Exploring the Epoch of Reionization with numerical simulations : the LoReLi database

Romain Meriot

3rd year PhD student at LERMA
Paris Observatory
Advisor : Benoit Semelin

February 28, 2024

The Epoch of Reionization

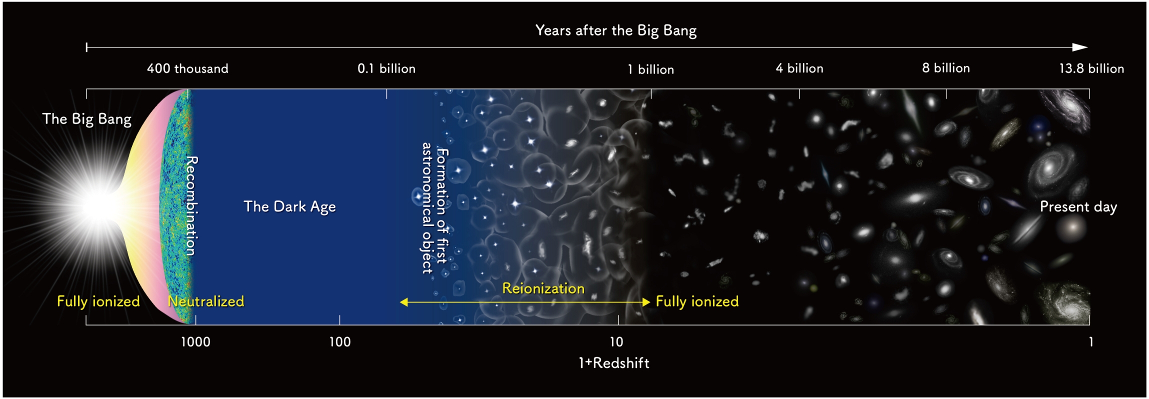


Figure: Illustration of the evolution of the universe

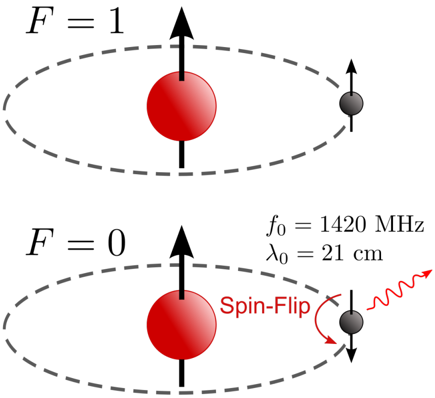
The 21 cm line of neutral Hydrogen

- Hyperfine transition at the HI ground state
- Observed by radiotelescopes: $\lambda \sim 1m$
- Seen in contrast with the Cosmic Microwave Background (CMB)

Brightness temperature

$$\delta T_b \propto 27 x_{HI} (1+\delta) \left[\frac{T_s - T_{CMB}}{T_s} \right] \text{ mK} \quad (1)$$

- x_{HI} : neutral fraction
- δ : overdensity
- T_s : spin temperature, $\frac{n_1}{n_0} = 3e^{-T_{21}/T_s}$ *depends on gas temperature and local Ly-α flux*



Observations (nope ¯_(ツ)_/¯)

Full 3D Signal

Square Kilometer Array
(∼ 2030)

Meanwhile : Attempts at measuring the power spectrum of the fluctuations of the signal (e.g HERA, LOFAR, NenuFAR in France)
We only have upper limits...

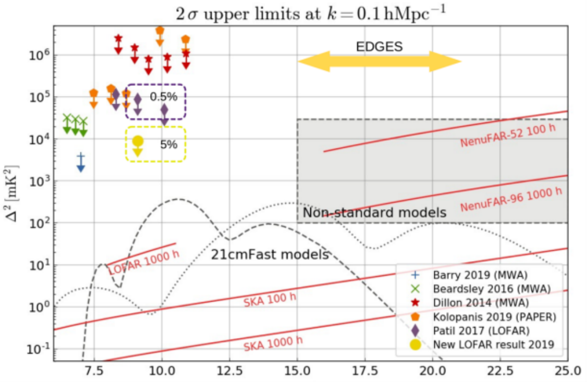


Figure: Upper limits on the power spectrum, expected instrumental sensitivity (*red lines*), predictions of the semi-analytical code 21cmFast, EDGES detection range. *Credit : Florent Mertens.*

The LICORICE simulation code

N-body simulation

- Coupling between dynamics and **3D radiative transfer**
- **Gravity** : TREE+Smoothed Particle Hydrodynamics
- **Gas physics** (photoionization, collisional ionization, recombination, cooling...)

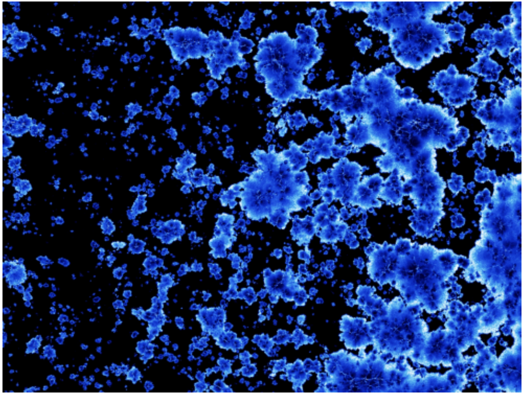


Figure: Temperature slice from a LICORICE simulation.Time flows from the left to the right; Light blue regions are ionized.

Purpose of the PhD

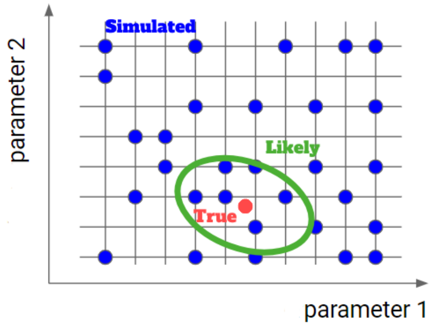


Figure: Illustration of parameter inference : one wants to find out which parameters correspond to the observations ("True")

Parameter inference

$\longrightarrow$ **MCMC** with high-res simulations : $< 10^{11}$ *cpu hours*

$\longrightarrow$ **Machine learning** : *100x fewer simulations* ($\sim 10^3$, Doussot et al. 2019)

Purpose of the PhD

- LICORICE simulations are expensive ! $\longrightarrow$ very low resolution (256^3 particles...)

Let's build a database of 10000 simulations and use it to train a neural network to perform the inference

Star formation in LICORICE

Without subgrid modeling, gaz particles denser than $100 \sim 200 \langle \rho_0 \rangle$ form stars :

$$df_* = (1 - f_*) \frac{\mathrm{dt}}{\tau_{SF}} \tag{2}$$

- f_* : stellar mass fraction
- τ_{SF} : *typical star formation time*

Problem : resolution limit

- At best : $\sim 300\,\mathrm{Mpc}$ on a side, 2048^3 particles
  $\longrightarrow$ resolved halos : $\sim 4 \times 10^9\,\mathrm{M}_\odot$
- smallest star-forming halos : $\sim 10^8\,\mathrm{M}_\odot$

At 256^3 resolution...

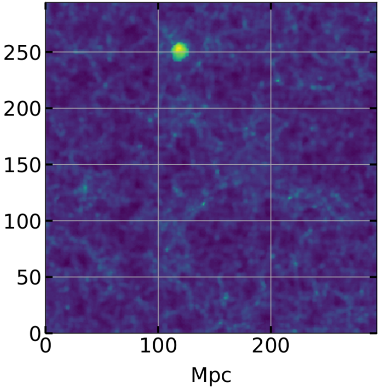


Figure: ionization map of a 256^3 particles simulation at $z \sim 7$

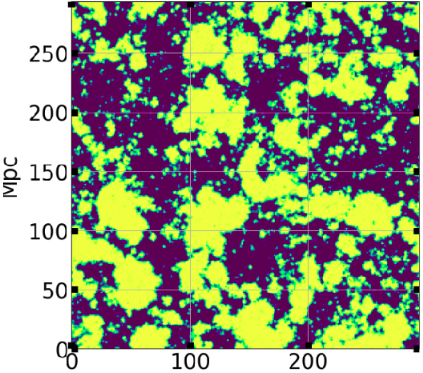


Figure: temperature map of a 2048^3 particles simulation at $z \sim 7$

$\longrightarrow$ Let's add subgrid models to improve star formation (and perhaps a few other aspects too...)

Solution : estimating the mass of non-resolved halos with Conditional Mass Functions (CMF)

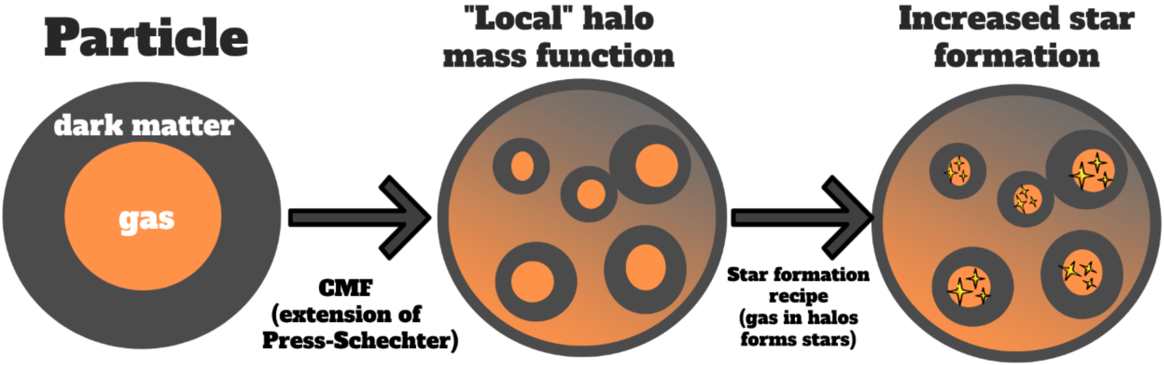


Figure: Illustration of the implementation : for each particle, we calculate its conditional mass function n_c, then its collapsed fraction f_{coll}, then form stars with $df_* = (f_{coll} - f_*)\frac{\mathrm{dt}}{\tau_{SF}}$

Star formation rate

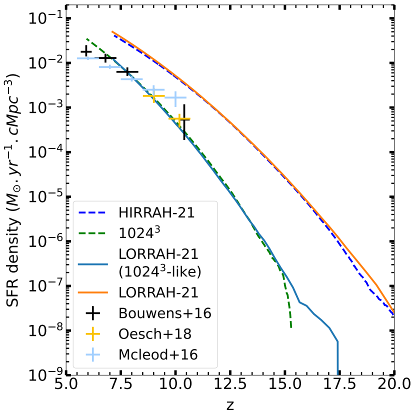


Figure: Star formation rate of a 256^3 simulation (solid) with parameters close to those of HIRRAH (2048^3, dotted)

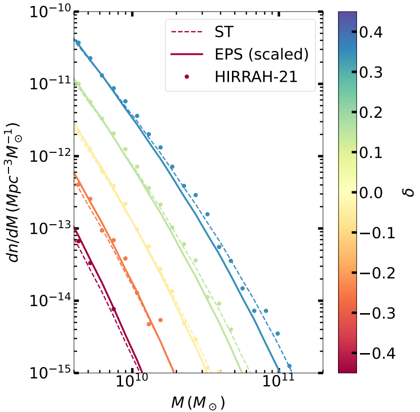


Figure: CMF of HIRRAH (2048^3) and theoretical predictions at different overdensities

With the subgrid modelling:

We can accurately model the EoR at low-res ! Reminder : low resolution simulations are 10000 times faster !

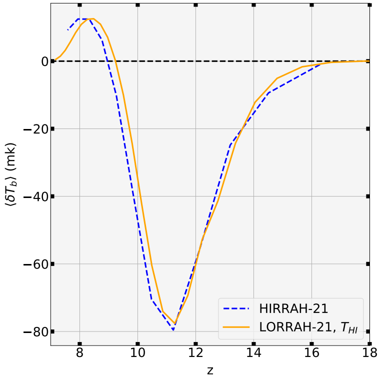
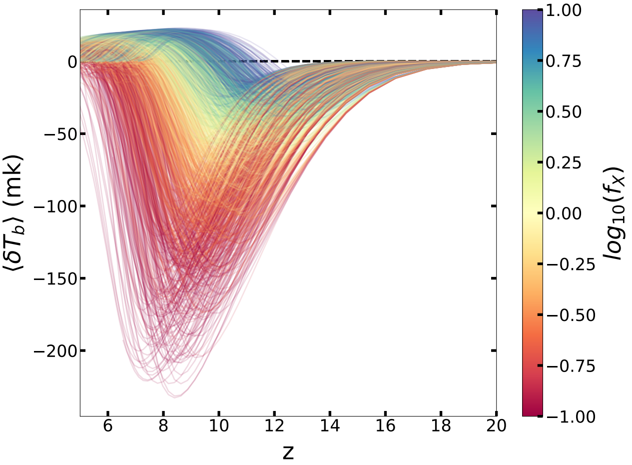


Figure: Average brightness temperature in 2048^3 (HIRRAH) and 256^3 (LORRAH) simulations with the same parameters.

Database of simulations : LoReLi II

- **10000 simulations**
- $\sim 5 \times 10^6$ cpuh, 1.5PB of data
- 5 parameters : escape fraction, X-ray emissivity, hard/soft X-ray ratio, 2 SFR Parameters
- Calibrated using observational constraints on star formation rate and reionization timings
- Cubes of : f_{coll}, δ, f_{ion}, $\langle T \rangle$, T_{HI}, x_α, δT_b
- powerspectra, global signals, raw data+cubes to calculate additional summaries



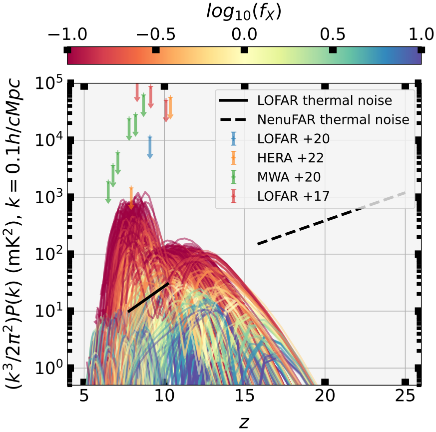


Figure: LoReLi power spectra as a function of z, $k \sim 0.1h/cMpc$

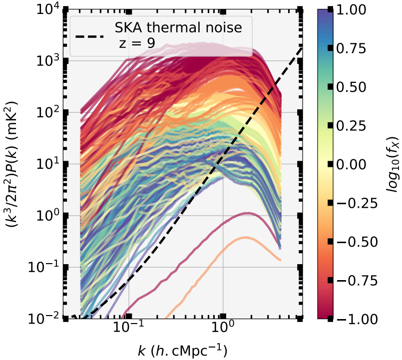


Figure: LoReLi spectra as a function of k, $z = 9$

Neural network emulator of the code

- **Train a neural network to emulate the simulation** (3x512 neurons MLP, $\sim 5\%$ error)
- Plug the emulator into an MCMC framework
- Perform classical MCMC inference

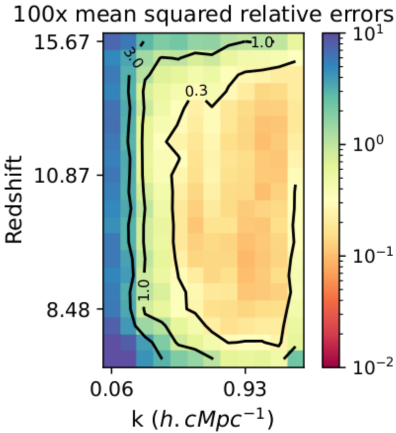


Figure: $100 \times \langle (\frac{P_{sim} - P_{emu}}{P_{sim}})^2 \rangle$ of the emulator

Inference with emulator

- Perform MCMC inference on some emulated signal
- uncertainty : noise (SKA 100h), cosmic variance, model error
- 10 emulators were trained with different random weight initializations
- **approximate** Gaussian likelihood : $logL = \sum_{k,z} \frac{(P_{obs}(k,z) - P_{pred}(k,z))^2}{\sigma^2}$

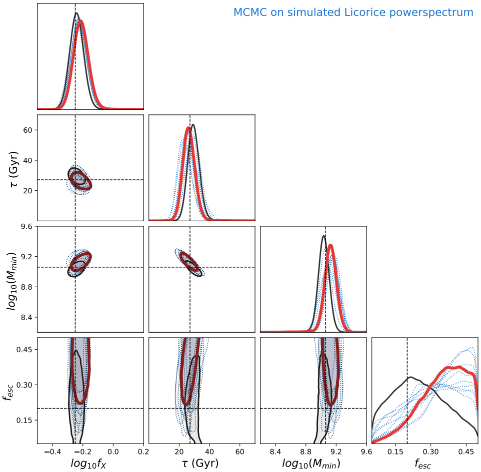


Figure: MCMC inference with the LoReLi emulators. **Average prediction in red** (Meriot & Semelin 2024)

Inference on actual HERA data !

- MCMC inference with the emulator on data from HERA Collaboration 2022a
- Consistent results : **exclusion of cold reionization models (small f_X)**

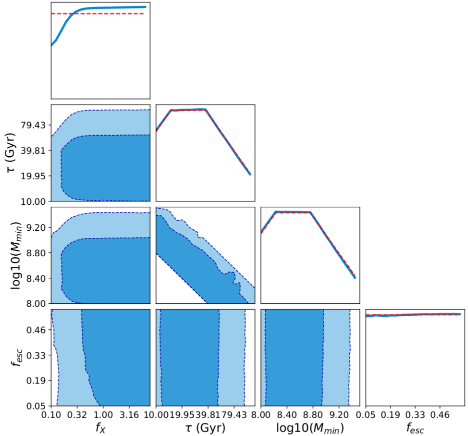


Figure: MCMC inference with the LoReLi emulator on HERA data. Red : prior, blue : posterior. (Meriot & Semelin 2024)

Conclusions

- Modelling of the EoR with fast simulations + subgrid models
- LoReLi I (760 simulations, https://21ssd.obspm.fr/), **LoReLi II (10000 simulations) !**
- **We do inference on real data with 3D RT simulations for the first time !**
- Next steps : Full comparison between inference techniques on LoReLi II *(emulator vs Likelihood-free Inference vs Bayesian Neural Networks, WIP :))*

The End

Neural network emulator of the code

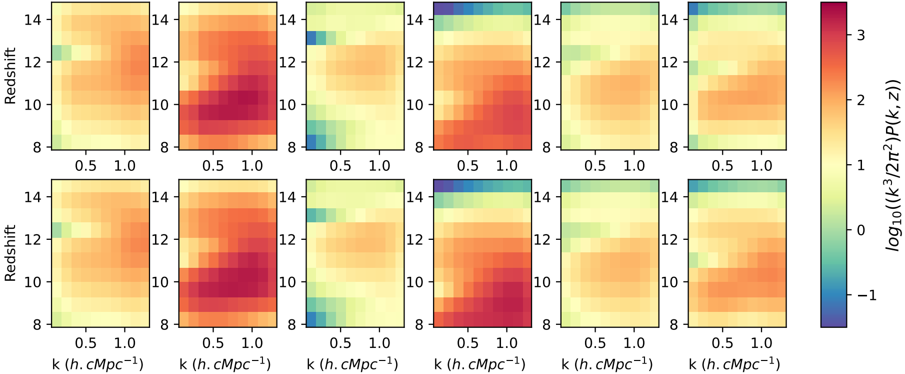


Figure: **Top** : randomly selected simulated power spectra. **Bottom** : corresponding emulated power spectra (Meriot & Semelin 2024)

Likelihood free inference

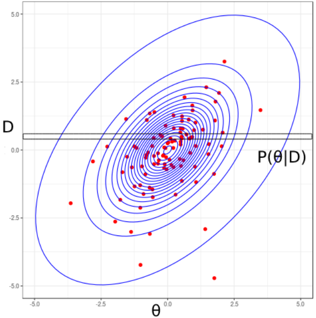


Figure: Illustration of Likelihood free inference : fitting the joint probability density allows direct evaluation of the posterior

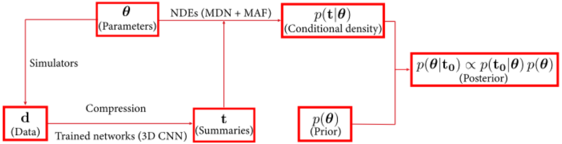


Figure: DELFI workflow (from Zhao et al 2022)

Direct inference on noised signals using Bayesian Neural Networks

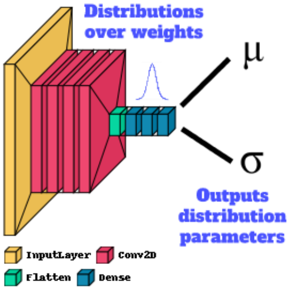


Figure: BNN architecture : 4 conv layer 1 dense, 1 dense variational layer, output layer. **Weights are drawn from learned gaussian distributions, and the network outputs a distribution of physical parameters. Here : means + covariance matrix**

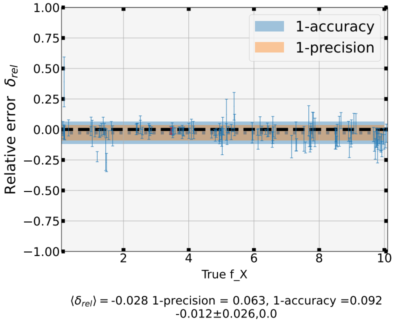
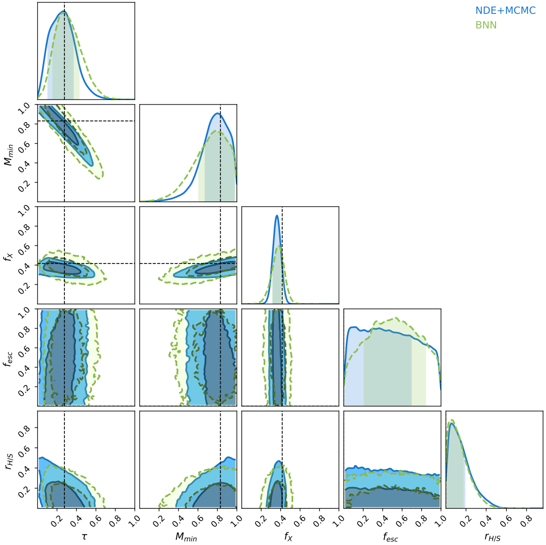


Figure: Relative error for each value of f_x, 100h SKA noise. Epistemic (= suboptimal training set) uncertainty : $\sim$ 25% of total uncertainty

Comparison between inference methods **(WIP !)**

- Train power spectrum emulator on LoReLi II and perform MCMC inference on some emulated signal + noise (SKA 100h)
- Train BNN on LoReLi II, forward pass to get posteriors
- Train NDE on LoReLi II, MCMC on signal + noise



Ionization maps

Large structures recovered, but ionization more uniform at low-res

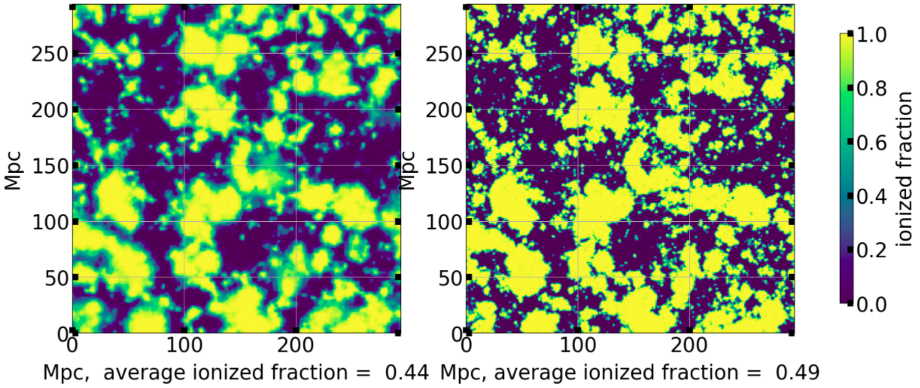


Figure: Ionization maps for a 256^3 simulation (*left*), and the 2048^3 reference (*right*), at $\sim 45\%$ ionization on average

Why do we care ?

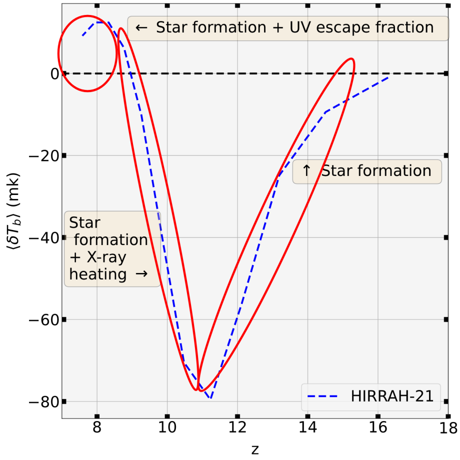
Brightness temperature

$$\delta T_b \propto 27 x_{HI} \left[\frac{T_s - T_{CMB}}{T_s} \right] \text{mK} \quad (3)$$

T_s : spin temperature

$$T_s{}^{-1} \approx \frac{T_{CMB}^{-1} + x_\alpha T_{gas}{}^{-1}}{1 + x_\alpha} \quad (4)$$

- $z \gtrsim 11$: Star formation, $x_\alpha \nearrow$
- $z \lesssim 11$: X-ray heating, $T_{gas} \nearrow$
- $z \sim 6$: Reionization, $x_{HI} \searrow$



Average signal in a high resolution simulation

What's next ? Likelihood free inference

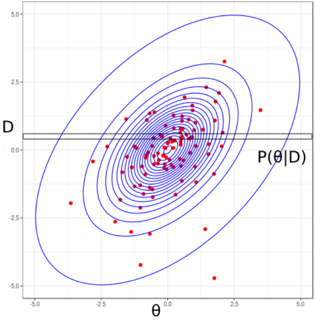


Figure: Illustration of Likelihood free inference : fitting the joint probability density allows direct evaluation of the posterior

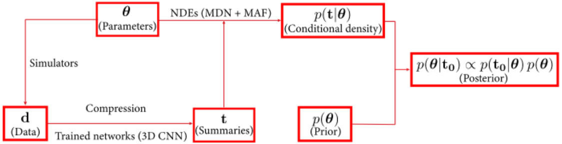


Figure: DELFI workflow (from Zhao et al 2022)

What about NenuFAR/LOFAR ?

- LOFAR : still too high
- NenuFAR : constraints only on **exotic** models : strong SFR, **very strong excess radio background** ($Ar \sim 1000$ **!!**)

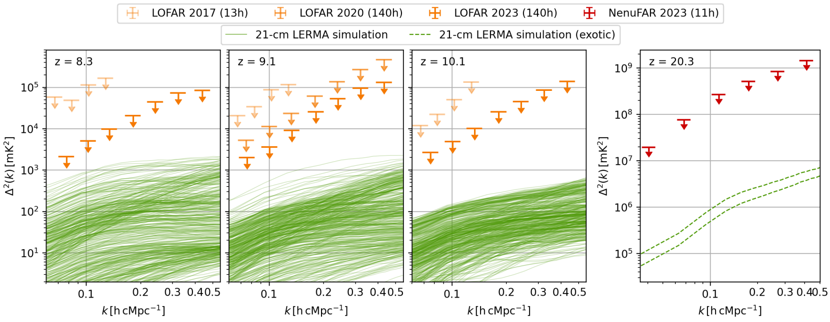


Figure: *Credit : Florent Mertens.* Most recent upper limits from LOFAR (3 leftmost panels) and NenuFAR (right panel). Green solid : standard LoReLi models. Green dashed : exotic LoReLi models with **Ar = 1000**, $M_{min} = 10^8\,\mathrm{M}_\odot$, $M_{min} = 4.10^7\,\mathrm{M}_\odot$

Bayesian inference !

Bayes equation

$$P(\theta|D) \propto P(D|\theta) \times P(\theta) \tag{5}$$

- Observed data D, astrophysical parameters θ
- $P(\theta)$ **Prior** knowledge of parameters
- $P(D|\theta)$ **Likelihood**
- $P(\theta|D)$ **Posterior**

We determine the posterior by evaluating the likelihood for many (carefully chosen) values of θ

Instantaneous star formation (like in semi analytical codes)

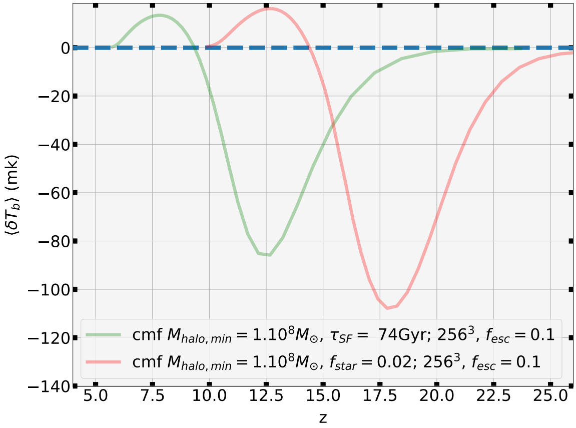


Figure: Signals of two Licorice simulations : with time dependent SF (green) and with instantaneous SF (red)

Parameter space

Parameters

- Star formation : M_{min}, τ_{SF}
- X-rays : f_X, $r_{H/S}$
- escape fraction : initial, threshold, final

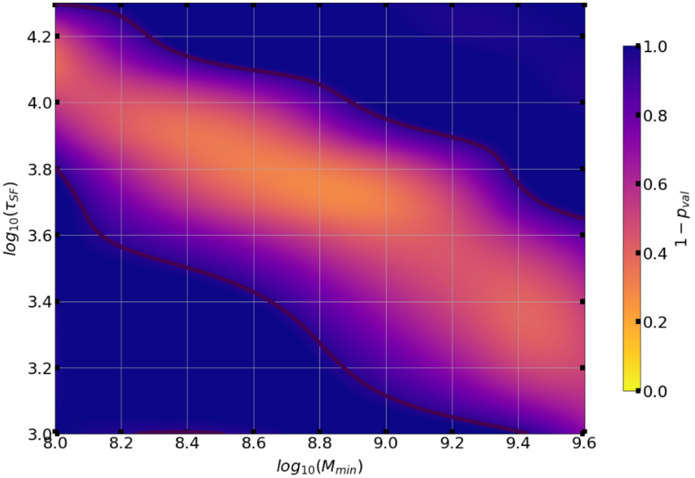


Figure: Explored parameter space in the M_{min}/τ_{SF} plane. (in practice : 20 points)

Building the database !

Compatible with observations ?

$\longrightarrow \chi^2$ goodness-of-fit statistical test

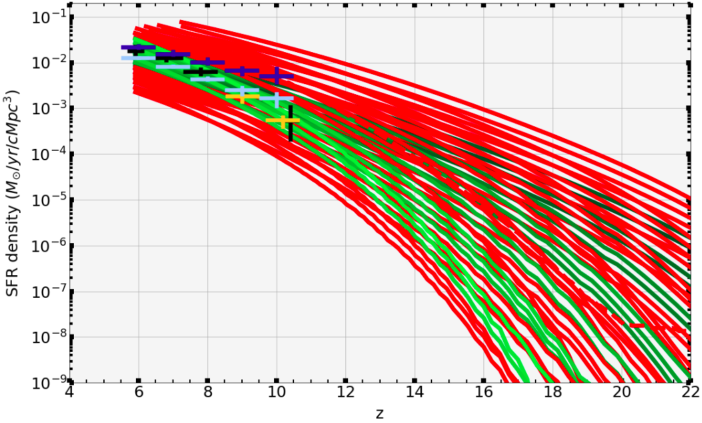


Figure: Star formation rate in simulations with $log(M_{min}) \in [8, 9.6]$ and $log(\tau_{SF}) \in [3, 4.3]$. Red : rejected by the χ^2 test.

Building the database !

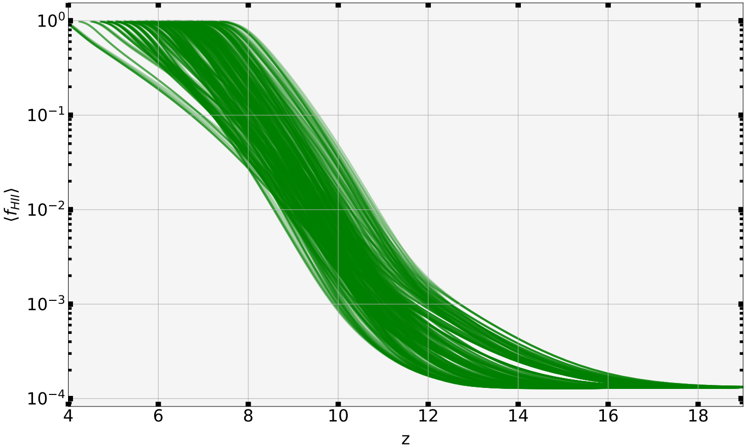


Figure: Average ionized fraction of 600 models.

HI temperature

Temperature of partially ionized particles polluted by the ionized part (unresolved ionization fronts)

- Ross et al. 2016 : "twin simulations" with and without Xrays, post-processing : ∼Ok results
- 21cmFAST : No partial ionization

We now calculate the HI temperature for each particle during the simulation (Adiabatic cooling + X-rays)

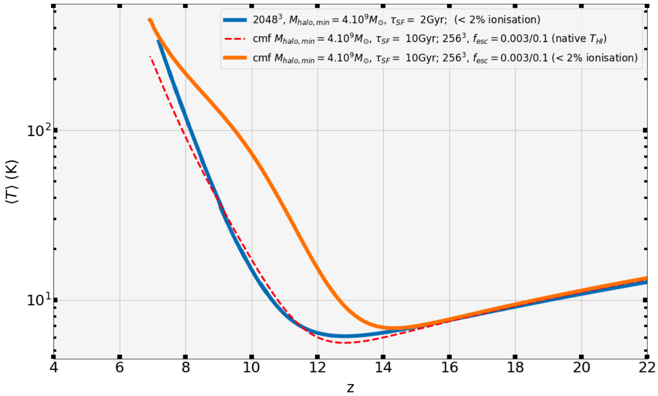


Figure: Temperature of the neutral gas : average T of weakly ($< 2\%$) ionized particles (*solid lines*) and native temperature of their neutral phase (*dashed lines*)

Solution : estimating the mass of non-resolved halos with Conditional Mass Functions (CMF)

We affect to each particle the mass of unresolved halos depending on its density δ_0 and its volume ($\leftrightarrow \sigma_0$).

mass fraction of the region that lies in halos

$$f_{coll} = V \int_{M_{min}}^{M_{region}} n_c(M) \mathrm{dM}/\mathrm{M_{region}} \quad (6)$$

Basic version : *Extended Press-Schechter :*

Lacey&Cole 1993

$$n_c(M) = \sqrt{\frac{2}{\pi}} \frac{d\sigma}{dM} \frac{\rho_0}{M} \frac{(\delta_c - \delta_0)\sigma}{(\sigma^2 - \sigma_0^2)^{3/2}} e^{\frac{(\delta_c - \delta_0)^2}{2(\sigma^2 - \sigma_0^2)}} \quad (7)$$

$$df_* = (f_{coll} - f_*) \frac{\mathrm{dt}}{\tau_{SF}}$$

The recombination issue

Recombination rate

$$Rec \propto \langle n_e n_{HII} \rangle \propto \langle n_{HII}^2 \rangle \quad (8)$$

However : limited resolution, $\langle n_{HII}^2 \rangle \neq \langle n_{HII} \rangle^2$

Theoretical solution ?

$$\mathrm{Clumping\ factor\ C} = \langle \mathrm{n_e n_{HII}} \rangle / \langle \mathrm{n_{HII}} \rangle^2 \quad (9)$$

$$Rec \propto C \langle n_{HII} \rangle^2 \quad (10)$$

Weak recombination : similar SFR $\rightarrow$ earlier reionization at low-res

Many implementations !

Kaurov Gnedin 2013, Mao et al 2019, Bianco et al 2020, Chen et al 2020...
$\longrightarrow$ None of them work (No coupling with temperature, high number of parameters, increasing error...

Escape fraction model

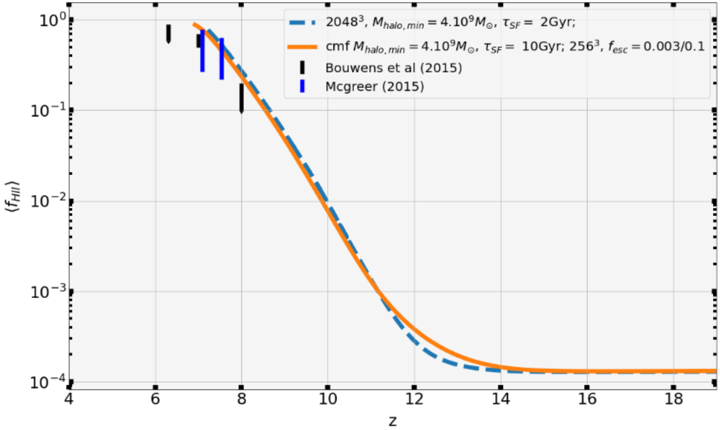


Figure: Ionized fraction of a 256^3 simulation with parameters close to those of the 2048^3. 2 different values of f_{esc} are used depending on the local ionized fraction.

Variable escape fraction : small if local ionized fraction below a given threshold.