The Square Kilometre Array
Massive Data Challenges at the Frontiers of Astronomy, Physics, & Astrobiology

Joseph Lazio
Project Scientist, SKA Program Development Office
&
Jet Propulsion Laboratory, California Institute of Technology
Square Kilometre Array

The Global Radio Wavelength Observatory

- Originally: “Hydrogen telescope”
  Detect H I 21-cm emission from Milky Way-like galaxy at z ~ 1

- SKA science much broader
  ⇒ Multi-wavelength, multi-messenger

- On-going technical development
  ➢ Cyber-infrastructure and “big data”

- International involvement
SKA Key Science

International working group
- Strong-field Tests of Gravity with Pulsars and Black Holes
- Galaxy Evolution, Cosmology, & Dark Energy
- Emerging from the Dark Ages and the Epoch of Reionization
- The Cradle of Life & Astrobiology
- The Origin and Evolution of Cosmic Magnetism

With design philosophy of *Exploration of the Unknown*

New Worlds, New Horizons in Astronomy & Astrophysics Decadal Survey conducted by National Research Council

• A Future Radio-Millimeter-Submillimeter System
  ... A second principle is provision for the long term future through a staged program leading towards major participation in all three components of the international Square Kilometer Array, which has enormous scientific potential and enthusiastic support around the globe.

• Recommendations for New Ground-Based Activities—Large Projects
  Priority 2 (Large, Ground). Mid-Scale Innovations Program [2 of 8 possible initiatives, alphabetical order]
  • Hydrogen Epoch of Reionization Array (HERA)
    Could evolve into low frequency component of SKA
  • North American Nanohertz Observatory for Gravitational Waves (NANOGrav)
    Could evolve into intermediate frequency component of SKA

Exploring the Universe with the world’s largest radio telescope
## NWNH Science Program and the SKA

<table>
<thead>
<tr>
<th>Category</th>
<th>Question</th>
<th>Project Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discovery</strong></td>
<td>Gravitational wave astronomy</td>
<td>“Strong-field Probes of Gravity with Pulsars and Black Holes”</td>
</tr>
<tr>
<td></td>
<td>Epoch of Reionization</td>
<td>“Emerging from the Dark Ages and the Epoch of Reionization”</td>
</tr>
<tr>
<td><strong>Origins</strong></td>
<td><em>What were the first objects to light up the Universe and when did they do it?</em></td>
<td>“Emerging from the Dark Ages and Epoch of Reionization”</td>
</tr>
<tr>
<td></td>
<td>How do circumstellar disks evolve and form planetary systems?</td>
<td>“The Cradle of Life and Astrobiology”</td>
</tr>
<tr>
<td><strong>Understanding the Cosmic Order</strong></td>
<td><em>How do baryons cycle in and out of galaxies and what do they do while they are there?</em></td>
<td>“Galaxy Evolution, Cosmology, and Dark Energy”</td>
</tr>
<tr>
<td><strong>Frontiers of Knowledge</strong></td>
<td><em>What controls the masses, spins and radii of compact stellar remnants?</em></td>
<td>“Strong-field Probes of Gravity with Pulsars and Black Holes”</td>
</tr>
</tbody>
</table>

*Exploring the Universe with the world’s largest radio telescope*
20th Century: We discovered our place in the Universe.
21st Century: We understand the Universe we inhabit.

Cosmology & Fundamental Physics

- Gravity
  - Can we observe strong gravity in action?
  - What is dark matter and dark energy? (dark energy and BAOs with H I galaxies)
- Magnetism
- Strong force
  Nuclear equation of state

Galaxies Across Cosmic Time, The Galactic Neighborhood, Stellar and Planetary Formation

- Galaxies and the Universe
  - How did the Universe emerge from its Dark Ages?
  - How did the structure of the cosmic web evolve?
  - Where are most of the metals throughout cosmic time?
  - How were galaxies assembled?
- Stars, Planets, and Life
  - How do planetary systems form and evolve?
  - What is the life-cycle of the interstellar medium and stars? (biomolecules)
  - Is there evidence for life on exoplanets? (SETI)
Evolution of the Universe

H I brightness temperature signal (w.r.t. CMB)

- X-ray heating, EoR
- First Stars
- Dark Ages
- SKA

(Pritchard & Loeb 2008)
Evolution of the Universe Epoch of Reionization

EDGES
(Bowman et al. 2008)

PAPER

SKA objective: Image the IGM transition in the H I (21-cm) line


Furlanetto et al.; Gnedin
Galaxy Assembly
Stars *and* Gas

- Gas content and dynamics becoming critical part of simulations.
  - N-body simulations themselves can lead to data challenges!
- Astronomy is an *observational* science.
- Need *observations* of gas content —over cosmic time—to understand galaxy formation!

---

Keres et al.

Eris simulation
(Guedes et al.)
NGC 6946 (T. Oosterloo)
Galaxy Assembly
The Role of Mergers

- Mergers are recognized as important aspect of galaxy evolution and formation.
- Gas can be sensitive tracer of interactions, long after original event took place.
  E.g., Holwerda et al. with THINGS.

(Moster et al. arXiv:1104.0246)
Astrobiology at Long Wavelengths

\( \lambda > 1 \text{ cm} \)

- Not affected by dust
- Complex molecules have transitions at longer wavelengths
- “Waterhole” (1.4–1.7 GHz)
- Magnetically-generated emissions from extrasolar planets

Complex organic molecules detected at radio wavelengths

EVL A 6 cm observations of protoplanetary disks; PEBBLES on e-MERLIN coming soon
**21st Century Astrophysics**

20th Century: We discovered our place in the Universe.

21st Century: We understand the Universe we inhabit.

---

**Cosmology & Fundamental Physics**

- **Gravity**
  - Can we observe strong gravity in action?
  - What is dark matter and dark energy? (dark energy and BAOs with H I galaxies)

- **Magnetism**

- **Strong force**
  Nuclear equation of state

---

**Galaxies Across Cosmic Time, The Galactic Neighborhood, Stellar and Planetary Formation**

- **Galaxies and the Universe**
  - How did the Universe emerge from its Dark Ages?
  - How did the structure of the cosmic web evolve?
  - Where are most of the metals throughout cosmic time?
  - How were galaxies assembled?

- **Stars, Planets, and Life**
  - How do planetary systems form and evolve?
  - What is the life-cycle of the interstellar medium and stars? (biomolecules)
  - Is there evidence for life on exoplanets? (SETI)

---

*Exploring the Universe with the world’s largest radio telescope*
Did Einstein Have the Last Word on Gravity?

\[ G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}/c^4 \]

Relativistic binaries probe
1. Equivalence principle
2. Strong-field tests of gravity
- Neutron star-neutron star and neutron star-white dwarf binaries known
- Black hole-neutron star binaries?

Kramer et al.

PSR J0737-3039
SKA: Gravitational Wave Detector

Test masses on lever arm
- **Pulsar Timing Array** = freely-falling **millisecond** pulsars
- LIGO = suspended mirrors
- LISA = freely-falling masses in spacecraft

Pulsar timing arrays starting to provide results from ensemble of pulsars
- EPTA (van Haasteren et al., *above*)
- PPTA (Yardley et al.)
- NANOGrav (*Demorest et al.*)

Exploring the Universe with the world's largest radio telescope
Origin & Evolution of Cosmic Magnetic

- Magnetic fields are fundamental, but poorly constrained
  - Affects galaxy, cluster evolution?
  - Affects propagation of cosmic rays in ISM and IGM
- All-sky rotation measure surveys provide $B$ fields along lines of sight
- Continuum in I, Q, and U!
Magnetic Fields and Cosmic Rays

- Are ultra-high energy cosmic rays (UHECRs) produced in nearby AGN?
- Galactic magnetic field influences cosmic ray propagation

➤ Different models of Galactic field imply different arrival directions
  - Axi-symmetric vs. bi-symmetric?
  - Field directions above and below the Galactic plane
  - Effect of turbulence?
  - ...?
Cosmology and Gravity

\[ G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}/c^4 \]

Origin and Fate of the Universe

- Era of “precision cosmology”
  … or precision ignorance
- Need to sample a substantial volume of the Universe
- Volume \( \sim D^2 \Delta D \Omega \)
  - \( D \) – distance; \( \Omega \) – solid angle
  - Surveying to larger \( D \) is difficult ➔ need larger telescopes
    “square kilometre” of SKA
  - Surveying larger sky areas \( \Omega \)
    “just” requires more observing time

Composition of the Universe

Exploring the Universe with the world’s largest radio telescope
Cosmology and Sky Surveys

- Image the sky, locating galaxies
  Analysis of locations compared with cosmological models to constrain parameters

- Two broad classes of surveys
  - Continuum: e.g., NVSS, FIRST, ASKAP/EMU, WSRT/APERTIF/WODAN
  - Spectroscopic: SDSS, Arecibo ALFALFA, ASKAP/WALLABY, SKA H i survey

  Spectroscopic surveys locate in 3-D space! very powerful

- Ultimate goal: spectroscopic survey of 1 billion galaxies

SDSS Simulated Sky
Detection of weak lensing (E modes) from FIRST (Chang et al.)

Radio observations should have fewer (different) systematics.
21\textsuperscript{st} Century Astrophysics

<table>
<thead>
<tr>
<th>Fundamental Forces and Particles</th>
<th>Origins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gravity</td>
<td>Galaxies and the Universe</td>
</tr>
<tr>
<td>Magnetism</td>
<td>Stars, Planets, and Life</td>
</tr>
<tr>
<td>Strong force</td>
<td></td>
</tr>
</tbody>
</table>

"The Universe is patiently waiting for our wits to grow sharper."

- Photon frequency / wavelength / energy
- Time
- Polarization
- Sensitivity
- Field of View
- Angular Resolution

Exploring the Universe with the world’s largest radio telescope
The Dynamic Radio Sky

- Neutron stars
  - Magnetars
  - Giant pulses
  - Short GRBs?
- Microquasars
- Tidal Disruption Events

- GRBs (γ-ray loud; γ-ray quiet?)
  - Afterglows
  - Prompt emission?
- Sub-stellar objects
  - Brown dwarfs
  - Extrasolar planets?
- Scintillation
- GW counterparts
- UHECRs
- ETI
- Exploding black holes
- ???

Rotating Radio Transients (RRATS)

Pulsating Brown Dwarfs
Imaging with Arrays

Fourier transform \((u-v)\) plane

Image plane

\[ N_{\text{data}} \sim N_{\text{antenna}}^2 N_{\text{frequency}} N_{\text{time}} \]

Exploring the Universe with the world’s largest radio telescope
Imaging Surveys

Requirements
- Many antennas in order to provide sensitivity and image quality.
  - $N_{\text{antenna}}$ must be large.
- Spectral resolution because of wide-field effects, line emission from galaxies, or both.
  - $N_{\text{frequency}}$ must be large.
- Long integrations in order to obtain adequate signal-to-noise ratio.
  - $N_{\text{time}}$, e.g., 1 hr at 1 s sampling?

$$N_{\text{data}} \sim N_{\text{antenna}}^2 N_{\text{frequency}} N_{\text{beams}} N_{\text{time}}$$

<table>
<thead>
<tr>
<th></th>
<th>ASKAP</th>
<th>SKA Phase 1</th>
<th>SKA Phase 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{\text{antenna}}$</td>
<td>30</td>
<td>$\sim 250$</td>
<td>$\sim 1000$</td>
</tr>
<tr>
<td>$N_{\text{beams}}$</td>
<td>30</td>
<td>1</td>
<td>$1?$</td>
</tr>
<tr>
<td>$N_{\text{frequency}}$</td>
<td>$\sim 16k$</td>
<td>$\sim 16k?$</td>
<td>$\sim 16k?$</td>
</tr>
</tbody>
</table>
## Imaging Surveys II

<table>
<thead>
<tr>
<th></th>
<th>ASKAP</th>
<th>SKA Phase 1</th>
<th>SKA Phase 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{\text{antenna}}$</td>
<td>30</td>
<td>250</td>
<td>1000</td>
</tr>
<tr>
<td>$N_{\text{beams}}$</td>
<td>30</td>
<td>1</td>
<td>1?</td>
</tr>
<tr>
<td>$N_{\text{frequency}}$</td>
<td>$\sim 16k$</td>
<td>$\sim 16k?$</td>
<td>$\sim 16k?$</td>
</tr>
<tr>
<td>$N_{\text{time}}$</td>
<td>$\sim 4k$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_{\text{data}}$</td>
<td>$1.8 \times 10^{12}$</td>
<td>$4 \times 10^{12}$</td>
<td>$65 \times 10^{12}$</td>
</tr>
<tr>
<td>$N_{\text{OPS}}$</td>
<td>$18 \times 10^{15}$</td>
<td>$40 \times 10^{15}$</td>
<td>$650 \times 10^{15}$</td>
</tr>
</tbody>
</table>

- Imaging is more than “just” an FFT.
  Gridding, deconvolution, wide-field corrections, self-calibration, …
- Community estimates are $10^4$ to $10^5$ ops per visibility datum(!).
- Leads to significant power challenges
  - Related to moving data on/off chips
  - Careful design can yield significant savings, e.g., D’Addario (SKA Memo 130)
Fundamental Physics with Radio Pulsars

Arrival times of pulses from radio pulsars can be measured with phenomenal accuracy

- Better than 100 ns precision in best cases
- Enables high precision tests of fundamental physics
  - Theories of gravity, gravitational waves, nuclear equation of state
  - 1993 Nobel Prize in Physics

- **Problem**: Not all pulsars are equal!
- Good “timers” < 10% of total population
- Need to find **many more**!
  - All-sky survey
Pulsar Surveys I

Requirements

- Large bandwidths because pulsars are faint
- Long integration times because pulsars are faint
- Rapid time sampling in order to resolve pulse profile
- Narrow frequency channelization in order to mitigate interstellar scattering

- For a “pixel” on the sky, accumulate data for a time $\Delta t$ over a bandwidth $\Delta \nu$
  
  Suppose $\Delta t = 20$ min., $\Delta \nu = 800$ MHz
- Time sampling $\delta t$ with frequency channelization $\delta \nu$
  
  For GBT GUPPI, $\delta t = 81.92$ $\mu$s, $\delta \nu = 24$ kHz

  60 GB data sets per pixel …
For GBT
• At 800 MHz, “pixel” \( \sim 16' = 0.3^\circ \)
• About 350 kpixels in the sky
• 20 PB data set

For SKA
• At 800 MHz, “pixel” = 1.2'
• About 76 Mpixels in the sky
• 4.6 EB data set
Data Intensive Astronomy
(“There is nothing new under the Sun.”)

Data Volumes

- Hipparchus
  - ca. 135 BCE
  - Stellar catalog with 850 entries
  - SKA pulsar survey

Computational Limitations

- Harvard computers
  - Production of stellar plates and spectra (“data rate”) was increasing enormously
  - Examined and classified telescope output
  - SKA all-sky survey

Exploring the Universe with the world’s largest radio telescope
SKA Pathfinding

- SKA is ultimate goal, though long-term program
- Precursors and many pathfinders in existence or under construction
  - Data challenges before SKA comes on-line
  - Scalability could be an issue
Square Kilometre Array

The Global Radio Wavelength Observatory

- Originally: “Hydrogen telescope”
  Detect H I 21-cm emission from Milky Way-like galaxy at z ~ 1

- SKA science much broader
  ⇒ Multi-wavelength, multi-messenger
- On-going technical development
  ➢ Cyber-infrastructure and “big data”
- International involvement