ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Laboratoire d'Astrophysique
Ecole Polytechnique Fédérale de Lausanne
Switzerland

http://lastro.epfl.ch

EPFL - LABORATOIRE
D'ASTROPHYSIQUE

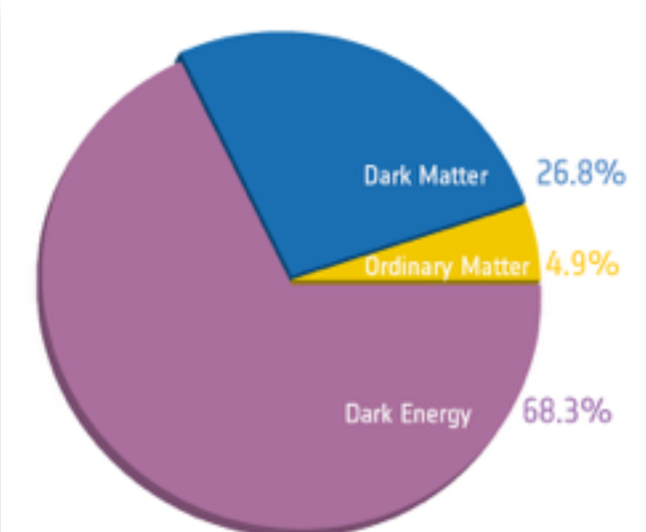
Measuring Galaxy Shapes for Cosmology

Impact of Big Data & Machine Learning

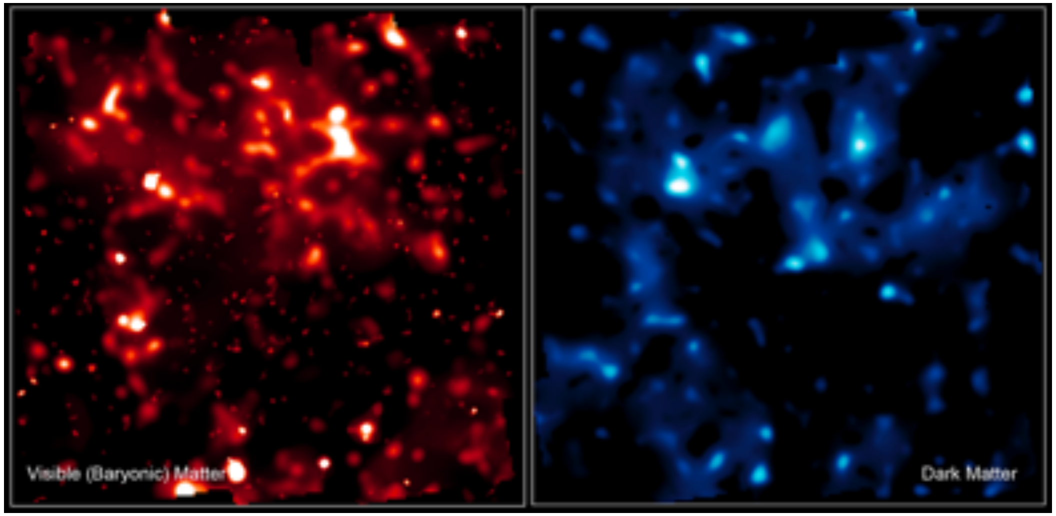
Marc Gentile, Sept. 29, 2016

Standard cosmology: successes and challenges

- The standard model is very successful in reproducing the dynamics and observed structures of the Universe
- But it is confronted to the evidence for a “dark” Universe



Planck Collaboration



Credit: NASA, ESA and R. Massey

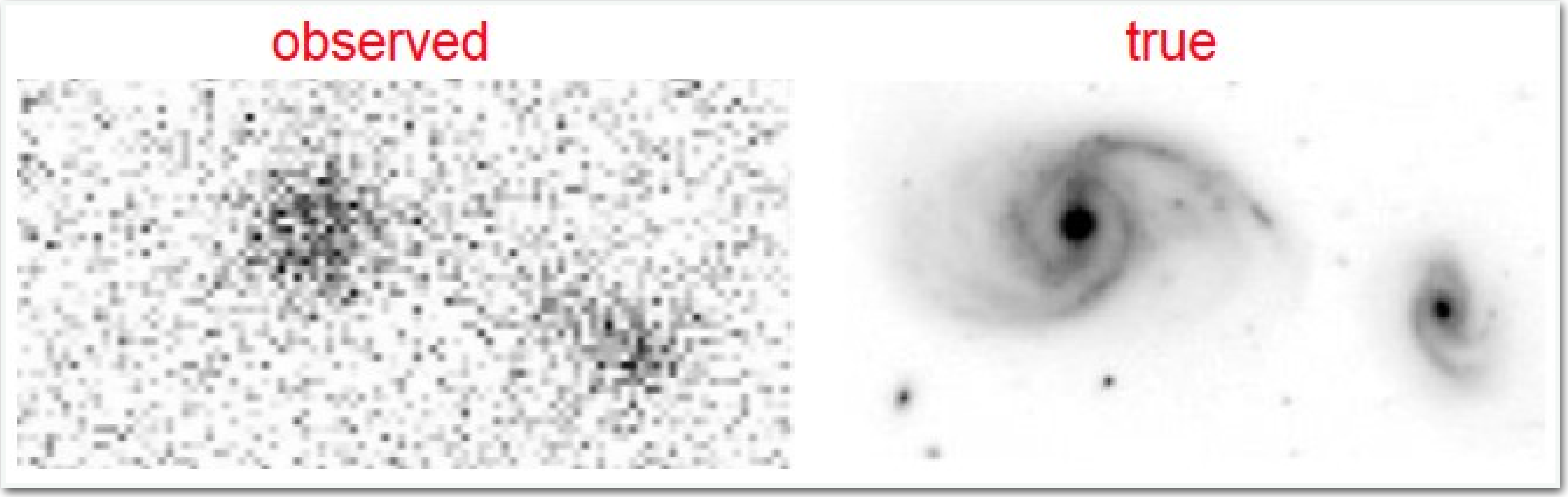
Weak Gravitational Lensing and the Cosmic Shear

- Light from distant sources is *slightly* deflected by the gravitational field of intervening matter along the line of sight
- Two main applications
  - Study dark matter on various scales
  - Constrain cosmology: using “Cosmic Shear”
- “Cosmic Shear”: weak lensing of background galaxies by foreground large-scale structures

Original galaxy **Magnification** **Shear**

...measure the slight modification of observed galaxy shapes for ~2 billions of galaxies

Measuring the Cosmic Shear: difficult!

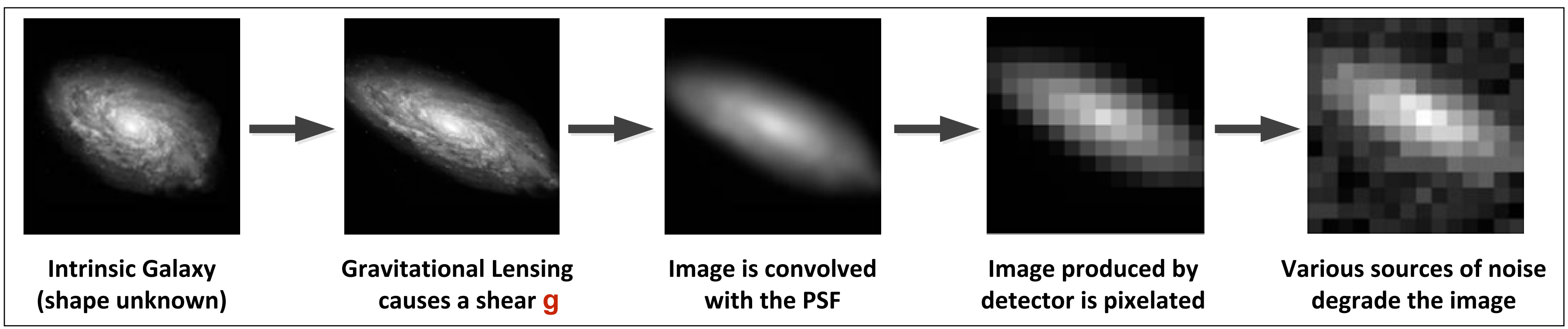


From H. Hoekstra

Observed galaxies: small, blurred, noisy, undersampled!

Measuring the Cosmic Shear: challenges

- Accurate shear measurement is difficult
  - Billions of galaxies / large sky areas to reduce statistical errors
  - Very weak signal, easily corrupted by many systematic effect => requires extremely tight control on systematic errors
  - Observed galaxies are faint, under-sampled, low S/N ratio
  - An inverse problem: no unique solution



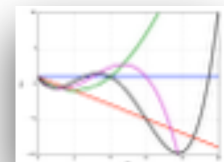
Bridle et al., 2008

Shear measurement methods: main approaches

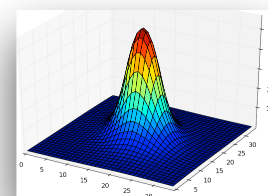
- Estimation from second-brightness moments
  - KSB/KSB+

$$e_1 = \frac{Q_{xx} - Q_{yy}}{Q_{xx} + Q_{yy}}, \quad e_2 = \frac{2Q_{xy}}{Q_{xx} + Q_{yy}}$$

- Analytical decomposition of galaxy images/shapes
  - Shapelets, Reglens



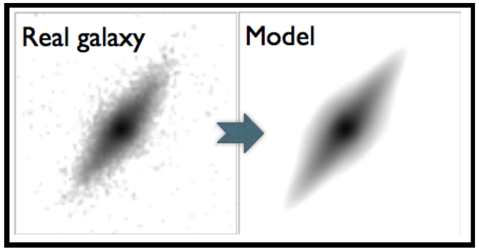
- Fit parametric model to galaxy profile
  - im3shape, lensfit, StackFit, **gFit**



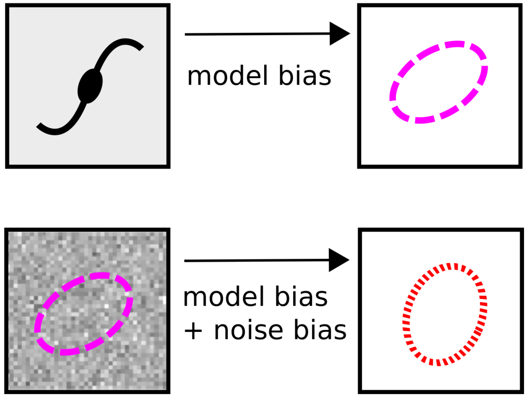
- Lensing community: STEP, GREAT initiatives to help improve algorithms through blind simulations

Typical shear measurement issues

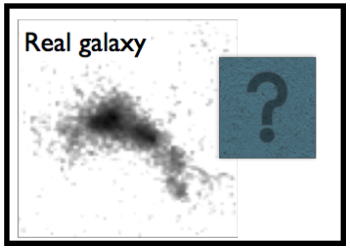
- Shear measurement biases



Mandelbaum et al.,



Kacprzak et al, 2013



- Applying the method on real data
  - Images are never perfectly clean, sensitivity to artefacts
  - Account for color, eliminate galaxy blends, etc.
- Processing time per galaxy
  - About 1sec / galaxy, too large even on a parallel cluster

Next-generation Weak Lensing Surveys

- In a few years time massive WL surveys will begin
- From Space:
  - ESA **Euclid**, 1.2m telescope (~2020)
  - NASA **WFIRST**, 2.4m telescope (~2020)
- From the ground:
  - **LSST** 8.4m Telescope (~2023)
  - **SKA** (Square Kilometre Array) Radio Telescope (~2020)
- Goal: measure cosmological parameters within 1% error
- Will shear measurement methods be ready?

Bias Requirements for Euclid-like surveys

- Additive and Multiplicative Biases

$$\gamma_i - \gamma_i^{true} = m_i \gamma_i^{true} + c_i \qquad (i = 1, 2)$$

- ESA Euclid: Bias requirements to estimate cosmological parameters within ~1%

$Survey$	$Area$	n_{gal}	z	m	c
Current	200	15	0.8	2×10^{-2}	1×10^{-3}
Upcoming	5000	15	0.8	4×10^{-3}	6×10^{-4}
Future (Euclid)	15000	35	0.9	$\boxed{1 \times 10^{-3}}$	$\boxed{3 \times 10^{-4}}$

- A *tenfold* improvement in bias required till the launch of Euclid telescope in ~2020…

Big Data in Weak Lensing

- Volume
  - **Euclid** imaging survey: 15’000 deg^2 ~4 times in mult. bands
    ~10 Petabytes raw space data
- Velocity
  - Large Synoptic Telecope (**LSST**)
    20’000 deg^2 in 6 bands, ~15 Terabytes / day raw data rate
  - **SKA** radio survey: ~3 Terabytes / second
- Variety
  - Heterogeneous datasets, ground+space, variable depth, multiple bands, different cameras, etc.
- Veracity
  - Low S/N, missing/incomplete/abnormal data, etc.

Big Data challenges for shear measurement

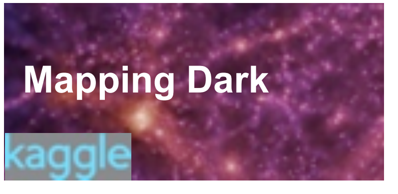
- Algorithms currently too slow / high volumes
  - Typical method processes ~1 galaxy per second
  - Either speed-up algorithms or apply massively distributed computing techniques (e.g. map-reduce)
- Algorithms must cope with data variety
  - Handle missing/incomplete/abnormal information
  - Exploit value in data from different source/nature
- But, Big Data also provides opportunities
  - Accessing more data can potentially yield better accuracy
  - More high S/N deep data, better calibration
    => improved bias correction
  - Applying new technologies/paradigms - Lessons from industry

Machine Learning (ML) can help

- All methods rely on calibrations for correcting biases
  - Requires expert knowledge in astrophysics
  - Requires expert knowledge of the method itself
  - Potential for improvements poorly explored so far
- Possibly the use of Machine Learning can help
  - Entirely new approaches for shape measurement
  - Help improve existing methods: speed, accuracy
- Machine Learning for calibration shown to work

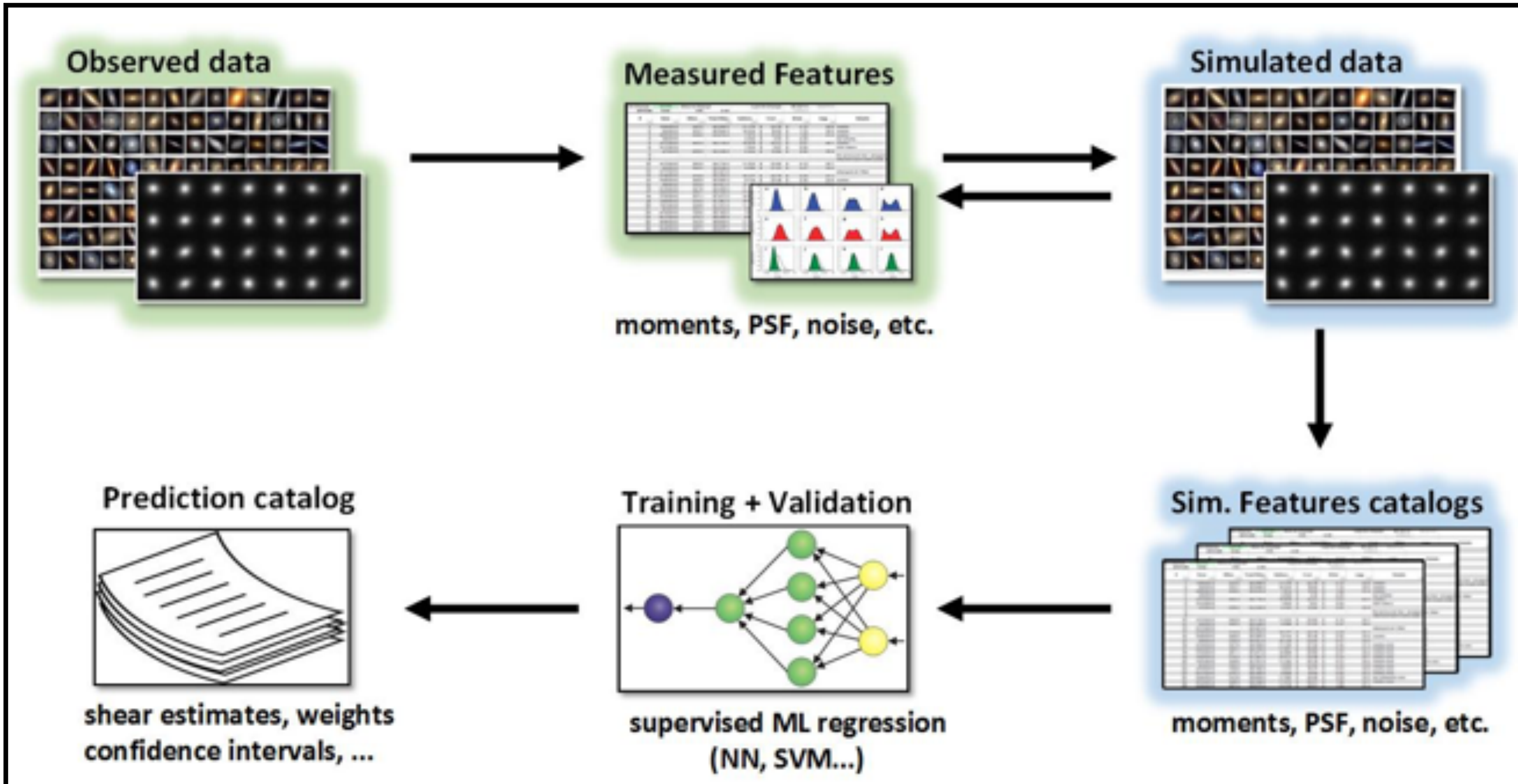
GREAT10

GREAT3



Machine Learning (ML) can help

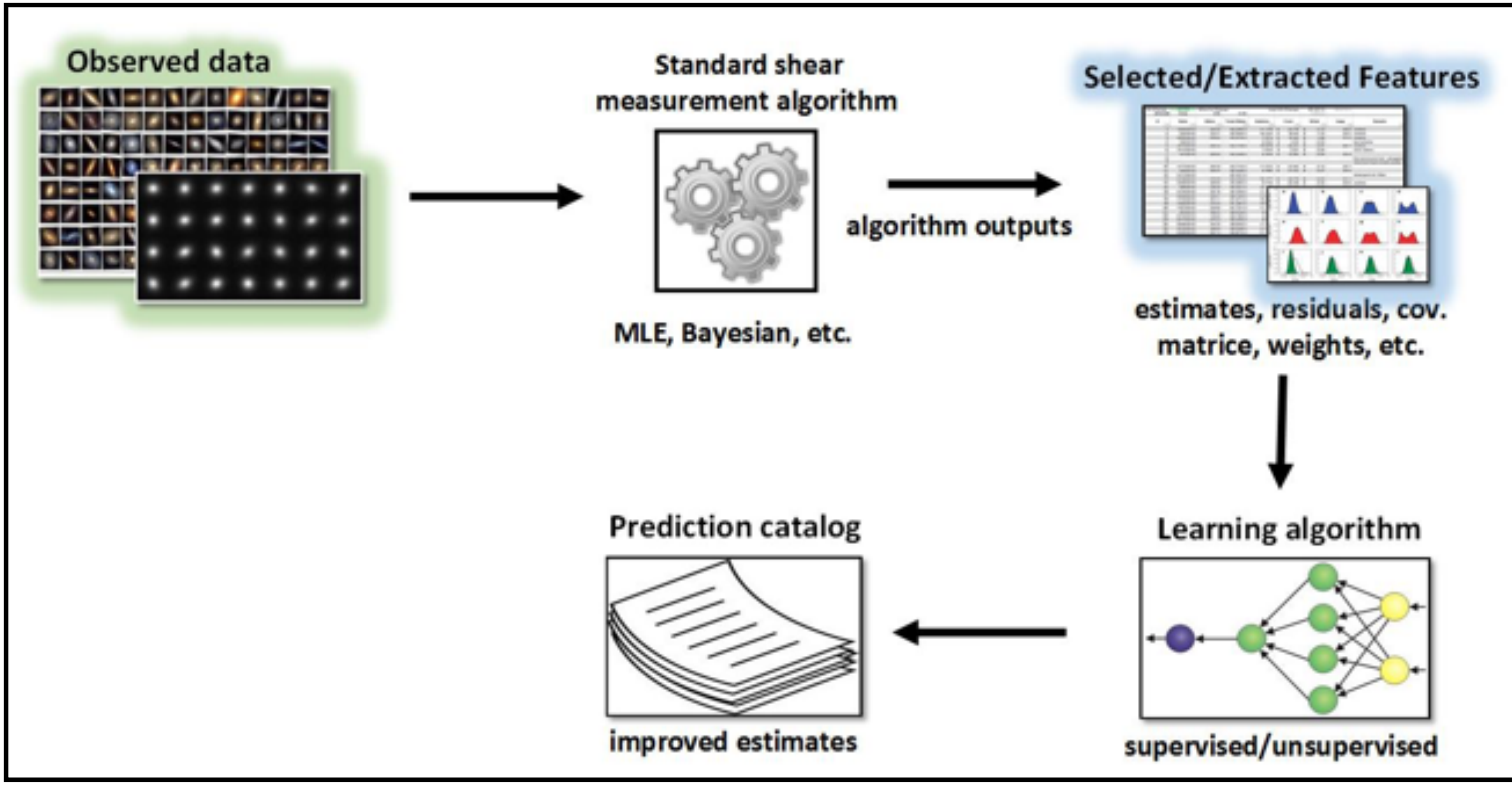
- Example: “Pure” ML approach based on simulations



Machine Learning as Direct Measurement

Machine Learning (ML) can help

- Example: "Hybrid" ML approach: automated bias calibration



Machine Learning as Calibration Scheme

Summary

- High potential of weak lensing and shear measurement for cosmology
- But accurate shear measurement is extremely difficult
- Current algorithms are still insufficiently accurate to help discriminate cosmological model within 1% precision
- Algorithms are too slow, especially with upcoming massive surveys (Big Data)
- One can learn lessons from the industry or other fields
- Machine Learning techniques can help improving algorithms accuracy and speed