Gravitational Wave Parameter Estimation Using Machine Learning

Presented by -

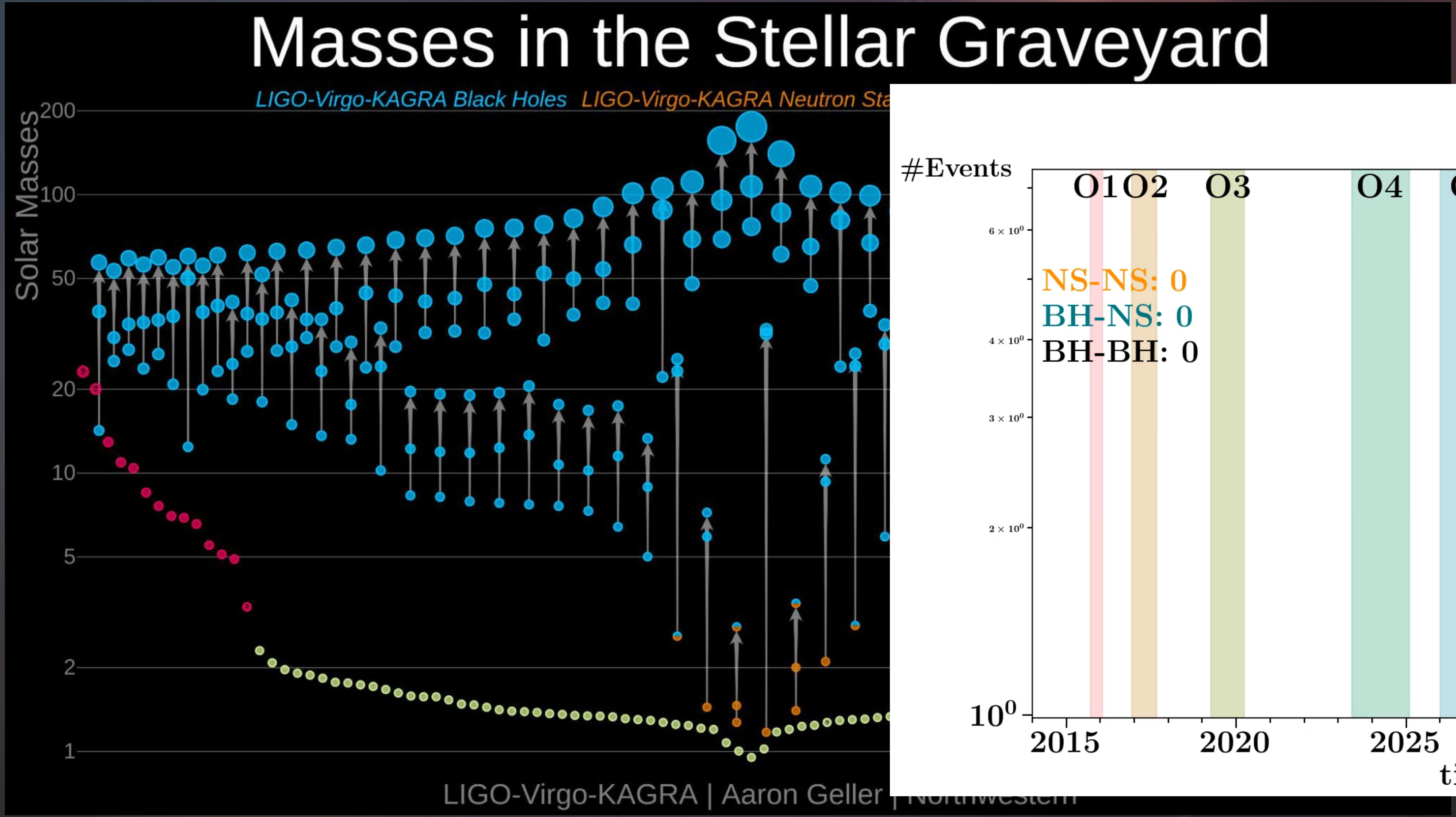
Chayan Chatterjee

OzGrav Winter School - 2023

The University of Western Australia,

20th July, 2023

Current State of GW Astronomy



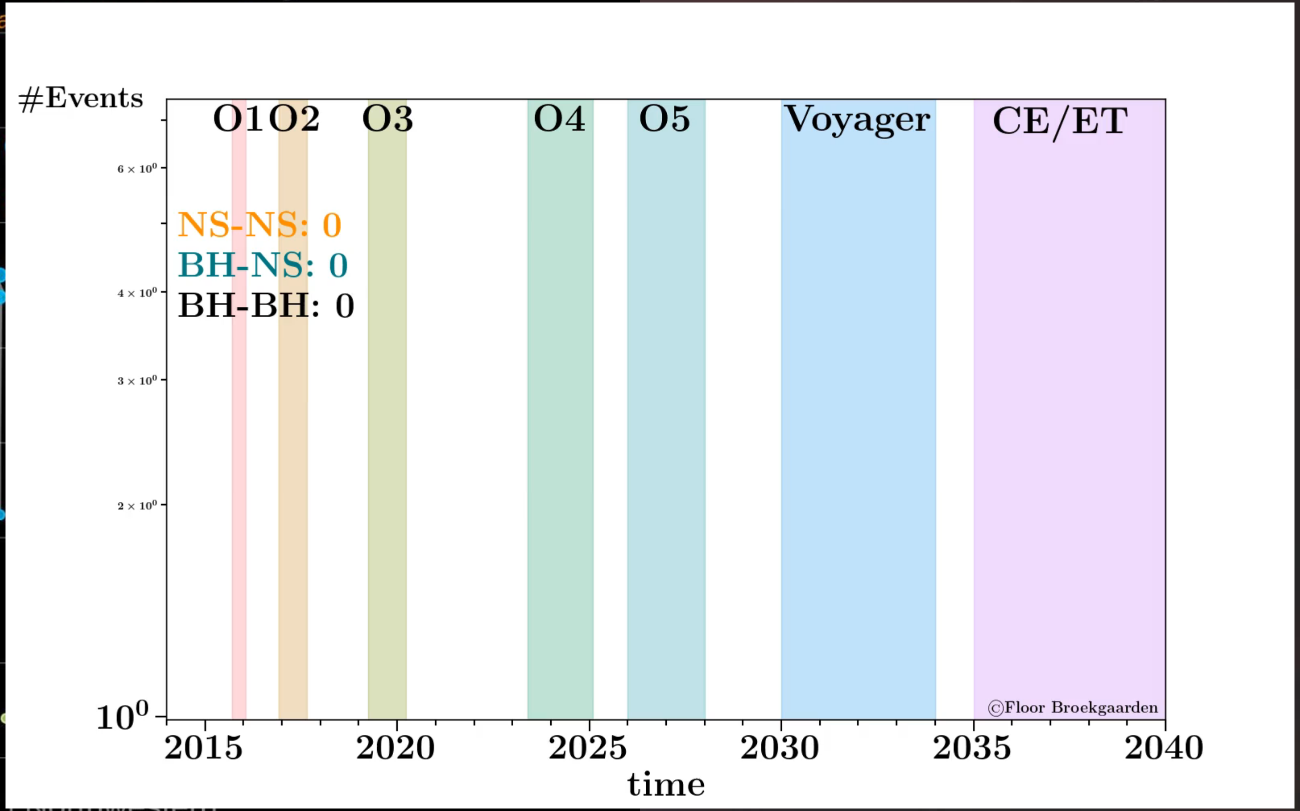
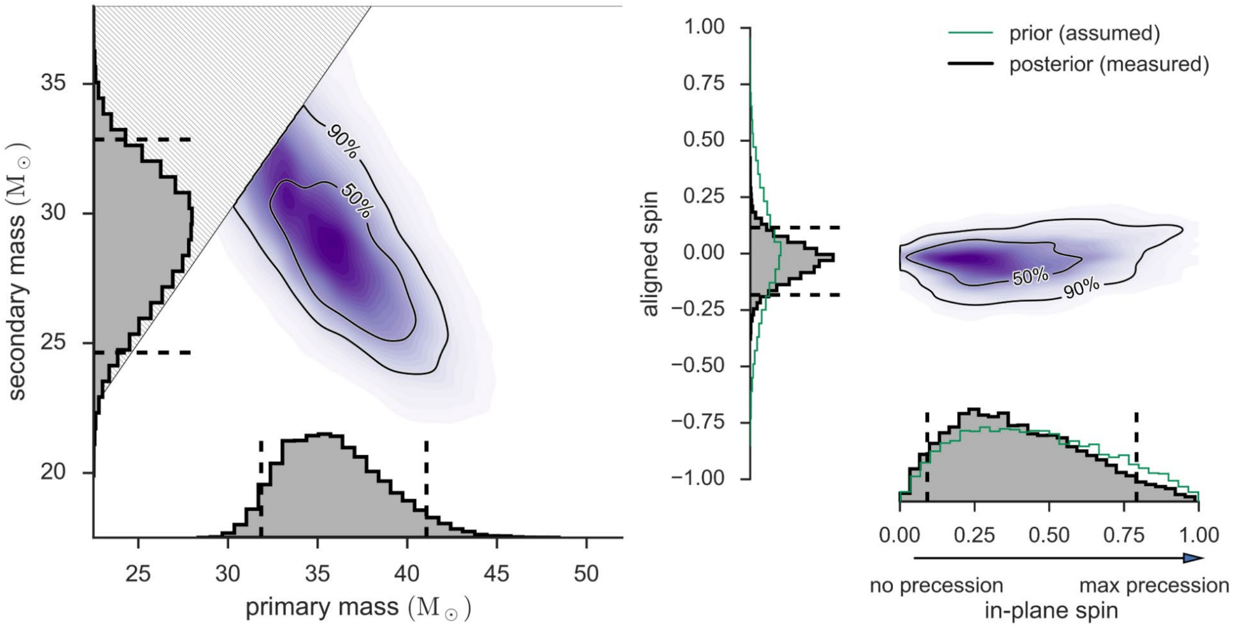


Illustration of observed black holes and neutron stars from gravitational and electromagnetic waves

Number of detected mergers expected from binary black holes, black hole-neutron stars, and binary neutron stars. Courtesy – Dr. Floor Broekgaarden

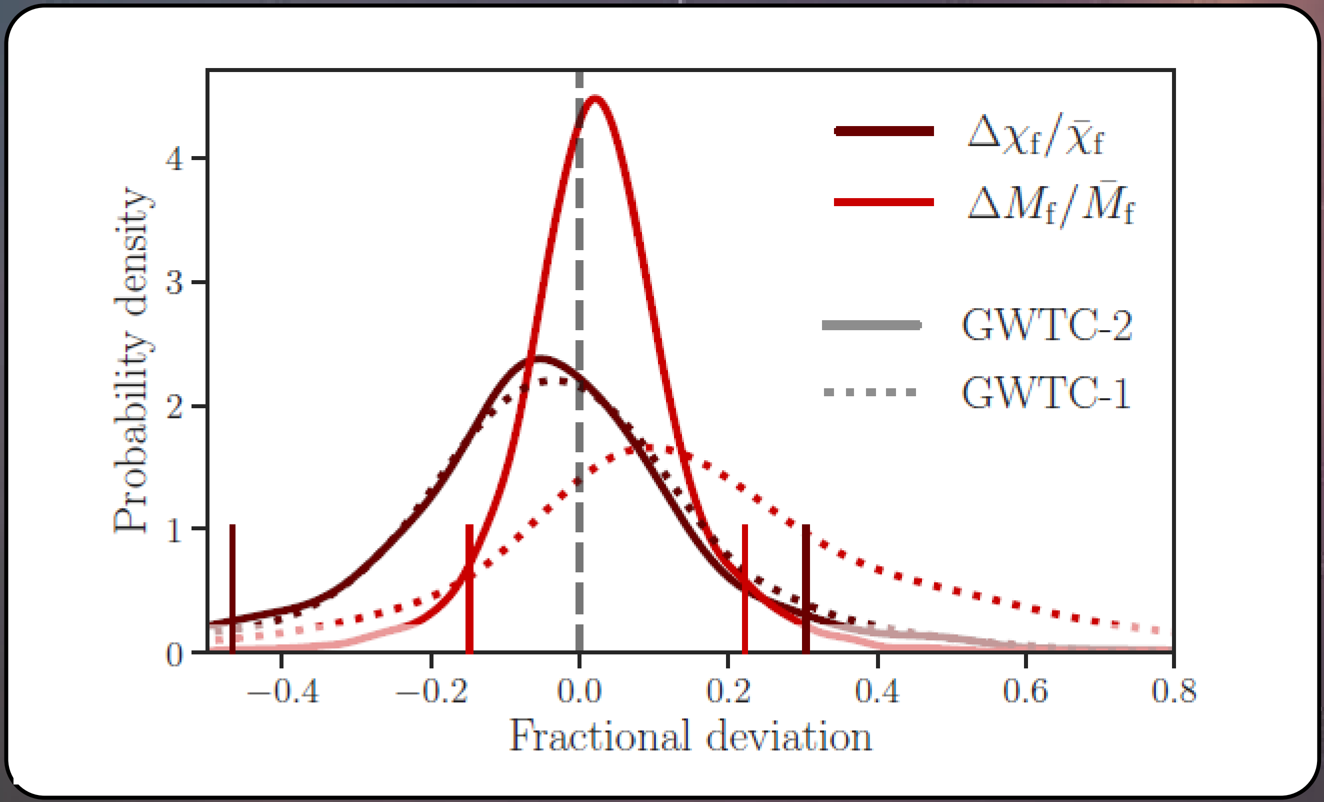
Beyond Detection



Ref: Abbott *et al.* (2016)

- Study single event source properties:
  - Component masses (m_1 , m_2).
  - Spins
  - Sky directions (ra, dec).
  - Luminosity distance.

Beyond Detection

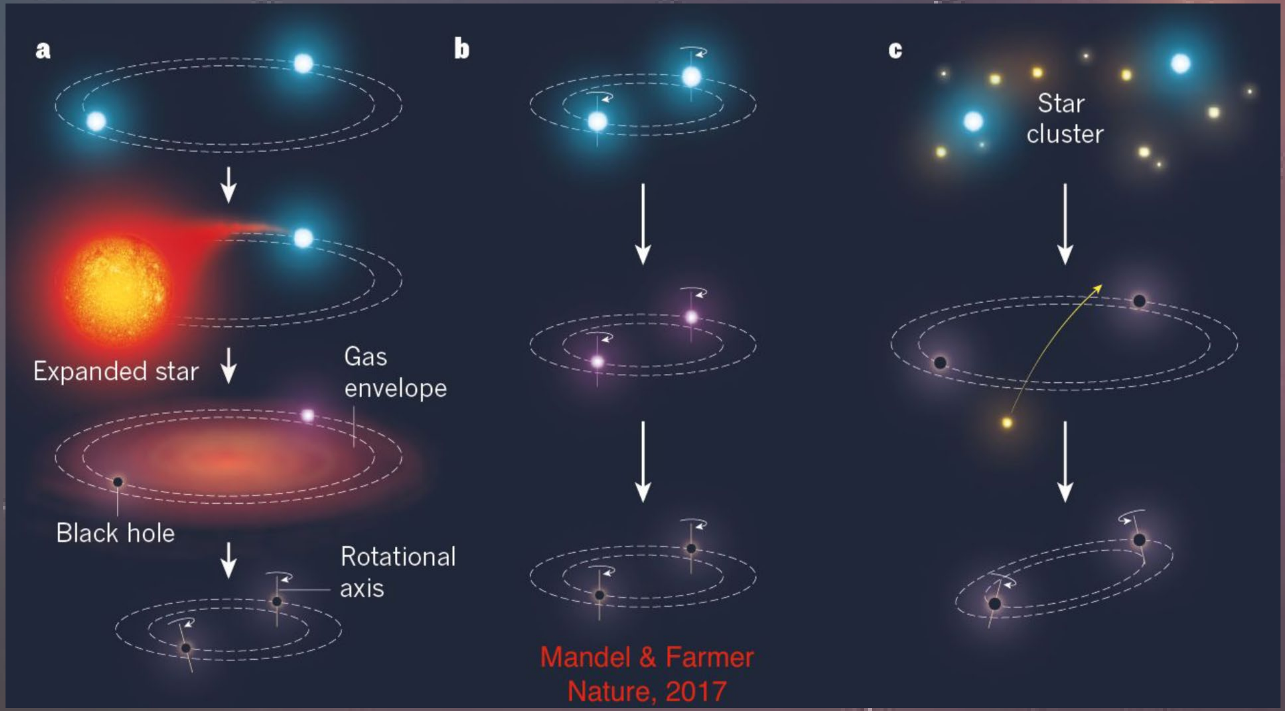


Posterior distributions on the mass-spin deviation parameters (black and purple curves) from GWTC-1 and GWTC-2.

Ref: Krishnendu and Ohme (2022)

- Study single event source properties:
  - Component masses (m_1, m_2).
  - Spins
  - Sky directions (ra, dec).
  - Luminosity distance.
- Model selection:
  - Test of presence of signal after merger (hypermassive NS or BH?).
  - Test of polarisation states (tensor/scalar/vector polarisations?)
  - Testing no-hair theorem.
  - Alternative theories of gravity.

Beyond Detection



Possible mechanisms for the merger of binary black holes (**Ref:** Mandel and Farmer, 2017):

a. Isolated evolution
b. Chemically homogenous evolution
c. Dynamical formation

- Study single event source properties:
  - Component masses (m_1 , m_2).
  - Spins
  - Sky directions (ra, dec).
  - Luminosity distance.
- Model selection:
  - Test of presence of signal after merger (hypermassive NS or BH?).
  - Test of polarisation states (tensor/scalar/vector polarisations?).
  - Testing no-hair theorem.
  - Alternative theories of gravity.
- Population studies:
  - Binary formation channels (isolated or dynamical?).

Gravitational Wave Parameter Estimation

Posterior distribution - for system parameters 'θ' (masses, spins, sky position, etc.) given detector strain data 's'

Likelihood – based on the assumption of stationary Gaussian noise.

$$p(\theta | s) = \frac{p(s | \theta) p(\theta)}{p(s)}$$

$$p(s | \theta) \propto \exp\left(-\frac{1}{2} \sum_{I} (s_I - h_I(\theta) | s_I - h_I(\theta)) \right)$$

$$(a | b) = 2 \int_0^\infty \mathrm{d}f \, \frac{\hat{a}(f)\hat{b}(f)^* + \hat{a}(f)^*\hat{b}(f)}{S_n(f)}$$

Prior – based on beliefs of system parameters before looking at data.

Gravitational Wave Parameter Estimation

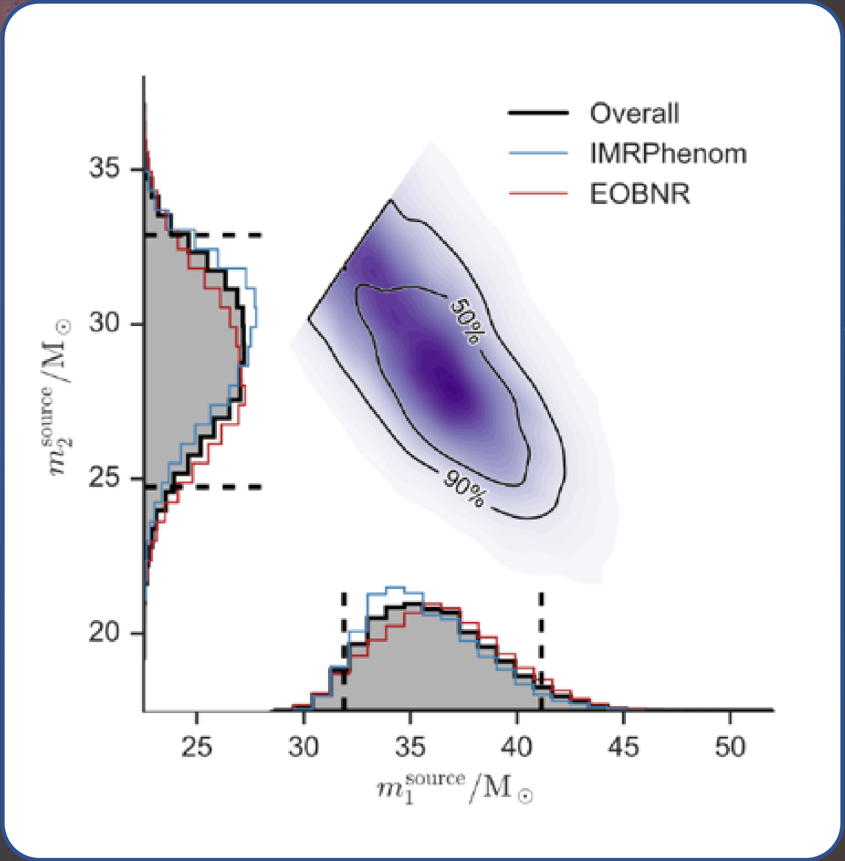
- **Bayesian Inference –** a way of computing probability density of things given data and model uncertainty.

- **Input:**
  - Experimental data
  - Model of gravitational wave signal.
  - Model of detector noise.

- **Output: Probability distribution**

  - Hypothesis: Signal or Noise? GR or not GR etc.

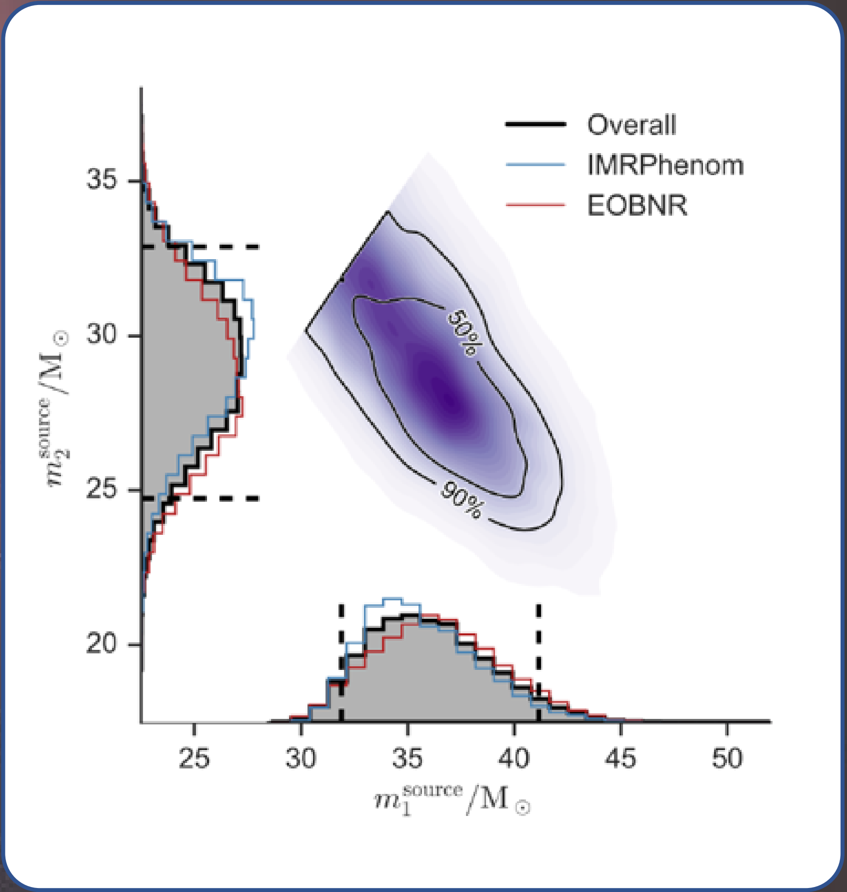
  - Source properties of compact binaries given noisy data and uncertainty in population model.



Ref: Abbott et al. (2016)

Gravitational Wave Parameter Estimation

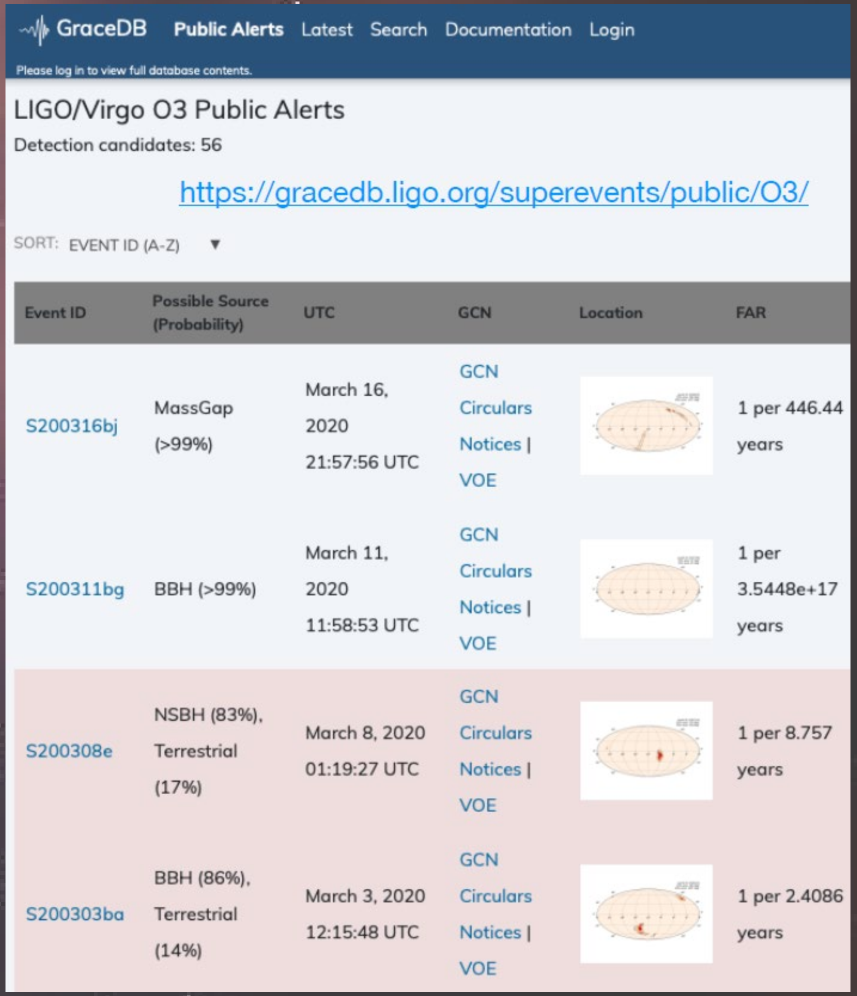
- We need to draw samples from the posterior.
- Iterative samplers like Markov Chain Monte Carlo (MCMC) and Nested Sampling are used to obtain posterior samples
- Move around in parameter space and compare strain data to waveform models.



Ref: Abbott et al. (2016)

Gravitational Wave Parameter Estimation

- Many likelihood evaluations are required for each independent sample.
- Likelihood is slow, requires a waveform to be generated.
- Typical analyses for O3 have taken between 6 hours and 5 days.
- Alternative – Machine Learning
  - Inference ~ milli-seconds.
  - Likelihood-free.
  - No posterior samples.
  - Only requires an ability to simulate training data (Simulation-based inference).



GraceDB Public Alerts Latest Search Documentation Login

Please log in to view full database contents.

LIGO/Virgo O3 Public Alerts

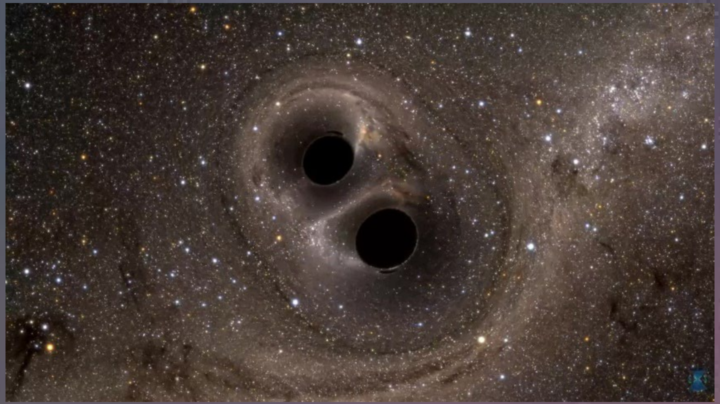
Detection candidates: 56

https://gracedb.ligo.org/superevents/public/O3/

SORT: EVENT ID (A-Z) ▼

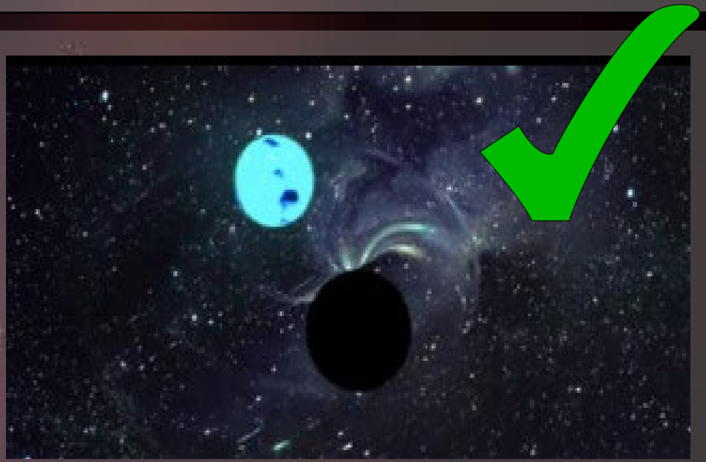
Event ID	Possible Source (Probability)	UTC	GCN	Location	FAR
S200316bj	MassGap (>99%)	March 16, 2020 21:57:56 UTC	GCN Circulars Notices | VOE		1 per 446.44 years
S200311bg	BBH (>99%)	March 11, 2020 11:58:53 UTC	GCN Circulars Notices | VOE		1 per 3.5448e+17 years
S200308e	NSBH (83%), Terrestrial (17%)	March 8, 2020 01:19:27 UTC	GCN Circulars Notices | VOE		1 per 8.757 years
S200303ba	BBH (86%), Terrestrial (14%)	March 3, 2020 12:15:48 UTC	GCN Circulars Notices | VOE		1 per 2.4086 years

The Need for Rapid GW Discovery



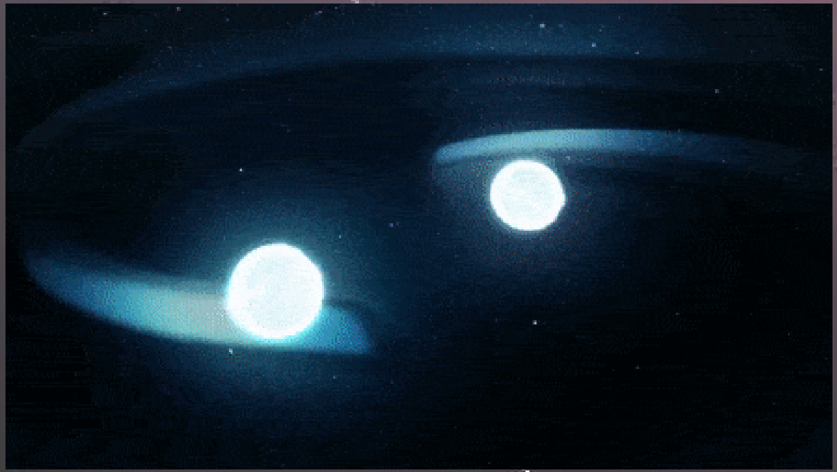
Binary Black Holes

Binary Neutron Stars



Neutron Star-Black Hole

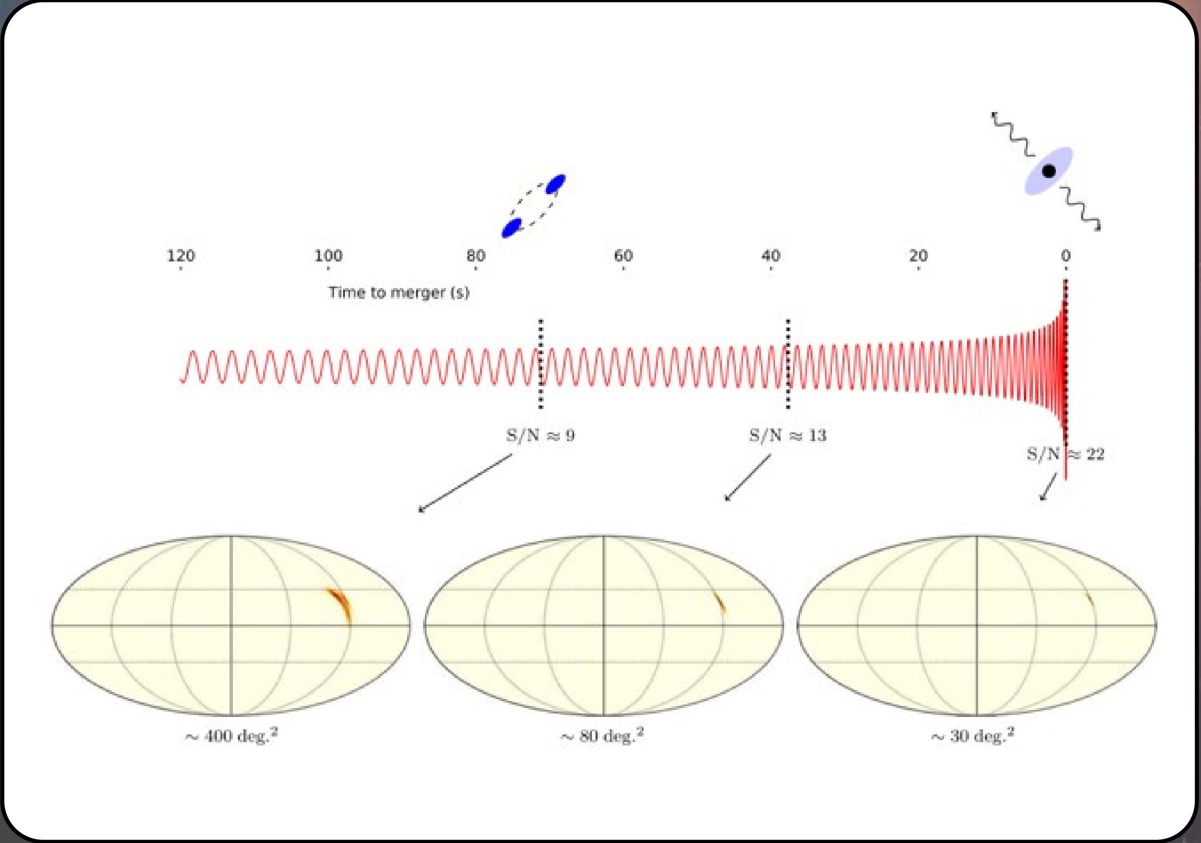
Compact Binary Coalescences (CBC)



GW + sGRB!

- Early electromagnetic observations are necessary for studying the following:
  - r-process nucleosynthesis and shock-heated ejecta (Optical & Ultraviolet).
  - Final state of the remnant (X-ray).
  - Premerger magnetosphere interactions (Radio).
  - Test models of BNS mergers as possible precursors of fast radio bursts (Radio).

Gravitational Wave 'Early Warning'

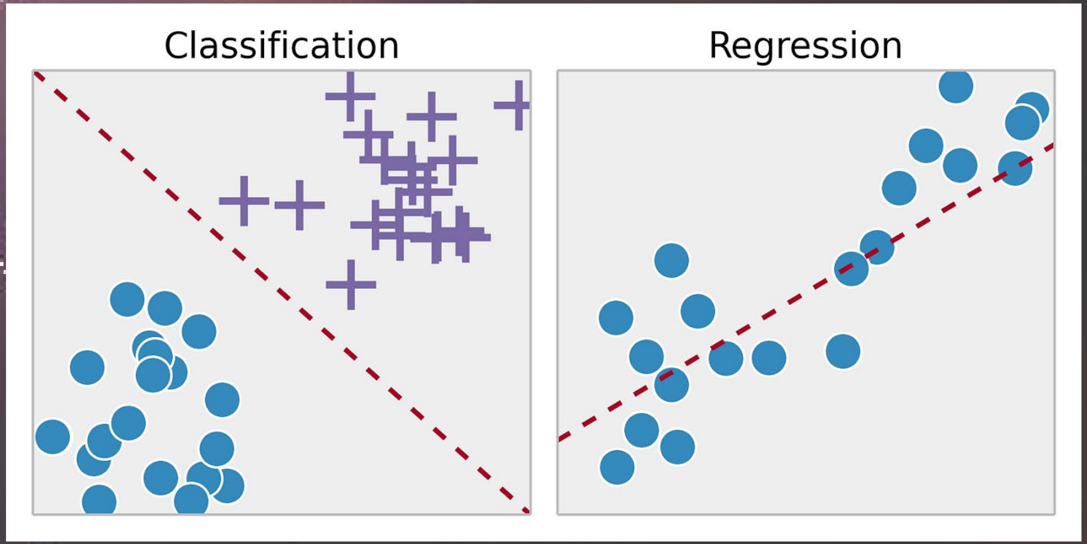


Ref: Magee *et al.* (2021), ApJL, 910, L21

- Binary neutron stars will spend 10–15 minutes in the band of Advanced LIGO and Virgo detectors at design sensitivity.
- ~ 7% (49%) of the total detectable BNS events will be detected 60 s (10 s) before the merger.
- About 2% will be detected before merger and localized to within 100 deg^2. – Sachdev *et al.* (2020), Kovalam *et al.* (2022).

Introduction to Machine Learning

- Machine Learning uses computers to learn patterns from data.
  - Typically used to solve problems that are hard to program in conventional ways.
- Typically we have a dataset $\{x^{(i)}\}$ consisting of many data points. The data points may or may not have associated labels $y^{(i)}$.
- Unsupervised learning: learn $p(x)$
  - Examples: Density estimation, sampling.
- Supervised learning: learn $p(y|x)$
  - Examples: Classification, regression



Machine Learning Recipe

- Build a dataset of training examples $(x^{(i)}, y^{(i)})$.
- Define a parametric probabilistic model of the data:

$$p(y|\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu(\boldsymbol{x}))^2}{2\sigma^2}\right)$$

where $\mu(\boldsymbol{x}) = \boldsymbol{\theta}\cdot\boldsymbol{x}$; σ fixed.

- Choose a measure of performance (loss function):

$$J(\theta) = -\sum_{i=1}^{N} \log p(y^{(i)}|\boldsymbol{x}^{(i)};\boldsymbol{\theta})$$

$$= \frac{N}{2}\log 2\pi\sigma^2 + \sum_{i=1}^{N} \frac{(y^{(i)}-\mu(\boldsymbol{x}^{(i)}))^2}{2\sigma^2}$$

$\propto$ mean squared error

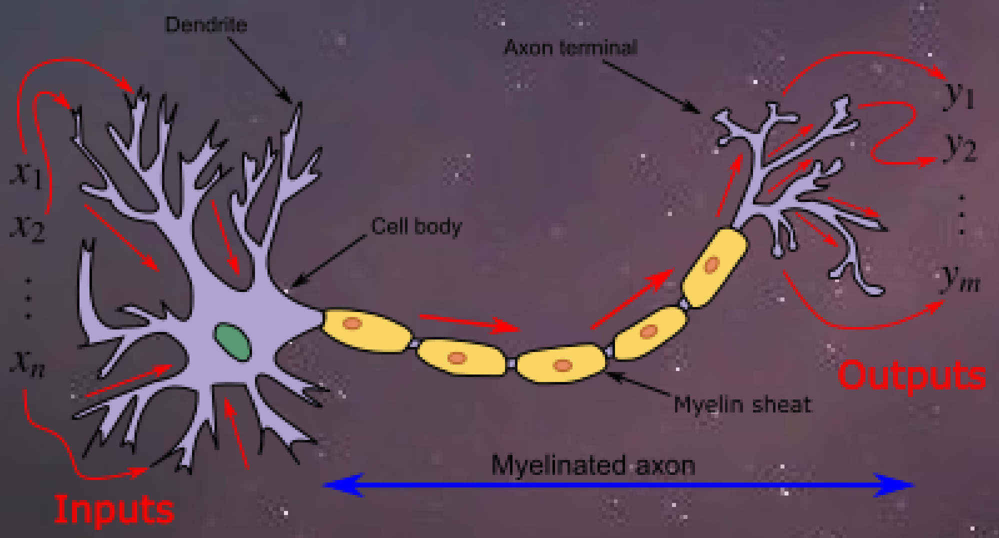
For greater flexibility, use a neural network

- Choose θ by fitting model to data according to loss function (maximum likelihood):

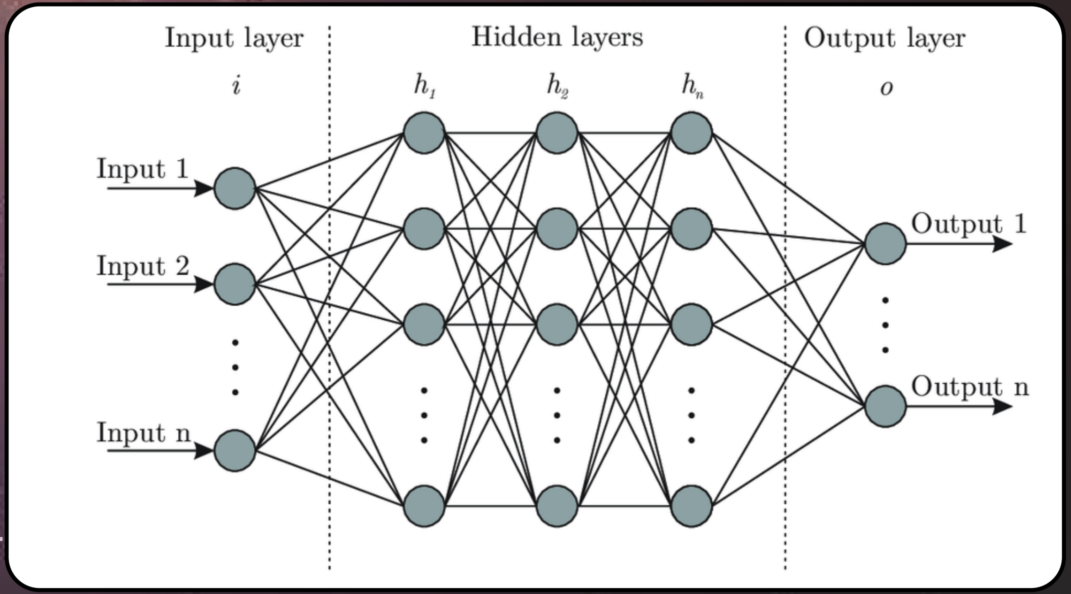
$$\nabla_{\boldsymbol{\theta}} J = 0 \quad \Longrightarrow \quad \boldsymbol{\theta}_{\text{ML}} = \left(X^\top X\right)^{-1} X^\top \boldsymbol{y}$$

Neural Networks

A Biological Neuron

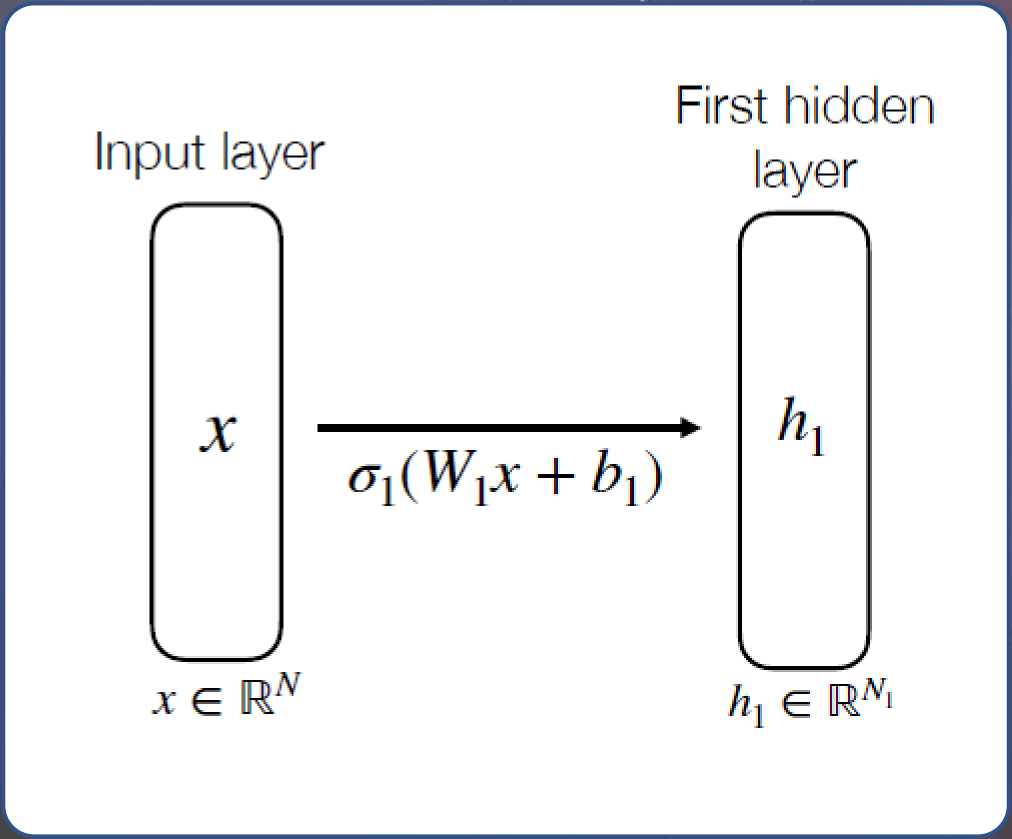


An Artificial Neural Network



Deep learning – a type of machine learning process that uses interconnected nodes or neurons in a layered structure that resembles the human brain.

Neural Networks



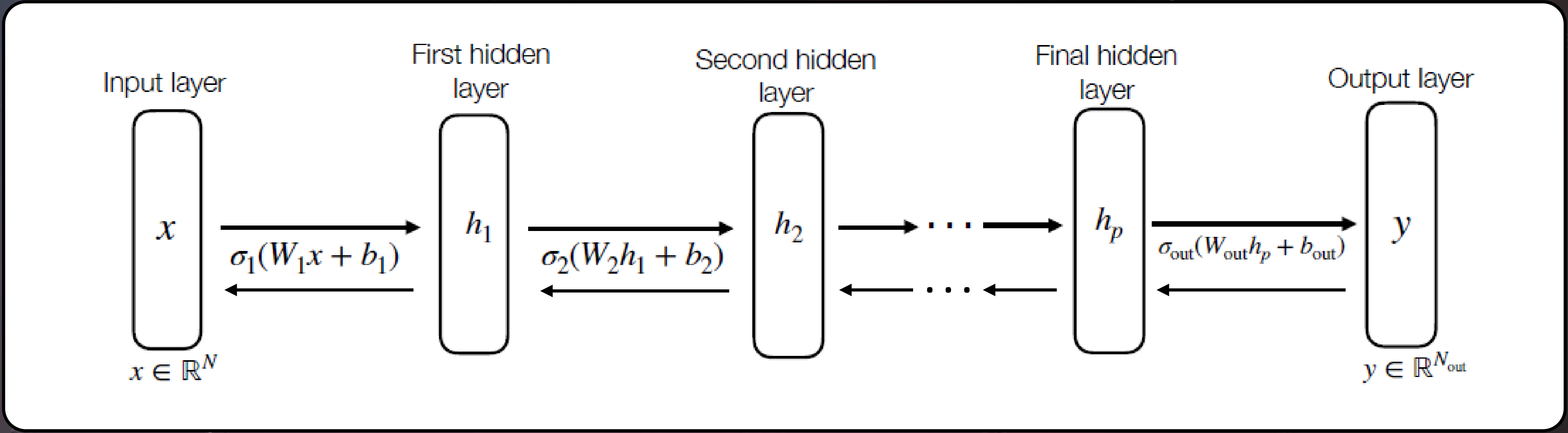
- Consists of:
  - A linear transformation –

$$Z_1 = W_1 x + b_1$$

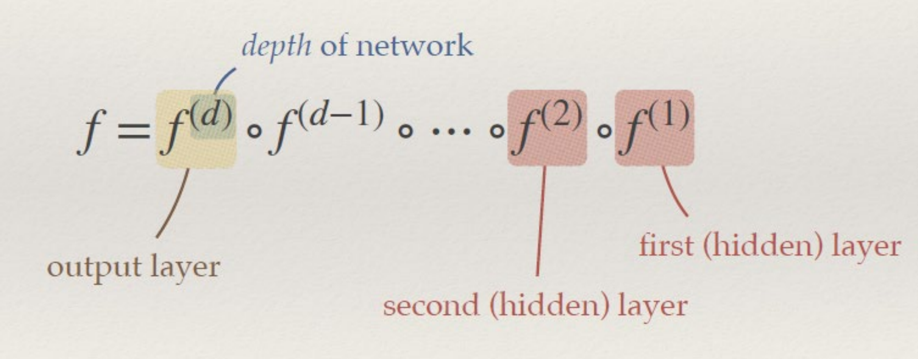
  - Simple element-wise non-linear mapping:

$$\sigma_1(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

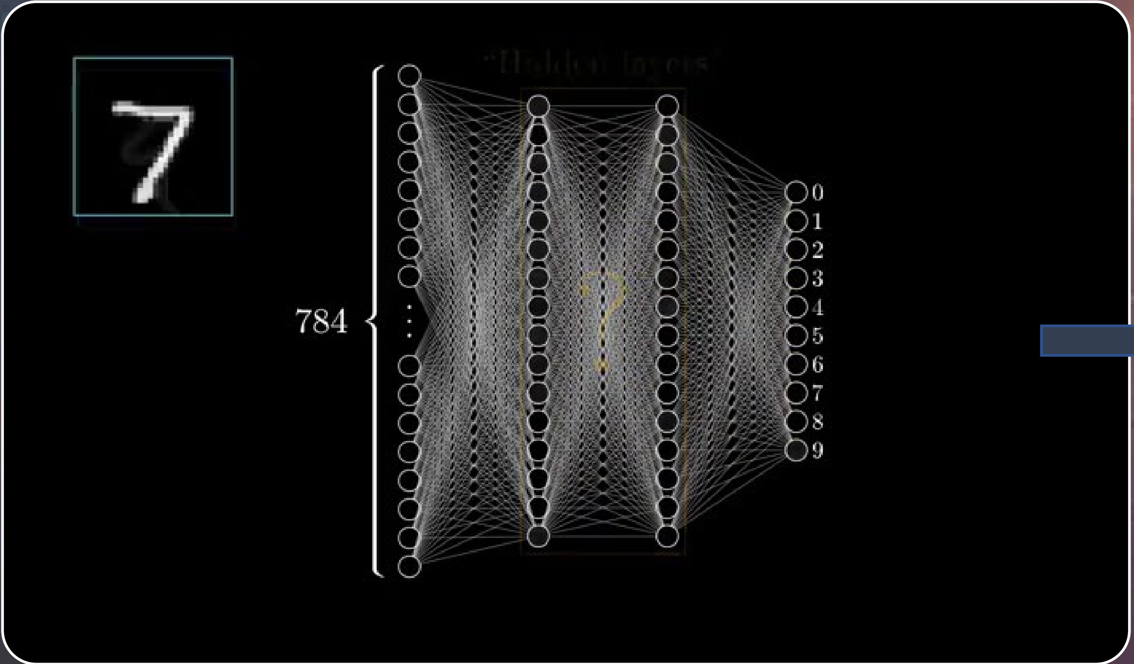
Neural Networks

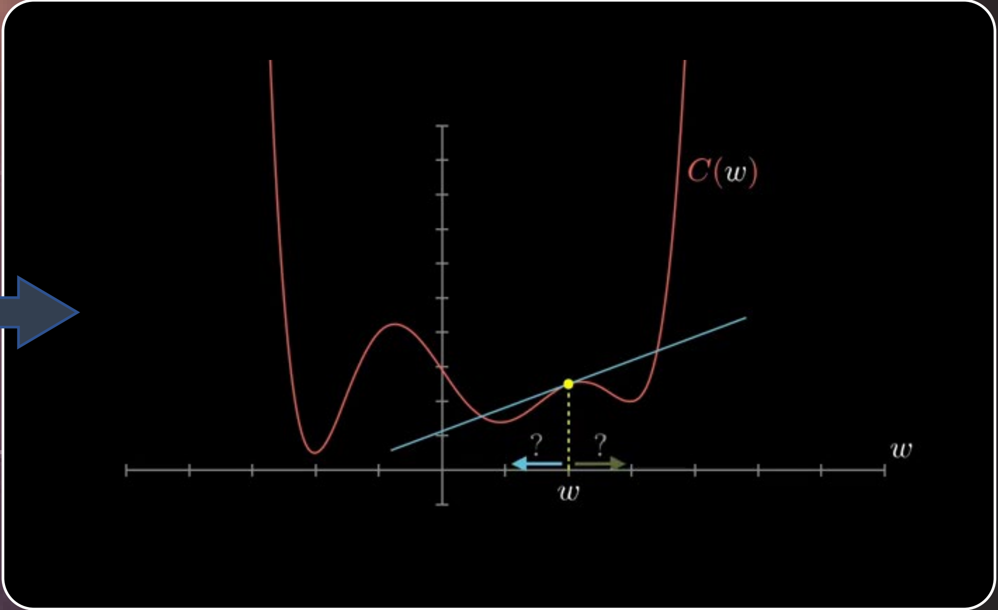


- Can be interpreted as a composition of mappings.
- Train network by tuning the weights W and biases b to minimize loss function.
- Stochastic Gradient descent combined with chain rule ("backpropagation") to adjust W and b.



Neural Networks – Optimizing a Loss Function





Neural Network –
Extracts features from input data

Stochastic Gradient Descent

$$\boldsymbol{\theta}_1 = \boldsymbol{\theta}_0 - \epsilon \nabla_{\boldsymbol{\theta}} J|_{\boldsymbol{\theta}_0}$$

Learning rate

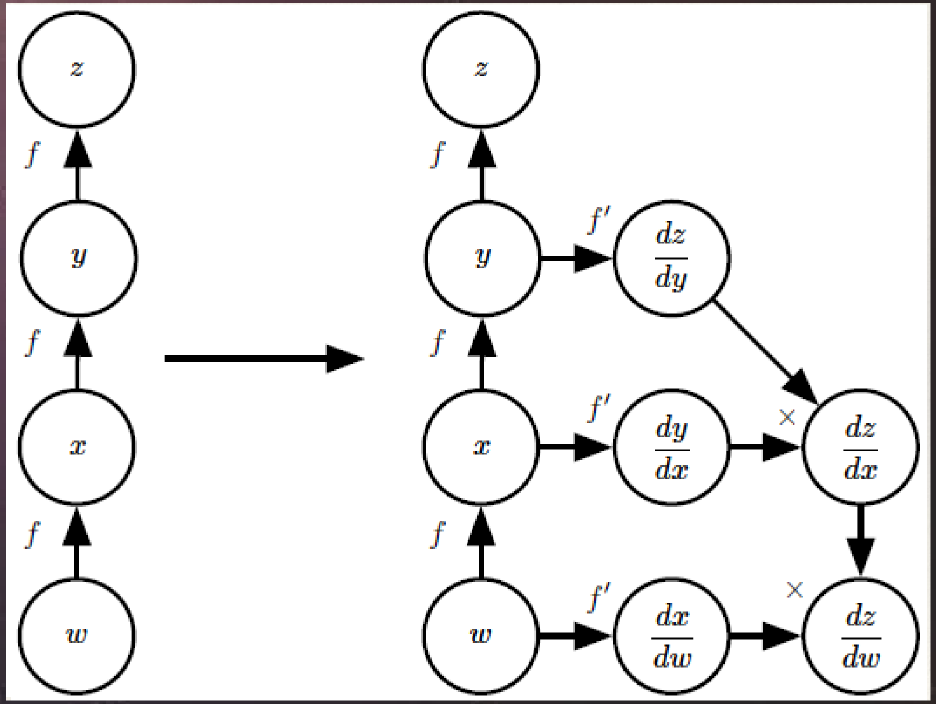
Backpropagation Algorithm

- To train the network, we need to efficiently compute gradients with respect to all of the network parameters. This is accomplished using a form of automatic differentiation called backpropagation.

- Relies on the compositional nature of neural networks, plus the chain rule of calculus and the differentiability of all operations.

$$
\begin{aligned}
\frac{\partial z}{\partial w} &= \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}\frac{\partial x}{\partial w} \\
&= f'(y)f'(x)f'(z) \\
&= f'(f(f(w)))f'(f(w))f'(w)
\end{aligned}
$$

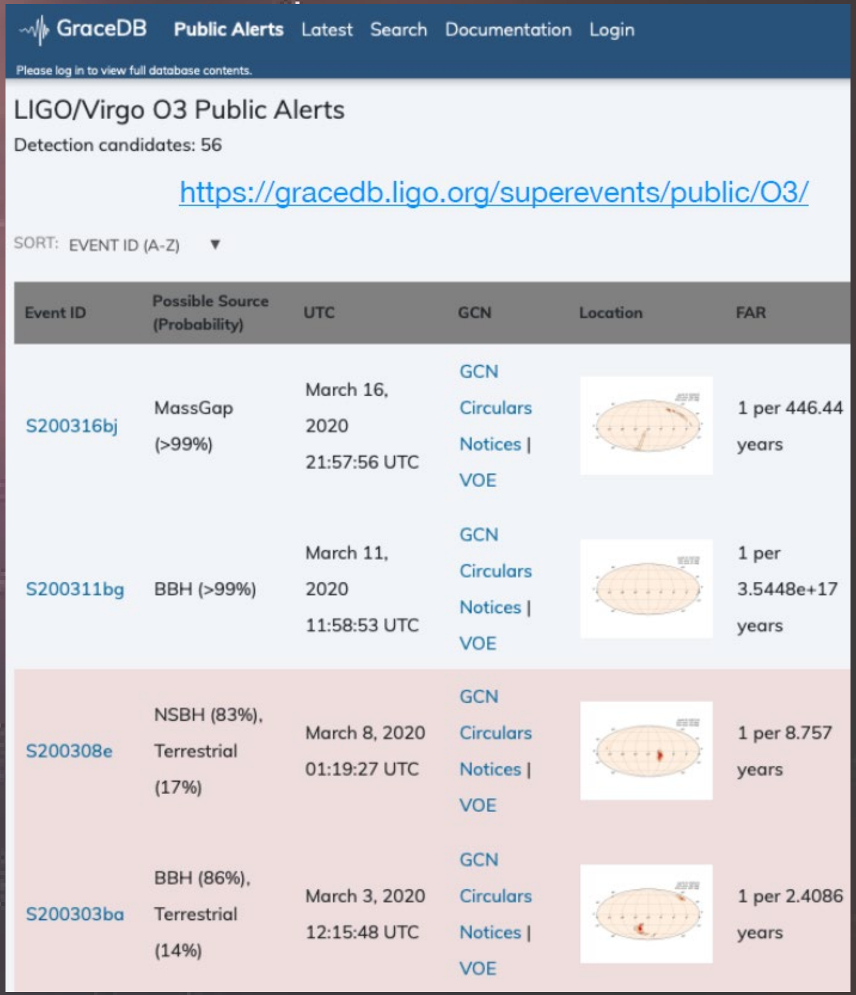
Efficient implementations in every deep-learning framework (PyTorch, TensorFlow, JAX, …)



Ian Goodfellow (2016)

Gravitational Wave Parameter Estimation

- Many likelihood evaluations are required for each independent sample.
- Likelihood is slow, requires a waveform to be generated.
- Typical analyses for O3 have taken between 6 hours and 5 days.
- Alternative – Machine Learning
  - Likelihood-free.
  - No posterior samples.
  - Only requires an ability to simulate training data (Simulation-based inference).



GraceDB Public Alerts Latest Search Documentation Login

Please log in to view full database contents.

LIGO/Virgo O3 Public Alerts

Detection candidates: 56

https://gracedb.ligo.org/superevents/public/O3/

SORT: EVENT ID (A-Z) ▼

Event ID	Possible Source (Probability)	UTC	GCN	Location	FAR
S200316bj	MassGap (>99%)	March 16, 2020 21:57:56 UTC	GCN Circulars Notices | VOE		1 per 446.44 years
S200311bg	BBH (>99%)	March 11, 2020 11:58:53 UTC	GCN Circulars Notices | VOE		1 per 3.5448e+17 years
S200308e	NSBH (83%), Terrestrial (17%)	March 8, 2020 01:19:27 UTC	GCN Circulars Notices | VOE		1 per 8.757 years
S200303ba	BBH (86%), Terrestrial (14%)	March 3, 2020 12:15:48 UTC	GCN Circulars Notices | VOE		1 per 2.4086 years

Likelihood-free Inference with Neural Networks

- Objective:

$$p(\theta\,|\,s) \to p_{\text{true}}(\theta\,|\,s)$$

Intractable to evaluate posterior for each sample

- Cross entropy:

$$L = -\int \mathrm{d}s\, p_{\text{true}}(s) \int \mathrm{d}\theta\, p_{\text{true}}(\theta\,|\,s)\, \log p(\theta\,|\,s)$$

$p_{\text{true}}(s)\, p_{\text{true}}(\theta|s) =$
$p_{\text{true}}(\theta)\, p_{\text{true}}(s|\theta)$

- Use Baye's theorem and Monte Carlo approximation:

$$L = -\int \mathrm{d}\theta\, p_{\text{true}}(\theta) \int \mathrm{d}s\, p_{\text{true}}(s\,|\,\theta)\, \log p(\theta\,|\,s)$$

$$\approx -\frac{1}{N}\sum_{i=1}^{N} \log p(\theta^{(i)}|s^{(i)}),$$

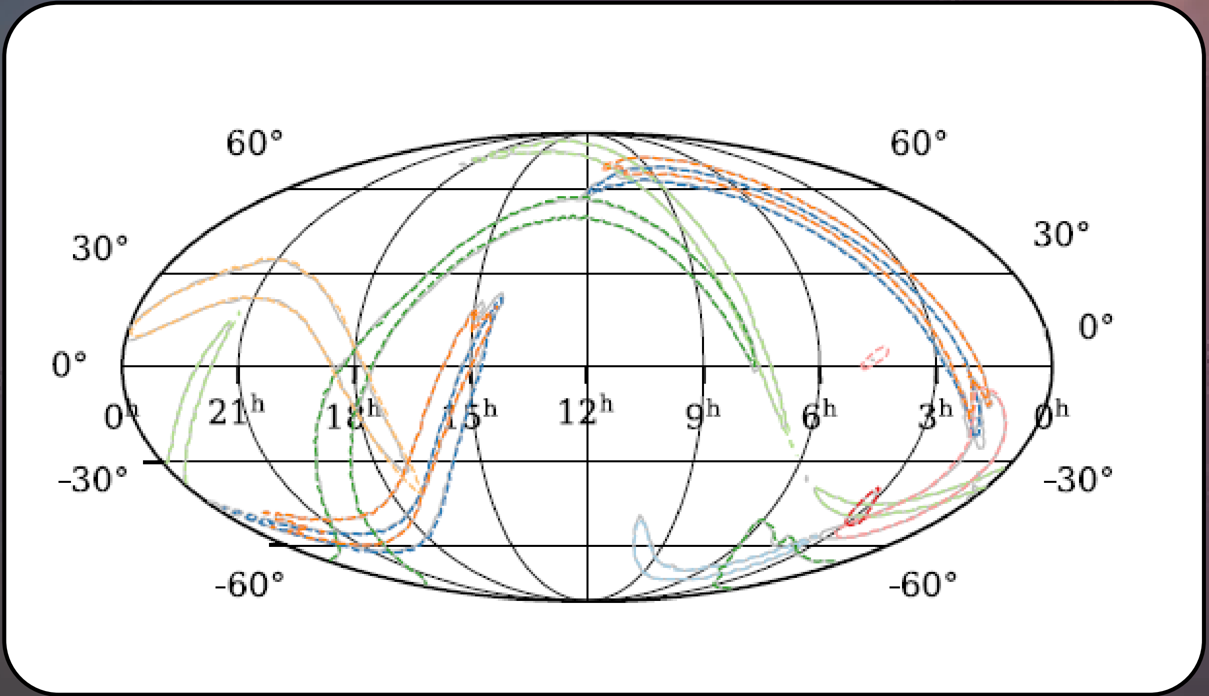
where $\theta^{(i)} \sim p_{\text{true}}(\theta)$,
$s^{(i)} \sim p_{\text{true}}(s|\theta^{(i)})$

No posterior samples.
No expensive likelihood evaluation.

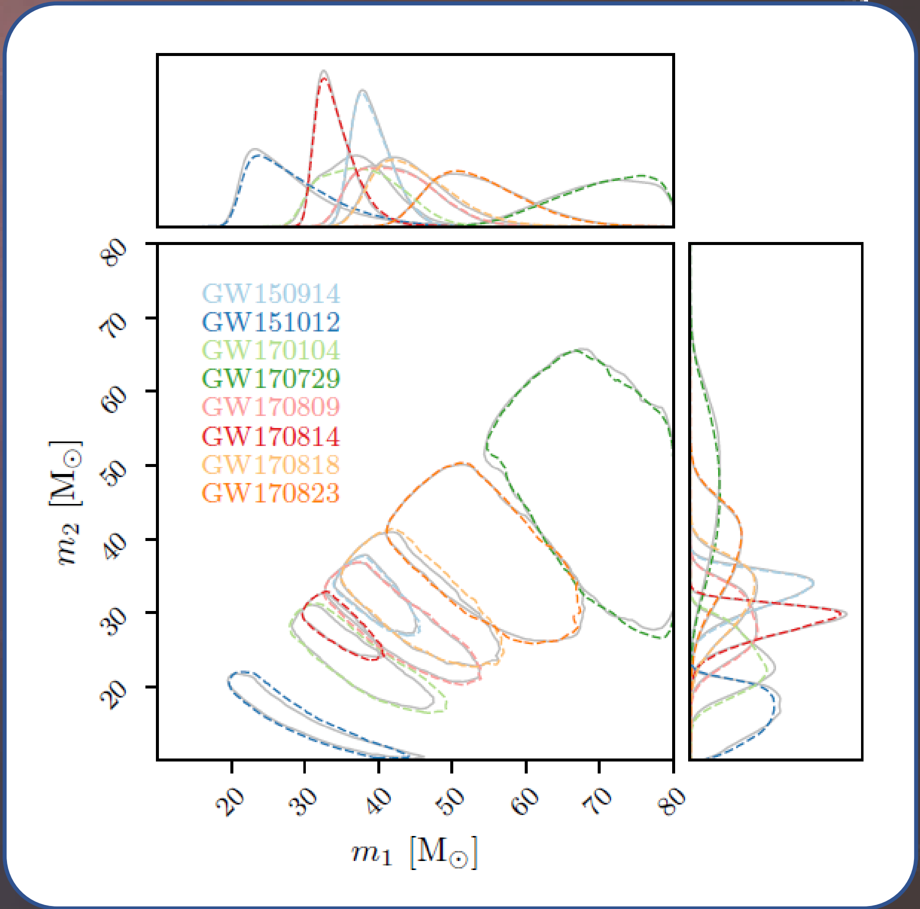
Deep Learning for BBH Parameter Estimation

Ref: Dax *et al.*, Phys. Rev. Lett. 127, 241103 (2021)

Sky Localization

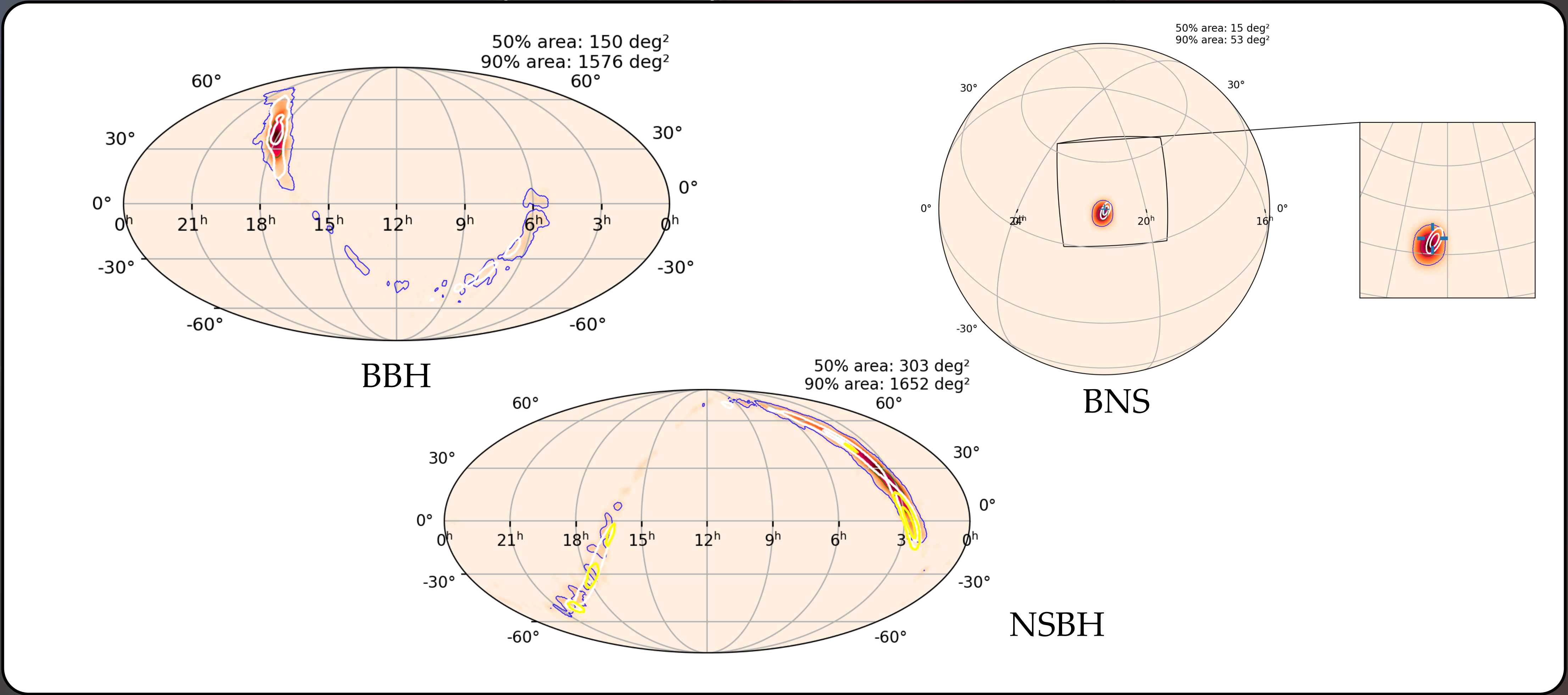


Mass Estimation



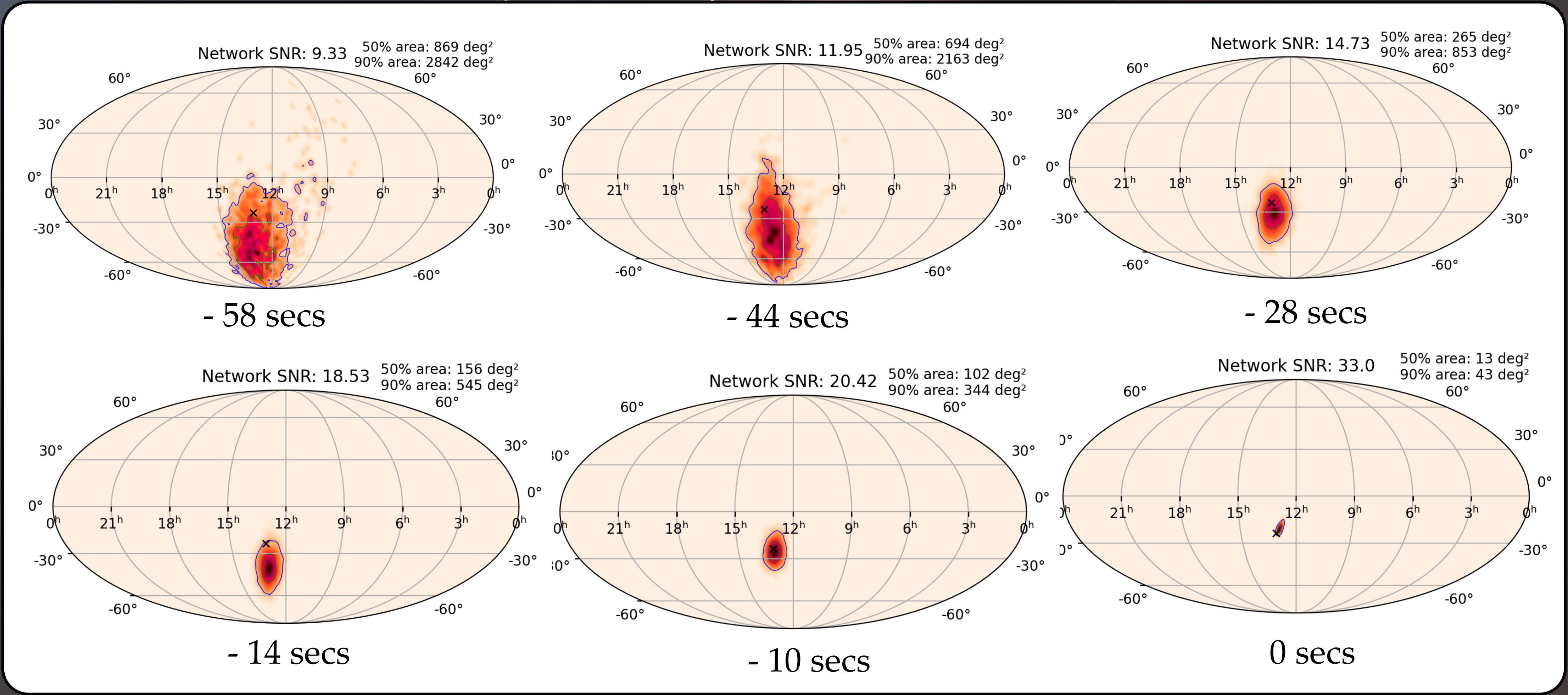
Analyzed all events consistent with prior $m_1, m_2 > 10$ solar mass

Deep Learning for GW Sky Localization



Ref: Chatterjee *et al.* , 2022 (Under review)

Deep Learning for Pre-merger Localization of GW170817



Ref: Chatterjee *et al.*, 2023 (accepted for publication in ApJ)