

Joint Survey Processing (JSP) of Euclid, LSST and WFIRST

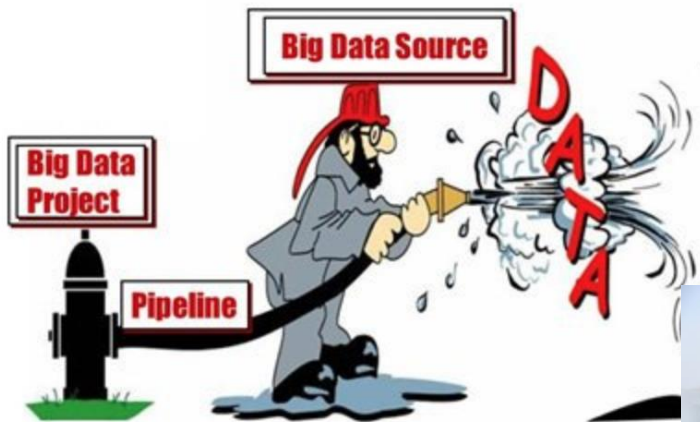
R. Chary (Caltech/IPAC) for the Joint Survey Processing Working Group

George Helou, Gabriel Brammer, Peter Capak, Andreas Faisst, Sergio Fajardo-Acosta, Dave Flynn, Steven Groom, Henry C. Ferguson, Carl Grillmair, Shoubaneh Hemmati, Anton Koekemoer, BoMee Lee, Sangeeta Malhotra, Hironao Miyatake, Peter Melchior, Ivelina Momcheva, Jeffrey Newman, Joseph Masiero, Roberta Paladini, Abhishek Prakash, Benjamin Rusholme, Nathaniel Stickley, Arfon Smith, Michael Wood-Vasey, Harry I. Teplitz

~40 scientists/engineers from 10 different institutions doing the groundwork

The Big Data Flood

Drinking from a FIREHOSE



Euclid/Firehose (2022+)



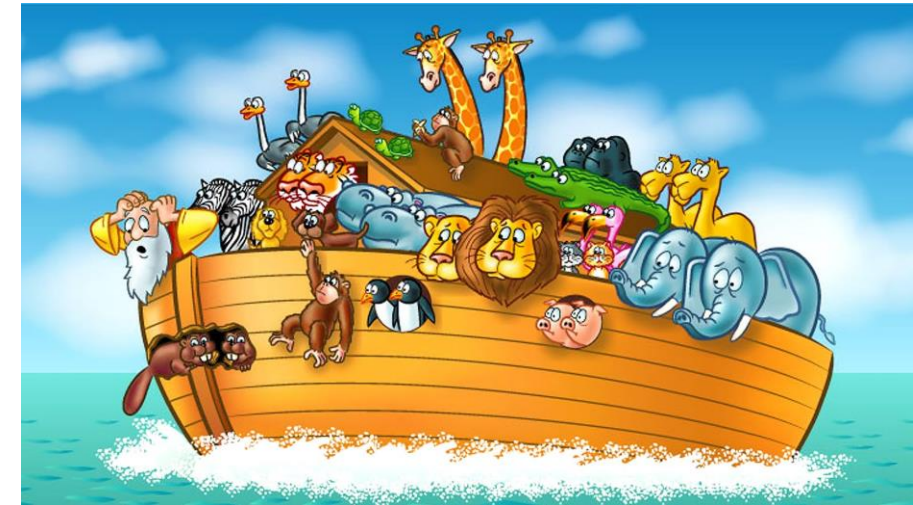
LSST/Iguazu (2023+)

WFIRST/Victoria Falls (2025+)



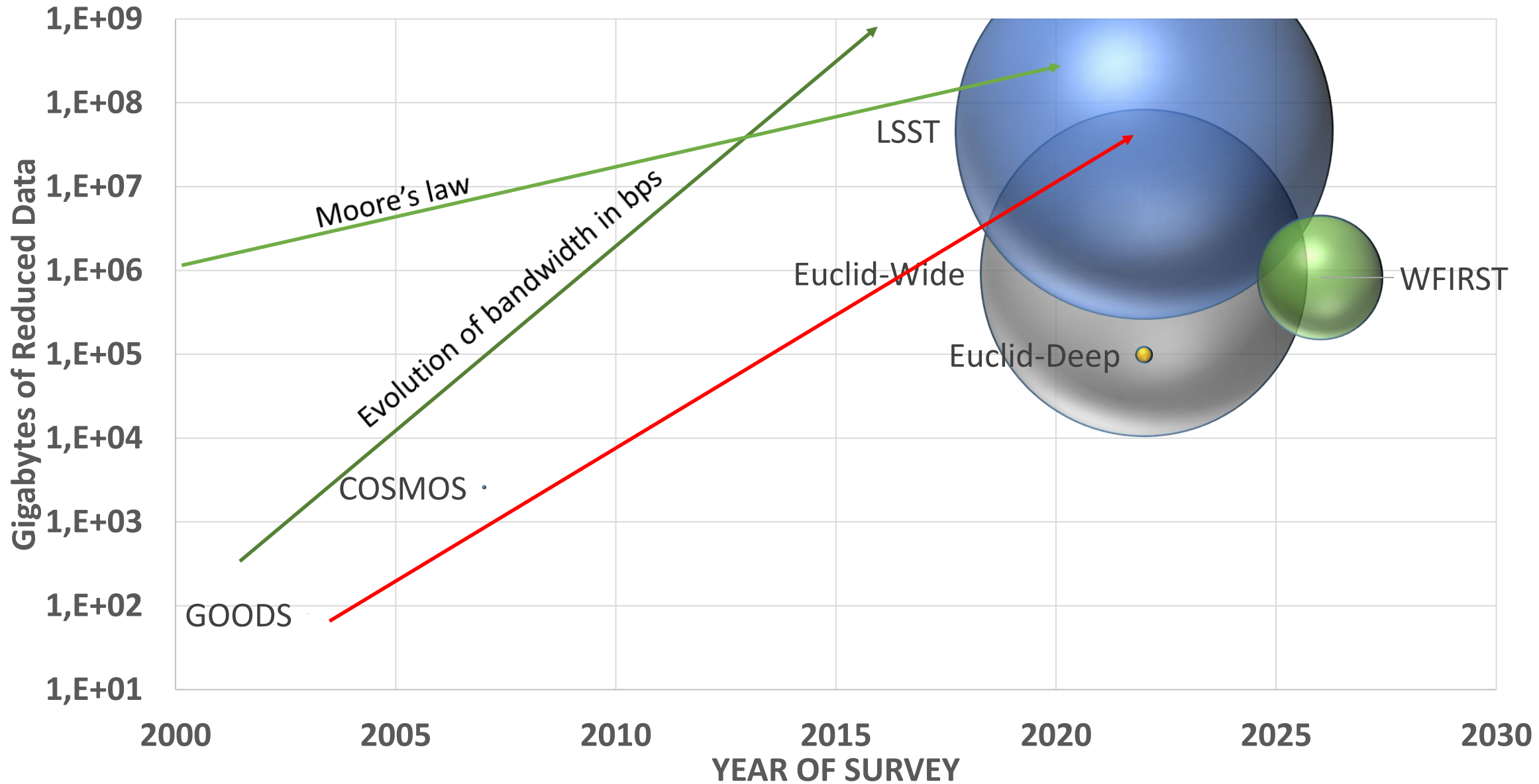
Chary, ADASS, Groningen, Oct 2019

Option 1: Fend for Yourself



Option 2: Channelize and reap the fruit
i.e. Joint Survey Processing

The evolution of multi-wavelength astronomical optical/NIR surveys



Area of bubble = relative area of survey

This naturally argues strongly for:

1. Multiplexing of compute nodes i.e. HPC
2. Getting voluminous data on high b/w connections to compute nodes

But WHY?

Key Scientific Issues

- Increasing source confusion in ground-based data with depth
 - Affects shapes for weak-lensing shear measurements
 - Astrometric offsets
 - Photometric bias due to mismatched isophotes
- Increasing need for precise photometry
 - Photometric redshifts of galaxies for dark energy studies
 - Lensed quasar time delays e.g. H0LICOWS
 - Accurate resolution matched color selection for EoR studies
 - Stellar and asteroidal variability studies
- Astrophysics in the time domain
 - Supernovae, stellar motions, AGN variability

Source confusion in 0.8'' seeing-limited data

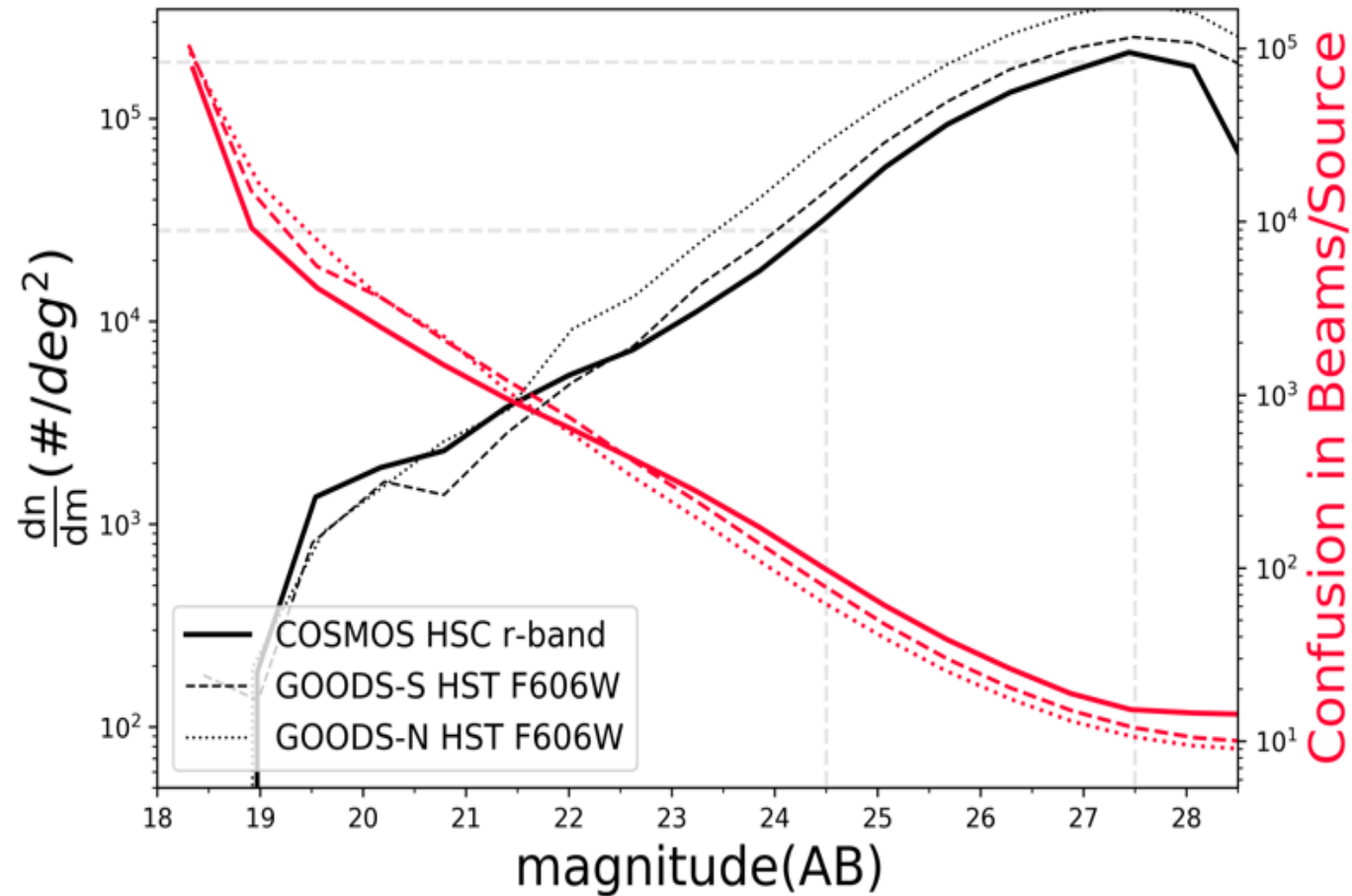
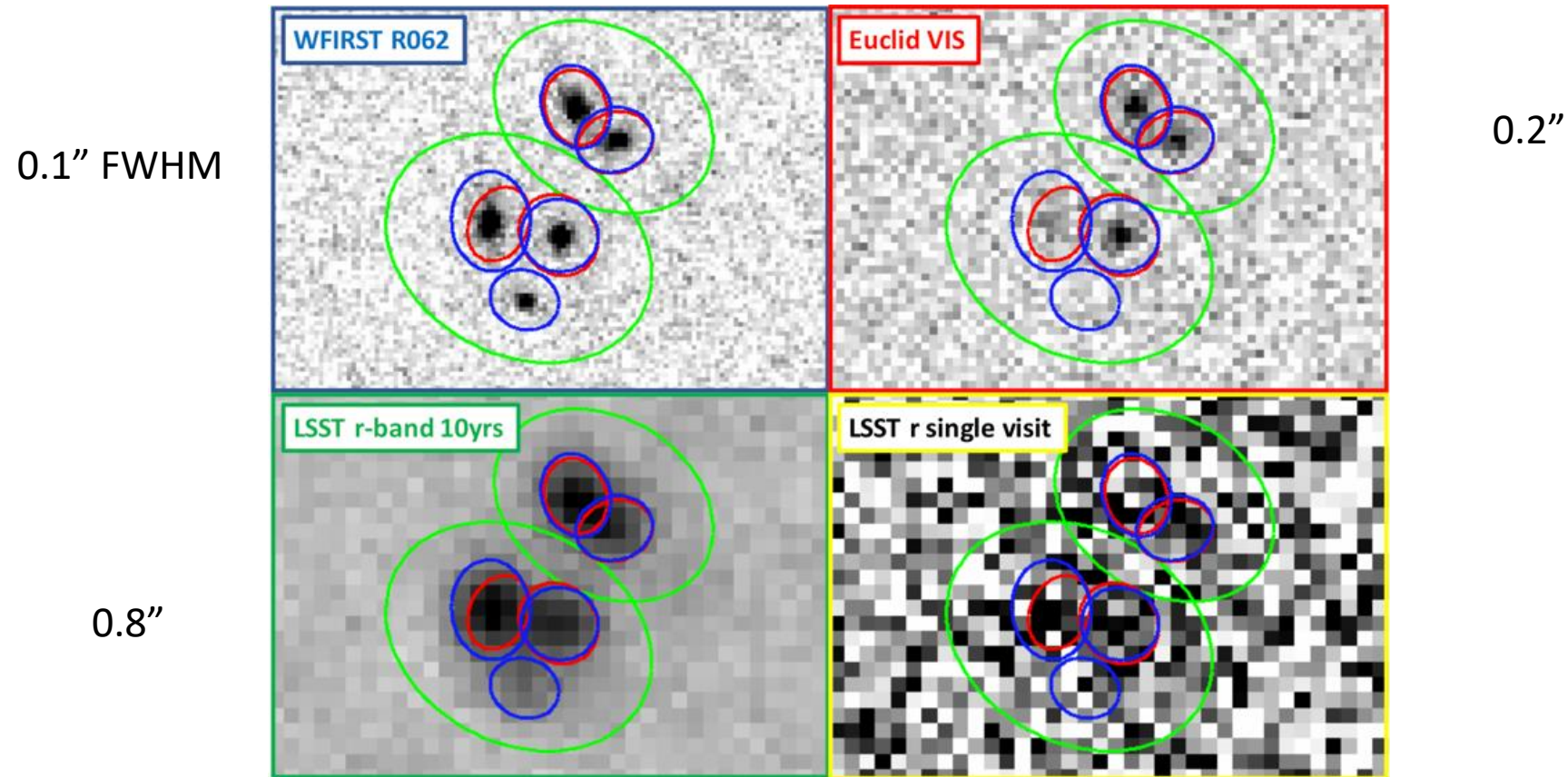
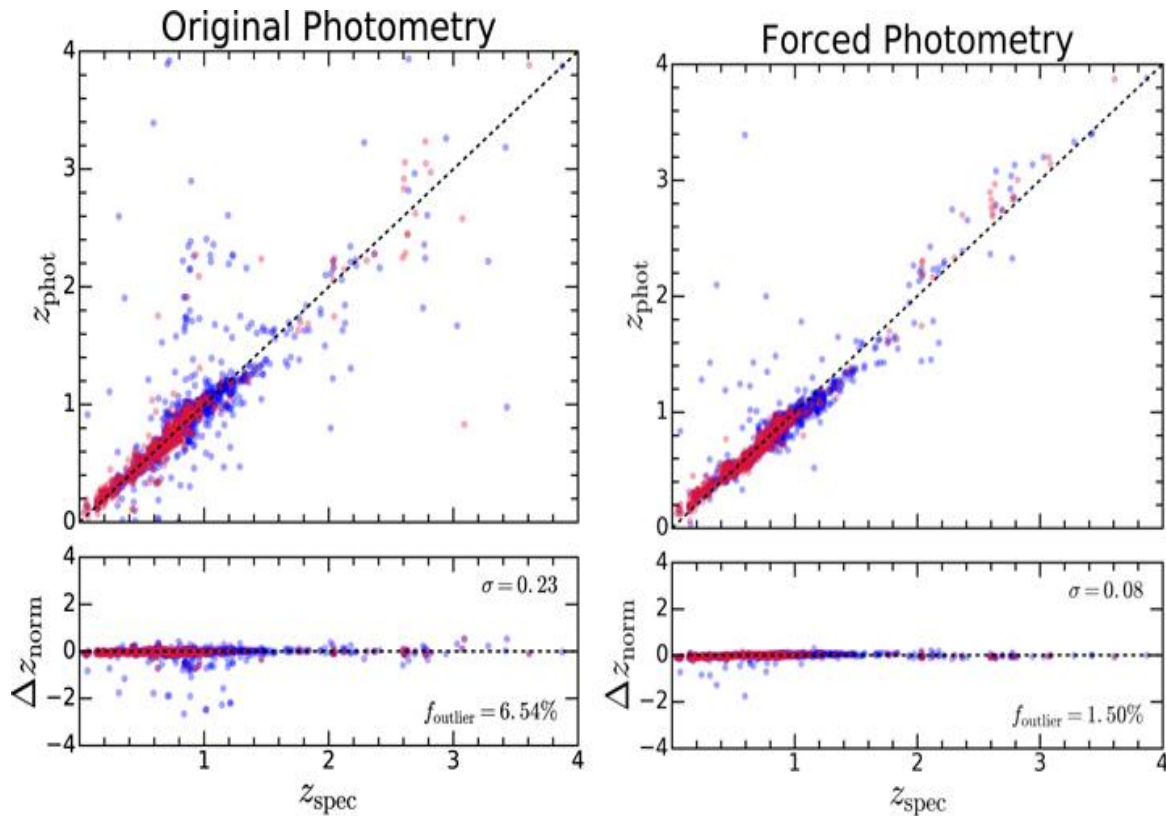


Figure by S. Hemmati

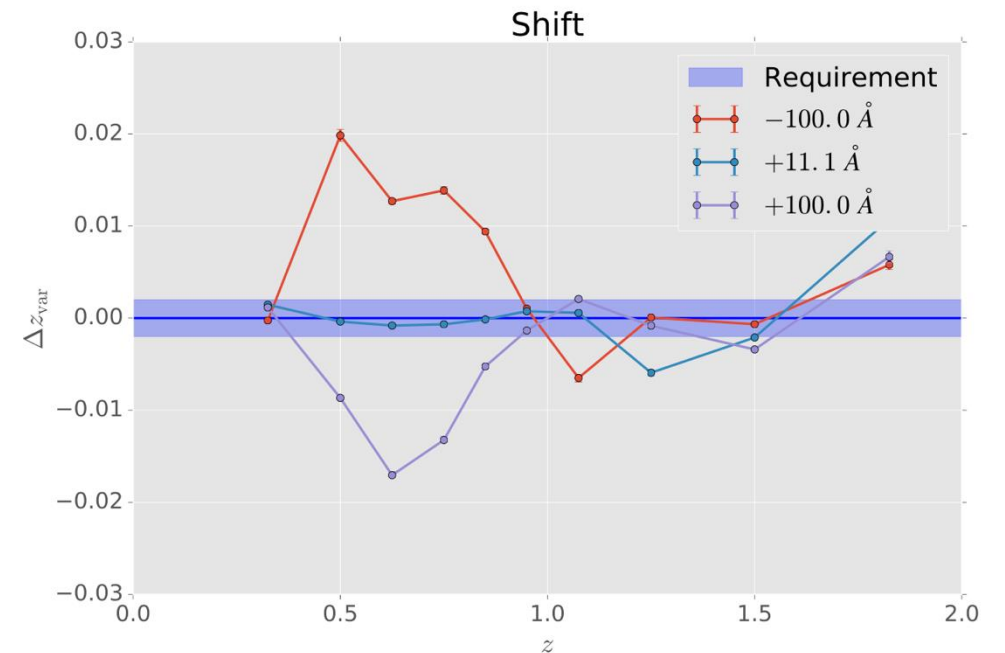
Matched isophotes, deconfusion and precision



Impact of Photometric Bias on Redshifts

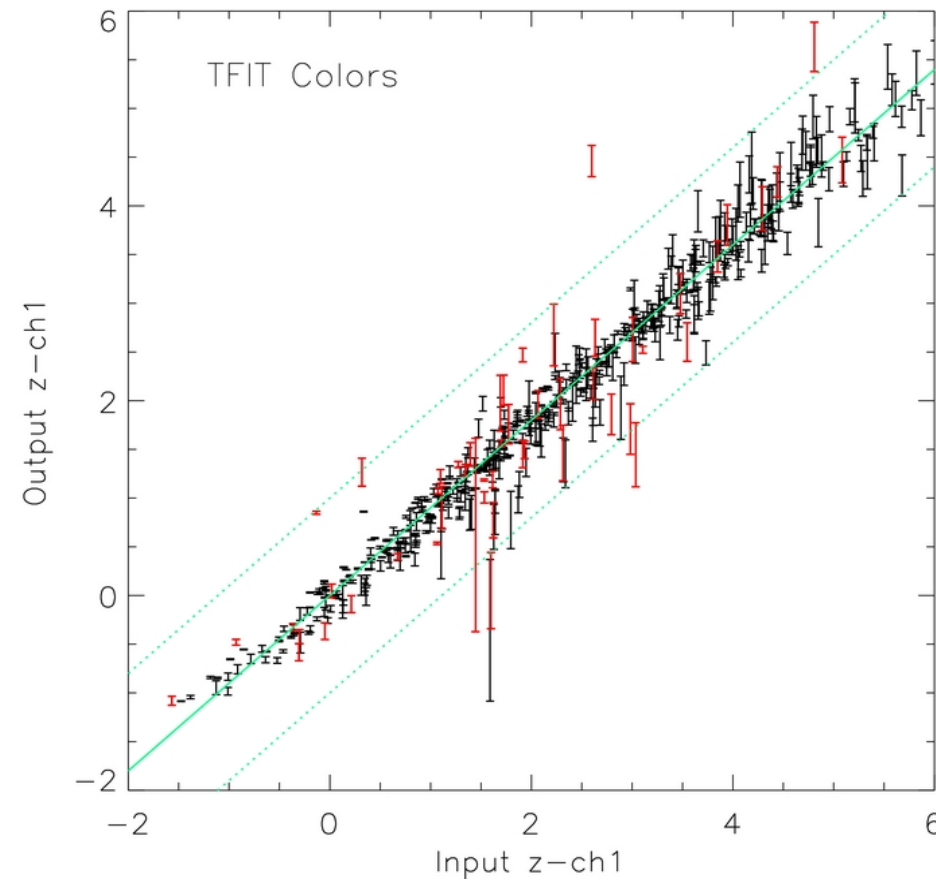
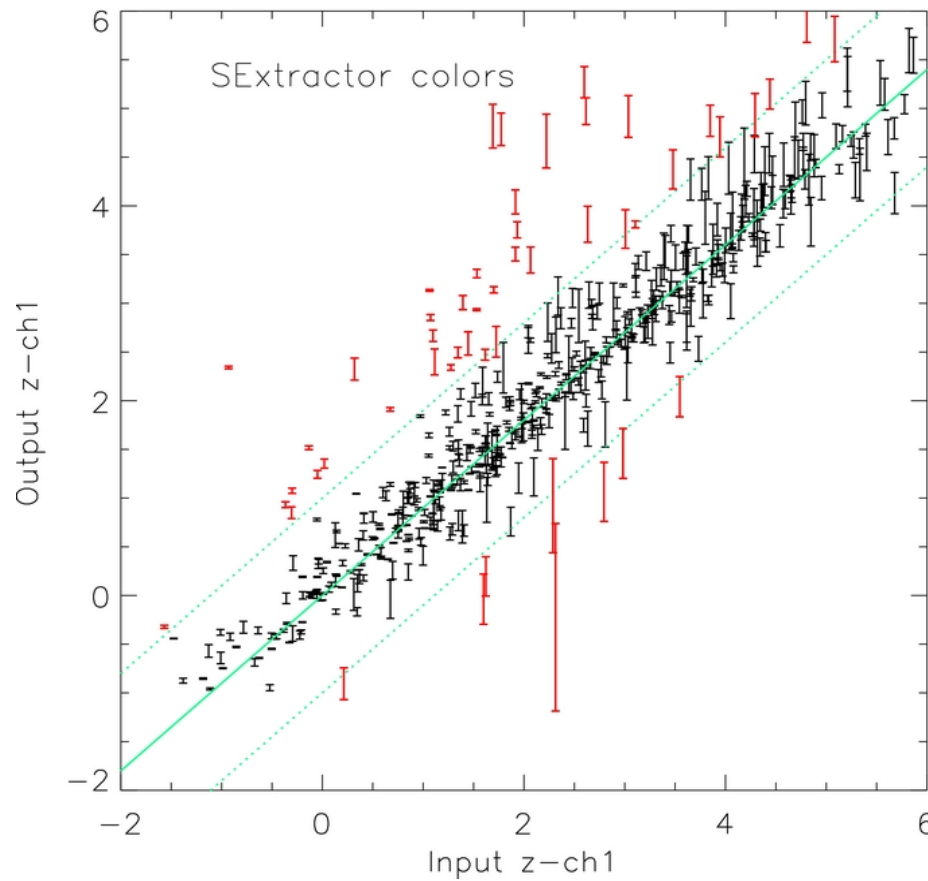


K. Nyland et al. 2017



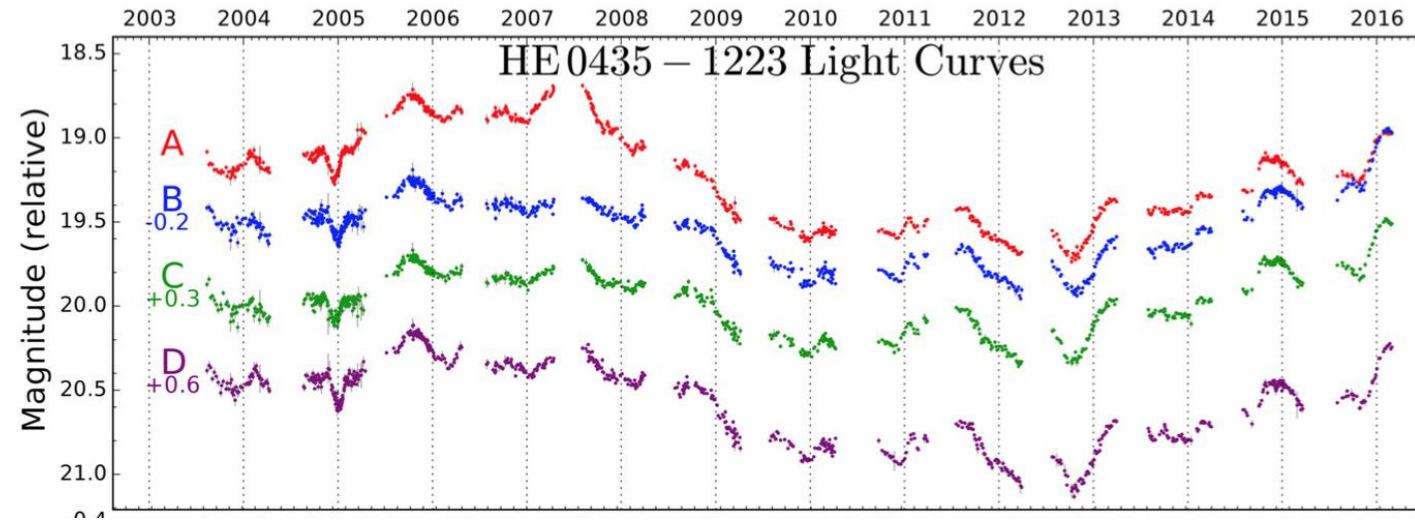
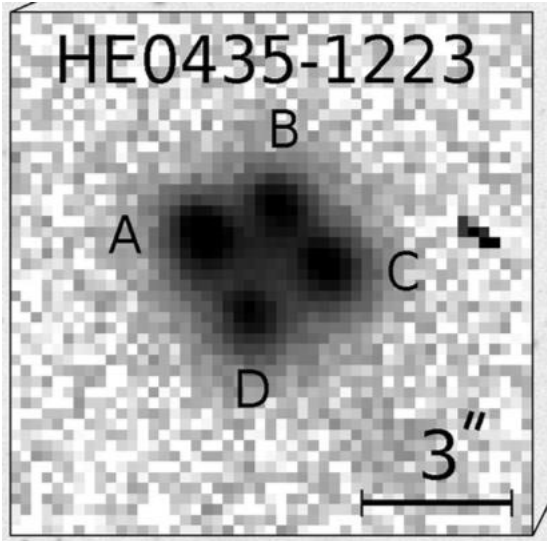
J. Coupon et al. Euclid internal report

Impact on accurate EoR color selections



K.-S. Lee et al. 2012

Strong-lens time delays in Euclid/LSST deep fields



H0liCows Team: Bonvin et al., Suyu et al., Treu & Marshall

Cadence of 6-10 days over many years

LSST will provide seeing limited data over similar timescales – photometry can be fit using priors

$$\Delta t_{ij} = \frac{D_{\Delta t}}{c} \left[\frac{(\theta_i - \beta)^2}{2} - \psi(\theta_i) - \frac{(\theta_j - \beta)^2}{2} + \psi(\theta_j) \right],$$

And deblending in slitless grism spectra

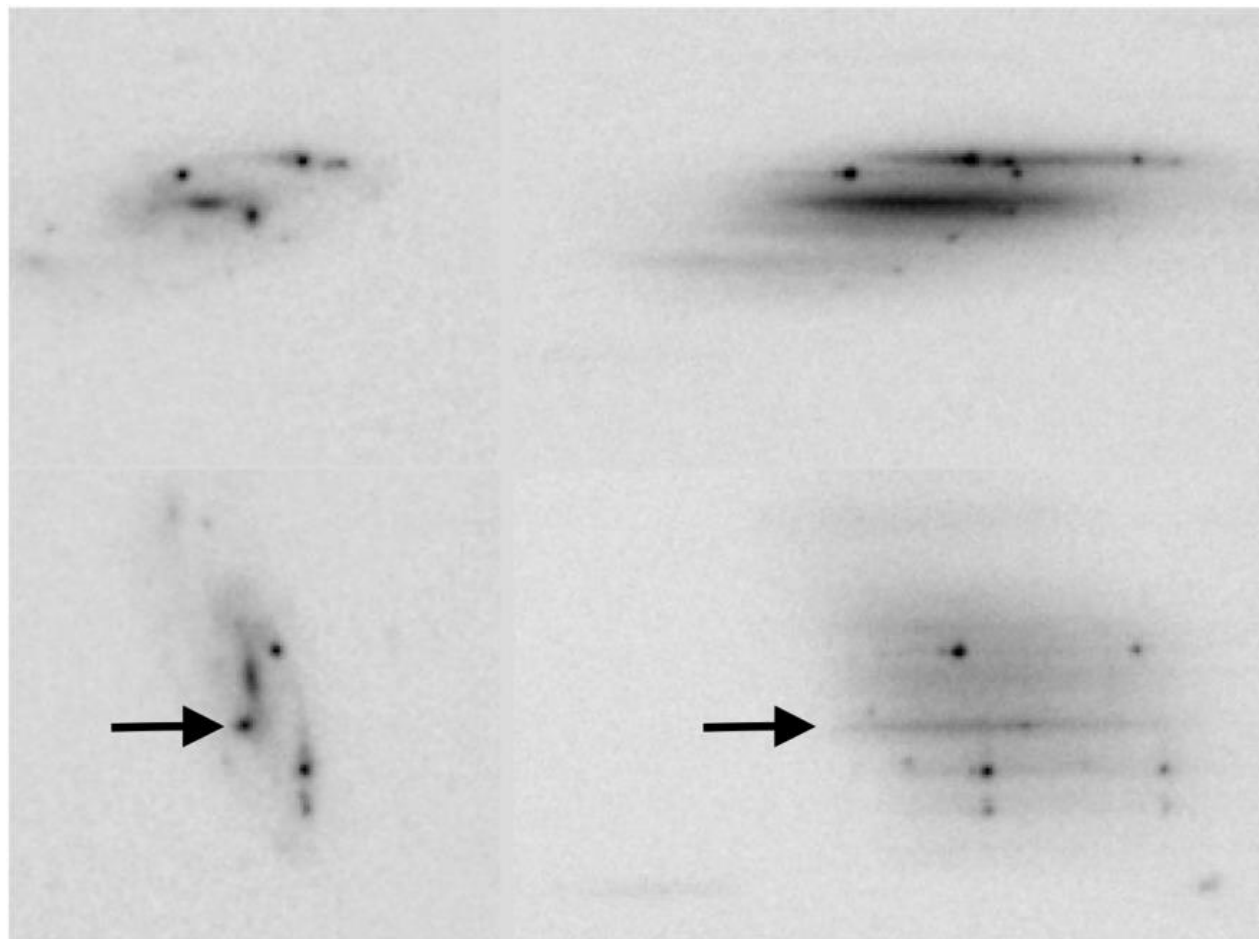
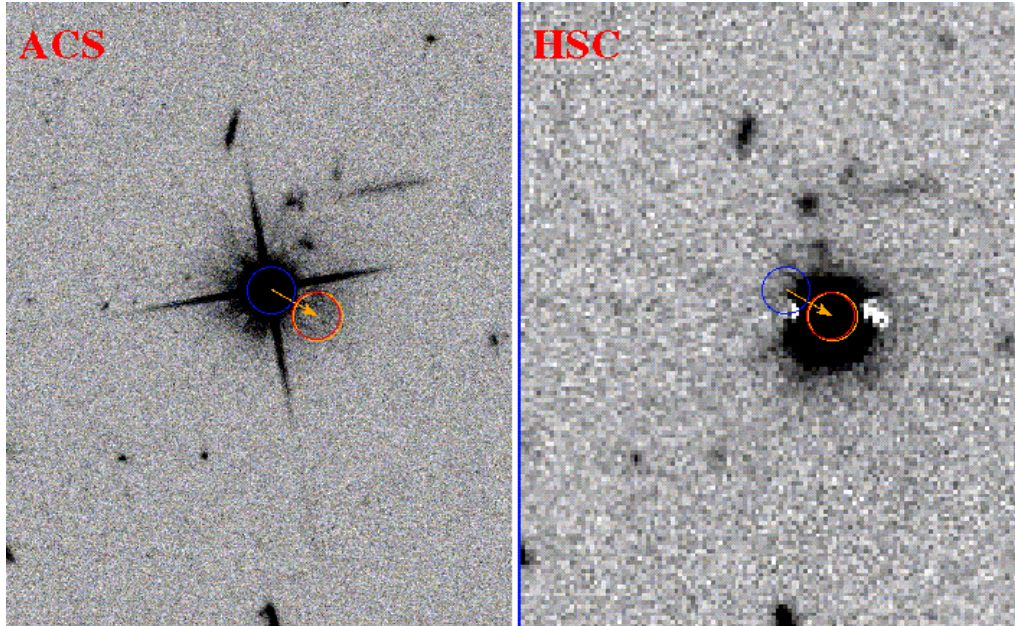
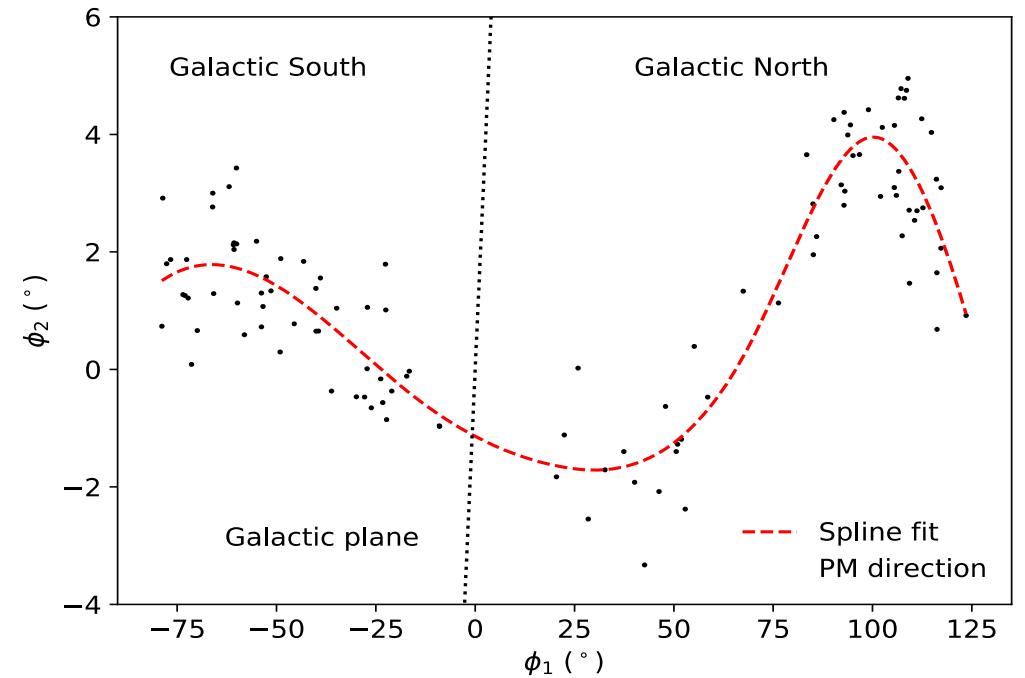


Figure from S. Malhotra

For stellar motion aficionados....

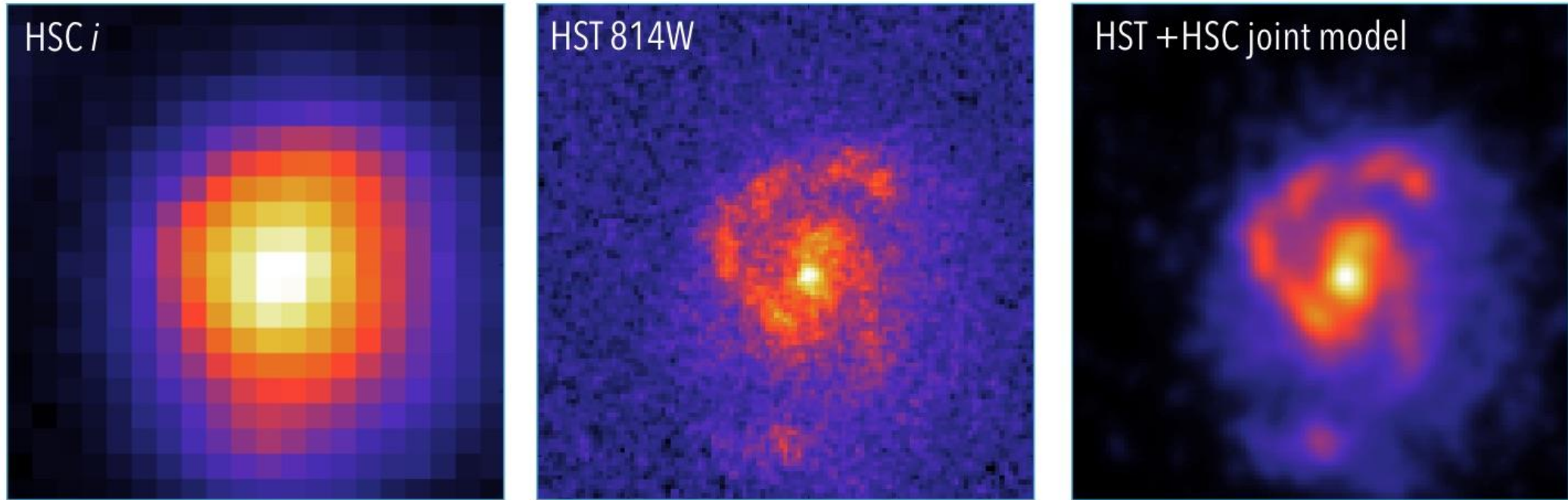


Joint astrometric studies can enable
<10mas/epoch positional measurements on individual stars



Proper motions of RR Lyrae in the Orphan Stream,
Erkal et al. 2019 indicating the tug of the LMC on the
Stream (Gaia result)

The Need: To be able to manipulate pixel level data



Joint analysis of space-based and ground-based data sets

R. Joseph & P. Melchior

Paradigm (Whole \gg Σ parts)

Details



Joint Survey Processing

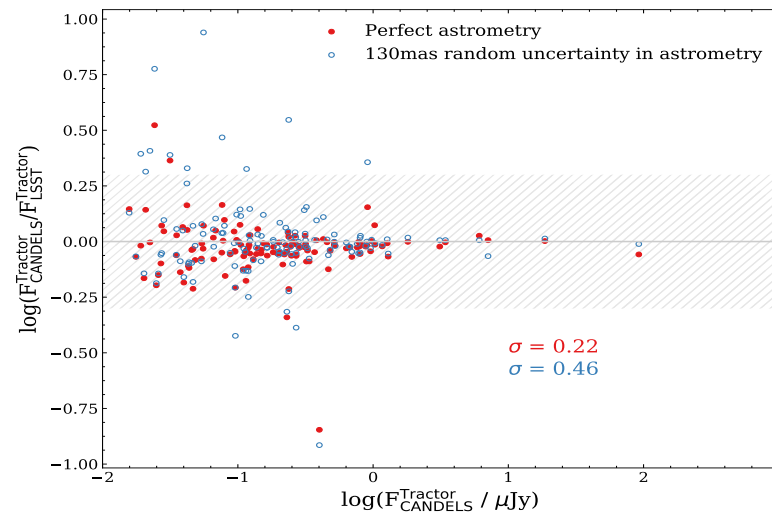
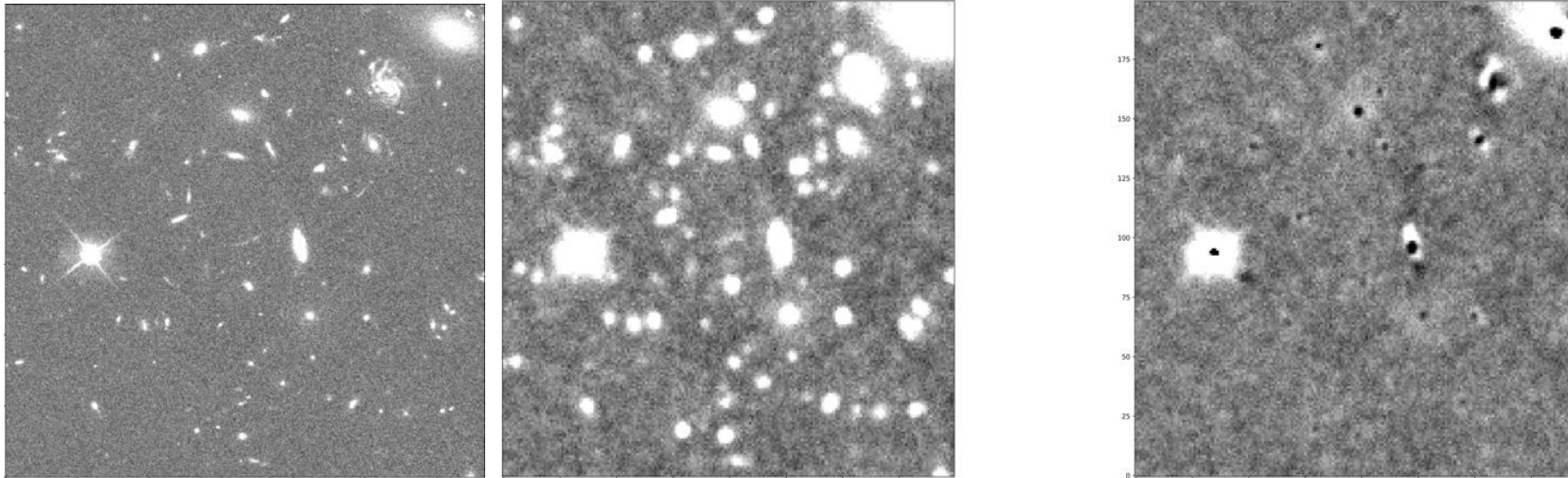
- Implementation requires
 - Algorithms e.g. TPHOT, TRACTOR, SCARLET which take as input position/color/time- dependent PSFs
 - A uniform, highly functional and scalable system architecture.

Task	a.i	a.ii	a.iii	a.iv	b.i	b.ii	b.iii	c.i	c.ii	c.iii	c.iv	c.v
	Requirements				Algorithms			Infrastructure, networking, hardware, interface				
Topics	Best phot-z for 3D WL, Breaking z-degeneracies	Time domain astronomy/SN /AGN transients/rapid response	Galactic, Solar System, Stellar Streams, Reionization	Microlensing, time baselines for SSOs	LSST Deconfusion, optimal photometry for phot-z, cross-mission color selection, dust corrections			Infrastructure, networking, hardware, interface				
Lead(s)	Newman (Upitt.)	Chary (IPAC)	Paladini (IPAC)	Helou (IPAC), Dawson (LLNL)	Lupton (Princeton) & Ferguson (STScI)	Lupton & Ferguson	Melchior (Princeton)	Dawson & A. Smith (STScI)	Dawson & A. Smith	Rusholme (IPAC)	Appleton (IPAC)	Teplitz (IPAC)
Members	Momcheva, Ferguson, Schneider, Prakash, Chary, Capak	Ferguson, Juric, Momcheva, Prakash, Capak, Armus, Wood-vasey	Ferguson, Momcheva, Wachter, Kirkpatrick, Chary, Grillmair	van der Marel, Carey, Grillmair	Dawson, Melchior, Schneider, Schulz, B. Lee, Appleton	Dawson, Melchior, Ferguson, Schneider, B. Lee, Grillmair, Armus	Dawson, Ferguson, Koekemoer, Lupton, Schulz	Schneider, Fox, Groom, Ebert	Fox, Flynn, Ebert	Smith, Fox, Groom, Berriman	Smith, Fox, Wachter, B. Lee, Rusholme, Berriman	Smith, Fox, Wachter, Groom, Rusholme, Berriman

Ways of Combining Datasets

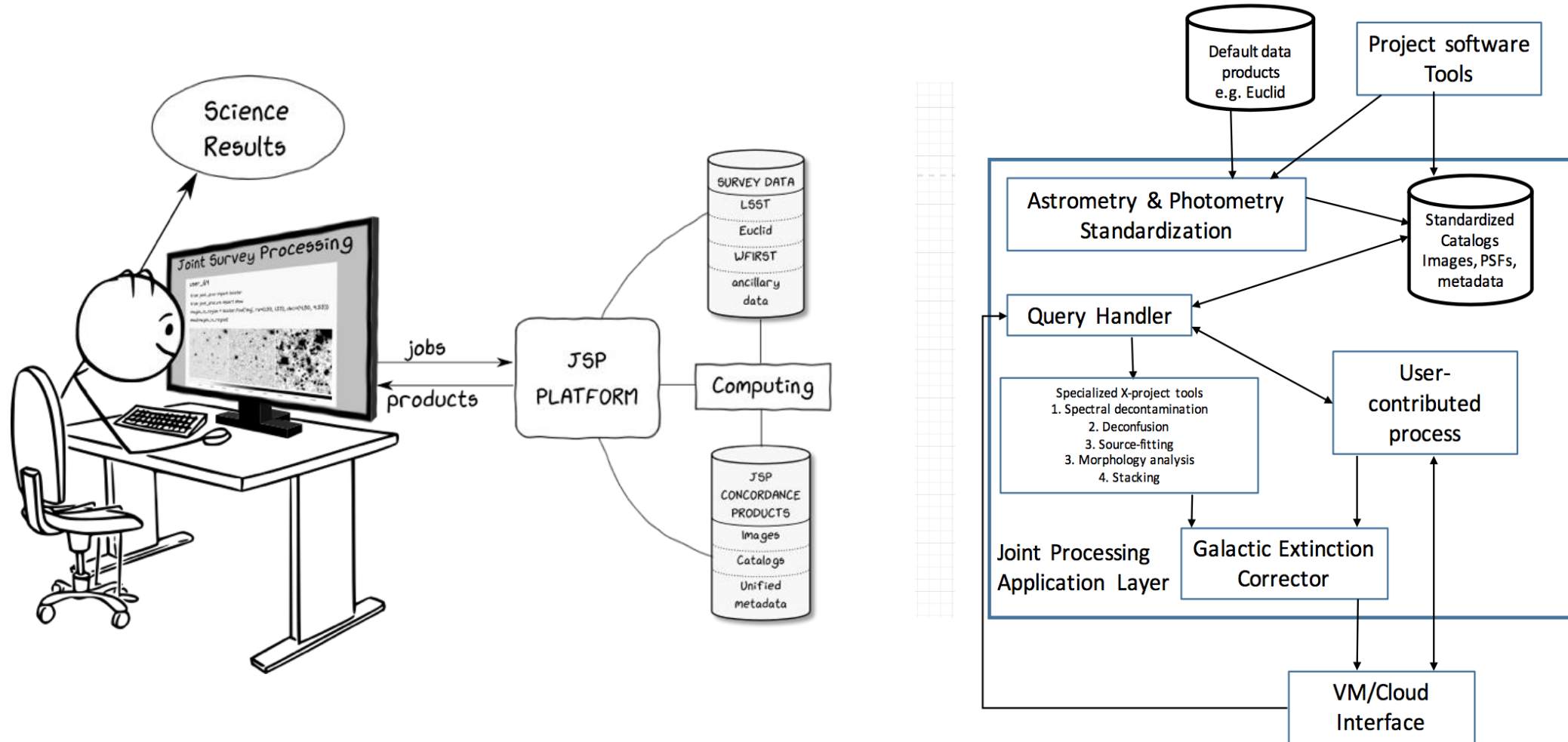
- Catalog matching [NOT joint processing]
 - Easy to handle, will be done by the projects at basic level
 - Photometric techniques from each project is different
 - Confusion and mis-association will plague the matching
- Joint fits in stacked mosaic pixel level data to photometry/positions/sizes of sources
 - Space based data can be used as a template for deconfusion
 - Consistent analysis across wavelengths, including extraction of limits
- Joint fits in single-epoch pixel level data to photometry/positions of sources
 - Astrometry from each epoch will be properly handled
 - Seeing variations will be factored into account during photometry
 - Transient sources will be properly handled, enabling selective stacks
 - Allows combining spectral grism data and imaging data
 - Most voluminous and challenging for infrastructure

Illustrative (not optimal) joint cataloging runs



It took a millimeter wave telescope (ALMA) to inform the optical/IR community that astrometry was off at 0.1-0.2''

Basic Architecture: Leveraging both existing HPC resources and cloud environment



The JSP timeline

- Started scoping activity July-2017, built team with representatives from the three projects as well as external people
- Submission to NASA, NSF, DOE in Mar 2019 and Astro2020 WPs
- Work initiated in Sep 2019
- Using DOE/NERSC supercomputing resources (100 khrs)
- Functional interface in FY2023, just in time for Euclid and LSST data.
- Updates are synchronized a few (~3-6) months after public data releases from the projects.

Summary

- We should be (very) scared of the upcoming data volume.
- Catalog matching is dead. Long live pixel level processing!
- Requires capabilities for pixel level manipulation and rapid visualization of single epoch images – the goal of JSP.
- All projects should work together with the bigger Legacy projects to have a uniform and scalable data manipulation interface with smart software for easy integration.
- Build up >>Gbps networking capabilities.
- Will need >Billion CPU hours for basic JSP, probably 10 times that for all science projects with these datasets.

Summary of Tasks

- Standardize the astrometry and photometry of the image-level data products and associated metadata available from each of these projects;
- Leverage these standardized products to generate the ultimately precise, deconfused, extinction-corrected, photometric catalogs over the entire overlapping sky area;
- Support the generation of such catalogs with other ancillary data sets that will be taken in support of Euclid and WFIRST e.g. with CFHT, Hyper-SuprimeCAM;
- Provide an interface for astronomical manipulation (e.g. shape fitting, diffuse emission extraction) of the standardized, calibrated frames from Euclid/LSST/WFIRST;
- Provide an interface for integration of the standardized Euclid/LSST/WFIRST catalogs to ancillary, lower-resolution, all-sky data products such as GALEX and WISE;
- Develop a high performance computing and networking environment which will enable manipulation of these data and allow monte-carlo simulations to be performed using the data sets from these projects.

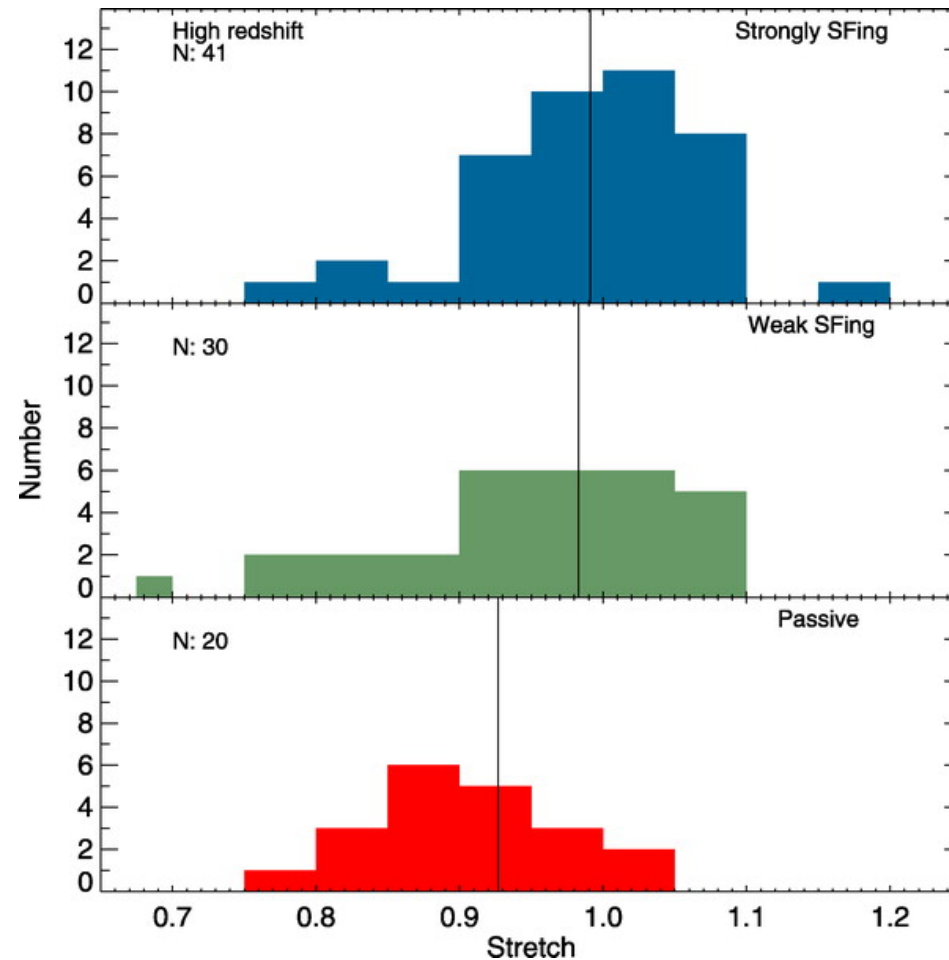
Paradigm (Whole > Σ parts)

- Well-documented set of standardized joint products
- Clear division of tasks between project, joint-processing and community and avoiding duplication
- Enables new multi-wavelength science while doing “old-science” better, reduces systematics
- Analysis and simulations are run on single copies of data located at the data processing institutions rather than moving (vast amounts) data between multiple institutions
- Provides a way for brain-power of projects to support community analysis – crucial when data-taking and processing is out of hands of end-user

Challenges

- Cannot get cross-project consistent simulations
- Data volumes are high
 - LSST ~ 70 PB of calibrated data and meta-data
 - Euclid ~ 2 PB
 - WFIRST ~ 20 PB for the whole survey
- Networking/infrastructure upgrades will be required otherwise on-the-fly analysis will take ~month

Separating transients by nature of host



Need to derive properties of hosts in which they occur
e.g. Sullivan et al. 2010 for Type Ia SNe

Microlensing

- LSST will likely undertake a microlensing survey (Rachel Street, Rosanne di Stefano)
- Both towards the bulge and out of plane. Bulge will be strongly affected by confusion, requiring WFIRST or a dedicated Euclid survey to provide priors.
- High spatial resolution extinction map will allow luminosities and spectral types of lensing systems
- Prior based deblending allows pushing detection thresholds for planets by ~ 2 mag i.e. super-Earths

