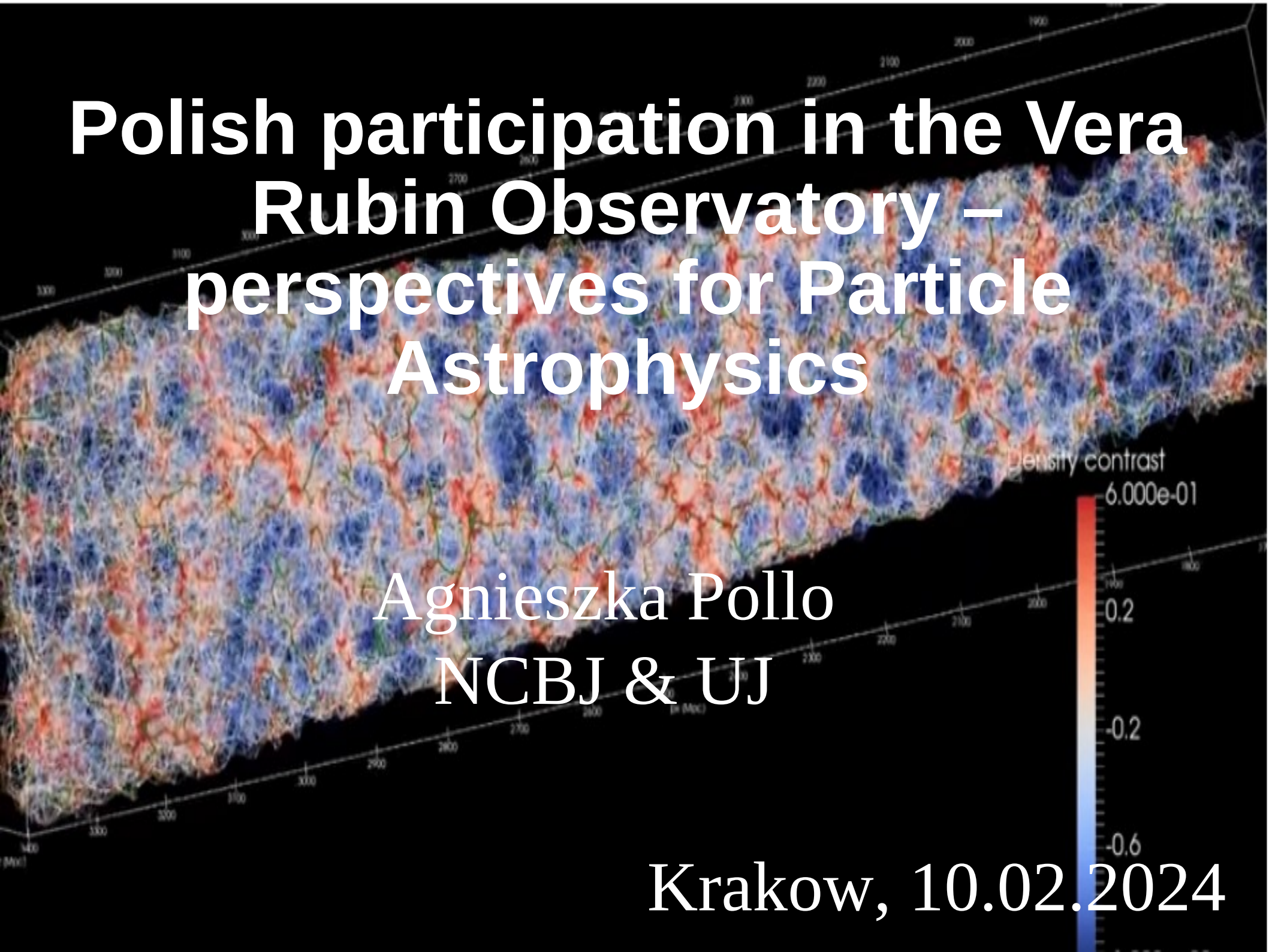
Polish participation in the Vera Rubin Observatory – perspectives for Particle Astrophysics

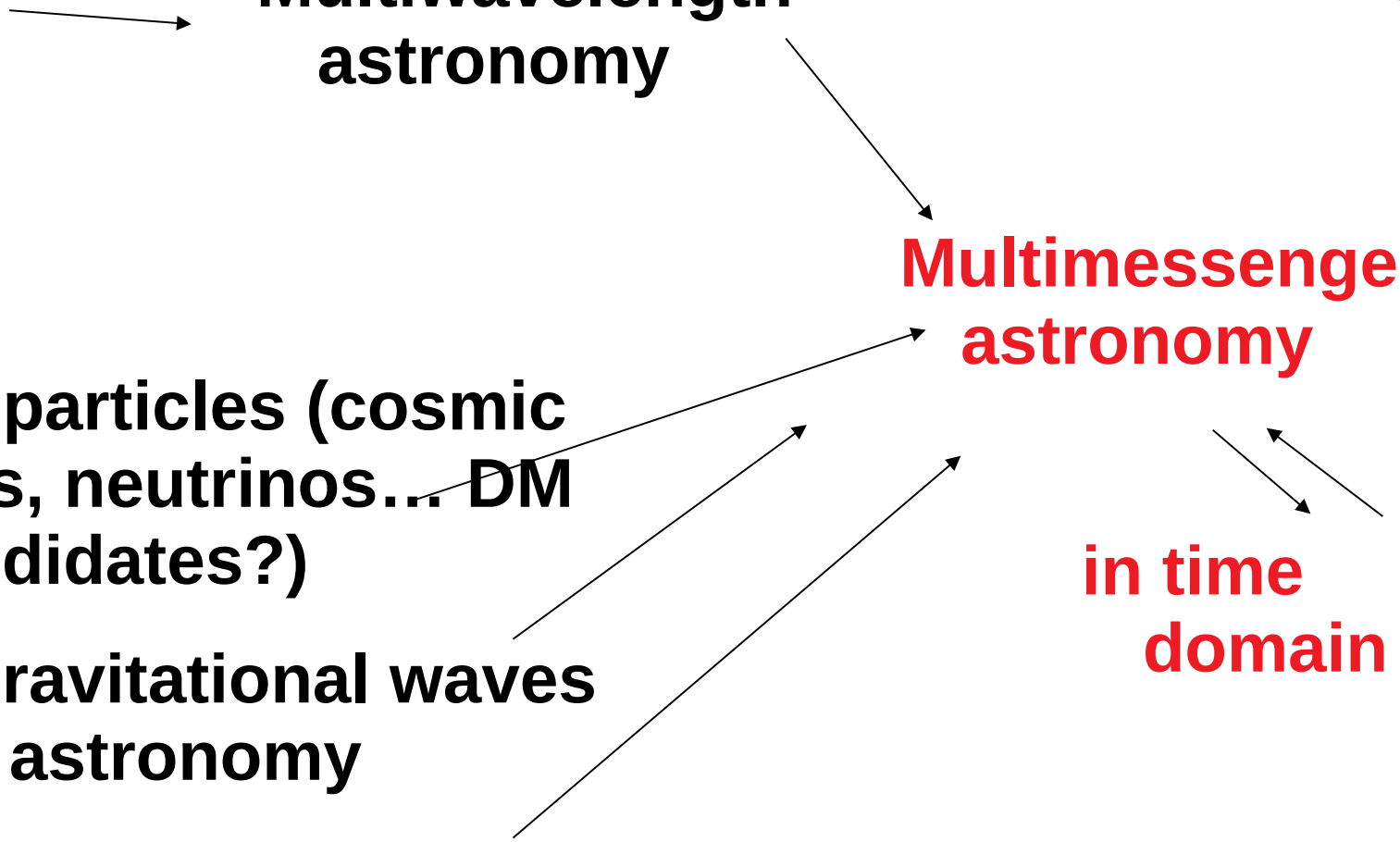


Agnieszka Pollo
NCBJ & UJ

Krakow, 10.02.2024

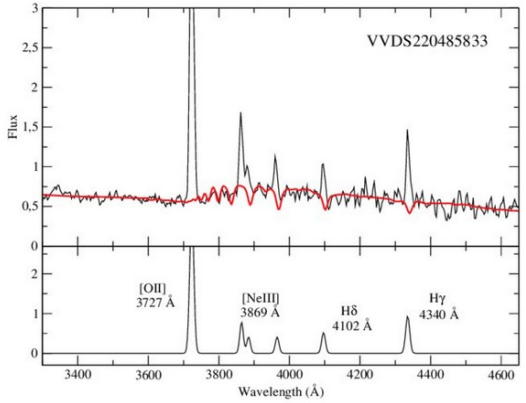
Modern astronomy

- Radio
- mm
- IR
- vis
- UV
- X
- gamma



Modern astronomy as “data science”: – all data need to be cross-correlated and treated together

- photometry + imaging (in different spectral ranges – from gamma rays to radio)



- spectra

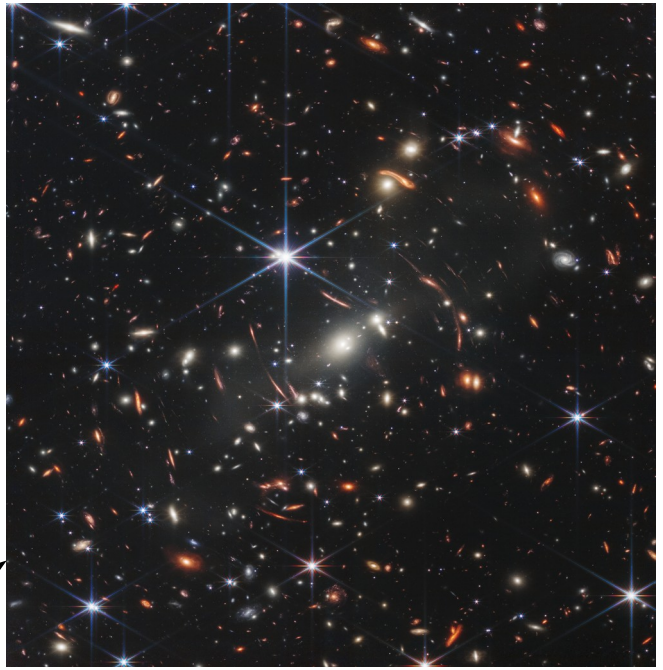
„Multimessenger time domain astronomy”

- time variability
- polarization

Modern astronomy as “data science” – knowledge transfer between very different domains

JWST (space)

CFHT (ground-based)



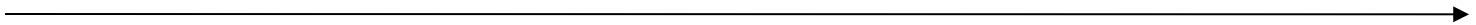
Illustris (simulation)

Modern astronomy as (astronomicaly) big data science

To compare:
human genom - 3.1 Gb

1980 1990 2000 2010 2020 2030

MB GB TB PB EB

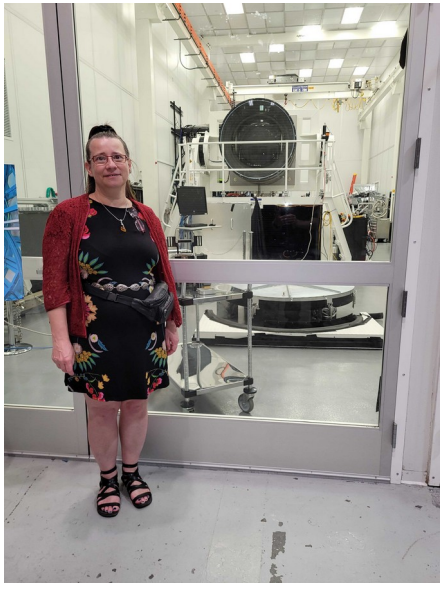


CCD → larger and deeper sky surveys → -> LSST, SKA...
tools for join analysis of different types of data
(Virtual Observatory)

All astronomical data existing today: ~500 PB
In a short few years – 10xmore will be produced yearly.

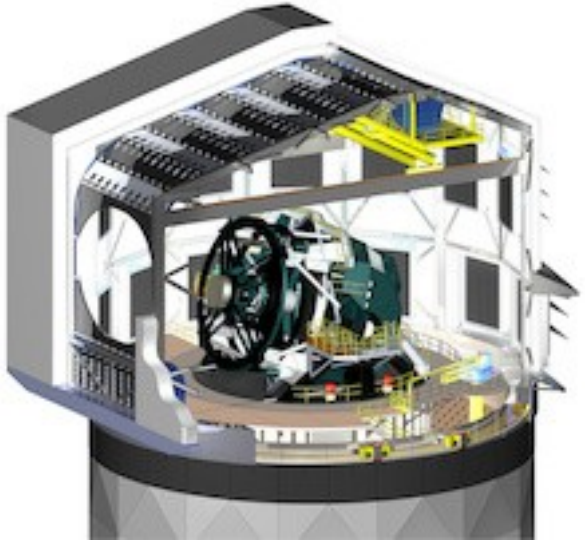
Large astronomical data of near future: Vera Rubin Observatory

First light expected in 1st half of 2025



- Large Survey of Space and Time (LSST)
- The deepest and widest sky survey ever + time domain
- mirror 8.4-m; 3200 megapixel camera
- **37 mld** stars and galaxies
- 10-years survey
- **15-30 Terabytes (~all SDSS) per night**
- After 10 years:

~200 petabytes of data

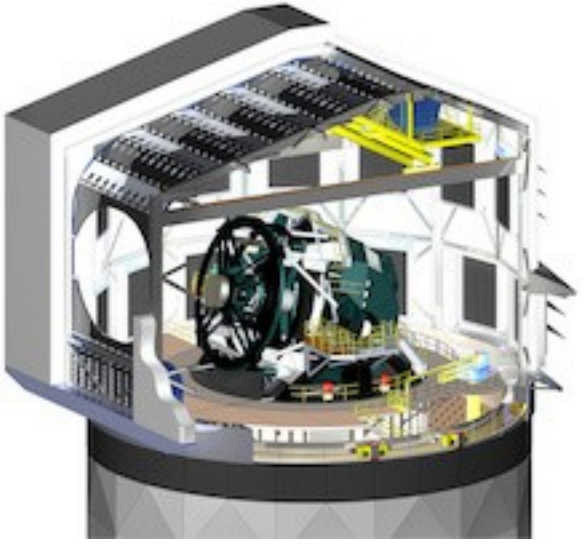


2019:
The Large Synoptic Survey Telescope (LSST)

→ Vera Rubin Observatory hosting Simonyi Survey Telescope, on which Legacy Survey of Space and Time (LSST) will be made

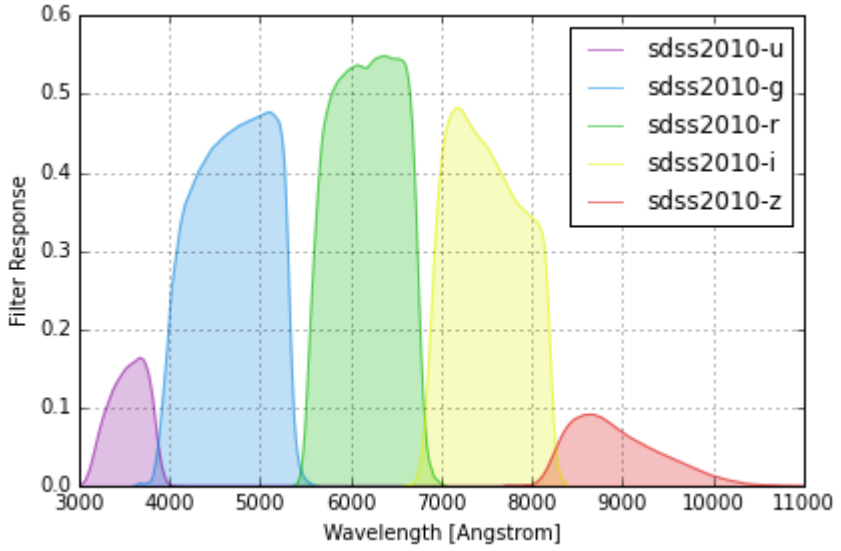
LSST
Politics: basics

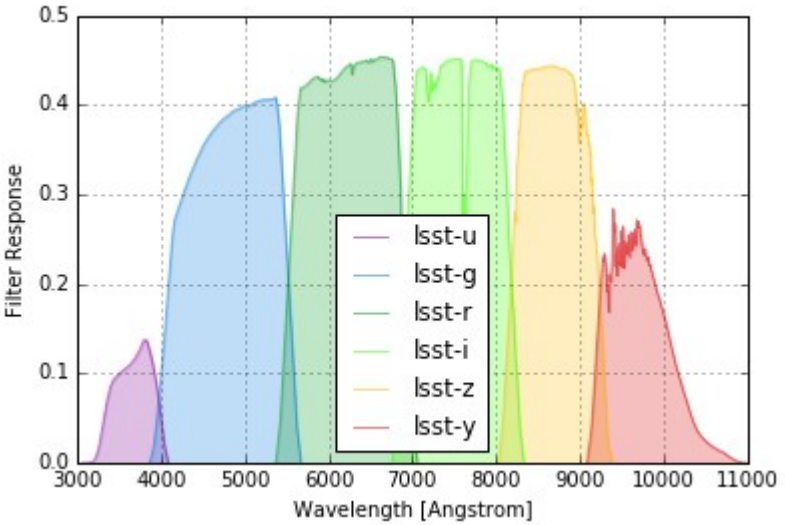
- large survey telescope in Chile
- US project funded by US funding agencies + private funds + “international contributors”
- **one of the biggest surveyers** for the next decade, with synoptic capabilities



The Vera Rubin Observatory
Technicalities

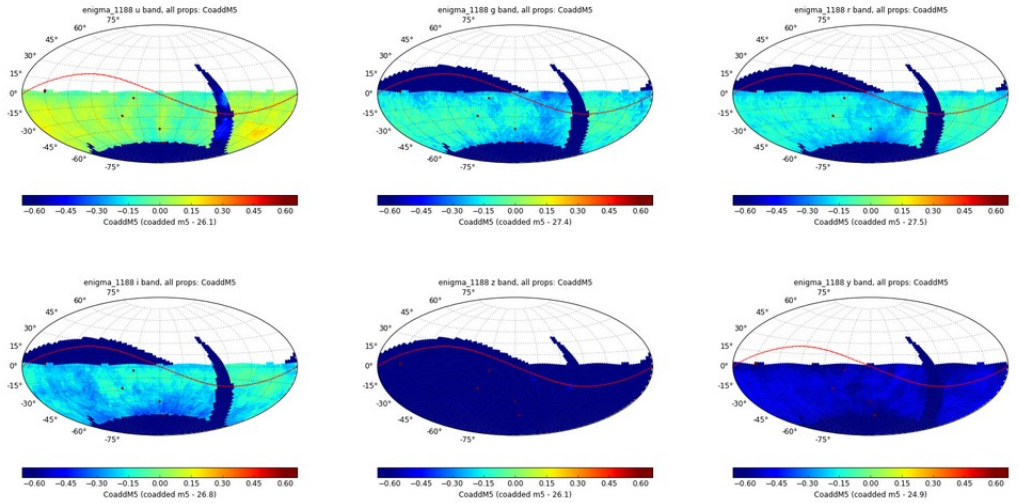
- 3200 megapixel camera with 3.5-degree field of view, 10 μm pixels capable of 0.2 arcsecond sampling
  - changing filters: ugrizy photometry





Legacy Survey of Space and Time (LSST)

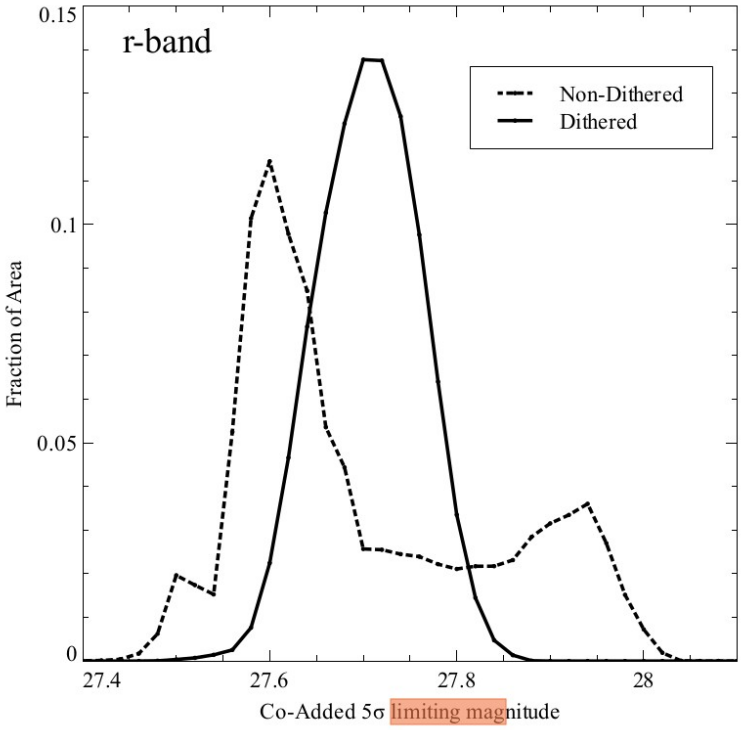
Technicalities: cadence



- >5 mln exposures in 10 years
- taken as pairs of back-to-back, 15-second exposures to aid in cosmic-ray rejection
- sky scanned ~every 3 days
- total of ~30,000 sq deg covered
- 18,000 deg2 main survey via “universal cadence” (with uniform coverage – the real cadence is more complex and still evolving)
- additional runs to facilitate concrete science cases
- ~10% of time dedicated for observation of Deep Drilling Fields

The Large Synoptic Survey Telescope (LSST)

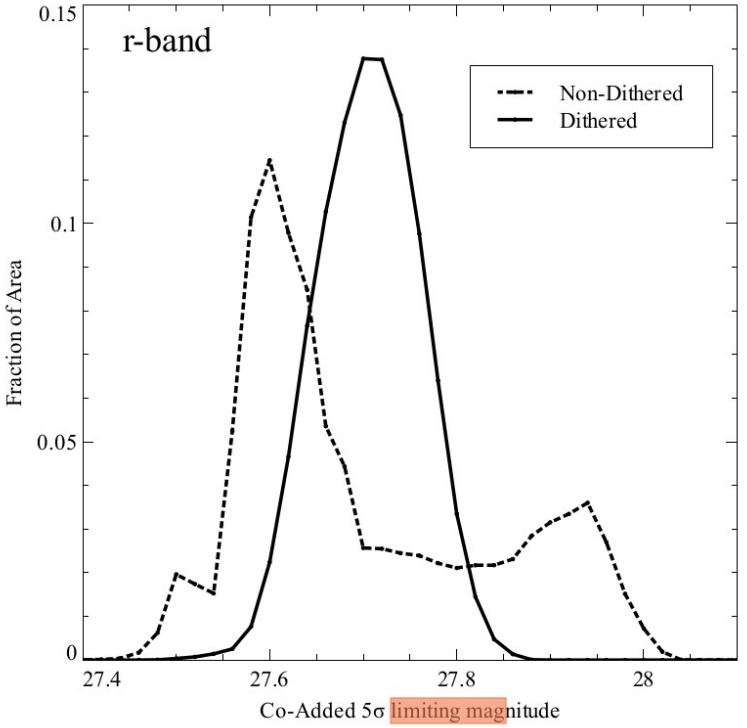
Data and data products



- **37 billion** stars and galaxies
- down to r~28 limiting mag
- 10 million alerts, 1000 pairs of exposures,
  15-30 Terabytes of data expected every night

The Large Synoptic Survey Telescope (LSST)

Data access model

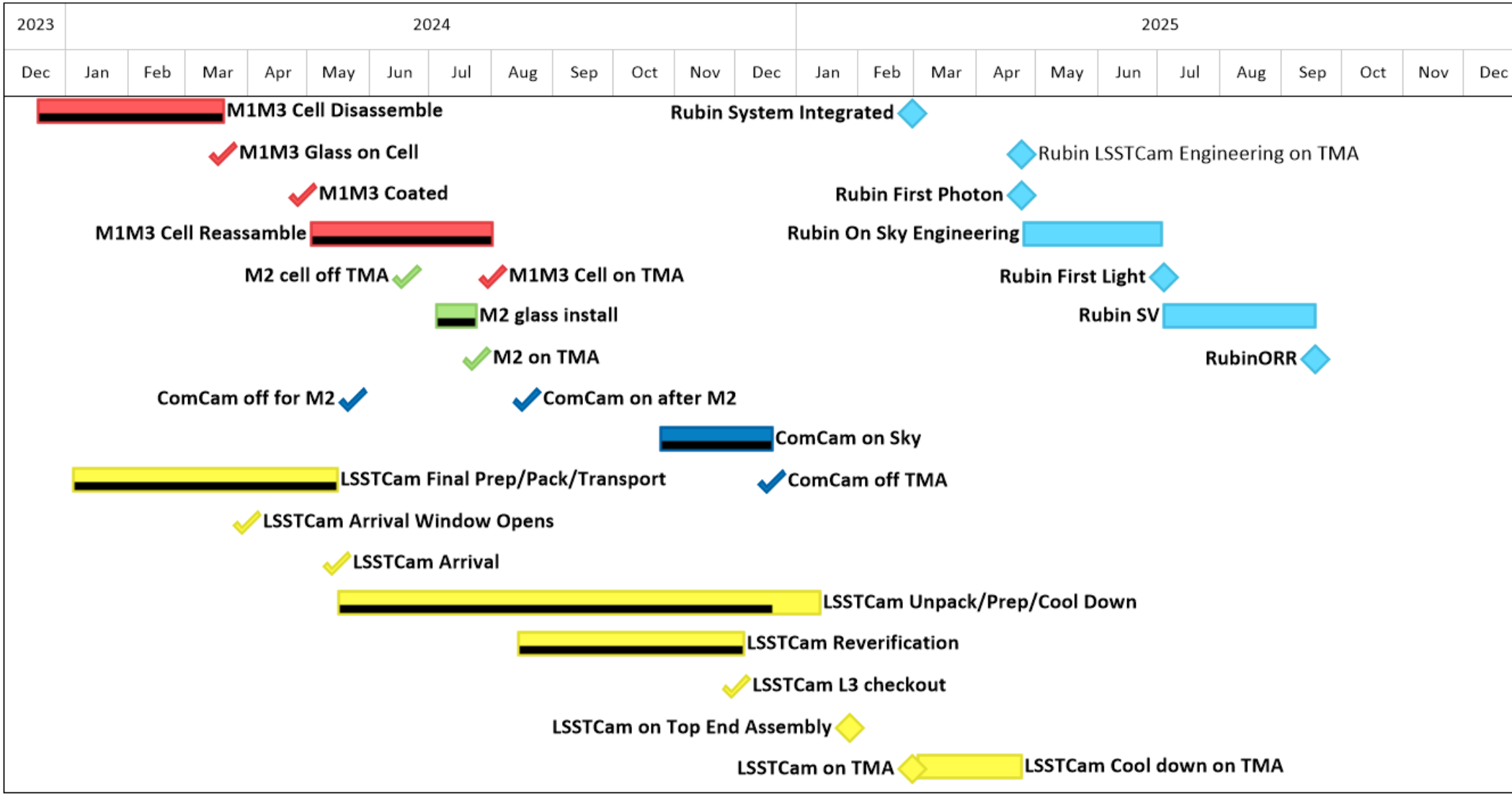


- **Prompt data products:**
  - ➔ stream of alerts publicly available released with 60-second latency → for all the world (through brokers)
  - ➔ source catalogs derived from difference images and image data released with 24-hour latency **(for data right holders)**
  - ➔ yearly full data releases **(for data right holders)**

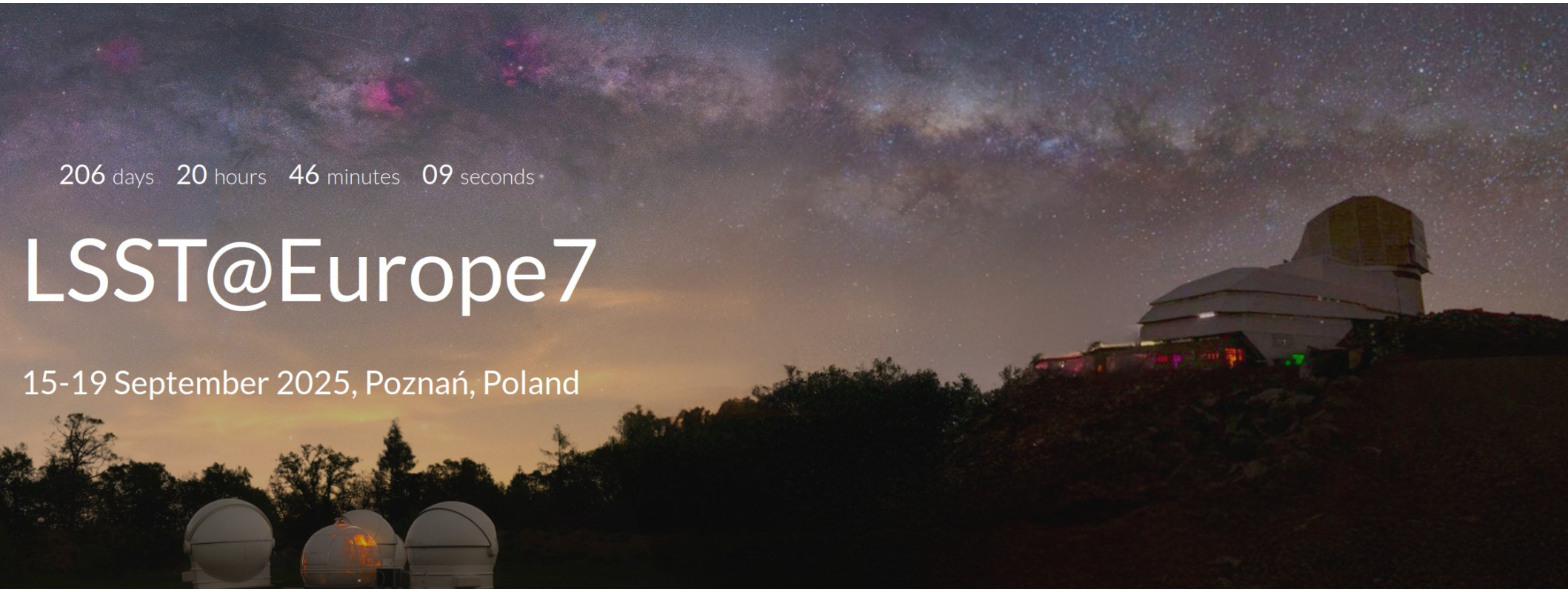
The Legacy Survey of Space and Time (LSST)

Starting this year

Rubin Key Activities December 2024

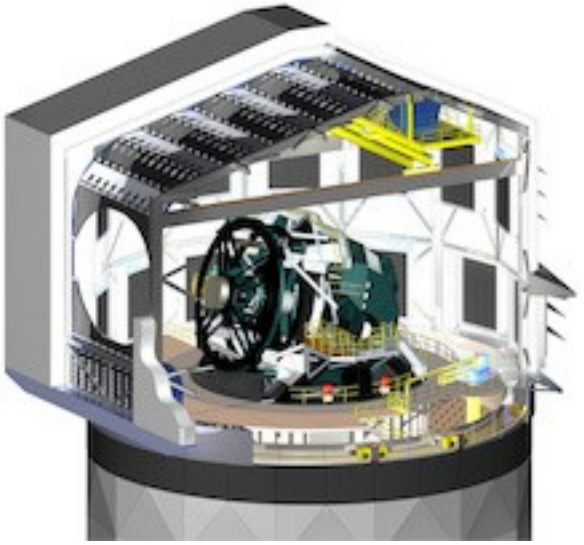


The nearest conference from the LSST@Europe series will be in Poland (Poznań)



https://lsst-europe7.syskonf.pl/

LSST
Science: basics



Main scientific topics to be covered:

- Constraining Dark Energy & Dark Matter
- Taking an Inventory of the Solar System
- Exploring the Transient Optical Sky
- Mapping the Milky Way

+discovery potential for the unknown

LSST
Politics: structure

•US project funded by US funding agencies + private funds + “international contributors”

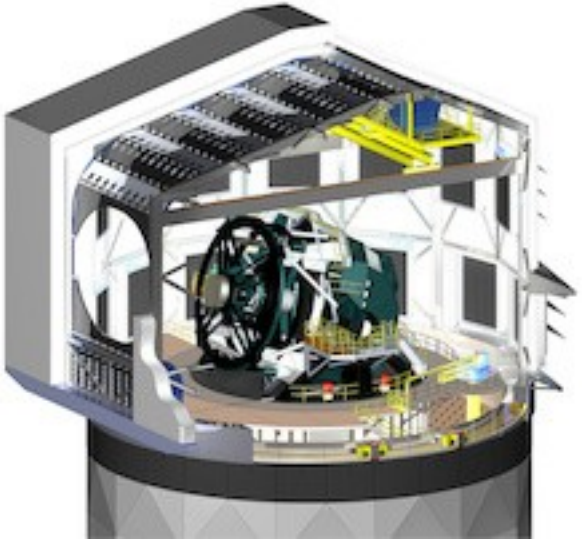
•LSST – project

•LSST consortium → Discovery Aliance

•LSST Science Collaborations (SCs):

- Galaxies
- Stars, Milky Way, and Local Volume
- Solar System
- Dark Energy
- Active Galactic Nuclei
- Transients/variable stars
- Strong Lensing
- Informatics and Statistics
- Strong Lensing

- LSST UK
- LSST Chile
- LSST France



LSST
Politics of participation rules

- In-kind contribution in exchange for a given number of “data right holders” (“a unit” = 1 “senior PI” and 4 “junior co-Is”).
- LSST-PL consortium – NCBJ (coortinating), CAMK PAN, CFT PAN, UMK, UJ, UW, UAM, UWr
- Presently: 15 senior PIs, circa 70 data right holders in total, covering practically all the range of SCs
- A list updated +/- once per year
- Final MoA still to be signed (but close)
- Come and talk if interested

LSST in-kind

- PI: Agnieszka Pollo (NCBJ & UJ) // Project Manager: dr hab. Paweł Piertrukowicz (UW)
- Six main points:
  - S1. “Lite IDAC” - Independent Data Access Center (CL: dr Krzysztof Nawrocki, NCBJ)
  - S2. “Near-Term Directable computing infrastructure effort for Dark Energy Science Collaboration providing a link between the NCBJ Lite IDAC, the NCBJ computing resources and the DESC computing infrastructure” (CL: dr Adam Zadrożny, NCBJ) (3x1.5 FTE)
  - S3. Science Pipeline Development in the LSST Galaxies Science Collaboration (3x0.5 FTE) (CL: dr hab Katarzyna Małek, NCBJ)
  - S4. Science Pipeline Development in the LSST Dark Energy Science Collaboration (3x1FTE) (CL: dr hab Maciej Bilicki, CFT PAN)
  - S5. Science Pipeline Development in the LSST Dark Energy Science Collaboration (2x0.5 FTE) (CL: dr hab Wojciech Hellwing, CFT PAN) → now merged with S2
  - S6. Science Pipeline Development in the LSST AGN Collaboration (3x0.5 FTE) (CL: prof. Bożena Czerny, CFT PAN)

S1. (Lite) Independent Data Access Center (IDAC)

- 500 CPU cores (Intel Xeon or AMD Epyc; might change based on market situation)
- 5 Petabytes of storage (HDD disk type)
- provision for the period FY23-FY35 (now: → 2025-3037), including two complete storage refreshes, such that 5 PB continuous storage and a final total of 56M CPU-hrs are provided
- fully compliant with the requirements spelled out in the Guidelines for Rubin Independent Data Access Centers, including:
  - Authentication/Authorization system inline with Rubin Observatory Access
  - IDAC broadly accessible to all Data Rights holders
  - Database system capable of handling 4^10 rows
  - Compliant with IVOA TAP interface, MyDB and Table Upload, CAOM support
  - Availability of about 500TB of disk for catalogs + MyDBs
  - Professional support staff
  - Sufficient connectivity to support users
  - Robust and effective communication with NOIRLab and the other IDACs in the Rubin IDAC network, to assure the software compatibility, and synchronization of the LSST catalog data releases.

Light IDAC - Installation Site

- Poznań Supercomputing and Networking Center (PSNC) in Poznań (Poland)
- Light IDAC planned as a part of a larger system KMD3 / PraceLab2 - in total:
  - 25PB storage
  - ~ 6k CPU physical cores system + some GPU
- KMD3 (National Data Repository)/PraceLab2 are run by consortium of Polish supercomputing centres (including NCBJ)
- Hardware already there, installattions ongoing (moving from the testbed)

Courtesy: K. Nawrocki

Specification of Light-IDAC (POL-NCB S1)

Specification of light IDAC-PL

- 500 CPU cores
  - 20-30 cores for system/db
  - 470-480 cores for users
- 2 x 0.25 FTE support stuff
- proposed center could serve 500-1000 users in total
  and 250 users using center concurrently

As specified in guidelines: https://rtn-003.lsst.io/

Courtesy: K. Nawrocki

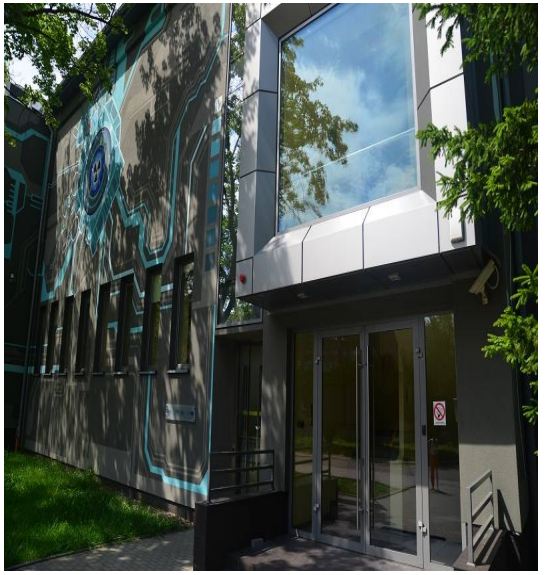
Specification of Light-IDAC (POL-NCB S1) +

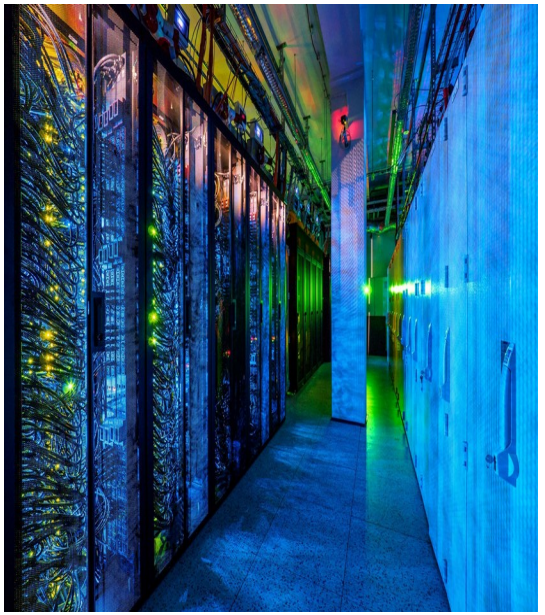
Specification of Light IDAC-PL +

- present plan: storage of (lite) catalog data
- new present-day needs: images (at least co-adds)

Plan

- additional storage
- access to GPUs for ML
- access to more CPUs
- → funding for project for extra storage and GPU confirmed (for now out of scope of the Light IDAC - an addition)





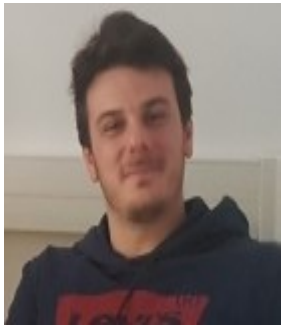
Courtesy: K. Nawrocki

IDAC - advantages

- Easier data access to the LSST data for local users
- Easier distribution and promotion of “secondary data products” created in Poland
- Possible point of crosscorrelation with data other projects → multiwavelength astronomy center
- New type of database philosophy in astronomy – still in development
- Possibility for fine-tuning our IDAC to our ”science cases”

Finally, some science

Will photometric LSST data be enough?



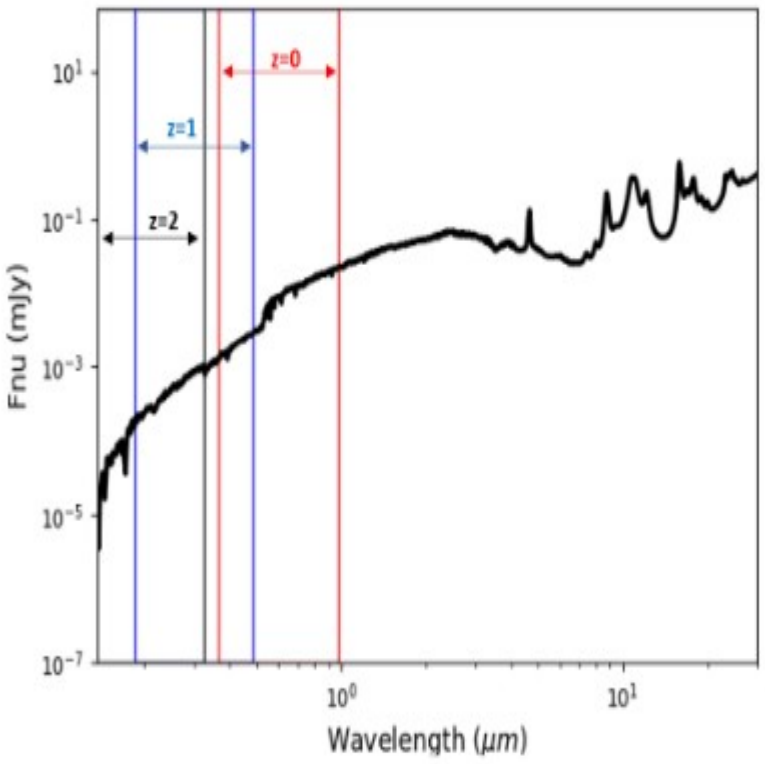


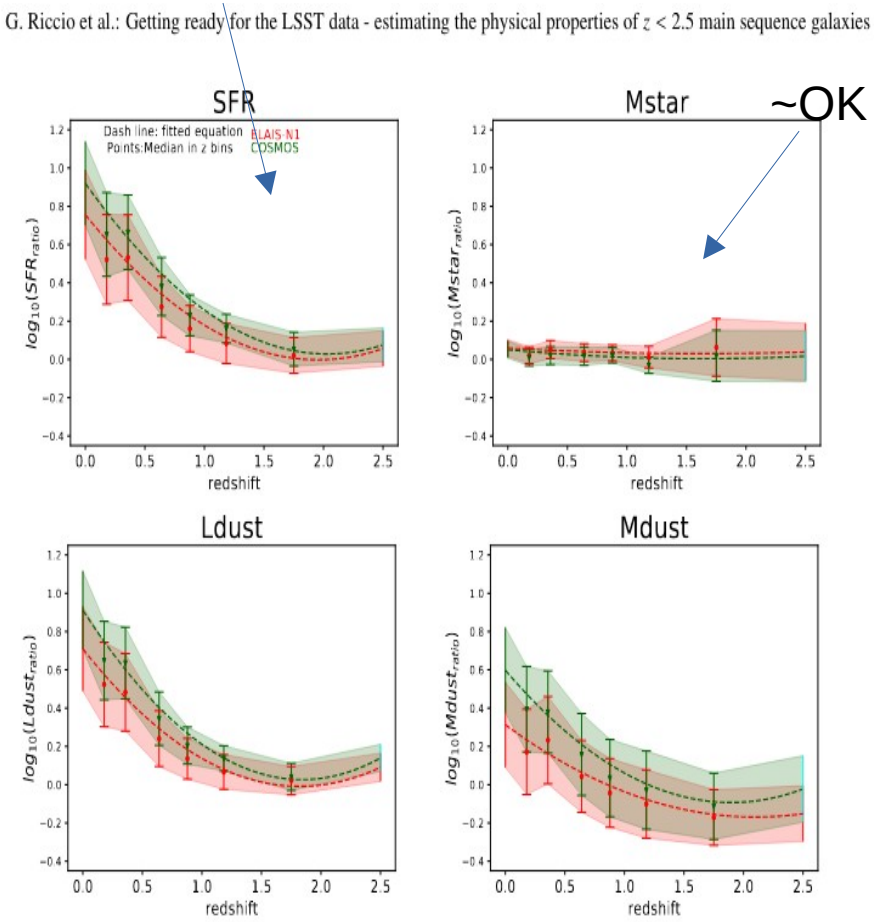
Fig. 8. LSST coverage of an example SED at different redshifts, indicated in the panel.

- … e.g. to measure physical properties of star forming galaxies?
- Data: Hershel Extragalactic Legacy Project (HELP):
  - real SF galaxies from ELAIS1 and COSMOS field with IR Herschel photometry
  - simulated LSST-only output for the same galaxies
- SED fitting (CIGALE) for
  - real data and
  - “LSST-like” simulation

Riccio, Malek et al. 2021

Will photometric LSST data be enough?

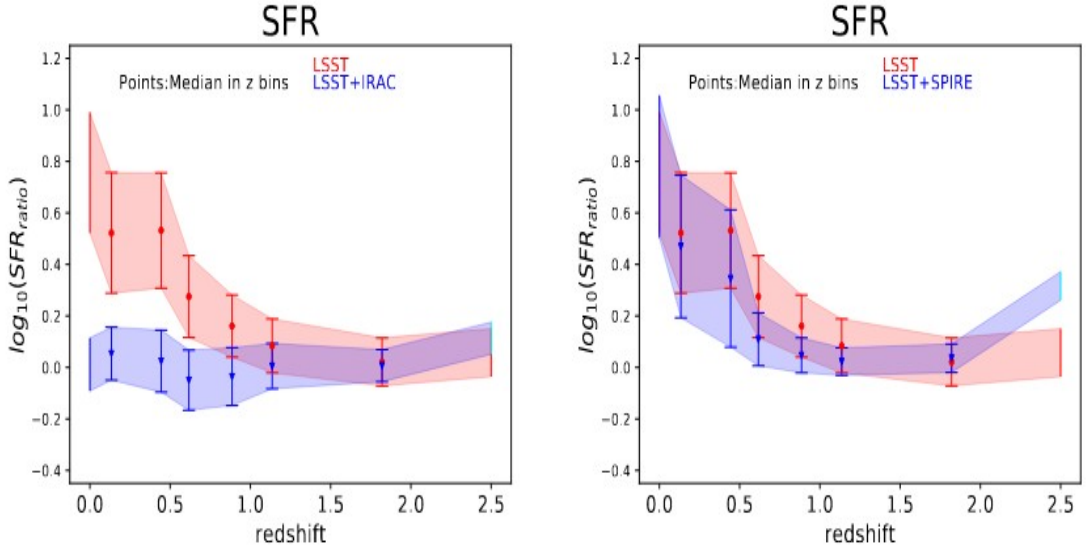
Very much not enough… and in a redshift-dependent way



Reason: **dust attenuation**

Cures:

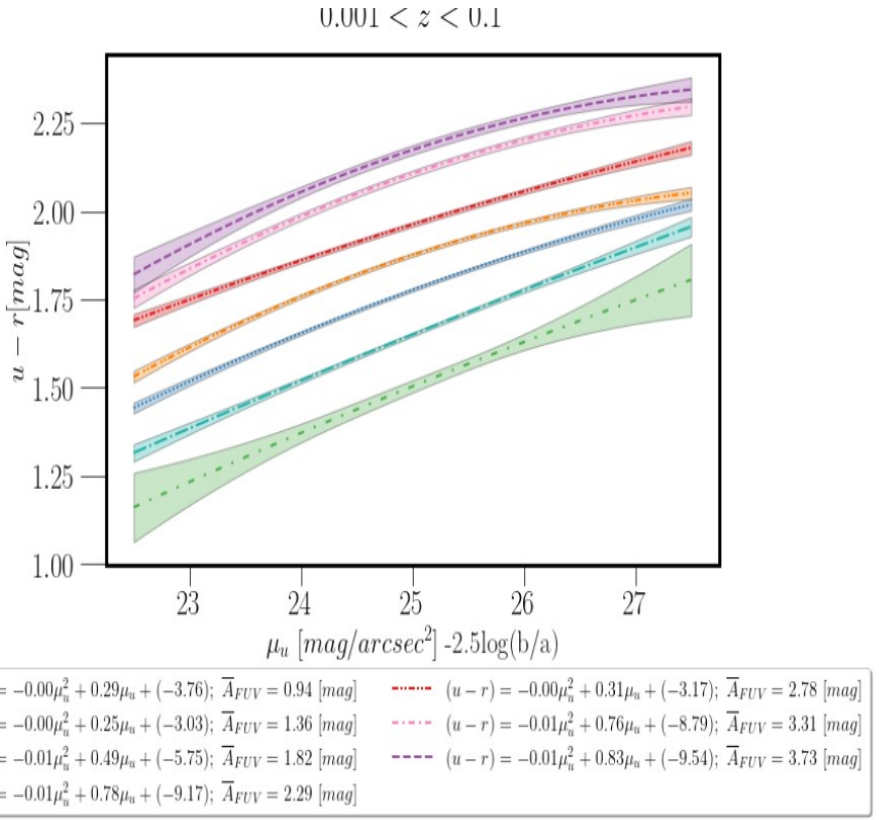
- IR data
- callibration (being careful about dfferent galaxy types)
- **what else…?**



Riccio, Malek et al. 2021

Recovery of properties of SF galaxies without IR

- Galaxy properties with CIGALE: a part of our (Polish) in-kind programme
- Is it possible to account for attenuation within the LSST-only parameter space?
- → Surface brightness in *u*
- Preliminary results: we can introduce a correction that works for SDSS star forming galaxies



Malek, Junais et al. 2024

Towards low surface brightness Universe

NARODOWE CENTRUM NAUKI

I. North Ecliptic Pole: deep multiwavelength data

- deep data from HSC, AKARI, HERSHEL, SCUBA2, CFHT, Spitzer… soon Euclid
- unique for multiwavelength studies of faint objects (including Low Surface Brightness Galaxies - LSBGs)

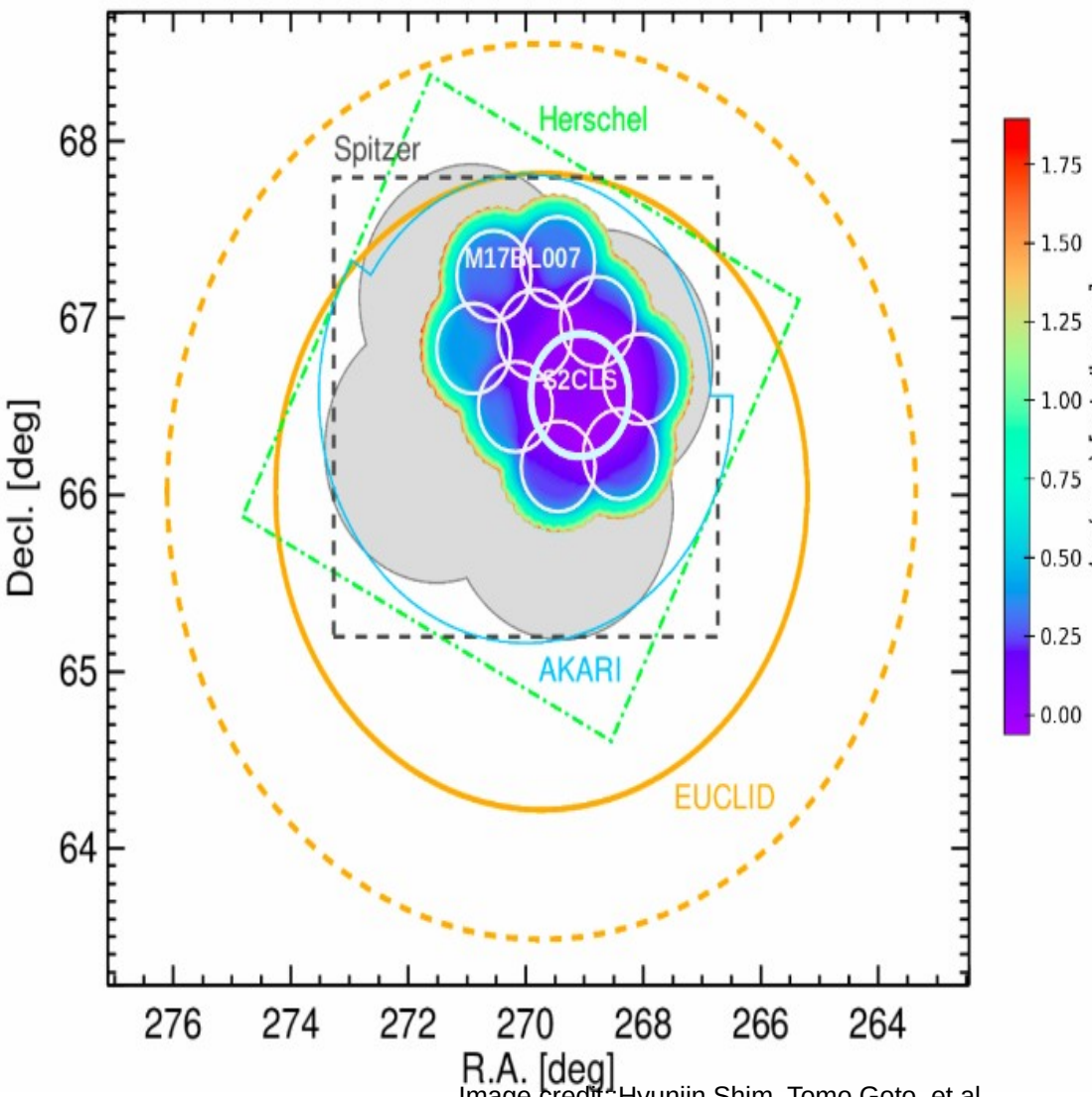
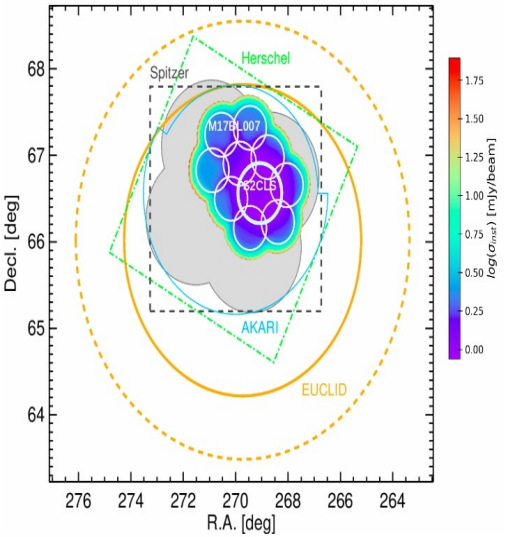
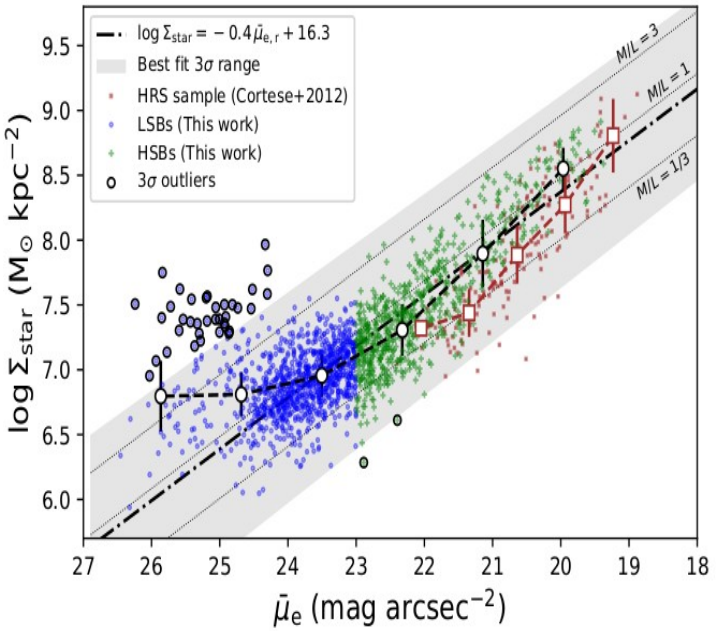


Image credit: Hyunjin Shim, Tomo Goto et al.

North Ecliptic Pole: dusty LSBGs?



- 1003 LSBGs + 628 faint HSBs
  - SED fitting and analysis of properties
- it is generally assumed that LSBGs are dust-free (but IR observational data are scarce)
- NEP: 36 dusty LSBGs (2.5% of the sample) with non-negligible attenuation



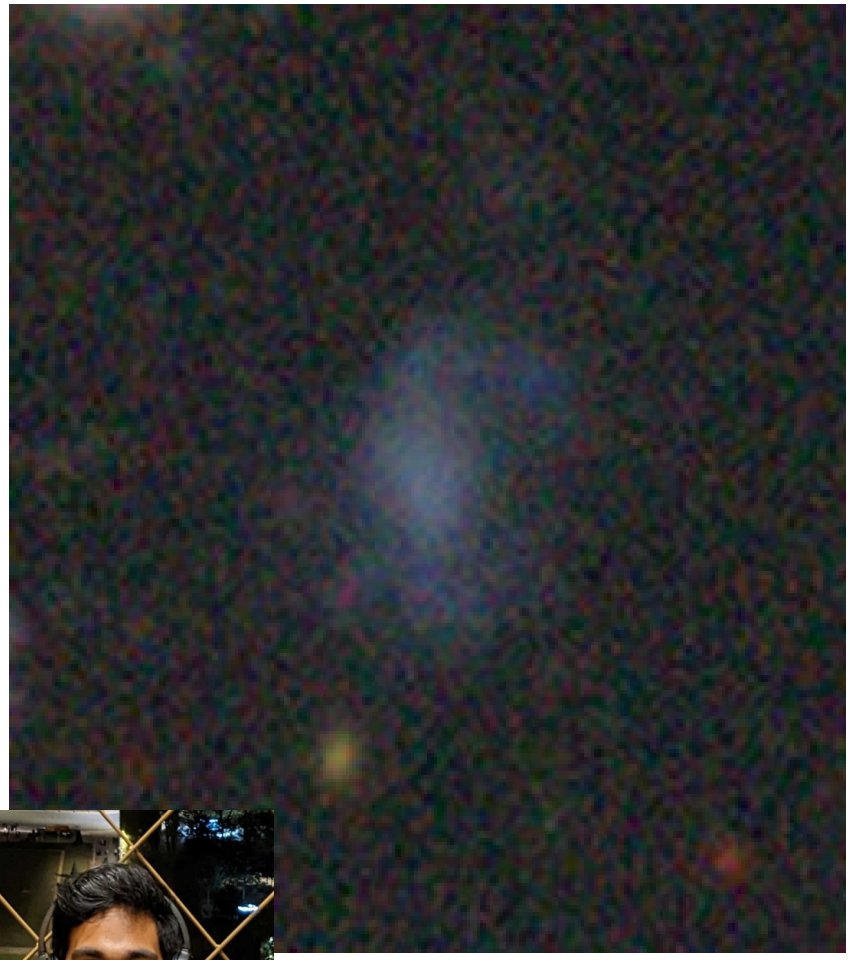
Junais et al. 2023

DES Y3 Gold: new catalog of LSBGs

- goal: addition/improvement over a sample selected by Tanoglidis et al. (2021)
- method: transformers // self-attention-based encoder models
- 27,000 LSBGs, among them 4083 new (mostly blue + extreme red, as compared to Tanoglidis et al. 2021)
- among them, 317 UDG candidates, including 276 new ones
- Continuation: see

  Michal's Vrabel

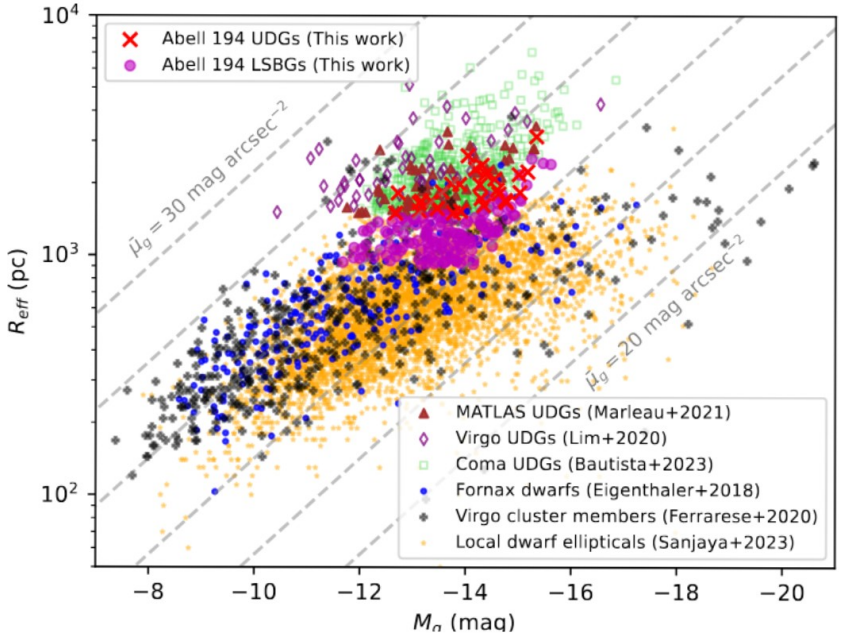
  talk later today!

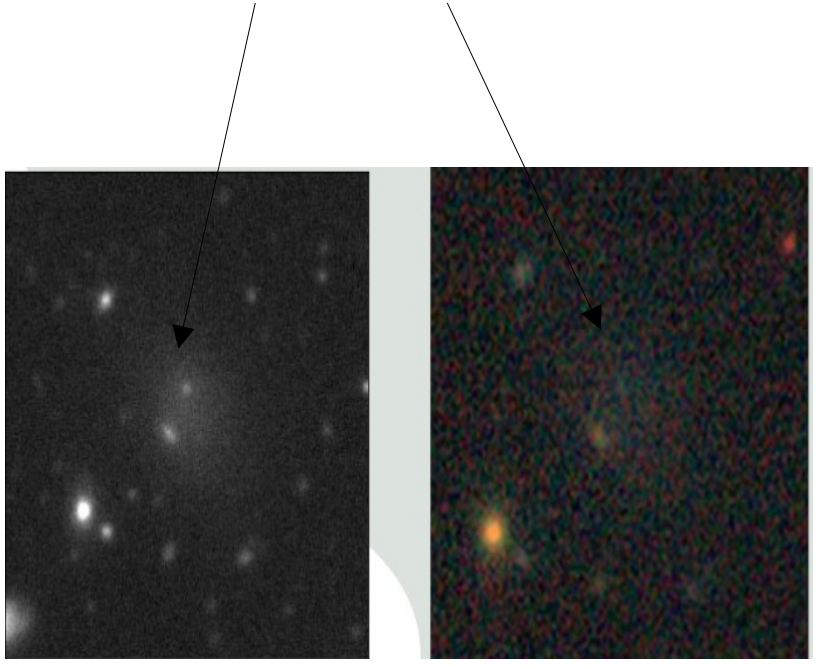


Thurutupilly et al. 2024

II. DES → HSC (the closest present counterpart of LSST)

New census of LSBGs in ABEL 194 → successful transfer learning from DES to HSC.





Thurutupilly et al. 2025 (just accepted: 2025arXiv250203142T)

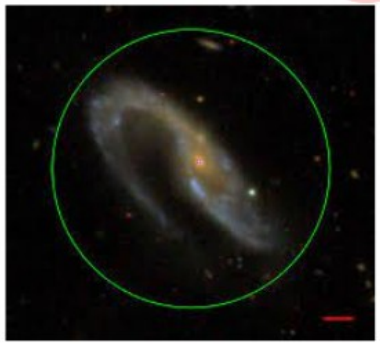
Summary

- Vera Rubin Observatory/ LSST coming soon (first light this year!)
  - Polish consortium active and growing
  - In-kind program including software contribution and data access center
  - Will be a real big data challenge
  - Will need auxiliary data and follow up for meaningful analysis
- If you want to engage, contact us now – you need time to get familiar and prepare for these data
- IDAC in Poland will (hopefully) make our tasks easier
- Exciting science (e.g. low surface brightness Universe) expected and already ongoing

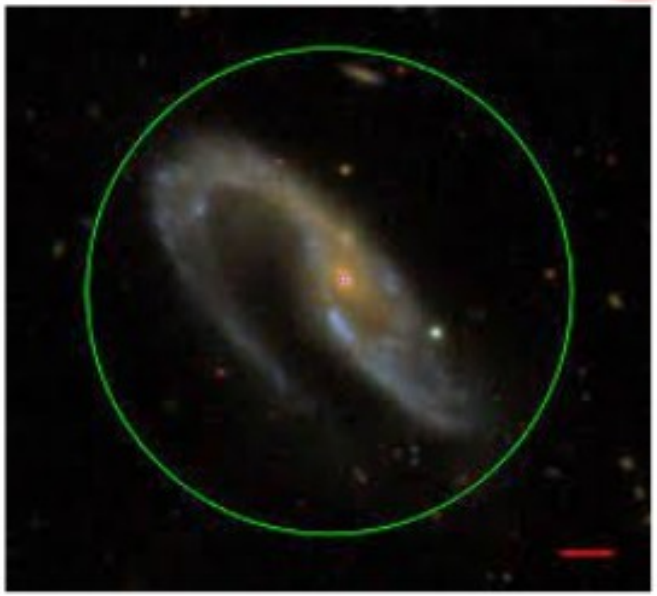
Galaxy mergers, or low surface brightness features of normal galaxies

How to automatically find merging galaxies?

→ Usually: image-based algorithms (Deep Learning, CNN..)
→Concept: see if we can do any good (but faster/easier/more interpretable) with photometry only (fluxes, colours, errors)
→ SDSS (DR6, because of Galaxy Zoo data available)
→ result: the most important piece of photometric data are… (fiber) errors on magnitudes

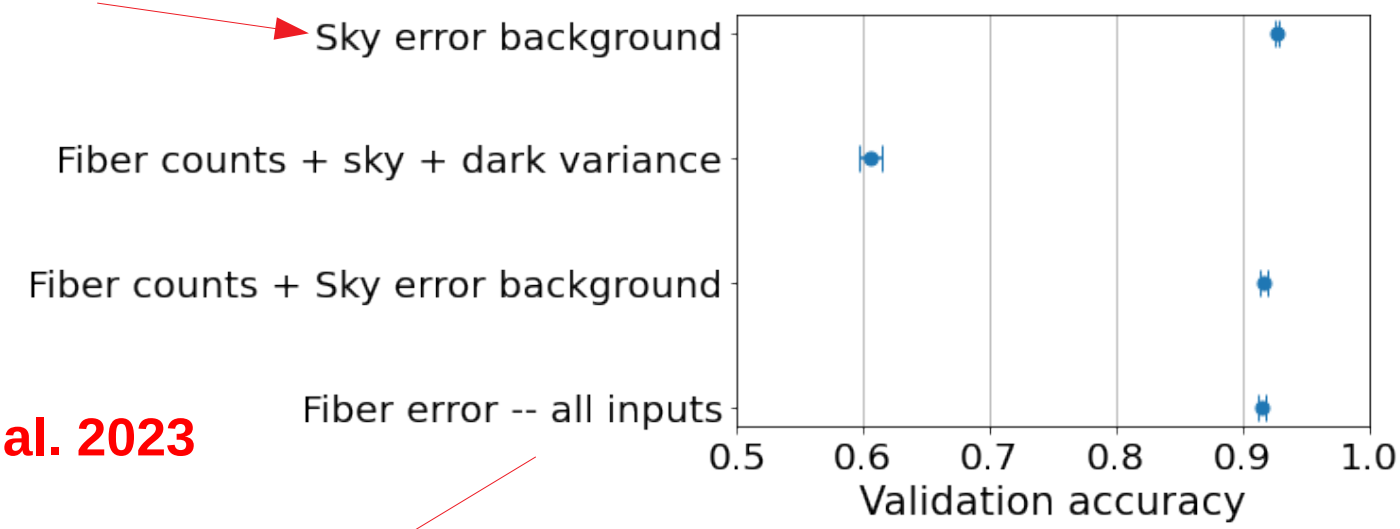


Suelves et al. 2023

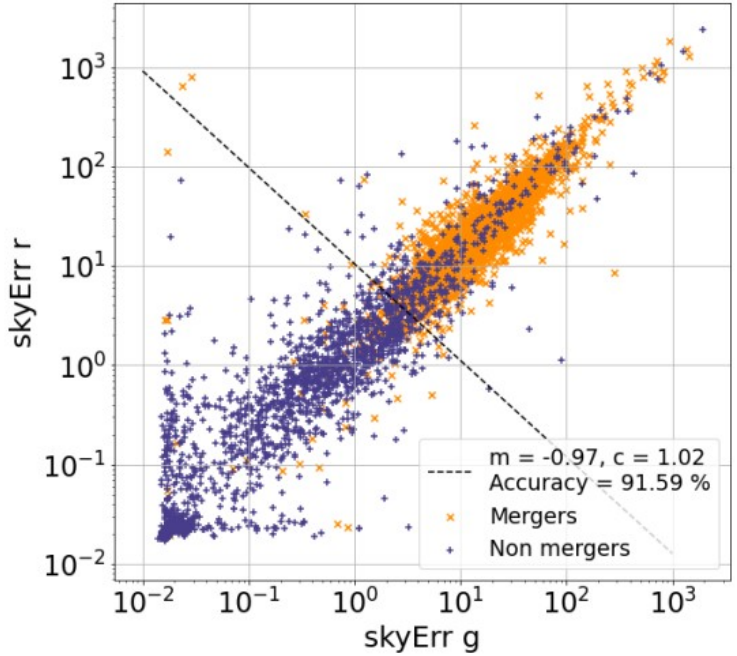


How to automatically find merging galaxies?

→ What is a magical ingredient of fiber errors?



Suelves et al. 2023



→ We can get ~92% accuracy with sky background errors only

→ Physical implications: merging galaxies at low z may not differ that much from other galaxies – we may need to look at their (LSB) surroundings to look for tidal features below sky background level

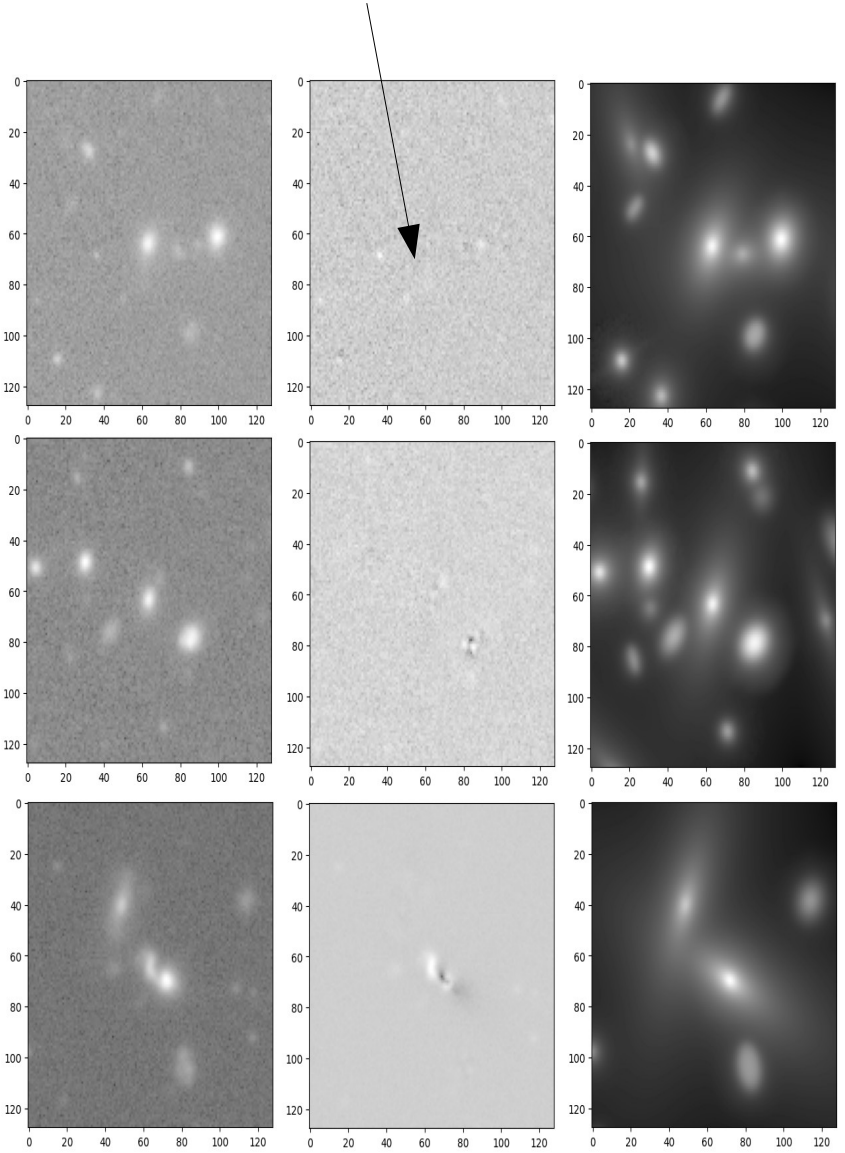
How to automatically find merging galaxies - continuation

→ Search for galaxies (galaxy mergers) without galaxies (with CNN)

→ first in simulated data (IllustrisTNG) – promising results (at least: without a galaxy in an image you can do as well as when keeping a galaxy)

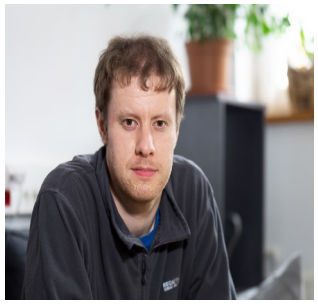
→ next, real data (at first HSC data from the NEP)

→ Find a more Universal presciption to use background noise data (NEP – we just got new Galaxy Zoo data).



Margalef-Bentabol, B. et al. 2024; Chudy et al. in prep.; Suelves et al. in prep.

Galaxy and Mass Assembly Survey: where do mergers happen?



Pearson et al. 2019

→ Galaxy merger catalogs in the GAMA survey (selected → by ML and → according to the Gini parameter)

→ Method: correlation function and marked correlation function

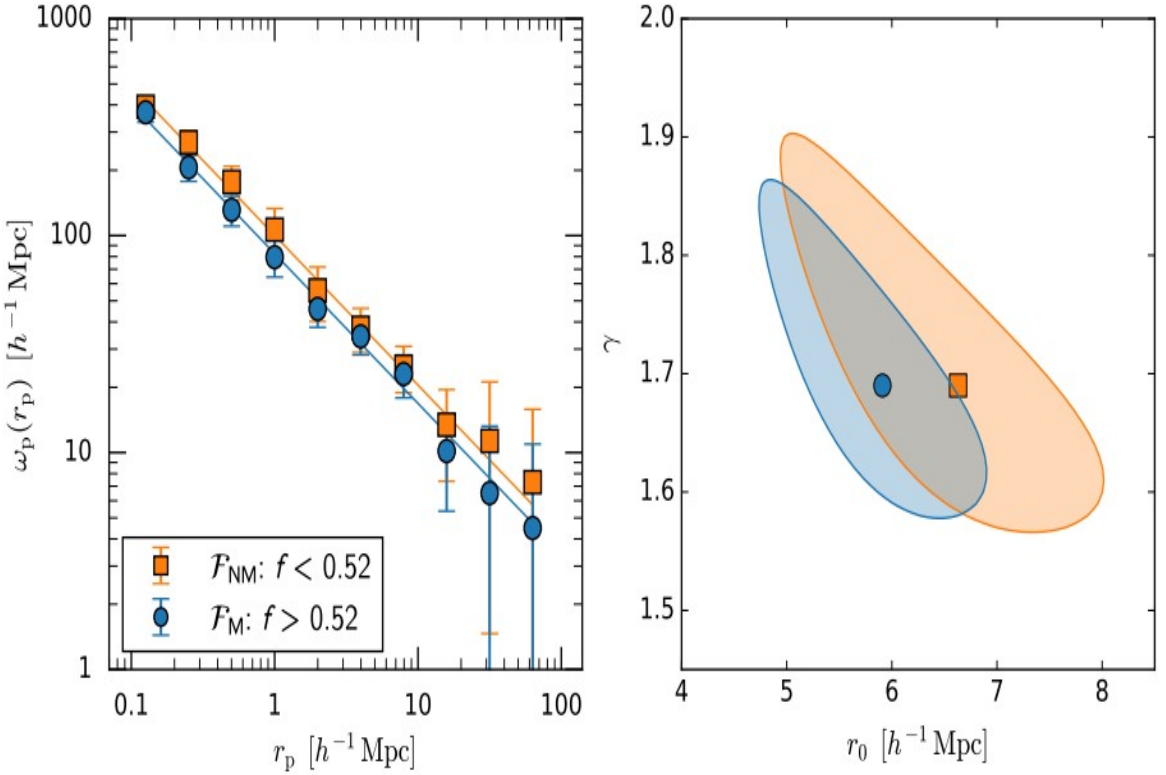
→ concept: probability of a galaxy to be a merger (according to CNN) can be regarded as a measure of galaxy “mergeriness” and then used as a weight (“mark”)

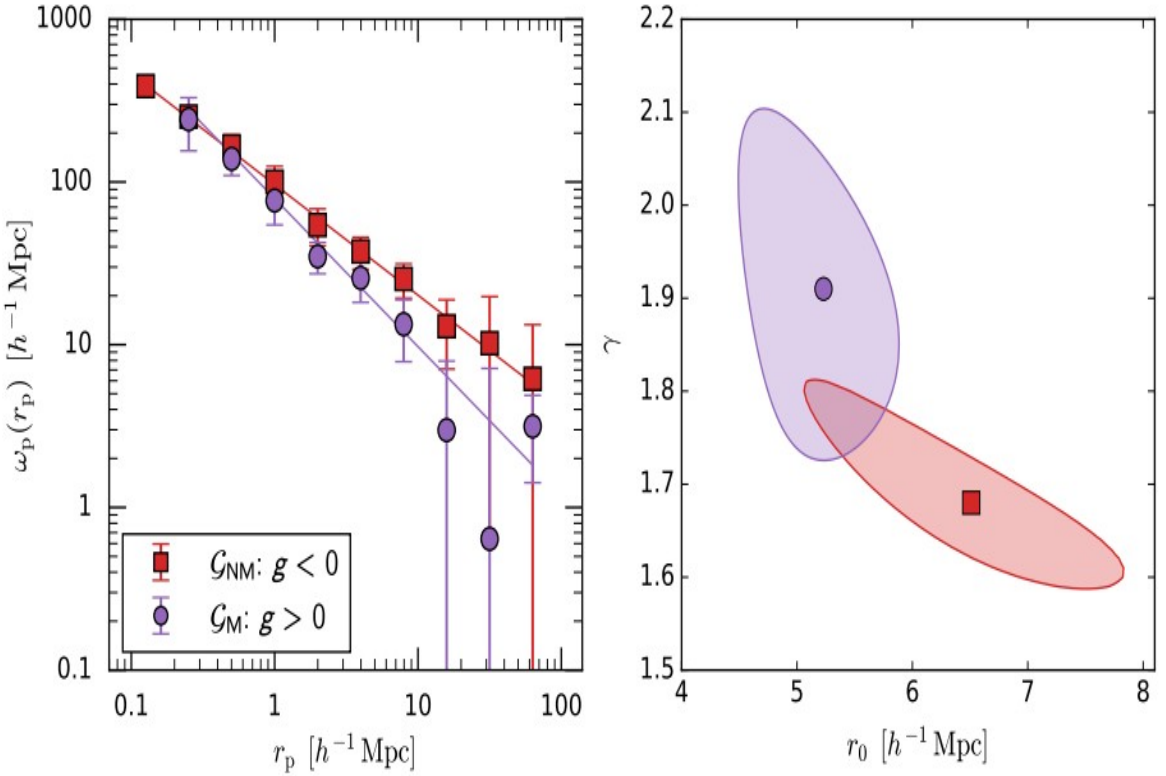
→ 0.1<z<0.16, volume limited sample(s)

Sureshkumar et al. 2024

Galaxy and Mass Assembly Survey: mergers in the large scale structure, or where do mergers happen?

Sureshkumar et al. 2024





Galaxy and Mass Assembly Survey: where do mergers happen?

→ Merging galaxies in the present day Universe prefer underdense environments (GAMA: Sureshkumar et al. 2024, NEP: Pearson et al. 2024)

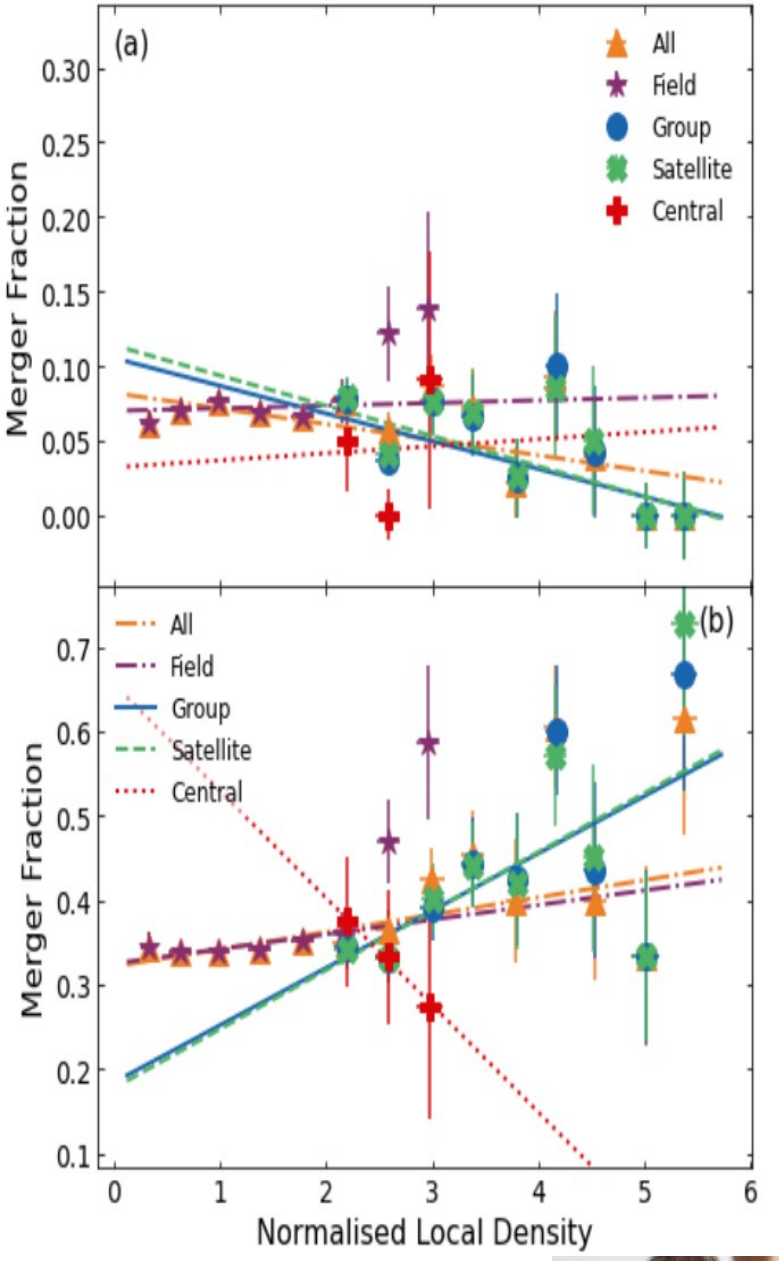
→ No significant rise in SFR w/r to similarly massive galaxies (Pearson et al. 2019, Pearson et al. 2024)

→ Most important is the invisible (i.e. low surface brightness features around).

Sureshkumar et al. 2024

North Ecliptic Pole: mergers also like underdense regions

- An ML-selected catalog with merger candidates with probabilities
- Density field based on photo-zs
- Mergers more abundant in lower density environments



Pearson et al. 2022

Pearson et al. 2024