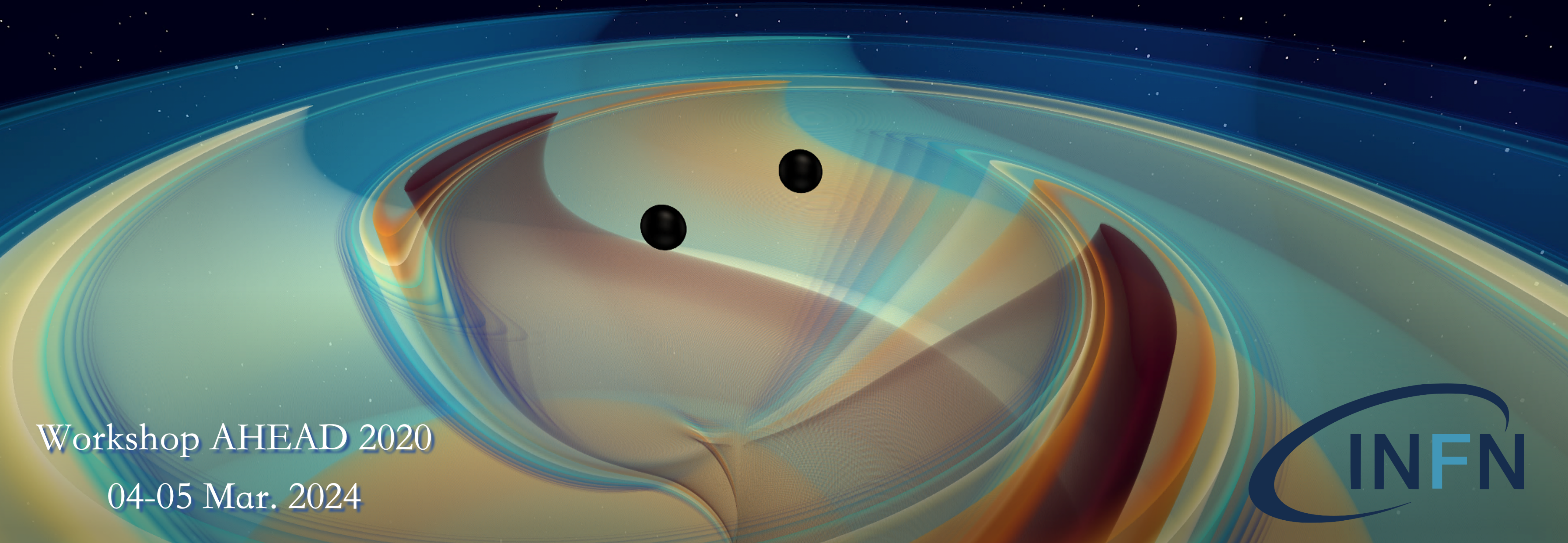
Credit: N. Fischer

Deep Learning to detect gravitational waves from binary Close Encounters:

Fast parameter estimation with Normalizing Flows

F. De Santi[1,2], M. Razzano[1,2], F. Fidecaro[1,2] , L. Papalini[1,2], L. Muccillo[1,2], B. Patricelli[1,2]

[1] Università di Pisa, [2] INFN Pisa



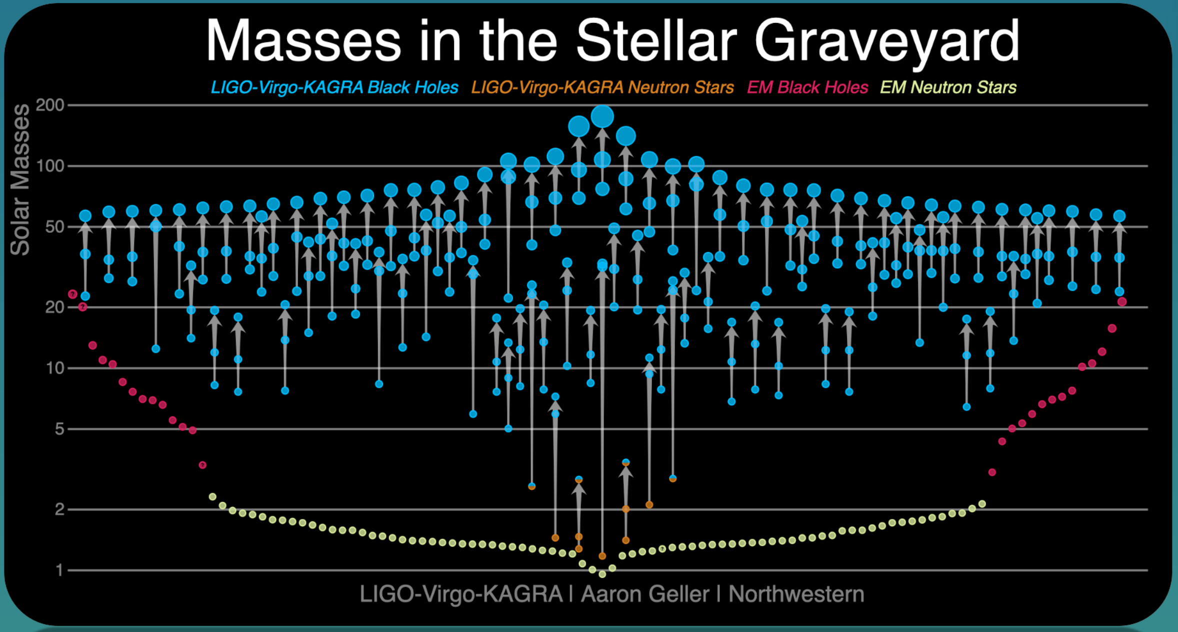
Workshop AHEAD 2020

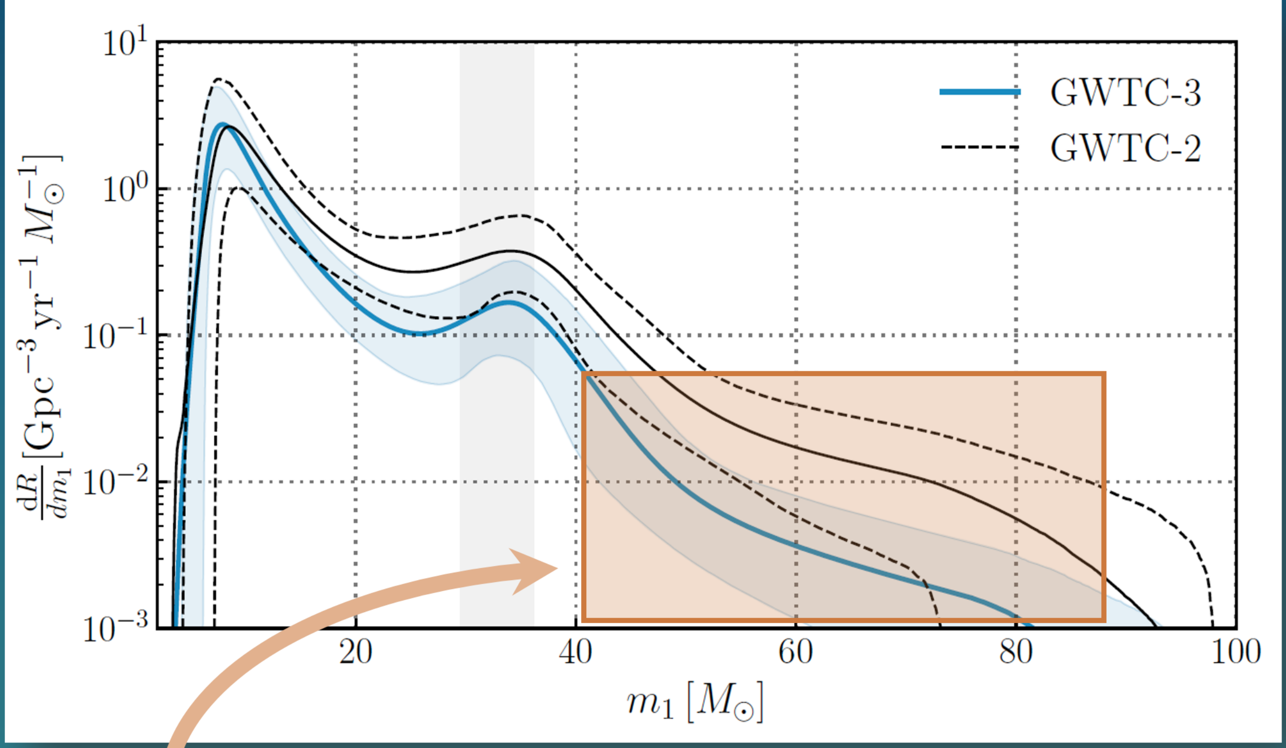
04-05 Mar. 2024

What has been observed so far ?

- Gravitational Wave Transient Catalogue 3 :
  - 90 transients (CBCs)
    - 2 Binary Neutron Stars
    - 3 Neutron Star + Black Hole

Abbot R. et al. (2021)





Abbot R. et al. (2023)

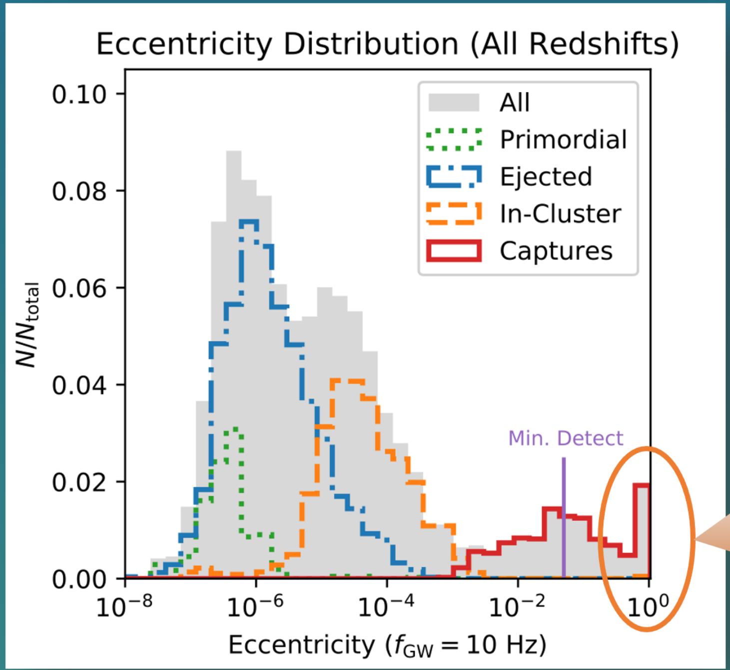
Hint to dynamical formation channels

Dynamical Interactions in dense stellar environments

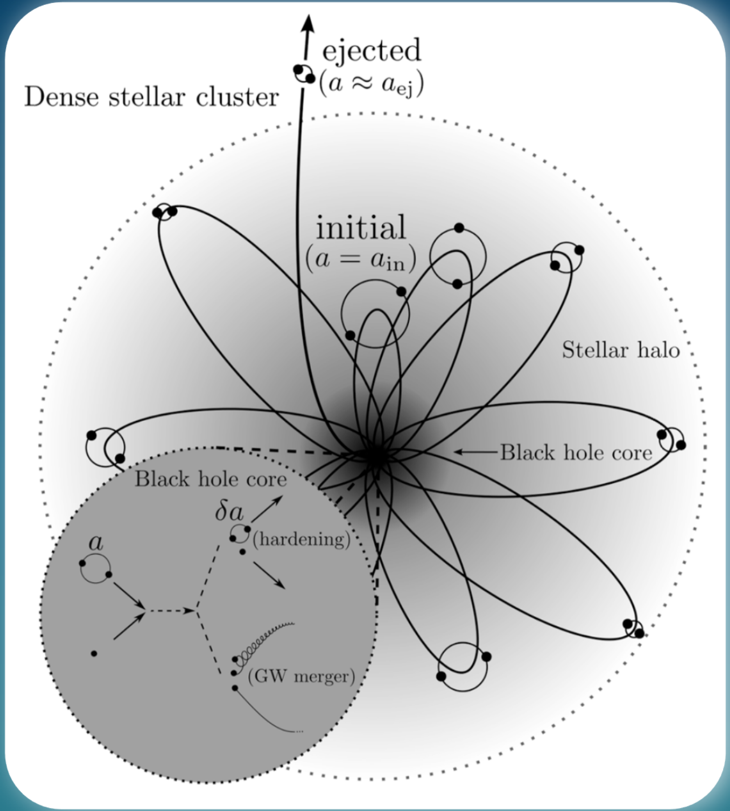
N – body interactions

(binary-single + binary-binary)

- Ejected inspirals: $f_{GW} < 10^{-2}$Hz
- In-cluster mergers: $e \sim 0$
- GW captures: $f_{GW} \geq 10^{-1}$Hz / $e \geq 0.05$



Rodriguez C. et al. (2018)



Samsing J. (2018)

Close Encounters

Subpopulation $e \sim 1$ at $f_{GW} \sim 10$ Hz

Rates (BHs): $1 - 2\ \mathrm{Gpc}^{-3}\mathrm{yr}^{-1}$

Close Encounters Waveforms: The Effective Fly-by framework

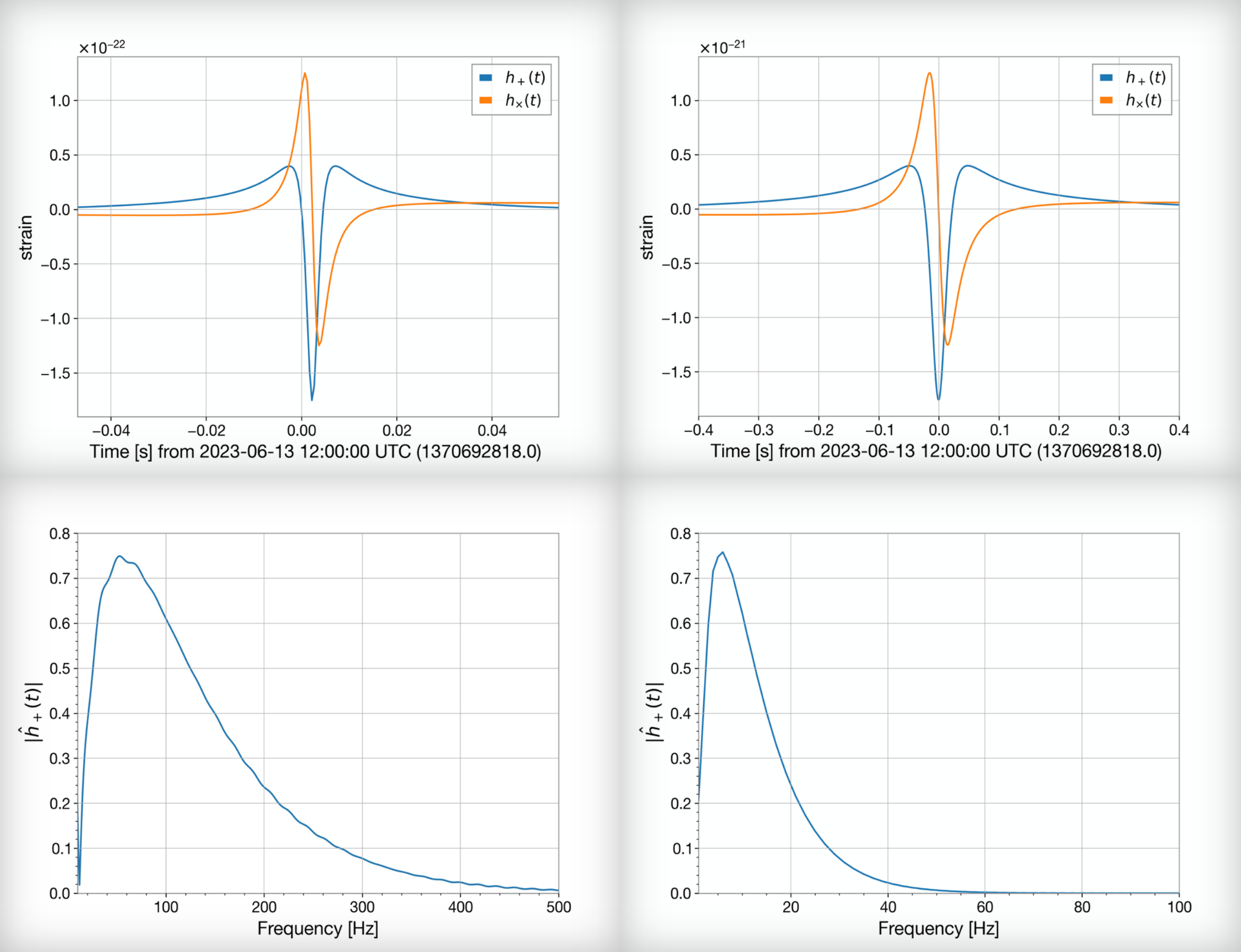
Loutrel N. (2020)

- **Single-burst** description (2.5 PN)
- parabolic orbit (fly-by) as $e \to 1$

$$h_{+,\times} \propto \frac{M^2 \eta}{p\, d_L}$$

Parameters:

- M total mass
- η symmetric mass ratio
- p semi latus-rectum
- e eccentricity
- d_L luminosity distance
- t_p time of periastron passage



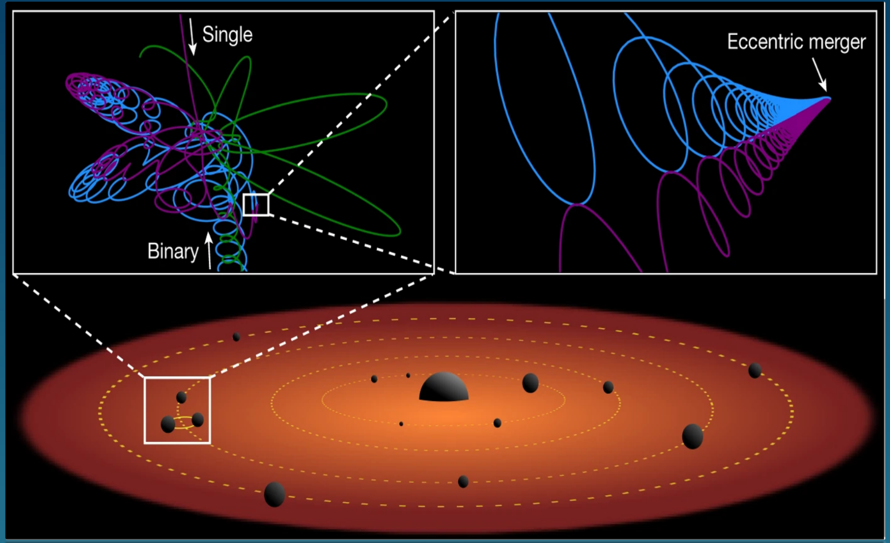
$10\, M_\odot + 10\, M_\odot$ $100\, M_\odot + 100\, M_\odot$

- **Multi-burst** sequences ⟺ known **timing model**

$$\theta_{i+1} = \theta_i + \left(\frac{d\theta}{dt}\right)_{\theta_i} T_i$$

Implications of GW Observations

Samsing J. (2022)



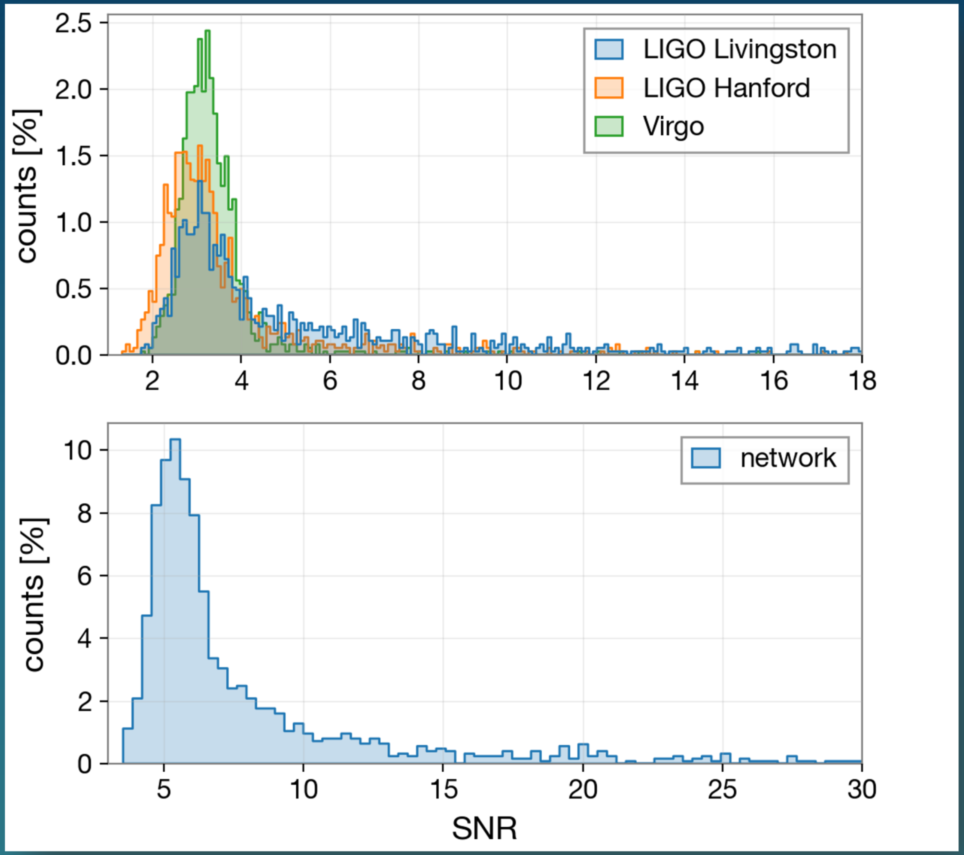
- High eccentricity ($e \sim 1$):
  - $\Longrightarrow$ GW emission: Repeated Bursts at each periastron
  - $\Longrightarrow$ $v \sim 0.7c \rightarrow$ strong field regime / GR Tests
  - $\Longrightarrow$ f-modes excitation in Neutron Stars → Equation Of State constraints
  - $\Longrightarrow$ Dynamical formation channel → Astrophysics
  - $\Longrightarrow$ Possible e.m. counterparts → Multimessenger Astronomy

Detection & Parameter Estimation required !

Close Encounters Detection

➢ Challenges :

- Low SNR with current sensitivities
- Burst-like signal with small frequency evolution



➢ Approaches :

- Power stacking of multi-bursts $SNR \sim N^{1/4}$
- Machine Learning (e.g. *Sorrentino 2023*)

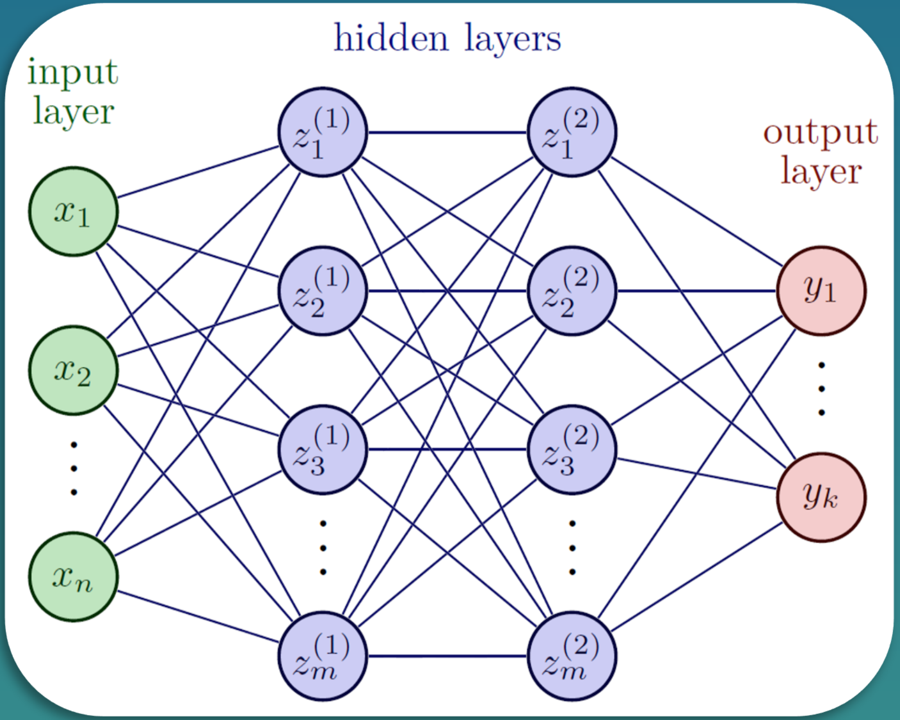
$$SNR^2 = 4\int_0^\infty df \frac{|\hat{h}(f)|^2}{S_n(f)}$$

$$SNR_{net} = \sqrt{\sum_d SNR_d^2}$$

Machine Learning & Deep Neural Networks

Artificial Neuron

$$y_j^{(L)} = \varphi^{(L)}\left[b_j + \sum_k w_{jk}^{(L)} x_k^{(L-1)}\right]$$



- Algorithms that are trained to perform specific tasks

Supervised Learning

- Classification (Noise vs Signal)
- Regression (Parameter Estimation)

- Traditional ML → point estimates
- Probabilistic ML → probability distributions

Normalizing Flows

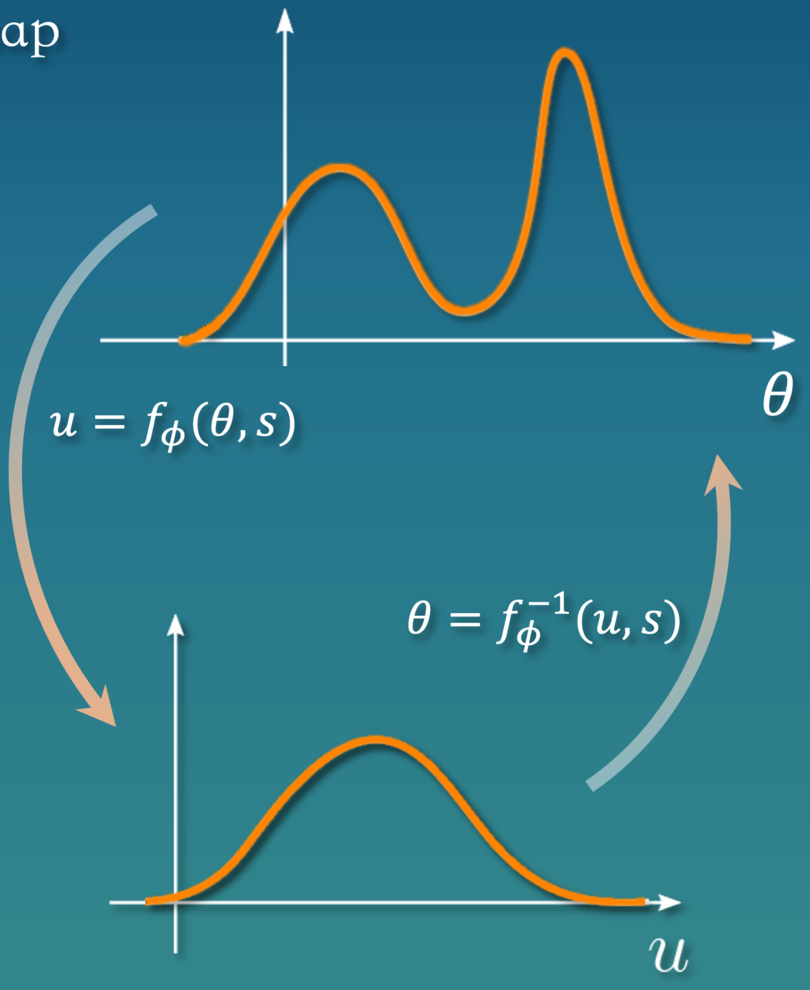
Pamakarios G. (2019)

- **Key idea:** parametrize an analytical function $f\colon \mathbb{R}^N \to \mathbb{R}^N$ to map GW parameters into a simpler distribution

$$\theta \xrightarrow{\quad f(s) \quad} u \sim \pi(u) = \mathcal{N}(0, \mathbb{I})$$

- f must:
  - be invertible and differentiable
  - depend on parameters ϕ ← NN output

$$p(\theta|s) \approx q(\theta|s) = \pi\left(f_\phi(\theta, s)\right)\left|\det J\left(f_\phi(\theta, s)\right)\right|$$



Training a Normalizing Flow

Kullback S., Leibler R.A. (1951)

- $p(\theta|\, s) \approx q(\theta|\, s) \Rightarrow$ define the *loss function* as the Kullback-Leibler divergence (distance) between p and q

$$\mathbb{KL}[p||q] = \int ds\, p(s) \int d\theta\, p(\theta|s) \log\left(\frac{p(\theta|\, s)}{q(\theta|\, s)}\right) =$$

unknown posterior

$$= \int ds\, p(s) \left[- \underbrace{\int d\theta\, p(\theta|s) \log q(\theta|s)}_{\mathbb{H}[p||q]} + \underbrace{\int ds\, p(\theta|s) \log p(\theta|s)}_{\mathbb{H}[p||p]} \right] =$$

(*using Bayes Theorem*)

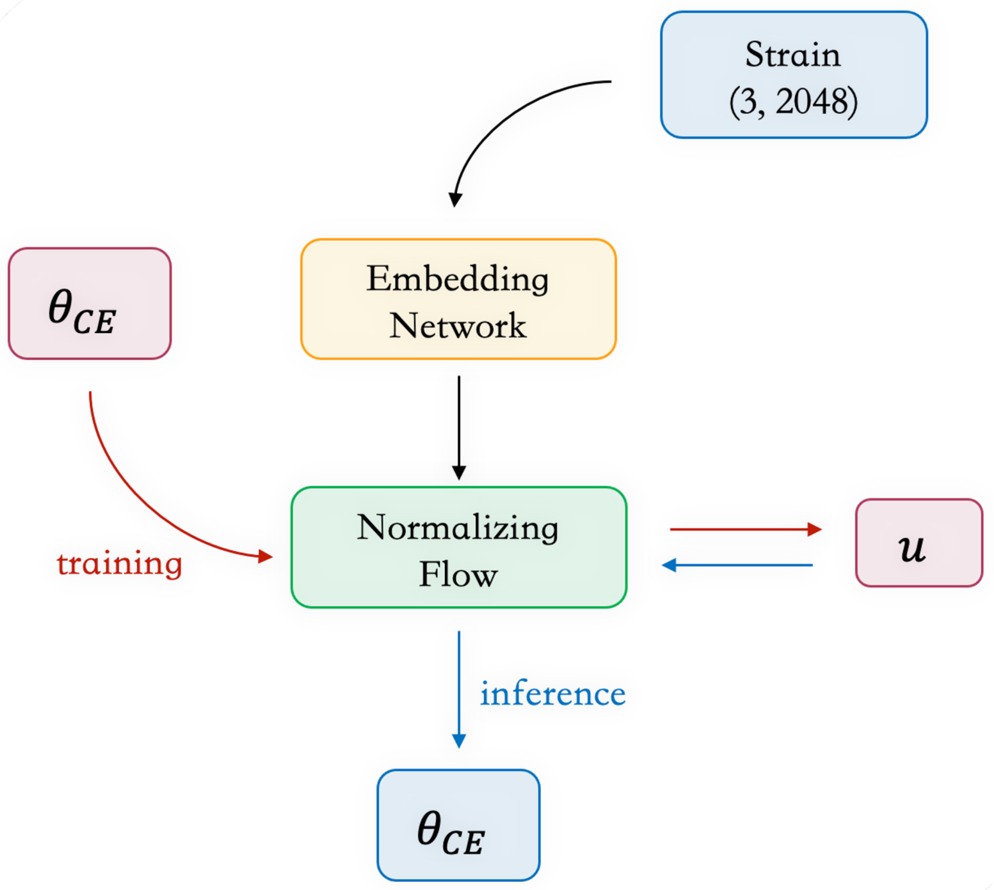
$$\simeq - \int d\theta\, p(\theta) \int ds\, p(s|\theta) \log q(\theta|s) \simeq -\frac{1}{N}\sum_{i=1}^{N} \log q(\theta_i \mid s_i)$$

Training dataset: $\begin{cases} \theta_i \sim p(\theta) \\ s_i \sim p(s|\theta_i) \end{cases}$

- *No Likelihood evaluation required*

HYPERION: Model Overview

➢ HYPer fast close EncounteR Inference from Observations with Normalizing flows



➢ Developed in Python + PyTorch

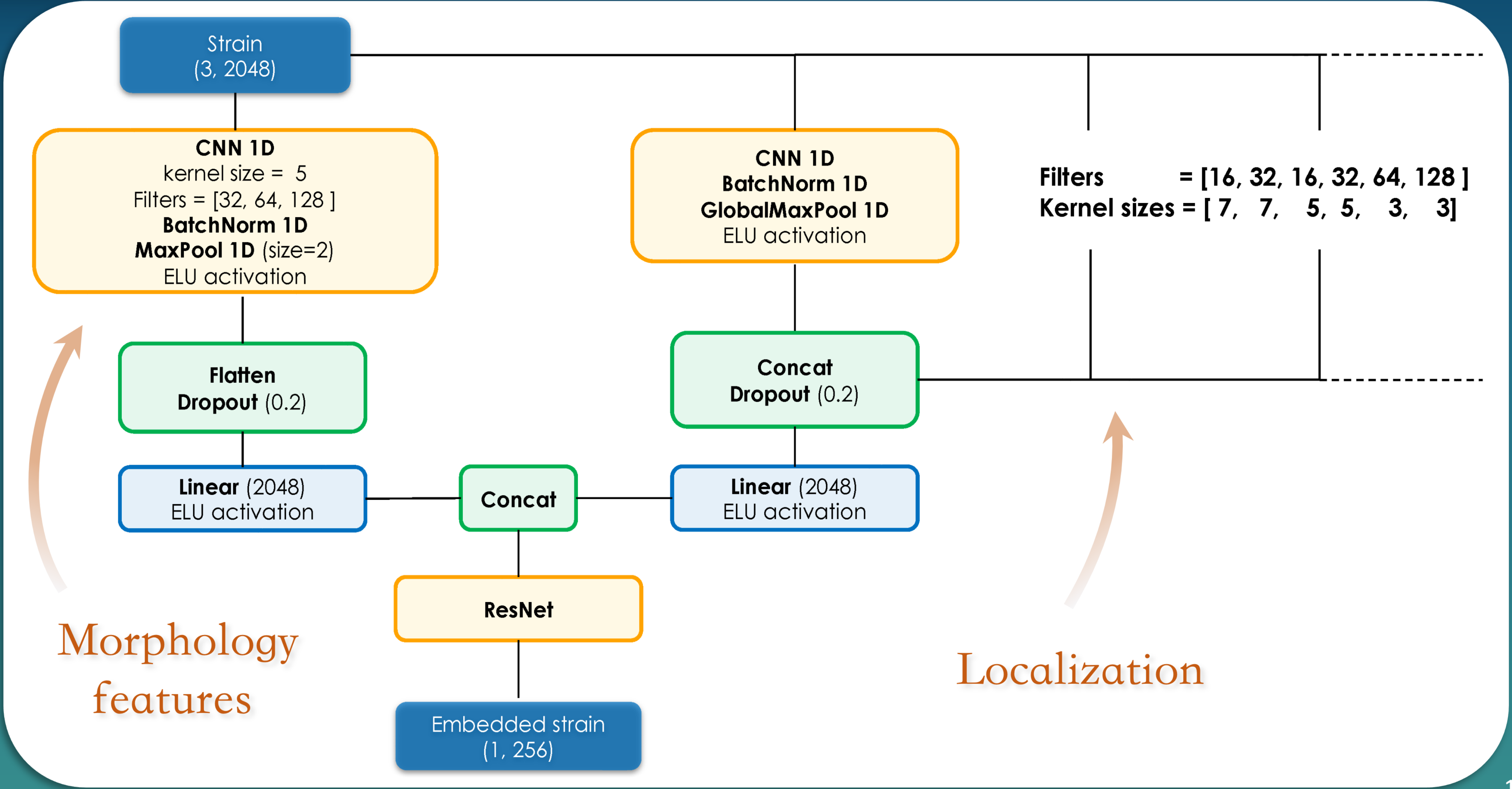
➢ Input: 1s strain at 2 kHz (3 detectors)

➢ Output: $\theta_{CE} = \{M, q, e_0, \bar{p}_0, d_L, \delta t_p, \alpha, \delta\}$

➢ Core modules:

- Embedding Network: CNN, extracts features
- Normalizing Flow: coupling layers, reconstructs posterior distribution
- $\sim$ 180 M parameters

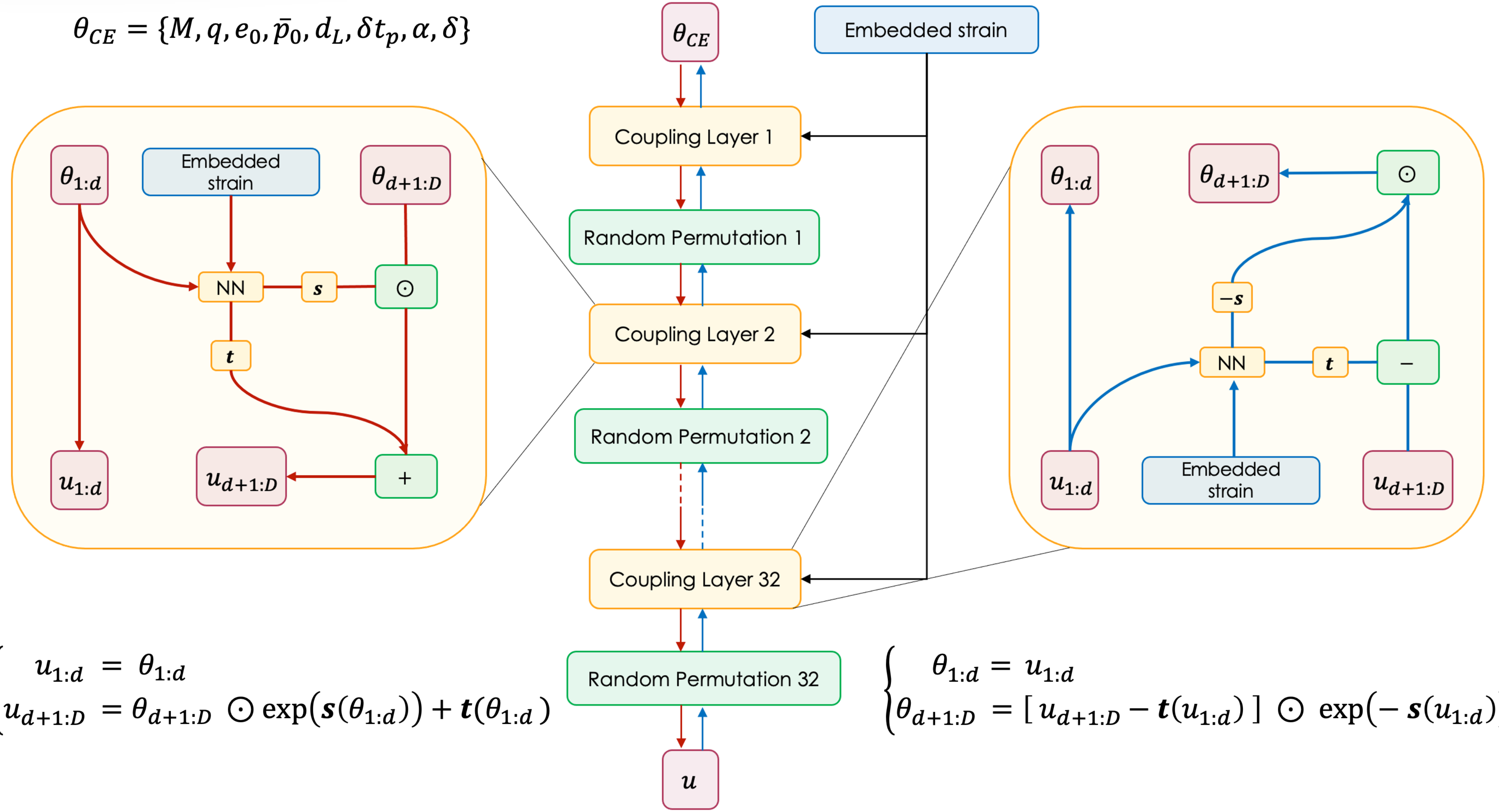
HYPERION: Embedding Network



HYPERION: Flow Architecture

➤ Affine Coupling Layers (*Dihn L. et al. (2017)*)

$$\theta_{CE} = \{M, q, e_0, \bar{p}_0, d_L, \delta t_p, \alpha, \delta\}$$

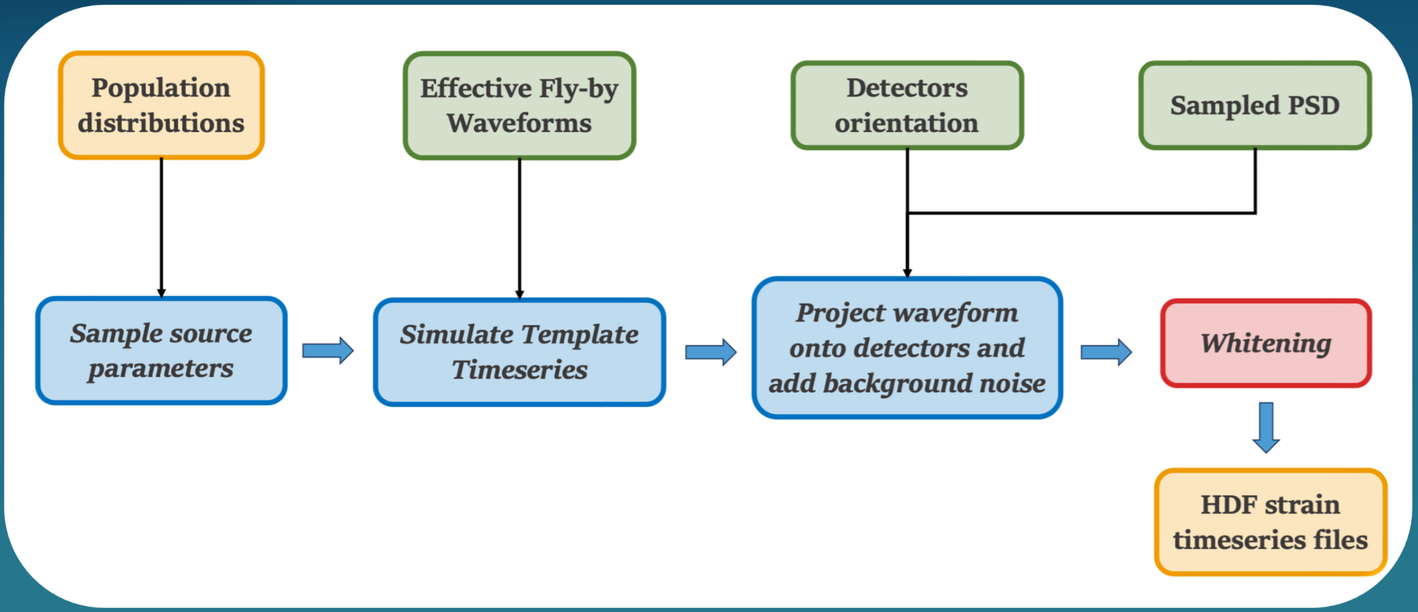


$$\begin{cases} u_{1:d} = \theta_{1:d} \\ u_{d+1:D} = \theta_{d+1:D} \odot \exp(\boldsymbol{s}(\theta_{1:d})) + \boldsymbol{t}(\theta_{1:d}) \end{cases}$$

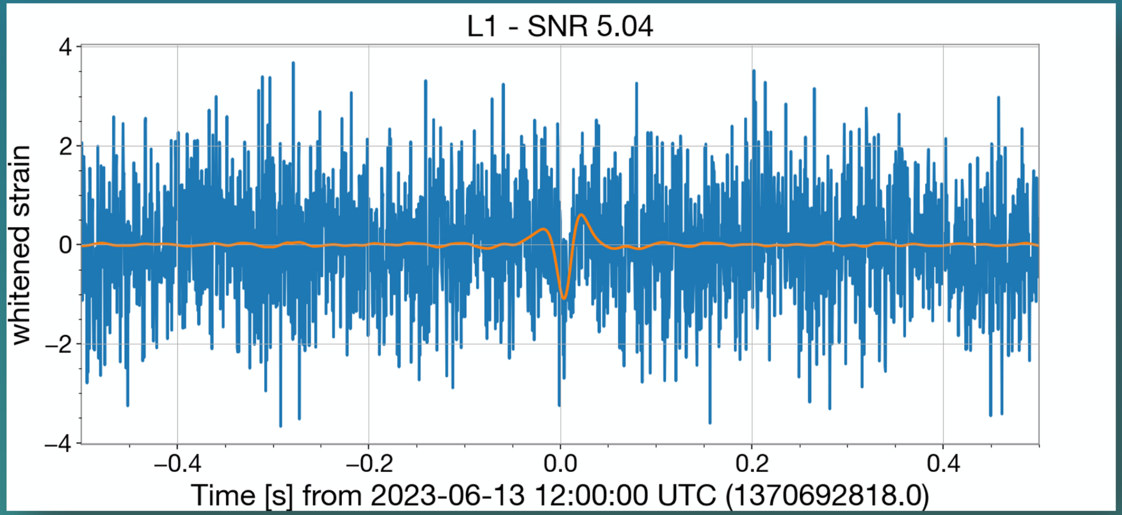
$$\begin{cases} \theta_{1:d} = u_{1:d} \\ \theta_{d+1:D} = [\, u_{d+1:D} - \boldsymbol{t}(u_{1:d}) \,] \odot \exp(-\boldsymbol{s}(u_{1:d})) \end{cases}$$

Simulations and preprocessing

θ_{CE}	distribution	min	max
m_1 [$M_\odot$]	uniform	10	100
$m_2 \leq m_1$ [$M_\odot$]	uniform	10	100
$\bar{p}_0$	uniform	13	25
d_L [Mpc]	uniform	100	2000
e_0	uniform	0.85	0.95
α	uniform	0	2π
δ	cos	$-\pi/2$	$\pi/2$
δt_p [s]	uniform	-0.25	0.25
ψ	uniform	0	π
ι	sin	0	π
GPS time	fixed	1370692818.0	

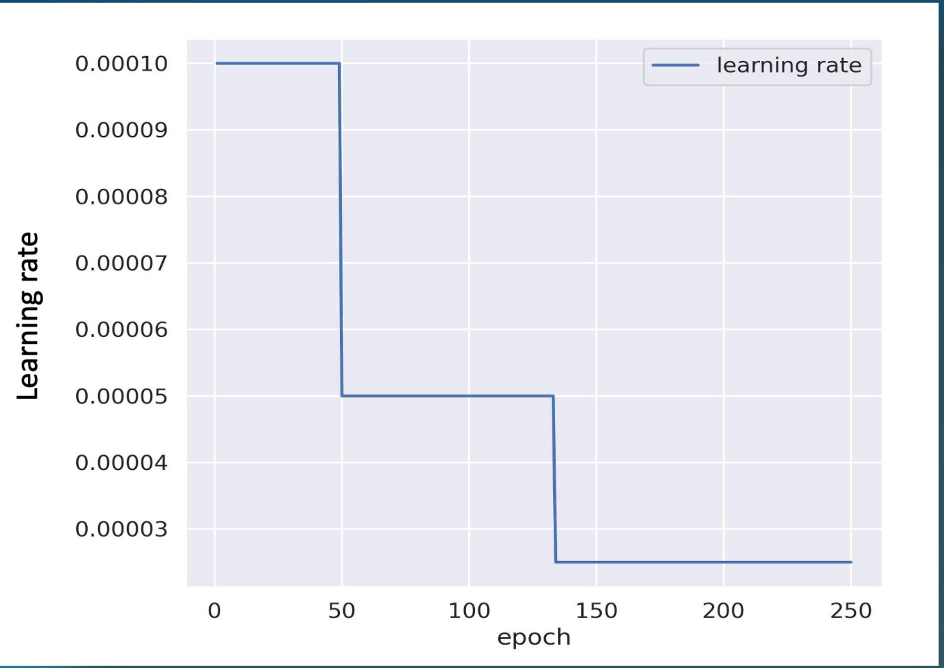


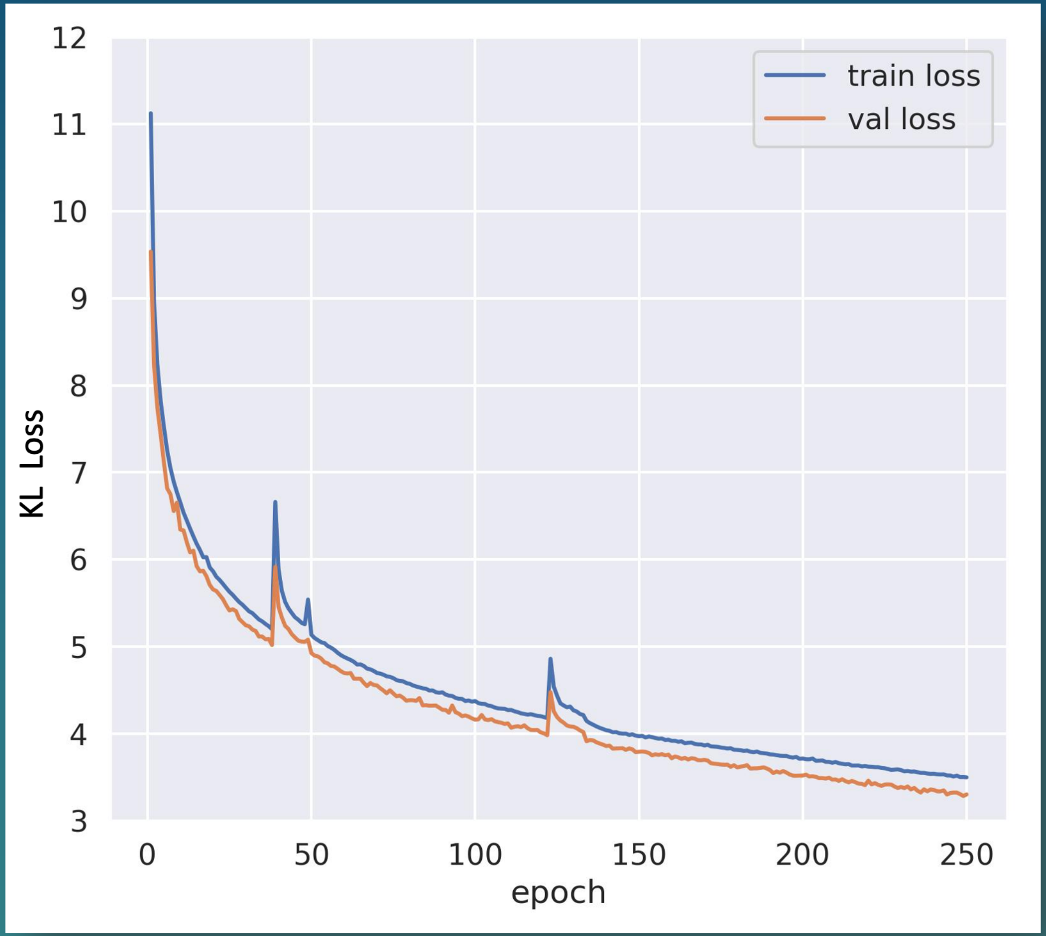
$5{\times}10^6$ training samples



Training

➢ ~ 20 hr on Nvidia A30 GPU





- Trained for 250 epochs
- Batch size = 512
- Training Dataset split: 90% training + 10% validation
- ADAM optimizer
- δt_p augmentation
- Learning rate reduced by 50% if validation loss did not improve after 10 epochs

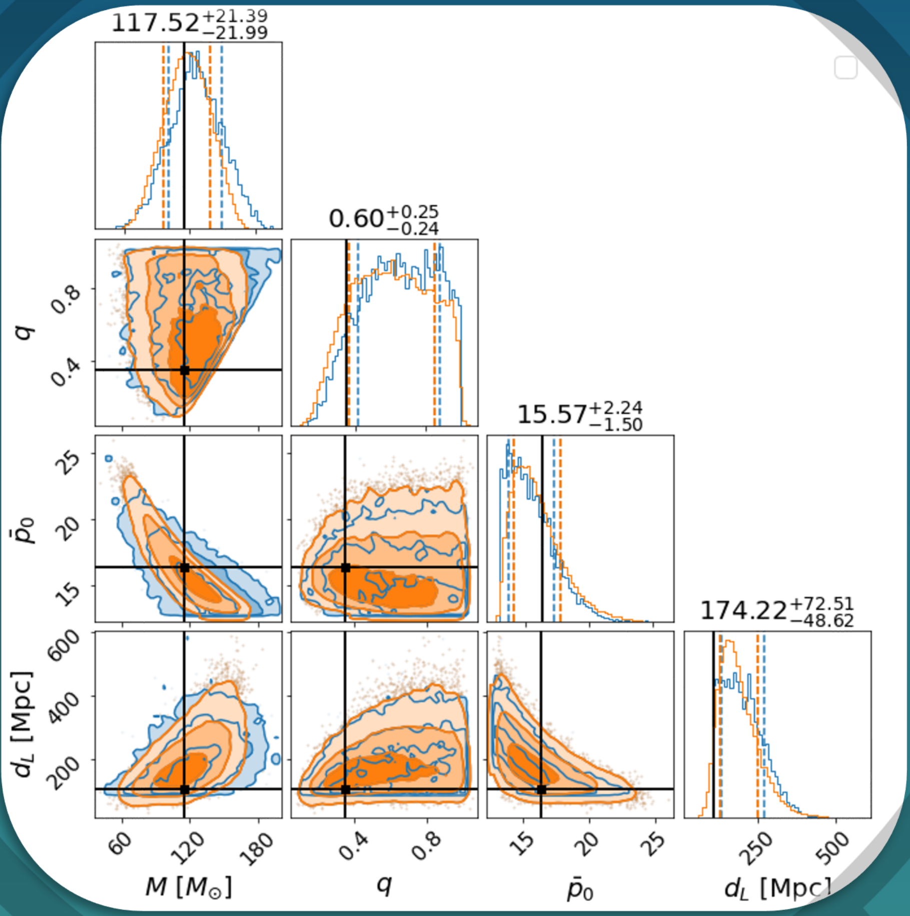
Results and Performance

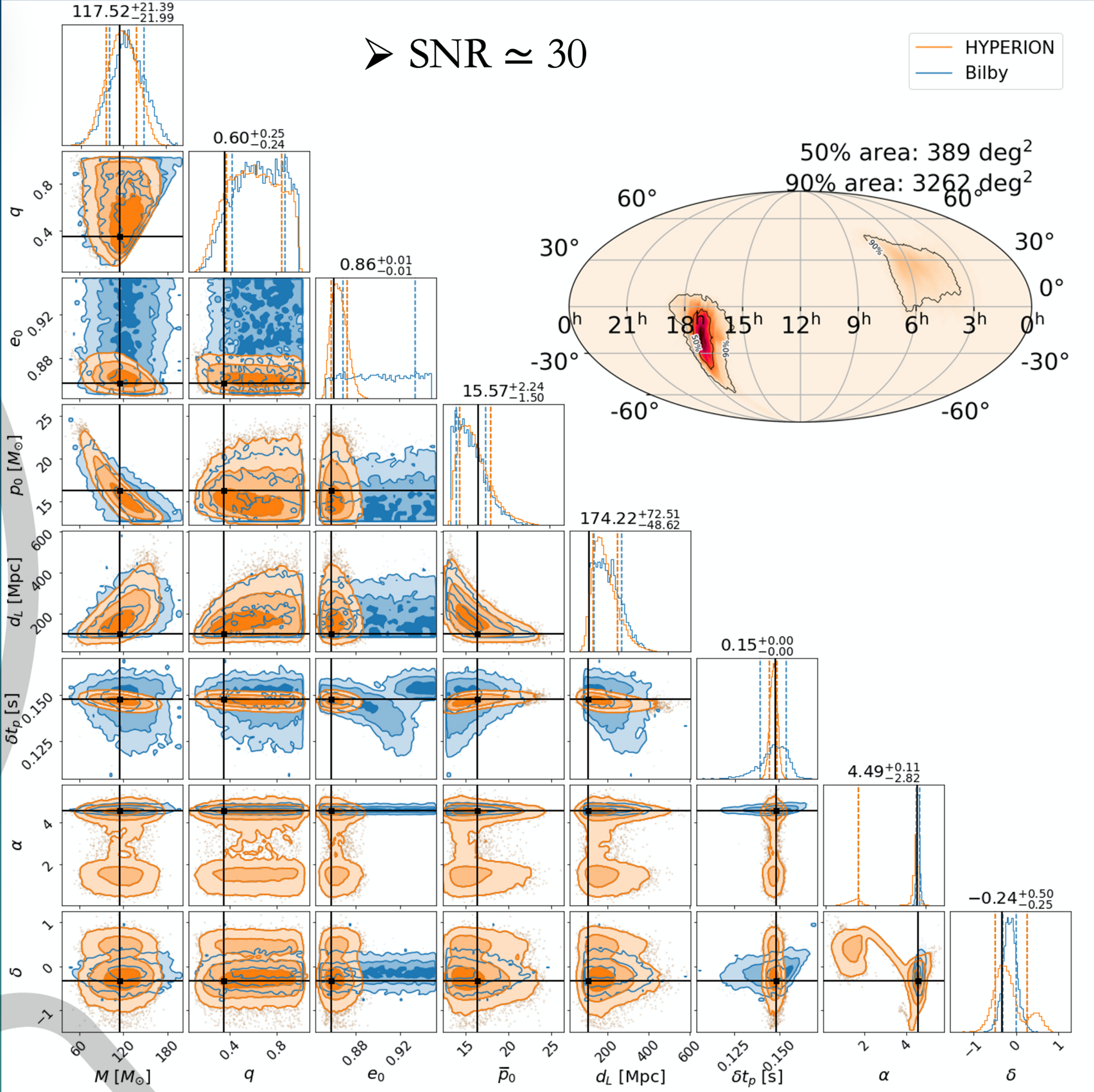
➢ Comparison with standard PE tools e.g. Bilby: (*Ashton G. et al. (2018)*) :

- Dynesty Sampler
- nact = 50
- nlive = 1000
- N CPU = 42

	Posterior samples	Inference time
Bilby	$\sim 5\times10^3$	$\sim$ 10 h
HYPERION (CPU)	5×10^4	$\sim$ 16 s
HYPERION (GPU)	5×10^4	$\sim$ 0.5 s

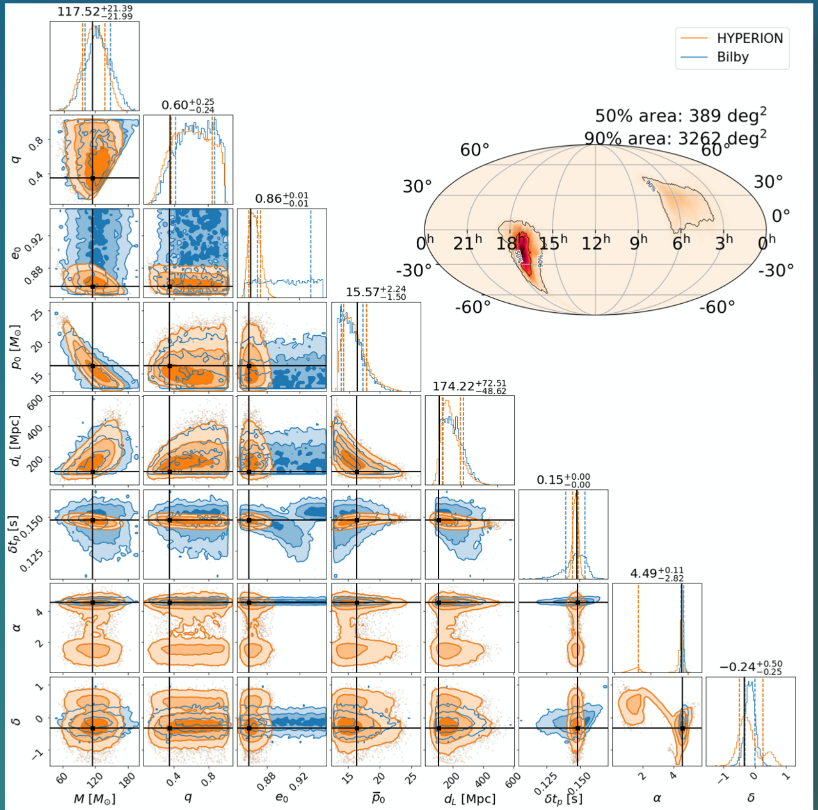
Results and Performance



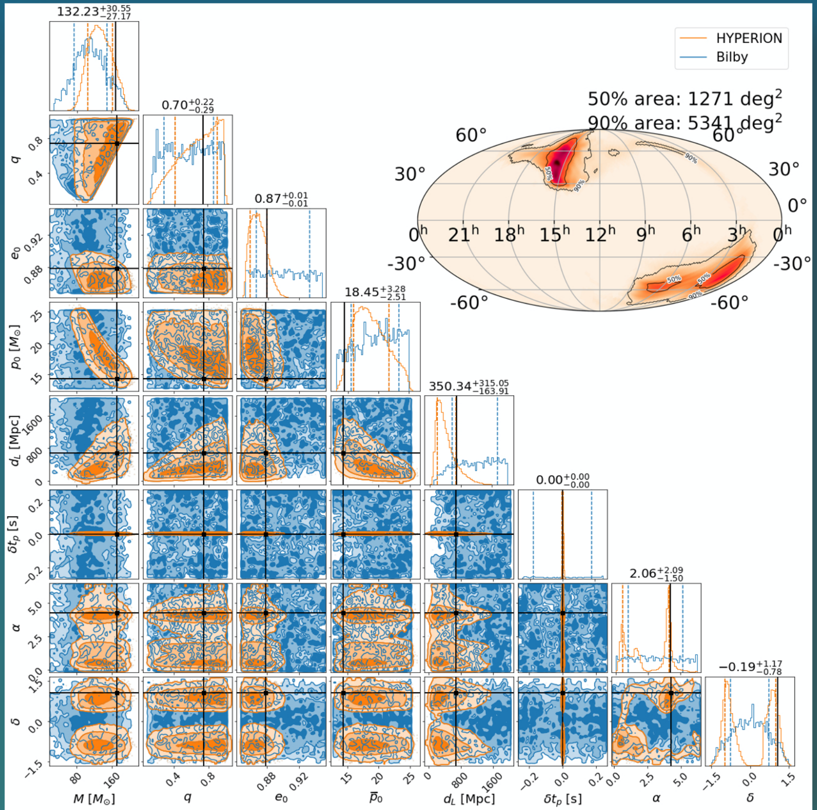


Results and Performance

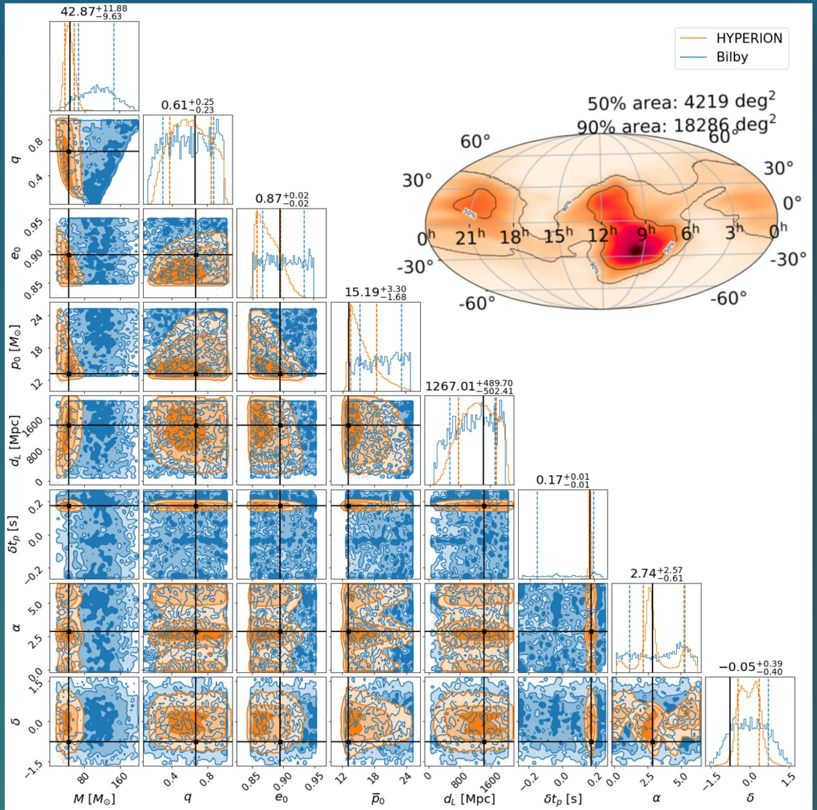
- ➤ SNR ≃ 30



- ➤ SNR ≃ 12



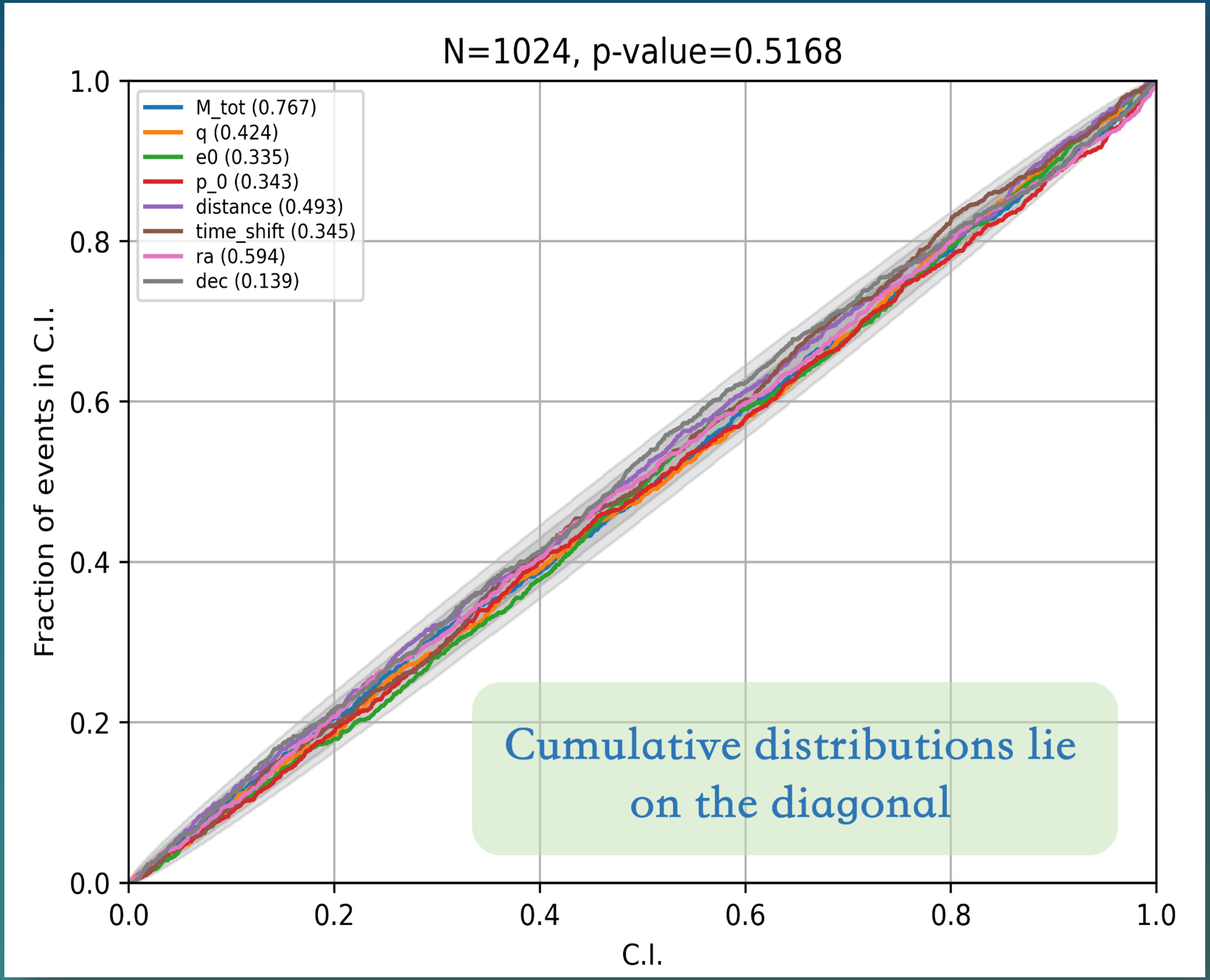
- ➤ SNR ≃ 6



Results and Performance: Probability - Probability Plot

1. Obtain posterior for 1024 events from the test dataset
2. For each parameter compute the percentile score of the true value in the marginalized posterior
3. Take the cumulative distributions
4. Kolmogorov-Smirnov Test (95%)

➢ Promising approach on simulations



Importance Sampling & Bayes Factors

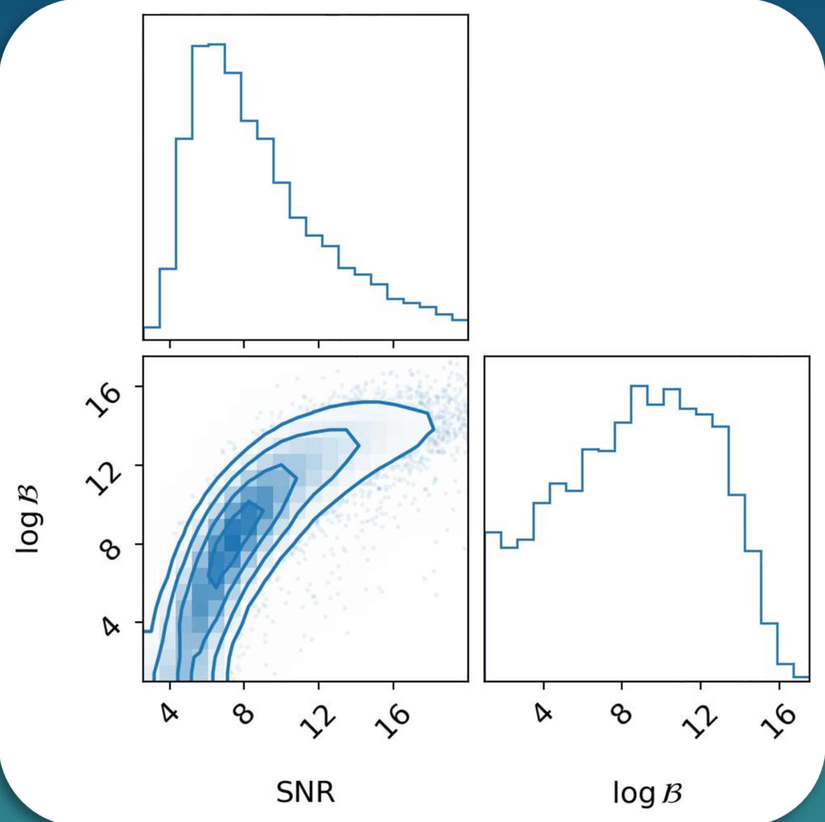
Murphy K. (2023)

➢ Bayes Factor

$$\mathcal{B}_{12} = \frac{\mathcal{Z}_1}{\mathcal{Z}_2}$$

$\mathcal{Z}_1$ → signal + noise

$\mathcal{Z}_2$ → noise



➢ Importance Sampling

$$\mathcal{Z}_1 = \int d\theta\, p(\theta)\, p(s|\,\theta) = \int d\theta\, \frac{p(\theta)\, p(s|\theta)}{q(\theta|s)}\, q(\theta|s)$$

$$\hat{\mathcal{Z}}_1 = \frac{1}{N}\sum_i \frac{p(\theta_i)\, p(s_i|\theta_i)}{q(\theta_i|s_i)} = \frac{1}{N}\sum_i w_i \qquad \text{(Unbiased estimator)}$$

Conclusions and future works

- Normalizing Flows for Close Encounters :
  - CEs are interesting but challenging sources
  - Likelihood – free approach
  - $\sim 10^4$ times faster than traditional methods → low latency → e.m. follow-up
  - To our knowledge, first application to CE

- Future work :
  - Test detection capabilities on real data
  - Perform systematic searches over O3

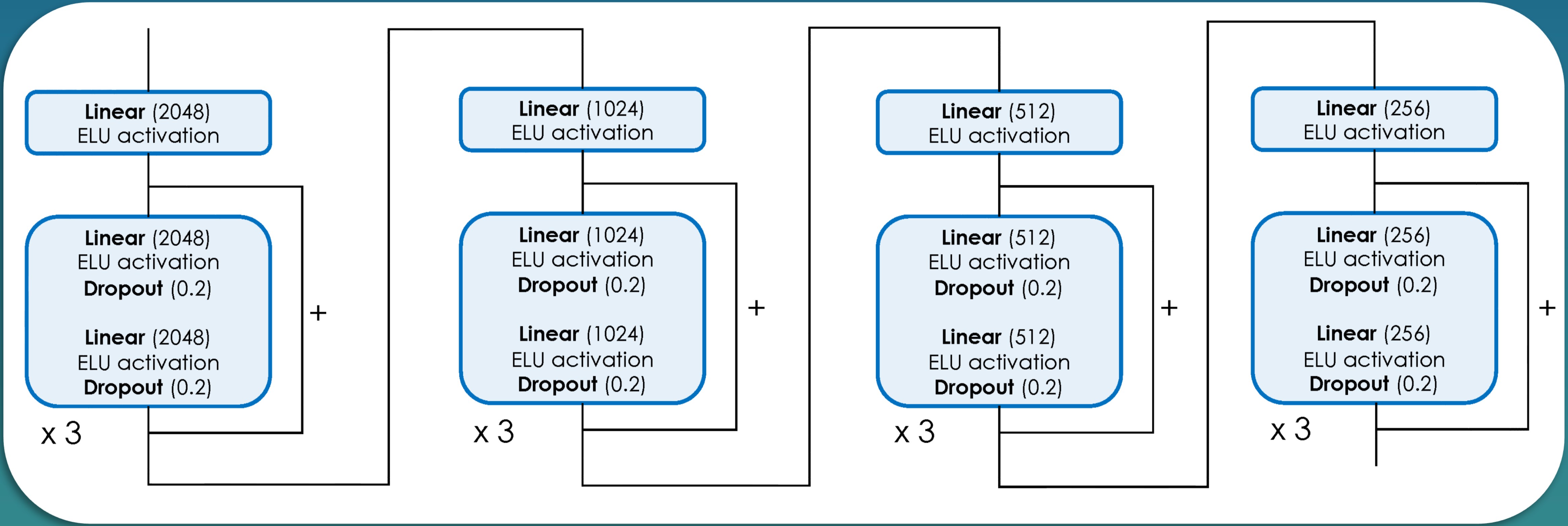
- Method Paper: VIR-0011B-24 , DCC: P2400041 (PRD submitted)

Thank You for the Attention !

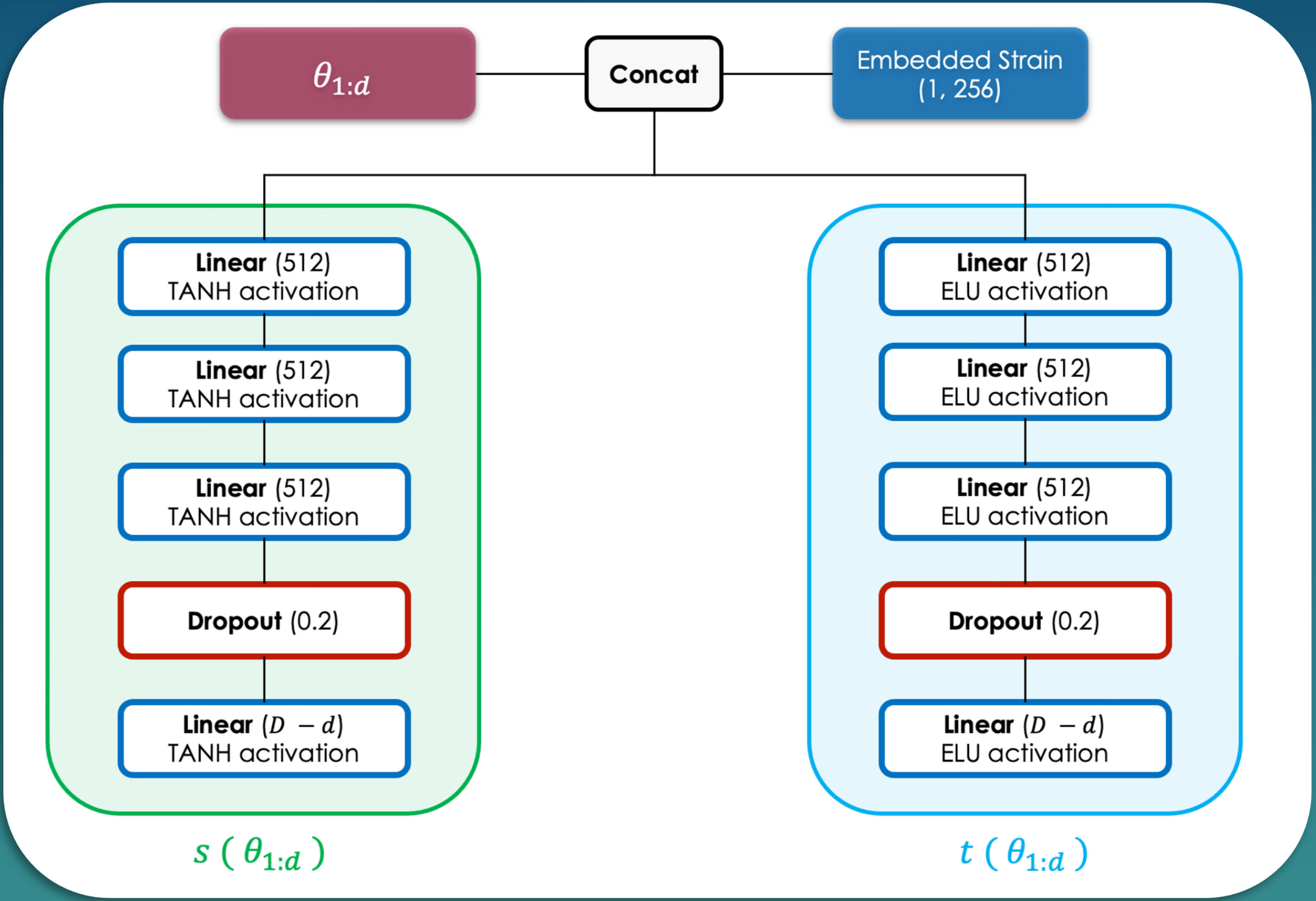
Backup slides

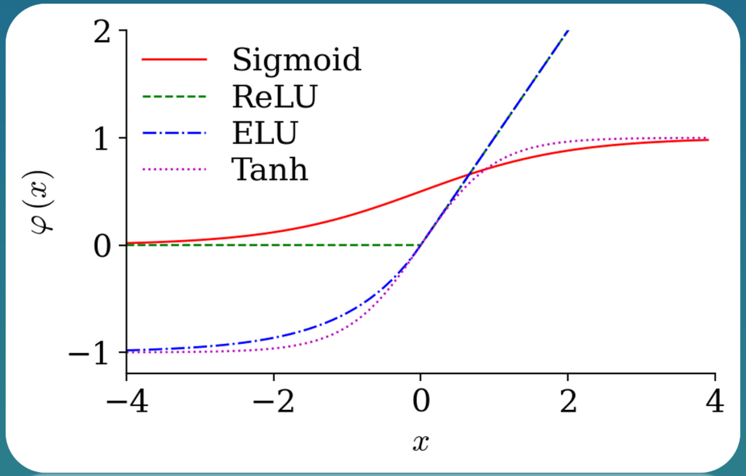
HYPERION: Embedding Network – ResNet block

- ResNet ⇒ more efficient information compression



HYPERION: Flow Architecture





Effective Fly-by Waveforms

$$h_{+,\times}(t) = -\frac{M^2\eta}{p[\ell(t)]d_L}\sum_{k=0}^{6}\sum_{n=0}^{2}\epsilon^n\Phi^{(n,k)}(\iota,\psi)$$

$$e(t) = e(\ell=0) + \left(\frac{de}{d\ell}\right)_{\ell=0} + O(\ell^2) =$$
$$= e_0 - \frac{304}{15}\frac{\eta e_0}{\bar{p}_0^{5/2}}\left(1+\frac{121}{304}e_0^2\right)\ell(t) + O(\ell^2)$$

$$\bar{p}(t) = \bar{p}(\ell=0) + \left(\frac{d\bar{p}}{d\ell}\right)_{\ell=0} + O(\ell^2) =$$
$$= \bar{p}_0\left[1 - \frac{64}{5}\frac{\eta}{\bar{p}_0^{5/2}}\left(1+\frac{7}{8}e_0^2\right)\ell(t) + O(\ell^2)\right]$$

$$\ell(t) = \frac{n_0}{2\pi F_{rr}}\exp\{[2\pi F_{rr}\,(t-t_p)]-1\}$$

$$F_{rr} = \frac{96}{10\pi}\frac{\eta}{M\,\bar{p}_0^4}(1-e_o^2)^{3/2}\left(1+\frac{73}{24}e_0^2+\frac{37}{96}e_0^4\right)$$

$$n_0 = M^{-1}\left(\frac{1-e_0^2}{\bar{p}_0}\right)^{3/2}$$