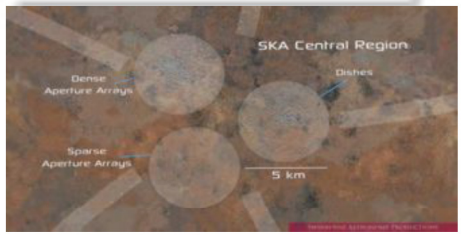

# The Square Kilometre Array Massive Data Challenges at the Frontiers of Astronomy, Physics, & Astrobiology

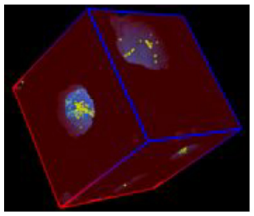
Joseph Lazio
Project Scientist, SKA Program Development Office
&
Jet Propulsion Laboratory, California Institute of Technology

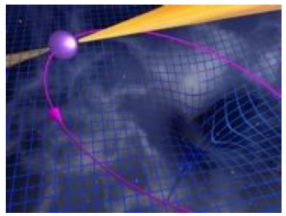
# Square Kilometre Array

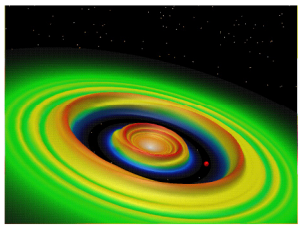
## The Global Radio Wavelength Observatory

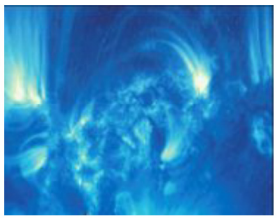
- Originally: “Hydrogen telescope”
  - Detect H I 21-cm emission from Milky Way-like galaxy at $z \sim 1$
- SKA science much broader
  - ⇒ Multi-wavelength, multi-messenger
- On-going technical development
  - ➢ Cyber-infrastructure and “big data”
- International involvement







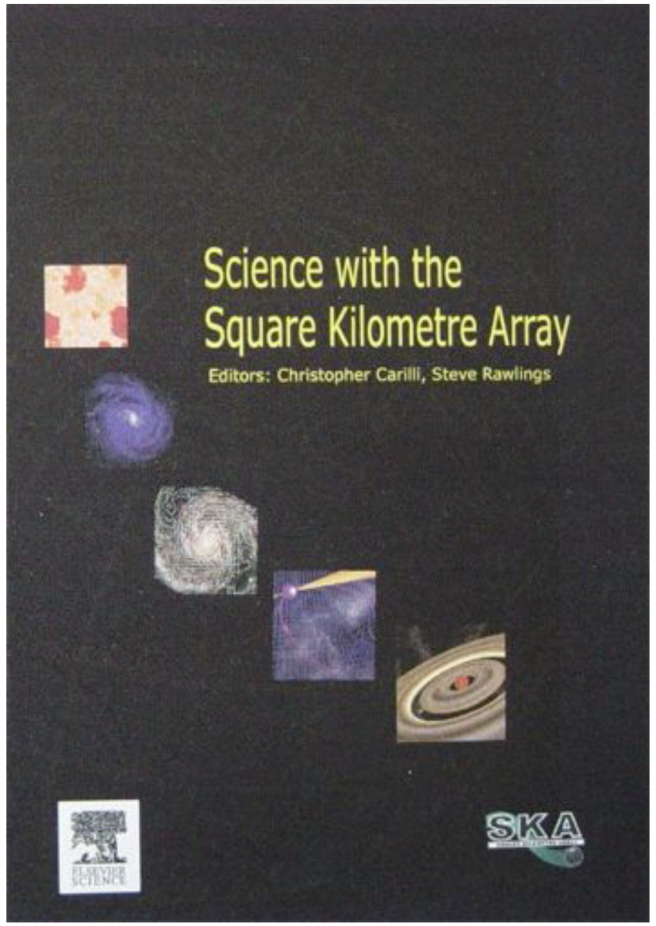


# SKA Key Science

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*




*Science with the Square Kilometre Array* (2004, eds. Carilli & Rawlings, *New Astron. Rev.*, **48**)

# SKA and the U.S.

*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)

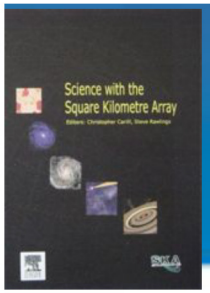
    Could evolve into low frequency component of SKA

  - North American Nanohertz Observatory for Gravitational Waves (NANOGrav)

    Could evolve into intermediate frequency component of SKA

# *NWNH* Science Program and the SKA


| | | |
|---|---|---|
| **Discovery** | Gravitational wave astronomy | "Strong-field Probes of Gravity with Pulsars and Black Holes" |
| | Epoch of Reionization | "Emerging from the Dark Ages and the Epoch of Reionization" |
| **Origins** | *What were the first objects to light up the Universe and when did they do it?* | "Emerging from the Dark Ages and Epoch of Reionization" |
| | *How do circumstellar disks evolve and form planetary systems?* | "The Cradle of Life and Astrobiology" |
| **Understanding the Cosmic Order** | *How do baryons cycle in and out of galaxies and what do they do while they are there?* | "Galaxy Evolution, Cosmology, and Dark Energy" |
| **Frontiers of Knowledge** | *What controls the masses, spins and radii of compact stellar remnants?* | "Strong-field Probes of Gravity with Pulsars and Black Holes" |




# 21st Century Astrophysics

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

**21st Century**: We understand the Universe we inhabit.

## Cosmology & Fundamental Physics

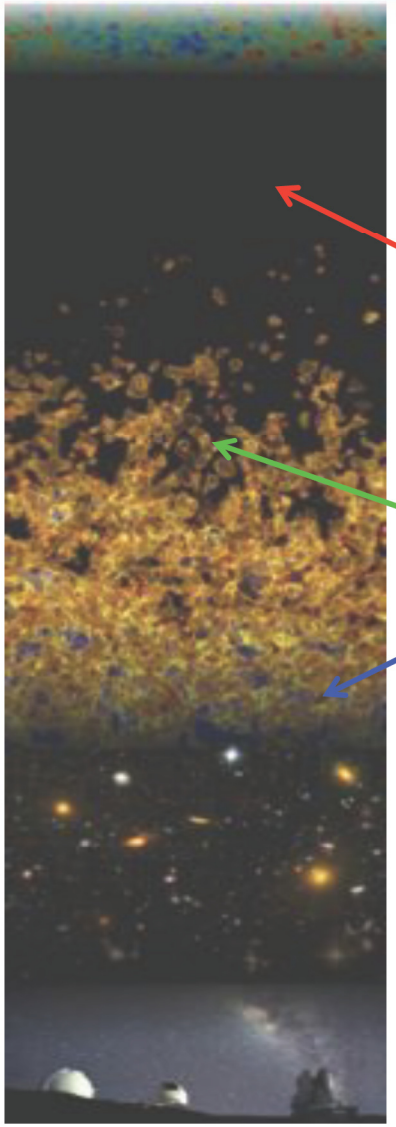
- Gravity
  - Can we observe strong gravity in action?
  - What is dark matter and dark energy? (dark energy and BAOs with H I galaxies)
- Magnetism
- Strong force

  Nuclear equation of state

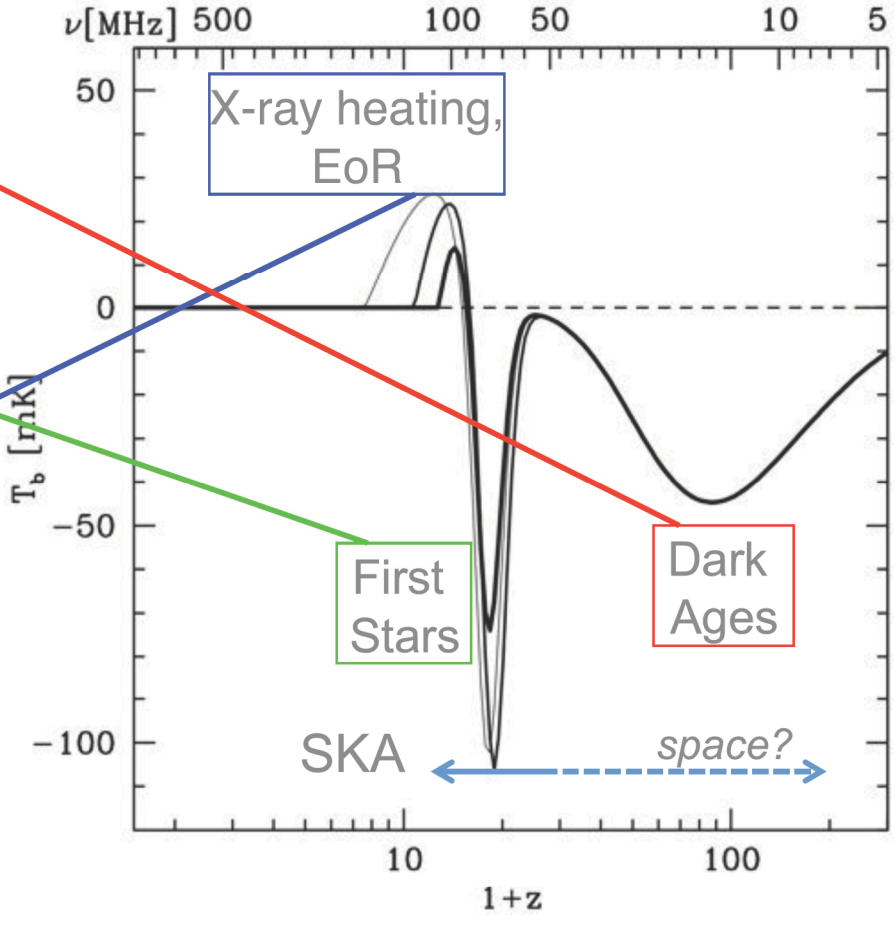
## Galaxies Across Cosmic Time, The Galactic Neighborhood, Stellar and Planetary Formation

- 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)

# Evolution of the Universe



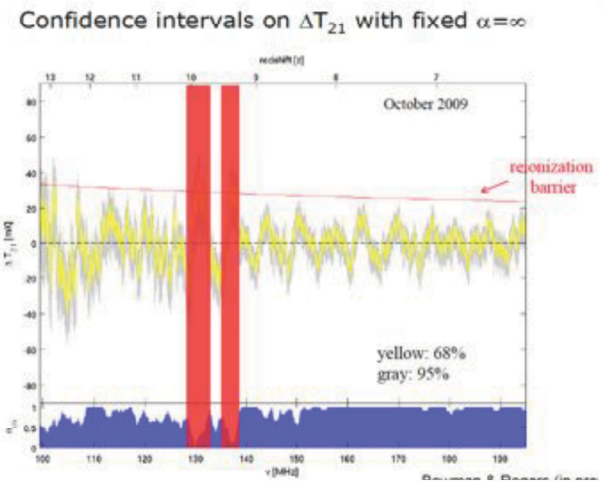
H I brightness temperature signal
(w.r.t. CMB)



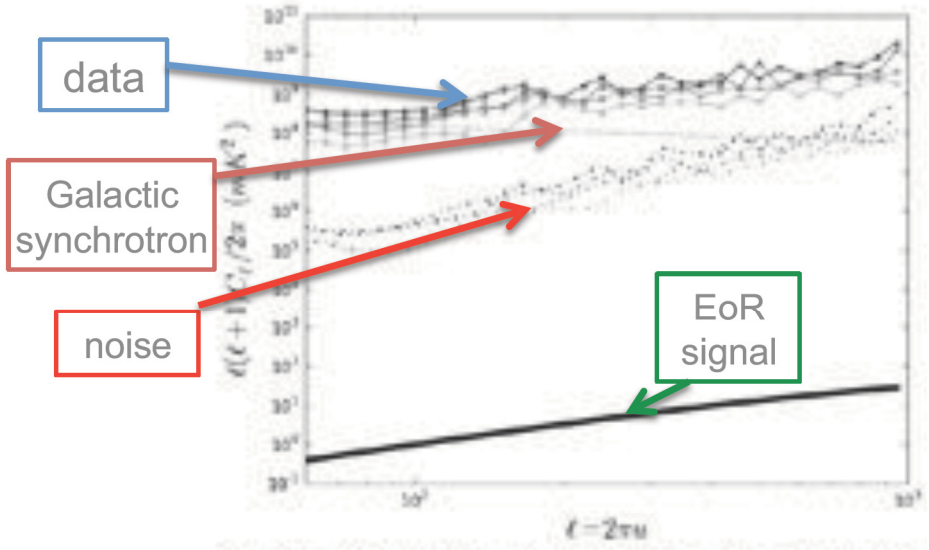

(Pritchard & Loeb 2008)

# Evolution of the Universe
## Epoch of Reionization

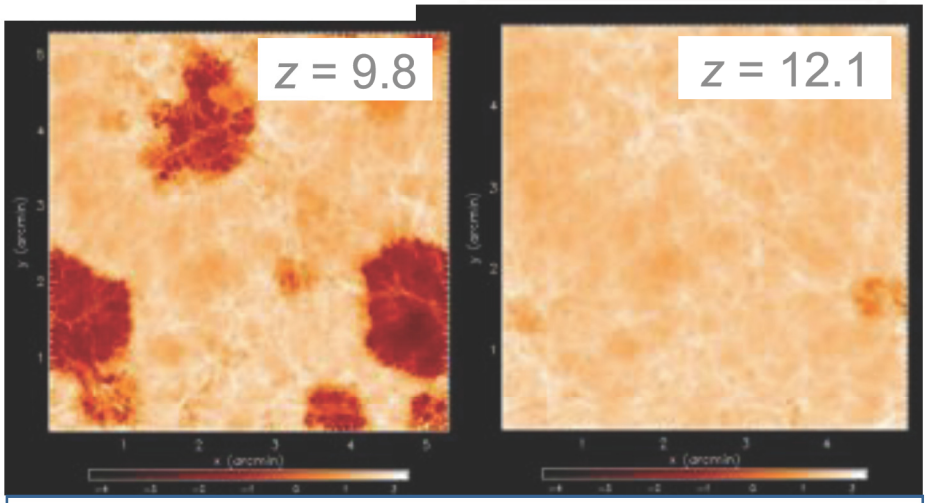
EDGES
(Bowman et al. 2008)



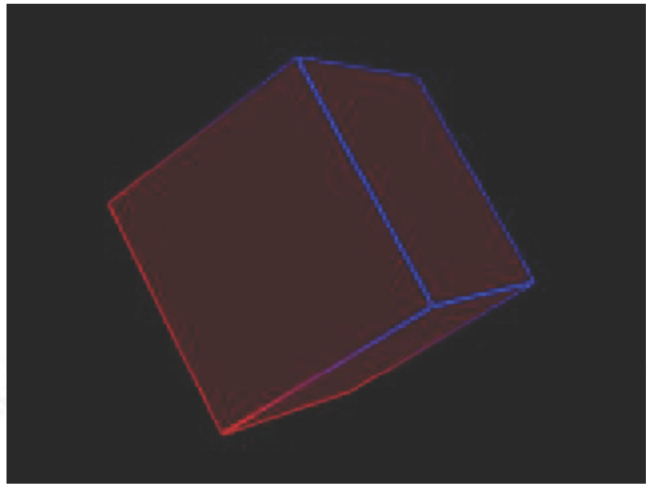

PAPER




Parsons et al. 2009; arXiv:0904.2334




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

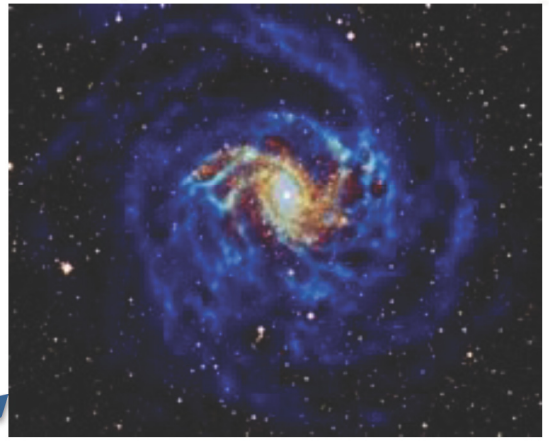


Furlanetto et al.; Gnedin

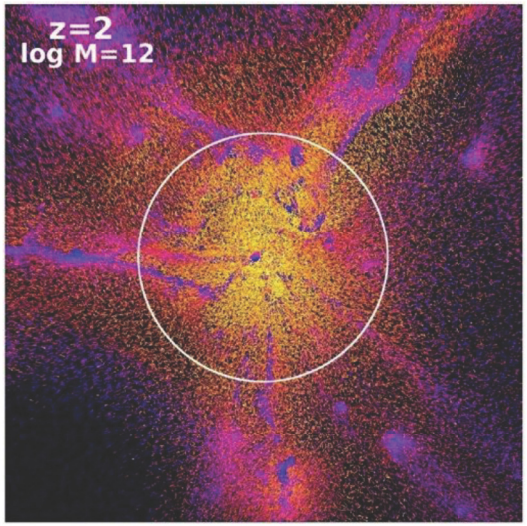
# Galaxy Assembly
## Stars *and* Gas

- Gas content and dynamics becoming critical part of simulations.
  - N-body simulations themselves can lead to data challenges!
- Astronomy is an *observational* science.

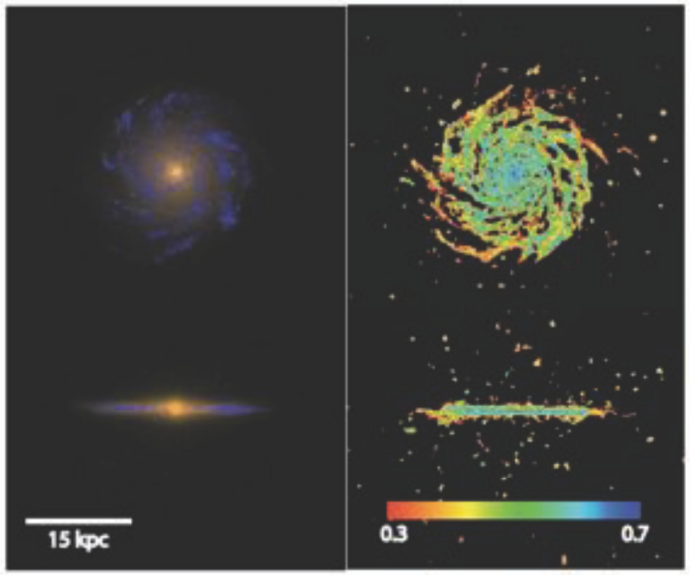
➢ Need observations of gas content —over cosmic time—to understand galaxy formation!



observation vs. simulation



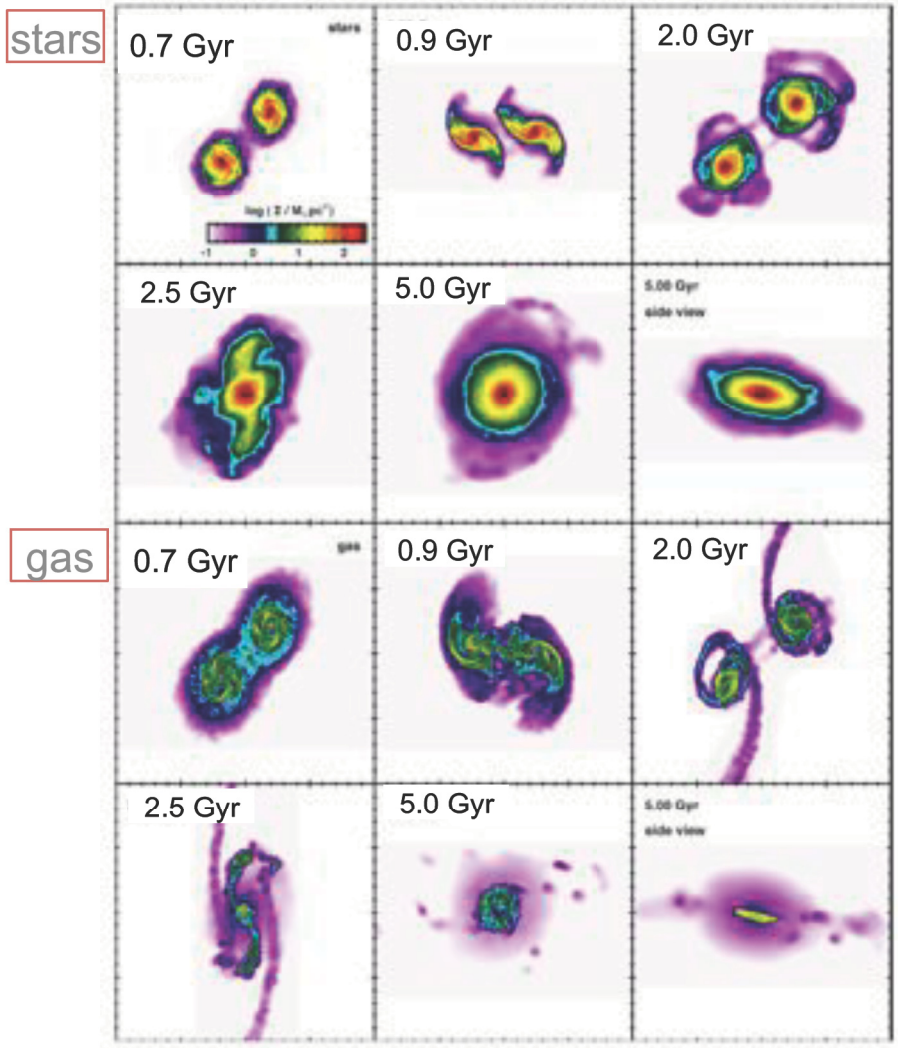

Keres et al.




Eris simulation (Guedes et al.)

NGC 6946 (T. Oosterloo)

# Galaxy Assembly
## The Role of Mergers




(Moster et al. arXiv:1104.0246)

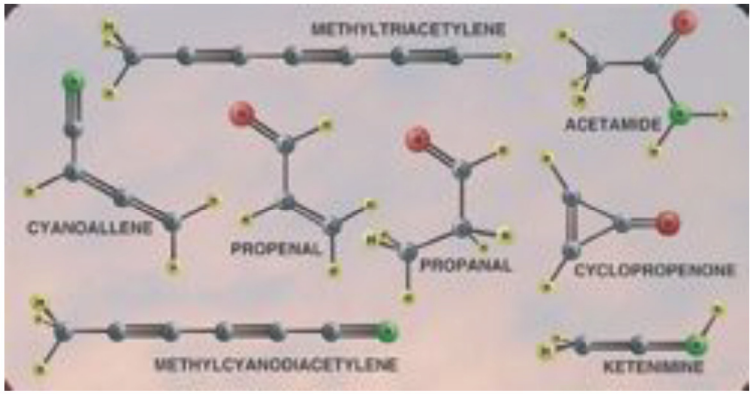

- Mergers are recognized as important aspect of galaxy evolution and formation
- Gas can be sensitive tracer of interactions, long after original event took place

  E.g., Holwerda et al. with THINGS

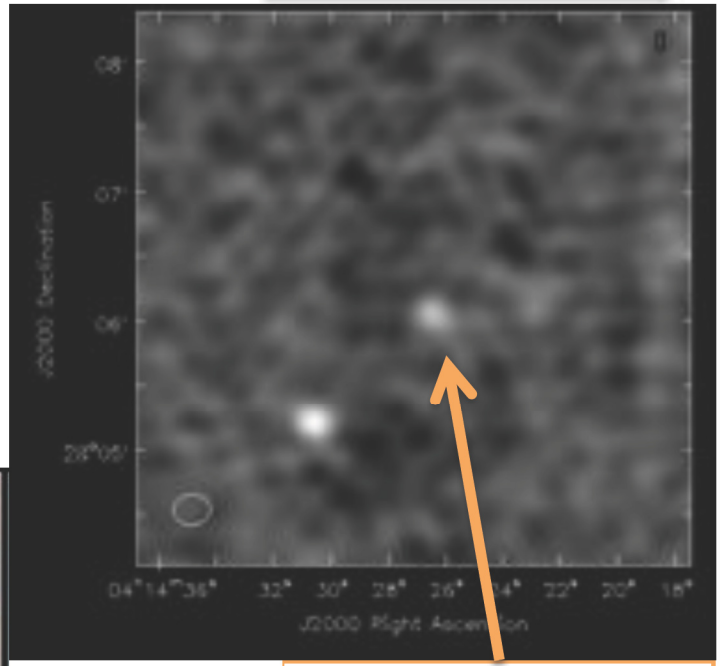
# Astrobiology at Long Wavelengths

## $\lambda > 1$ cm

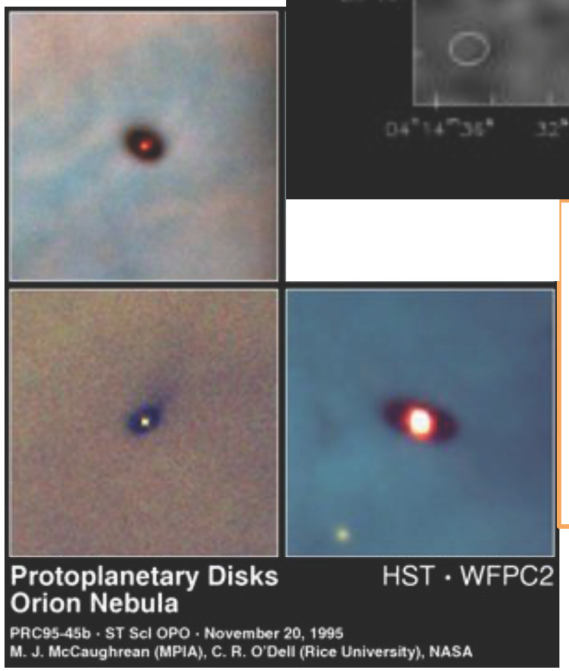
- Not affected by dust
- Complex molecules have transitions at longer wavelengths
- “Waterhole” (1.4–1.7 GHz)
- Magnetically-generated emissions from extrasolar planets

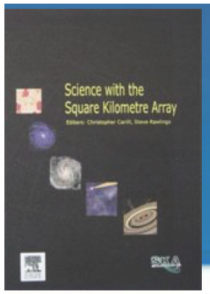



Complex organic molecules detected at radio wavelengths



EVLA 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

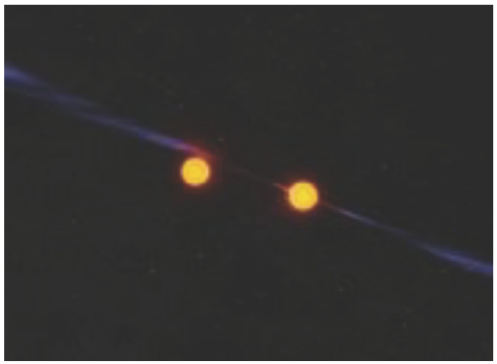
  Nuclear equation of state

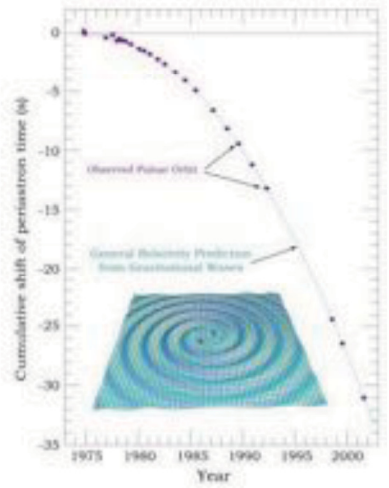
## Galaxies Across Cosmic Time, The Galactic Neighborhood, Stellar and Planetary Formation

- 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)

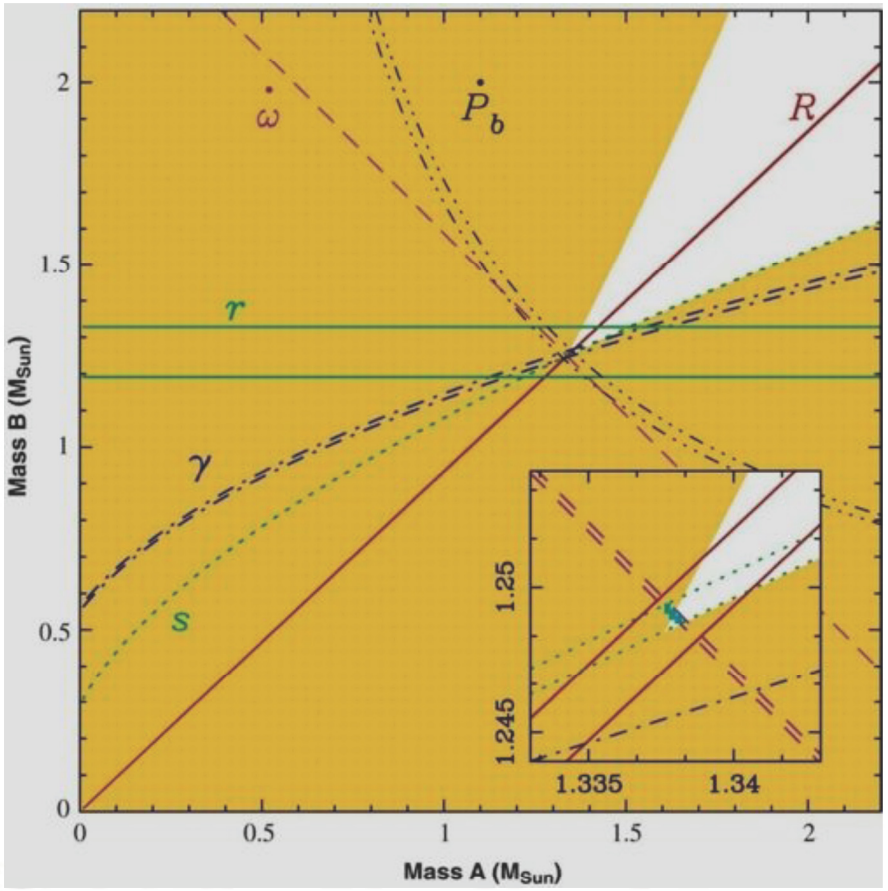
# Did Einstein Have the Last Word on Gravity?

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






PSR J0737-3039




Kramer et al.

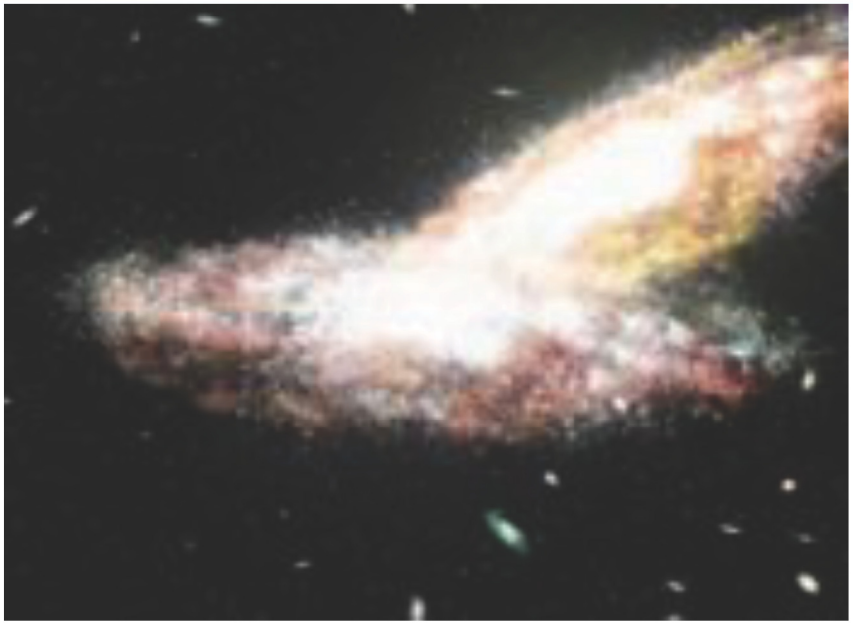
Relativistic binaries probe

1. Equivalence principle
2. Strong-field tests of gravity

- Neutron star-neutron star and neutron star-white dwarf binaries known

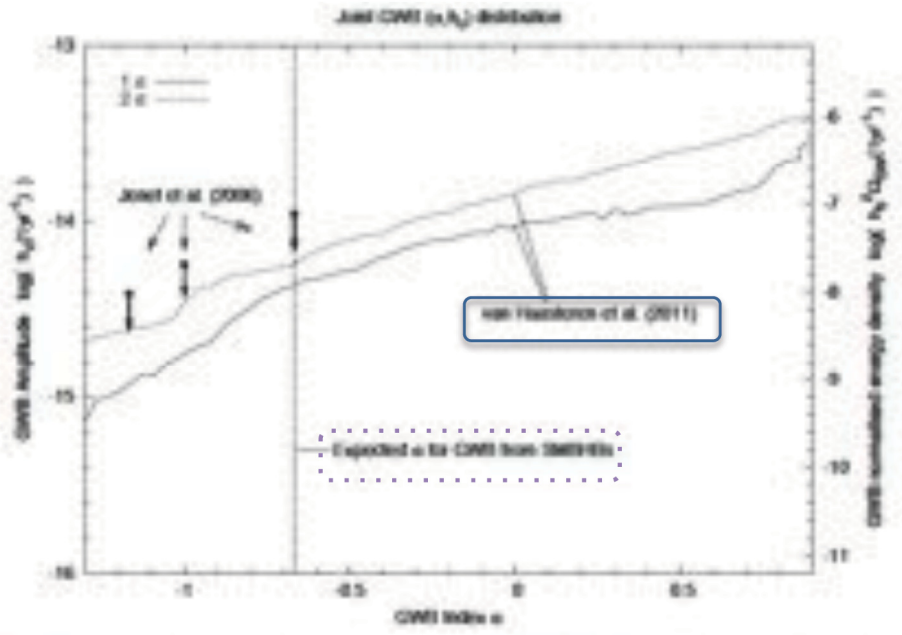
? Black hole-neutron star binaries?

# 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

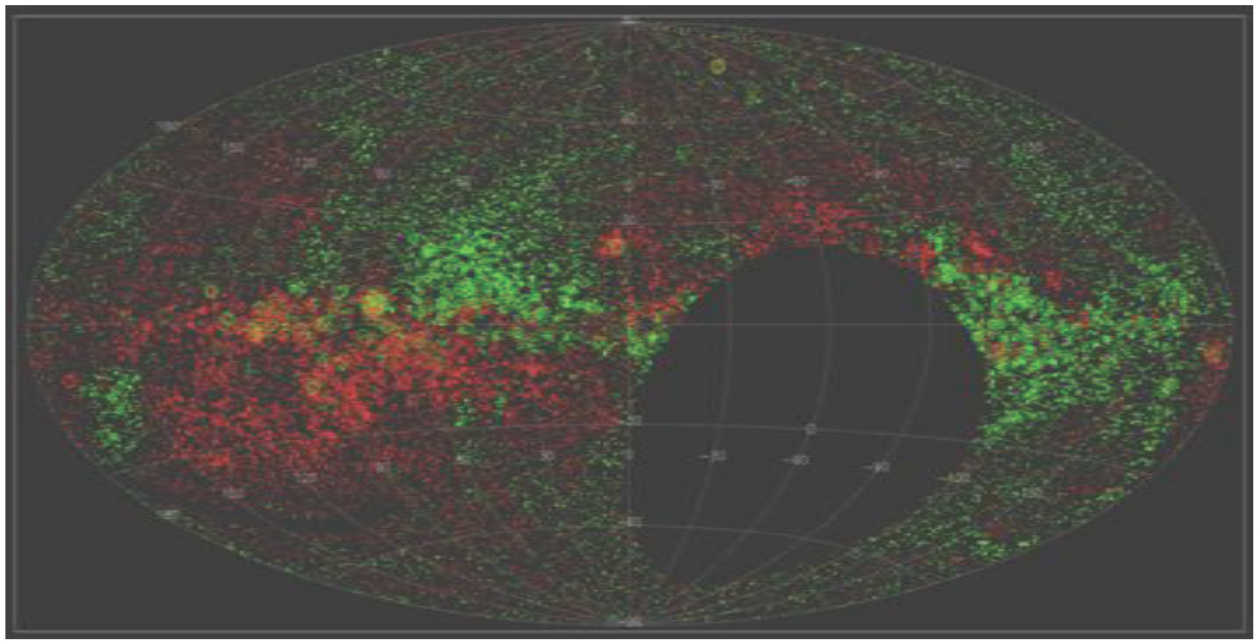


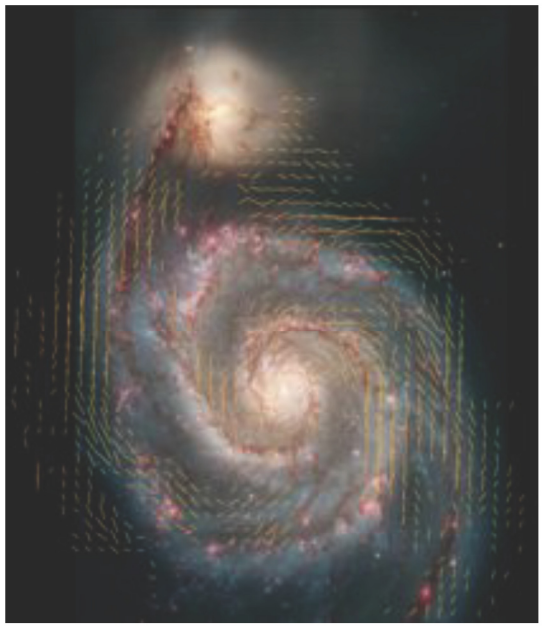

Pulsar timing arrays starting to provide results from ensemble of pulsars

- EPTA (van Haastern et al., *above*)
- PPTA (Yardley et al.)
- *NANOGrav (Demorest et al.)*

# Origin & Evolution of Cosmic Magnetic

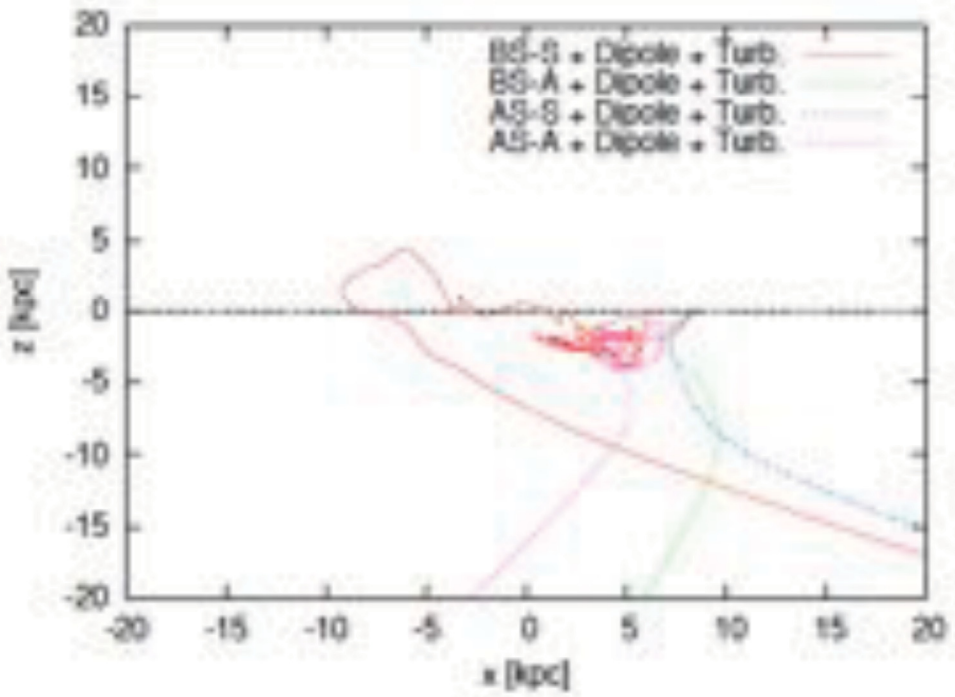
- Magnetic fields are fundamental, but poorly constrained
  - Affects galaxy, cluster evolution?
  - Affects propagation of cosmic rays in ISM and IGM
- All-sky rotation measure surveys provide **B** fields along lines of sight
- Continuum in I, Q, and U!





# 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?
  - – …?



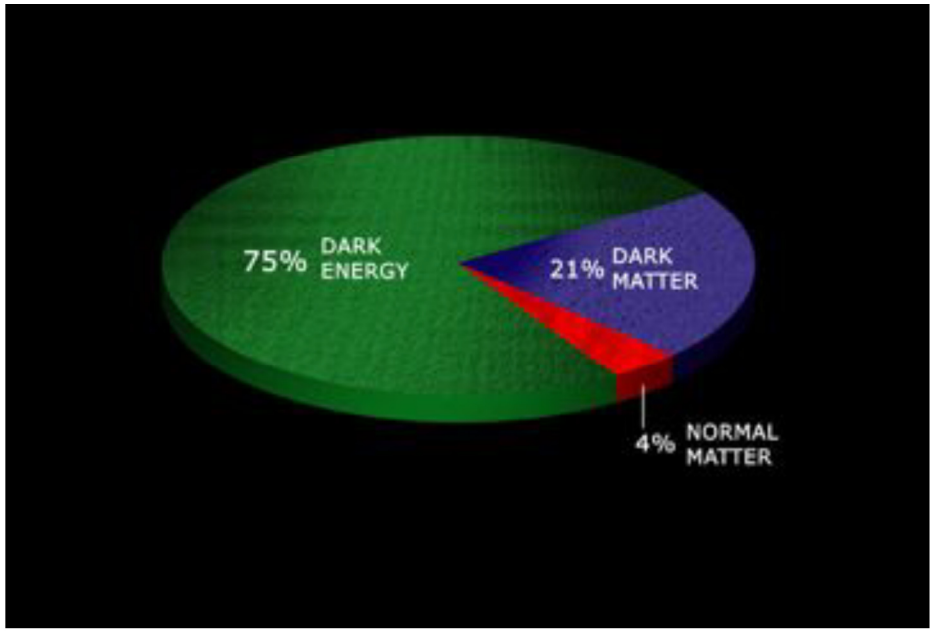

Takami, arXiv:1104.0278

# Cosmology and Gravity

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

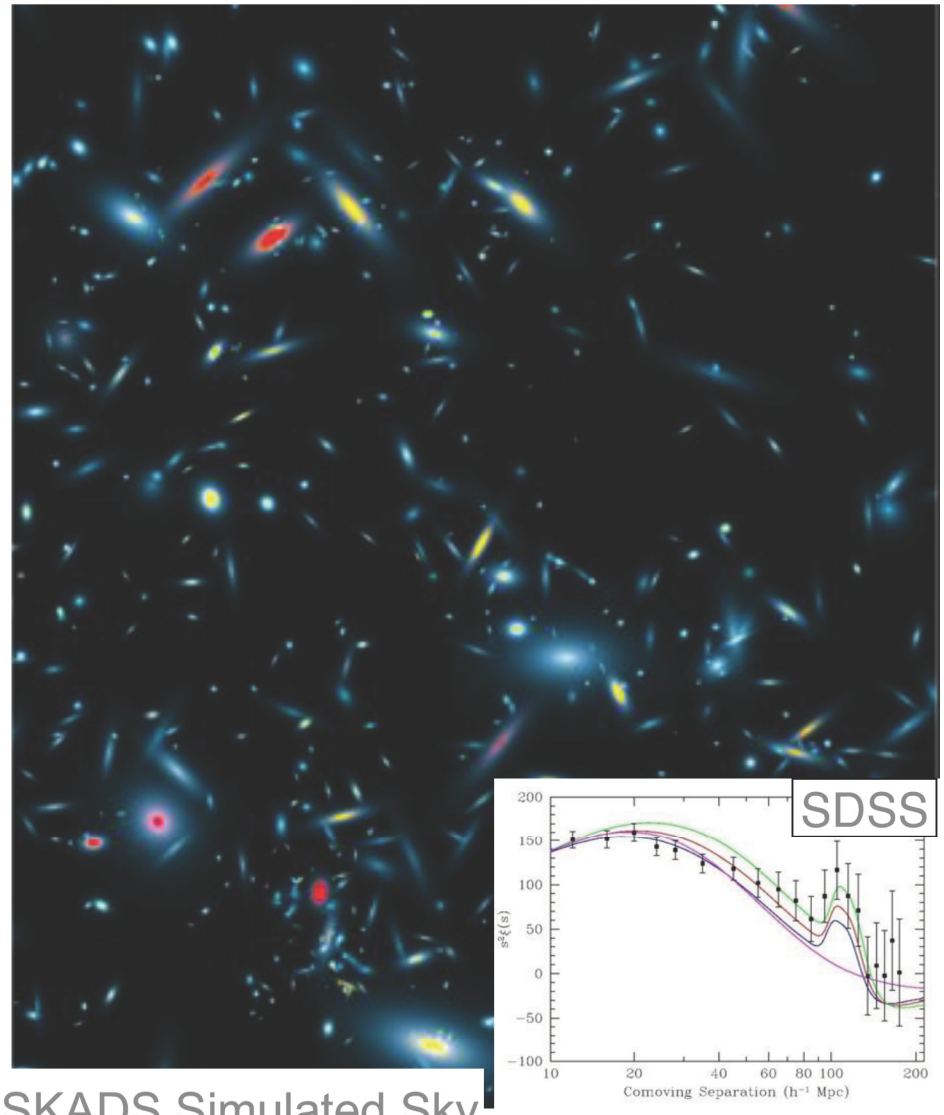
Origin and Fate of the Universe

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




Composition of the Universe

# Cosmology and Sky Surveys

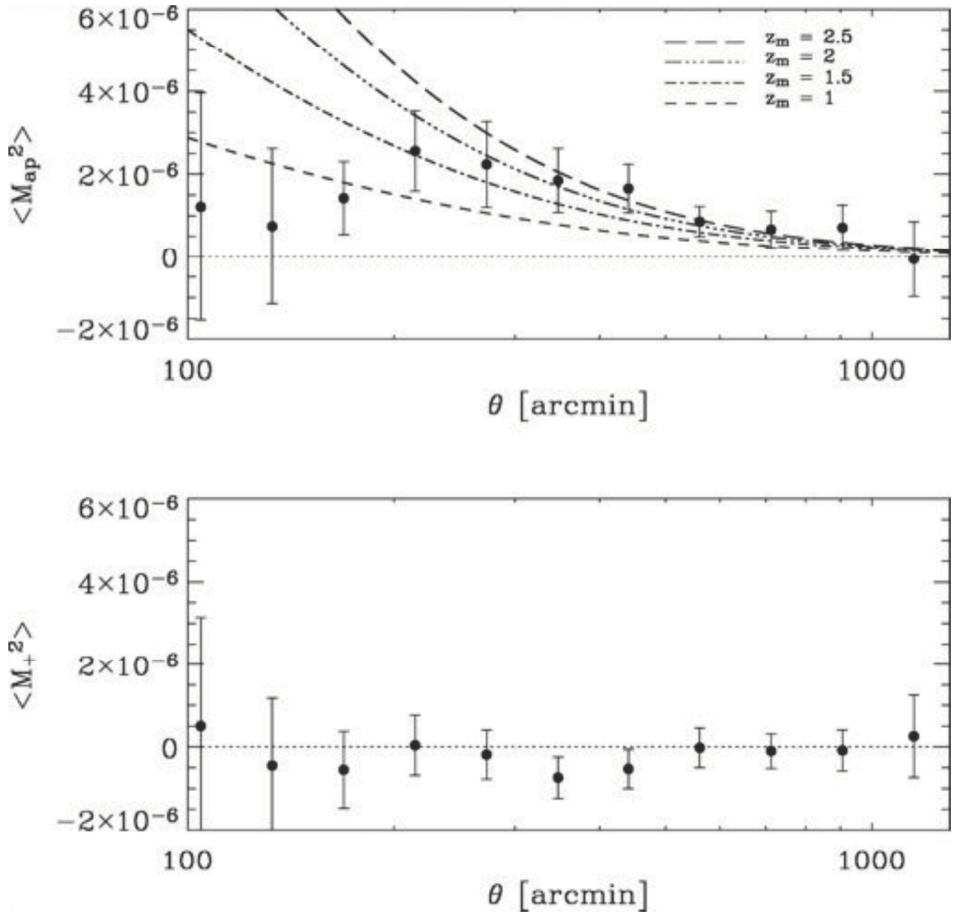


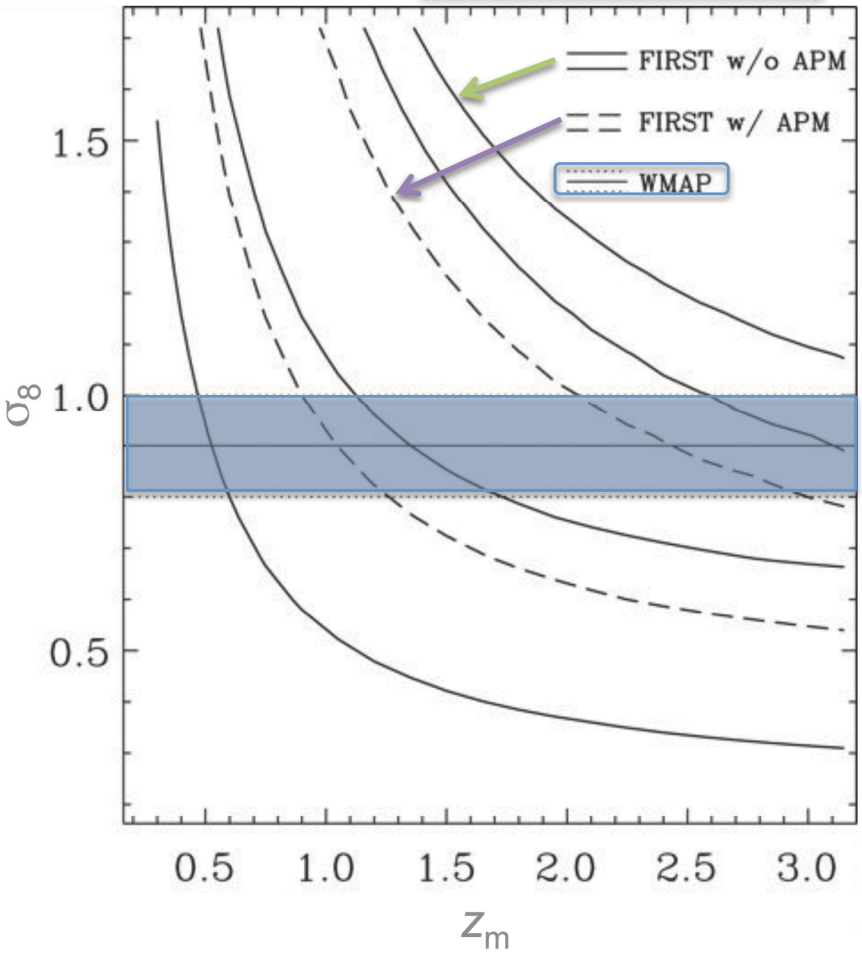

SKADS Simulated Sky

- 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

# Cosmology and Gravity

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




Radio observations should have fewer (different) systematics

# 21st Century Astrophysics

## Fundamental Forces and Particles

- Gravity
- Magnetism
- Strong force

## Origins

- Galaxies and the Universe
- Stars, Planets, and Life

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

Photon frequency / wavelength / energy

Time

Polarization

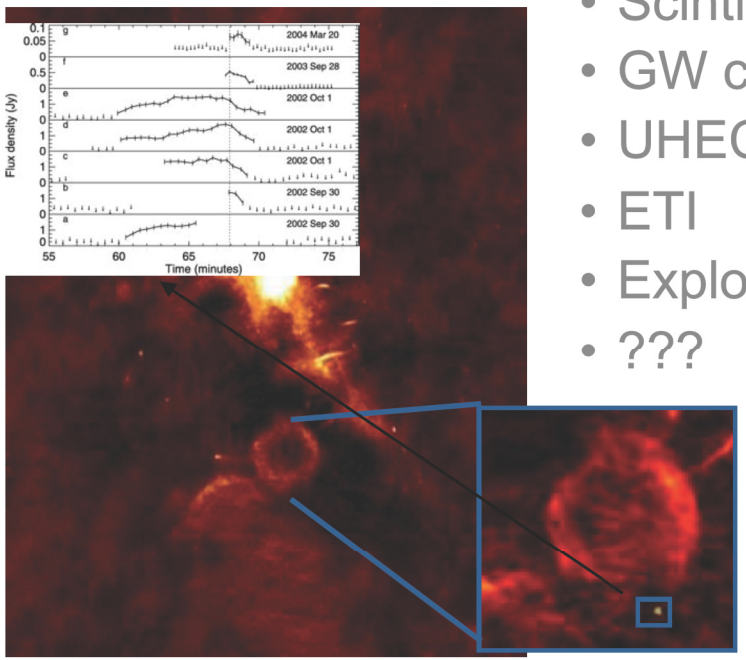
Sensitivity

Field of View

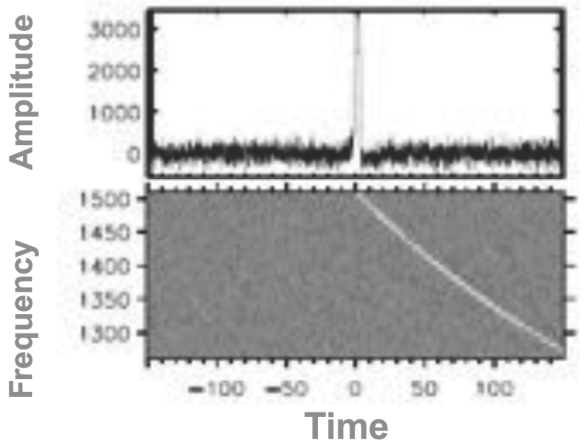
Angular Resolution

# The Dynamic Radio Sky

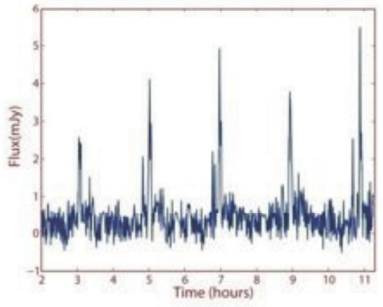
- Neutron stars
  - Magnetars
  - Giant pulses
  - Short GRBs?
- Microquasars
- Tidal Disruption Events

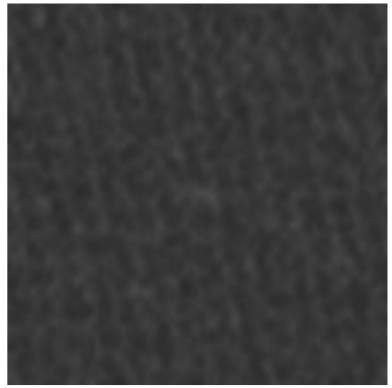



- GRBs ($\gamma$-ray loud; $\gamma$-ray quiet?)
  - Afterglows
  - Prompt emission?
- Sub-stellar objects
  - Brown dwarfs
  - Extrasolar planets?
- Scintillation
- GW counterparts
- UHECRs
- ETI
- Exploding black holes
- ???

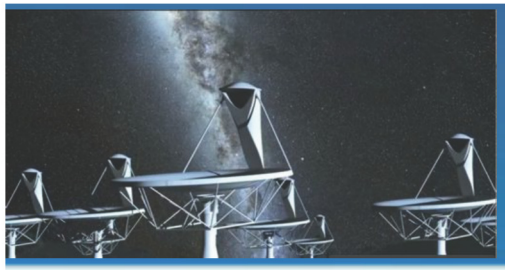



Rotating Radio Transients (RRATS)



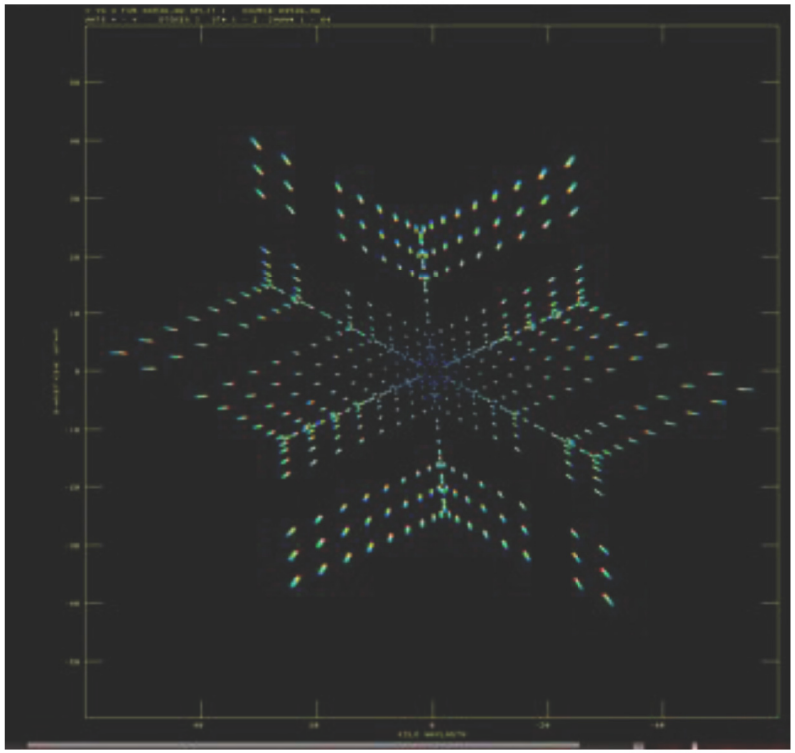



Pulsating Brown Dwarfs



# Imaging with Arrays

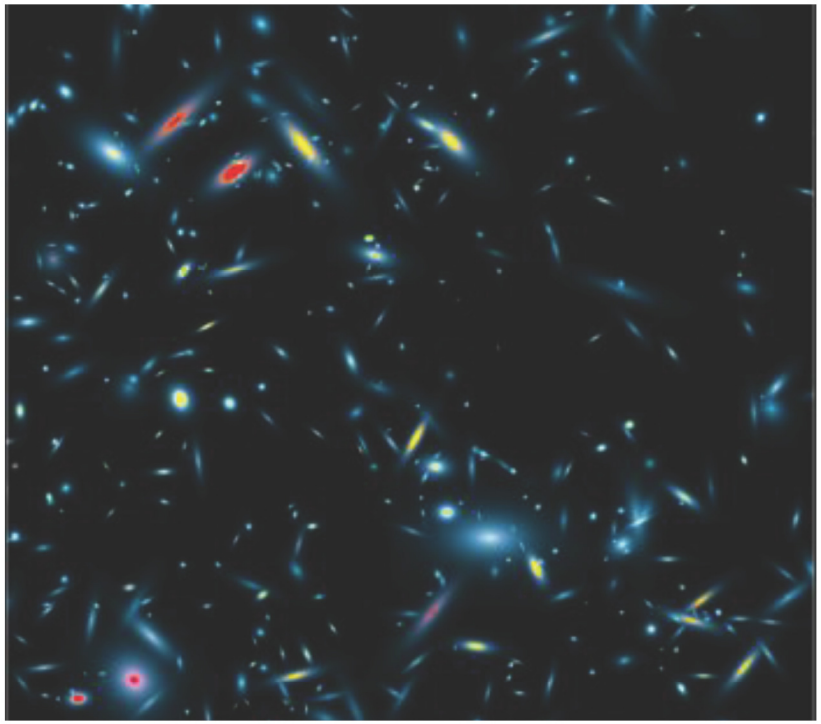
## Fourier transform (*u*-*v*) plane



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

Fourier Transform ↔

## Image plane



# Imaging Surveys

## Requirements

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

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

| ASKAP | SKA Phase 1 | SKA Phase 2 |
|---|---|---|
| $N_{antenna} = 30$ | $N_{anntena} \sim 250$ | $N_{antenna} \sim 1000$ |
| $N_{beams} = 30$ | $N_{beams} = 1$ | $N_{beams} = 1?$ |
| $N_{frequency} \sim 16k$ | $N_{frequency} \sim 16k?$ | $N_{frequency} \sim 16k?$ |

# Imaging Surveys II

| ASKAP | SKA Phase 1 | SKA Phase 2 |
|---|---|---|
| $N_{antenna} = 30$ | $N_{anntena} \sim 250$ | $N_{antenna} \sim 1000$ |
| $N_{beams} = 30$ | $N_{beams} = 1$ | $N_{beams} = 1?$ |
| $N_{frequency} \sim 16k$ | $N_{frequency} \sim 16k?$ | $N_{frequency} \sim 16k?$ |
| $N_{time} \sim 4k$ | | |
| $N_{data} \sim 1.8 \times 10^{12}$ | $N_{data} \sim 4 \times 10^{12}$ | $N_{data} \sim 65 \times 10^{12}$ |
| **$N_{OPS} \sim 18 \times 10^{15}$** | **$N_{OPS} \sim 40 \times 10^{15}$** | **$N_{OPS} \sim 650 \times 10^{15}$** |

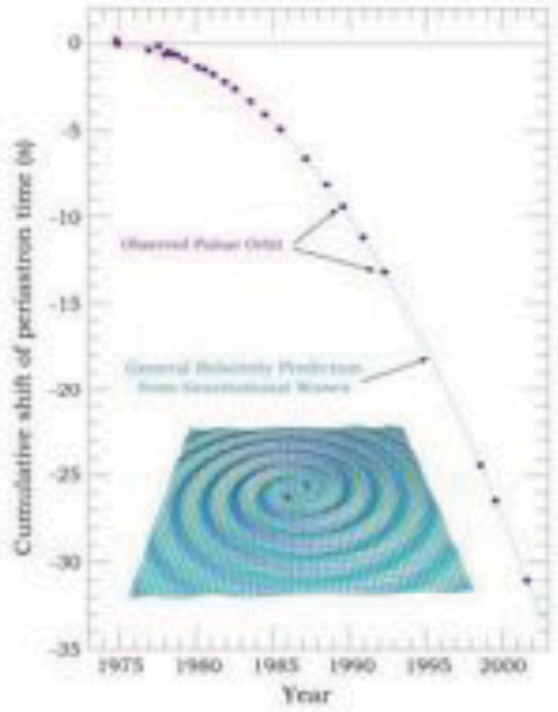
- Imaging is more than "just" an FFT.
  Gridding, deconvolution, wide-field corrections, self-calibration, …
- Community estimates are $10^4$ to $10^5$ ops per visibility datum(!).
- Leads to significant power challenges
  - Related to moving data on/off chips
  - Careful design can yield significant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

# Fundamental Physics with Radio Pulsars

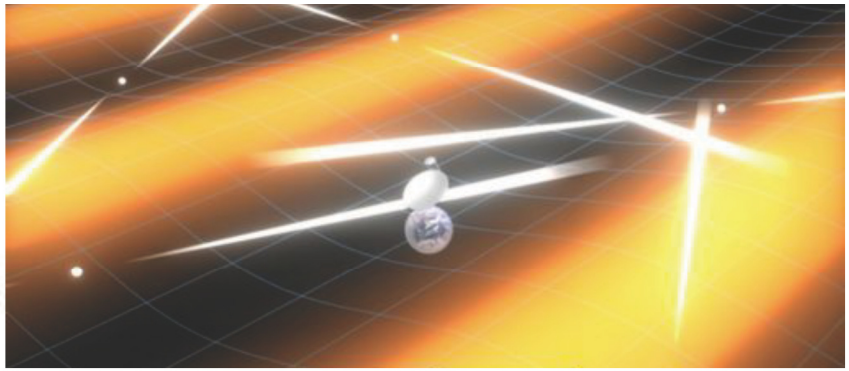
Arrival times of pulses from radio pulsars can be measured with phenomenal accuracy

- 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



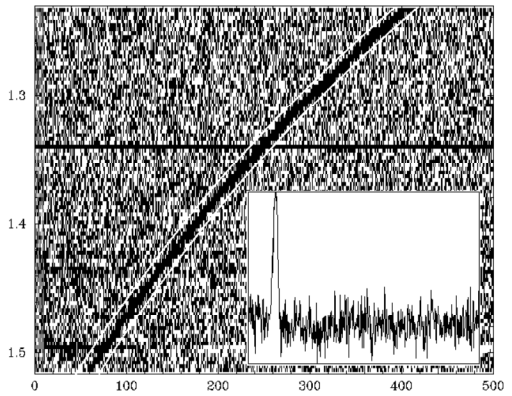

← Ultra-relativistic binaries & gravitational wave studies



# 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 …

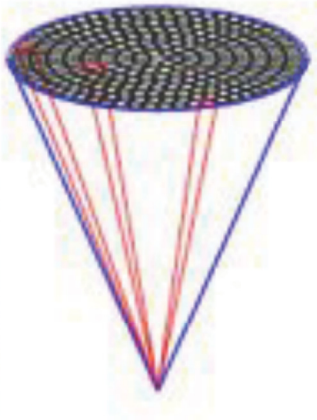
# Pulsar Surveys II

**For GBT**

- At 800 MHz, “pixel” ~ 16' = 0.3°
- 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

Ιππαρχος (Hipparcus)
- ca. 135 BCE
- Stellar catalog with 850 entries

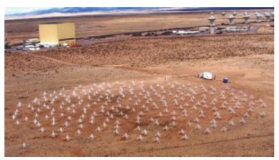
➢ **SKA pulsar survey**

## Computational Limitations

Harvard computers
- Production of stellar plates and spectra ("data rate") was increasing enormously
- Examined and classified telescope output

➢ **SKA all-sky survey**

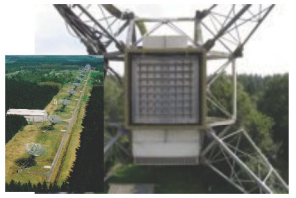
# SKA Pathfinding

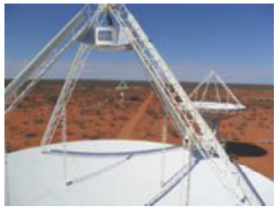


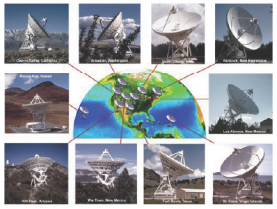
- SKA is ultimate goal, though long-term program
- Precursors and many pathfinders in existence or under construction

➢ 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 $z \sim 1$
- SKA science much broader
  - ⇒ Multi-wavelength, multi-messenger
- On-going technical development
  - ➢ Cyber-infrastructure and “big data”
- International involvement