# Processing Big (Radio Astronomy) Data with GPUs

Laura Spitler & Marina Berezina
Max-Planck-Institut für Radioastronomie, Bonn

8 October 2014

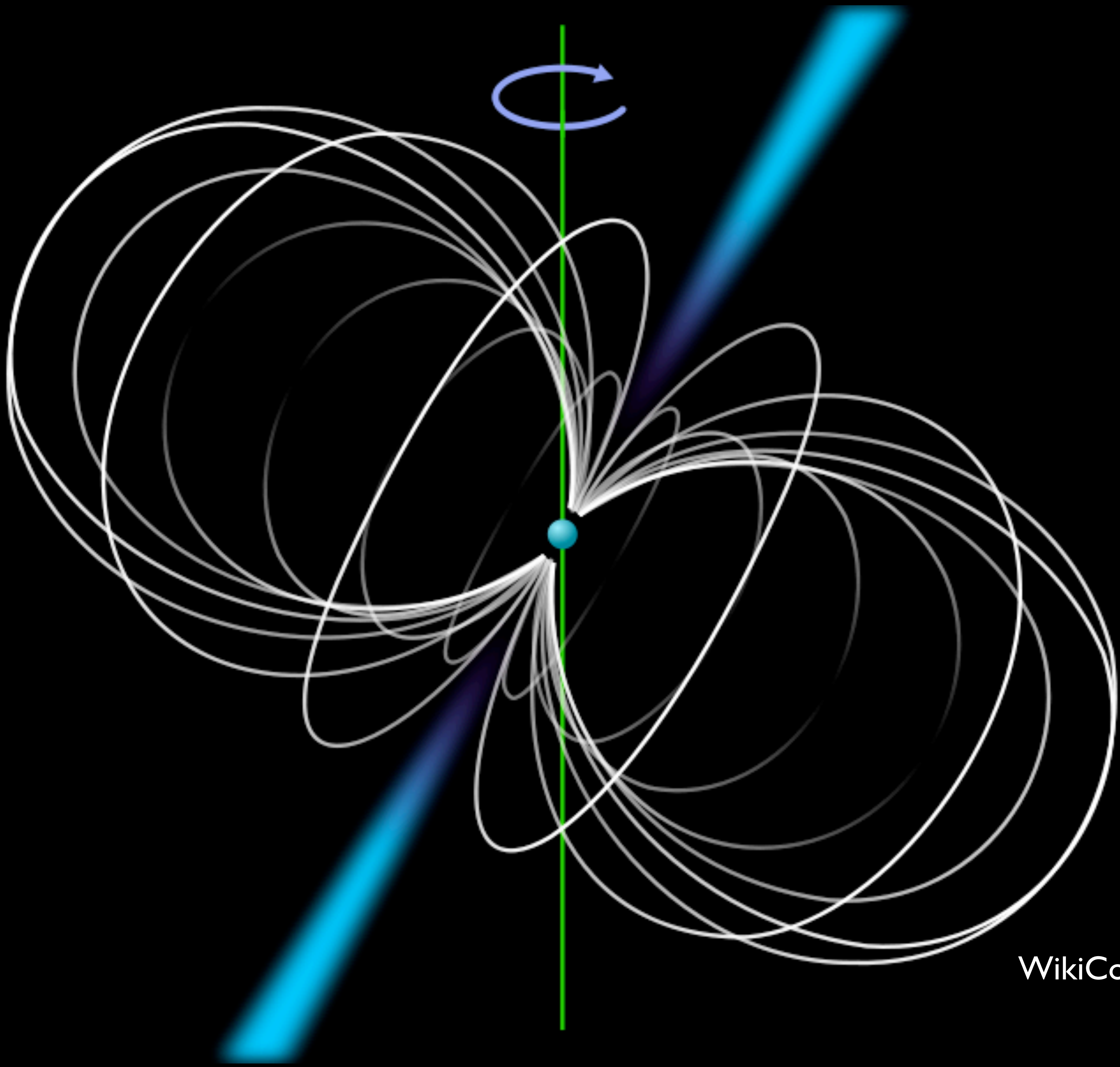
Using **radio pulsars** to study fundamental physics...

...and using **fast radio bursts** to study astrophysics.

# Pulsar

“pulsating radio source”

Spinning neutron star with a very strong magnetic field

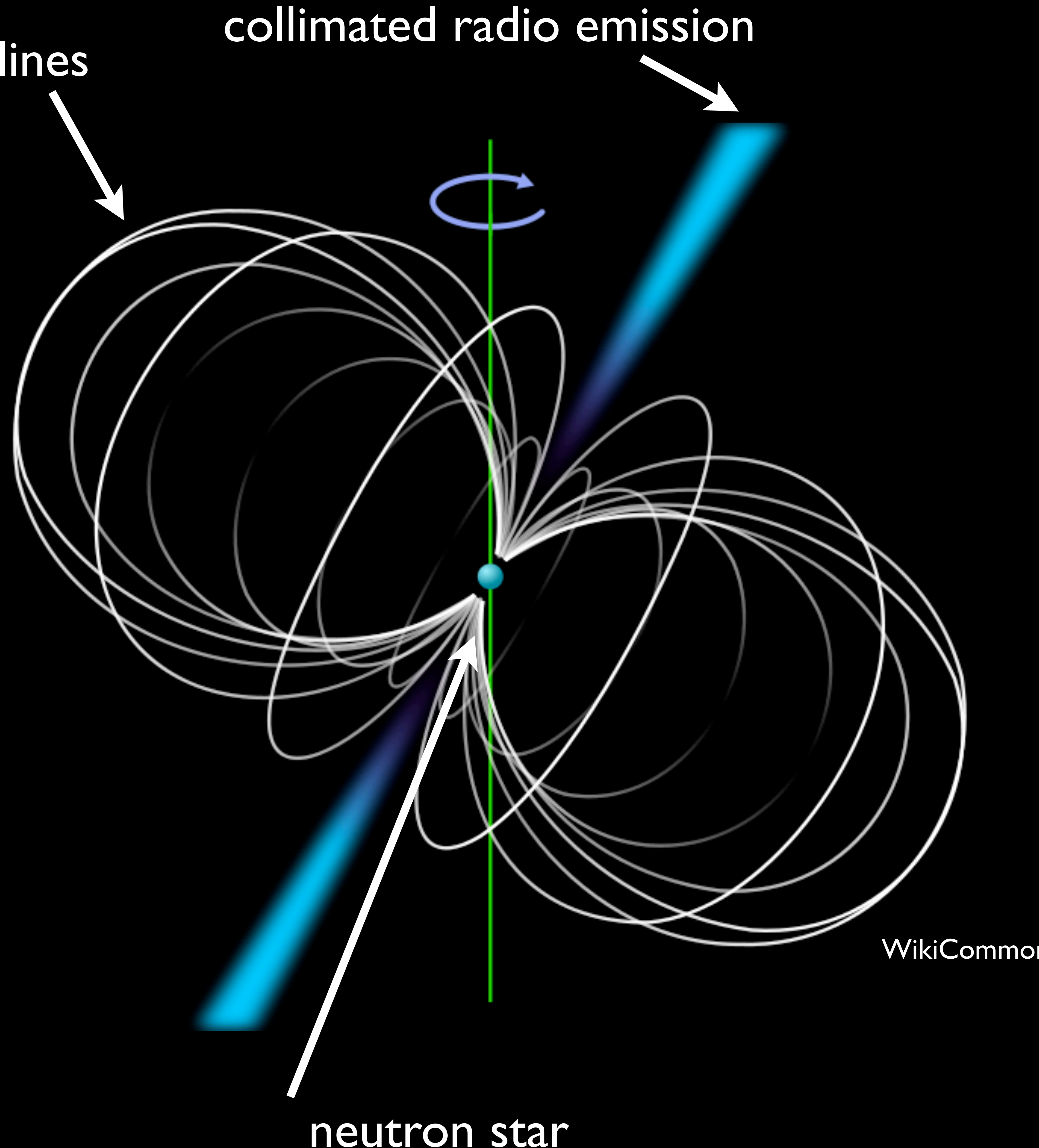


WikiCommon

# Pulsar

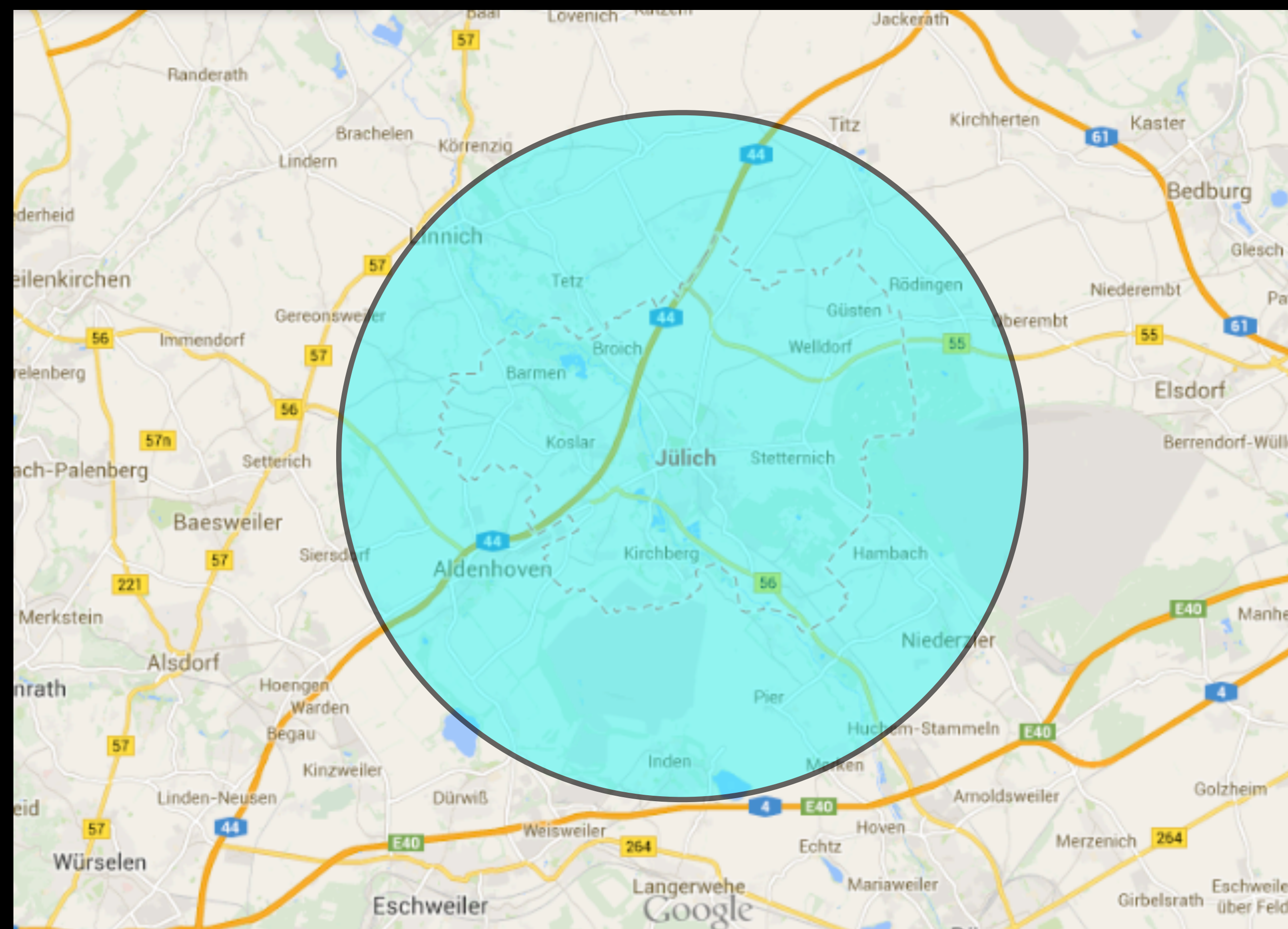
"pulsating radio source"

Spinning neutron star with a very strong magnetic field



# Neutron Stars

The same mass as the sun but with the diameter of 20 km



GoogleMaps

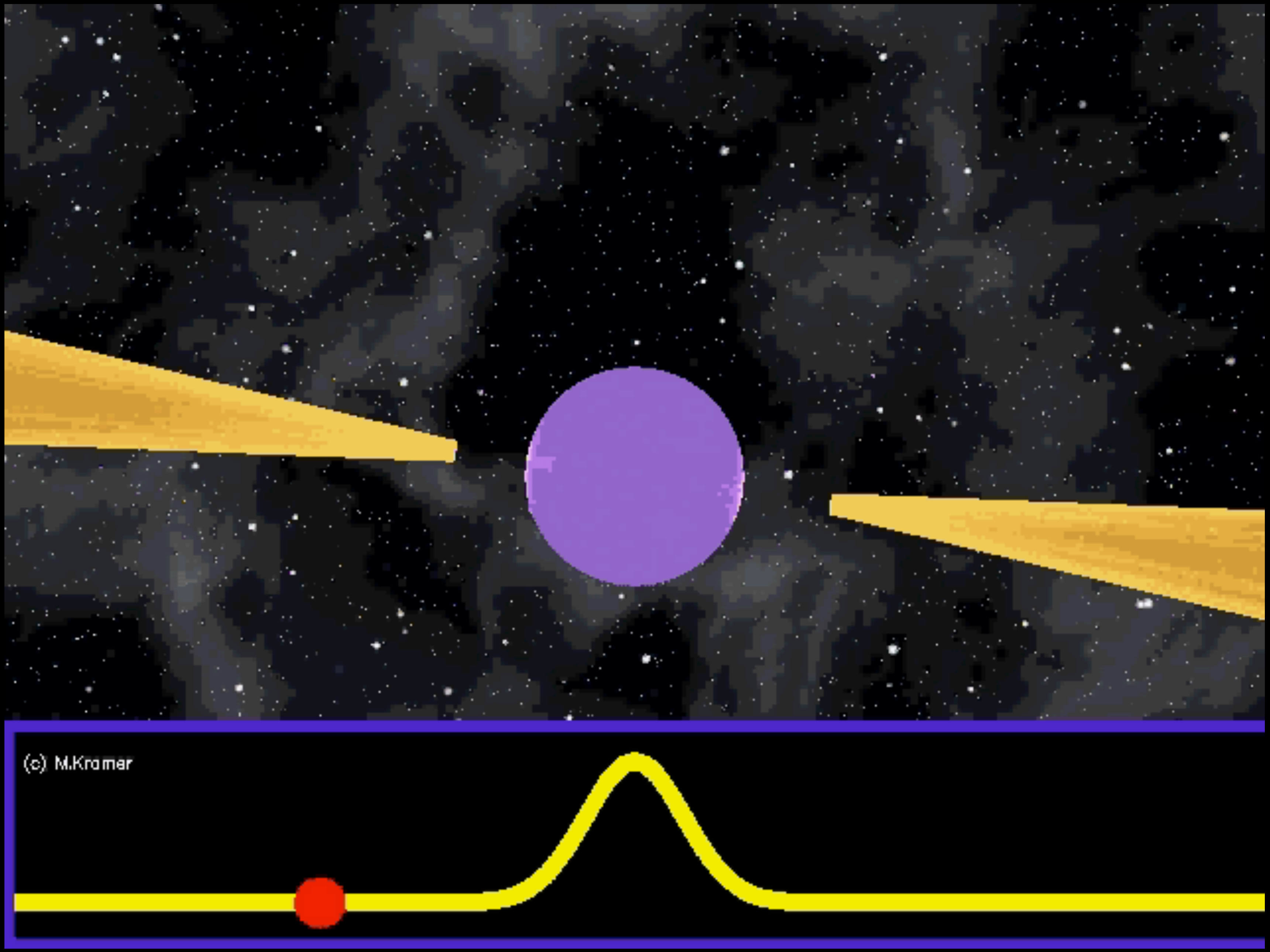
Magnetic fields $10^{10}$ and $10^{15}$ times as strong as the Earth's

The density of an atomic nucleus

The escape velocity from the surface is 70% the speed of light

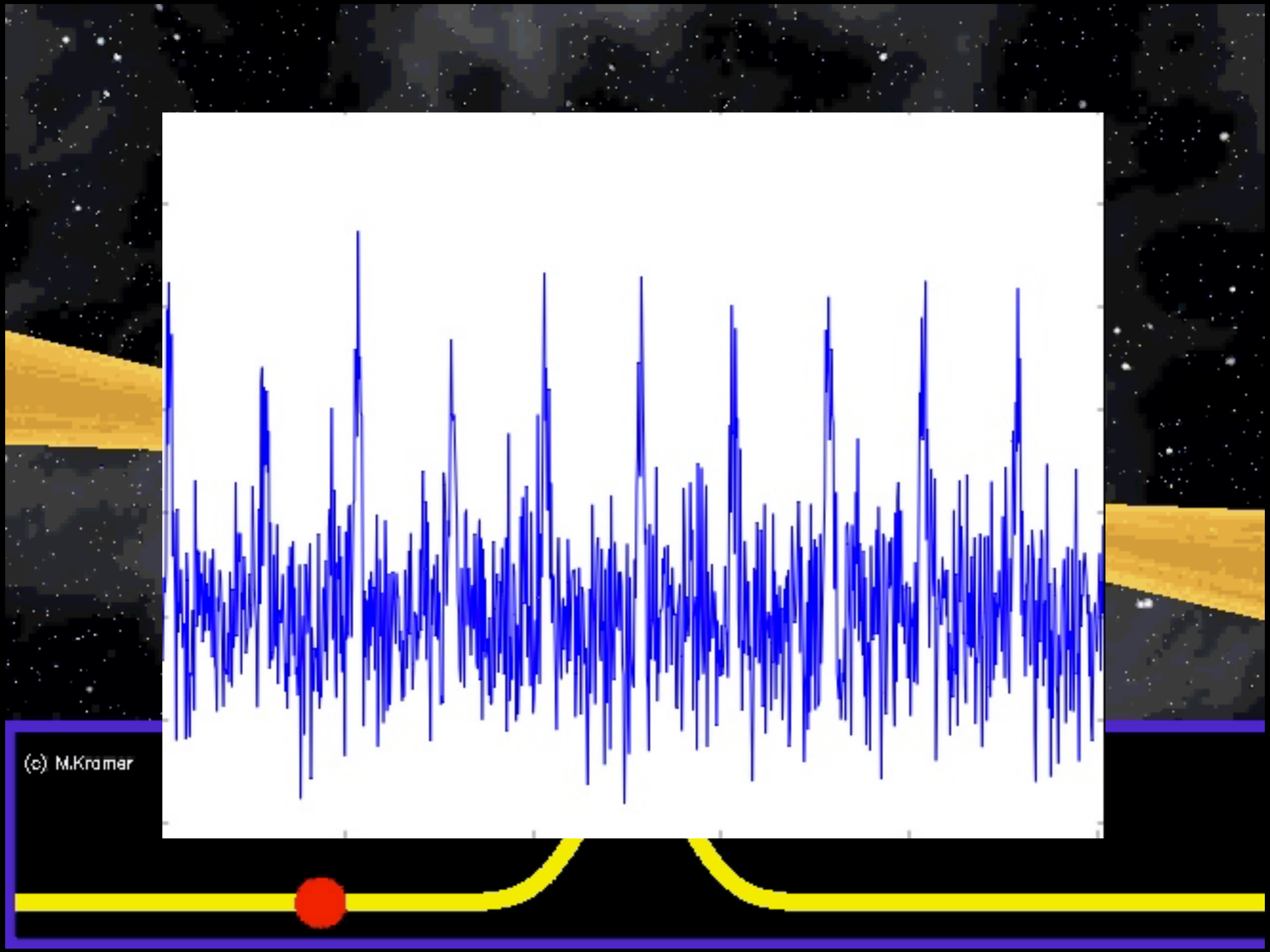
Fastest spinning pulsar rotates 700 times a second

# “Lighthouse” Model




Michael Kramer (MPIfR/JBCA)

# “Lighthouse” Model




Michael Kramer (MPIfR/JBCA)

# Precise Astronomical Clocks

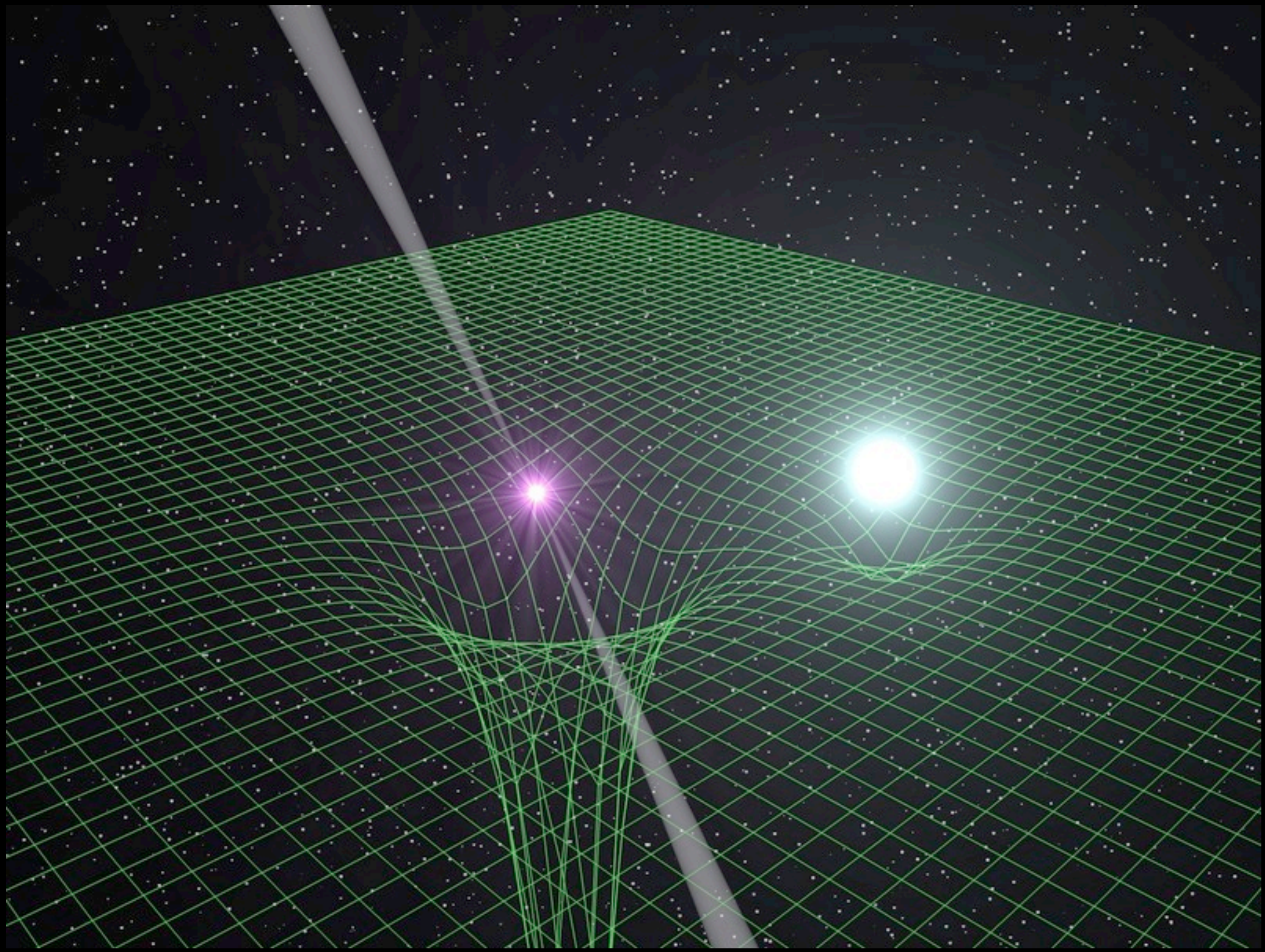
The spin frequency of J0348+0432 on 14 Mar 2012:

25.5606361937675(4) Hz

The spin frequency of J0348+0432 on 08 Oct 2014:

25.56063618102021 Hz

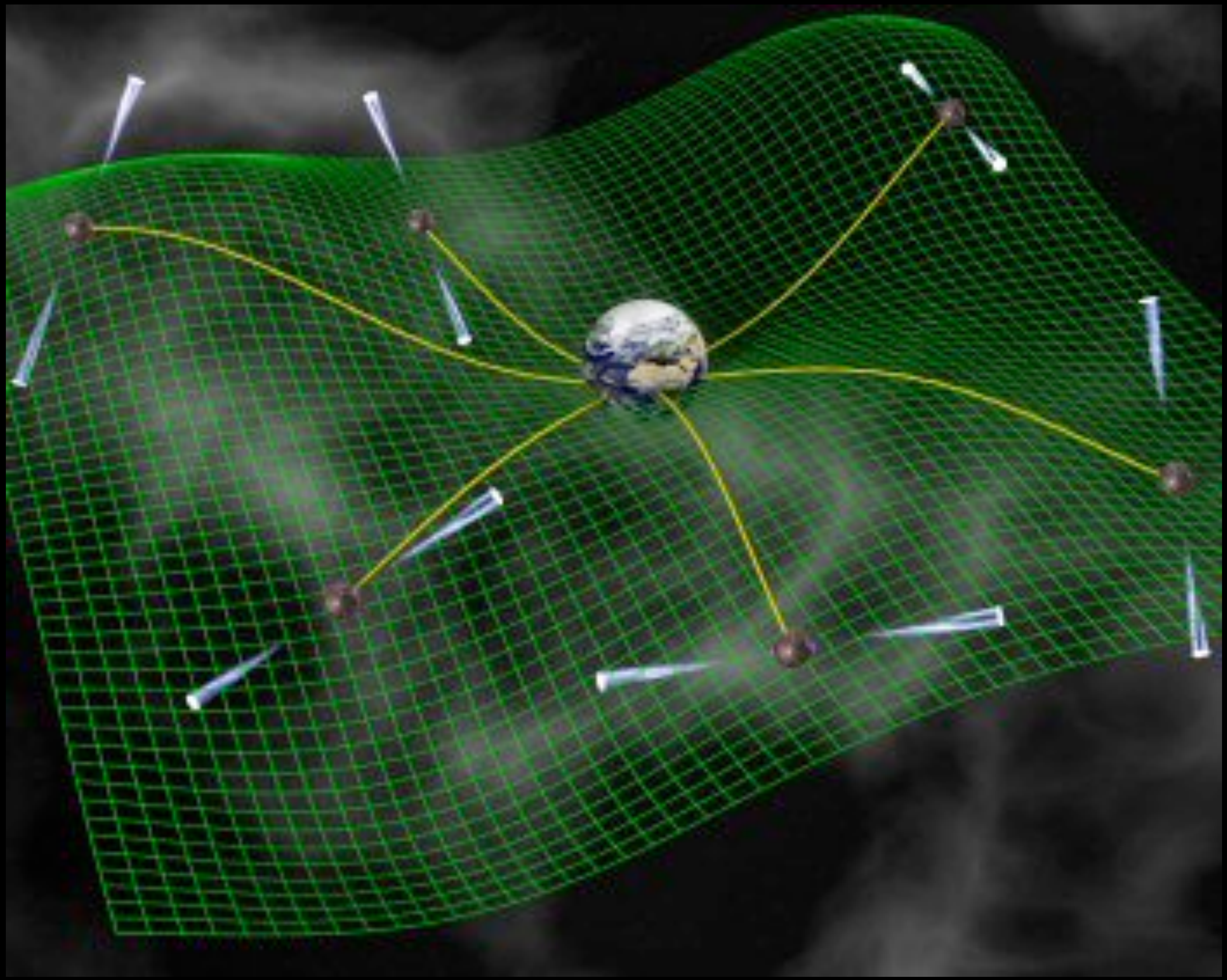
Pulsars are tools to test our understanding of gravity



Science / J. Antoniadis (MPIfR)

# Gravitational Tests with Pulsars

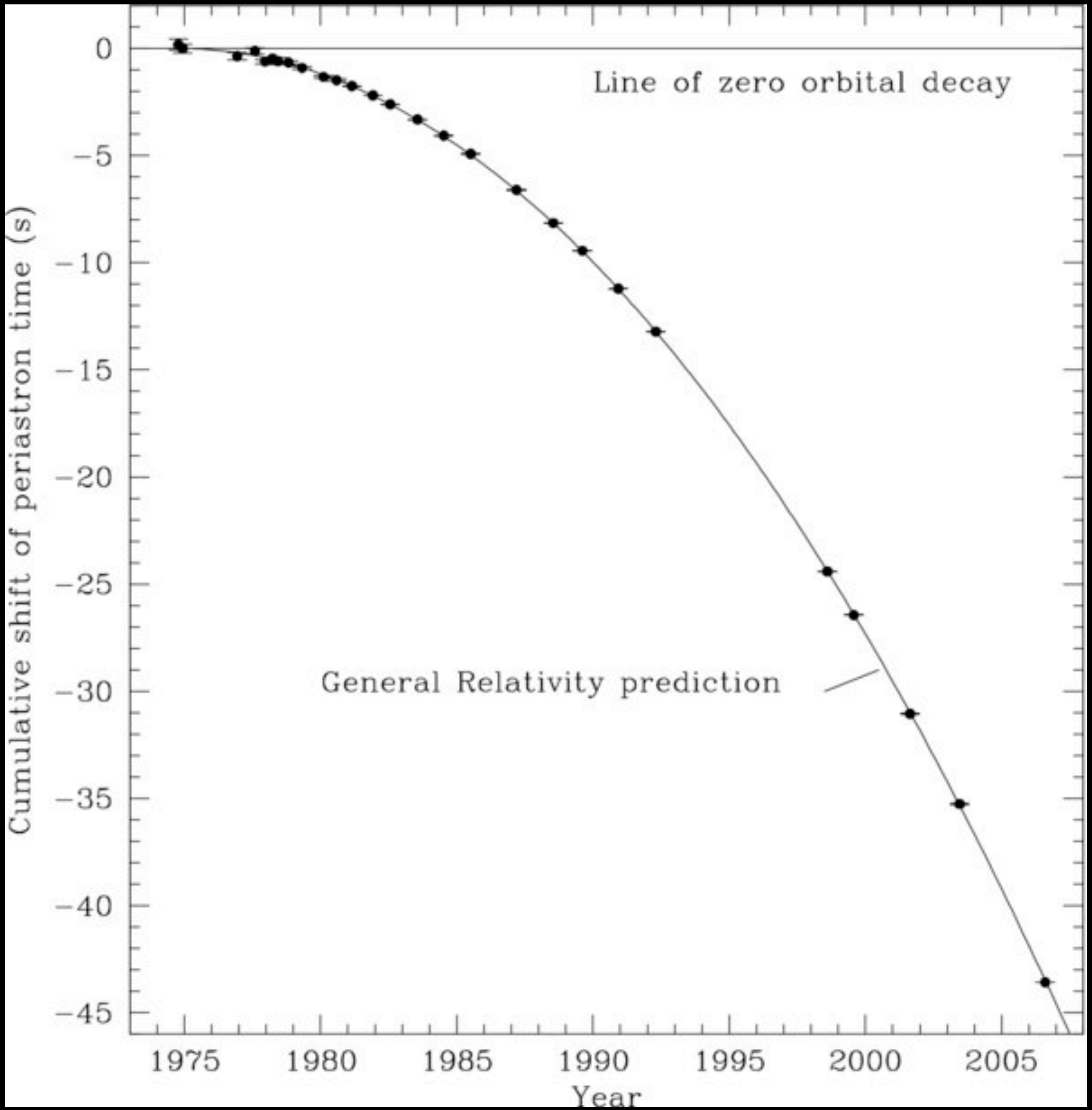
- First indirect evidence of the existence of gravitational waves
- Networks of pulsars can detect the low frequency gravitational wave background
- Pulsars in binary systems provide best tests of alternative theories of gravity



David Champion

# Gravitational Radiation

Binary pulsar B1913+16 provided first indirect evidence for gravitational waves.




Weisberg, Nice, Taylor, ApJ, 722, 2010

# Search for even more suitable pulsars

(holy grail: pulsar orbiting a black hole)

# Radio Telescopes

Effelsberg 100m

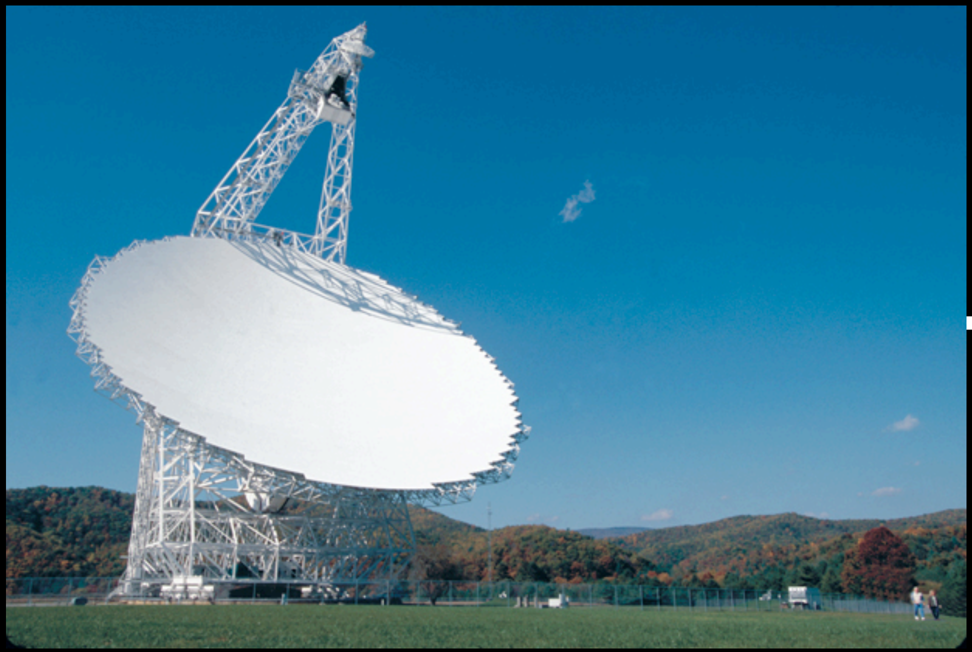
MPIfR

LOFAR

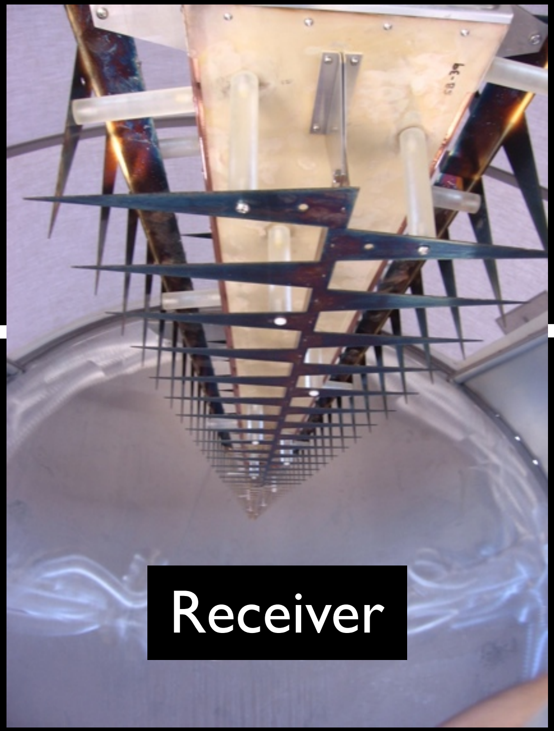
Hans Hordijk

Jansky Very Large Array

NRAO/AUI/NSF



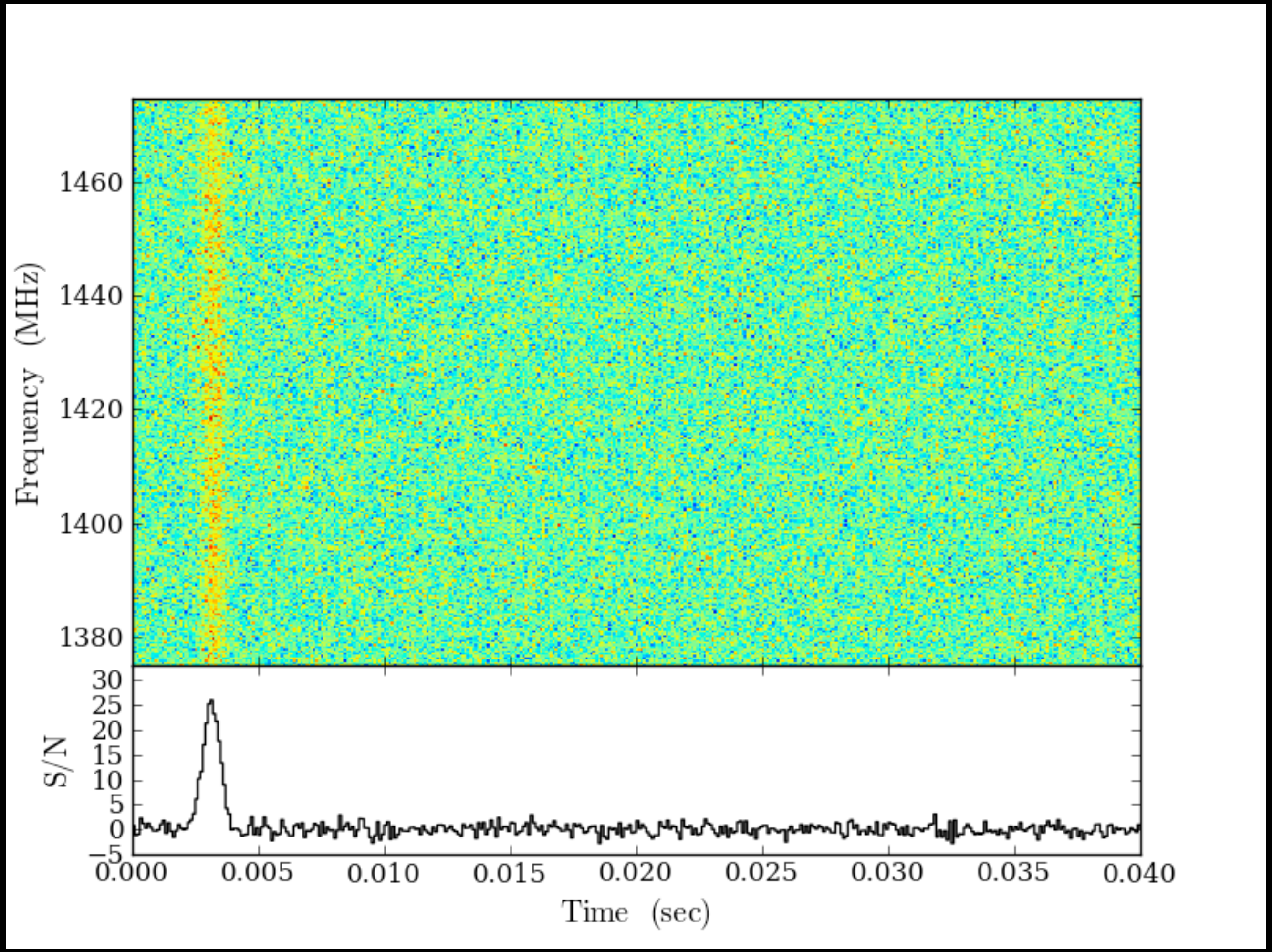
Green Bank Telescope, USA NRAO/AUI/NSF

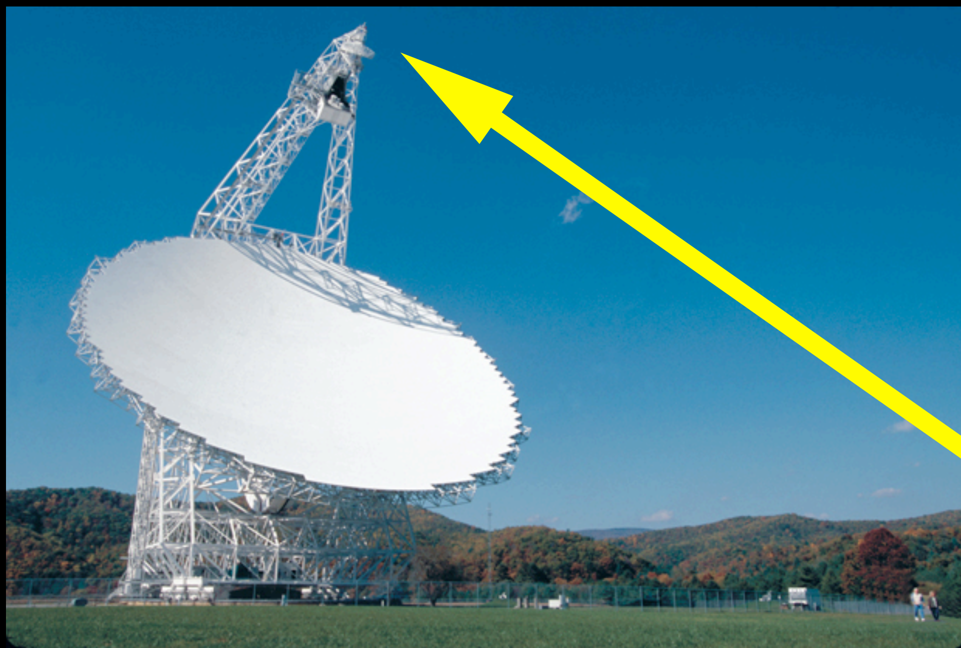



Spectrometer (FFT)

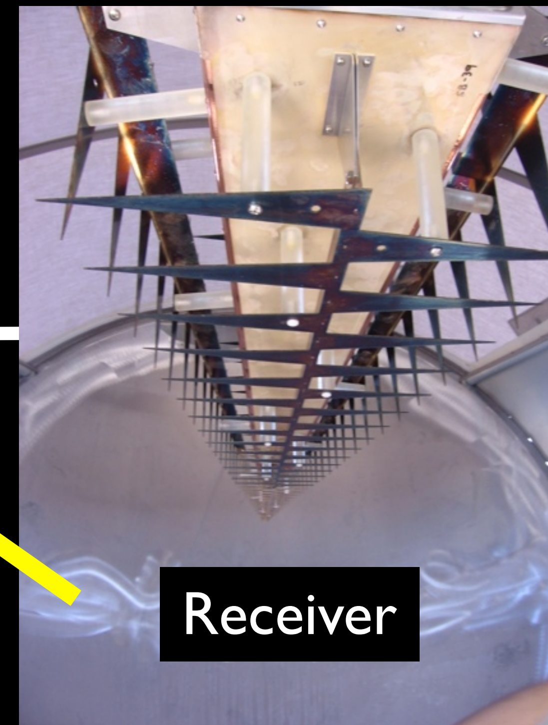
UC Berkeley, CASPER group, H. Chen

Pulsar survey data:
2D array of time,
frequency, and intensity





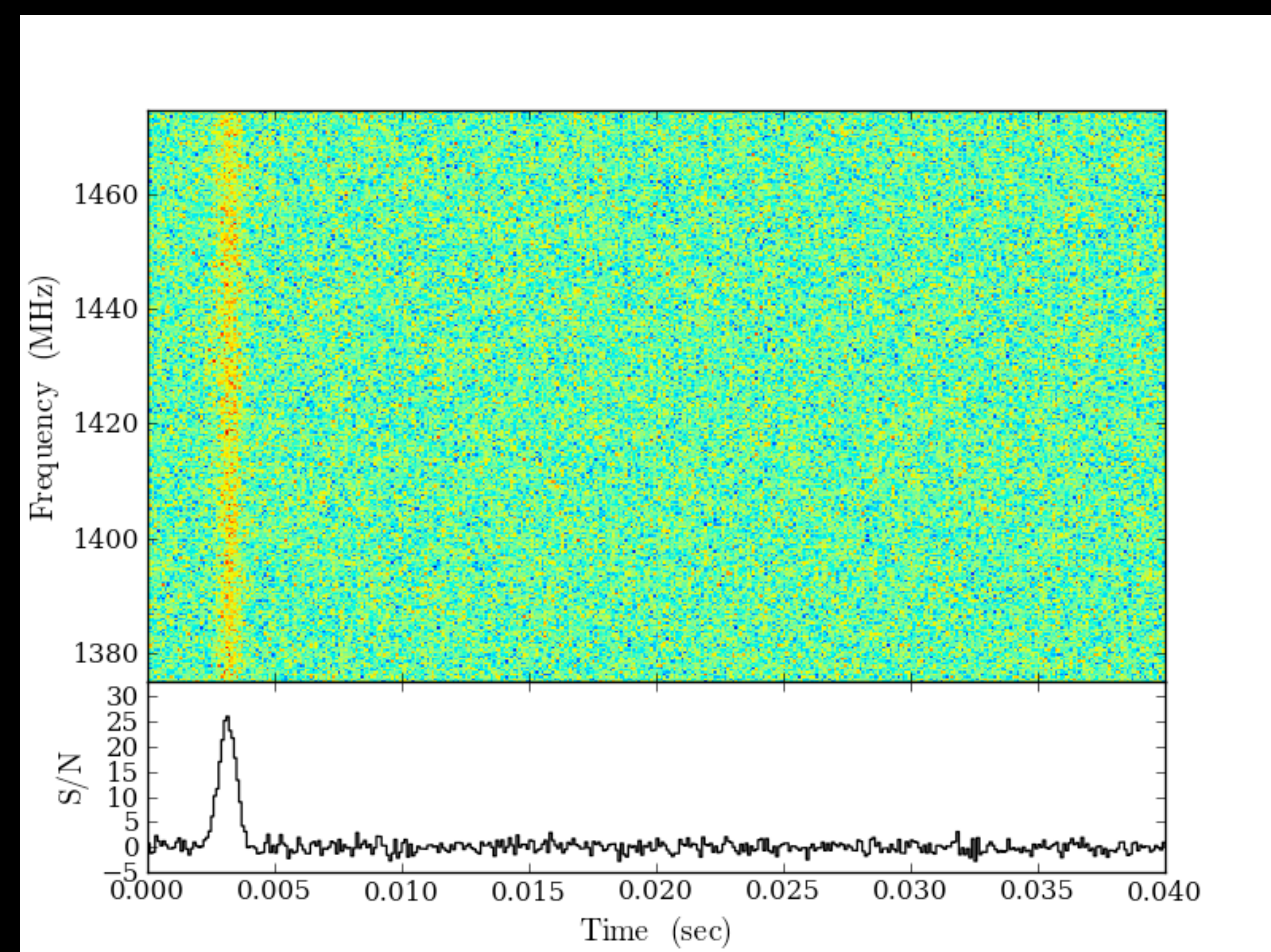
Green Bank Telescope, USA NRAO/AUI/NSF




## Spectrometer (FFT)

UC Berkeley, CASPER group, H. Chen

# Pulsar survey data: 2D array of time, frequency, and intensity

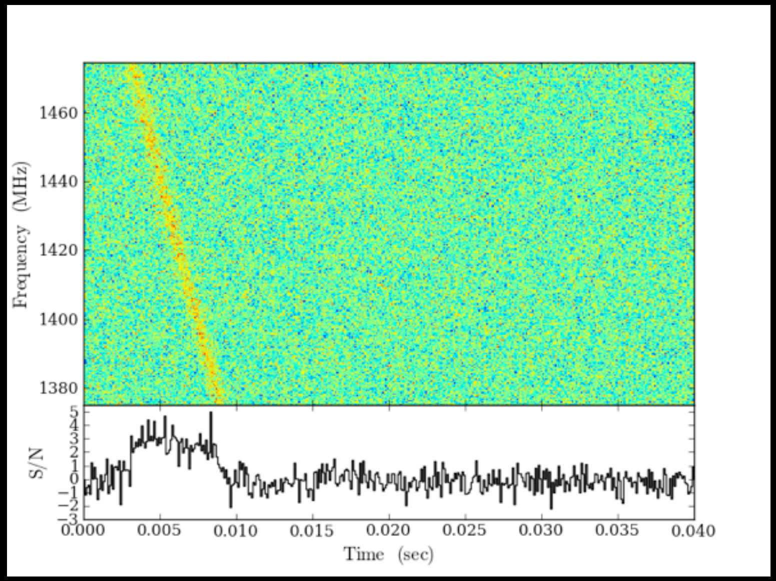


# `peasoup`

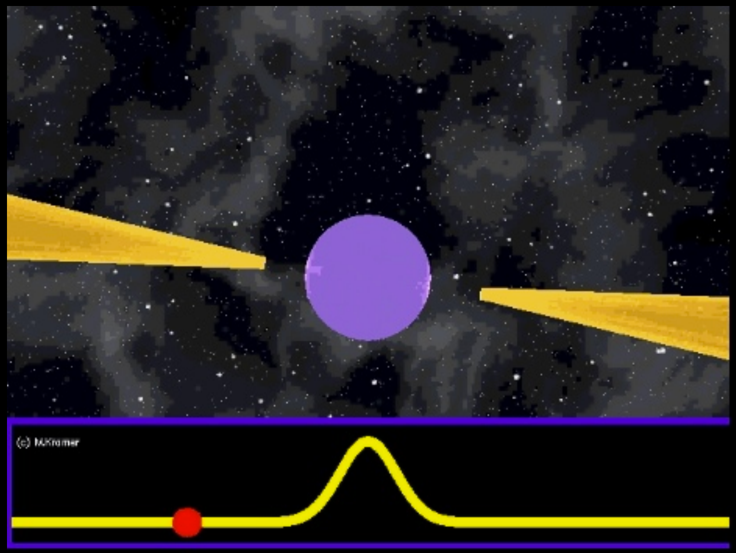
- Full GPU-based pulsar search pipeline
- Written by Ewan Barr (U of Swinburne)
- Uses `dedisp` dedispersion library (Ben Barsdell)
- Multi-GPU capability

# Algorithms search over three parameters:

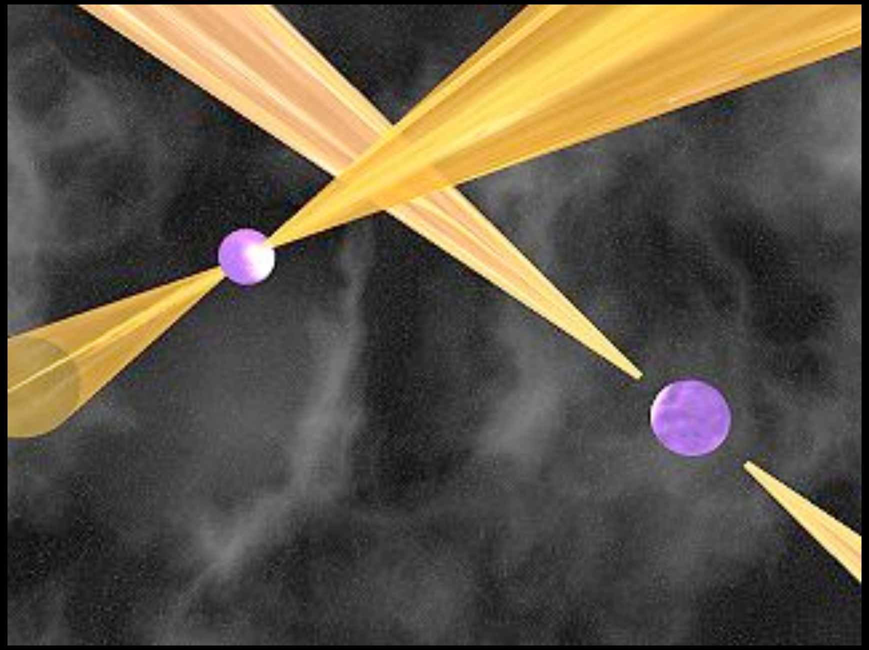
<u>Dispersion Measure</u>




<u>Rotation Period</u>



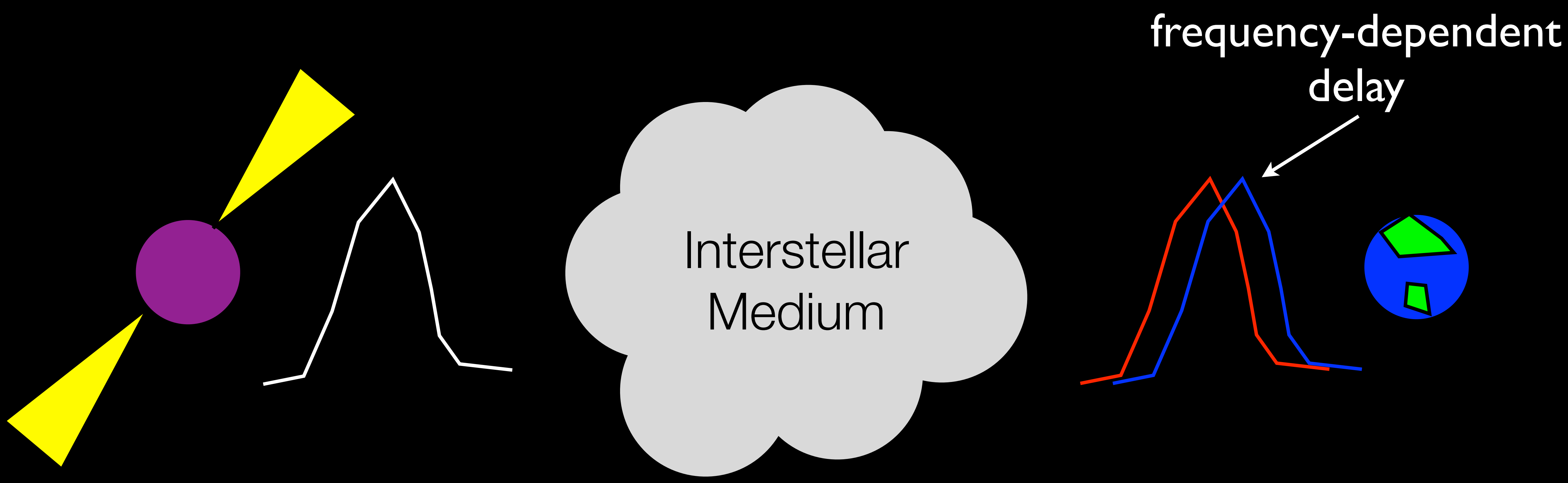
<u>Orbital Acceleration</u>



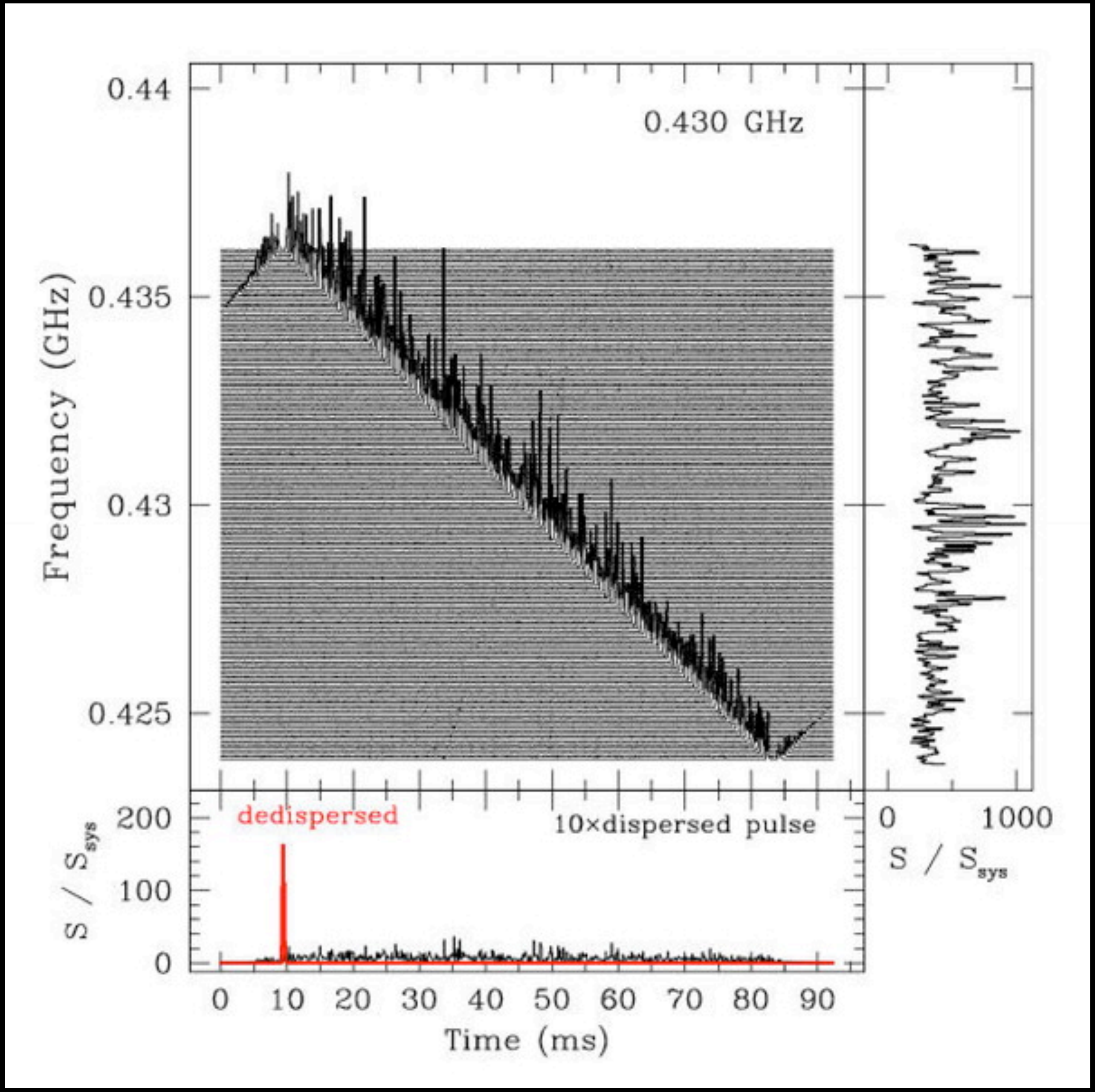
Michael Kramer (MPIfR/JBCA)

Further unknown: position on the sky

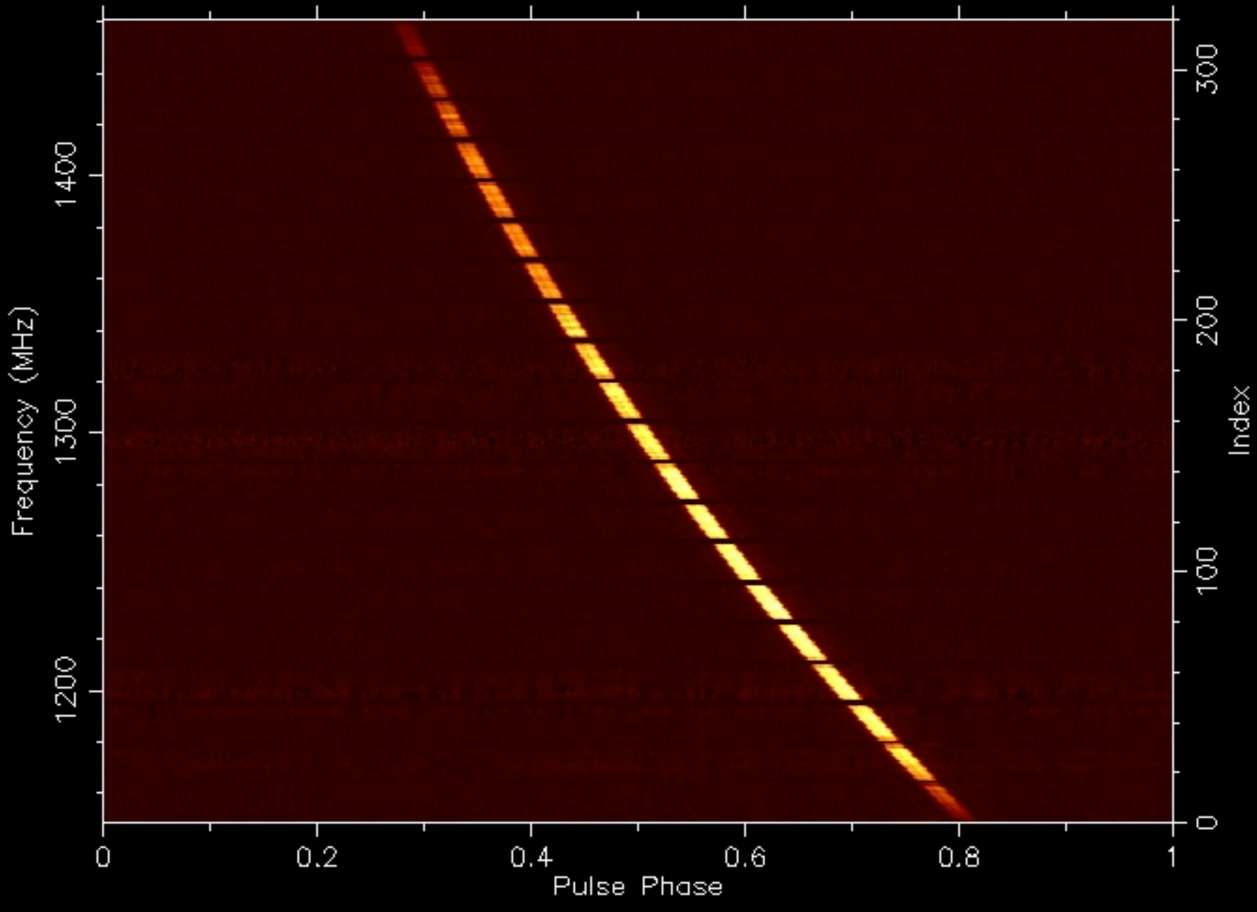
# Dispersion




The delay is proportional to the number of electrons along the line-of-sight.
The magnitude of the delay is quantified by the **dispersion measure** (DM).

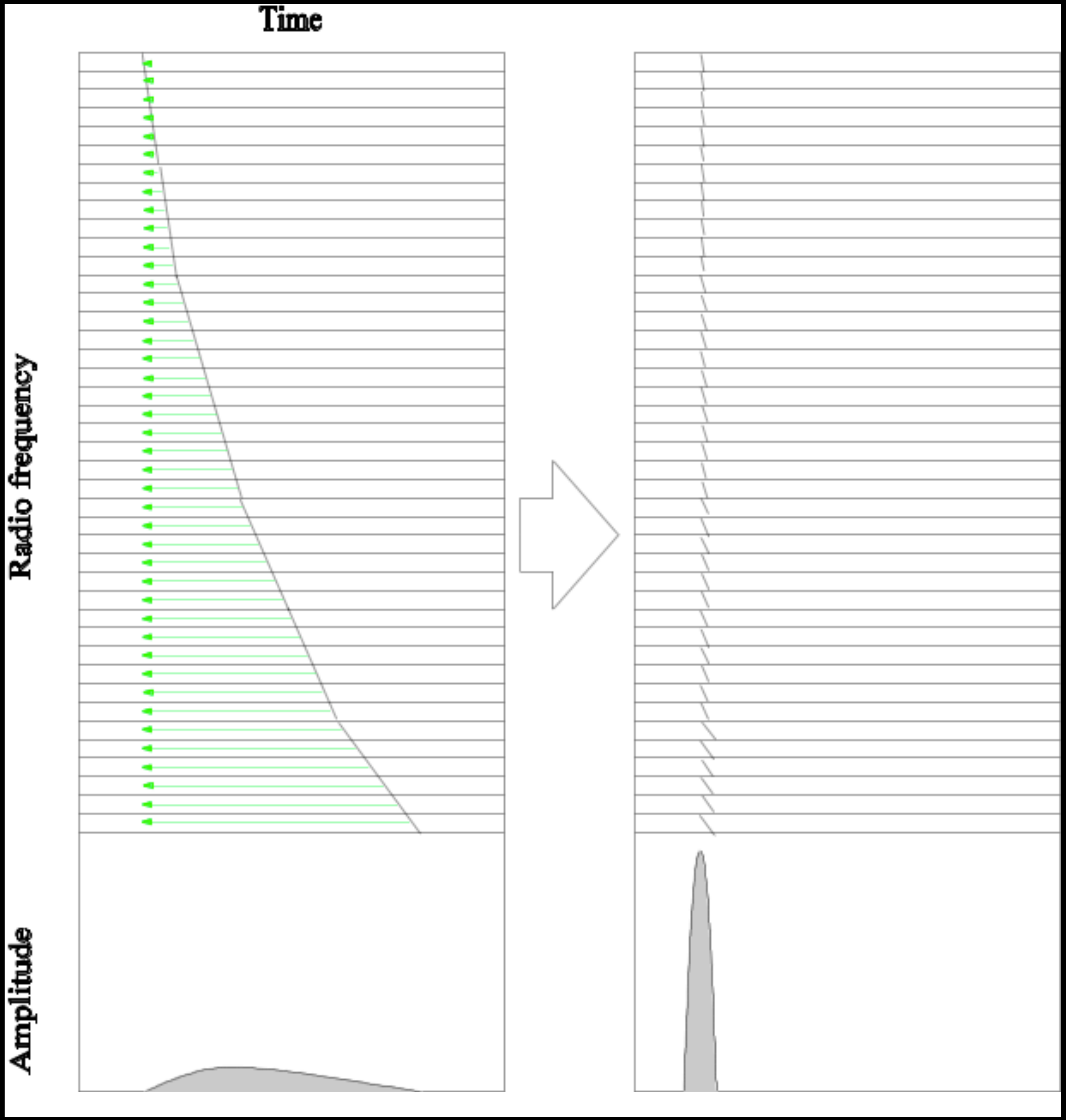



Cordes et al (2004)



# Dedispersion

Must try a large (~1000) number of trial DMs to find the correct value for the pulsar.



Pulsar survey data $A_{\nu,t}$

with $N_t$ time samples and $N_\nu$ frequency channels

Mathematically: $$D_{\mathrm{DM,t}} = \sum_{\nu}^{N_\nu} A_{\nu,t+\Delta t(\mathrm{DM},\nu)}$$

(for a given trial DM)

Operational Complexity: $\mathcal{O}(N_t N_\nu N_{\mathrm{DM}})$

`dedisp` library:

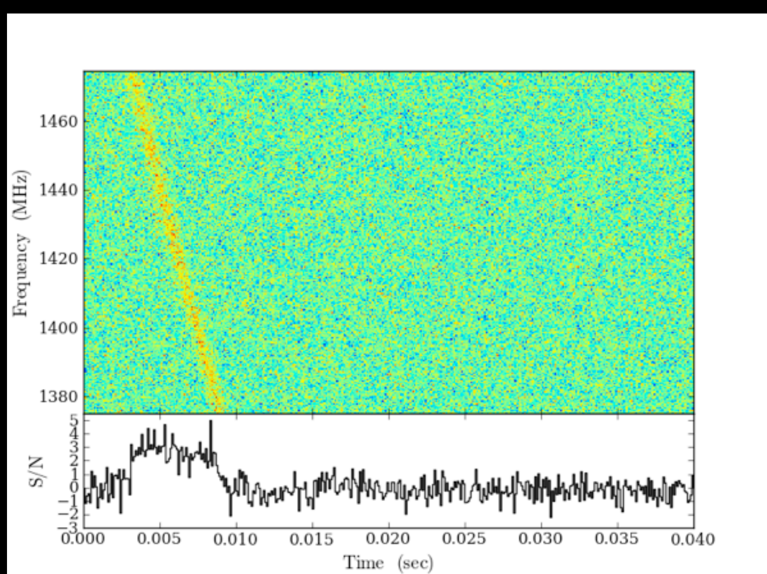
Parallelized over dimensions $N_t$ and $N_{\mathrm{DM}}$

Barsdell et al, MNRAS, 443, 2012

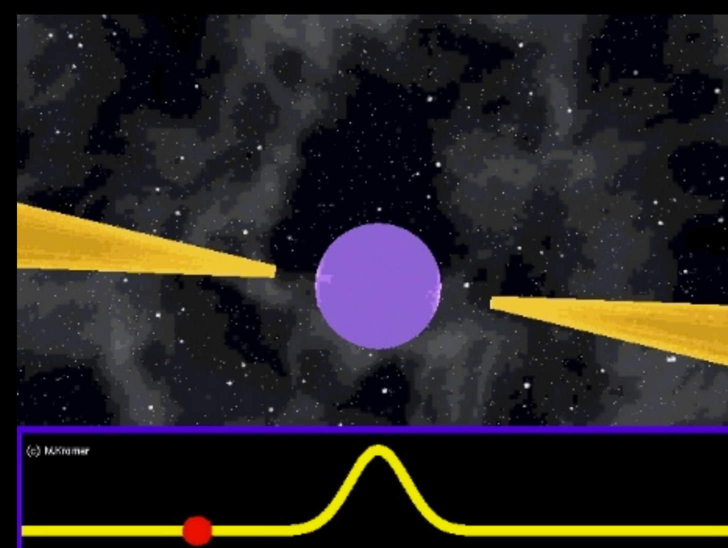
http://dedisp.googlecode.com/ (or sourceforge)

# Algorithms search over three parameters:

<u>Dispersion Measure</u>




<u>Rotation Period</u>



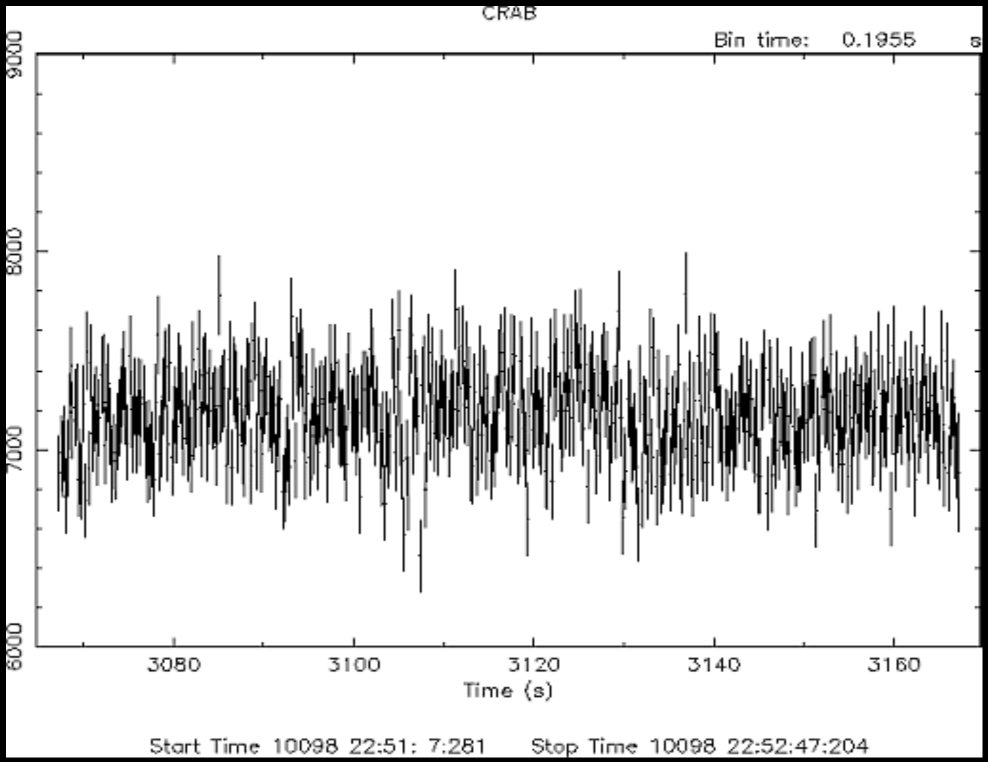
Further unknown: position on the sky

# Searching for periodic signals

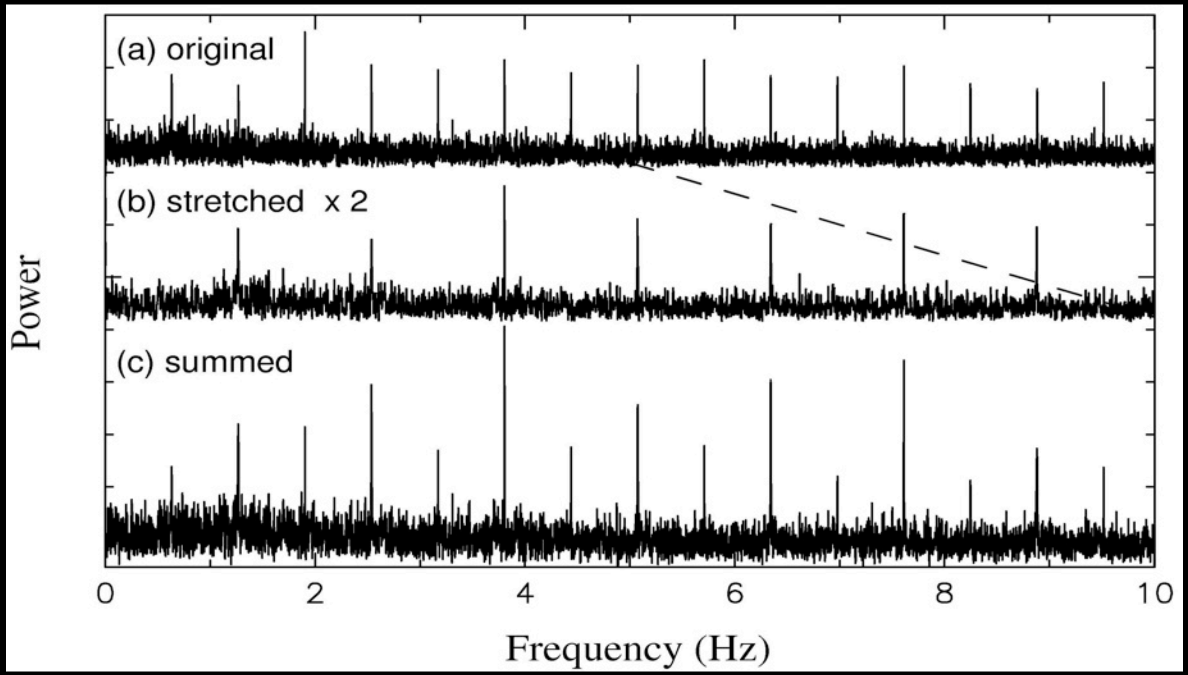
Use Fourier a transform to efficiently calculate many trial rotation periods

Pulsar signal spread across many harmonics of the pulsar spin frequency

Peak finding algorithm identifies statistically significant bins



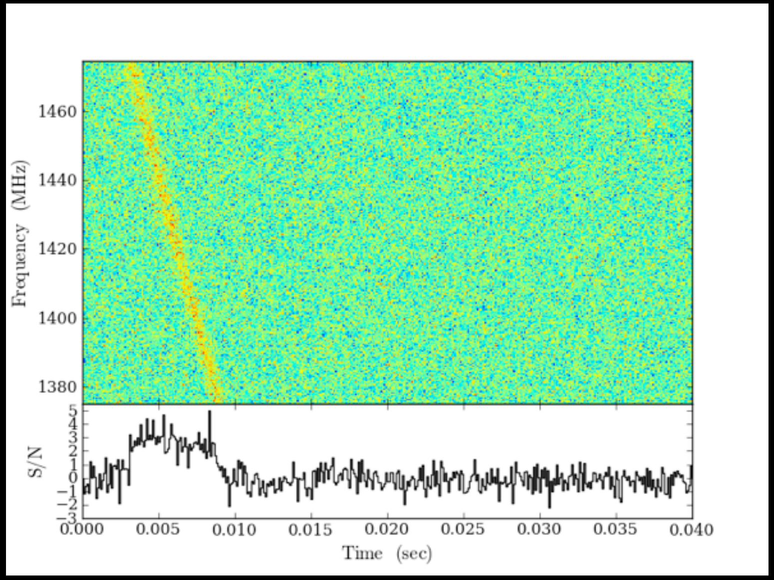

FFT the time series



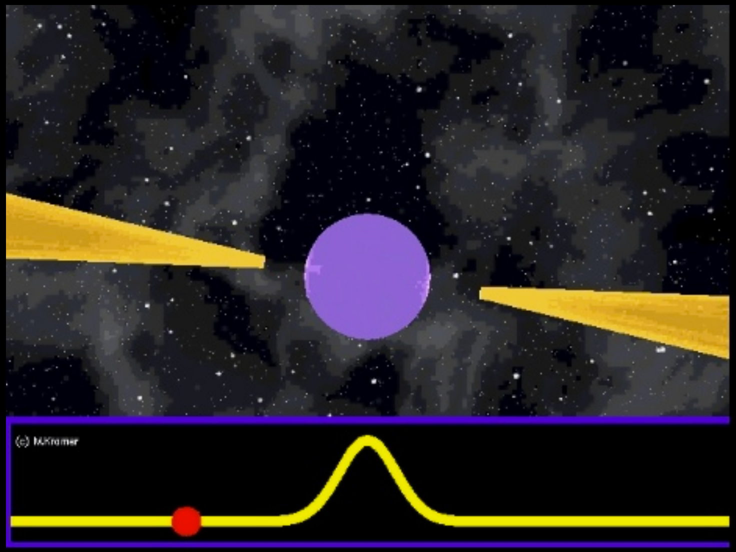

Harmonic summation

# Algorithms search over three parameters:

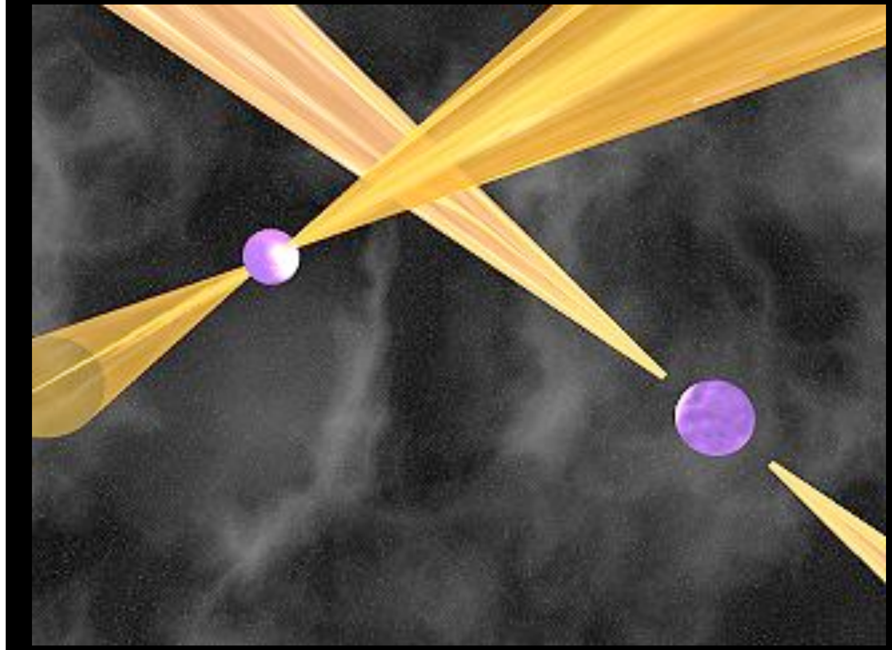
Dispersion Measure




Rotation Period



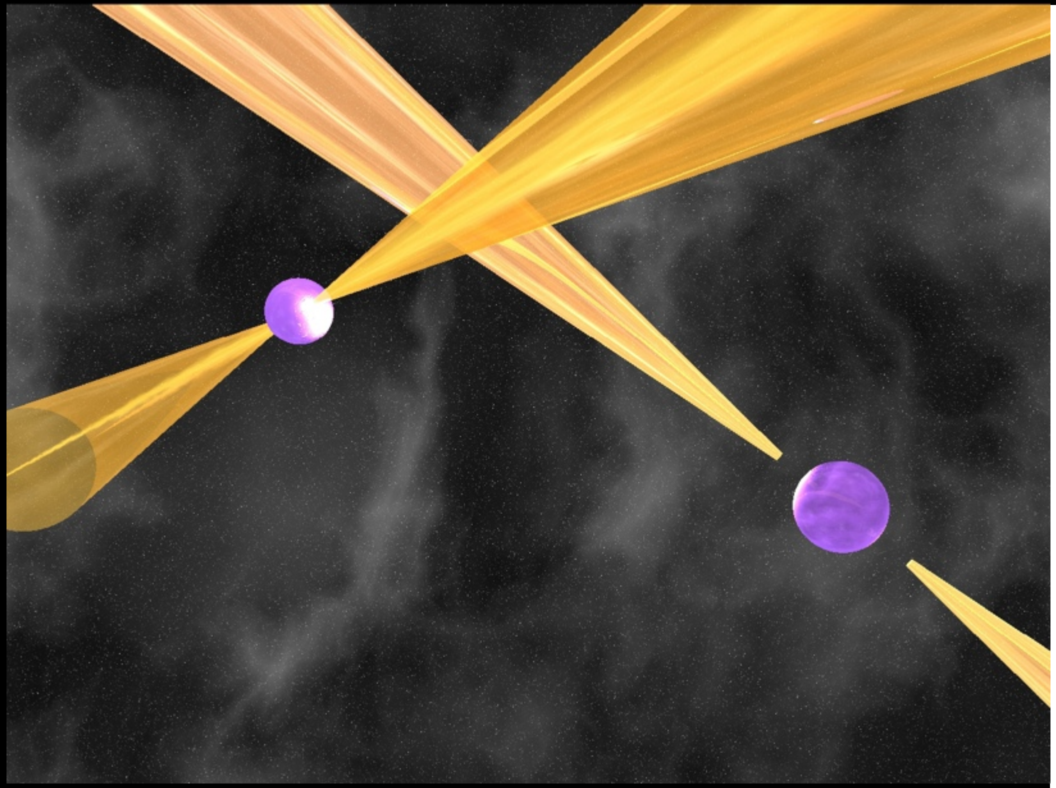
Orbital Acceleration



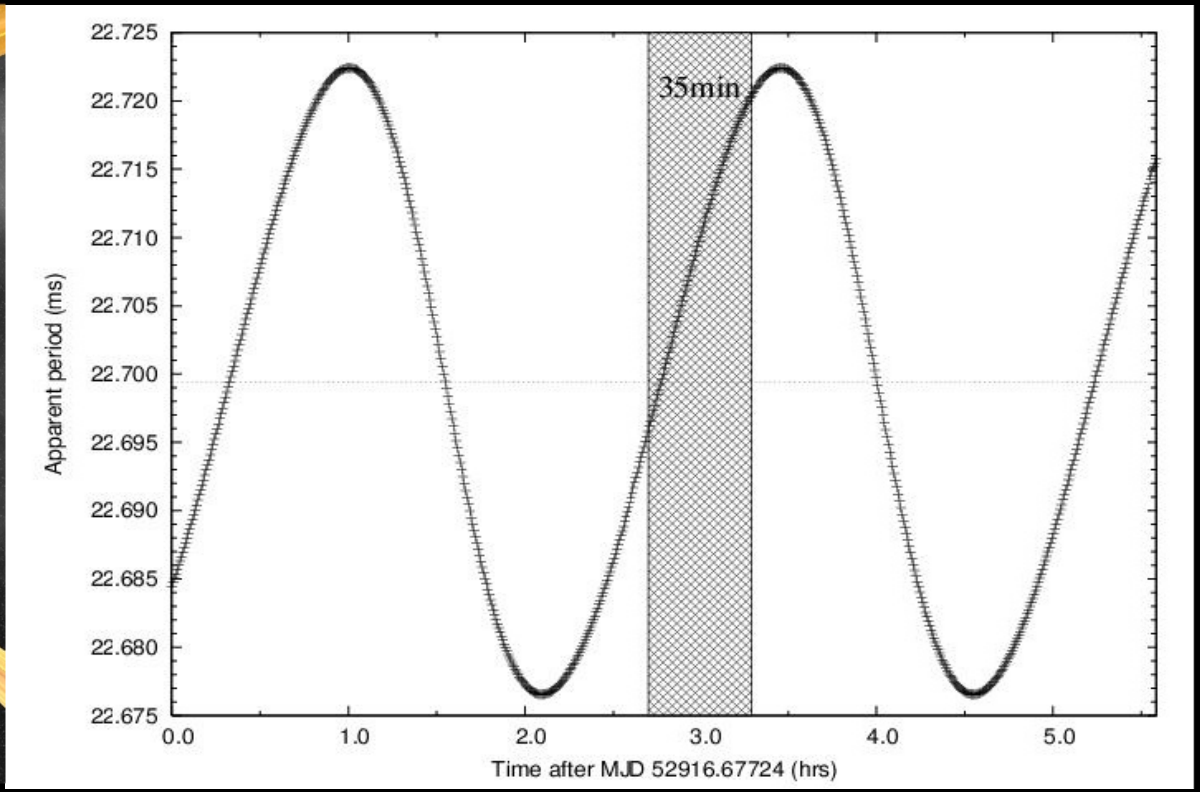
Michael Kramer (MPIfR/JBCA)

Further unknown: position on the sky

# Searching for Accelerated Signals



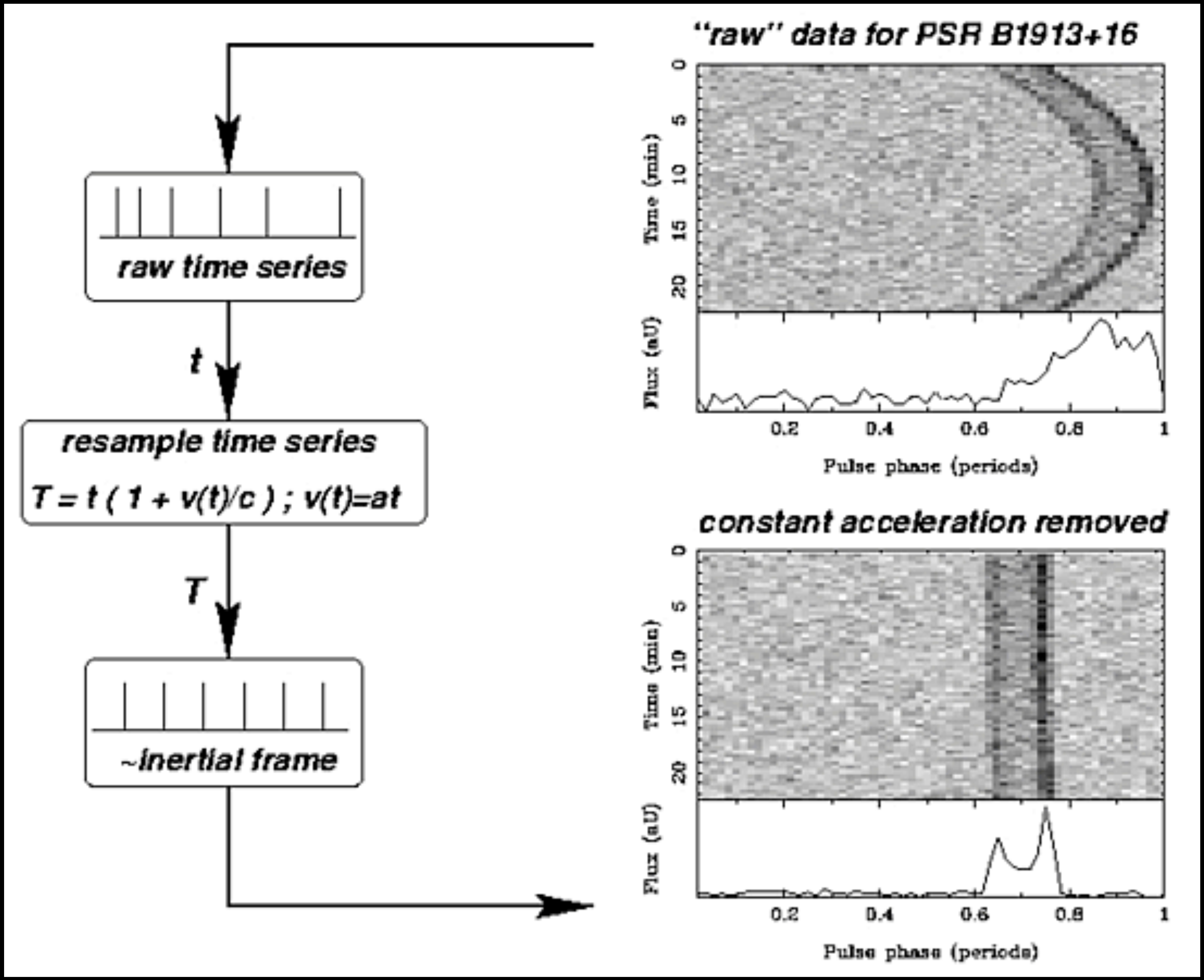
M. Kramer (MPIfR)




Observed spin pulsar spin period changes over the coarse of its orbit due to the Doppler effect

The S/N of peak is reduced in the FFT of the time series

Correct with "time domain resampling"

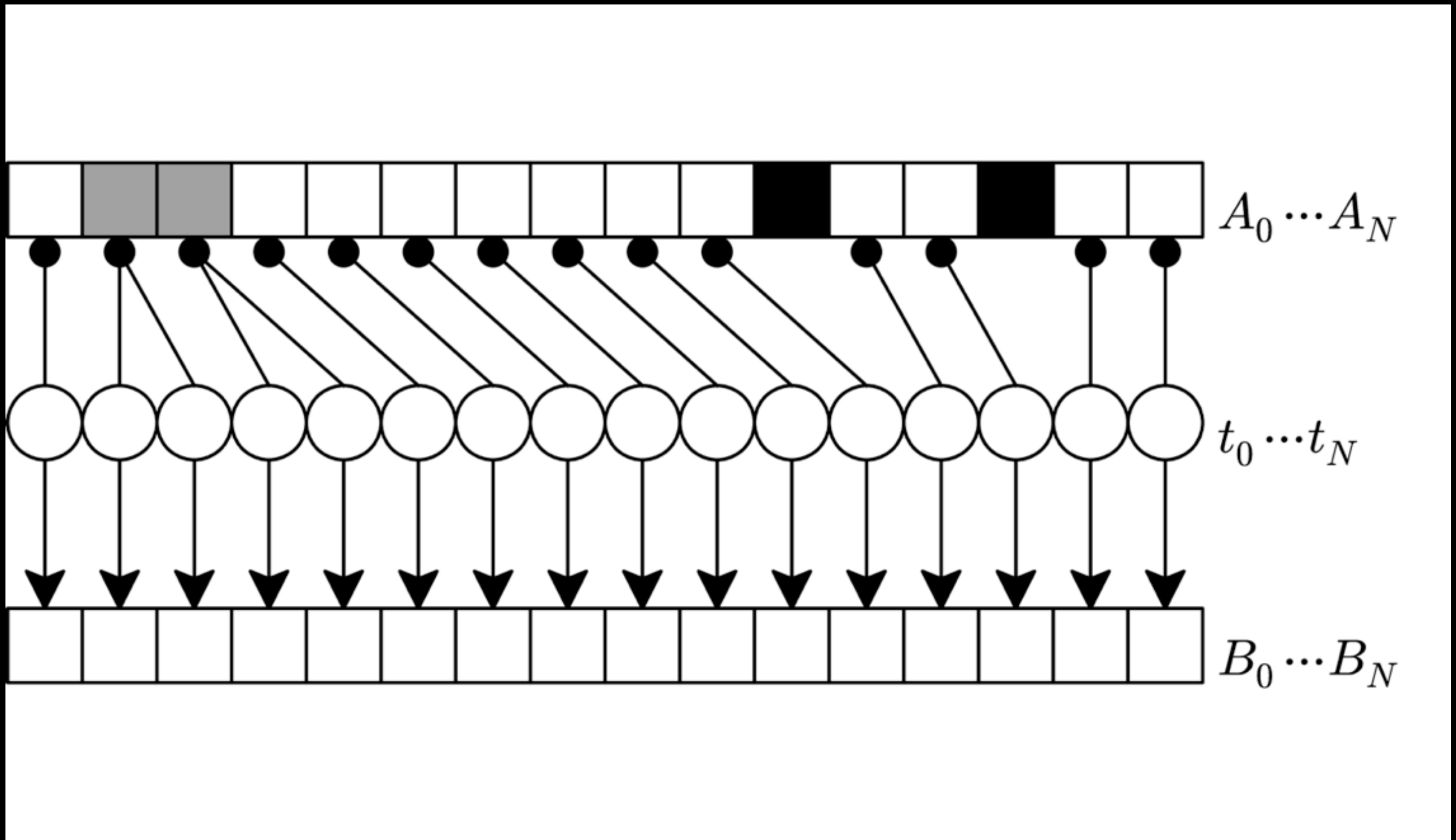



http://www.jb.man.ac.uk/distance/frontiers/pulsars/section6.html

# Correcting for Acceleration

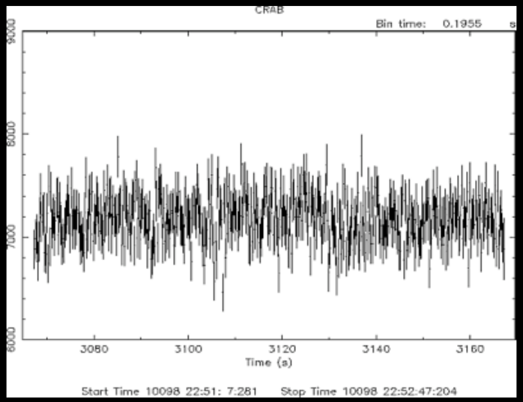
Stretch and squeeze time series for a given Doppler acceleration

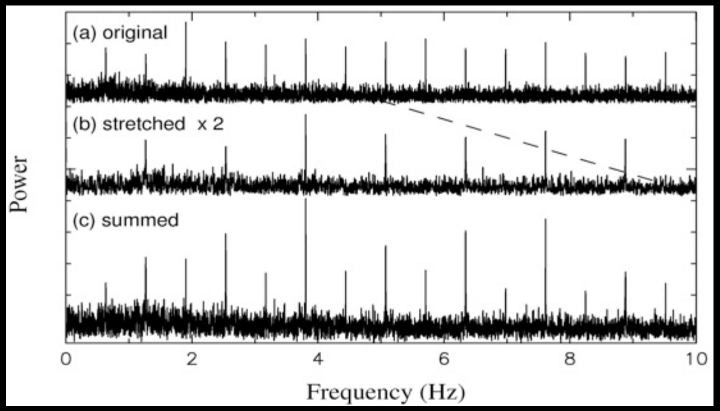
Must try a range of trial acceleration values (10s - 100s) and is calculated for each trial DM



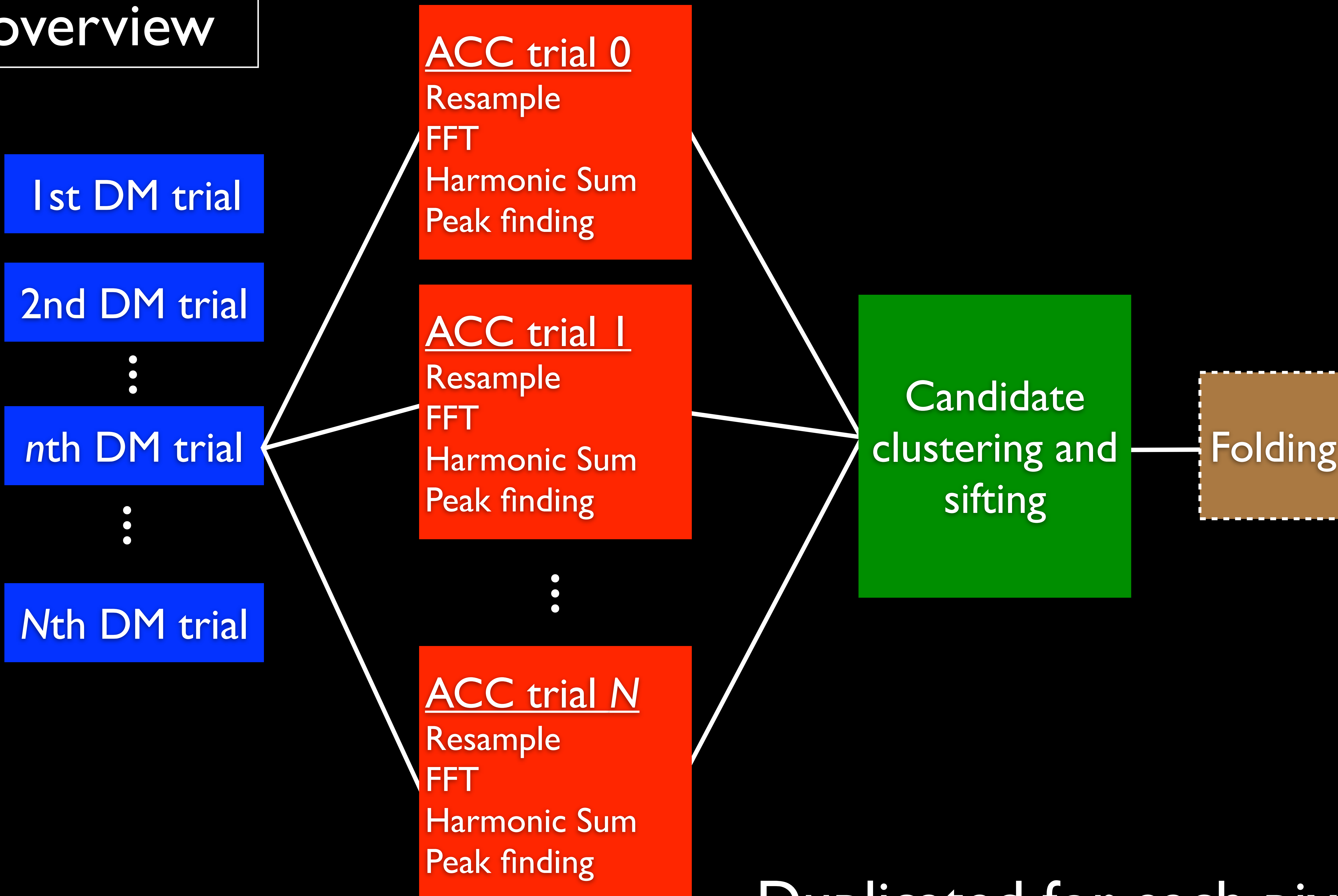

$A_n$ and $B_n$ are original and resampled time series
$t_n$ are threads

FFT + HS + Peak finding



# Pipeline overview




Duplicated for each pixel

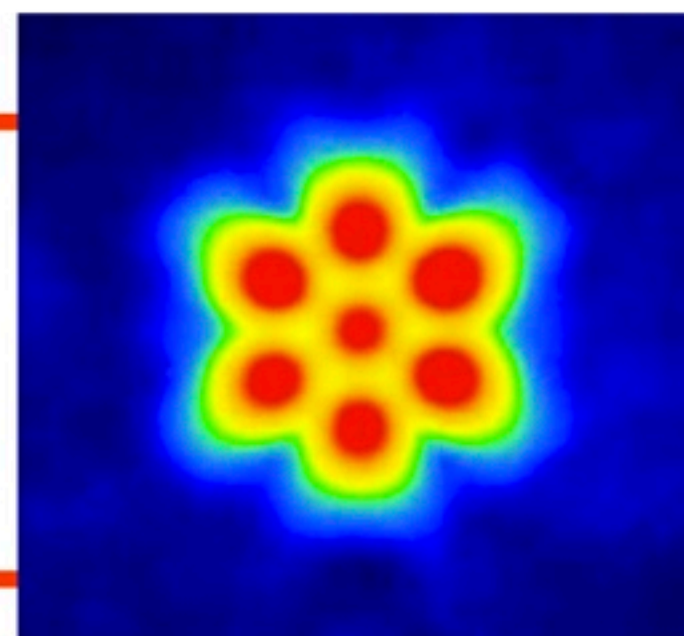
# THE HIGH TIME RESOLUTION UNIVERSE SURVEY (HTRU)

**All-sky hunt for pulsars with the Effelsberg and Parkes radio telescopes**

# HTRU-North: **Setups**

- The first all-northern-sky survey in more than 20 years!
- Has much higher time and frequency resolution than previous surveys
- Will probe deep into the Galactic plane in the hunt for exotic pulsars

Center frequency: 1.36 GHz
Bandwidth: 300 MHz
Sampling time: 54 μs
7-beam receiver



~ 180,000 sky pointings
1.5 million beams
76 million candidates
2.5 Petabytes of data

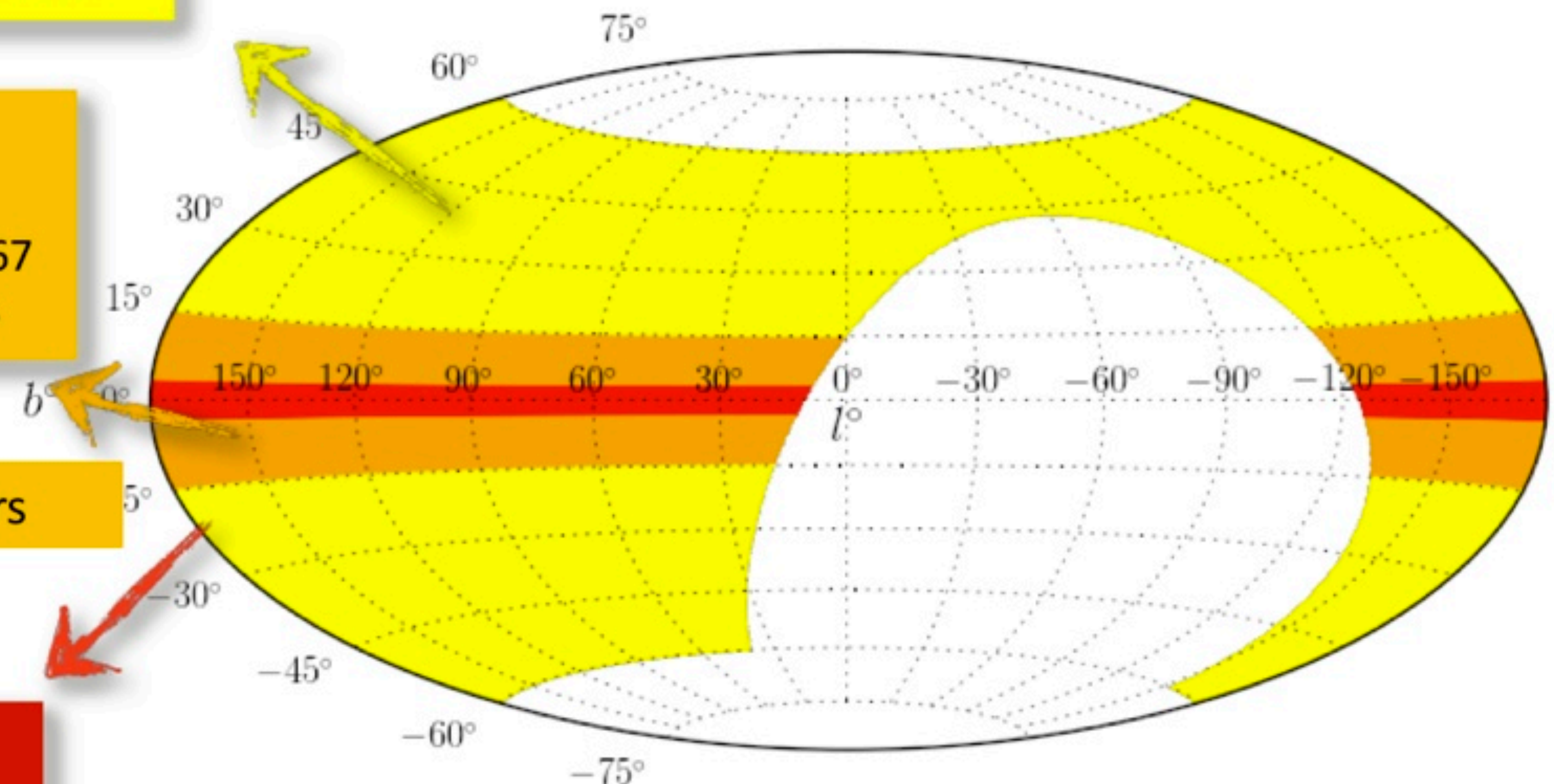
# HTRU-North: Strategy

$|b| > 15^\circ$
tobs = 90 s
Number of beams = 1066135
Data total (TB) = 818.1

bright pulsars and both Galactic and extragalactic transients

$|b| < 15^\circ$
tobs = 180 s
Number of beams = 375067
Data (total) (TB) = 575.6

bright millisecond pulsars



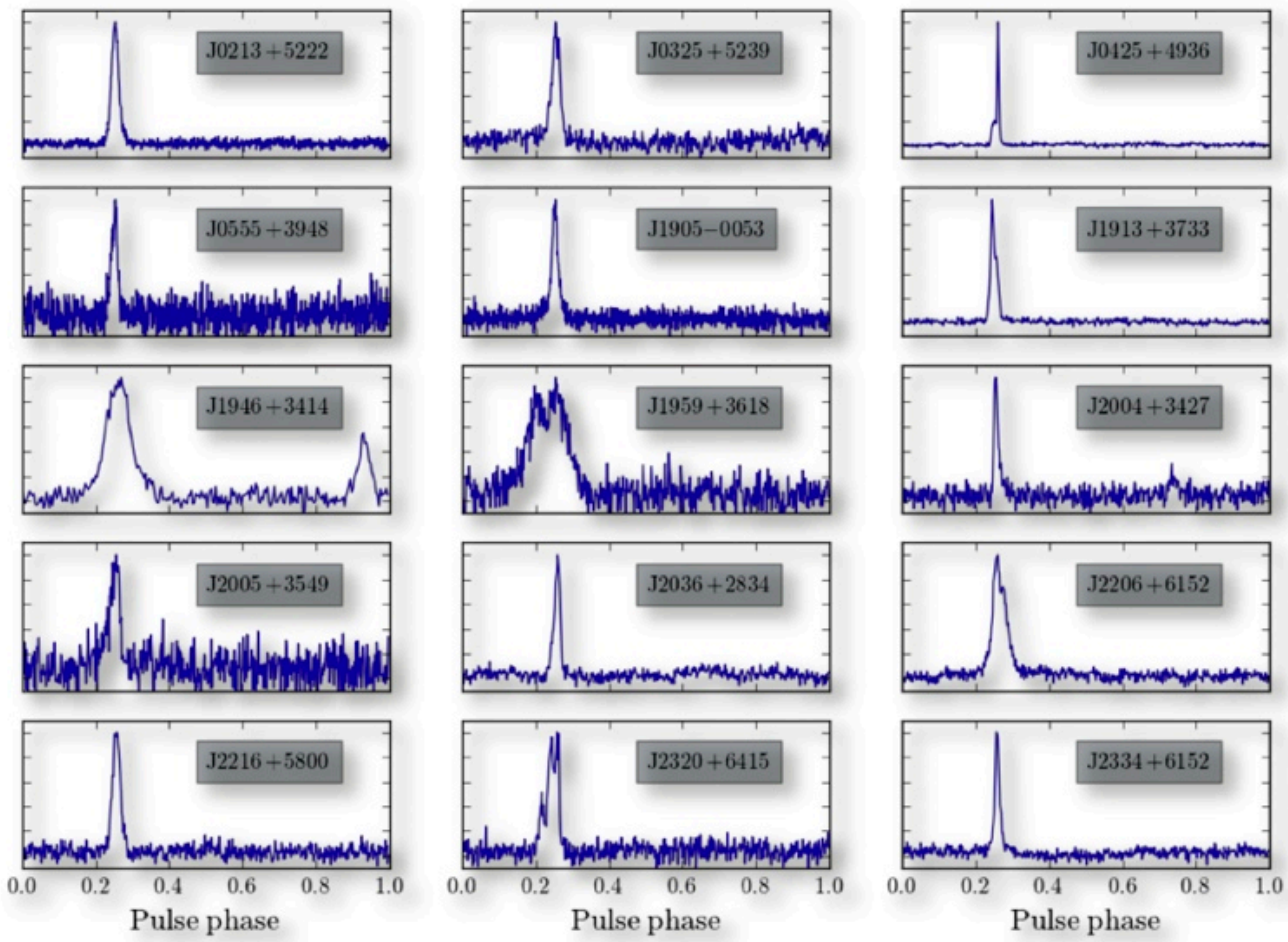

$|b| < 3.5^\circ$
tobs = 1500 s
Number of beams = 87395
Data (total) (TB) = 1117.76

double neutron stars and others exotic objects

# Data processing in the early days

| Fast pipeline | Full pipeline |
|---|---|
| Computers at Effelsberg | VLBI cluster at MPIfR (10 computing nodes) |
| Low-resolution version of the data | Full-resolution data |
| Almost in real time | Takes a **long** time |
| Sensitive to 90% of normal pulsars from the data and transients | Sensitive to pulsars in binary systems |
| Periodicity search and single-pulse search | Periodicity search with acceleration search |

# HTRU-North: First discoveries



# PSR J1946+3417: The most massive pulsar ever?

Spin period: **3.17 ms**

Orbital period: **27.02 days**

Orbital eccentricity: **0.134**

**( highly eccentric orbit!)**

Companion mass: **0.27 $M_{Sun}$**

Nature of companion: **white dwarf**

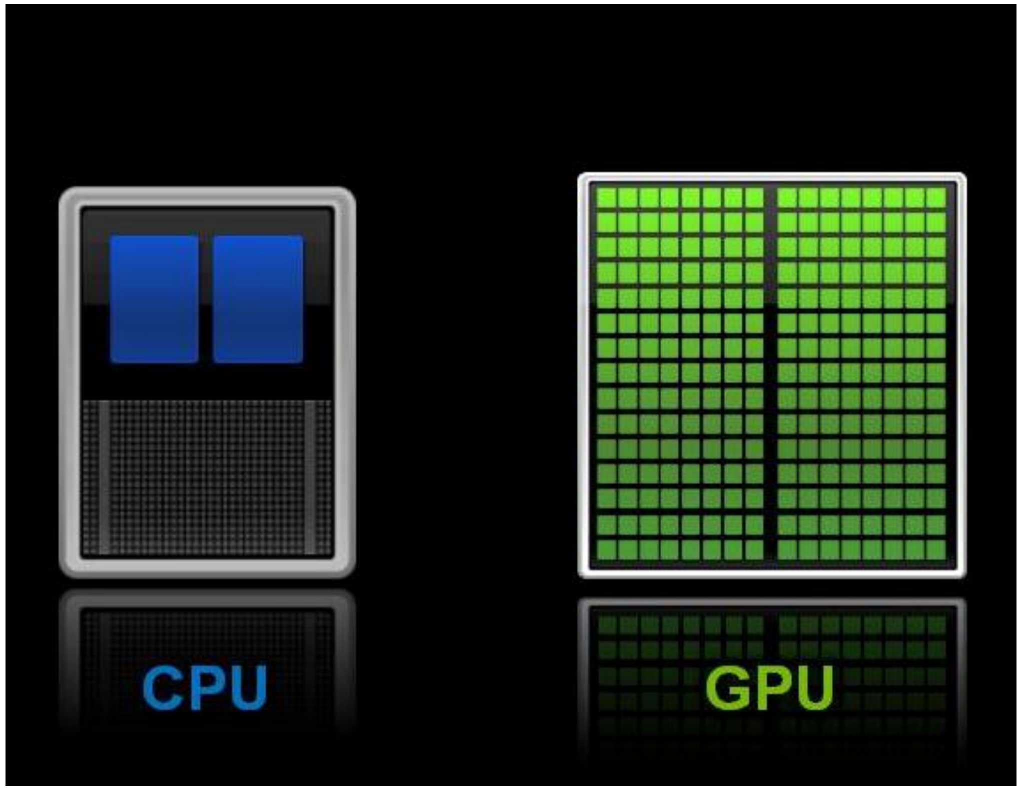
Total mass of the system: **2.2 $M_{sun}$**

**$M_p$=1.93 ± 0.07 $M_{Sun}$**

JÜLICH
FORSCHUNGSZENTRUM

# GPU pipeline: a new word in pulsar data processing

A new pulsar processing pipeline that runs on GPUs.
Large speed-ups in processing over traditional CPUs




**Juropa3 Experimental Cluster**

**Four GPU nodes**

**Two Kepler GPU cards each**

**PEASOUP**

# GPU processing: optimistic (future) numbers

- **900** GPU-hours per week (night time, weekends)
- **~4.5** minutes for processing one beam on a single GPU
- **12000** beams per **week** using **two** GPUs on every node (**8** GPUs)**!!!**

**CPU PIPELINE:** **12000** mid-lat beams in > **one year**

**GPU PIPELINE:** **100000** mid-lat beams can be processed in

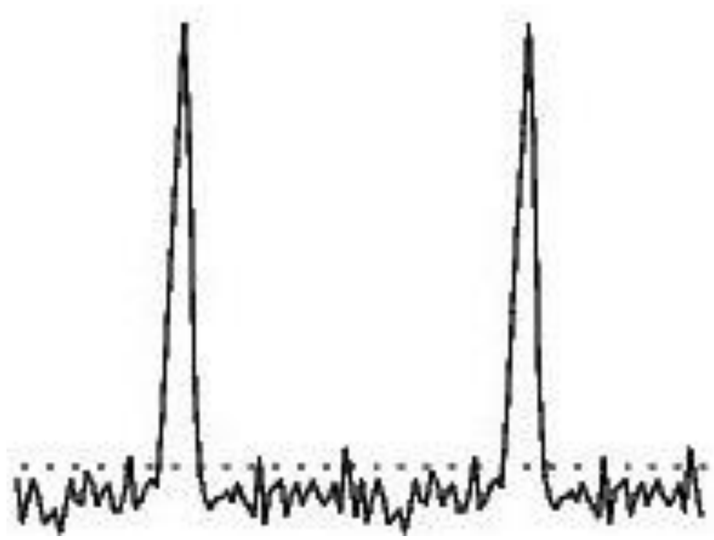
**9 GPU-weeks!!!**

Number of mid-lat beams processed with GPU: **13500** (~8% of the mid-lat beams observed by now ≈ 3% of the mid-lat beams planned to be observed in the survey)

# What has been found: Two new bright MSPs

**(and tens of known pulsars :)**

**PSR J2045+3633**



**Spin period:** 31.68 ms

**Orbital period:** 32.3 days

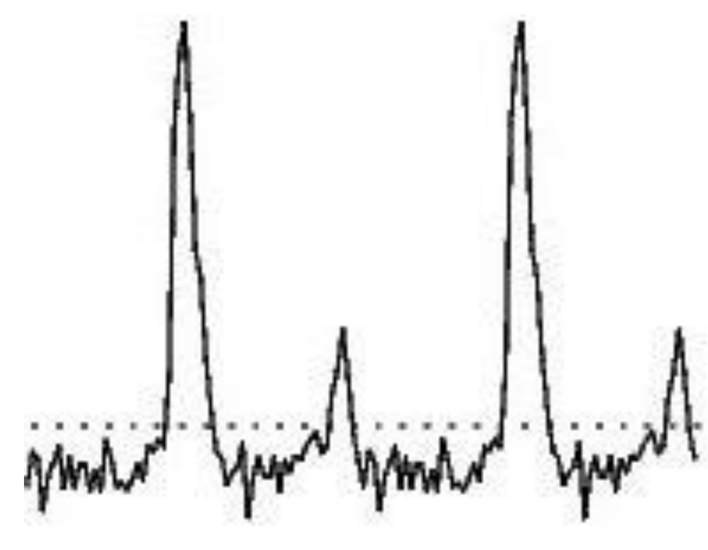
**Orbital eccentricity:** 0.017

**(unusually eccentric orbit!)**

**Minimum companion mass**: 0.8 $M_{Sun}$

**Nature of companion**: massive white dwarf or neutron star (?!)

**PSR J2053+4650**



**Spin period:** 12.58 ms

**Orbital period:** 2.45 days

**Orbital eccentricity:** 0.000008

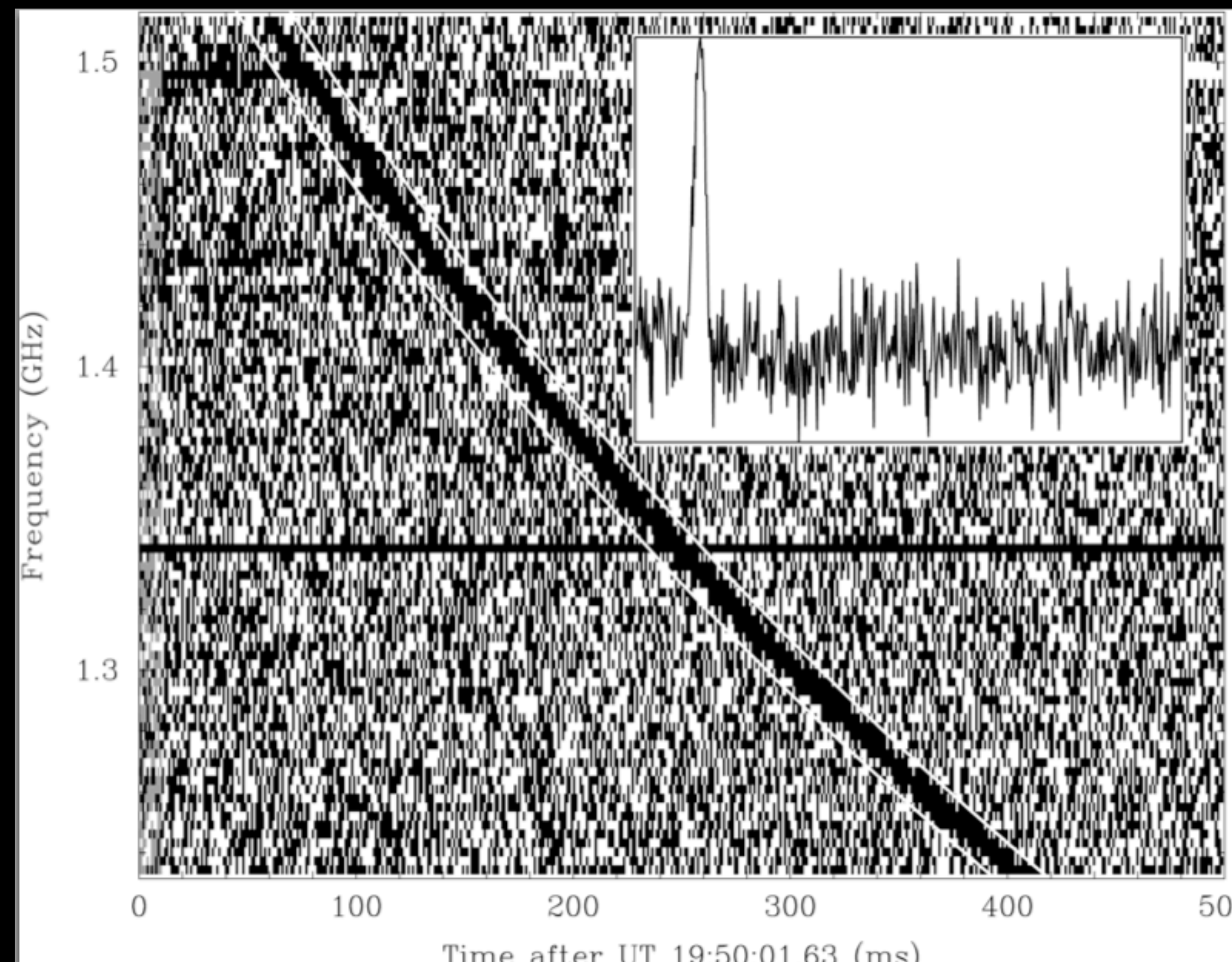
**(circular orbit!)**

**Minimum companion mass**: 0.8 $M_{Sun}$

**Nature of companion**: massive white dwarf

# Fast Radio Bursts

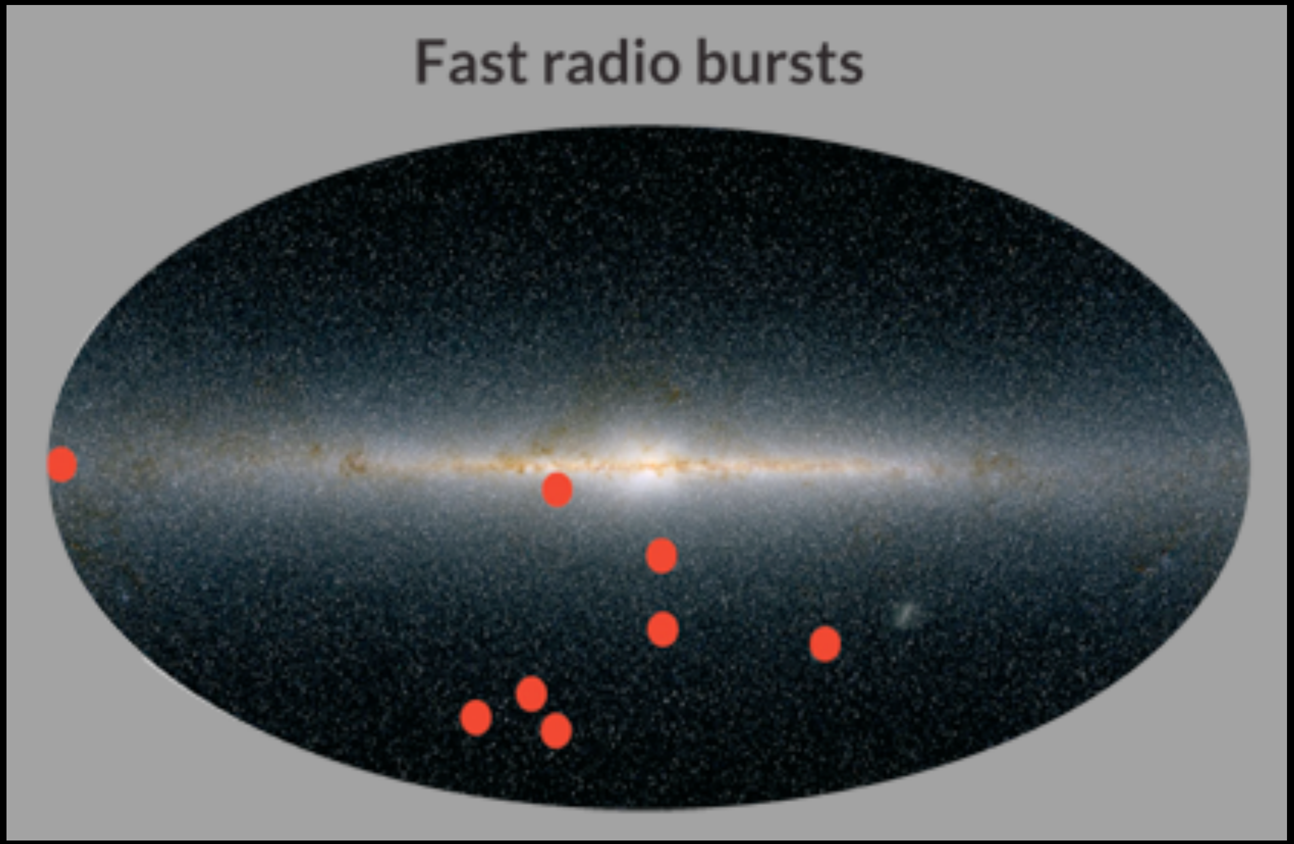
- Short-duration radio bursts
- Discovered in archival pulsar data
- Implied distances place the sources in other galaxies
- Apparently non-repeating



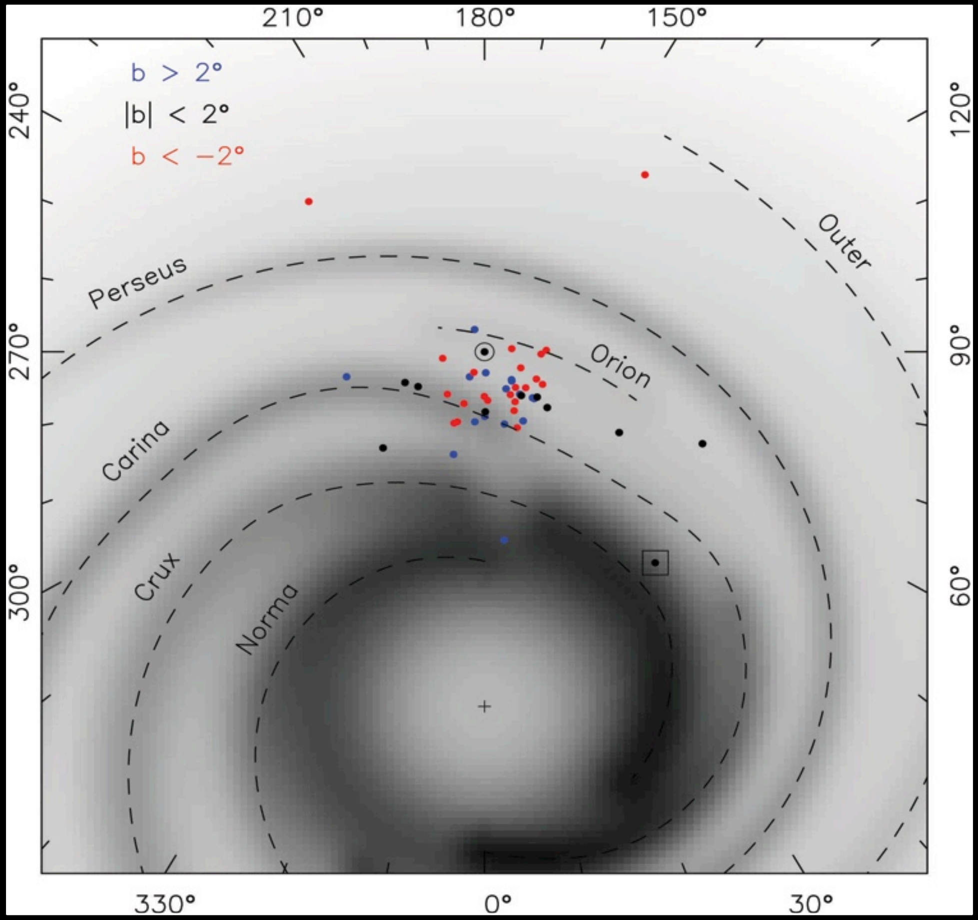

Lorimer et al 2007

# Extragalactic?

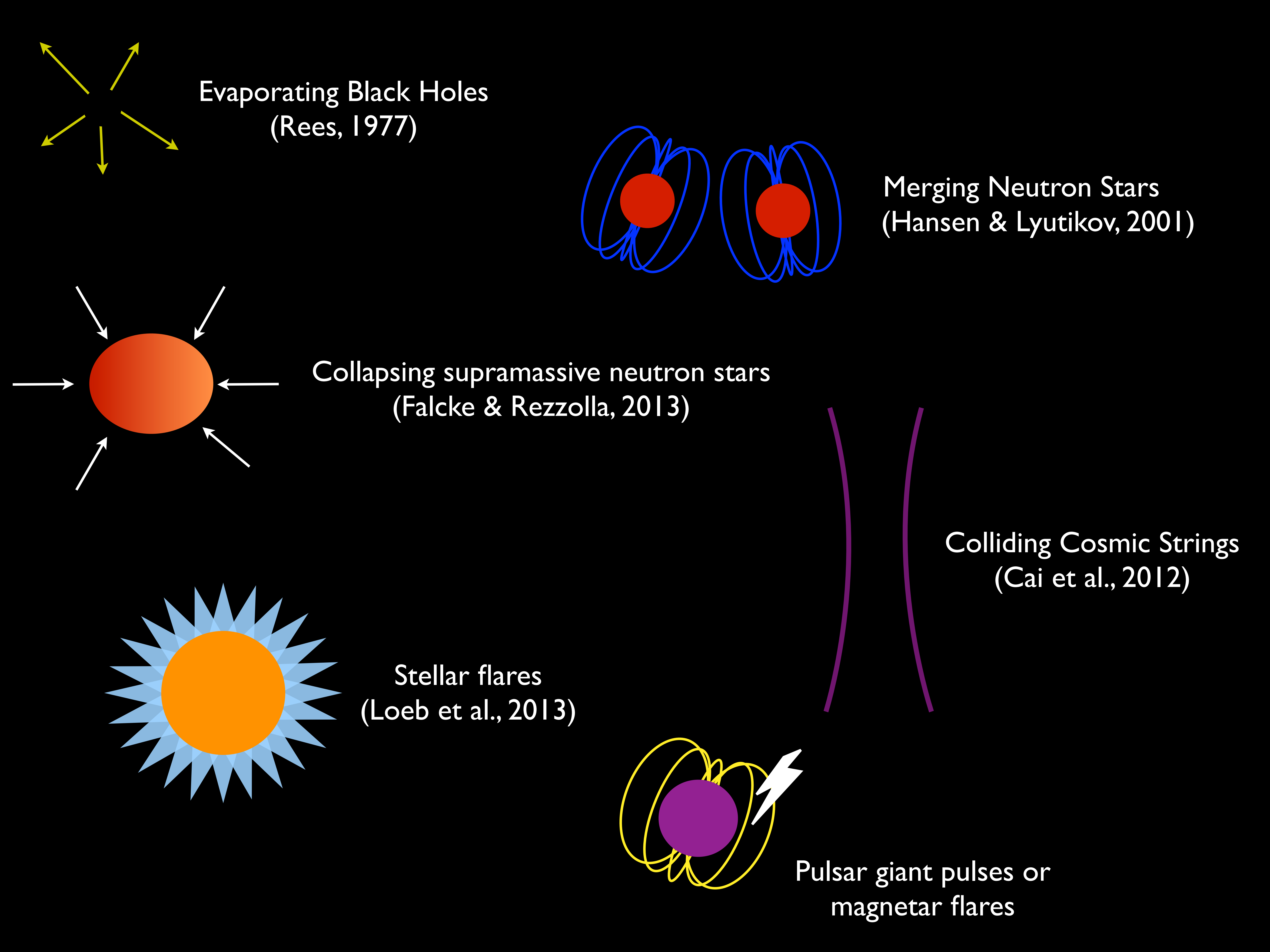
Observed dispersion measure is too large for a source in our Galaxy




J.Carpenter, T.H. Jarrett/2MASS, R. Hurt, C.




From observations of pulsars, we can model the distribution of electrons in our Galaxy


Evaporating Black Holes
(Rees, 1977)
Merging Neutron Stars
(Hansen & Lyutikov, 2001)
Collapsing supramassive neutron stars
(Falcke & Rezzolla, 2013)
Colliding Cosmic Strings
(Cai et al., 2012)
Stellar flares
(Loeb et al., 2013)
Pulsar giant pulses or
magnetar flares

# Goal: Discover FRBs in realtime in order to trigger follow-up

LOFAR

Hans Hordijk

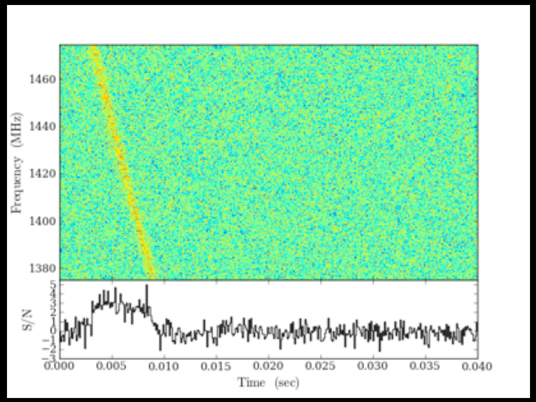
Non-radio follow-up

SDSS Team, Fermilab Visual Media Services

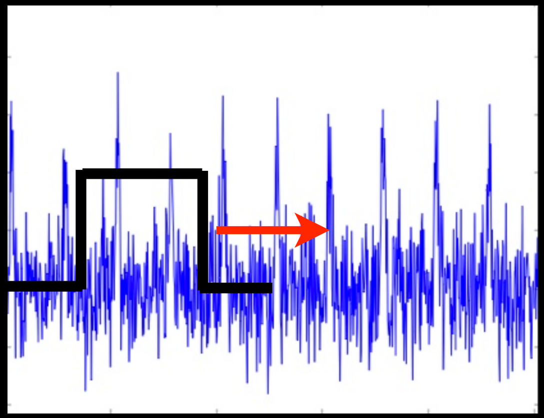
# “heimdall” transient detection code

- Written by Ben Barsdell
- All processing done in GPU
- Uses the `dedisp` library

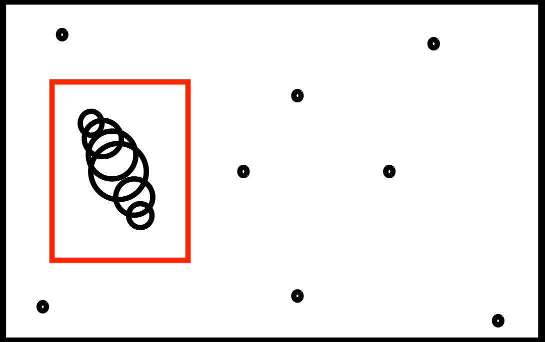
Dedispersion
(loop over trials)




Matched filtering
(loop over boxcar filter widths)



Peak detection
and clustering



Processing archival HTRU-N data on JUROPA3

Real-time system running in parallel to HTRU-N observations at Effelsberg

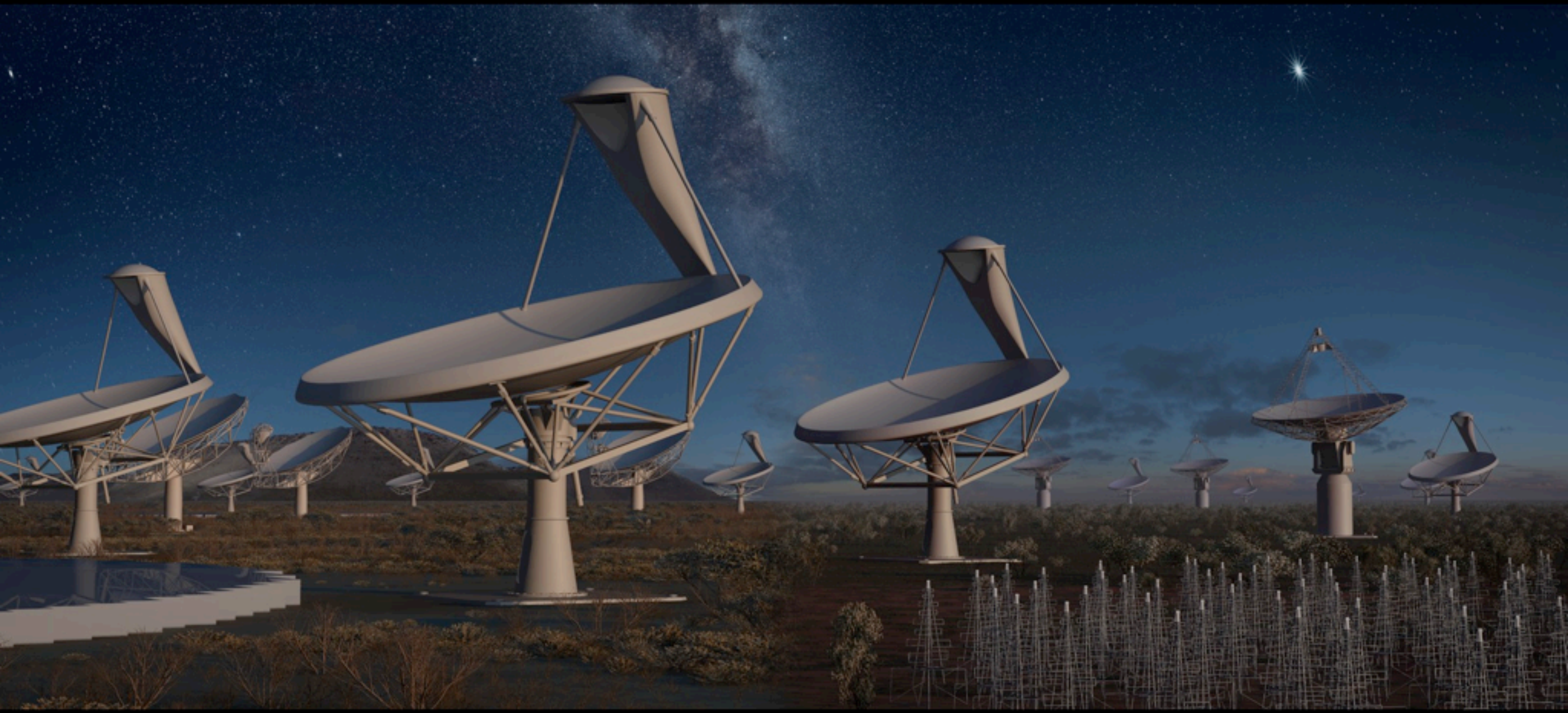
Processing archival HTRU-N data on JUROPA3

Processed ~8% of data
Working on visualizing and digesting results
Expecting 2-3 fast radio bursts

Real-time system running in parallel to HTRU-N observations at Effelsberg

Currently working on interfacing spectrometer with GPUs

# Square Kilometer Array



To be built in South Africa and Australia

Three configurations with different technologies

# Computational challenge

Exaopt-scale realtime processing

GPUs will likely be play a role in the processor

| | Present | SKA |
|---|---|---|
| $N_{beams}$ | 7 to 13 | 2000 |
| $N_{DM}$ | ~1000 | ~6000 |
| $N_{acc}$ | ~70 | ~700 |

... plus sifting, candidate identification, folding

Questions?
Fragen?