RadioLensfit: an HPC Tool for Accurate Galaxy Shape Measurement with SKA

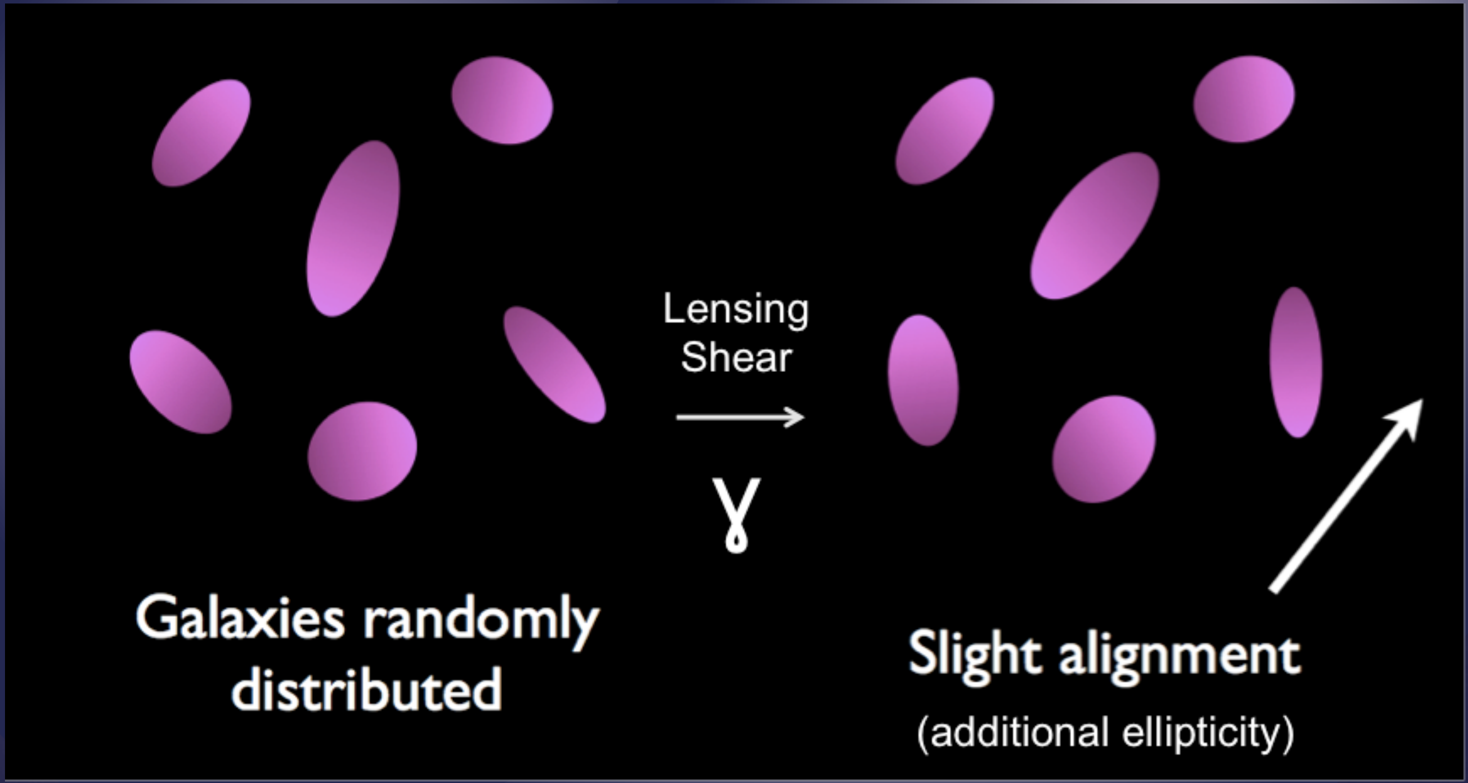
Marzia Rivi
IRA-INAF

5 October 2021
The Third National Workshop on the SKA project

Galaxy Shapes for Weak Lensing

The measure of gravitational lensing is a powerful technique for estimating mass distribution of dark matter.

Weak lensing regime (large scales), i.e. $|\gamma| \ll 1$:



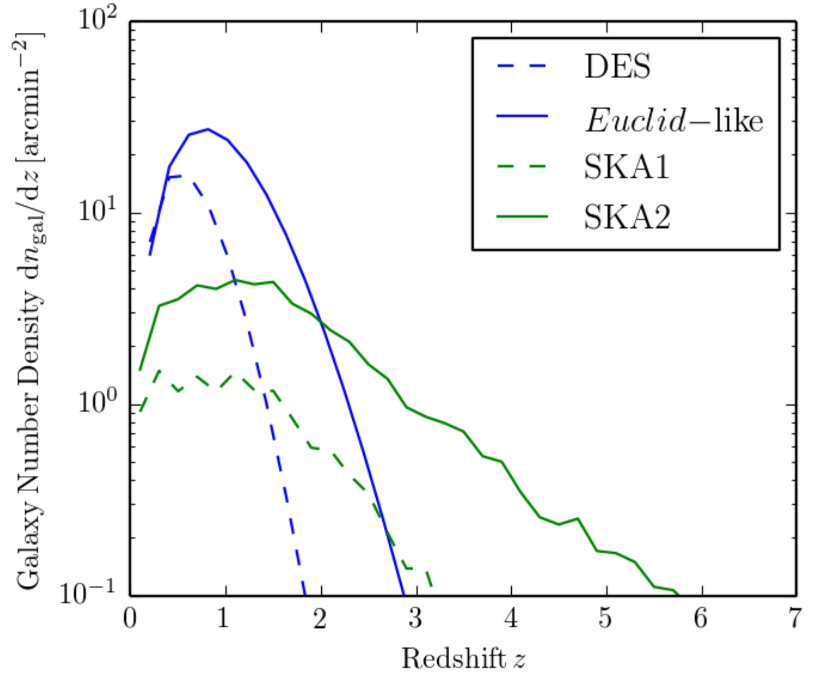
For a **large sample** of background SF galaxies, **<e>** can be used as an *estimator of the weak lensing shear* γ assuming sources randomly oriented, i.e. $\langle \mathbf{e}^s \rangle = 0$.

Weak Lensing with SKA

The *Square Kilometre Array* (SKA) will reach **sufficient sensitivity** and **angular resolution** to provide large number density of faint star-forming galaxies.

Experiment	$A_{\rm sky}$ [deg^2]	$n_{\rm gal}$ [arcmin^{-2}]
SKA1	5,000	2.7
DES	5,000	12
SKA2	30,000	10
Euclid-like	15,000	30

Radio significant **high-redshift tail** compared with optical correspondent surveys



Harrison+ 2016

Well-known and deterministic knowledge of the PSF (most instrumental systematic effect)

Unique radio approaches may provide estimate of galaxy's intrinsic orientation allowing mitigation of intrinsic alignments:

- Polarization (**Brown & Battye 2011, Whittaker+ 2015**)
- HI rotational velocity (**Morales 2006, Huff+ 2013**)

Weak Lensing with SKA

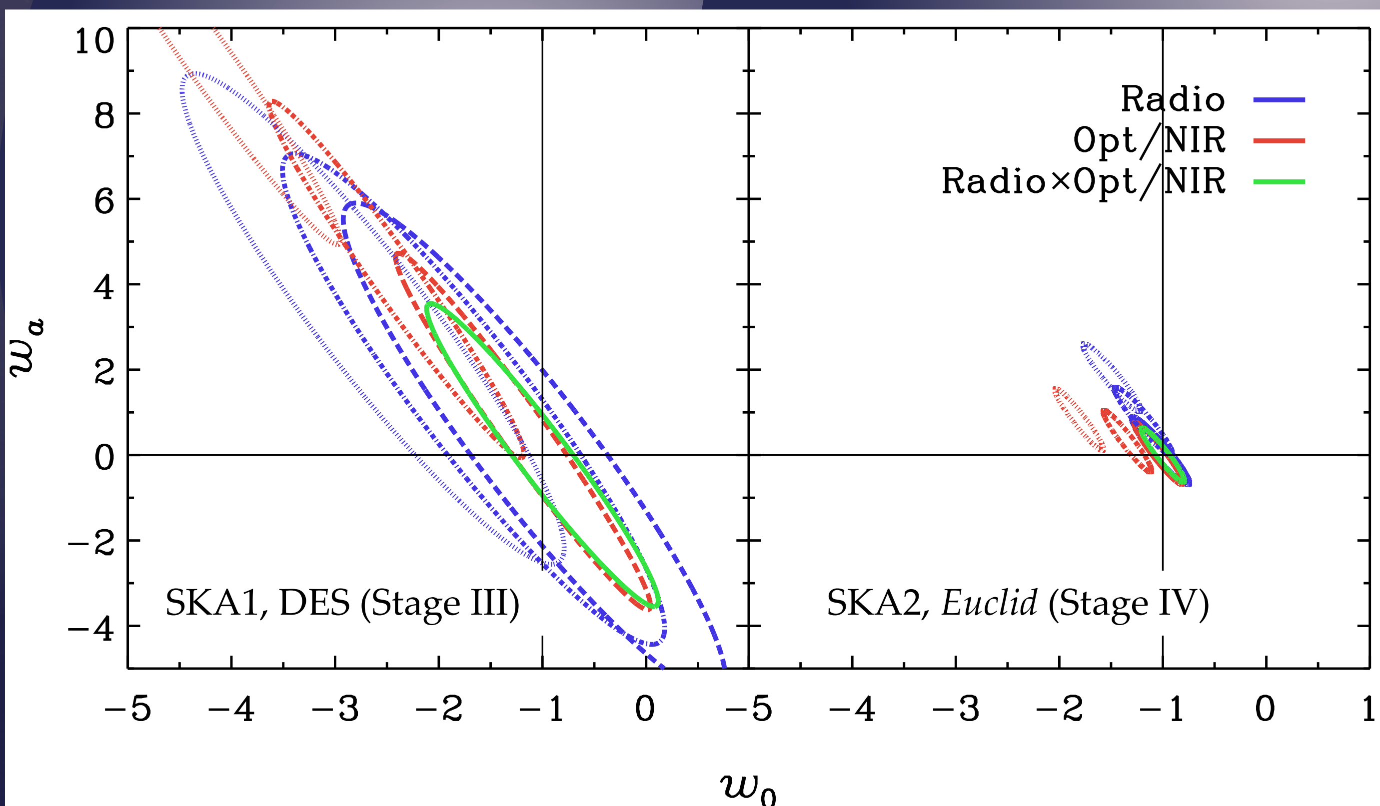
Removal of additive experimental systematics by radio-optical cross-correlation

$$\langle\tilde{\gamma}_o\tilde{\gamma}_r\rangle = \langle\gamma\gamma\rangle + \langle\gamma_o^i\gamma\rangle + \langle\gamma_r^i\gamma\rangle + \langle\gamma_o^i\gamma_r^i\rangle + \langle\gamma_o^s\gamma_r^s\rangle$$

Cosmic shear signal

Intrinsic correlations in galaxy shapes (GI & II signals)

Systematics will be uncorrelated for optical and radio telescopes



Simulation for DETF experiments

residual systematics modelling (**Amara & Refregier 2008**):

$$C_\ell^{\rm add} = A_{\rm add}\frac{n_{\rm add}\log(\ell/\ell_{\rm add})+1}{\ell(\ell+1)}$$

$$\sigma^2_{\rm sys} = \int\frac{{\rm d}\ln\ell}{2\pi}\,\ell(\ell+1)\left|C_\ell^{\rm add}\right|$$

$\sigma^2_{\rm sys} = 10^{-7}, 10^{-6}, 5\times10^{-5}$

$n_{\rm add} = -1$

$l_{\rm min} = 20$, $l_{\rm max} = 3000$

Camera+ 2017

Galaxy Shape Measurement

Non linear imaging procedure introduces non-negligible systematics and highly correlated noise. Essential to work directly with the original data!

complex visibilities ⇒ big data ⇒ **Joint analysis in the two domains**
sources not localized

IMAGE DOMAIN

- source detection and classification
- position, flux (and possibly size) measurement
- ellipticity initial guess

VISIBILITY DOMAIN for accurate ellipticity measurement following optical approach: **model fitting** of the surface brightness profile

1. **Joint fitting** of all sources
   - Shapelets invariant under Fourier transform (**Chang+ 2002**), but already dismissed in the optical domain due to model bias
   - Hamiltonian Monte Carlo Markov Chains (**Rivi+ 2019**), very accurate but very time-consuming even for small datasets

2. **Single source model fitting**
   - RadioLensfit - adaptation of *lensfit* method used in optical surveys, (**Rivi+ 2016, Rivi & Miller 2018, 2021**)

Galaxy shape model

Radio model: defined by the synchrotron radiation from the ISM in the disk

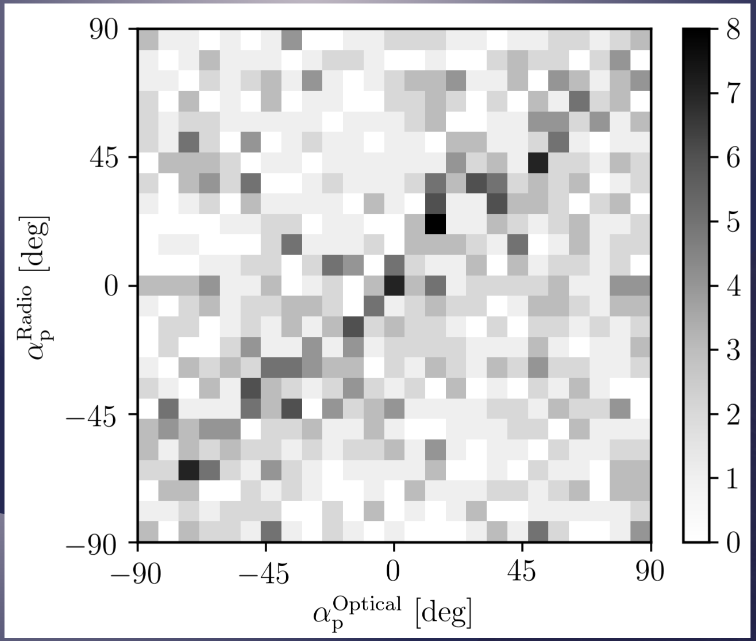
- **single component**
- **exponential brightness profile**: $I(r) = I_0 e^{-r/\alpha}$

First confirmation of this assumption from radio-optical shape comparison of the cross-matching sources of the COSMOS field at 3 GHz

optical: HST-ACS
radio: VLA

- at 1.4 GHz (PSF FWHM ≈ 1.45 arcsec) almost no correlation (**Tunbridge+ 2016**) Resolution is too low!

- at 3 GHz (PSF FWHM ≈ 0.75 arcsec) a correlation of position angles is found (**Hillier+ 2019**)



Optical: good performance in the GREAT Challenges of methods using Se'rsic models, they reduce model bias observed with shapelets.

RadioLensfit: Methodology

- **Source visibilities extraction** (by flux/SNR decreasing order):
  - **sky model** made of round sources to isolate the visibilities of a single source
  - **faceting** to restrict the f.o.v around the source: averaging visibilities within grid cells of size $\Delta_u = 1/\theta$ (wavelength units), θ dependent on source size
- **Analytical visibilities model**:

$$V(u,v,w) = \left(\frac{\lambda_{\rm ref}}{\lambda}\right)^{\beta} \frac{S_{\lambda_{\rm ref}} e^{2\pi i\left(ul+vm+w(\sqrt{1-l^2-m^2}-1)\right)}}{\left(1+4\pi^2\alpha^2|\mathbf{A}^{-T}\mathbf{k}|^2\right)^{3/2}}.$$

$$\mathbf{A} = \begin{pmatrix} 1-e_1 & -e_2 \\ -e_2 & 1+e_1 \end{pmatrix}$$

- **Bayesian marginalization** of the likelihood over position (analytically), flux (semi-analytically) using uniform prior, and size (numerically) with log-normal prior $\Rightarrow L(e_1, e_2)$
- **Likelihood sampling**: ML + adaptive grid around the maximum point

$$\hat{\mathbf{e}} = \text{mean}, \qquad \sigma^2_{\hat{\mathbf{e}}} = \sqrt{\det(cov)}$$

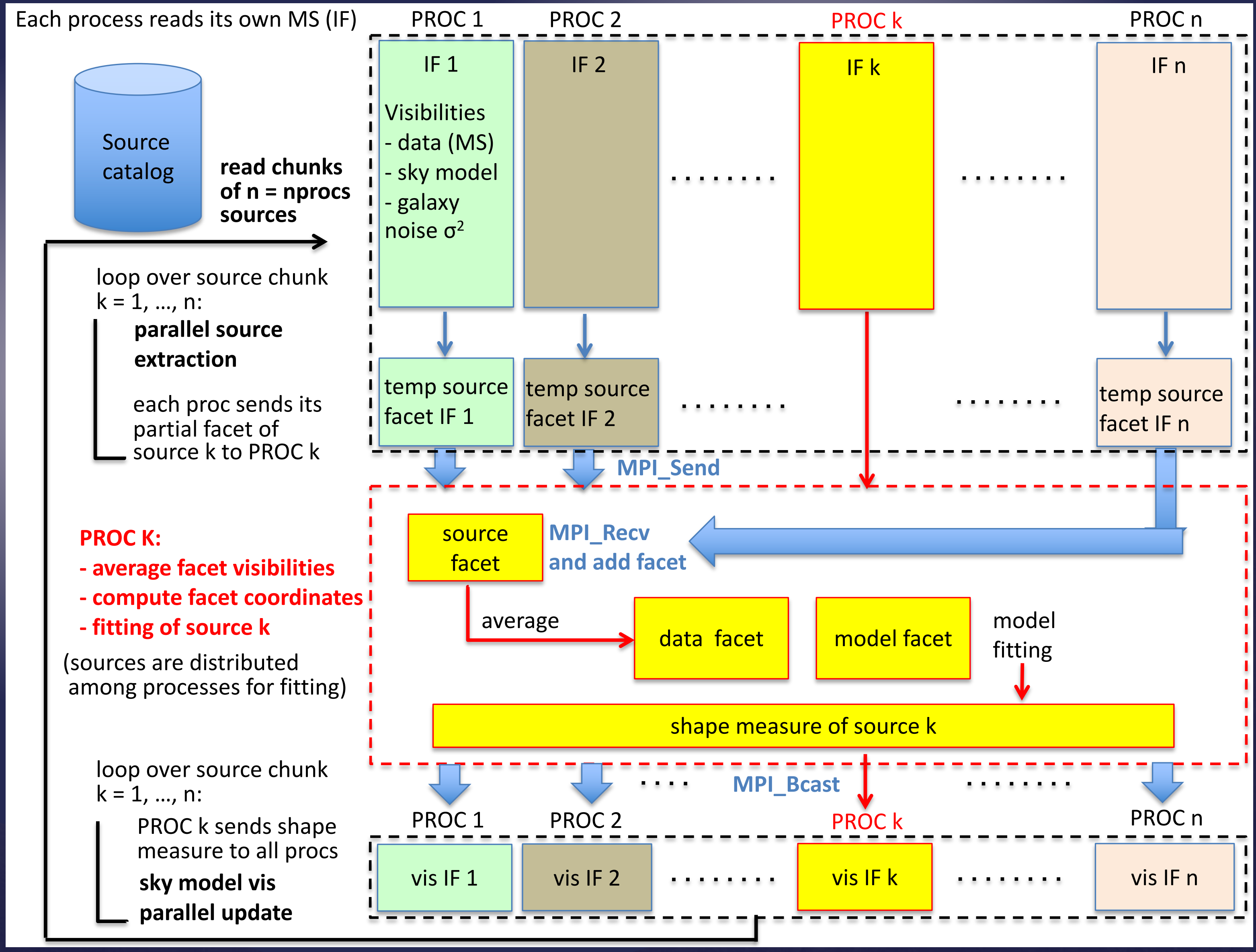
- **Refinement of the sky model:** replace the ellipticity of the current source with the measured one.

RadioLensfit: Parallelization

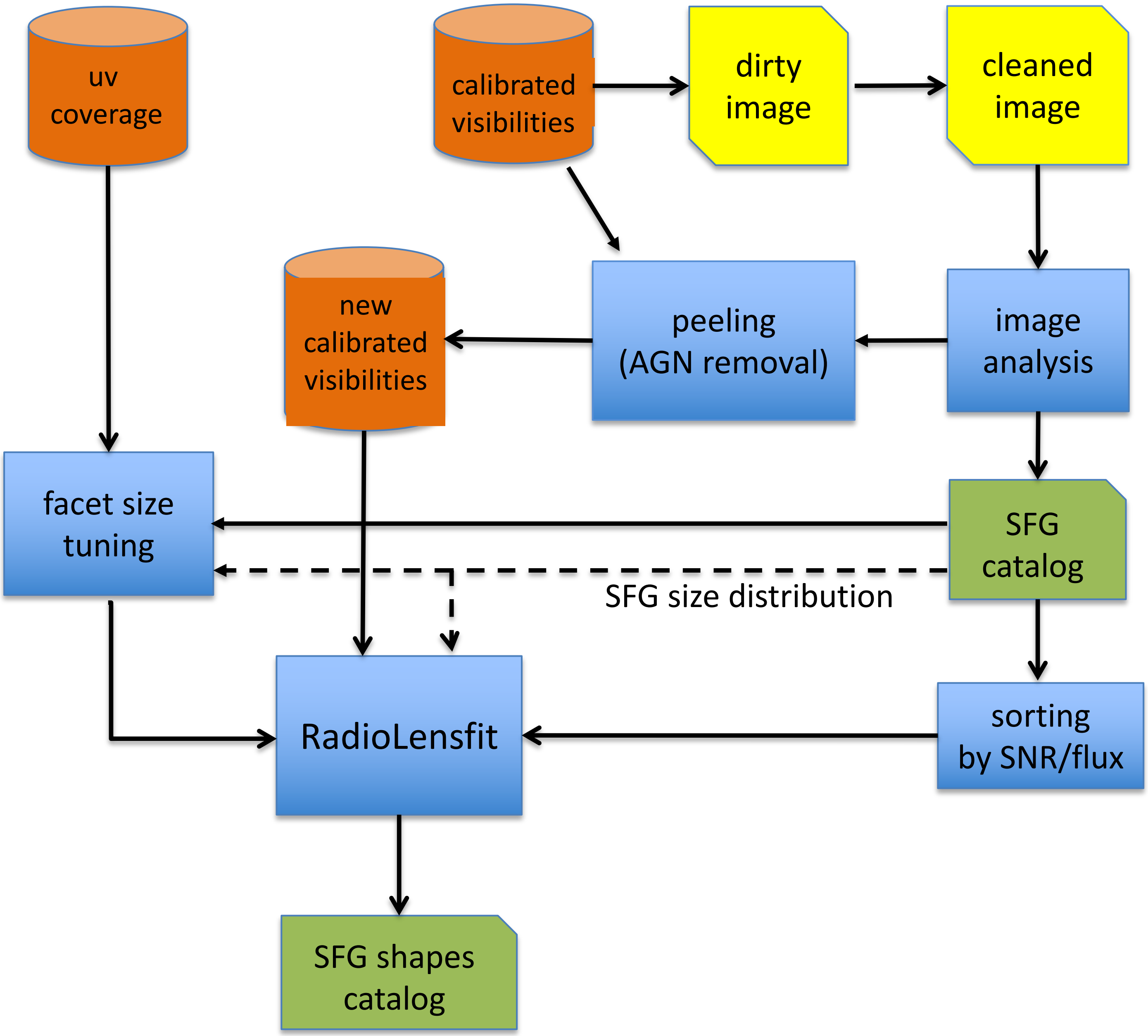
Hybrid parallelization: MPI + OpenMP

MPI:

OpenMP:
- model visibilities
- log-likelihood $L(\alpha, e_1, e_2)$



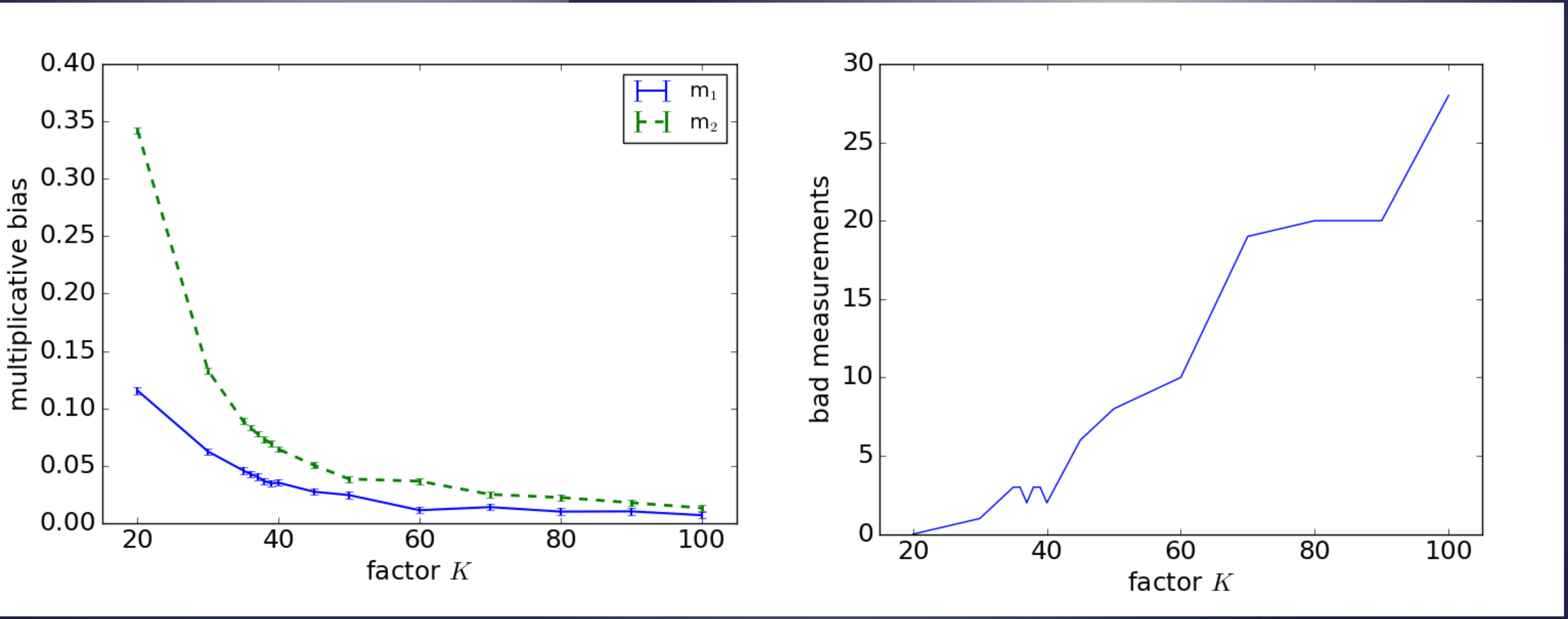
Radio weak lensing pipeline



RadioLensfit: Faceting tuning

- If not available, estimate source scalelength correlation with source flux: $\alpha = f(S)$. Typically a linear relation between log α_{med} and log$S$
- Facet field of view $\theta = K\alpha$, K constant dependent on the uv coverage

Linear bias model: $$\tilde{e}_i - e_i = m_i e_i + c_i, \qquad i = 1, 2,$$



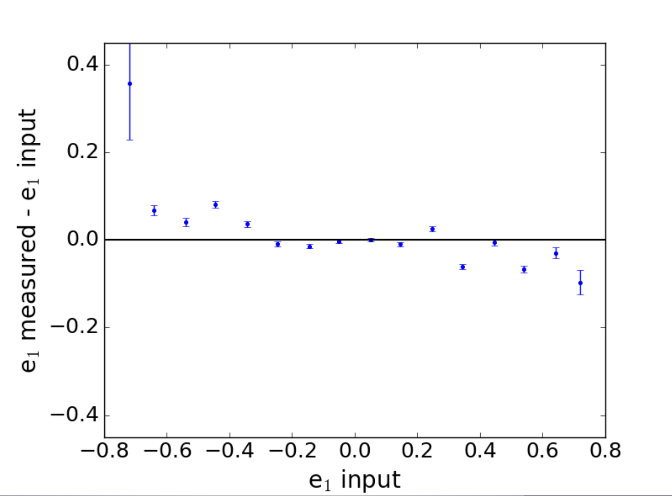
SKA1-MID simulation of 1000 single sources with SNR > 20

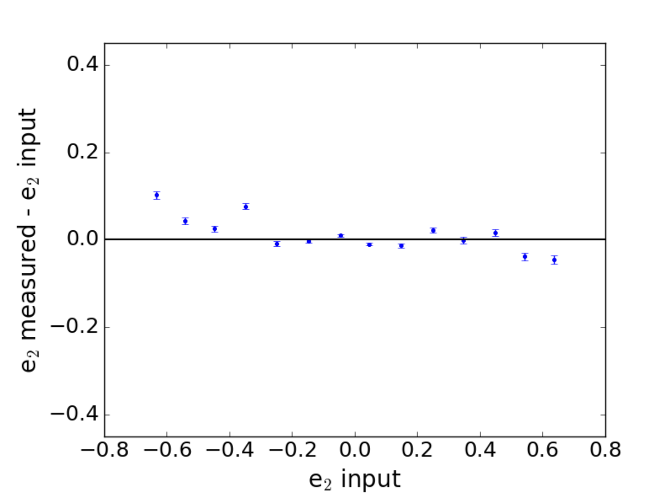
Rivi & Miller 2021

RadioLensfit: Performance

SKA1-MID simulation at 2.7 gal/armin^2 density

1000 sources with flux 10 – 200 μJy (SNR > 10)
single channel at 1.4 GHz, 240 MHz bandwidth, 8-h observation, 60 s sampling time





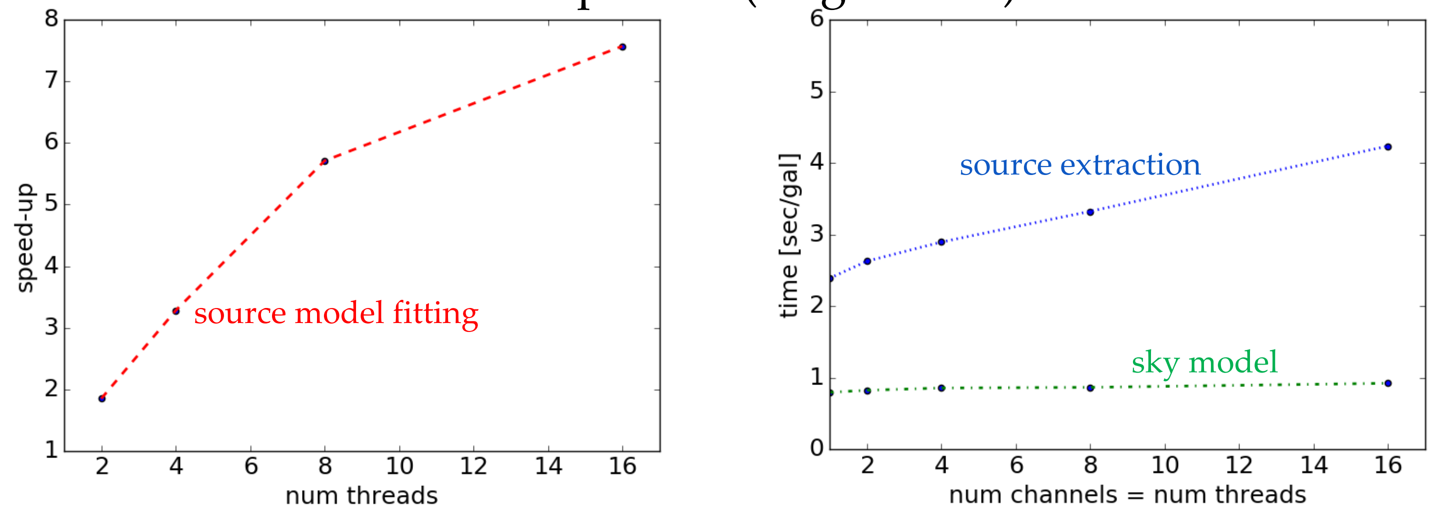
Test	bad	shape bias			
		m_1	c_1	m_2	c_2
RL single	13	0.0322 ± 0.0056	0.0041 ± 0.0015	0.0321 ± 0.0051	0.0062 ± 0.0015
RL all + sizes	10	0.0402 ± 0.0057	0.0057 ± 0.0016	0.0393 ± 0.0051	0.0001 ± 0.0015
RL all	14	0.0748 ± 0.0056	-0.0005 ± 0.0015	0.0620 ± 0.0051	0.0067 ± 0.0015
HMC all	0	0.0296 ± 0.0043	0.0001 ± 0.0010	0.0282 ± 0.0040	-0.0002 ± 0.0010

RadioLensfit results **comparable with joint fitting** (HMC) **but much faster**:

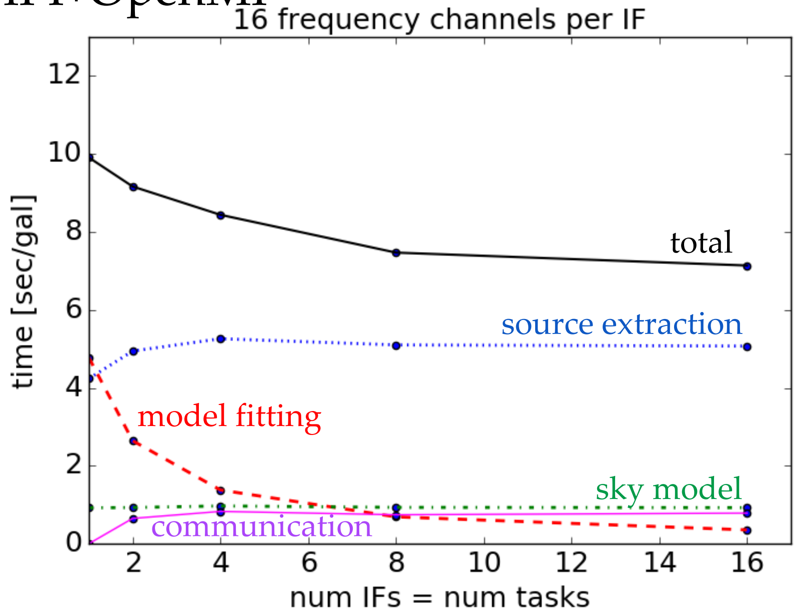
RL on 16-core Intel Xeon E5-2650 at 2.00 GHz: 1h 35 min **(Rivi & Miller 2021)**
HMC on same CPU + 2 NVIDIA Tesla K40 GPUs: ~ 9 weeks **(Rivi+ 2019)**

RadioLensfit: Scalability

OpenMP (single node)



MPI+OpenMP



INAF HOTCAT cluster
SKA1_MID simulation ~ 10^7 uv points

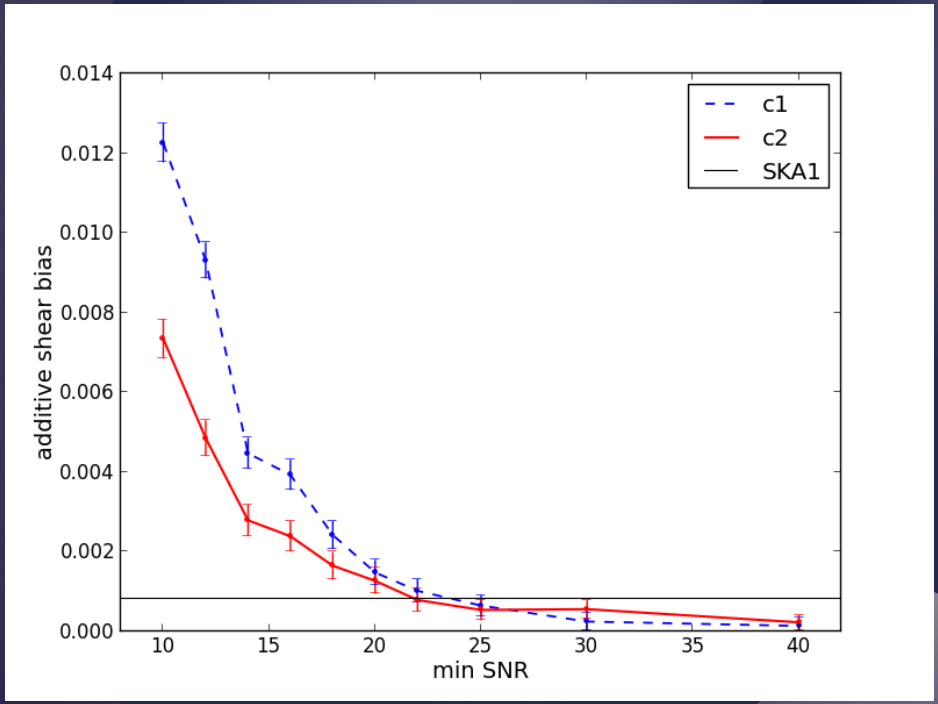
Overall **super-linear weak scalability** despite tasks synchronization overhead thanks to the strong scalability of the model fitting

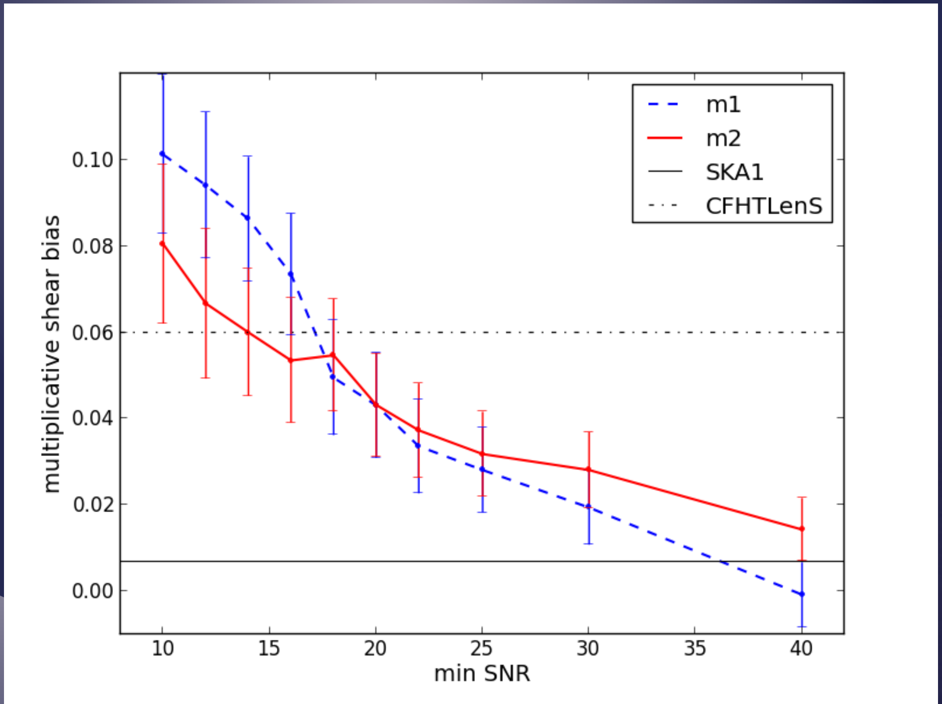
Rivi & Miller 2021

RadioLensfit: Shear bias

ideal case (no neighbourood bias), 10^4 single galaxies

shear bias requirements for SKA1: $m < 6.4 \times 10^{-3}$, $c < 8.0 \times 10^{-4}$





Removable by isotropization of the PSF

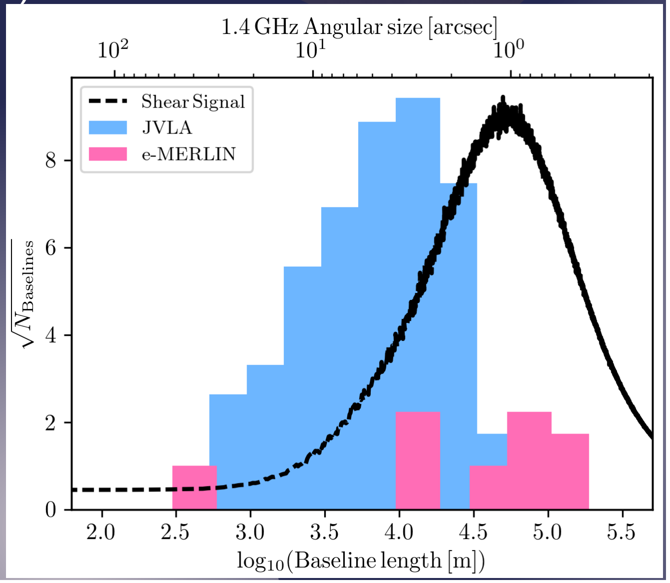
Calibration is required as in the optical surveys

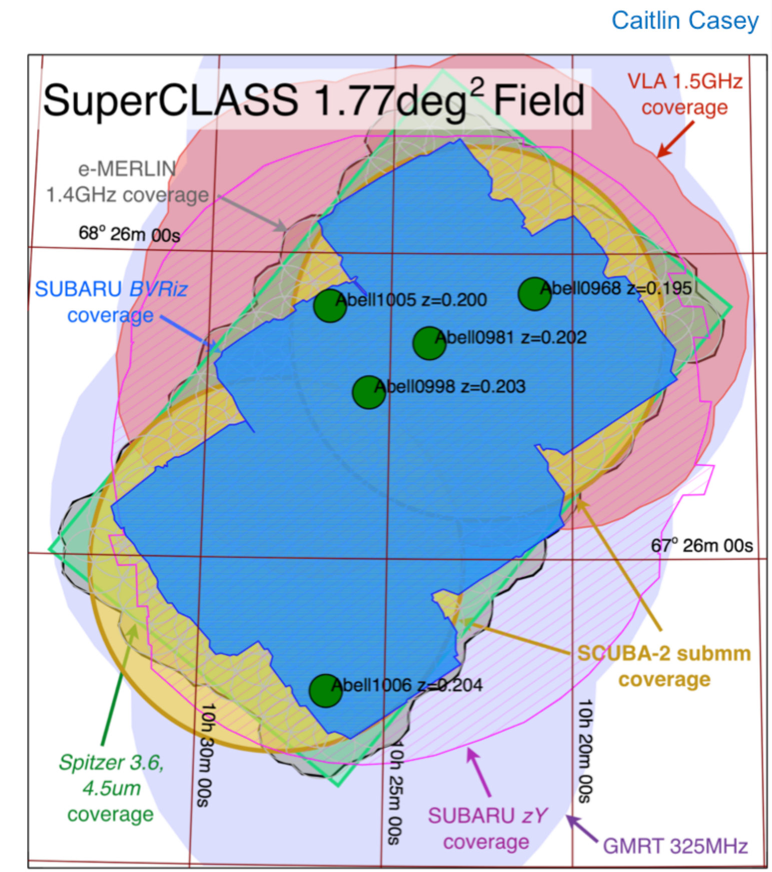
Rivi+ 2016

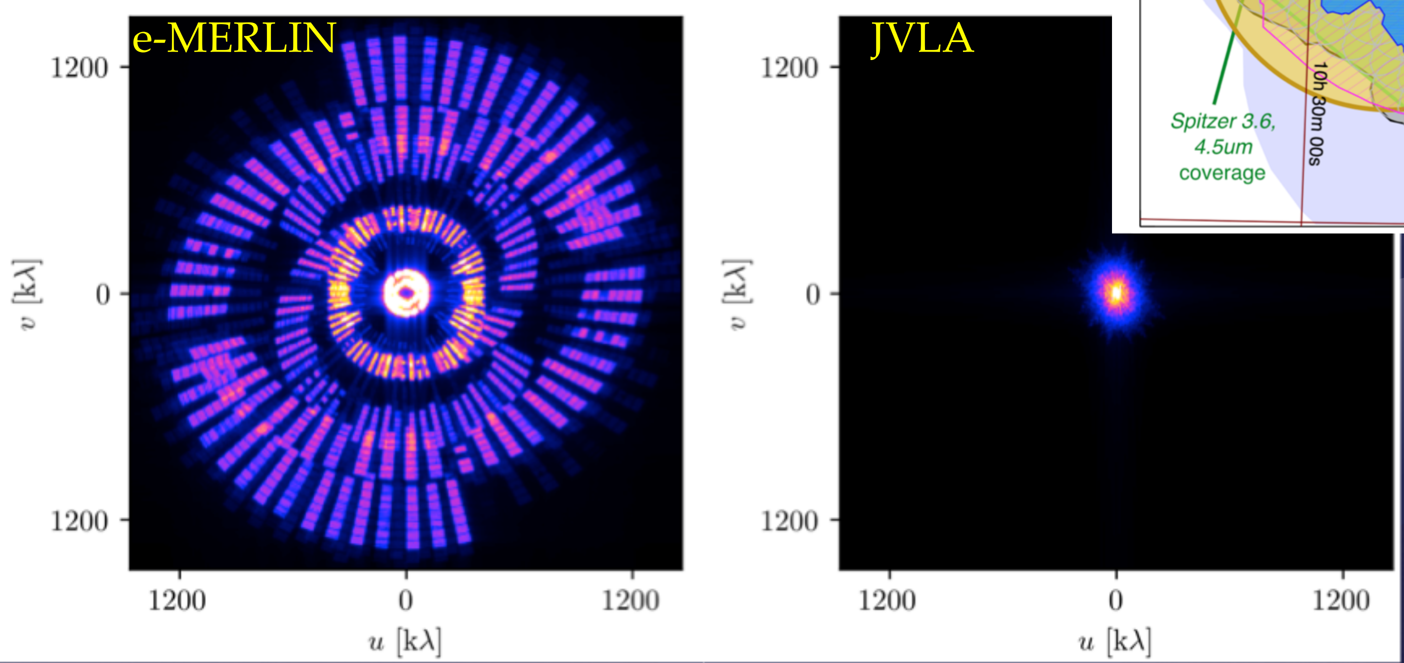
SuperCLuster Assisted Shear Survey (SuperCLASS)

e-MERLIN (1.4 GHz)
JVLA (1.5 GHz)

Combination of e-MERLIN and JVLA baselines to access relevant spatial scales







SuperCLASS: DR1 galaxy shapes

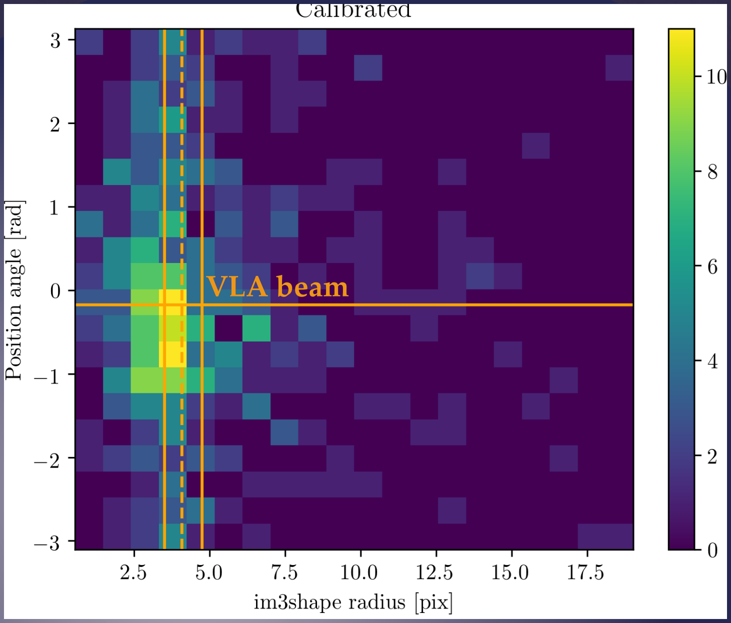
Data Release 1: 0.26 deg^2

Only *image-plane* shape measurement (IM3SHAPE) with calibration simulations on a source-by-source basis (SuperCALS method)

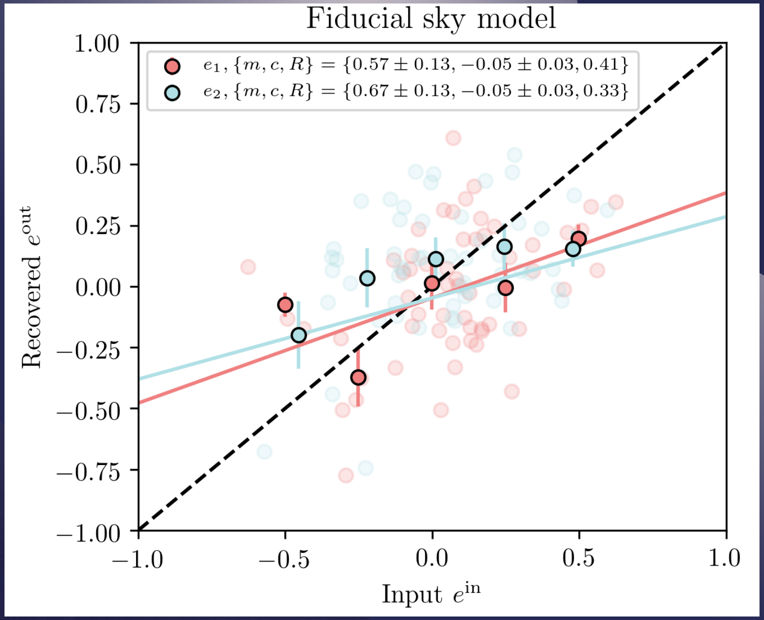
VLA: 440 SF galaxies (0.47 gal/armin^2) with SNR > 7

E-MERLIN: 56 matched sources (≈ 0.06 gal/arcmin^2)

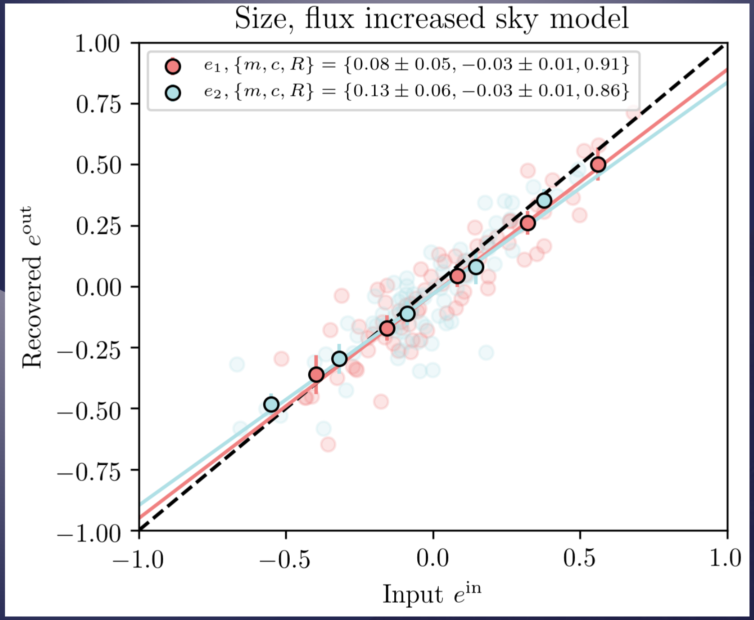
size and shape of the beam is dominating the morphology of recovered sources



Simulated SFG of J28 pointing of the VLA DR1



flux × 100, size × 3



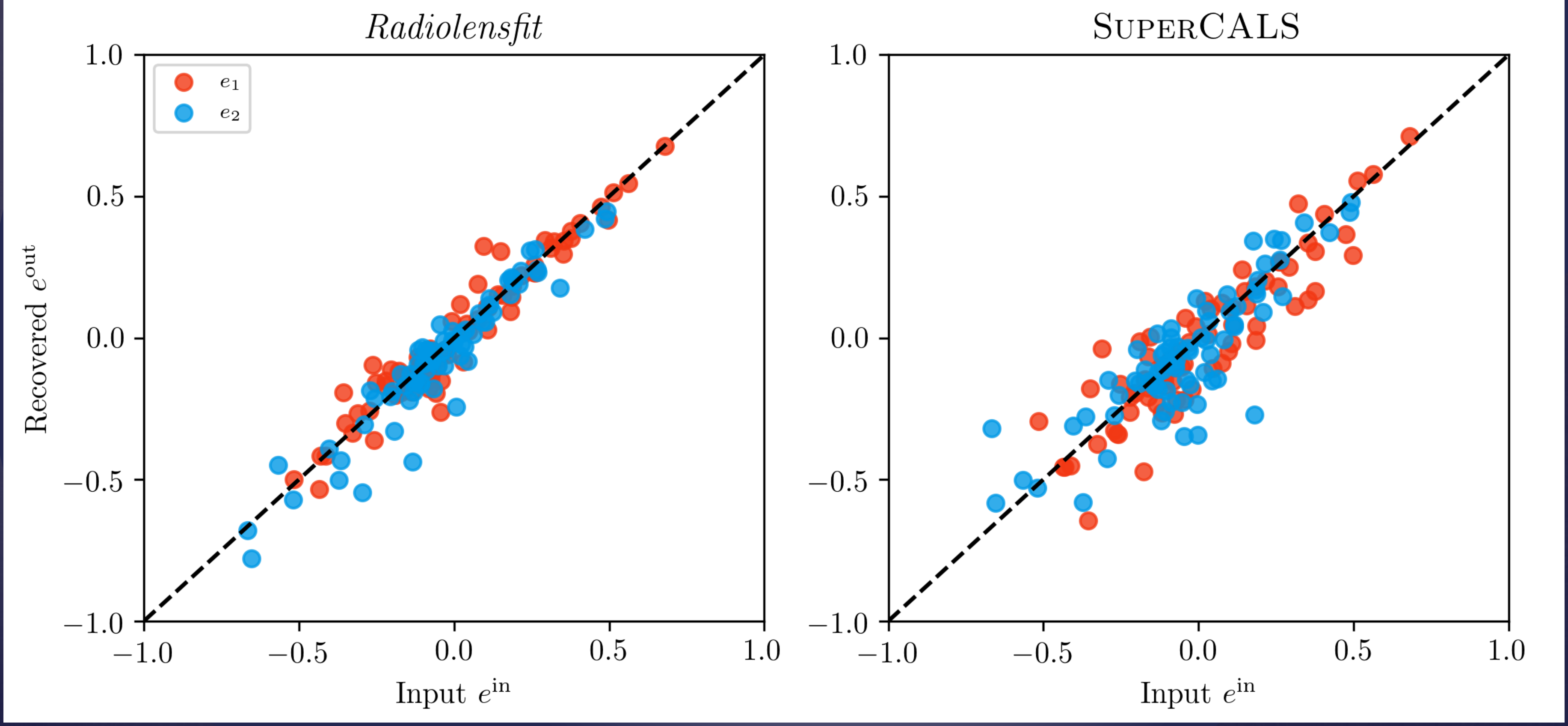
Harrison & the SuperCLASS Collaboration 2020

SuperCLASS: Visibilities vs Image

Simulated SFG of J28 pointing of the VLA DR1, size and flux increased:

$m_1 = 0.042 \pm 0.002$, $c_1 = 0.0048 \pm 0.0003$
$m_2 = 0.039 \pm 0.002$, $c_2 = -0.0045 \pm 0.0003$

$m_1 = 0.08 \pm 0.05$, $c_1 = -0.03 \pm 0.01$
$m_2 = 0.13 \pm 0.06$, $c_2 = -0.03 \pm 0.01$



Much tighter correlation between input and recovered ellipticities

Bias reduced by almost one order of magnitude!

Conclusions

- SKA will open weak lensing observations to the radio band
- New methods for accurate SFG shape measurements are required

- **Joint analysis** in the image/visibility domains **may be required for SKA**. Image analysis may be used for:
  - sources position and flux (and possibly size)
  - initial guess on shape parameters
  - sources classification (for removal of AGN)

- **RadioLensfit** features:
  - working in the visibility domain
  - model fitting of a source at a time
  - exploitation of multi-core and multi-node systems
  - results comparable with the joint fitting approach

- **SuperCLASS** precursor WL survey (JVLA + e-MERLIN)
  - superCALS (image domain) vs RadioLensfit (visibility domain) methods
  - RadioLensfit may reduce the shape bias by almost an order of magnitude