The (obscene) Challenges of Next-Generation Pulsar Surveys

Scott Ransom
NRAO / Univ. of Virginia
Charlottesville, VA

Haslam 408MHz Survey
Summary: The Pulsar Search Problem

- Pulsars are faint – we are sensitivity limited
- Only good way to get sensitivity is collecting area
- Large improvements in $A_{\text{eff}}$ mean arrays
- Coherent beam forming causes a data explosion
  \[ N_{\text{beams}} \sim \left( \frac{D_{\text{dish}}}{D_{\text{core}}} \right)^2 \sim 4000 \]  
  (15m dishes in 1km core)
- To find MSPs, a single beam is $\sim$50-200 MB/s
- Cannot record the data, must process in realtime
- This is a problem with any array (i.e. GMRT, JVLA, LOFAR, MeerKAT), not just SKA
What are the radio properties of pulsars?

• **Broadband continuum**
• **Steep spectra** (0.1-3 GHz)
• **Highly linearly polarized**
• **Point sources:**
  • ISM effects (scintillation and scattering: freq dependent)
  • Spatial resolution irrelevant
• **No confusion or beam dilution**
• **Very weak average flux density** (~mJy or less) – We are sensitivity starved
Only 2-3% of known pulsars are “interesting” for basic/astro physics individually.

In Galaxy, we know:
~160 binary MSPs
~40 isolated MSPs
~40 binary part-recyc
~20 isolated part-recyc

Definitions:
Part-recycled:
P > 20 ms, B < 3x10^{10} G
MSP:
P < 20 ms, B < 10^9 G
Pulsars are an SKA KSP (and a Phase 1 “headline science”)

• **Strong Field Tests of Gravity** (PSR-NS, PSR-BH)
  - Was Einstein right?, Cosmic Censorship Conjecture (i.e. Naked singularities), No-hair theorem

• **Detection of a Stochastic Gravitational Wave Background** (MSP timing)

• **Equation of State of Matter at Supra-Nuclear Density**

• **Lots of other astrophysics**
  - NS masses, ISM structure, Galactic magnetic fields, plasma physics, binary evolution, SNR kicks...

(See Kramer et al. In "Science with the Square Kilometer Array", eds. C. Carilli and S. Rawlings)
The Double Pulsar J0737-3039

- 7(!) post-Keplerian orbital terms / effects
- 0.05% test of general relativity in “strong”—field
- Timing may eventually allow measurement of the neutron star moment of inertia

Kramer et al., 2006, *Science*, 314, 97
Direct Gravitational Wave Detection with a Pulsar Timing Array

- Looking for nHz freq gravitational waves from super massive black hole binaries
- Need good MSPs:
  - Significance scales with the number of MSPs being timed
- Must time 20+ pulsars for 5-10 years at precision of ~100 nanosec!
- North American, European (EPTA), and Australian (PPTA) efforts

For more information, go to nanograv.org
Physics at or beyond Nuclear Density

1.97(4) M⊙ neutron star
Orbit within 1 deg of edge-on
Measured with Shapiro Delay
Strongly constrains “soft” equations of state
And (recent) exotic systems...

- **Double pulsar** J0737-3039 (Lyne et al., *Science*, 2004)
- **Radio magnetar** XTE J1810-197 (Camilo et al., *Nature*, 2006)
- **P-dot changing PSR** B1931+24 (Kramer et al., *Science*, 2006)
- **Eccentric MSP** J1903+0327 (Champion et al., *Science*, 2008)
- **“Missing Link” MSP** J1023+0038 (Archibald et al., *Science*, 2009)
- **2-Msun MSP** J1614-2230 (Demorest et al., *Nature*, 2010)
- **“Diamond Planet”** J1719-1438 (Bailes et al., *Science*, 2012)
- **Massive NS** J0348+0432 (Antoniadis et al., *Science*, 2013)
- **MSP-LMXB switching** M28I (Papitto et al., *Nature*, 2013)
- **MSP in triple system** J0337+1715 (Ransom et al., *Nature*, 2014)
- **Future?**: MSP-MSP, PSR-BH, sub-MSP, ultra-massive, ….
A PSR Renaissance from DSP & HPC

- **Until very recently**: pulsar observations were severely limited by our instrumentation, not by our telescopes
- **Now**: digital signal processing and computing are finally allowing us to fully use our telescopes

Plus: Arecibo, GBT, Parkes, Jodrell Bank, Nancay, Effelsberg, GMRT, Westerbork, Urumqi...
New All-Sky Pulsar Surveys

- All major radio telescopes are conducting all-sky pulsar surveys
- These generate lots of data:
  - 1000s of hrs, 1000s of channels, ~15 kHz sampling: gives more than a Petabyte
- Requires huge amounts of high performance computing
  - Many times real-time to process
  - Creates millions of candidates
New Millisecond Pulsars

Numbers have:
more than quadrupled in last 10 yrs
doubled in last ~4 years
What does the future hold?

- Single dishes (e.g. GBT, Arecibo, and FAST in China) with large numbers of simultaneous beams (40-100)

- Searches with arrays (e.g. LOFAR, MeerKAT, EVLA, SKA) dramatically increase the data rate

  - JVLA survey $\sim 10$ GB/s
  - SKA1 survey $\sim 1$ TB/s!
SKA Sensitivity Summary

- SKA1-mid (~250 dishes x 15m): timing (search)
  - 1600 m^2k^{-1} (~900 m^2k^{-1} within 1 km diam)

- SKA2-mid (~2500(?) x 15m):
  - 16000(?) m^2k^{-1} (~2000 m^2k^{-1} within 1 km diam)

FAST: 1800 m^2k^{-1} > Factor of ~2

Areccibo: 1000 m^2k^{-1} > Factor of ~4

GBT / MeerKAT: 250 m^2k^{-1} > Factor of ~3

Parkes: 80 m^2k^{-1}
SKA Phase 1 Pulsar Searching

- See Smits et al. 2009
- ~20,000 each of potentially visible normal pulsars, RRATs, and MSPs
- SKA1 has the potential to find a large fraction (~50%?) of these pulsars
- Survey speed for Phase 1 with 15m dishes and fully sampled primary beam is: 54x Parkes MB, 180x GBT, 70x Arecibo, 30x FAST

Simulation by J. Cordes
Problem in a nutshell... data rates

The future of radio astronomy is arrays...
Pulsar searching basics:

- Normal pulsars come “free” if you can find MSPs
- MSPs are in binaries and distributed isotropically
- Slow, relativistic binaries (i.e. NS-NS, NS-BH) should be close to the Galactic Plane, where ISM effects are stronger
- Cannot trade sensitivity for integration time (need single, coherent integrations to find binaries)
- Searches have to be done per sky pixel (use fewer pixels!)
- Four main parts to the search process:
  - (Beam forming,) De-dispersion, Acceleration searching, Folding

\[
\text{Sensitivity} \propto \frac{\text{Collecting Area}}{T_{\text{sys}} \sqrt{t_{\text{int}}} \text{ Bandwidth}}
\]

\[
\text{Computations} \propto f_{\text{spin}}^3 t_{\text{int}}^3
\]
Coherent Beam Forming

Spatial resolution is the main disadvantage to searching with arrays over a single dish: fill primary beam with synthesized or “tied array” beams

$$N_{\text{beams}} \sim \left(\frac{D_{\text{core}}}{D_{\text{dish}}}\right)^2 \sim 4000$$

(MeerKAT / SKA using only ~1km diam core)

$$N_{\text{ops}} \sim 1$$ Pops for all beams

Output \(\sim N_{\text{beams}} \times 300 \text{ MB/s} \sim 10 \text{ Tbits/s}$$

Every beam is independent with ~10-15K chans @ 20kHz
Dispersion

Lower frequency radio waves are delayed with respect to higher frequency radio waves by the ionized interstellar medium.

\[ \Delta t \propto DM \nu^{-2} \]

(DM = Dispersion Measure)

- Need \( \sim 10^4 \) frequency channels
- DM for undiscovered pulsar is unknown
- Must search over \( \sim \text{few} \times 10^4 \) trial DMs!
- This multiplies data rate by factor of few
- \( \sim 0.1 \) Pops for SKA1
- De-dispersion is very I/O intensive

Barsdell et al 2012
“Acceleration Searches”

- Most interesting PSRs are in binaries: Doppler effect
- If the orbital period $\gg$ observation time, then the acceleration is approx constant during the observation
- A “chirp” with small $\Delta f/f$ (phase changes quadratically)
- Correction done in time or freq domain

$P_{orb} = 10 T_{obs}$
Acceleration Search Logistics

- Fourier-based searches use **harmonic summing** (increases sensitivity by 2-5x over continuum imaging)
- CPU + memory intensive, but embarrassingly parallel
- Datarate multiplied ~100x: **Need real-time searches!**
- Dominates processing: ~10 PFLOPS for SKA1

![Image](image.png)
- 3ms binary MSP in 1.8hr orbit, with 25-min observation
- First 3 harmonics in “F/Fdot plane”
Candidate “Folding”

- Surveys generate millions of candidate pulsar signals with associated metadata. Candidate “sifting” is a big problem...

- For best candidates, “fold” raw data modulo candidate spin period at the detected DM. Extremely I/O intensive
Single Pulse Searches

- Many PSRs (or RRATs) have highly variable pulse amplitudes
- Possible new transients... “Lorimer Bursts”, FRBs
Possible SKA1 Searches

- **Very Early Science: Targeted Searches**
  - SNRs, gamma-ray sources, PWNe (0.5-3 GHz)
  - Globular Clusters (0.5-3 GHz)

- **Early Science: Limited Surveys**
  - Local Galaxies (LMC, SMC; 1.5-3 GHz)
  - Gal Ctr Region (1.5-3 GHz)
  - Low Latitude Galactic Plane (<|1deg|, ~1.5-3 GHz)

- **Full Science: Large Area Surveys**
  - Galactic Plane (<|5deg|, 0.8-1.6 GHz, ~10-30 min)
  - All-sky survey (0.45-0.9 GHz, ~10 min)
  - These are real-time: need to buffer ~0.5 PB!
After discovery, they all need timed!

- Timing tells us if the pulsars are interesting
- Follow up timing will take more telescope time than the surveys (although spread over more time)
- Timing can use *all* of SKA instead of the core, *or*, the SKA can be sub-arrayed (this is extremely inefficient, though)...  
- Large FoV will allow simultaneous timing of multiple PSRs (large number of PSRs will *require* it)
  - Not the case for high-precision MSPs – too sparse
- Imaging capabilities should provide excellent starting positions for the pulsars (requiring fewer timing obs for “boring” pulsars)
- Triage of pulsars will be crucial
So what's happening now?

- GPU-ized de-dispersion (i.e. Barsdell, Magro...)
- GPU-ized acceleration searches
  - PRESTO's `accelsearch` by Jintao Luo
  - Ewan Barr's PEASOUP
- Experiments with de-dispersion and acceleration searching (with harmonic summing) on FPGAs
- Real-time RFI detection and removal
- Many efforts in machine learning for candidate ID
- Need high-speed, parallel folding algorithms for multiple candidates (one-pass through raw data)
- Completely new search algorithms?
Summary

PSR searching with arrays is hard, primarily because of synthesized beams (want most compact core $D_{\text{core}}$).

Expand capabilities over time as computing improves by using more of the primary beam and/or increasing $D_{\text{core}}$.

Only need more sensitivity than Parkes in the South to start with targeted surveys:

- Core dishes of MeerKAT will give ~GBT sens.
- FAST will see some of the South with 2xArecibo sens.

SKA -1 will find ~5000-10000 PSRs. All will need to be timed. Some will be timed long term (and with full sensitivity) with SKA-2.

No ability to re-do analysis: Re-do = Re-observe.

All-sky SKA1 surveys will be real-time and very difficult.
JVLA Pulsar Survey?

Would use fast-dump correlation mode. At L-band, resulting VLA survey speed is \(~10\)x single-pixel GBT. Sounds good, except for...

Problem #1:

\[
\text{Data Rate (B/s)} = N_{\text{pol}} \times N_{\text{chan}} \times \frac{N_{\text{dish}}(N_{\text{dish}} - 1)}{2} \times \frac{\text{Bytes}}{\text{vis}} \times \frac{1}{dt} \\
= 2 \times 1000 \times \frac{351}{1} \times 1 \times 10000 \\
= \textbf{7 GB/s!} \quad \text{(compare to \(~10\)MB/s for 1 beam of the GBT)}
\]

and Problem #2:

Computation Rate \(\propto T_{\text{int}}^3 \log_2(T_{\text{int}}) \times D_{\text{array}}^2\)

D-configuration needs \(~100\)x more computation than a similar single-dish survey. E-config necessary to make this reasonable!