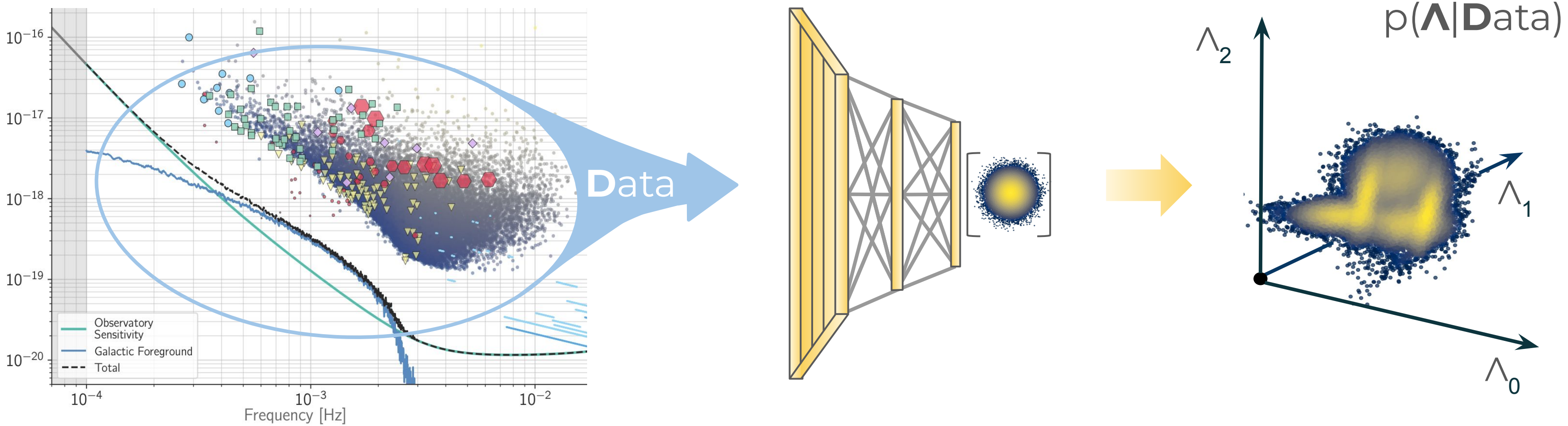
SISSA DATASCIENCE
Machine Learning for the Natural Sciences

SISSA
Scuola Internazionale Superiore di Studi Avanzati

Simulation Based Population Inference

Galactic Binaries in LISA



Rahul **Srinivasan**
Postdoc, SISSA, Italy

In collaboration with Enrico **Barausse,** Natalia **Korsakova**, Roberto **Trotta**

EuCAIF

The Galactic Double White Dwarf Population

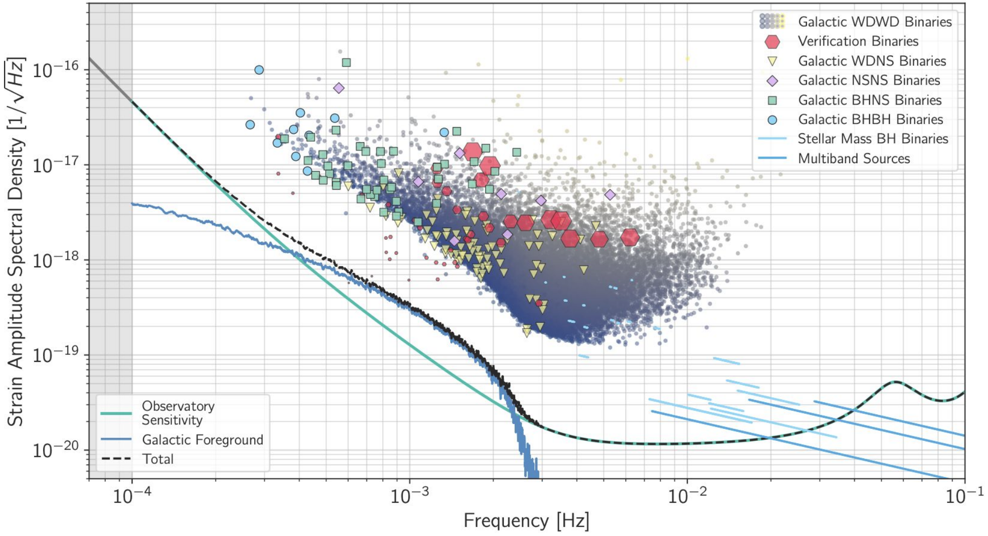
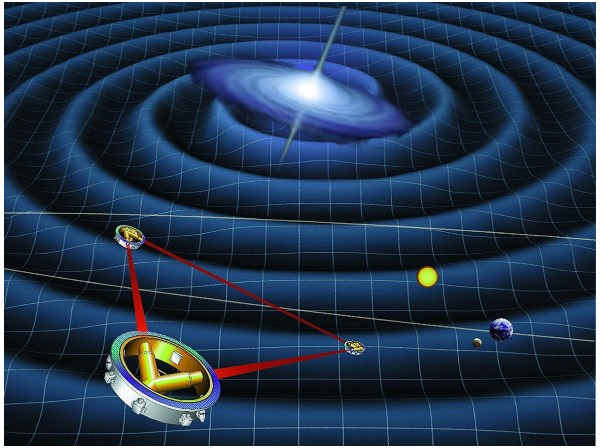


Figure from the LISA red book

LISA



The Galactic Double White Dwarf Population

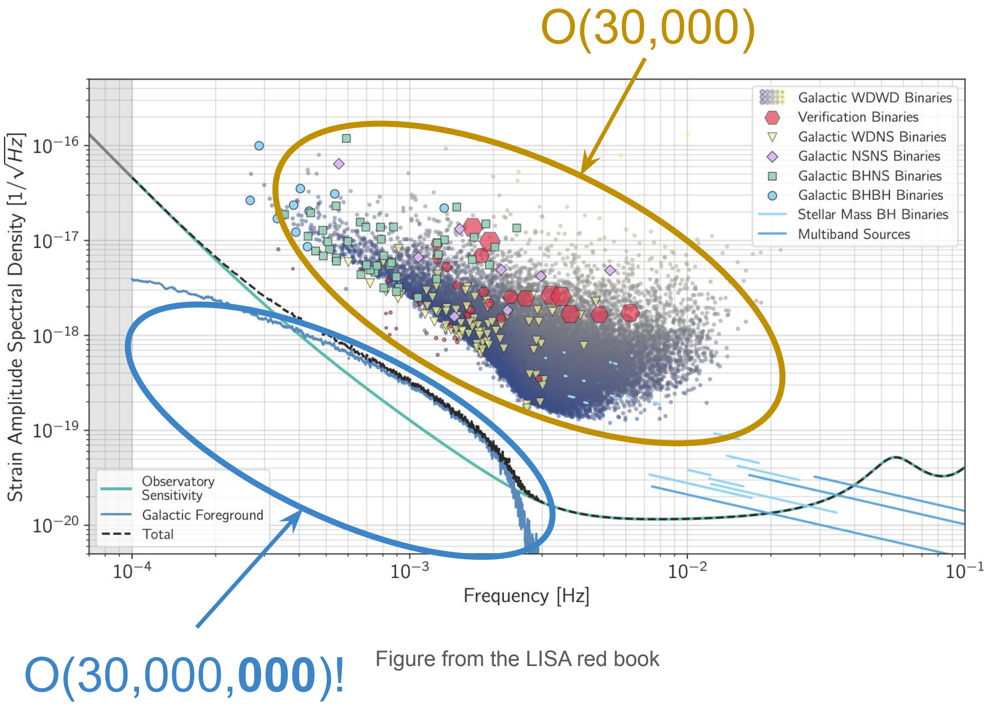
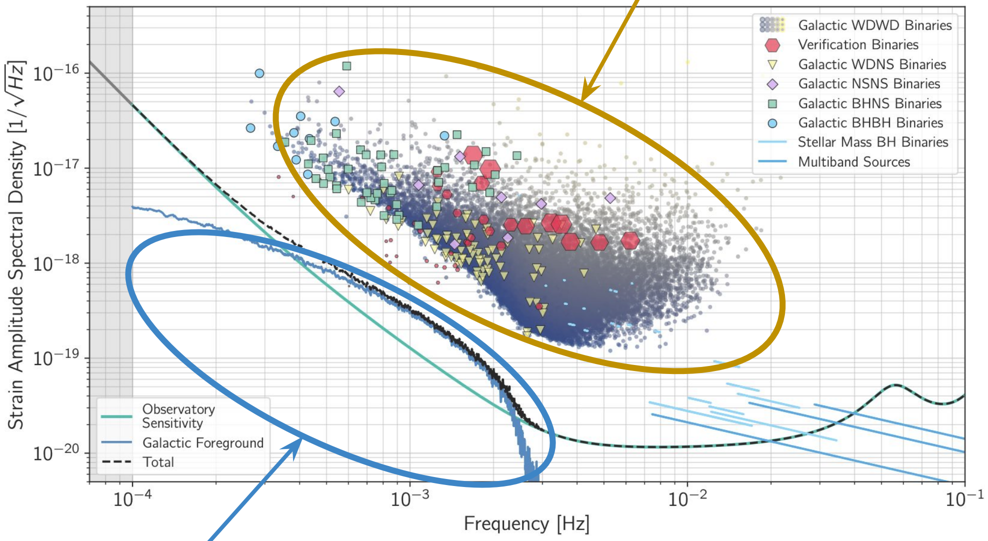


Figure from the LISA red book

A cocktail party of a few, loud, resolvable sources, and an overlapping, cacophonous, incoherent foreground.

The Galactic Double White Dwarf Population

O(30,000)



O(30,000,**000**)!

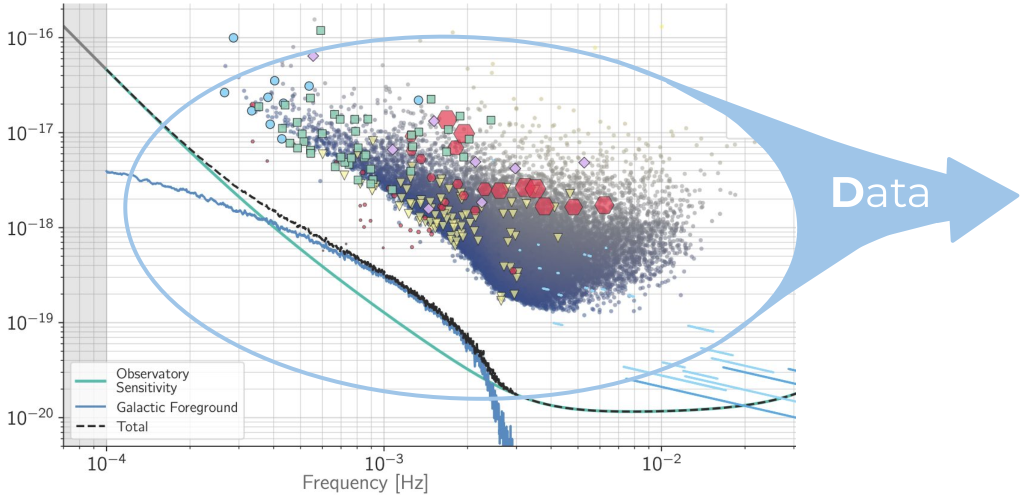
Figure from the LISA red book

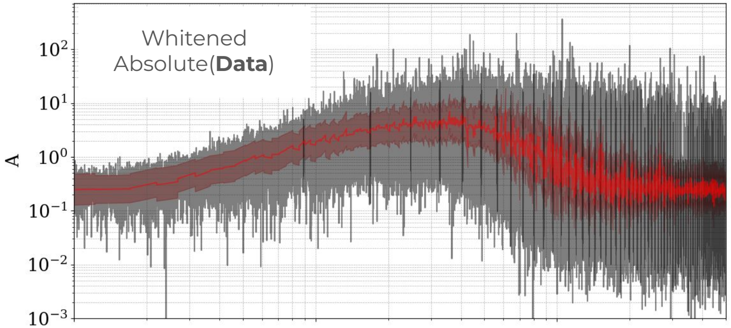
Where lies the population signature:
the few resolvable sources[1] or
the many unresolvable sources[2]?

[1] Korol, Rossi, Barausse, MNRAS 483 no. 4 5518–5533
[2] Smith et al 2020; 2004.09700

The Data

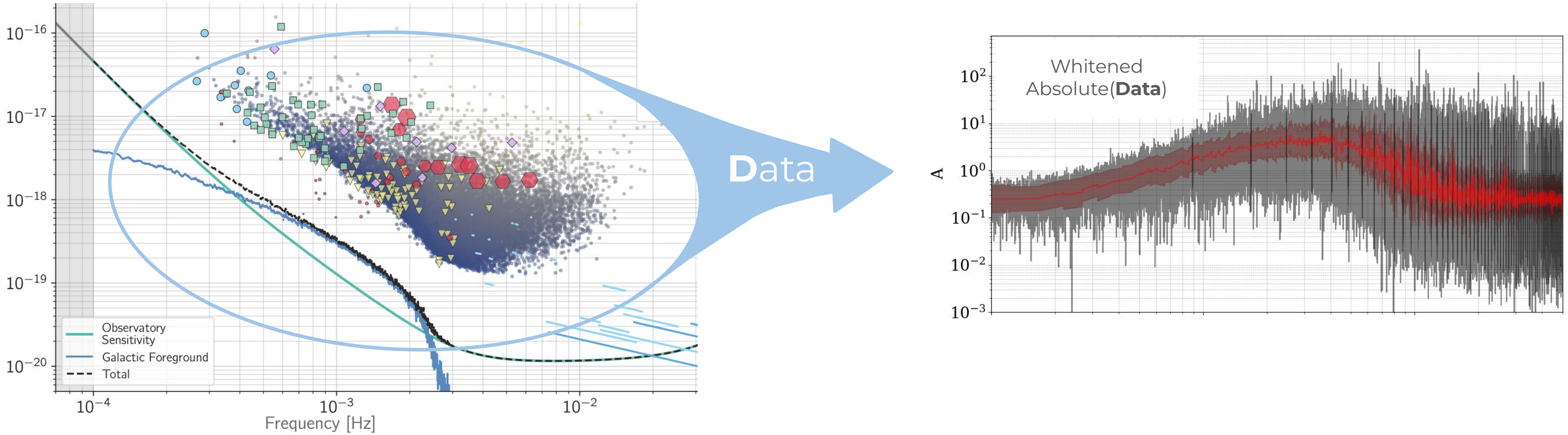
GPU-leveraged forward simulator





The Data

O(10^6) frequency bins!
Computational *prohibitive* likelihood evaluation.
Major bottleneck for Bayesian samplers.



SBI might be a *necessity*.

Simulation-based Inference

Population $\boldsymbol{\Lambda}$: = $\rho_{\text{Distribution}}$(Mass, Separation), N_{Sources}, Milky $\text{Way}_{\text{Scale}}$

Simulation-based Inference

Population $\boldsymbol{\Lambda}$: = $\rho_{Distribution}$(Mass, Separation), $N_{Sources}$, Milky Way_{Scale}

$\downarrow$ **Simulator**

Source $\boldsymbol{\theta}$: = Mass, Separation, Sky-position, Distance

Simulation-based Inference

Population $\boldsymbol{\Lambda}$: = $\rho_{Distribution}$(Mass, Separation), $N_{Sources}$, Milky Way_{Scale}

Source $\boldsymbol{\Theta}$: = Mass, Separation, Sky-position, Distance

Simulator

Data $\mathbf{D}$: = $\sum_i GW_i + n_{LSIA}$

Simulation-based Inference

SBI

Population $\mathbf{\Lambda}$: = $\rho_{\text{Distribution}}$(Mass, Separation), N_{Sources}, Milky $\text{Way}_{\text{Scale}}$

Source $\mathbf{\Theta}$: = Mass, Separation, Sky-position, Distance

Simulator

Data $\mathbf{D}$: = $\sum_i GW_i + n_{LSIA}$

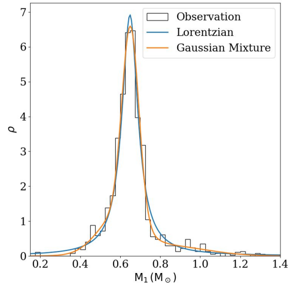
Given examples of $\mathbf{\Lambda} \rightarrow \mathbf{D}$,

SBI solves the *inverse problem* $\mathbf{D} \rightarrow \mathbf{\Lambda}$, i.e., $p(\mathbf{\Lambda}|\mathbf{D}, \mathbf{M}_{\text{Simulator}})$

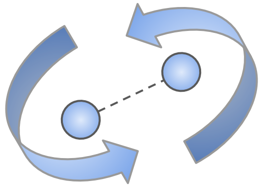
Λ : Parameterizing the Population

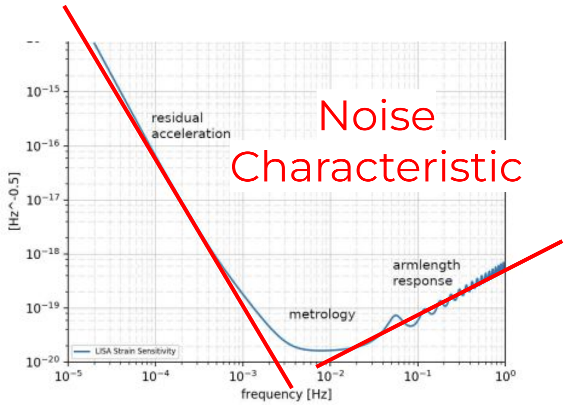
Total DWD Number	: 1
Primary DWD Mass Dist.	: 2
Separation Dist.	: 1
MW Disc	: 2
MW Bulge	: 2
LISA Noise Parameters	: 2
Total	: 10d

M_1 distribution

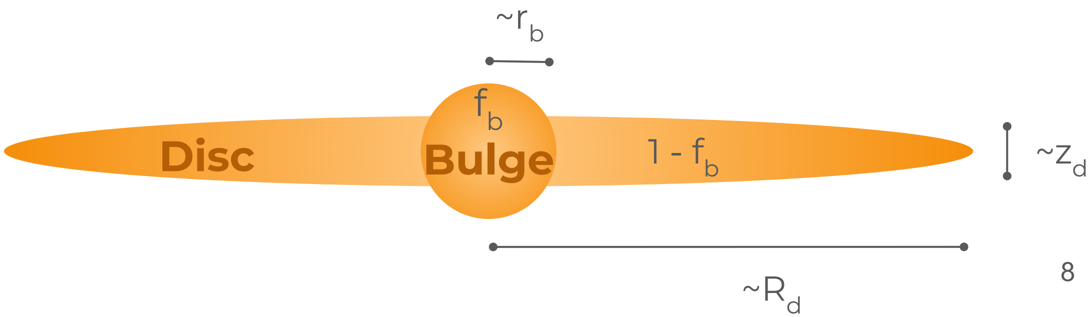


Separation Index

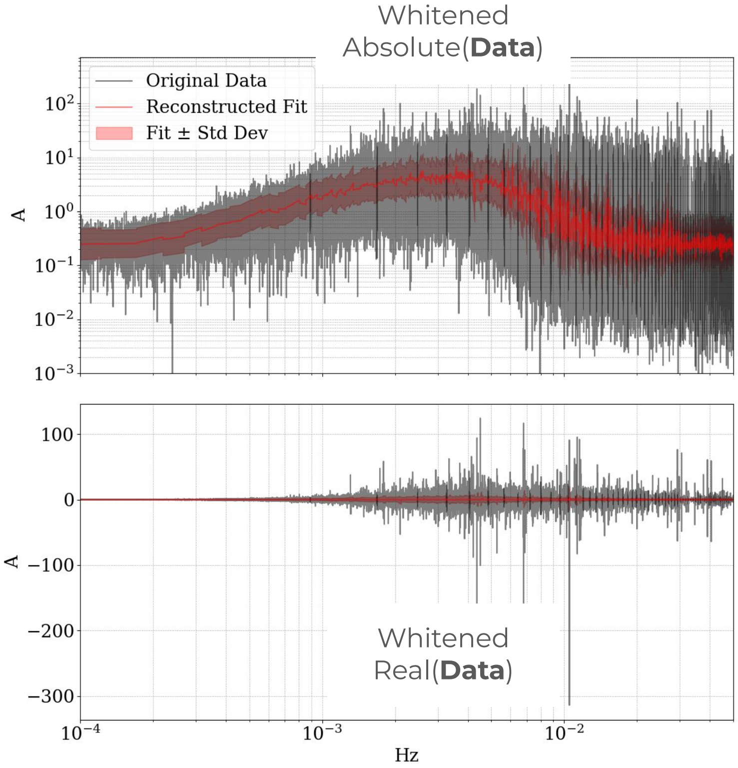




Milky Way Scale



Data summary



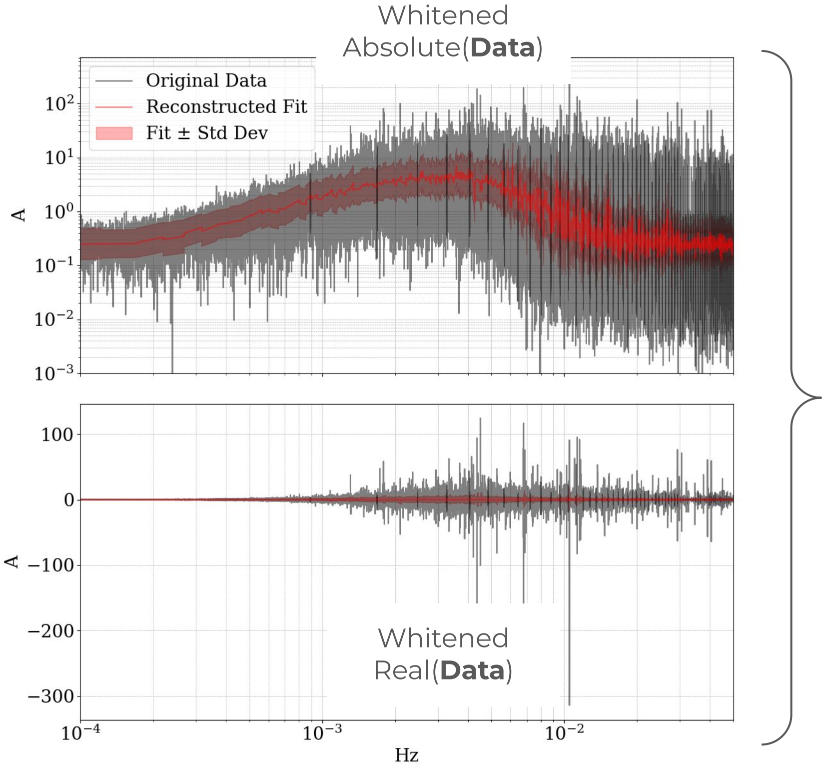
A network with O(10^6) input neurons is very memory intensive and data hungry.
Need to summarize the data.

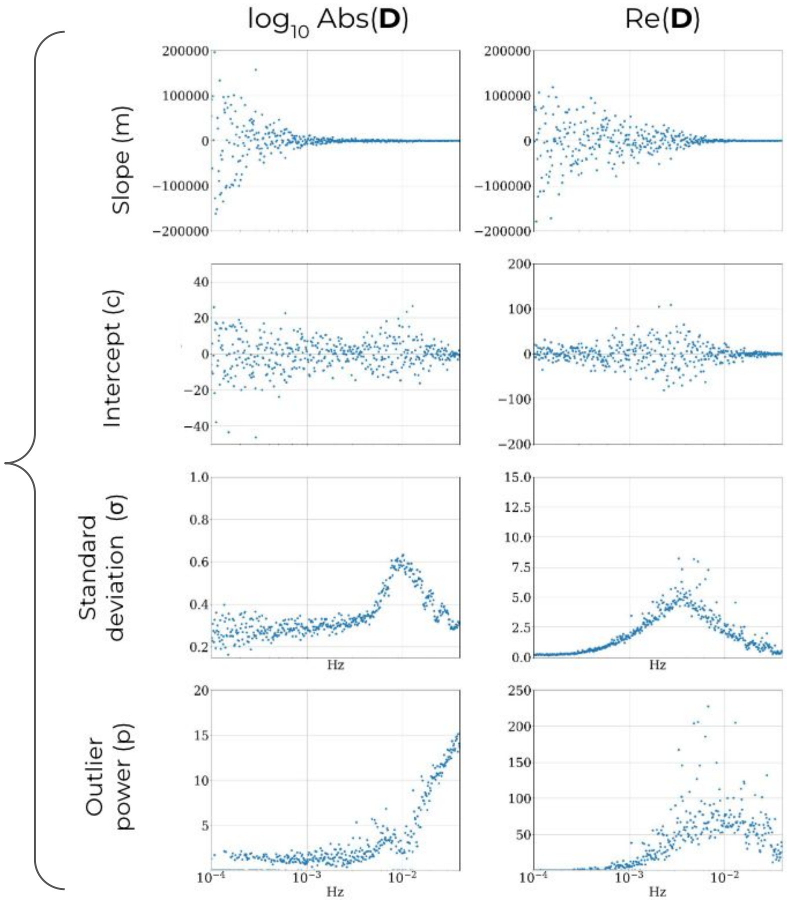
Our *current* prescription:

Linear fit of data within frequency batches
Summary: fit and residual parameters.

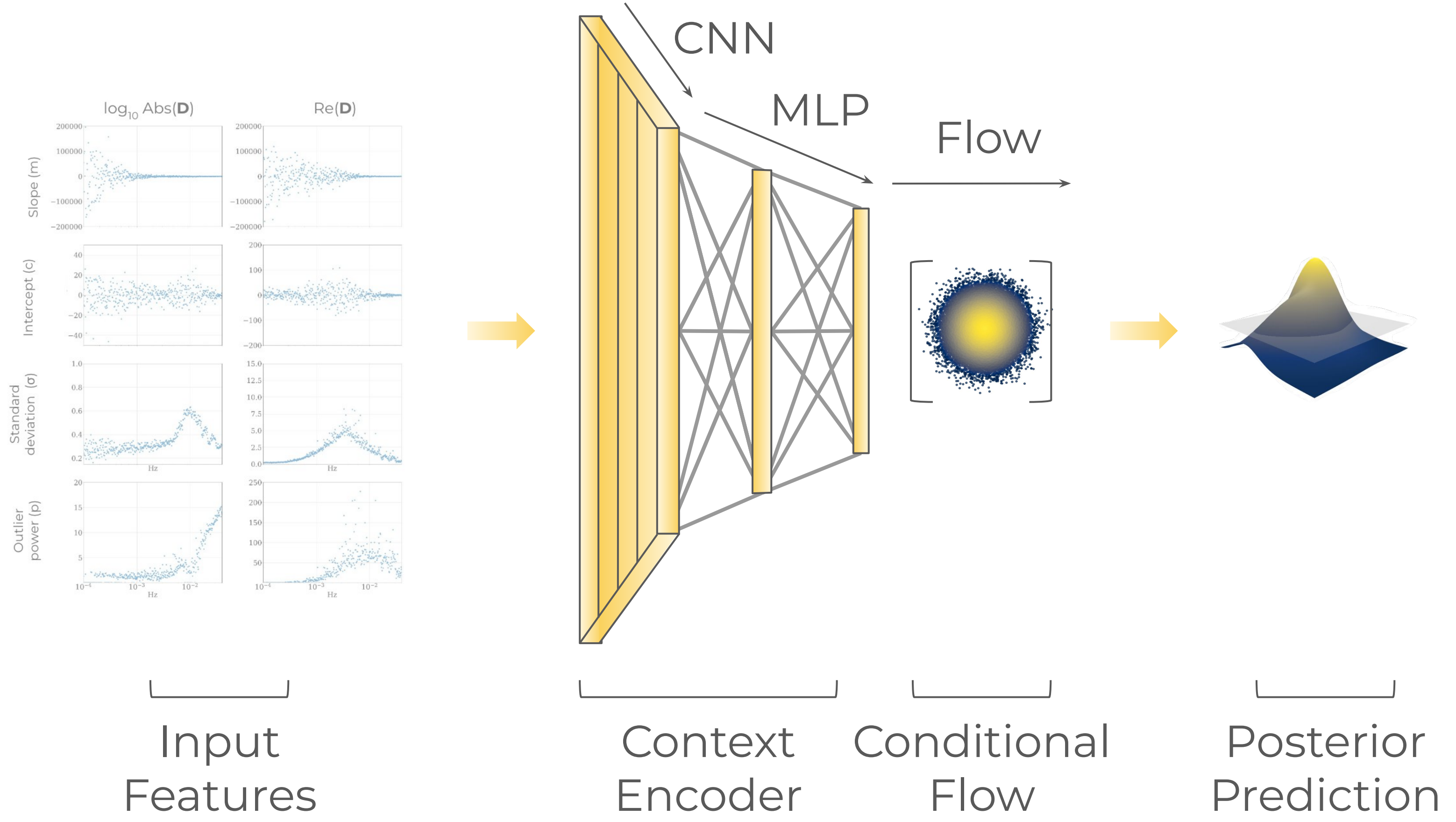
Data reduction: O(10^6) → O(10^3)

Data summary

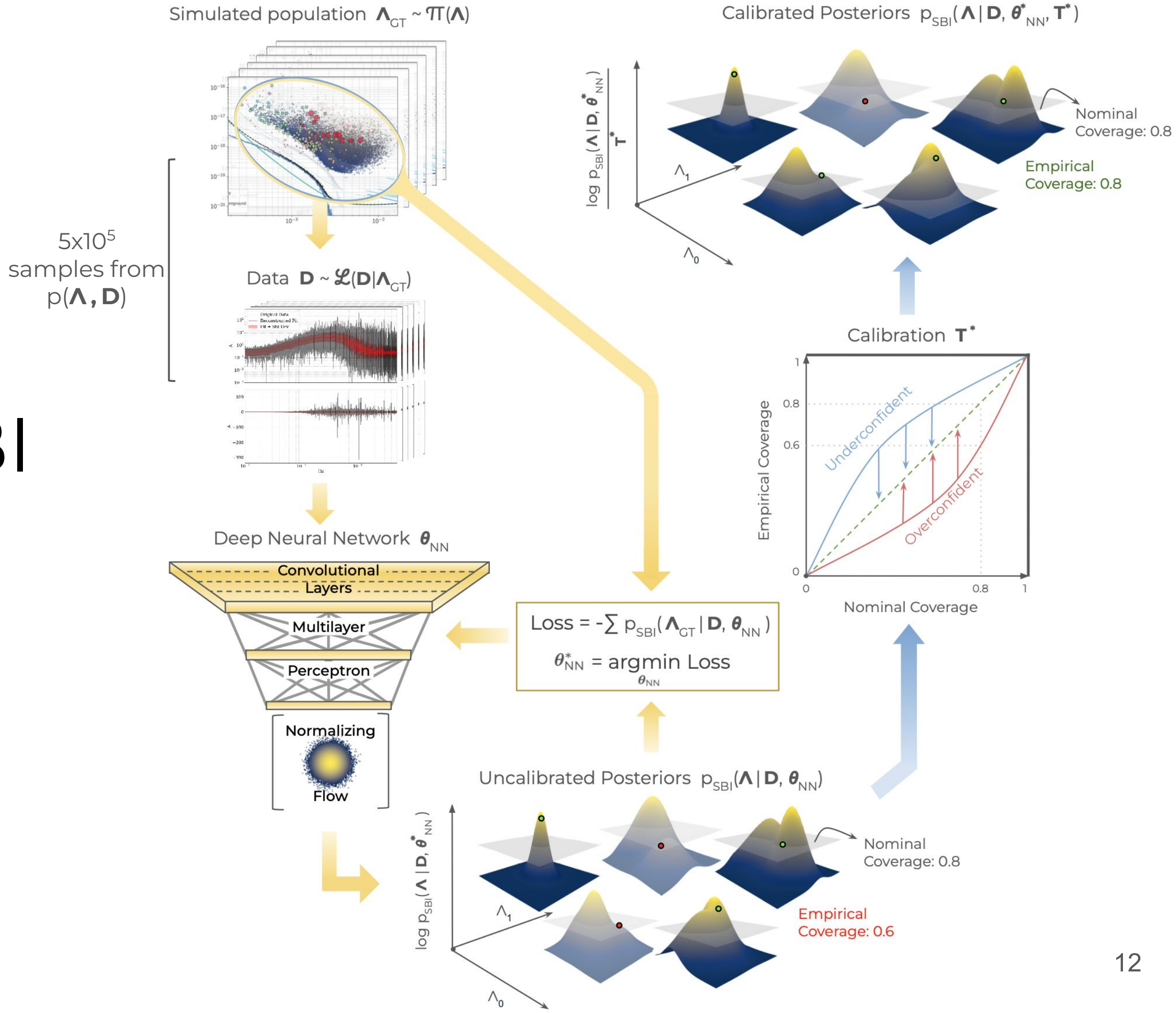


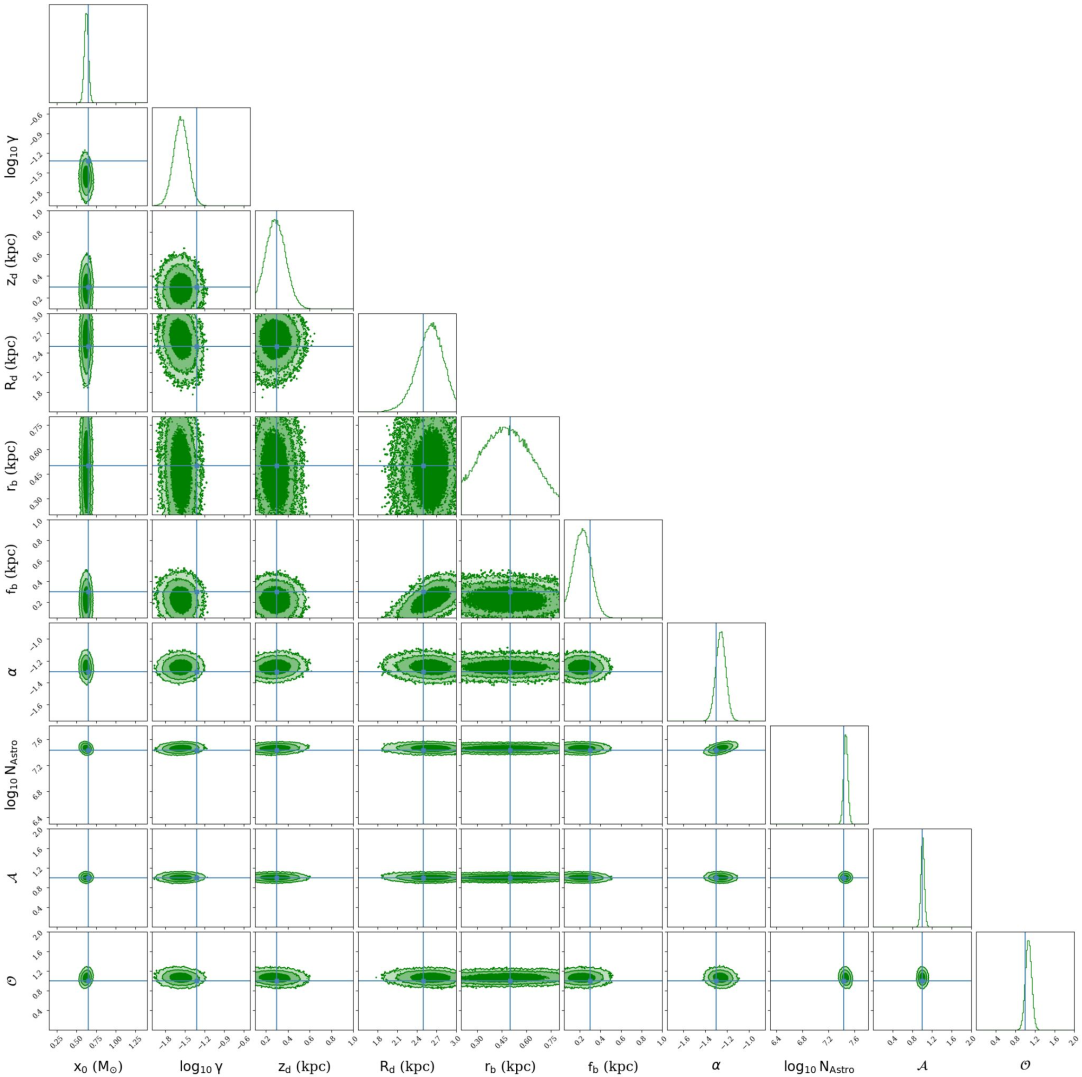


Network Architecture

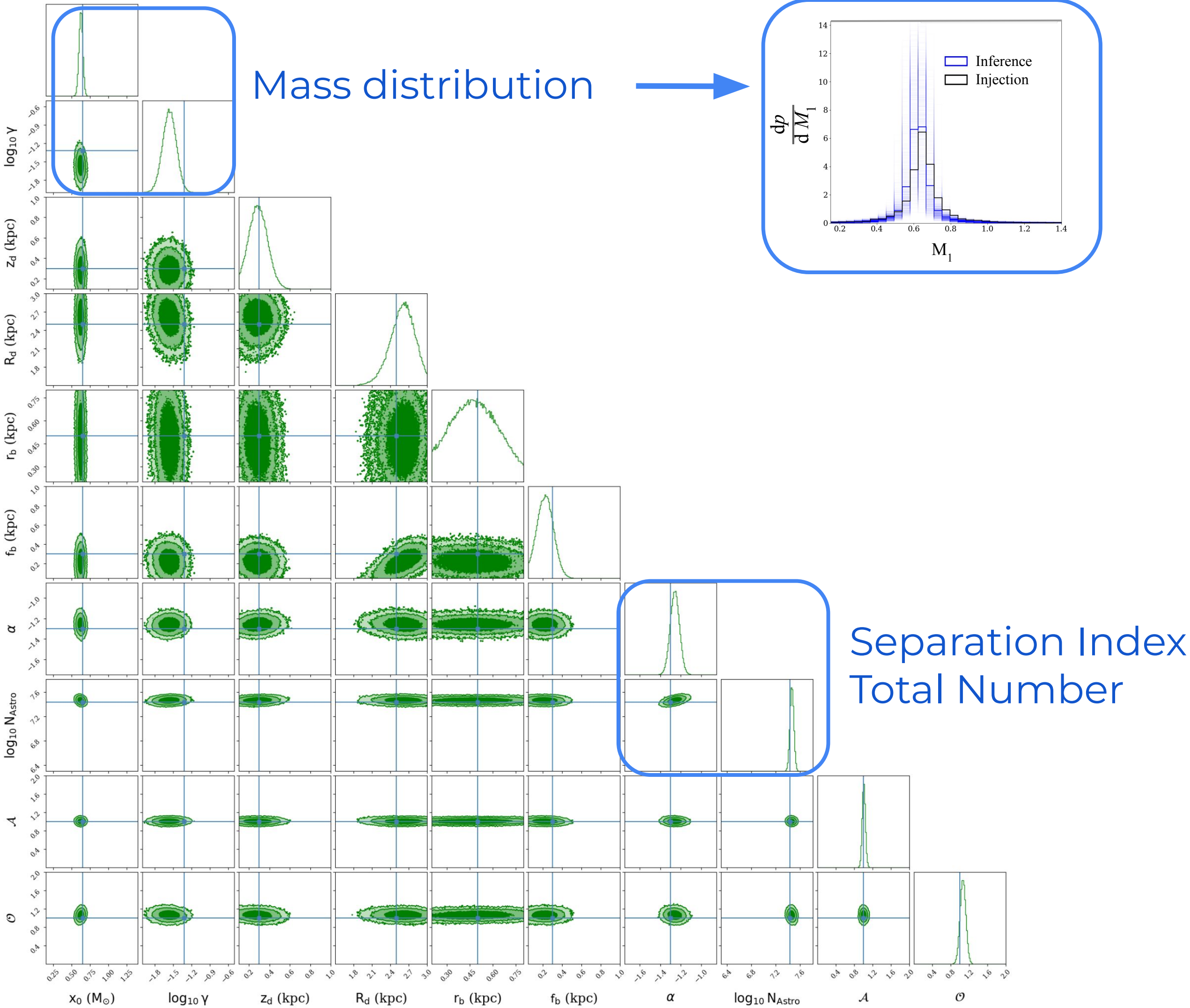


Training the SBI

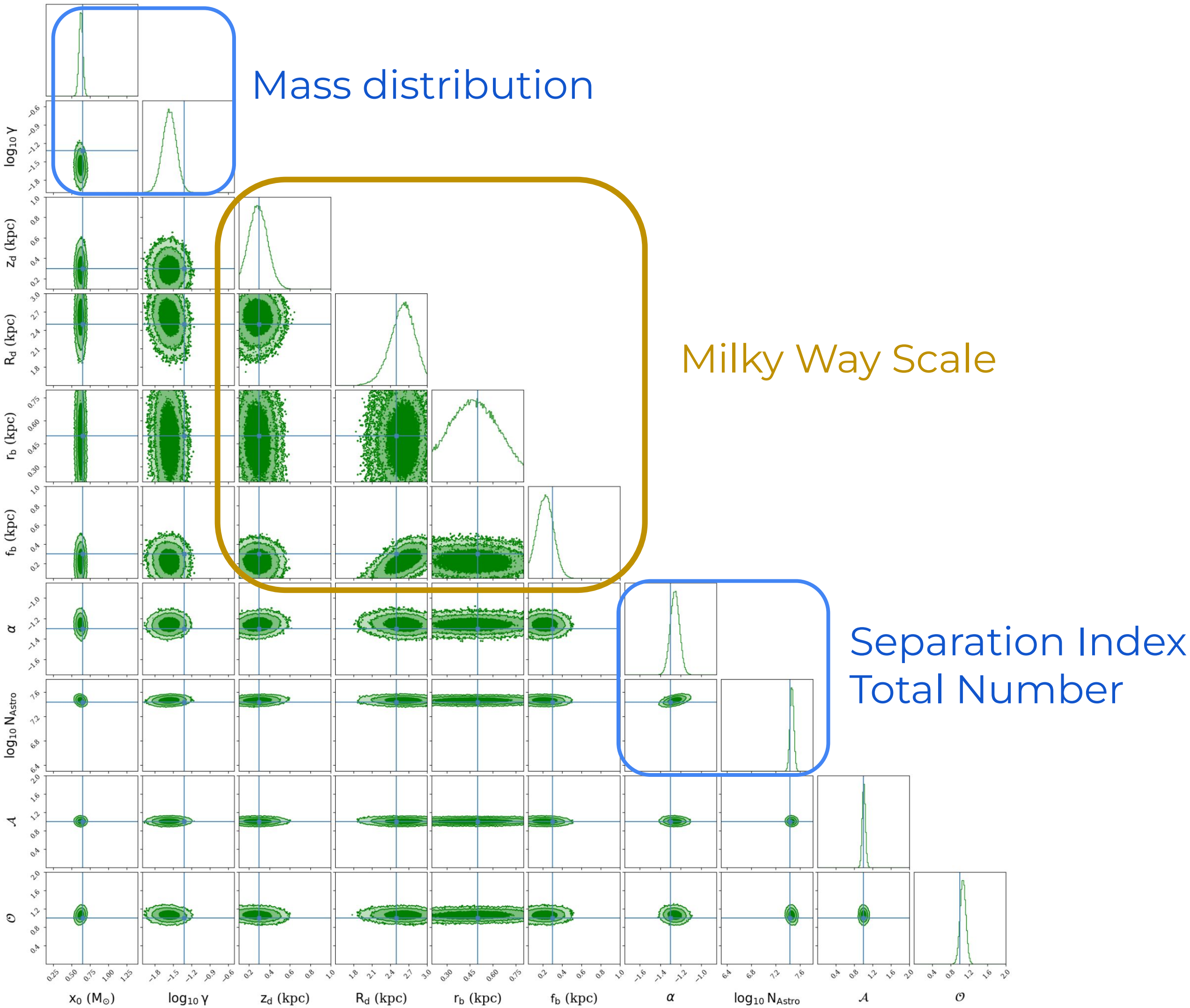




Galactic Double White Dwarf Population Inference

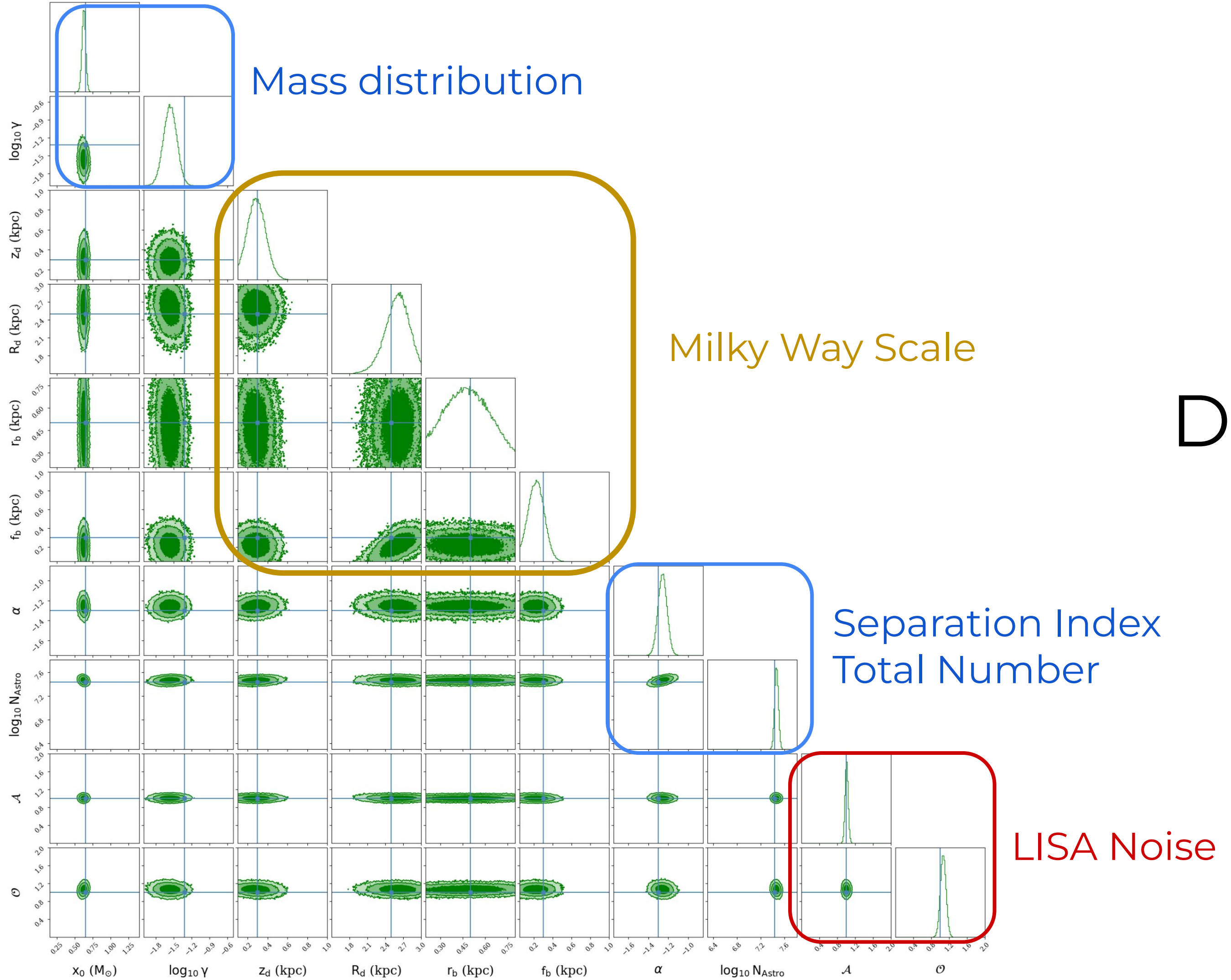


Galactic Double White Dwarf Population Inference



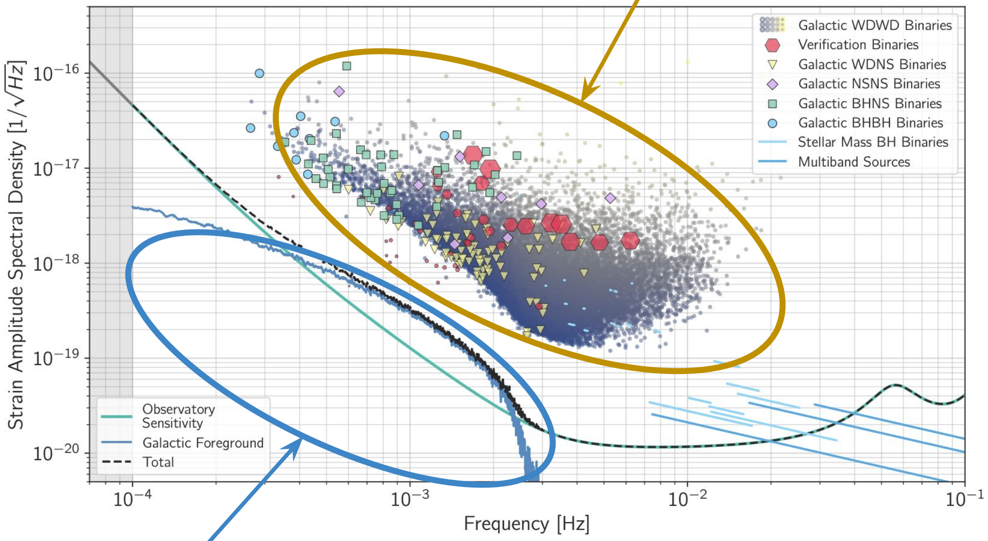
Galactic Double White Dwarf Population Inference

Galactic Double White Dwarf Population Inference



But, where lies the information?

O(30,000)



O(30,000,**000**)!

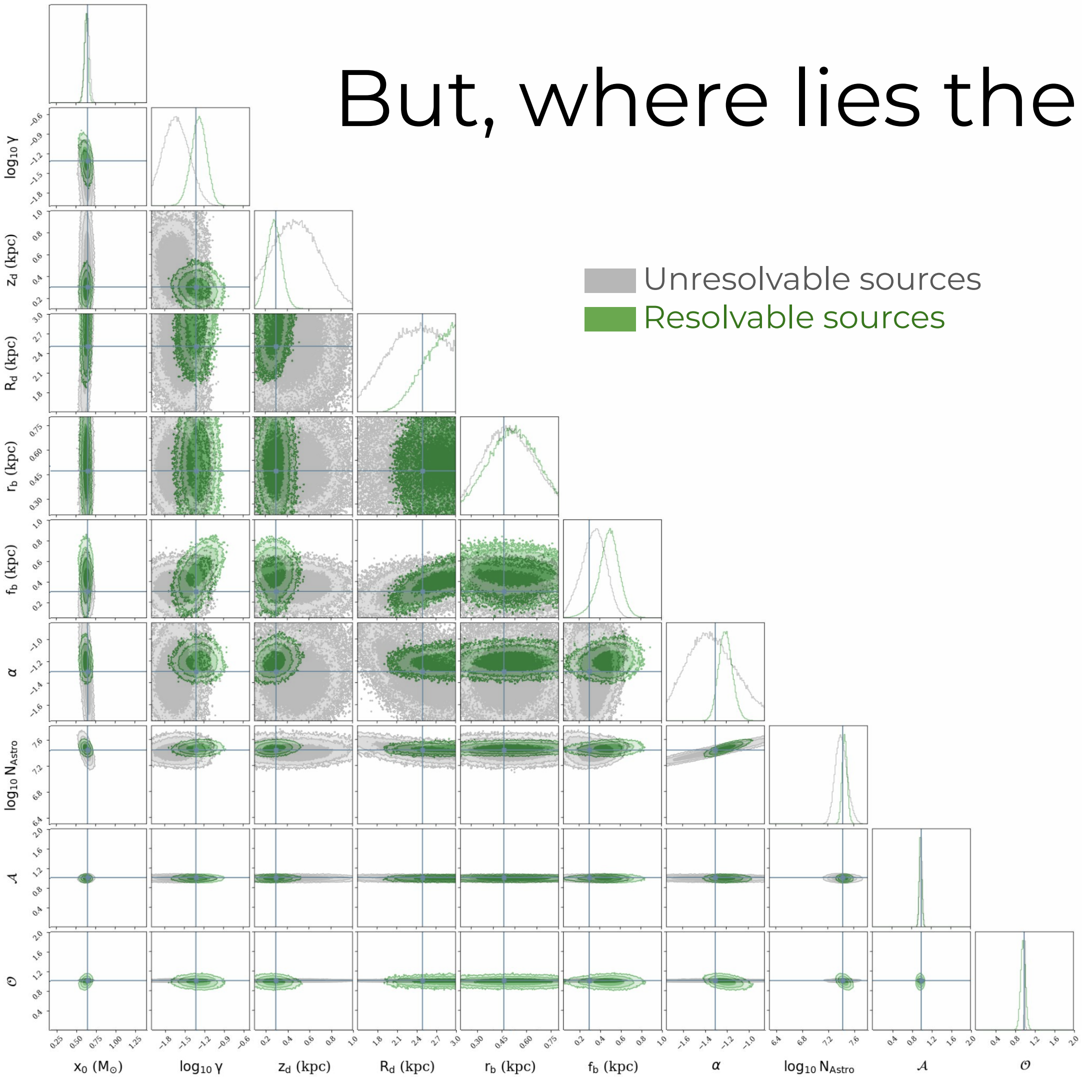
Figure from the LISA red book

Where's the population signature?

the few resolvable sources or

the many unresolvable sources ?

But, where lies the information?



Probably the resolvable !

(at least for this particular data summary)

Paper & codes soon to be made public!

Caveats, caveats, caveats

* The inference can only be as good as the simulator/model $p(\mathbf{\Lambda}|\mathbf{D}, \mathbf{M}_{\text{Simulator}})$.

* The inference can only be as good as the data summary.

* The intractable (or at least hard-to-pose) likelihood prevents testing SBI for convergence to an "optimal" posterior.

* The SBI can often be underconfident: Calibration <u>required</u>.