Rubin Project and Community Workshop 2022

(De)Blending: Plans and Challenges

James J. Buchanan
Lawrence Livermore National Laboratory

August 11, 2022

837263
LLNL-PRES-XXXXXX
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

Plan for This Session

- James Buchanan (Me) – Overview
- Erfan Nourbakhsh – Effect of unrecognized blends on cosmic shear inference
- Fred Moolekamp – Scarlet Lite: Status and plans
- Ismael Mendoza – BlendingToolKit, and probabilistic catalogs
- General discussion

Plan for This Session

- The first few talks: Highlight work that's been done/published/released in the past couple of years
- Ismael's talk: Shout outs to some ongoing work as well
- Synthetic Source Injection breakout earlier today
  - Won't be covered here
- The speakers will focus mainly on Dark Energy science, but blending affects all kinds of Rubin science
- This is meant to be an incitement to discussion
  - Questions! Comments! Requests!

Prevalence of Blending

- Remember: Blending is the rule, not the exception

Among distinct objects detected in HSC Wide survey, 58% have to be deblended from other objects (Bosch et al 2018)

- LSST will go deeper and wider than previous surveys -> the problem is worse than ever before

62% of galaxies detected in the full LSST survey depth will overlap another source over > 1% of their total flux (Sanchez et al 2021)

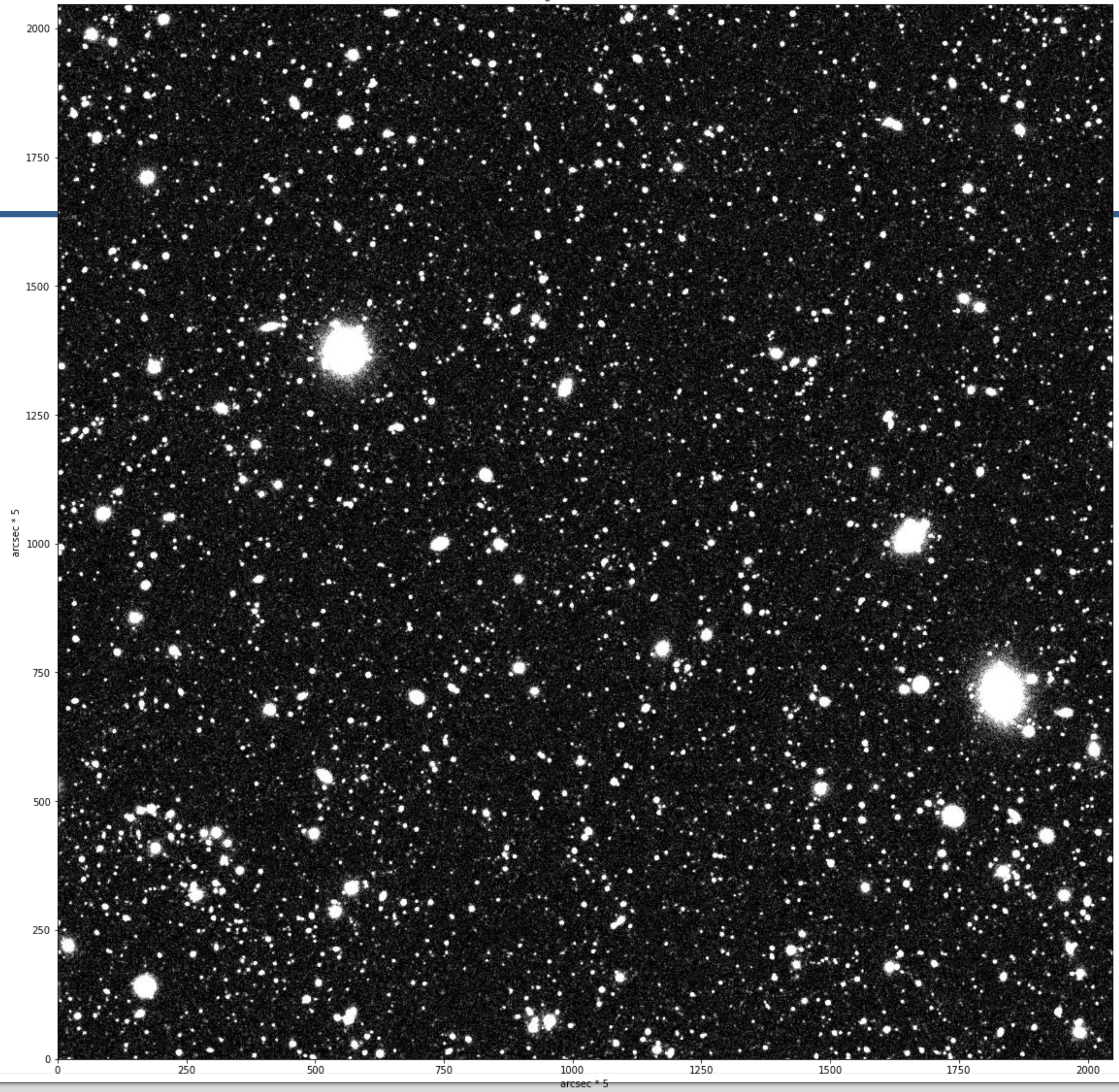
What we do now

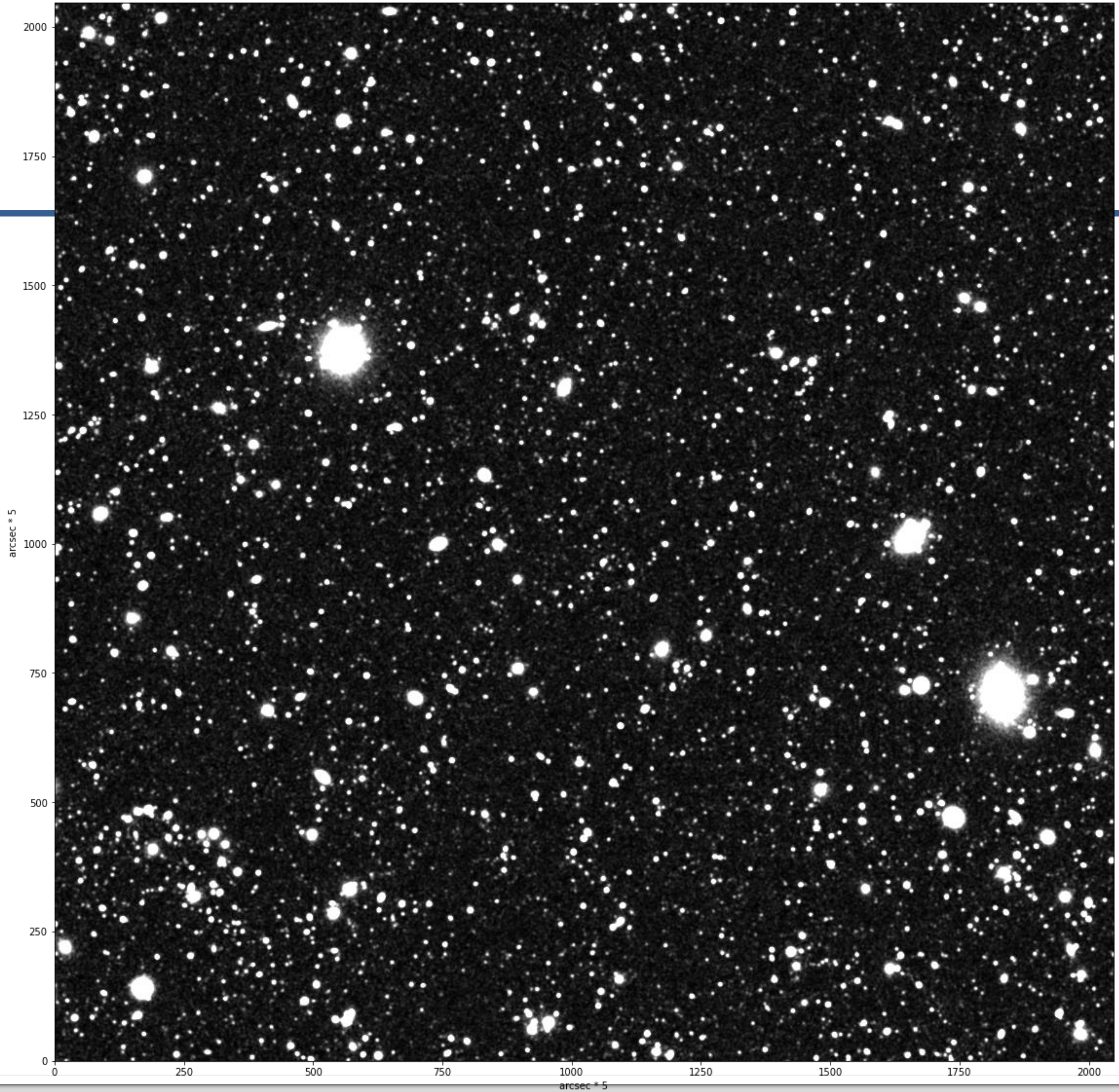
- In the standard LSST Science Pipelines workflow
- Step 0: Make coadds
- **Step 1: Make footprints**

Convolve the coadd with (a Gaussian approximation of) the PSF -> S/N image

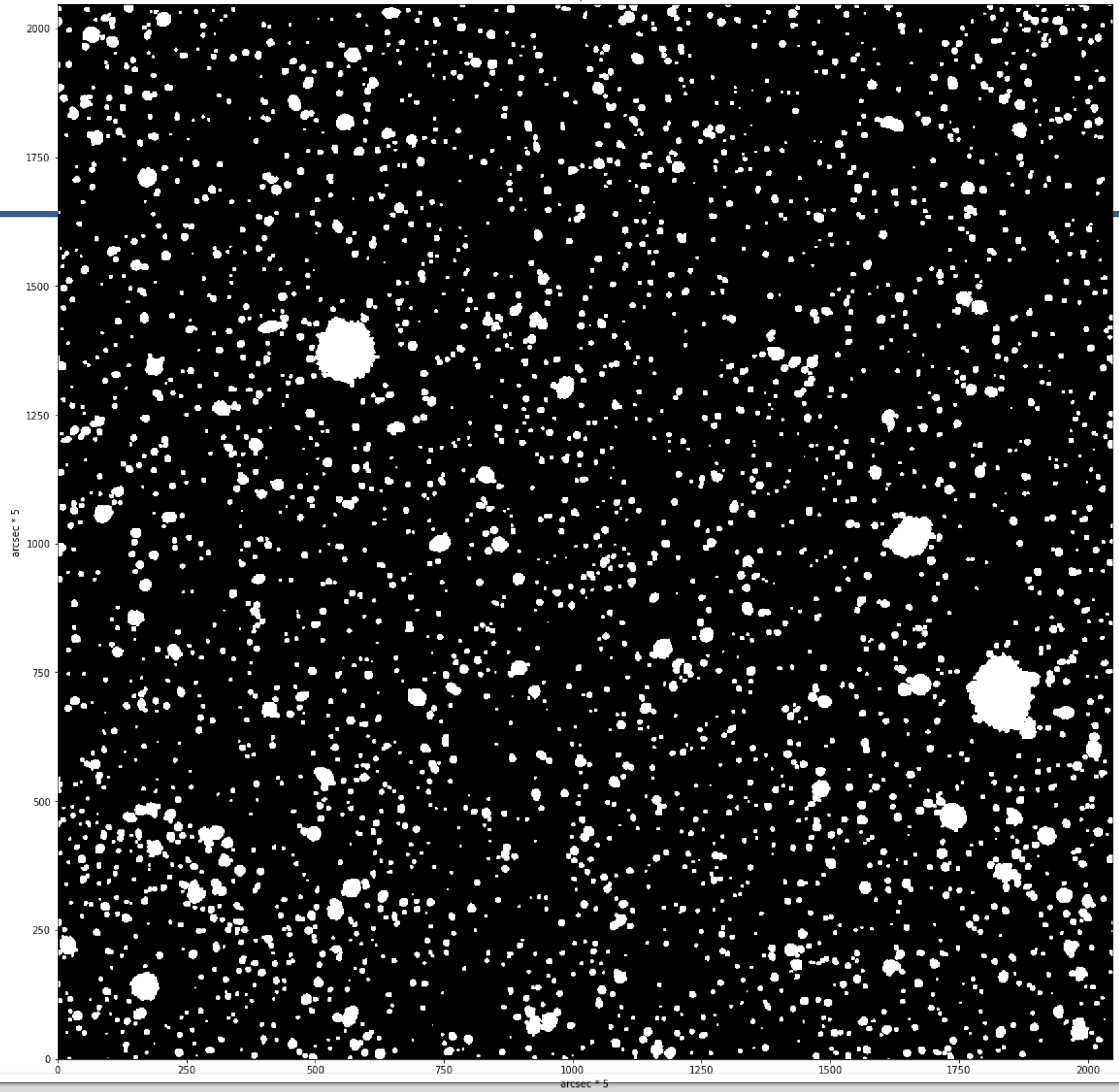
Identify pixels in the S/N image above a 5σ threshold -> "proto-footprints"

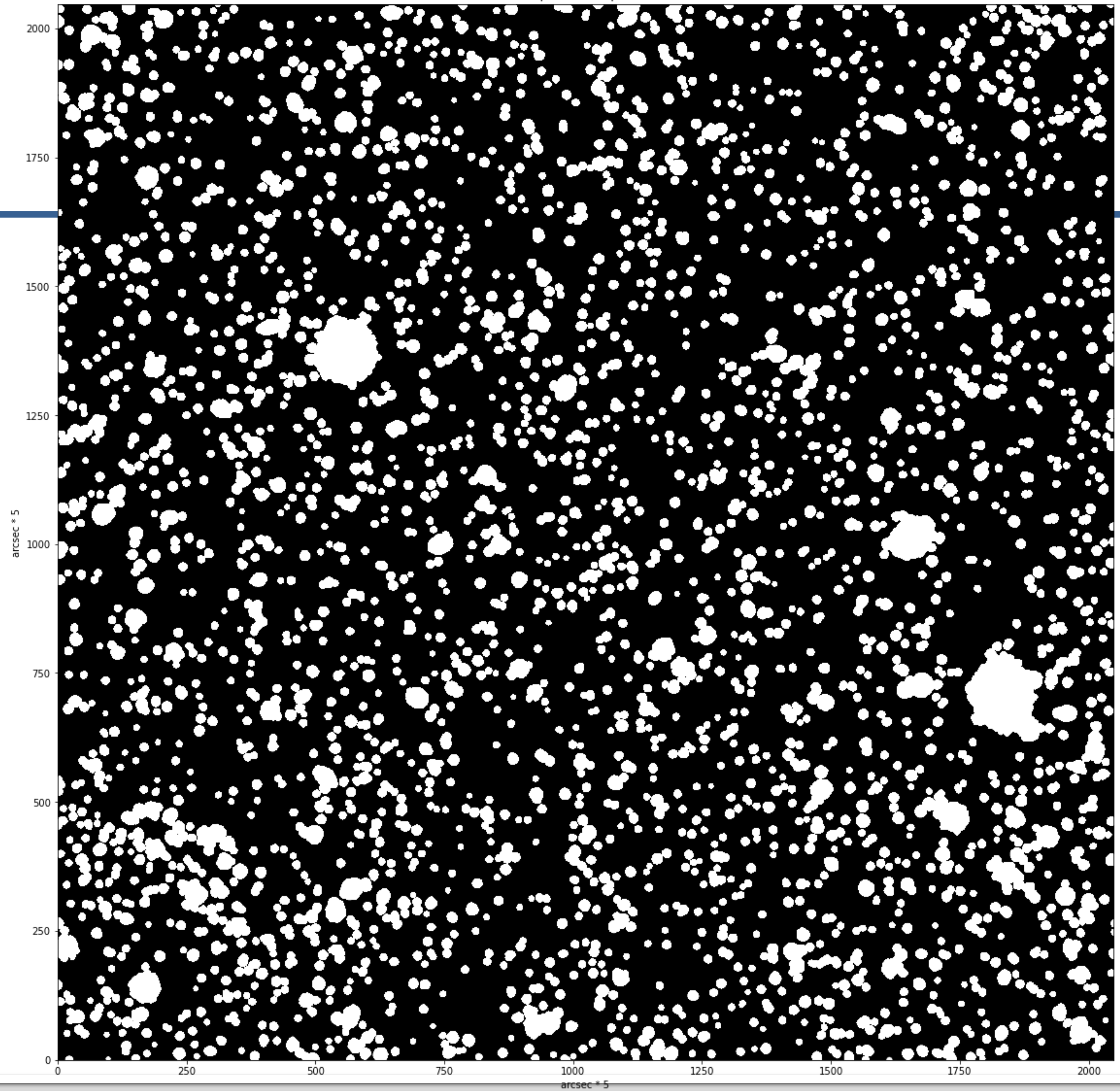
Expand each pixel in the proto-footprints by a diamond shape, 2.4 times the radius of the PSF, and merge together -> footprints


2000
1750
1500
1250
1000
750
500
250
0
arcsec * 5
0
250
500
750
1000
1250
1500
1750
2000
arcsec * 5



arcsec * 5
2000
1750
1500
1250
1000
750
500
250
0
0
250
500
750
1000
1250
1500
1750
2000
arcsec * 5


arcsec * 5
arcsec * 5
0
250
500
750
1000
1250
1500
1750
2000


2000
1750
1500
1250
1000
750
500
250
0
arcsec * 5
0
250
500
750
1000
1250
1500
1750
2000
arcsec * 5

What we do now

- **Step 2: Find peaks**

Within each footprint, peaks in the S/N image are associated to distinct objects

Temporary background oversubtraction reduces number of spurious peaks at the fringes of bright footprints

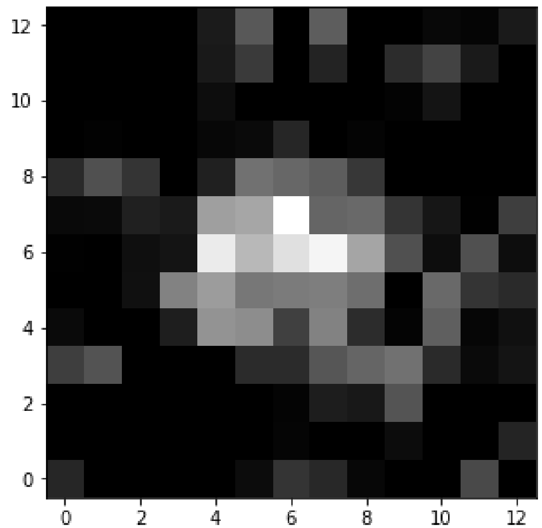
- **Step 3: Merge footprints and peaks across bands**

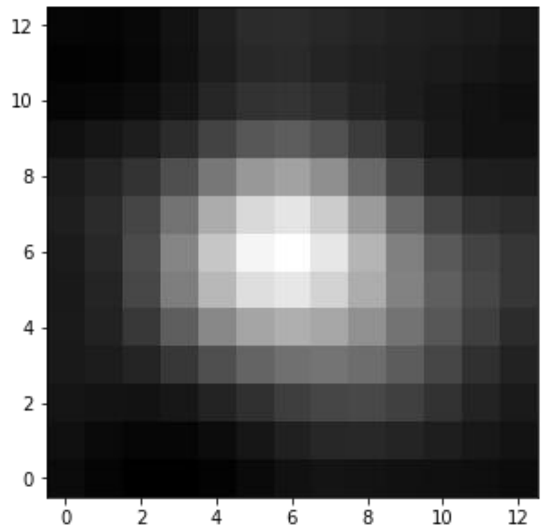
Merge peaks in “priority order” starting from i-band: peaks in different bands are the same if < 0.3”, different if > 1”

- **Step 4: If a footprint contains 2 or more peaks, deblend** (Fred’s talk) -> Finally, make measurements on the deblended outputs

Peak finding

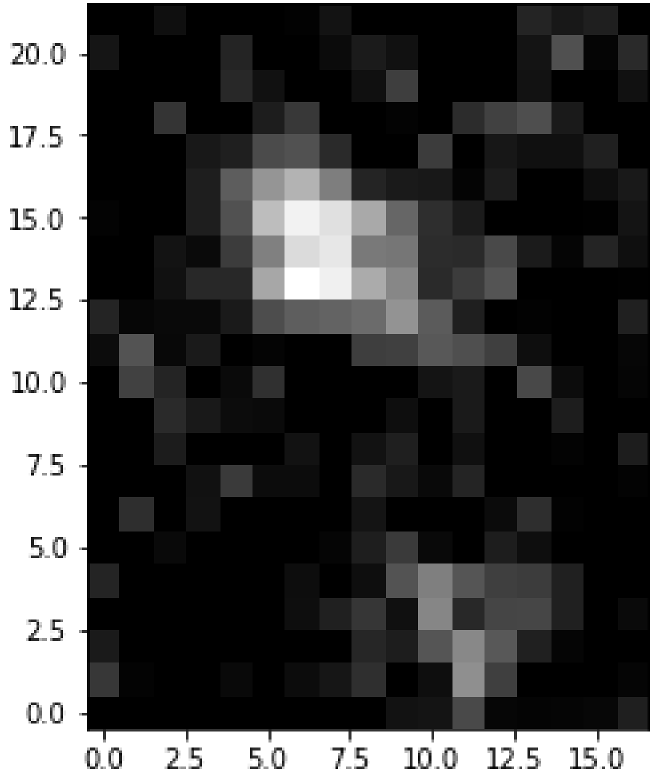
- Peak: Pixel in the smoothed image that is at least as bright as all 8 of its neighbors

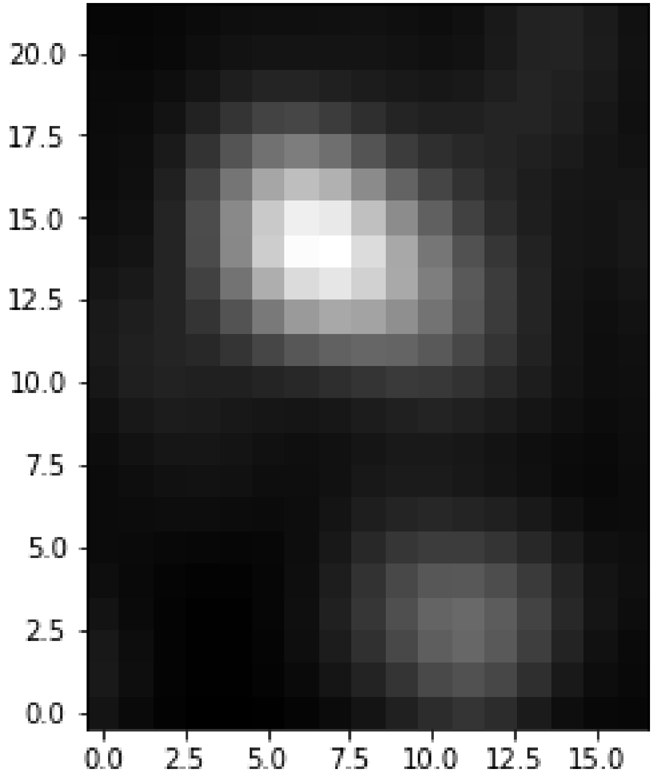




Peak finding

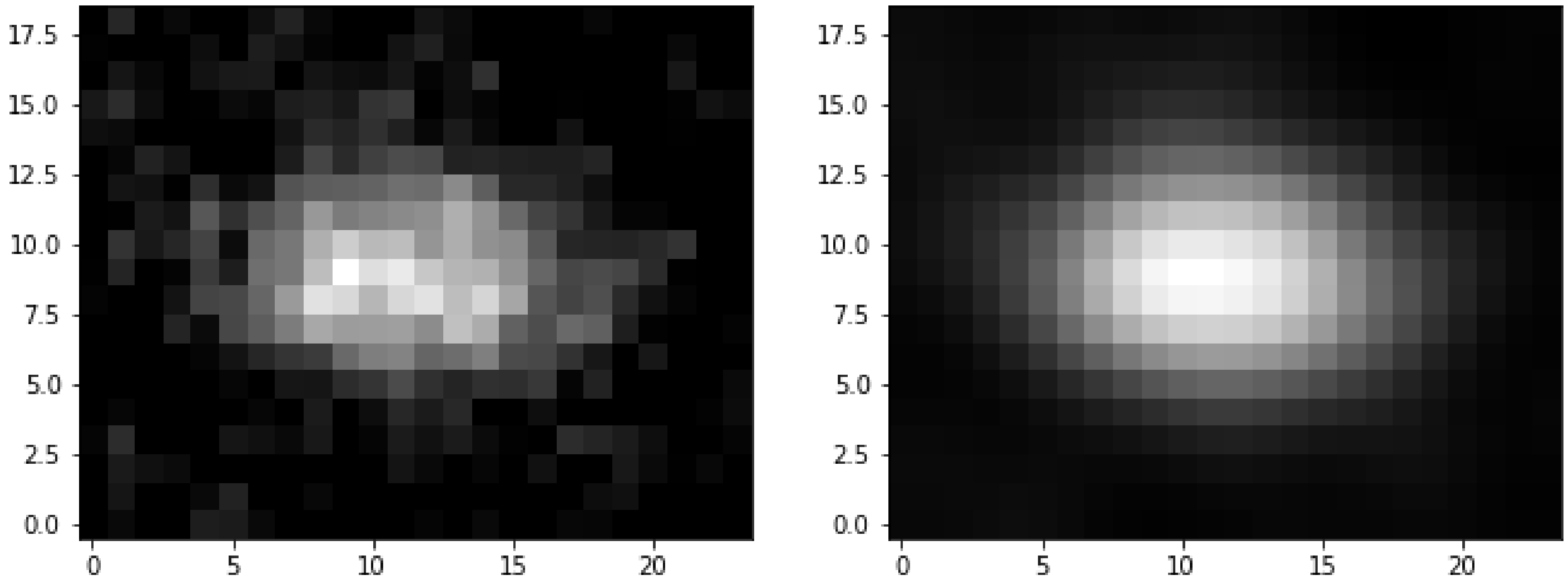
- Peak: Pixel in the smoothed image that is at least as bright as all 8 of its neighbors





Peak finding

- Peak: Pixel in the smoothed image that is at least as bright as all 8 of its neighbors



Some Reasonable Questions

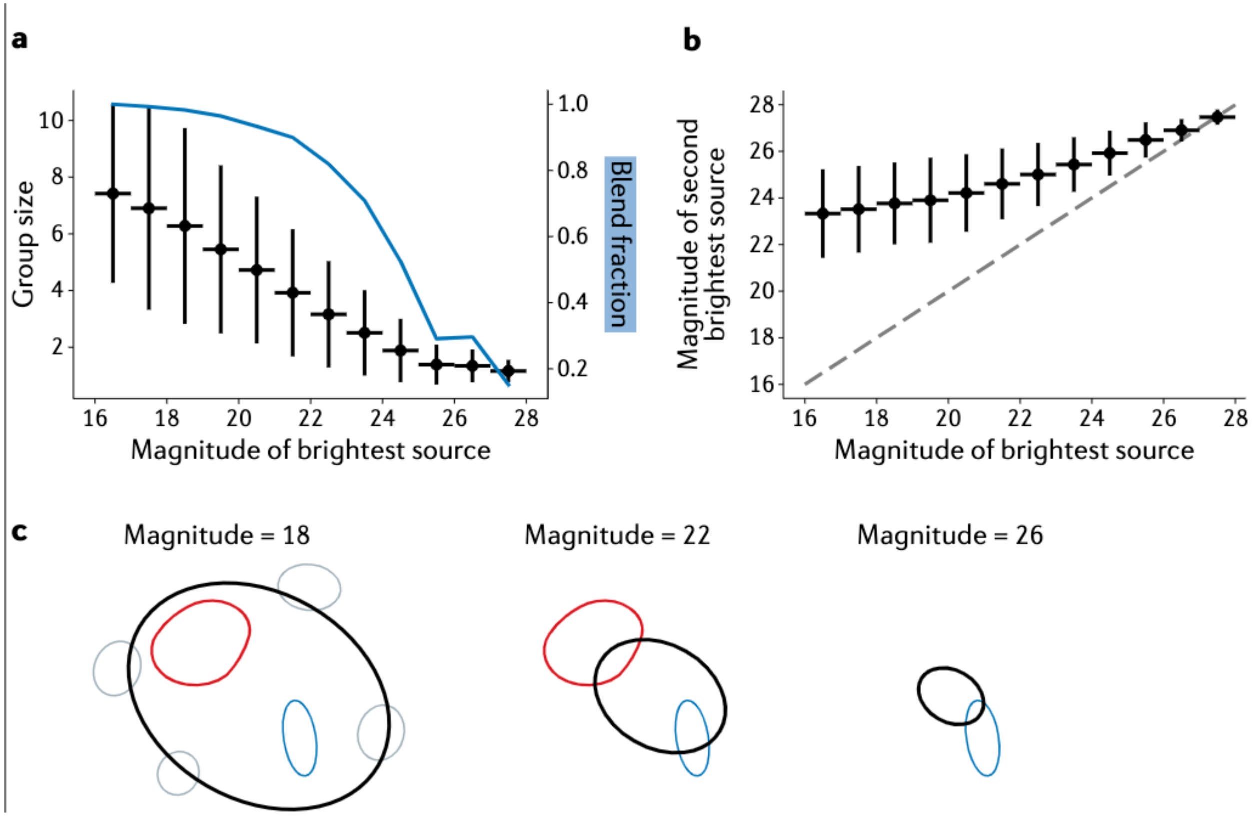
- Is a 5σ detection threshold optimal for our science?
  - Lowering this threshold would tend to create more + wider footprints
- Are we expanding footprints in the optimal way?
  - Diamond shape? 2.4?
  - This interacts with the detection threshold: reducing the amount of expansion partly compensates for a lower detection threshold
- Is this the best way to detect objects?
  - Is peak finding the best way to count the number of objects in a footprint?
- One paper that looks at all of these: Buchanan et al 2022

What Does “Blended” Mean?

- Broadly speaking: An object is significantly blended with another if it isn’t isolated, i.e. its properties cannot be well-measured without special compensation (“deblending”)

What Does “Blended” Mean?

- Melchior et al 2021: **objects** are blended if they’re in the **same detection footprint**



What Does “Blended” Mean?

- LSST Science Pipelines: **detected objects** are declared blended if they’re in the **same footprint**

- Buchanan et al 2022: A ***footprint*** is “**truly**” blended if it contains 2 or more actual objects w/ “sufficient” true flux
  - About 1/3 of LSST *footprints* are blended in i-band

- Sanchez et al 2021: An ***object*** is “**truly**” blended if “enough” of its true flux overlaps another object
  - About 2/3 of LSST *objects* are blended

What Does “Blended” Mean?

- Blending is not all-or-nothing – some blends are worse for science than others

What Does “Blended” Mean?

- Blending affects not only the estimated measurements of objects, but also the estimated **number** of objects

- A large, bright galaxy will:

prevent faint overlapping galaxies from being detected or well-measured

produce spurious detection peaks in its outer edges

  – corrected for by local background oversubtraction during detection

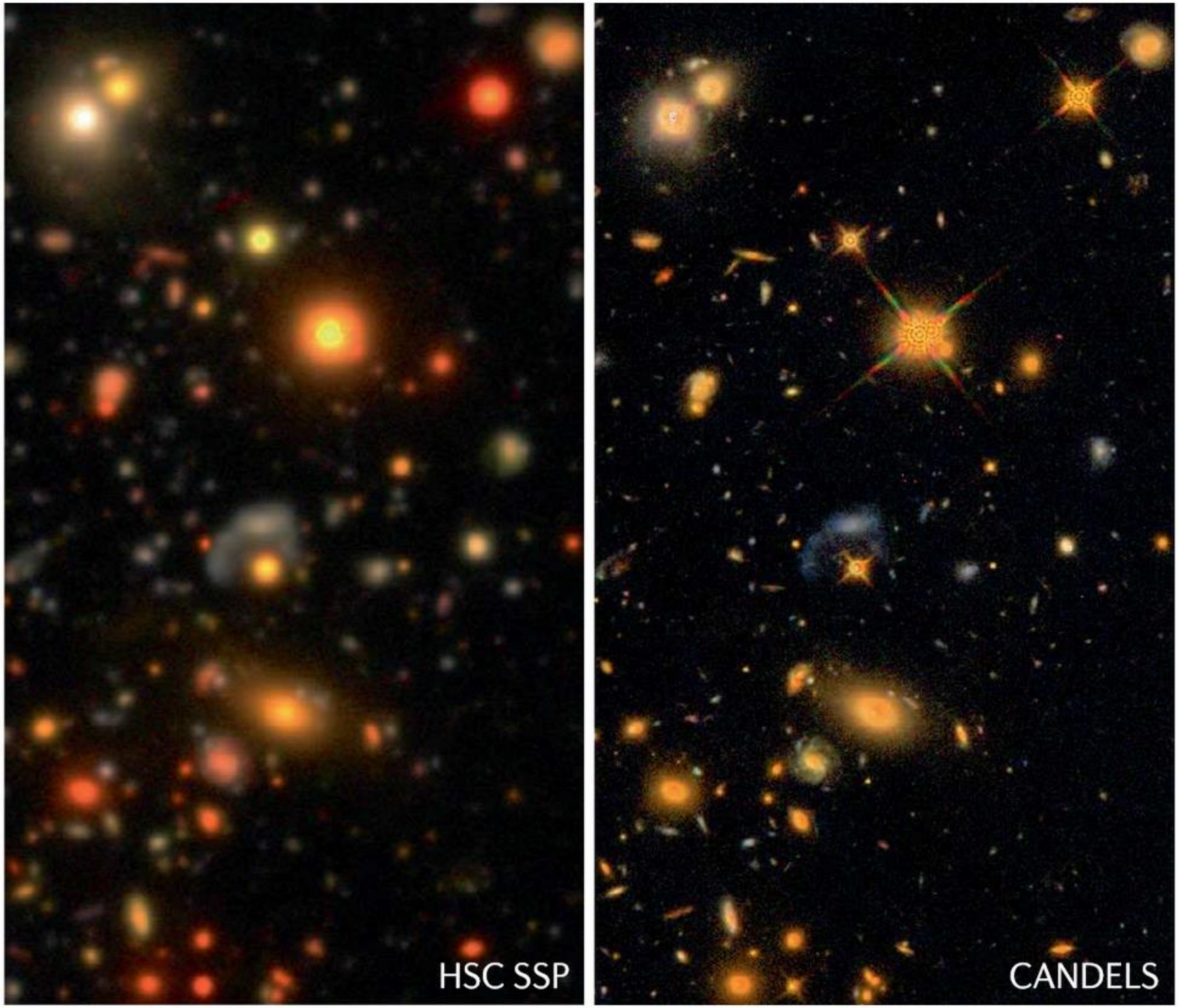
What Does “Blended” Mean?

- Too many peaks in a footprint -> Spurious detections/“shredding”
- Too few peaks -> Unrecognized blends
- See e.g. Dawson et al 2016, Buchanan et al 2022, Nourbakhsh et al 2022 for estimates of the scope of the problem, its impact on cosmic shear inference, and how we might mitigate it

What Does “Blended” Mean?

- In ground-based surveys, PSF blurring often causes objects to appear blended even when they don’t “truly” overlap

(Image from [Melchior et al 2021])



What Does “Blended” Mean?

- When objects don’t truly overlap, the light from their separate images simply adds together
  - Common assumption in blending algorithms, image simulations
  - Used in DC2 simulations
- When they do overlap, have to account for **obscuration/occultation** by foreground dust (see e.g. Gaztanaga et al 2021)

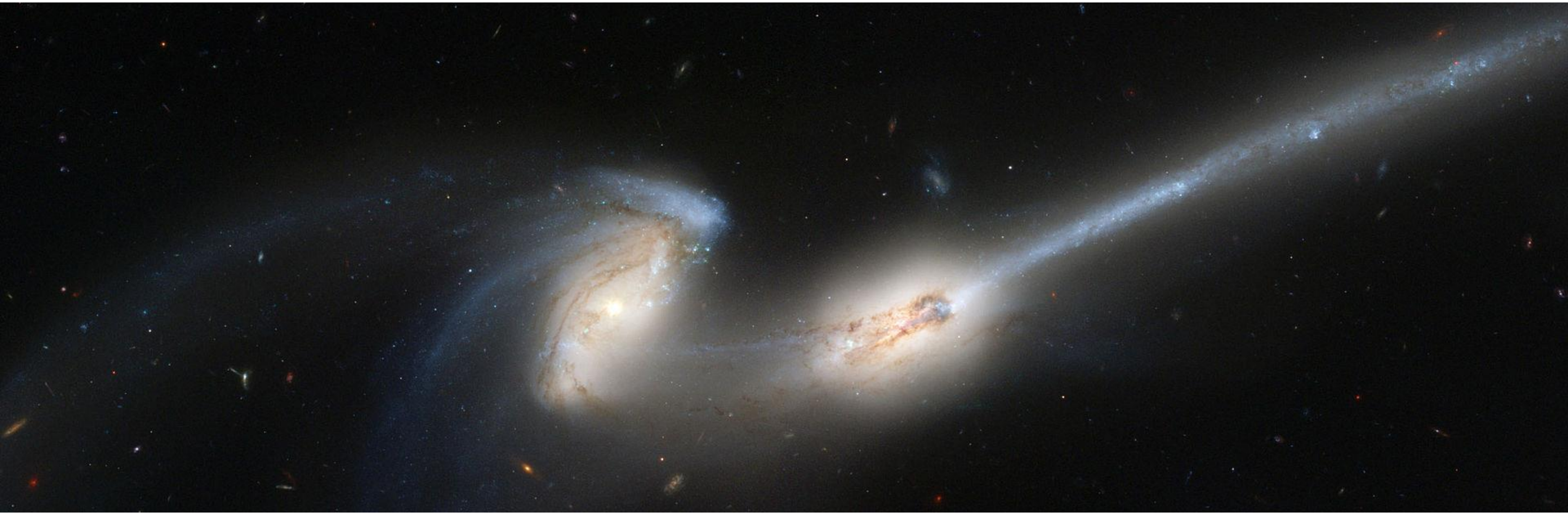
What Does “Blended” Mean?

- Real occultations are difficult to identify (see e.g. Keel et al 2013, Holwerda et al 2015)
- Currently don’t see much community effort to model this in LSST blending studies
  - Reiman & Göhre 2019, for simulated blend images, use the max pixel intensity of two galaxies instead of adding them together

Hubble image of NGC 253

What Does "Blended" Mean?

- Genuine galaxy mergers occur with some regularity (e.g. Walmsley et al 2022)
- How do deblenders treat these? How do we want them to be treated?



Hubble image of NGC 4676 A&B

So You Think You Can Build a Better Deblender

- The Science Pipelines perform detection and deblending in a very specific way – your deblending studies are *most* relevant to early LSST science if you use the Pipelines for comparisons
  - i.e. footprints + Scarlet
  - As opposed to:
    SExtractor, Fisher matrix forecasts

So You Think You Can Build a Better Deblender

- Several great "proof of concept" deblender studies published, with more under development. These are always welcome! We need to keep exploring creative new possibilities to take maximal advantage of our data.

- Some relatively recent efforts:

- Arcelin et al 2021

- Liu et al 2021

- Hausen 2021

So You Think You Can Build a Better Deblender

- Several great “proof of concept” deblender studies published, with more under development. These are always welcome! But to do better science in the end, we need to demonstrate several important things:

So You Think You Can Build a Better Deblender

- Several great "proof of concept" deblender studies published, with more under development. These are always welcome! But to do better science in the end, we need to demonstrate **value**:
- Science impact studies: Does the method improve some final measurements?
  - Without degrading other measurements?
- Ongoing studies of blending on photo-z estimation, clustering inference, etc.
- Today: Impact of blending on cosmic shear inference of S_8 (Nourbakhsh et al 2022)

So You Think You Can Build a Better Deblender

- Several great “proof of concept” deblender studies published, with more under development. These are always welcome! But to do better science in the end, we need to demonstrate **reliability**:
- Validation: Does the method yield stable, well-understood results on all sorts of inputs?
- Today: A community tool for systematically testing deblenders on all sorts of inputs (BlendingToolKit)

So You Think You Can Build a Better Deblender

- Several great "proof of concept" deblender studies published, with more under development. These are always welcome! But to do better science in the end, we need to demonstrate **reliability**:

- Validation: Does the method yield stable, well-understood results on **realistic data**?
  - Realistic galaxy shapes, realistic number densities at different wavelengths
  - Works on full images, in addition to idealized cutouts
  - Robust to variations in PSF, background level, noise level
  - Handles junk inputs gracefully
  - Handles different priors gracefully (for Bayesian methods)

So You Think You Can Build a Better Deblender

- Several great “proof of concept” deblender studies published, with more under development. These are always welcome! But to do better science in the end, we need to demonstrate **tractability**:

- Concrete implementation strategy: Specific plans for **who** runs it (Rubin, DESC, someone else), what **resources** are needed to run on the full survey, what specific input data is needed, **how input data is accessed**, **what specific data products are produced**, **how those are distributed**, how those are used for scientific inference

- Today: Current status and plans for Scarlet in the Science Pipelines

That's Blending in a Nutshell

- For more of an introduction, check out this Nature Reviews Physics Perspective article: Melchior et al 2021
- Get in touch with the DESC Blending Working Group
  - Conveners: Me and Cyrille Doux
  - DESC Confluence page