Polgraw-allsky: GPU-accelerated all-sky blind search for periodic gravitational waves

Michał Bejger (Nicolaus Copernicus Astronomical Center),
in collaboration with Paweł Ciecieląg
for the POLGRAW-Virgo group

Wigner GPU days, 2.6.16

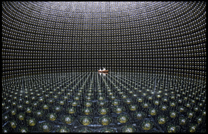
Outline

- ⋆ Gravitational waves and the detection principle,
- ⋆ Astrophysical sources of gravitational waves,
- ⋆ Search for ”gravitational pulsars” - polgraw-allsky pipeline for almost-monochromatic signals.

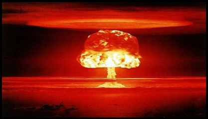
Four fundamental interactions

- ⋆ Electromagnetism (light, X-rays, radio, heat, material properties...) — ⋆ γ photon

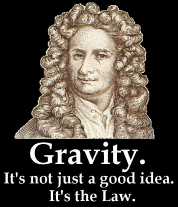
- ⋆ Weak interactions (radioactivity, decays, neutrinos) — ⋆ $W^{\pm}$ and Z^0 bosons



- ⋆ Strong interactions (stability of nuclei) — ⋆ gluons



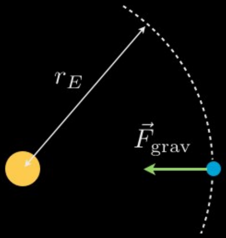
- ⋆ Gravitation — ⋆ graviton (?)



Gravitation by Newton and Einstein

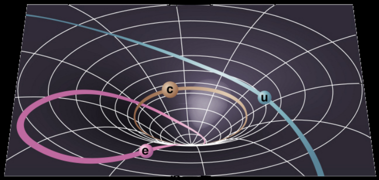
Newton: Gravitational *force* acting immediately at a distance:

$$F = G\frac{m_1 m_2}{r^2}$$

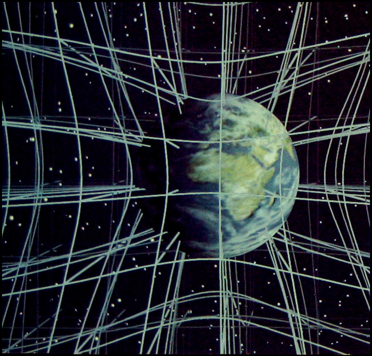


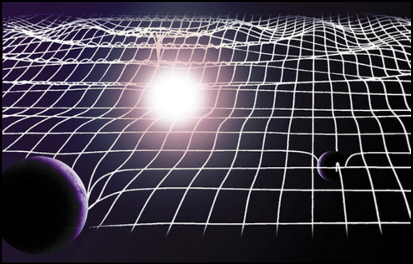
Einstein: free motion of bodies along „straightest" trajectories in *curved* spacetime.

Gravity propagates at the speed of light.

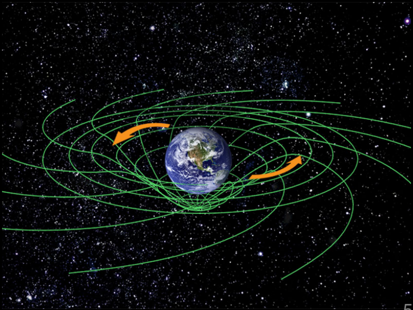


Einstein (1915): gravitation *is* the geometry of spacetime

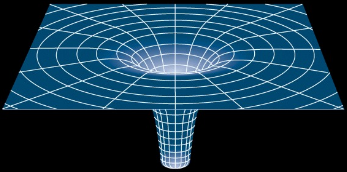


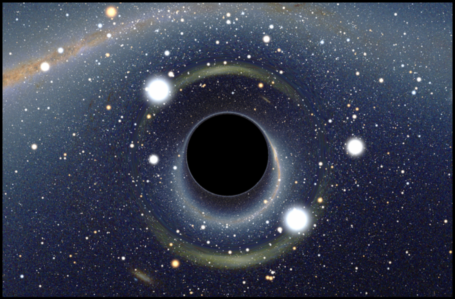


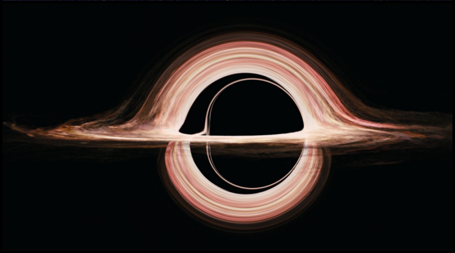
"Mass tells spacetime how to curve, and spacetime tells mass how to move."



Schwarzschild (1916): black holes

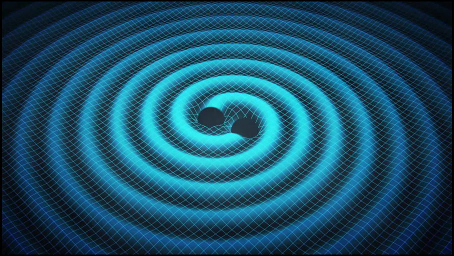






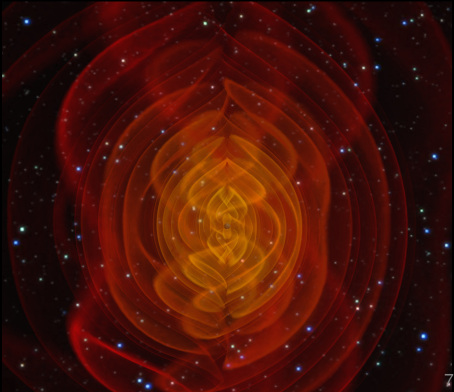
- ⋆ Region of spacetime curved so strongly that the escape velocity from it is larger than *c*.
- ⋆ Objects with masses between a few $M_\odot$ and billions of $M_\odot$.
- ⋆ Black holes may rotate (Kerr 1963).

Gravitational waves



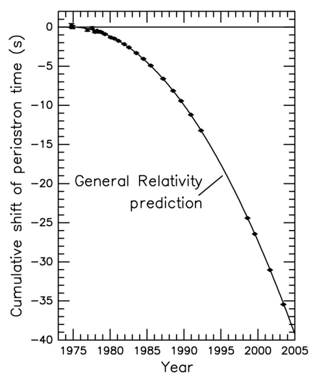
Einstein (1916) - there are wave solutions to GR equations *(time-varying distortions of the curvature propagating with the speed of light)*:

- ⋆ transverse in nature,
- ⋆ generated by accelerated motion of masses,
- ⋆ two polarizations ($+$ i $\times$).



Gravitational waves: indirect evidence

70s - observations of pulsars in relativistic binary systems (e.g. Hulse-Taylor pulsar):



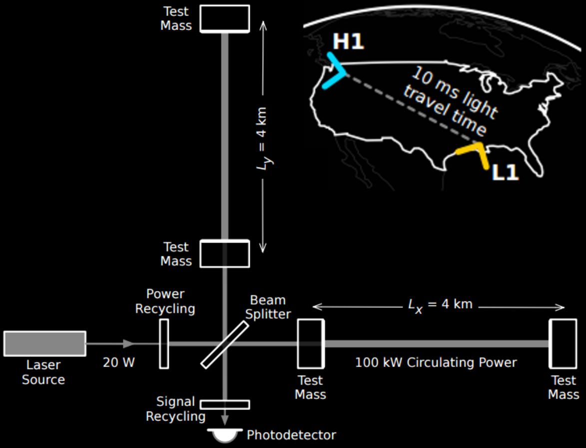
System is losing energy by emitting gravitational waves.

60s - remarkable insight of Bohdan Paczyński:

- *"Gravitational Waves and the Evolution of Close Binaries", AcA 1967* - orbital period evolution of WZ Sge and HZ29 driven by the GW emission.

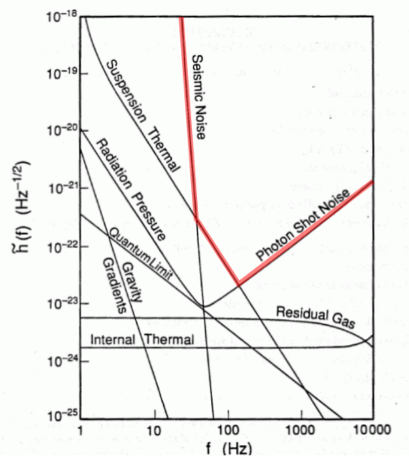
Detection principle: laser interferometry

"How to measure distance when the ruler also changes length?"



Changes in arms length are **very** small: $\delta L_x - \delta L_y = \Delta L < 10^{-18}$ m (smaller than the size of the proton). Wave amplitude $h = \Delta L / L \leq 10^{-21}$.

How the sensitivity curve looks like?



Initial LIGO proposal (1989)

- ⋆ Range of frequencies similar to human ears:

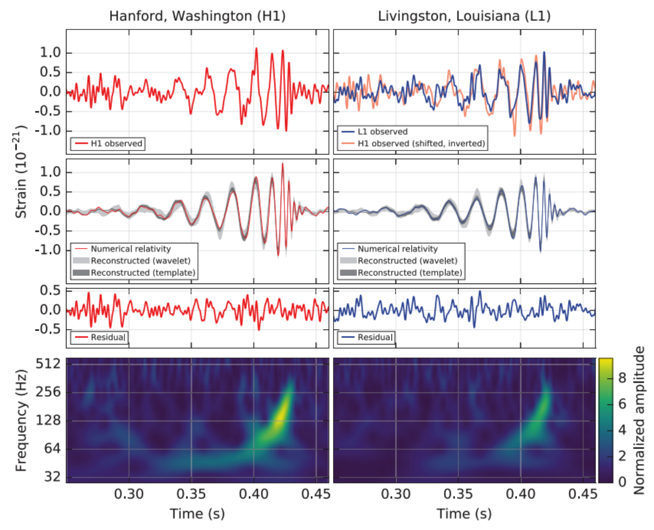
From 20 Hz (H0) to a few thousands Hz (3960 Hz, H7) - 8 octaves.

- ⋆ Poor, like for an ear, angular resolution.

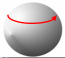
LIGO-Virgo detector network

GW150914: black hole binary system

14th of September 2015 r. both LIGO detectors (first Livingston, Hanford 7 ms later) registered the same signal:

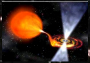


- Periodic sources
  - Binary Pulsars, Spinning neutron stars, Low mass X-ray binaries

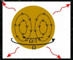


"Mountain" on neutron star

Wobbling neutron star

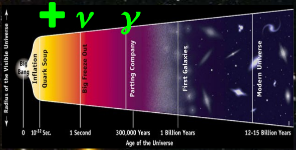


Accreting neutron star

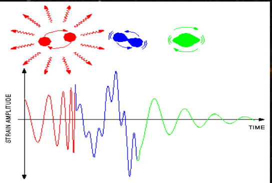


R-modes

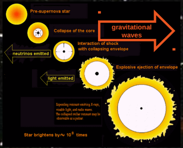
- Stochastic background
  - Primordial Big Bang (t = 10^{-22} sec)
  - Continuum of sources

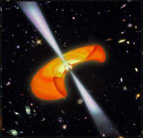


- Coalescing compact binaries
  - Classes of objects: NS-NS, NS-BH, BH-BH
  - Physics regimes: Inspiral, merger, ringdown
  - Numerical relativity will be essential to interpret GW
  - waveforms



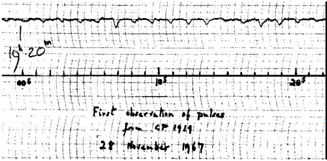
- Burst events
  - e.g. Supernovae with asymmetric collapse

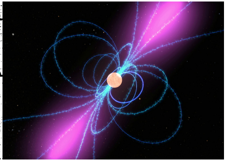


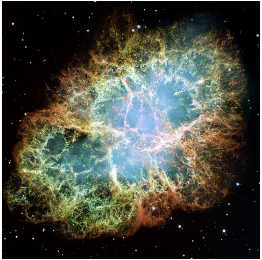


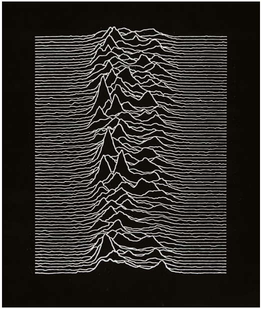
- *The Unexpected!*

Neutron stars = very dense, magnetized stars

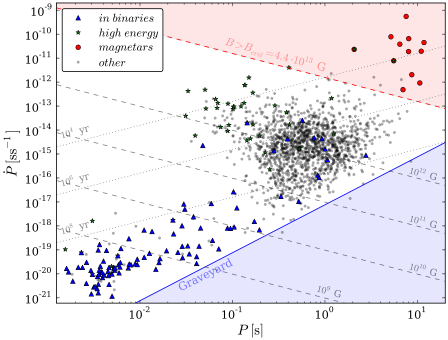




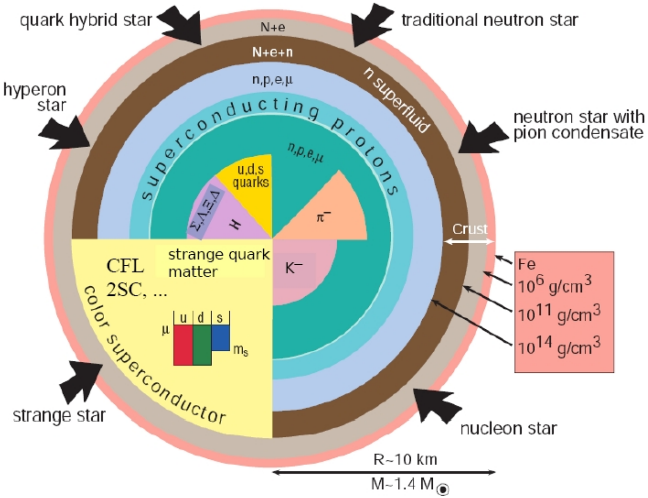




- ⋆ The most relativistic, material objects in the Universe: compactness $M/R \simeq 0.2 - 0.4$.



The mystery of neutron star interiors



(Courtesy: F. Weber)

Dense matter in conditions impossible to obtain on Earth!

Continuous GWs from spinning neutron stars

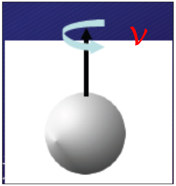
Characteristics:

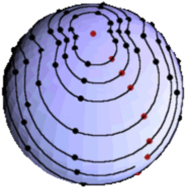
1. Long-lived: $T > T_{obs}$

2. Nearly periodic: $f_{GW} \sim \nu$

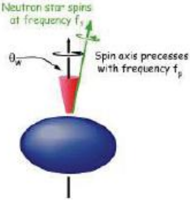
Generation mechanisms (we need a time varying quadrupole moment):

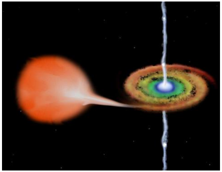
1. *Mountains*
   (elastic stresses, magnetic fields)
2. *Oscillations*
   (r-modes)
3. *Free precession*
   (magnetic field)
4. *Accretion*
   (drives deformations from r-modes, thermal gradients, magnetic fields)





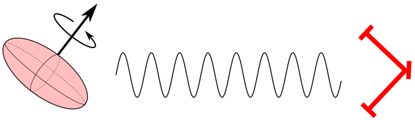
Courtesy: B. J.Owen

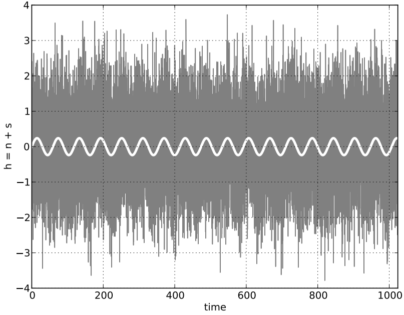


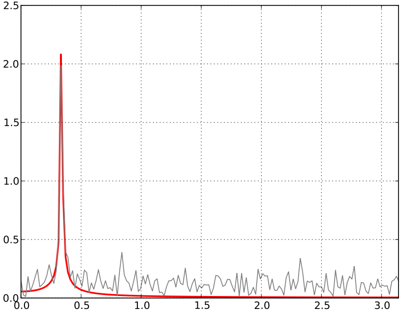


Courtesy: McGill U.

Example: weak monochromatic signals hidden in the noise







In this case Fourier transform is sufficient to detect the signal (a matched filter method):

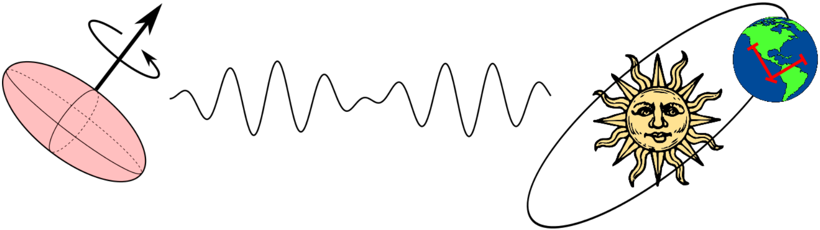
$$F = \left| \int_0^{T_0} x(t) \exp(-i\omega t) dt \right|^2$$

Signal-to-noise

$$SNR = h_0 \sqrt{\frac{T_0}{S_0}}$$

In reality: signal is modulated

Since the detector is on Earth, influence of planets and Earth's rotation changes the signal's amplitude and phase.



- ⋆ Signal is almost monochromatic: pulsars are slowing down,
- ⋆ To analyze, we have to demodulate the signal (detector is moving),
- → precise ephemerids of the Solar System used.

F-statistic

To estimate how well the amplitude and phase modulated model matches with the data $x(t)$, we calculate $\mathcal{F}$ (Jaranowski, Krolak, Schutz 1998),

$$\mathcal{F} = \frac{2}{S_0 T_0} \left(\frac{|F_a|^2}{\langle a^2 \rangle} + \frac{|F_b|^2}{\langle b^2 \rangle} \right)$$

where S_0 is the spectral density, T_0 is the observation time, and

$$F_a = \int_0^{T_0} x(t) a(t) \exp(-i\phi(t)) dt, F_b = \ldots$$

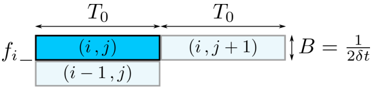
$a(t)$, $b(t)$ - amplitude modulation functions that depend on sky position of the source $(\alpha, \ \delta)$,

$\phi(t)$ - phase modulation function that depends on $(f, \dot{f}, \alpha, \delta)$

Methods of data analysis

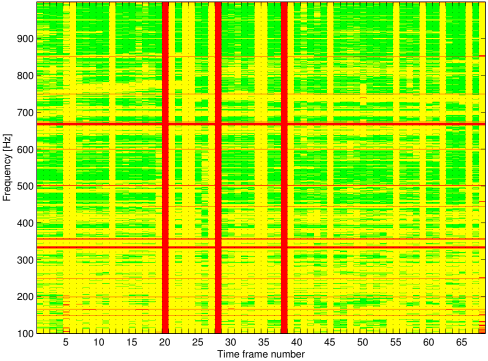
Computing power $\propto T_0^5 \log(T_0)$. Coherent search of $T_0 \simeq 1\ yr$ of data would require zettaFLOPS (10^{21} FLOPS) $\rightarrow$ currently impossible $\ddot{\frown}$

Solution: divide data into shorter length time frames ($T_0 \simeq 2$ days)



- $\star$ narrow frequency bands - sampling time $\delta t = 1/2B$, number of data points $N = T/\delta t \rightarrow N = 2TB$

$\rightarrow$ feasible on a petaFLOP computer.



Example search space (Virgo Science Run 1). Red: no data, yellow: bad data, green: good data.

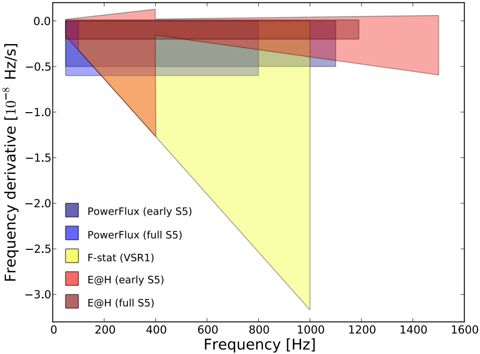
Typical all-sky search: parameter space

- ⋆ Narrow (1 Hz) frequency bands f: [100 – 1000] Hz,
- ⋆ Spin-down f_1 range proportional to f:

$$[-1.6 \times 10^{-9} \frac{f}{100\text{Hz}},\ 0]\ \text{Hz s}^{-1}$$

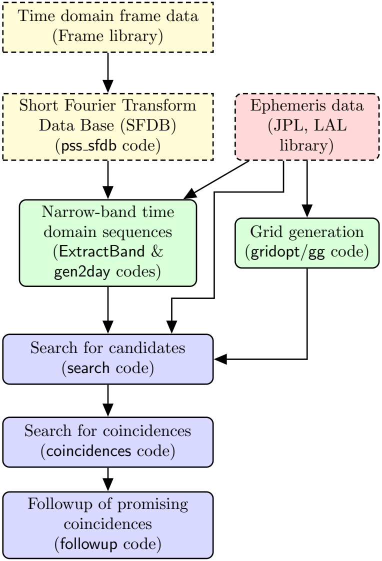
- ⋆ All-sky search: number of sky positions $\alpha(f), \delta(f) \propto f$.

Comparison of the $f - \dot{f}$ plane searched (yellow) with that of other recent all-sky searches:



In all-sky astrophysical applications, the 4-dim parameter space $(f, \dot{f}, \alpha, \delta)$ is big (in VSR1: $\simeq 10^{17}$ F-statistic evaluations)

All-sky pipeline https://github.com/mbejger/polgraw-allsky.git

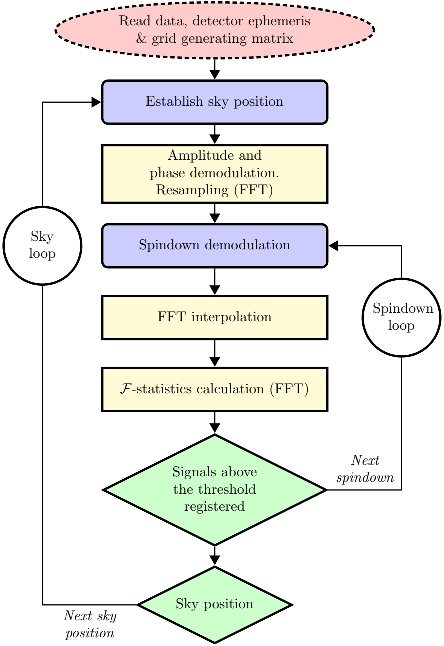


- ⋆ Input data generation (Raw time domain data $\sim$ *PB*)
- ⋆ Pre-processing $\rightarrow \sim$ *TB* (input time series, detector ephemerids and grid of parameters),
- ⋆ Stage 1: F-statistic search for candidate GW signals (the most time-consuming part of the pipeline)

$\rightarrow$ 10^{10} candidates/detector, 100 TB of output.

- ⋆ Stage 2: Coincidences among candidate signals from different time segments,
- ⋆ Stage 3: Followup of interesting coincidences - evaluation of F-statistic along the whole data span.

Most expensive part: search for candidate signals

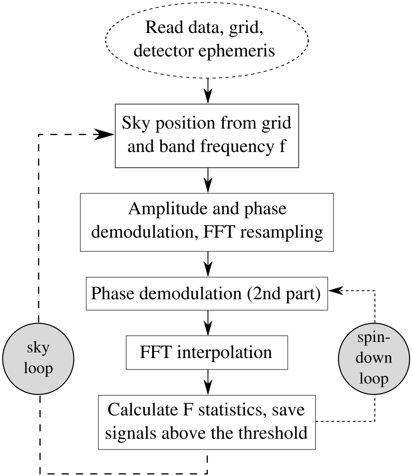


- ⋆ Suitable algorithms that allow for Fast Fourier Transforms,
- ⋆ Optimized grid of parameters - minimum number of operation to reach desired sensitivity,

  → partial demodulation before the inner spindown loop (only once per sky position),
- ⋆ Sky positions completely indepedent of each other

  → "Embarasingly parallel problem"

F-stat all-sky search description: `NVidia CUDA` implementation



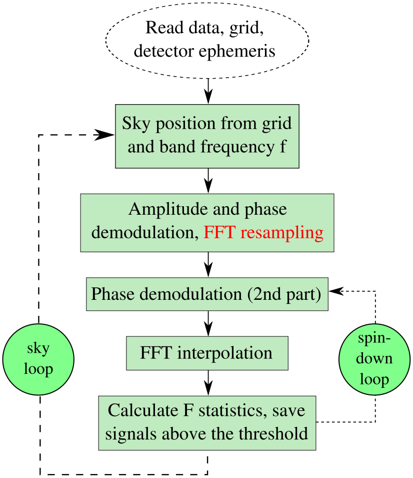
Main parameters in coherent search for continuous wave signals:

- ⋆ bandwidth 0.25 Hz
- ⋆ sampling time 2 s
- ⋆ data length N = 86164 (two sideral days)

⋆ 4D grid: α, δ, f, $\dot{f}$ - sky positions, frequency and spindown

⋆ Uses the F-statistic defined in Jaranowski, Królak & Schutz (1998), algorithm described and tested in Astone et al. (2010)

⋆ No. of F-statistic evaluations $\propto f^3$ (no. of sky positions $\propto f^2$, spindown $\propto f$)



Basically the whole loop over sky (α, δ) can be computed in parallel since the sky positions are independent of each other

The majority of computing is spent on

- ⋆ calculating the phase (trigonometric functions, $\gtrapprox$ 20%)
- ⋆ FFT ($\gtrapprox$ 70%)

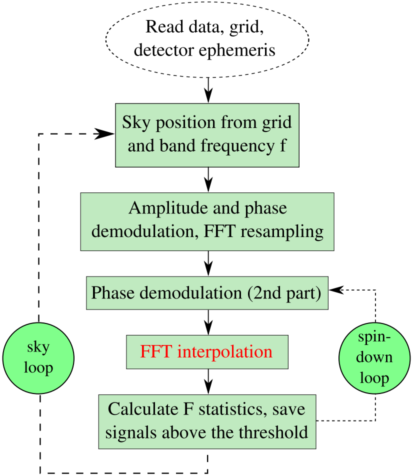
Efficient FFT requires 2^N data points ($N_{data} = 86164 < 2^{17}$) $\rightarrow$ padding with zeros to $N = 2^{17}$

FFT: time resampling

- ⋆ Resampling to barycentric time - FFT and inverse:
  - ⋆ nearest-neighbour ($\simeq$ 5% error),
  - ⋆ splines ($\simeq$ 0.1% error), custom implementation

The only part that currently has to be done in double-precision.

F-stat: parallelization strategy



- ⋆ How to do FFT with GPU:
  - ⋆ write custom kernel for FFT, launch concurrently.
  - ⋆ use CUDA cuFFT library:
    - ☺ well-optimized (Cooley-Tukey, Bluestein), 1D/2D/3D double precision complex/real transforms, multiple transforms, in- and out-of-place transforms,
    - ☹ cannot launch many instances at the same time (at least not with every card/CUDA version).
- ⋆ cuSPARSE (sparse matrix routines)

Results of implementation on GPUs

- ⋆ Input data loaded to device once. For each detector (V1, L1 & H1),
  - ⋆ time-series ($N \times sizeof(double) = 674$ KB)
  - ⋆ ephemerids ($3N \times sizeof(double) = 2$ MB)

  + a grid-generating matrix (388 B).
- ⋆ Sequence of kernels (mostly FFT) launched in a loop from CPU (kernel concurency needs work!),
- ⋆ Time resampling using double precision, everything else (main spindown loop) using single precision (needs work!),
- ⋆ Non-optimized usage of constant memory (needs work!),
- ⋆ Asynchronous output transfer to host.

Current GPU results: $\sim \times 10$ speedup **with respect to the optimized CPU version** that uses `AVX/AVX2` events and `YEPPP!` library

Estimated time τ to match one template:

- ⋆ CPU (Intel(R) Xeon(R) CPU E5-2680 v2 @ 2.80GHz) $\simeq 5 \times 10^{-3}$ s
- ⋆ GPU (GeForce GTX Titan) $\simeq 3 \times 10^{-4}$ s

Performance scaling - favorably for high frequencies (no. of spindowns $\propto f$).

Summary/references

We have two search codes for candidate signals for a network of detectors:

- ⋆ a well-optimized CPU code,
- ⋆ a working GPU version that is now optimized.

- P. Astone, K. M. Borkowski, P. Jaranowski, M. Piętka and A. Królak, PRD, **82**, 022005 (2010)
- `https://developer.nvidia.com/cuFFT`
- P. Jaranowski, A. Królak, and B. F. Schutz, PRD **58**, 063001 (1998).
- Polgraw-allsky github repository: `https://github.com/mbejger/polgraw-allsky.git`