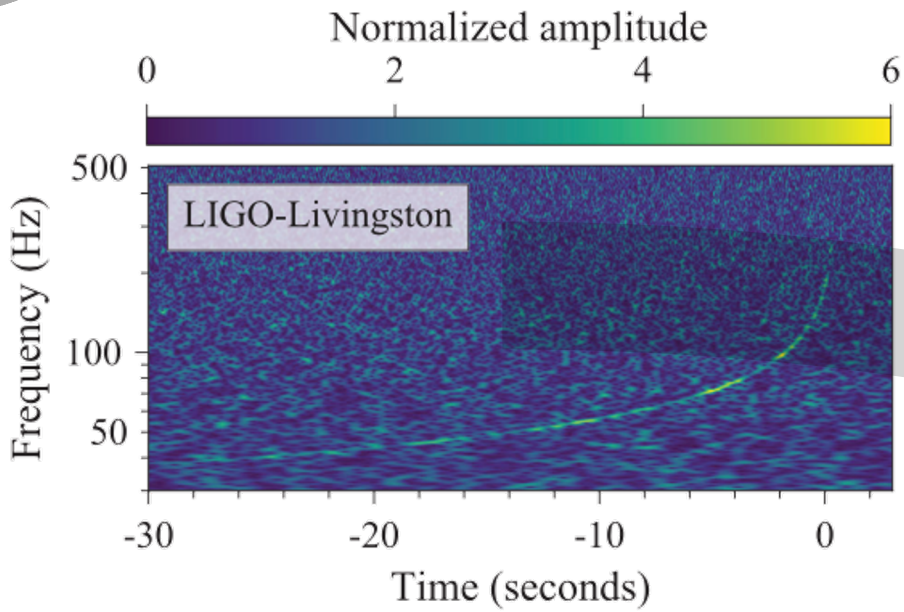
Machine Learning for Gravitational Wave Astronomy

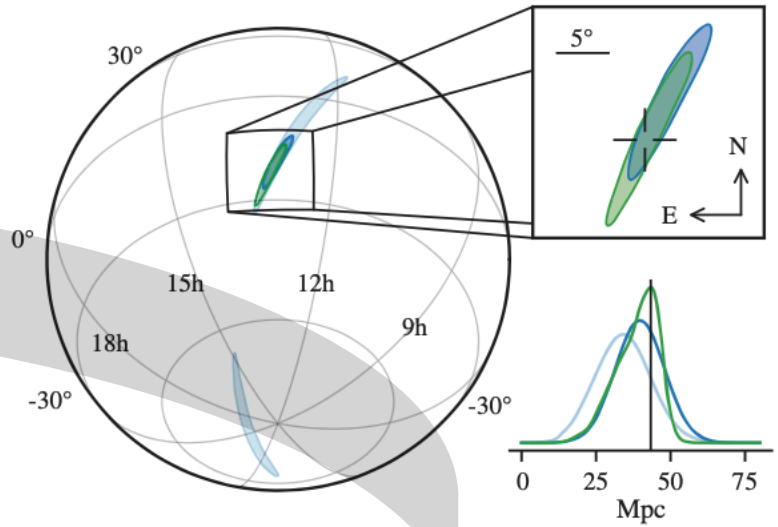
Gabriele Vajente

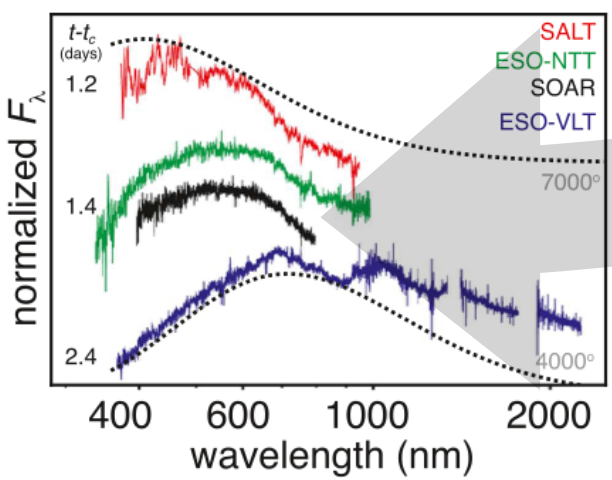
LIGO Laboratory – California Institute of Technology

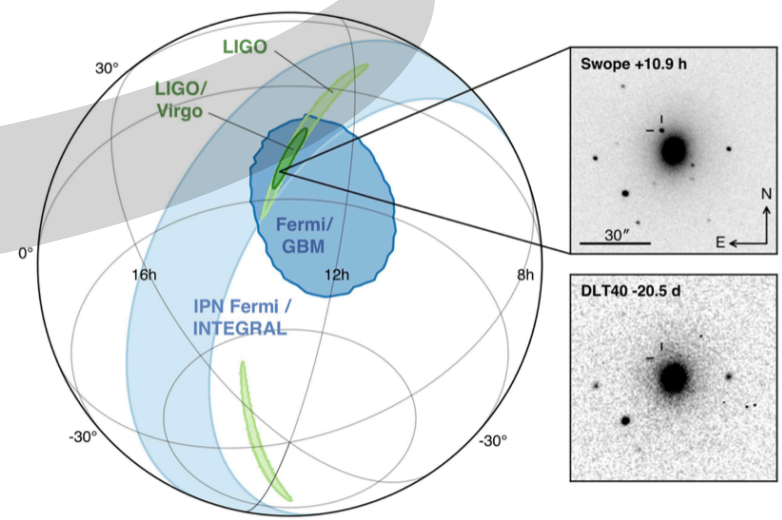
GWMESS 2021 - Institut Henri Poincaré

Multi-messenger (GW) astronomy



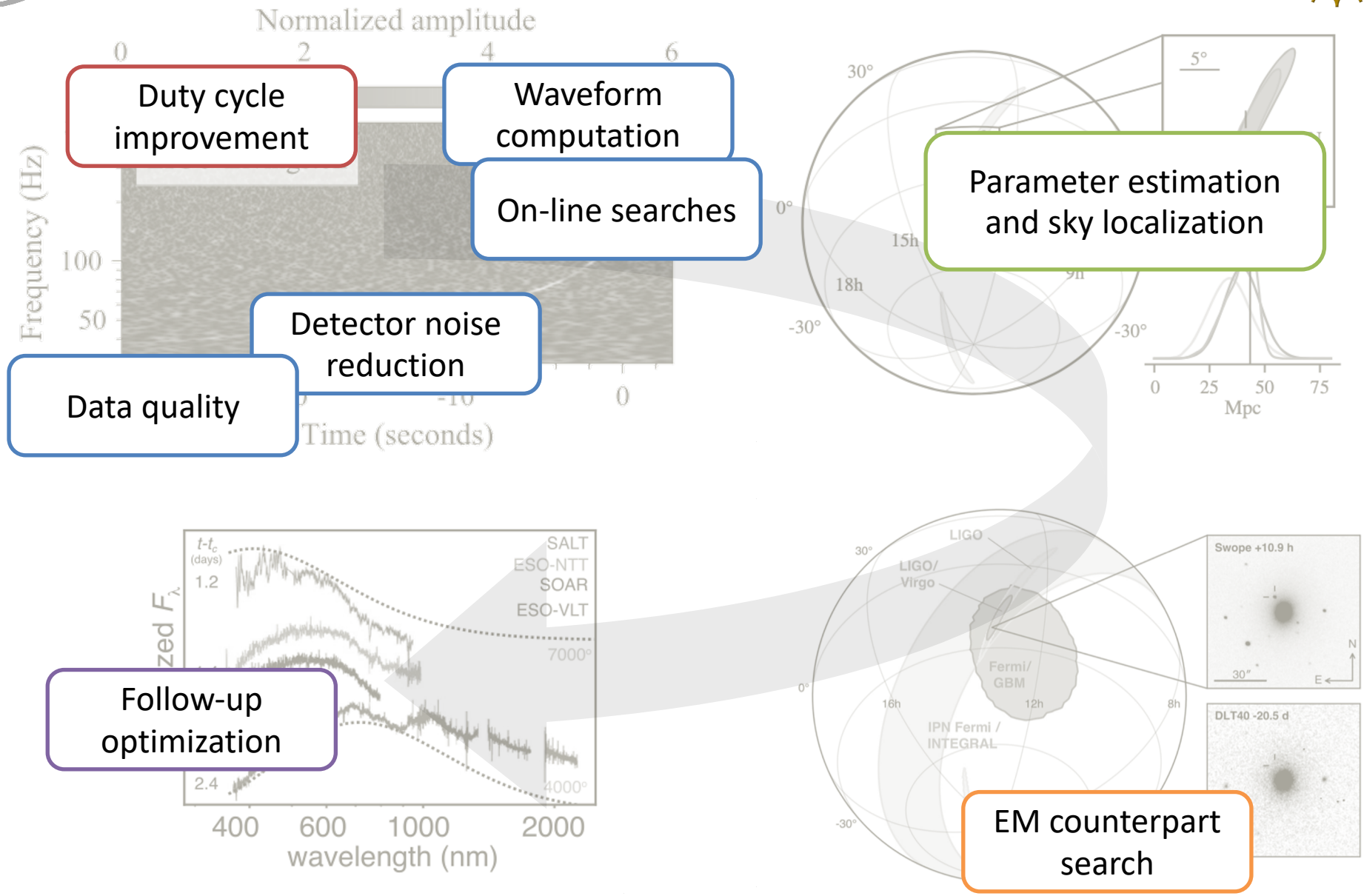






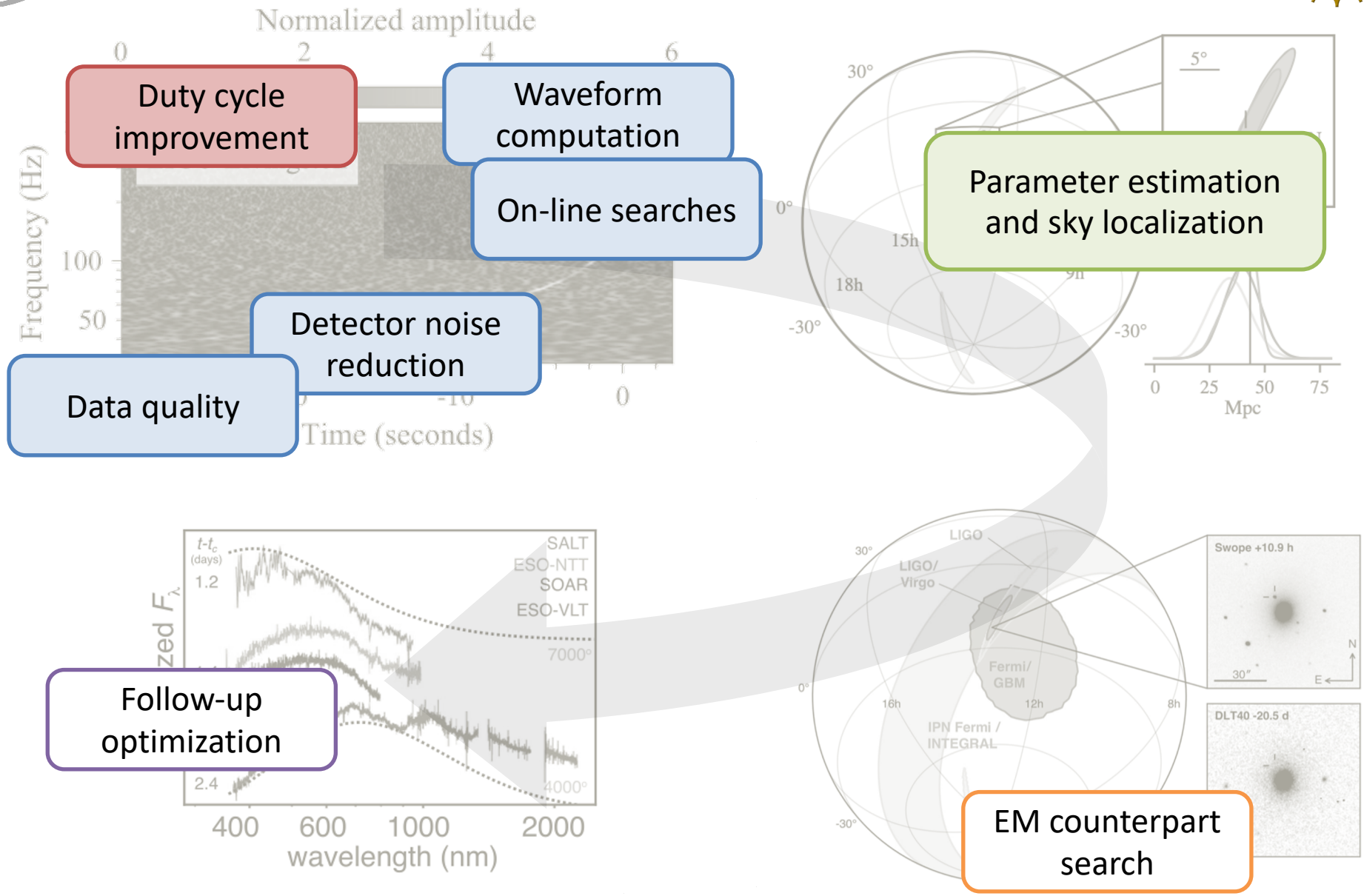
PRL 119, 161101 (2017), AJL, 848:L12, 2017

Machine Learning



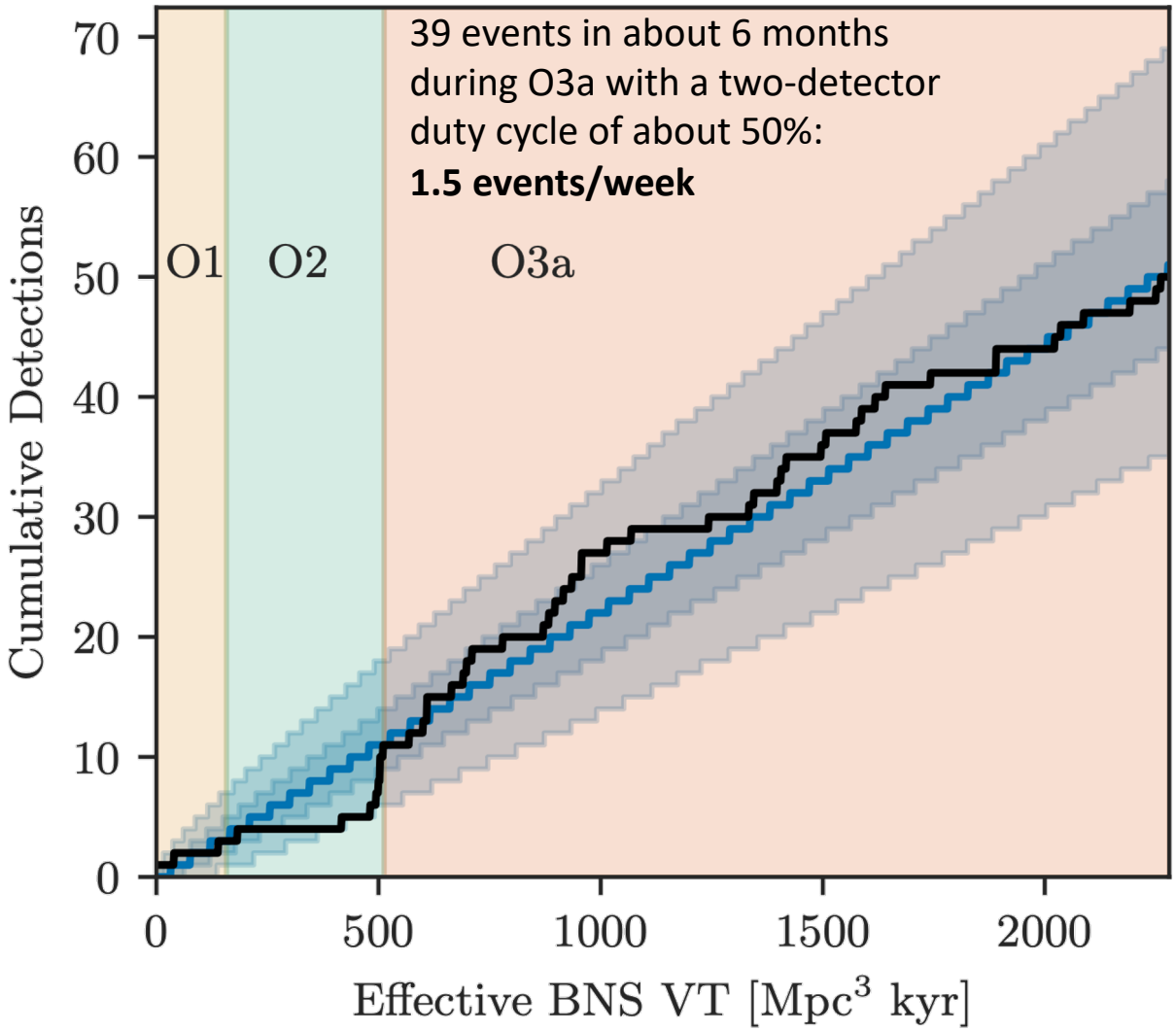
Mach. Learn.: Sci. Technol. **2** 011002 (2021) Nature Reviews Physics 1, 600–608(2019)

Machine Learning



Mach. Learn.: Sci. Technol. **2** 011002 (2021) Nature Reviews Physics 1, 600–608(2019)

Event rate will increase in the future



Number of events depends on

- Observation time (duty cycle)
- Observation distance (sensitivity)
- Astrophysical search sensitivities (background glitches, efficiency and coverage)

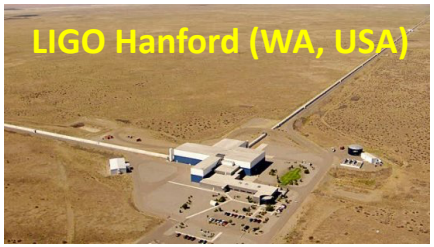
Science payoff depends on

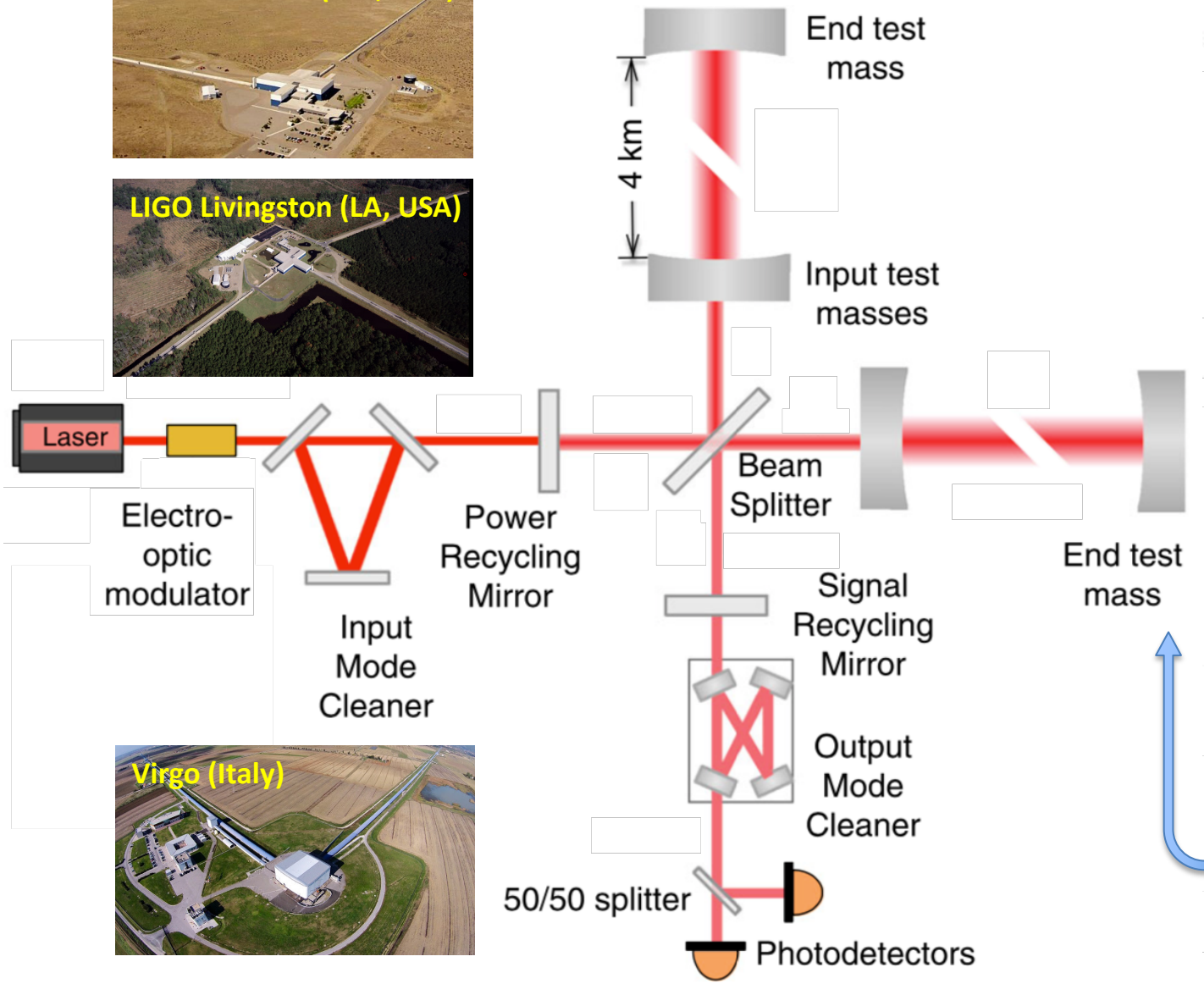
- Number of events
- Precision of waveform and parameter estimation
- Fast alerts for EM counterparts

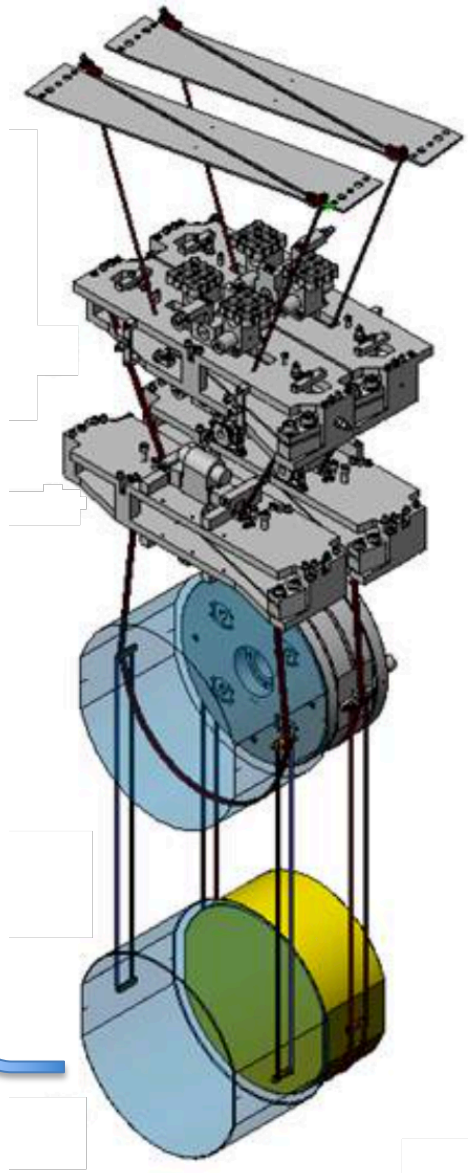
arXiv:2010.14527v2

ML TO IMPROVE THE DETECTORS

Advanced GW Detectors

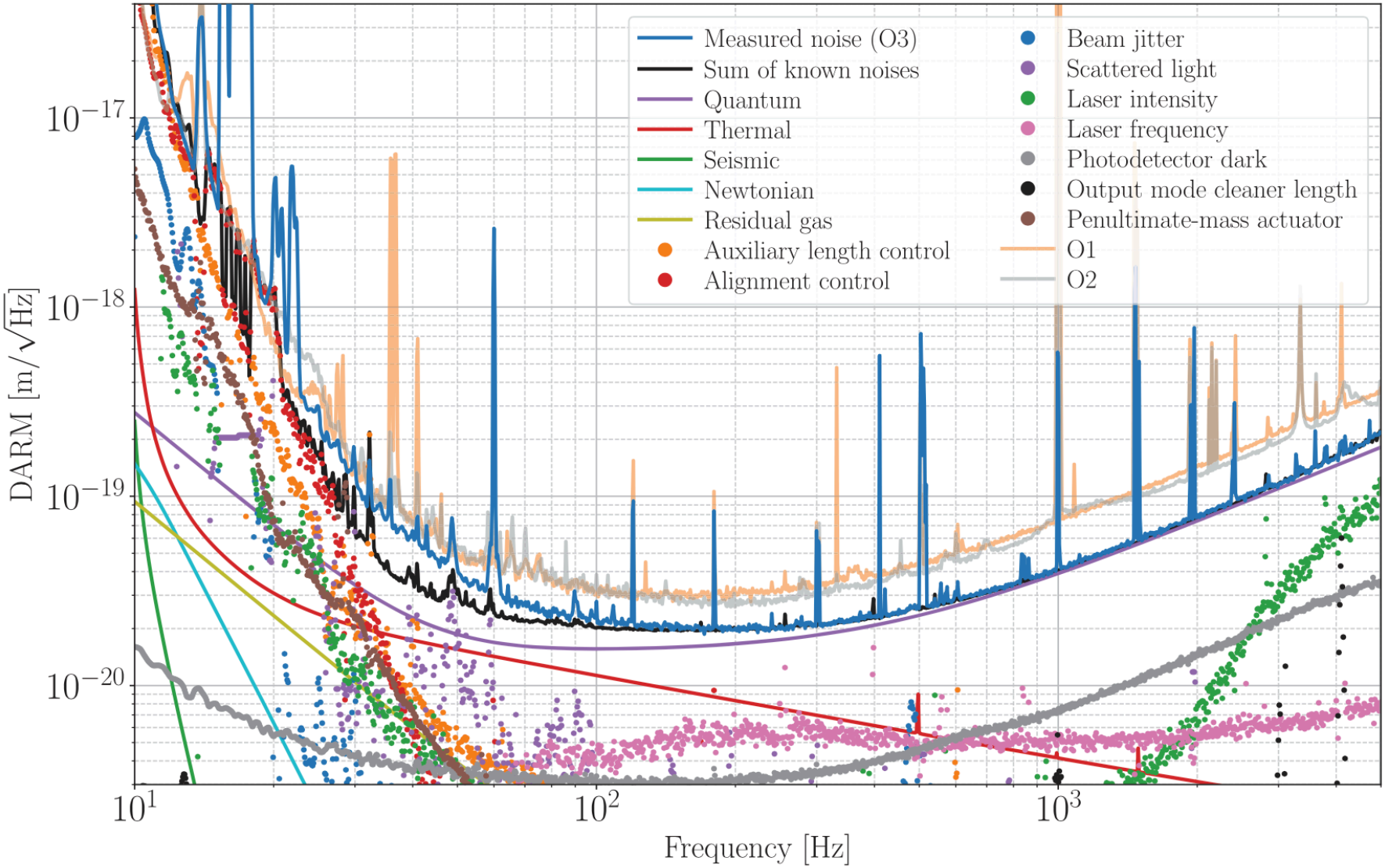






Phys. Rev. D 93, 112004 (2016)

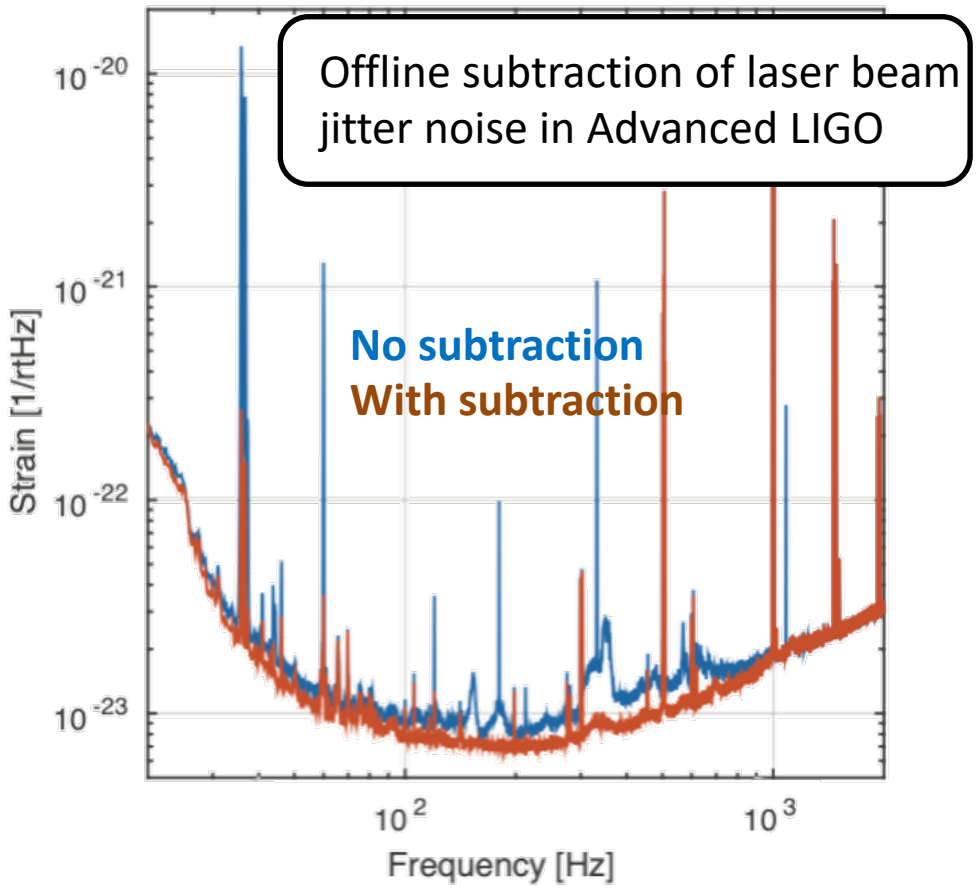
Noises



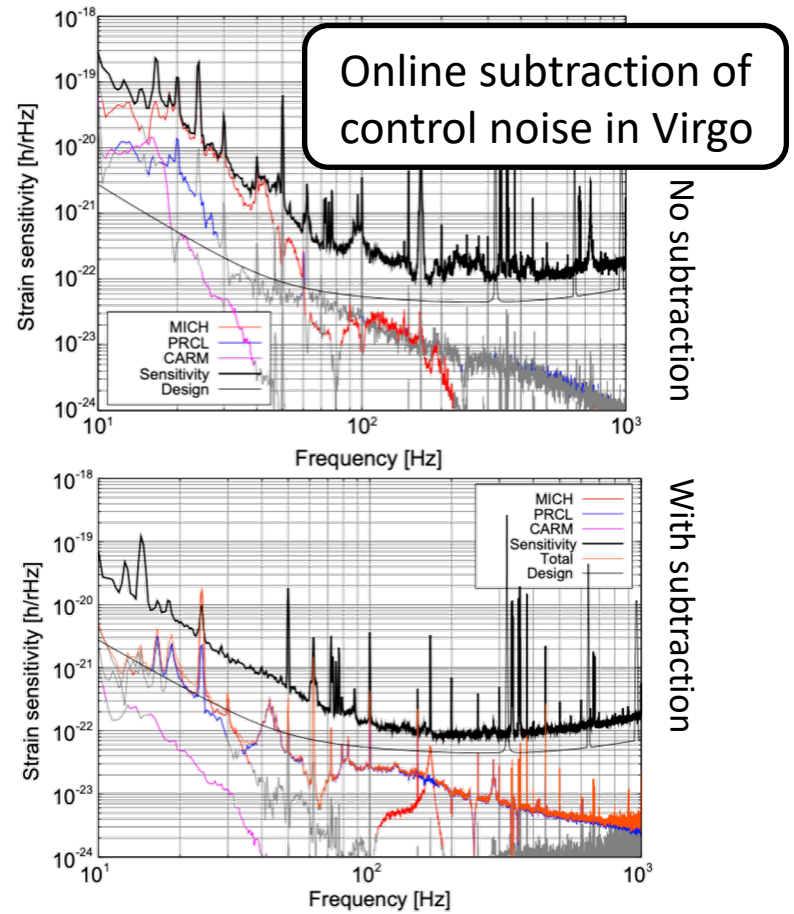
PHYS. REV. D 102, 062003 (2020)

Linear Noise Subtraction

Phys. Rev. D 99, 042001 (2019)



Astroparticle Physics 33 (2010) 75



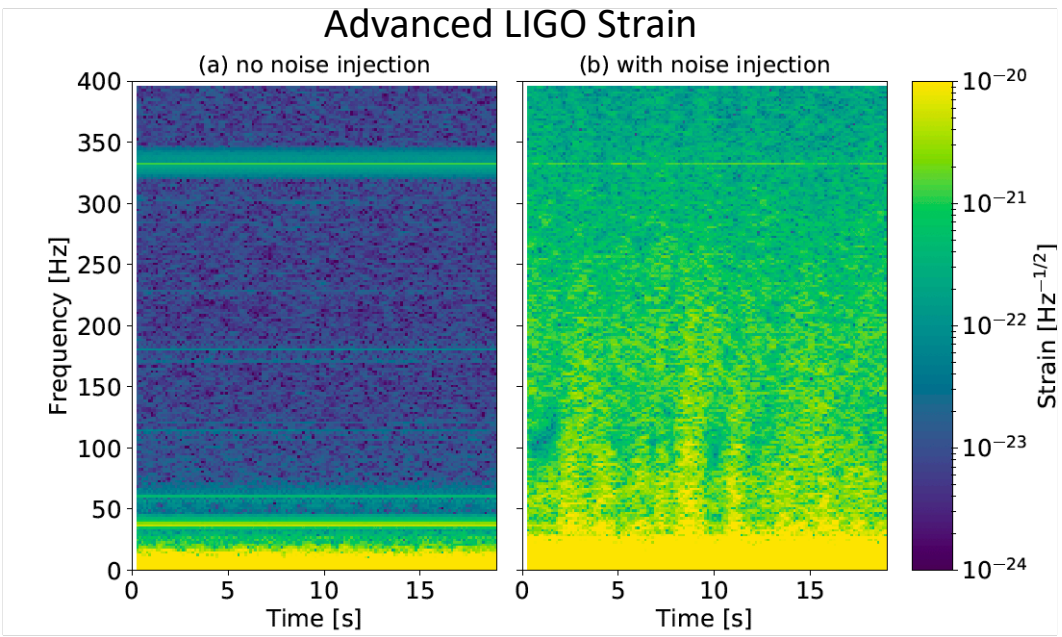
- Unexpected noise couplings can limit the detector sensitivity
- If linear and stationary:
  - Coherence / Transfer Function analysis, Wiener filter, online or offline

Non-Linear and Non-Stationary

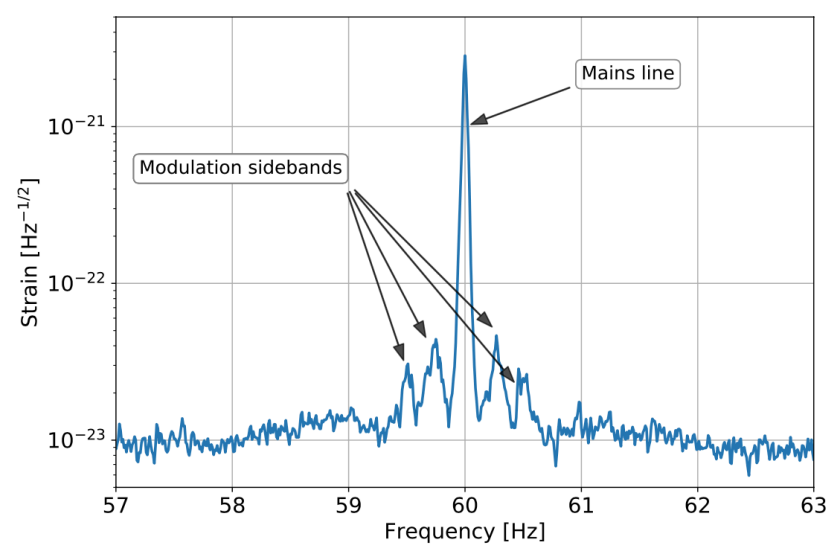
- Noise in auxiliary channels can couple in a **non-linear** way

$$h(t) = \varepsilon_{\mathrm{B}}(t) + \mathcal{F}\left[w_1(\tau < t), \ldots, w_N(\tau < t)\right]$$

- However, (in most cases*) we expect small deviations from linearity: quadratic coupling or non-stationary coupling



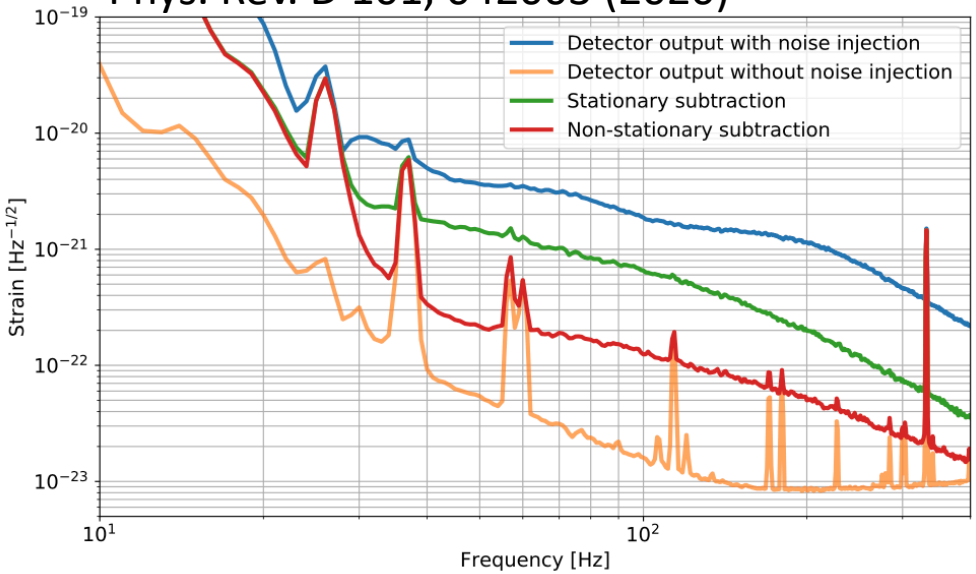
Phys. Rev. D 101, 042003 (2020)

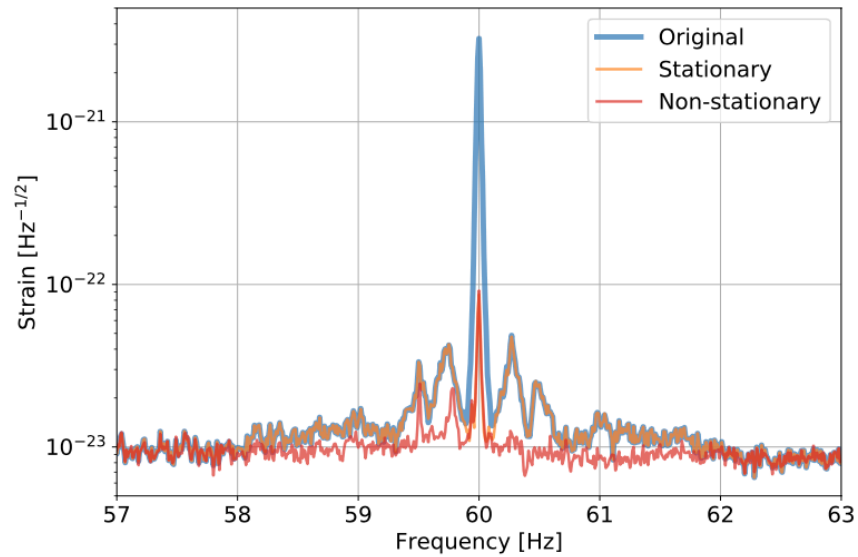


*scattered light is an exception: arXiv:2007.14876 (2020), Opt. Expr. 2 10546 (2013)

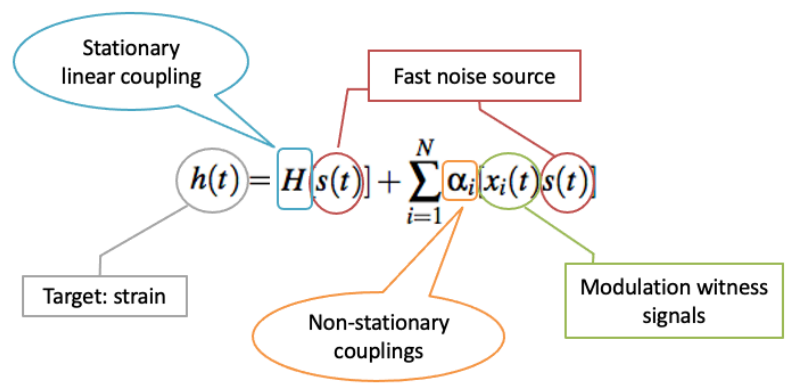
Physics-informed Machine Learning Models

Phys. Rev. D 101, 042003 (2020)





- Non-stationary noise modelling and subtraction **outperforms** the linear subtraction
- Not the most general non-linear model, but "**good enough**" to improve Advanced LIGO sensitivity
- **Interpretable results**: we can learn what degree of freedom modulates the coupling

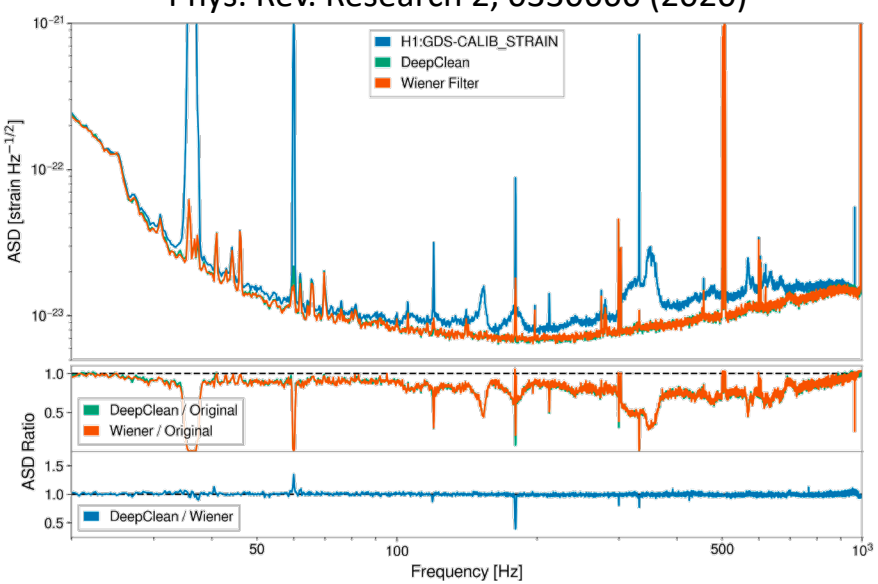


Why not a Neural Network?

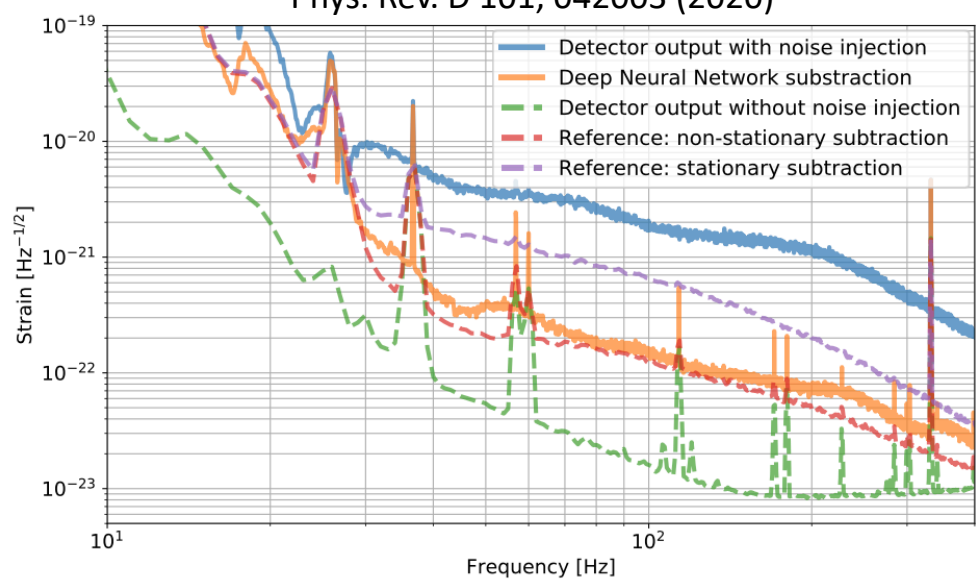
- In theory, a Deep Neural Network can express arbitrarily complex non-linear functions
- Need to include time-dependency and history, therefore Convolutional NN or Recurrent NN
- DNN shown to reproduce linear subtraction performance
- Although a RNN is in principle capable of expressing a slowly modulated coupling, it's hard to train, at the end of training it underperforms, and it is not interpretable

Same sampling problem as in the control applications: need fast sampling to capture noise and long times to capture variations. CNN and RNN do not scale well with very long time series

Phys. Rev. Research 2, 0330666 (2020)



Phys. Rev. D 101, 042003 (2020)



ML FOR DATA QUALITY

Data Quality

Class. Quantum Grav. **33** 134001

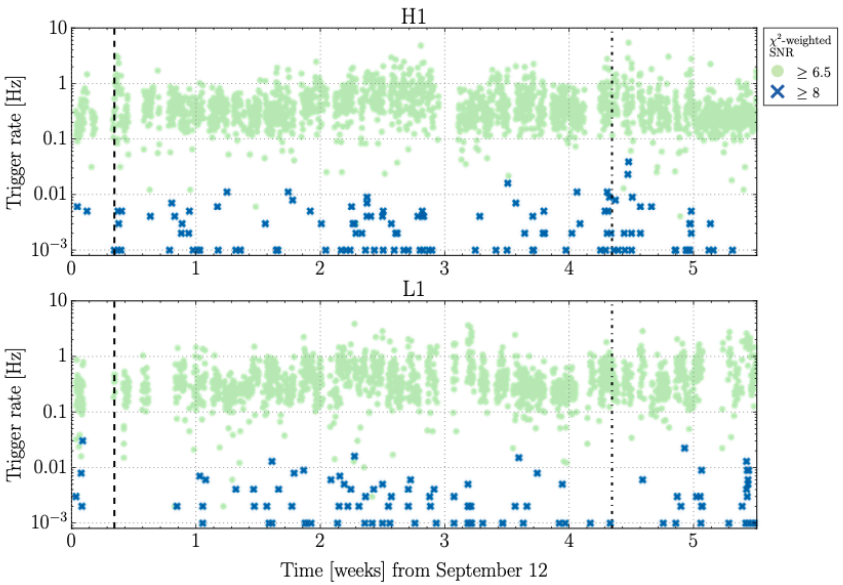


Figure 5: The rate of single interferometer background triggers in the CBC search for H1 (above) and L1 (below), where color indicates a threshold on the detection statistic, χ^2-weighted SNR. Each point represents the average rate over a 2048 second interval. The times of GW150914 and LVT151012 are indicated with vertical dashed and dot-dashed lines respectively.

Class. Quantum Grav. **35** 065010

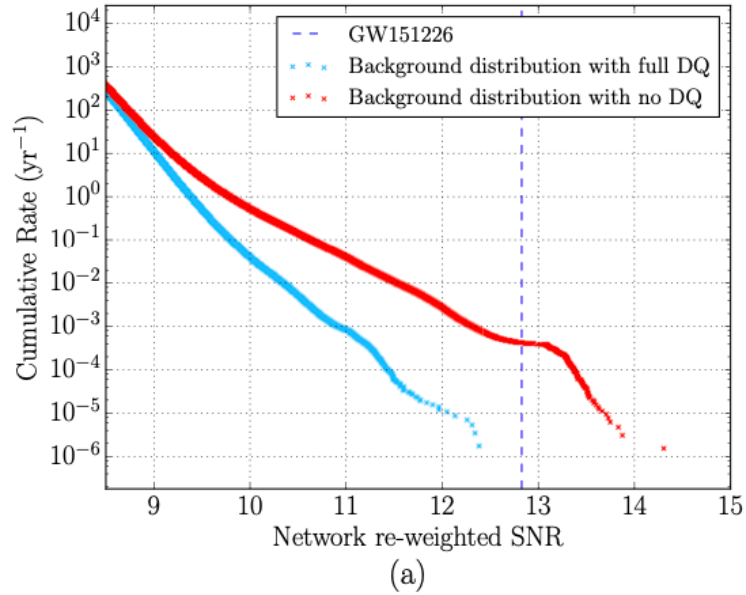
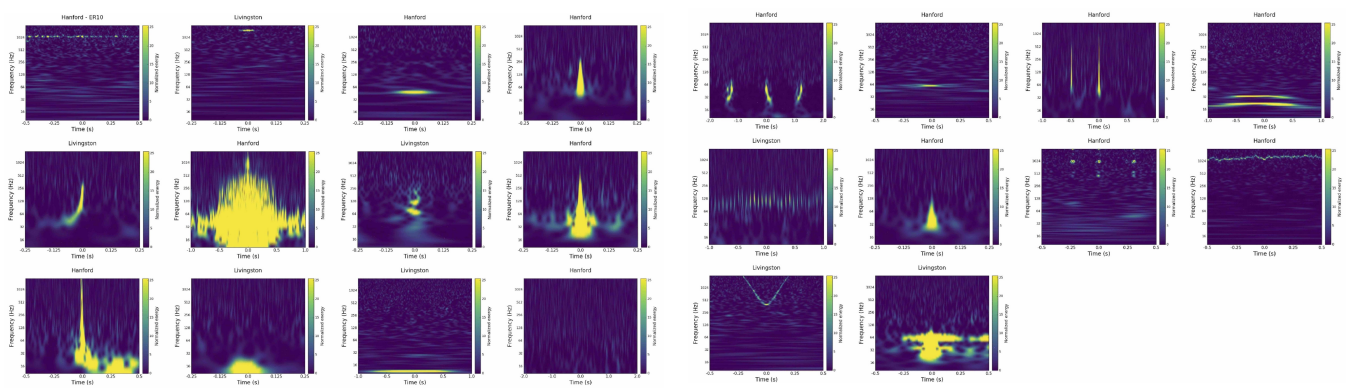


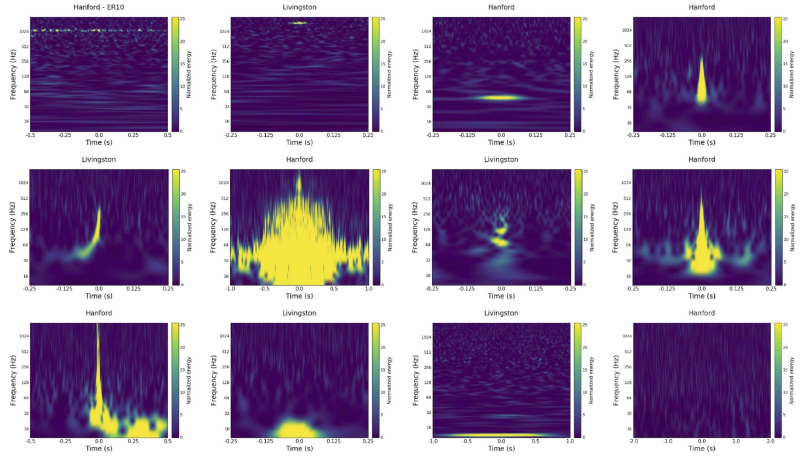
Figure 8. The background distribution in the bulk bin before and after applying DQ vetoes for the analysis containing GW151226. (a) The cumulative rate of background triggers in the bulk bin as a function of re-weighted SNR.

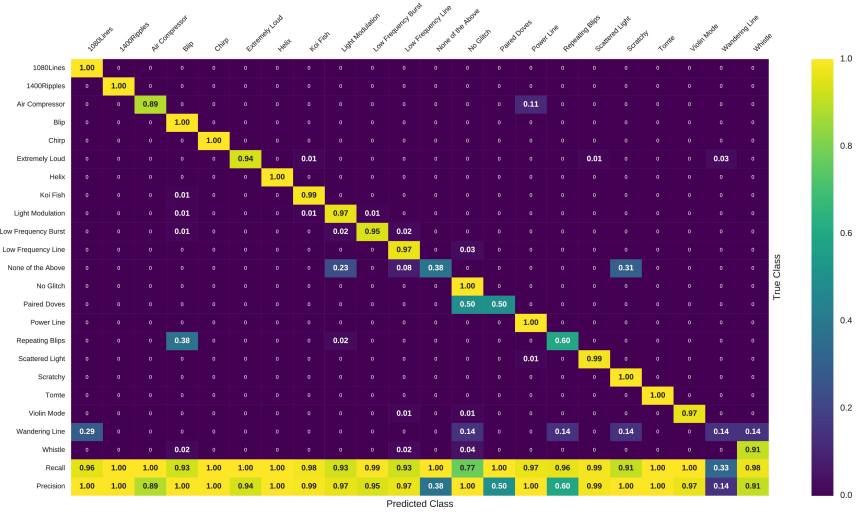
Information Sciences
444 (2018), 172-186



ML for data quality

Gravity Spy classification / Citizen Science





Information Sciences 444 (2018), 172-186
Class. Quantum Grav. 34 (2017) 064003

iDQ veto generation based on aux channels

Class. Quantum Grav. **30** 155010
arXiv:2005.12761v1

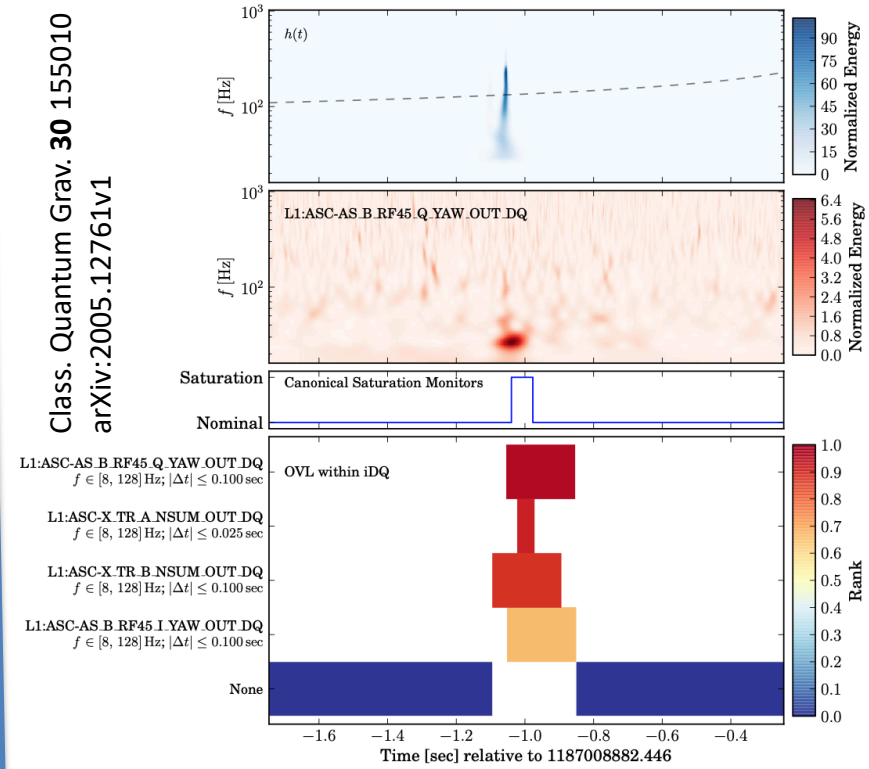


FIG. 7. iDQ's low-latency predictions surrounding the non-Gaussian noise transient coincident with GW170817 in the LIGO Livingston detector. (*top panel*) Time-frequency decomposition of the GW strain chosen to highlight the noise transient's short duration; GW170817's inspiral track is shown for reference (*dashed line*), with times measured relative to the coalescence time. (2nd *panel*) Time-frequency representation of the top-ranked auxiliary witness active at this time, autonomously identified by iDQ as correlated with non-Gaussian noise without *a priori* knowledge about the type of noise present in the detectors. (3rd *panel*) A canonical saturation monitor, which identified the time as problematic. (*bottom panel*) OVL's feature importance, showing multiple veto configurations (Appendix B) active in coincidence with the glitch. Color denotes OVL's rank, with rank→ 1 indicating high confidence in the presence of a glitch.

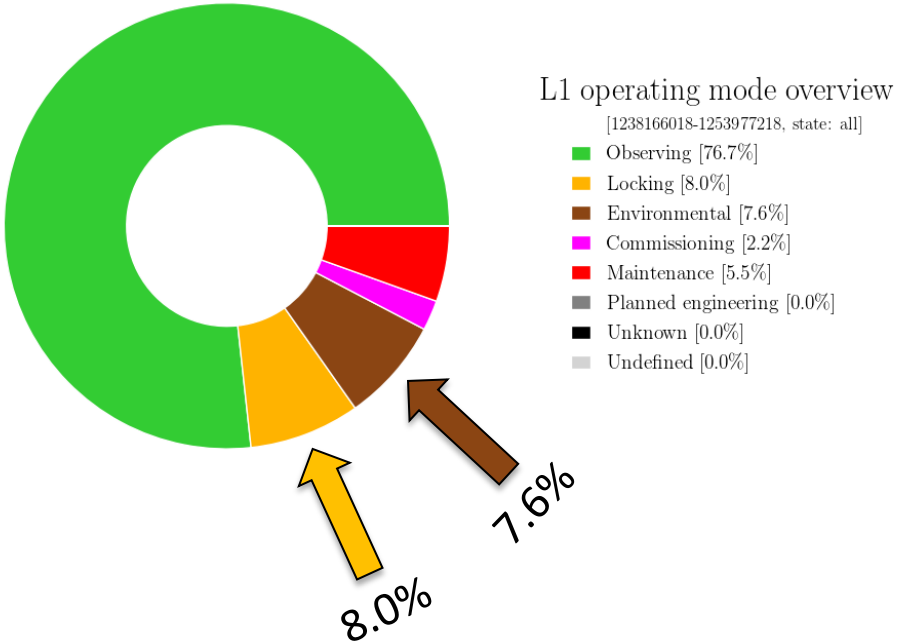
More ML for data quality

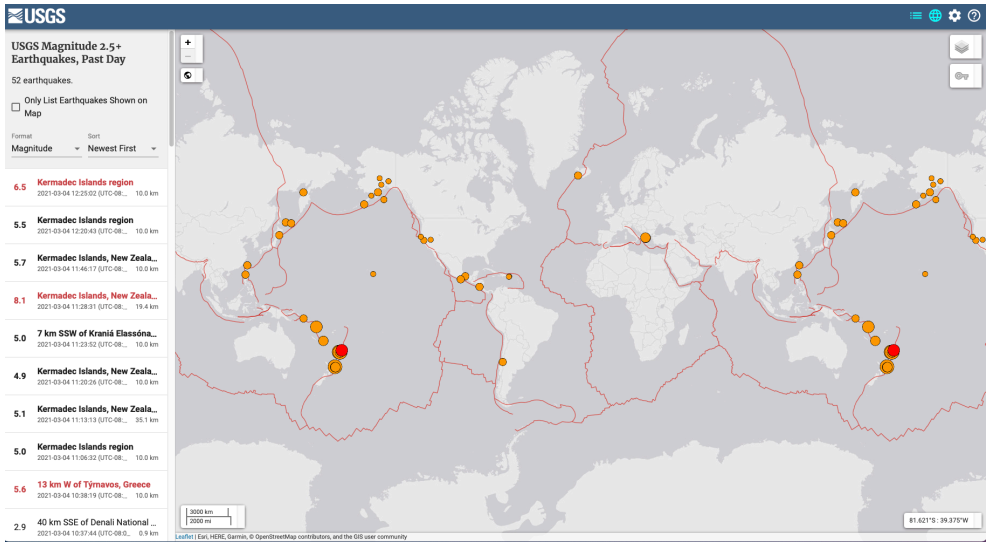
- An example of big data
  - Find correlations between spurious glitches in the main GW strain signal with other auxiliary channels
  - Auxiliary channels: a few thousands fast, many tens of thousands slow
  - Correlation could be
    - "linear" glitch in aux channel coincident with glitch in main channel
    - "non linear" some feature in aux correspond to a glitch in main channel (saturation, level crossing, derivative value, etc…)
  - Need for efficient and robust data mining algorithms and classification algorithms

- Commun. Comput. Phys., 25 (2019), pp. 963-987
- Phys. Rev. D **101**, 102003 (2020)
- Phys. Rev. D 97, 101501(R) (2018)
- Proceedings of 25th ACM SIGKDD Workshop https://doi.org/10.475/123_4

ML FOR DETECTOR ROBUSTNESS

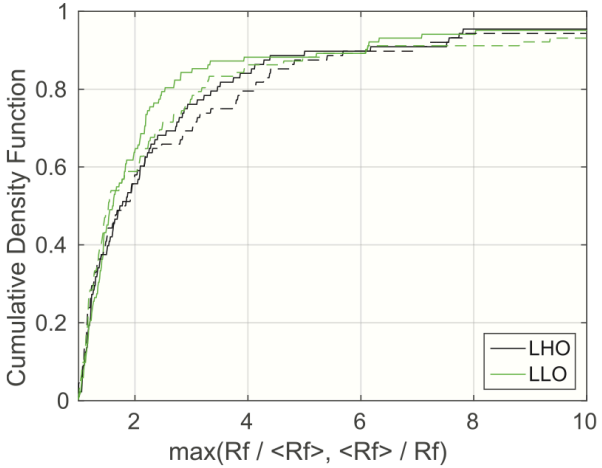
Duty cycle improvement





(1) Predict and prevent

- Compute time of arrival of seismic waves and amplitudes, based on available USGS data



Class. Quantum Grav. **34** 044004
Class. Quantum Grav. 35 055004

Figure 7. Performance of estimation of peak velocities seen during O1 at the interferometers (LHO and LLO) using fit parameters estimated from S5-S6 data. The *x*-axis gives the maximum of the ratio between the estimated and measured peak ground velocities and vice versa. The solid lines use the final earthquake parameter estimates while the dashed lines use the preliminary earthquake estimates. About 90% of events are within a factor of 5 of the predicted value. The difference in fit parameters due to use of preliminary notices is minimal.

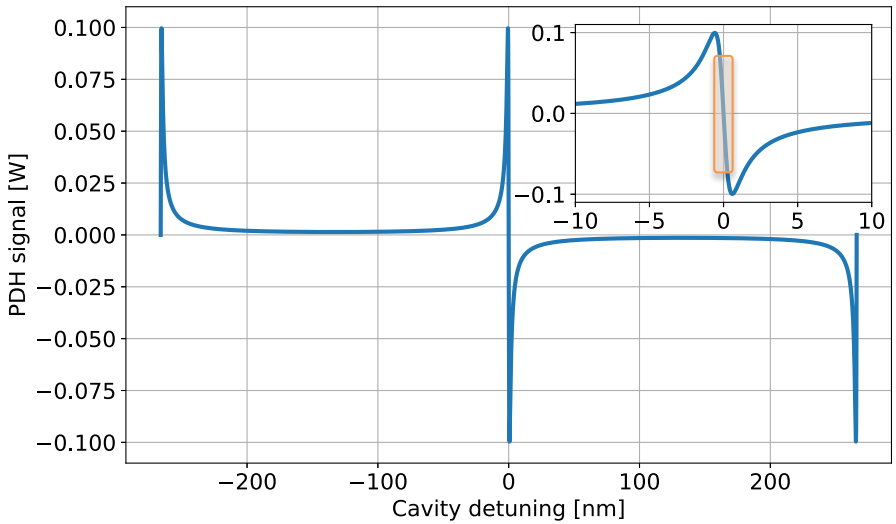
(2) Be faster at recovering

- Improve the "lock acquisition" sequence and make it more robust and faster

Sensing and Control

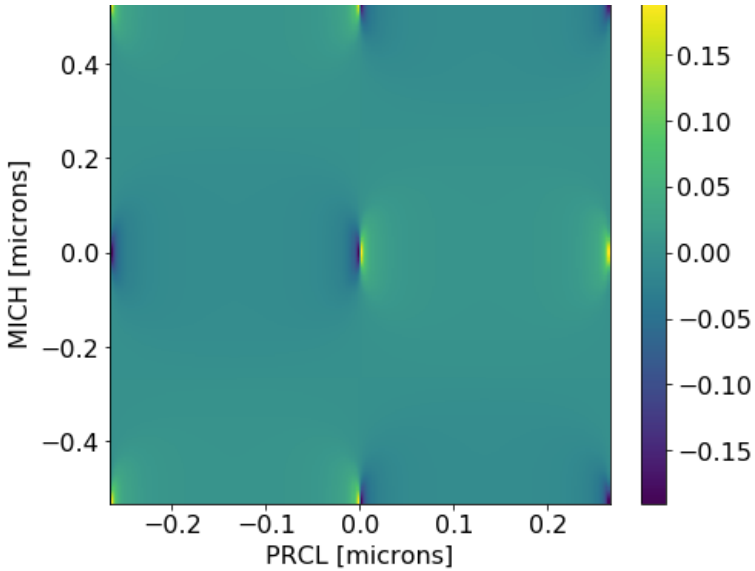
- Radiofrequency modulation-demodulation scheme to measure the deviation from resonance
- Small linear range
  - 5x longitudinal degrees of freedom
  - Linear response exists for small fraction (~ nm) of the (periodic by $\lambda/2 \simeq 500$ nm) values for each d.o.f.
  - Fraction of phase space volume where linear control is possible < **10^{-9}**
- Distance varying by microns on a few seconds time scale due to ground seismic motion

For a single arm cavity:



Applied Physics B volume 31, 97 (1983)

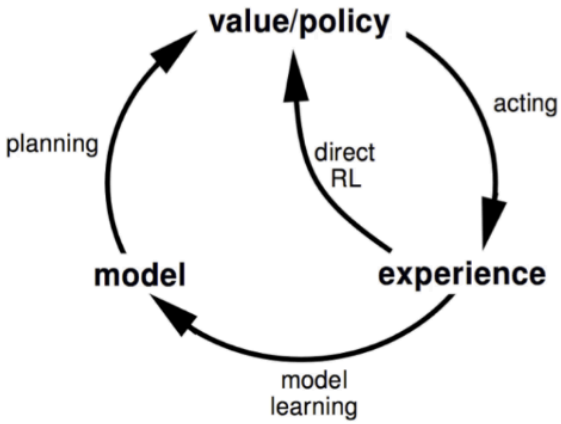
For a simple three-mirror interferometer:

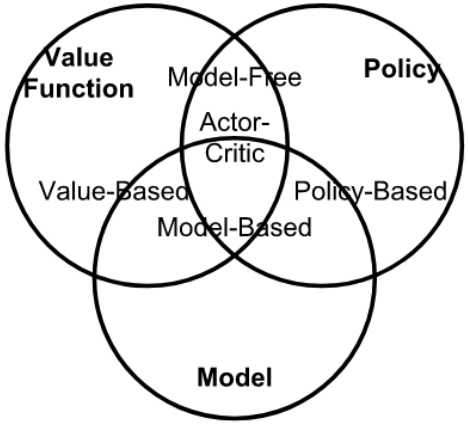


Machine Learning for Lock Acquisition

Non-linear control problem: drive the system into a narrow region of the phase space, where linear control is possible

- **Reinforcement Learning**
  - Episodic task
  - Reward easy to set up, as time to reach operating point, or time spent near the operating point
- **Pros**:
  - Could be model agnostic
  - But could also use limited model knowledge and improve the model
- **Cons**:
  - Time consuming: typically need many attempts on the real system (expensive)
  - Risky: random policies could perform unwanted actions on a delicate system

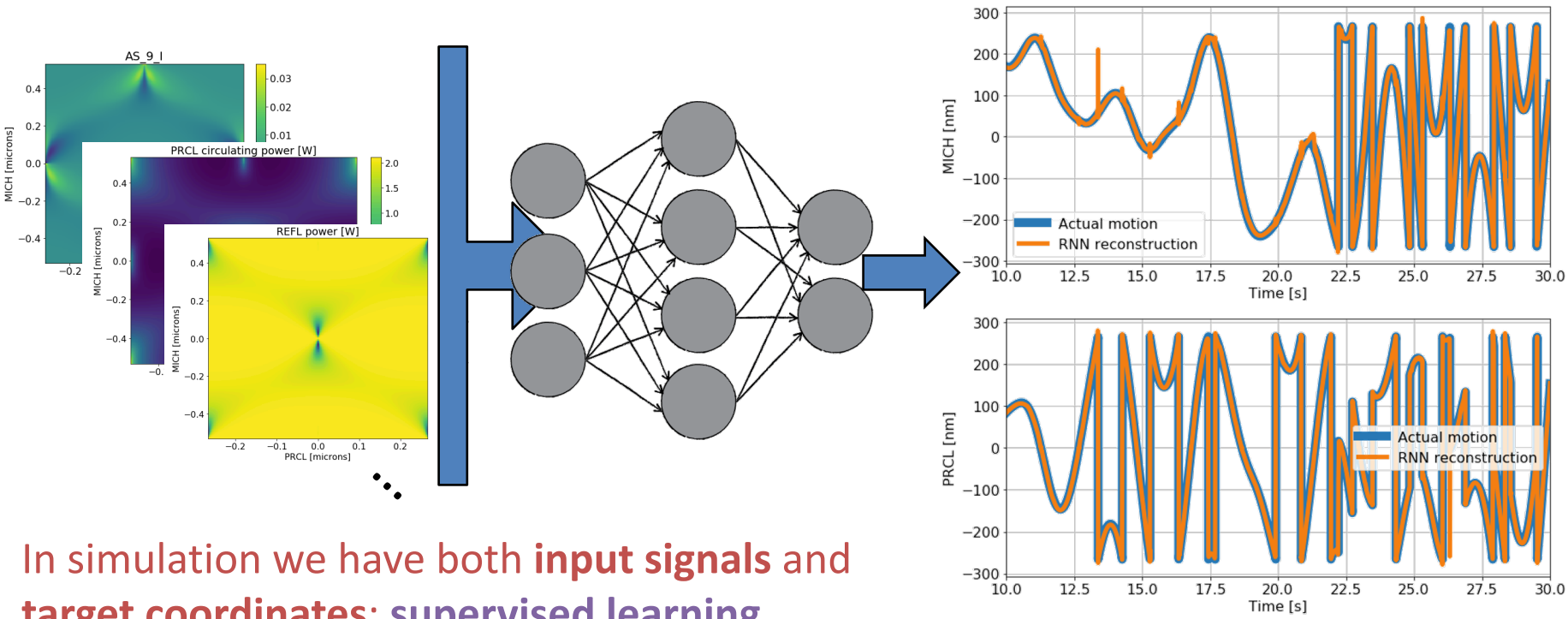




Machine Learning for Lock Acquisition

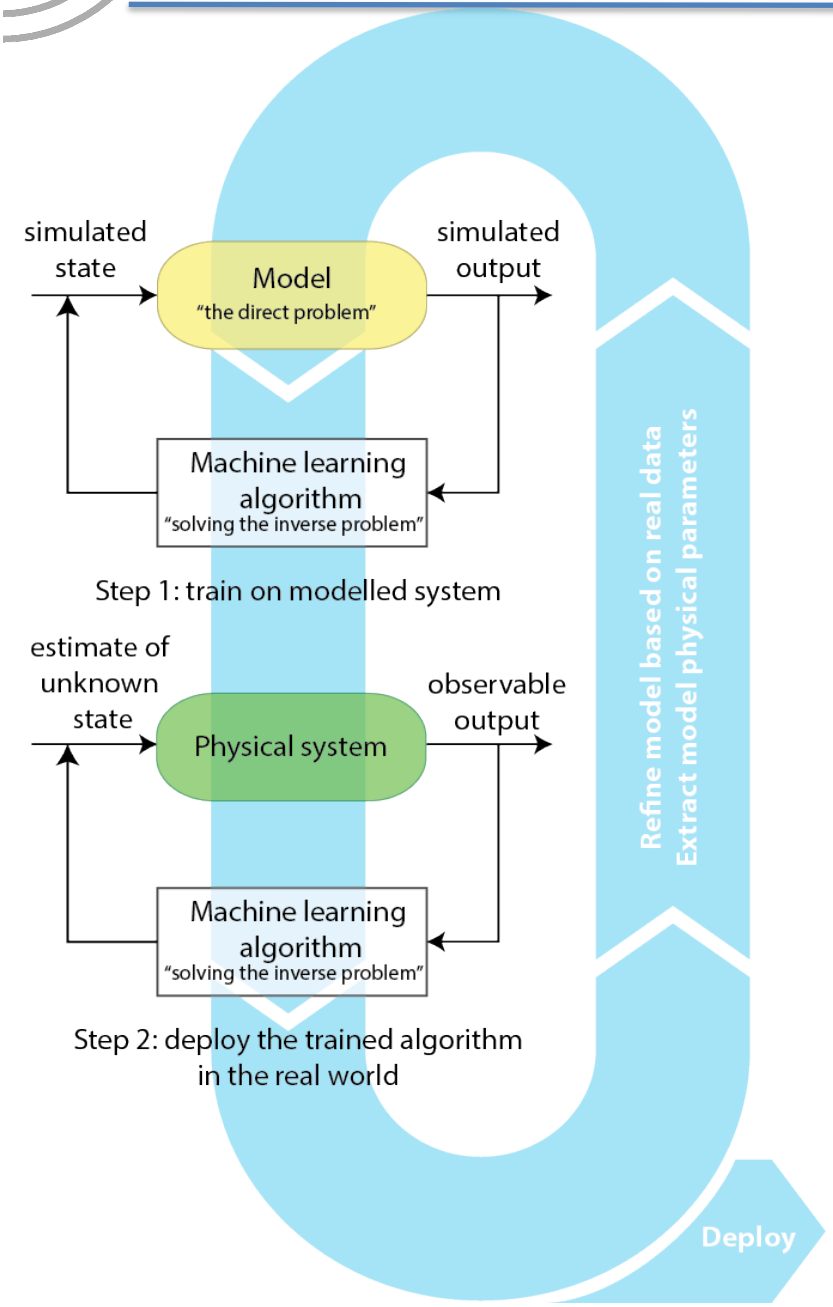
Non-linear control problem: drive the system into a narrow region of the 5d phase space, where linear control is possible

- Construct a **non-linear state estimator**: use all available signal as input, build an estimate of the degree of freedom position that works everywhere



In simulation we have both **input signals** and **target coordinates**: **supervised learning**

Machine Learning for Lock Acquisition

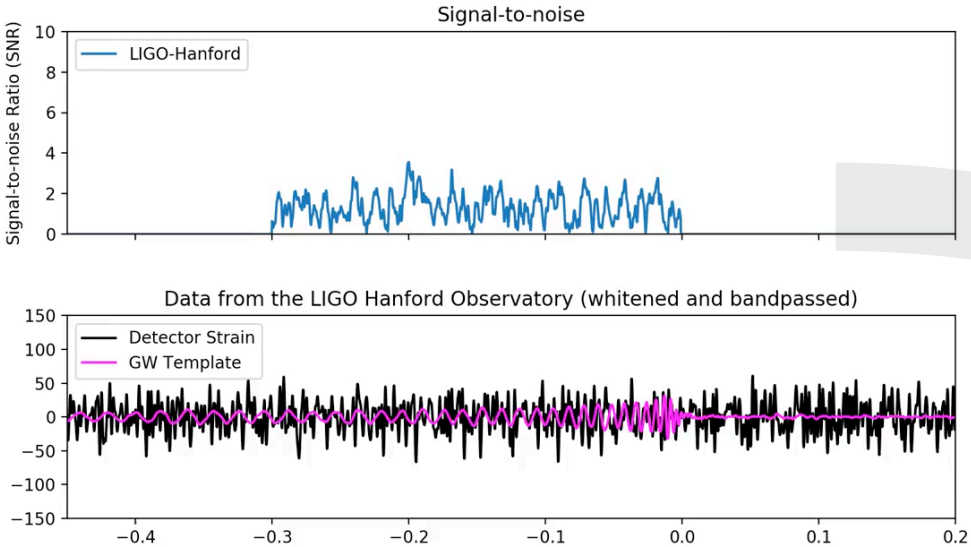


- Deep Neural Network learning needs a lot of training examples (**10^5 - 10^6**)
- Not practical to do it online (and **we don't have the targets** in the real system!!)
- Use a **simulation** of the system as accurate as possible (including uncertainties)
- **Train on the simulated data**
- Deploy on the real system and test the performance
- Fine tune if needed

ML TO IMPROVE DATA ANALYSIS

Matched filtering

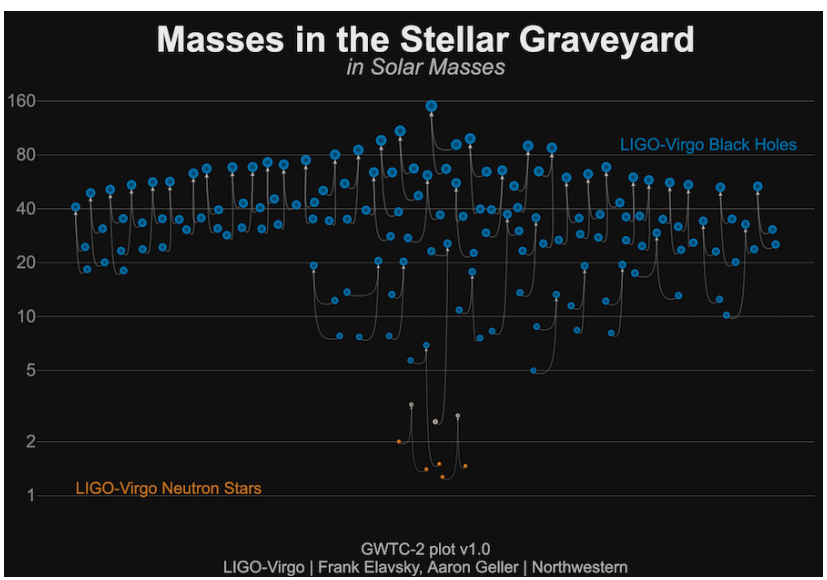
Alex Nitz https://youtu.be/bBBDR5jf9oU

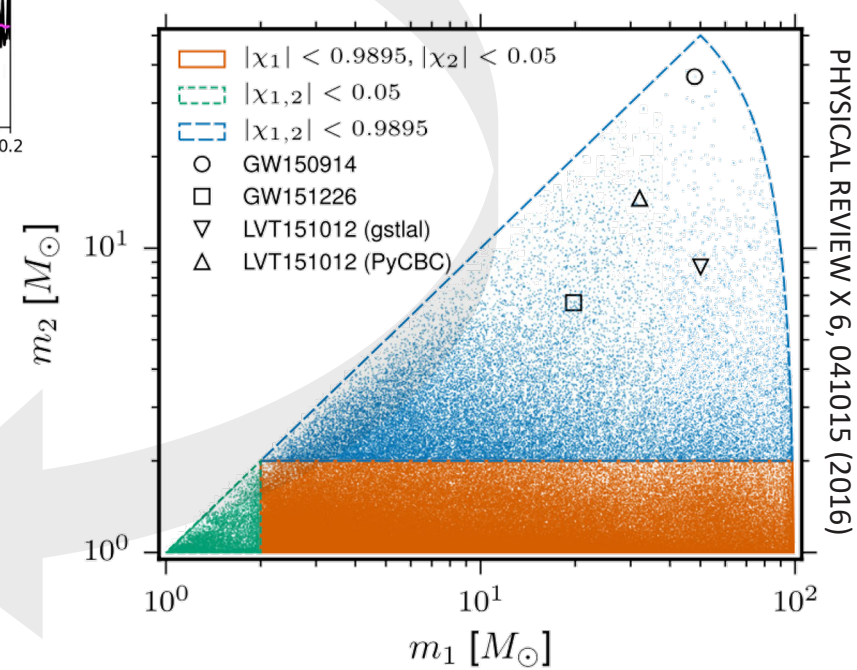


Matched filtering is the optimal detection statistics in Gaussian stationary noise

$$x(t_0) = 2\int_{-\infty}^{\infty} \frac{\tilde{s}(f)\tilde{h}^*_{\text{template}}(f)}{S_n(f)}\,df$$

arXiv:2010.14527v2





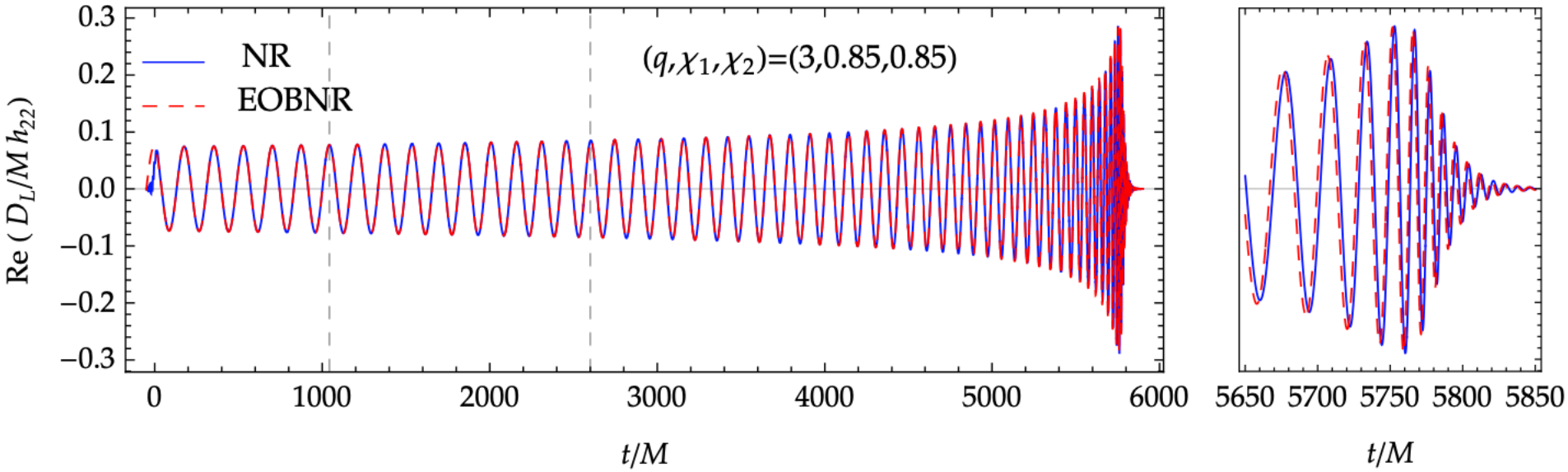
PHYSICAL REVIEW X 6, 041015 (2016)

Class. Quantum Grav. **37** 055002 (2020)

PHYSICAL REVIEW D 85, 122006 (2012)

ML for waveform simulation

Phys. Rev. D 95, 044028 (2017)



Accurate waveform rely on analytical inspiral models
merged / tuned with numerical relativity computation

Very **computationally demanding**
Need large number to cover parameter space (~500k) [arXiv:2010.14527v2]
Used surrogate / reduced order models [Phys. Rev. Lett. 114, 071104]

Machine Learning could improve the interpolation / model fitting

Phys. Rev. D **100**, 043005 (2019)
Class. Quantum Grav. **37** 075012 (2020)

ML for waveform simulation

Class. Quantum Grav. **37** (2020) 075012

Phys Rev D 96, 123011 (2017)
Gaussian Process Regression

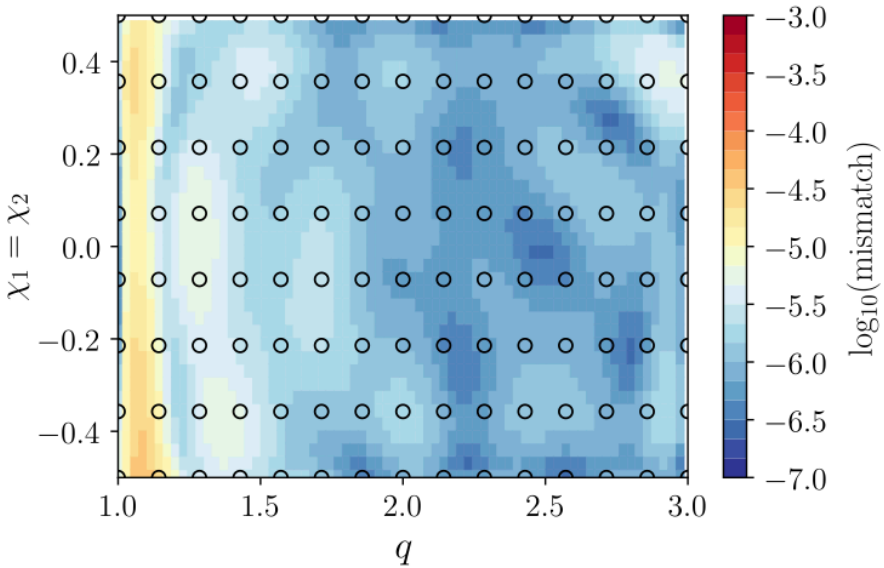


FIG. 9. Mismatch between IMRPhenomD waveforms and GPR mean waveforms with a regularly-gridded training set. The black circles show the locations of training waveforms from IMRPhenomD used to train the GPR. There are $15 \times 8 = 120$ training points on this grid, and the maximum mismatch in the region is 4.3×10^{-3}.

Phys Rev Lett 122, 211101 (2019)
Neural Network

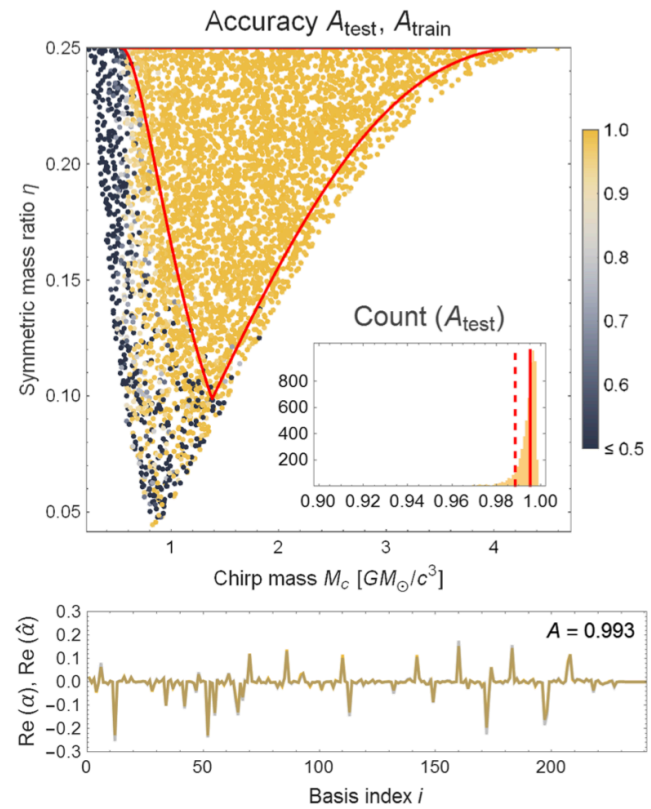
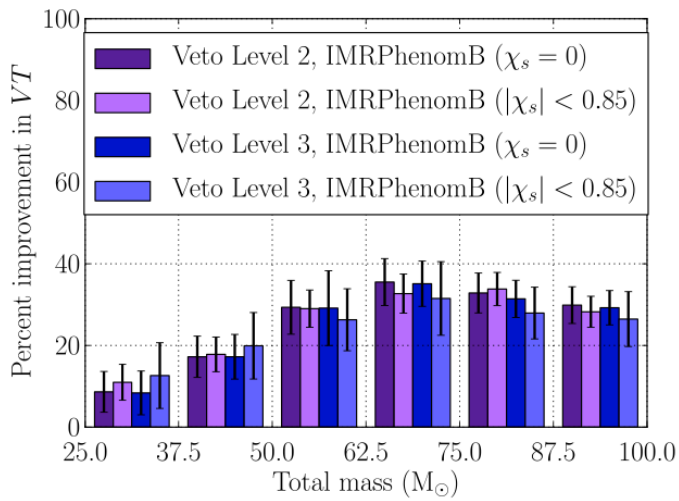


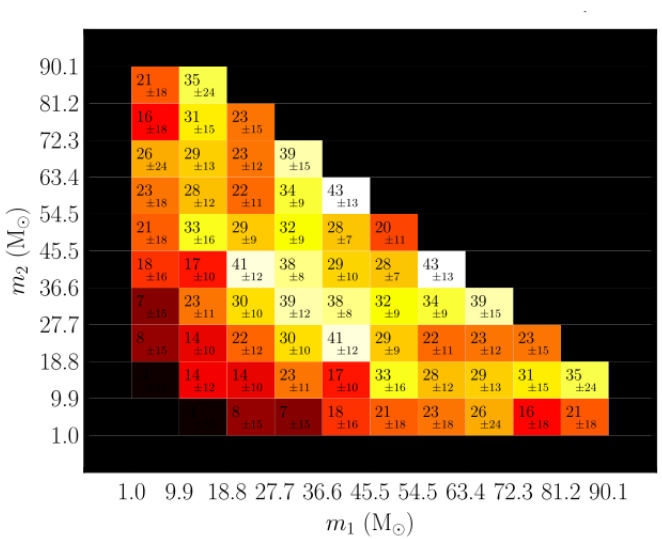
FIG. 2. Top: Plot of accuracy A as a function of (M_c, η) for test set (inside red border) and for 3000 training examples with (M_c, η) outside the domain of interest. Inset: Histogram of test-set accuracy values with tenth percentile (dashed line) and median (solid line) indicated. Bottom: Visualization of typical coefficient vectors $\mathrm{Re}(\alpha)$ (yellow) and $\mathrm{Re}(\hat{\alpha})$ (gray).

Class. Quantum Grav. **37** (2020) 075012

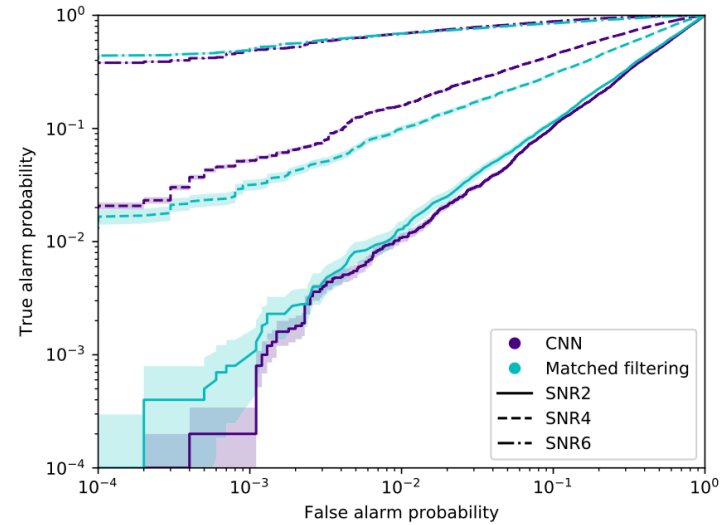
ML for Compact Binary Searches

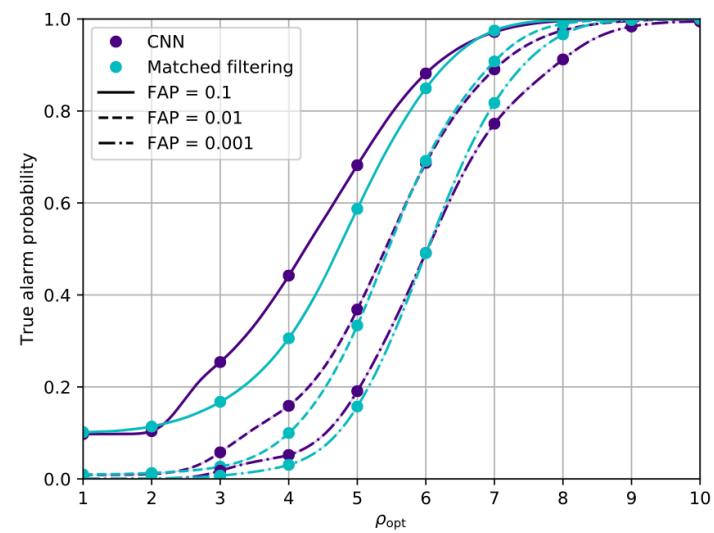
Phys. Rev. D **91**, 062004 (2015)
Random Forests of Bagged Decision Trees





Phys. Rev. Lett. 120, 141103 (2018)
Deep Convolutional Neural Network

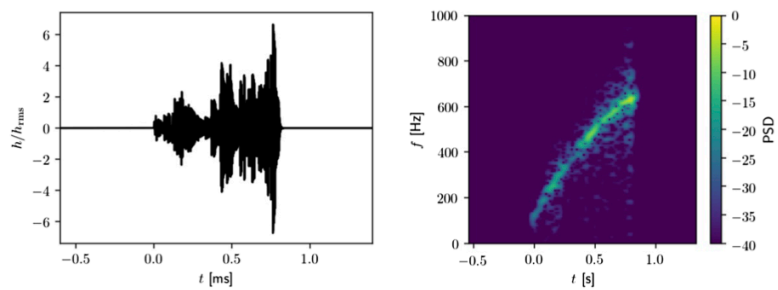




ML for other searches

Burst / Supernovae

- Reduce the effect of glitches using ML [Class. Quantum Grav. **34** 094003]
- Use CNN to search for modeled SN signals [Phys. Rev. D 98 122002]
- ML approach to reduce background [Mach. Learn.: Sci. Technol. 1 015005]



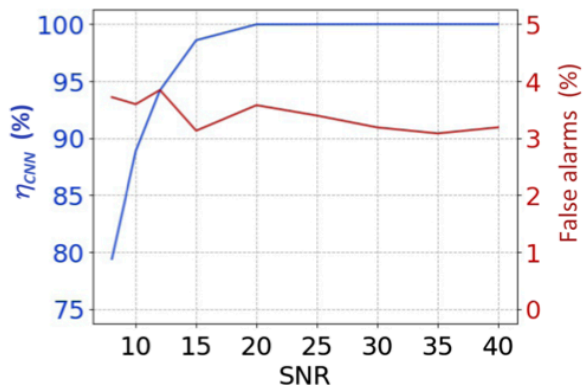


FIG. 6. η_{CNN} (Eq. (8) for different SNRs (blue line) computed during the validation process at threshold $\theta_{\mathrm{CNN}} = 0.5$. The red line is the false alarm probability [$\mathrm{FAR}_{\mathrm{CNN}}$ Eq. (8)] associated to the various SNRs computed at the same threshold. The used data set is based on 20000 signals and noise events for each SNR, half used for training and half for validation.

Continuous Waves

- DL for classification of CW sources [Mach. Learn.: Sci. Technol. 1 025016]
- Searches for unknown CW sources [Phys. Rev. D 100 044009]
- Hidden Markov models [Phys. Rev. D D93 123009]

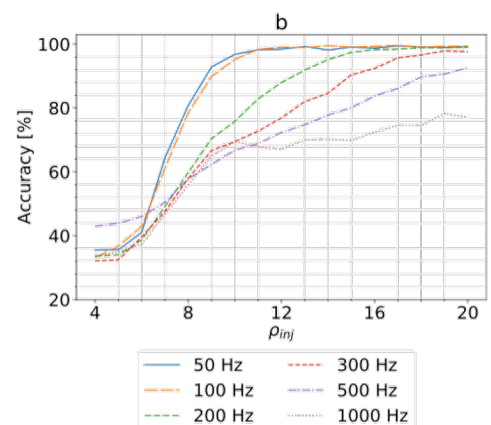
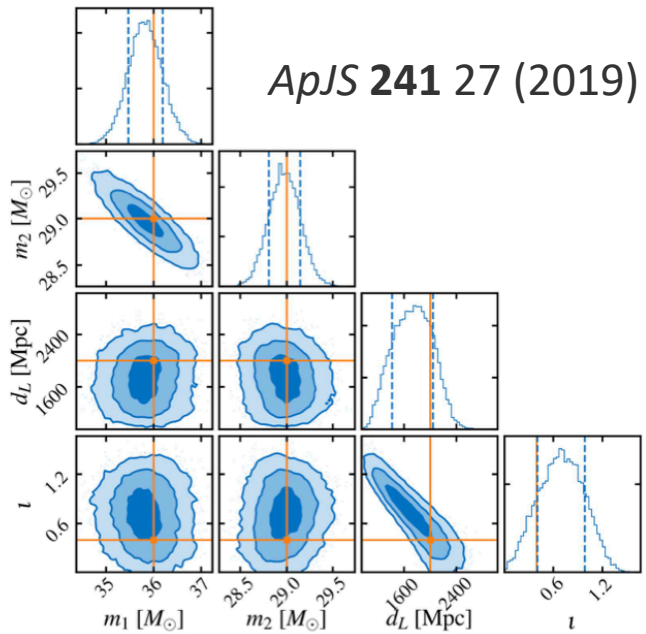


Figure 9. Evolution of accuracy as the function of the injected SNR ρ_{inj} for 1D CNN (a) and 2D CNN (b). The first model achieved a maximum level of accuracy of $\rho_{\mathrm{inj}} = 10-12$ and maintained its value for the whole range of frequencies. The 2D version varied significantly in relation to the frequency with the maximum accuracy being gradually shifted toward larger values of ρ_{inj}. The characteristic shift in accuracy (upper plot) between the lower frequencies (50 and 100 Hz) and the rest was associated with the density of signal candidates distributions. The **cgw** and **line** instances were easier to separate from **noise** since their distributions of parameters had very sparse character (see Fig 2 for comparison)—the **noise** signal candidates were not grouping around fluctuations in the frequency domain (the background of figure 2(a)), allowing easier classification than for higher frequencies.

ML for parameter estimation

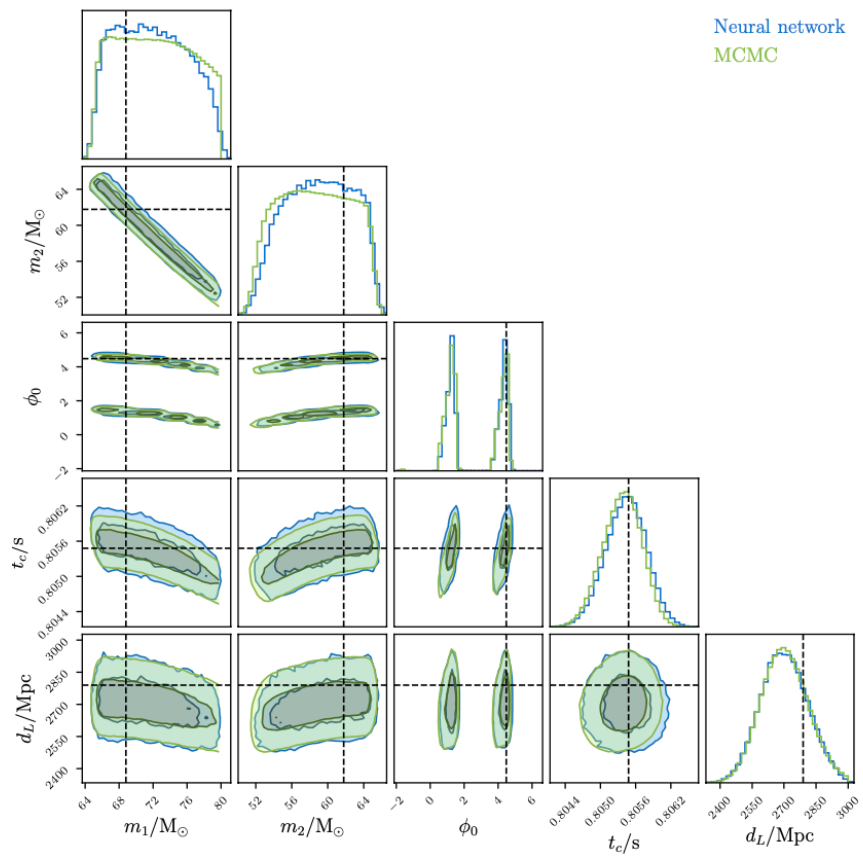


ApJS **241** 27 (2019)

Bayesian Inference is very computationally demanding

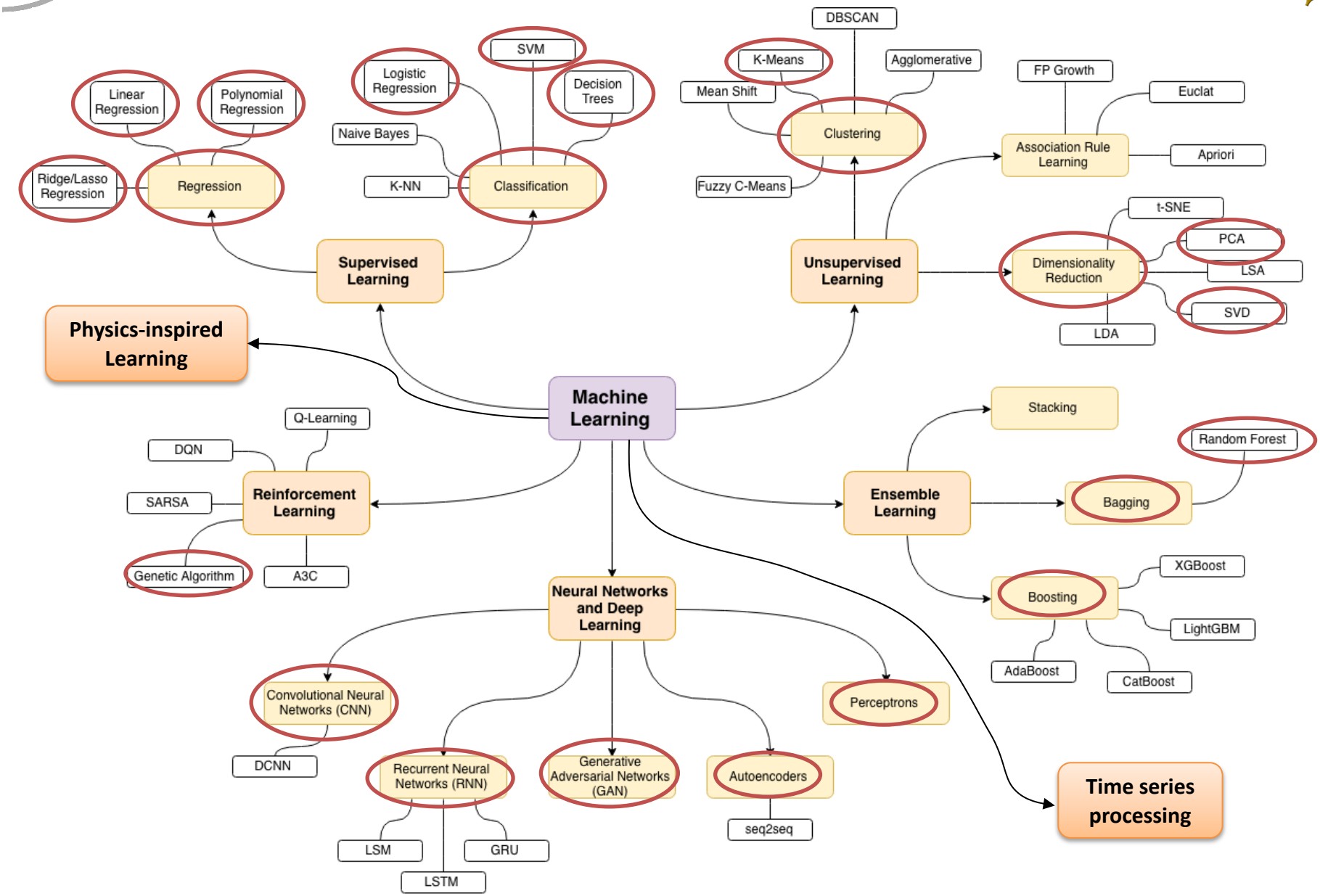
Not scalable to 3rd generation detectors with multiple events per day and very long signals (minutes instead of seconds)

Deterministic and Bayesian Neural Networks
arXiv:1903.01998 [gr-qc]
Variational Autoencoders
arXiv:1909.06296 [astro-ph.IM]
Autoregressive Neural Network Flow
arXiv:2002.07656 [astro-ph.IM]



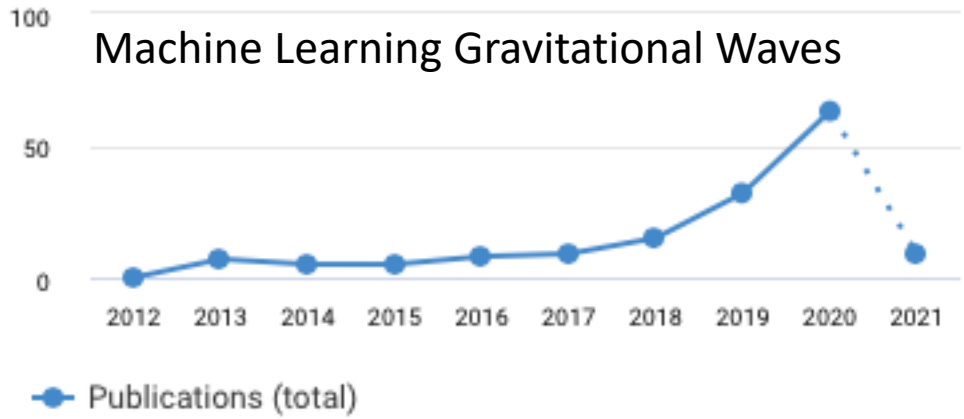
CONCLUSIONS

ML and GW

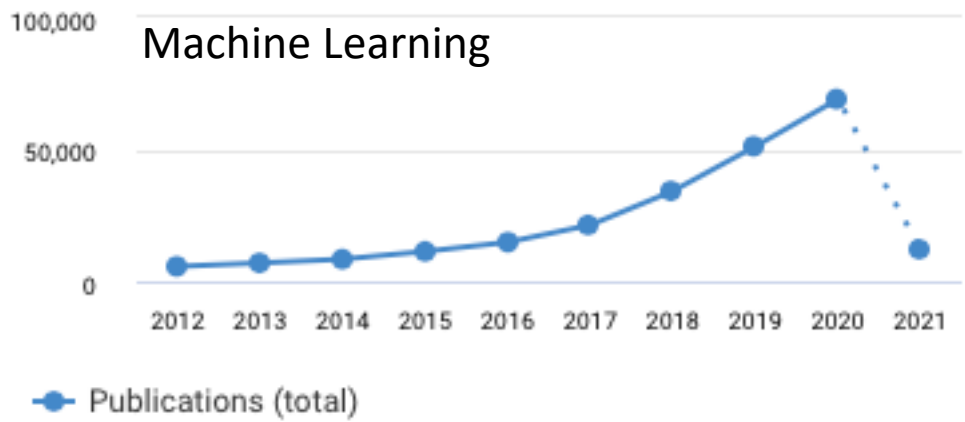


https://i.pinimg.com/originals/d9/ba/26/d9ba26bb7ee690d54b2a6e057ed87481.png

ML and GW



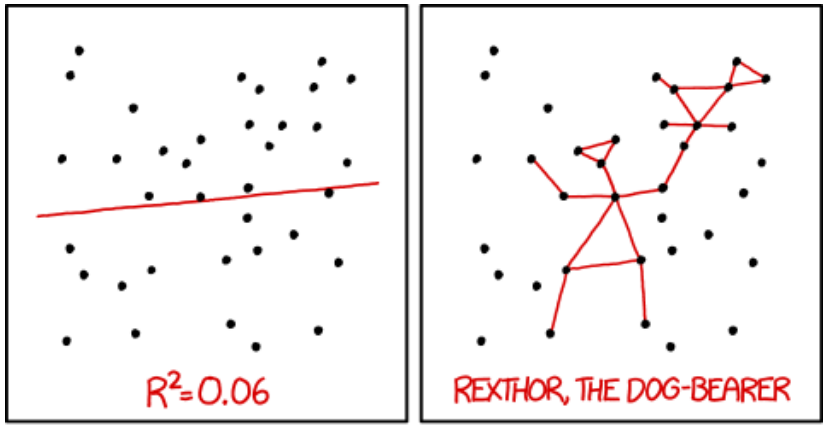
PUBLICATION YEAR	
2021	9
2020	63
2019	32
2018	15
2017	9
2016	8
2015	5
2014	5
2013	7
2010	1



PUBLICATION YEAR	
2021	12,146
2020	69,048
2019	51,167
2018	34,234
2017	21,363
2016	14,831
2015	11,339
2014	8,400
2013	6,977
2012	5,659

https://app.dimensions.ai/discover

The right tool for the job



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.