Data Analysis Challenges

C. Palomba, T. Regimbau, G. Prodi,
J. Veitch, *T. Dent* + guests

Pax VI, 29th May 2019

(Attempted) Schedule

- ~40-50m ‘structured time’
  - 1 ‘question’ or ‘open issue’ per panellist
  - Write down spinoff topics, return to them in ...
- remainder : ‘unstructured time’
  - will prioritize *unsolved* questions

Only one question:
Will we ever detect a CW signal?

Cristiano Palomba – INFN Roma

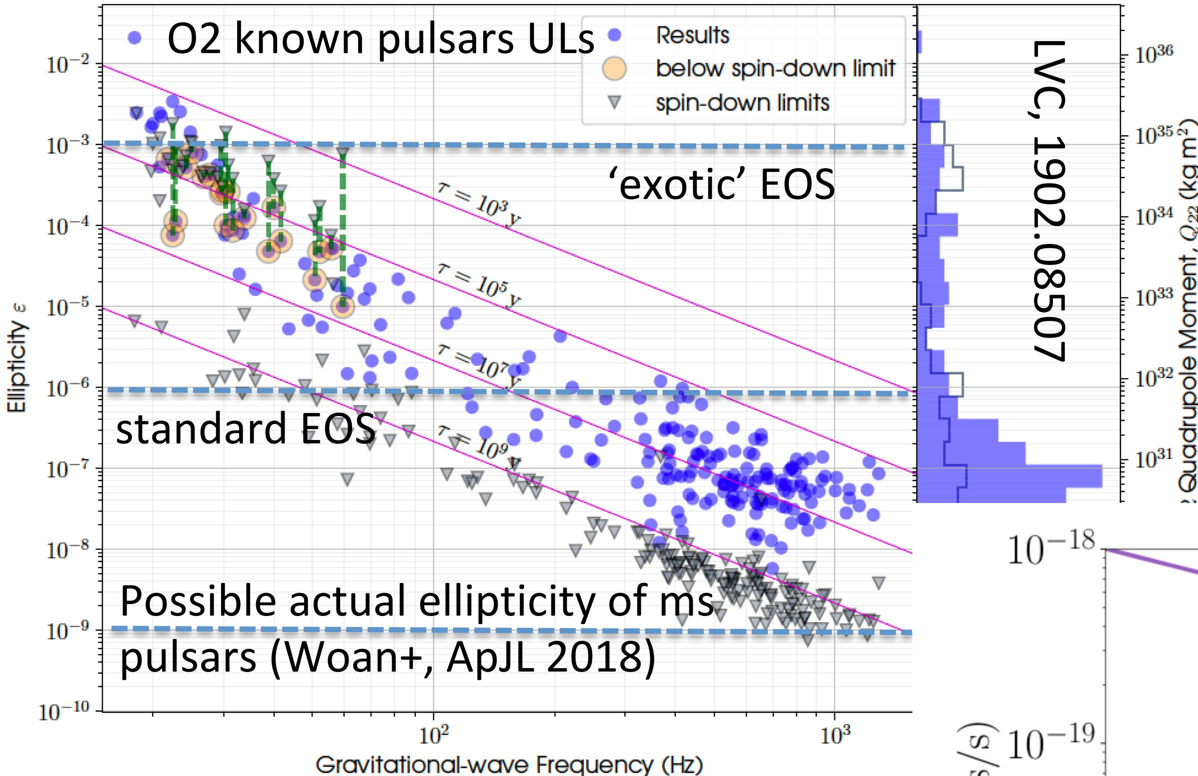
Only one question:
Will we ever detect a CW signal?

A) Expected Signal amplitudes vs detector sensitivity: where do we stand?

B) How can we improve analysis methods sensitivity?

C) Analysis methods robustness: What if there are deviations from the signal model? Noise impact?

Cristiano Palomba – INFN Roma

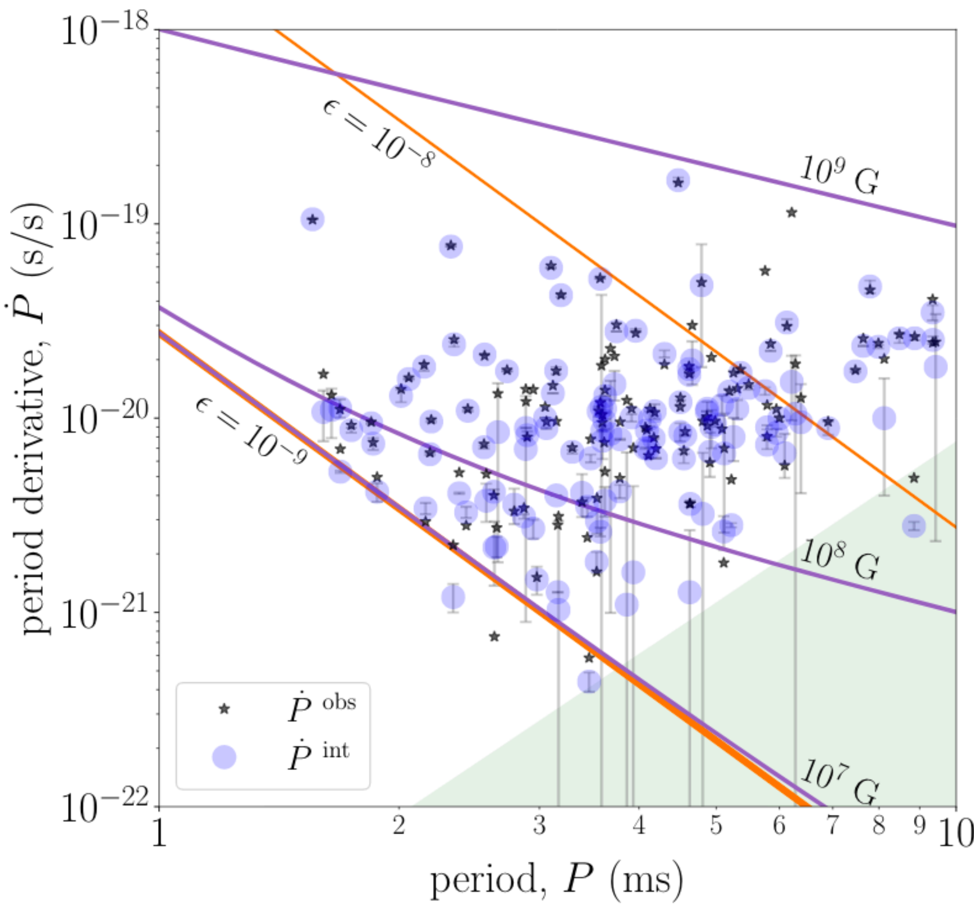


A: where do we stand?

$$h_0 = \frac{4\pi^2 G}{c^4}\frac{I_{zz}\varepsilon f_0^2}{d}$$

CW emission constrained for 21 pulsars

In O3 we could constrain emission also for J0711-6830, J0437-4715



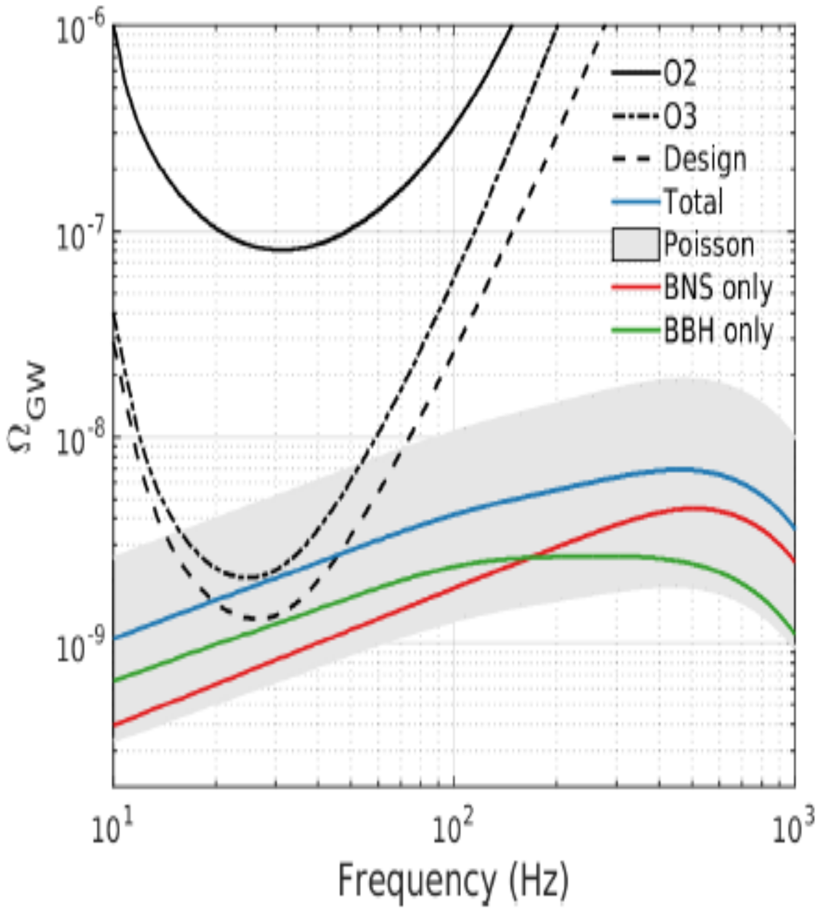
So, what can we do?

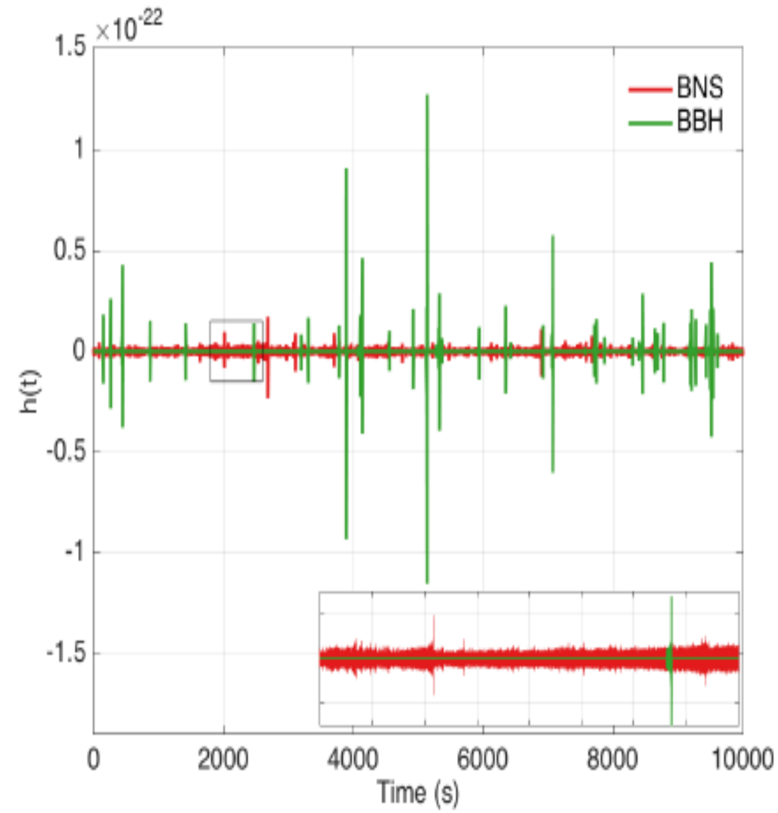
- Keep improving analysis pipelines (next questions)
- Search for ‘new’ potentially interesting sources (e.g. ultra-light bosons clouds around BHs)
- Wait for better and better detector sensitivity
- Keep observing….

Stochastic background challenges

- The events we detect now are loud individual sources at close distances ($z \sim 0.1$-0.5 for BBHs and $z \sim 0.01$ for the BNS).
- Many more sources at larger distances contribute to create a stochastic background that dominates in the band of LIGO/Virgo.
- With 2G the goal is to detect this background. How do we separate the different contributions BBH/BHNS/BNS or field/dynamical/primordial?
- With 3G the goal is to subtract it and observe the cosmological background below.

Separate different contributions?





Abbott et al. PRL, 120.091101 (2017)

Remove the CBC background

- Total background: $\Omega_{\text{tot}} = \Omega_{\text{cbc}} + \Omega_{\text{astro}} + \Omega_{\text{cosmo}}$

- 3G detectors will be able to resolve a large number of CBCs. If we remove the estimated waveforms from the data:

$$\Omega_{\text{tot}} - \hat{\Omega}_{\text{cbc}} = \Delta\Omega_{\text{cbc}}^{\text{resolved}} + \Delta\Omega_{\text{cbc}}^{\text{unresolved}} + \Omega_{\text{astro}} + \Omega_{\text{cosmo}}$$

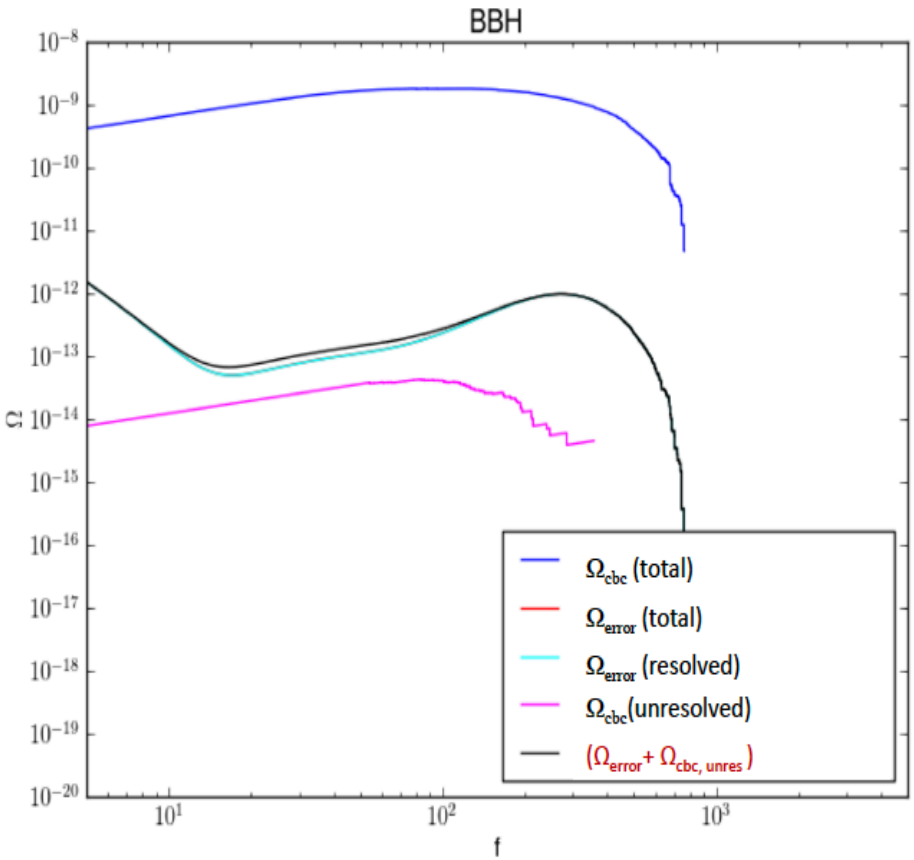
- In the ideal case the foreground can potentially be subtracted to the level $\Omega_{gw} = 2 \times 10^{-13}$ (Regimbau et al., PhysRevLett.118.151105)

- More realistically using Fisher matrix PE (S. Sachdev et al. in preparation)

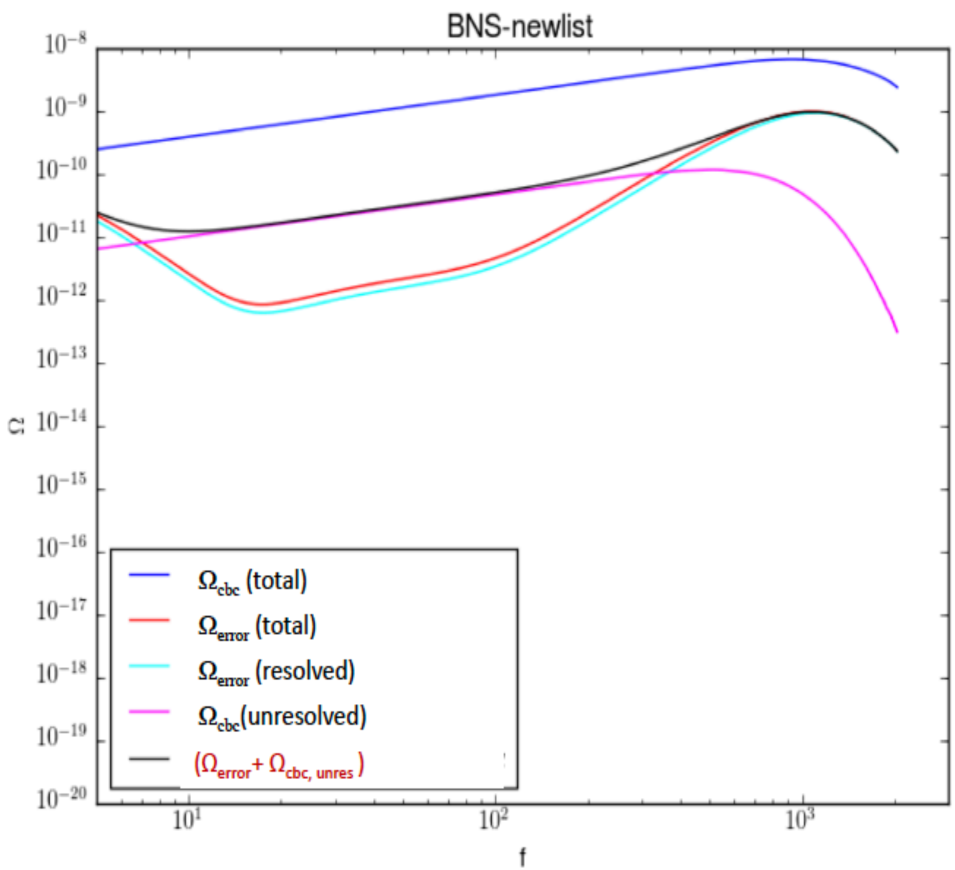
Remove the CBC background

S. Sachdev et al. in preparation

BBH: residual dominated by PE uncertainty

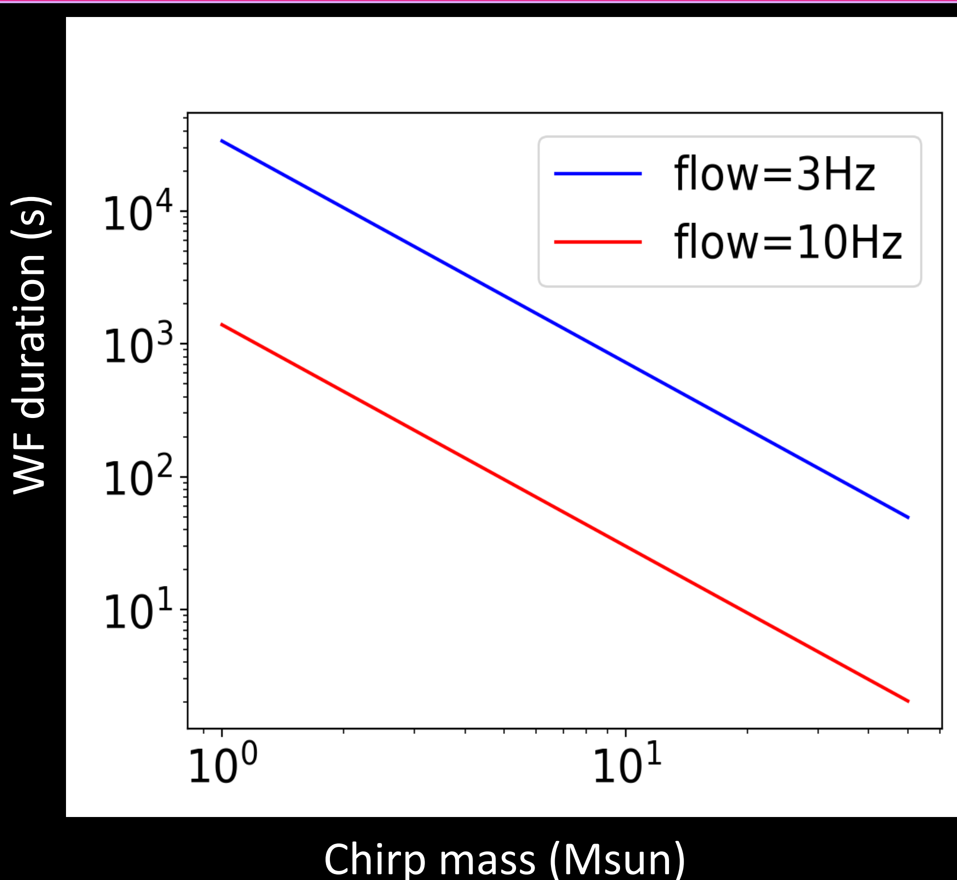


BNS: residual dominated by unresolved sources



Guest appearance from audience: S. Vitale

Low frequency – computational challenges

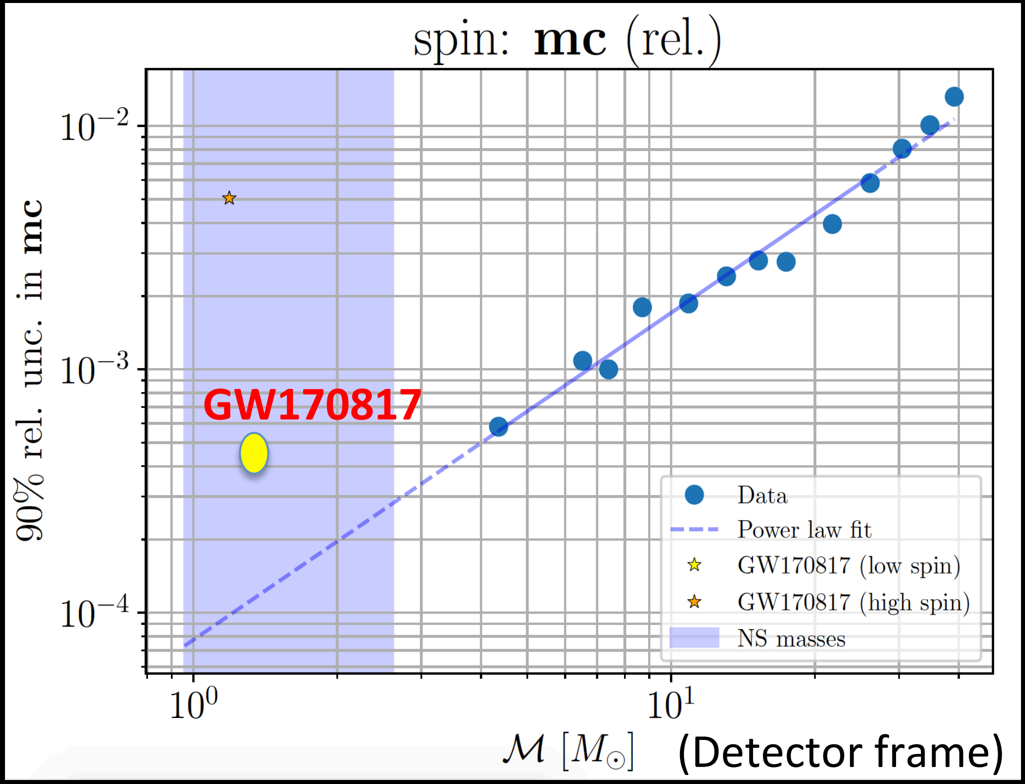


- We would love to know how well CBC parameters can be measured in 3G. However…
- Duration of waveforms (& hence computational time required) blows up
  - As f_low decreases
  - As chirp mass decreases
- With current methods we **cannot** run parameter estimation codes on BNS in 3G

Chirp mass extrapolation

SNR=32, CE

flow=10

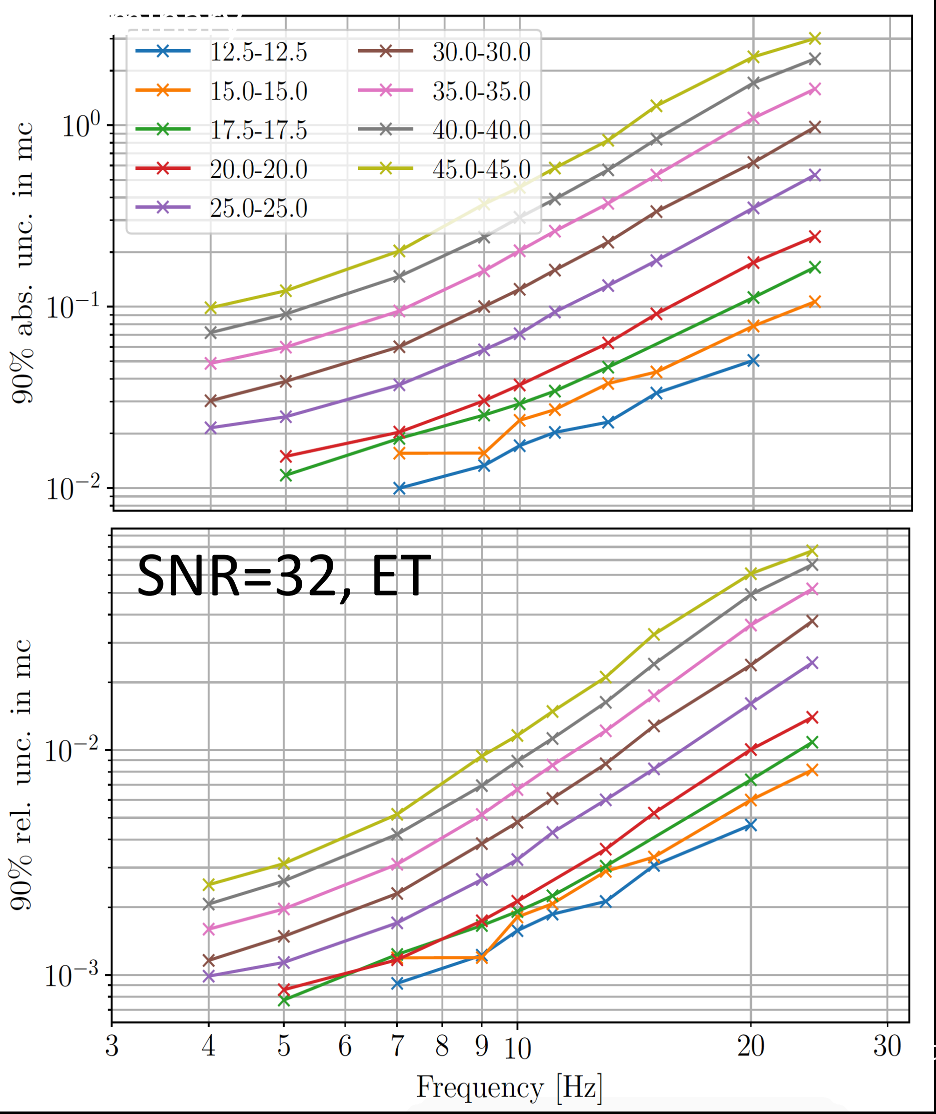


Whittle+, preliminary

- Run actual PE (MCMC not Fisher matrix) for decreasing true chirp mass
- See if results can be extrapolated to the BNS mass region
  - Sathya working on similar ideas

Lower-frequency extrapolation

Whittle+,

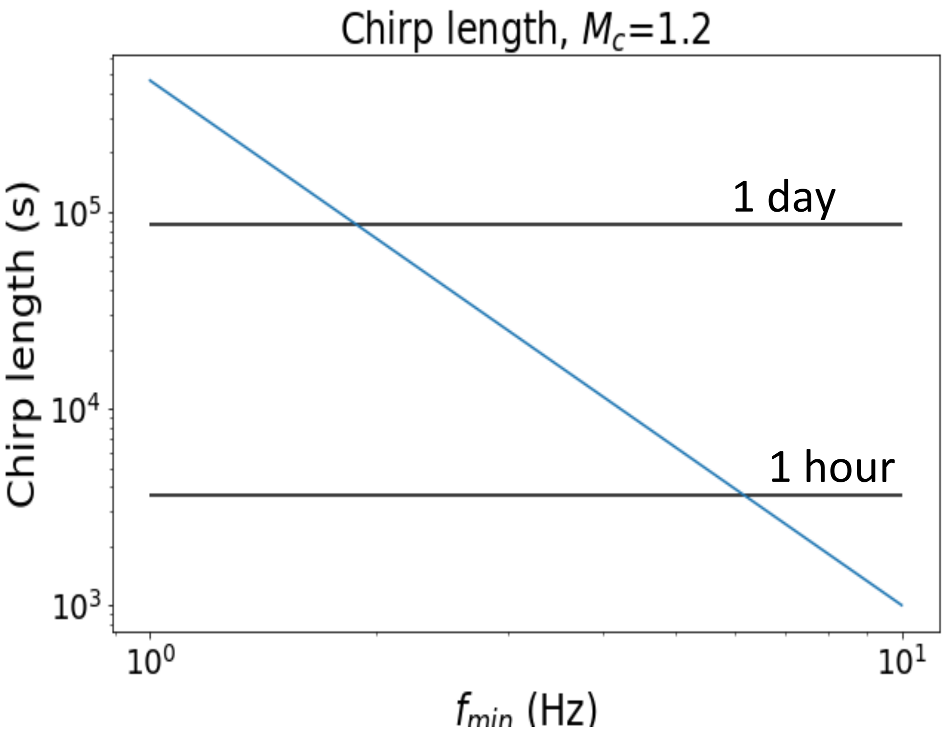


- Run actual PE (MCMC, not Fisher matrix) for decreasing f_low
- See if results can be extrapolated to the BNS mass region
- Ongoing : 2D mass/f_low extrapolation
- Caveat : don't have time dependent antenna patterns
  - Does anybody?

Looooong signals

In 3G and LISA

Long Signals (BNS / 3G)

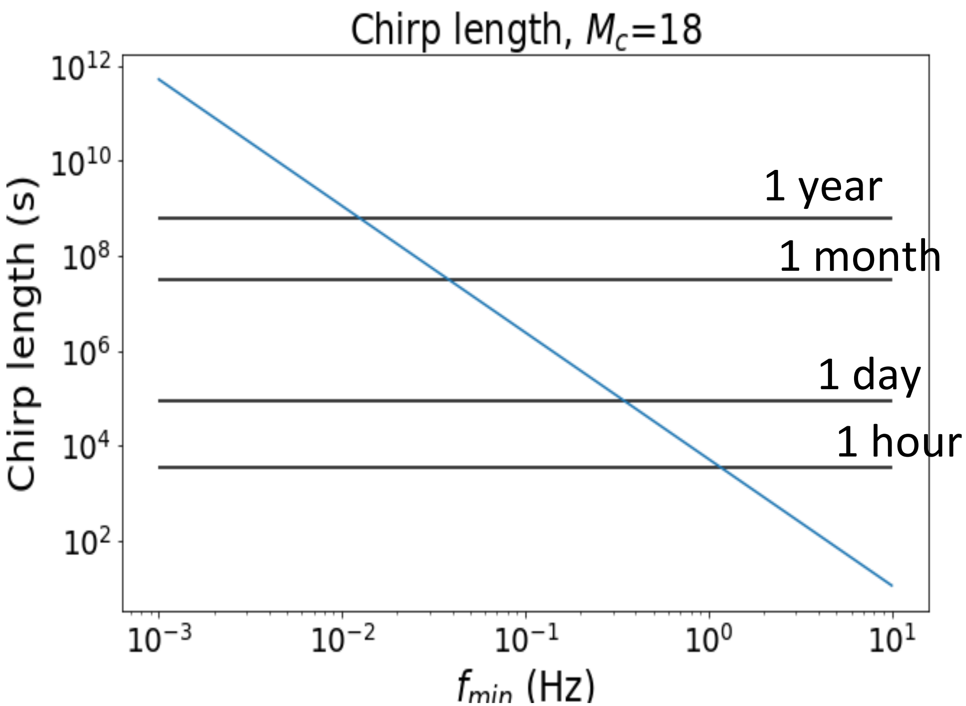


Signal length $\propto f^{-8/3} \, M_c^{-5/3}$

Data analysis challenges:

- Sheer data volume - FFTs O(NlogN)
- Template generation & memory
- Template bank size (M.F. searches)
  - Solved for search pipelines?
- Amplitude and phase modulation from Earth's rotation
  - Confusion with precession / subtle phasing effects?
- Chance of glitches during signal approaches 100%
  - Non-gaussian noise
- Power spectrum changes during signal!
  - Non-stationary noise

Long Signals (LISA / smBBH)



LISA analysis for binaries?

Simulation of LISA response is computationally expensive (TDI 2)

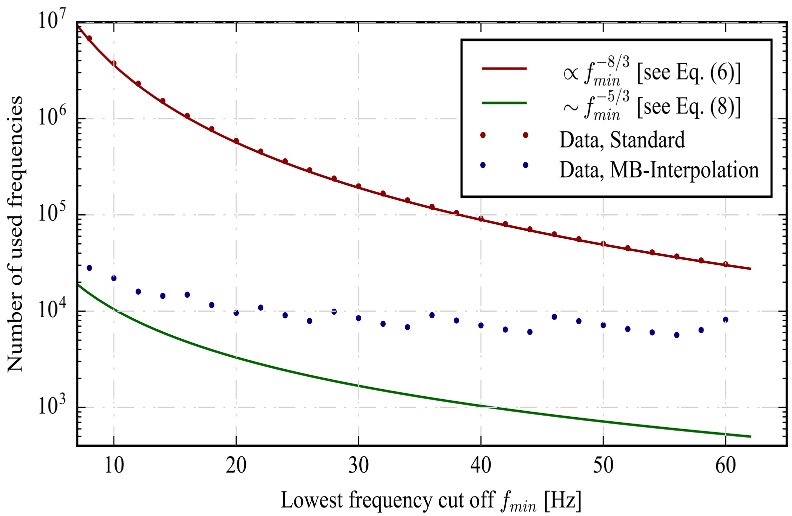
Non-stationarity, non-gaussianity?

Overlapping signals (DWD + CBC + EMRI…)

How best to simulate multi-band analyses?

Computational cost: Solutions?

- Multi-band / sparse sampling?
  - Sub-Nyquist: Acceleration scales as f_max / f_min at best
  - Super-Nyquist: aliasing, data compression schemes?
- ROQ?
  - Creating the basis is extremely expensive (asymptotic order N^2??)
  - Number of bases increases like template bank density
  - Compatibility with sky-dependent amplitude modulation?
- GPU computing?
  - One-off benefit. Needs to be ~million times faster
- Machine learning
  - Possible, but how to train?



Vinciguerra+ 1703.02062

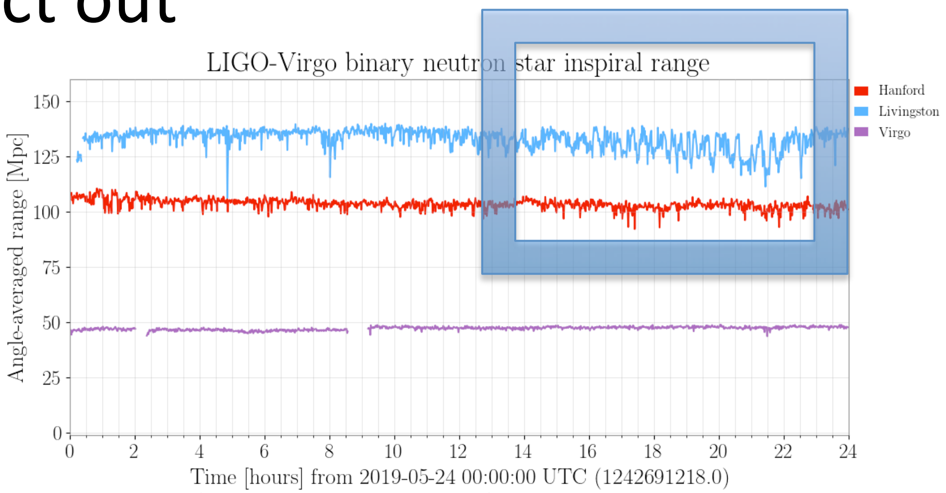
Analysis Method Challenges

- Superimposed signals / glitches
  - CLEAN: fit and subtract (then repeat)
  - Global fit for multiple signals
    - Computational cost issues with very long signals
    - RJMCMC algorithm
      e.g. Umstätter for WD binaries, Cornish/Littenberg BayesWave
  - Other ideas? Viterbi / unmodelled as first step in hierarchical analysis
- Varying PSD over signal timescale
  - Time domain analysis?
  - Can we compute optimal basis? (ROQ-like methods)
    - Location-dependent?
  - Knock-on impact on PE from differing PSD estimates?

Guest appearance #2 : Sathya / Early Warning

Glitch detection / removal

- Operational definition of glitch : Excess power that is not coherent between detectors
- For physically separated detectors, can spot glitches individually & gate/subtract out
  - what is glitch 'duty cycle' ?
  - some days show near continuous scattering over ~hours



- Co-located ifos may have partly 'coherent' glitches?
- ET-style null stream may not catch all glitches?

Only one question:
Will we ever detect a CW signal?

A) Expected Signal amplitudes vs detector sensitivity: where do we stand?

B) How can we improve analysis methods sensitivity?

C) Analysis methods robustness: What if there are deviations from the signal model?

Cristiano Palomba – INFN Roma

B) How can we improve analysis methods sensitivity?

Cannot do better than matched filter (ideally)

Increase available computing power/code efficiency

Restrict/optimize parameter space

Improve sensitivity of the incoherent step in semi-coherent analysis

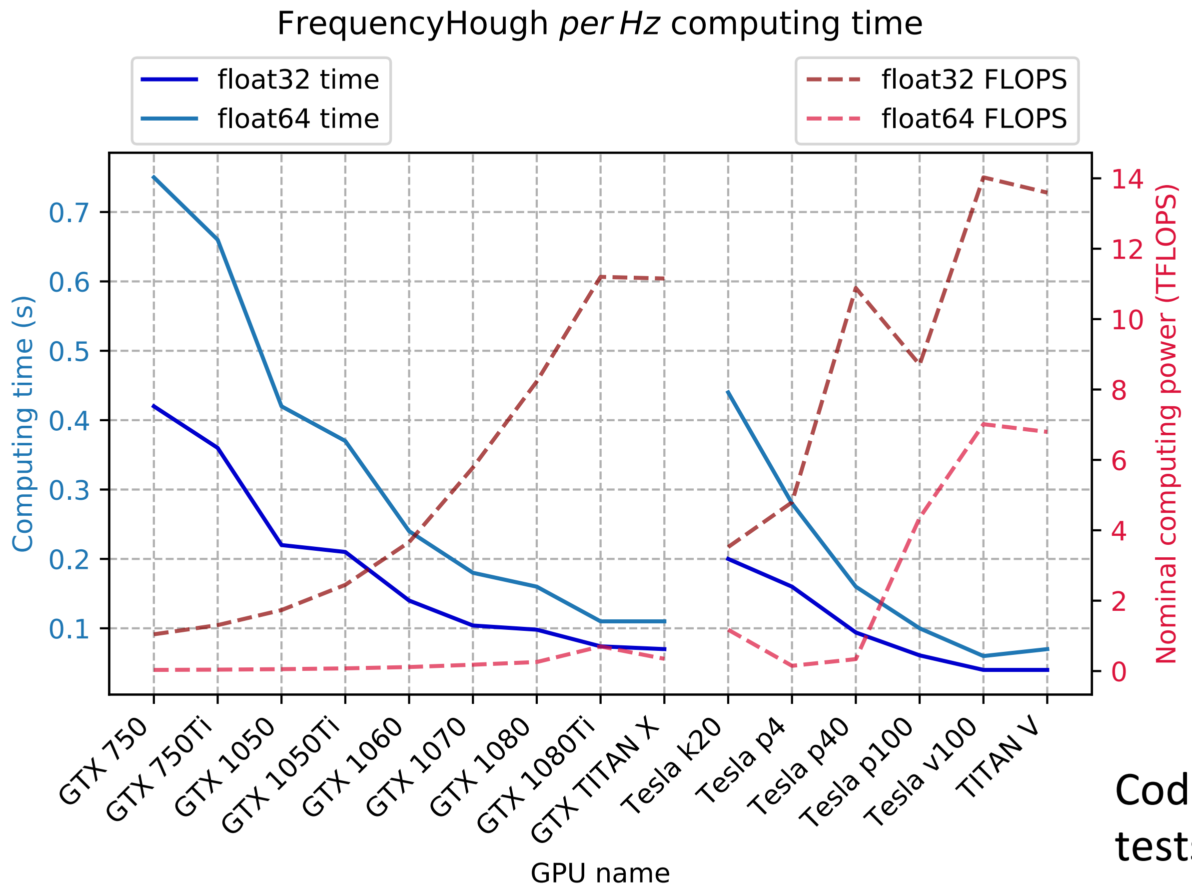
Where it is better to spend our precious computing power?

With EM input (see MM session)

GPUs are a possibility we are considering.

Tests using the FrequencyHough Transform code have shown a speed-up of 15-20 on a ‘good’ GPU

Still space for improvement



Code porting and tests by Iuri La Rosa

Attempts to narrow down GW parameter ranges using EM observations [e.g. Rowlinson+ 2013, Lasky+ 2017]

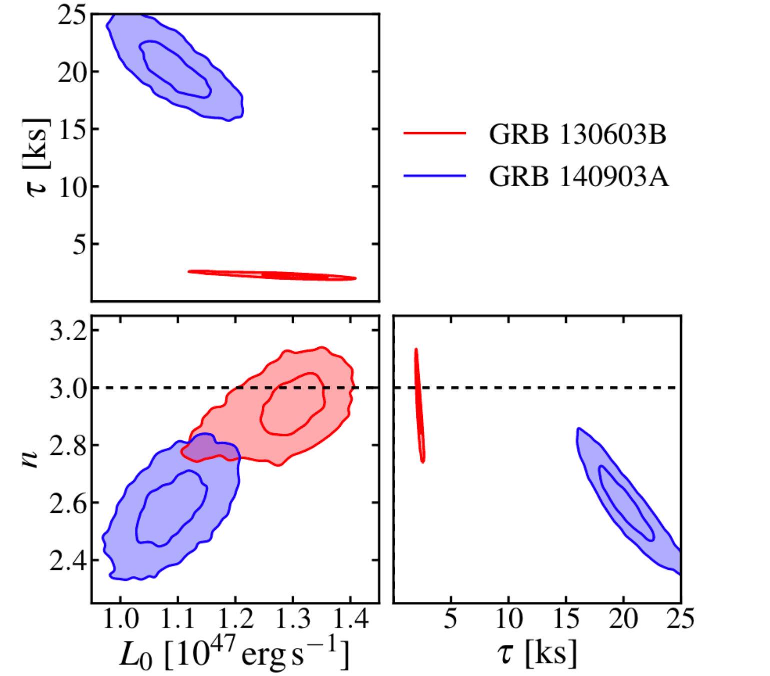
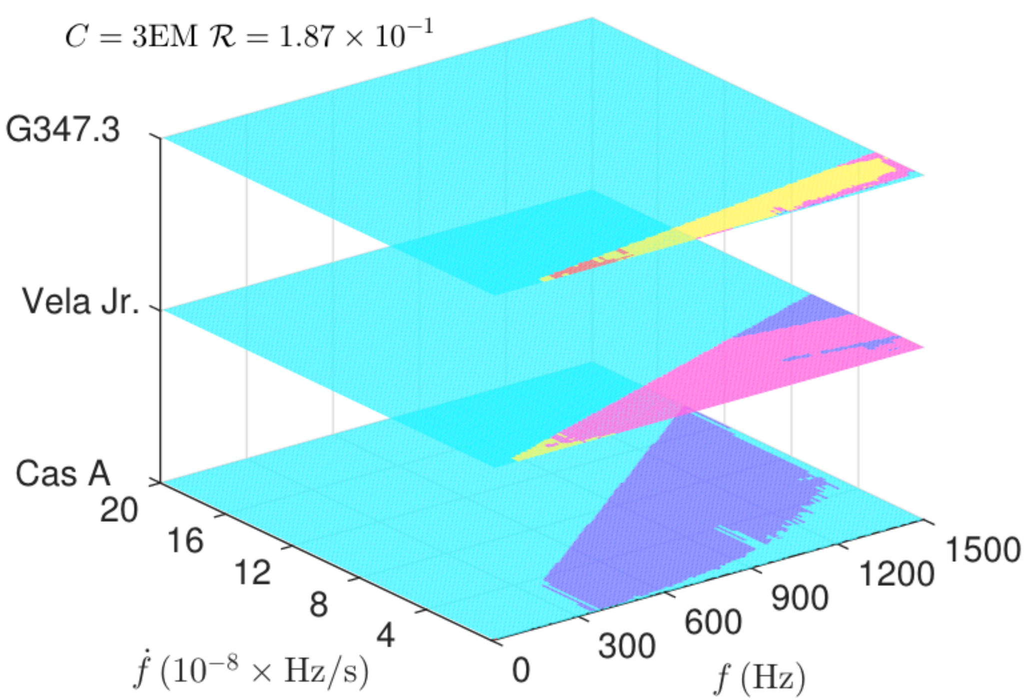


FIG. 2.— Posterior probability distributions for the parameters in Eq. (2) for GRB 130603B (red) and GRB 140903A (blue). The contours show the one- and two-sigma confidence intervals, and the dashed line indicates the fiducial value of $n = 3$.

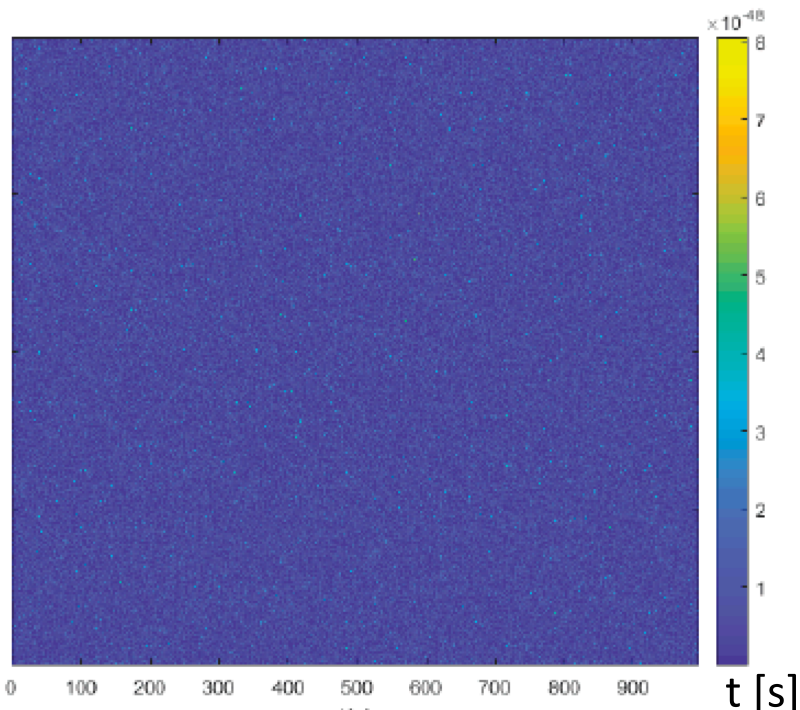


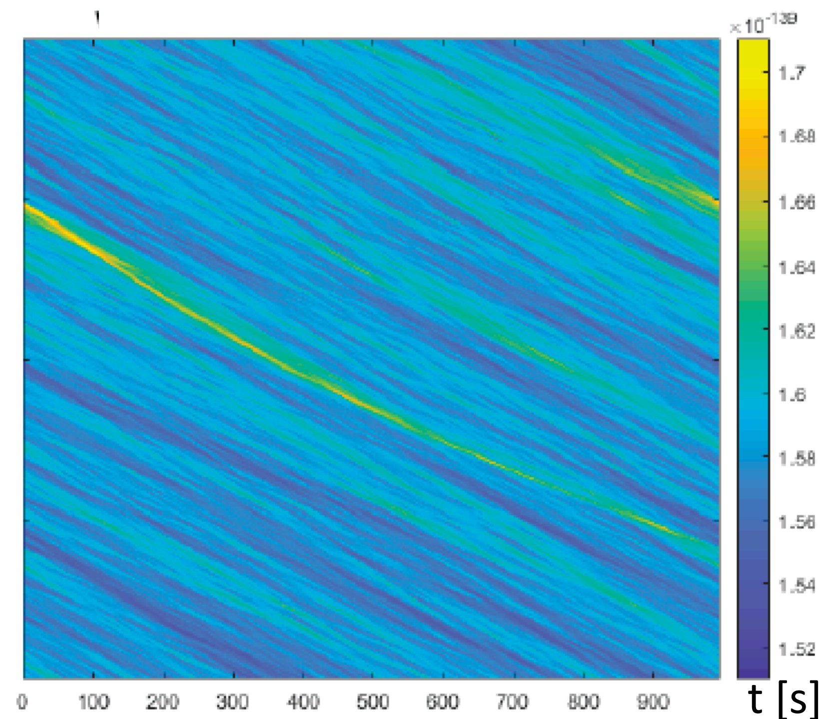
An optimization scheme for directed searches has been presented in Ming+ 1708.02173

In a semi-coherent search we typically select 'peaks' in a spectrograms, which are then processed

By increasing the probability of selecting signal peaks we increase the overall sensitivity

Image processing techniques, eg based on 2D FT, appear promising [Pierini+, in prep]



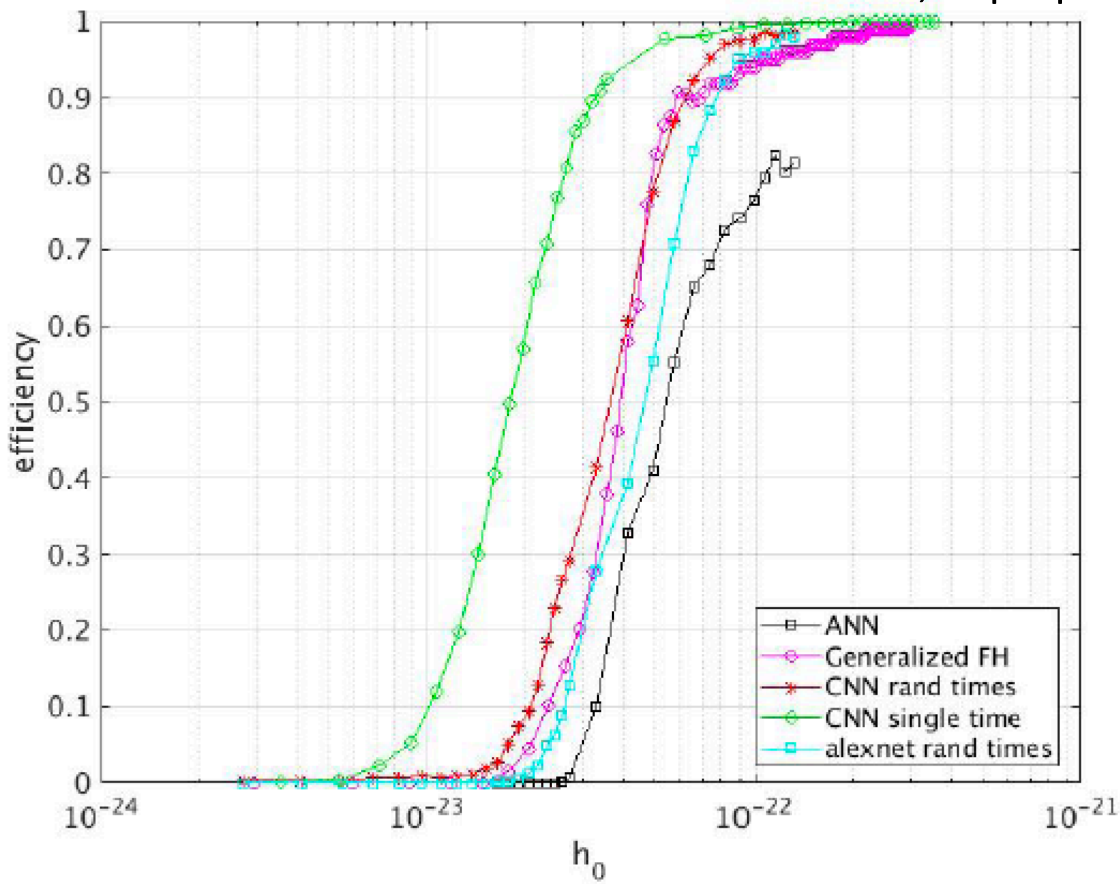


C) Analysis methods robustness: What if there are deviations from the signal model? Noise impact?

- Robustness is crucial
- Different paths are being investigated
- Machine learning seems promising for long-transient searches (duration of ~hours)
- Image processing + machine learning could be promising for ‘standard’ CW searches

Detection efficiency for the search of long-transients with various ML implementations and a 'classical' modeled search (Generalized FrequencyHough)

Miller+, in prep



- Exploit signal features, e.g. sidereal modulation, to discriminate among signals and noise

- Follow-up of candidates can be prone to even small deviations from expected models: going ‘deeper’ can be dangerous

- Robust methods, like those based on Viterbii algorithm, have been implemented but are somewhat limited in sensitivity