Fast gravitational-wave parameter estimation for compact binary coalescences using efficient parameterizations and normalizing flows

Speaker: Kazuki Takada[A]

Collaborator: Soichiro Morisaki[A], Vivien Raymond[B], Masaki Iwaya[B]

A: ICRR, The University of Tokyo, B: Cardiff University

The purpose GW data analysis

Observational GW is determined by the parameters below.

$$\theta = (m_1, m_2, \boldsymbol{s}_1, \boldsymbol{s}_2, d_L, \alpha, \delta, \theta_{JN}, \psi, t_c, \phi_c)$$

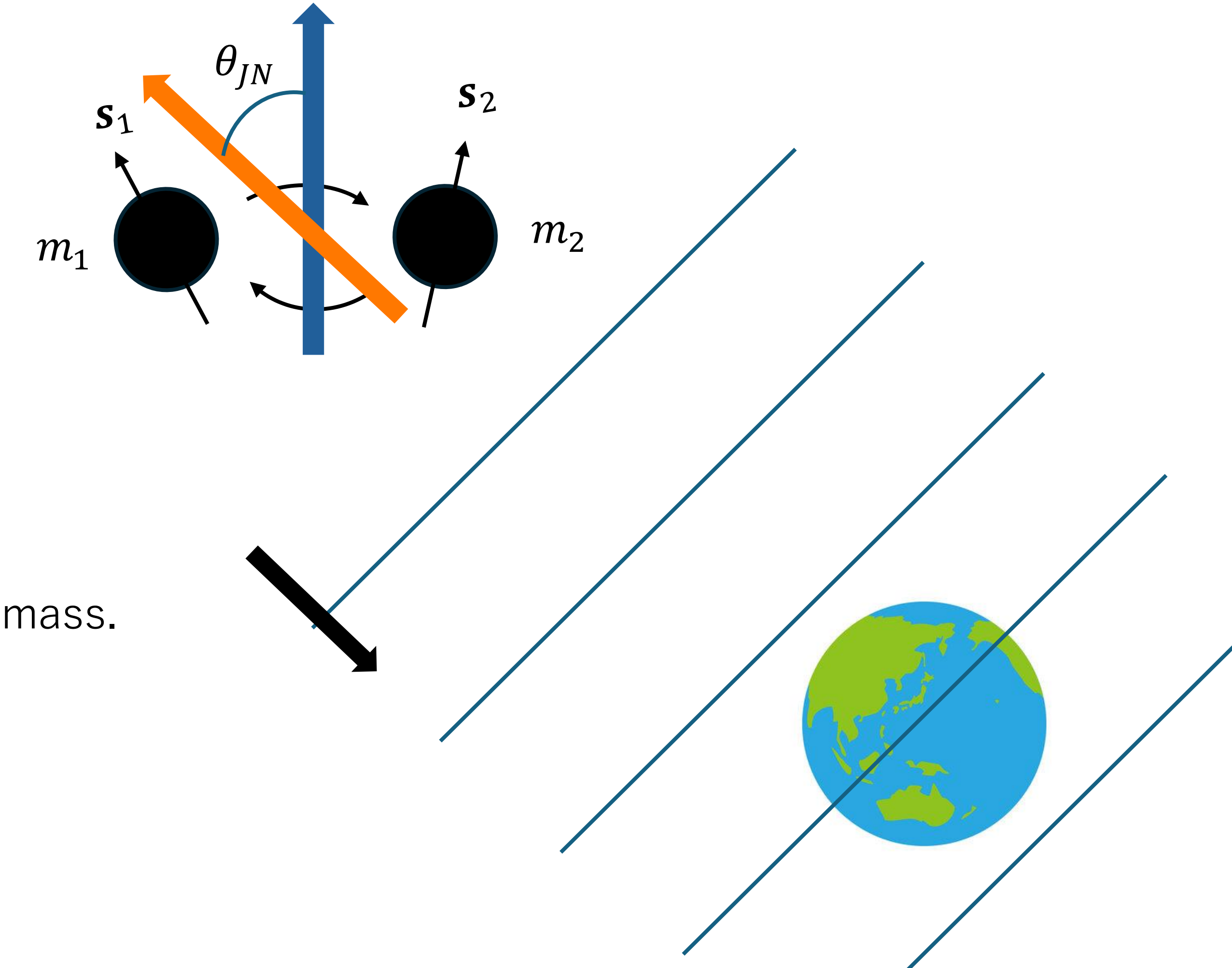
m_1, m_2: Mass for each component

$\boldsymbol{s}_1, \boldsymbol{s}_2$: Spin for each component

d_L, α, δ: Sky position

t_c, ϕ_c: Coalscence time & phase

θ_{JN}, ψ: Inclination & Polarization angle



We also use well observed parameter below about mass.

$M_c = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}$:Chirp mass

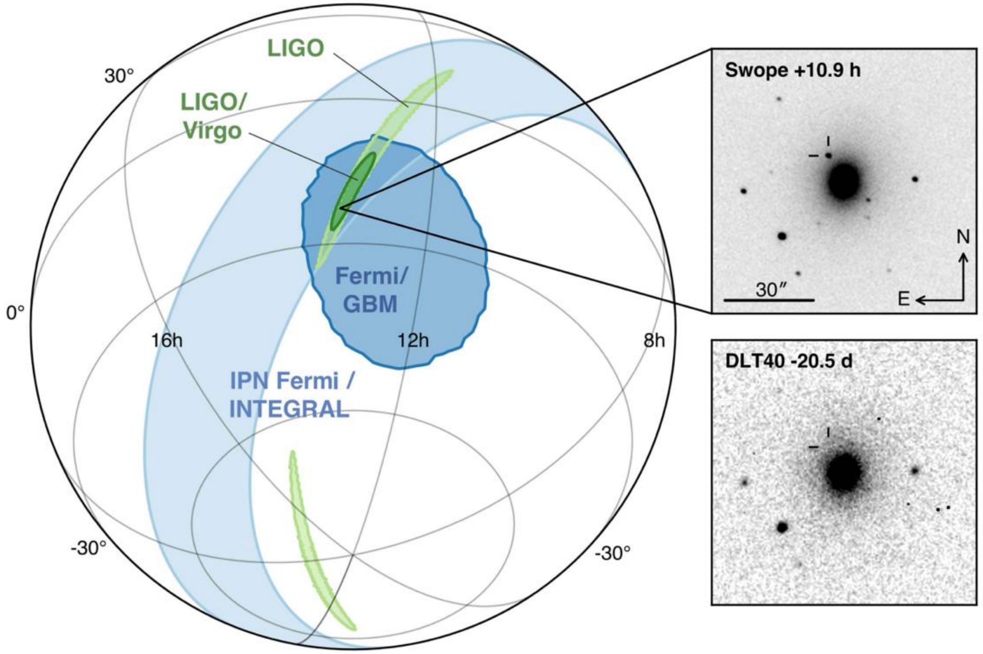
$q = \frac{m_2}{m_1}$: Mass ratio

Golden Event for Multi Messenger Astronomy：GW170817

GWs provide key information for electromagnetic follow-up.

・Mass, Phase evolution → Neutron Star or Black hole?

・Distance, Sky coordinates → Identify the follow-up area



Abbott et al., Astrophys. J. Lett. 848, L12 (2017)

・Analyzed LIGO-Virgo data

→ Inferred sky area: $\sim 30\mathbf{deg}^2$

$<< 1000\mathbf{deg}^2$ (by Fermi)

・Following Search by other telescopes

→ The host galaxy was identified 11h later

Parameter Estimation

GW source properties are traditionally inferred using Bayesian inference.

$$p(\theta|d) = \frac{L(d|\theta)\pi(\theta)}{p(d)}$$

$p(\theta|d)$: Posterior

$\pi(\theta)$: Prior

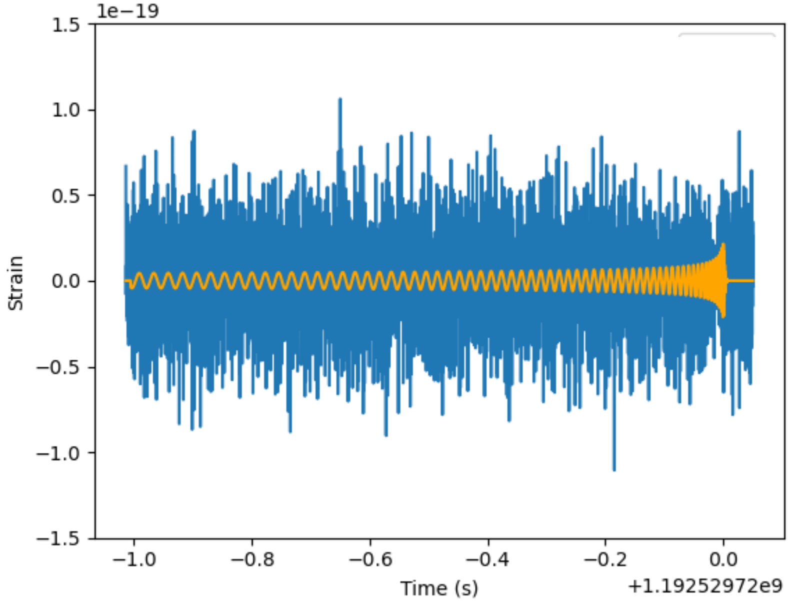
$L(d|\theta)$: Likelihood

$p(d)$: Evidence

θ: Parameter

- m_1, m_2: Mass
- $\boldsymbol{s}_1, \boldsymbol{s}_2$: Spin
- d_L, α, δ: Sky position
- t_c, ϕ_c: Time & Phase
- θ_{JN}, ψ: Inclination & Polarization

d: Observed data



The Problems of Bayesian Inference

① In BNS case, the signal is very long.

Time to merger from $f/2$[Hz] orbital frequency:

$$T = \frac{5}{256}\left(\frac{GM_c}{c^3}\right)^{-\frac{5}{3}}(\pi f)^{-\frac{8}{3}}$$

In the case of $f = 20$[Hz],

- (BBH) $m_1 = m_2 = 30M_\odot$: ~1sec
- (BNS) $m_1 = m_2 = 1.4M_\odot$:~150 sec

② We have to evaluate the likelihood as many as $10^7 \sim 10^8$ times.

$$L(d|\theta) = \exp\left\{-2\int_{f_{\min}}^{f_{\max}} \frac{|d(f) - h(f;\theta)|^2}{S_n(f)}df\right\}$$

→ It’s computationally expensive and takes much time.

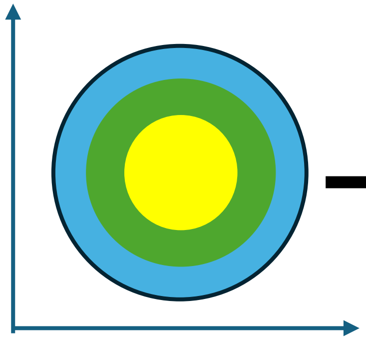
Various methods have been proposed to reduce the inference cost.

Towards Rapid Parameter Estimation: Machine Learning

- Machine learning is a hot topic on GW data analysis.

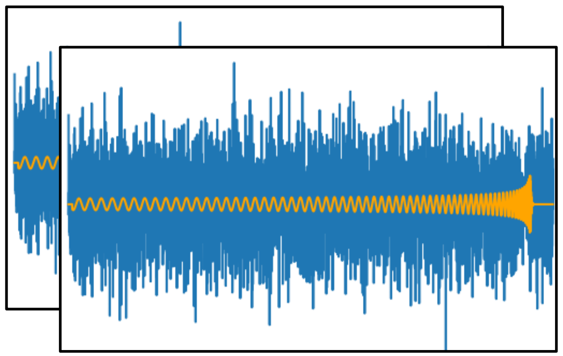
  Ref. Dax+ PRL 2021; Gabbard+ Nat. Phys. 2022; Bhardwaj+ PRD 2023; Raymond+ MNRAS 2025; Marx+ PRD 2026; Roulet+ PRD 2026 ⋯

- Basic idea: use neural networks(NN) to model complex distributions.



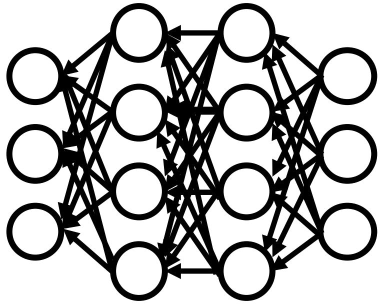
①: Sample θ from prior

$$\theta \sim p(\theta)$$

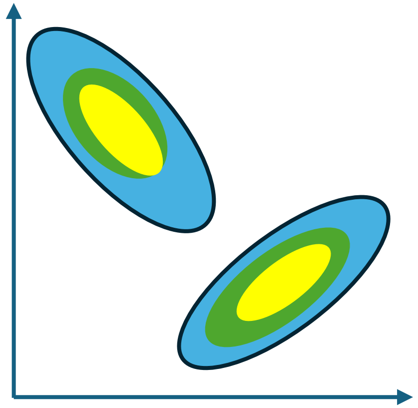


②:Generate mock data

$$d \sim p(d|\theta)$$



③: Train the NN by (θ, d) datasets



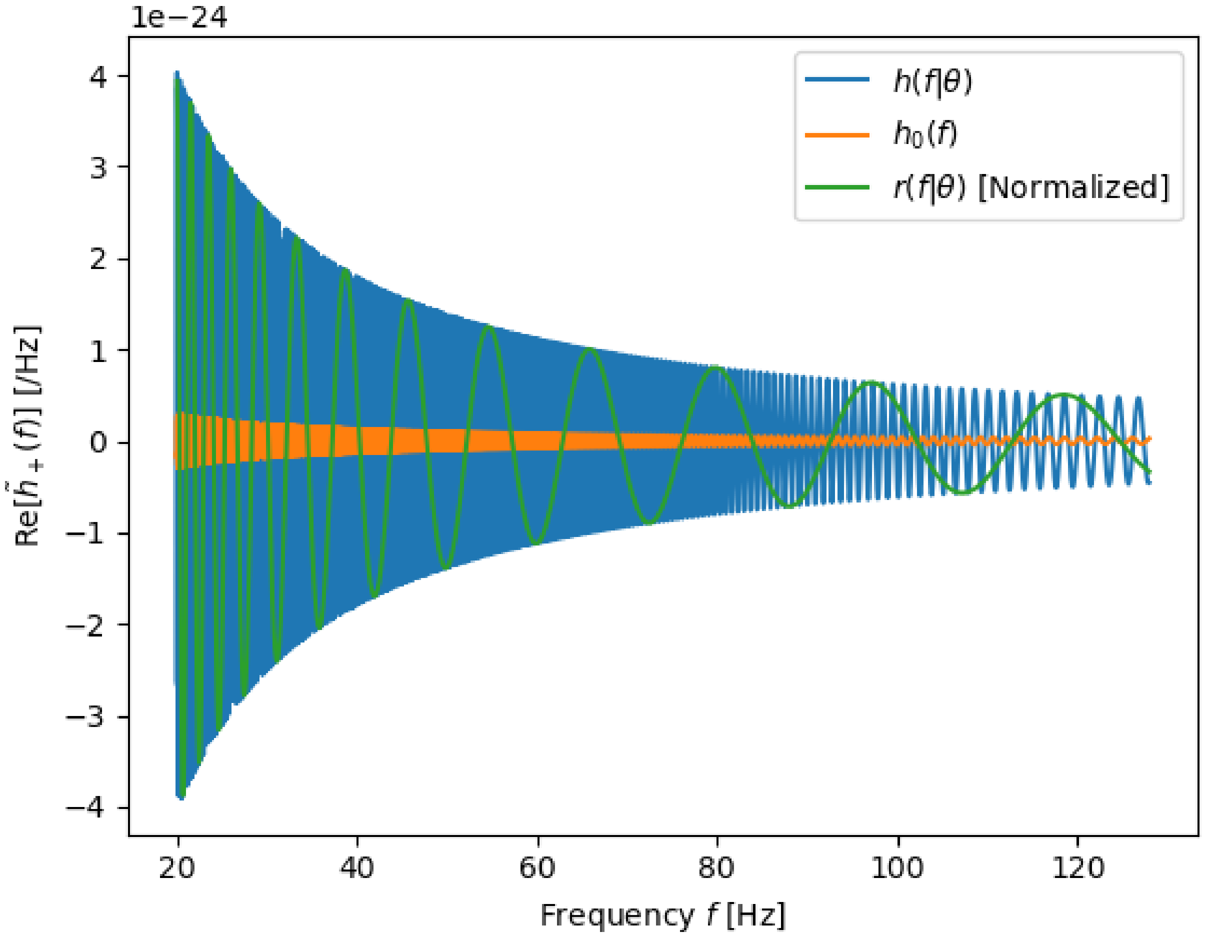
④: After training, we can get posteriors in seconds

Previous work

- Machine learning has already been applied to BNS GW analysis.

  Dax et al. Nature 639, 49–53 (2025), Roulet et al. arXiv:2604.08897 (2026), Iwaya et al. arXiv:2604.22380(2026)

- Data compression via "Relative Binning" enables training for BNS data.



Ref. Zackay et al. (2018) arxiv:1806.08792

Waveform ratio: $r(f|\theta) = \dfrac{h(f|\theta)}{h_0(f)}$

→ Approximate the ratio linearly within each frequency interval.

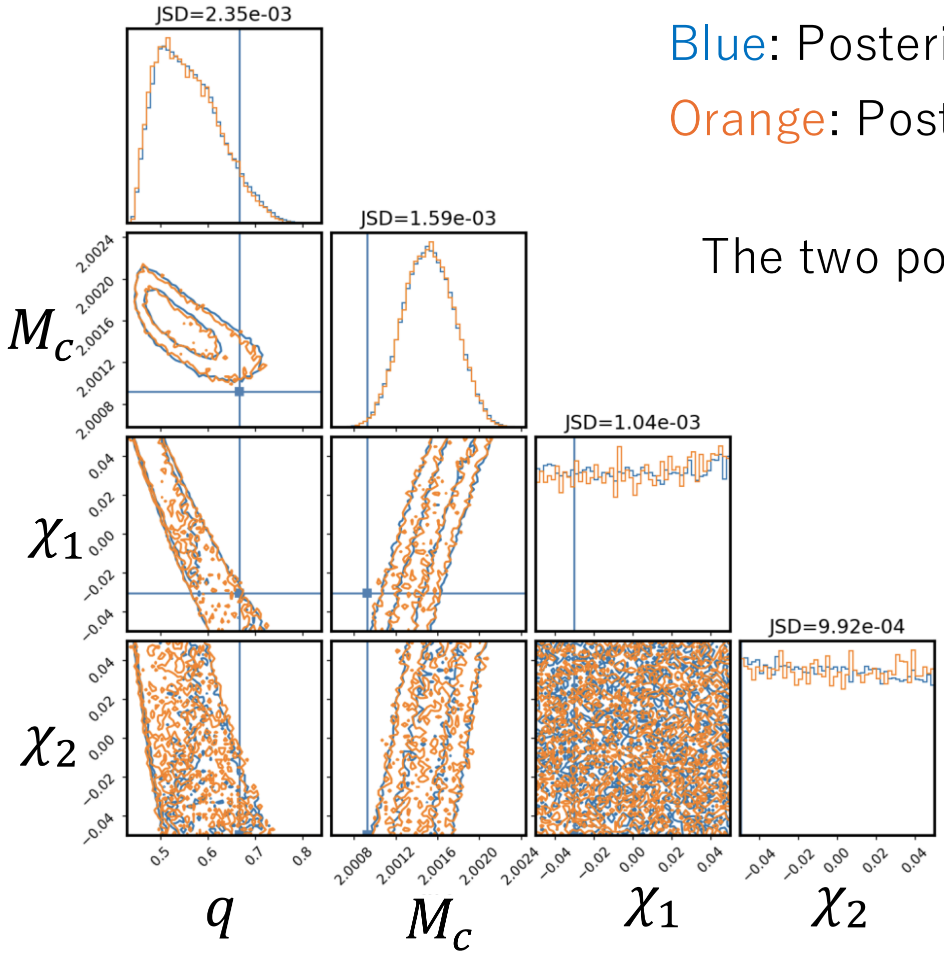
→ Compress data dimension
$O(10^5) \rightarrow O(10^2)$ points

$\lesssim$ 0.5 % of original data size

Previous work

Ref. M.Iwaya, V.Raymond, S.Morisaki, **K.T.** (2026), arXiv:2604.22380

- Machine learning(SBI) has already been applied to BNS GW analysis.
- Data compression via “Relative Binning” enables training for BNS data.



Blue: Posterior from SBI

Orange: Posterior from Bilby(Baysian inference)

The two posteriors agree well.

We successfully developed an inference model for BNS combined with ”Relative Binning”.

Inference time:

Bilby (Bayesian inference): 2-3 days

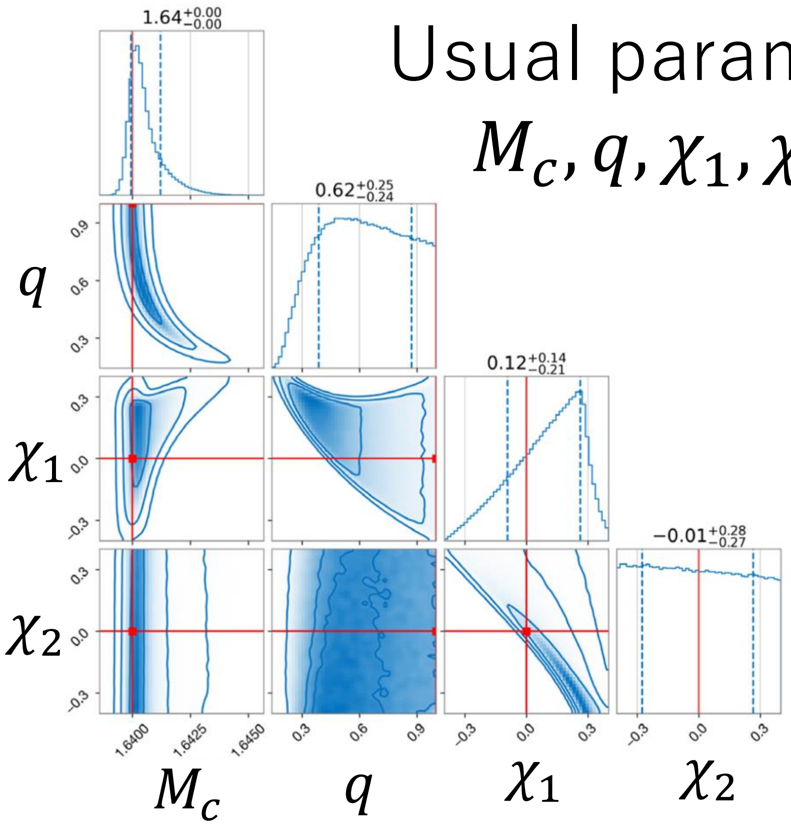
SBI: 2-3 seconds

Towards Rapid Parameter Estimation: Efficient Parameterization

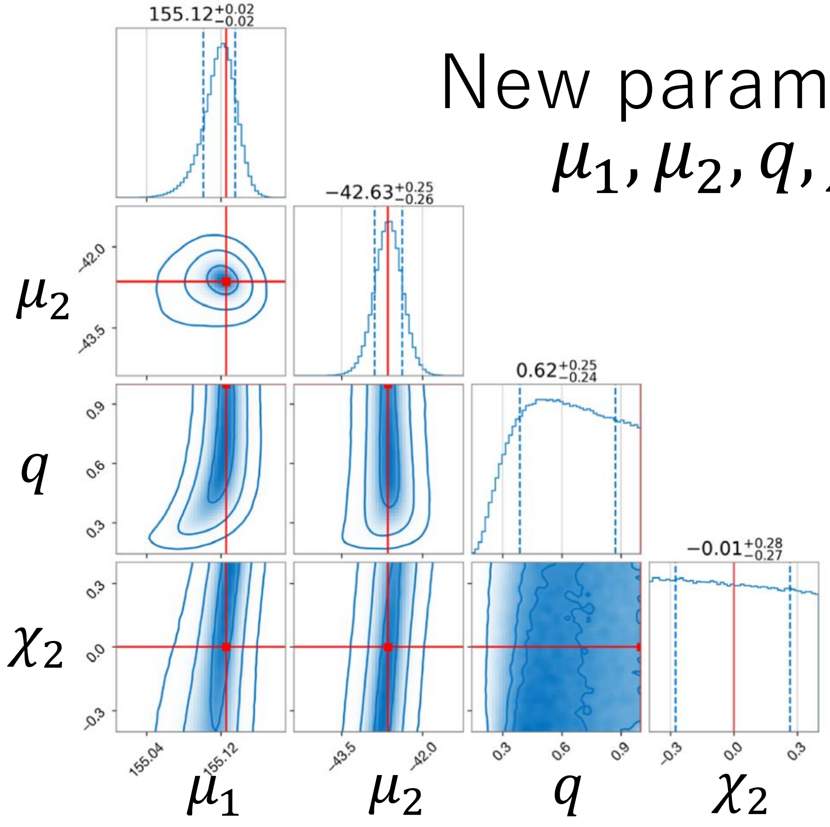
New parameters μ_1 and μ_2 were proposed to simplify the posterior.

Ref. Lee et al., PRD 105, 124057 (2022)

Usual parameters
M_c, q, χ_1, χ_2



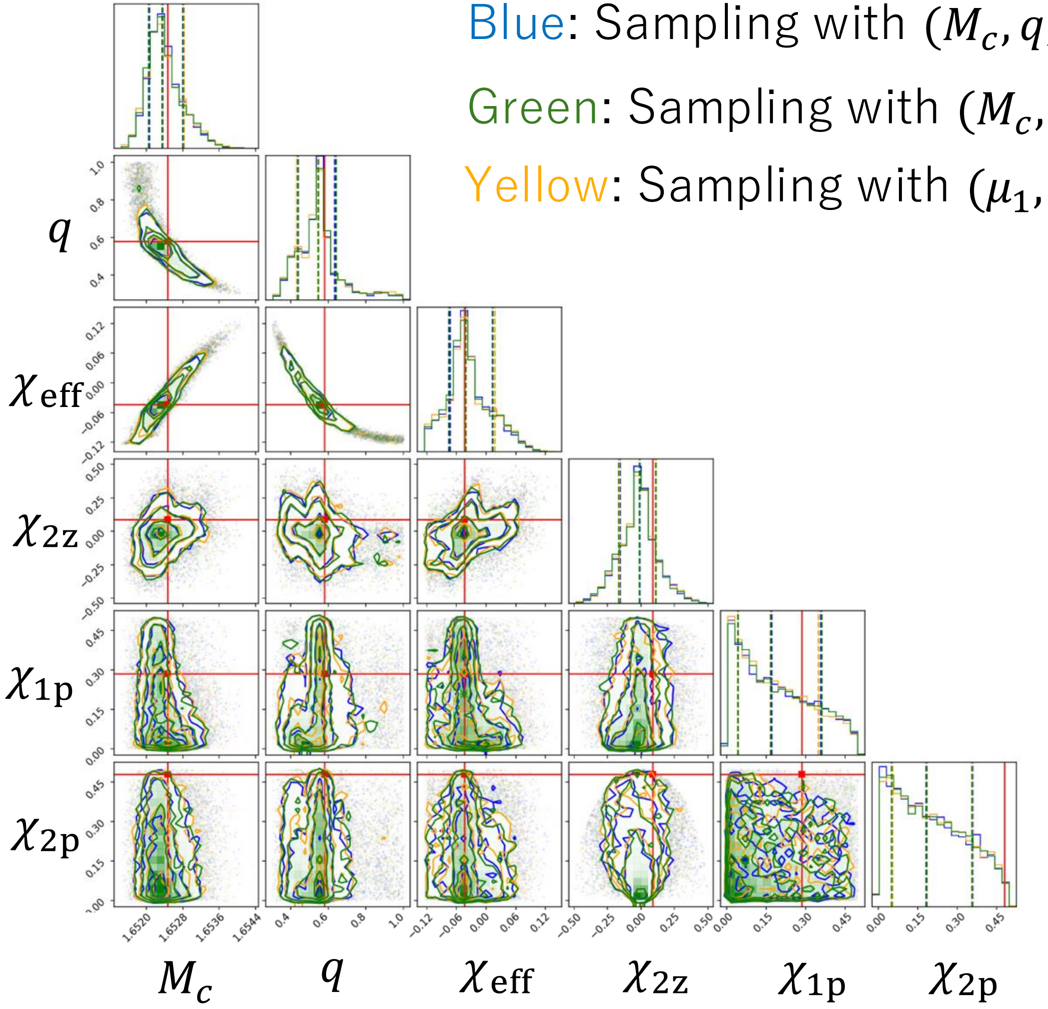
New parameters
μ_1, μ_2, q, χ_2



➡ New parameters reduces the posterior’s correlation.
It might be easier for NN to learn.

Towards Rapid Parameter Estimation: Efficient Parameterization

We found that nested sampling converges faster with $(\mu_1, q, \chi_{\mathrm{eff}}, \chi_{2z}, \chi_{1p}, \chi_{2p})$ than with $(M_c, q, a_1, a_2, \theta_1, \theta_2)$.

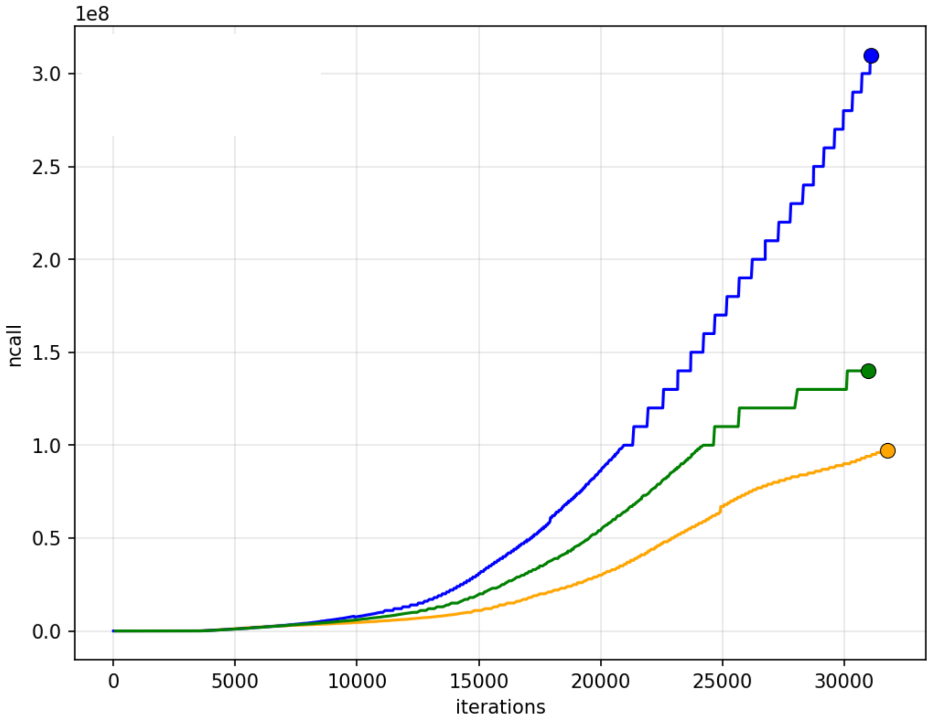


Blue: Sampling with $(M_c, q, a_1, a_2, \theta_1, \theta_2)$.

Green: Sampling with $(M_c, q, \chi_{\mathrm{eff}}, \chi_{2\mathrm{z}}, \chi_{1\mathrm{p}}, \chi_{2\mathrm{p}})$.

Yellow: Sampling with $(\mu_1, q, \chi_{\mathrm{eff}}, \chi_{2\mathrm{z}}, \chi_{1\mathrm{p}}, \chi_{2\mathrm{p}})$.

~2.7 times faster on average



Building a NN Model for BNS using Efficient Parameters

We construct a NN model learned with efficient parameters (μ_1, μ_2).

Training parameters (θ)

- μ_1, μ_2, q, χ_2: Reparameterized mass & spin
- d_L, α, δ: Sky position
- t_c, ϕ_c: Coalscence Time & Phase
- θ_{JN}, ψ: Inclination & Polarization

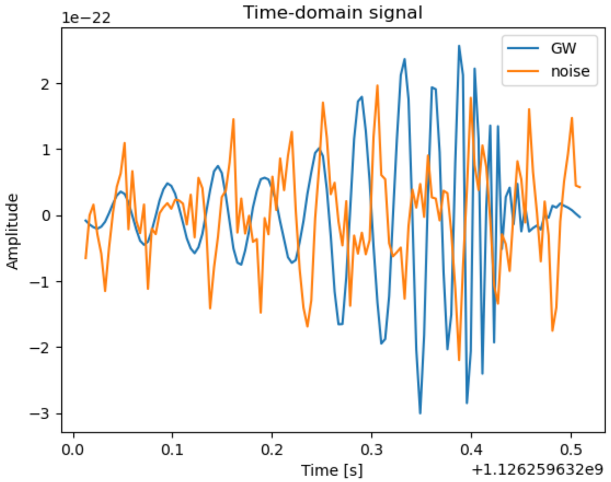
Parameters' range

$$\begin{cases} 2.0 \leq M_c \leq 2.01 \\ 0.3 \leq q \leq 1.0 \\ -0.05 \leq \chi_{1,2} \leq 0.05 \\ 50 \leq d_L \leq 150 \end{cases}$$

We prepared 10 million training data sets (θ, d).

Simulated data (d)

- $d(f) = h(f) + n(f)$
- Assume gaussian noise, 1 detector, aligned spin
- Data compression using relative binning.

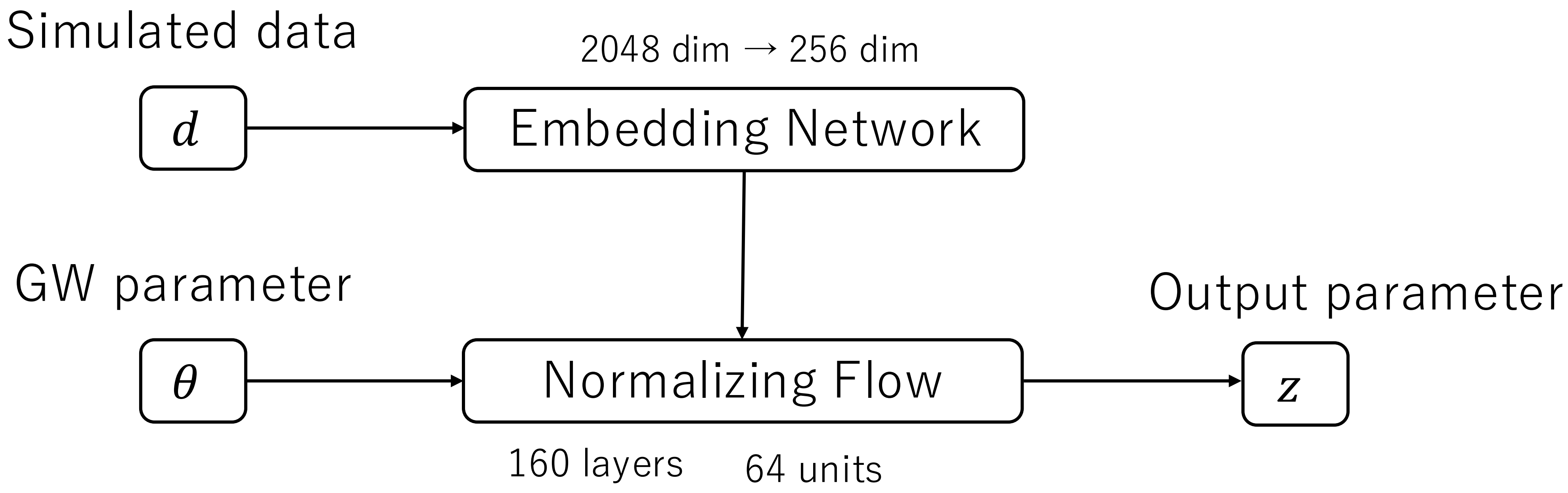


: Data example (Time domain)

Blue: GW

Orange: noise

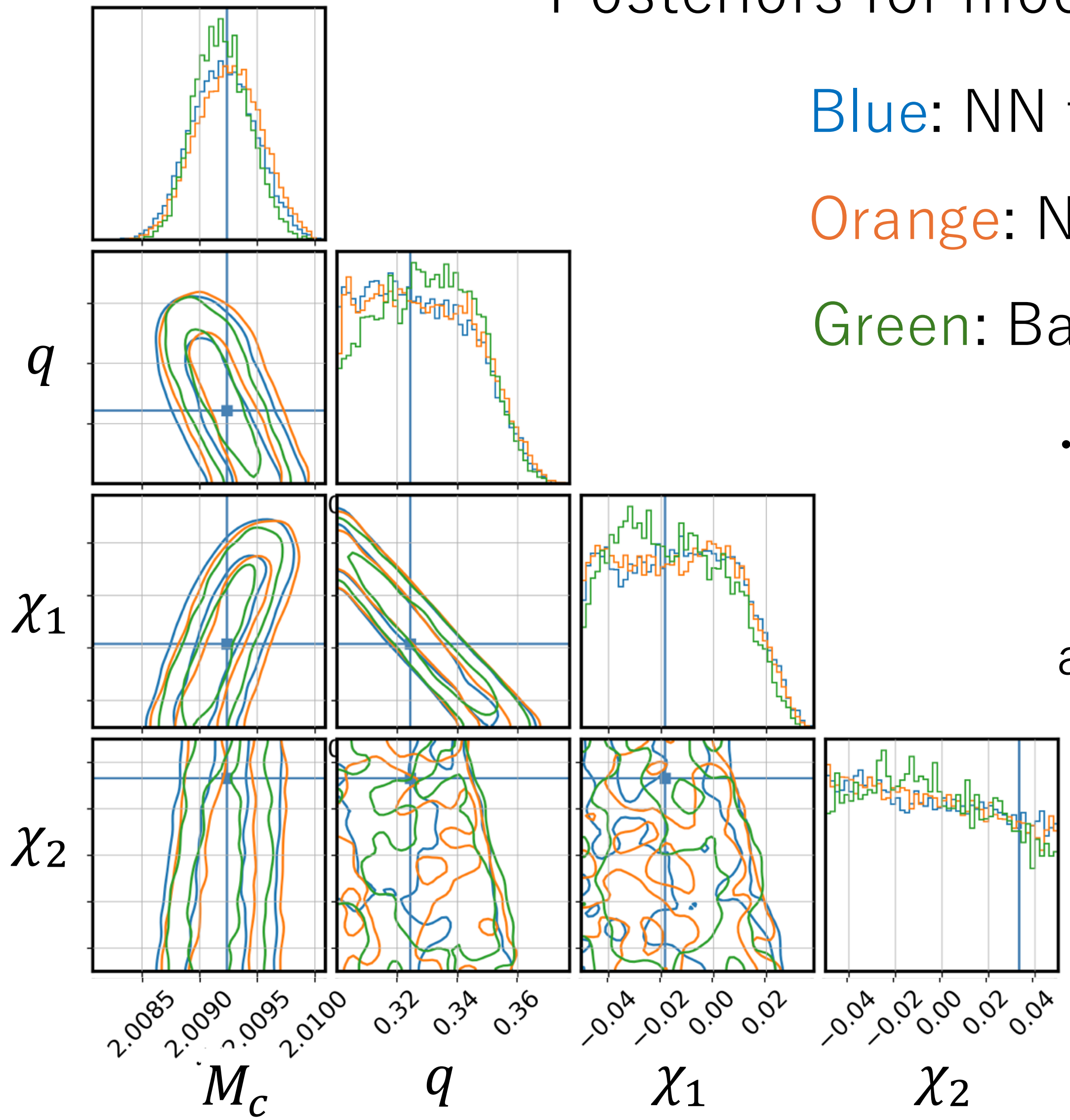
Set up for Neural Network



- The NN has ~28 million trainable parameters.
- Training takes 3-4 days on NVIDIA V100 GPU.

Results: Posterior Comparison

Posteriors for mock data with SNR~15



Blue: NN trained with new params.

Orange: NN trained with usual params.

Green: Bayesian Inference

Line: Injection value

- Both NNs' results broadly agree with Bayesian inference.
- Discrepancies arise because the NN learns an approximation to the true posterior.
- These discrepancies can be reduced by reweighting.

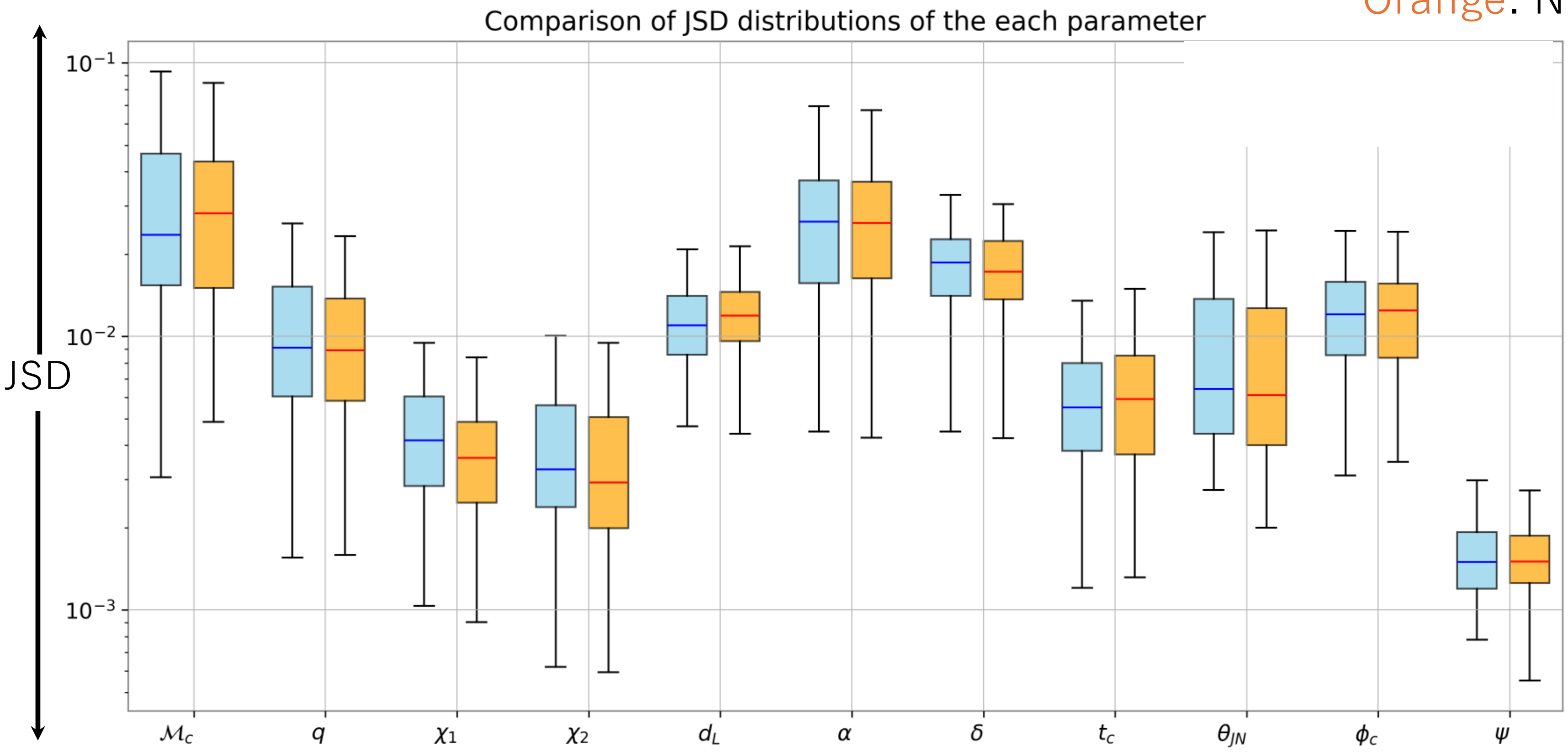
Results: Performance Comparison

Compared the NN and Bayesian posteriors for 100 simulated events using the Jensen–Shannon divergence (JSD[nat]).

SNR: 20~30

Blue: NN trained with new params

Orange: NN trained with normal params



- Both models show similar JSD distributions.

→ No significant difference in inference accuracy is observed.

- Both models take 2-3 seconds for inference.

Summary & Future Work ①

We trained a normalizing-flow model using efficient waveform parameters and evaluated its performance on BNS gravitational-wave data.

- Inference takes only ~2–3 seconds, much faster than Bayesian inference.
- For simulated events with relatively high SNR ($\gtrsim 15$), the NN produces unbiased posteriors consistent with Bayesian inference.

However ⋯,

- No clear difference in inference accuracy was found between the reparameterized and standard models.

・This may be because the NN is sufficiently expressive, reducing the impact of efficient parameterization.

・Following Roulet et al. (2026), we are introducing a new scaling method based on the mean and variance of the data.

Summary & Future Work ②

We also developing efficient parameterizations for nested sampling.

- Sampling directly with $(\mu_1, q, \chi_{eff}, \chi_2, \chi_{1p}, \chi_{2p})$ performs better.
  - μ_1 is less correlated with q.
  - χ_{eff} is well measured than usual spin parameters $(a_1, a_2, \theta_1, \theta_2)$
  - On average, ~2.7 times faster than usual sampling with $(M_c, q,\ a_1, a_2, \theta_1, \theta_2)$.

- We are also working to implement efficient extrinsic parameters $\left(t_d, \widehat{\phi_c},\ \widehat{d_L},\ \theta_{net}, \phi_{net}, \widehat{\theta_{JN}}, \hat{\psi}\right)$ following Roulet et al., PRD 106, 123015 (2022).

- We are considering implementing them in Bilby.

Appendix

Waveform

Frequency-domain inspiral waveform up to 3.5PN order (non-precessing)

$$\tilde{h}(f) = A(f;\theta_x)e^{i\Psi(f;\lambda_i)} \qquad x = \frac{f}{f_0}$$

L. Blanchet, 2013

$$\Psi = 2\pi f_0 x t_c - \phi_c + \lambda_0 x^{-5/3} + \lambda_2 x^{-1} + \lambda_3 x^{-2/3} + \lambda_4 x^{-1/3} + \lambda_{5L}\log(x) + \lambda_6 x^{1/3} + \lambda_{6L}\log(x)x^{1/3} + \lambda_7 x^{2/3},$$

$$\lambda_0 = \frac{3}{128}(\pi\mathcal{M}f_0)^{-5/3},$$

$$\lambda_2 = \frac{5}{96\eta^{2/5}}\left(\frac{743}{336} + \frac{11}{4}\eta\right)(\pi\mathcal{M}f_0)^{-1},$$

$$\lambda_3 = -\frac{3\pi}{8\eta^{3/5}}\left(1 - \frac{1}{4\pi}\beta\right)(\pi\mathcal{M}f_0)^{-2/3},$$

$$\lambda_4 = \frac{15}{64\eta^{4/5}}\left(\frac{3058673}{1016064} + \frac{5429}{1008}\eta + \frac{617}{144}\eta^2 - \sigma\right)(\pi\mathcal{M}f_0)^{-1/3}$$

$$\lambda_{5L} = \frac{3}{128\eta}\left(\frac{38645\pi}{756} - \frac{65\pi}{9}\eta\right)$$

$$\lambda_6 = \frac{3}{128\eta^{6/5}}\left(\frac{11583231236531}{4694215680} - \frac{640\pi^2}{3} - \frac{6848}{21}\left(\gamma_E + \log 4 - \frac{1}{5}\log\eta + \frac{1}{3}\log(\pi\mathcal{M}f_0)\right) - \frac{15737765635}{3048192}\eta + \frac{2255\pi^2}{12}\eta + \frac{76055}{1728}\eta^2 - \frac{127825}{1296}\eta^3\right)(\pi\mathcal{M}f_0)^{1/3}$$

$$\lambda_{6L} = -\frac{1}{128\eta^{6/5}}\frac{6848}{21}(\pi\mathcal{M}f_0)^{1/3}$$

$$\lambda_7 = \frac{3}{128\eta^{7/5}}\left(\frac{77096675\pi}{254016} + \frac{378515\pi}{1512}\eta - \frac{74045\pi}{756}\eta^2\right)(\pi\mathcal{M}f_0)^{2/3},$$

The coefficients are determined by the chirp mass Mc, the mass ratio q, and the z-components of the component spins, χ_1 & χ_2.

0PN: M_c

1PN: M_c, q

1.5PN: M_c, q, χ_1, χ_2

$$\eta = \frac{q}{(1+q)^2}$$

$$\beta = \frac{1}{12}\sum_{i=1}^{2}\left[113\left(\frac{m_i}{m_1+m_2}\right)^2 + 75\eta\right]\chi_i,$$

Definition of μ_1, μ_2

Inspiral waveform up to 1.5PN order

$$\tilde{h}(f) = \mathcal{A}\left(\frac{f}{f_{\rm ref}}\right)^{-\frac{7}{6}} e^{-i\Psi(f)},$$

$$\Psi(f) = \psi_1\left(\frac{f}{f_{\rm ref}}\right)^{-\frac{5}{3}} + \psi_2\left(\frac{f}{f_{\rm ref}}\right)^{-1} + \psi_3\left(\frac{f}{f_{\rm ref}}\right)^{-\frac{2}{3}} + \psi_4 + \psi_5\left(\frac{f}{f_{\rm ref}}\right)$$

$$\psi_1 = \frac{3}{128}(\pi G\mathcal{M}f_{\rm ref}/c^3)^{-\frac{5}{3}},$$

$$\psi_2 = \frac{55}{384}\left(\eta + \frac{743}{924}\right)\eta^{-\frac{2}{5}}(\pi G\mathcal{M}f_{\rm ref}/c^3)^{-1},$$

$$\psi_3 = \frac{3}{32}(\beta - 4\pi)\eta^{-\frac{3}{5}}(\pi G\mathcal{M}f_{\rm ref}/c^3)^{-\frac{2}{3}},$$

$$\psi_4 = -2\phi_c - \frac{\pi}{4},$$

$$\psi_5 = 2\pi f_{\rm ref} t_c,$$

Posterior of $\psi_1, \psi_2\ \psi_3$:

$$p(\{\psi_1, \psi_2, \psi_3\} \mid d) \;\propto\; p(\{\psi_1, \psi_2, \psi_3\}) \exp\left[-\frac{1}{2}\sum_{i,j} \underline{\tilde{\Gamma}_{ij}}\, \Delta\psi_i\, \Delta\psi_j\right]$$

Diagonalize the Fisher matrix

$$\tilde{\Gamma}_{ij} = \sum_{m,n} U^T_{im}\, \lambda_m\, \delta_{mn}\, U_{nj}$$

$$\lambda_1 > \lambda_2 > \lambda_3 \geq 0$$

$$\underline{\mu_n \equiv \sum_i U_{ni}\, \psi_i,} \quad (n = 1, 2, 3)$$

Definition

Posterior of $\mu_1, \mu_2\ \mu_3$:

$$p(\mu \mid d) \propto p(\mu) \prod_n e^{-\frac{1}{2}\lambda_n \Delta\mu_n^2}$$

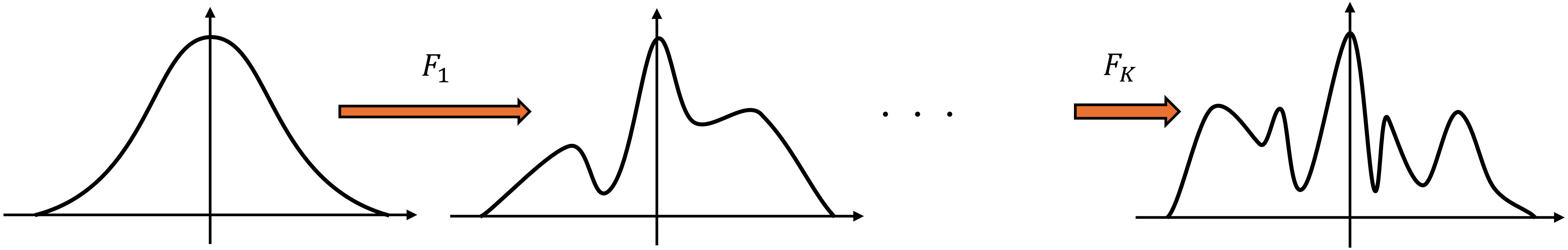
➡ Roughly Gaussian distribution

Normalizing Flow

Ref. Papamakarios et al., JMLR 2021

・Transform samples from a normal distribution to represent complex target distributions.

・Model the transformation functions (F_i) using neural networks.



$u \sim \mathcal{N}(u|0, I)$ $\theta_1 \sim q_1(\theta_1|d)$ $\theta_K \sim q_K(\theta_K\,|d)$

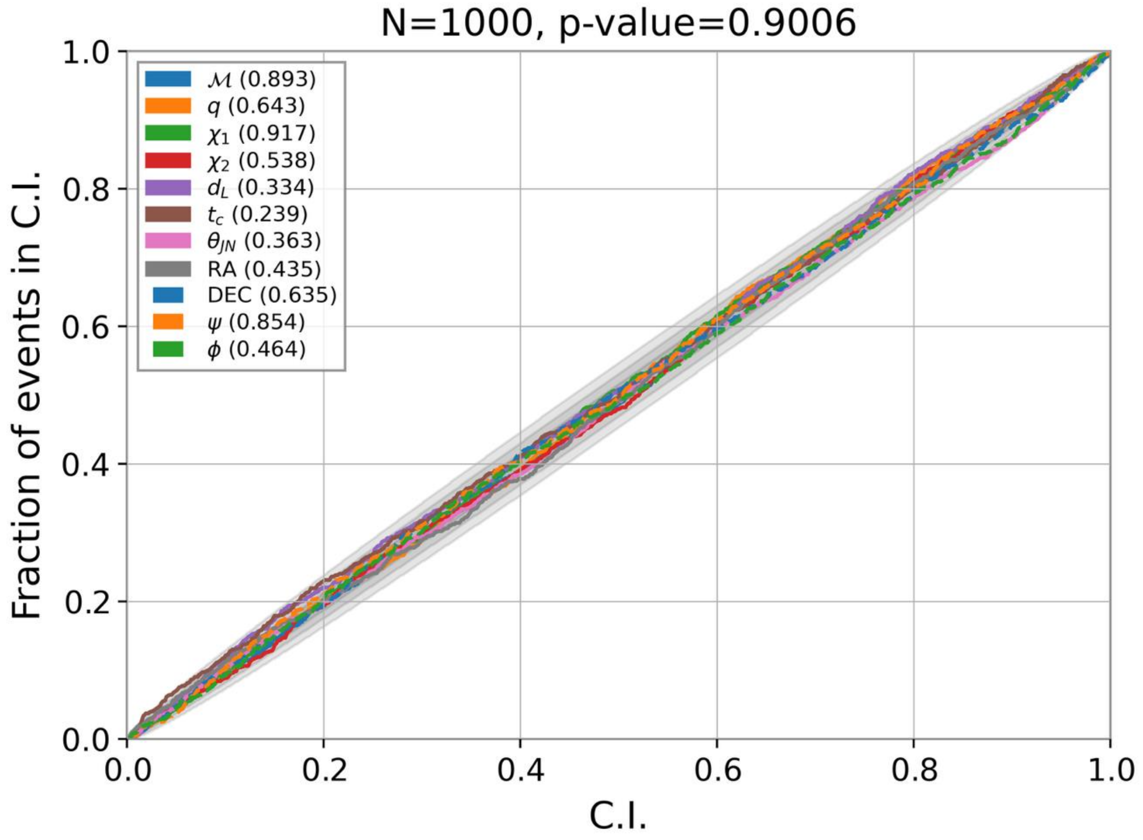
Transform: $\theta_K = F_K \circ \cdots \circ F_2 \circ F_1(u)$

Density: $q_K(\theta_K|d) = q_0(u) \cdot \prod_{i=1}^{K} \left| det\left(\frac{\partial F_i^{-1}(\theta_i)}{\partial \theta_i}\right) \right|$

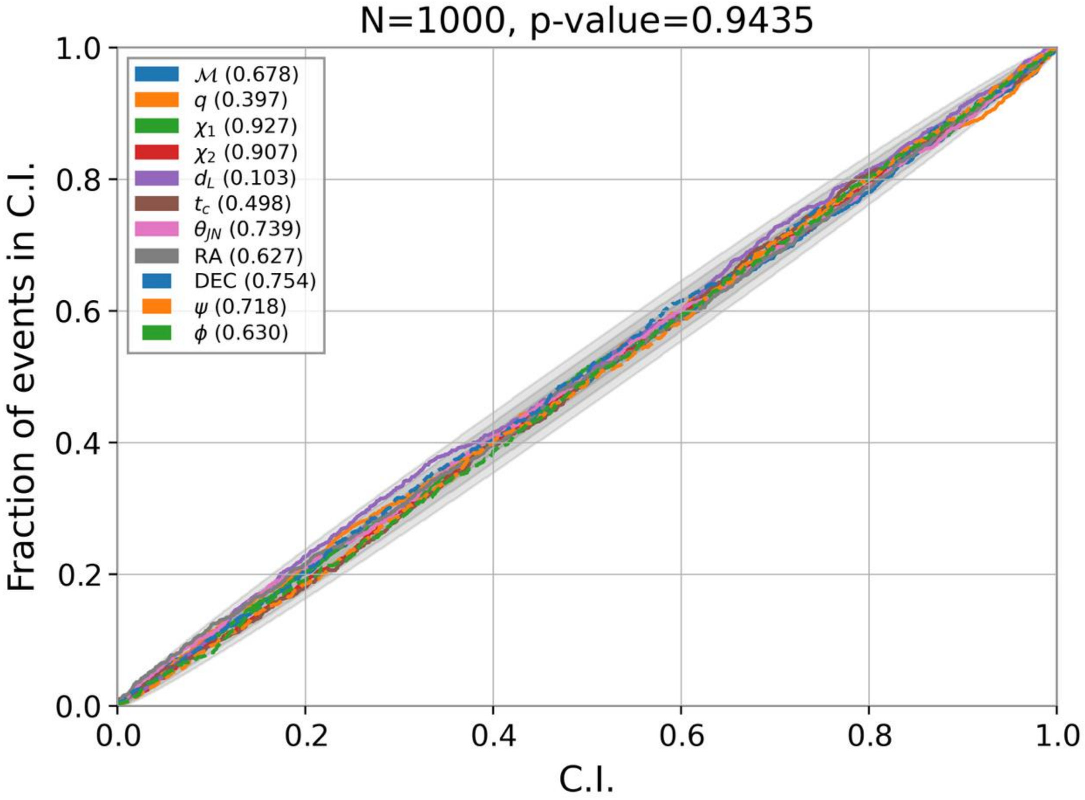
Evaluation as a Posterior Estimator

Applied the NN model to 1,000 simulated events and plotted the fraction of posteriors containing the true values (P–P plot). $(20 \leq \mathrm{SNR} \leq 25)$

NN model with new params



NN model with usual params



→ No significant bias is observed.