

ASTERICS & Work Flows within the SRC

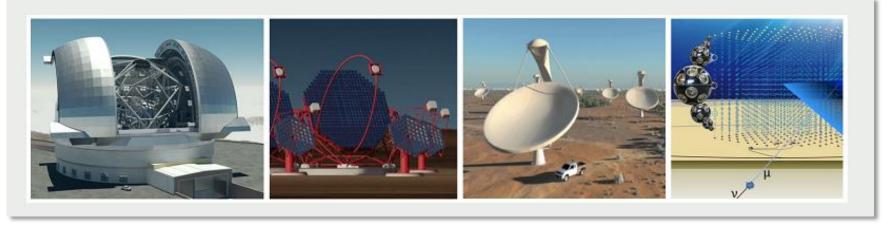
Paul Alexander



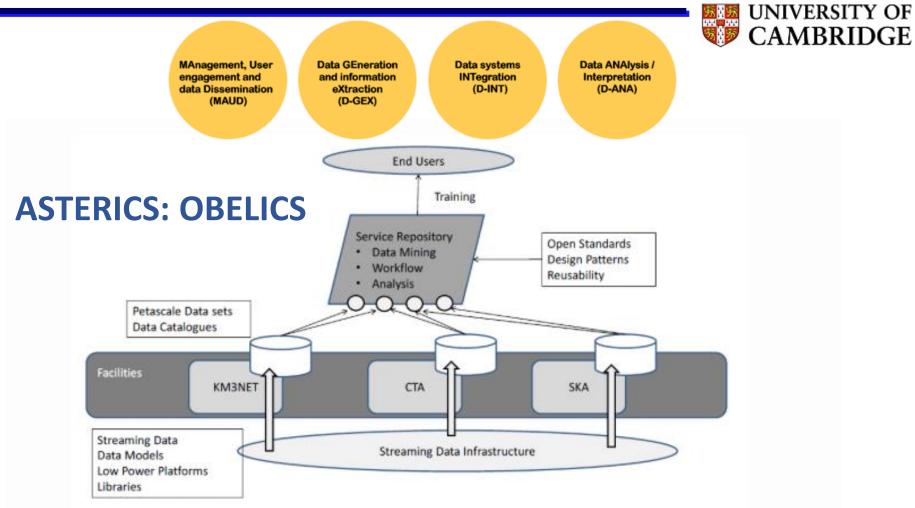
ASTERICS

ASTERICS:

Support and accelerate the implementation of the ESFRI telescopes, to enhance their performance beyond the current state-of-the-art, and to see them interoperate as an integrated, multi-wavelength and multi-messenger facility



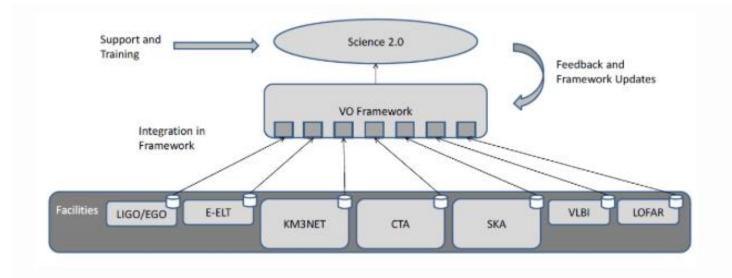
E-ELT CTA SKA KM3NET



- Maximise software re-use and co-development of technology for Petascale data
- Adapt and optimise extremely large database systems to fulfil the requirements of the ASTERICS ESFRI projects.
- Study and demonstrate data integration across ASTERICS ESFRI and pathfinder projects using data mining tools and statistical analysis techniques on Petascale data sets.



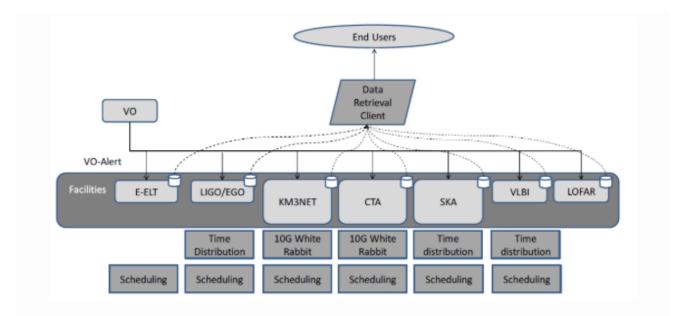
ASTERICS: DADI



 Adapt the VO framework and tools to the ESFRI project needs, and make sure European astronomers remain lead actors in the IVOA, influencing it in the interest of the European infrastructures and the European scientific community



ASTERICS: CLEOPATRA



- Develop technology for the enabling of long-haul and many-element time and frequency distribution over fibre connections.
- Develop methods for relaying alerts, which will signal transient event detections between the facilities and enable joint observing programmes, including scientific strategies and methods for joint observing.
- Foster the development of advanced scheduling algorithms, using AI approaches for optimal usage of the ESFRI facilities. CLEOPATRA's tasks reflect a consistent set of enhancements of the facilities based on developments in connectivity and data transport.

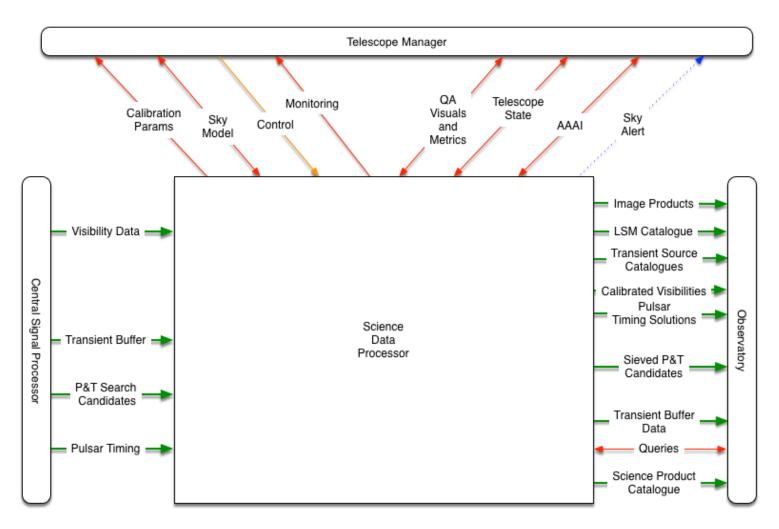


Data Products: From Science Case

	SKA1	SKA2	
The Cradle of Life & Astrobiology	Proto-planetary disks; imaging inside the snow/ice line (@ < 100pc), Searches for amino acids.	Proto-planetary disks; sub-AU imaging (@ < 150 pc), Studies of amino acids.	
The oracle of the d Astronology	Targeted SETI: airport radar 10^4 nearby stars.	Ultra-sensitive SETI: airport radar 10^5 nearby star, TV ~10 stars.	
Strong-field Tests of Gravity with	1st detection of nHz-stochastic gravitational wave background.	Gravitational wave astronomy of discrete sources: constraining galaxy evolution, cosmological GWs and cosmic strings.	
Pulsars and Black Holes	Discover and use NS-NS and PSR-BH binaries to provide the best tests of gravity theories and General Relativity.	Find all ~40,000 visible pulsars in the Galaxy, use the most relativistic systems to test cosmic censorship and the no-hair theorem.	
he Origin and Evolution of Cosmic	The role of magnetism from sub-galactic to Cosmic Web scales, the RM-grid @ 300/deg2.	The origin and amplification of cosmic magnetic fields, the RM-grid @ 5000/deg2.	
Magnetism	Faraday tomography of extended sources, 100pc resolution at 14Mpc, 1 kpc @ z ≈ 0.04.	Faraday tomography of extended sources, 100pc resolution at 50Mpc, 1 kpc @ z ≈ 0.13.	
Galaxy Evolution probed by Neutral	Gas properties of 10^7 galaxies, <z> ≈ 0.3, evolution to z ≈ 1, BAO complement to Euclid.</z>	Gas properties of 10^9 galaxies, <z> ≈ 1, evolution to z ≈ 5, world-class precision cosmology.</z>	
Hydrogen	Detailed interstellar medium of nearby galaxies (3 Mpc) at 50pc resolution, diffuse IGM down to N_H < 10^17 at 1 kpc.	Detailed interstellar medium of nearby galaxies (10 Mpc) at 50pc resolution, diffuse IGM down to N_H < 10^17 at 1 kpc.	
The Transient Radio Sky	Use fast radio bursts to uncover the missing "normal" matter in the universe.	Fast radio bursts as unique probes of fundamental cosmological parameters and intergalactic magnetic fields.	
	Study feedback from the most energetic cosmic explosions and the disruption of stars by super-massive black holes.	Exploring the unknown: new exotic astrophysical phenomena in discovery phase space.	
alaxy Evolution probed in the Radio	Star formation rates (10 M_Sun/yr to z ~ 4).	Star formation rates (10 M_Sun/yr to z ~ 10).	
Continuum	Resolved star formation astrophysics (sub-kpc active regions at z ~ 1).	Resolved star formation astrophysics (sub- kpc active regions at z ~ 6).	
	Constraints on DE, modified gravity, the distribution & evolution of matter on super- horizon scales: competitive to Euclid.	Constraints on DE, modified gravity, the distribution & evolution of matter on super- horizon scales: redefines state-of-art.	
Cosmology & Dark Energy	Primordial non-Gaussianity and the matter dipole: 2x Euclid.	Primordial non-Gaussianity and the matter dipole: 10x Euclid.	
Cosmic Dawn and the Epoch of	Direct imaging of EoR structures (z = 6 - 12).	Direct imaging of Cosmic Dawn structures (z = 12 - 30).	
Reionization	Power spectra of Cosmic Dawn down to arcmin scales, possible imaging at 10 arcmin.	First glimpse of the Dark Ages (z > 30).	



Scope of the SKA Science Data Processor (SDP)





SKA Data Products

Table 1: SDP Data Products

Data Product	Description (For each product a QA and Processing Log will also be maintained)				
Image Products 1: Image Cubes	 Imaging data for Continuum, as cleaned restored Taylor term images (n.b. no image products for Slow Transients detection have been specified – maps are made, searched and discarded) Residual image (i.e. residuals after applying CLEAN) in continuum Clean component image (or a table, which could be smaller). Spectral line cube after continuum subtracted Residual spectral line image (i.e. residuals after clean applied) Representative Point Spread Function for observations (cutout, small in size compared to the field of view (FOV)) 				
Image Products 2: UV-grids	 Calibrated visibilities, gridded at the spatial and frequency resolution required by the experiment. One grid per facet (so this grid is the FFT of the dirty map of each facet). c.f. ECP150007 [AD02] Accumulated Weights for each uv cell in each grid (without additional weighting applied). 				
Calibrated Visibilities	Calibrated visibility data (for example for EoR experiments) and direction-dependent calibration information, with time and frequency averaging performed as requested to reduce the data volume.				
LSM Catalogue	Catalogue of a subset of the Global Sky Model (GSM) containing the sources relevant for the scheduling block being processed. These are the sources in the FOV, as well as, potentially, strong sources outside of the current FOV. Initially, the LSM is filled from the GSM; during the data processing the sources found in the images are added to the LSM.				

Transient Source Catalogue	Time-ordered catalogue of candidate transient objects pertaining to each detection alert from the real-time, so-called, Fast Imaging.
Pulsar Timing Solutions	For each of the observed pulsars the output data from the pulsar timing section will include the original input data as well as averaged versions of these data products (either averaged in polarisation, frequency or time) in PSRFITs format. The arrival time of the pulse. The residuals from the current best-fit model for the pulsar. An updated model of the arrival times.
Transient Buffer Data	Voltage data passed through from the CSP when the transient buffer is triggered.
Sieved Pulsar and Transient Candidates	A data cube which will be folded and dedispersed at the best Dispersion Measure (DM), period and period derivative determined from the search. A single ranked list of non-imaging transient candidates from each scheduling block. For those transients deemed of sufficient interest, the associated "filterbank" data will also be archived. A set of diagnostics/heuristics that will include metadata associated with the scheduling block and observation. If a sufficiently interesting pulsar is discovered this will generate an alert as well as being recorded in a Log. (While we have a requirement to report single pulse events, it is not clear whether we have to provide alerts for anything other than single pulses. This is being referred to the Telescope Teams.)
Science Alerts Catalogue	Catalogue of Science Alerts produced and communicated by the SDP. The alerts themselves are IVOA alerts; this catalogue provides a searchable and retrievable record of past alerts.
Science Product Catalogue	A database relating to all Science Products processed by the SDP. It includes associated scientific metadata that can be queried and searched and includes all information so that the result of a query can lead to the delivery of data.

Image-based types of Analysis

General Image Processing

- Visualisation
- Noise estimation
- Basic image operations, weighted averaging
- Filtering and masking
- Deconvolution and smoothing
- Standard image-based feature detection and extraction

Specialised Image Processing

- Source Detection in noisy data
- Faraday and rotation-measure imaging

• Feature detection Classification

- Model fitting (simple or statistically robust)
- Representation in a new basis (wavelets, compressed sensing, ...)
- Direct machine learning classification in image space
- High-dimensional classification prior to scientific statistical analysis
 - E.g. based on image features + spectral features + polarization features

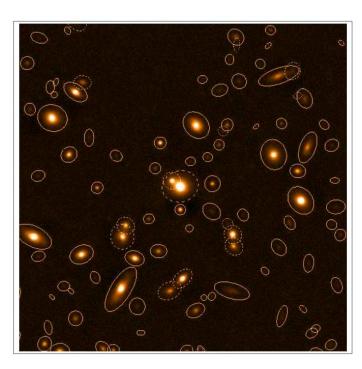
Multi-dimensional image data size up to 50k x 50k x 64 k

Spatial Spectral / velocity Polarization



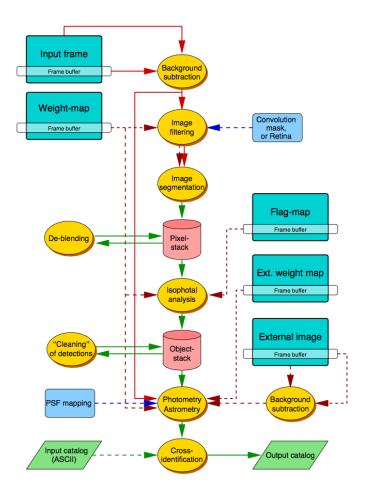


Source detection and feature extraction



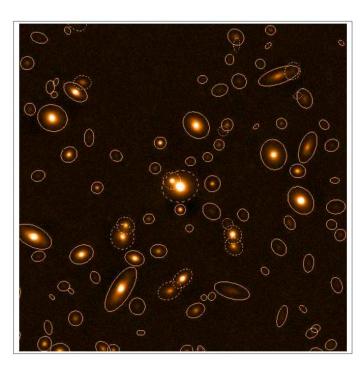
Sextractor very widely used algorithm Key aspects of algorithm

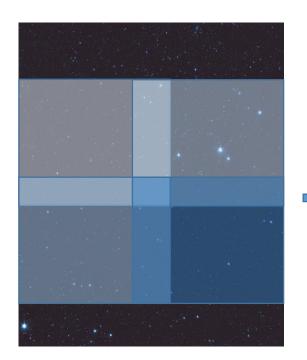
- Filtering
- Noise measurement
- Segmentation
- Object identification
- Classification





Source detection and feature extraction







Sextractor very widely used algorithm Key aspects of algorithm

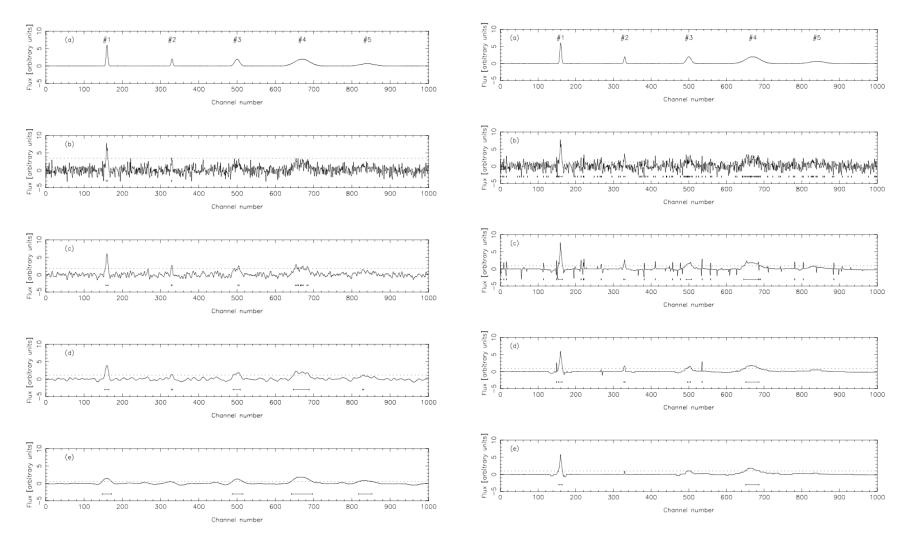
- Filtering
- Noise measurement
- Segmentation
- Object identification
- Classification

Embarrassingly parallel

- Distribute with simple guard region
- Simple extension to multi dimensions



Extension to multi-scale: Duchamp

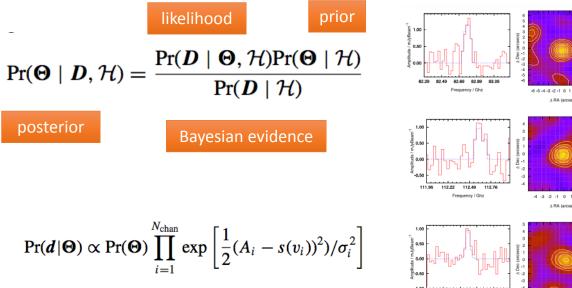


Multi-scale smoothing

Wavelet representation



Full Bayesian Source Finding



Finding galaxies in Alma fields

Frequency / Ghz

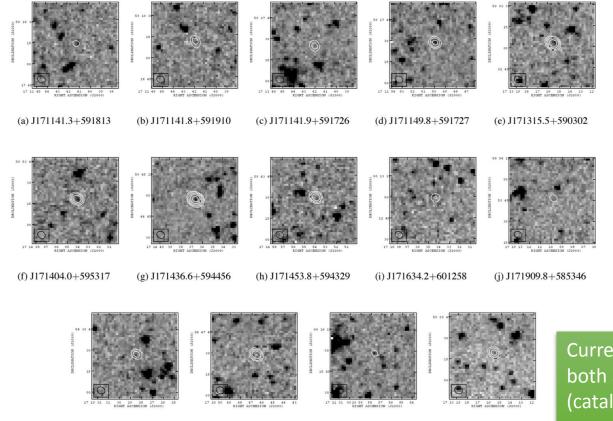
Full Bayesian analysis

- Model for the source simple or complex
- Compare evidence is there evidence of 1, 2, ... N sources
- Quantifies detectability
- Can get very computationally challenging
- Image-space parallel
- Visibility space very computationally expensive

Building source detection as part of deconvolution step using Bayesian / MEM approach



Comparison between wavebands



(k) J171928.4+583927

(l) J172046.5+584729

(m) J172057.6+601558

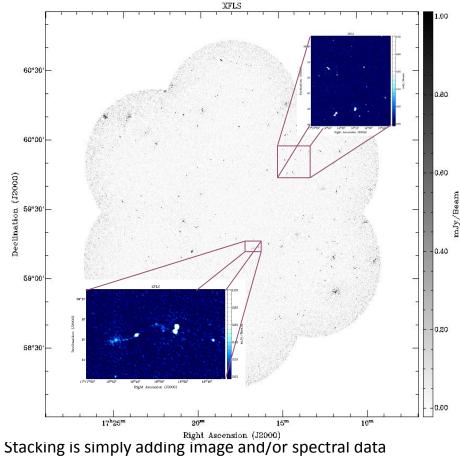
(n) J172315.5+590919

Currently source detection in both and compare lists (catalogues)

ASTERICS task is to develop joint Bayesian analysis

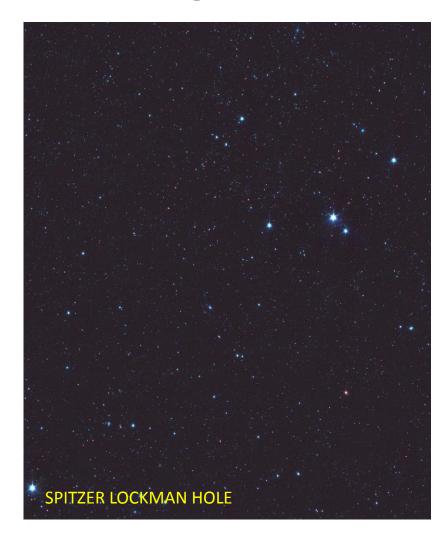


Comparison between wavebands: Stacking



Stacking is simply adding image and/or spectral dat referenced to give better sensitivity

- E.g. emission detected in one waveband for a class of object and not in another
- Use positions from detected band for referencing





Comparison between wavebands: Stacking

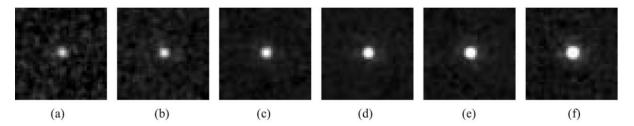


Figure 2. Median stacked 1.4-GHz radio images for the faintest six flux density bins. All images have a size of $61.5 \times 61.5 \operatorname{arcsec}^2 (41 \times 41 \operatorname{pixel})$. The 24-µm flux density range of sources used to create the stacked image are given below each sub-image, with further details in Table 2. The grey-scale ranges between -2 and 20 µJy beam⁻¹. (a) 150–201 µJy. (b) 201–268 µJy. (c) 268–359 µJy. (d) 359–479 µJy. (e) 479–641 µJy. (f) 641–857 µJy.

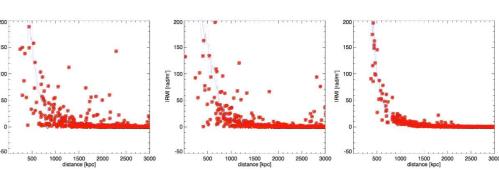
ASTERICS task is to develop Bayesian approach Extension of joint analysis

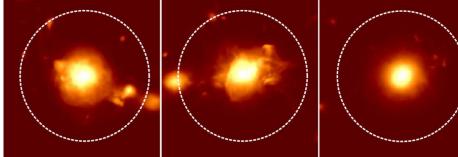
Stacking is simply adding image and/or spectral data referenced to give better sensitivity

- E.g. emission detected in one waveband for a class of object and not in another
- Use positions from detected band for referencing

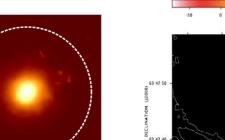


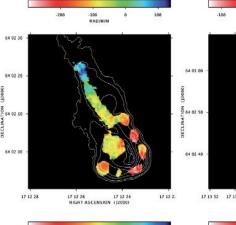
Rotation Measure Synthesis

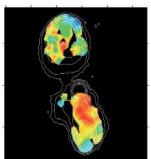






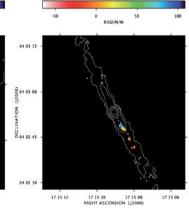






RAD/M/M





Govoni et al. 2006

17 13 17 17 13 16 RIGHT ASCENSION (J2000) 17 13 15

17 13 18

RAD/M/M

$$\tilde{P}(\lambda^2) = W(\lambda^2) \int_{-\infty}^{+\infty} F(\phi) e^{2i\phi\lambda^2} d\phi$$

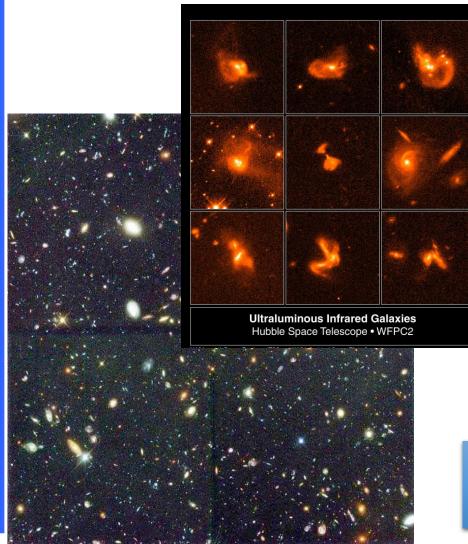
This can be inverted (note the addition of λ_0^2):

$$\tilde{F}(\phi) = K \int_{-\infty}^{+\infty} \tilde{P}(\lambda^2) e^{-2i\phi(\lambda^2 - \lambda_0^2)} d\lambda^2 = F(\phi) * R(\phi) \neq F(\phi)$$

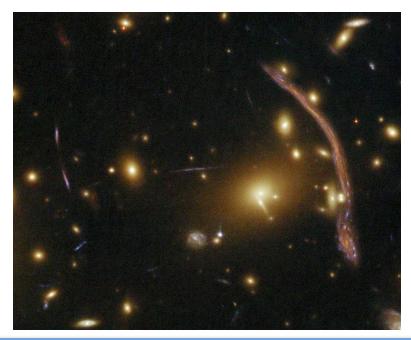
- Use RM Grid to probe evolution of field in clusters
- RM Synthesis gives much more information



Feature extraction and classification: galaxy types



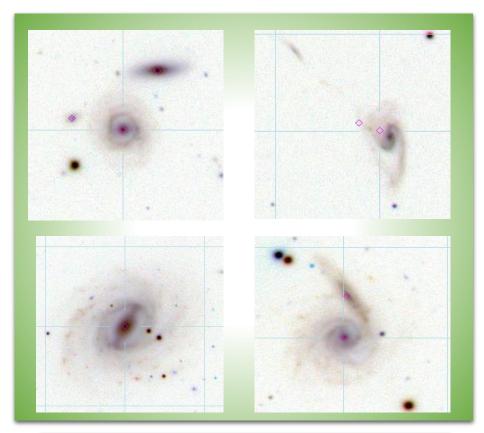
- Representation in a basis
- Machine learning
 - Deep neural nets
 - Gaussian processes



Demanding use case is weak lensing

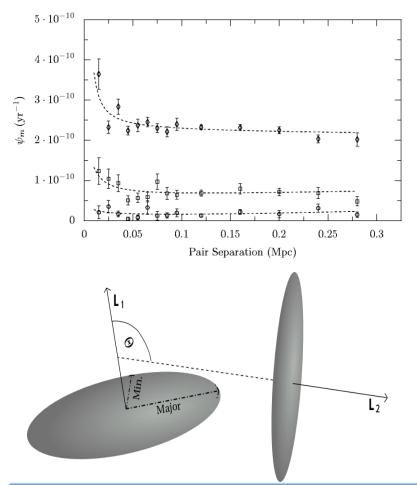
• Determine systematic background image distortion around a foreground gravitating mass

Feature extraction and classification: dynamics



Identify in image-velocity space physically close pairs

- Estimate dynamics from imaging
- Measure Star formation from spectroscopy or multi-waveband images

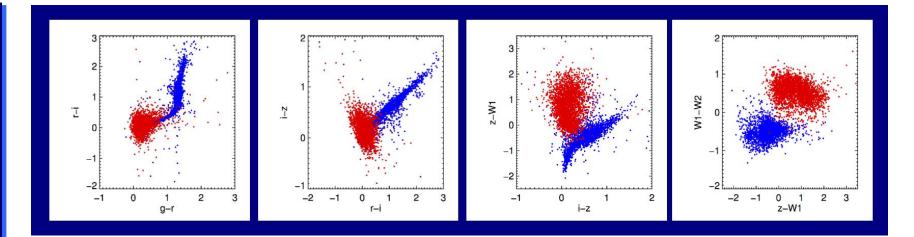


UNIVERSITY OF CAMBRIDGE

Star formation increased by spin-aligned interactions



Classification based on features



- Simple problems tackled via PCA
- Machine learning preferred approach
- Gaussian processes particularly suited to this sort of classification

Deep learning is potentially very computationally intensive

• Likely to be a major activity to extract science once initial analysis is done

Visibility-based types of Analysis

• Imaging

- Re-imaging from "visibility data" similar to SDP processing
- Re-imaging from gridded "visibility data"

Power-spectral estimation and model comparison

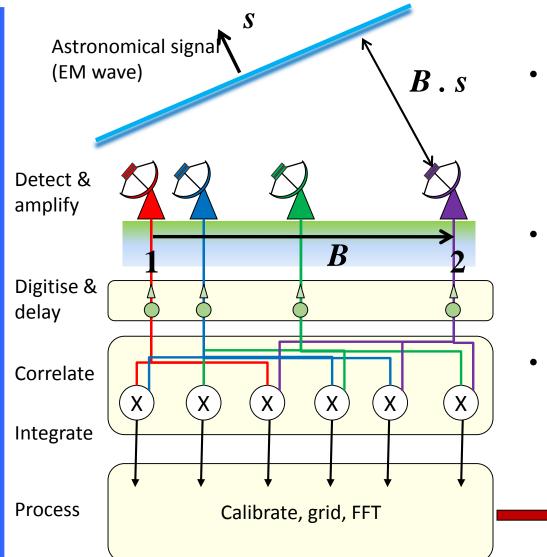
- Estimate information direct in the visibility space
- Model fitting
- Bayesian analysis

Extremely compute intensive Re-imaging needed in most demanding cases and learning about instrument





Imaging



Visibility:

$$V(B) = E_1 E_2^*$$

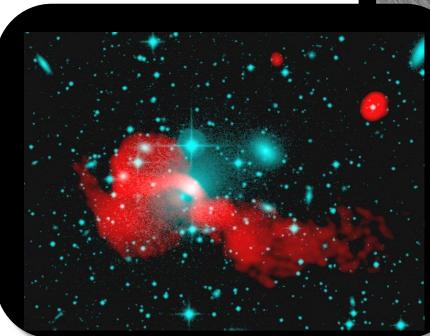
= $I(s) \exp(i \omega B \cdot s/c)$

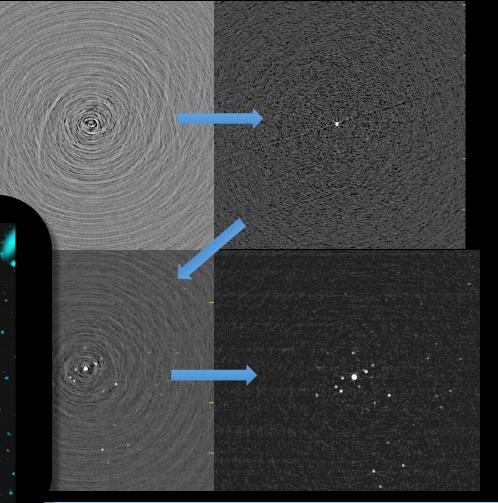
- Resolution determined by maximum baseline $\theta_{max} \sim \lambda / B_{max}$
- Field of View (FoV) determined by the size of each dish $\theta_{\rm dish} \sim \lambda$ / D



Imaging

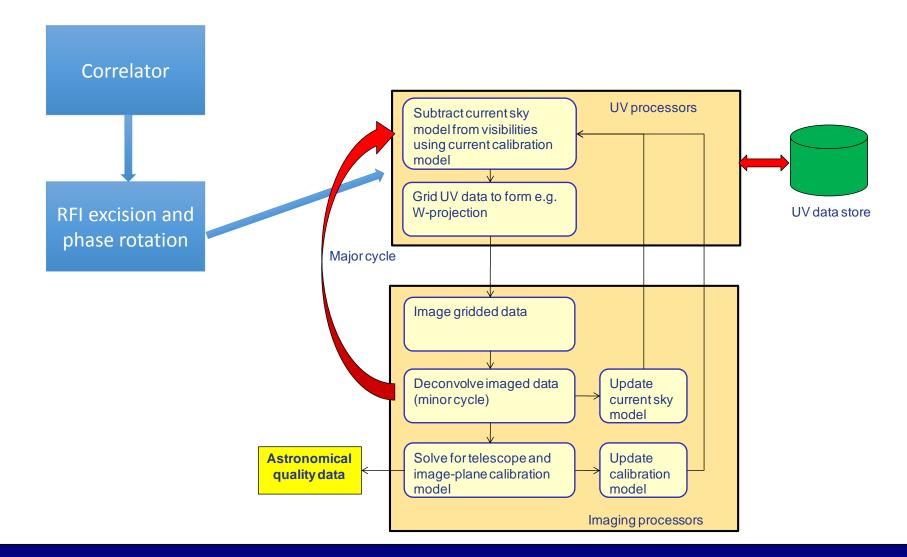
- Images formed by an inverse algorithm similar to that used in MRI
- Typical image sizes up to:
- 30k x 30k x 64k voxels

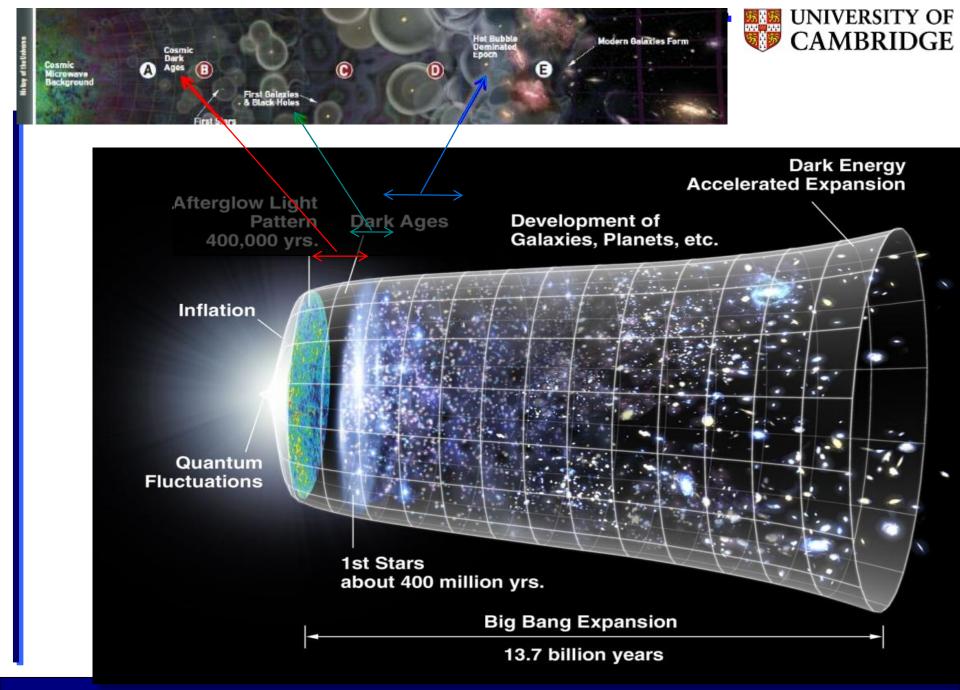


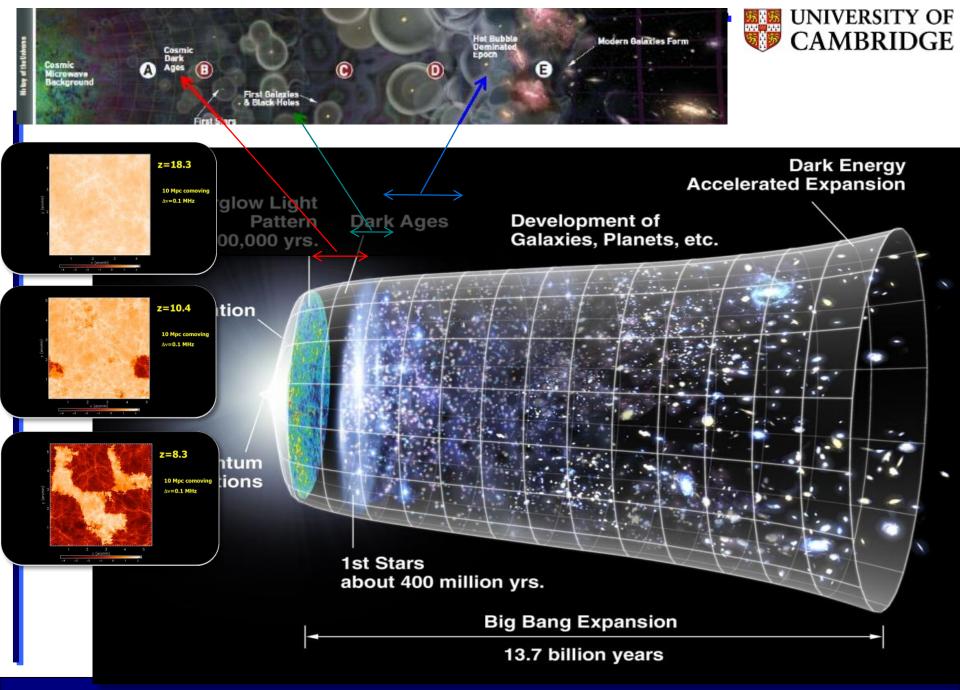


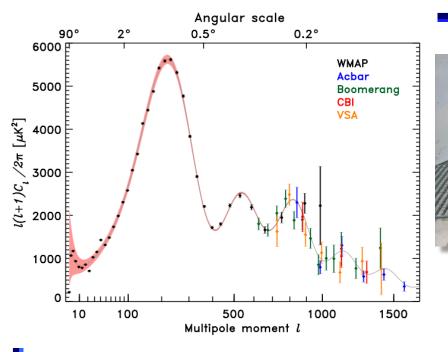


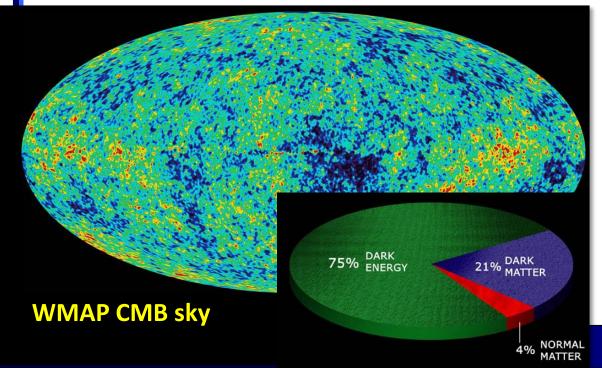
Imaging Workflow









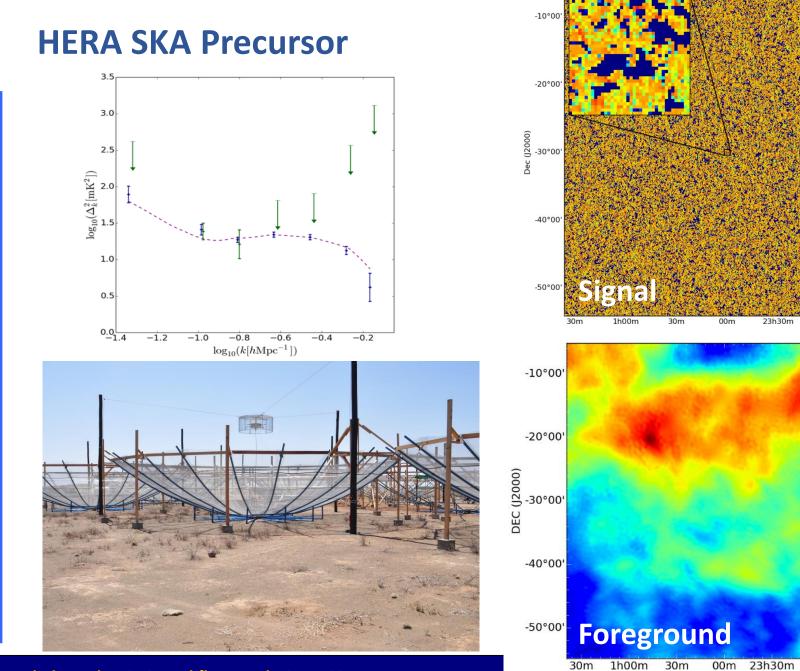




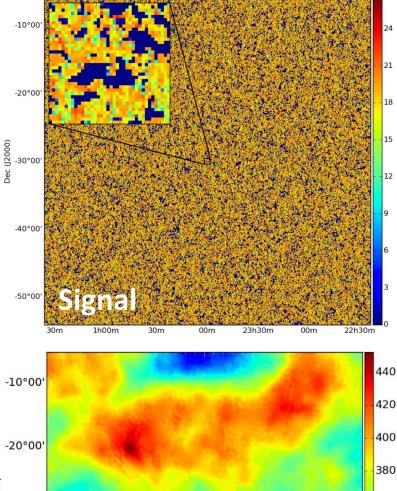
VSA experiment

Measurements of CMB powerspectrum have established era of "precision cosmology"

Universe has a flat geometry $\Omega = 1$ Matter contribution $\Omega_m \sim 0.25$ Normal (baryonic) matter $\Omega_B \sim 0.04$ Cold Dark Matter (CDM) $\Omega_m \sim 0.21$ Dark energy $\Omega_\Lambda \sim 0.75$



Paul Alexander: SRC Workflows and ASTERICS



RA (J2000)

360

340

320

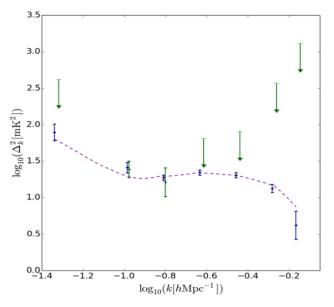
300

280

22h30m

00m

HERA SKA Precursor

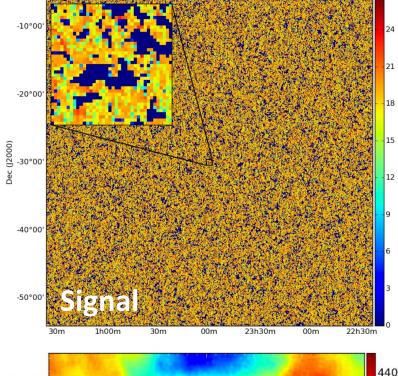


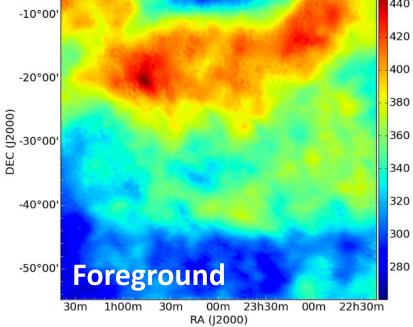
Work from visibility data

- Improve calibration via imaging-like workflow
- Solve for time varying phase screen
- Remove foreground emission

Direct Bayesian estimation of the power spectrum

- Model is a spherically symmetric
- Marginalise over residual foreground





Science analysis and data mining

• Science Analysis

- Visualisation
- Standard data analytics
- Model fitting including model generation

• Data mining

- Data-base like analytics on SKA and other data
- Importing data from multiple sources
- Context based searches
 - Users will expect to be able to do the same searches as they see on Google / Facebook
 - Find me all images including extended spiral larger than 100 arcsec ...

Total SKA Archive can grow at 300 Pbytes / yr

Even "Standard" analysis now challenging

Could be very challenging



Paul Alexander: SRC Workflows and ASTERICS

Experiment definition

Science Simulation

Simulation of science target space

• Experimental simulation

- Simulate experimental response
- Simulate analysis procedure
 - E.g. co-adding (averaging) many separate observations on a field → does the experiment need to average in image or UV-space?

How much simulation will need to be done at SRC?

Experimental simulation requires running a data generator plus the SDP environment as run at the observatory

$$\langle \mathbf{V}_{p,q} \rangle = \sum_{s} \mathbf{K}_{p,s} \mathbf{E}_{p,s} \mathbf{G}_{p,s} \mathbf{P}_{p,s} \langle \mathbf{B}_{s} \rangle \mathbf{P}_{q,s}^{H} \mathbf{G}_{q,s}^{H} \mathbf{E}_{q,s}^{H} \mathbf{K}_{q,s}^{H}$$





Possible SKA1 Continuum Survey Strategy

Deep / Multi-tier

- Star formation & BH accretion history
- Role of AGN feedback over cosmic Time
- Evolution of FIR-Radio correlation
- Role of environment

Wide / All Sky

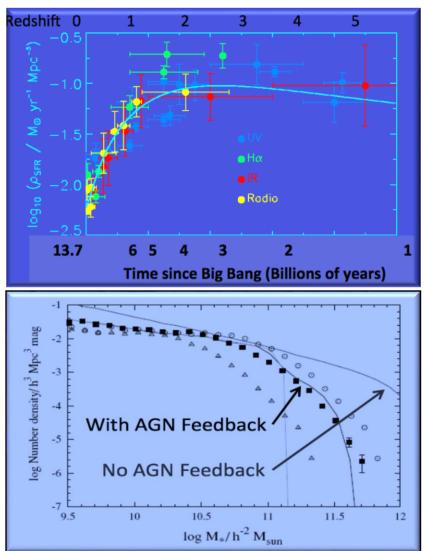
- First galaxies, BHs & protoclusters
- Galaxy clusters, cosmic web
- RL AGN physics/lifecycle
- RQ/RL AGN dichotomy
- ISM and SF physics in nearby galaxies
- Origin of FIR-Radio correlation
- Strong lensing

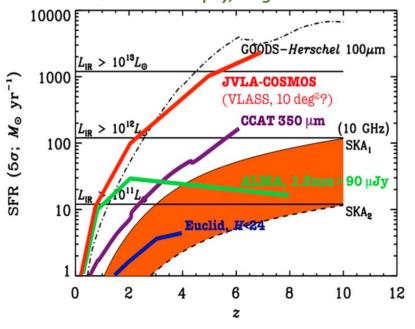
Science Drivers	Freq.	Tier	Rms (full BW)	Area	Res.	Science/ Commensality
SFHU Non-thermal (gal/AGN co-evol.)	~1 GHz Band 1/2	Ultra Deep	50 nJy	1 deg2	~0.5"	AGN/gal co-evol.
		Deep	200 nJy	10-30 deg2	~0.5"	AGN/gal co-evol. High-z Magnetism HI deep field (B1)
		Wide	1 uJy	1000 deg2	~0.5"	Weak/Strong Lensing
		Ultra Deep	40 nJy	0.008 deg2	~0.1"	AGN/gal co-evol.
SFHU Thermal (gal/AGN co-evol.)	~10 GHz Band 5	_				
		Deep	300 nJy	1 deg2	~0.1"	AGN/gal co-evol.
Legacy Strong Lensing (rare populations)	~1 GHz Band 2	All-sky	4 uJy	31000 deg2	~2" 0.5"	Magnetism Cosmology tests Transients (beam forming) HI surveys Our Galaxy
Clusters (RL AGNs)	~120 MHz	All-sky	20 uJy (confusion)	31000 deg2	8"	EoR



Galaxy Evolution: Star formation and feedback

Hopkins et al. 2004

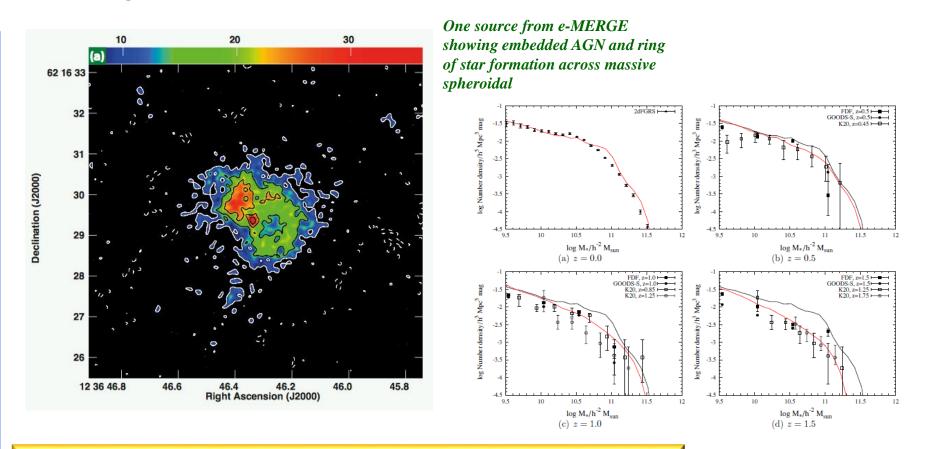




- Radio continuum excellent tracer of SF in SKA era
- AGN Feedback radio essential

Murphy, Sargent et al 2015

Galaxy Evolution: Star formation and feedback

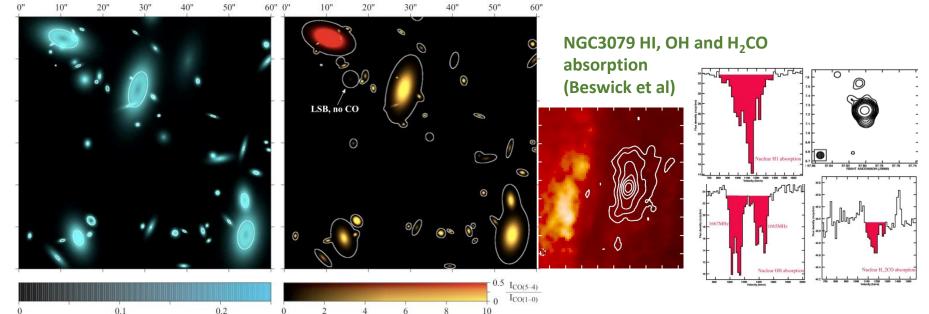


UNIVERSITY OF CAMBRIDGE

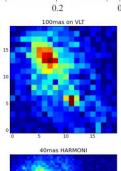
- e-MERLIN demonstrated the power of resolving AGN and star-forming contributions
- e-MERLIN follow up of brighter galaxies to probe detailed starformation / AGN interactions and feedback
- Large fraction of wide-field sample will be accessible to e-MERLIN

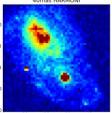
CAMBRIDGE

Galaxy Evolution: SKA / ALMA / E-ELT /e-MERLIN synergy



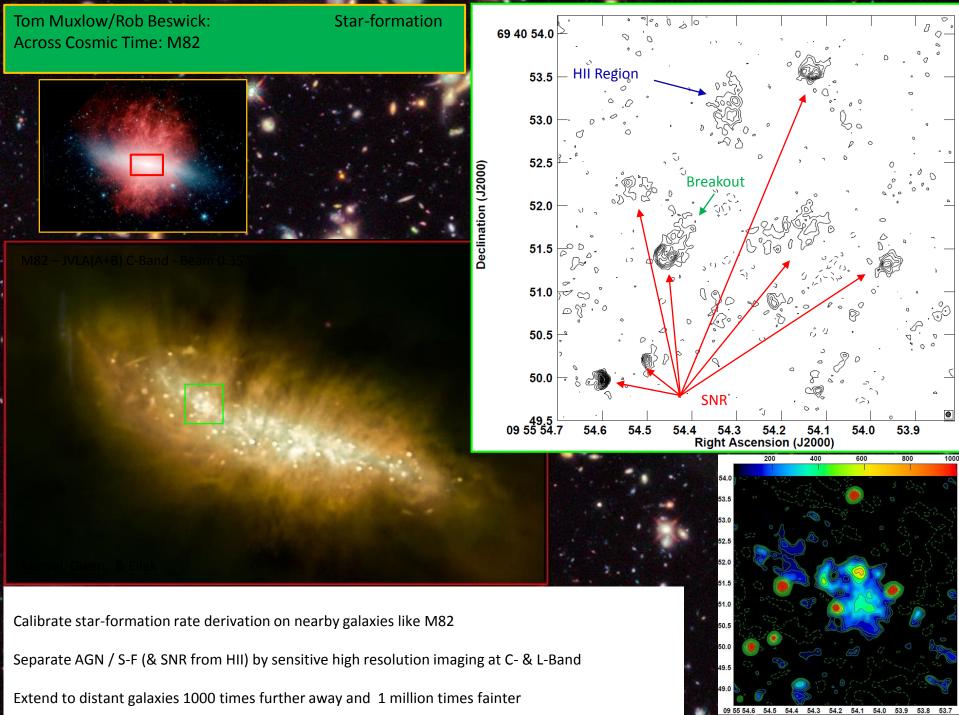
CO(1-0) flux density / [Jy km s⁻¹ arcmin⁻²]

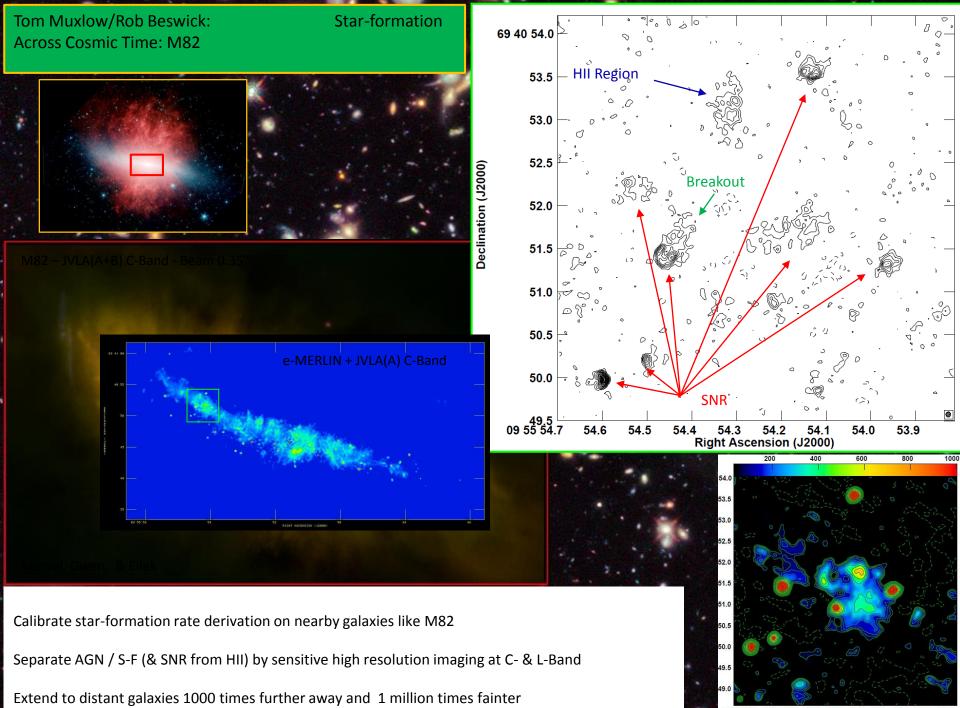




ALMA, SKA and e-MERLIN are natural IFUs at different scales and transitions

- E-ELT IFU maps, e.g. Halpha
- Continuum imaging e-MERLIN + VLBI

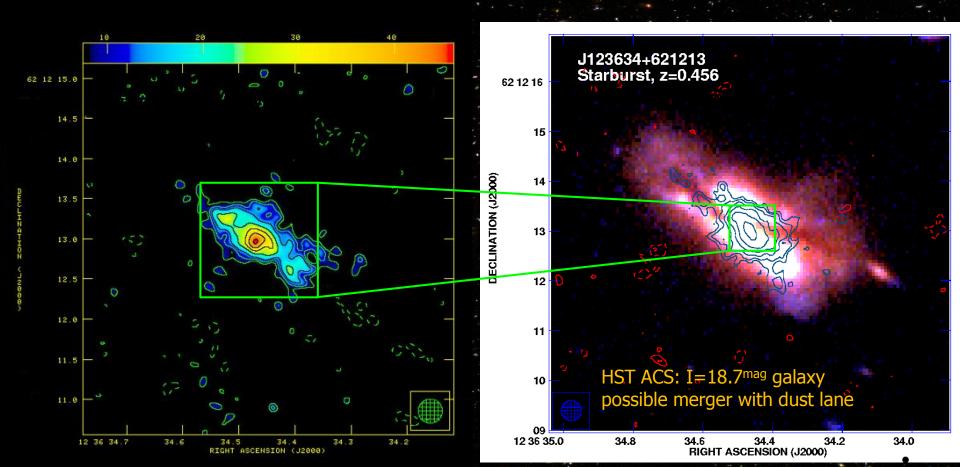




09 55 54.6 54.5 54.4 54.3 54.2 54.1 54.0 53.9 53.8 53.7

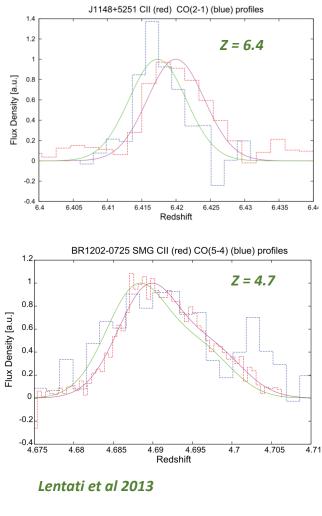
Star-formation Across Cosmic Time: e-MERGE Tier 1



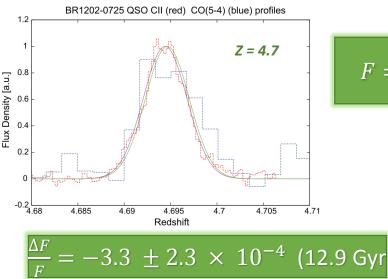


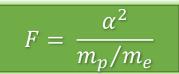


Cosmology: Variation of fundamental constants?



DF/F = -3.3

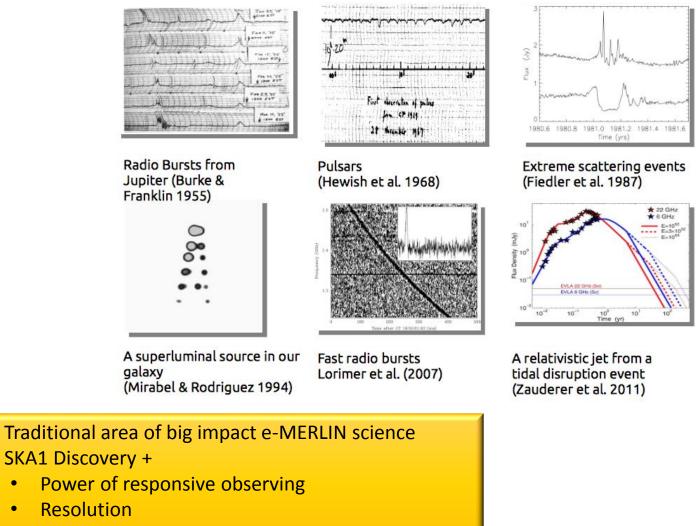




- Observe different types of transitions (e.g. electronic vs rotational) in galaxies at range of z
- Use differences in inferred redshits of the lines to look for evidence of variation in fundamental constants
- Problem
 - Are we tracing same gas in the two transitions?
 - Do we observe astrophysics or physics?
- Here resolution of e-MERLIN very helpful
 - HI / OH
 - More transitions



Transient Universe



Link to EVN •

•

•



Transient Universe

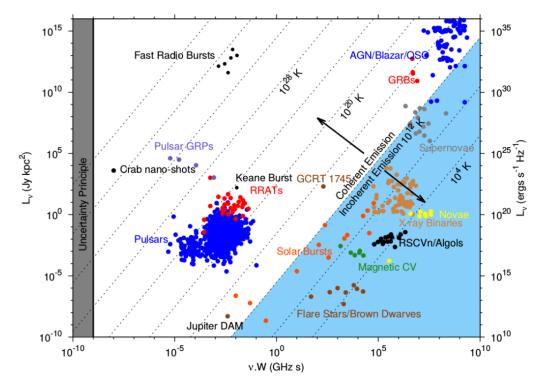
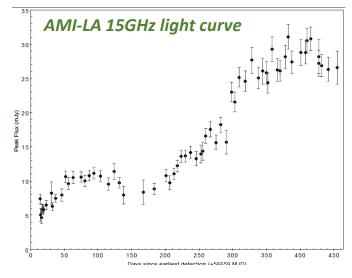


Figure 4: Transients parameter space expanded to include coherent sources. From Pietka, Fender & Keane (2015), following a long line of similar plots (e.g. Cordes, Lazio & McLaughlin 2004).

- Traditional area of big impact e-MERLIN science
- SKA1 Discovery +
 - Power of responsive observing
 - Resolution
 - Link to EVN

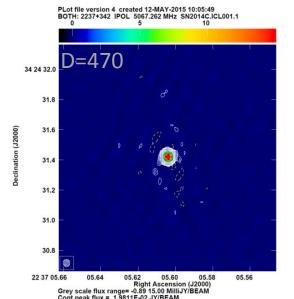


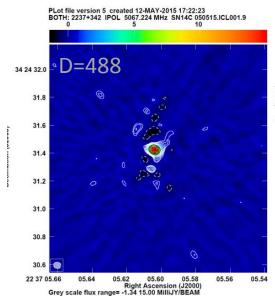
Transient Universe: SN2014C

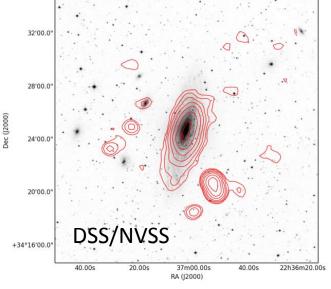


SN2014C (Lick discovery)

- D=15.1Mpc in NGC7731
- Unusual double peak
- Indicative of shell interaction with dense CSM
- Possibly radio bright SN1b
- Now trigger VLBI follow-up pending
- e-MERLIN resolution vital to remove confusion







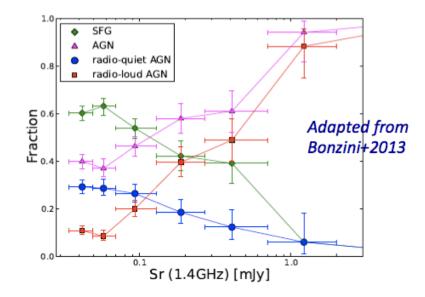
Anderson, Fender et al

Paul Alexander: SRC Workflows and ASTERICS

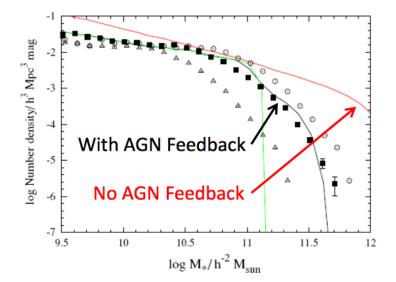


AGN Physics

- Physics of RL AGN physics/lifecycle
- Physics of RQ/RL AGN dichotomy
- How do RL / RQ AGN provide feedback
- Jet Physics: origin and propagation

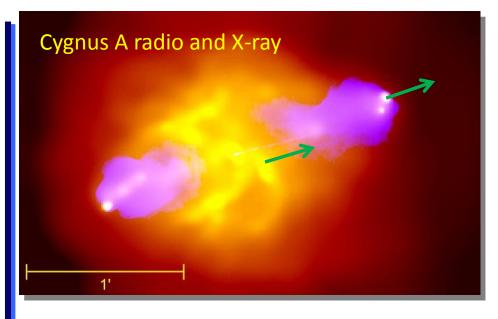


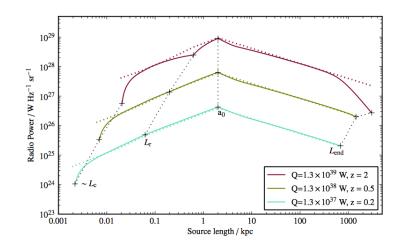
- Radio Astronomy and AGN Physics !!
- Much progress but still many unsolved problems
- Interesting in own right as well as importance in overall galaxy evolution
- AGN physics is interesting





AGN Physics





տրմ пΓ 30 10 20 n 31 26 46.8 46.7 46.6 46.5 46.5 46.4 20 46.4 20 46.3 46.3 46.3 46.3 46.3 Global VLBI + MERLIN + VLA (1.35GHz 46.2 17.86 17.84 17.82 RIGHT ASCENSION (J2000) 13 52 17.94 17.92 17.90 17.88 17.80 17.78 17.76 13100 13100 13300 13400

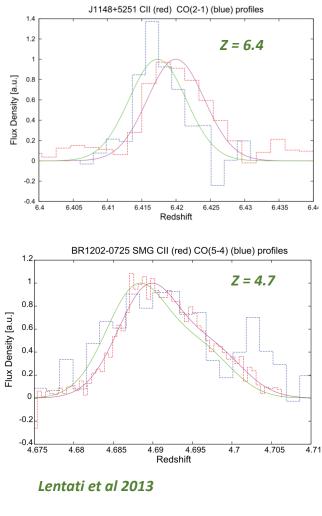
Neutral hydrogen absorption against 3C293

(Beswick, Peck, Taylor & Giovannini, 2004 MNRAS 352, 49)

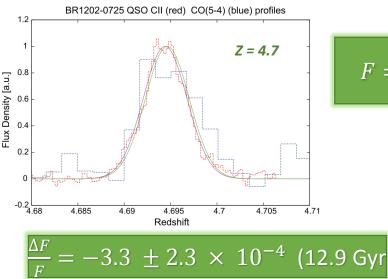
Paul Alexander: SRC Workflows and ASTERICS

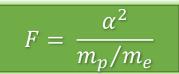


Cosmology: Variation of fundamental constants?



DF/F = -3.3





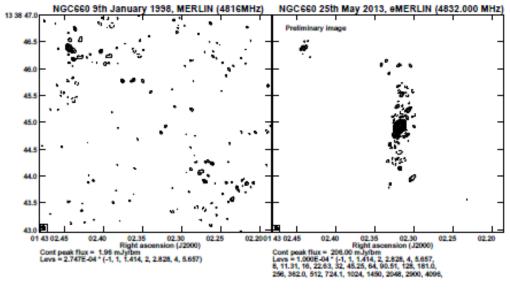
- Observe different types of transitions (e.g. electronic vs rotational) in galaxies at range of z
- Use differences in inferred redshits of the lines to look for evidence of variation in fundamental constants
- Problem
 - Are we tracing same gas in the two transitions?
 - Do we observe astrophysics or physics?
- Here resolution of e-MERLIN very helpful
 - HI / OH
 - More transitions

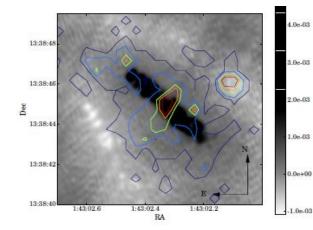


A new period of activity in NGC660

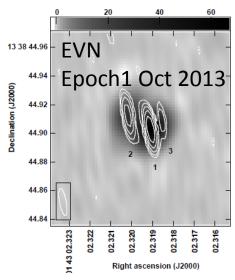
Witnessing the **(re-)birth** of an AGN Joint eMERLIN/EVN/WSRT study of spectral lines and continuum

- New source discovered in Arecibo monitoring (2008-2010)
- Nuclear continuum source peaked at ~0.5Jy – now in steady decline





Pre-outburst MERLIN Lband and post outburst X-ray (Chandra)



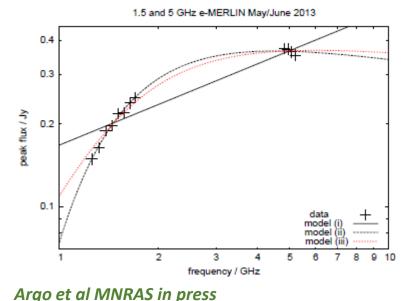


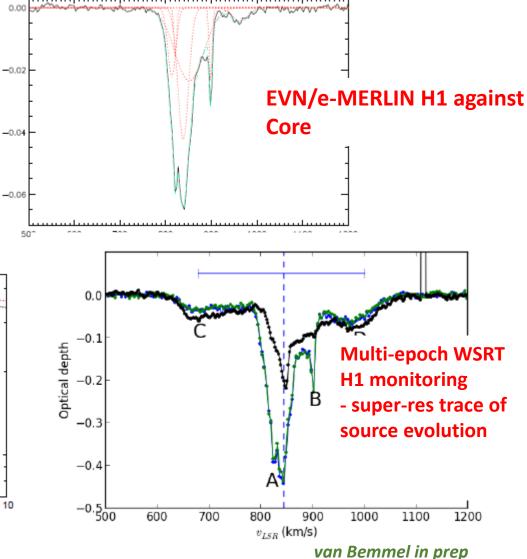
Nuclear source & Gas?

-0.02

4bsorption (Jy/bm)

- Outburst provide chance of high-res/sen line observations of neutral molecular ISM
- Wide-band continuum SED
- New GPS source







Paul Alexander: SRC Workflows and ASTERICS