Bayesian Parameter Estimation using Conditional Variational Autoencoders for Gravitational Wave Astronomy

Hunter Gabbard, PhD Candidate, University of Glasgow

Bayesian Deep Learning for Cosmology and Gravitational waves,
PCCP, APC laboratory, Université de Paris

March 5th, 2020

Talk Overview

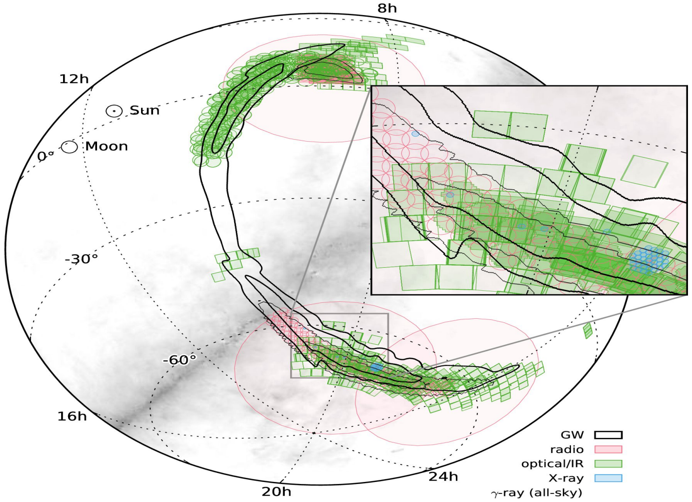
- Motivation for this work.
- What are conditional variational autoencoders?
- What did we do?
- Where do we go from here?

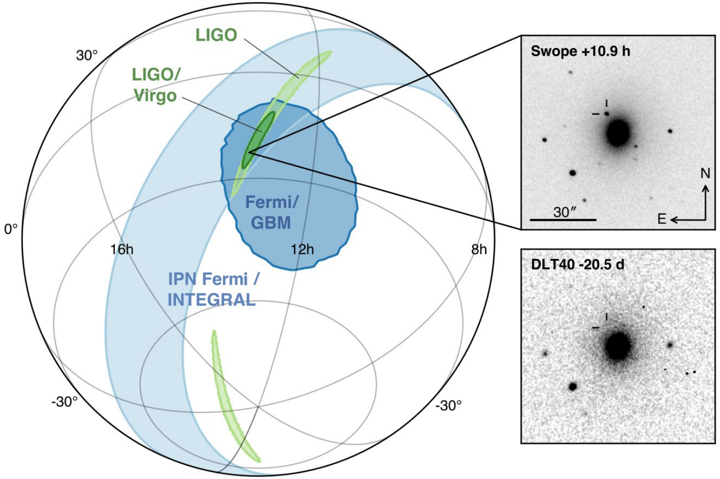
Motivation

- Existing Bayesian parameter estimation is optimal but very slow

- For transient GW events and multi-messenger astronomy it is crucial that we produce data products very quickly.

LIGO-Virgo Collaboration, ApJ 848, 2, L12, 59 (2017)





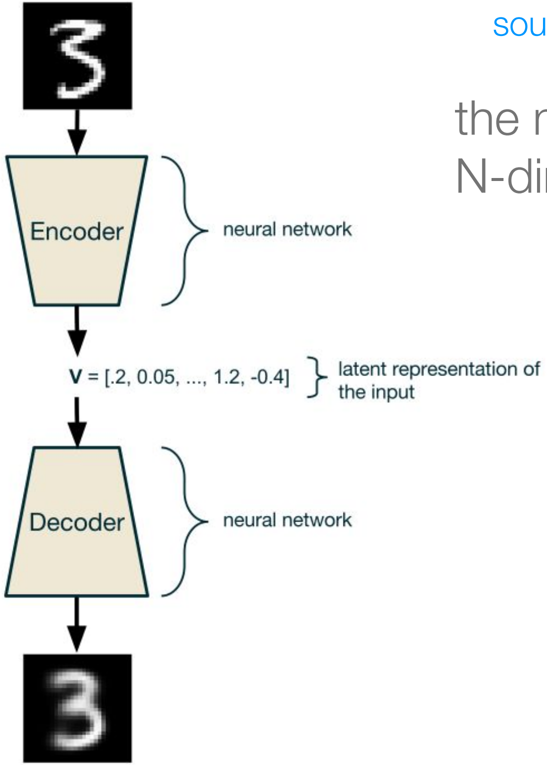
LIGO-Virgo Collaboration, ApJ Volume 826, 1, L13, 8 (2016)

Conditional Variational Autoencoders

(it’s not magic, I promise)

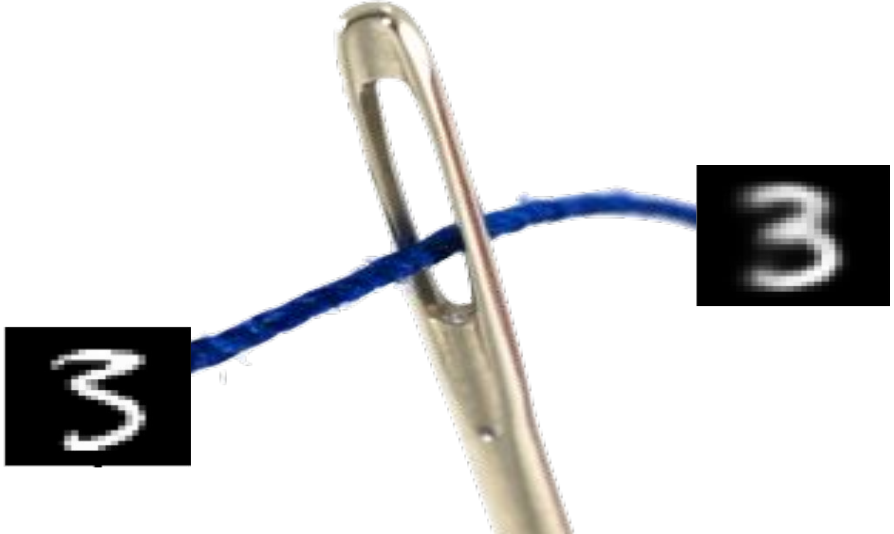
First - an autoencoder

loss based on input/output similarity

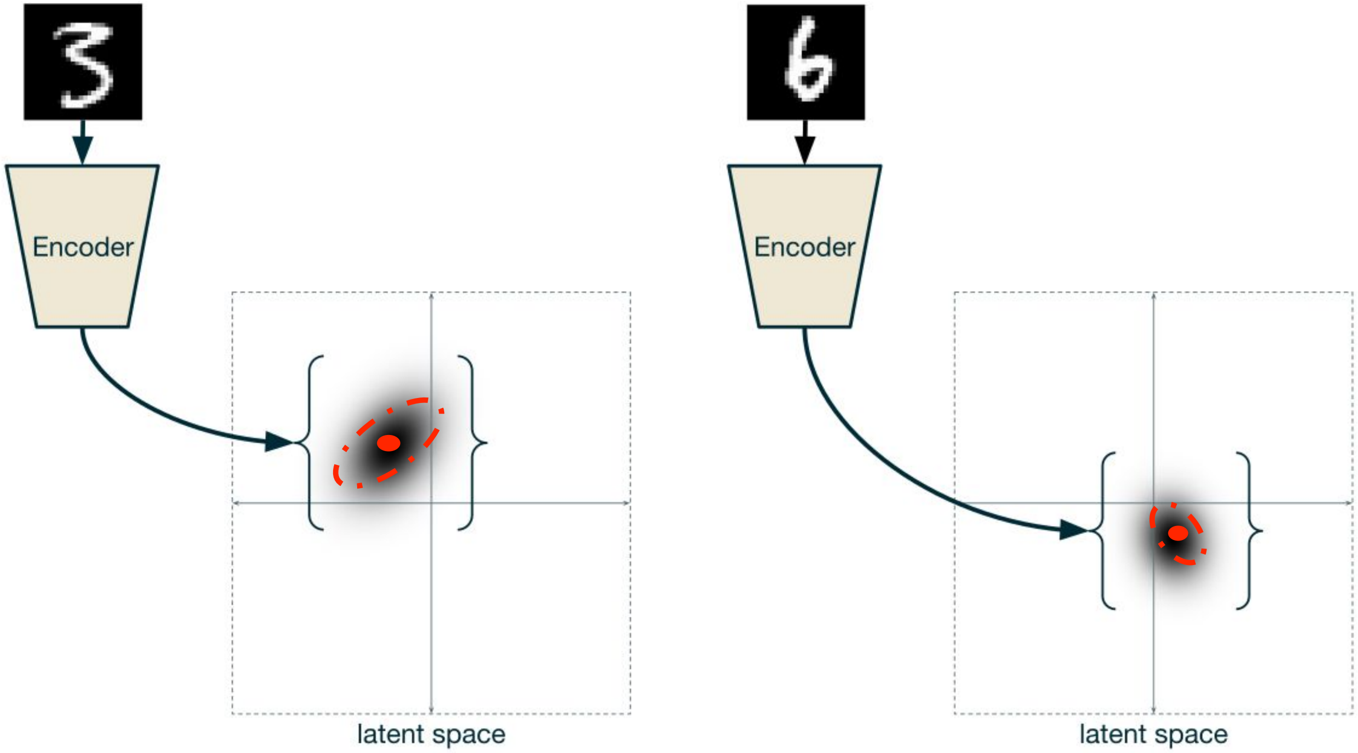


source: https://ijdykeman.github.io/ml/2016/12/21/cvae.html

the mysterious "latent" space is jargon for some N-dimensional non-physical parameter space in which to represent your data



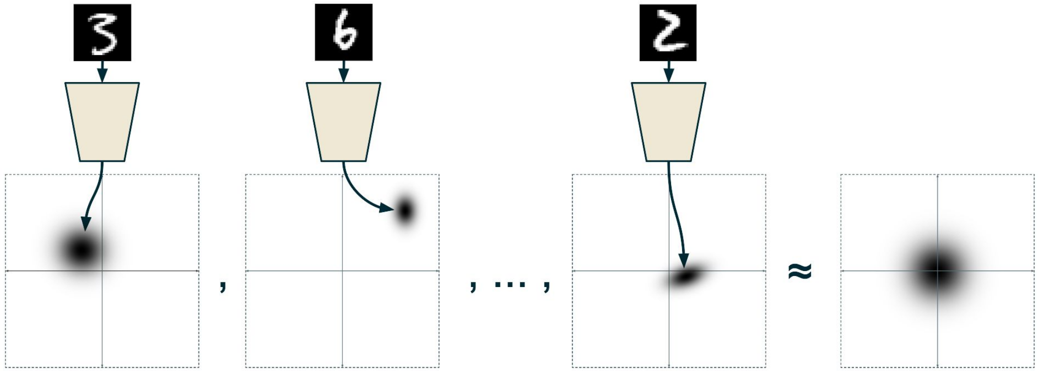
Next - a variational autoencoder

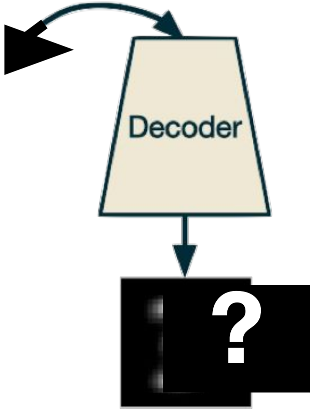


source: https://ijdykeman.github.io/ml/2016/12/21/cvae.html

Next - a variational autoencoder

The loss function incorporates a KL-divergence term testing the Gaussianity of the total distribution in the latent space

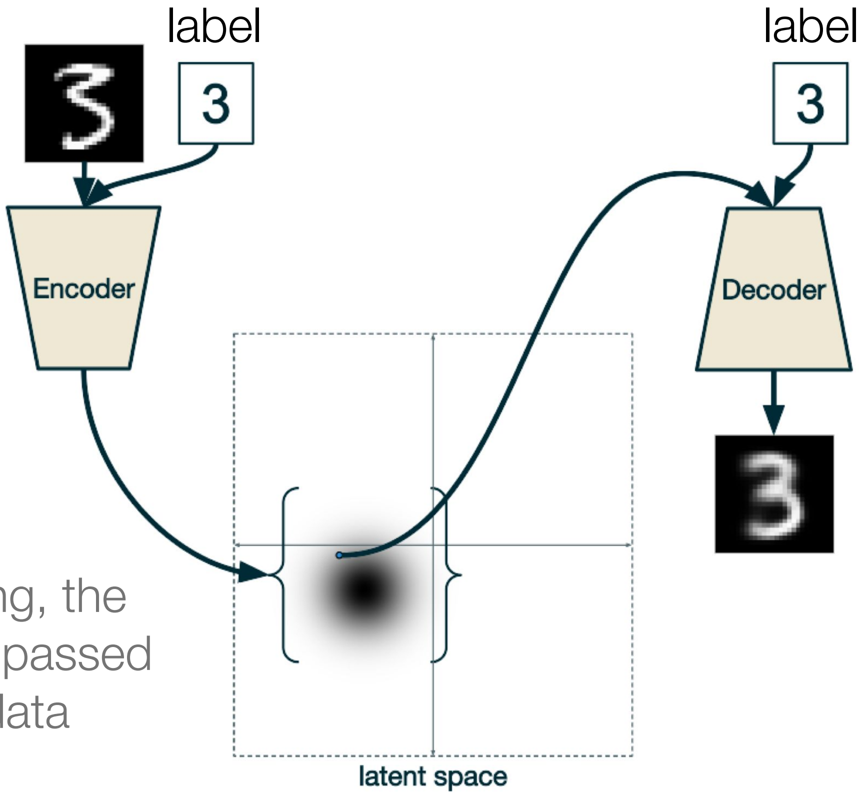




random draw
can't produce particular numbers on command

source: https://ijdykeman.github.io/ml/2016/12/21/cvae.html

Then - a conditional variational autoencoder



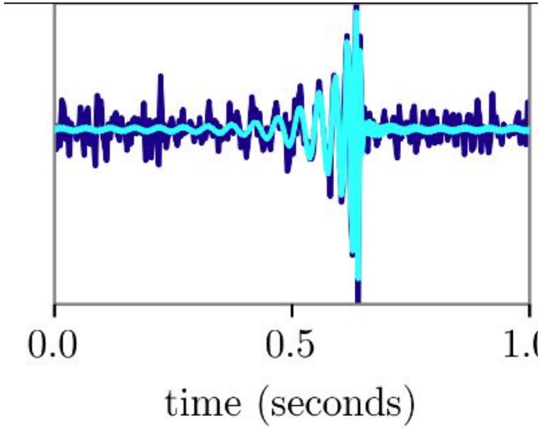
Within training, the true “label” is passed with the data

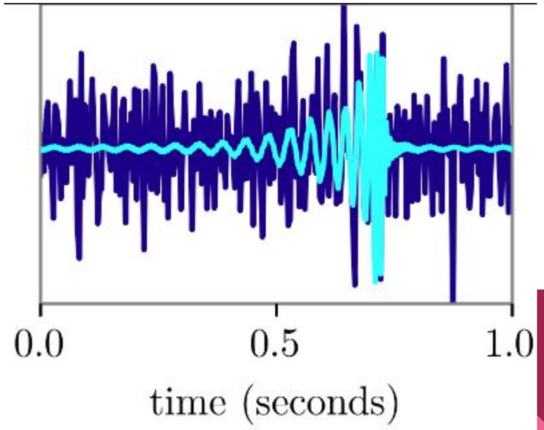
now you can ask for a random “7” or “4”, etc…

source: https://ijdykeman.github.io/ml/2016/12/21/cvae.html

Analysis overview

- CVAE trained on whitened binary black hole time series in Gaussian noise and the true parameter values
  - **(posteriors are NOT used in training)**
- 1 million training data, and 256 test samples.
- Prior ranges (uniform):
  - m_1/m_2: 35 - 80 M
  - t_0: last 65 - 85 % of time 1s window
  - Distance: 1 Gpc - 3 Gpc
  - Phase: 0 - 2pi
- Tune network.
- We produce posteriors, not point estimates.
- Compare with Bayesian inference (Bilby samplers).





Some of the components of our scheme

The GW data y — we generate lots of these

The GW parameters x

The prior $p(x)$ — from this

The likelihood $p(y|x)$ — we never have to evaluate this

The latent space parameters z

The posterior $p(x|y)$ — we want this

The recognition function $q_\phi(z|x,y)$

The target distribution $p_\theta(x|y) = \int dz\, r_{\theta_1}(z|y) r_{\theta_2}(x|z,y)$

we are going to use 3 neural networks to model each of these 3 functions

Gabbard *et al*, arXiv:1909.06296 (2019)

The scheme

Loss function:

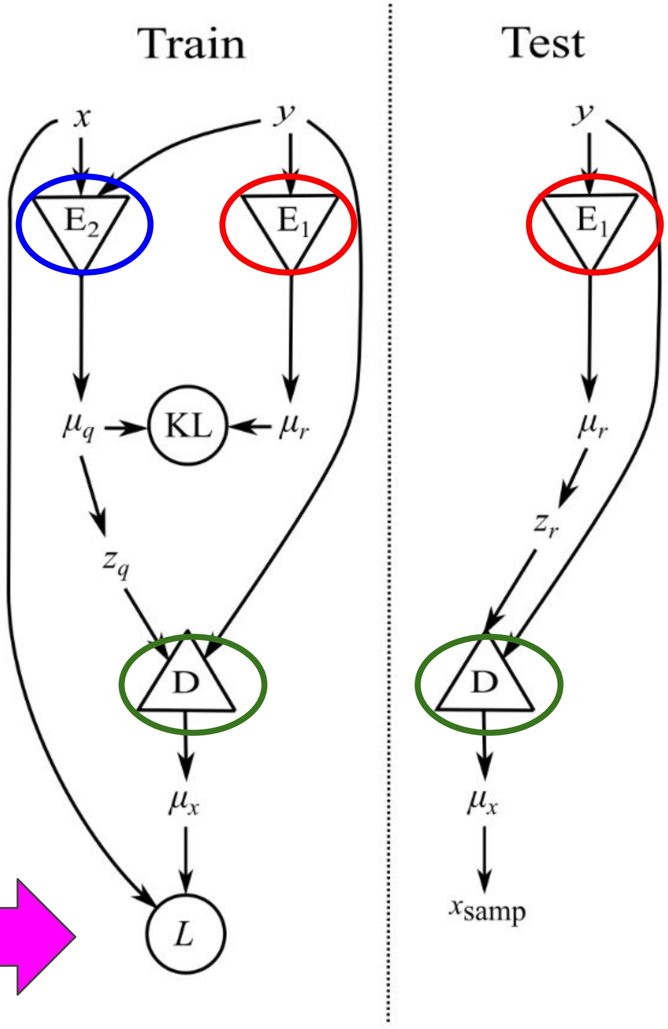
- Start of derivation

$$H(p,r) = -\int dx\, p(x|y) \log r_\theta(x|y)$$

... Here be math dragons ...

- End of derivation

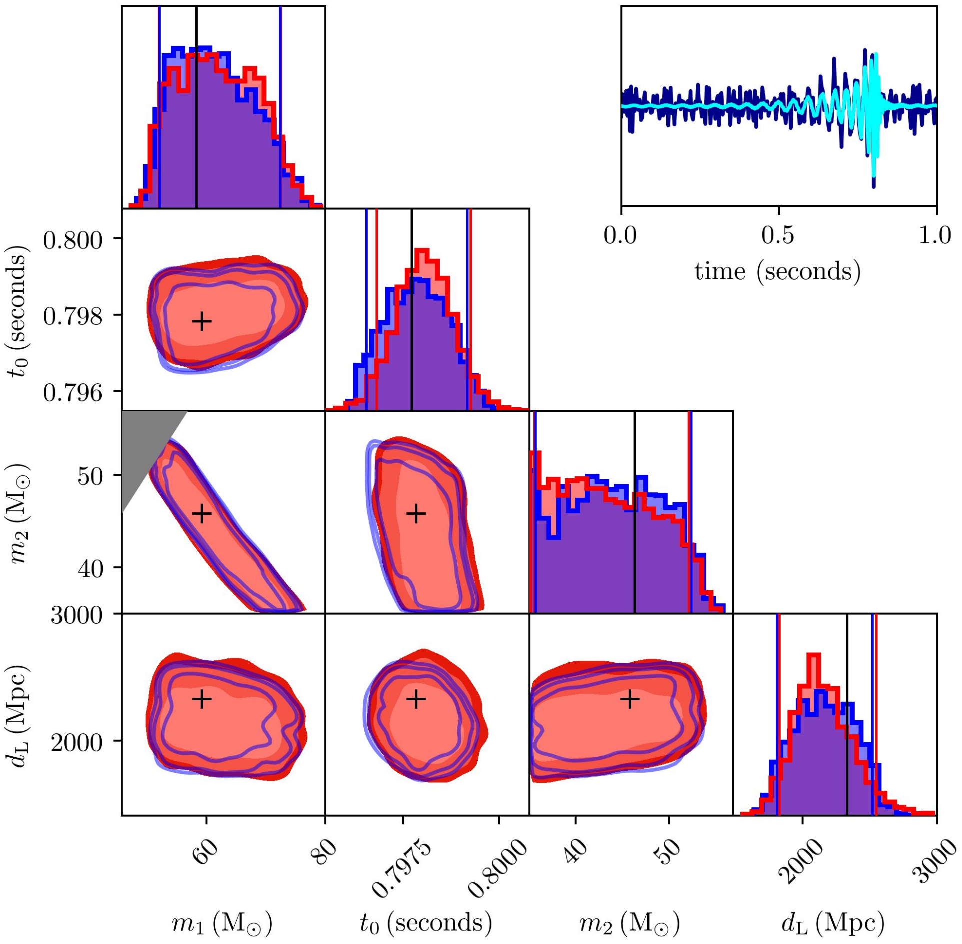
$$H \lesssim -\frac{1}{N}\sum_{n=1}^{N}\left[\log r_{\theta_2}(x_n|z_n, y_n) - \mathrm{KL}\left[q_\phi(z|x_n, y_n)\|r_{\theta_1}(z|y_n)\right]\right].$$



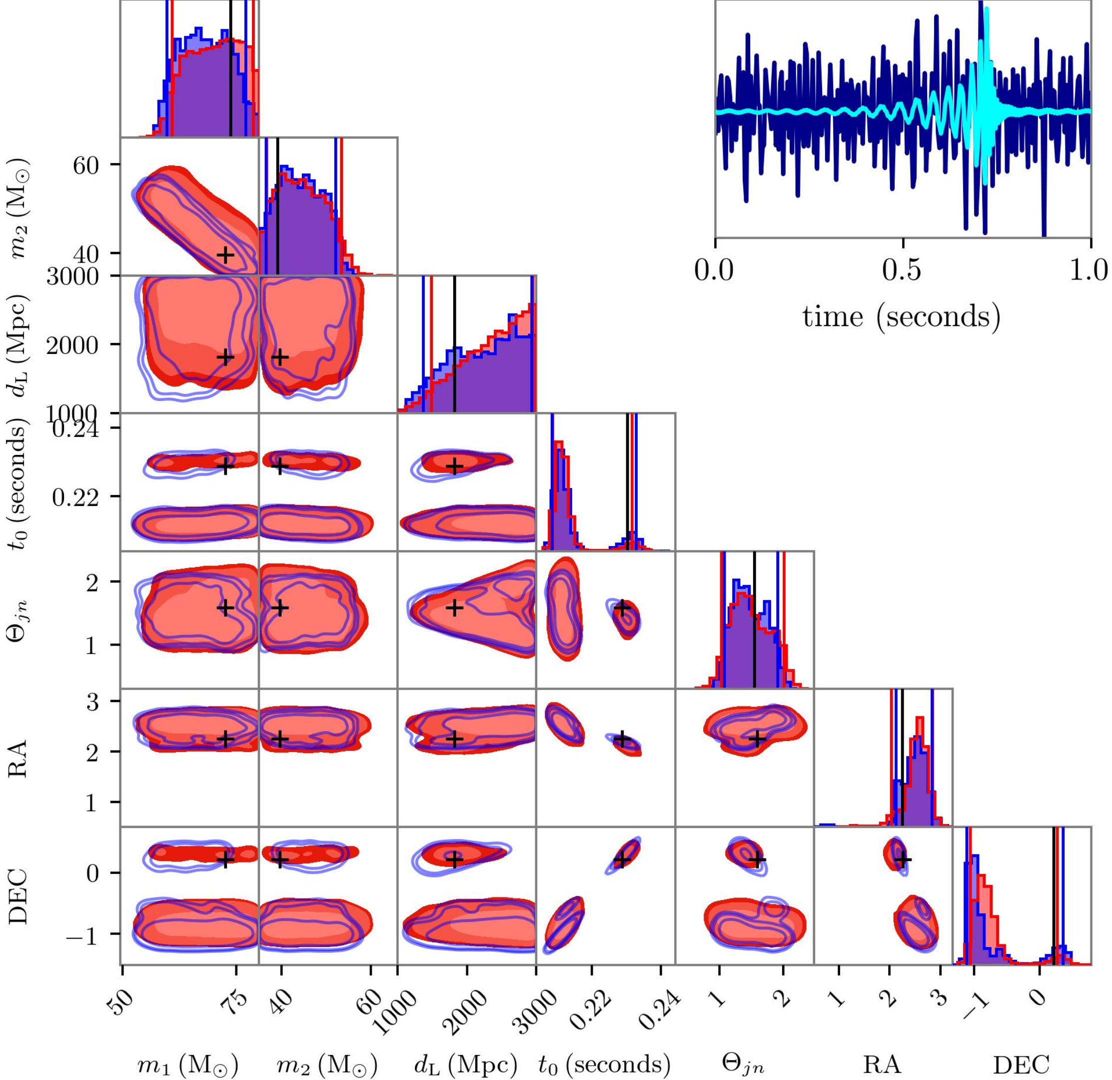
What did we find out?

(it’s fast and accurate)

Posterior comparisons



Red: Our method
Blue: Bilby Dynesty Sampler



7 parameter case

The speed

TABLE I. Durations required to produce samples from each of the different posterior sampling approaches.

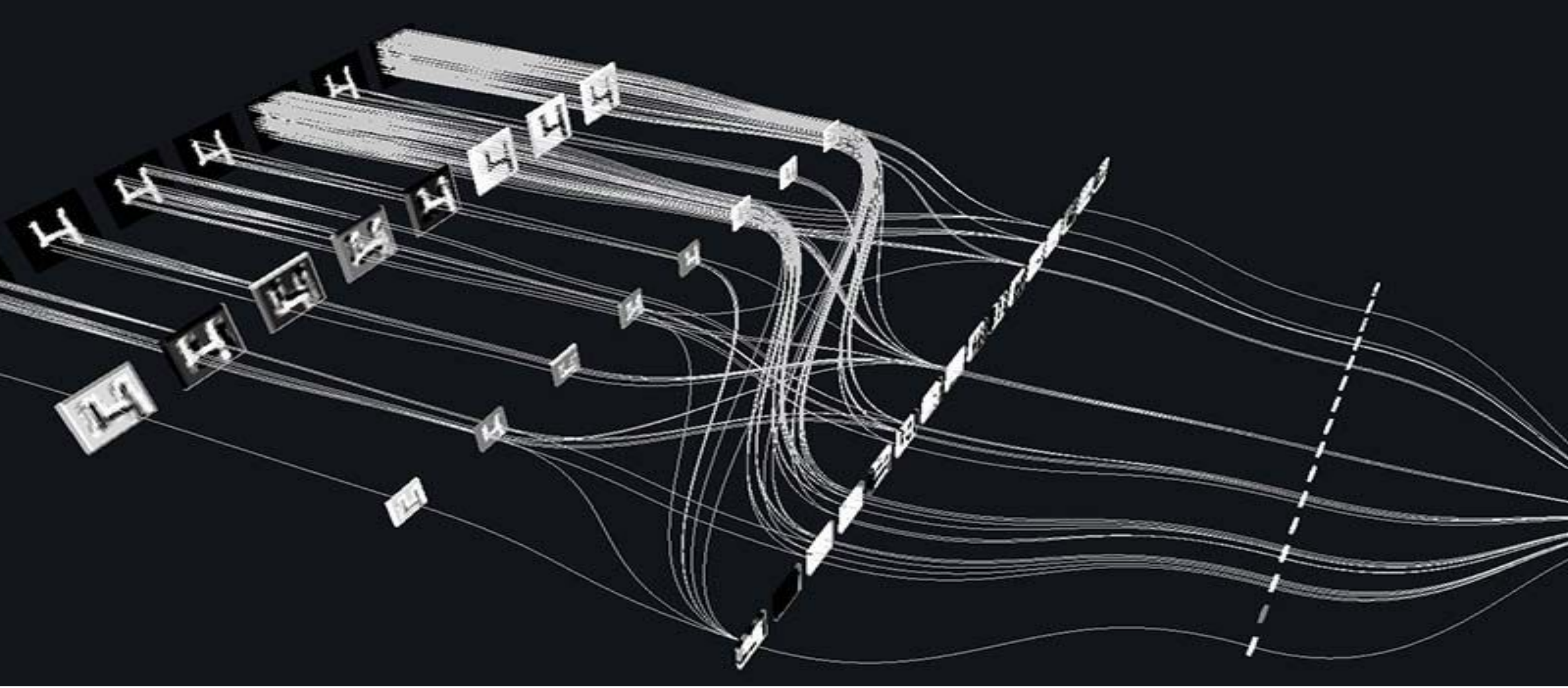
sampler	run time (seconds) min	max	median	ratio $\frac{\tau_{\text{VItamin}}}{\tau_X}$
Dynesty[a]	602	1538	774[b]	2.6×10^{-6}
Emcee	2005	11927	4351	4.6×10^{-7}
Ptemcee	3354	12771	4982	4.0×10^{-7}
Cpnest	1431	5405	2287	8.8×10^{-7}
`VItamin`[c]		$\mathbf{2 \times 10^{-3}}$		1

[a] The benchmark samplers all produced $\mathcal{O}(3000 - 10000)$ samples dependent on the default sampling parameters used.

[b] The reader may note that benchmark sampler run times are a few orders of magnitude lower than what is typical of a complete BBH analysis ($\mathcal{O}(10^5 - 10^6)$ seconds). This is primarily due our use of a reduced parameter space, low sampling rate and choice of sampler hyperparameters.

[c] For the **`VItamin`** sampler 3000 samples are produced as representative of a typical posterior. The run time is independent of the signal content in the data and is therefore constant for all test cases.

Gabbard *et al*, arXiv:1909.06296 (2019)



Conclusions

The take home message

Variational Inference - Future work

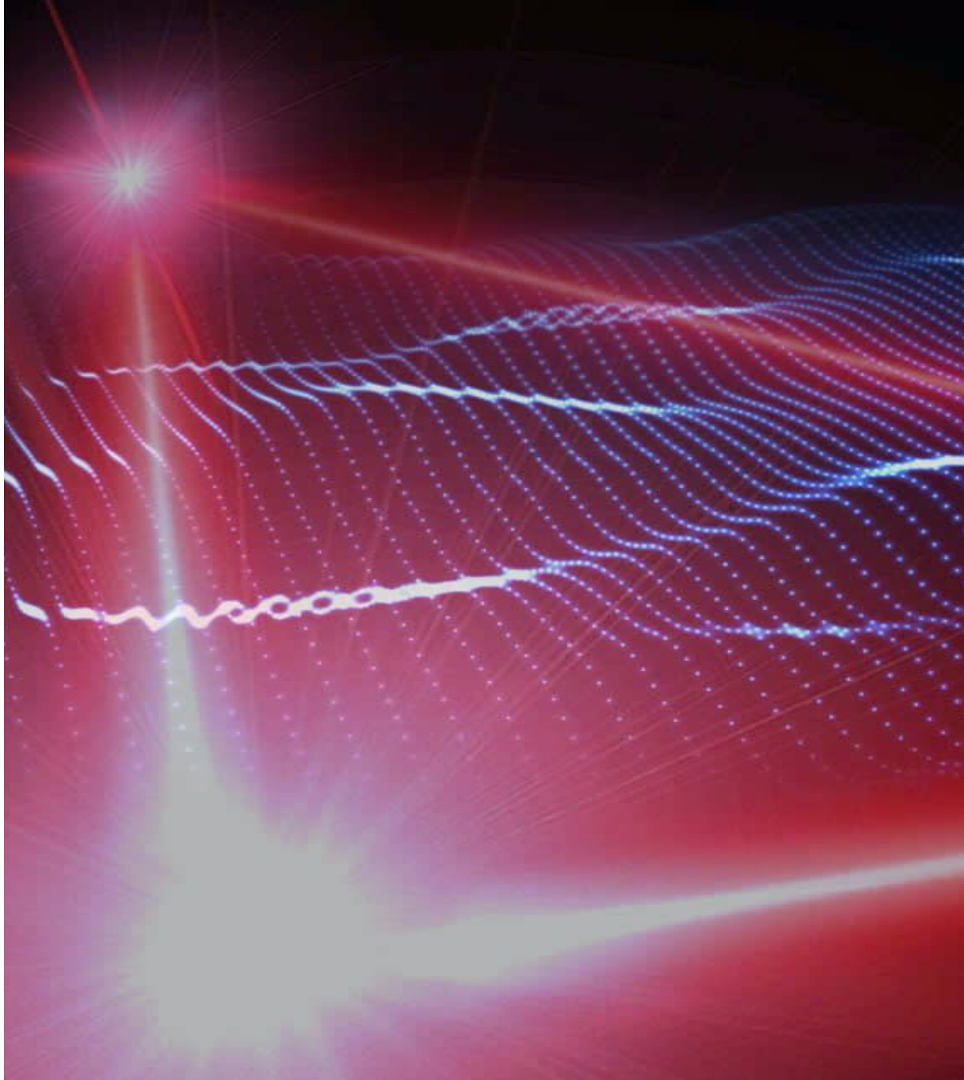
- We’ve shown (and so have other groups) [Chua & Valisneri arXiv:1909.05966 (2019), Green et al. arXiv:2002.07656 (2020)] that variational inference is a powerful tool.

- Extending this to more realistic cases is the next step.

- The ultimate aim is to have this working on binary neutron stars which emit electromagnetic radiation.

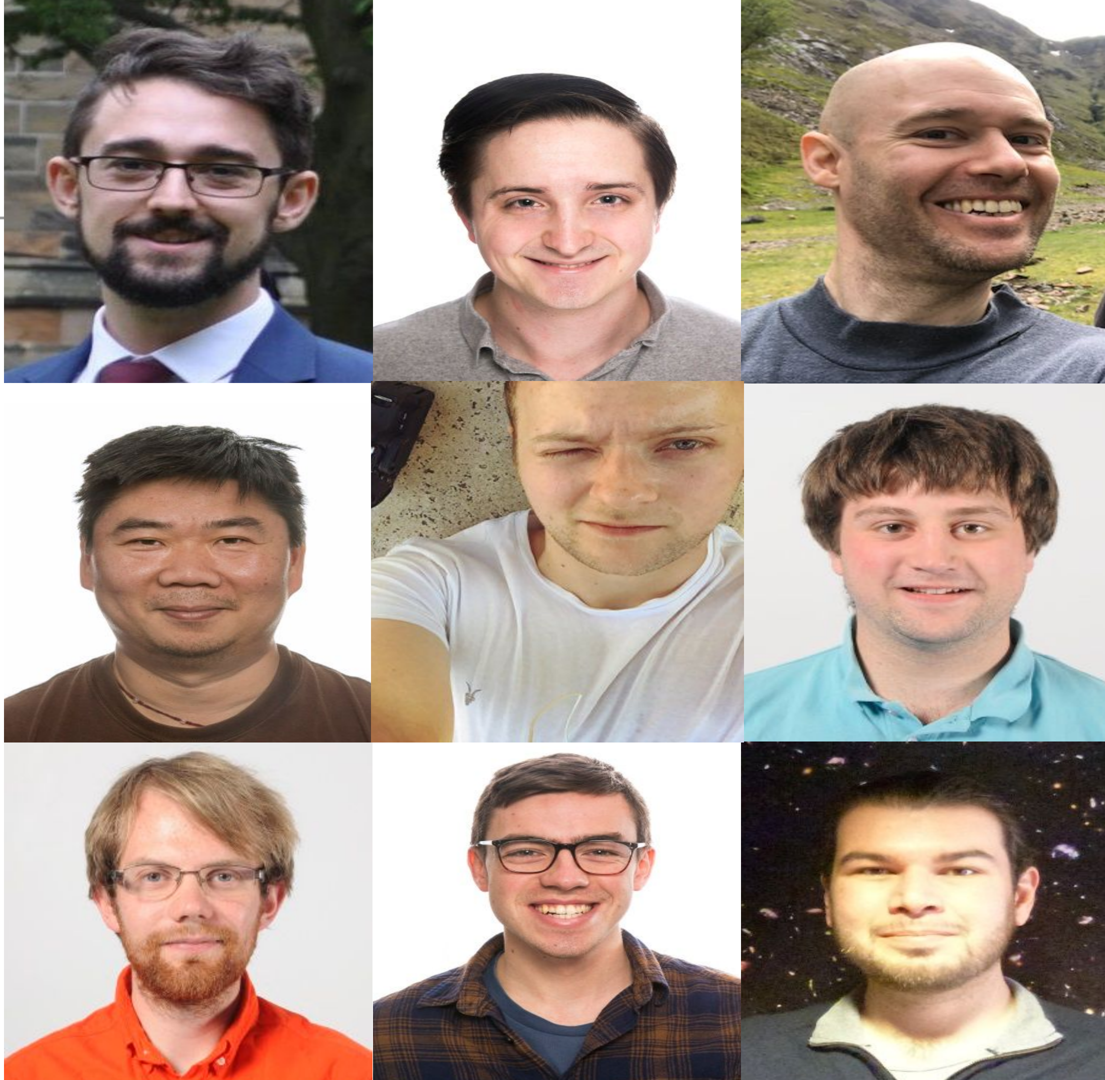
- Our pipeline is called VItamin and is available to play with here

https://github.com/hagabbar/VItamin



Summary

- We provided motivation for decreasing the latency of producing GW posteriors.
- We covered variational autoencoders.
- We finished off with variational inference for Bayesian parameter estimation.
- Paper is on arXiv and currently with referees at Nature Physics (arXiv:1909.06296)



Thank you for your attention!