary trispectrum of the 21-cm tracer field

With this improved modelling of the tracer field and the newly obtained secondary bi- and trispectrum, we assess the information content of 21-cm fluctuations to constrain primordial physics, under the most optimal conditions. More specifically, we perform a Fisher analysis to determine the error bars on the amplitude of various primordial bi- and trispectra in a potential 21-cm Dark Ages experiment, marginalizing over the parameters of secondary non-Gaussianity. From this analysis we draw the following conclusions:

- although secondary non-Gaussianity swamps the primordial signal, the primordial signal can be extracted without a big loss of signal-to-noise both for the bispectrum and the trispectrum
- tomography of 21-cm fluctuations can in principle improve constraints on primordial non-Gaussianity by several orders of magnitude, already for relatively low $k_{\text{max}}$, as noted before in refs. [11, 12, 14]
- the enhanced scaling of some trispectra with $k_{\text{max}}$ [15] indeed manifests itself in 21-cm observations.
- our Fisher forecasts demonstrate the importance of marginalizing over primordial shapes as compared to secondary shapes when considering the massive particle exchange templates

Our results clearly show the potential of the 21-cm signal from the Dark Ages to improve constraints on primordial non-Gaussianity and thereby our understanding of the infant universe. Many interesting directions of research remain in order to better understand the intricacies of the 21-cm Dark Ages signal. We would like to reiterate some of the assumptions made in this work that can be improved upon as well as some other possible directions of future research:

- we have included baryonic pressure effects to the secondary bispectrum along the lines of [35]. One of the assumptions made there is the constancy of the baryonic Jeans scale $k_J$. In reality this scale depends on time (redshift) and a more accurate modelling of secondary non-Gaussianity of the 21-cm signal should include this time-dependence
- Our expression for the 21-cm trispectrum can be used to include non-Gaussian covariance in an analysis of the 21-cm power spectrum such as [36] (see [37] for a similar analysis in the context of LSS). In principle non-Gaussian covariance will also impact the forecasts presented in this work. We leave this to future work.
• By modelling the effects of the first stars’ Lyman-α radiation on the 21-cm brightness temperature, one could extend our analysis from the Dark Ages into the epoch of Cosmic Dawn. Earth-based experiments measuring the 21-cm signal from this epoch are already being prepared and could provide constraints on primordial non-Gaussianity in the nearer future [38, 39].

Finally it’s imperative to once more mention the practical and technological challenges involved in order to measure such a signal. Although the 21-cm signal from the epoch of reionization can be observed from earth, once we pass into the Dark Ages \( (z \geq 30) \), the signal has redshifted into wavelengths that are largely blocked by earth’s ionosphere, requiring us to design and build a lunar or space based observatory. Nevertheless, serious efforts are being made for establishing a lunar based observatory with future NASA and ESA moon missions [13, 40, 41]. Furthermore, a host of foreground sources obscure the pristine 21-cm signal, of which synchrotron emission in our own galaxy the most dominant and the amplitude of these foregrounds increasing towards lower frequencies (higher redshifts). In theory, foregrounds can be avoided when looking at fluctuations as the foreground mode appears initially as a line of sight mode on large scales. Interferometers then introduce additional mode coupling that leads to the famous wedge [42–45]. The future of 21-cm as a competitive cosmological probe will rely critically on whether we are capable of extracting the signal away from the wedge (although it might be possible to reconstruct some of the modes in the wedge, see e.g. [46]). However these challenges are part of the ongoing efforts in this field, both at low redshifts [47–50], reionization redshifts [38, 39, 51–53] and high redshift probes [38, 54, 55].\(^\text{14}\) Finally, we have ignored any form of noise as due to the experimental setup (e.g. thermal noise, systematics etc.) as well as any technical challenges such as integration time. We leave a more realistic forecast including these considerations to future work. The main goal of this work is to show the potential of the 21-cm signal at high redshift and to establish the importance of this probe as a way to explore the nature of the early universe in the future. Reaching sensitivities that allow us to answer questions related to new particles and new fields will be limited by the number of available modes [15] and the epoch of the dark ages in principle contains the largest number of these pristine modes.

Acknowledgments

The authors thank Matteo Biagetti, Alba Kalaja, Eiichiro Komatsu and Guilherme Pimentel for insightful discussions. We thank Julian Muñoz and Moritz Münchmeyer for discussions and providing feedback on the manuscript. We thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster. T.F is supported by the Fundamentals of the Universe research program within the University of Groningen. P.D.M acknowledges support from the Netherlands organization for scientific research (NWO) VIDI grant (dossier 639.042.730). L.V.E.K acknowledges the financial support from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 884760, "CoDEX").

\(^\text{14}\) Besides challenges related to signal extraction for the very large interferometers there are also computational challenges which will have to be addresses due to the shear number of baselines that can be correlated, see e.g. [50] and references therein.
A Evolution of perturbations

In this appendix, we will provide the explicit evolution equations for perturbations in the free electron fraction and gas temperature, used to numerically solve for the coupling coefficients $C^x_{i,j}$ (defined in equation (2.29) and equation (2.30)).

Free electron fraction

To obtain the evolution equations for $\delta^x_n$, we substitute the expansions for $\delta_x$ and $\delta_T$ (equation (2.28)) into the evolution equation for $\delta^x_{ex}$ (equation (2.26)). Order by order, we then obtain the following evolution equations:

\[ \dot{\delta}_1^T = -\Gamma_R (\delta_1^T + A_1 \delta_1^T + \delta_1), \quad (A.1) \]
\[ \dot{\delta}_2^T = -\Gamma_R \left[ \delta_2^T + A_2 \delta_2^T + \delta_2 + (\delta_1^T)^2 + 2 \delta_1^T \delta_1 + A_1 \delta_1^T (\delta_1 + \delta_1^T) + A_2 (\delta_1^T)^2 \right], \quad (A.2) \]
\[ \dot{\delta}_3^T = -\Gamma_R \left[ \delta_3^T + A_3 \delta_3^T + \delta_3 + 2 \delta_1^T \delta_1 + 2 (\delta_1^T)^2 \delta_1 + 2 (\delta_1^T)^2 + (\delta_1^T)^2 \right] + 2 A_1 (\delta_1^T \delta_2^T + \delta_2^T \delta_1^T) + A_1 \delta_1^T \delta_1^T (\delta_1^T + 2 \delta_1) + A_1 (\delta_1^T \delta_2 + \delta_2^T \delta_1) + A_2 (\delta_1^T)^2 (\delta_1 + 2 \delta_1^T) + 2 A_2 \delta_1^T \delta_2^T + A_3 (\delta_1^T)^3. \quad (A.3) \]

Gas temperature

Similarly, we obtain the evolution equations for $\delta_n^T$ by substituting equation (2.28) into equation (2.20) to get:

\[ \dot{\theta}_1^T = \Theta_1, \quad (A.4) \]
\[ \dot{\theta}_2^T = \Theta_2 + \frac{2}{3} H \delta_1 (\delta_1^T - \delta_1) - \Gamma_C \delta_1^T \delta_1^T, \quad (A.5) \]
\[ \dot{\theta}_3^T = \Theta_3 + \frac{2}{3} H \delta_1 \left[ (\delta_1^T)^2 + \delta_2^T - 3 \delta_2 \right] + \frac{2}{3} H \left[ 2 \delta_2 - (\delta_1^T)^2 \right] \delta_1^T - \Gamma_C [\delta_1^T \delta_1^T + \delta_1^T \delta_1^T], \quad (A.6) \]

where we have used that $\dot{\delta}_n = n H \delta_n$ and defined $\Theta_n$ as:

\[ \Theta_n \equiv \Gamma_C \left[ \left( T_\gamma / T_{\text{gas}} - 1 \right) \delta_n^T - \left( T_\gamma / T_{\text{gas}} \right) \delta_n^T \right] + \frac{2}{3} H n \delta_n. \quad (A.7) \]

Coupling coefficients

Note that the evolution equations for $\delta_n^T$ are coupled linear differential equations, we can turn them into coupled differential equations for the coupling coefficients. Compacty, the evolution equations for the couplings can be written as:

\[ \frac{d}{da} C_{n,m}^a = \Delta_{n,m}^a + \Sigma_{n,m}^a, \quad (A.8) \]

in terms $a \equiv (1 + z)^{-1}$ as evolution variable. In the compact notation above, $\Delta_{n,m}^a$ denote all terms contributing to the evolution at the order $n$, whereas $\Sigma_{n,m}^a$ encode the (combined) contribution of lower order perturbations. Explicitly, $\Delta_{n,m}^a$ are defined as:

\[ \Delta_{n,m}^T \equiv - \left[ n + \frac{\Gamma_C}{H} \frac{T_\gamma}{T_{\text{gas}}} \right] \frac{C_{n,m}^T}{a} + \frac{\Gamma_C}{H} \left[ \frac{T_\gamma}{T_{\text{gas}} - 1} \right] \frac{C_{n,m}^T}{a} + \frac{2n}{3a} \delta_{n,m}, \quad (A.9) \]
\[ \Delta_{n,m}^x \equiv - \left[ n + \frac{\Gamma_R}{H} \right] \frac{C_{n,m}^x}{a} - \frac{1}{a} \frac{\Gamma_R}{H} \left[ A_1 C_{n,m}^T + \delta_{n,m} \right]. \quad (A.10) \]
By construction, the evolution of the coefficients $C_{n,m}^a$ is only sourced by $C_{n,m}^a$, so that $\Sigma_{n,m}^a = 0$. The cases $n \neq m$ are non-zero and read:

$$\Sigma_{1,1}^T = -\frac{1}{a} \Gamma C \alpha \beta C^T_{1,1},$$

$$\Sigma_{3,1}^T = -\frac{1}{a} \Gamma C \frac{2}{3a} \left[ C_{1,1}^T + C_{1,2}^T \right],$$

$$\Sigma_{3,2}^T = -\frac{1}{a} \Gamma C \frac{2}{3a} \left[ C_{1,1}^T + C_{1,2}^T \right] - 2C_{1,1}^T C_{1,2}^T + 2C_{1,2}^T C_{1,1}^T,$$

$$\Sigma_{3,3}^T = -\frac{1}{a} \Gamma R \left[ C_{1,1}^T (C_{1,1}^T + 1) + A_1 C_{1,1}^T (1 + 2C_{1,1}^T) - 2C_{1,2}^T C_{1,1}^T, 2A_1 C_{1,1}^T + A_2 C_{1,1}^T \right],$$

$$\Sigma_{3,1}^x = -\frac{1}{a} \Gamma R \left[ 2C_{1,1}^T (1 + 2C_{1,1}^T) + 2C_{1,3}^T (1 + 2C_{1,1}^T) + A_1 C_{1,1}^T (1 + 2C_{1,1}^T) + A_1 C_{1,2}^T (1 + 2C_{1,1}^T) + A_2 C_{1,1}^T \right],$$

$$\Sigma_{3,2}^x = -\frac{1}{a} \Gamma R \left[ 2C_{1,2}^T (1 + 2C_{1,1}^T) + A_1 C_{1,3}^T (1 + 2C_{1,1}^T) + A_1 C_{1,1}^T (1 + 2C_{1,1}^T) + A_2 C_{1,1}^T \right].$$

### Initial conditions

To solve the above system of equations, we set the initial conditions following the reasoning of [19]. At high redshifts ($z_{1} = 1000$) we have $T_{\text{gas}} = T_{\gamma}$ to high accuracy. Since we neglect fluctuations in the photon temperature, we conclude that at $z_{1}$ gas temperature fluctuations must vanish as well, $\delta_{T}(z_{1}, \mathbf{x}) = 0$, or equivalently $C_{n,m}^T(z_{1}) = 0$. Regarding electron fraction fluctuations $\delta_{e}$ (or equivalently $C_{n,m}^e(z_{1}) = 0$), one should in principle start computing the evolution at earlier times to get the appropriate initial conditions at the initial redshift $z_{1}$. However, we are interested in the effect of $\delta_{e}$ at late times (during the Dark Ages) on the 21-cm signal, which only enters via the coupling to gas temperature fluctuations. At late times, this coupling is insensitive to the initial value $\delta_{e}(z_{1})$, since perturbations grow after recombination and the initial value is quickly forgotten [19]. Therefore, we may set $\delta_{e}(z_{1}) = 0$, corresponding to $C_{n,m}^e(z_{1}) = 0$.

### B Fisher information matrix

We summarize the relevant equations used to perform the forecasts in this paper. Following [36] consider a cosmic variance limited survey over 14 redshift bins between $30 \leq z \leq 100$, assuming these are sufficiently uncorrelated. For the 21-cm bispectrum the (cosmic variance limited) Fisher information matrix for a redshift bin $z_{i}$ is defined by [12, 56, 57]:

$$F_{\alpha\beta}(z_{i}) = \int_{k_{1} > k_{2} > k_{3}} \frac{\partial B_{\delta T}(k_{1}, k_{2}, k_{3}, z_{i})}{\partial p_{\alpha}} \frac{(2\pi)^{6} \delta_{D}^{2}(k_{1} + k_{2} + k_{3})}{P_{\delta T}(k_{1}, z_{i}) P_{\delta T}(k_{2}, z_{i})} \frac{\partial B_{\delta T}(k_{1}, k_{2}, k_{3}, z_{i})}{\partial p_{\beta}}$$

(B.1)

where $p_{\alpha}, p_{\beta}$ are the bispectrum parameters of interest. The total Fisher information is then simply given by the sum over redshift bins. When considering primordial bispectra one sees that all $z$ and line-of-sight momentum dependence in the numerator cancel against the denominator and we find the simplified Fisher information matrix:

$$F_{\alpha\beta} = V_{\text{tot}} \int_{k_{1} > k_{2} > k_{3}} \frac{\partial B_{\zeta}(k_{1}, k_{2}, k_{3})}{\partial p_{\alpha}} \frac{(2\pi)^{6} \delta_{D}(k_{1} + k_{2} + k_{3})}{P_{\zeta}(k_{1}) P_{\zeta}(k_{2}) P_{\zeta}(k_{3})} \frac{\partial B_{\zeta}(k_{1}, k_{2}, k_{3})}{\partial p_{\beta}}$$

(B.2)
where we used $\delta_D(0) = V/(2\pi)^3$ where $V_{\text{tot}}$ is the total comoving survey volume. Now no binning of redshift is necessary. Furthermore, for (isotropic) primordial bispectra, the fisher matrix can be further simplified by integrating out the delta function, to read [58]

$$
F_{\alpha\beta} = \frac{V_{\text{tot}}}{8\pi^2} \int_{k_{\text{min}}}^{k_{\text{max}}} dk_1 \int_{k_1/2}^{k_1} dk_2 \int_{k_{\text{min}}}^{k_\perp} dk_3 \frac{\partial B_\delta^\perp(k_1, k_2, k_3)}{\partial p_\alpha} \frac{k_1 k_2 k_3}{P_\zeta(k_1) P_\zeta(k_2) P_\zeta(k_3)} \frac{\partial B_\delta^\perp(k_1, k_2, k_3)}{\partial p_\beta}
$$

(B.3)

where $k_{\text{max}} = \text{Max}(k_{\text{min}}, k_1 - k_2)$. Using this expression we determine the total amount of information in 4. Contrary to the primordial bispectrum, the 21-cm secondary bispectrum (see e.g. (3.32)) depends explicitly on the line-of-sight momentum through the angles $\mu(k) = k_\parallel/k$. Furthermore, the redshift dependence in the numerator and denominator of the Fisher matrix equation no longer cancel. Starting from equation (B.1) and integrating out the delta function we find

$$
F_{\alpha\beta}(z_i) = \frac{1}{3!} \frac{V_i}{(2\pi)^5} \int_{k_{\text{min}}}^{k_{\text{max}}} d^2 k_1 \perp d^2 k_2 \perp dk_1 \perp dk_2 \perp \frac{\partial B_\delta^\perp(k_1, k_2, k_3, z_i)}{\partial p_\alpha} \frac{1}{P_\delta^\perp(k_1, z_i) P_\delta^\perp(k_2, z_i) P_\delta^\perp(k_3, z_i)} \frac{\partial B_\delta^\perp(k_1, k_2, k_3, z_i)}{\partial p_\beta}
$$

(B.4)

where now $k_3 = -k_1 - k_2$ and the factor $1/3!$ in front is there to count every triangle configuration only once. Going to cylindrical coordinates, we can get rid of one more (angular) degree of freedom by writing $k_3^2 = k_1^2 + k_2^2 + 2k_1 k_2 \cos(\phi_{12})$ and invoking isotropy in the perpendicular direction:

$$
F_{\alpha\beta}(z_i) = \frac{V_i}{3!(2\pi)^5} \int_0^{2\pi} d\phi_{12} \int_{k_{\text{min}}}^{k_{\text{max}}} d\phi_{12} d^2 k_1 \perp d^2 k_2 \perp dk_1 \perp dk_2 \perp \frac{\partial B_\delta^\perp(k_1, k_2, k_3, z_i)}{\partial p_\alpha} \frac{1}{P_\delta^\perp(k_1, z_i) P_\delta^\perp(k_2, z_i) P_\delta^\perp(k_3, z_i)} \frac{\partial B_\delta^\perp(k_1, k_2, k_3, z_i)}{\partial p_\beta}
$$

(B.5)

We use this expression to evaluate the Fisher matrix for the 21-cm secondary bispectrum. The integration ranges can be straightforwardly modified to distinguish between $k_\parallel$ and $k_\perp$ as we do in section 4.3.

Analogous to the bispectrum, the Fisher matrix for the trispectrum is defined as:

$$
F_{\alpha\beta}(z_i) = \int_{k_1 > k_2 > k_3 > k_4} \frac{\partial T_\delta^\perp(k_1, k_2, k_3, k_4, z_i)}{\partial p_\alpha} \frac{(2\pi)^6 \delta_D^2(k_1 + k_2 + k_3 + k_4)}{P_\delta(k_1, z_i) P_\delta(k_2, z_i) P_\delta(k_3, z_i) P_\delta(k_4, z_i)} \frac{\partial T_\delta^\perp(k_1, k_2, k_3, k_4, z_i)}{\partial p_\beta}
$$

(B.6)

When considering only primordial trispectra, direction- and z-dependence again drop out. The quadrilateral can then be parametrized by the length of four sides and two diagonals and the Fisher matrix can be shown to take the following form [58]:
\[
F_{\alpha\beta} = \frac{V_\text{tot}}{4! (2\pi)^3} \int_{\mathcal{V}_T} dk_1 \, dk_2 \, dk_3 \, dk_4 \, ds \, dt \, \frac{\partial T_\xi(k_1, k_2, k_3, k_4, s, t)}{\partial \rho_\alpha} \times \frac{\partial T_\xi(k_1, k_2, k_3, k_4, s, t)}{\partial \rho_\beta} \times k_1 k_2 k_3 k_4 \delta t
\]

where \(\mathcal{V}_T\) is the tetrahedral region spanned by the quadrilateral, which can be enforced by triangle conditions on every side of the tetrahedron \([59]\). Furthermore, the function \(g_1\) is given by \([58]\):

\[
g_1 = s^2 t^2 \left( \sum_i k_i^2 - s^2 - t^2 \right) - s^2 \kappa_{23} \kappa_{14} + t^2 \kappa_{12} \kappa_{32} - (k_1^2 k_3^2 - k_2^2 k_4^2)(\kappa_{12} + \kappa_{34}) \quad (B.7)
\]

Using equation (B.7) we forecast the amount of information available to constrain the trispectrum from the Dark Ages in section 4. For the 21-cm secondary trispectrum, we again distinguish between perpendicular and line-of-sight momenta. One way to parametrize these quadrilaterals is using three of its sides:

\[
F_{\alpha\beta}(z_i) = \frac{1}{4!} \frac{V_i}{(2\pi)^3} \int d^2 k_1 \, d^2 k_2 \, d^2 k_3 \, dk_{1\parallel} \, dk_{2\parallel} \, dk_{3\parallel} \, \frac{\partial T_\delta(k_1, k_2, k_3, z_i)}{\partial \rho_\alpha} \left( \prod_{j=1}^4 \frac{1}{P_\delta(k_j, z_i)} \right) \frac{\partial T_\delta(k_1, k_2, k_3, z_i)}{\partial \rho_\beta} \quad (B.9)
\]

where it is now understood that \(k_4 = -k_1 - k_2 - k_3\). Furthermore, in cylindrical coordinates we can write

\[
k_{4\perp}^2 = k_{1\perp}^2 + k_{2\perp}^2 + k_{3\perp}^2 + 2 k_{1\perp} k_{2\perp} \cos (\phi_1 - \phi_2) + 2 k_{1\perp} k_{3\perp} \cos (\phi_1 - \phi_3) + 2 k_{2\perp} k_{3\perp} \cos (\phi_2 - \phi_3) \quad (B.10)
\]

and we can get rid of one angular degree of freedom by isotropy, leading to the following expression for the Fisher matrix:

\[
F_{\alpha\beta}(z_i) = \frac{1}{4!} \frac{V_i}{(2\pi)^3} \int_0^{2\pi} d\phi_2 \, d\phi_3 \int_{k_{\text{min}}}^{k_{\text{max}}} dk_{1\perp} \, dk_{2\perp} \, dk_{3\perp} \, \frac{1}{P_\delta(k_j, z_i)} \frac{\partial T_\delta(k_1, k_2, k_3, z_i)}{\partial \rho_\alpha} \left( \prod_{j=1}^4 \frac{1}{P_\delta(k_j, z_i)} \right) \frac{\partial T_\delta(k_1, k_2, k_3, z_i)}{\partial \rho_\beta} \quad (B.11)
\]

Using this expression we evaluate the Fisher information matrix including the 21-cm secondary trispectra. Note that the integration ranges can be modified to distinguish between \(k_{\text{max}}\) and \(k_{\text{max}}^\perp\) as we do in section 4.3.

Some of the trispectra considered in this work, such as the \(r_N^{\text{loc}}\) and intermediate shape, have a diverging behaviour in the collapsed limit, where one of the diagonals becomes very small. In order for the VEGAS integration algorithm \([60]\) to be able to properly evaluate
the Fisher matrix for these trispectra, it is more favorable to parametrize the quadrilateral in terms of one side and two diagonals instead, leading to the alternative expression:

\[ F_{\alpha\beta}(z_i) = V_i \frac{V_i}{4(2\pi)^8} \int_0^{2\pi} d\phi_s \, d\phi_t \int_{k_{\min}}^{k_{\max}} \, dk_{1\perp} \, ds_{\perp} \, dt_{\perp} \, dk_{1\parallel} \, ds_{\parallel} \, dt_{\parallel} k_{1\perp} s_{\perp} t_{\perp} \]

\[
\frac{\partial T_{ST}(k_1, k_2, k_3, k_4, z_i)}{\partial p_{\alpha}} \left( \prod_{j=1}^{4} \frac{1}{P_{ST}(k_j, z_i)} \right) \frac{\partial T_{ST}(k_1, k_2, k_3, k_4, z_i)}{\partial p_{\beta}}
\]

(B.12)

where the sides of the quadrilateral now read:

\[
k_{2\perp} = |s_{\perp} - k_{1\perp}|^2 = s_{\perp}^2 + k_{1\perp}^2 - 2s_{\perp} k_{1\perp} \cos(\phi_1 - \phi_s),
\]

\[
k_{2\parallel} = s_{\parallel} - k_{1\parallel}
\]

\[
k_{3\perp} = |k_{1\perp} - s_{\perp} - t_{\perp}|^2 = k_{1\perp}^2 + s_{\perp}^2 + t_{\perp}^2 - 2k_{1\perp} s_{\perp} \cos(\phi_1 - \phi_s) - 2k_{1\perp} s_{\perp} \cos(\phi_1 - \phi_s) - 2s_{\perp} t_{\perp} \cos(\phi_s - \phi_t)
\]

\[
k_{3\parallel} = k_{1\parallel} - s_{\parallel} - t_{\parallel}
\]

\[
k_{4\perp} = |t - k_{1\perp}|^2 = t_{\perp}^2 + k_{1\perp}^2 - 2t_{\perp} k_{1\perp} \cos(\phi_1 - \phi_t),
\]

\[
k_{4\parallel} = s_{\parallel} - k_{1\parallel}
\]

(B.13)

Once the Fisher information matrix has been evaluated, the error with which we can determine the parameter \( p_{\alpha} \) is then given by:

\[ \sigma_{p_{\alpha}} = F_{\alpha\alpha}^{-1/2} \]  

(B.14)

while the signal-to-noise is given by its inverse:

\[ \left( \frac{S}{N} \right)_{p_{\alpha}} = \sqrt{F_{\alpha\alpha}} \]  

(B.15)

When trying to constrain multiple parameters the error for each parameters is instead determined by marginalising over the parameter space, which in practise means inverting the Fisher information matrix to obtain the marginalised error:

\[ \sigma_{p_{\alpha}}^{\text{marg}} = \sqrt{(F^{-1})_{\alpha\alpha}}. \]  

(B.16)

The amount of signal that is lost due to the marginalisation can then be quantified by the signal-to-noise degradation (SND) factor:

\[ \text{SND}_{p_{\alpha}} = \frac{\sigma_{p_{\alpha}}^{\text{marg}}}{\sigma_{p_{\alpha}}} = \sqrt{(F^{-1})_{\alpha\alpha} F_{\alpha\alpha}}. \]  

(B.17)

which depends on how much parameters are correlated. This correlation can be quantified by the Fisher information matrix elements for the parameters considered:

\[ C_{\alpha\beta} = \frac{F_{\alpha\beta}}{\sqrt{F_{\alpha\alpha} F_{\beta\beta}}}. \]  

(B.18)

\footnote{Note that we use a slightly different definition of the SND factor as compared to [12], where the SND factor is defined as \( \text{SND} = \sqrt{(F^{-1})_{\alpha\alpha} F_{\alpha\alpha}} - 1 \).}
which will take a value between $-1$ and $1$ and is often referred to as the overlap or cosine between two parameters. As an example, for two parameters the SND factor and correlation matrix are then simply related by:

$$\text{SND}_{\alpha\beta} = \sqrt{\frac{1}{1 - (C_{\alpha\beta})^2}}. \quad \text{(B.19)}$$

Hence, an overlap of about $C_{\alpha\beta} = 0.995$ is needed to reduce the sensitivity by a factor of 10.

## C 21-cm secondary trispectrum

We present the explicit expressions of the 11 contributions to the secondary trispectrum of 21-cm brightness temperature fluctuations. The contributions from two second order fluctuations are:

$$T_{\delta_1 \delta_2 \delta_3} = 4c_1^{(1)}(k_1)c_1^{(1)}(k_2)c_1^{(2)}(k_3)c_1^{(2)}(k_4) \left[F_2^{(s)}(k_1, k_24)F_2^{(s)}(-k_2, k_{24})P_1 P_2 P_{24} + k_1 \leftrightarrow k_2 \right] + 5 \text{ p.}$$

$$T_{\delta_1 \delta_2 \theta_2} = 4c_1^{(1)}(k_1)c_1^{(1)}(k_2)c_2^{(2)}(k_3)c_2^{(2)}(k_4) \left[G_2^{(s)}(k_1, k_24)G_2^{(s)}(-k_2, k_{24})P_1 P_2 P_{24} + k_1 \leftrightarrow k_2 \right] + 5 \text{ p.}$$

$$T_{\delta_1 \delta_2 \theta_3} = 4c_1^{(1)}(k_1)c_1^{(1)}(k_2)c_2^{(2)}(k_3)c_2^{(2)}(k_4) \left[F_2^{(s)}(k_1, k_24)G_2^{(s)}(-k_2, k_{24})P_1 P_2 P_{24} + k_1 \leftrightarrow k_2 \right] + 11 \text{ p.}$$

$$T_{\delta_1 \delta_2 |\delta_1|2} = 2c_1^{(1)}(k_1)c_1^{(1)}(k_2)c_2^{(2)}(k_3) \left\{ \left[c_3^{(2)}(k_3, k_{23}) + c_3^{(2)}(k_3, -k_2) \right] F_2^{(s)}(k_1, k_{23})P_1 P_2 P_{23} + k_1 \leftrightarrow k_2 \right\} + 11 \text{ p.}$$

$$T_{\delta_1 \delta_2 |\delta_1|2} = 2c_1^{(1)}(k_1)c_1^{(1)}(k_2)c_2^{(2)}(k_3) \left\{ \left[c_3^{(2)}(k_3, k_{23}) + c_3^{(2)}(k_3, -k_2) \right] G_2^{(s)}(k_1, k_{23})P_1 P_2 P_{23} + k_1 \leftrightarrow k_2 \right\} + 11 \text{ p.}$$

$$T_{\delta_1 |\delta_1|2 |\delta_1|2} = c_1^{(1)}(k_1)c_1^{(1)}(k_2) \left\{ \left[c_3^{(2)}(k_3, k_{13}) + c_3^{(2)}(k_3, -k_1) \right] \left[c_3^{(2)}(k_4, k_{24}) + c_3^{(2)}(k_4, -k_2) \right] \right\} \times P_1 P_2 P_{13} + k_3 \leftrightarrow k_4 \right\} + 5 \text{ p.} \quad \text{(C.1)}$$

The contributions from one third order fluctuation are:

$$T_{\delta_1 \delta_1 \delta_3} = 6c_1^{(1)}(k_1)c_1^{(1)}(k_2)c_1^{(1)}(k_3)c_1^{(3)}(k_4) F_3^{(s)}(k_1, k_2, k_3)P_1 P_2 P_3 + 3 \text{ p.}$$

$$T_{\delta_1 \delta_1 \delta_3} = 6c_1^{(1)}(k_1)c_1^{(1)}(k_2)c_1^{(1)}(k_3)c_1^{(3)}(k_4) G_3^{(s)}(k_1, k_2, k_3)P_1 P_2 P_3 + 3 \text{ p.}$$

$$T_{\delta_1 \delta_1 |\delta_1|3} = c_1^{(1)}(k_1)c_1^{(1)}(k_2)c_1^{(1)}(k_3) \left\{ \left[c_3^{(2)}(k_4, -k_2, -k_3) + k_2 \leftrightarrow k_3 \right] + 2 \text{ c.p.} \right\} P_1 P_2 P_3 + 3 \text{ p.}$$

$$T_{\delta_1 \delta_1 |\delta_1|2} = 2c_1^{(1)}(k_1)c_1^{(1)}(k_2)c_1^{(1)}(k_3) \left[c_5^{(3)}(k_4, k_{14})F_2^{(s)}(k_2, k_3) + 2 \text{ c.p.} \right] P_1 P_2 P_3 + 3 \text{ p.}$$

$$T_{\delta_1 \delta_1 |\delta_1|2} = 2c_1^{(1)}(k_1)c_1^{(1)}(k_2)c_1^{(1)}(k_3) \left[c_5^{(3)}(k_4, k_{14})G_2^{(s)}(k_2, k_3) + 2 \text{ c.p.} \right] P_1 P_2 P_3 + 3 \text{ p.} \quad \text{(C.2)}$$

where we used the condensed notation $P_i = P_i(k_i)$ for the baryonic power spectra, c.p. denotes cyclic permutation.
References

[1] Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini et al., Planck2018 results, Astronomy & Astrophysics 641 (Sep, 2020) A10.

[2] PLANCK collaboration, Y. Akrami et al., Planck 2018 results. IX. Constraints on primordial non-Gaussianity, Astron. Astrophys. 641 (2020) A9, [1905.05697].

[3] P. D. Meerburg et al., Primordial Non-Gaussianity, 1903.04409.

[4] J. M. Maldacena, Non-Gaussian features of primordial fluctuations in single field inflationary models, JHEP 05 (2003) 013, [astro-ph/0210603].

[5] P. Creminelli, On non-Gaussianities in single-field inflation, JCAP 10 (2003) 003, [astro-ph/0306122].

[6] P. Creminelli and M. Zaldarriaga, Single field consistency relation for the 3-point function, JCAP 10 (2004) 006, [astro-ph/0407059].

[7] N. Arkani-Hamed and J. Maldacena, Cosmological Collider Physics, 1503.08043.

[8] H. Lee, D. Baumann and G. L. Pimentel, Non-Gaussianity as a Particle Detector, JHEP 12 (2016) 040, [1607.03735].

[9] D. Baumann, G. Goon, H. Lee and G. L. Pimentel, Partially Massless Fields During Inflation, JHEP 04 (2018) 140, [1712.06624].

[10] A. Cooray, 21-cm Background Anisotropies Can Discern Primordial Non-Gaussianity, Phys. Rev. Lett. 97 (2006) 261301, [astro-ph/0610257].

[11] A. Pillepich, C. Porciani and S. Matarrese, The bispectrum of redshifted 21-cm fluctuations from the dark ages, Astrophys. J. 662 (2007) 1–14, [astro-ph/0611126].

[12] P. D. Meerburg, M. Münchmeyer, J. B. Muñoz and X. Chen, Prospects for Cosmological Collider Physics, JCAP 03 (2017) 050, [1610.06559].

[13] J. Silk, The limits of cosmology: role of the Moon, Phil. Trans. A. Math. Phys. Eng. Sci. 379 (2021) 20190561, [2011.04671].

[14] J. B. Muñoz, Y. Ali-Haïmoud and M. Kamionkowski, Primordial non-gaussianity from the bispectrum of 21-cm fluctuations in the dark ages, Phys. Rev. D 92 (2015) 083508, [1506.04152].

[15] A. Kalaja, P. D. Meerburg, G. L. Pimentel and W. R. Coulton, Fundamental limits on constraining primordial non-Gaussianity, JCAP 04 (2021) 050, [2011.09461].

[16] PLANCK collaboration, N. Aghanim et al., Planck 2018 results. VI. Cosmological parameters, Astron. Astrophys. 641 (2020) A6, [1807.06209].

[17] S. Furlanetto, S. P. Oh and F. Briggs, Cosmology at Low Frequencies: The 21 cm Transition and the High-Redshift Universe, Phys. Rept. 433 (2006) 181–301, [astro-ph/0608032].

[18] A. Lewis and A. Challinor, The 21cm angular-power spectrum from the dark ages, Phys. Rev. D 76 (2007) 083005, [astro-ph/0702600].

[19] Y. Ali-Haïmoud, P. D. Meerburg and S. Yuan, New light on 21 cm intensity fluctuations from the dark ages, Phys. Rev. D 89 (2014) 083506, [1312.4948].

[20] A. Lewis, Linear effects of perturbed recombination, Physical Review D - Particles, Fields, Gravitation and Cosmology 76 (2007) 1–5, [0707.2727].

[21] P. J. E. Peebles, Recombination of the Primeval Plasma, Astrophys. J. 153 (1968) 1.

[22] S. Seager, D. D. Sasselov and D. Scott, How exactly did the universe become neutral?, Astrophys. J. Suppl. 128 (2000) 407–430, [astro-ph/9912182].
[23] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan and L. Senatore, The Effective Field Theory of Inflation, *JHEP* **03** (2008) 014, [0709.0293].
[24] X. Chen and C. Ringeval, Searching for Standard Clocks in the Primordial Universe, *JCAP* **08** (2012) 014, [1205.6085].
[25] X. Chen, M. H. Namjoo and Y. Wang, Models of the Primordial Standard Clock, *JCAP* **02** (2015) 027, [1411.2349].
[26] X. Chen and M. H. Namjoo, Standard Clock in Primordial Density Perturbations and Cosmic Microwave Background, *Phys. Lett. B* **739** (2014) 285–292, [1404.1536].
[27] X. Chen, M. H. Namjoo and Y. Wang, Quantum Primordial Standard Clocks, *JCAP* **02** (2016) 013, [1509.03930].
[28] X. Chen and Y. Wang, Large non-Gaussianities with Intermediate Shapes from Quasi-Single Field Inflation, *Phys. Rev. D* **81** (2010) 063511, [0909.0496].
[29] T. Suyama and M. Yamaguchi, Non-Gaussianity in the modulated reheating scenario, *Phys. Rev. D* **77** (2008) 023505, [0709.2545].
[30] K. M. Smith, L. Senatore and M. Zaldarriaga, Optimal analysis of the CMB trispectrum, [1502.00635].
[31] L. Senatore and M. Zaldarriaga, The Effective Field Theory of Multifield Inflation, *JHEP* **04** (2012) 024, [1009.2093].
[32] L. Senatore and M. Zaldarriaga, A Naturally Large Four-Point Function in Single Field Inflation, *JCAP* **01** (2011) 003, [1004.1201].
[33] V. Assassi, D. Baumann and D. Green, On Soft Limits of Inflationary Correlation Functions, *JCAP* **11** (2012) 047, [1204.4207].
[34] F. Bernardeau, S. Colombi, E. Gaztanaga and R. Scoccimarro, Large scale structure of the universe and cosmological perturbation theory, *Phys. Rept.* **367** (2002) 1–248, [astro-ph/0112551].
[35] M. Shoji and E. Komatsu, Third-order Perturbation Theory With Non-linear Pressure, *Astrophys. J.* **700** (2009) 705–719, [0903.2669].
[36] X. Chen, P. D. Meerburg and M. Münchmeyer, The Future of Primordial Features with 21 cm Tomography, *JCAP* **09** (2016) 023, [1605.09364].
[37] D. Bertolini, K. Schutz, M. P. Solon, J. R. Walsh and K. M. Zurek, Non-Gaussian Covariance of the Matter Power Spectrum in the Effective Field Theory of Large Scale Structure, *Phys. Rev. D* **93** (2016) 123505, [1512.07630].
[38] *Square Kilometer Array*. https://www.skatelescope.org/.
[39] J. C. Pober et al., What Next-Generation 21 cm Power Spectrum Measurements Can Teach Us About the Epoch of Reionization, *Astrophys. J.* **782** (2014) 66, [1310.7031].
[40] J. O. Burns, Transformative science from the lunar farside: observations of the dark ages and exoplanetary systems at low radio frequencies, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **379** (Nov, 2020) 20190564.
[41] L. V. E. Koopmans et al., Peering into the dark (ages) with low-frequency space interferometers: Using the 21-cm signal of neutral hydrogen from the infant universe to probe fundamental (Astrophysics, *Exper. Astron.* **51** (2021) 1641–1676, [1908.04296].
[42] A. R. Parsons, J. C. Pober, J. E. Aguirre, C. L. Carilli, D. C. Jacobs and D. F. Moore, A Per-Baseline, Delay-Spectrum Technique for Accessing the 21cm Cosmic Reionization Signature, *Astrophys. J.* **756** (2012) 165, [1204.4749].
[43] A. Liu, A. R. Parsons and C. M. Trott, *Epoch of reionization window. I. Mathematical formalism*, Phys. Rev. D **90** (2014) 023018, [1404.2596].

[44] A. Liu, A. R. Parsons and C. M. Trott, *Epoch of reionization window. II. Statistical methods for foreground wedge reduction*, Phys. Rev. D **90** (2014) 023019, [1404.4372].

[45] A. Ghosh, F. Mertens and L. V. E. Koopmans, *Deconvolving the wedge: maximum-likelihood power spectra via spherical-wave visibility modelling*, Mon. Not. Roy. Astron. Soc. **474** (2018) 4552–4563, [1709.06752].

[46] C. Modi, M. White, A. Slosar and E. Castorina, *Reconstructing large-scale structure with neutral hydrogen surveys*, JCAP **11** (2019) 023, [1907.02330].

[47] J. R. Shaw, K. Sigurdson, U.-L. Pen, A. Stebbins and M. Sitwell, *All-sky interferometry with spherical harmonic transit telescopes*, The Astrophysical Journal **781** (Jan, 2014) 57.

[48] X. Chen, *Radio detection of dark energy—the Tianlai project*, Scientia Sinica Physica, Mechanica & Astronomica **41** (Jan., 2011) 1358.

[49] L. B. Newburgh, K. Bandura, M. A. Bucher, T.-C. Chang, H. C. Chiang, J. Cliche et al., *Hirax: a probe of dark energy and radio transients*, Ground-based and Airborne Telescopes VI (Aug, 2016).

[50] PUMA collaboration, A. Slosar et al., *Packed Ultra-wideband Mapping Array (PUMA): A Radio Telescope for Cosmology and Transients*, Bull. Am. Astron. Soc. **51** (2019) 53, [1907.12559].

[51] LOFAR, *Low-Frequency Array*. https://www.astro.rug.nl/lofareor/scibgd/scibgd.html.

[52] LWA, *Long Wavelength Array*. http://lwa.phys.unm.edu/science.html.

[53] MWA, *Murchison Widefield Array*. https://www.mwatelescope.org/science/epoch-of-reionization-eor.

[54] F. G. Mertens, B. Semelin and L. V. E. Koopmans, *Exploring the Cosmic Dawn with NenuFAR*, in Semaine de l’astrophysique française 2021, 9, 2021. 2109.10055.

[55] C. DiLullo, G. B. Taylor and J. Dowell, *Using the Long Wavelength Array to Search for Cosmic Dawn*, J. Astron. Inst. **09** (2020) 2050008, [2005.10669].

[56] R. Scoccimarro, E. Sefusatti and M. Zaldarriaga, *Probing primordial non-Gaussianity with large - scale structure*, Phys. Rev. D **69** (2004) 103513, [astro-ph/0312286].

[57] T. Baldauf, M. Mirbabayi, M. Simonović and M. Zaldarriaga, *LSS constraints with controlled theoretical uncertainties*, 1602.00674.

[58] J. R. Fergusson, D. M. Regan and E. P. S. Shellard, *Rapid Separable Analysis of Higher Order Correlators in Large Scale Structure*, Phys. Rev. D **86** (2012) 063511, [1008.1730].

[59] X. Chen, B. Hu, M. X. Huang, G. Shiu and Y. Wang, *Large primordial trispectra in general single field inflation*, Journal of Cosmology and Astroparticle Physics **2009** (5, 2009).

[60] G. P. Lepage, *Adaptive multidimensional integration: VEGAS enhanced*, J. Comput. Phys. **439** (2021) 110386, [2009.05112].