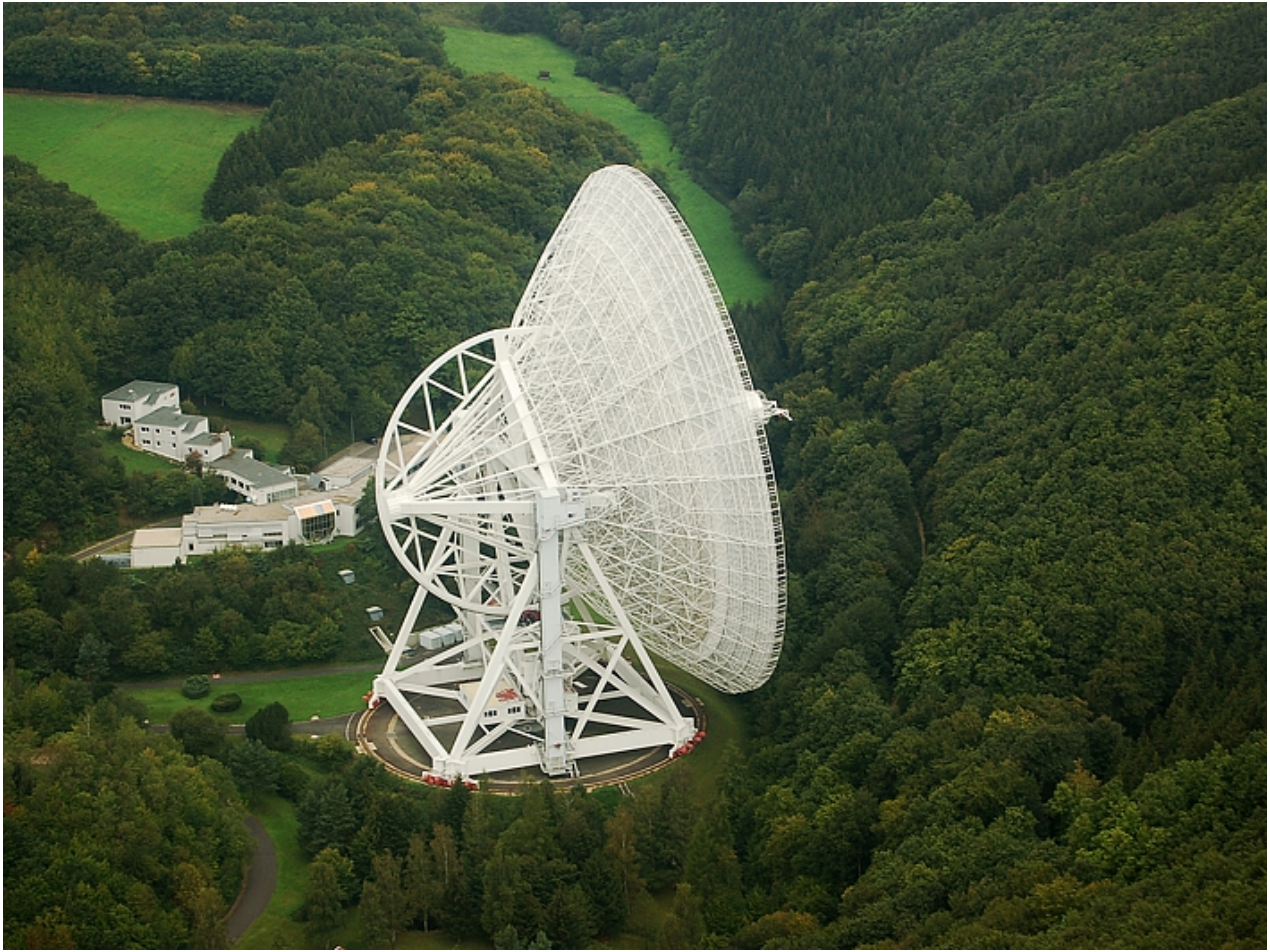
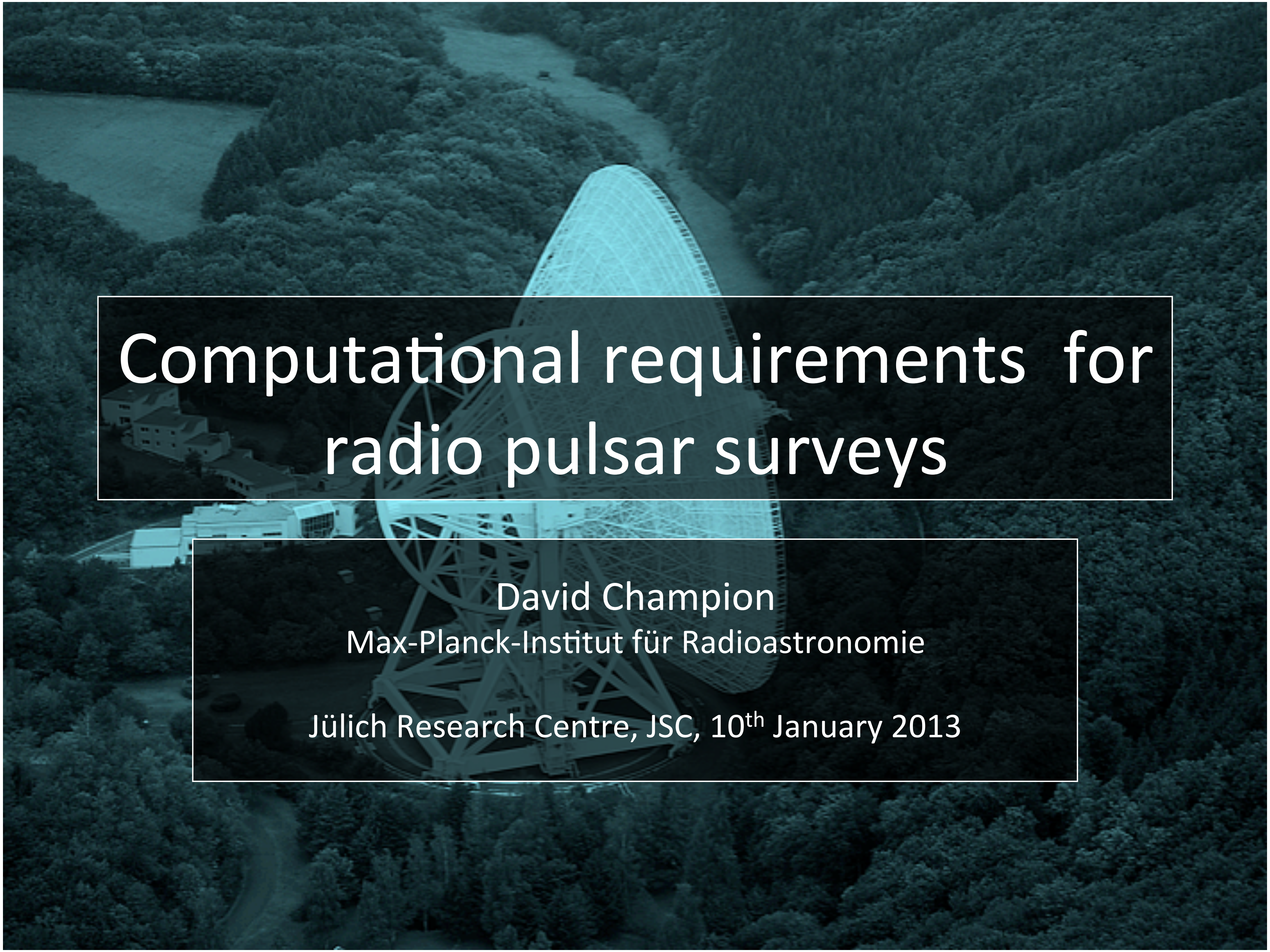
Computational requirements for radio pulsar surveys

David Champion
Max-Planck-Institut für Radioastronomie

Jülich Research Centre, JSC, 10th January 2013





Computational requirements for radio pulsar surveys

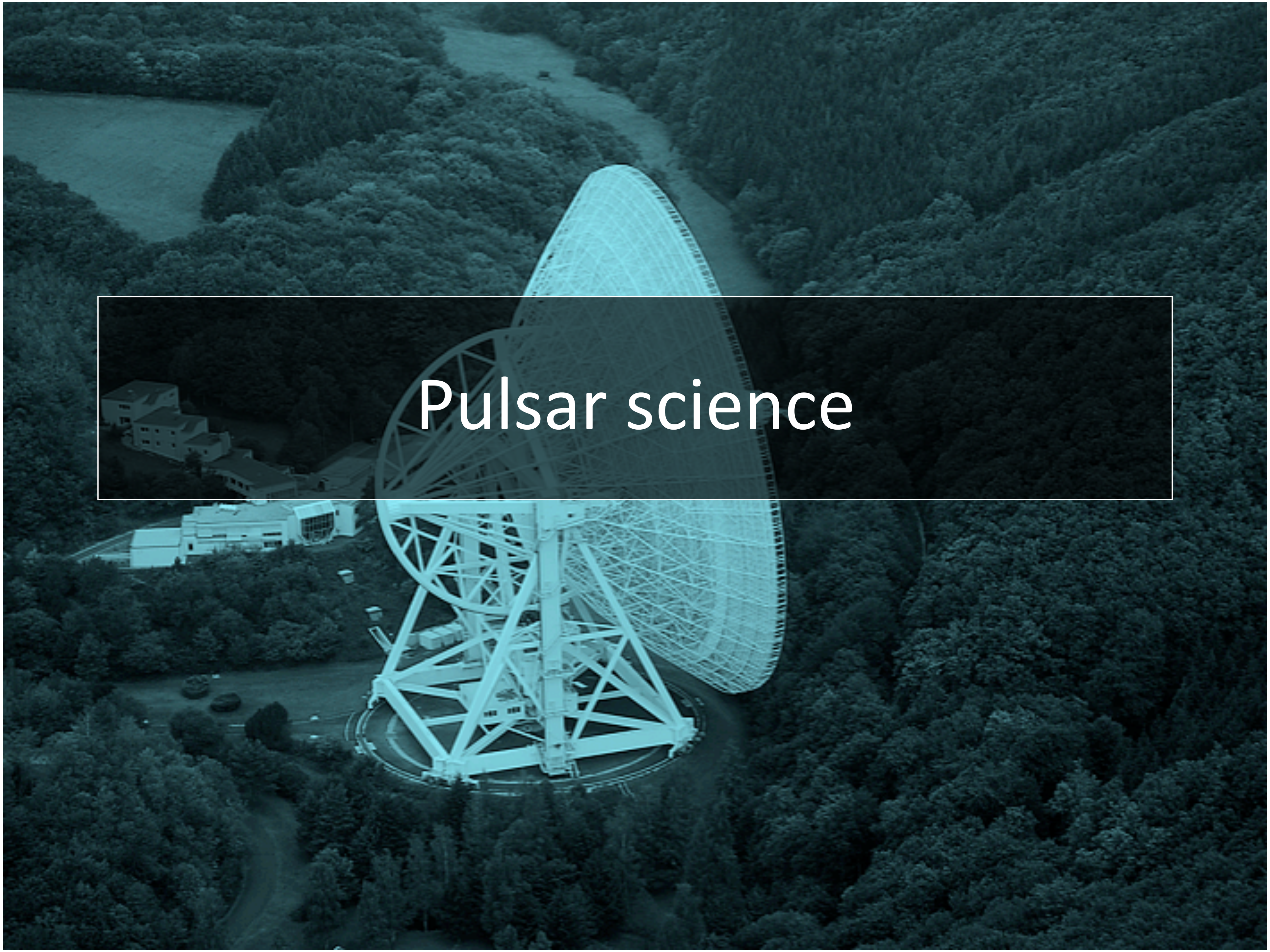
David Champion
Max-Planck-Institut für Radioastronomie

Jülich Research Centre, JSC, 10th January 2013

Outline

- Pulsar science
- Pulsar searching
  - Rapid time sampling
  - Pulse dispersion
  - Binary searching
  - Current survey details
- Near future (this year)
  - Phased Array Feeds
  - Ultra-broadband receivers
- The Square Kilometre Array

Pulsar science



What is a pulsar?

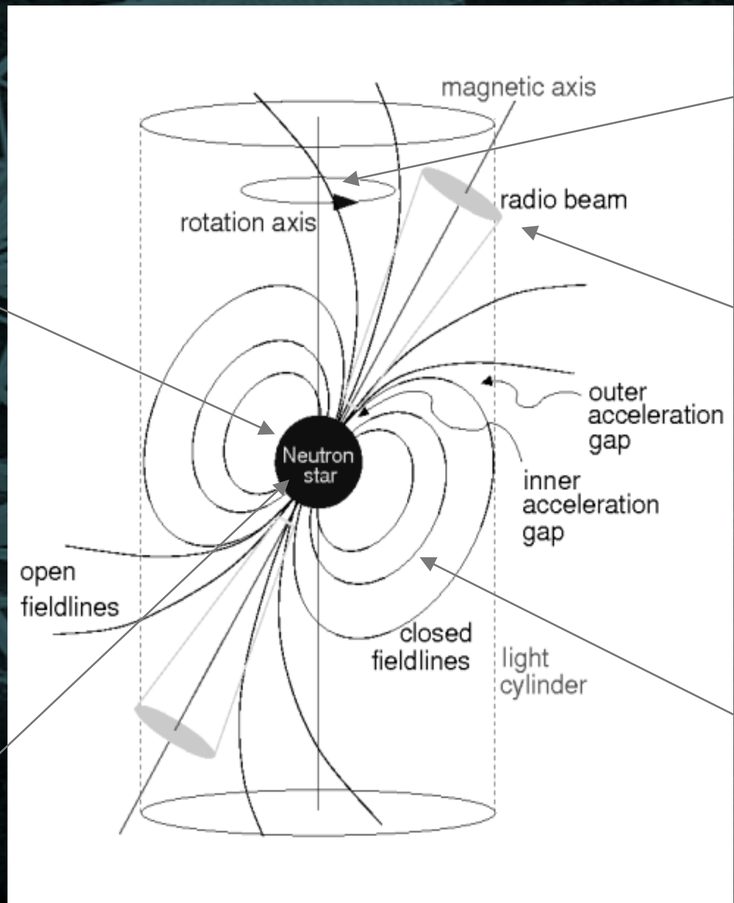
The end of a massive star....

....the birth of a pulsar

What is a pulsar?

Discovered 1967 by PhD student Jocelyn Bell-Burnell and supervisor Antony Hewish at Cambridge.

Diameter = 20km



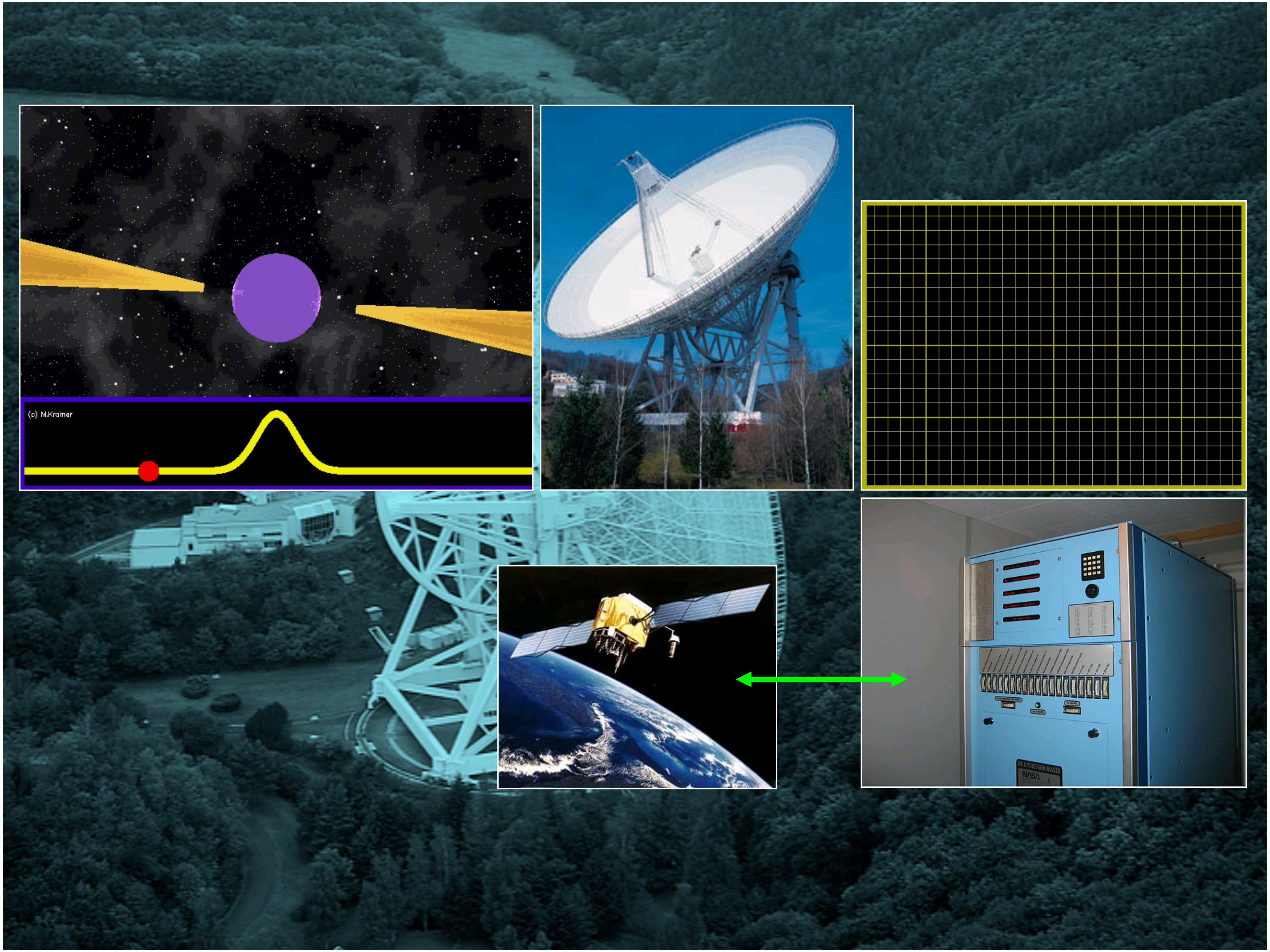
Rapidly rotating, up to 700 times per second.

Cosmic lighthouse.

Neutron star more dense than nucleus of atom, $10^8 - 10^{11}$ kg/cm^3. Mass = 1.4 solar mass.

=

Strong magnetic fields of order 10^7 - 10^8 Tesla c.f. most powerful Earth based magnets few Tesla.


(c) M.Kramer

Pulsar Populations

- More than 1800 pulsars discovered
  - ~100 have rotational periods < 30 ms (MSPs)
  - ~100 are in binary systems
  - ~30 null or mode switch
  - 8 are double neutron star systems (DNS)
  - Only 1 double pulsar system
  - No BH-PSR systems… Yet!
- Various populations allow insights into several different areas of research
- The most interesting pulsars are the rarest

Pulsars are precise...

Rotational period of PSR J1713+0747 (31.7.2007, 0:00 UTC):

4.570136597647441 ms
± 0.000000000000006

The orbit of PSR J0737+3039 shrinks by
7.42 ± 0.02 mm/day

- Millisecond pulsars (MSPs) stability of pulse period rivals terrestrial atomic clocks.
- Pulsars are a unique and versatile tool for many physical applications e.g.

  - *Direct detection of gravitational waves.*
  - *Equation of state of extreme matter.*
  - *Detection of particles beyond the standard model.*
  - *Tests of strong field gravity using binary pulsars.....*

Testing Einstein with binary pulsars

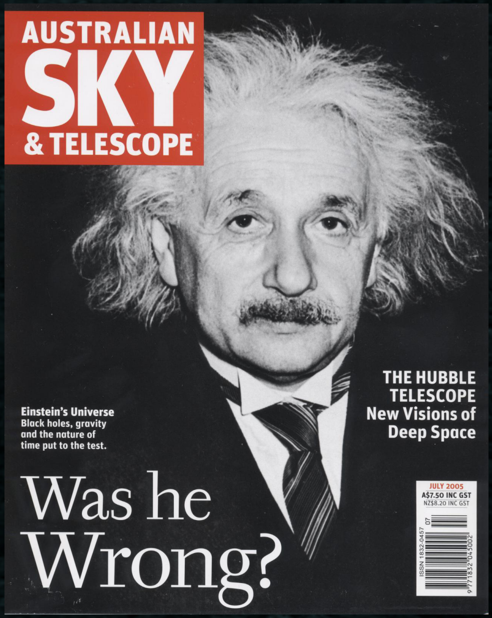
- General relativity (GR) describes the force of gravitation.

- Is GR the full description of gravity?

Accelerated expansion of the Universe, dark energy.

Big Bang.

GR currently not compatible with quantum mechanics.



- There exist other theories of gravity that predict violations of GR. Is it possible to test GR's description of gravity?

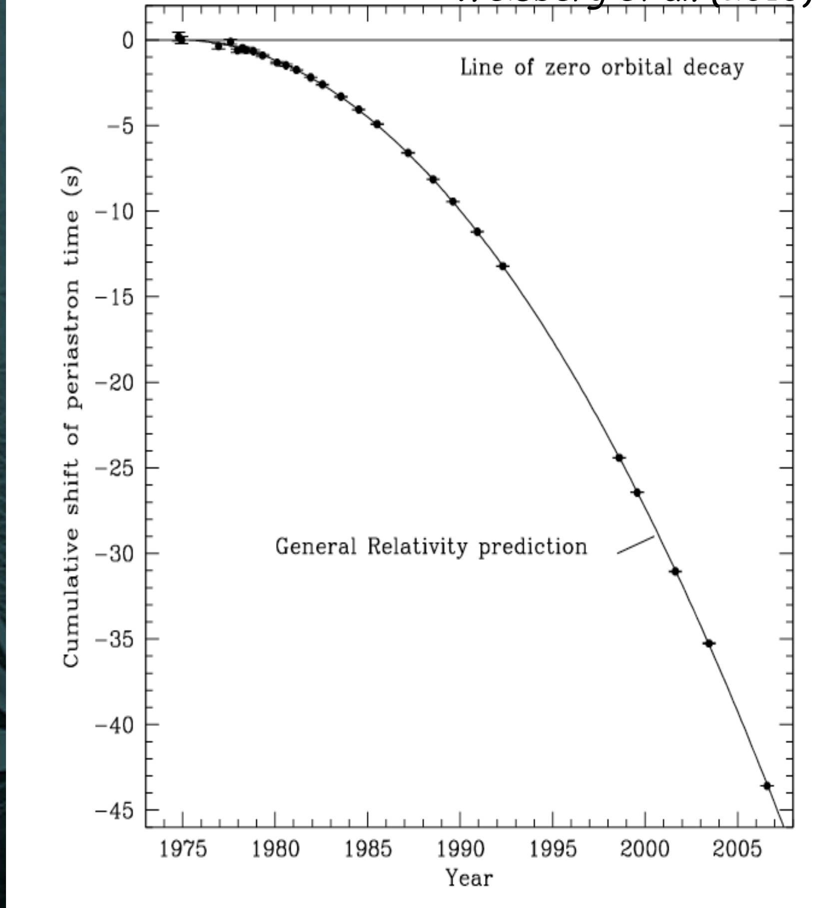
Testing Einstein with binary pulsars

- In 1974 Hulse & Taylor made the first discovery of a binary pulsar, B1913+16.
- Double neutron star (DNS) system, with orbital period of 7.8 hrs.
- Two point masses, one atomic clock, in a relativistic system!
- GR predicts modifications to the standard Keplerian orbit, so called Post Keplerian (PK) corrections.

Testing Einstein with binary pulsars

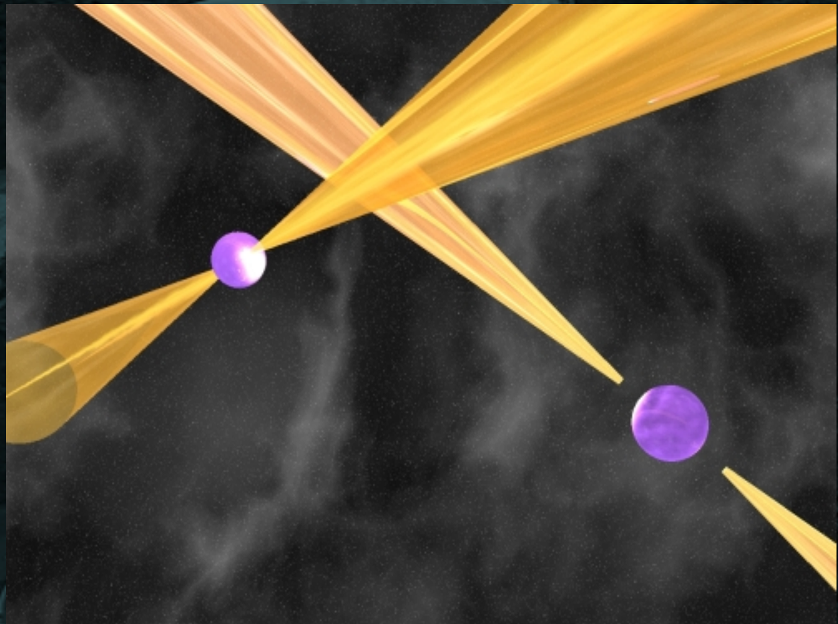
- In GR orbital period decays due to emission of energy in form of gravitational waves.
- Observations agreed with GR predictions (~ 0.2%) and inferred the existence of gravitational waves.
- In 1993 Hulse & Taylor awarded the Noble Prize in Physics.

Weisberg et al. (2010)



Testing Einstein with binary pulsars

- In 2003 the first discovery of a double pulsar, J0737-3039 A/B (Burgay *et al*. 2003, Lyne *et* al. 2004).

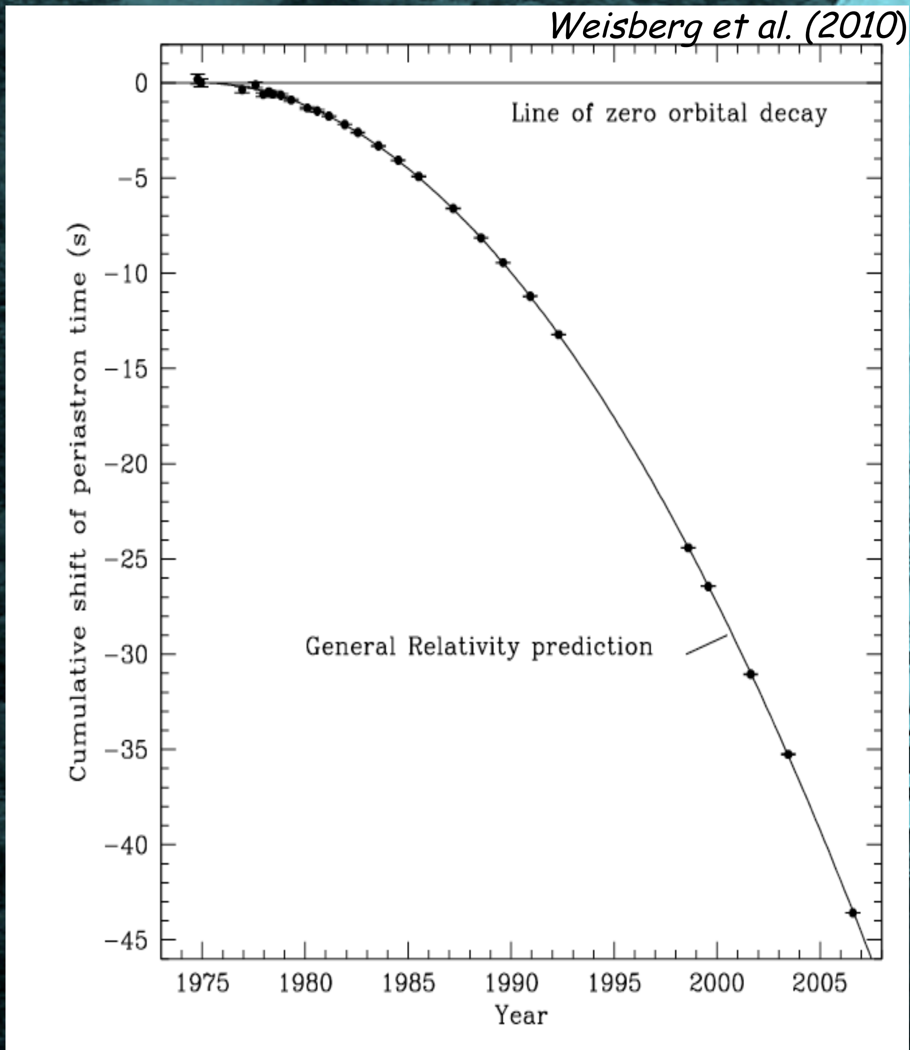


Credit: Michael Kramer

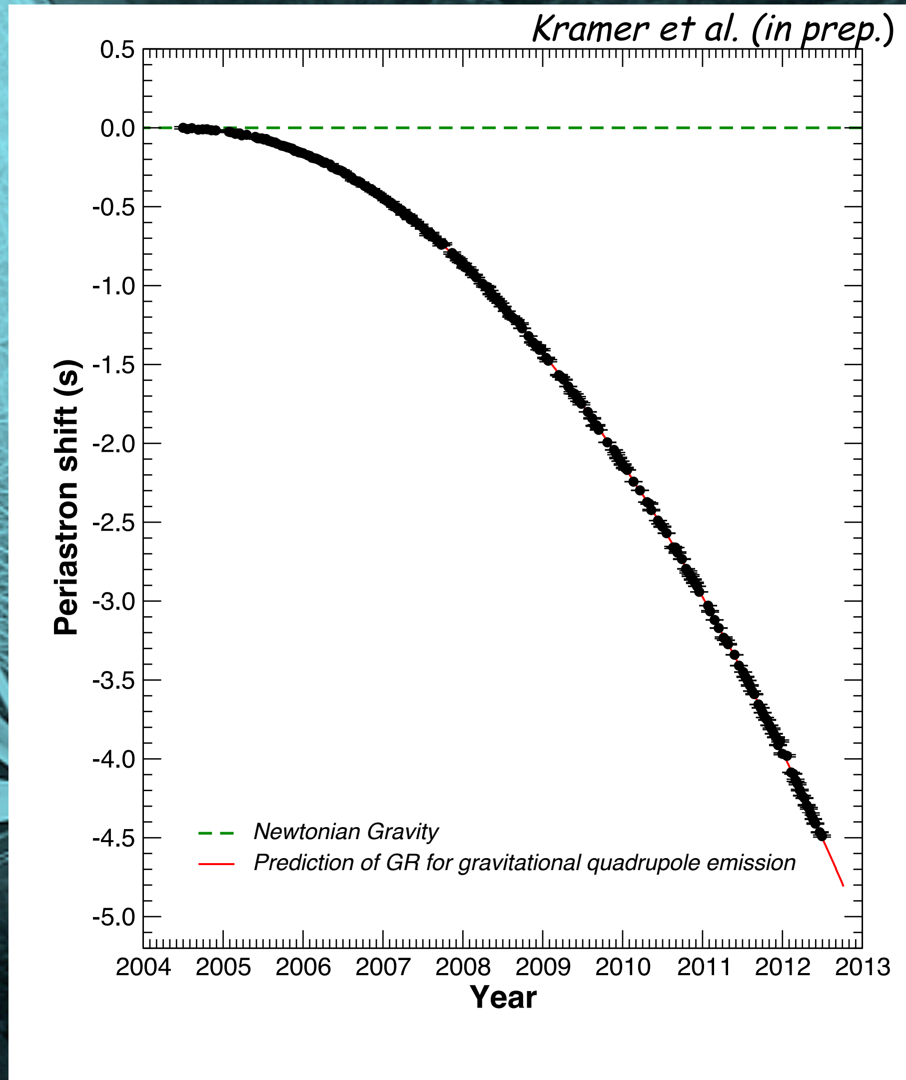
- Both neutron stars are visible as radio pulsars bound in a 2.45 hr orbit.

Gravitational wave emission

PSR B1913+16, 8-hr orbit



PSR J0737-3039, 2.4-hr orbit



New discoveries advance our knowledge

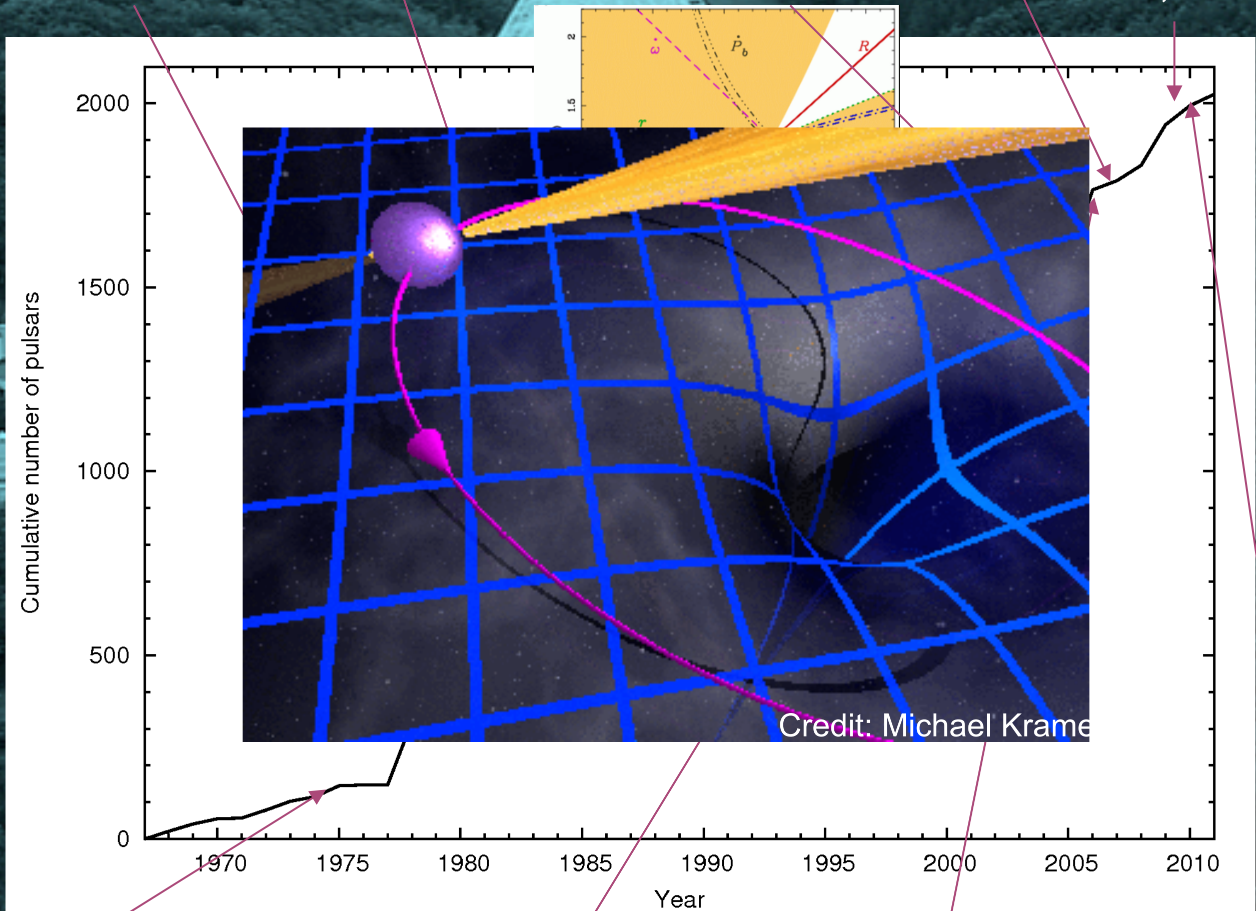
First millisecond PSR, 1982.

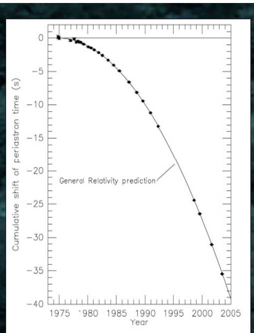
First globular cluster PSR, 1987.

Double PSR, 2004.
(most precise test of GR!)

Rotating radio transients, 2006.

Fermi points to 32 millisecon
PSRs, 2009...





Can we
do more?
PSR-BH?

First binary pulsar, 1974.
(Noble prize for inferring
the existence of GW!)

First exoplanet found
around a PSR, 1992

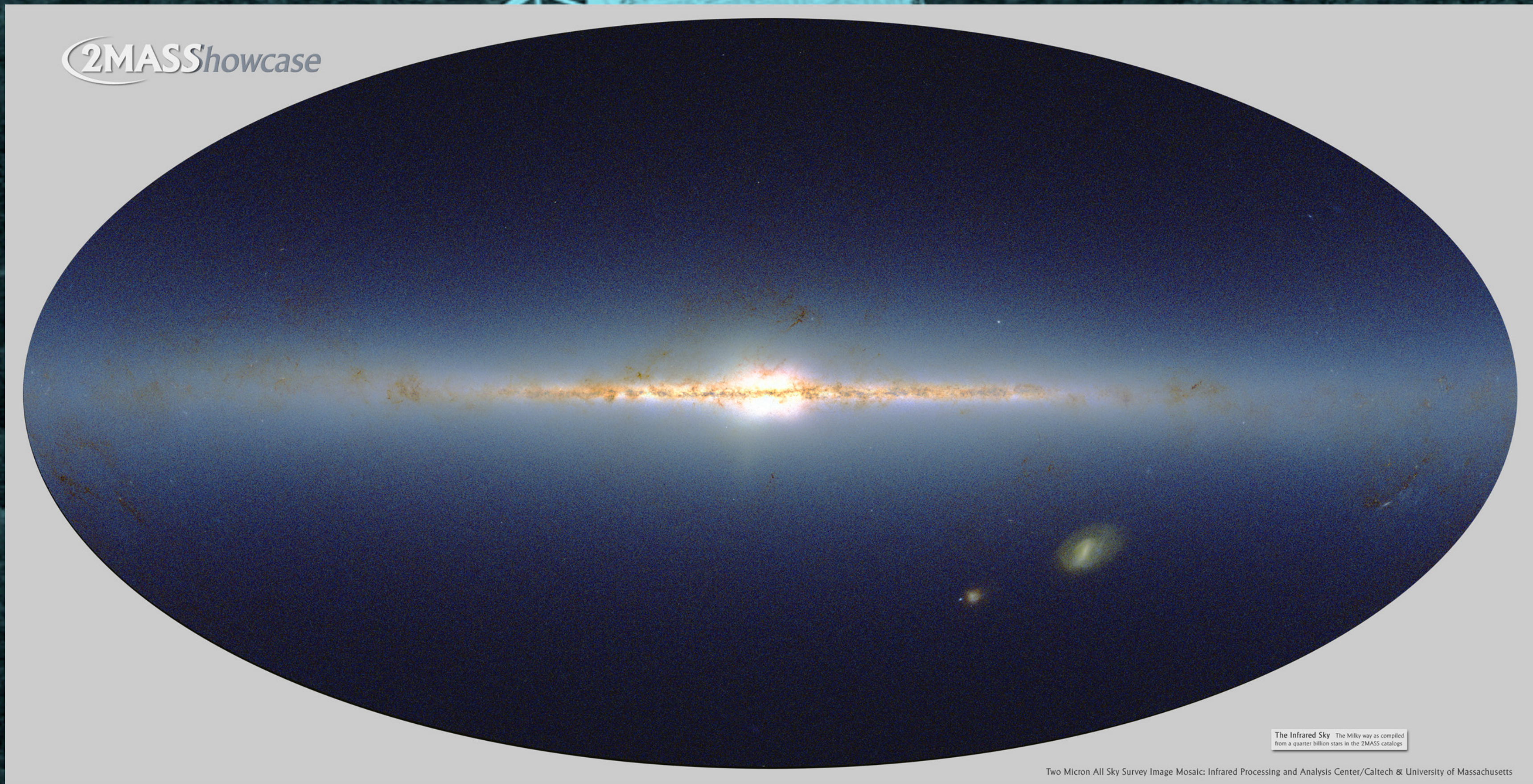
Radio magnetars, 2006.

First pulsar discovery by
the general public, 2010.

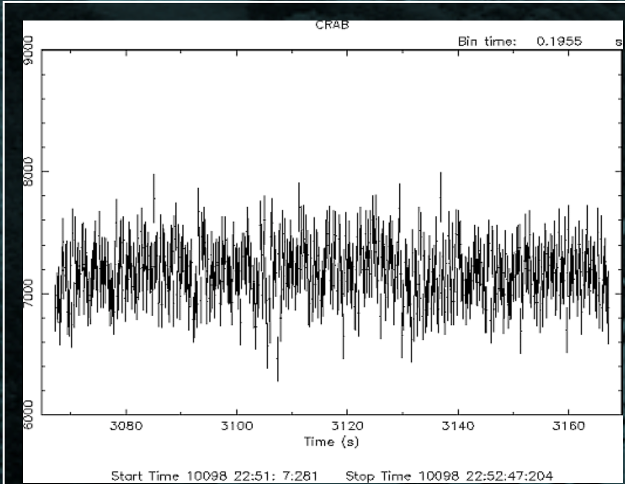

Searching for new pulsars

Where do you search?

- Pulsars are incredibly weak and can only be detected within our local Galaxy, The Milky Way (relatively nearby)

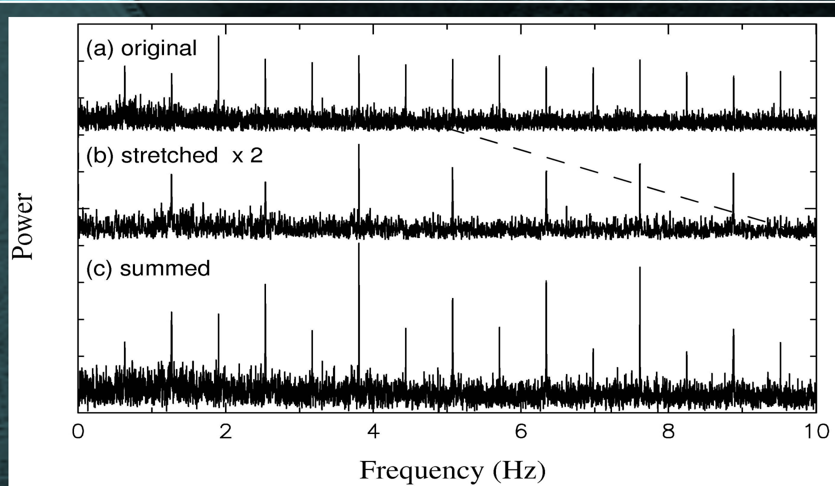


Searching for periodic signals



Time domain

FFT



Frequency domain

Hamonic summation

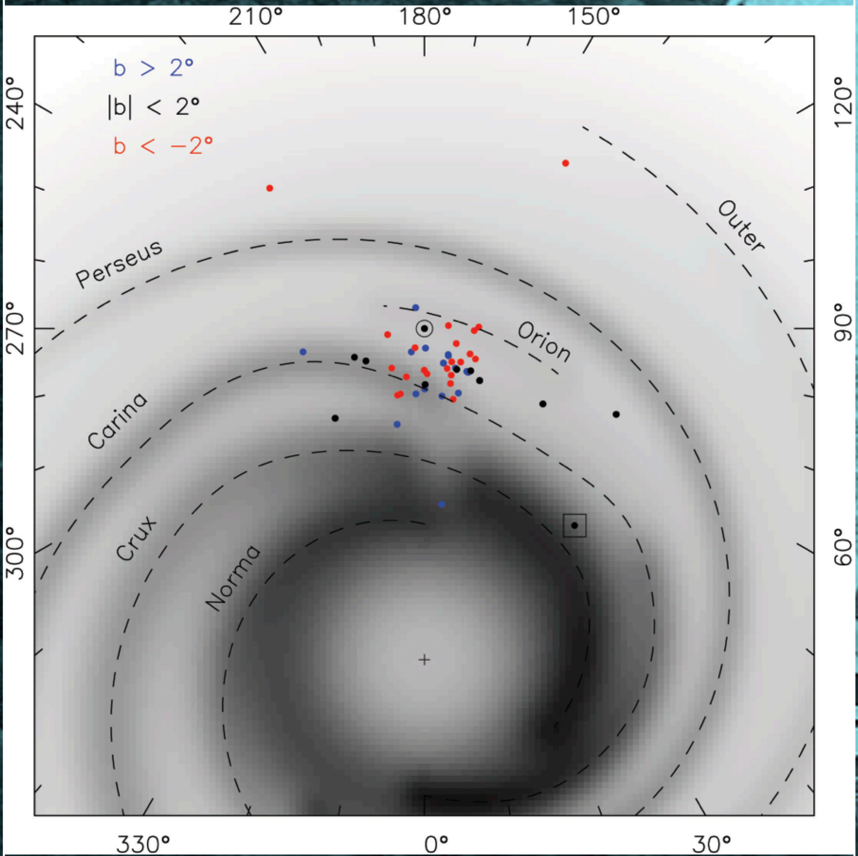
- Single pulses are very weak in the time series.
- FFTs are used to detect periodic signals.
- Harmonic summing increases delectability of low duty cycle pulsars.
- To detect MSPs (~500 Hz) typical sample times are ~64 µs (~7800 Hz).
- Increased observation time (4300 s) and bandwidth (~340 MHz) gives improved sensitivity.
- Typical FFT: 15,000,000 samples -> 60 MB
- Radiometer equation:

$$S_{\min} = \beta \frac{(S/N_{\min})T_{sys}}{G\sqrt{n_p t_{\text{int}} \Delta f}} \sqrt{\frac{W}{P-W}}$$

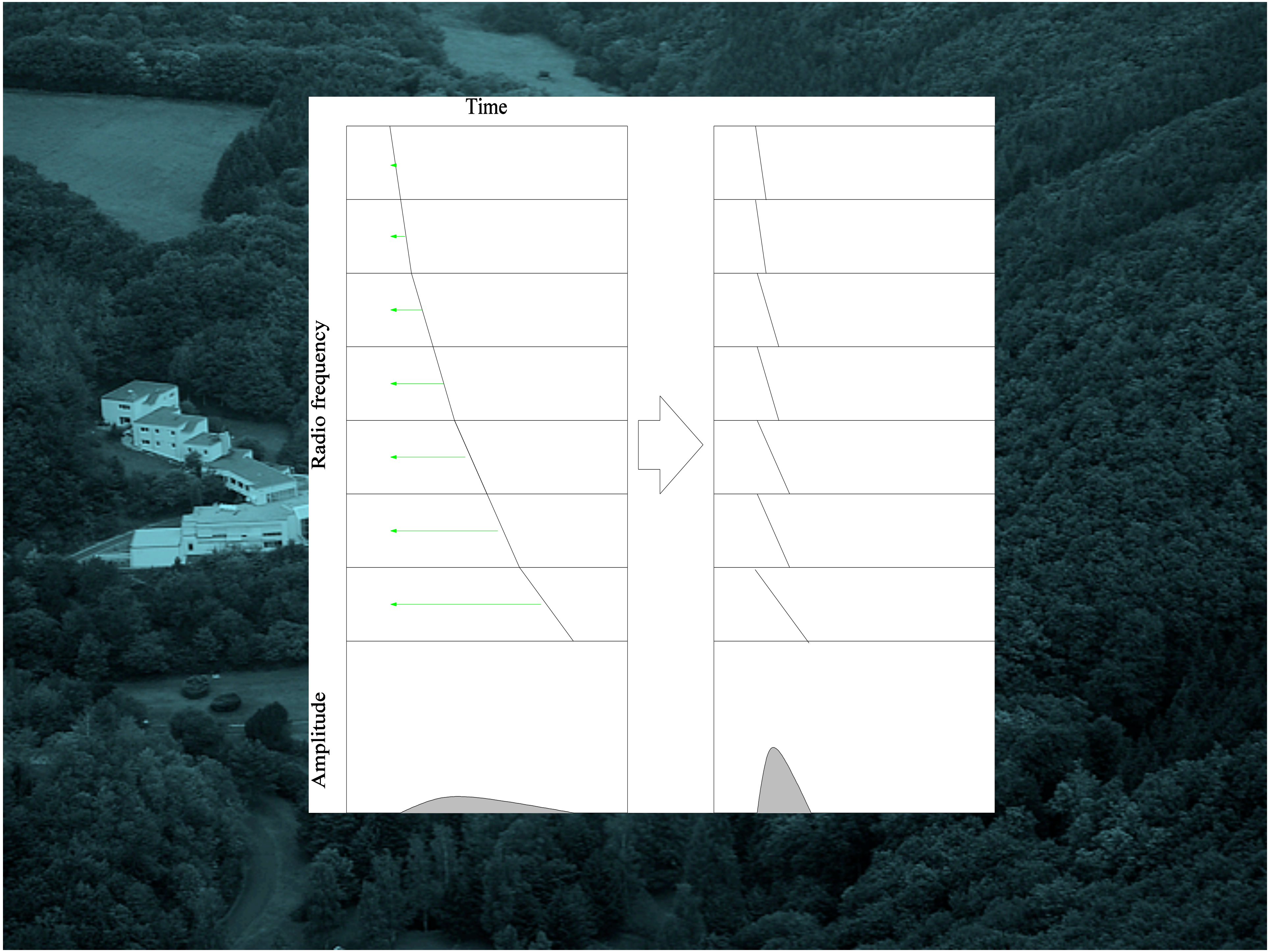
A pipeline

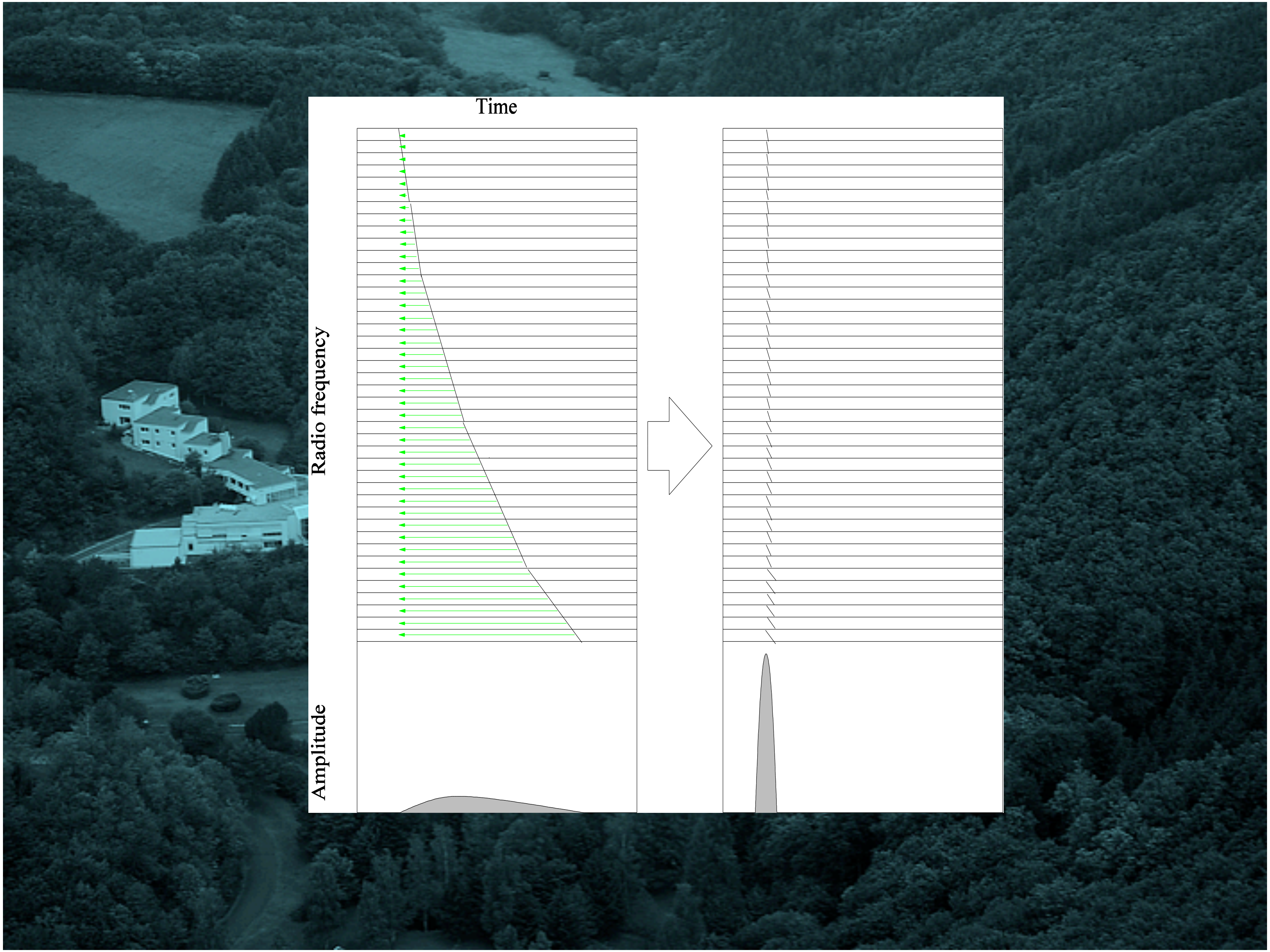
Prepare the data
↓
RFI removal
↓
FFT
↓
Harmonic summing
↓
Sift and fold the candidates

Dispersion



- Radio frequency dependent delay in the arrival time of pulses
- Caused by free electrons along the line of sight
- High frequencies arrive before low frequencies
- Strongest in the plane
- Limits the volume in which we are sensitive to pulsars

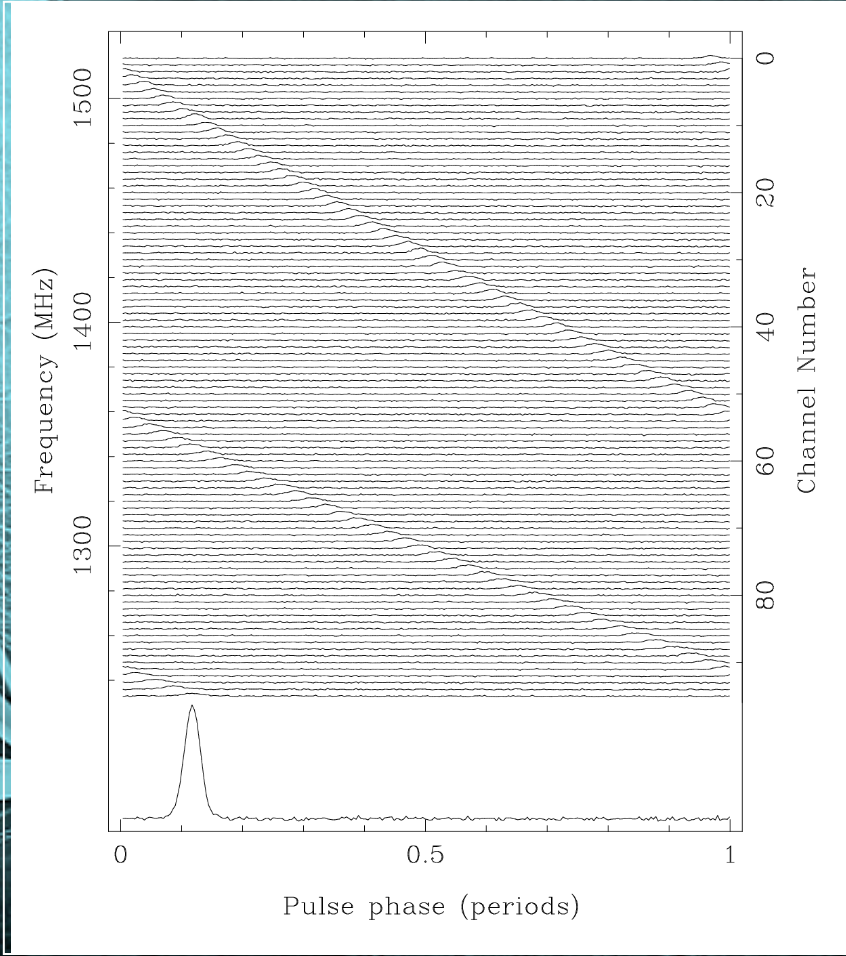

Time
Radio frequency
Amplitude



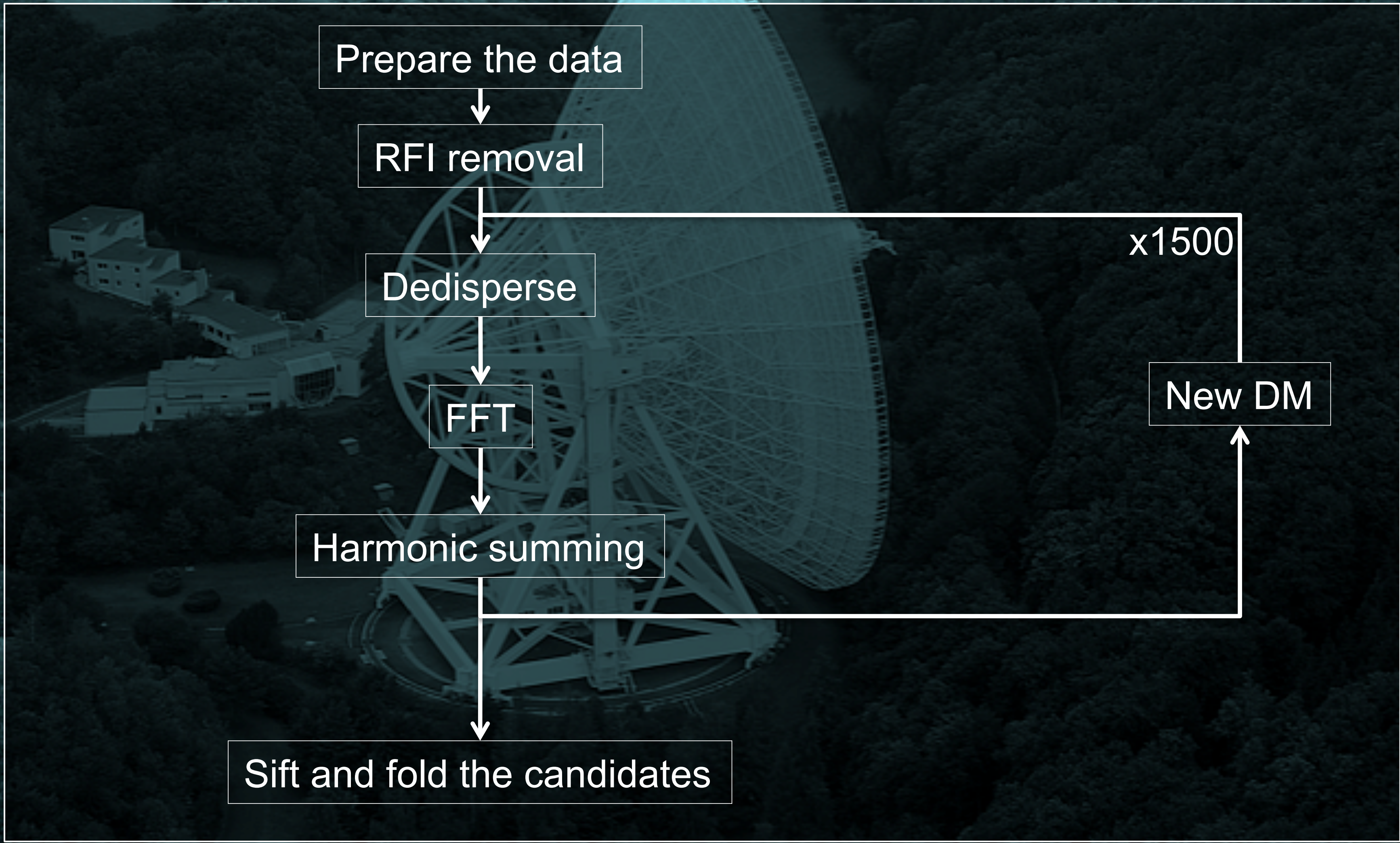
Time
Radio frequency
Amplitude

Dealing with dispersion

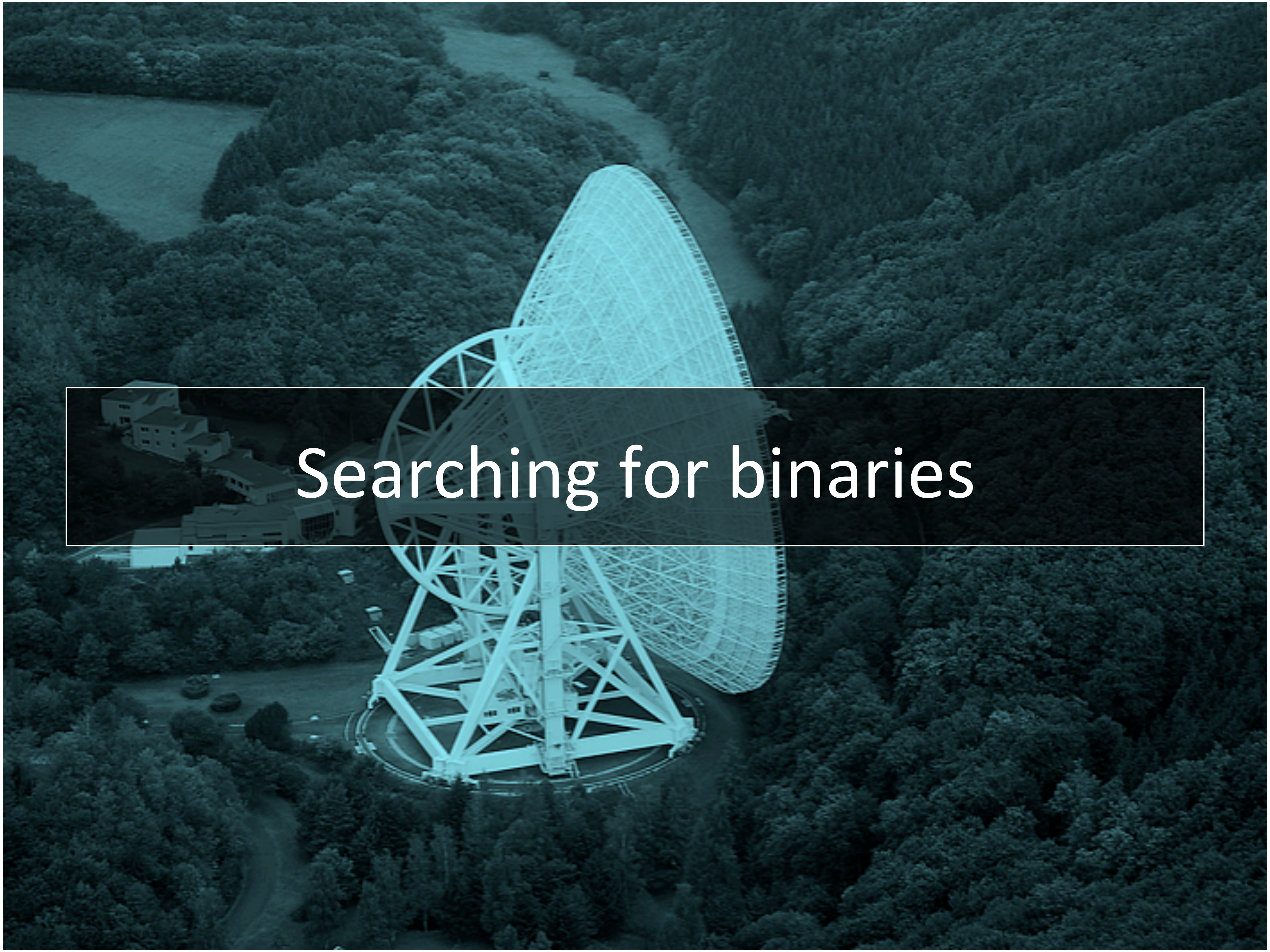
- Dispersion is initially unknown, ~1500 of trials required
- Large data files (18 GB) make dedispersion a slow process
- Sequential access to the files
- All trial time series are created at once
- Subbanding and treeing techniques radically speed up the process
- Downsampling occurs at the diagonal DM (when intra channel smearing = 2*tsamp)



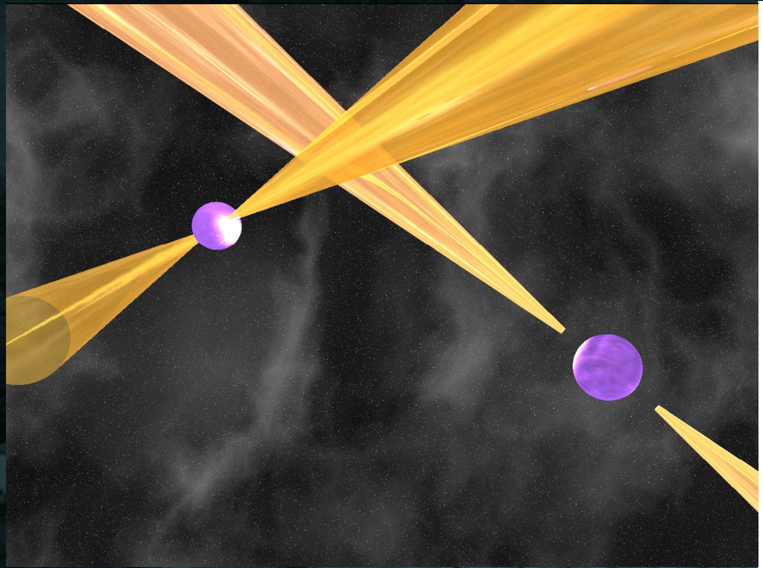
A pipeline

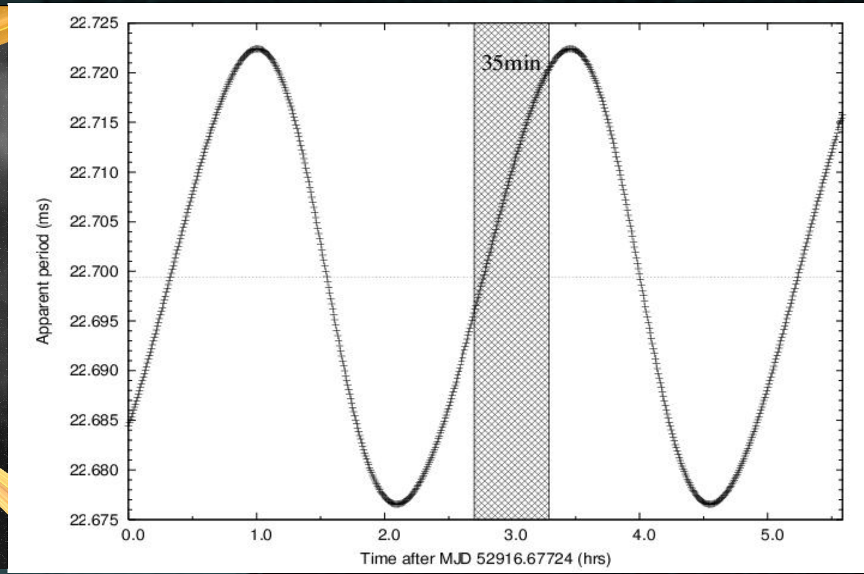

Prepare the data
RFI removal
Dedisperse
FFT
Harmonic summing
Sift and fold the candidates
x1500
New DM

Searching for binaries

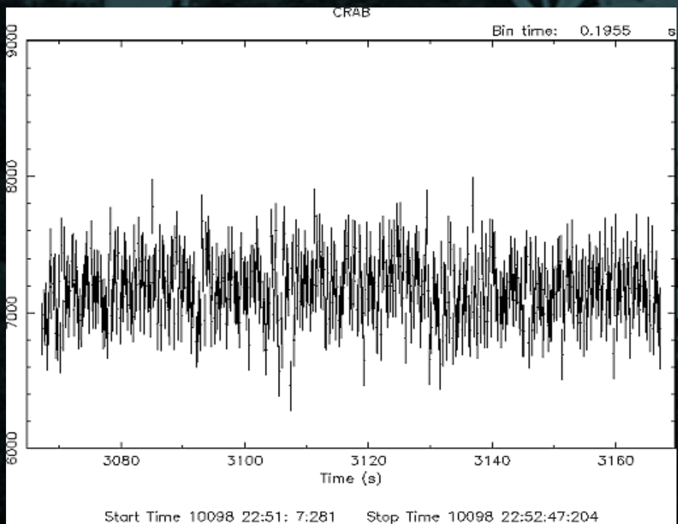


• In binary systems spin period can change over the observation due to the Doppler effect.

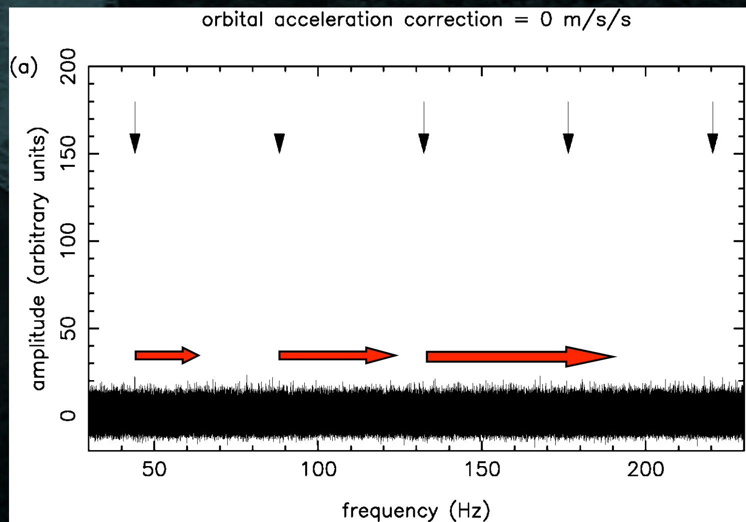




• In a standard pulsar search spin frequency is smeared over a number of spectral bins.



FFT



Acceleration searches

$$\nu(t) = \nu_0 (1 - V(t)/c)$$

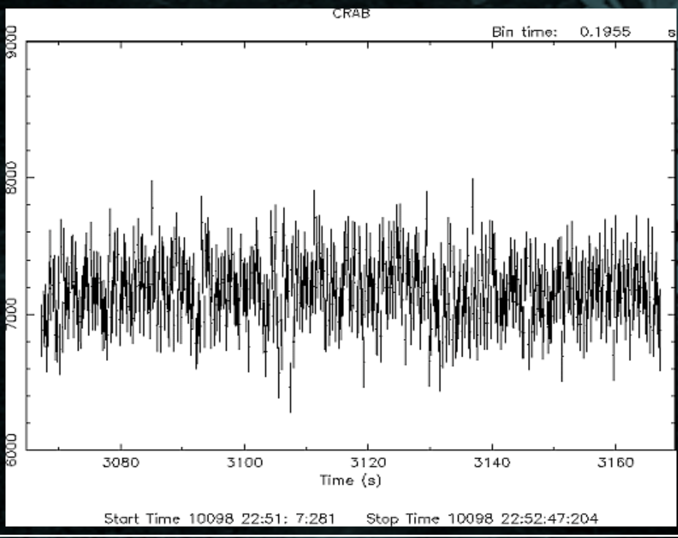
Doppler formula

$$V(t) = \Omega_b \frac{a_p \sin i}{\sqrt{(1-e^2)}} \left[\cos(\omega + A_T) + e \cos \omega\right]$$

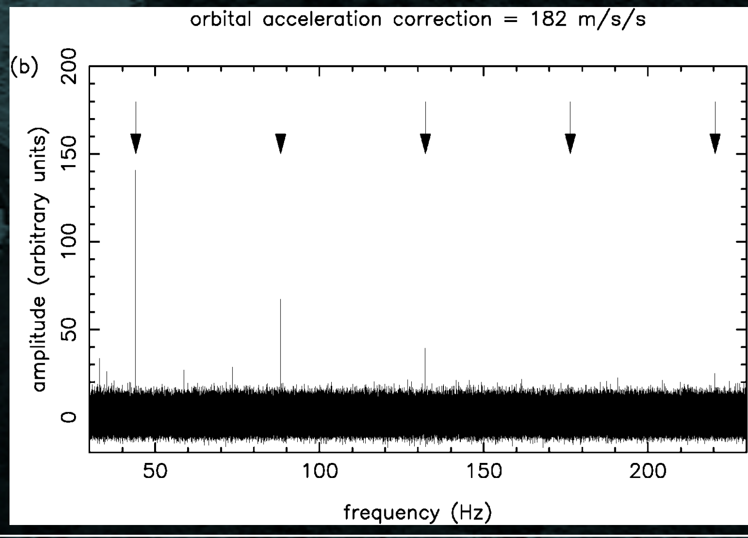
Line of sight velocity from Keplers laws. Search in five dimensional parameter space. Not possible!

$$V(t) = V_0 + a_0 t + j_0 \frac{t^2}{2!} + \ldots$$

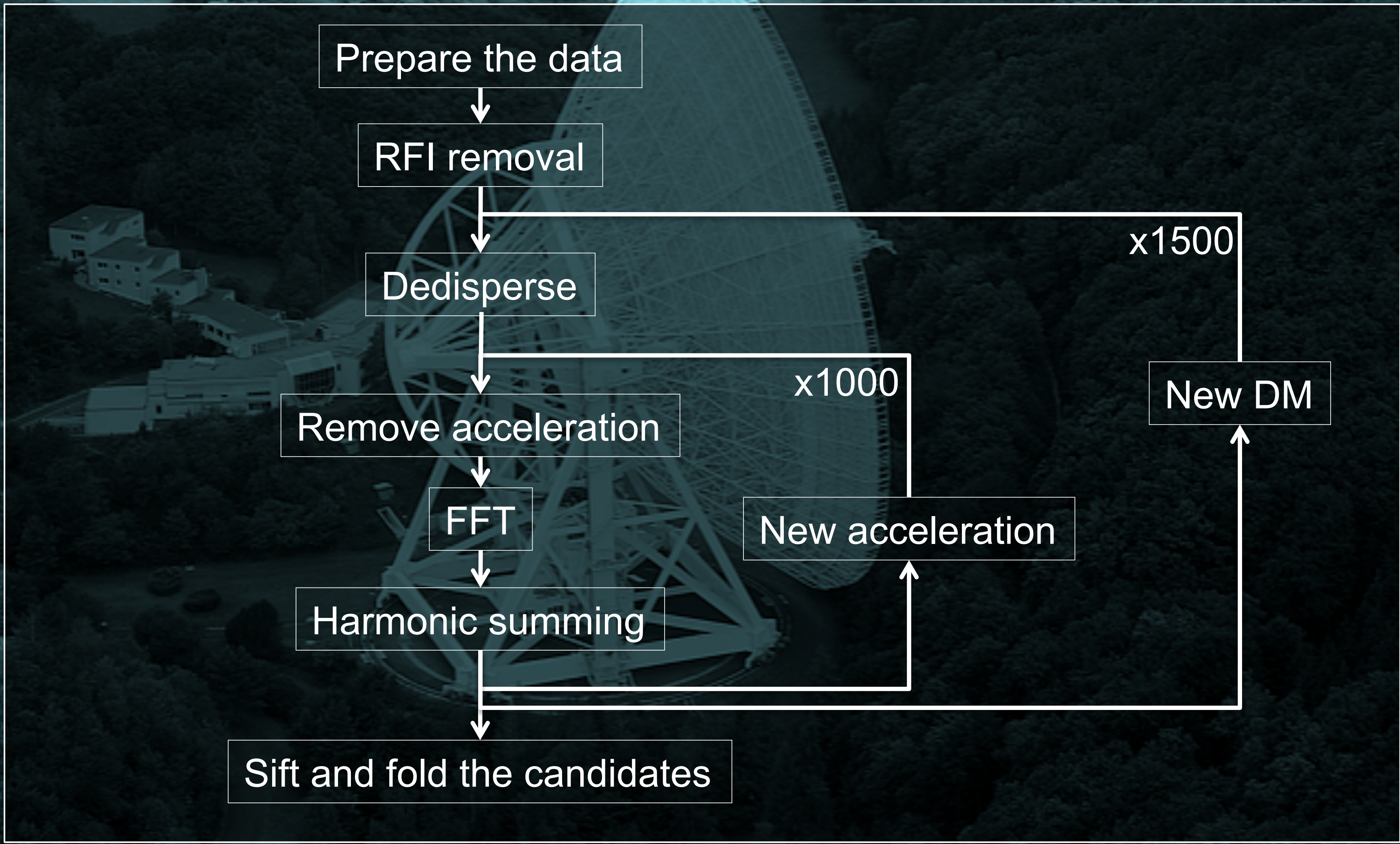
If the observation time is significantly shorter than the orbital period, can make linear approximation i.e. search for constant accelerations.



Resample then FFT



A pipeline



Prepare the data
RFI removal
Dedisperse
x1500
x1000
New DM
Remove acceleration
FFT
New acceleration
Harmonic summing
Sift and fold the candidates

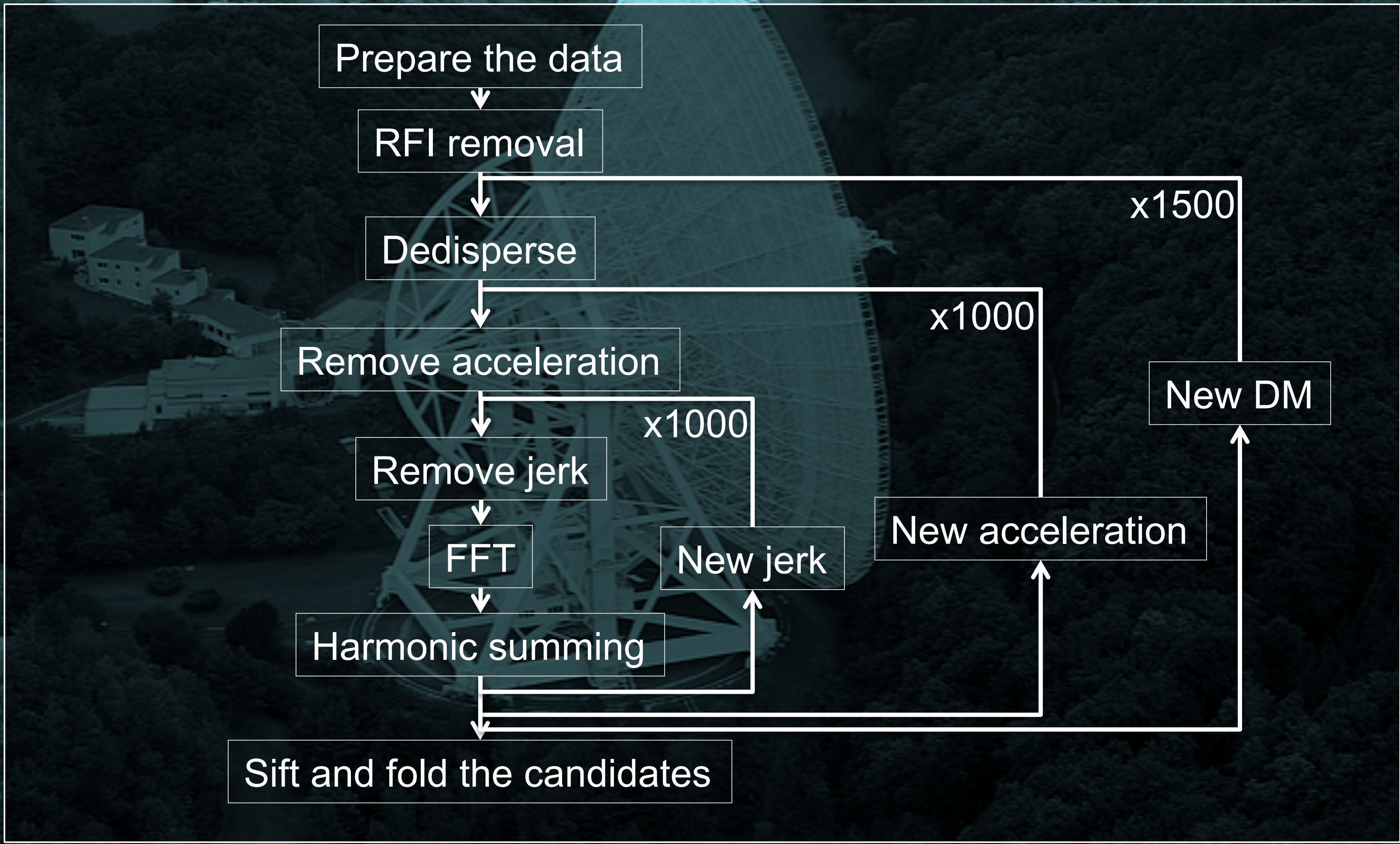
Extending the acceleration search

- The Porb/Tobs ratio can be increased by including the acceleration derivative, jerk

$$V(t) = V_0 + a_0 t + j_0 \frac{t^2}{2!} + \ldots$$

- Increases sensitivity by having a greater amount of data added coherently
- Number of trials is proportional to 1/Tsamp and Tobs^3
- There is some covariance with acceleration that will increase the trial spacing of both

A pipeline


Prepare the data
RFI removal
Dedisperse
Remove acceleration
Remove jerk
FFT
Harmonic summing
Sift and fold the candidates
New jerk
x1000
New acceleration
x1000
New DM
x1500

High time resolution Universe surveys

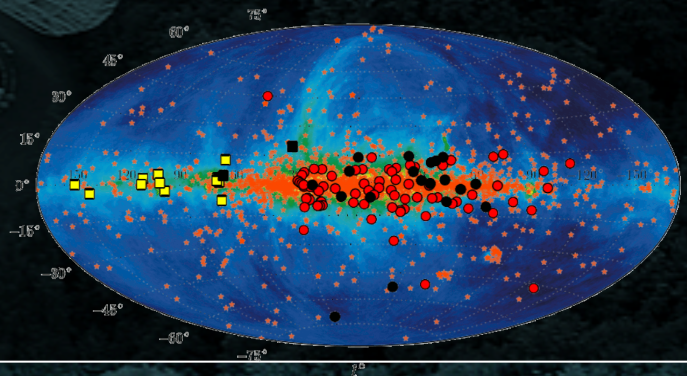
- Recent improvements in hardware that utilize field programmable gate arrays to produce high resolution digital filterbanks.
- Dispersion smearing across individual frequency channels reduced.

- Better sensitivity distant, highly dispersed MSPs!
- HTRU - Collaboration between pulsar groups in UK (Uni. Manchester), Australia (Swinburne/CASS), Germany (MPIfR-Bonn), Italy (Cagliari) to perform an all-sky survey to find highly dispersed MSPs missed in previous surveys!

High time resolution Universe surveys

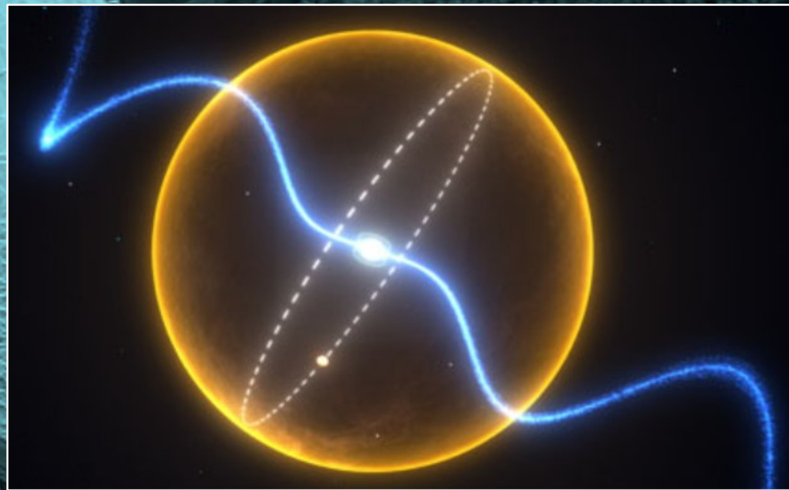
- Total data size: 6 PB
- 15990 beams in survey
- Beam file size: 17 GB (for largest files)
- Each producing 12 GB intermediate files
- And 6 GB results
- Processing (per beam) is **tuned** to:
  - 1584134 FFTs of 60 MB (15 MP) time series
  - Taking ~12k CPU hrs

We are finding new pulsars! 16 from Effelsberg, +100 at Parkes. Diamond planet, magnetar, Cosmological bursts, Eccentric binary MSP in the North….



Discovery highlights

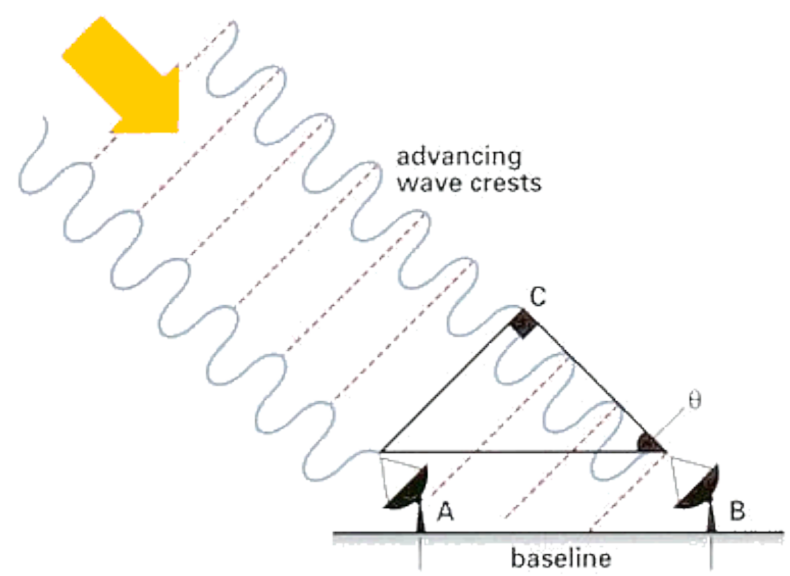
- A pulsar with a diamond planet
- A magnetar in the radio
- 3 extra-galactic bursts
- Several precise timers
- An MSP in an eccentric orbit



Near Future
(this year)

Interferometry

- Telescopes can be added coherently to form an array
- Delays need to be added for geometric distance difference
- Changing the delay moves the beam within the FOV of the telescopes
- Signal sampled at Nyquist frequency and correlated

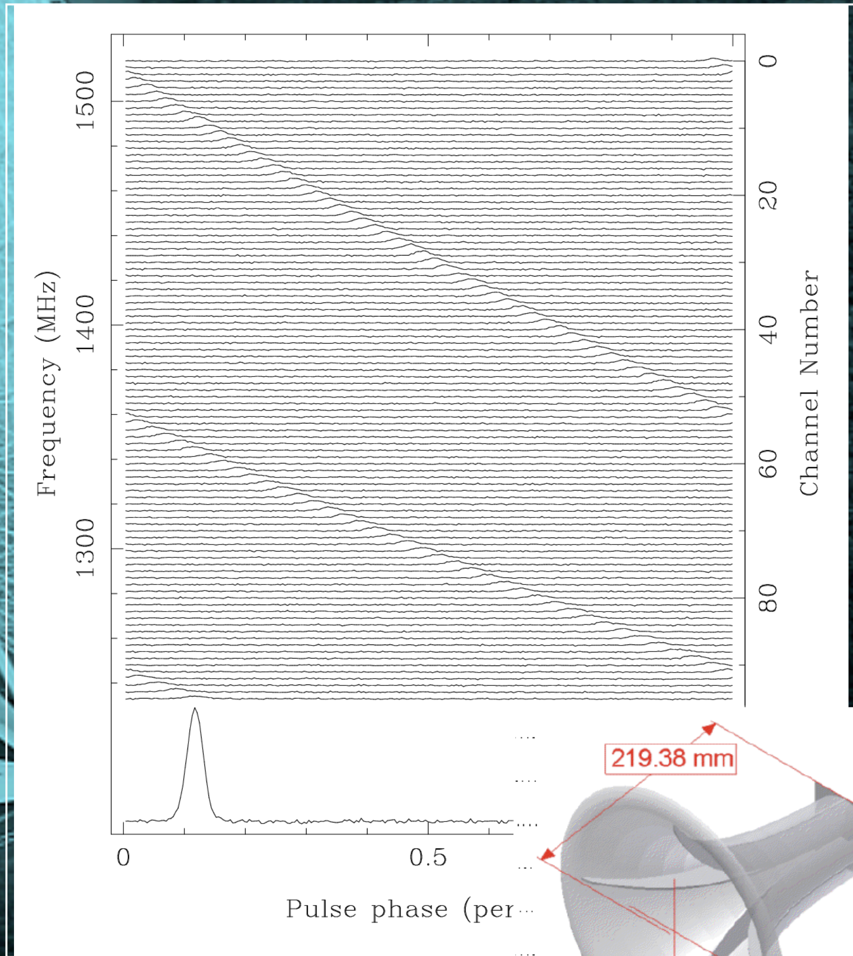


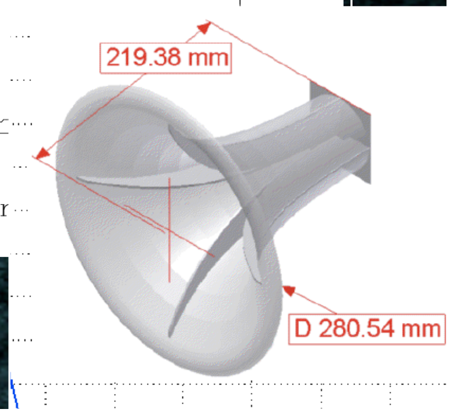
Phased Array Feeds

- Over 100 elements
- The elements are phased-up to produce ~32 beams instantaneously
- Increases survey speed by factor of 4.5
- Also increases data-rate
- Real-time transient detection to look for bursts of radio waves from cosmological sources (massive black-holes colliding).
- Real-time RFI rejection

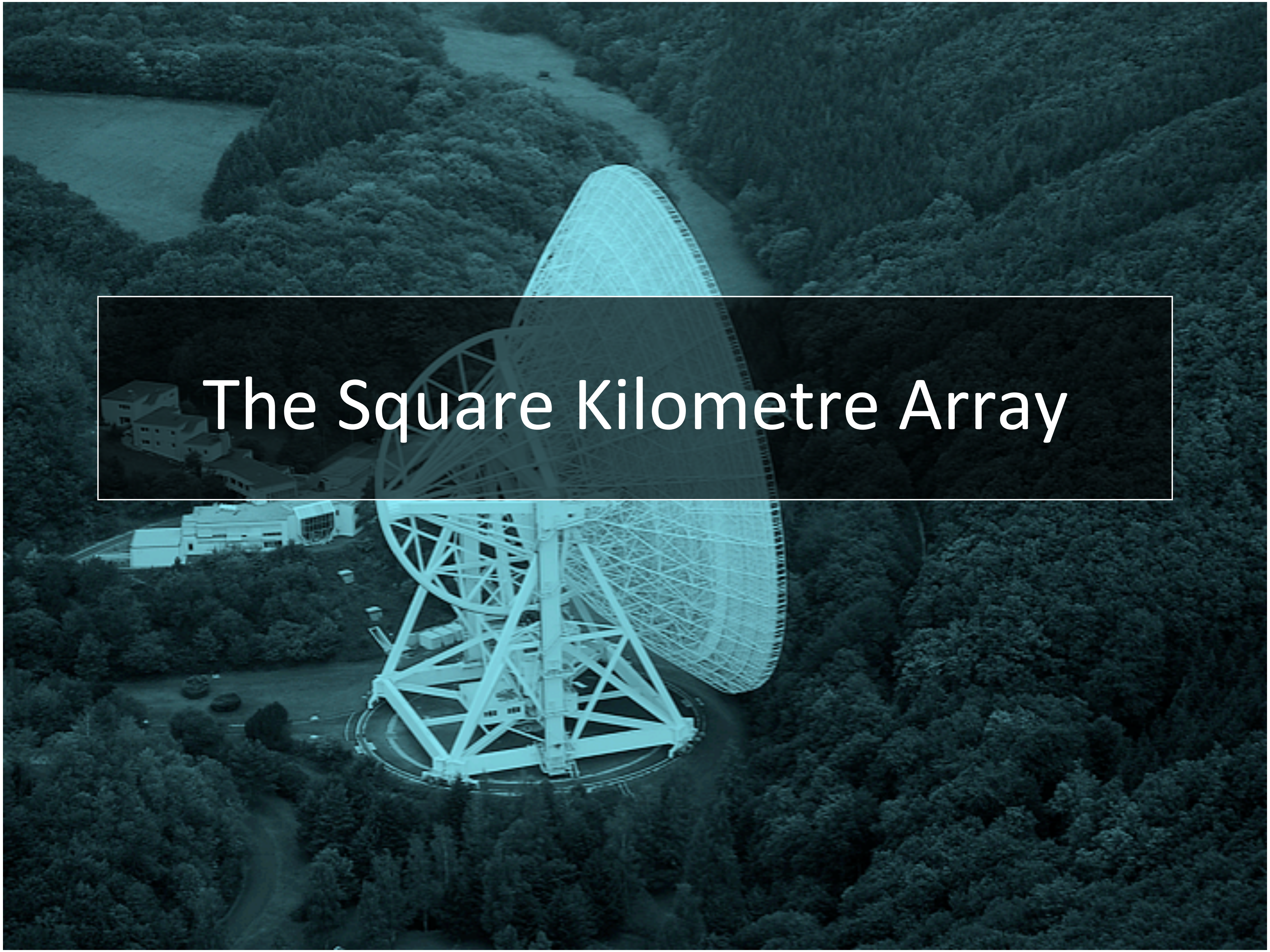
Ultra Broadband Receivers

- Traditional receivers have ~300 MHz BW
- UBB has nearly 2.5GHz BW
- Increases sensitivity to weak pulsars without increasing integration times
- Dispersion becomes a major factor
- Many trials required to counter effects
- Size of intermediate files increases dramatically





The Square Kilometre Array



The Square Kilometre Array

- An large array of small telescopes with 1sq km collecting area
- A mixture of technologies at different frequencies
- Located in Southern Africa and in Western Australia
- Germany is now a member of the SKA
- Located in two cores with smaller stations spread over the continents
- Exascale computing and large IO over great distances

Summary

Aggregate performance numbers

- What is the aggregate data volume for state-of-the art projects in your fields of research?
- 6 PB
- If possible, discuss problem size dependence.
- Bandwidth and number of beams dependent
- How is this expected to change in the future?
- Rate increase by x5, size by a few -- 10
- I/O intensity: What is the fraction of the time spent for I/O vs. total execution time for typical runs?
- Between ~5% and ~40%

I/O pattern analysis

- Could you describe the I/O patterns generated by your application(s), maybe by providing some pseudo-code?
- See flow diagram
- Is sequential or random access dominating?
- Almost always sequential
- Are read or write operations dominating or is it a mix of both?
- Read operations dominate (10:1)
- Is the fraction of small request sizes large?
- No

Parallel I/O

- Parallel I/O intensity: Is sequential or parallel I/O dominating the time spent on I/O operations?
- Usually sequential (depends on number of cores)
- I/O concurrency level: What fraction of the application tasks is involved in concurrent I/O tasks?
- Depends on RAM / Core ratio

Application details

- Which libraries and I/O interfaces have you been using?
- Standard C / C++
- Could you outline I/O challenges faced on current architectures? What are your expectations concerning future requirements of applications from your research area?
- Traditional Beowulf clusters struggle to meet our I/O requirements. This will only get worse with new receivers.
- Does your application provide opportunities to perform asynchronous data management and processing operations in an active storage subsystem? If yes, could you provide a high-level description of the processing operations?
- Not normally for searching but for potentially for candidate ranking