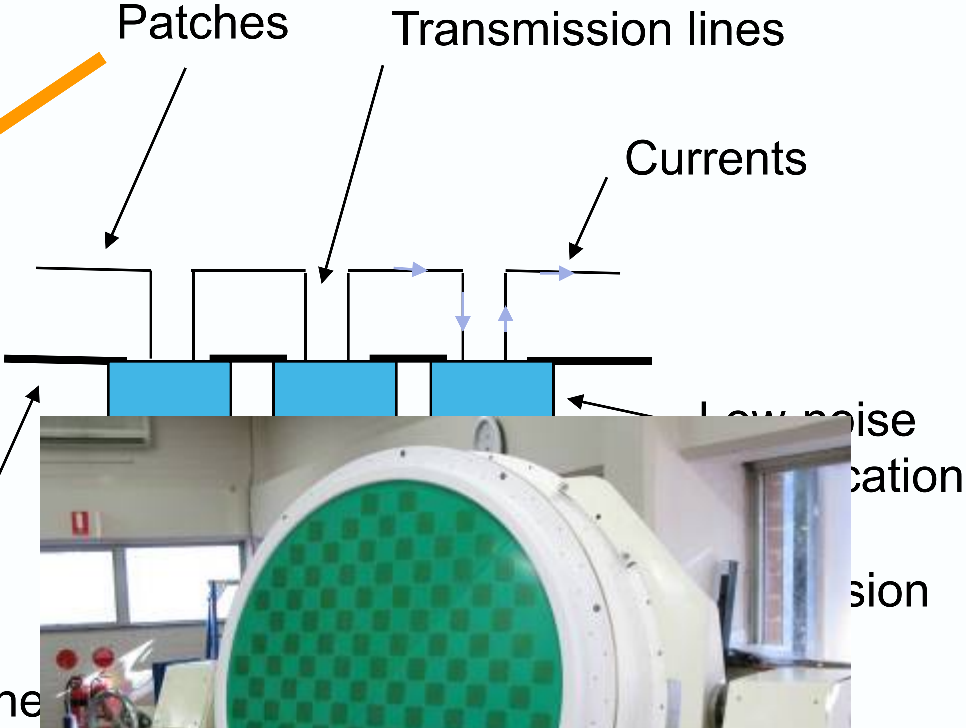
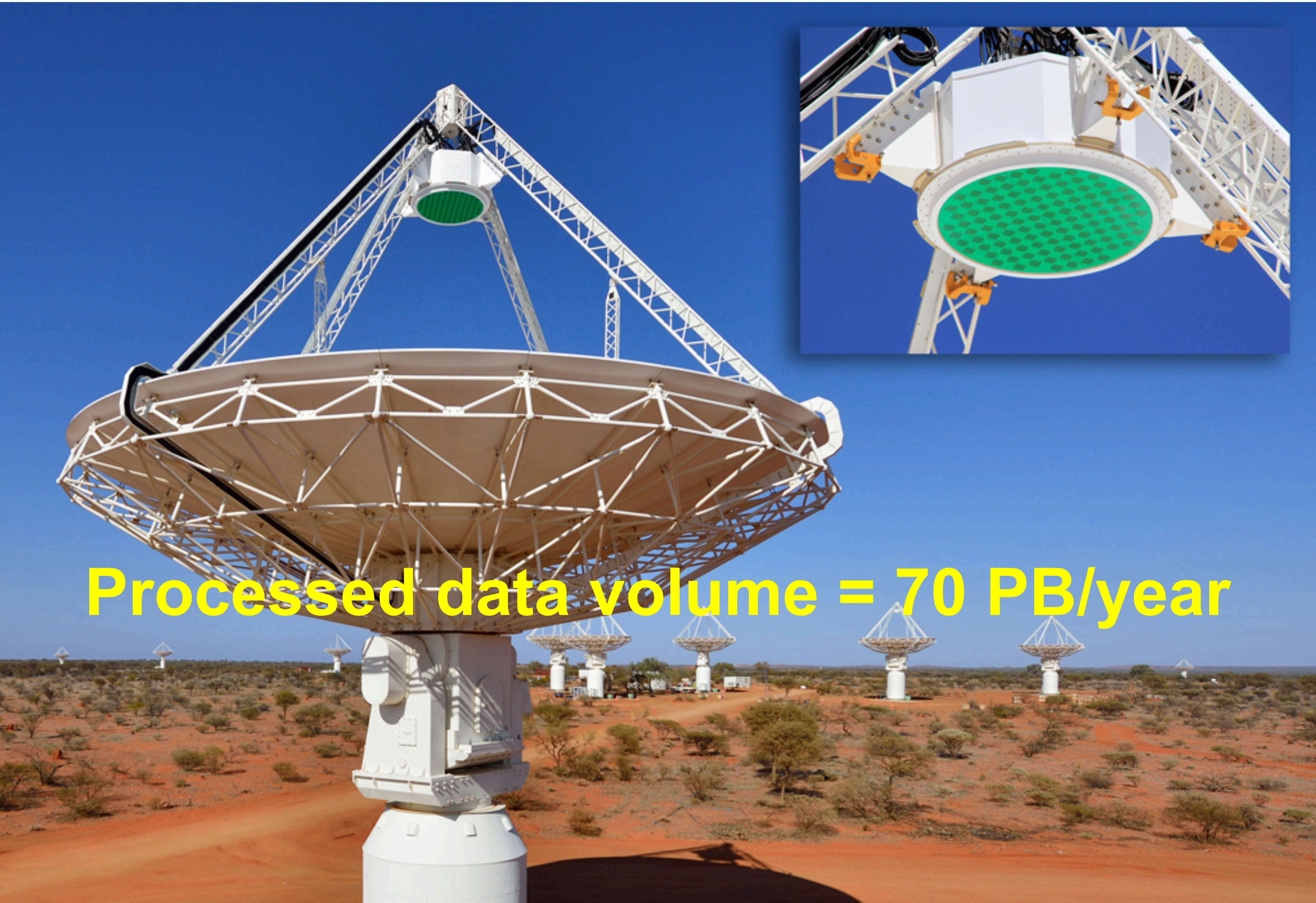
WTF? Evolutionary Map of the

ASKAP 188-element Phased Array Feed



- Connected checkerboard array
- Operating range defined by electromagnetics and LNA
  - 1800-700MHz
- LNA - High-impedance, differential

Total bandwidth = 2.1THz per antenna



Processed data volume = 70 PB/year

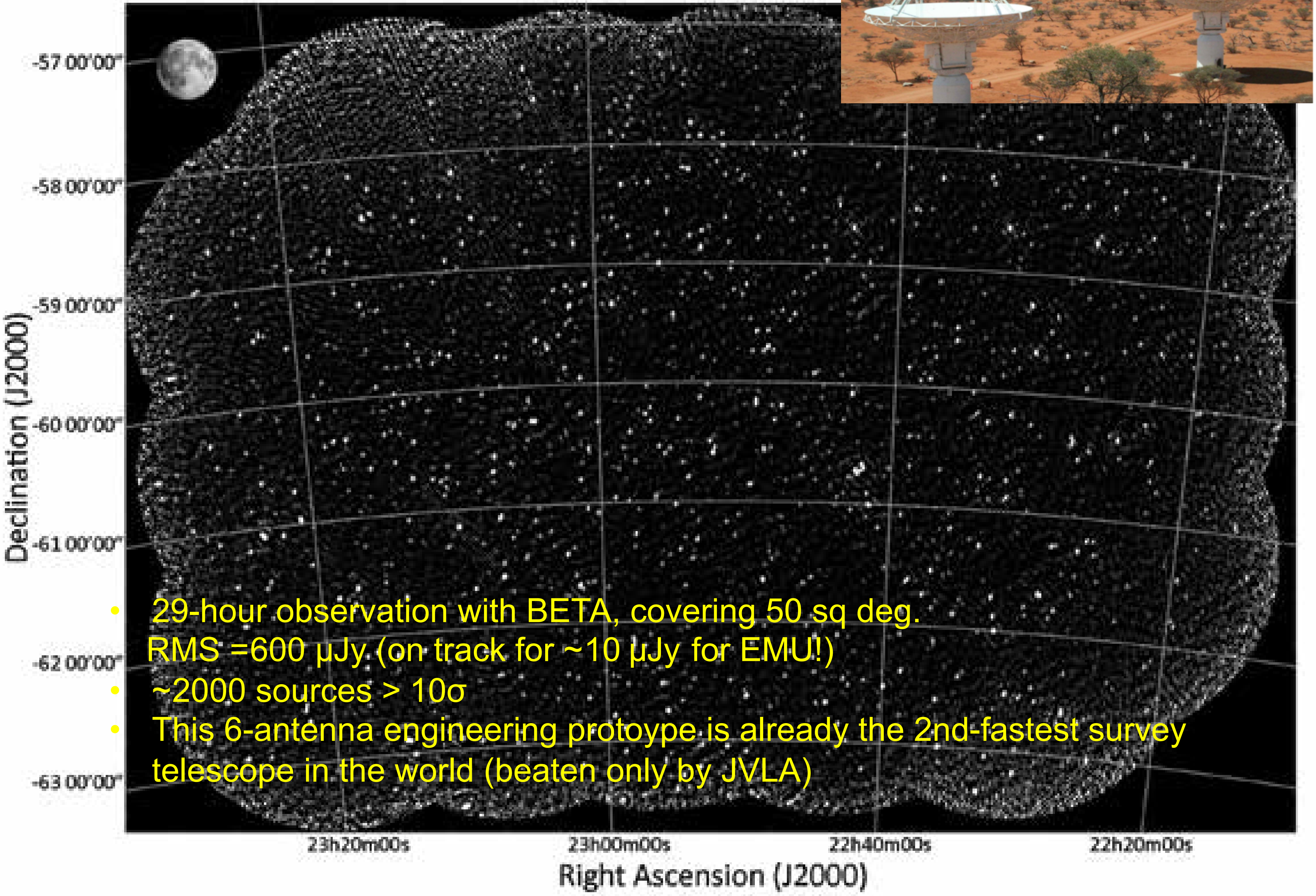
ASKAP Science Data Processor Platform

- The *galaxy* system at Pawsey
- 472 x Cray XC30 Compute Nodes
  - 200 TFlop/s Peak
- Cray Aries (Dragonfly topology)
- Cray Sonexion Lustre Storage
  - 1.4 PB usable
  - 480 x 4TB Disk Drives, RAID 6 + Hot Spares
  - Peak I/O performance: 30 GByte/s

Current ASKAP status

- All 36 antennas and infrastructure completed,
- Engineering prototype array ("BETA") currently operating with prototype PAFs on 6 antennas giving 9 beams
- 4 MkII PAFS currently installed
- **Planned schedule**
- 12 MkII PAFS installed by ~Dec 2015
- Mid 2016: "shared risk" ASKAP early science
- All 30-36 MkII PAFs installed by mid-2016
- 2017: Full EMU/WALLABY surveys start ??????

PAFs work (even with only 6 antennas)!



- 29-hour observation with BETA, covering 50 sq deg.
RMS =600 μJy (on track for ~10 μJy for EMU!)
- ~2000 sources > 10σ
- This 6-antenna engineering protoype is already the 2nd-fastest survey telescope in the world (beaten only by JVLA)

EMU Overview

Deep radio image of 75% of the sky (to declination +30°)

Frequency range: 1100-1400 MHz

40 x deeper than NVSS
- 10 μJy rms across the sky

5 x better resolution than NVSS (10 arcsec)

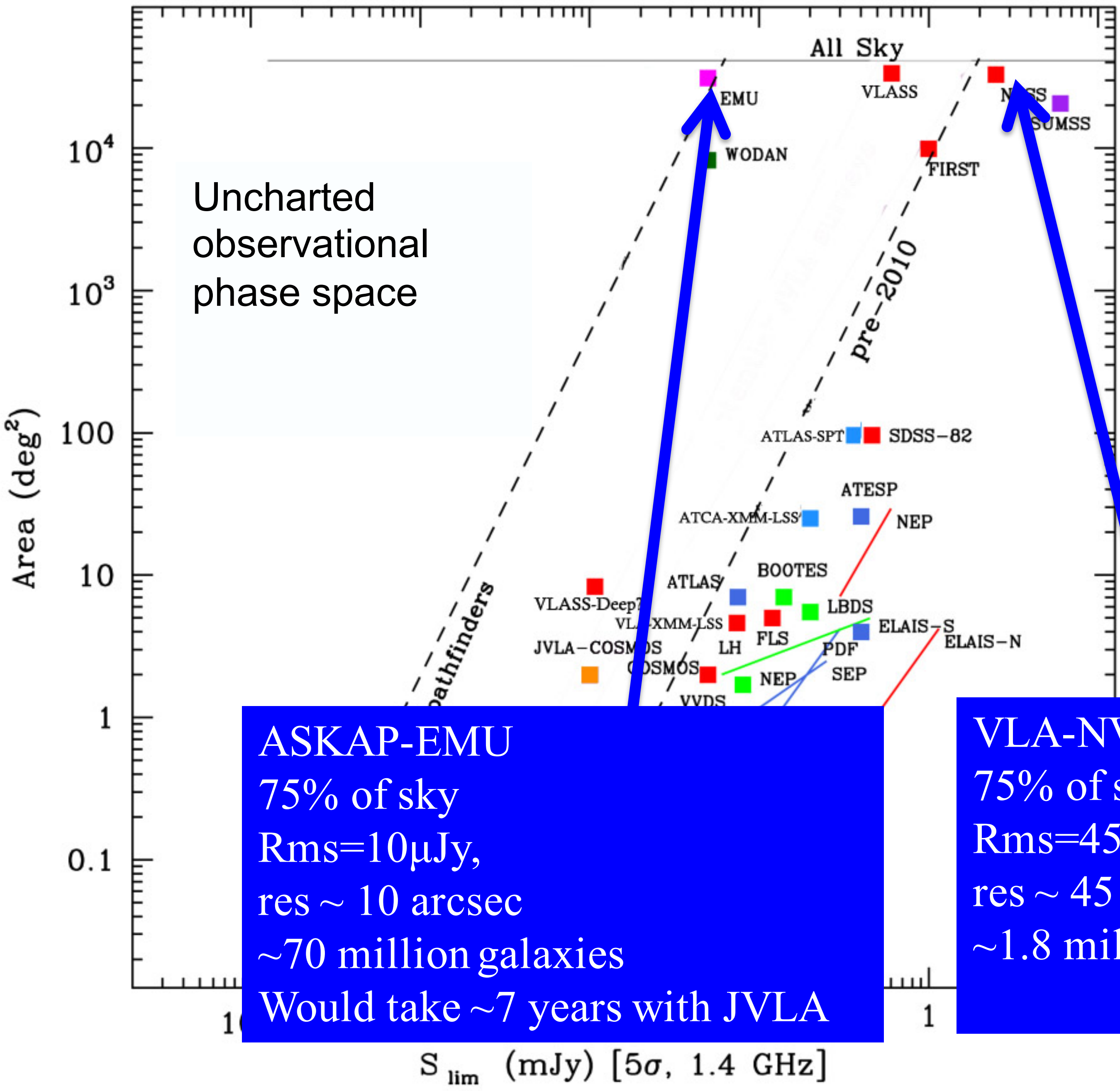
Better sensitivity to extended structures than NVSS

Will detect and image ~70 million galaxies at 20cm

All data to be processed in pipeline

Images, catalogues, cross-IDs, to be placed in public domain

Survey starts 2017(?)



Uncharted observational phase space

Surveys at other frequencies have been converted to 1.4 GHz equivalent assuming S α ν$^{-0.7}$

ASKAP-EMU
75% of sky
Rms=10μJy,
res ~ 10 arcsec
~70 million galaxies
Would take ~7 years with JVLA

VLA-NVSS
75% of sky
Rms=450μJy,
res ~ 45 arcsec
~1.8 million galaxies

How does EMU differ from earlier surveys?

1. Scale – increases the number of known radio sources by a factor of ~30

2. Will not be dominated by AGN – about half the galaxies will be normal SF galaxies

3. Ambition – includes:

- Cross-identification with optical/IR catalogues
- Ancillary data (redshifts etc)
- Key science projects as an integral part of the project

4. Explicitly includes “discovering the unexpected”

EMU and its pathfinders

Comparison: NVSS
3π sr
Rms=450 μJy
1.8 million galaxies

ATCA – ATLAS
(2006-2013)
6 antennas single-pixel

7 sq deg
Rms=15 μJy
6000 galaxies

ATCA – ATLAS - SPT
(2013-2015)
6 antennas single-pixel

100 sq deg
Rms=40 μJy
30,000 galaxies
300 clusters?

ASKAP – early science
(2016)
12 antennas MkII PAF

1000 sq deg
Rms=30 μJy
0.5 million galaxies

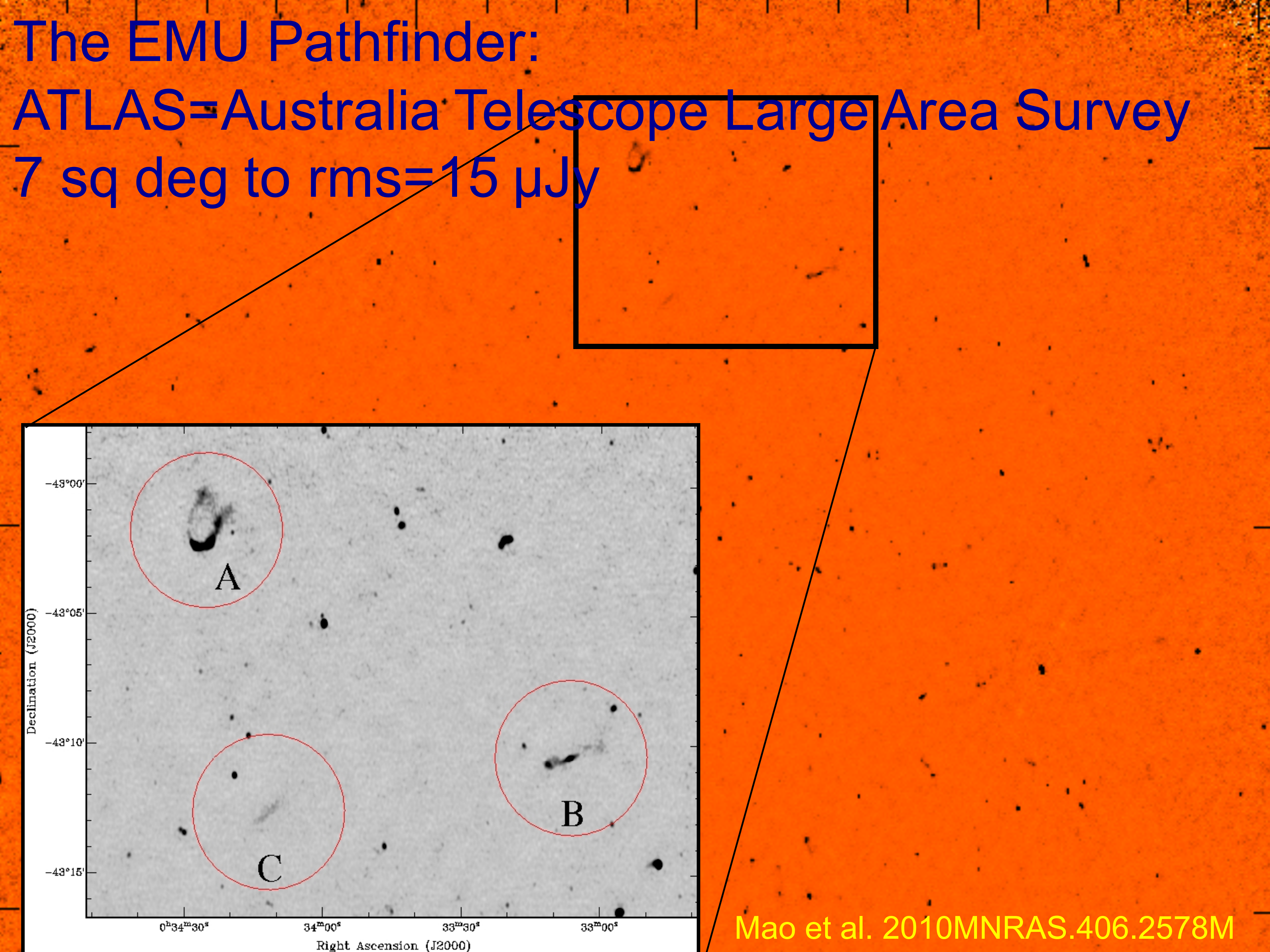
ASKAP – EMU
(2017-2018)
30-36 antennas MkII PAF

3π sr
Rms=10 μJy
70 million galaxies

The EMU Pathfinder:

ATLAS=Australia Telescope Large Area Survey

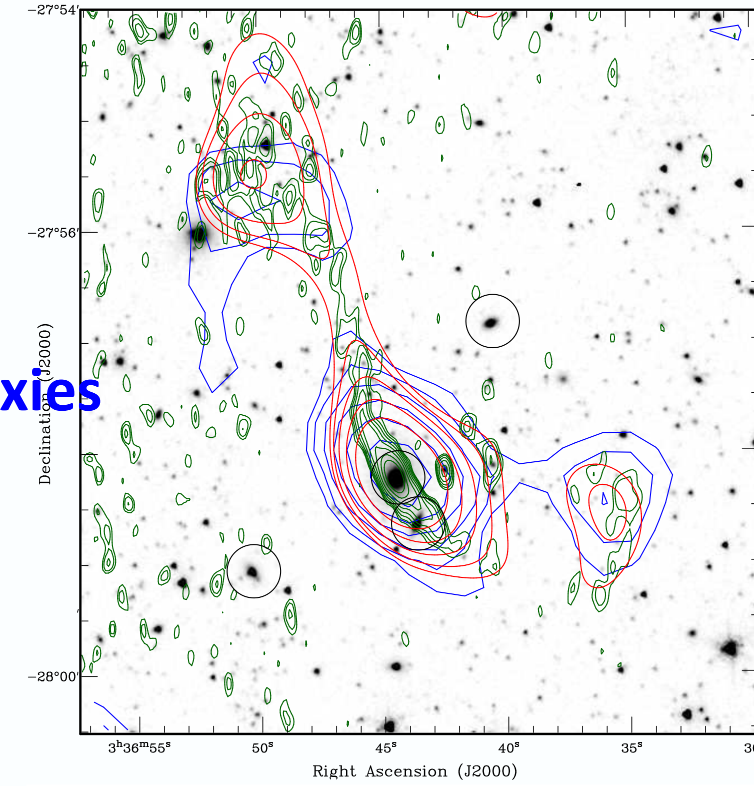
7 sq deg to rms=15 μJy



Mao et al. 2010MNRAS.406.2578M

KSP: Clusters

- In 4 sq deg of the ATLAS CDFS field, we find
  - 44 clusters via tailed galaxies (up to z~2)
  - 1 relic
  - 2 putative haloes
- Scaling this up to EMU…
  - 300,000 clusters detected via tailed galaxies
  - (maybe) 60,000 haloes
- c.f. eRosita ~100,000 expected

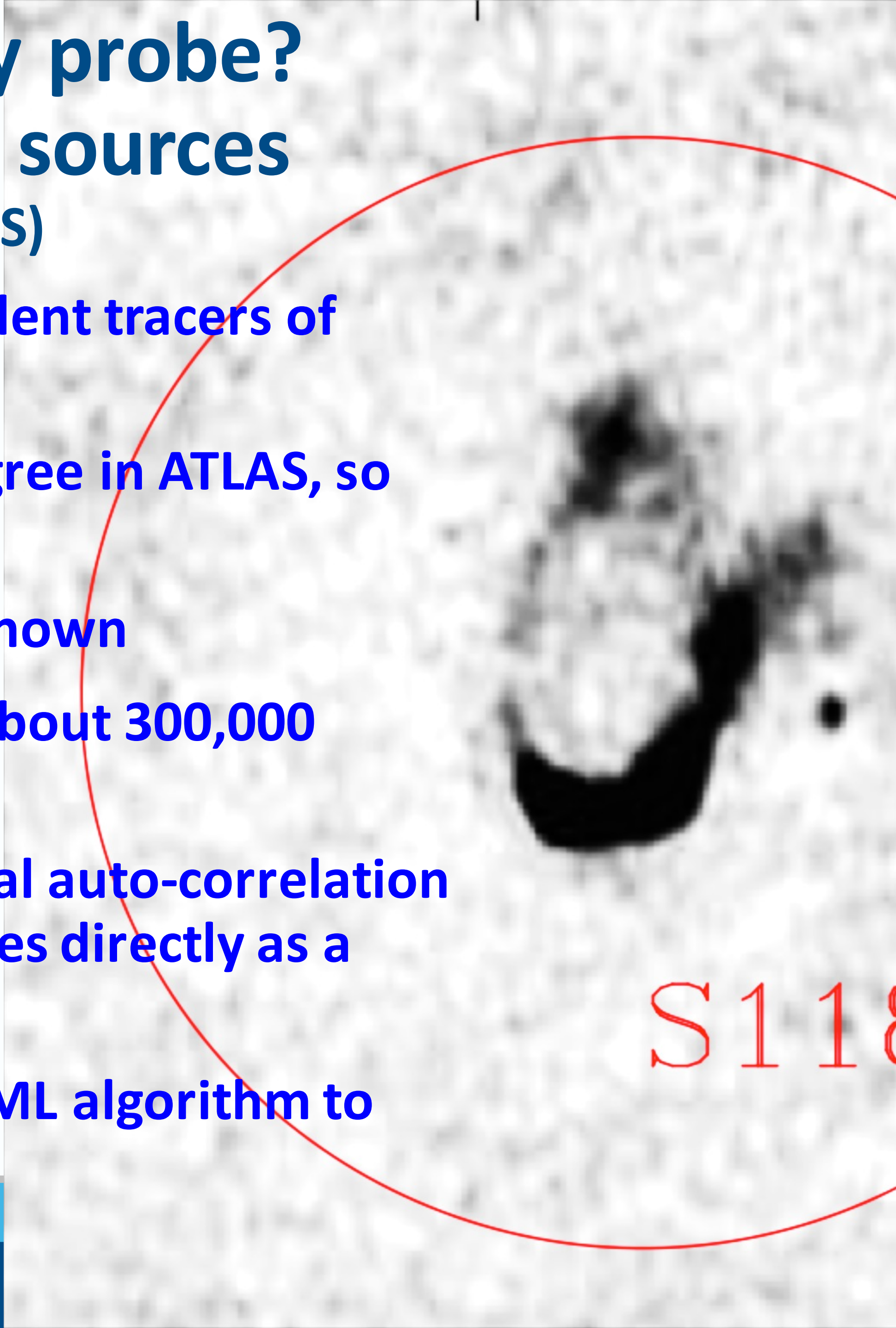


See Andrew O'Brien talk on Thursday

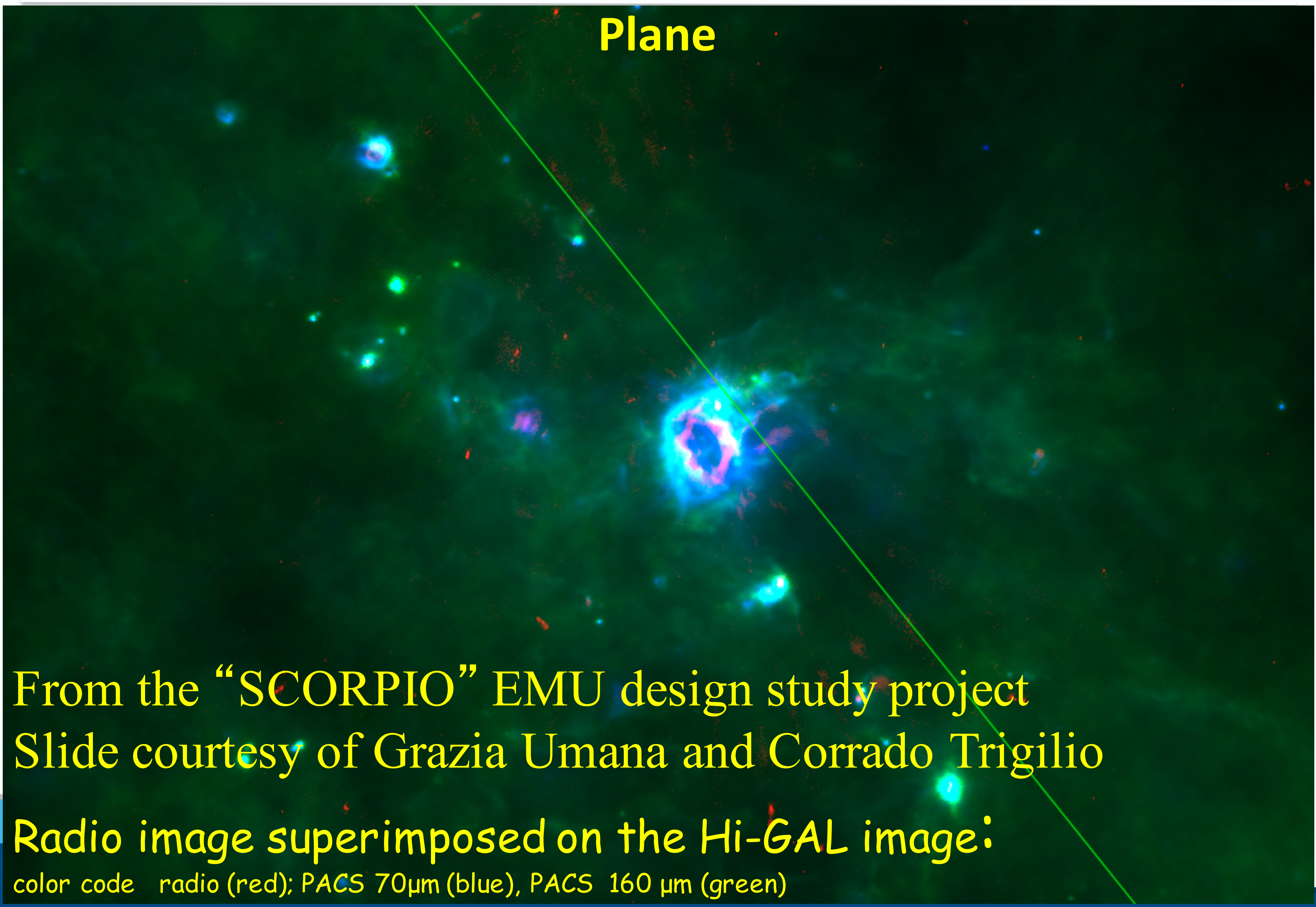
A fifth EMU cosmology probe? Finding bent-tail radio sources

(suggested by Bruce Bassett, AIMS)

- Bent-tail radio sources are excellent tracers of clusters
- We see about 10 per square degree in ATLAS, so expect ~ 300,000 in EMU
- c.f. ~a few thousand currently known
- Also eRosita expects to detect about 300,000 clusters in Xray
- (maybe) can also measure spatial auto-correlation function of bent-tail radio sources directly as a cosmological probe
- Should be fairly easy to build a ML algorithm to detect them



KSP: The deepest, highest resolution atlas yet of the Galactic Plane



From the "SCORPIO" EMU design study project

Slide courtesy of Grazia Umana and Corrado Trigilio

Radio image superimposed on the Hi-GAL image:

color code radio (red); PACS 70μm (blue), PACS 160 μm (green)

EMU Key Science Projects

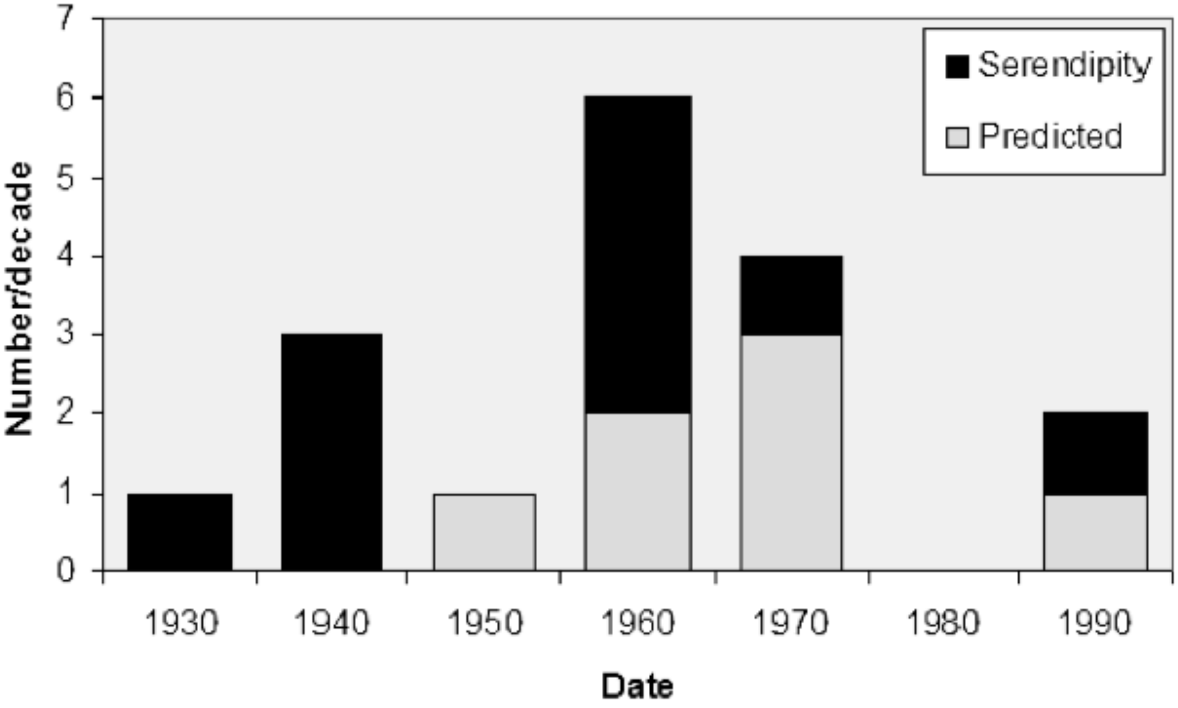
	Project Leaders
EMU Value-Added Catalogue	Nick Seymour
Characterising the Radio Sky	Ian Heyw
EMU Cosmology	Davi
Cosmic Web	
Clusters of Galaxies	le Johnston-Hollitt Chiara Ferrari
Cosmic star formation history	Andrew Hopkins
Radio-loud AGN	Anna Kapinska
Radio AGN in the EoR	Jose Afonso
Radio-quiet AGN	Isabella Prandoni
Local Univers	Josh Marvil, Michael Brown
The Gal le	Roland Kothes
SCO dio Stars	Grazia Umana
ning Data for the Unexpected	Ray Norris

How did galaxies form and evolve?

CSIRO

KSP: Mining large surveys for the unexpected

What fraction of discoveries in astronomy were "Popperian"?



(b) Predicted v Serendipity

+1 for dark energy (2012)

Serendipity:11
Predicted: 7

From Ekers (2009) PoS(sps5)007

Discoveries with HST

Project	Key project	Planned?	Nat. Geo. top ten?	Highly cited?	Nobel prize?
Use Cepheids to improve value of H0	✓	✓	✓	✓	
study intergalactic medium with uv spectroscopy	✓	✓			
Medium-deep survey	✓	✓			
Image quasar host galaxies		✓	✓		
Measure SMBH masses		✓	✓		
Exoplanet atmospheres		✓	✓		
Planetary Nebulae		✓	✓		
Discover Dark Energy			✓	✓	✓
Comet Shoemaker-Levy			✓		
Deep fields (HDF, HDFS, UDF, FF, etc)			✓	✓	
Proplyds in Orion			✓		
GRB Hosts			✓		

from Norris et al. 2013: arXiv1210.7521

Discoveries with HST (see e.g. Lallo: *arXiv:1203.0002)*

Project	Key	Planned?	Nat.	Highly	Nobel prize?
Use Cepheids to					
study intergalac uv spectroscopy					
Medium-deep s					
Image quasar hc					
Measure SMBH					
Exoplanet atmos					
Planetary Nebul					
Discover Dark E					✓
Comet Shoemak					
Deep fields (HDF					
Proplyds in Orio			✓		
GRB Hosts			✓		

Summary:

Of the "top ten" HST discoveries:

- **1 was a key project**
- **4 were planned by astronomers but were not key projects**
- **5 were totally unexpected (e.g. dark energy)**

CSIRO

The discovery of pulsars

Jocelyn Bell:

- explored a new area of observational phase space
- knew the instrument well enough to distinguish interference from signal
- observant enough to recognise a sidereal signature
- open minded – prepared for discovery
- within a supportive environment
- persistent

See Bell-Burnell (2009) PoS(sps5)014 for a personal perspective

Could Jocelyn Bell Discover the Unexpected in ASKAP data?

- Data volumes are huge – cannot sift by eye
- Instrument is complex – no single individual will be familiar with all possible artifacts
- ASKAP will be superb at answering well-defined questions (the "known unknowns")
- Humans won't be able to find the "unknown unknowns"
- Can we mine data for the unexpected, by rejecting the expected?

If not, ASKAP will not reach its full potential
i.e. it will not deliver value for money

What does ASKAP need to do to discover the unexpected?

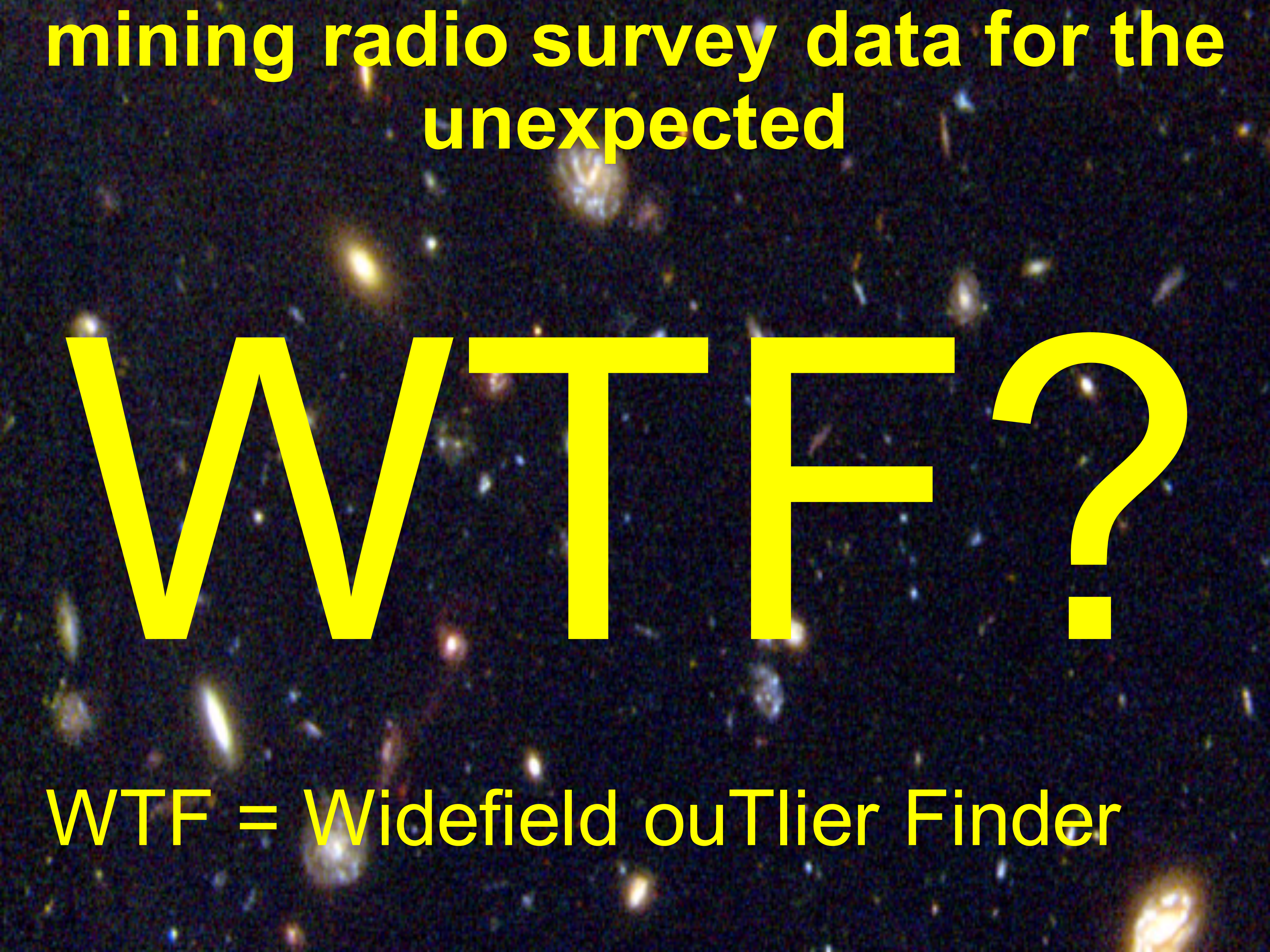
- **Maximise the volume of new phase space**
  - A good surrogate is to use # of known objects
  - Maximised by an all-sky survey

- **Retain flexibility**
  - don't optimise the telescope ONLY for science goals

- **Develop data mining software to search for the unexpected**

- This will be an important part of data-intensive research

mining radio survey data for the unexpected



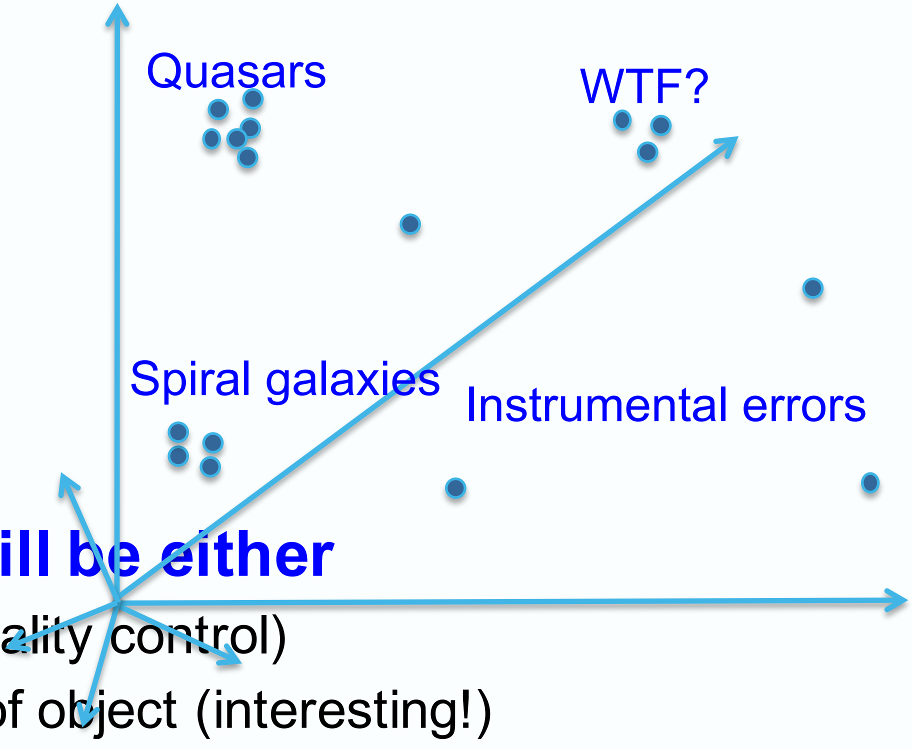
WTF?

WTF = Widefield ouTlier Finder

Mining large data sets for the unexpected

WTF will work by searching the n-dimensional (large n) phase space of observables, using techniques such as

- Decision tree approach
- Zoo approach
- Cluster analysis
- k-nearest-neighbours
- self-organised maps
- Bayesian approach to combine all the above



Identified objects/regions will be either

- processing artifacts (important for quality control)
- statistical outliers of known classes of object (interesting!)
- New classes of object (WTF)

WTF Phase 1a (July-October 2015)

- **Received a grant from Amazon Web Services to develop WTF on the AWS cloud platform**
- **Goals:**
  - **Implement WTF**
  - **Evaluate AWS platform as a collaborative research environment**
- **Approach**
  - **Set up challenges consisting of data (images or tables) with embedded “EMU eggs”**
  - **Data include both simulations and real data**
  - **Invite ML and other algorithm groups to discover the EMU eggs**
  - **Develop visualisation tools to understand the process and data**

WTF Phase 2 (January-June 2016)

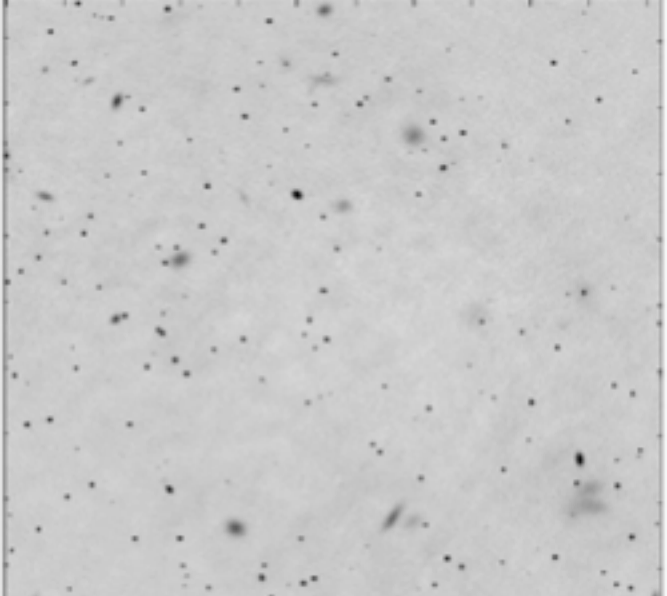
- Once we have the infrastructure set up, and have tested the process using in-house challengers, we will open it up to all interested groups
- Outcomes:
  - Test different approaches and algorithms, to see which are best at discovering WTFs
  - Perhaps even make a real discovery on the real data!

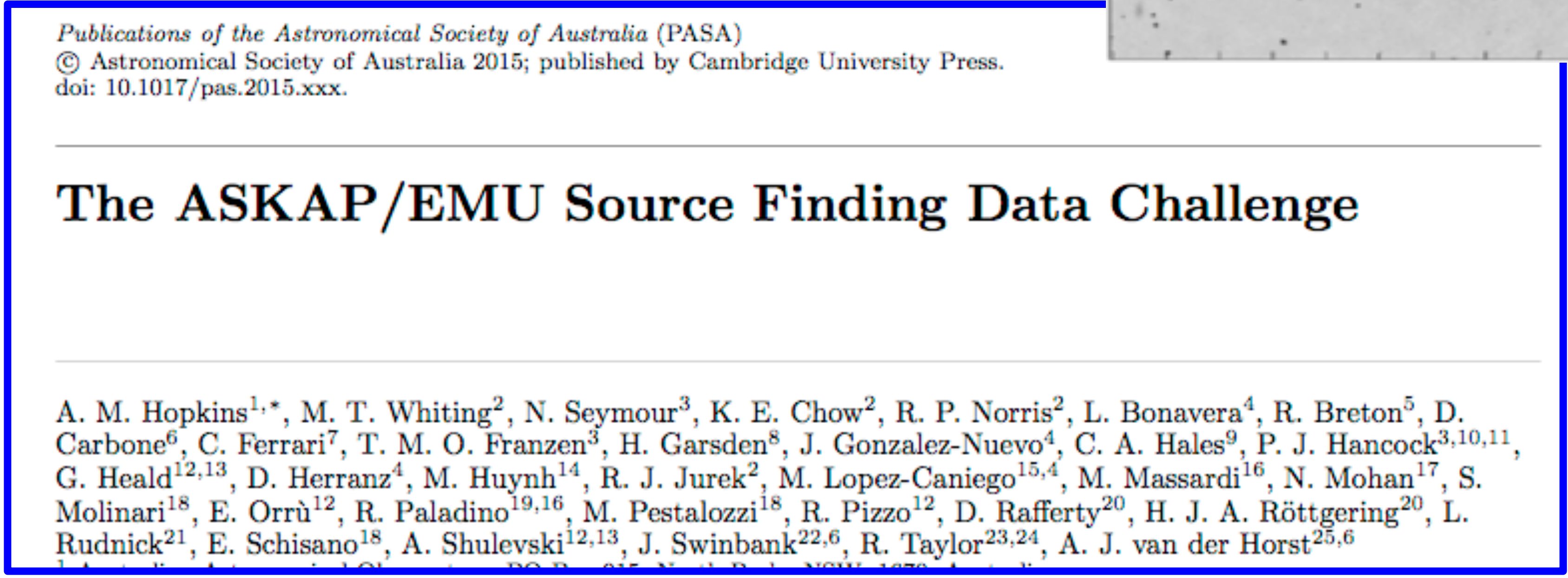
Other potential EMU machine learning projects

EMU Data Science Challenges: 1. Compact Source Extraction

Initial data challenge study showed existing source extraction algorithms not up to the job.

Need to build on this, identify problems, using fix existing algorithms or develop a better one.





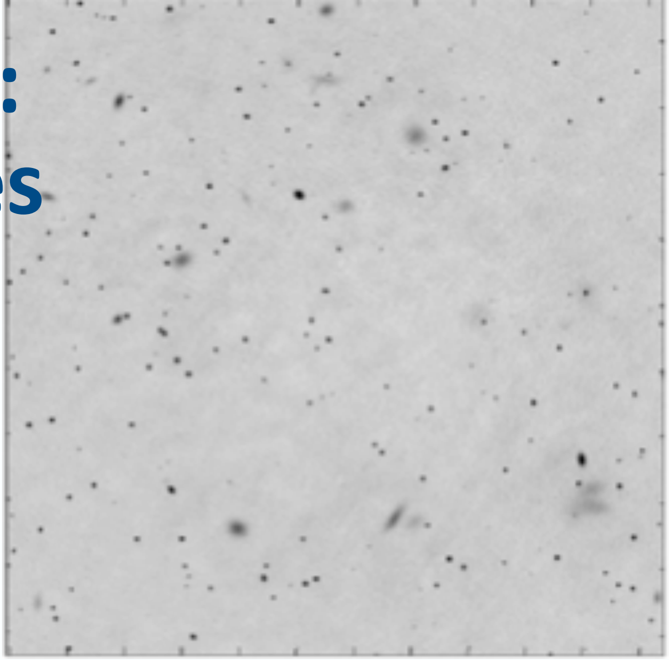
Publications of the Astronomical Society of Australia (PASA)
© Astronomical Society of Australia 2015; published by Cambridge University Press.
doi: 10.1017/pas.2015.xxx.

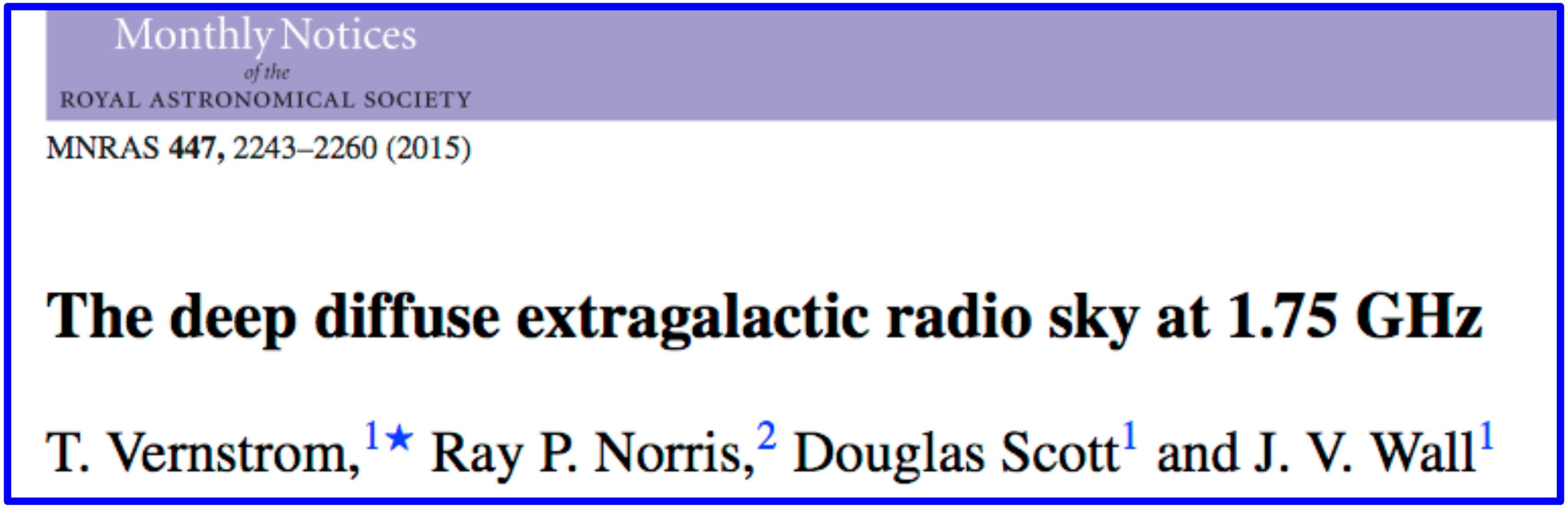
The ASKAP/EMU Source Finding Data Challenge

A. M. Hopkins[1,*], M. T. Whiting[2], N. Seymour[3], K. E. Chow[2], R. P. Norris[2], L. Bonavera[4], R. Breton[5], D. Carbone[6], C. Ferrari[7], T. M. O. Franzen[3], H. Garsden[8], J. Gonzalez-Nuevo[4], C. A. Hales[9], P. J. Hancock[3,10,11], G. Heald[12,13], D. Herranz[4], M. Huynh[14], R. J. Jurek[2], M. Lopez-Caniego[15,4], M. Massardi[16], N. Mohan[17], S. Molinari[18], E. Orrù[12], R. Paladino[19,16], M. Pestalozzi[18], R. Pizzo[12], D. Rafferty[20], H. J. A. Röttgering[20], L. Rudnick[21], E. Schisano[18], A. Shulevski[12,13], J. Swinbank[22,6], R. Taylor[23,24], A. J. van der Horst[25,6]

EMU Data Science Challenges:
2. Extraction of diffuse sources

- No existing algorithm can routinely extract diffuse sources
- A number of algorithms in development (and have been for years!)
- We need this soon for ATLAS-SPT!





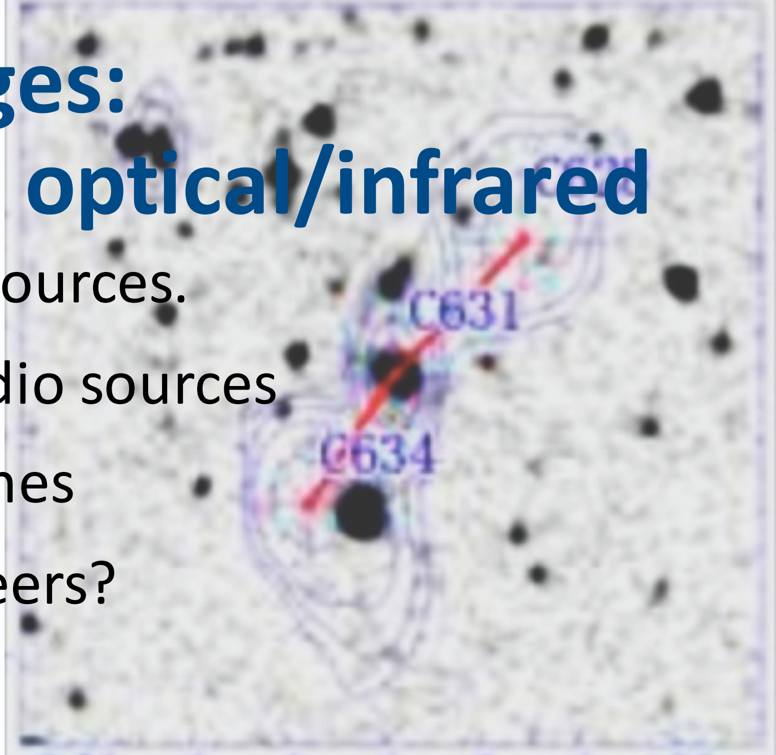
Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **447,** 2243–2260 (2015)

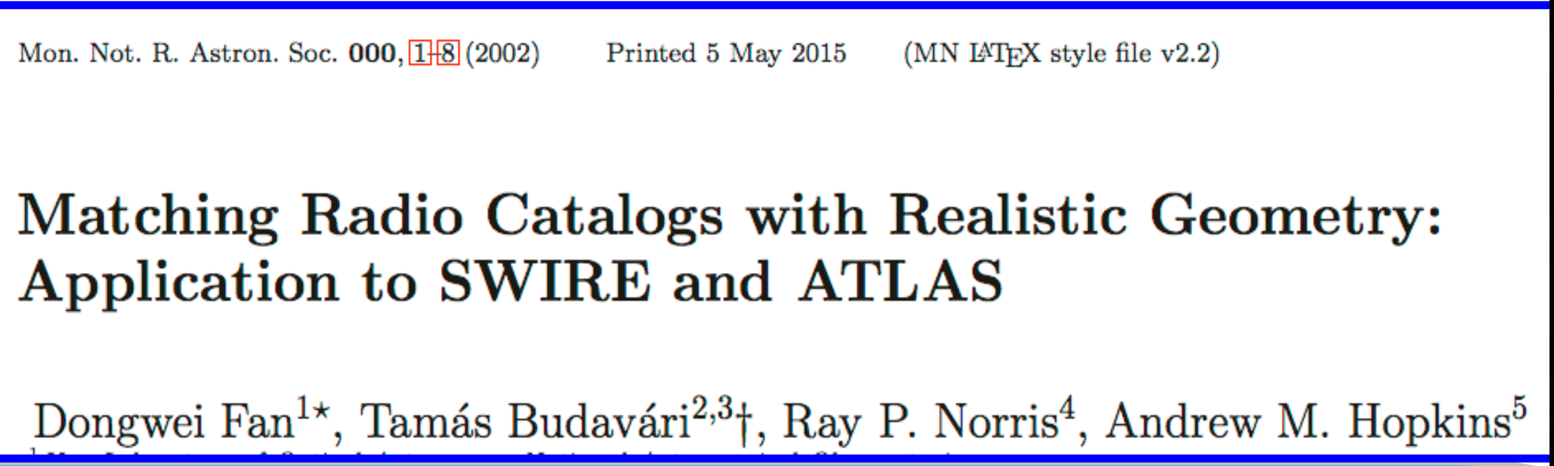
The deep diffuse extragalactic radio sky at 1.75 GHz

T. Vernstrom,[1★] Ray P. Norris,[2] Douglas Scott[1] and J. V. Wall[1]

EMU Data Science Challenges: 3. Cross-identification with optical/infrared



- Existing algorithms work well with point sources.
- They fail badly on typical core-jet-lobe radio sources
- Currently exploring a number of approaches
- Plan to start a ML approach – any volunteers?



Mon. Not. R. Astron. Soc. **000**, 1–8 (2002)   Printed 5 May 2015   (MN LaTeX style file v2.2)

Matching Radio Catalogs with Realistic Geometry: Application to SWIRE and ATLAS

Dongwei Fan[1]*, Tamás Budavári[2,3]†, Ray P. Norris[4], Andrew M. Hopkins[5]

Would you like to be part of EMU?
Email me on Ray.Norris@csiro.au