Gravitational-wave population models with deep learning

Matthew Mould
University of Birmingham, UK
mmould@star.sr.bham.ac.uk

arXiv:2205.12329

Davide Gerosa
(Milan-Bicocca)

Stephen R. Taylor
(Vanderbilt)

UNIVERSITY OF BIRMINGHAM

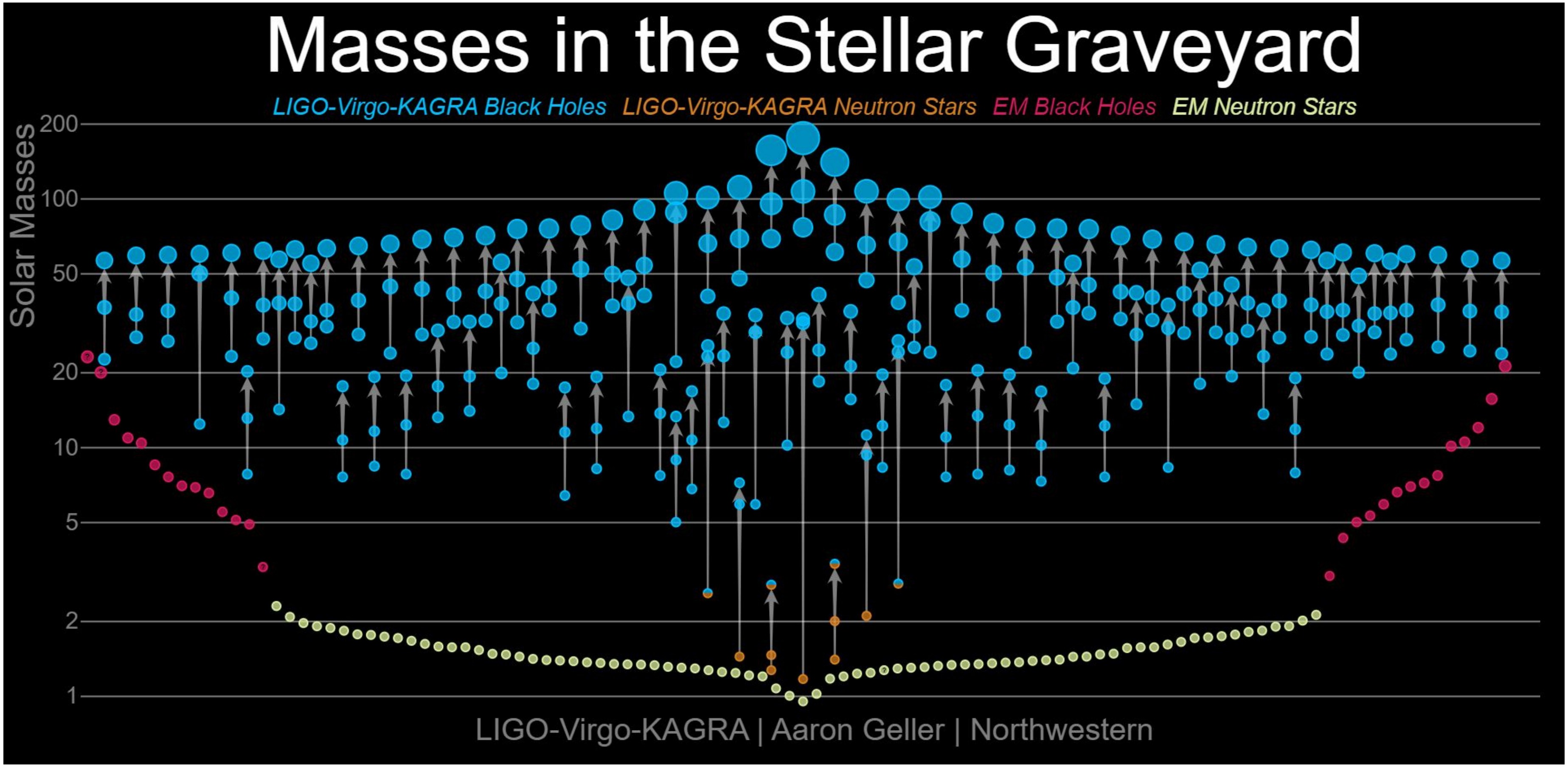
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

G2NET

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

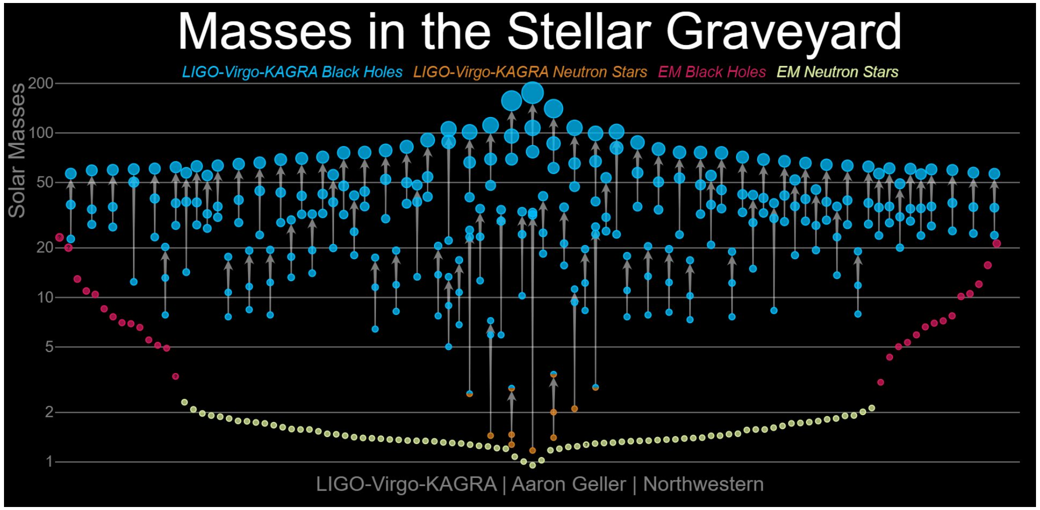
G2NET
next challenges
29/09/2022

Population inference with GWTC-3 data



[LIGO-Virgo-KAGRA / Aaron Geller / Northwestern]()

Population inference with GWTC-3 data



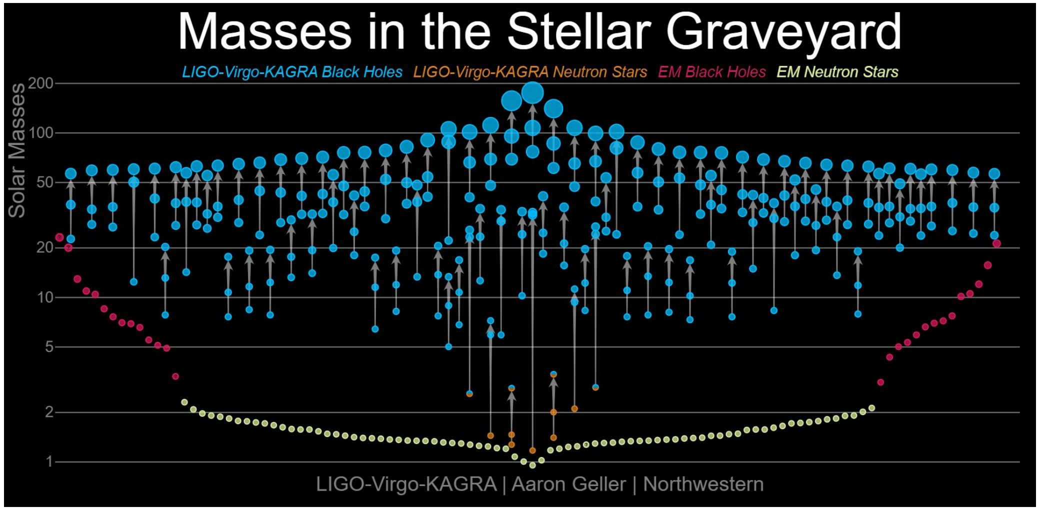
[LIGO-Virgo-KAGRA / Aaron Geller / Northwestern]

Infer the merger properties:

$$p(\boldsymbol{\theta}_n|\boldsymbol{d}_n) = \frac{\pi(\boldsymbol{\theta}_n)\mathcal{L}(\boldsymbol{d}_n|\boldsymbol{\theta}_n)}{\mathcal{Z}(\boldsymbol{d}_n)}$$

$$\boldsymbol{\theta}_n = (\text{masses}, \text{spins}, \ldots)_n$$

Population inference with GWTC-3 data



[LIGO-Virgo-KAGRA / Aaron Geller / Northwestern]

Infer the merger properties:

$$p(\boldsymbol{\theta}_n|\boldsymbol{d}_n) = \frac{\pi(\boldsymbol{\theta}_n)\mathcal{L}(\boldsymbol{d}_n|\boldsymbol{\theta}_n)}{\mathcal{Z}(\boldsymbol{d}_n)}$$

$$\boldsymbol{\theta}_n = (\text{masses}, \text{spins}, \ldots)_n$$

Stack events together and account for detection biases to **infer the population properties:**

$$\mathcal{L}(\{\boldsymbol{d}_n\}|\boldsymbol{\lambda}) = \prod_{n=1}^{N} \frac{\int \mathcal{L}(\boldsymbol{d}_n|\boldsymbol{\theta}_n)\pi(\boldsymbol{\theta}_n|\boldsymbol{\lambda})\mathrm{d}\boldsymbol{\theta}_n}{\int P(\text{detect}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})\mathrm{d}\boldsymbol{\theta}}$$

Population models head-to-head

An example **phenomenological** model: $\boldsymbol{\theta} = m\,,\ \boldsymbol{\lambda} = \alpha$

- Mass gap between neutron stars and black holes
- Mass gap above the PISN threshold
- IMF is a power law

$$\pi(m|\alpha) \propto m^{\alpha}$$

Population models head-to-head

An example **phenomenological** model: $\boldsymbol{\theta} = m\,,\ \boldsymbol{\lambda} = \alpha$

- Mass gap between neutron stars and black holes
- Mass gap above the PISN threshold
- IMF is a power law

$$\pi(m|\alpha) \propto m^{\alpha}$$

Simulation-based model: $\boldsymbol{\theta} = m\,,\ \boldsymbol{\lambda} = \alpha_{\mathrm{CE}}$

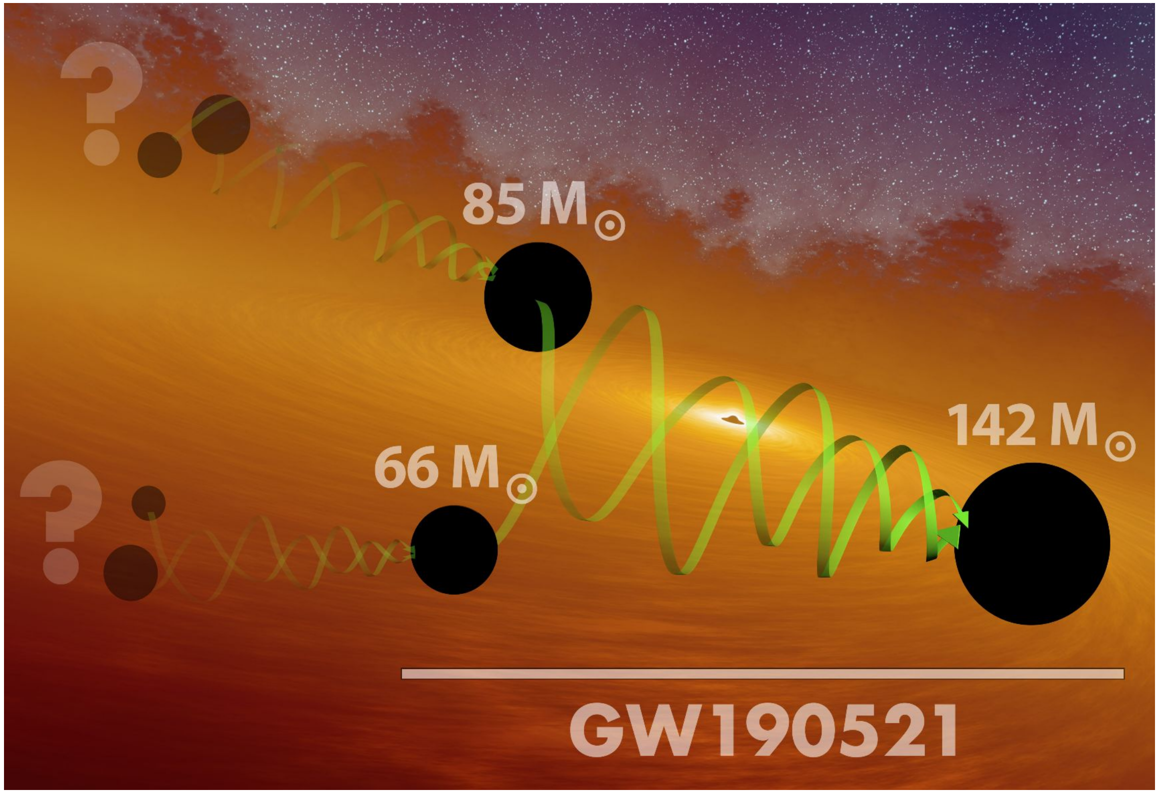
Interpolation

$$\begin{bmatrix} \alpha_{\mathrm{CE},1} \\ \vdots \\ \alpha_{\mathrm{CE},n} \end{bmatrix} \Rightarrow$$

Simulator
e.g.,
pop. synth.

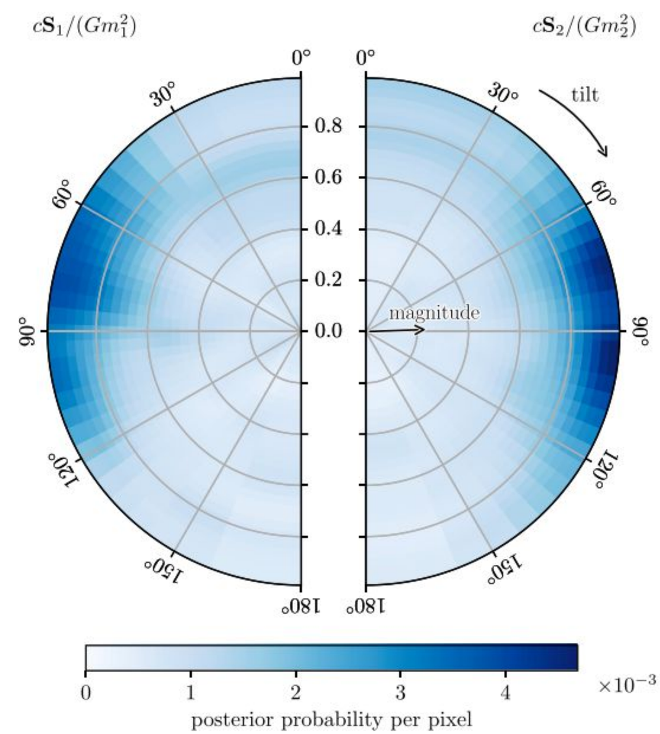
$$\Rightarrow \begin{bmatrix} \{m\}_1 \\ \vdots \\ \{m\}_n \end{bmatrix} \sim \begin{bmatrix} \pi(m|\alpha_{\mathrm{CE},1}) \\ \vdots \\ \pi(m|\alpha_{\mathrm{CE},n}) \end{bmatrix}$$

How do the heaviest BBHs form?

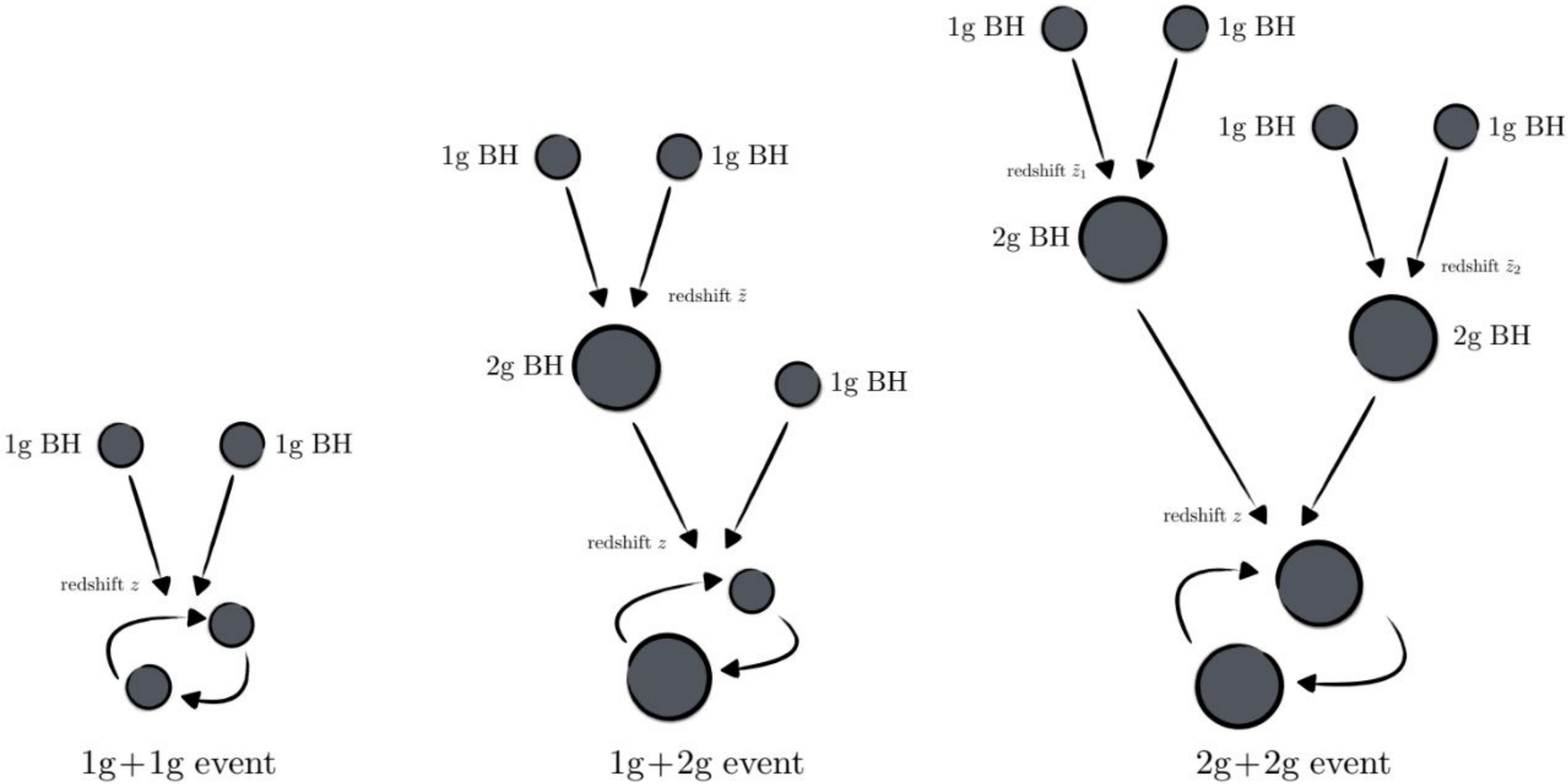


[LIGO/Caltech/MIT/R. Hurt (IPAC)]

[LVC 2020]



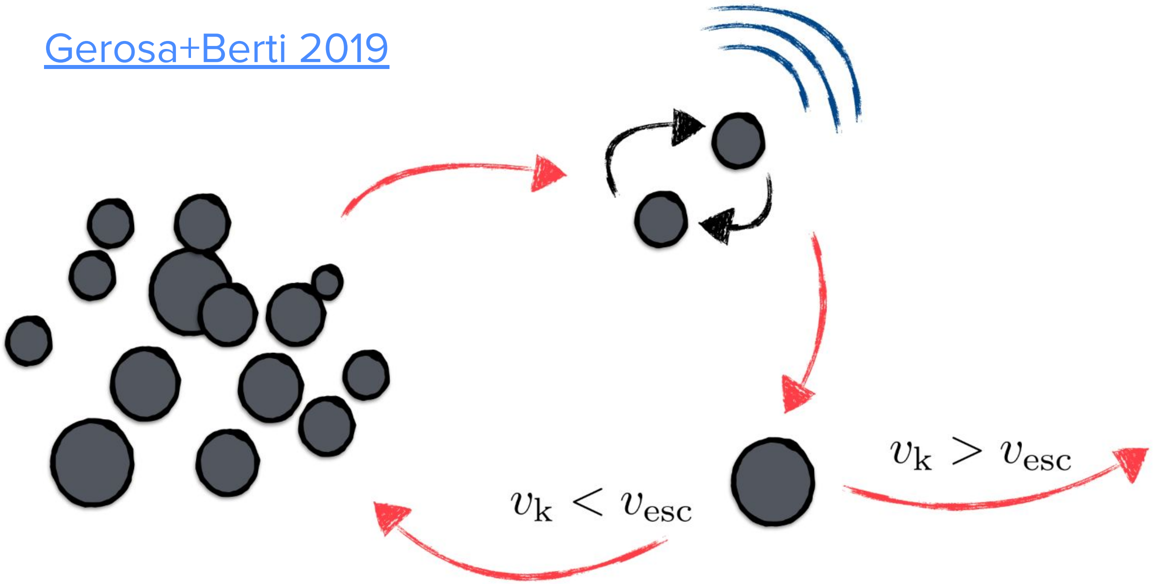
Are GW black holes hierarchical?



Gerosa+Berti 2017

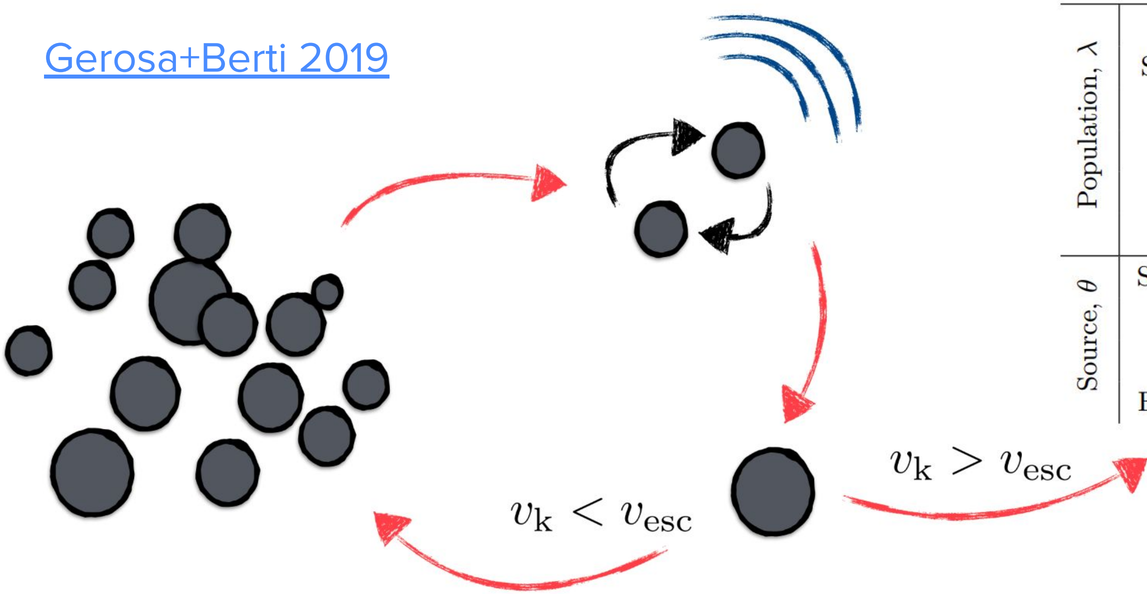
Simple model of hierarchical mergers

Gerosa+Berti 2019



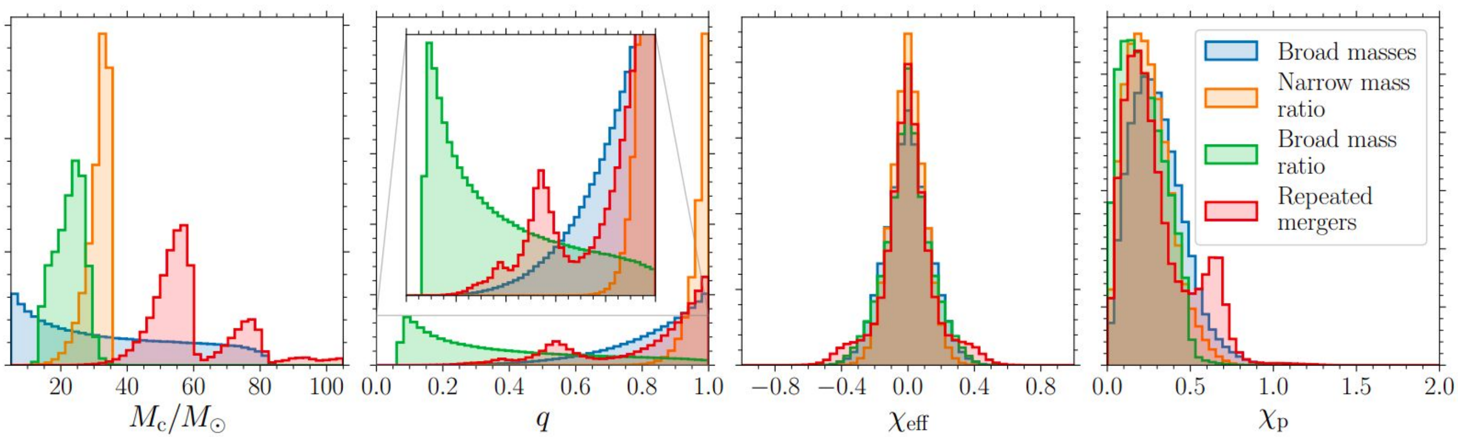
Simple model of hierarchical mergers

Gerosa+Berti 2019



	Parameter	Symbol	Range
Population, λ	Primary pairing slope	α	$[-10, 10]$
	Secondary pairing slope	β	$[-10, 10]$
	1g mass slope	γ	$[-10, 10]$
	Escape speed slope	δ	$[-10, 10]$
	Maximum 1g mass	m_{max}	$[30M_\odot, 100M_\odot]$
	Maximum 1g spin	χ_{max}	$[0, 1]$
Source, θ	Source-frame chirp mass	M_c	$[5M_\odot, 105M_\odot]$
	Mass ratio	q	$[0, 1]$
	Effective aligned spin	χ_{eff}	$[-1, 1]$
	Effective precessing spin	χ_p	$[0, 2]$

Gerosa, **MM**+ 2021



Emulating the population

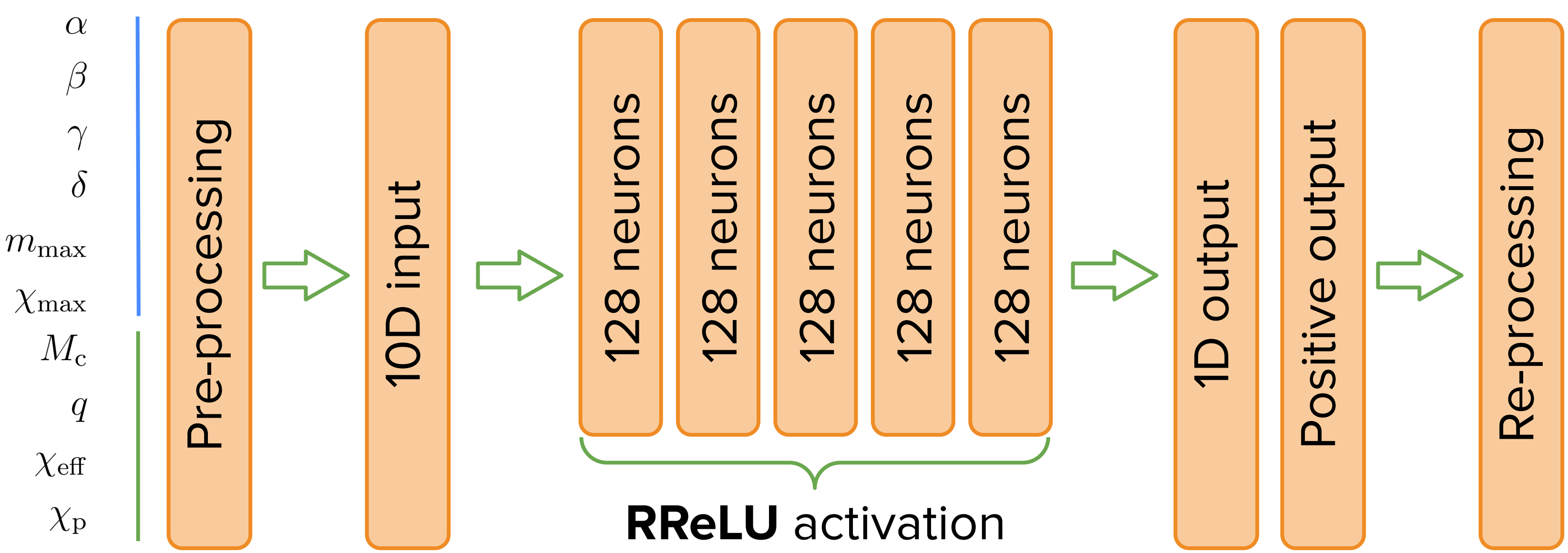
- Estimate each population density with KDE

$$\boldsymbol{\theta} \mapsto \pi(\boldsymbol{\theta}|\boldsymbol{\lambda}^i),\; i = 1, \ldots, 1000$$

(KDEpy)
(Botev+ 2010)

- Learn the full mapping with a **deep neural network** (TensorFlow)

$$\boldsymbol{\theta}, \boldsymbol{\lambda} \mapsto \pi(\boldsymbol{\theta}|\boldsymbol{\lambda})$$



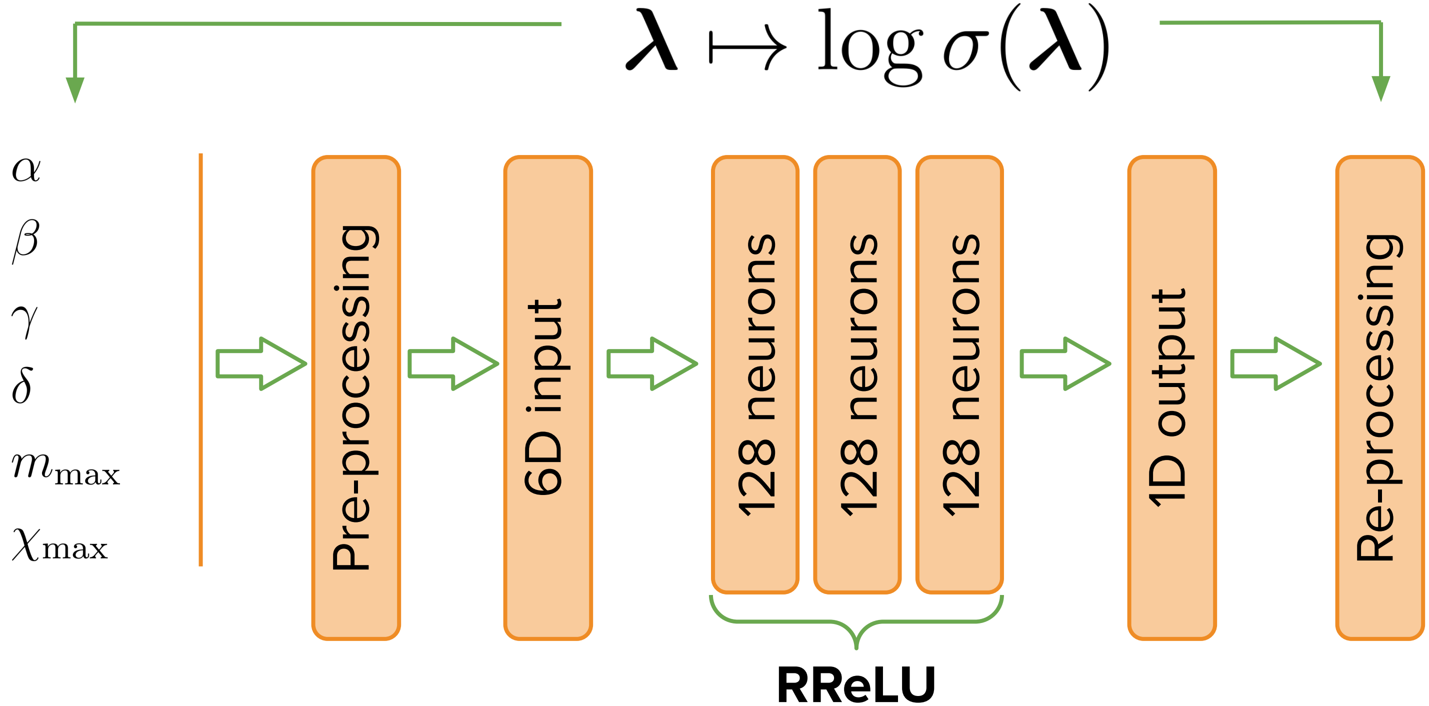
Emulating the detection efficiency

- Estimate detection efficiency for each simulation (Finn+Chernoff 1993)

$$\sigma(\boldsymbol{\lambda}^i) = \int P(\mathrm{detect}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda}^i)\mathrm{d}\boldsymbol{\theta} = \frac{1}{N^i}\sum_{j=1}^{N^i} P(\mathrm{detect}|\boldsymbol{\theta}_j^i),\ i = 1, \ldots, 1000$$

- Learn the full mapping with a deep neural network (TensorFlow)

$$\boldsymbol{\lambda} \mapsto \log\sigma(\boldsymbol{\lambda})$$



- Ultimately work with the *log*-likelihood
- Also results in better learning

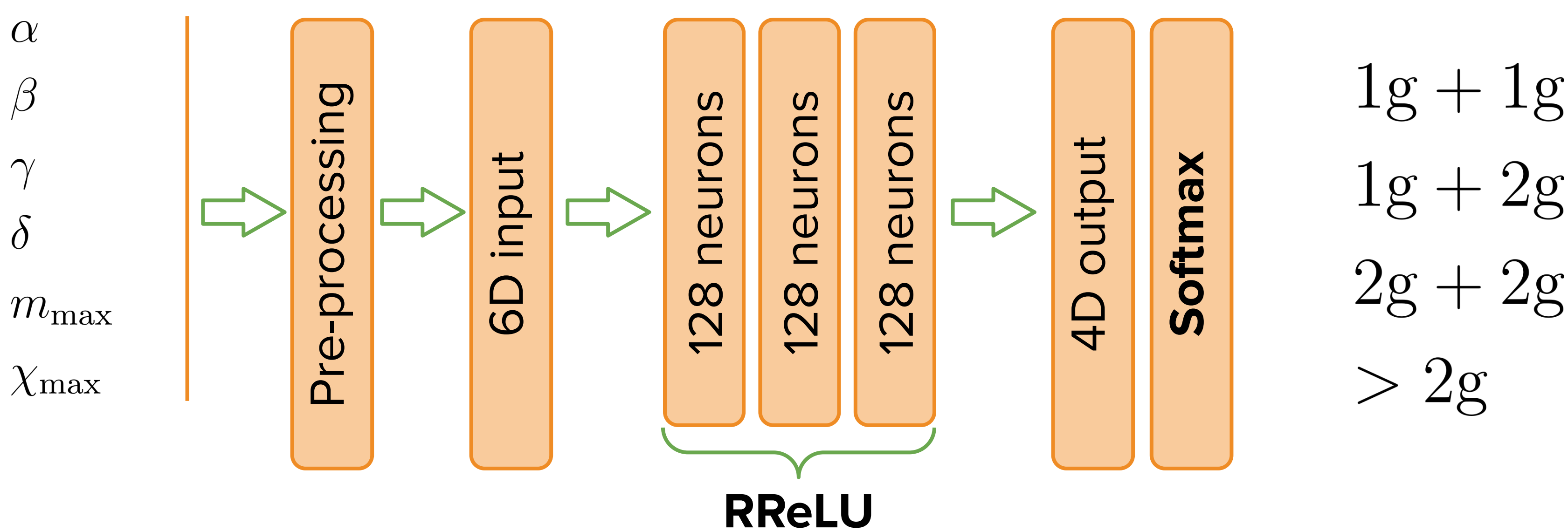
Emulating the branching fractions

- Compute fraction of sources in each merger generation $\sum_g f_g(\boldsymbol{\lambda}) \equiv 1$

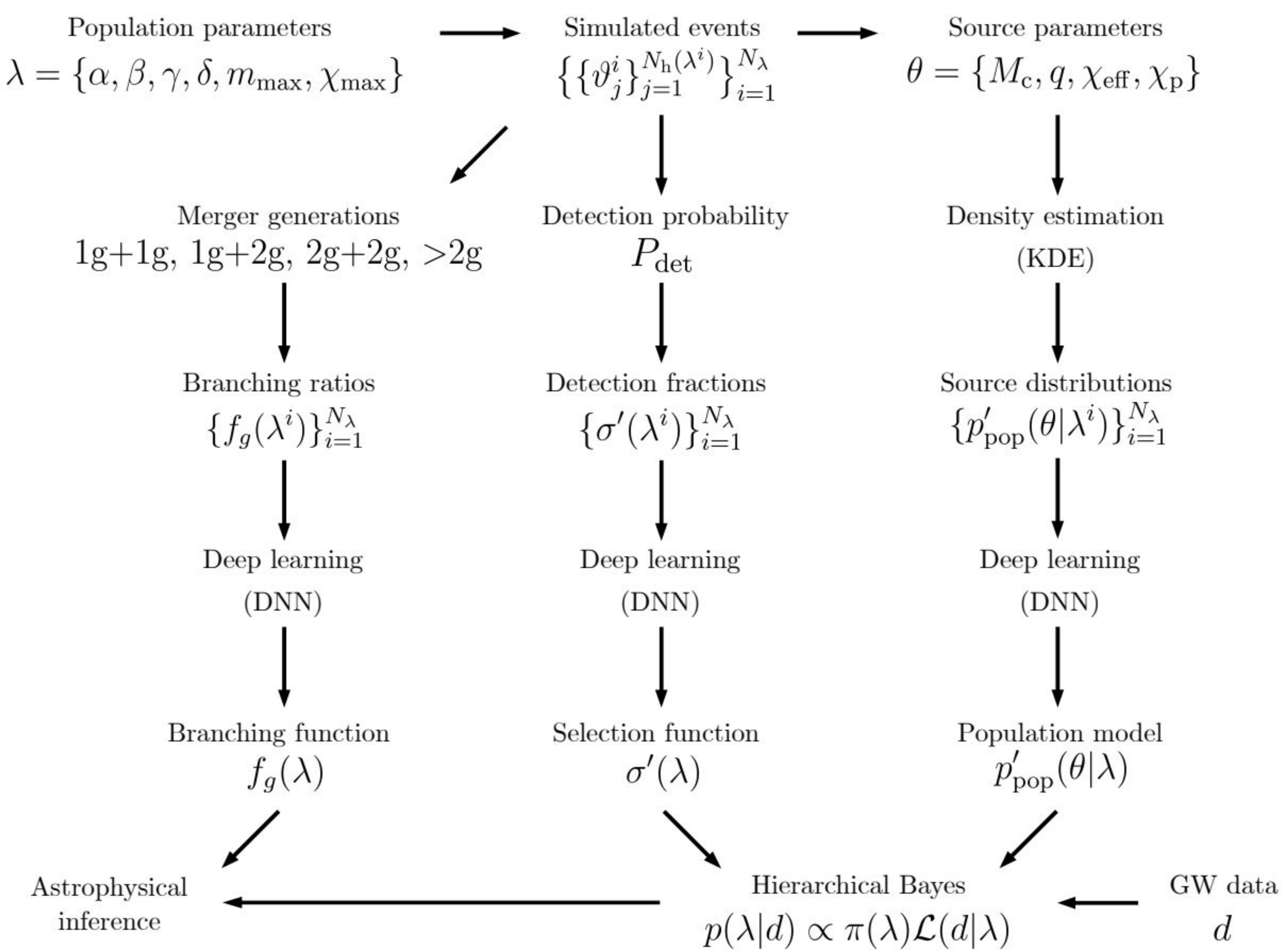
$$f_g(\boldsymbol{\lambda}^i) = N_g^i / N^i,\ i = 1, \ldots, 1000$$

- Learn the full mapping with a deep neural network [(TensorFlow)](TensorFlow)

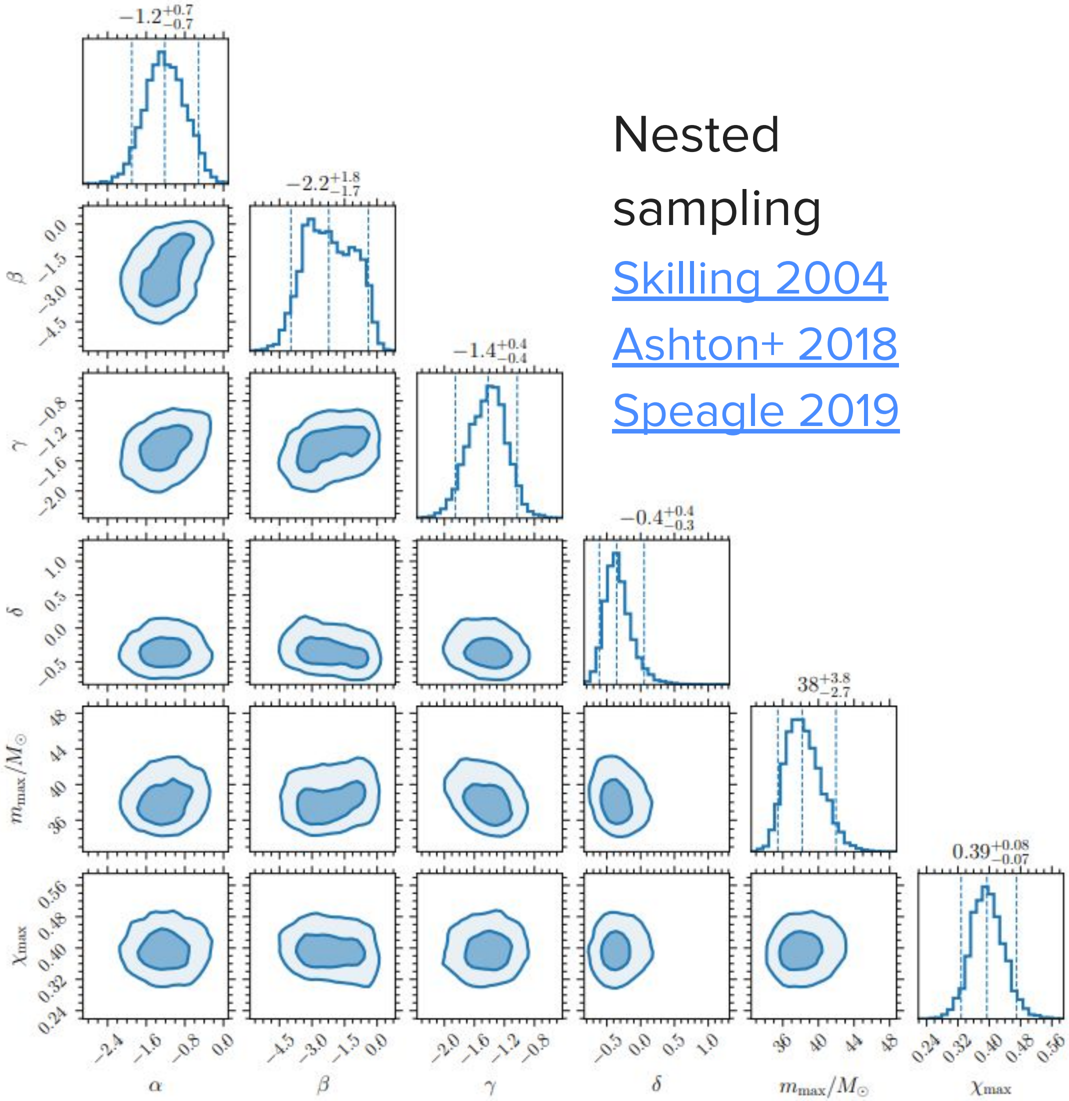
$$\boldsymbol{\lambda} \mapsto f_g(\boldsymbol{\lambda})$$



The full pipeline



Hyper-posterior from GWTC-3



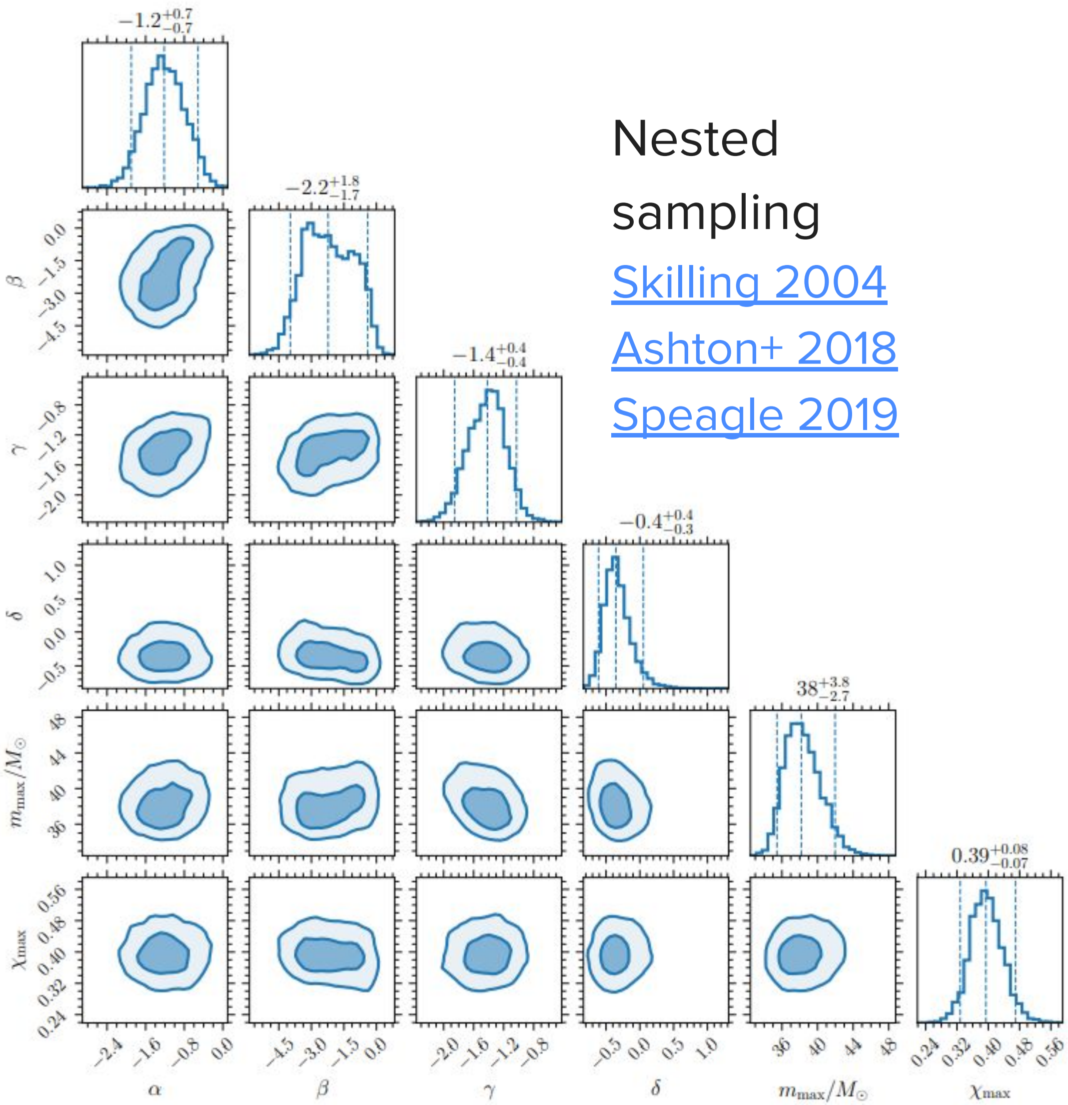
Nested sampling

[Skilling 2004](#)

[Ashton+ 2018](#)

[Speagle 2019](#)

Hyper-posterior from GWTC-3

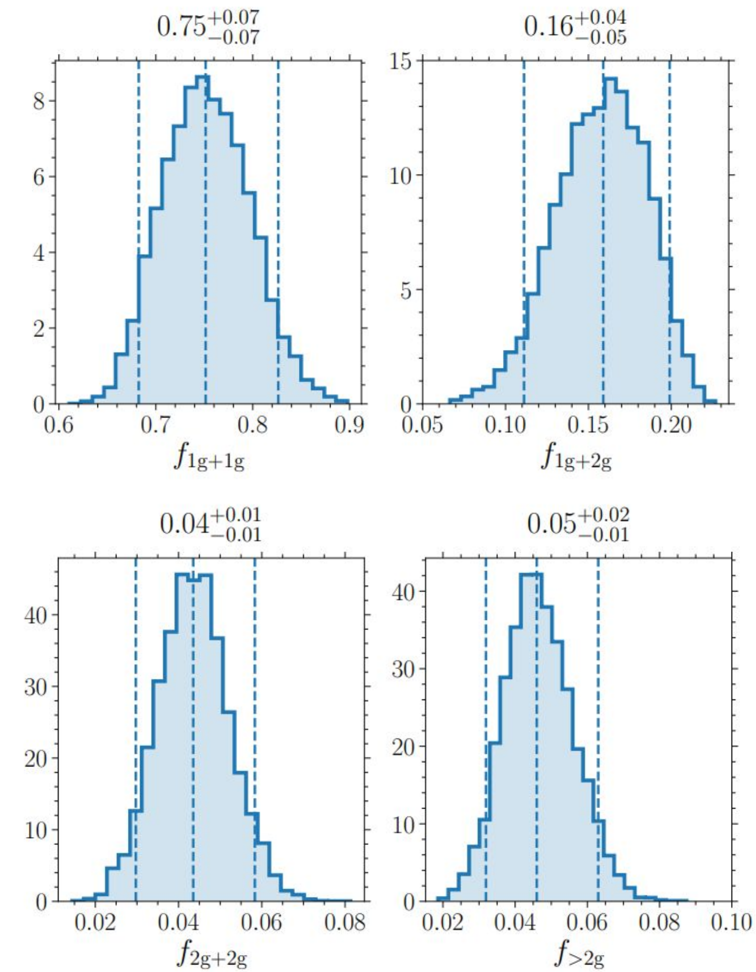


Nested sampling

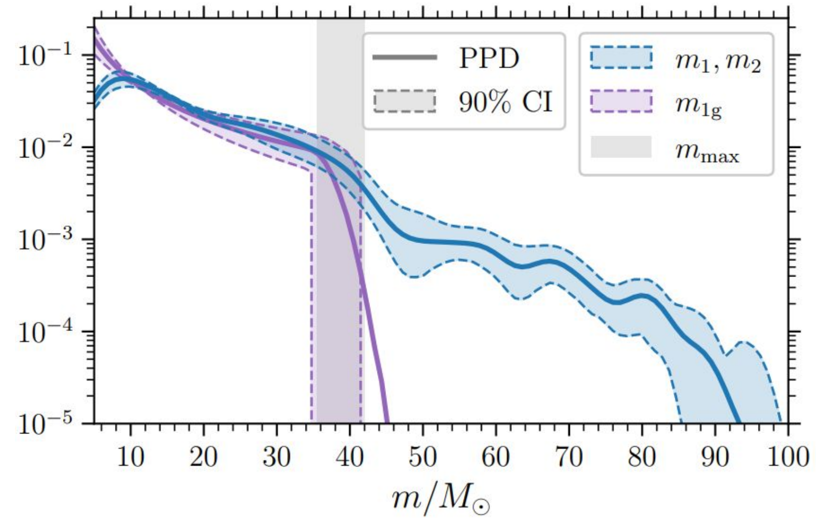
[Skilling 2004](#)

[Ashton+ 2018](#)

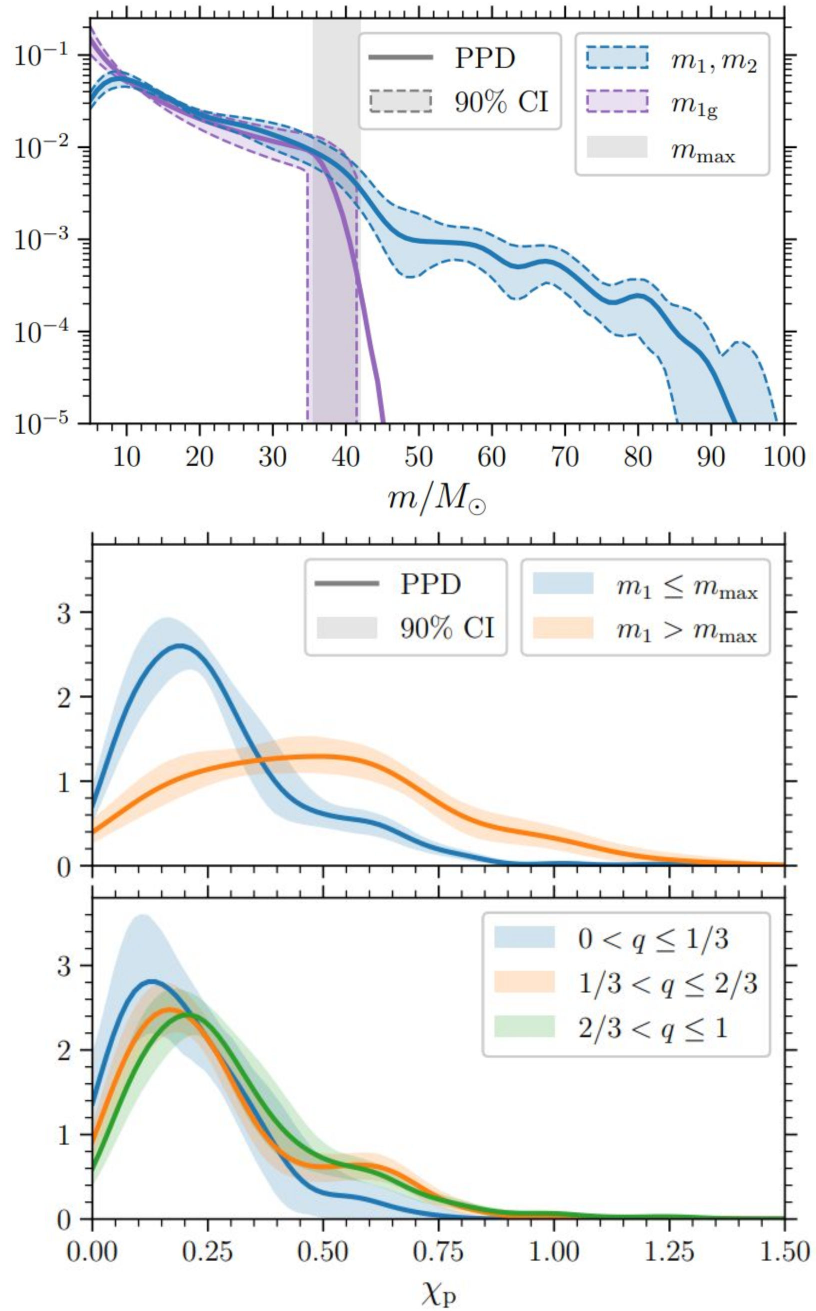
[Speagle 2019](#)



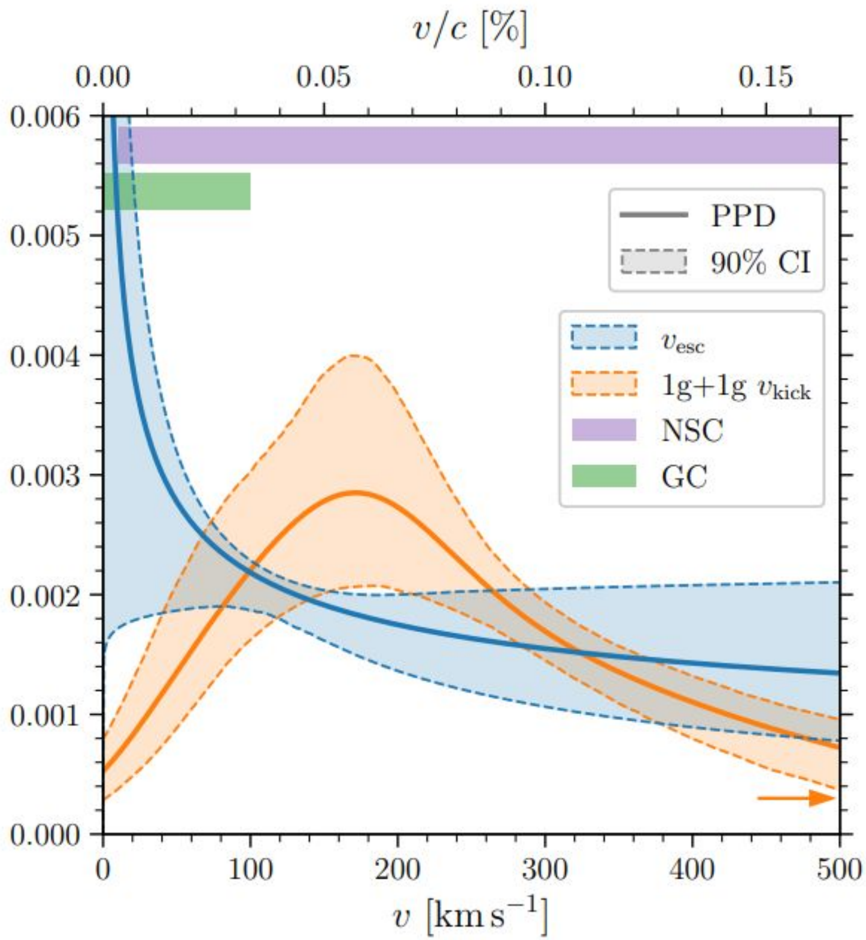
Population posterior from GWTC-3

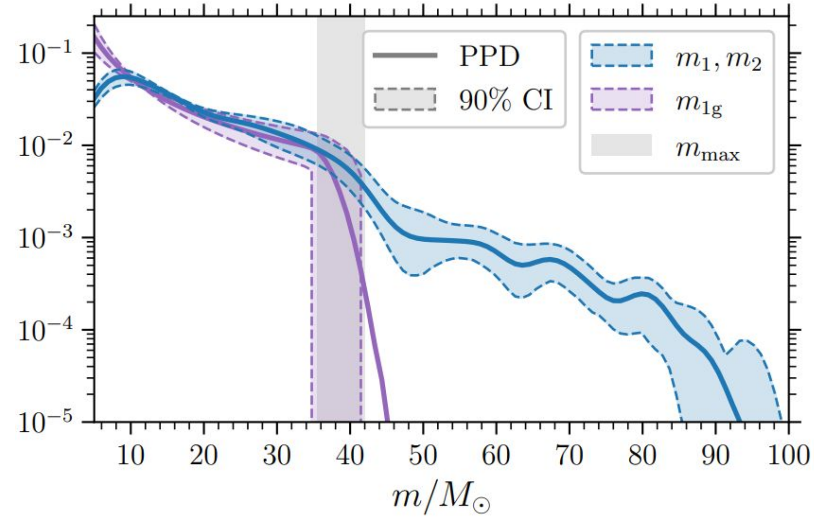


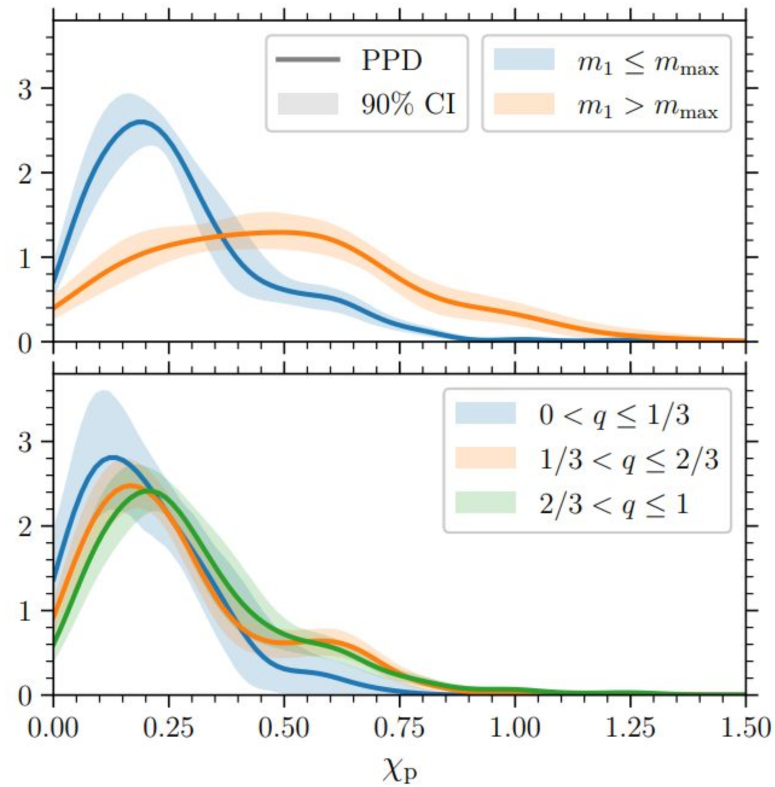
Population posterior from GWTC-3



Population posterior from GWTC-3







Gravitational-wave population models with deep learning

- **Emulation of simulated merger populations with deep learning.**
- **Some GWTC-3 features explanabile with hierarchical mergers.**
- **Simple numerical simulations for now, but easily applicable.**

- **Simulation-based sampling? Neural posterior estimation?**

  Only if the population distribution is not needed...
- **Model labelled subpopulations (e.g., generations) individually.**

  Un-normalized model / normalizing flow (Wong+ 2020) + classifier?
- **Learn the intrinsic population distributions without assumptions.**

  Non-parametric models (Rinaldi+ 2021), Bayesian normalizing flow?
- **Sayan Neogi (student): deep learning for merger remnants**

  Inverse model with uncertain predictions (Varma+ 2019, Haegel+ 2020)

Gravitational-wave population models with deep learning

Matthew Mould
University of Birmingham, UK
mmould@star.sr.bham.ac.uk

arXiv:2205.12329

Davide Gerosa
(Milan-Bicocca)

Stephen R. Taylor
(Vanderbilt)

UNIVERSITY OF BIRMINGHAM

UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

G2NET

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

G2NET next challenges
29/09/2022

Simulation summary

	Parameter	Symbol	Range
Population, λ	Primary pairing slope	α	$[-10, 10]$
	Secondary pairing slope	β	$[-10, 10]$
	1g mass slope	γ	$[-10, 10]$
	Escape speed slope	δ	$[-10, 10]$
	Maximum 1g mass	$m_{\max}$	$[30M_\odot, 100M_\odot]$
	Maximum 1g spin	$\chi_{\max}$	$[0, 1]$
Source, θ	Source-frame chirp mass	M_c	$[5M_\odot, 105M_\odot]$
	Mass ratio	q	$[0, 1]$
	Effective aligned spin	$\chi_{\rm eff}$	$[-1, 1]$
	Effective precessing spin	$\chi_{\rm p}$	$[0, 2]$

Perform 1000 simulations: generate hyperparameters with **Latin Hypercube Sampling**

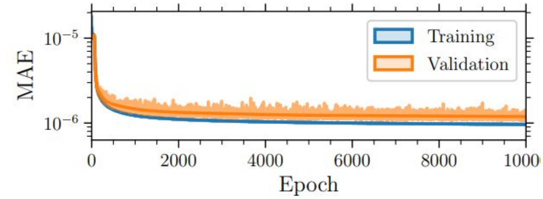
500 clusters per simulation

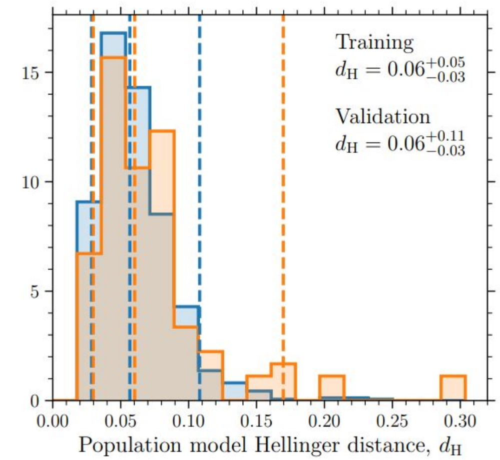
5000 seed black holes per cluster

(Gerosa+Berti 2019)

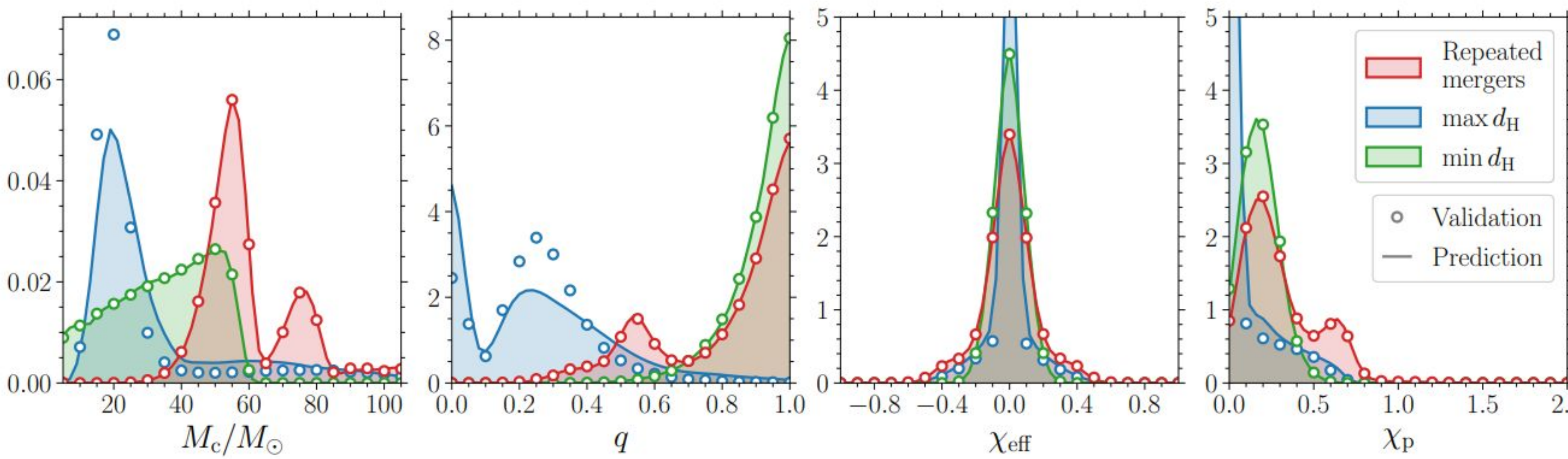
Deep-learned population

- NVIDIA A100 GPU (**Baskerville**)
- Trained for 10,000 epochs (4 days)
- 194,481,000 total samples (10D → 1D)
- 90%-10% training-validation split
- Batch size = 0.01% of training data
- Adam optimizer
- Learning rate = 0.0001
- Mean absolute error (MAE) loss function



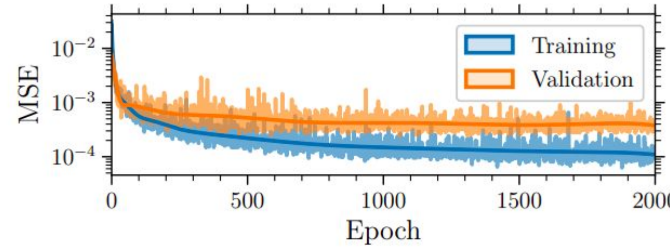


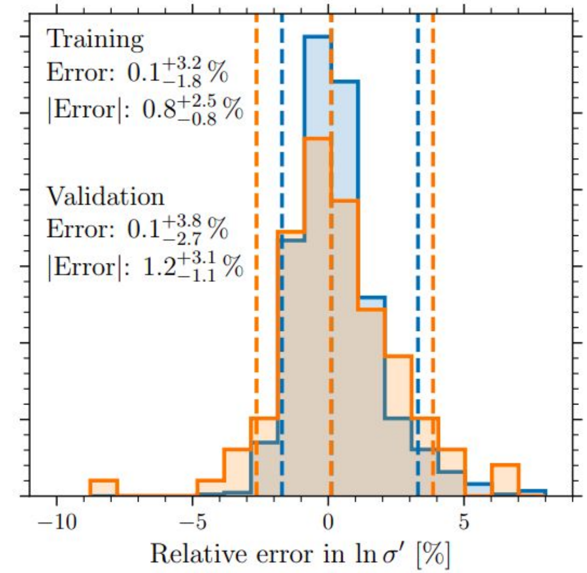
$$d_{\mathrm{H}}^2(p, q) = 1 - \int \sqrt{p(x)q(x)}\mathrm{d}x$$

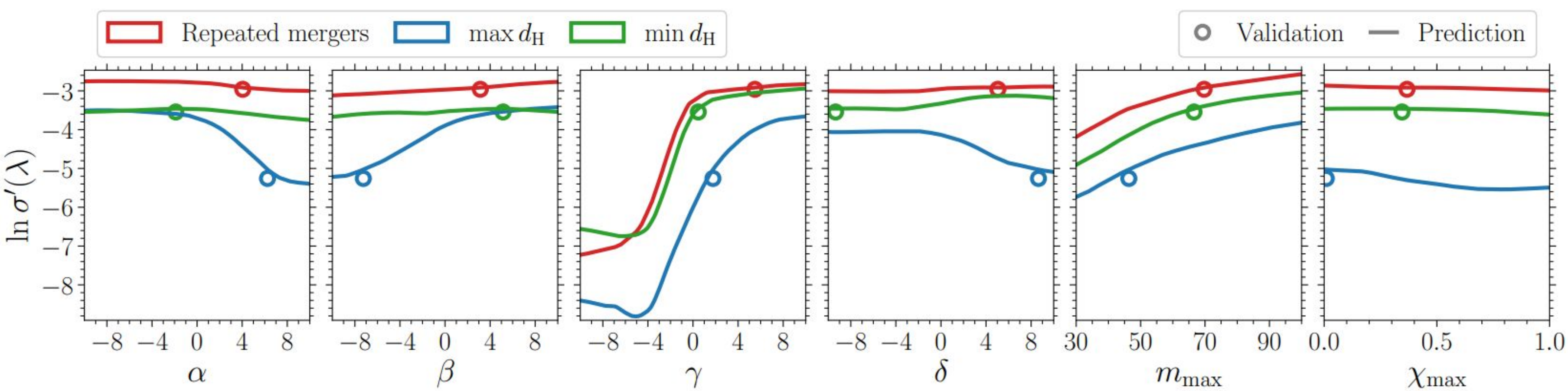


Deep-learned selection

- Trained on laptop
- 2,000 epochs (4 minutes)
- 1000 samples (6D → 1D)
- 90%-10% training-validation split
- Batch size = 1% of training data
- Adam optimizer
- Learning rate = 0.001
- Mean squared error (MSE) loss function



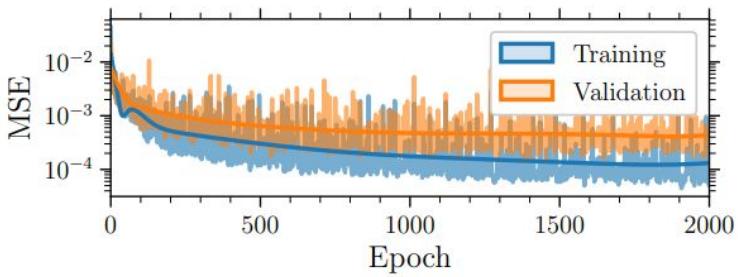


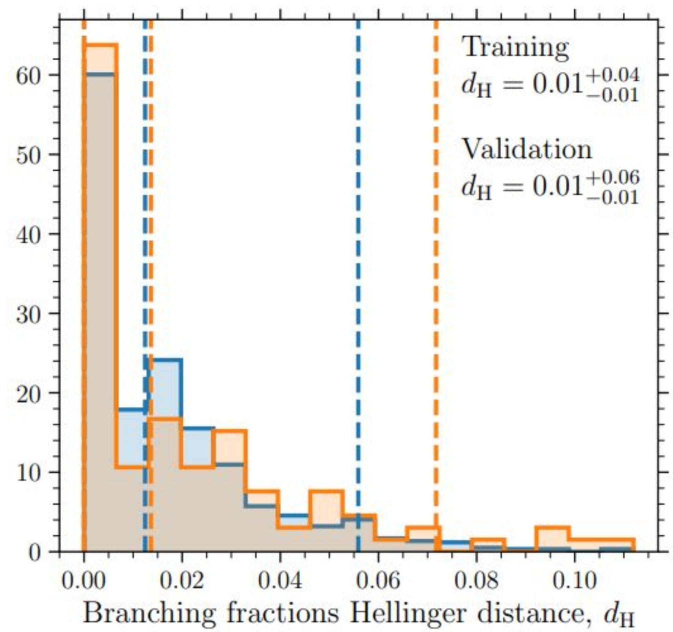


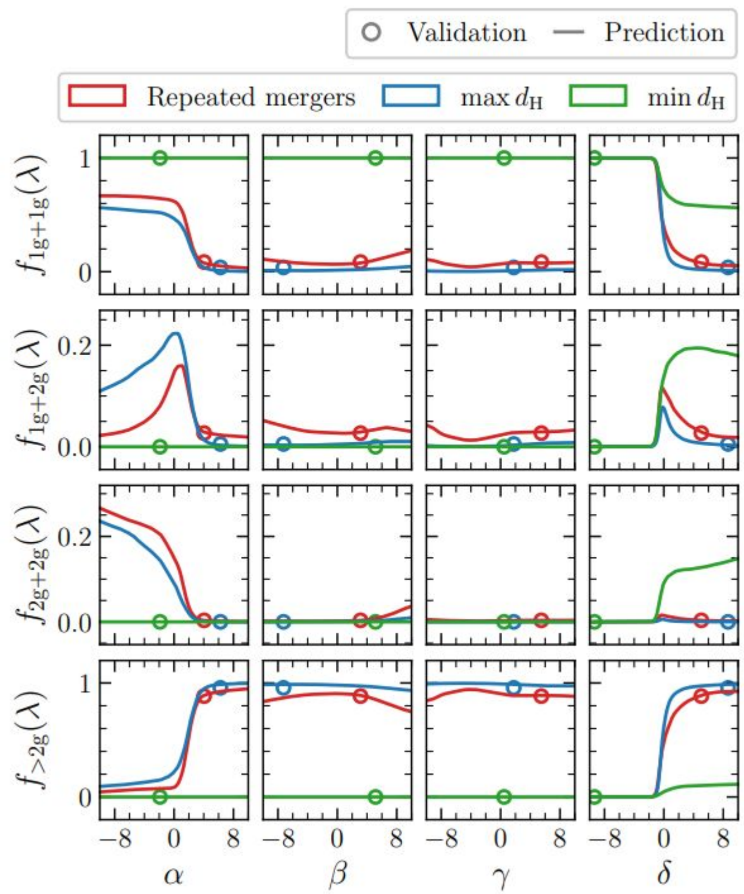
Deep-learned branching fractions

$$d_{\mathrm{H}}(p,q)^2 = 1 - \sum_i \sqrt{p_i q_i}$$

$$f_g(\boldsymbol{\lambda}) = N_g(\boldsymbol{\lambda})/N(\boldsymbol{\lambda})$$







Validation with mock catalogs