Deep-learning emulators and hierarchical Bayesian inference: application to gravitational-wave astronomy

Davide Gerosa

University of Milano-Bicocca

arXiv:2203.03651

with M. Mould, S. Taylor

davide.gerosa@unimib.it

www.davidegerosa.com

Mar 13, 2023

ML_INFN weekly meeting

(online)

90 waves and counting

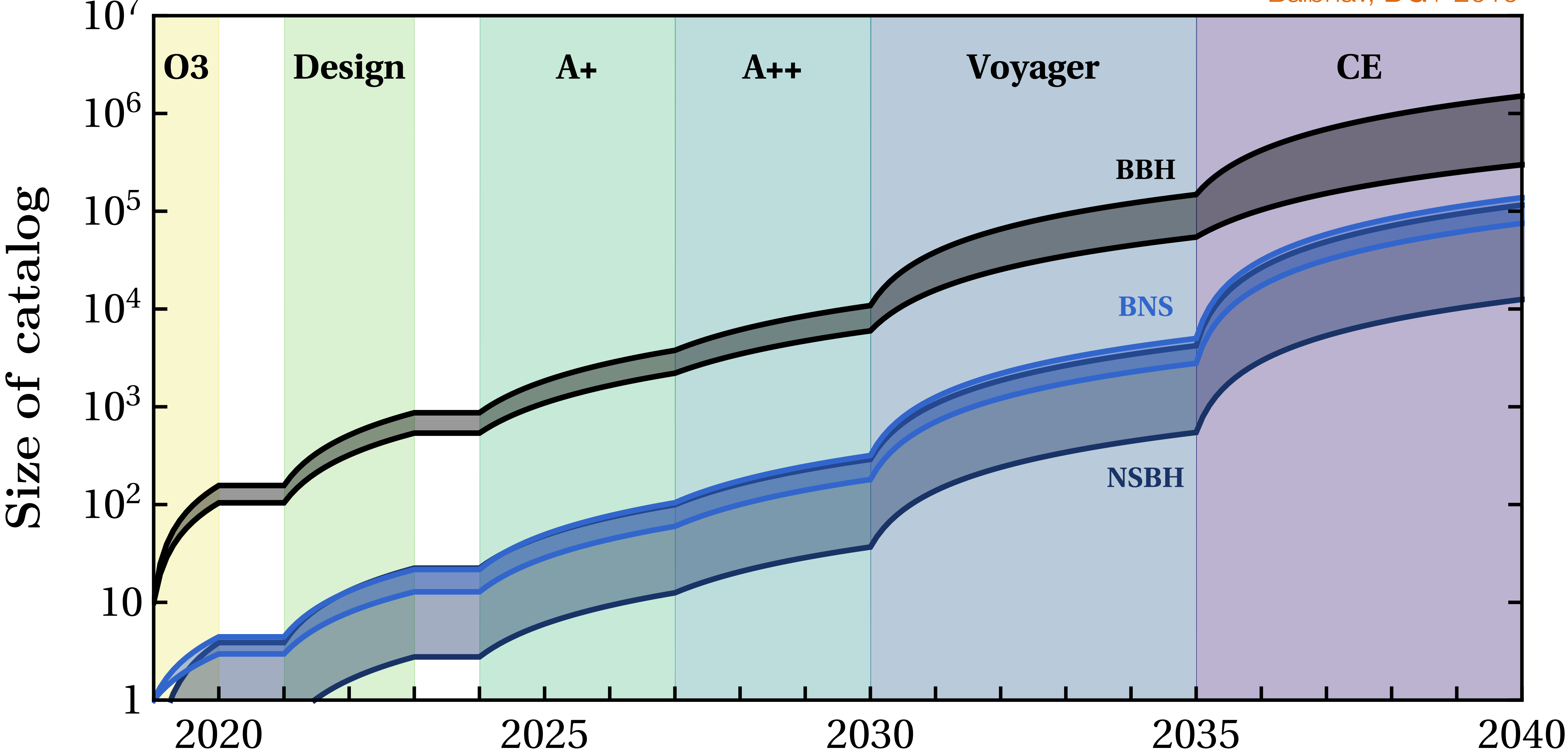
Discovering are piling up!

About 90 black-hole binary mergers detected so far.

LIGO 2021

Will become millions in ~20 years!

Baibhav, **DG**+ 2019



Can black holes really make it?

Power emitted in gravitational waves:

$$\frac{da}{dt} = -\frac{64}{5}\frac{G^3M^3}{c^5a^3}\frac{q}{(1+q)^2}$$

Peters 1964

GW-driven inspiral timescale

$$t_{\rm GW} \sim a\frac{dt}{da} \sim a^4$$

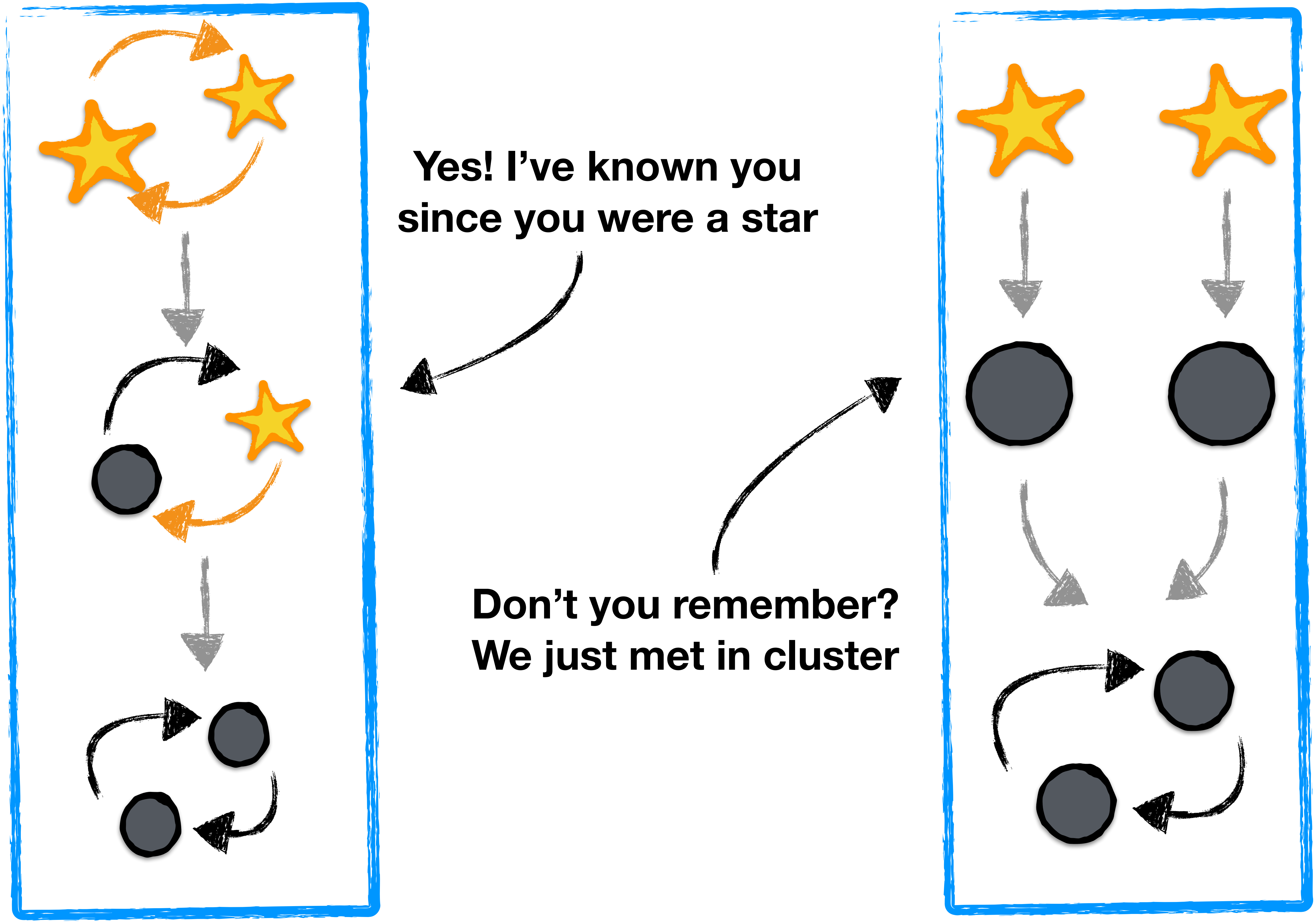
Gravitational waves are efficient below

$$a_{\rm GW} = 1.2\times10^{11}\left(\frac{t_{\rm GW}}{1.4\times10^{10}{\rm yr}}\right)^{1/4}\left(\frac{M}{M_\odot}\right)^{3/4}{\rm cm} \;\sim 10R_\odot$$

stellar-mass BHs

Relativity alone cannot explain the LIGO events,
we need some **astrophysics**

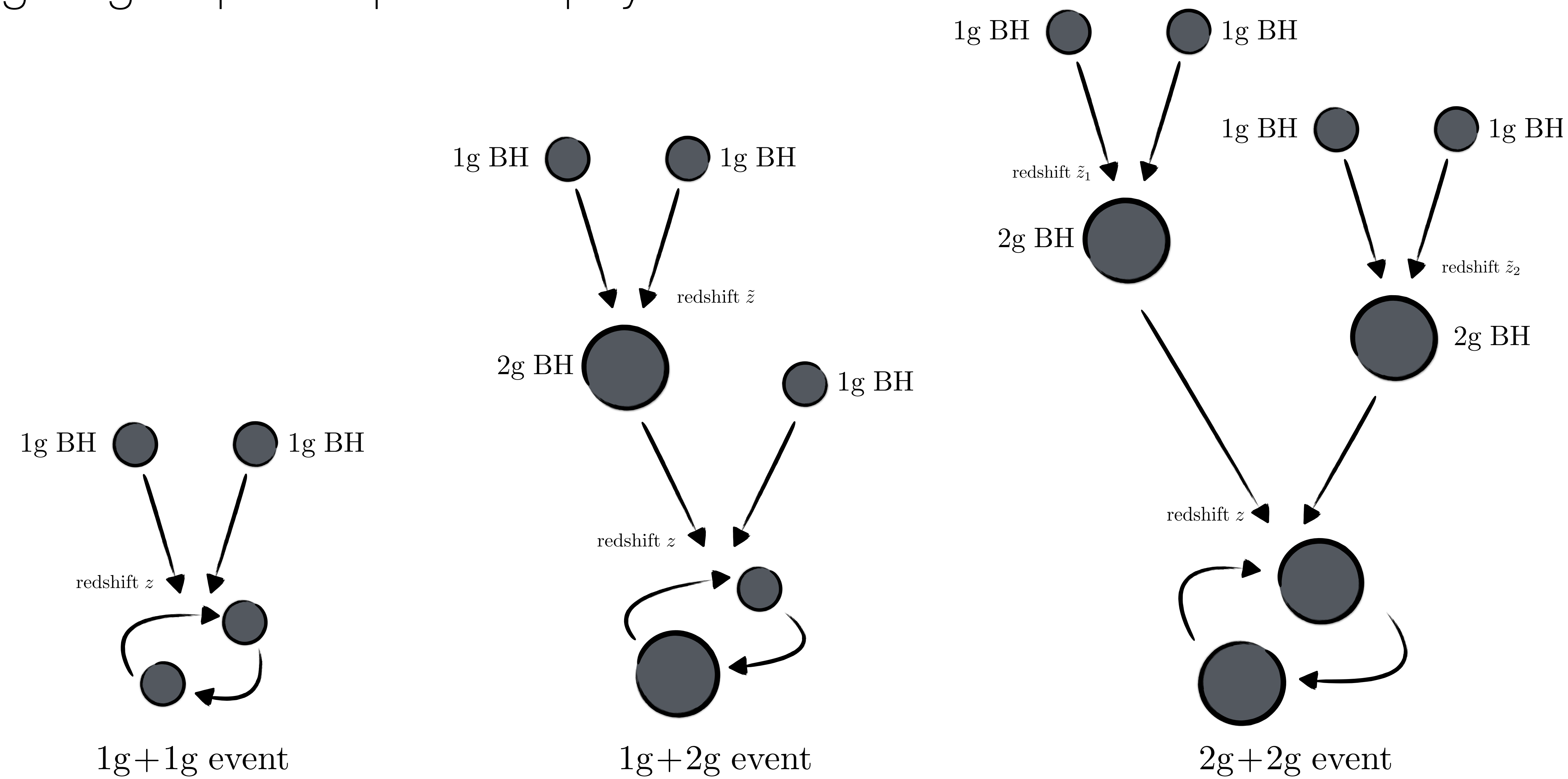
Have we been together for so long?



Hierarchical black-hole mergers

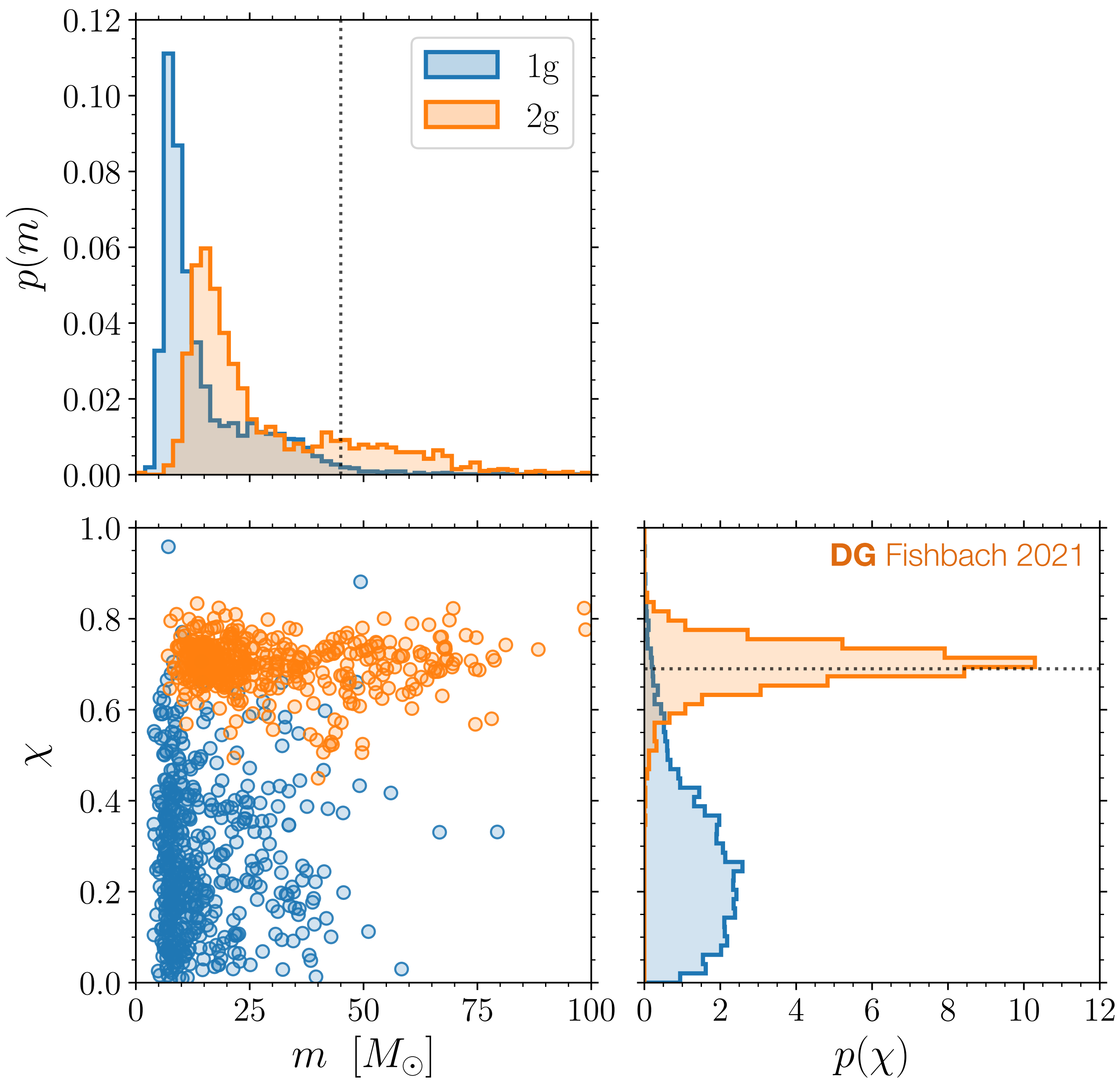
DG Berti 2017

Targeting a specific piece of physics here:



Orthogonal, but complementary to the usual field vs. cluster debate

An explosion of new predictions



- Masses in the pair-instability mass gap
  Heger+ 2003, Woosley+ 2007
- Peculiar spin distribution peaked at 0.7
  DG Berti 2017, Fishbach+ 2017
- But GW kicks require large escape speed
  DG Berti 2019
- Very frequent in AGNs
  Yang+ 2019, Tagawa+ 2020
- Promising for GW190412
  DG Vitale Berti 2020, Rogriguez+ 2020
- Leading explanation for GW190521
  LIGO/Virgo 2020
- Perhaps several events in the LIGO catalog?
  Kimball+ 2021
- An exclusion region
  DG Giacobbo Vecchio 2020
- … but don't overdo it!
  Zevin Holz 2022

And many more! Enough for a dedicated review **DG** Fishbach 2021

Populations, the Bayes way

θ	**Single-event parameters:** masses, spins, redshifts
λ	**Population parameters:** spectral index of mass distribution, cutoffs

Inhomogeneous Poisson process:

Loredo 2004, Mandel+ 2019,
Thrane, Talbot 2019, Vitale, **DG**+ 2022,

$$p(\lambda|d) \propto \pi(\lambda)\sigma^{-N}(\lambda)\prod_{i=1}^{N}\int p_{\rm pop}(\theta|\lambda)\mathcal{L}(d_i|\theta)\,\mathrm{d}\theta$$

Population posterior

Population prior

N events…

Population model

Single-event likelihood

Selection effects: $\sigma(\lambda) = \int p_{\rm pop}(\theta|\lambda)p_{\rm det}(\theta)\mathrm{d}\theta$

Detection probability

What model for the Universe?

Option 1: **Simple, parametrized functional forms**

LIGO/Virgo and many others

Evaluating $p_{\rm pop}(\theta|\lambda)$ is straightforward and can be done at each likelihood evaluation

<u>But</u>: *Astrophysicists put a lot of effort in simulating stellar evolution, clusters, AGNs, and all of that!*

Option 2:

Can we instead interpret GW data using cool astro predictions *directly*?

Evaluating $p_{\rm pop}(\theta|\lambda)$ now is a costly simulation…

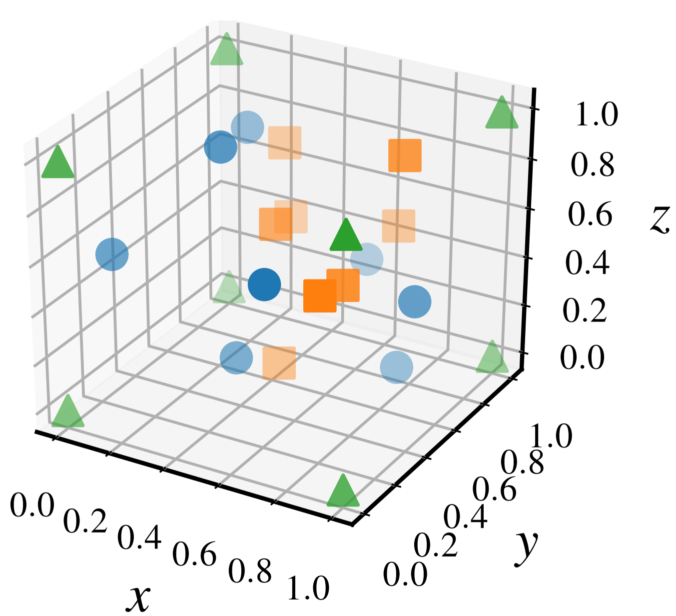
Ingredients in the blender

1. A population synthesis code

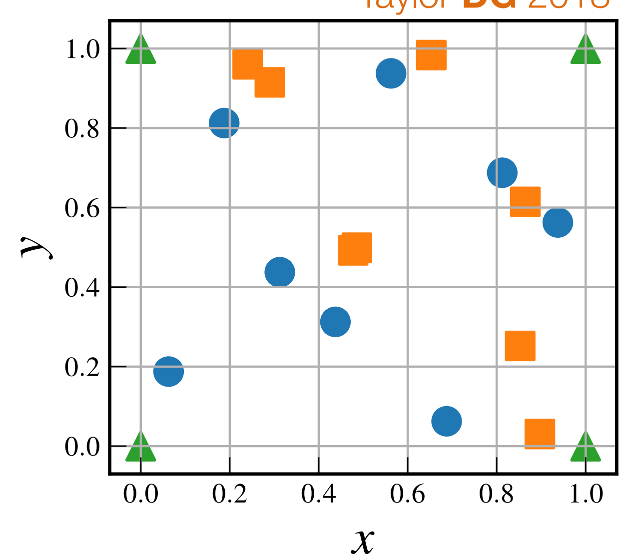
I'm not going to even try citing people here! So many excellent studies

- Early prototype with limited set of COMPAS runs Taylor **DG** 2018
- Current application: simple hierarchical merger populations Mould **DG** Taylor 2022
- Hopefully soon: full isolated formation channel inference
- Need help to do dynamics

2. Design a training bank. Space filling algorithms



Taylor **DG** 2018



- Latin hypercubes
- Now working on implementing progressive hypercube sampling

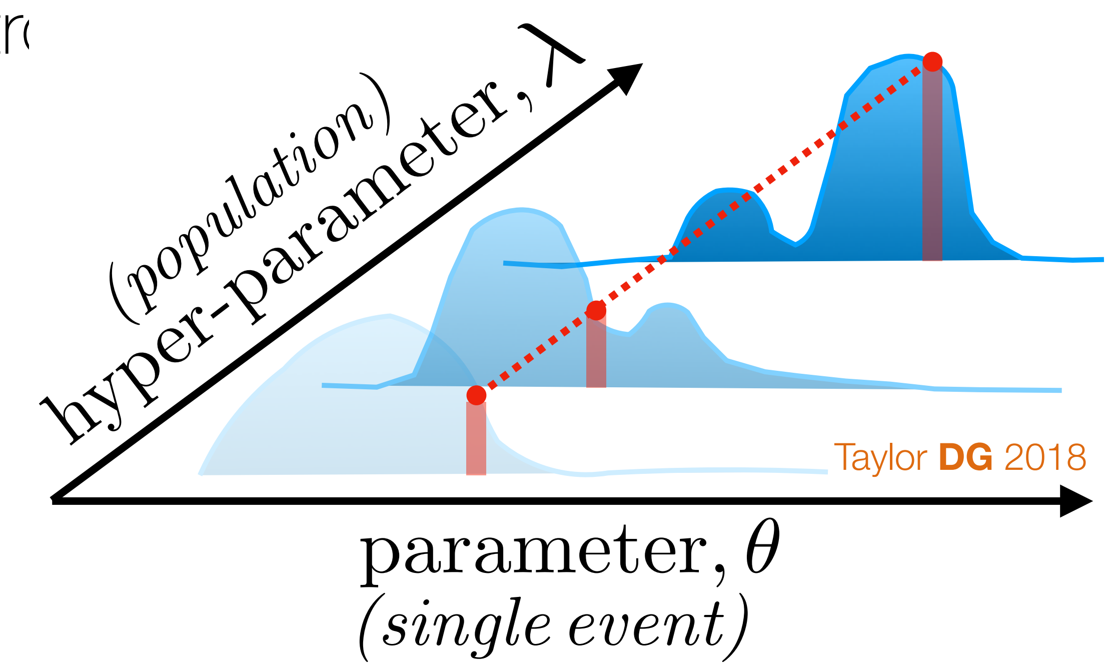
Ingredients in the blender

3. Some form of data compression

- Used principal component analysis successfully Taylor, **DG** 2018
- Tucker decomposition to avoid array raveling?
- Non-linear dimensionality reduction schemes?

4. A powerful conditional density estimation scheme $p_{\mathrm{pop}}(\theta|\lambda)$

- Gaussian process regression
  Taylor, **DG** 2018, Wong, **DG** 2019
- FFT-based KDE and a multilayer perceptr
  Mould **DG** Taylor 2022
- Autoregressive flows
  Wong, Contardo, Ho 2020



Ingredients in the blender

5. A model for the detector $p_{\rm det}(\theta)$

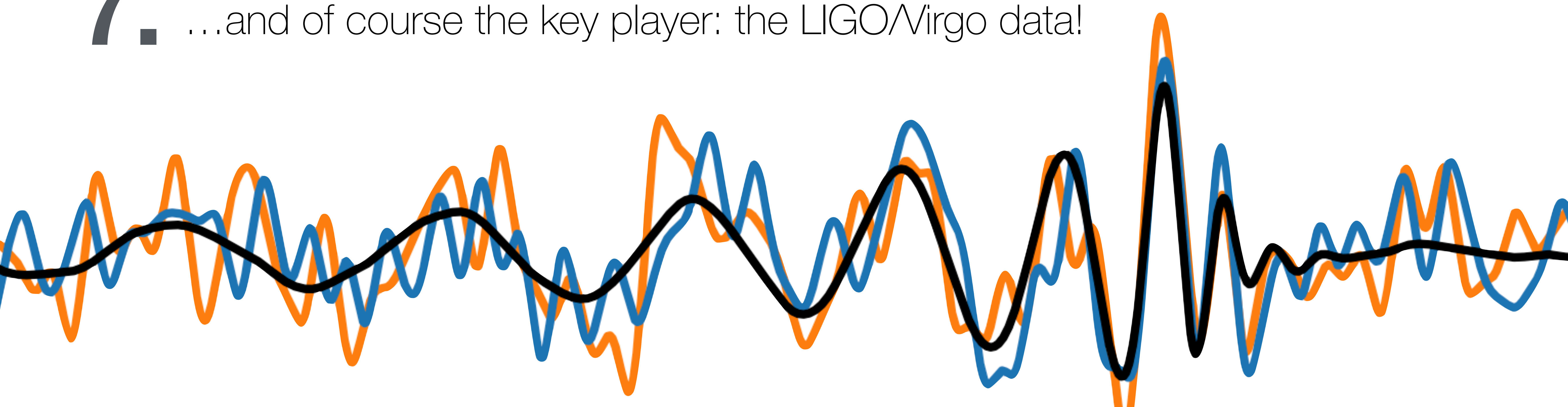
- A simple SNR cut? Finn Chernoff 1992
- Pipeline injections? LIGO/Virgo 2019, 2021
- Some attempts at machine-learn the GW detectability.

DG Pratten Vecchio 2020, Talbot Thrane 2022

6. A sampler for $p(\lambda|d)$

- A vanilla nested sampling for now… but should we?

7. …and of course the key player: the LIGO/Virgo data!



Just balls of black holes for now….

We need a population that is easy enough for now but non-analytic…

Key idea: **take a parametrized model but allow for hierarchical mergers**

In this talk a cluster is… a “thing” with a given escape speed $v_{\rm esc}$

DG, Berti 2019, **DG** Giacobbo Vecchio 2021, Zevin Holz 2022

- Masses: $p(m) \propto m^\gamma \qquad m \in [5M_\odot, m_{\rm max}]$
- Spins: $p(\chi) = \text{const} \qquad \chi \in [0, \chi_{\rm max}]$
- Pairing: $p_{\rm pair}(m_1) \propto m_1^\alpha$
  $p_{\rm pair}(m_2|m_1) \propto m_2^\beta$
- Clusters: $p(v_{\rm esc}) \propto v_{\rm esc}^\delta$

Just balls of black holes for now….

We need a population that is easy enough for now but non-analytic…

Key idea: **take a parametrized model but allow for hierarchical mergers**

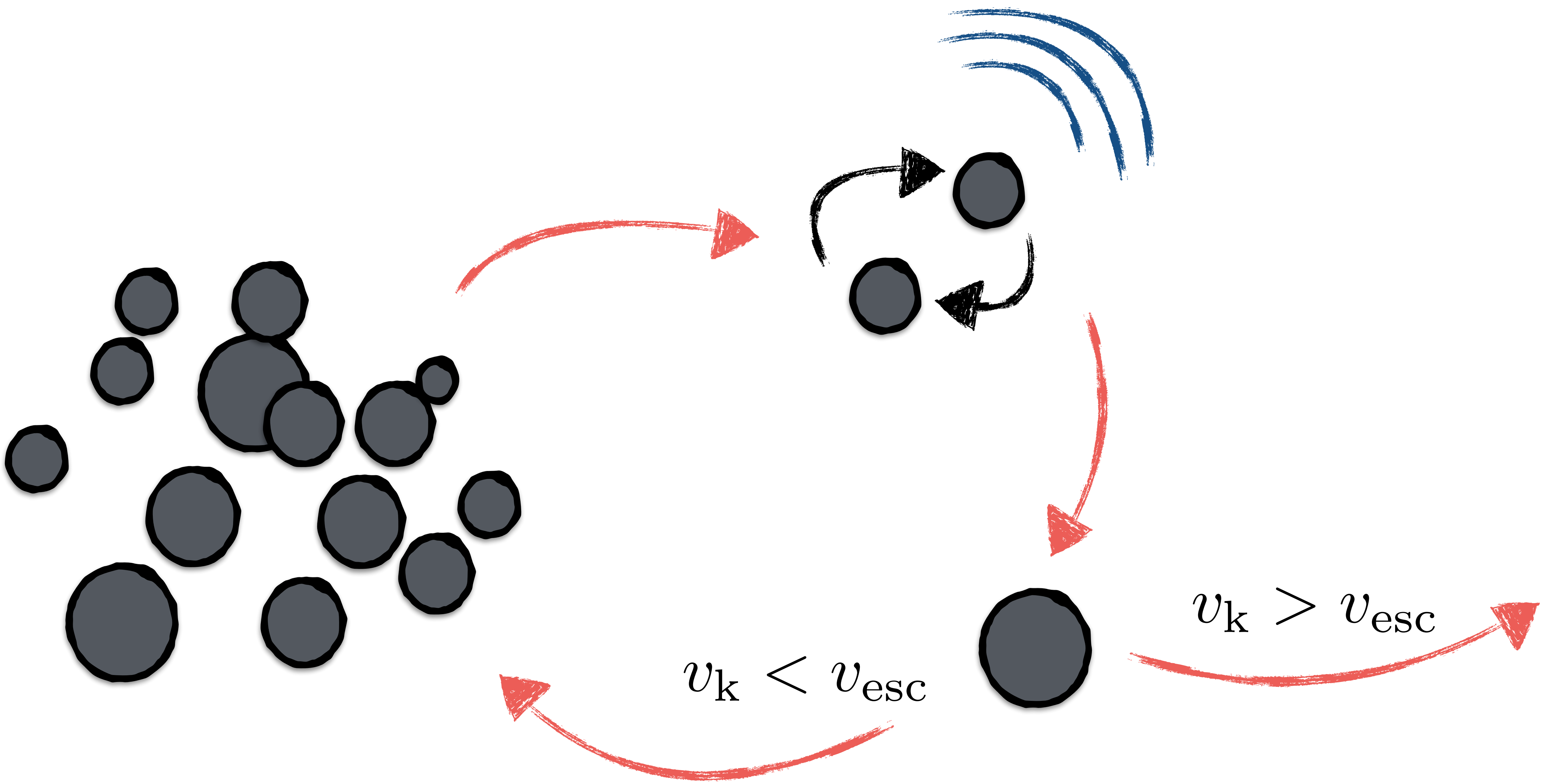
In this talk a cluster is… a “thing” with a given escape speed $v_{\rm esc}$

DG, Berti 2019, **DG** Giacobbo Vecchio 2021, Zevin Holz 2022

- Masses: $p(m) \propto m^{\boxed{\gamma}}$ $\quad m \in [5M_\odot, \boxed{m_{\rm max}}]$
- Spins: $p(\chi) = \rm const$ $\quad \chi \in [0, \boxed{\chi_{\rm max}}]$
- Pairing: $p_{\rm pair}(m_1) \propto m_1^{\boxed{\alpha}}$
  $p_{\rm pair}(m_2|m_1) \propto m_2^{\boxed{\beta}}$
- Clusters: $p(v_{\rm esc}) \propto v_{\rm esc}^{\boxed{\delta}}$

Six population parameters

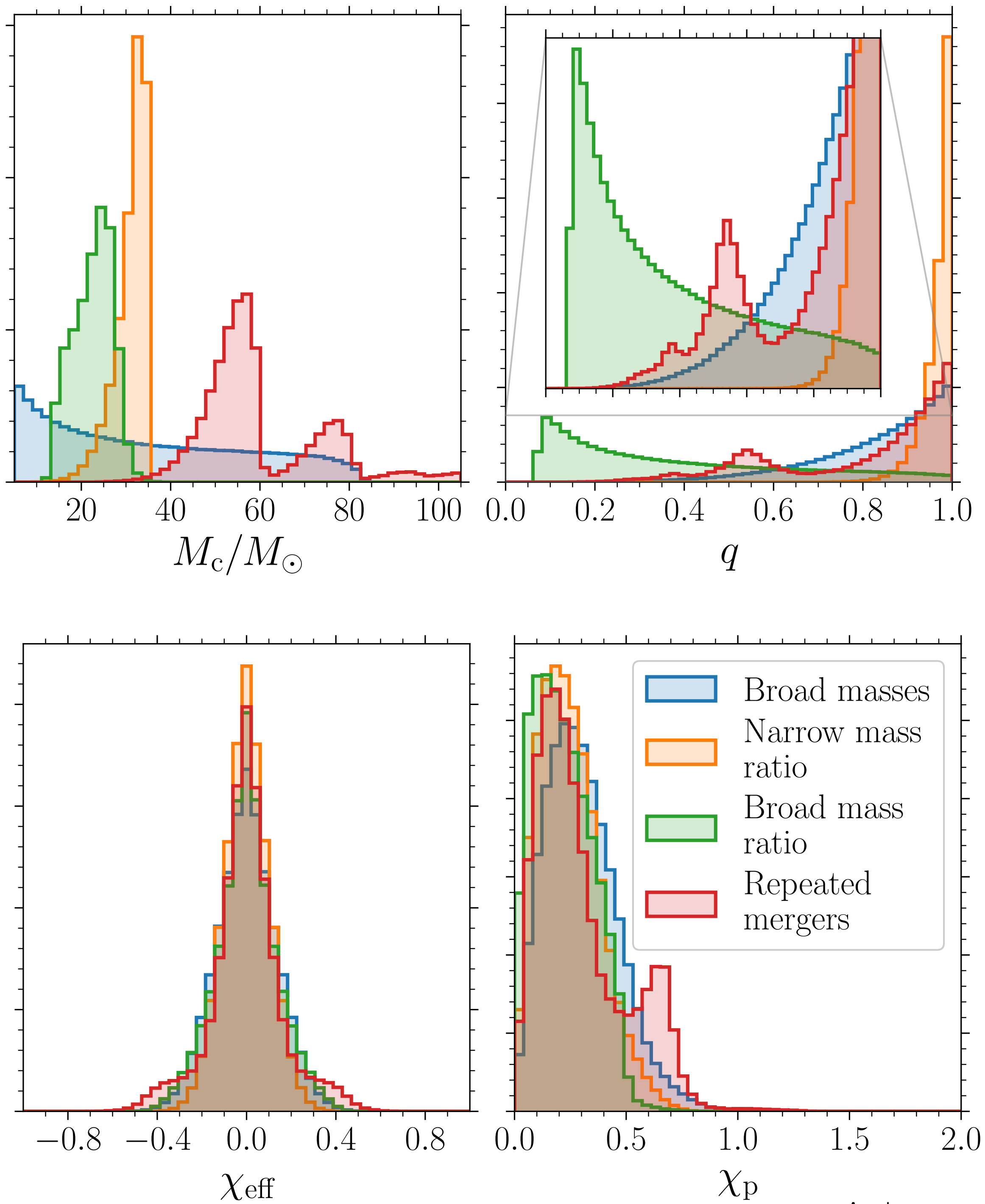
Out of the cluster, one kick at the time



Targeted populations

Tackling inference on four event parameters $\theta = \{M_c, q, \chi_{\mathrm{eff}}, \chi_{\mathrm{p}}\}$

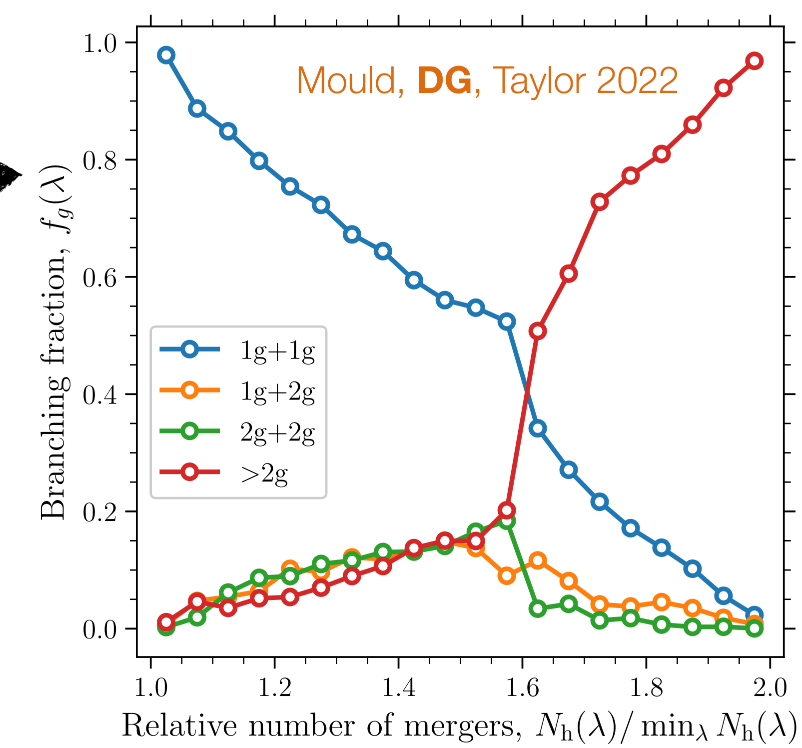
Mould, **DG**, Taylor 2022



Four representative hyperparameter locations

This is a hard problem!
Strong correlations, multimodalities, spikes, gaps, degeneracies

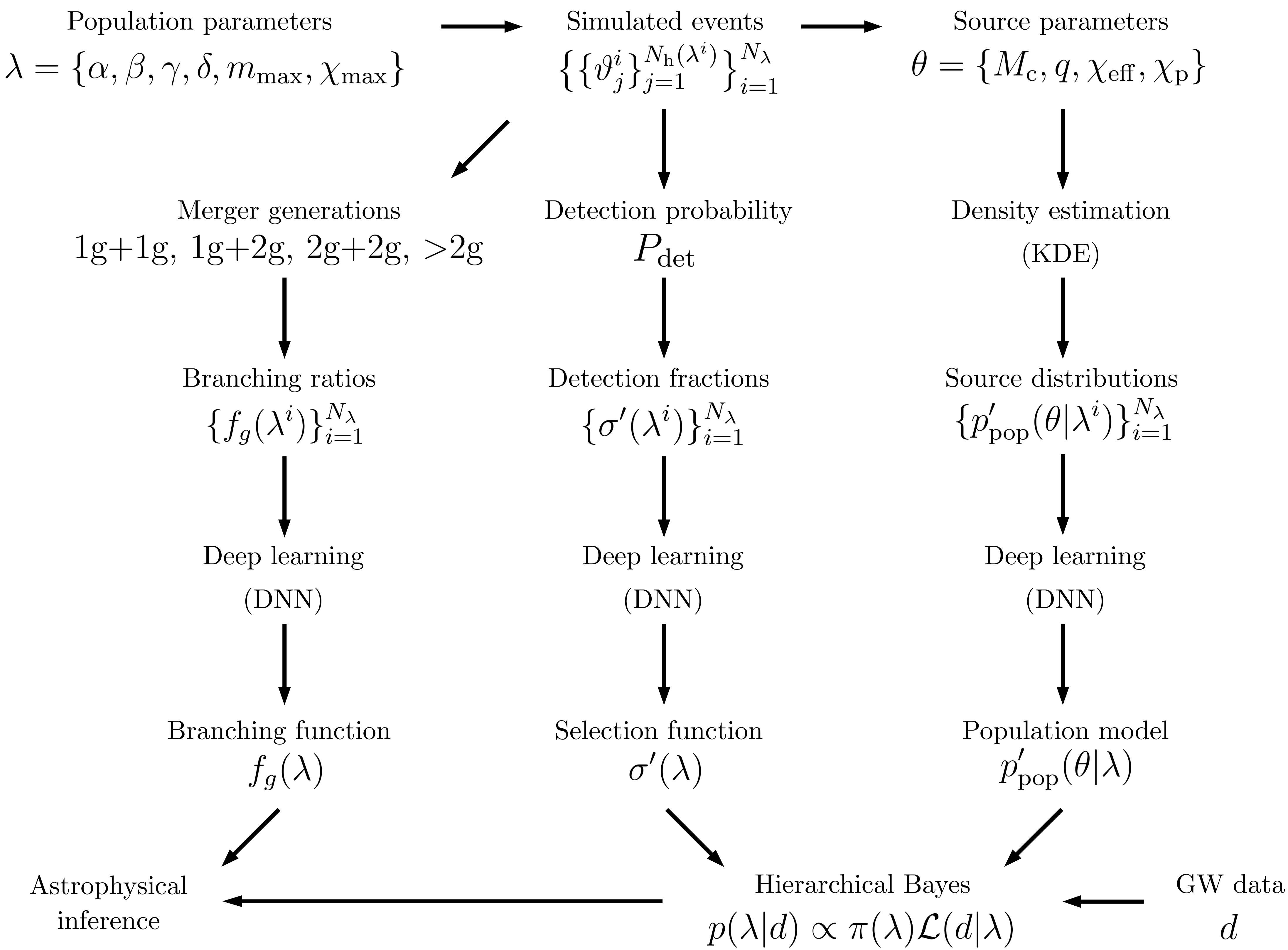
… and we keep track of the merger generation



Ask me at coffee why χ_{p} goes up to 2… **DG+** 2021

Mould, **DG**, Taylor 202

Full pipeline



Mould, **DG**, Taylor 2022

Full pipeline

Simulations

Population parameters
$\lambda = \{\alpha, \beta, \gamma, \delta, m_{\max}, \chi_{\max}\}$

Simulated events
$\{\{\vartheta_j^i\}_{j=1}^{N_\mathrm{h}(\lambda^i)}\}_{i=1}^{N_\lambda}$

Source parameters
$\theta = \{M_\mathrm{c}, q, \chi_\mathrm{eff}, \chi_\mathrm{p}\}$

Merger generations
1g+1g, 1g+2g, 2g+2g, >2g

Detection probability
P_det

Density estimation
(KDE)

Branching ratios
$\{f_g(\lambda^i)\}_{i=1}^{N_\lambda}$

Detection fractions
$\{\sigma'(\lambda^i)\}_{i=1}^{N_\lambda}$

Source distributions
$\{p'_\mathrm{pop}(\theta|\lambda^i)\}_{i=1}^{N_\lambda}$

Deep learning
(DNN)

Deep learning
(DNN)

Deep learning
(DNN)

Branching function
$f_g(\lambda)$

Selection function
$\sigma'(\lambda)$

Population model
$p'_\mathrm{pop}(\theta|\lambda)$

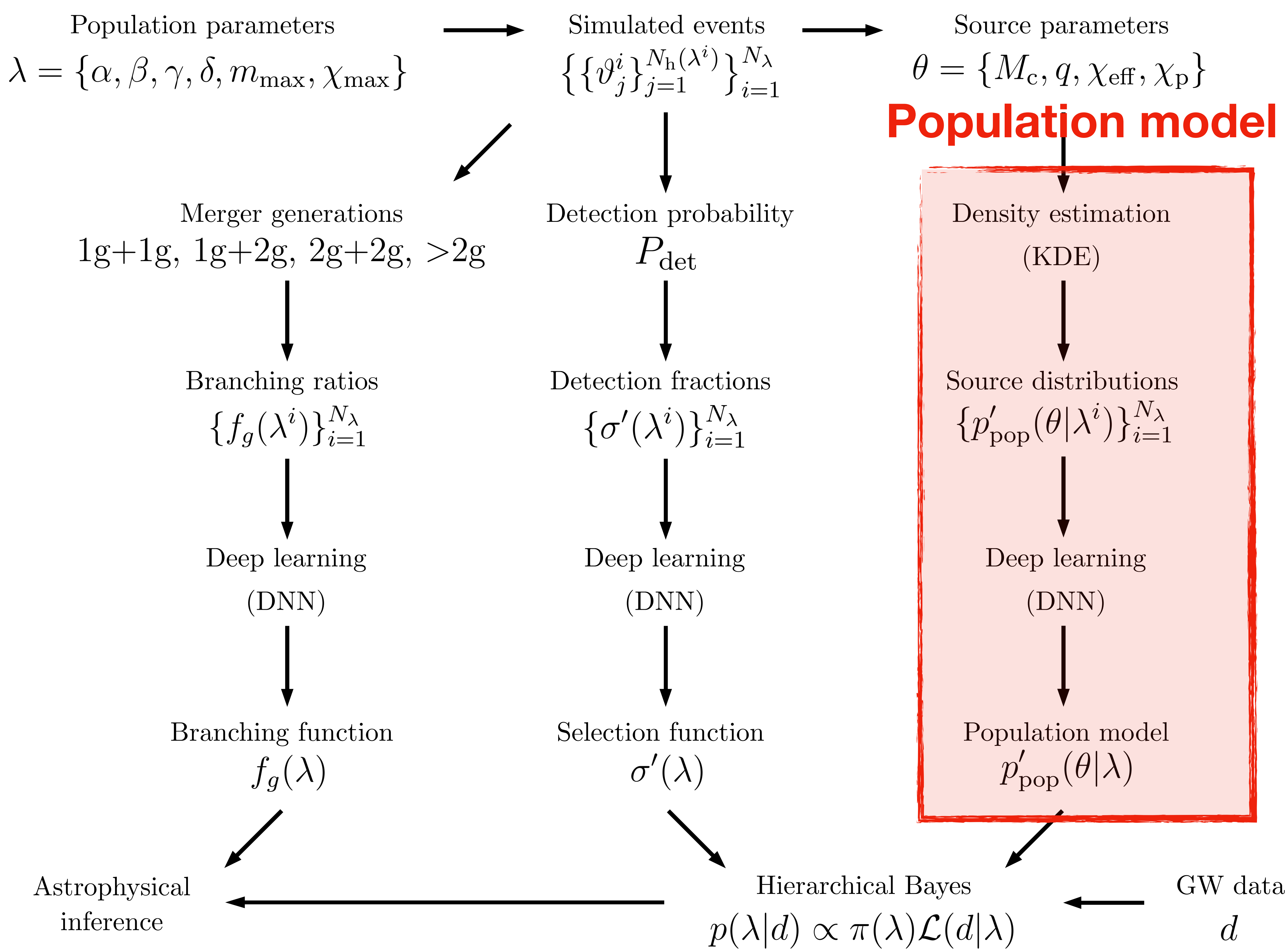
Astrophysical
inference

Hierarchical Bayes
$p(\lambda|d) \propto \pi(\lambda)\mathcal{L}(d|\lambda)$

GW data
d

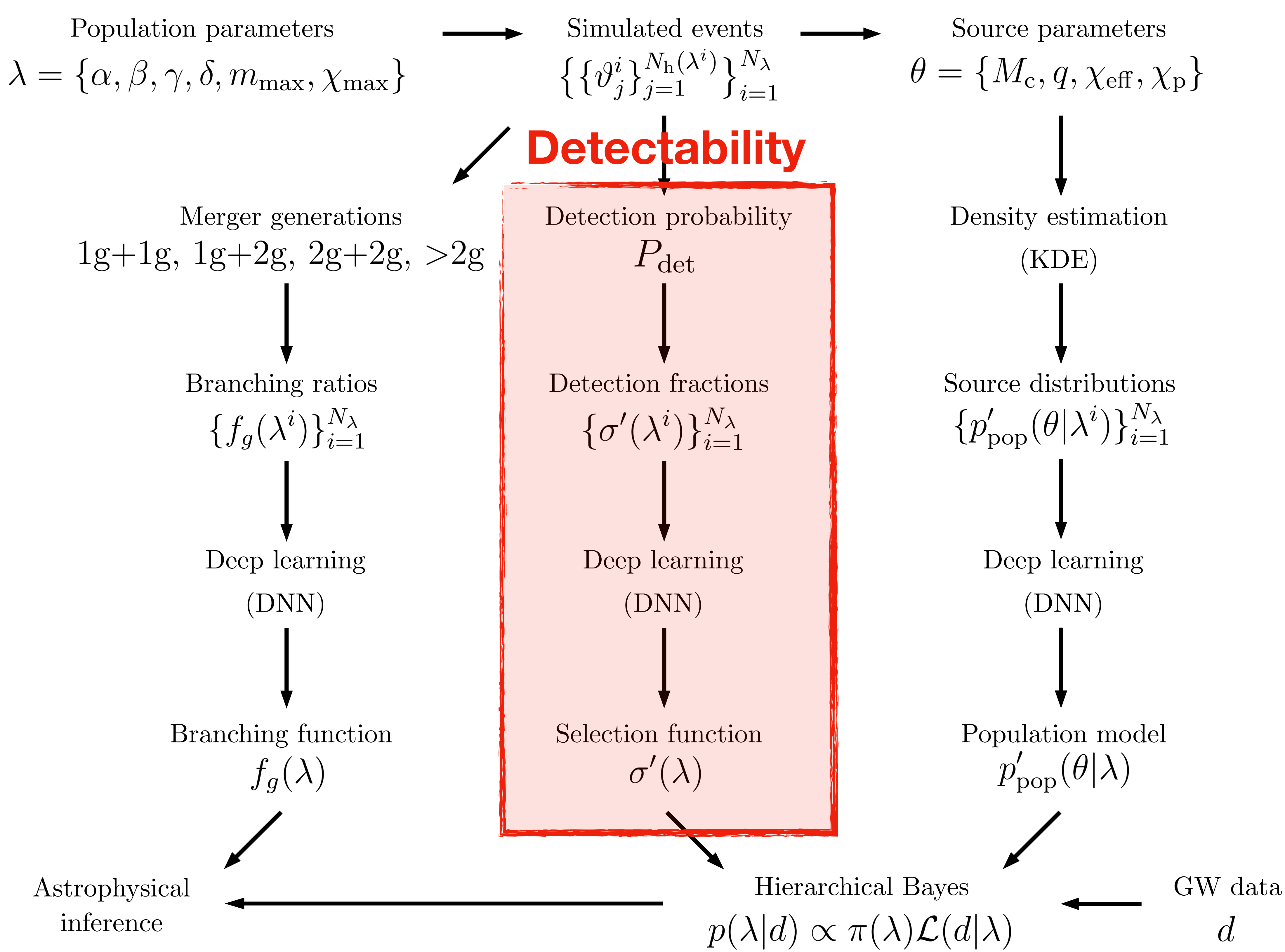
Full pipeline

Mould, **DG**, Taylor 2022



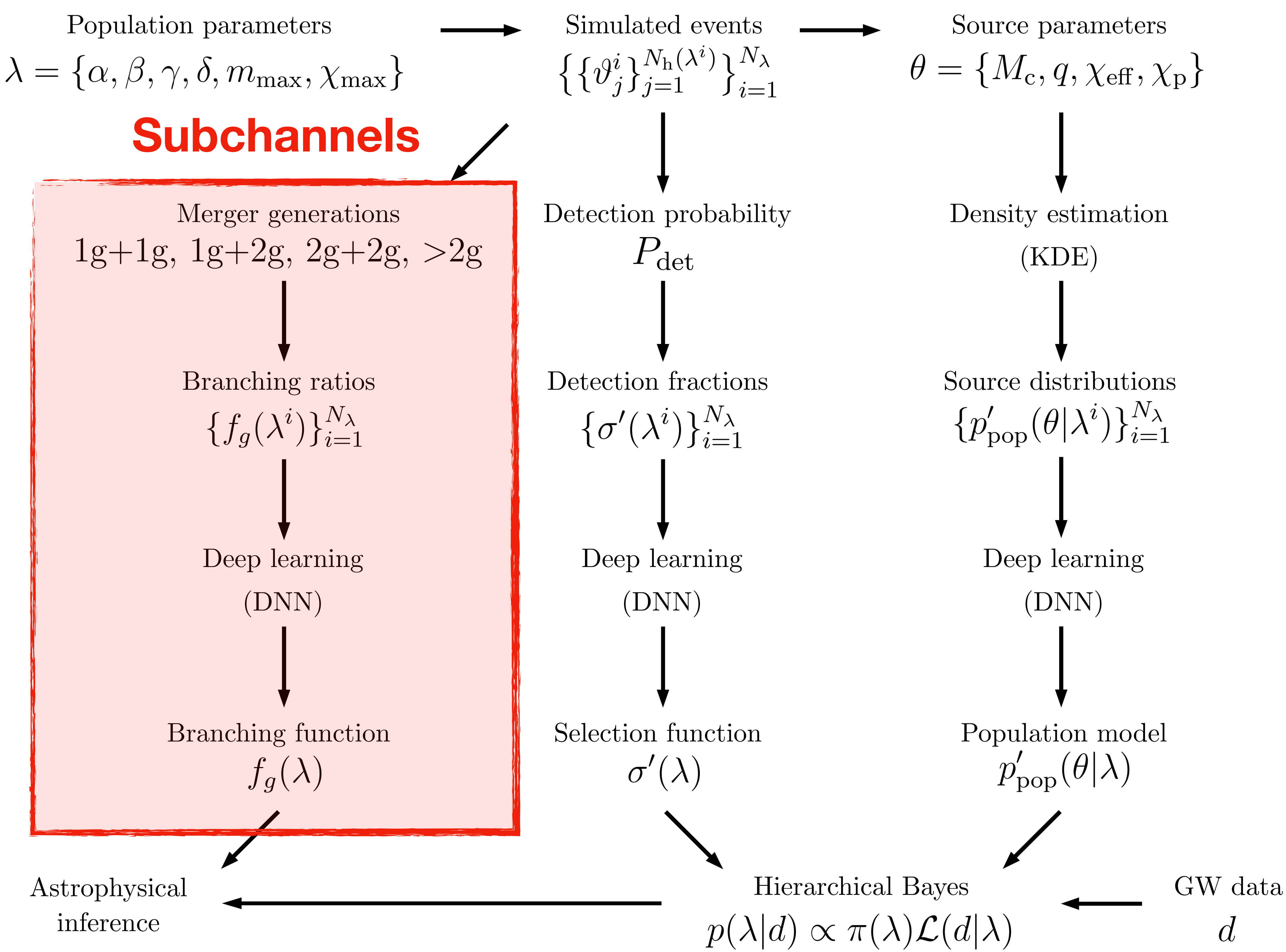
Full pipeline

Mould, **DG**, Taylor 2022



Full pipeline

Mould, **DG**, Taylor 2022



Full pipeline

Mould, **DG**, Taylor 2022

Population parameters
$\lambda = \{\alpha, \beta, \gamma, \delta, m_{\max}, \chi_{\max}\}$

Simulated events
$\{\{\vartheta^i_j\}_{j=1}^{N_{\rm h}(\lambda^i)}\}_{i=1}^{N_\lambda}$

Source parameters
$\theta = \{M_{\rm c}, q, \chi_{\rm eff}, \chi_{\rm p}\}$

Merger generations
1g+1g, 1g+2g, 2g+2g, >2g

Detection probability
$P_{\rm det}$

Density estimation
(KDE)

Branching ratios
$\{f_g(\lambda^i)\}_{i=1}^{N_\lambda}$

Detection fractions
$\{\sigma'(\lambda^i)\}_{i=1}^{N_\lambda}$

Source distributions
$\{p'_{\rm pop}(\theta|\lambda^i)\}_{i=1}^{N_\lambda}$

Deep learning
(DNN)

Deep learning
(DNN)

Deep learning
(DNN)

Branching function
$f_g(\lambda)$

Selection function
$\sigma'(\lambda)$

Population model
$p'_{\rm pop}(\theta|\lambda)$

Inference

Astrophysical inference

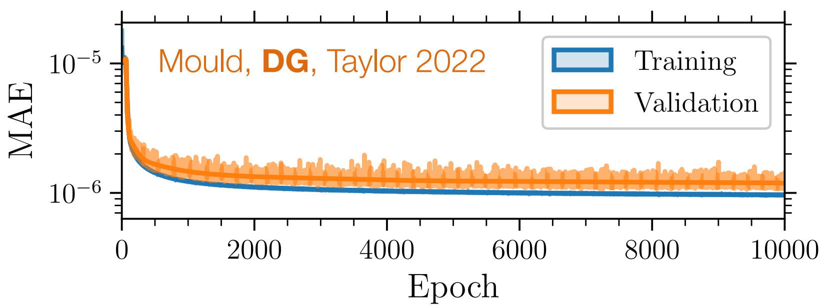
Hierarchical Bayes
$p(\lambda|d) \propto \pi(\lambda)\mathcal{L}(d|\lambda)$

GW data
d

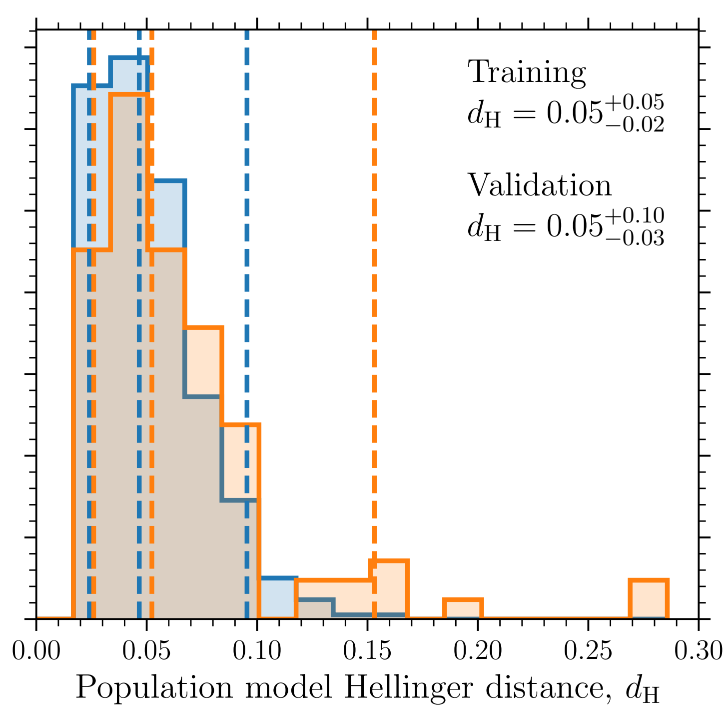
Population neural network

- A fully connected network
- A total of ~70k parameters!
- Implemented in Google's Tensorflow
- Fast (~days) training on GPU

Layer	Neurons	Activation	Parameters
Input	10	–	0
Dense 1	128	RReLU	1408
Dense 2	128	RReLU	16,512
Dense 3	128	RReLU	16,512
Dense 4	128	RReLU	16,512
Dense 5	128	RReLU	16,512
Output	1	Absolute value	129
		Total	67,585



No significant overfitting

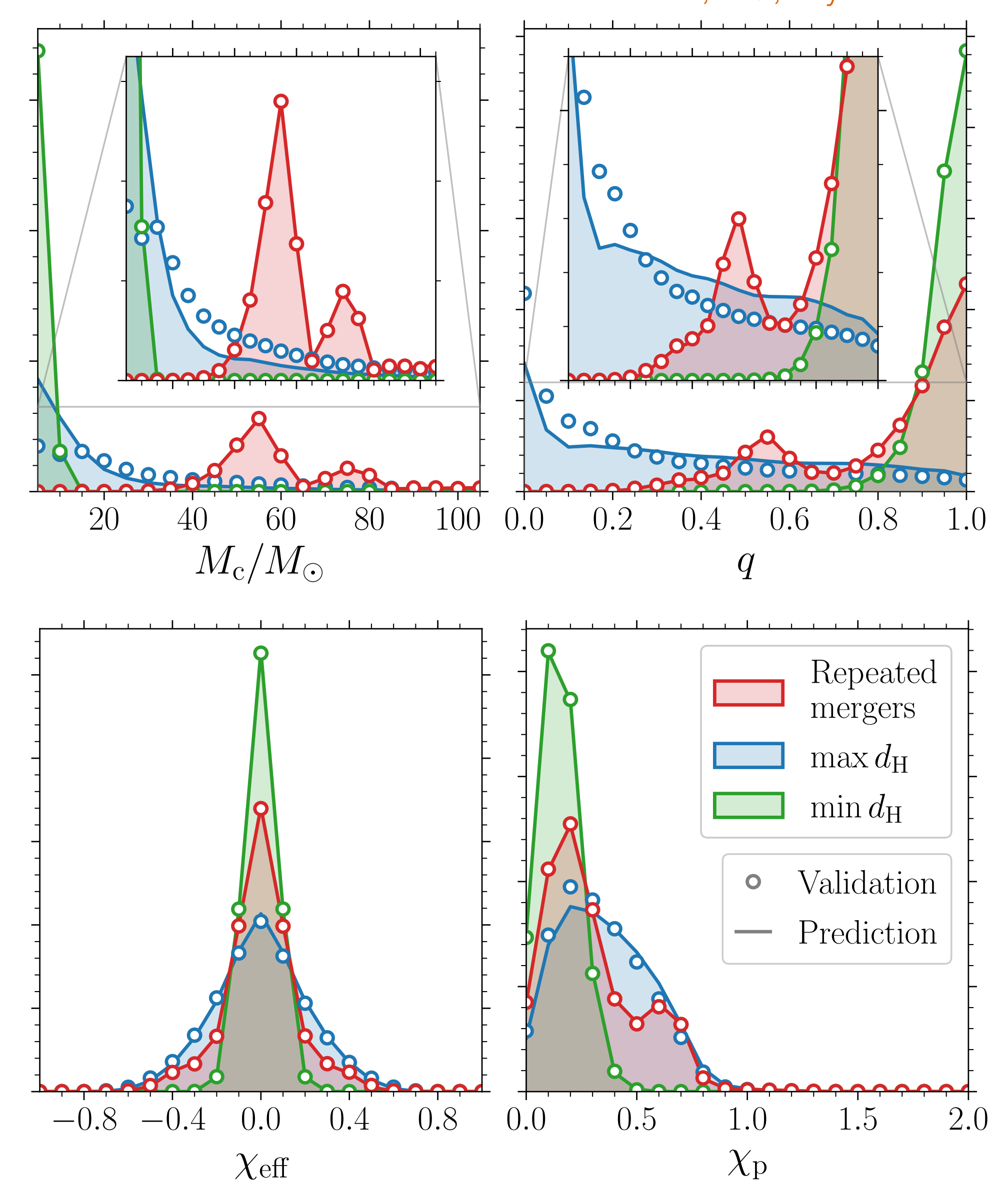


Interpolated distributions are statistically the same

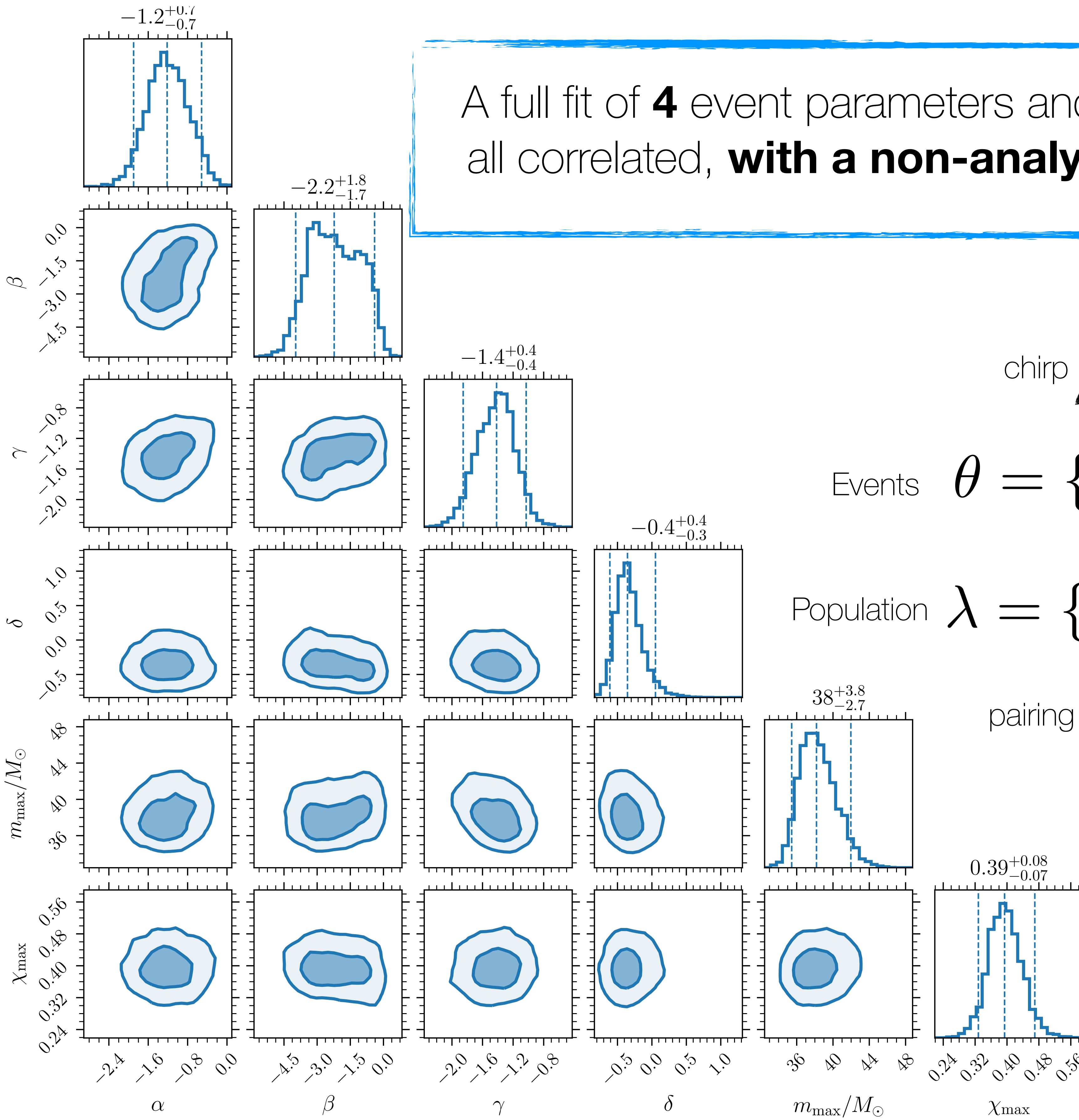
$$d_{\mathrm{H}}(p,q)^2 = 1 - \int \sqrt{p(x)q(x)}\mathrm{d}x\,.$$

Able to capture spikes and almost-discontinuous features

Mould, **DG**, Taylor 2022



Full GWTC3 results



A full fit of **4** event parameters and **6** population parameters, all correlated, **with a non-analytical population model**!

chirp mass, mass ratio, effective spin, precession

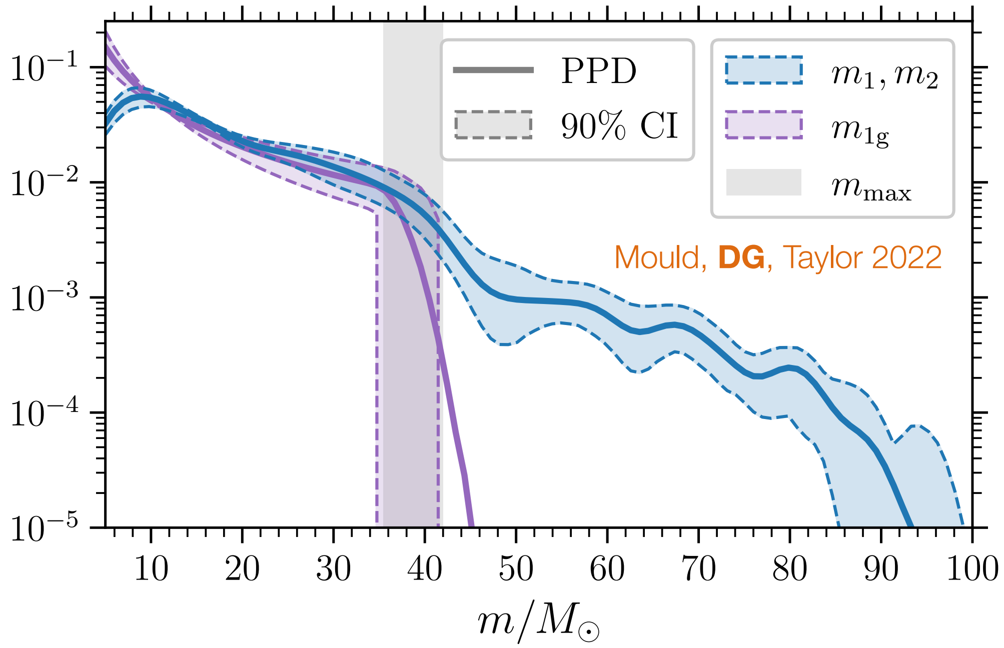
Events $\theta = \{M_c, q, \chi_{\rm eff}, \chi_{\rm p}\}$

Population $\lambda = \{\alpha, \beta, \gamma, \delta, M_{\max}, \chi_{\max}\}$

pairing, pairing, 1g mass slope, escape speeds, max 1g mass, max 1g spin

Mould, **DG**, Taylor 2022

Inference and predictions

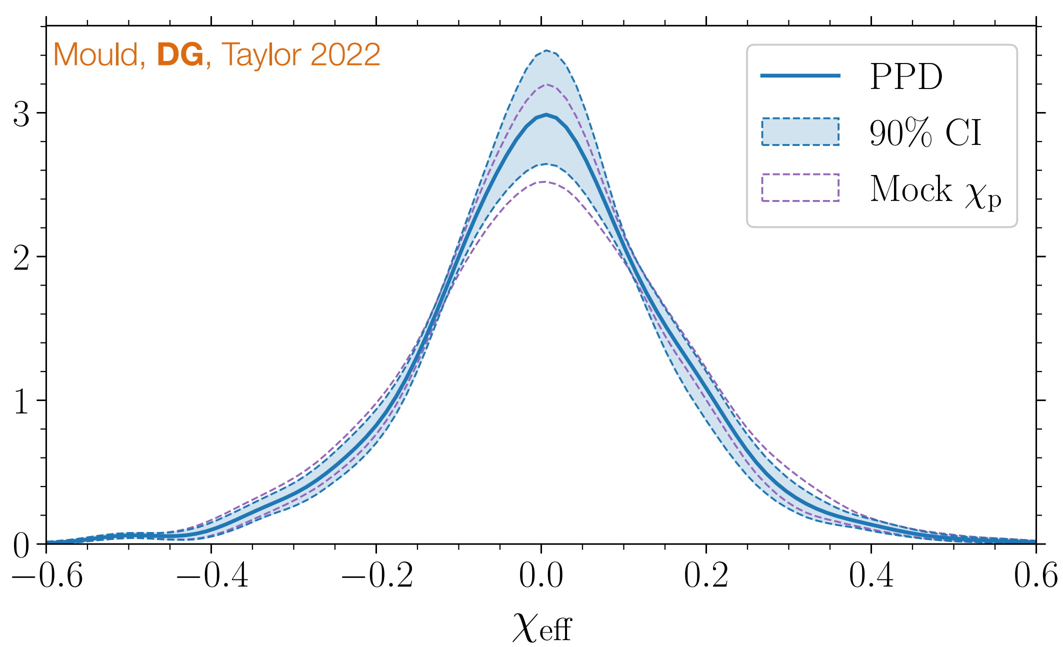


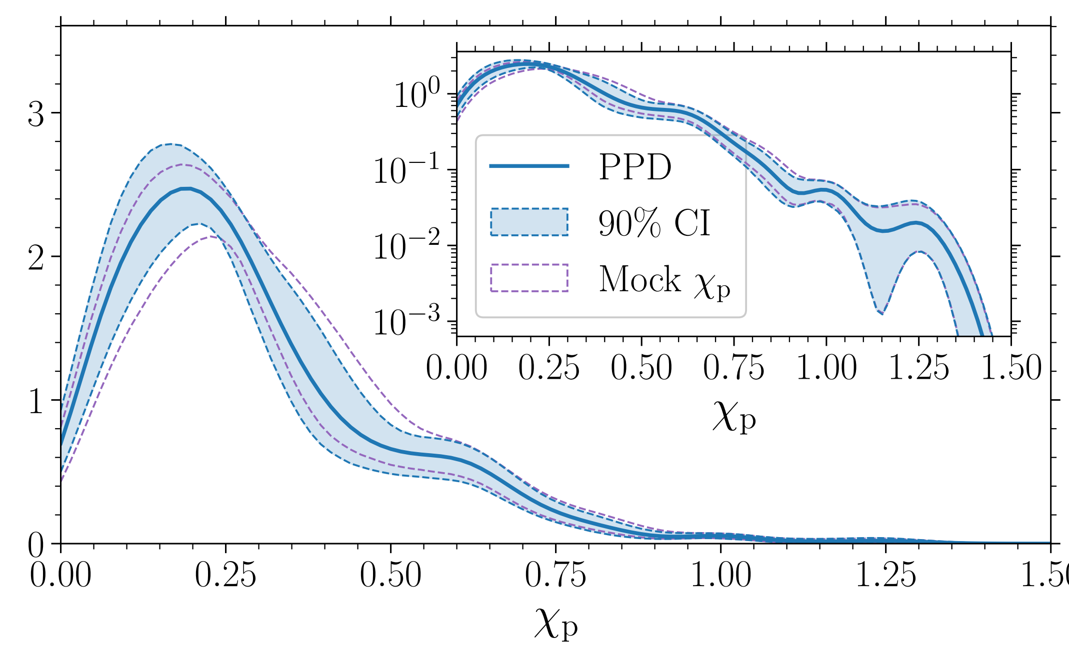
Masses

- Repeated mergers populate the upper mass gap
- 1g cutoff ok with pair instability SN?
- Additional structure in the gap due to higher generations

Spins

- Fat tails in the effective spin
- Fine structures in spin precession

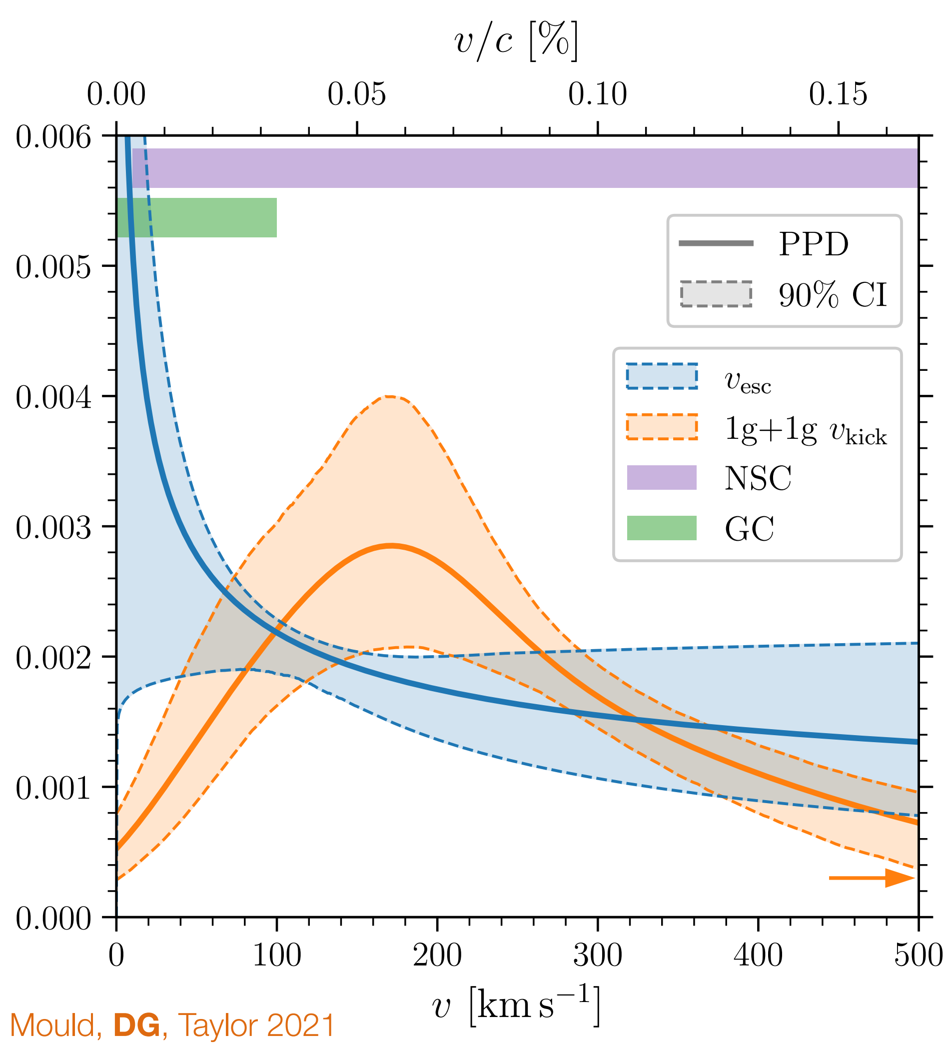




Inference and predictions

Escape speeds

- Easy to infer secondary population parameters (here the escape speed)
- But can go crazy! Metallicity, environments, etc

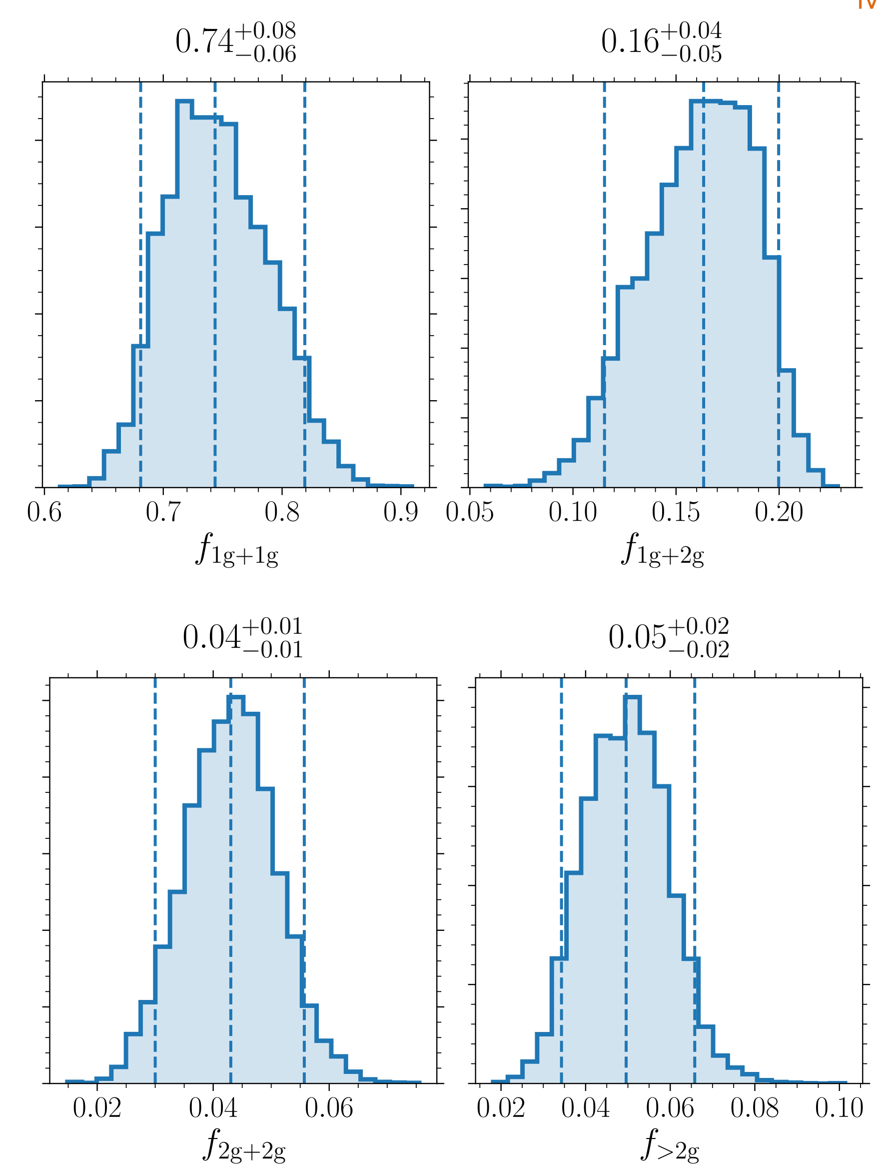


Mould, **DG**, Taylor 2021

Generations

- If we allow for hierarchical mergers, the fit wants to go there! cf e.g. Kimball+ 2021
- Easy to infer subchannels (here the generation)
- But can go crazy! Any label in the population…

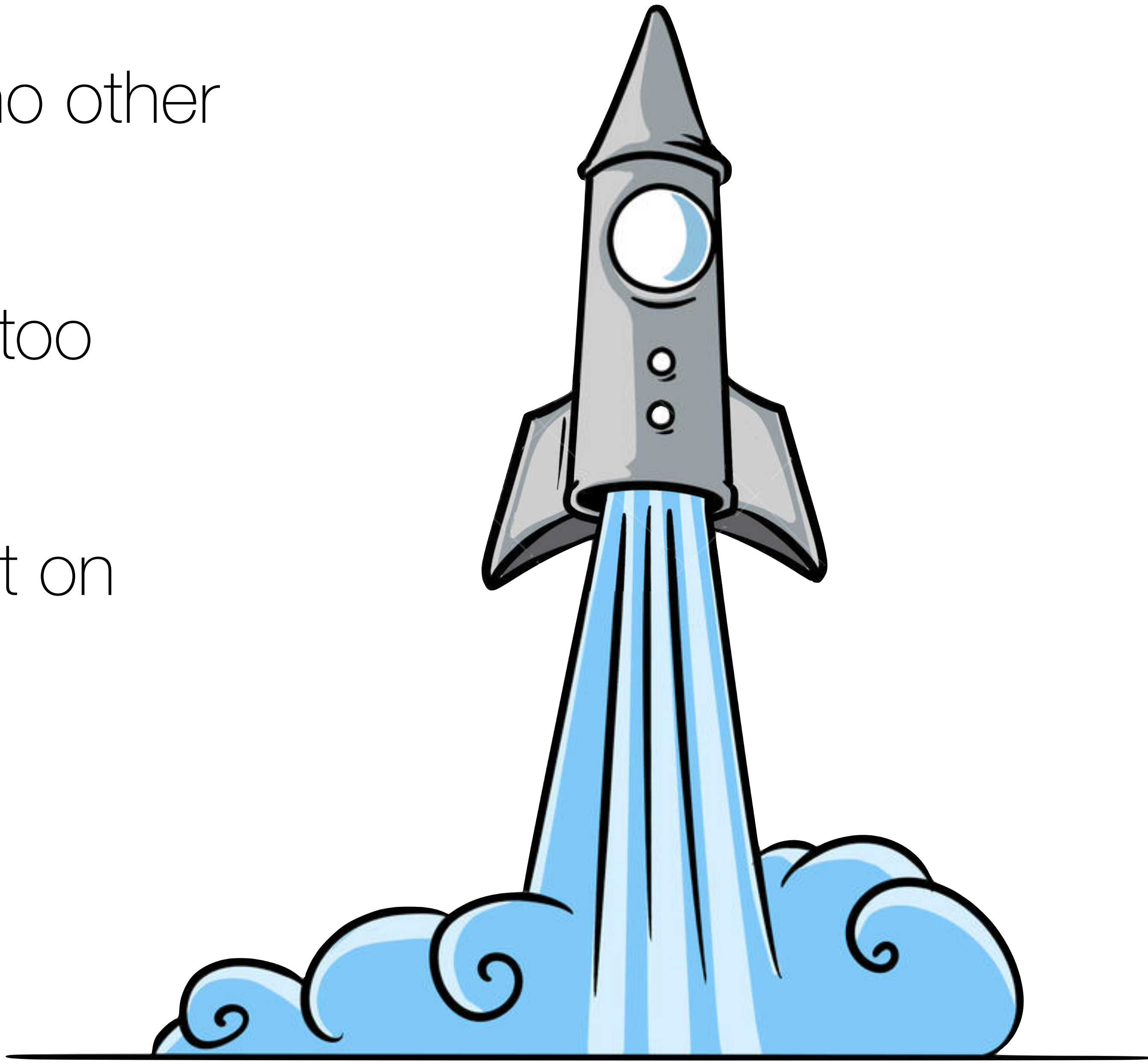
Mould, **DG**, Moore, in prep



Mould, **DG**, Taylor 2022

Ready for launch

- A complete, highly optimized population inference pipeline designed to digest outputs of astrophysical simulations and GW data
- **Deep learning is crucial here** (no other way, I think)
- Current astrophysics is admittedly too simple…
- …but we're ready to use this beast on state-of the art models!



Deep-learning emulators and hierarchical Bayesian inference: application to gravitational-wave astronomy

Davide Gerosa

University of Milano-Bicocca

arXiv:2203.03651

with M. Mould, S. Taylor

davide.gerosa@unimib.it
www.davidegerosa.com

Mar 13, 2023

ML_INFN weekly meeting

(online)

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

European Research Council

Fondazione CARIPLO

LEVERHULME TRUST

INFN

ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing