Impact of mass mapping for cosmological parameter estimation

June 6th, 2024

Andreas Tersenov

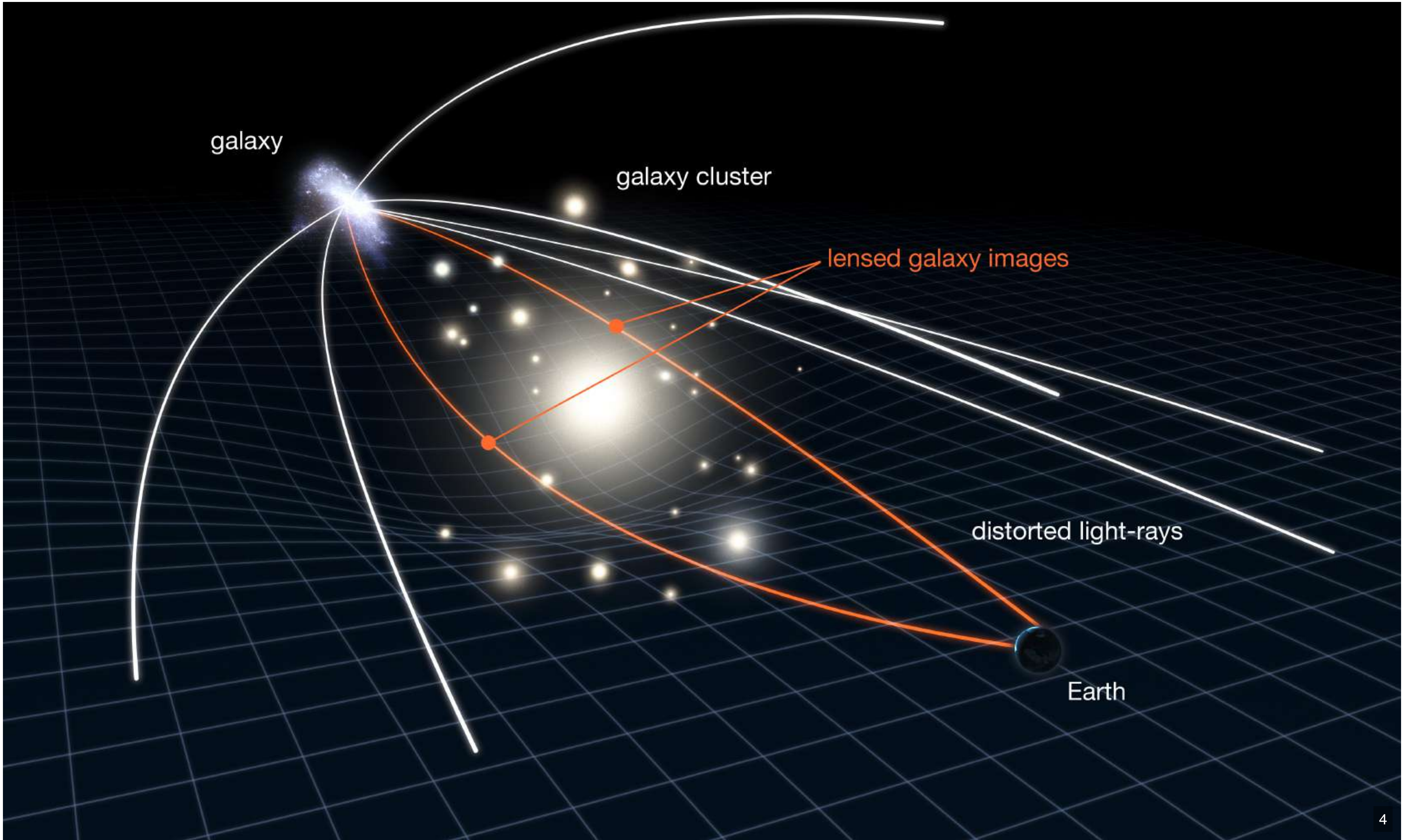
FORTH INSTITUTE OF COMPUTER SCIENCE
FORTH INSTITUTE OF ASTROPHYSICS

slides at andreastersenov.github.io/talks/TITAN_Crete_2024/

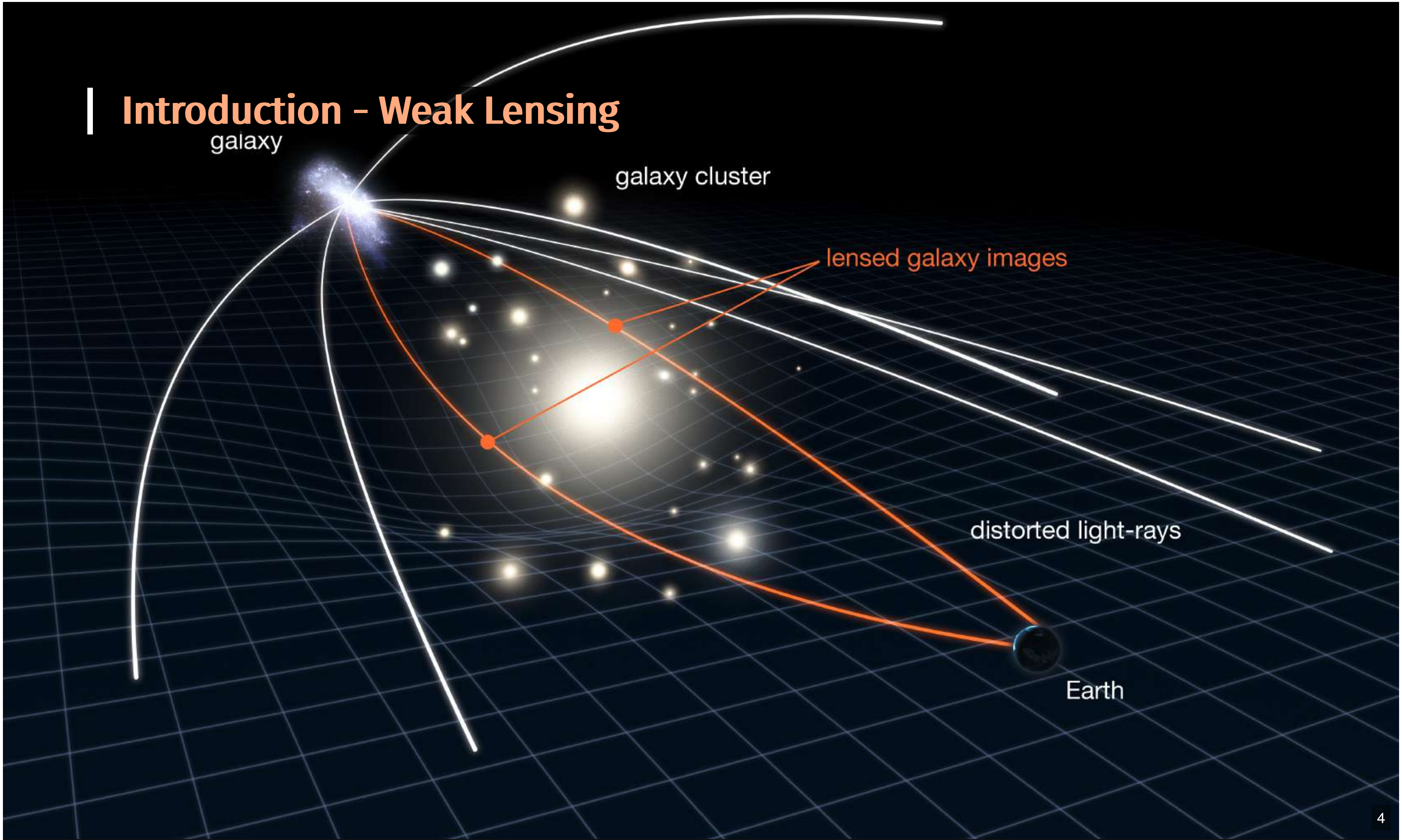
General Idea

Create a pipeline:

- Use it to investigate the impact of **mass mapping algorithms** on cosmology constraints.
- Use it on **UNIONS/CFIS** data to get observational constraints on cosmology.
- Use it to test **novel mass mapping methods**

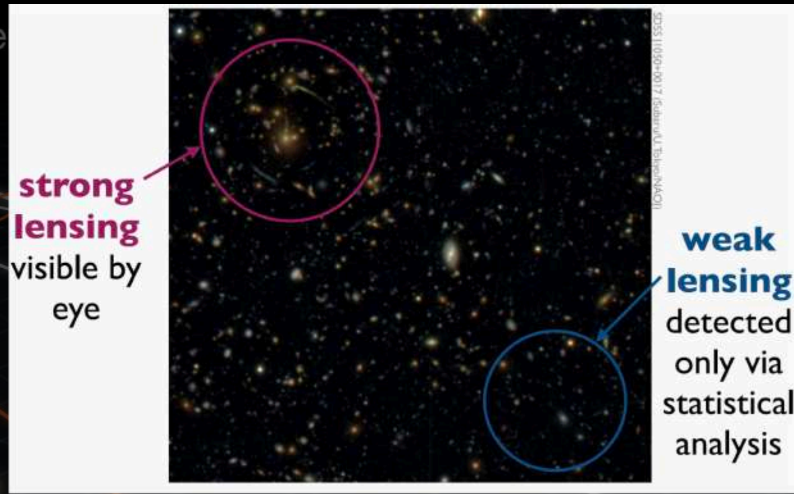

galaxy
galaxy cluster
lensed galaxy images
distorted light-rays
Earth

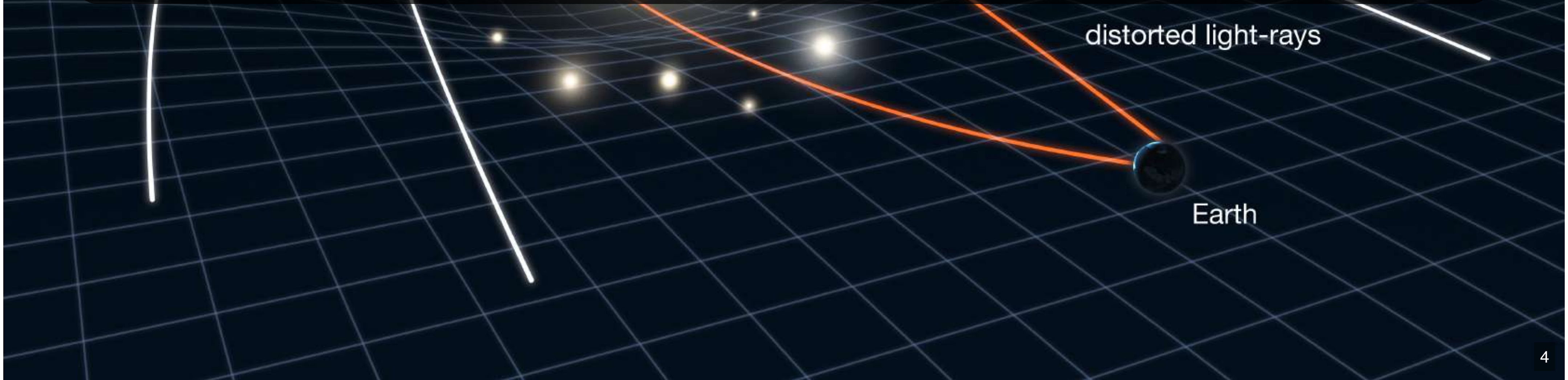
Introduction - Weak Lensing



Introduction – Weak Lensing

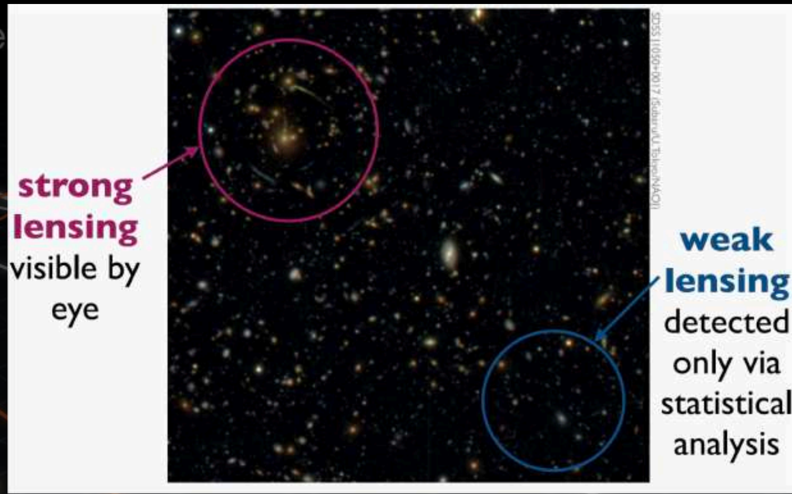
- **WL** = Observational technique in cosmology for studying the matter distribution in the universe
- **Principle:** deflection of light from distant galaxies by gravitational fields → causes image distortion
- **Weak** → subtle & coherent distortions of background galaxy shapes





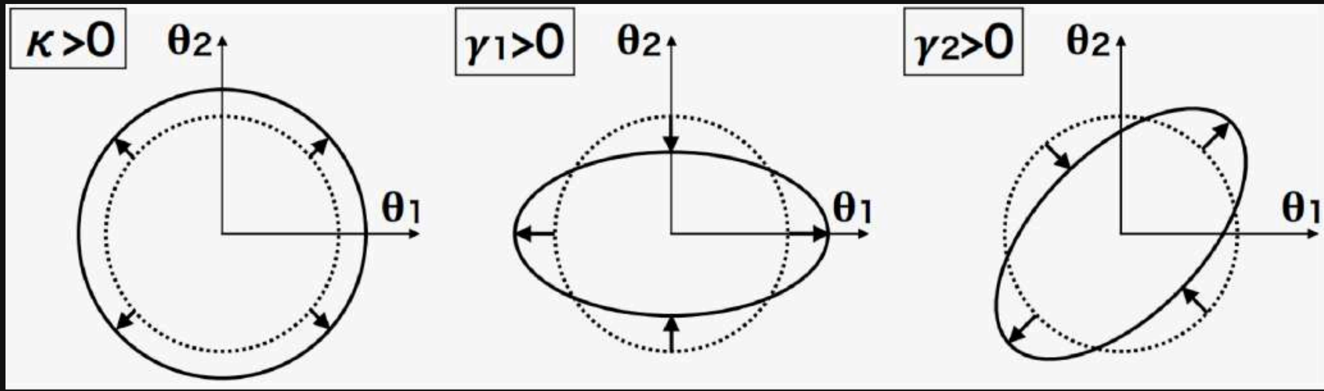
Introduction – Weak Lensing

- **WL** = Observational technique in cosmology for studying the matter distribution in the universe
- **Principle:** deflection of light from distant galaxies by gravitational fields → causes image distortion
- **Weak** → subtle & coherent distortions of background galaxy shapes

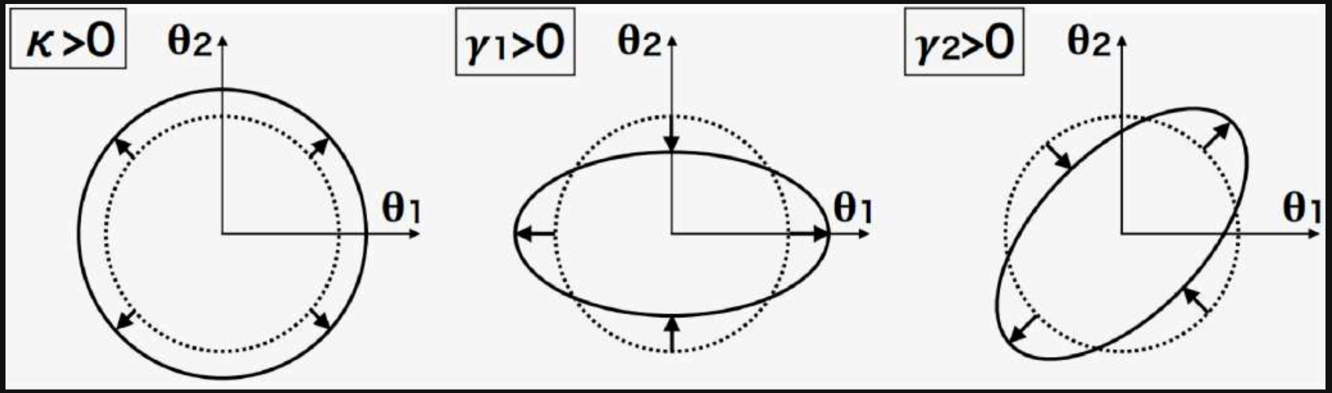


- WL provides a direct measurement of the gravitational distortion.
- WL enables us to probe the cosmic structure, investigate the nature of dark matter, and constrain cosmological parameters.

Shear & Convergence



Shear & Convergence



Convergence κ

$$\kappa = \frac{1}{2}(\partial_1\partial_1 + \partial_2\partial_2)\psi = \frac{1}{2}\nabla^2\psi$$

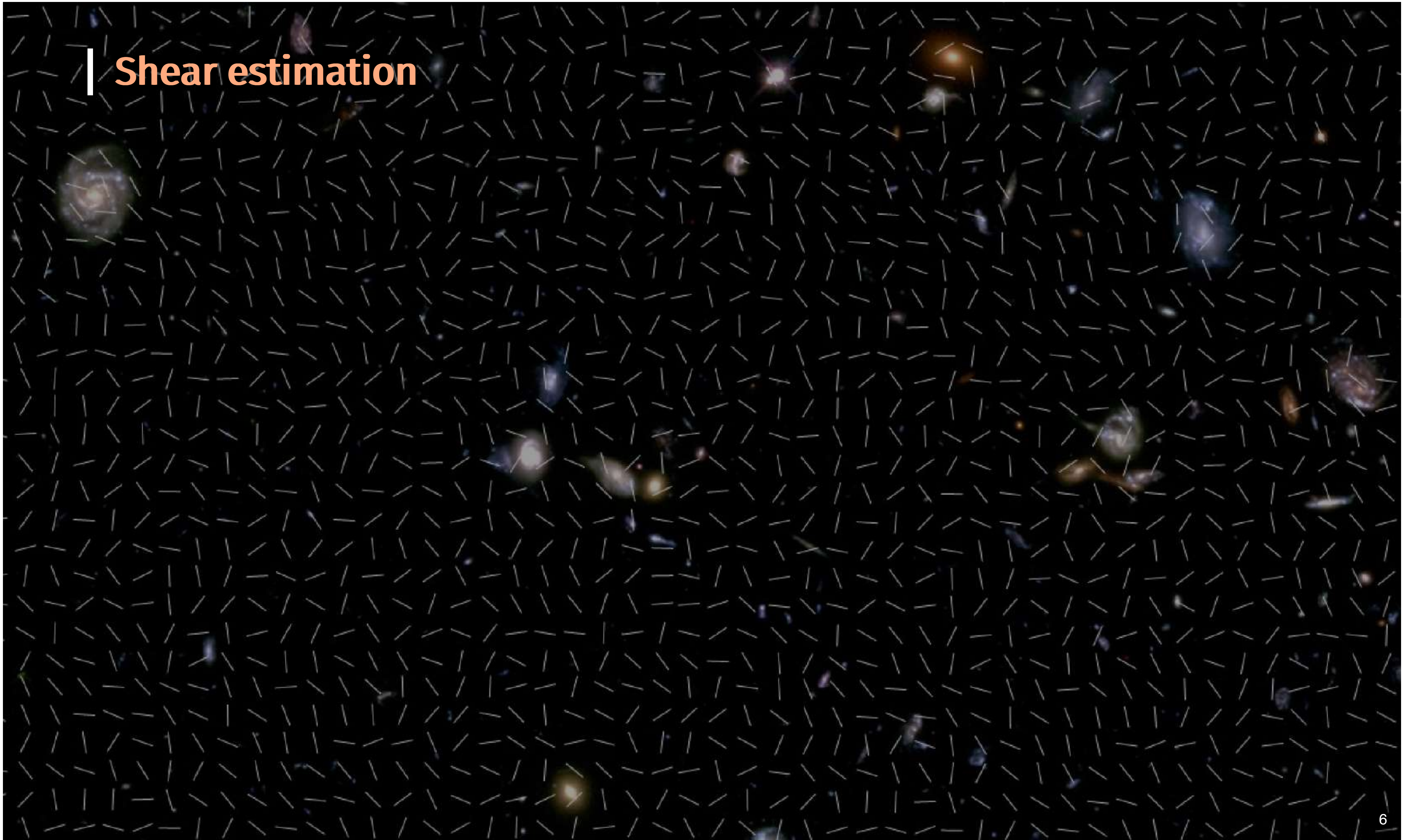
$\longrightarrow$ difficult to measure

Shear γ

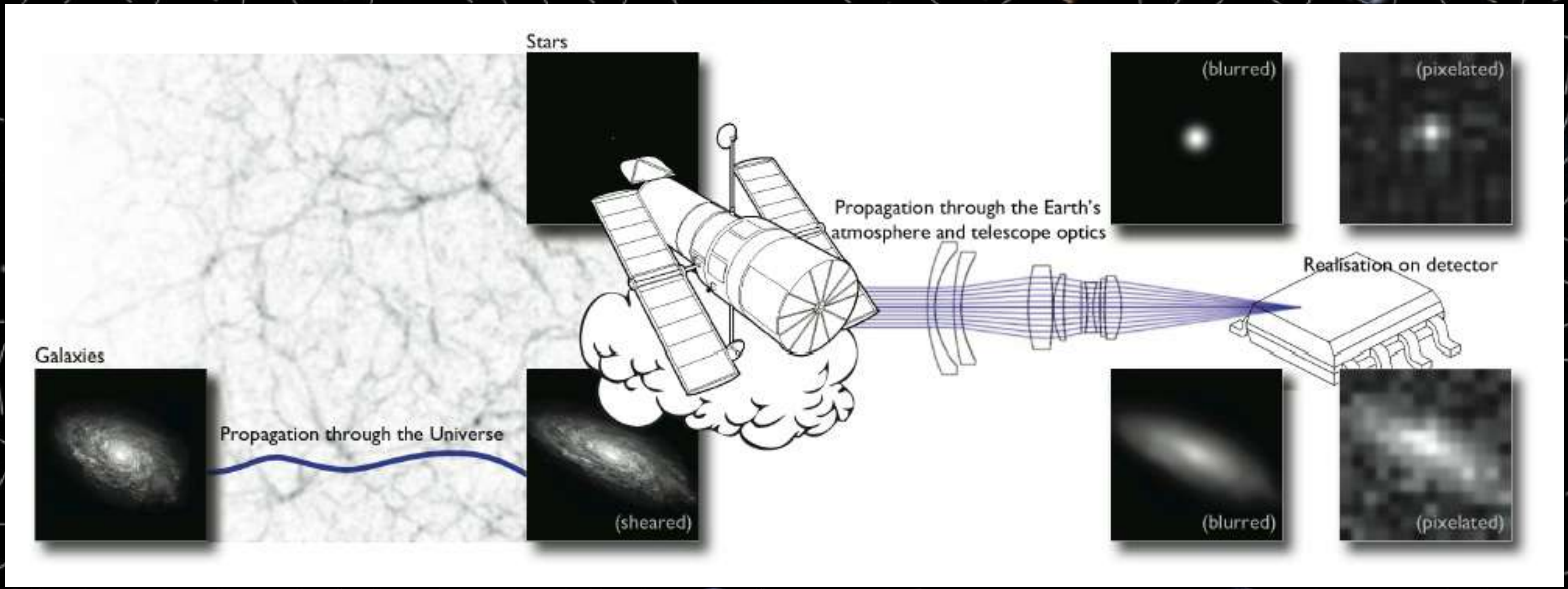
$$\gamma_1 = \frac{1}{2}(\partial_1\partial_1 - \partial_2\partial_2)\psi, \;\; \gamma_2 = \partial_1\partial_2\psi$$

$\longrightarrow$ can be measured by statistical analysis of galaxy shapes

Shear estimation



Shear estimation

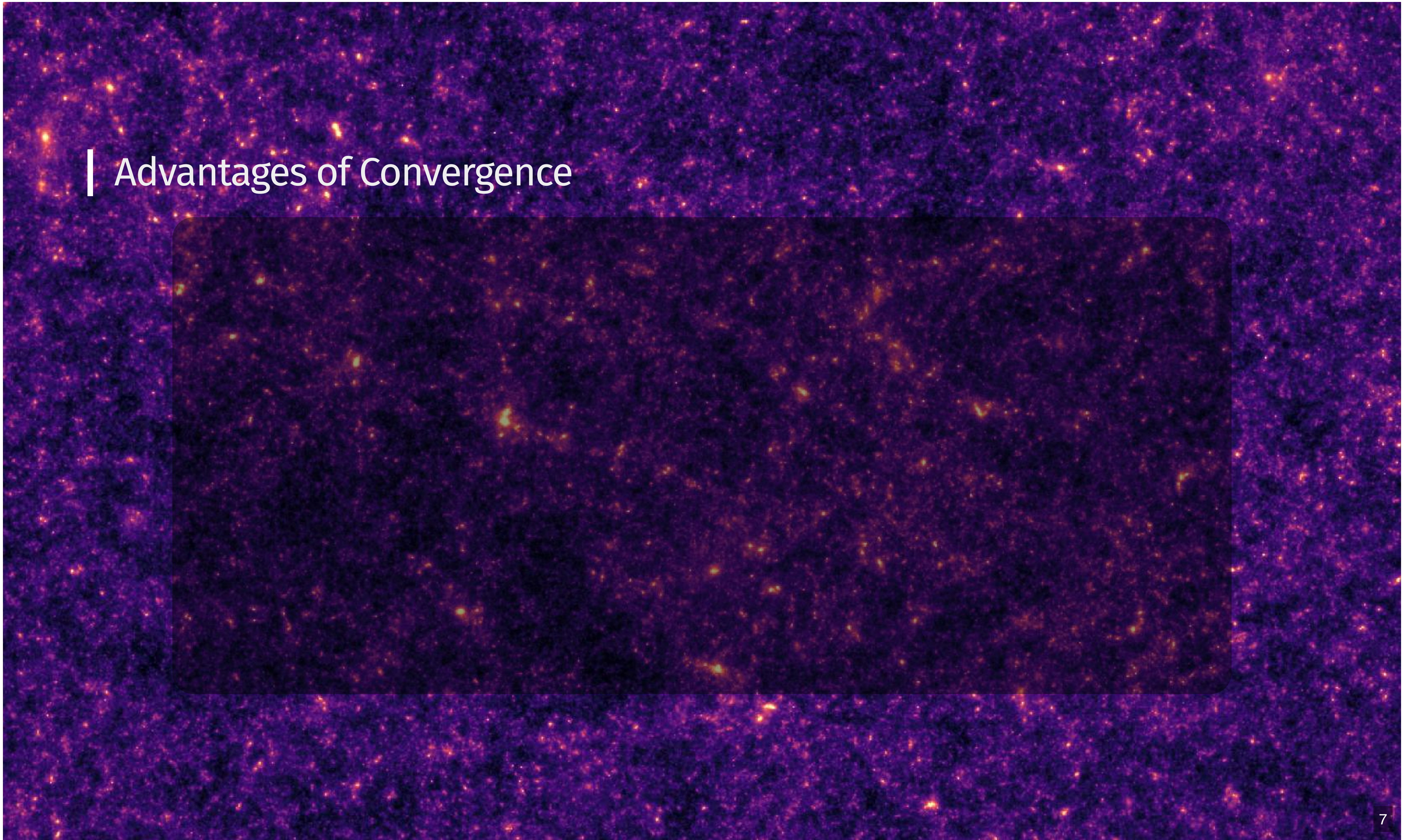


Galaxy shapes as estimators for gravitational shear

$$e = \gamma + e_i \qquad \text{with} \qquad e_i \sim \mathrm{N}(0, I)$$

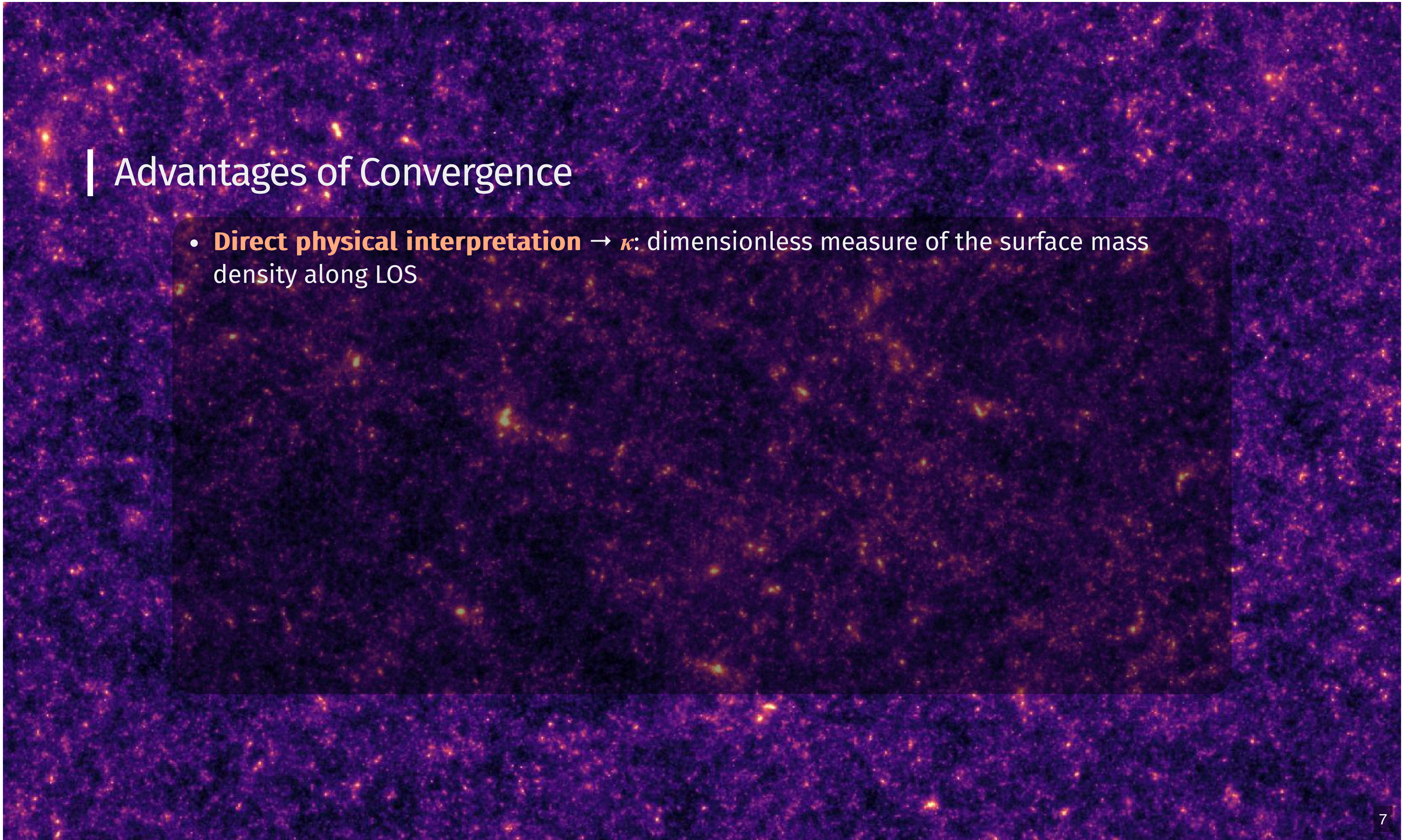
- We are trying the measure the **ellipticity** e of galaxies as an estimator for the **gravitational shear** γ

Advantages of Convergence



Advantages of Convergence

- **Direct physical interpretation** → κ: dimensionless measure of the surface mass density along LOS



Advantages of Convergence

- **Direct physical interpretation** → κ: dimensionless measure of the surface mass density along LOS
- **κ-map shows the LSS of the Universe** → allows comparisons with other tracers of matter distribution

Advantages of Convergence

- **Direct physical interpretation** → κ: dimensionless measure of the surface mass density along LOS
- **κ-map shows the LSS of the Universe** → allows comparisons with other tracers of matter distribution
- **Probes dark matter and its properties directly**

Advantages of Convergence

- **Direct physical interpretation** → κ: dimensionless measure of the surface mass density along LOS
- **κ-map shows the LSS of the Universe** → allows comparisons with other tracers of matter distribution
- **Probes dark matter and its properties directly**
- **Easier to work with than shear**: scalar quantity instead of spin-2 field

Advantages of Convergence

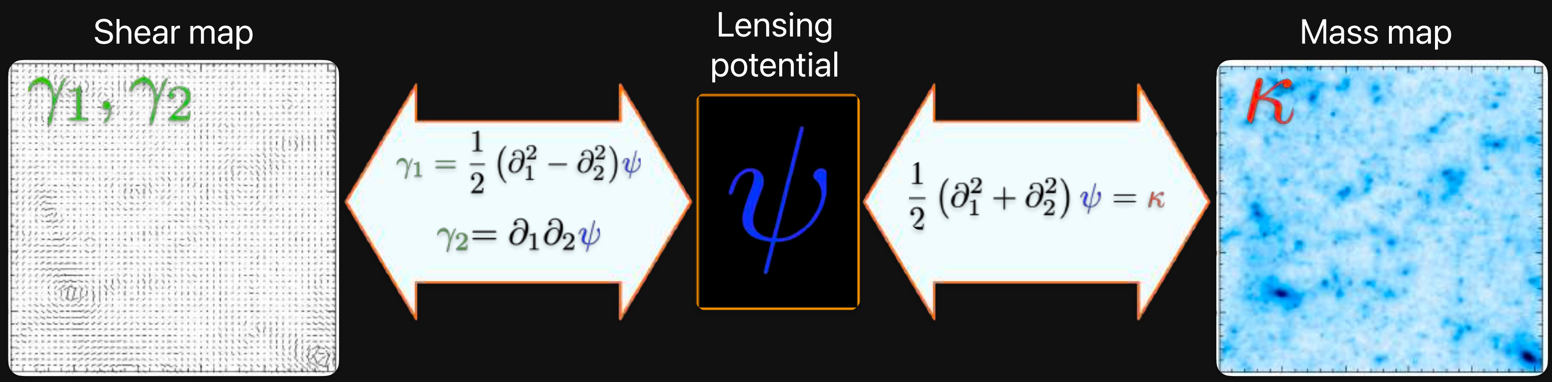
- **Direct physical interpretation** → κ: dimensionless measure of the surface mass density along LOS
- **κ-map shows the LSS of the Universe** → allows comparisons with other tracers of matter distribution
- **Probes dark matter and its properties directly**
- **Easier to work with than shear**: scalar quantity instead of spin-2 field
- **Allows higher order statistics** in a convenient way

Advantages of Convergence

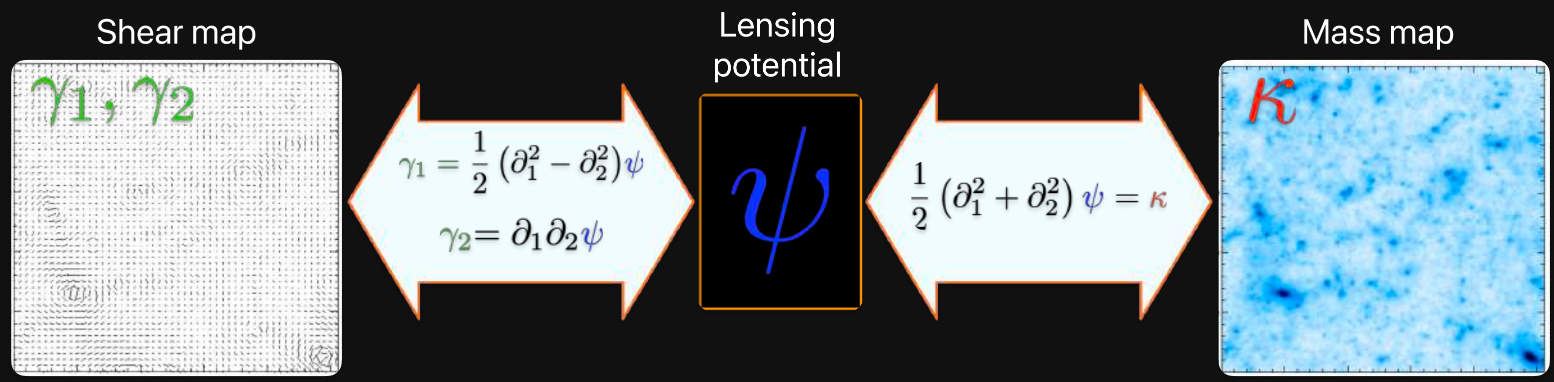
- **Direct physical interpretation** → κ: dimensionless measure of the surface mass density along LOS
- **κ-map shows the LSS of the Universe** → allows comparisons with other tracers of matter distribution
- **Probes dark matter and its properties directly**
- **Easier to work with than shear**: scalar quantity instead of spin-2 field
- **Allows higher order statistics** in a convenient way
- **Cosmological parameter estimations** → can be used to constrain cosmological parameters

Relation between κ and γ

Relation between κ and γ



Relation between κ and γ



- From **convergence** to **shear**: $\gamma_i = \hat{P}_i\kappa$
- From **shear** to **convergence**: $\kappa = \hat{P}_1\gamma_1 + \hat{P}_2\gamma_2$

$$\hat{P}_1(k) = \frac{k_1^2 - k_2^2}{k^2}, \quad \hat{P}_2(k) = \frac{2k_1k_2}{k^2}$$

Kaiser-Squires inversion

Kaiser-Squires inversion

Advantages:

- Simple ***linear*** operator
- Very easy to implement in Fourier space
- Optimal, in theory

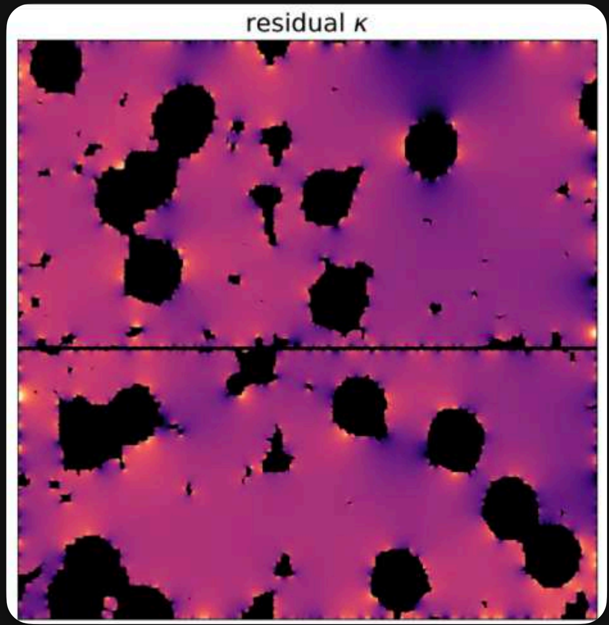
Kaiser-Squires inversion

Advantages:

- Simple ***linear*** operator
- Very easy to implement in Fourier space
- Optimal, in theory

Practical difficulties:

- Shear measurements are discrete, **noisy**, and **irregularly sampled**
- We actually measure the **reduced shear**: $g = \gamma/(1-\kappa)$
- Masks and integration over a subset of $\mathbb{R}^2$ lead to border errors ⇒ **missing data problem**
- Convergence is recoverable up to a constant ⇒ **mass-sheet degeneracy problem**



Bayesian reconstruction

Bayesian reconstruction

- Mass mapping problem → statistical inference problem
- **Goal**: infer most probable value of κ-field given observed shear data

$$\underbrace{p(\boldsymbol{\kappa} \mid \boldsymbol{\gamma}, \mathrm{M})}_{\text{Posterior}} \propto \underbrace{p(\gamma \mid \boldsymbol{\kappa}, \mathrm{M})}_{\text{likelihood}} \underbrace{p(\boldsymbol{\kappa} \mid \mathrm{M})}_{\text{prior}}$$

M: cosmological model

Bayesian reconstruction

- Mass mapping problem → statistical inference problem
- **Goal**: infer most probable value of κ-field given observed shear data

$$\underbrace{p(\boldsymbol{\kappa} \mid \boldsymbol{\gamma}, \mathrm{M})}_{\text{Posterior}} \propto \underbrace{p(\gamma \mid \boldsymbol{\kappa}, \mathrm{M})}_{\text{likelihood}} \underbrace{p(\boldsymbol{\kappa} \mid \mathrm{M})}_{\text{prior}}$$

M: cosmological model

- **Likelihood distribution**: prob. of observing γ data given true κ → encodes the forward process of the model → *contains the physics*

- **Prior distribution**: encodes the knowledge about the signal before observing data

  - **Log-likelihood** (when the noise is white Gaussian)

$$\log p(\gamma \mid \kappa) = -\frac{1}{2}\left(\gamma - \mathbf{F}^*\mathbf{PF}\kappa\right)^\dagger \Sigma_n^{-1}\left(\gamma - \mathbf{F}^*\mathbf{PF}\kappa\right) + \text{ constant}$$

  - **Maximum A Posteriori solution**

$$\hat{x} = \arg\max_x \ \log p(y|x) + \log p(x)$$

Wiener filter

- Assumes **prior** on κ → **Gaussian random field**

$$p_{\mathrm{Gauss}}(\kappa) = \frac{1}{\sqrt{\det 2\pi \mathbf{S}}} \exp\left(-\frac{1}{2}\tilde{\kappa}^{\dagger}\mathbf{S}^{-1}\tilde{\kappa}\right)$$

- Likelihood (assuming uncorrelated, Gaussian noise) → also **Gaussian**

$$p(\gamma|\kappa) = \frac{1}{\sqrt{2\pi \det \mathbf{N}}} \exp\left[-\frac{1}{2}(\gamma - \mathbf{A}\kappa)^{\dagger}\mathbf{N}^{-1}$$

- The Gaussian prior ***encodes the assumption*** that the fluctuations in the κ-field are well described by a Gaussian random field, with **power spectrum** given by the cosmological model

Convergence power spectrum

$$\langle \tilde{\kappa}(\boldsymbol{k})\tilde{\kappa}^{*}(\boldsymbol{k}')\rangle = (2\pi)^2\delta_D(\boldsymbol{k}-\boldsymbol{k}')P_{\kappa}(k)$$

Stat. measure of the spatial distribution of the convergence field → quantifies the amplitude of the fluctuations in κ as function of their spatial scale

Wiener filter

Wiener solution of the inverse problem

$$\hat{\kappa}_{\text{wiener}} = \arg\min_{\kappa} \left\| \Sigma^{-1/2} \left(\gamma - \mathbf{F}^{*}\mathbf{P}\mathbf{F}\kappa \right) \right\|_2^2 + \log p_{\text{Gaussian}} (\kappa$$

- This solution corresponds to the **maximum a posteriori** (MAP) solution under the assumption of a Gaussian prior on κ, and it matches the **mean** of the Gaussian posterior.

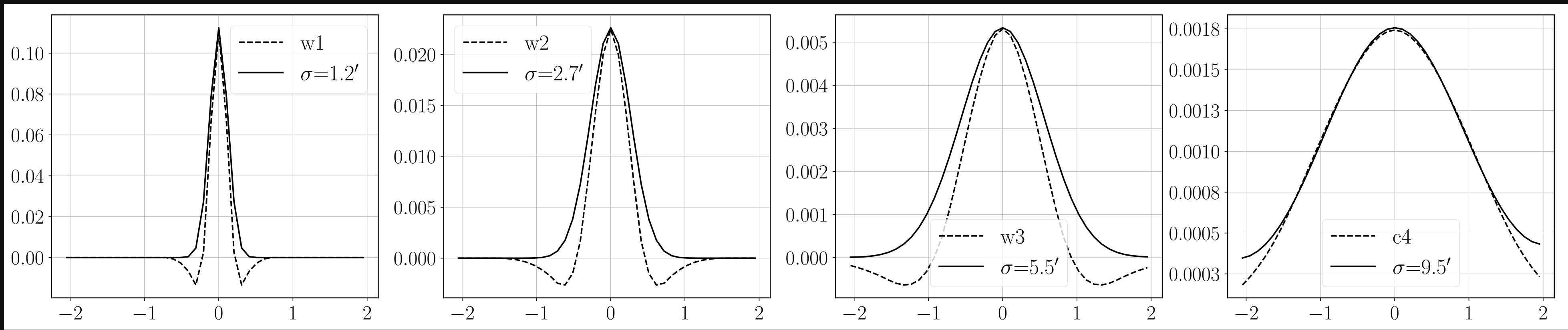
Wiener reconstruction

$$\hat{\kappa}_{\text{wiener}} = \mathbf{S}\mathbf{P}^{\dagger} \left[\mathbf{P}\mathbf{S}\mathbf{P}^{\dagger} + \mathbf{N} \right]^{-1} \tilde{\gamma}$$

Sparse recovery

- Decomposes the signal into another domain (dictionary), where it is **sparse**

- Implement the ***wavelet transform*** → decomposes the signal into **wavelet functions** (waveforms of limited duration with an average value of zero)



- Use **starlet wavelets** → represent well structures resembling the κ of a **DM halo** (positive & isotropic)

- The application of sparsity prior ***enforces*** a cosmological model where the matter field is a combination of spherically symmetric DM halos

MCALens

- Models κ-field as a sum of a **Gaussian** and **non-Gaussian** component

$$\kappa = \underbrace{\kappa_{\mathrm{NG}}}_{\text{Standard Wiener filter approach}} + \underbrace{\kappa_{\mathrm{G}}}_{\text{Modified wavelet approach}}$$

$$\min_{\kappa_G,\kappa_{NG}} \|\gamma - \mathbf{A}(\kappa_G + \kappa_{NG})\|^2_{\Sigma_n} + C_{\mathrm{G}}(\kappa_G) + C_{\mathrm{NG}}(\kappa_{NG})$$

- **MCA** (morphological Component Analysis) performs an alternating minimization scheme:
  - Estimate κ_{G} assuming κ_{NG} is known:

$$\min_{\kappa_G} \|(\gamma - \mathbf{A}\kappa_{NG}) - \mathbf{A}\kappa_G\|^2_{\Sigma_n} + C_{\mathrm{G}}(\kappa_G)$$

  - Estimate κ_{NG} assuming κ_{G} is known:

$$\min_{\kappa_{NG}} \|(\gamma - \mathbf{A}\kappa_G) - \mathbf{A}\kappa_{NG}\|^2_{\Sigma_n} + C_{\mathrm{NG}}(\kappa_{NG})$$

Overview of mass mapping methods

Overview of mass mapping methods

- **Wiener filter**: Prior on κ → Gaussian random field, Likelihood → Gaussian, Solution → corresponds to MAP for gaussian κ

Overview of mass mapping methods

- **Wiener filter**: Prior on κ → Gaussian random field, Likelihood → Gaussian, Solution → corresponds to MAP for gaussian κ
- **Sparse recovery**: Prior on κ → sparse in wavelet basis, Enforces model where matter field is a combination of DM halos

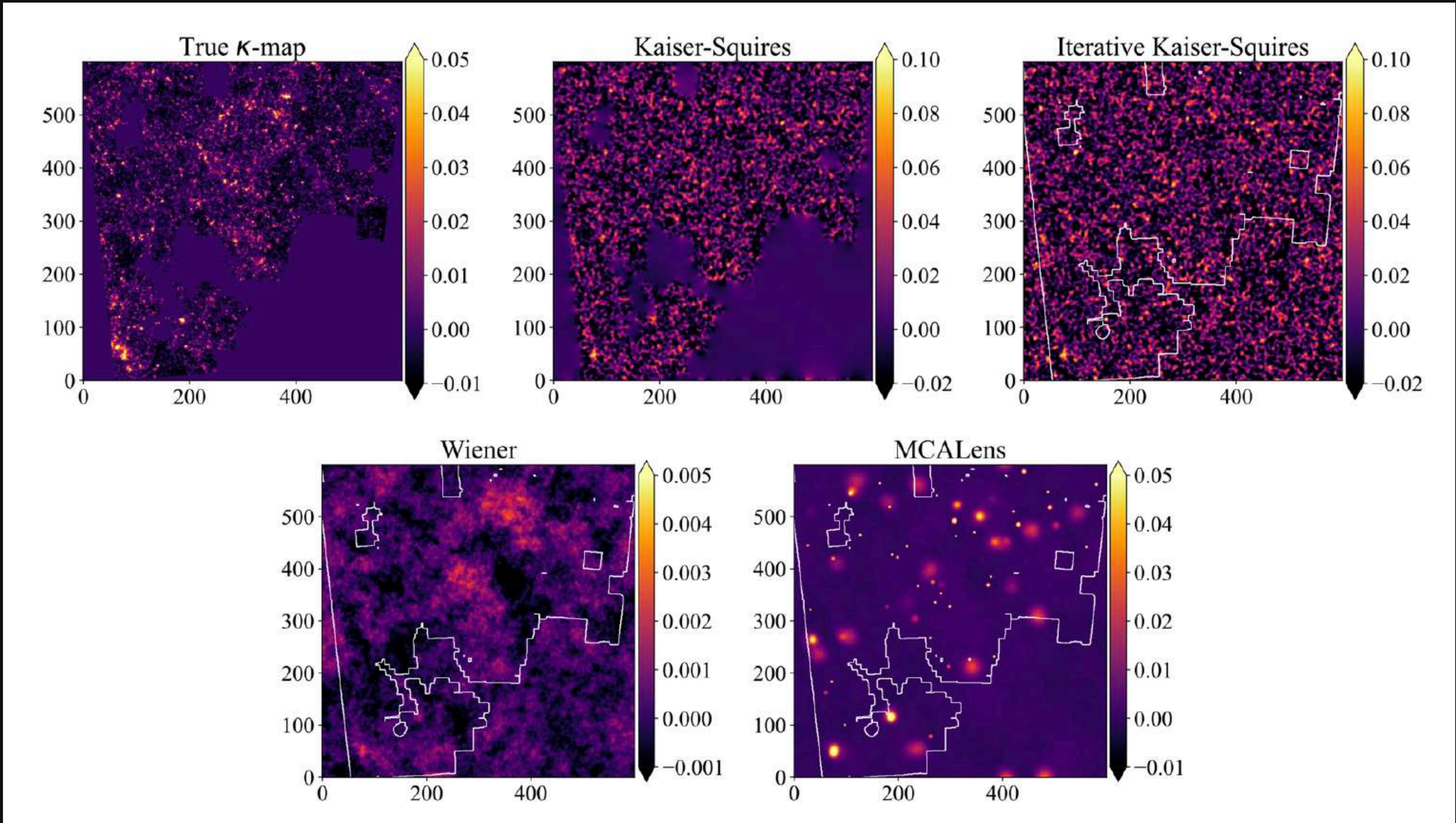
Overview of mass mapping methods

- **Wiener filter**: Prior on κ → Gaussian random field, Likelihood → Gaussian, Solution → corresponds to MAP for gaussian κ
- **Sparse recovery**: Prior on κ → sparse in wavelet basis, Enforces model where matter field is a combination of DM halos
- **MCALens**: Models κ-field as sum of a Gaussian and a non-Gaussian component, Uses Morphological Component Analysis

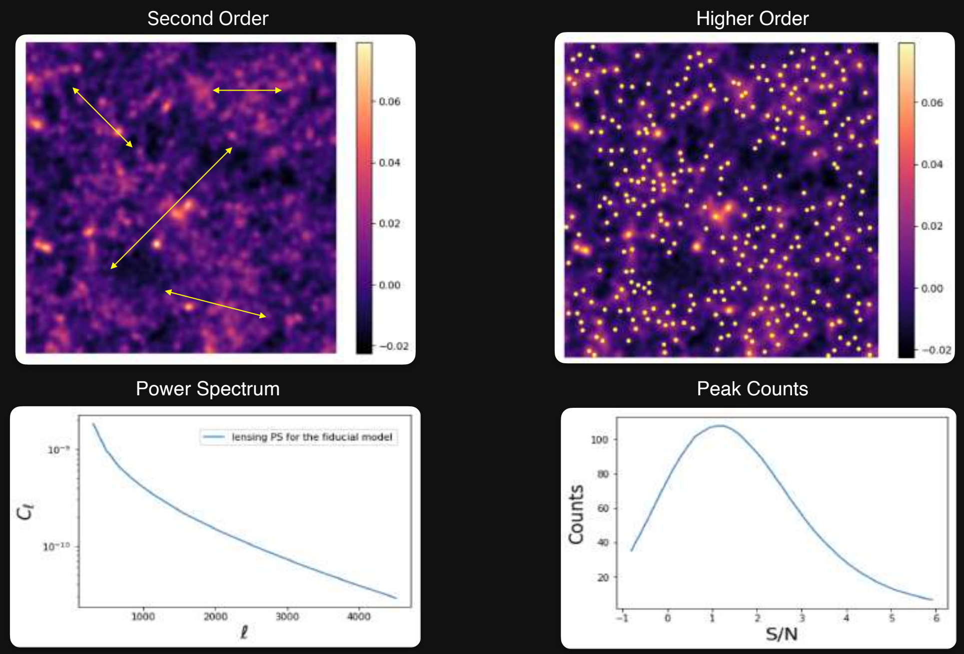
Overview of mass mapping methods

- **Wiener filter**: Prior on κ → Gaussian random field, Likelihood → Gaussian, Solution → corresponds to MAP for gaussian κ
- **Sparse recovery**: Prior on κ → sparse in wavelet basis, Enforces model where matter field is a combination of DM halos
- **MCALens**: Models κ-field as sum of a Gaussian and a non-Gaussian component, Uses Morphological Component Analysis
- **Deep Learning Methods**:
  - **DeepMass**: CNN with a U-Net-based architecture, prior from simulations
  - **DeepPosterior**: Probabilistic mass mapping with deep generative models, Prior from 2pt statistics modelling at large scales & Deep Learning on simulations for small scales, Sampling with Annealed HMC

Mass mapping methods:



Higher Order Statistics: Peak Counts



- **Peaks**: local maxima of the SNR field

$$\nu = \frac{(\mathcal{W} * \kappa)(\theta_{\mathrm{ker}})}{\sigma_n^{\mathrm{filt}}}$$

- Peaks trace regions where the value of κ **is high** → they are associated to **massive structures**

Wavelet peaks

Wavelet peaks

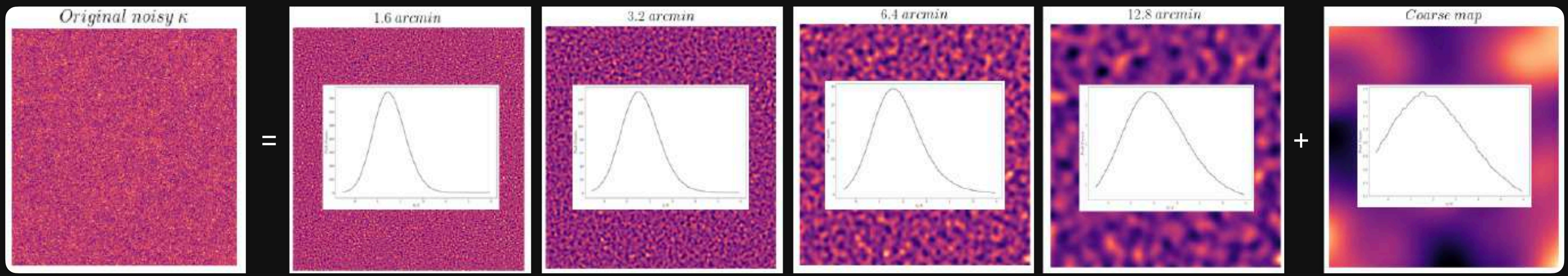
- We consider a **multi-scale analysis** compared to a single-scale analysis

Wavelet peaks

- We consider a **multi-scale analysis** compared to a single-scale analysis
- Apply (instead of Gaussian filter) a **starlet transform** → allows us to represent an image I as a sum of wavelet coefficient images and a coarse resolution image

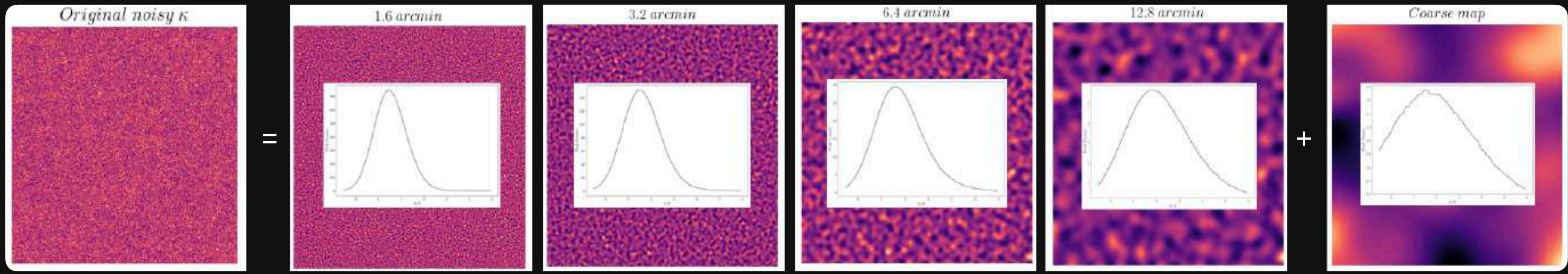
Wavelet peaks

- We consider a **multi-scale analysis** compared to a single-scale analysis
- Apply (instead of Gaussian filter) a **starlet transform** → allows us to represent an image I as a sum of wavelet coefficient images and a coarse resolution image



Wavelet peaks

- We consider a **multi-scale analysis** compared to a single-scale analysis
- Apply (instead of Gaussian filter) a **starlet transform** → allows us to represent an image I as a sum of wavelet coefficient images and a coarse resolution image



- Allows for the **simultaneous** processing of data at different scales → **efficiency**
- Each wavelet band covers a different frequency range, which leads to an almost **diagonal peak count covariance matrix**

What happens if we consider all pixels instead of selecting multi-scale minima and maxima?

Starlet ℓ_1-norm

Starlet ℓ_1-norm

- New higher order summary statistic for weak lensing observables

Starlet ℓ_1-norm

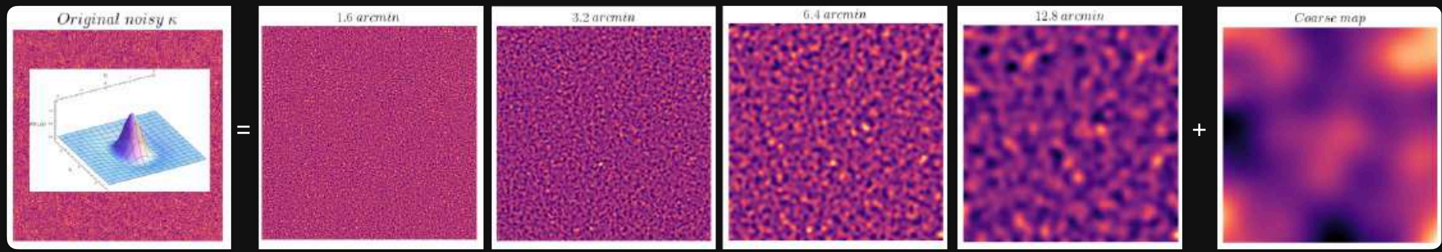
- New higher order summary statistic for weak lensing observables
- Provides a fast multi-scale calculation of the full void and peak distribution

Starlet ℓ_1-norm

- New higher order summary statistic for weak lensing observables
- Provides a fast multi-scale calculation of the full void and peak distribution
- **ℓ_1-norm** = sum of absolute values of the starlet decomposition coefficients of a weak lensing map

Starlet ℓ_1-norm

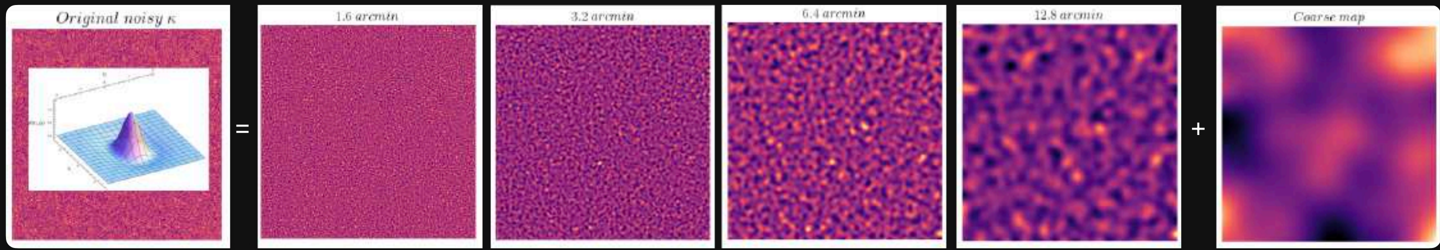
- New higher order summary statistic for weak lensing observables
- Provides a fast multi-scale calculation of the full void and peak distribution
- **ℓ_1-norm** = sum of absolute values of the starlet decomposition coefficients of a weak lensing map



$$\ell_1^{j,i} = \sum_{u=1}^{\# \mathrm{coef}(S_{i,j})} \left| S_{j,i}[u] \right| = \| S_{j,i} \|_1 , \ \ S_{j,i} = \{ w_{j,k} / B_i w_{j,k} B_{i+1} \}$$

Starlet ℓ_1-norm

- New higher order summary statistic for weak lensing observables
- Provides a fast multi-scale calculation of the full void and peak distribution
- **ℓ_1-norm** = sum of absolute values of the starlet decomposition coefficients of a weak lensing map



$$\ell_1^{j,i} = \sum_{u=1}^{\# \mathrm{coef}(\mathrm{S}_{i,j})} \left|\mathrm{S}_{j,i}[u]\right| = \|\mathrm{S}_{j,i}\|_1, \;\; \mathrm{S}_{j,i} = \{w_{j,k}/B_i w_{j,k} B_{i+1}\}$$

- Information encoded in **all** pixels
- Automatically includes peaks and voids
- Multi-scale approach
- Avoids the problem of defining peaks and voids

Inference with HOS

Inference with HOS

- Use **cosmoSLICS** simulations: suite designed for the analysis of WL data beyond the standard 2pt cosmic shear

Inference with HOS

- Use **cosmoSLICS** simulations: suite designed for the analysis of WL data beyond the standard 2pt cosmic shear

- **cosmoSLICS** cover a wide parameter space in $\left[\Omega_m, \sigma_8, w_0, h\right]$.

Inference with HOS

- Use **cosmoSLICS** simulations: suite designed for the analysis of WL data beyond the standard 2pt cosmic shear

- **cosmoSLICS** cover a wide parameter space in $\left[\Omega_m, \sigma_8, w_0, h\right]$.

- For **Bayesian inference** → use a **Gaussian likelihood** for a cosmology independent covariance, and a **flat prior**.

Inference with HOS

- Use **cosmoSLICS** simulations: suite designed for the analysis of WL data beyond the standard 2pt cosmic shear

- **cosmoSLICS** cover a wide parameter space in $\left[\Omega_m, \sigma_8, w_0, h\right]$.

- For **Bayesian inference** → use a **Gaussian likelihood** for a cosmology independent covariance, and a **flat prior**.

- To have a prediction of each HOS **given a new set of parameters**→ employ an interpolation with **Gaussian Process Regressor** (GPR)

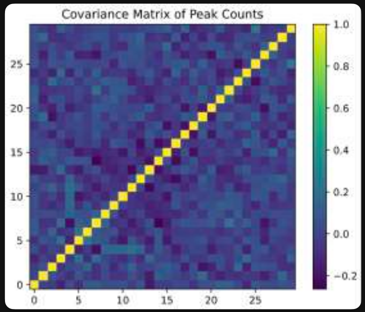
Noise & Covariance

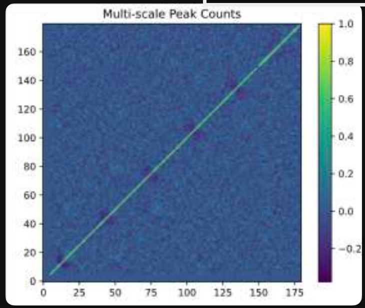
- We consider **Gaussian**, but **non-white** noise → the noise depends on the **number of galaxies** in each pixel

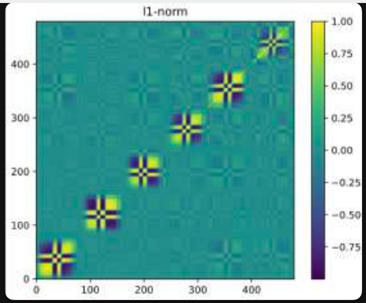
$$\sigma_n = \frac{\sigma_\epsilon}{\sqrt{2 n_{\text{gal}} A_{\text{pix}}}}$$

- We incorporate **masks**
- Calculate covariance:

$$\boxed{C_{ij} = \sum_{r=1}^{N} \frac{\left(x_i^r - \mu_i\right)\left(x_j^r - \mu_j\right)}{N-1}}, \quad \mu_i = \frac{1}{N} \sum_r x_i^r$$



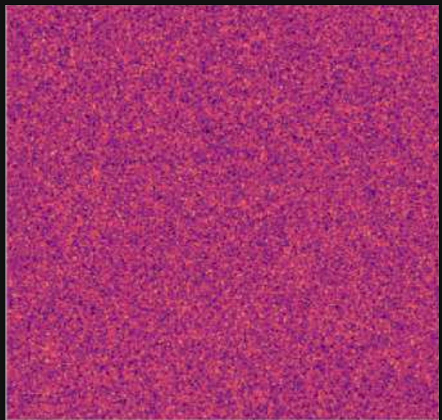




Starlet filter tends to make the covariance matrix more diagonal

From data to contours

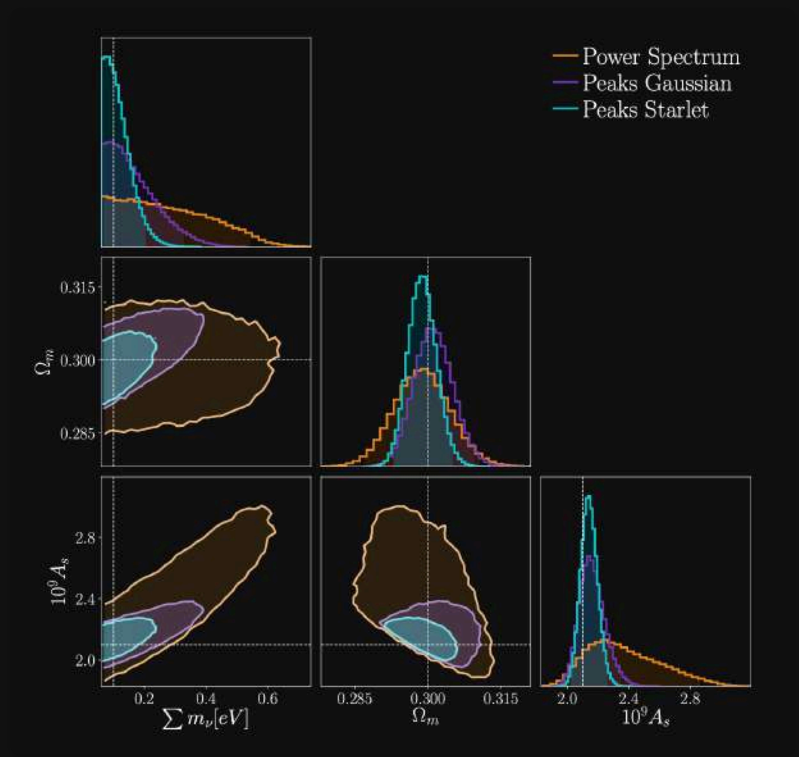
Data



$$\log \mathcal{L}(\theta) = -\frac{1}{2}\left[d - \mu(\theta)\right]^T C^{-1}\left[d - \mu(\theta)\right]$$

MCMC

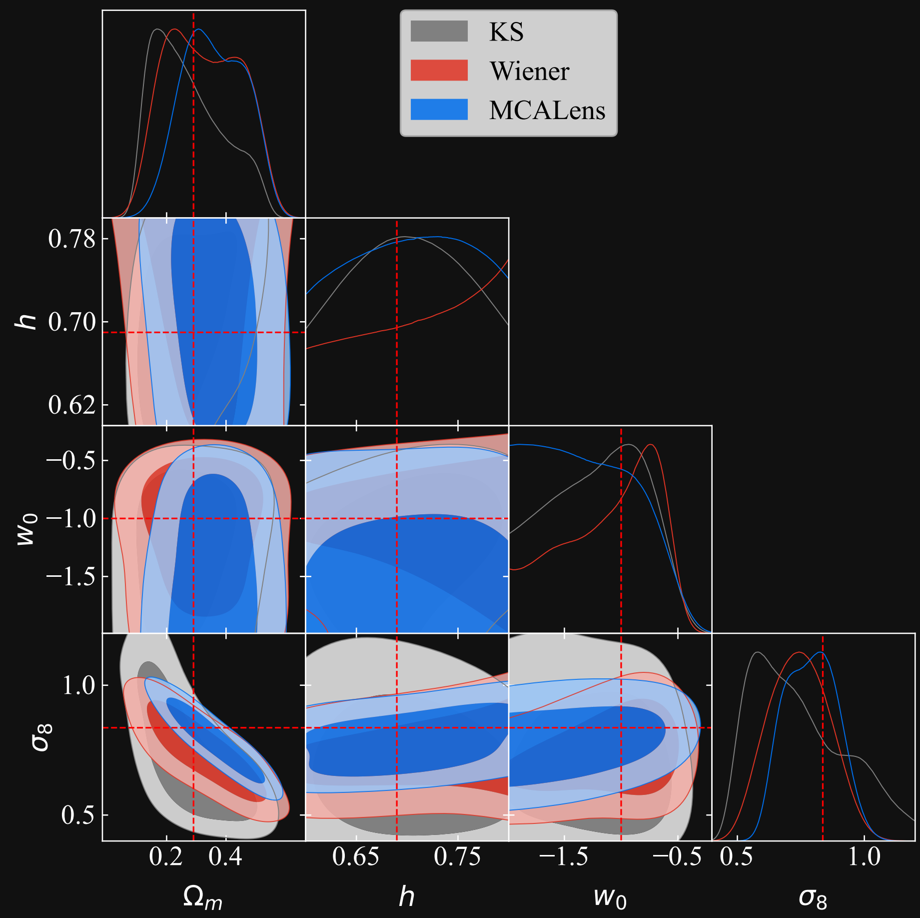
Constraints on parameters



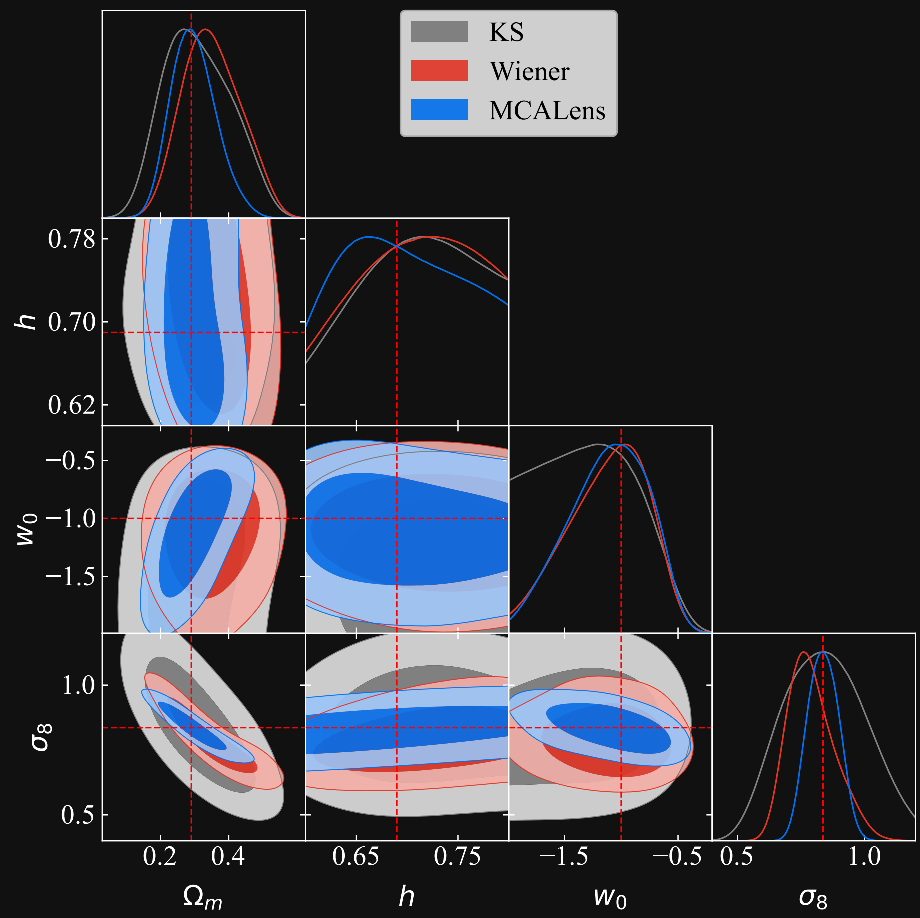
- μ: expected theoretical prediction, d: data array (mean over realizations of a HOS), C: covariance matrix

So does the choice of the mass mapping algorithm matter for the final constraints?

The (standard) mono-scale peak counts



Wavelet multi-scale peak counts



New mass mapping method

New mass mapping method

There is still no optimal mass mapping method for the analysis of large surveys. We need a method that is:

- **Accurate** (small MSE and error bars)
- **Fast** and **efficient** → **point estimates** instead of sampling from posterior
- **Uncertainty bounds**
- Does not need **retraining** for each new **mask** or **noise level**
- **Flexible** and **adaptable** to different cosmologies

New mass mapping method

There is still no optimal mass mapping method for the analysis of large surveys. We need a method that is:

- **Accurate** (small MSE and error bars)
- **Fast** and **efficient** → **point estimates** instead of sampling from posterior
- **Uncertainty bounds**
- Does not need **retraining** for each new **mask** or **noise level**
- **Flexible** and **adaptable** to different cosmologies

Our proposal

- **Create new method, based on Plug-and-Play algorithms**
- Framework for solving image recovery problems by combining **physical models** and **learned models**
- Use regularized optimization techniques + (deep) image denoiser, which is used to impose a prior on the solution.

Conclusions

Conclusions

- Mass mapping is a **challenging** problem in weak lensing
- Several methods have been developed, each with its **advantages** and **limitations**
- HOS provide **complementary** information to the standard 2pt statistics, and can help **extract more information** from the data & **break degeneracies**

Conclusions

- Mass mapping is a **challenging** problem in weak lensing
- Several methods have been developed, each with its **advantages** and **limitations**
- HOS provide **complementary** information to the standard 2pt statistics, and can help **extract more information** from the data & **break degeneracies**

Results

- Created pipeline for HOS analysis of weak lensing catalogs
- Implemented it to show that the **choice of the mass mapping algorithm has a significant impact on the cosmological constraints**

Conclusions

- Mass mapping is a **challenging** problem in weak lensing
- Several methods have been developed, each with its **advantages** and **limitations**
- HOS provide **complementary** information to the standard 2pt statistics, and can help **extract more information** from the data & **break degeneracies**

Results

- Created pipeline for HOS analysis of weak lensing catalogs
- Implemented it to show that the **choice of the mass mapping algorithm has a significant impact on the cosmological constraints**

Future work

- Add **DL methods** to the pipeline
- Use the pipeline for a HOS analysis of **UNIONS data**
- Develop **PnP mass mapping method**