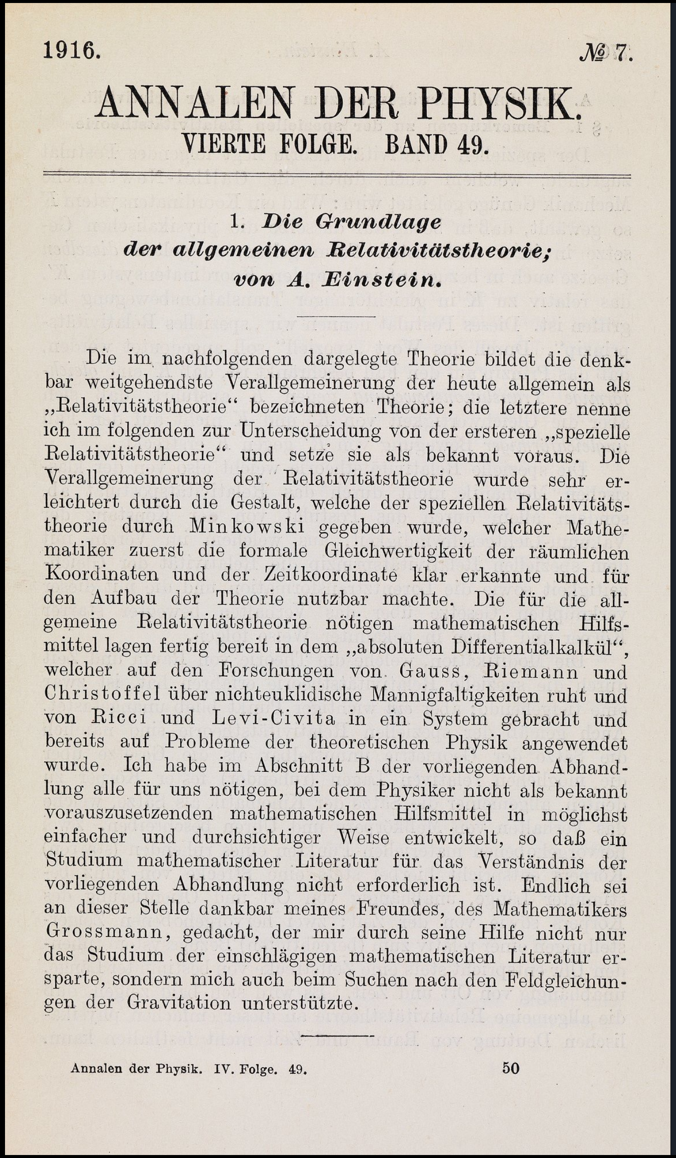
Overview of ML for Gravitational Waves

***Eric A. Moreno**, on behalf of A3D3*

Additional collaborators (non-exhaustive): Will Benoit, Deep Chatterjee, Hong-Yin(Andy) Chen, Chia-Jui Chou, Michael Coughlin, Alec Gunny, Philip Coleman Harris, Ethan Jacob Marx, Erik Katsavounidis, Ryan Raikman, Dylan Sheldon Rankin, Muhammed Saleem Cholayil, Li-Cheng Yang, et al.

ML4Jets November 3rd, 2022

Accelerated AI Algorithms for Data-Driven Discovery



1916. № 7.

ANNALEN DER PHYSIK.
VIERTE FOLGE. BAND 49.

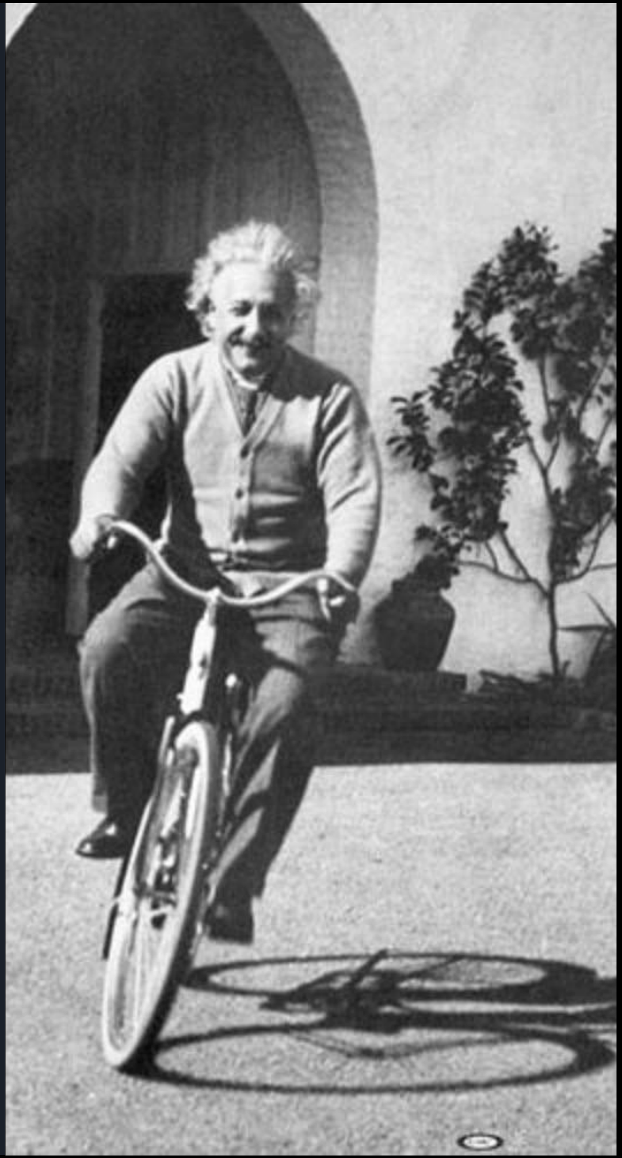
1. *Die Grundlage der allgemeinen Relativitätstheorie;*
von A. Einstein.

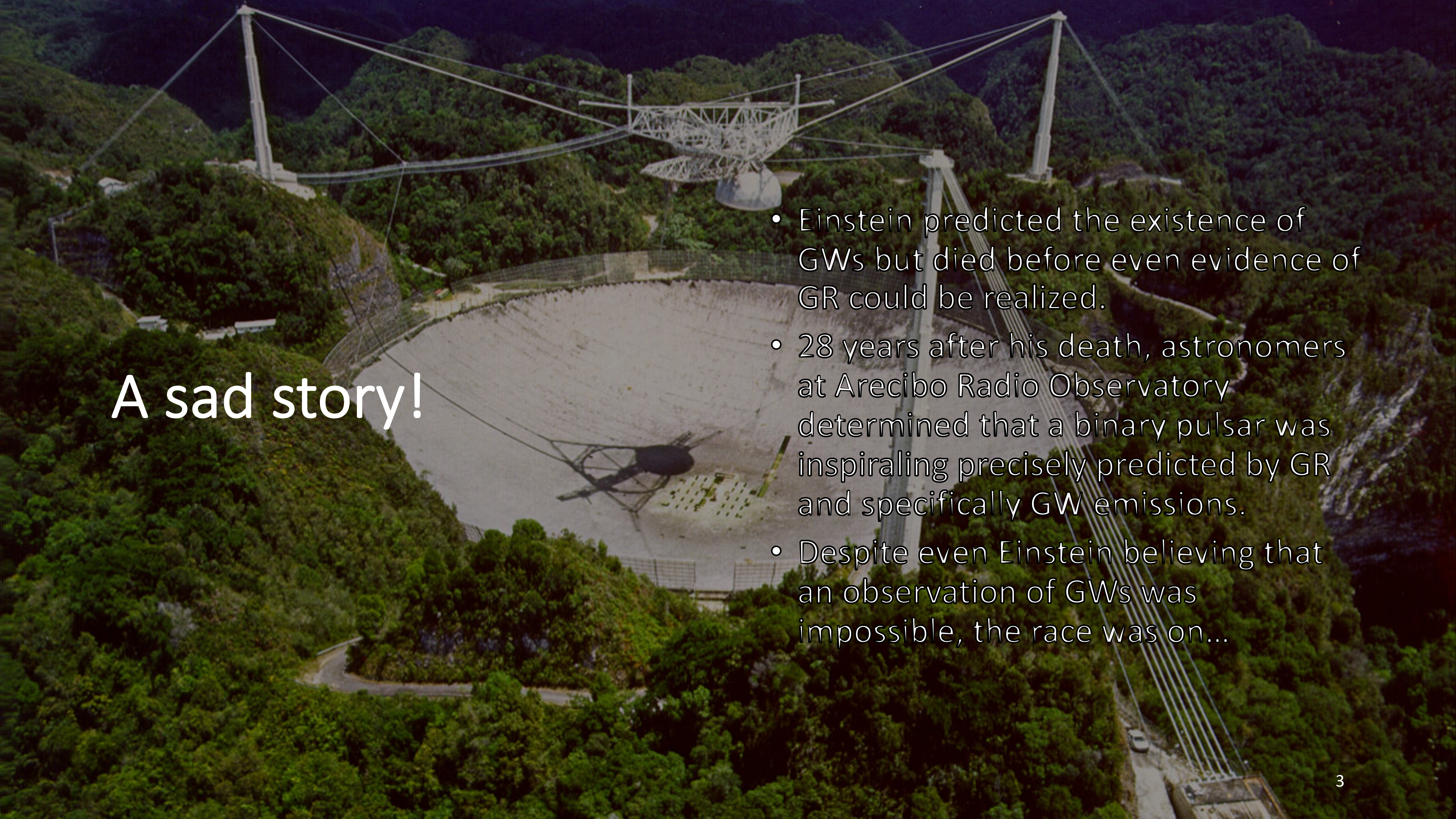
Die im nachfolgenden dargelegte Theorie bildet die denkbar weitgehendste Verallgemeinerung der heute allgemein als „Relativitätstheorie“ bezeichneten Theorie; die letztere nenne ich im folgenden zur Unterscheidung von der ersteren „spezielle Relativitätstheorie“ und setze sie als bekannt voraus. Die Verallgemeinerung der Relativitätstheorie wurde sehr erleichtert durch die Gestalt, welche der speziellen Relativitätstheorie durch Minkowski gegeben wurde, welcher Mathematiker zuerst die formale Gleichwertigkeit der räumlichen Koordinaten und der Zeitkoordinate klar erkannte und für den Aufbau der Theorie nutzbar machte. Die für die allgemeine Relativitätstheorie nötigen mathematischen Hilfsmittel lagen fertig bereit in dem „absoluten Differentialkalkül“, welcher auf den Forschungen von Gauss, Riemann und Christoffel über nichteuklidische Mannigfaltigkeiten ruht und von Ricci und Levi-Civita in ein System gebracht und bereits auf Probleme der theoretischen Physik angewendet wurde. Ich habe im Abschnitt B der vorliegenden Abhandlung alle für uns nötigen, bei dem Physiker nicht als bekannt vorauszusetzenden mathematischen Hilfsmittel in möglichst einfacher und durchsichtiger Weise entwickelt, so daß ein Studium mathematischer Literatur für das Verständnis der vorliegenden Abhandlung nicht erforderlich ist. Endlich sei an dieser Stelle dankbar meines Freundes, des Mathematikers Grossmann, gedacht, der mir durch seine Hilfe nicht nur das Studium der einschlägigen mathematischen Literatur ersparte, sondern mich auch beim Suchen nach den Feldgleichungen der Gravitation unterstützte.

Annalen der Physik. IV. Folge. 49. 50

A little history…

- It’s been 9 years since the “happiest thought” of Albert Einstein’s life sitting in the patent office in Bern.
- World War I is raging on in Europe in 1916 and the newly famous Einstein publishes the General Theory of Relativity
- Included in GR is the concept of ripples in spacetime that should be induced by accelerating objects – **Gravitational Waves (GW)!**





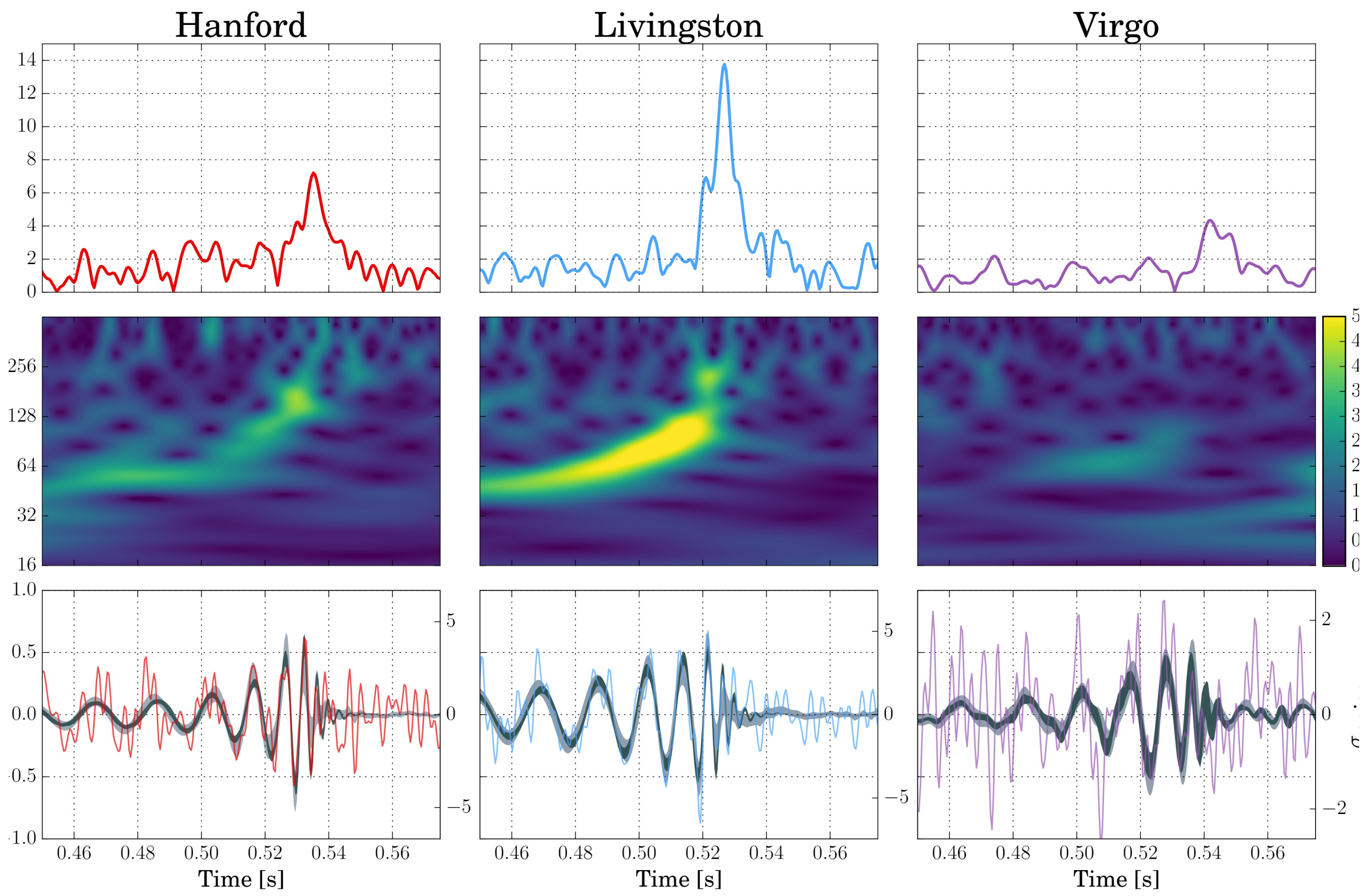
A sad story!

- Einstein predicted the existence of GWs but died before even evidence of GR could be realized.
- 28 years after his death, astronomers at Arecibo Radio Observatory determined that a binary pulsar was inspiraling precisely predicted by GR and specifically GW emissions.
- Despite even Einstein believing that an observation of GWs was impossible, the race was on…

A long road for GWs…

- 1980s – NSF Funds MIT and Caltech to research laser interferometers, LIGO, GEO founded
- 1990s – Construction begins on LIGO, VIRGO, GEO600
- 1999-2003 – LIGO/VIRGO/GEO inauguration
- Later 2000s-2010s – LIGO/VIRGO starts upgrading to Advanced LIGO/VIRGO
- September 2015 – aLIGO ready for first run
- September 14th, 2015 – aLIGO detects GWs from collision of two black holes
- 2017 - Multi-Messenger Astronomy (MMA) is realized (LIGO/VIRGO), additional GW experiments are built, funded, theorized
- 2020s+ – Golden era for GW astronomy! Detectors from all over the world are coming online

GW170817



International GW Collaborations

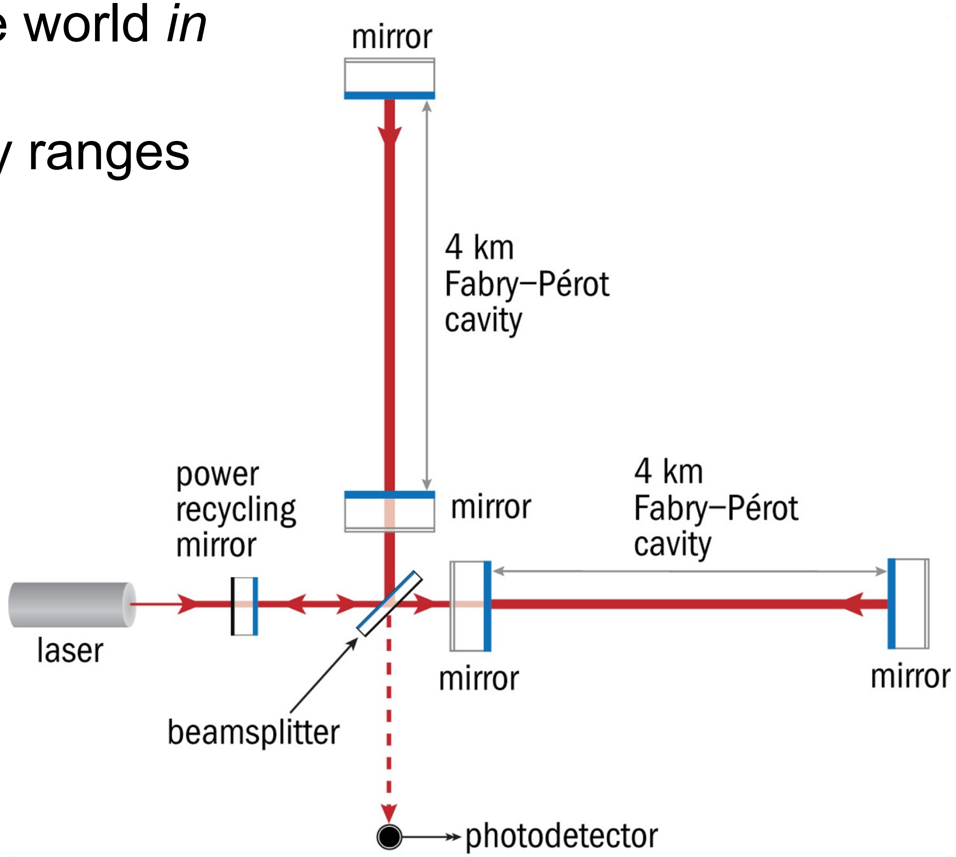
Observatories now (and in the future) include:

1. GEO600 (Germany)
2. LIGO (Hanford, Livingston USA)
3. VIRGO (Italy)
4. KAGRA (Japan)
5. LIGO-India (India 2020s)
6. Einstein Telescope (Europe 2030s)
7. LISA (Geosynchronous orbit 2030s)

General Concept of GW Interferometers

- Michaelson interferometers (at heart) placed across the world *in different orientations* for full sky coverage
- Different designs can be sensitive to different frequency ranges

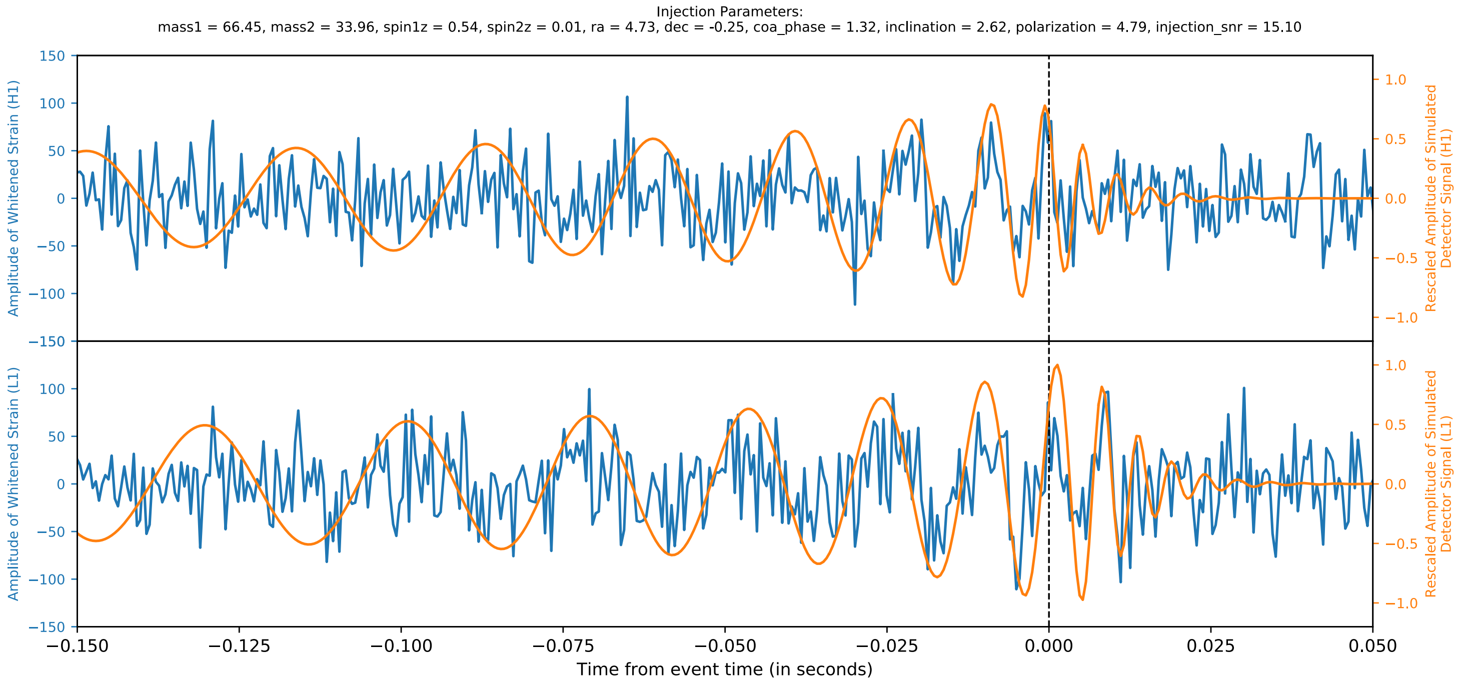
Source: Elena Cuoco - Real Time Classifier for transient signals in Gravitational Waves, From raw data to classified triggers



Produces: **time-series** [1-D strain + auxiliary channels]

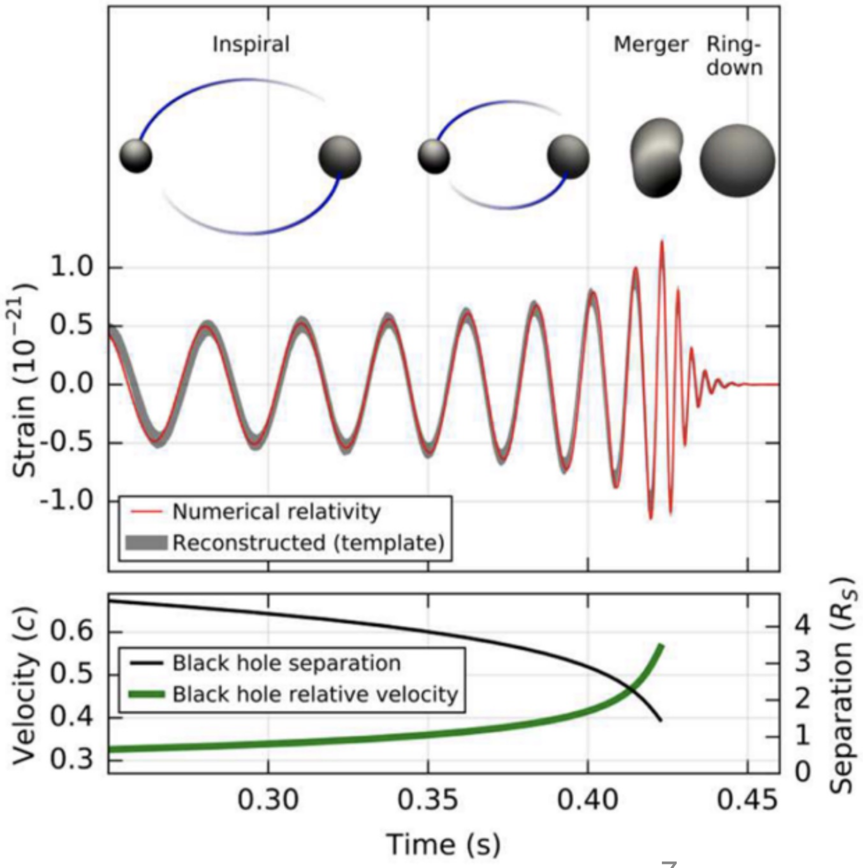
Sounds trivial!!

BBH Sample



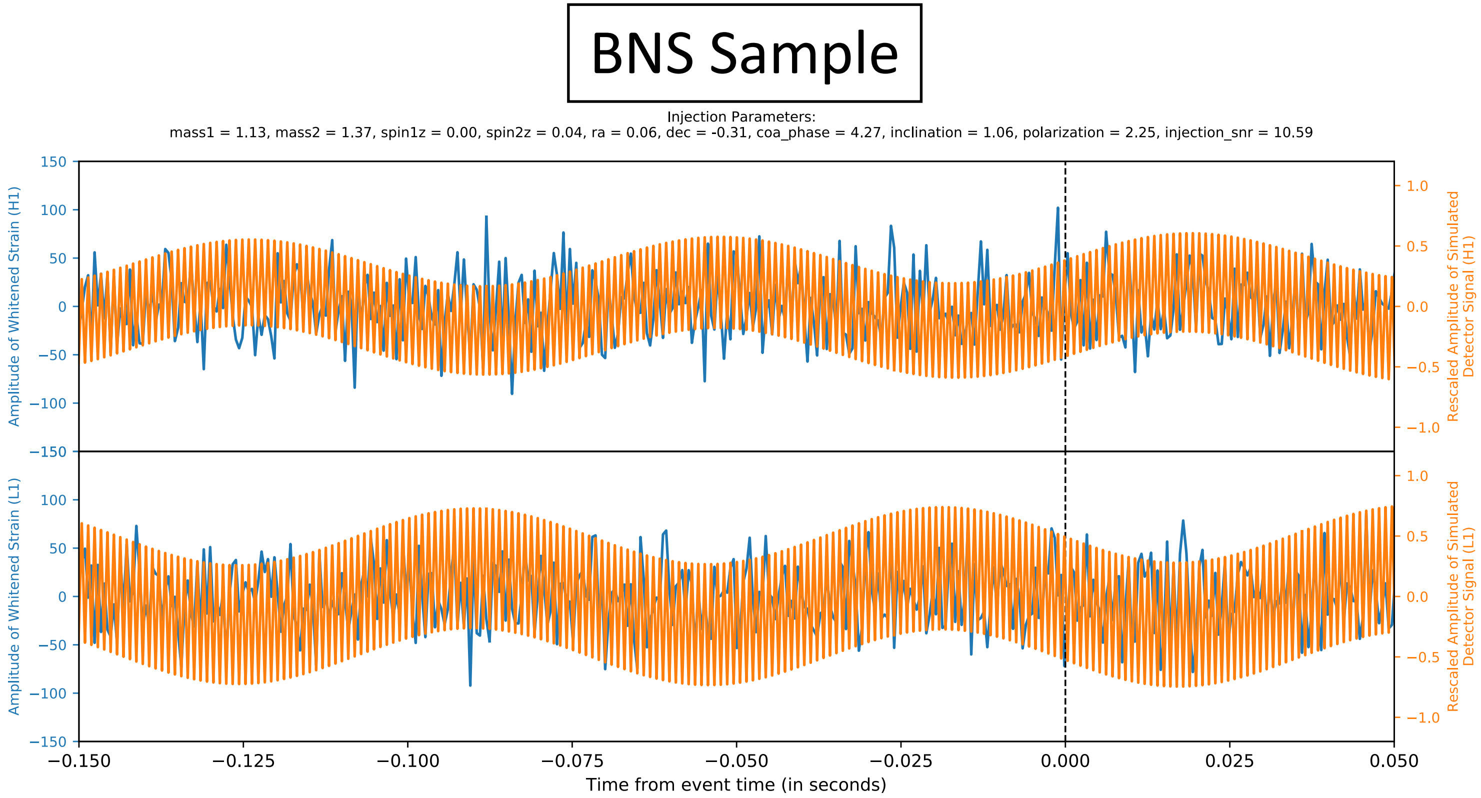
Source: github.com/timothygebhard/ggwd, https://www.gw-openscience.org/data/

Figure: LIGO/Virgo



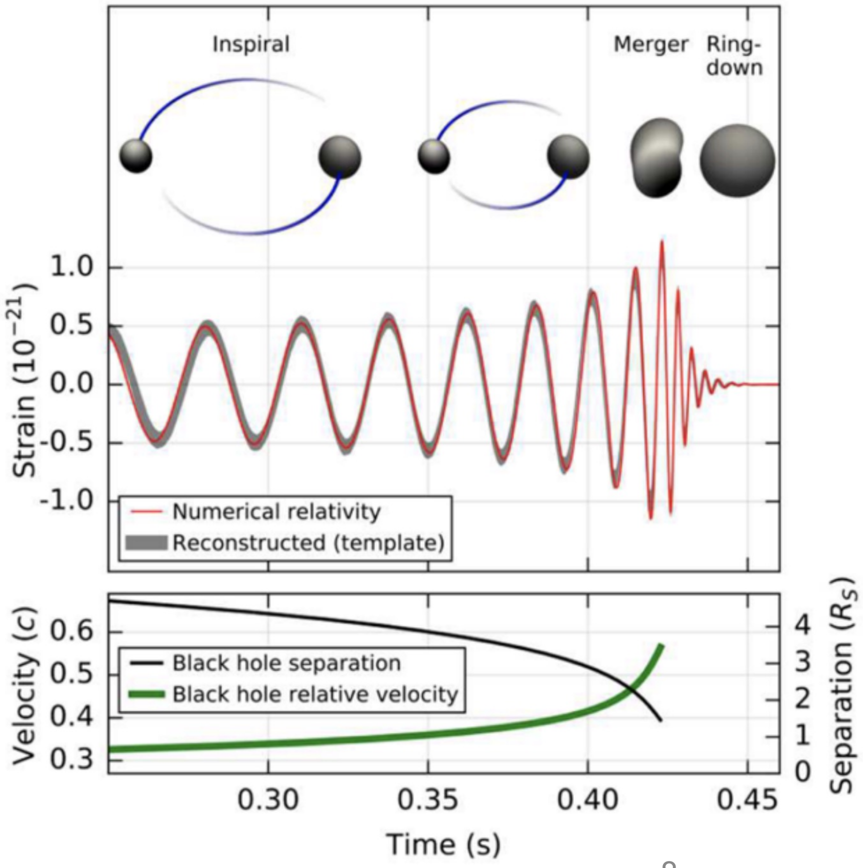
Produces: **time-series** [1-D strain + auxiliary channels]

Sounds trivial!!



Source: github.com/timothygebhard/ggwd, https://www.gw-openscience.org/data/

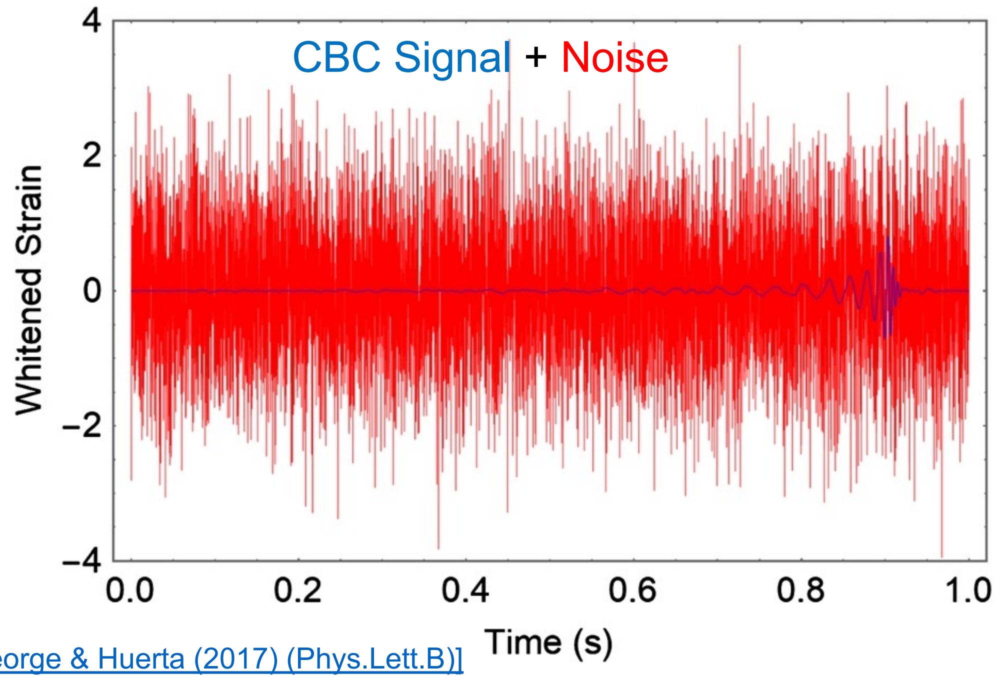
Figure: LIGO/Virgo



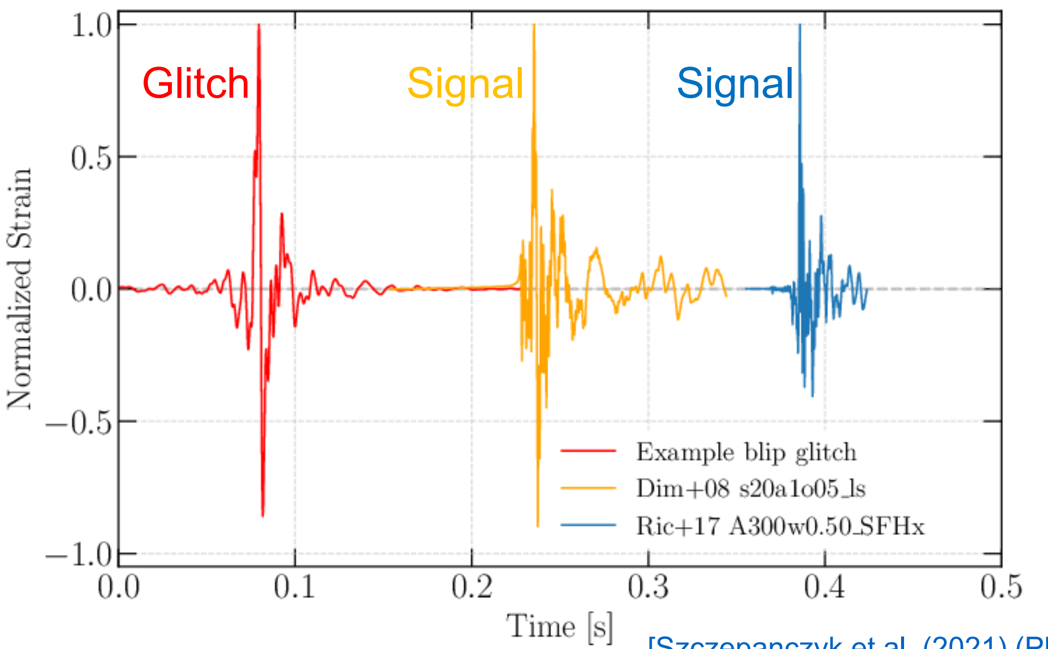
Produces: **time-series** [1-D strain + auxiliary channels]

GW Dataset - Nontrivial

- Sounds trivial, but isn't – length measurements are ~ $10^{-22}\ m$
- Constantly changing detector noise usually clouds signal
- Detector glitches occur every O(10 sec) – resembling GWs in excess power!



[George & Huerta (2017) (Phys.Lett.B)]



[Szczepanczyk et al. (2021) (Phys.Rev.D)]

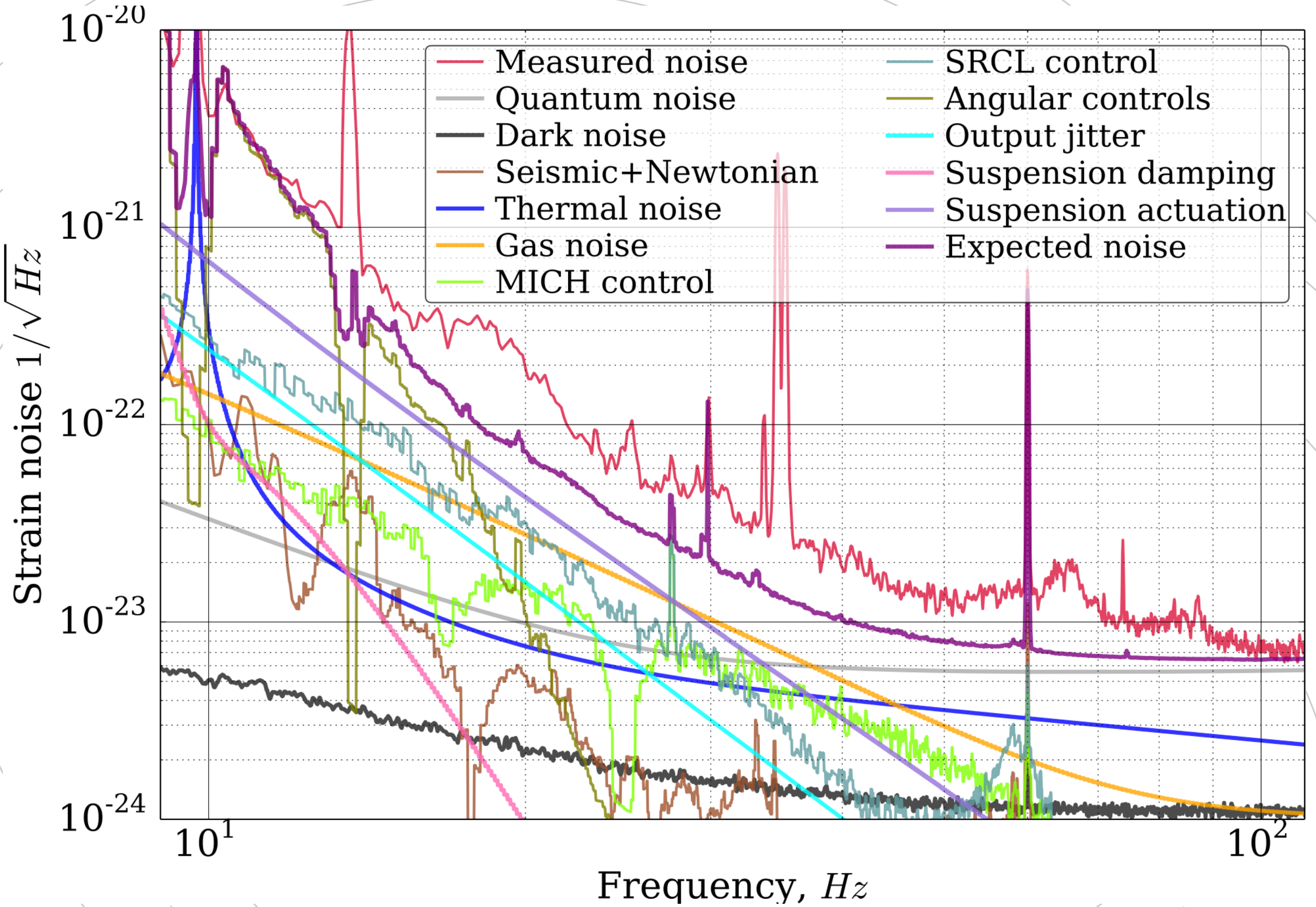
ML-Revolution in GW-physics

Check out: https://iphysresearch.github.io/Survey4GWML/ for full paper lists

Similar to HEP, AI can play an important role in real-time and offline data processing:

- Data Quality
- GW Detection, Real-time low latency MMA
- Parameter Estimation
- Anomaly Detection

Note: Advances in any collaboration propagate quickly to other GW facilities - lots of knowledge share, open-science 😉



Section 1: Data Quality

Data Cleaning

[Ormiston et al. (2020) (Phys.Rev.Res)]
https://indico.cern.ch/event/1156222/contributions/5062800/
https://indico.cern.ch/event/1156222/contributions/5062795/

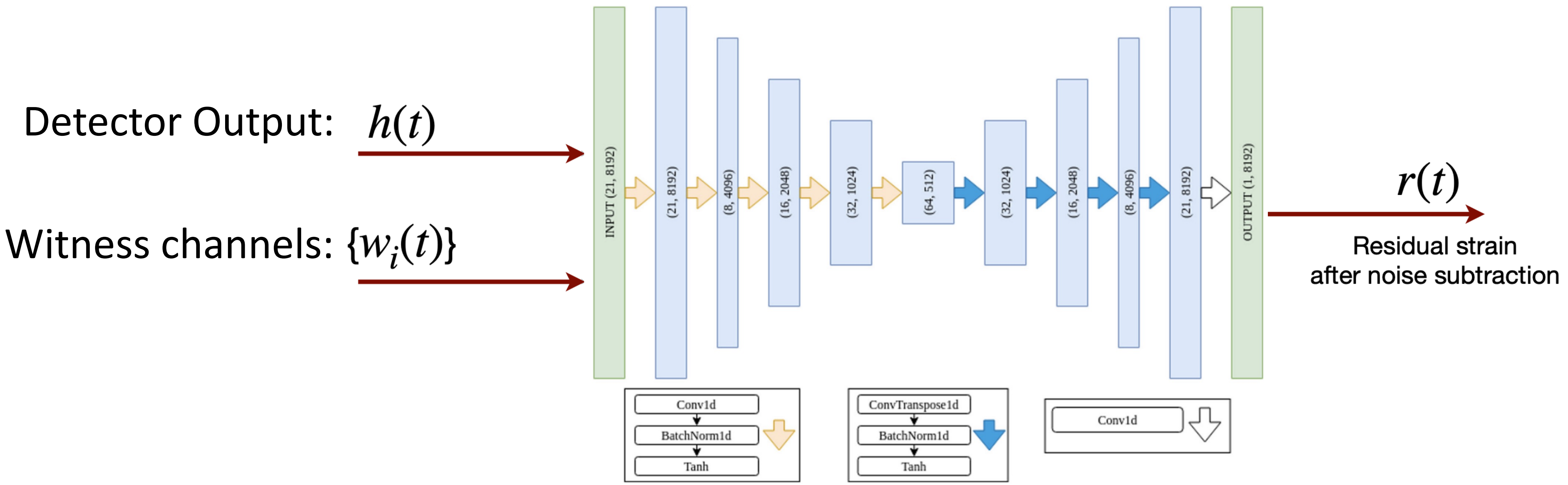
- The output reconstructed from an interferometer contains

$$h(t) = \underbrace{s(t)}_{\text{Possible GW Signal}} + \underbrace{n_w(t)}_{\text{Witness Noise}} + \underbrace{n_f(t)}_{\text{Fundamental Noise (non-removable)}}$$

- Objective: recover $s(t)$ with best possible signal-to-noise ratio by minimizing $n_w(t)$.
- Real-time/Offline noise reduction can provide quicker detections, more accurate GW parameters, find signals below the noise.

DeepClean

[Ormiston et al. (2020) (Phys.Rev.Res)]
https://indico.cern.ch/event/1156222/contributions/5062800/
https://indico.cern.ch/event/1156222/contributions/5062795/
[Bacon et al. (2022)]

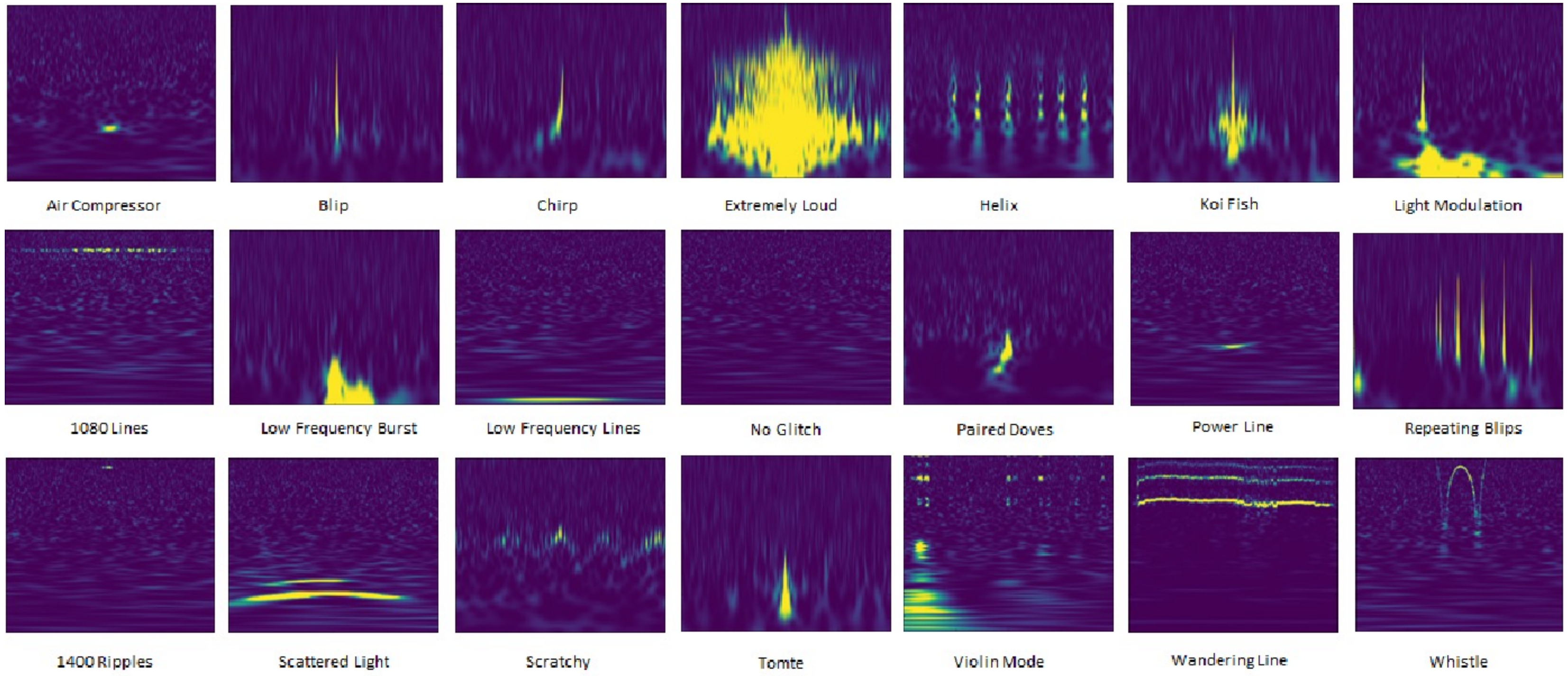


Fully convolutional auto-encoder maps the witness channels $\{w_i(t)\}$ into the noise predictions $n_w(t)$ which are then subtracted from detector output $h(t)$

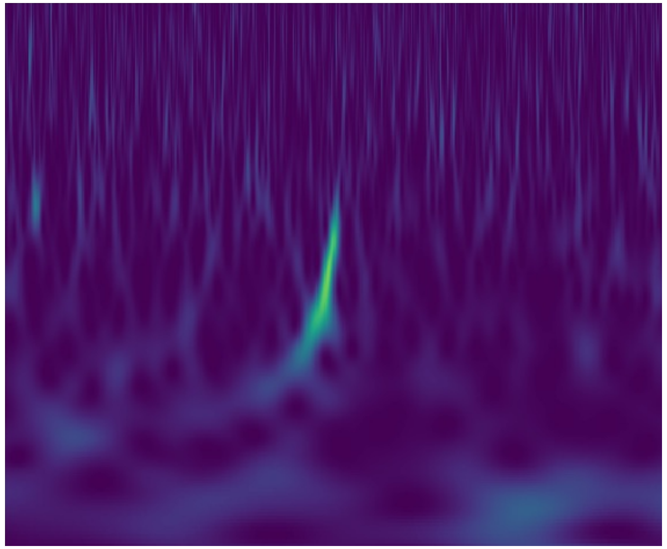
Glitches!, glitches, glitches?

https://indico.cern.ch/event/1156222/contributions/5084202/
https://madlab.cs.ucr.edu/papers/FEED_2019_paper_8.pdf

- As opposed to detector noise subtracted by DeepClean, glitches are short duration non-gaussian noise transients originated from instrumental or environmental couplings.
- Glitches are by far the culprit of most significant false alarms – excess power isn't enough!

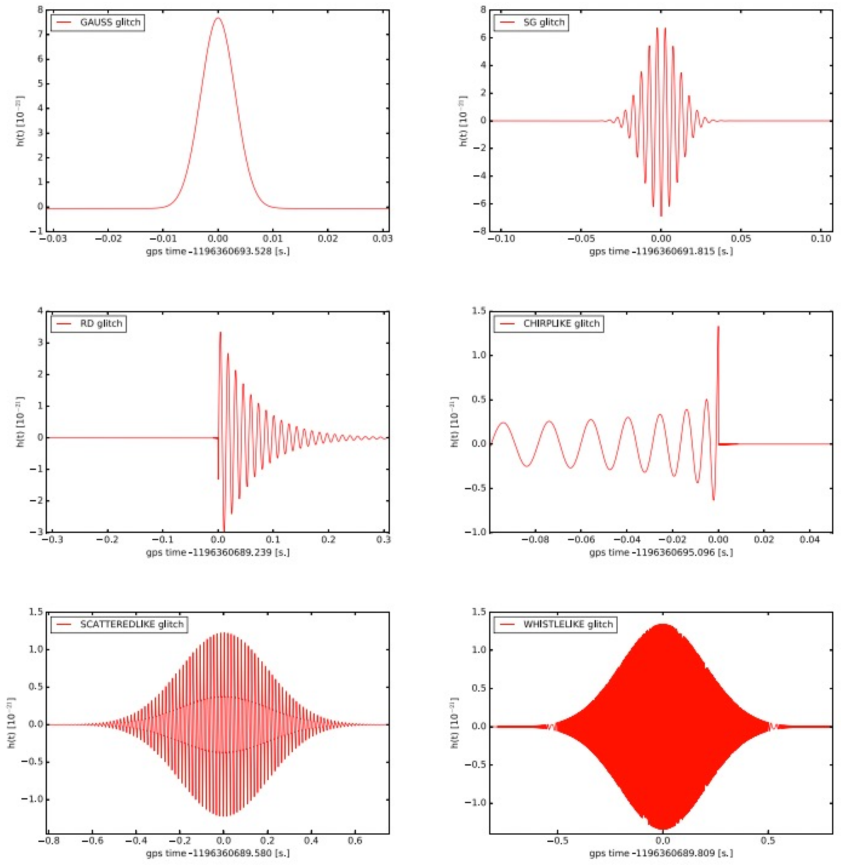


Signal

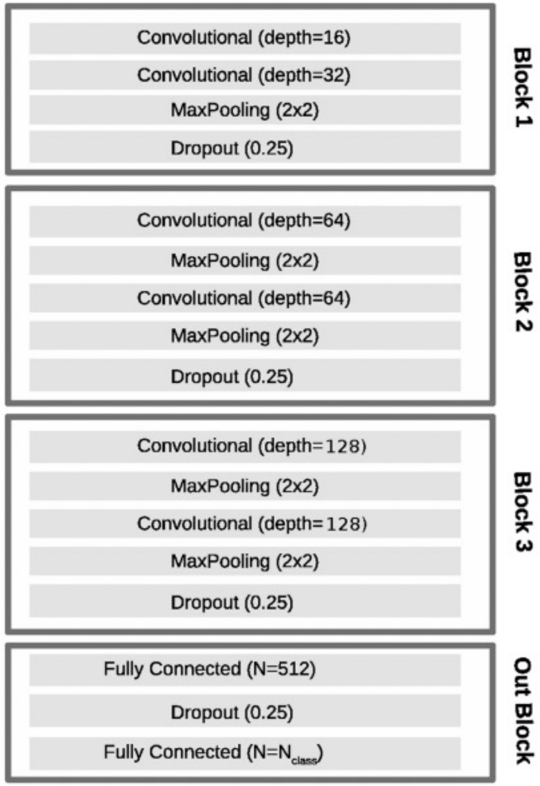


DL Glitch Classification

[George et al. (2017) (Phys.Rev.D)]
[Razzano & Cuoco (2018) (Class. Quantum Grav.)]



CNN



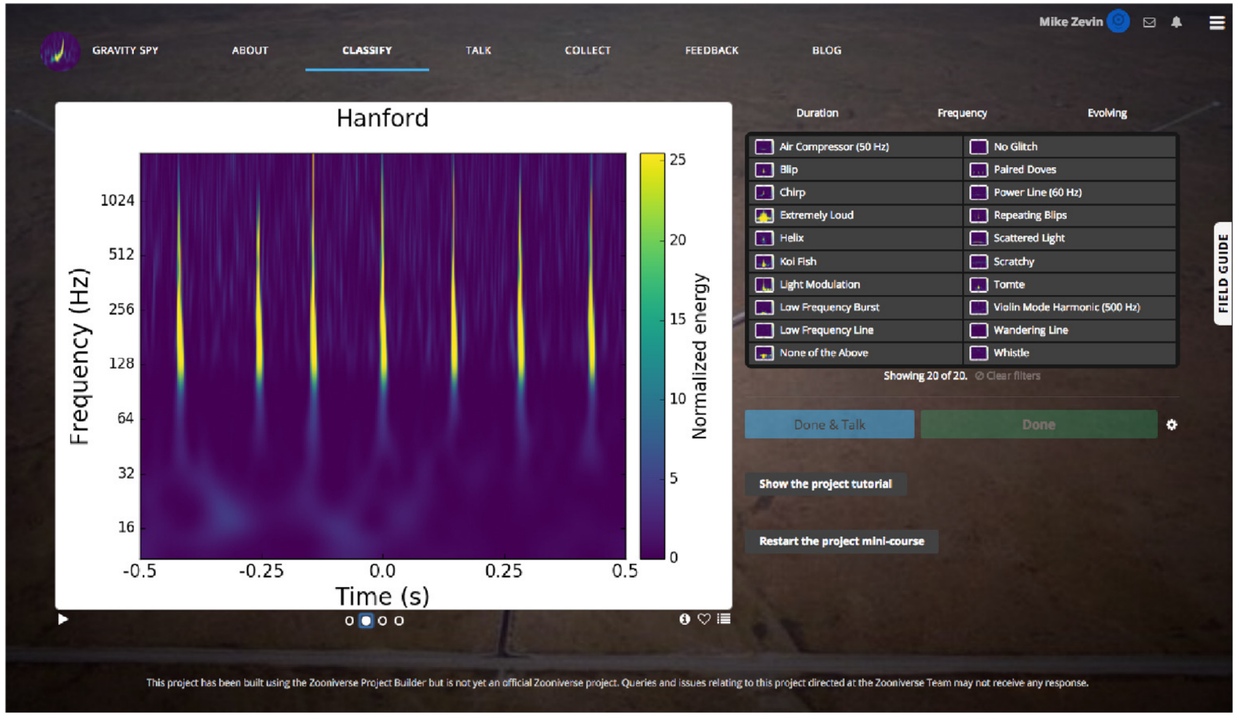
- Glitches affect data quality but do not have delayed propagation between detectors (unless by chance)
- Important to investigate classification of glitches simply to group them into families of morphologies
- Once glitch families are identified, study specific impact on data quality flags/detector performance

GravitySpy

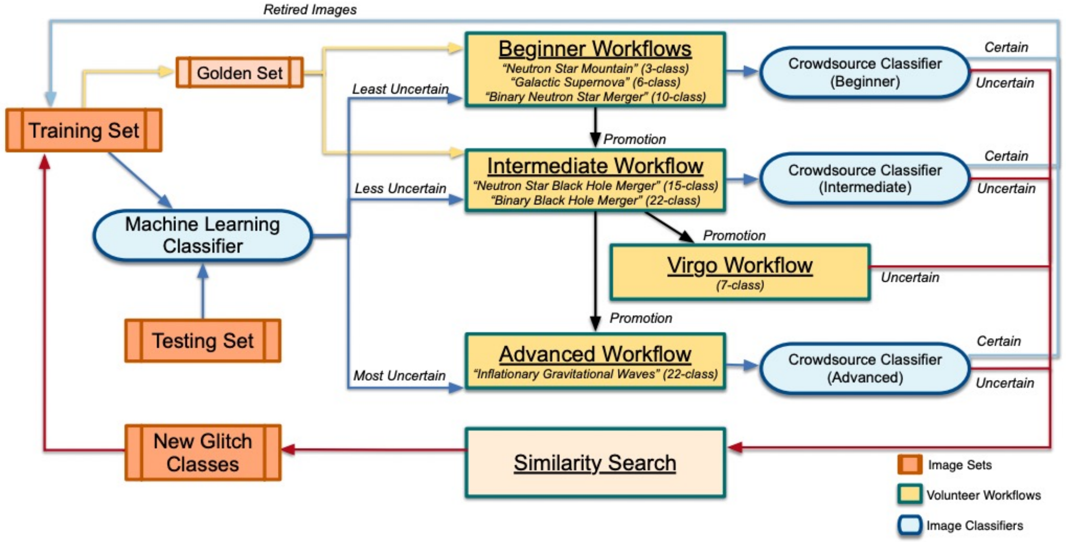
[Zevin et al. (2017) (Class. Quantum Grav.)]
[Coughlin et al. (2019) (Phys.Rev.D)]

Concept: Have humans label data (for fun) then can then be used for DL Glitch Classification
Citizen Science: https://www.zooniverse.org/projects/zooniverse/gravity-sp

Front-End



Back-End



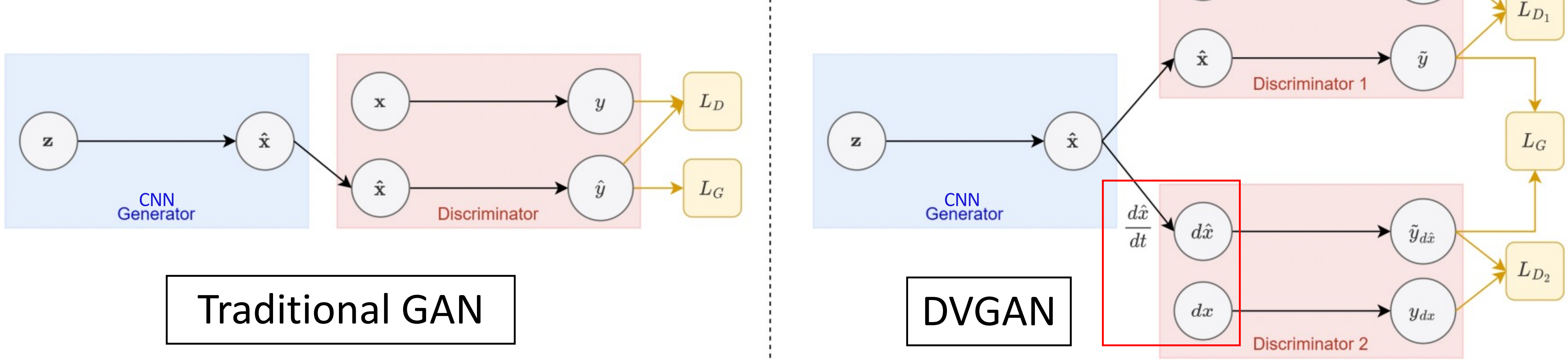
30,055 Volunteers

7,223,295 Classifications

GANs – Why not!

[Dooney et al. (2022)]

Wasserstein Distance Loss (Earth Movers Distance)



Encode human intuition - second discriminator (derivative) ensures continuous 1D signals!

Section 2: GW-Detection & Low Latency MMA

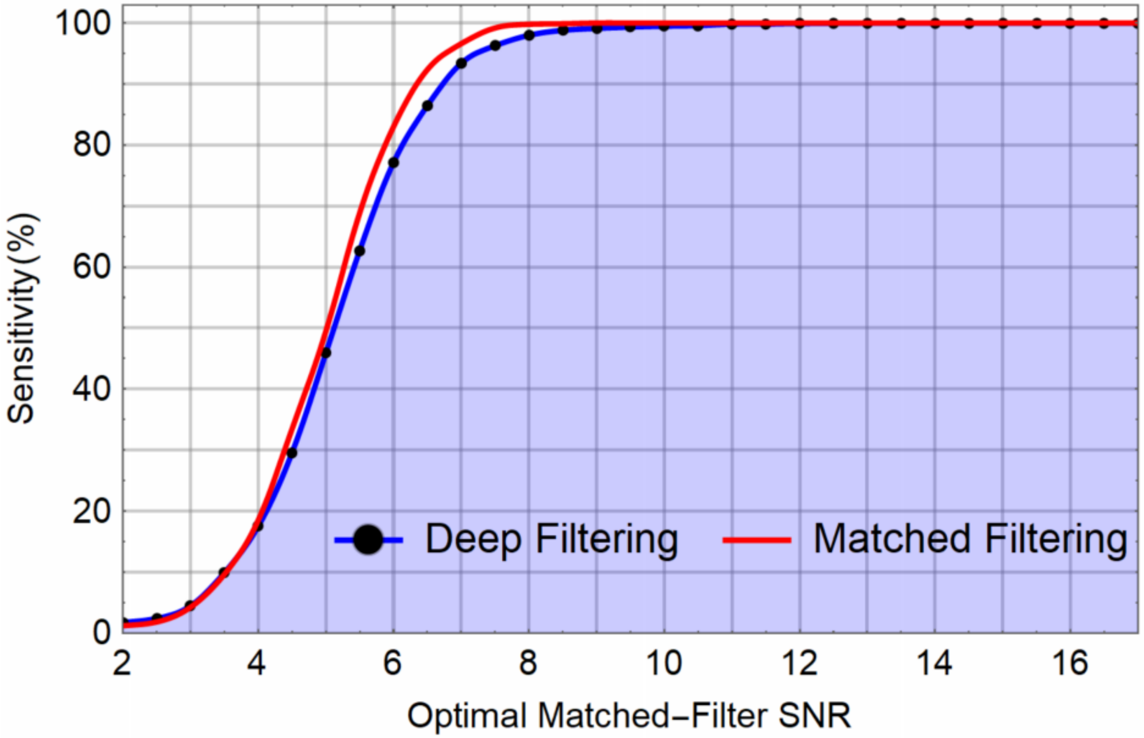
Compact Binary Searches

Matched Filtering (MF)

- **Current method** used by LIGO
- Theoretically perfect, with assumptions that do not exist in the real data
- Compares incoming GW data to bank of simulated waveforms
- Can only identify GWs that are available in GW banks (no exotic events)
- Computationally intensive!

Deep Learning

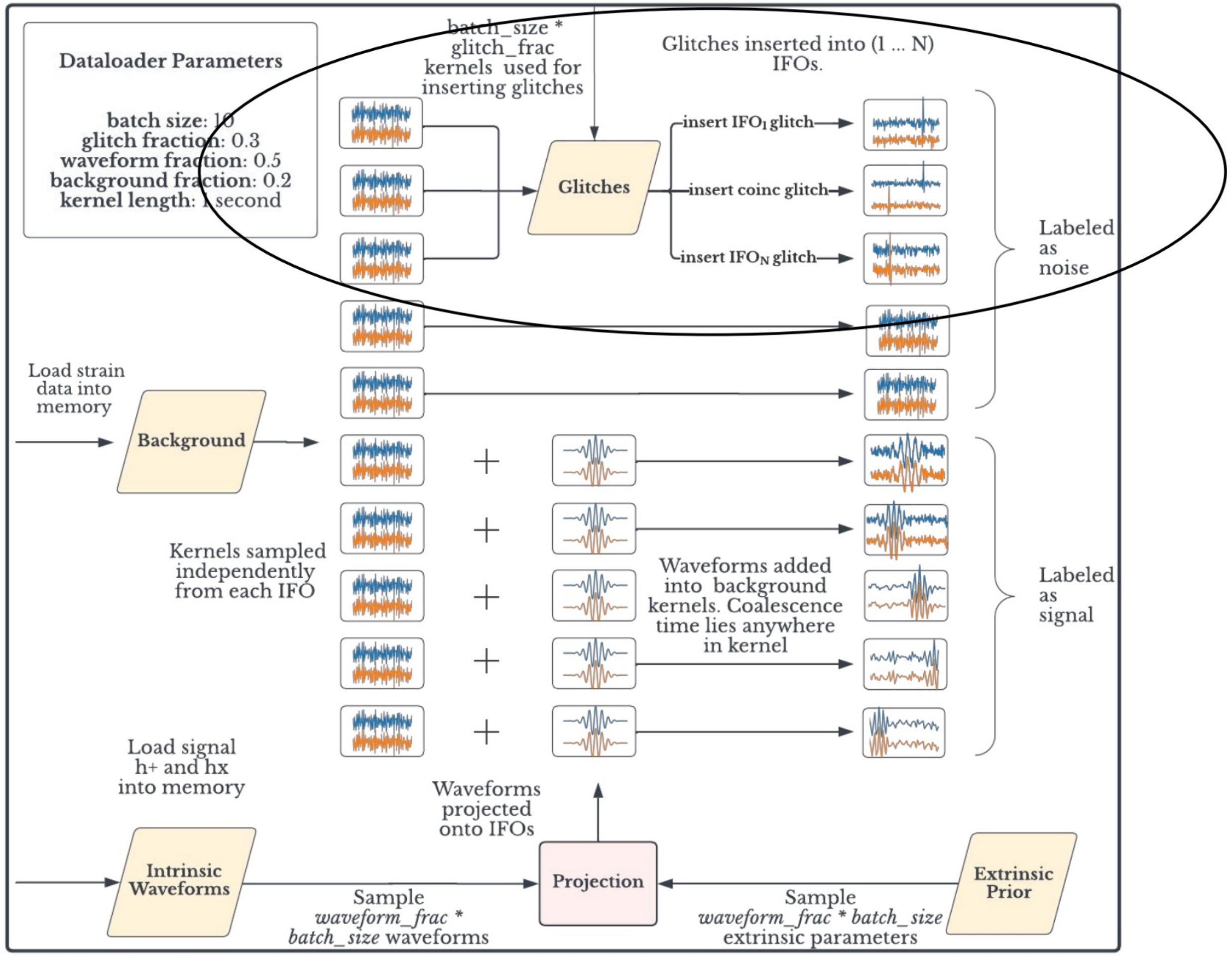
- Convolutional Neural Networks (CNNs)
- Take time-series inputs, can determine detections and estimate parameters of events
- Parallelizable
- **Performing about the same as MF – let's talk about this!**



Implementations:[George & Huerta (2017) (Phys.Lett.B)] , [Kim et al. (2015) (Class.QuantumGrav.], [Gabbard et al. (2018) (Phys.Rev.Lett)]

Complete Training Pipeline

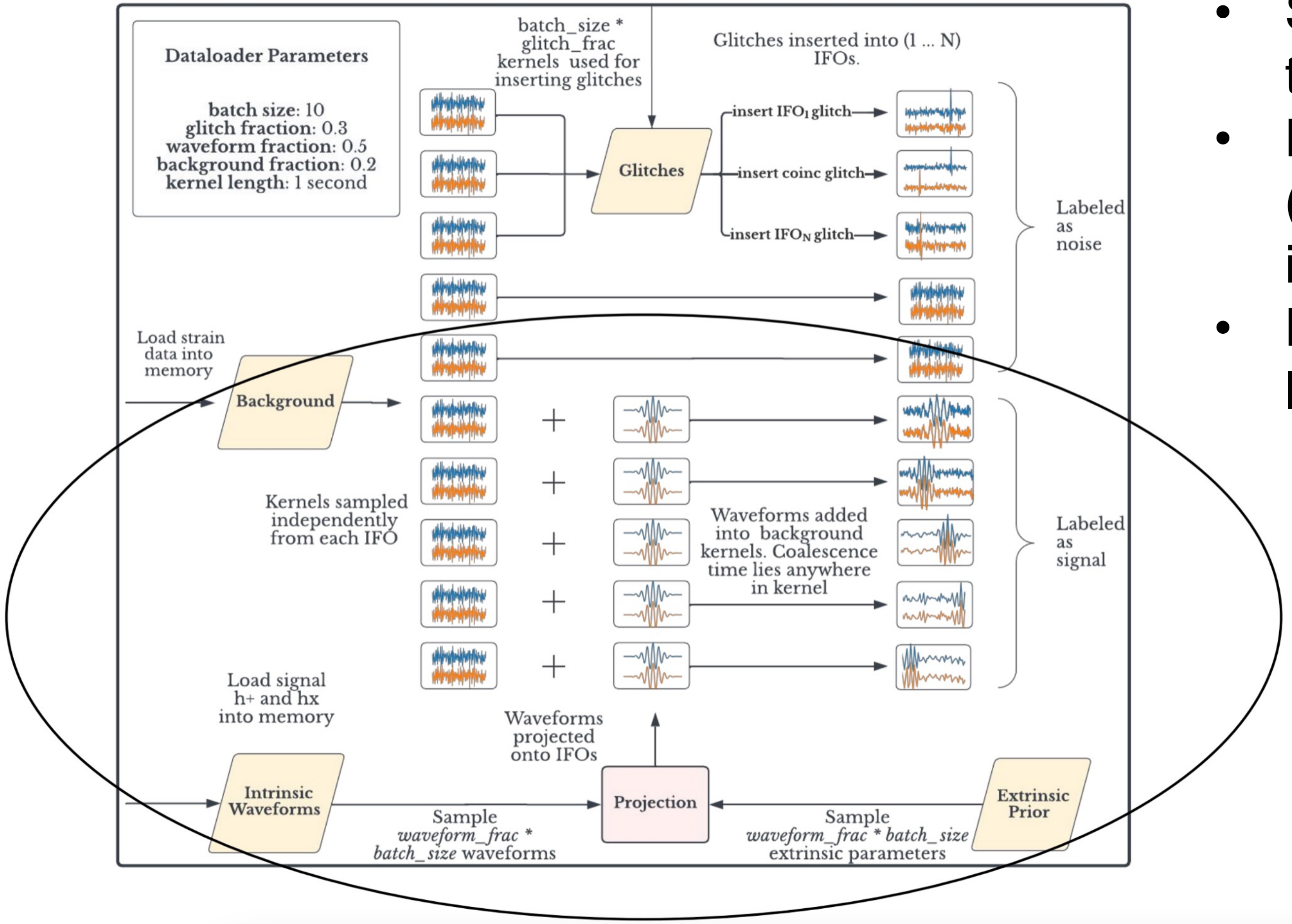
https://indico.cern.ch/event/1156222/contributions/5084202/
https://github.com/ML4GW/ml4gw



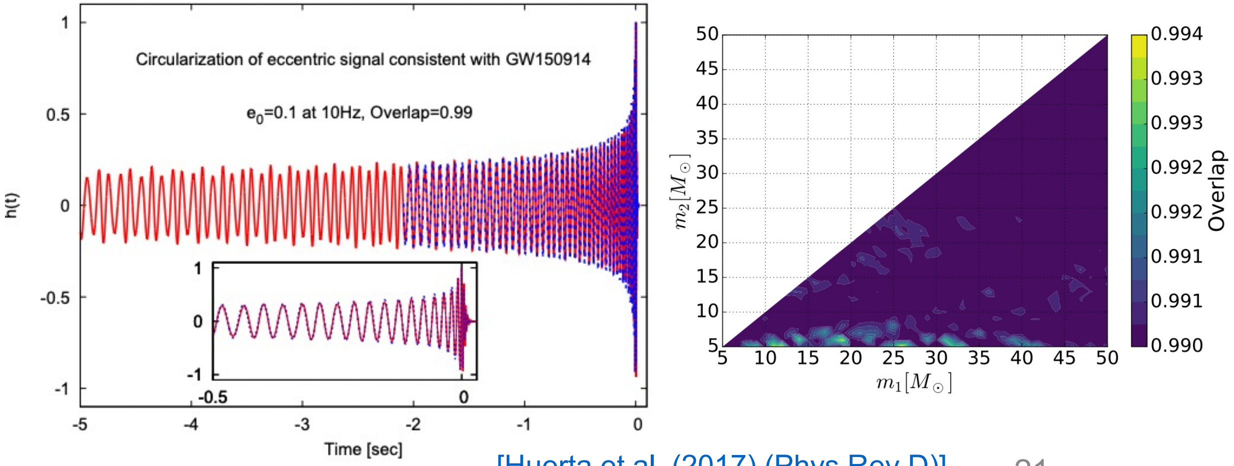
- Excess power algorithm/GravitySpy identifies glitches in training timeseries
- During training, oversample these glitches to force learning

Complete Training Pipeline

https://indico.cern.ch/event/1156222/contributions/5084202/
https://github.com/ML4GW/ml4gw



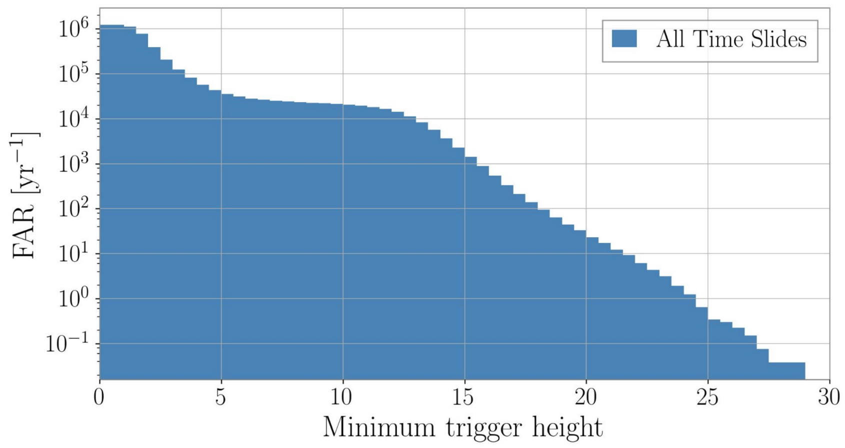
- Sample **intrinsic parameters** (masses, spins, tilts) and generate raw polarizations
- During training, sample **extrinsic parameters** (sky localization, distance) and project onto interferometers
- DL-work here as well: waveform modeling at a lower computational cost (waveform surrogates)

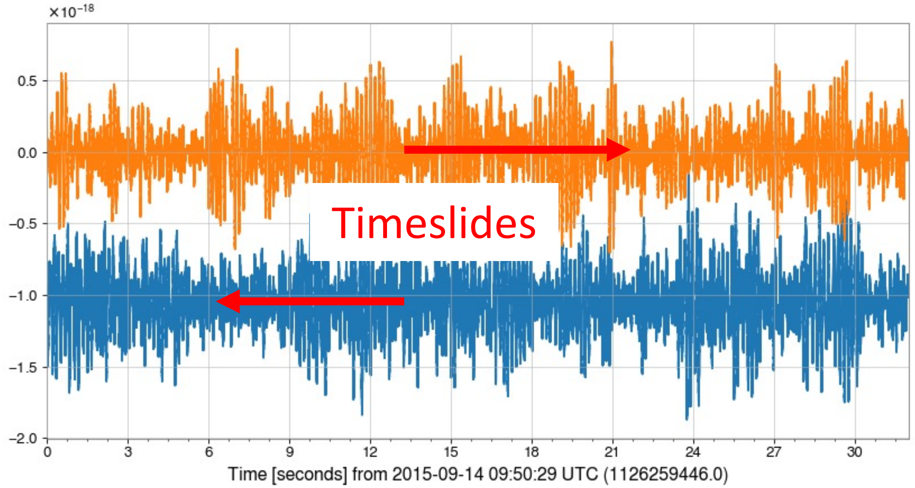


[Huerta et al. (2017) (Phys.Rev.D)]
[Blackman et al. (2017) (Phys.Rev.D)]

Scientifically Sound Validation

https://indico.cern.ch/event/1156222/contributions/5084202/
https://github.com/ML4GW/ml4gw

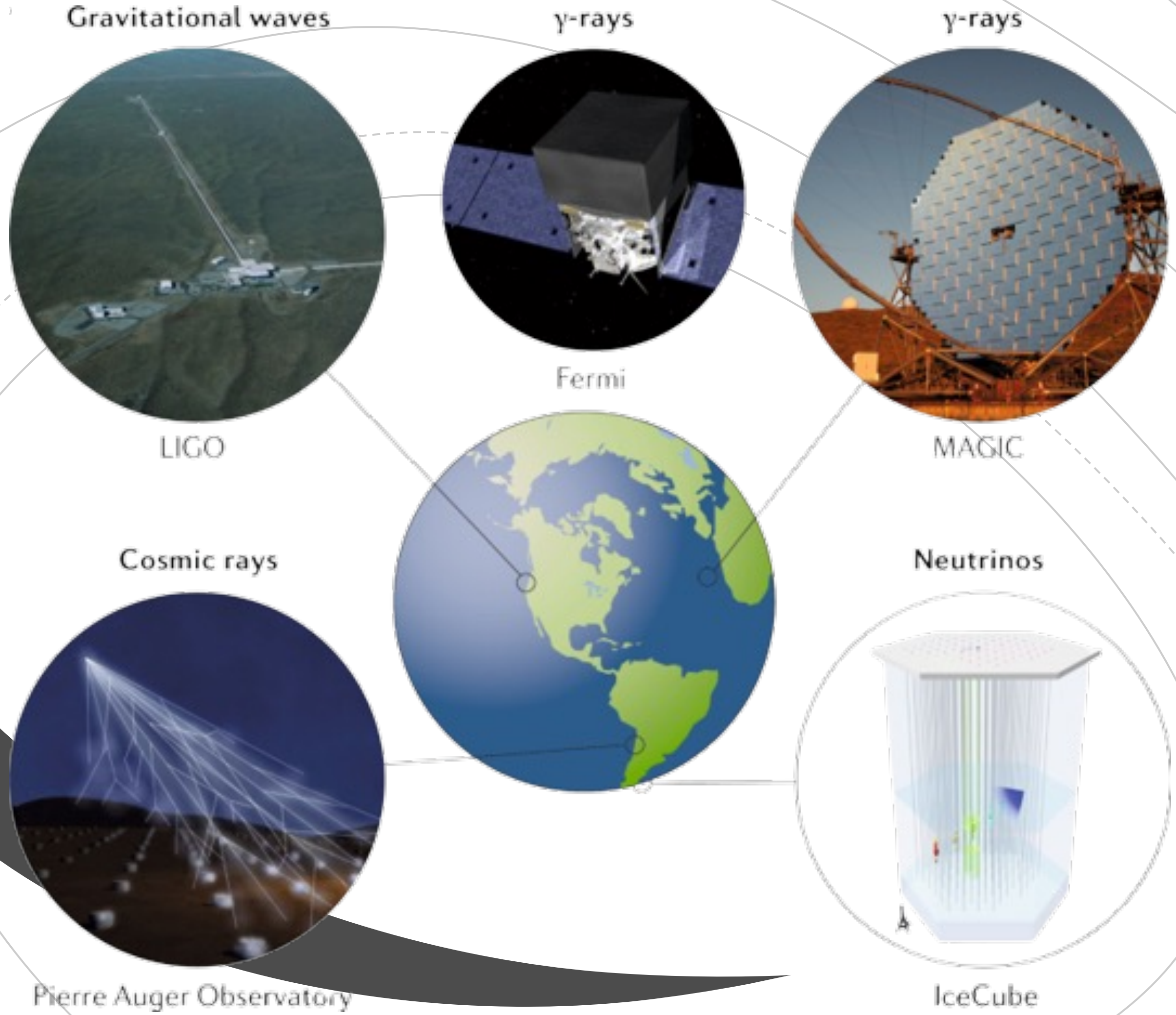




- GW search sensitivity evaluated by comparing to background events generated through **timeslides**
- Achieving high significance detections requires analyzing years of background

[GW-Science with ML Review [Cuoco et al. (2020) (MLST)]]

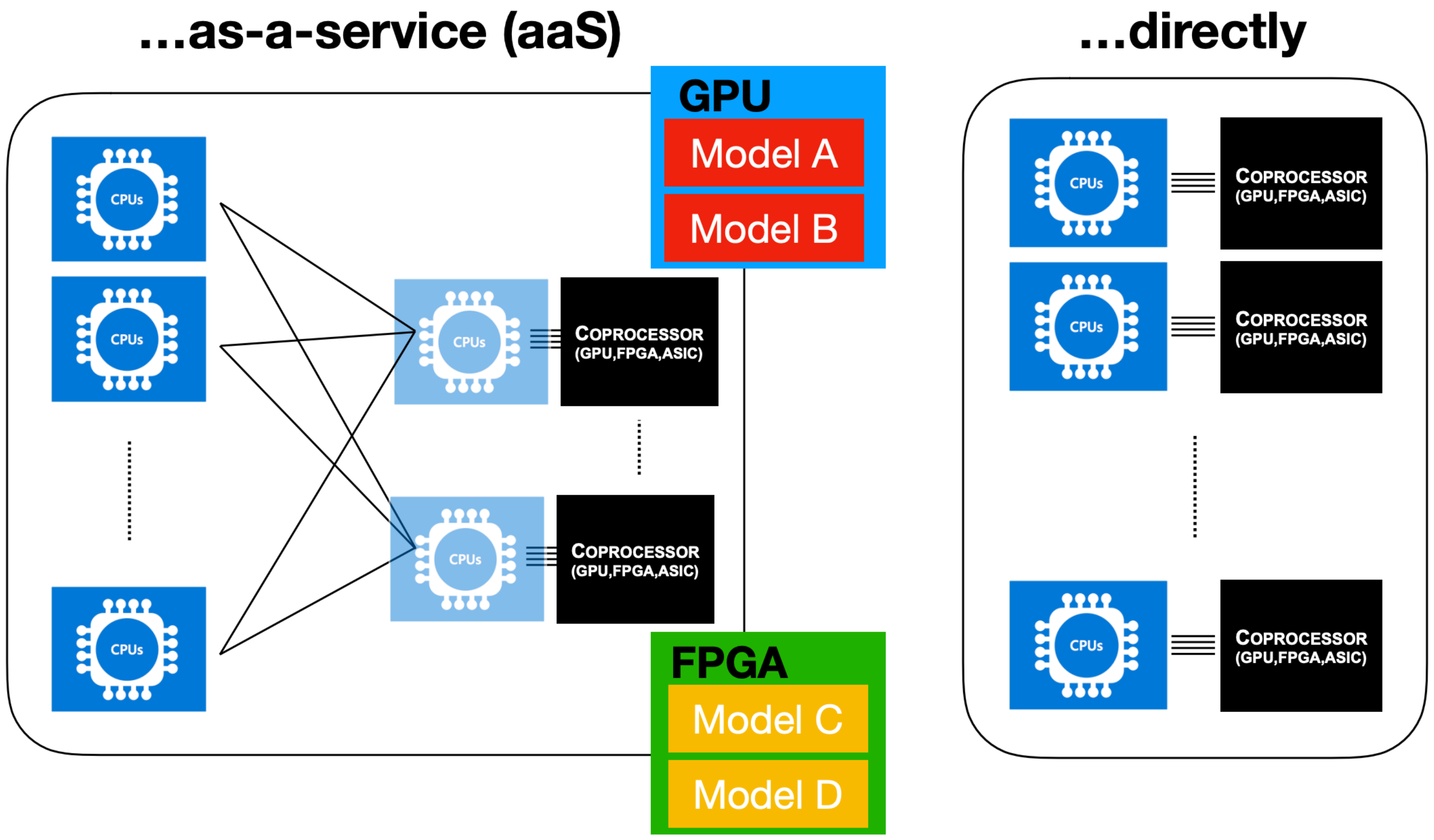
“CNN algorithms should implement an accurate statistical measure of the background to take current, ML-based CBC searches from proof-of-principle studies to production search codes.”



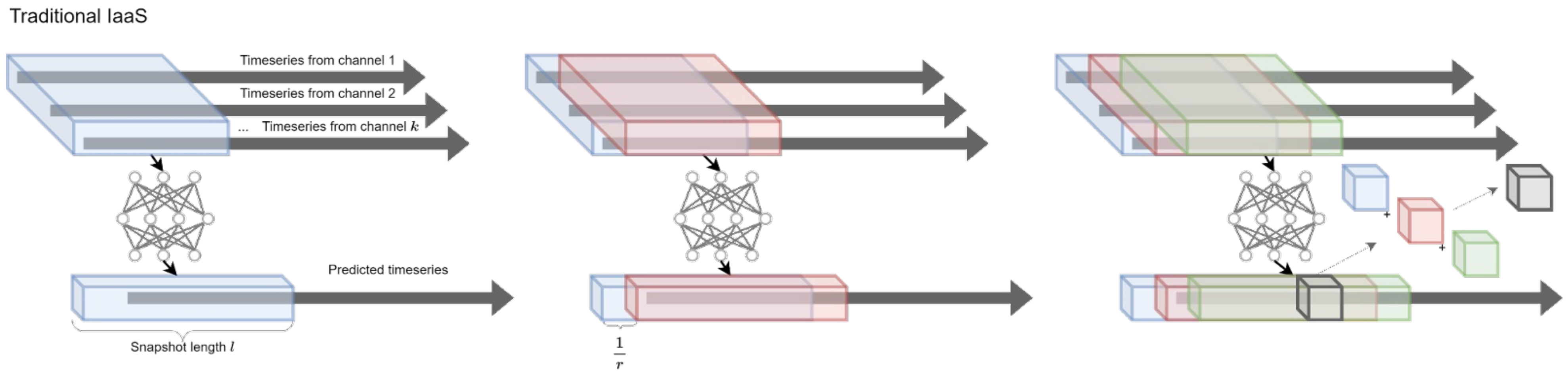
Section 2.5: Real-time Low Latency MMA

Inference-as-a-Service (IaaS)

- IaaS is becoming a common paradigm (also in HEP) to efficiently use compute resources
- Highly parallelizable
- Off-the-shelf solution: Triton inference server



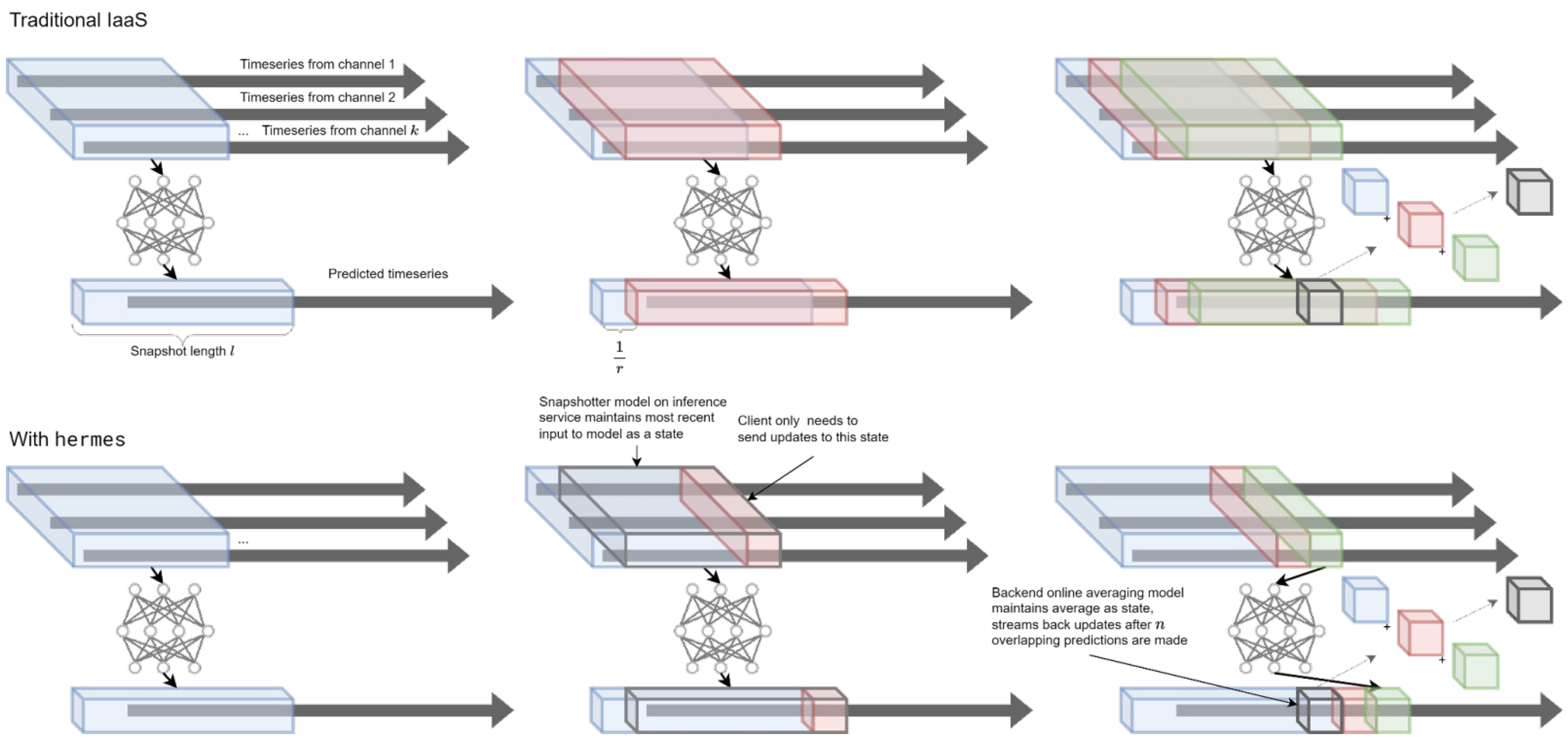
IaaS challenges timeseries streaming



HERMES – IaaS Timeseries Utility

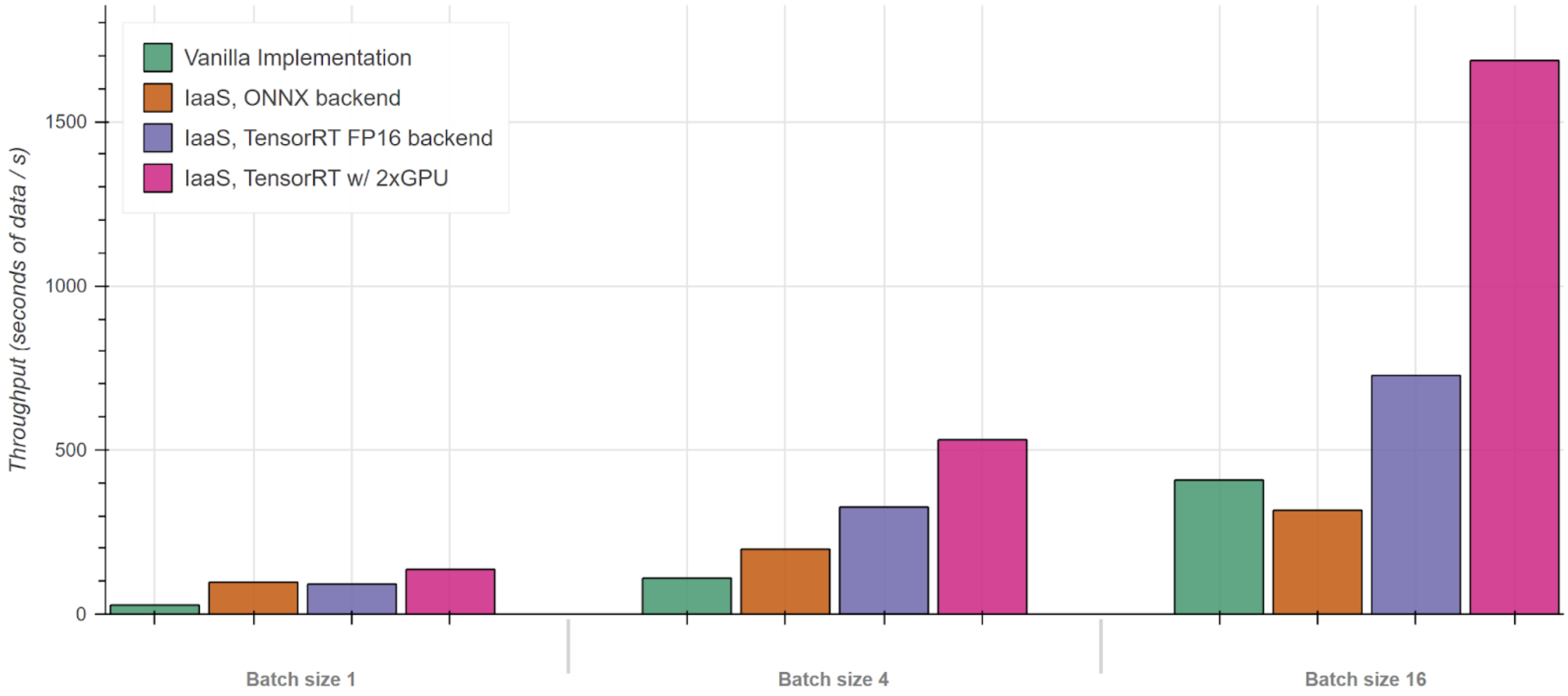
Heterogeneous Real-Time Messaging Service

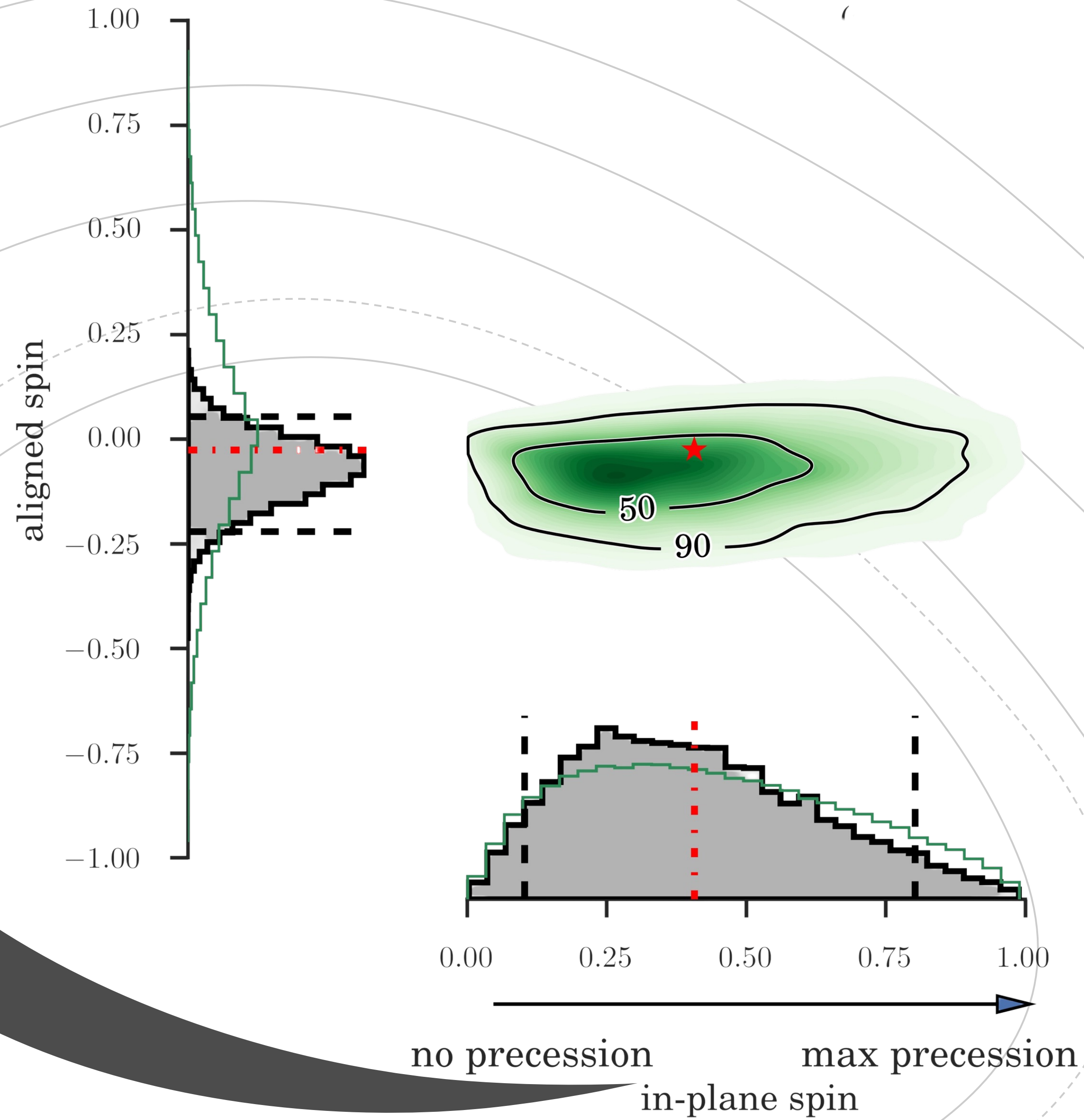
https://github.com/ml4gw/hermes



HERMES – Increased Throughput

https://github.com/ml4gw/hermes
https://alecgunny.github.io/hermes-examples/





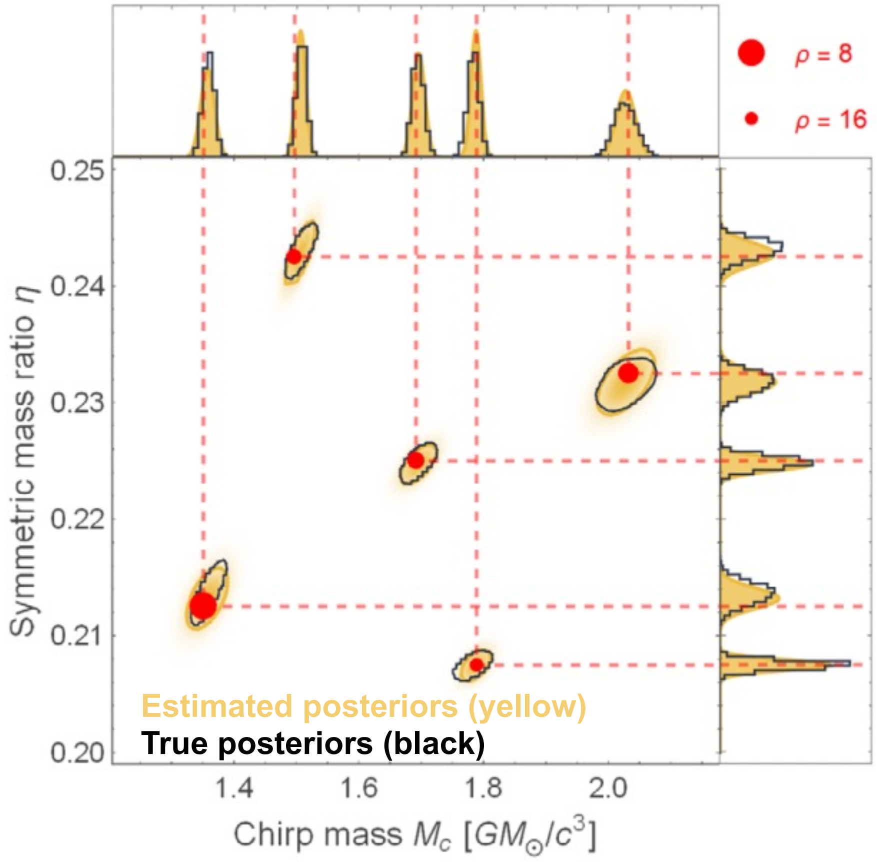
Section 3: Parameter Estimation

Bayesian Inference

[Gabbard et al. (2022) (Nature Phys.)]
[Chua & Vallisneri (2020) (Phys.Rev.Lett)]

- Understanding the astrophysics behind a signal requires parameter estimation
- Bayesian parameter estimation pipelines are active:
  - LALInference
  - Bilby
  - RIFT
- Bayesian evidence is computational costly - takes days or weeks to measure GW signal parameters

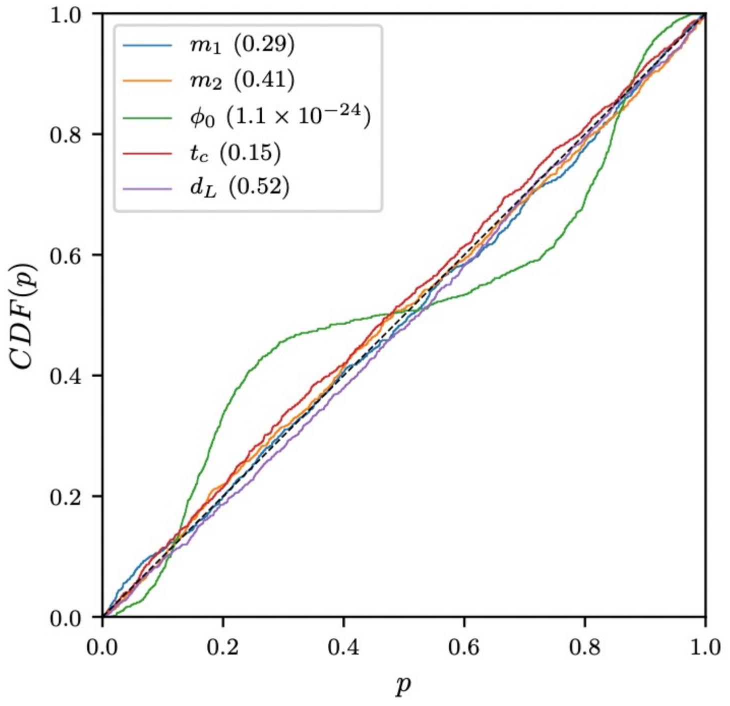
Reproduce GW Bayesian posterior using NN



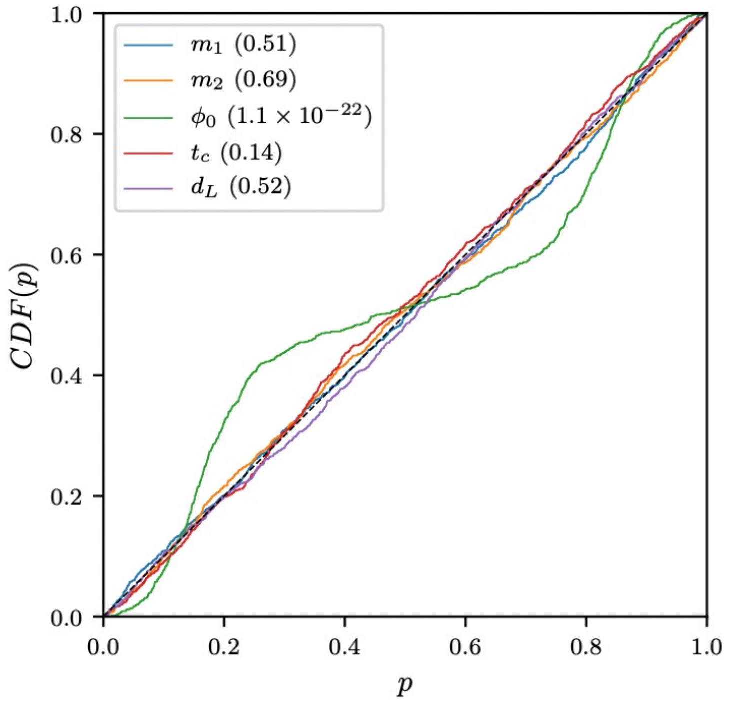
Bayesian Posterior – Normalizing Flows

[Green et al. (2020) (Phys.Rev.D)]
[Dax et al. (2021) (Phys.Rev.Lett)]

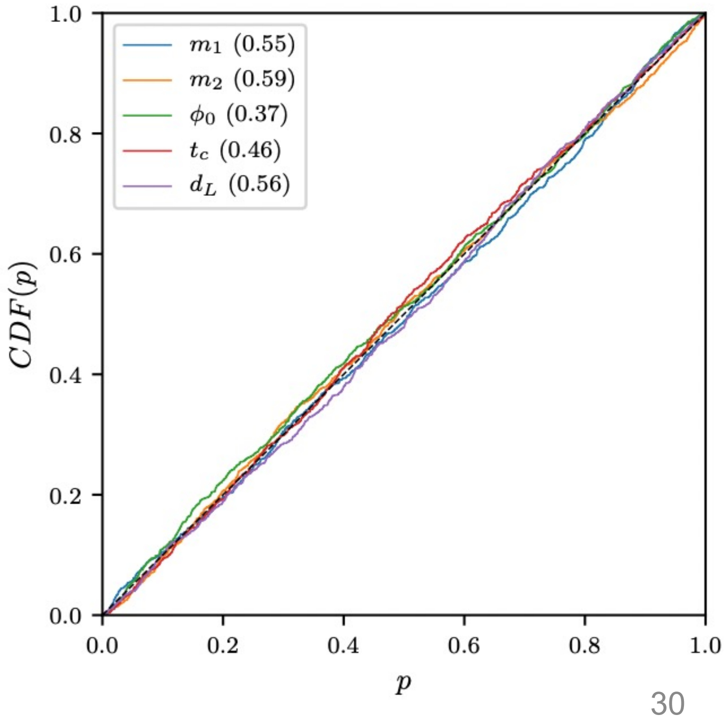
Convolutional Variational Autoencoder (CVAE)



Masked Autoregressive Flows



CVAE + Autoregressive Flows



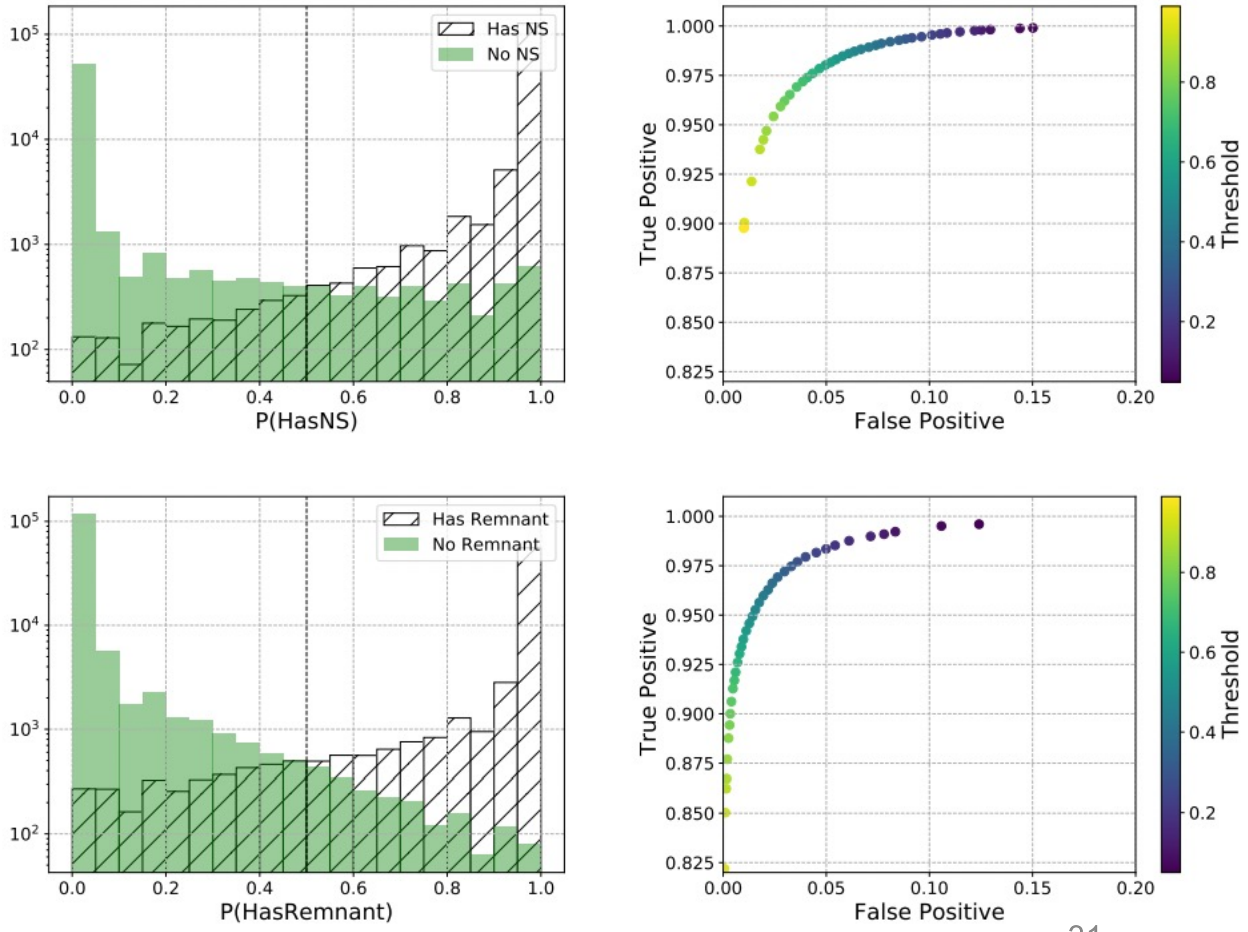
Low Latency EM-Bright

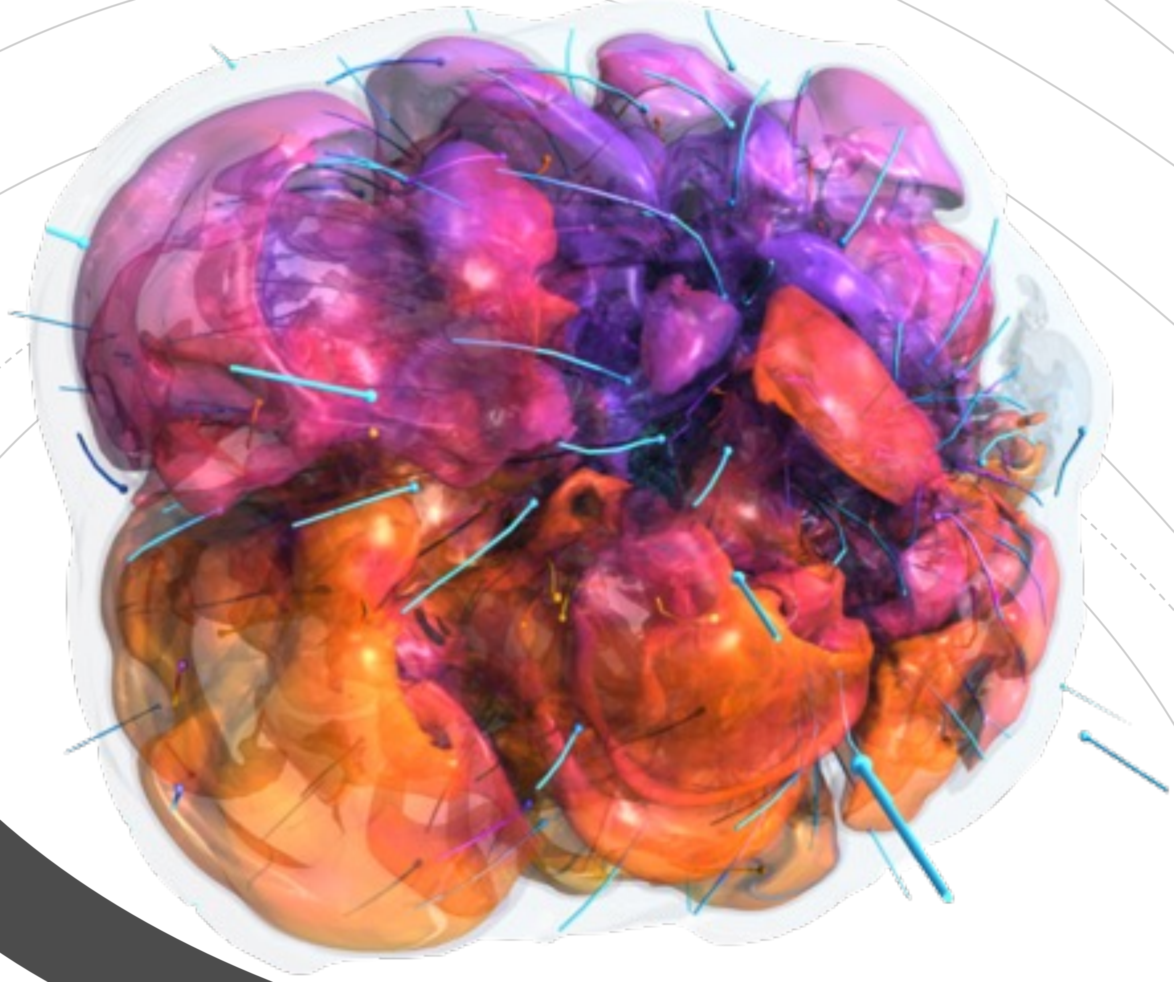
[Chatterjee et al. (2020) (Astrophys.J.)]

Beyond classification, source property inference EM-bright is provided to MMA community:

1. Probability that CBC system contains NS of mass less than 3.0 $M_{\odot}$ - **P(HasNS)**
2. Probability that the final coalesced object is surrounded by tidally-disrupted matter after the merger - **P(HasRemnant)**

Use K-neighbors-neighbors (KNN) vote on fast matched filter searches





Section 4: Anomaly Detection

What are anomalies?

Anomalies are unmodelled waveforms

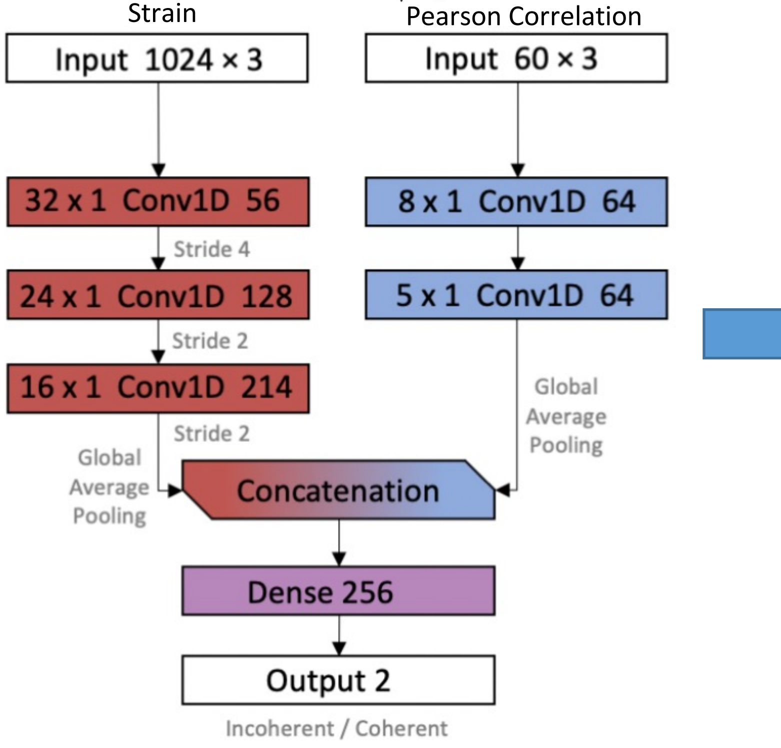
- GWs have been detected in Matched-Filter pipelines from:
  - Binary Black Holes (BBH)
  - Binary Neutron Stars (BNS)
  - Black Hole – Neutron Star (BHNS)
- Anomalous signals:
  - Core Collapse Supernovas (CCSNe)
  - Neutron Star Glitches
  - Cosmic Strings
  - NSs collapsing to BHs
  - Gravitational Bremsstrahlung
  - Other stochastic processes

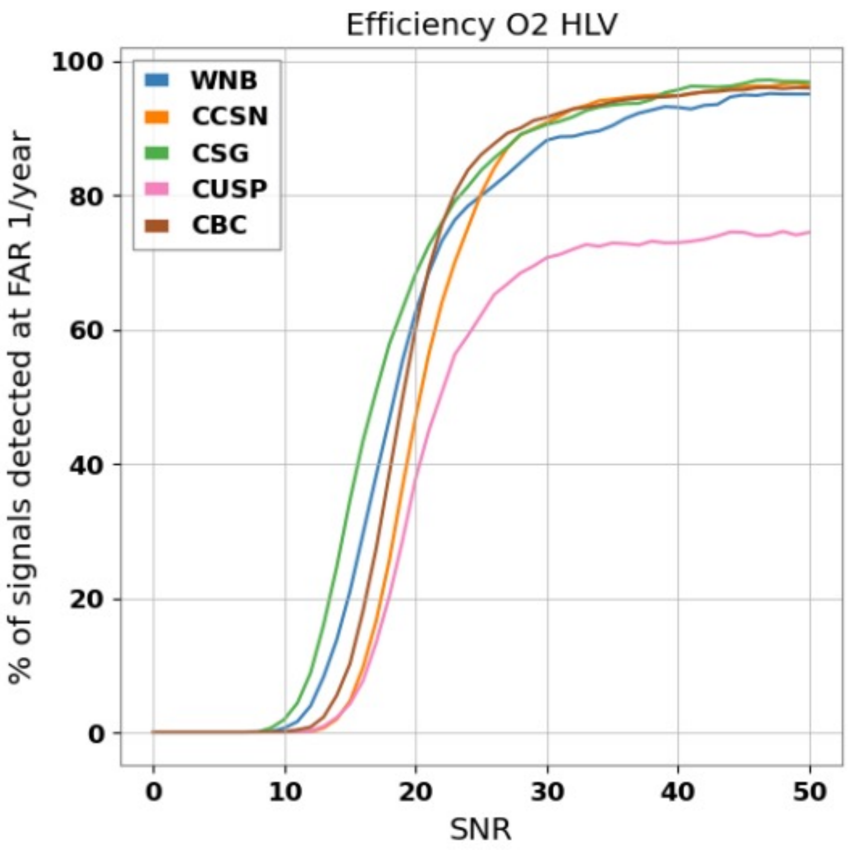
Correlation Detection

[Vasileios et al. (2022)]

Idea: The one thing we can count on is inter-detector correlation

$$r_{\alpha\beta}[n] = \frac{\sum_{i=1}^{N}(d_\alpha[i]-\bar{d}_\alpha)(d_\beta[i+n]-\bar{d}_\beta)}{\sqrt{\sum_{j=1}^{N}(d_\alpha[j]-\bar{d}_\alpha)^2\sum_{k=1}^{N}(d_\beta[k]-\bar{d}_\beta)^2}}$$

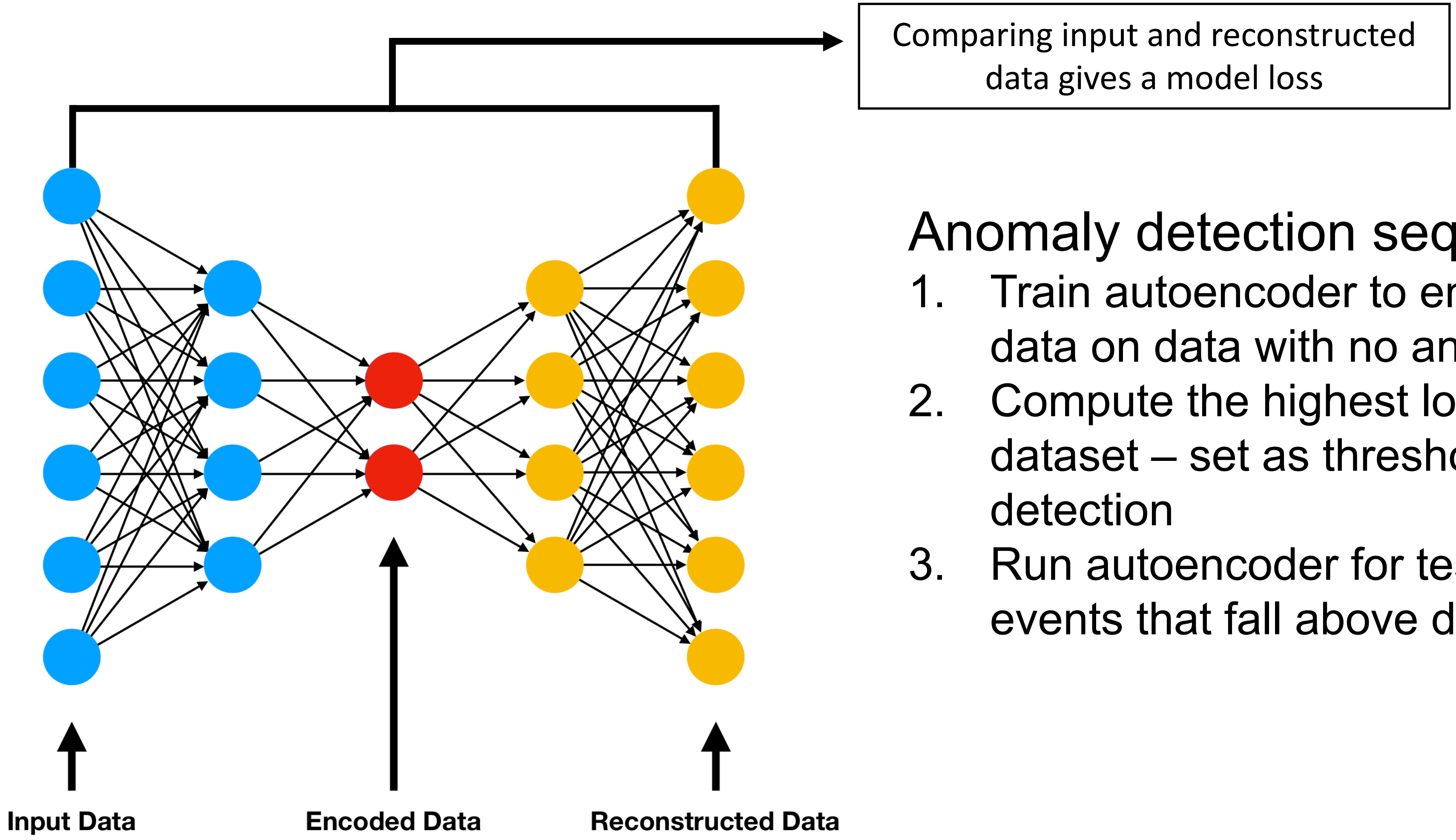




Trained on generic white-noise bursts (WNBs) as signal

Unsupervised Learning: Detection

[Moreno et al. (2022) (MLST)]



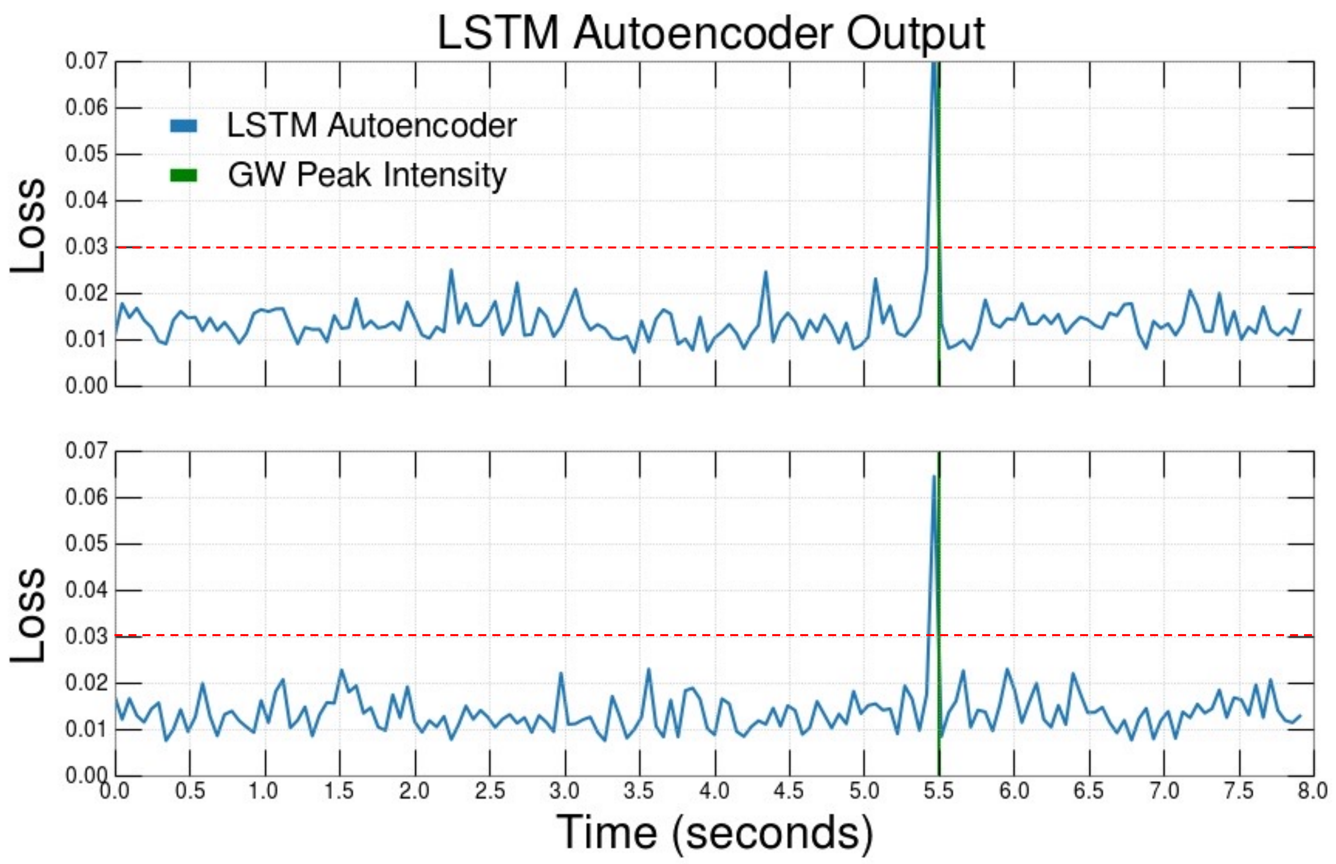
Anomaly detection sequence:

1. Train autoencoder to encoder and decode data on data with no anomalies.
2. Compute the highest loss on the training dataset – set as threshold for anomalous detection
3. Run autoencoder for test data, identify events that fall above detection threshold

Event Loss with Autoencoders

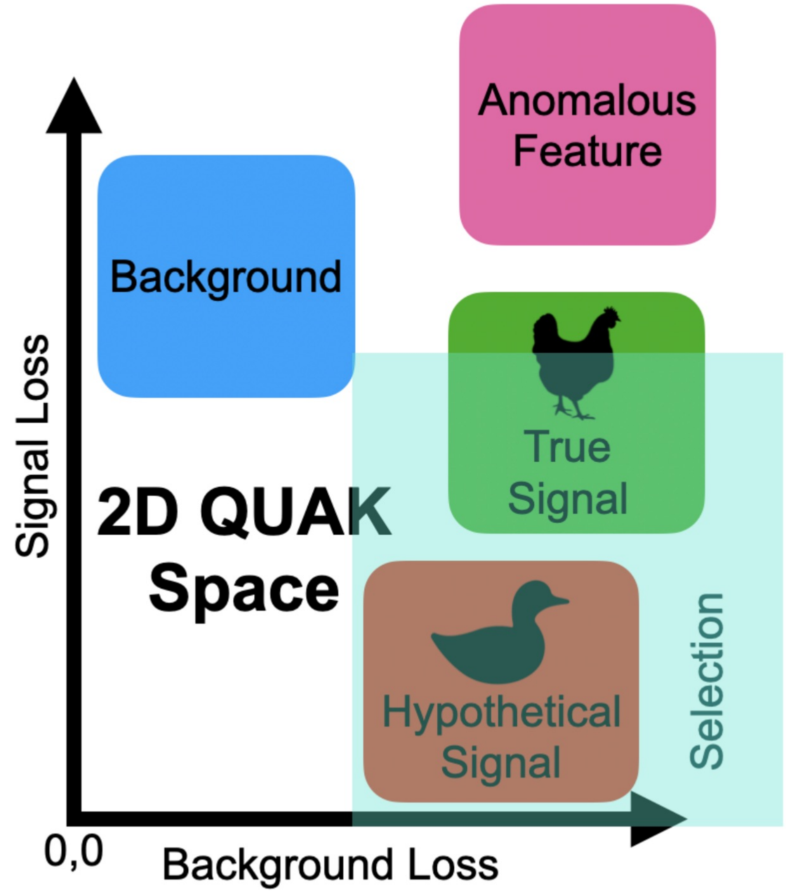
[Moreno et al. (2022) (MLST)]

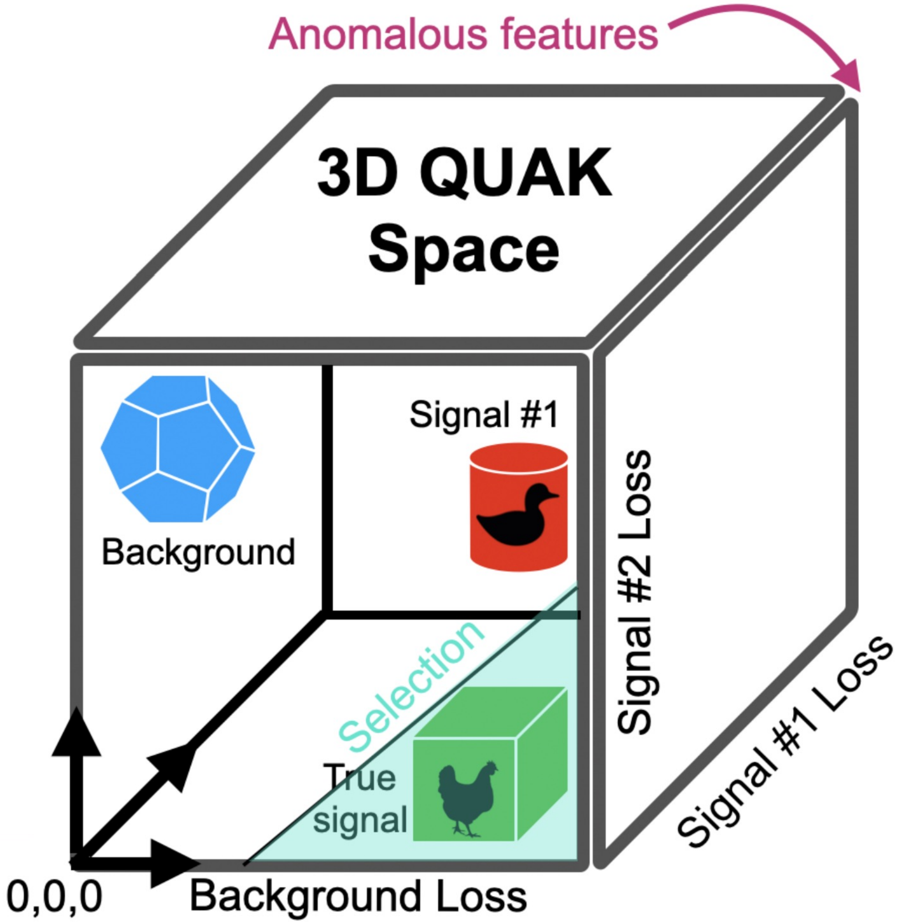
- LSTM (and now Transformer) AE evaluated on BBH and BNS events yields promising results
- **Red dotted line represents detection threshold** which can be determined according to FPR
- During training, **AE never receives information about any GW (signal) -> Source Agnostic**



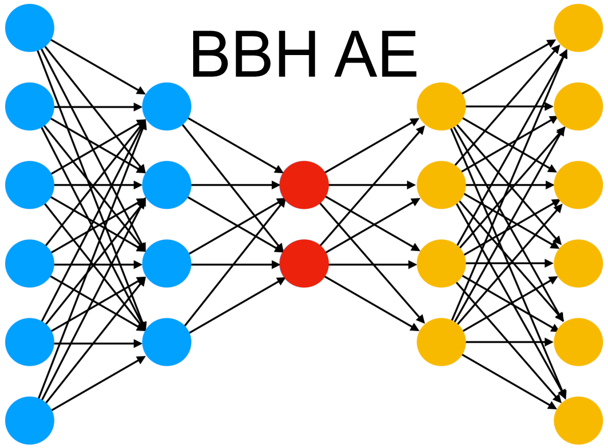
Quasi-Anomalous Knowledge - QUAK

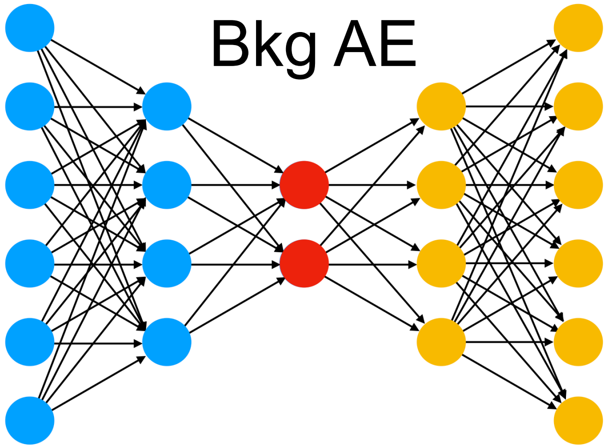
[Park et al. (2021) (JHEP)]

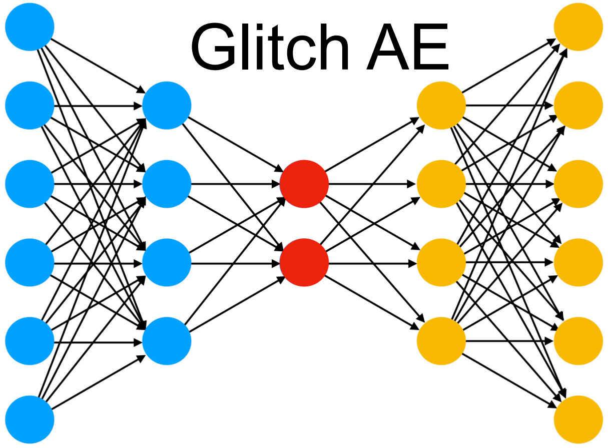


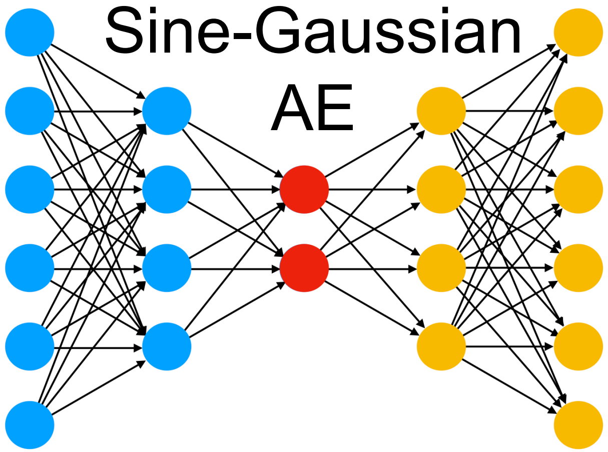


4D QUAK Space

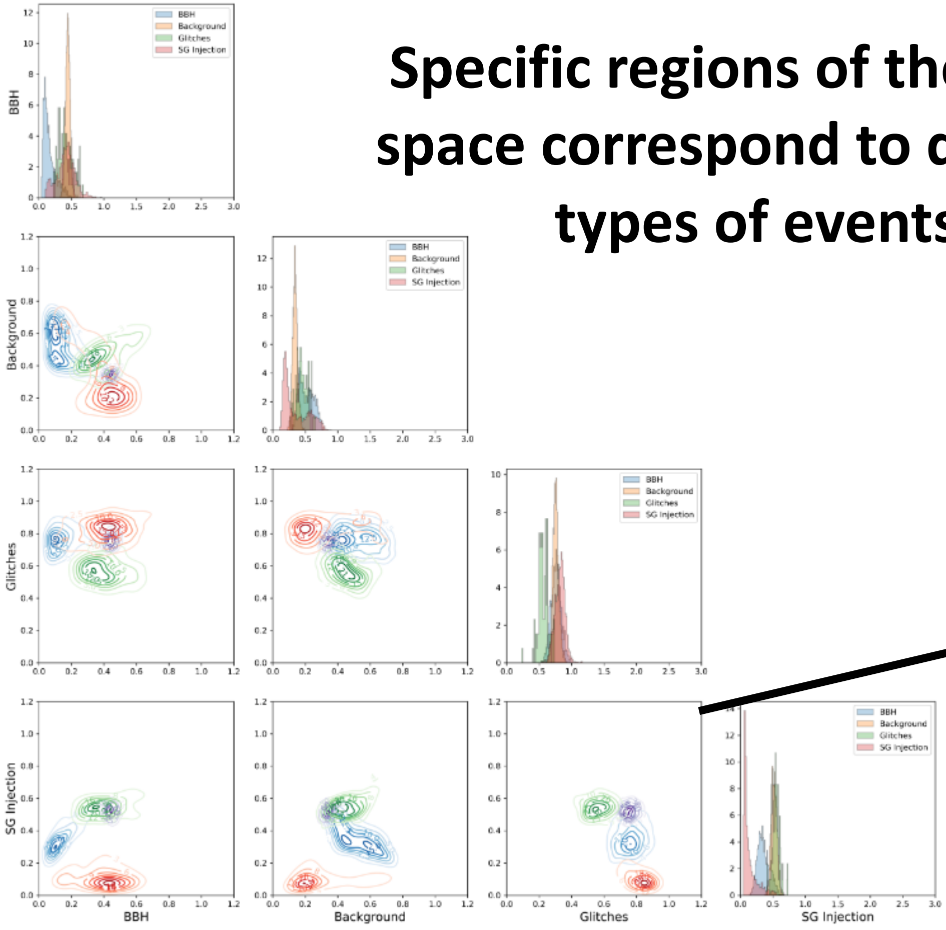




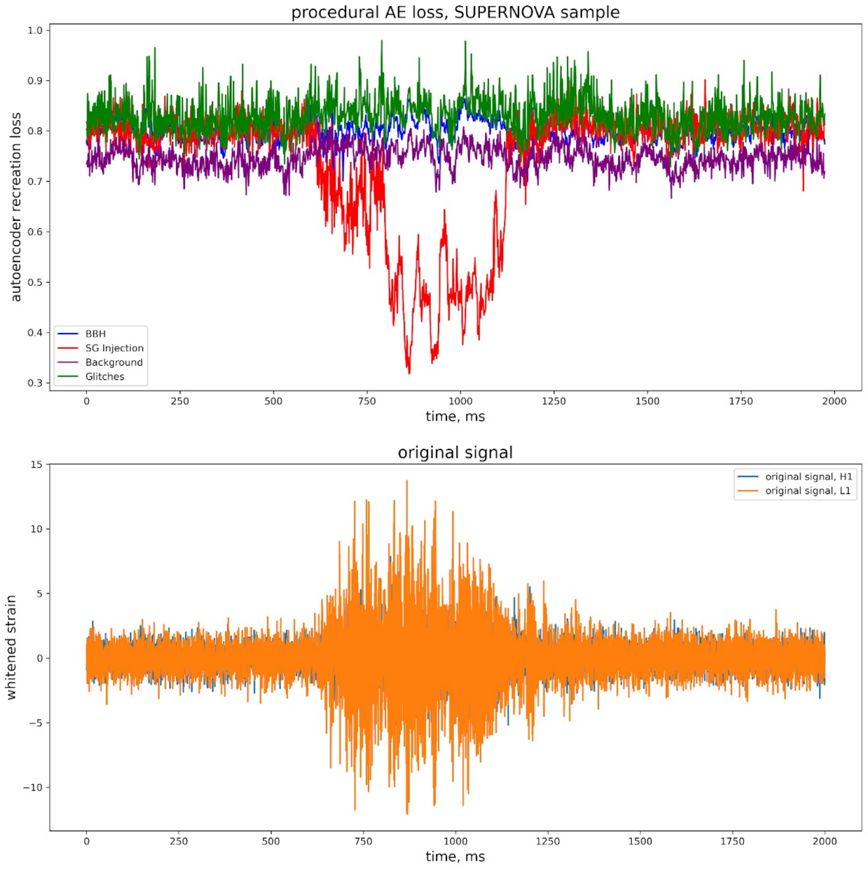




Trained QUAK Spaces



Specific regions of the QUAK space correspond to different types of events!



Also interested in other embedded spaces…..

Conclusions

Check out: https://iphysresearch.github.io/Survey4GWML/ for full paper lists

ML in GW-Physics is alive and well:

- Data Quality
- GW Detection, Real-time low latency MMA
- Parameter Estimation
- Anomaly Detection

From A3D3, keep a look out for:

1. **ML4GW Workshop @ MIT (January 2023)**
2. **Mock-data Challenge**

a3d3 Accelerated AI Algorithms for Data-Driven Discovery

Backup

ML-Revolution in GW-physics

Some tools being developed in our org: https://github.com/ML4GW

Similar to HEP, AI can play an important role in real-time and offline data processing:

1. Data Cleaning
   - Non-linear noise subtraction using autoencoders (DeepClean)
2. GW detection, parameter estimation
   - End-to-end optimized, validated, analysis-ready models (BBHNet)
   - Timeseries tools (Hermes, ML4GW)
3. Real-time low latency MMA
   - Parallel processing in Inference-as-a-Service pipelines (IaaS)
4. Anomaly Detection
   - Autoencoders, embedded spaces to detect outlying events

Different Kinds of Noise

https://arxiv.org/pdf/2005.06534.pdf
https://indico.cern.ch/event/1156222/contributions/5062800/
https://indico.cern.ch/event/1156222/contributions/5062795/

- Noise can be split into two kinds:

$$\mathrm{n}(t) = n_{nw}(t) + n_{w}(t)$$

- Non-removable (fundamental noise)
- Budgeted by system design
- Eg: photon shot noise, thermal noise
- Can only be reduced with upgraded design and technology

- Source of noise witnessed by dedicated system monitors (witness sensors)
- Environmental contamination or technical noise

DeepClean

https://arxiv.org/pdf/2005.06534.pdf
https://indico.cern.ch/event/1156222/contributions/5062800/
https://indico.cern.ch/event/1156222/contributions/5062795/
https://arxiv.org/abs/2205.13513

Detector Output: $h(t)$

Witness channels: $\{w_i(t)\}$

$$n_w(t) = \mathscr{F}(\{w_i\}; \vec{\theta})$$

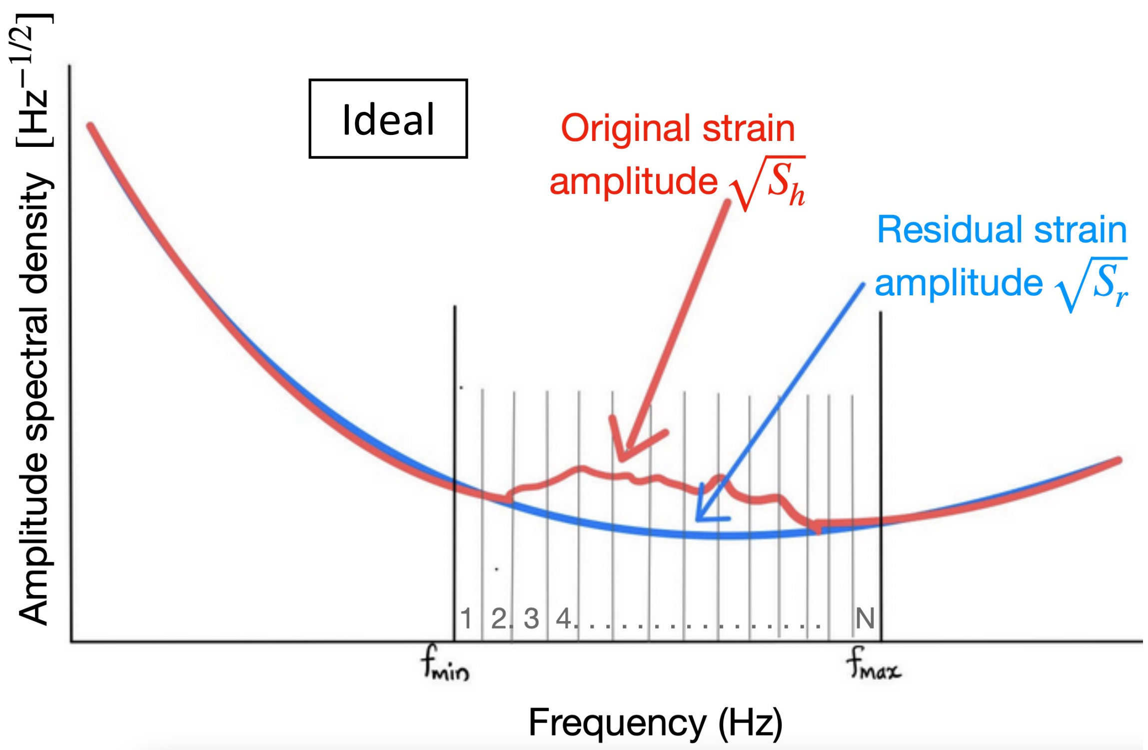
$$r(t) = h(t) - n_w(t)$$

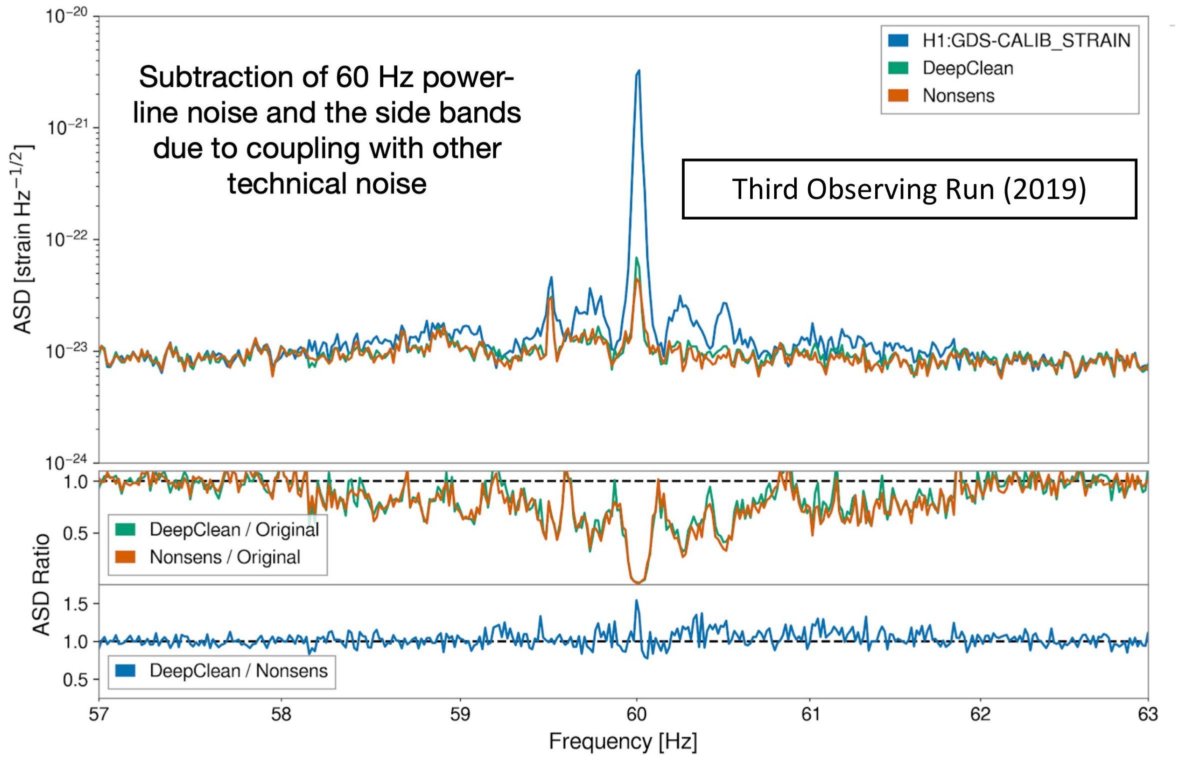
$r(t)$

Residual strain
after noise subtraction

DeepClean Performance

https://arxiv.org/pdf/2005.06534.pdf
https://indico.cern.ch/event/1156222/contributions/5062800/
https://indico.cern.ch/event/1156222/contributions/5062795/

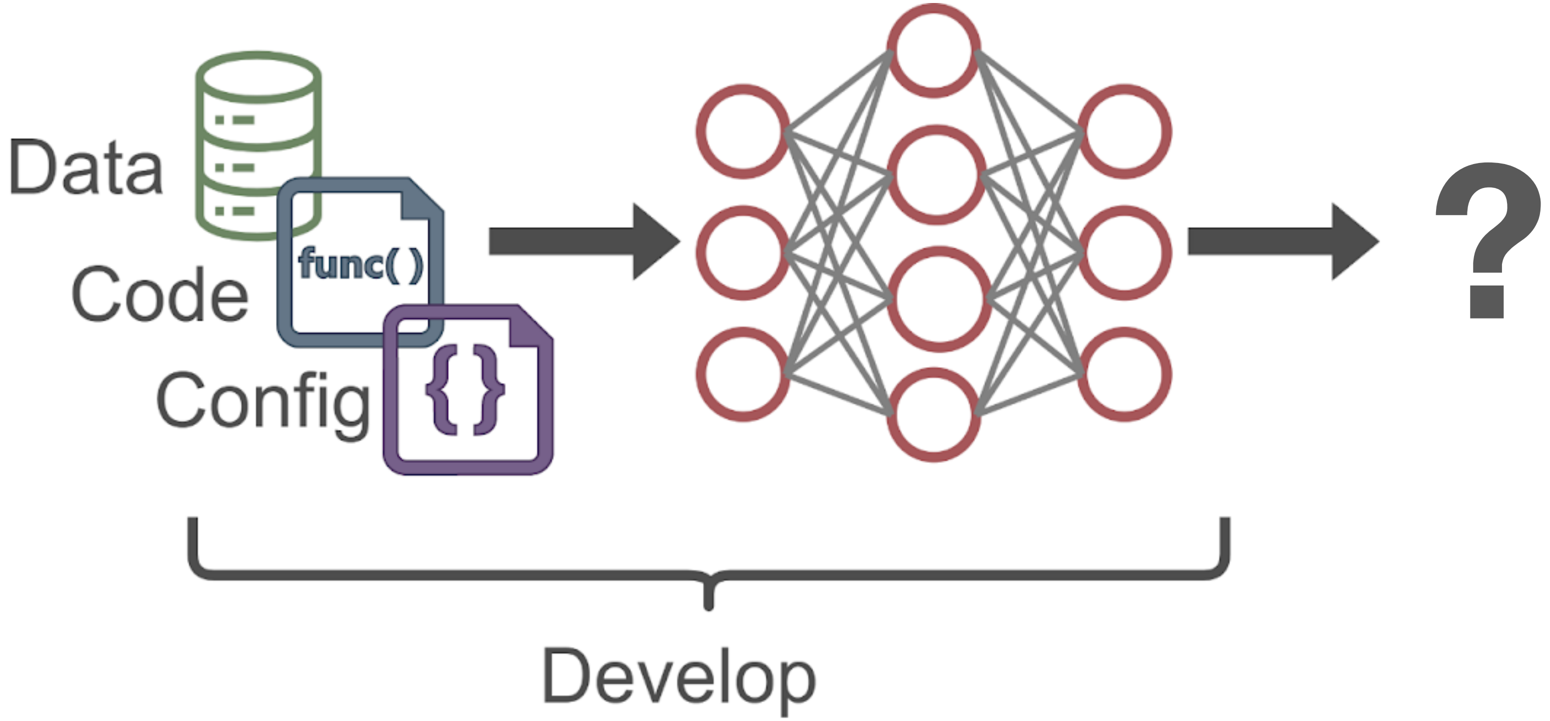




Fully convolutional auto-encoder maps the witness channels $\{w_i(t)\}$ into the noise predictions $n_w(t)$ which are then subtracted from detector output $h(t)$

Need for End-to-End ML

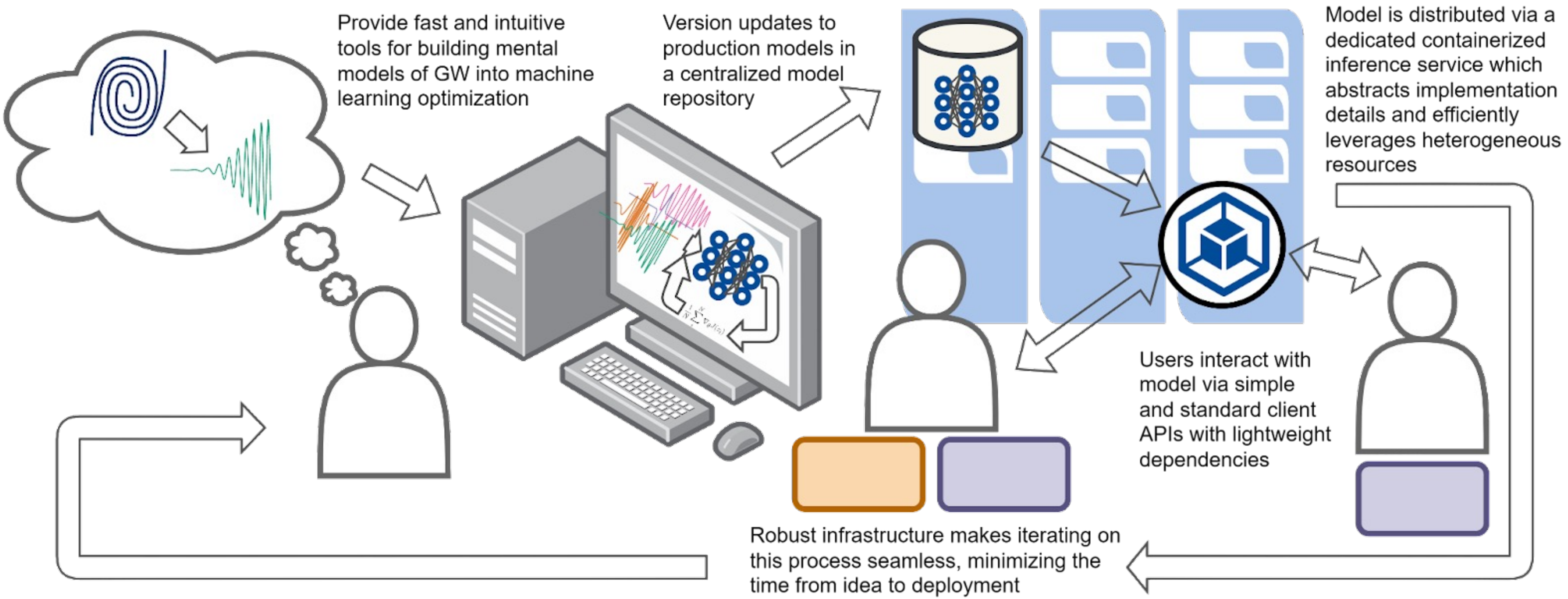
https://indico.cern.ch/event/1156222/contributions/5084202/
https://github.com/ML4GW



- How do offline metrics correspond to real-time performance?
- What happens when data goes bad?
- What happens when model goes stale?

End-to-End ML

https://indico.cern.ch/event/1156222/contributions/5084202/
https://github.com/ML4GW



How to actually train a model

https://indico.cern.ch/event/1156222/contributions/5084202/
https://github.com/ML4GW

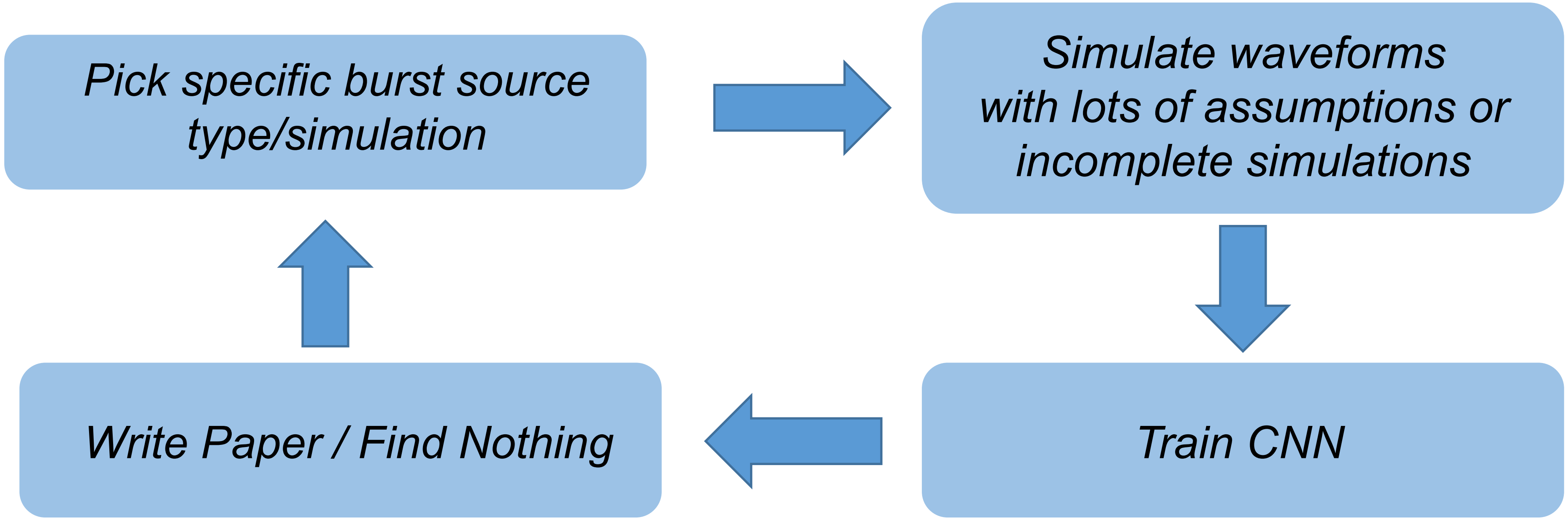
- It's easy enough to generate a toy dataset, train a CNN, detect GWs!
- In practice, this is almost completely useless.

- You need to have a complete infrastructure that can utilize **real** data :
  - Consistent preprocessing schemes
  - Data Loader
  - Data augmentation to sample glitches, oversample noise
  - Data generator to generate hypothetical signal
  - Validation on real data
  - Systematic implementation/attention to detail that can be used in REAL training pipelines

- Preferably, this should also be accessible to non-GW physicists!
- Preferably, this whole process should be automatic, requiring little user input.
- Hence: ML4GW Organization

Burst Searches

How do you find something when you don't know what it looks like?
Historically ML burst search pipelines have gone like this….

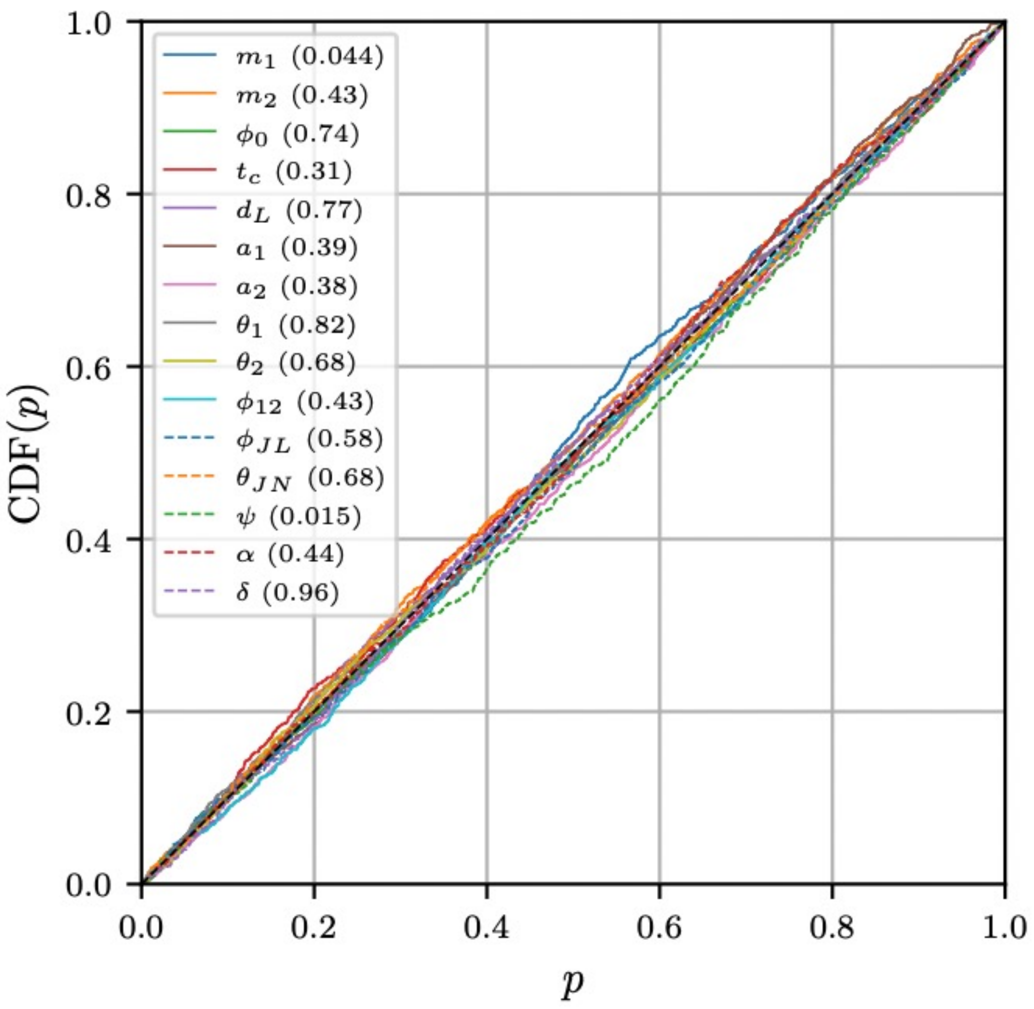


But there are some new ideas! Let's talk about this more in Sec. 4 anomaly detection…

Bayesian Posterior – Normalizing Flows

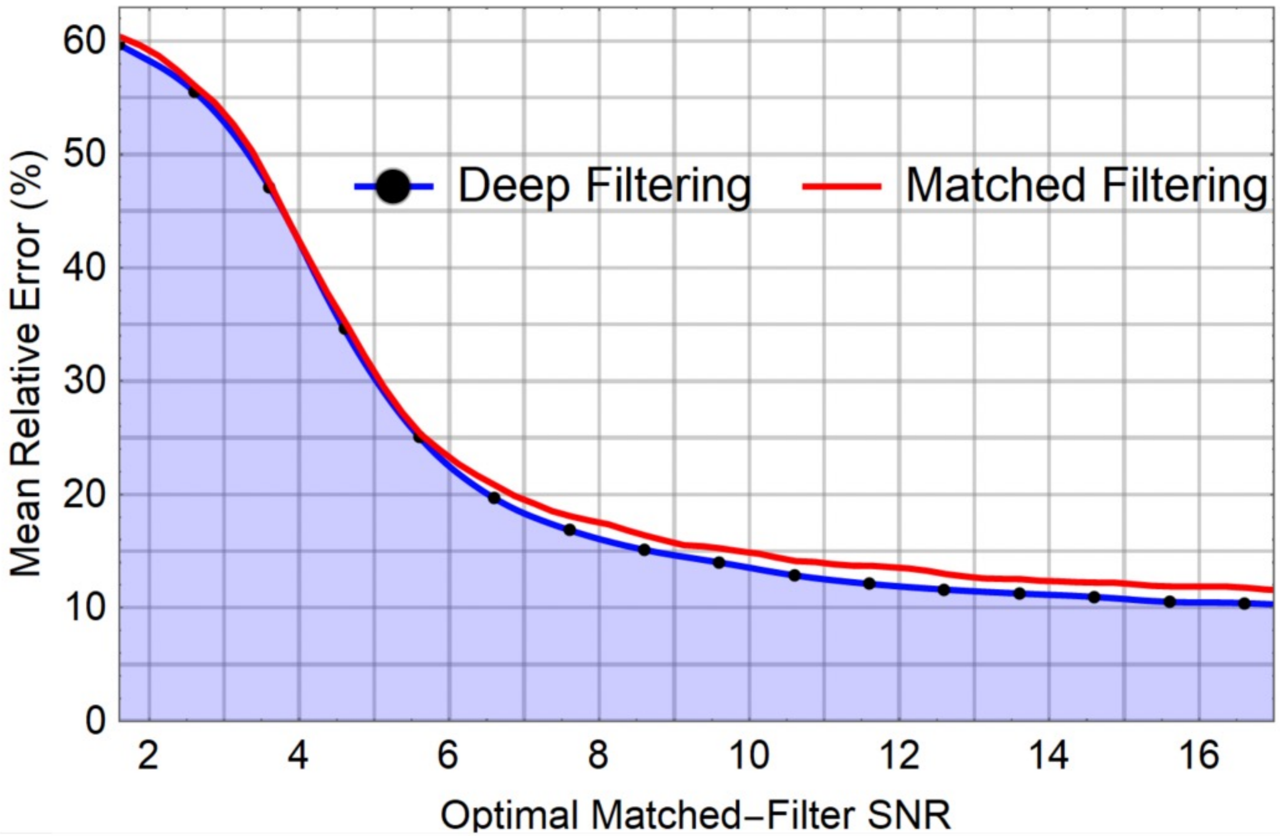
[Green et al. (2020) (Phys.Rev.D)]
[Dax et al. (2021) (Phys.Rev.Lett)]

Neural Posterior Estimation using Normalizing Flows



Direct Parameter Estimation

[George & Huerta (2017) (Phys.Lett.B)]



Quick estimation of binary masses* after a supervised detection using CNN

*Usually, you need full parameter set for MMA pipelines

QUAK Spaces on Supernovae