Title: Precision Cosmology from the Clustering of Galaxies

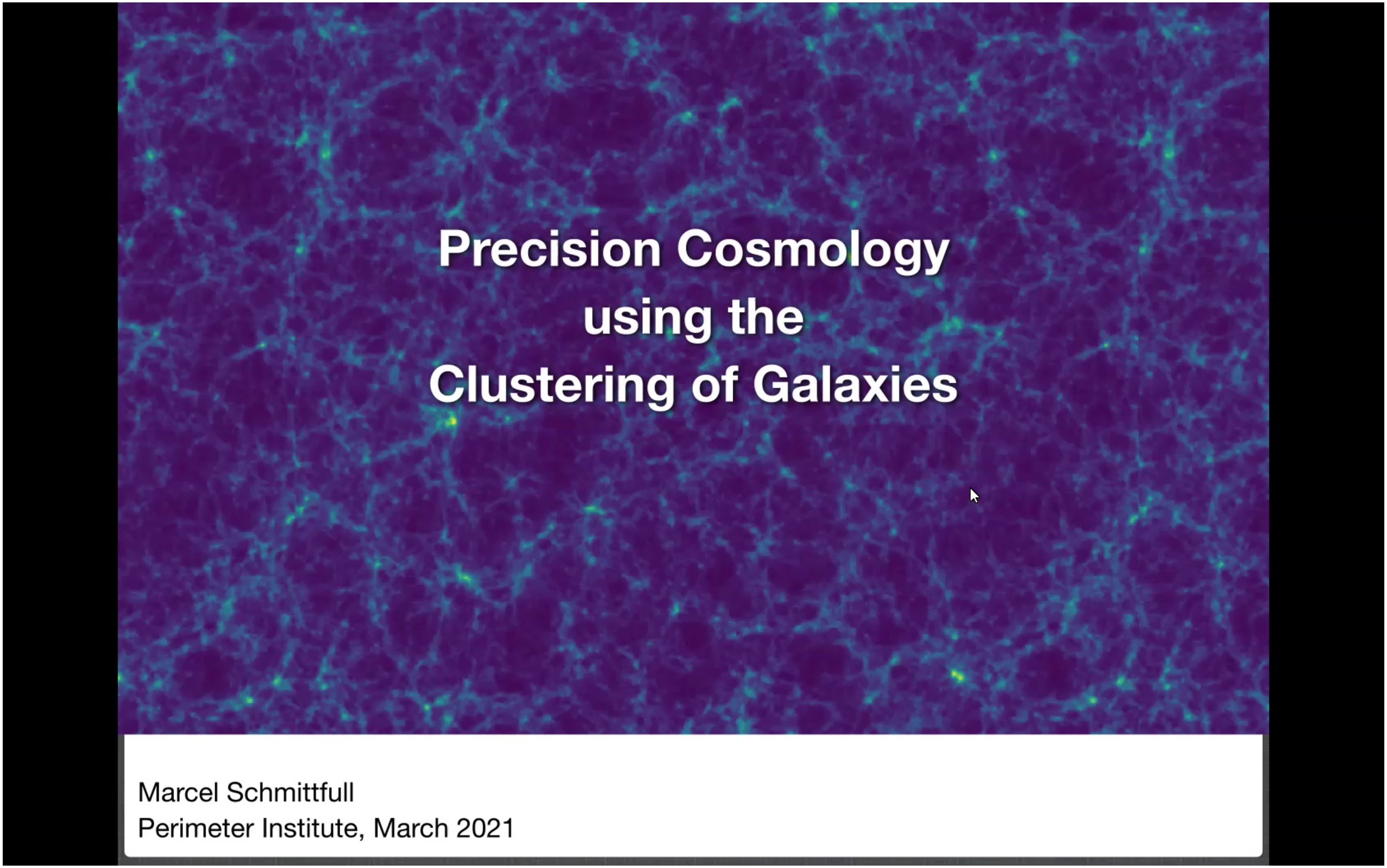
Speakers: Marcel Schmittfull

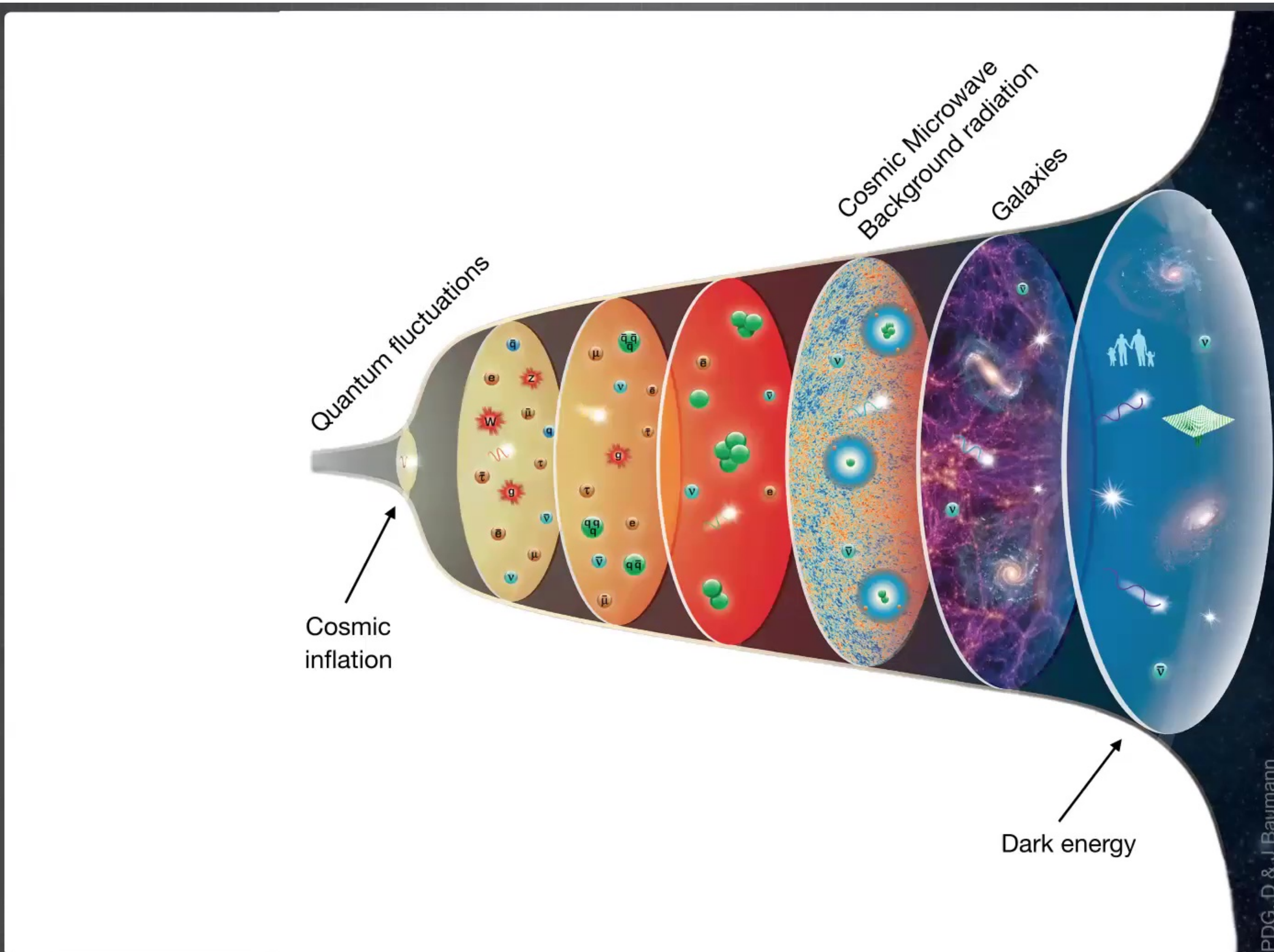
Series: Cosmology & Gravitation

Date: March 23, 2021 - 11:00 AM

URL: http://pirsa.org/21030038

Abstract: Large surveys of the positions of galaxies in the Universe are becoming increasingly powerful to shed light on some of the unsolved problems of cosmology, including the question of what caused the early Universe to expand. The analysis of the data is challenging, however, because the signal is small, the data is difficult to model, and its probability distribution is not fully known. I will present some recent ideas to approach these challenges.


Precision Cosmology
using the
Clustering of Galaxies
Marcel Schmittfull
Perimeter Institute, March 2021


Quantum fluctuations
Cosmic Microwave
Background radiation
Galaxies
Cosmic
inflation
Dark energy
PDG, D & J Baumann

Nature of each building block is unknown

Cosmic inflation

What particle physics model led to the rapid expansion?

How did our Universe begin?

Dark energy

What causes this 2nd epoch of rapid expansion?

Is it a cosmological constant? Is General Relativity broken?

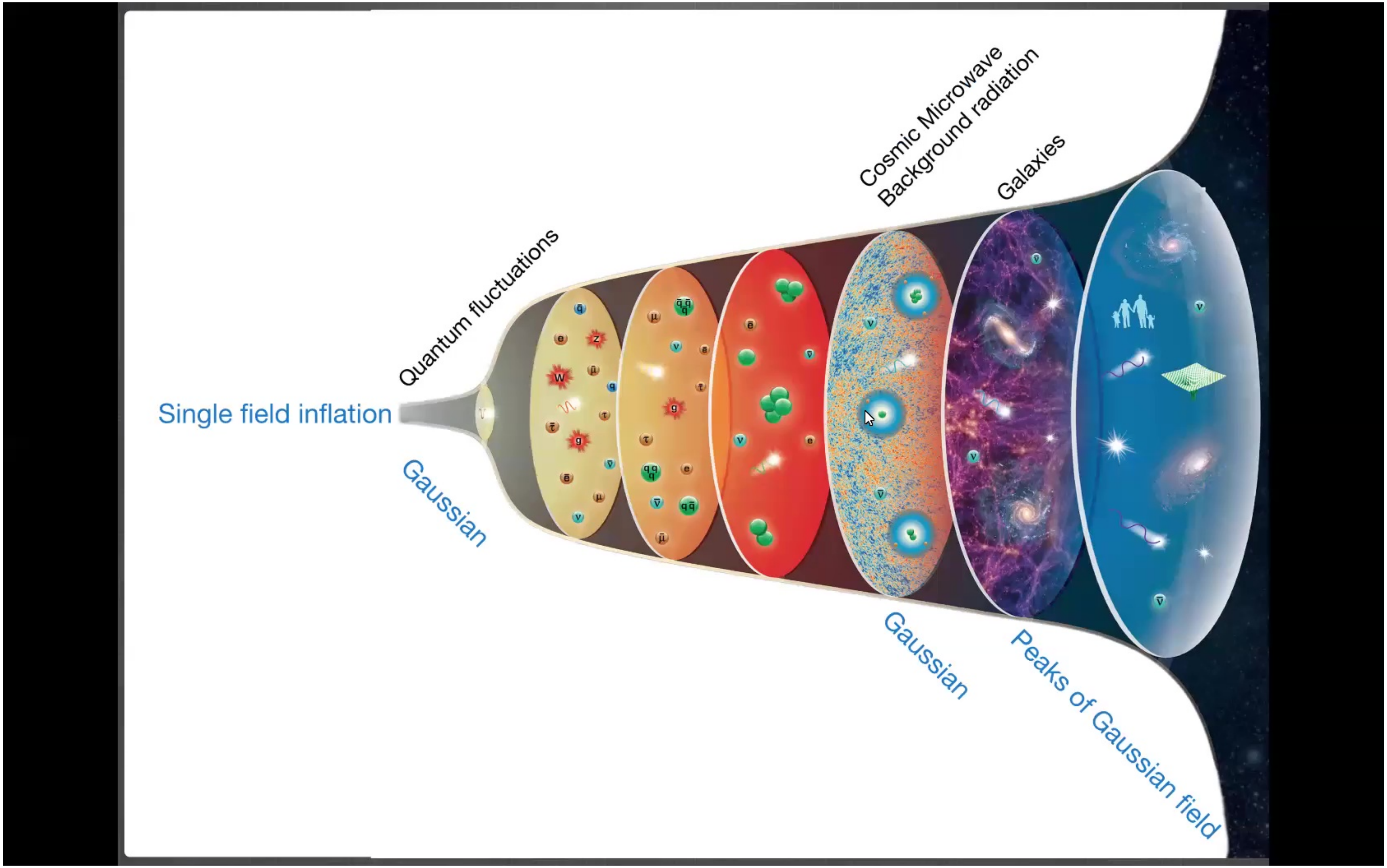
Dark matter

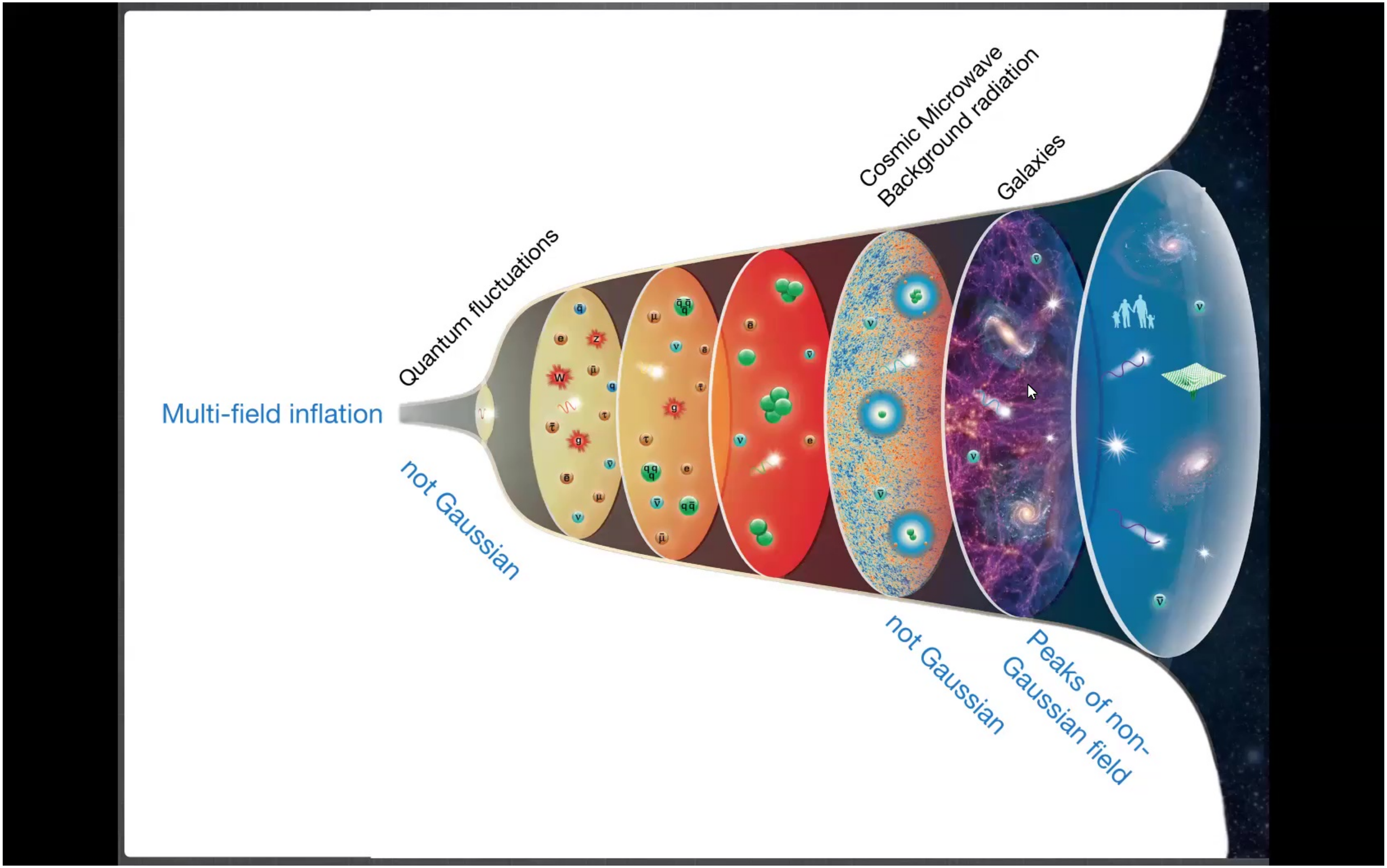
What particle(s) is it made of?

Relativistic particles

What is the mass (hierarchy) of neutrinos?

Are there additional light particles?


Single field inflation
Gaussian
Quantum fluctuations
Cosmic Microwave
Background radiation
Galaxies
Gaussian
Peaks of Gaussian field

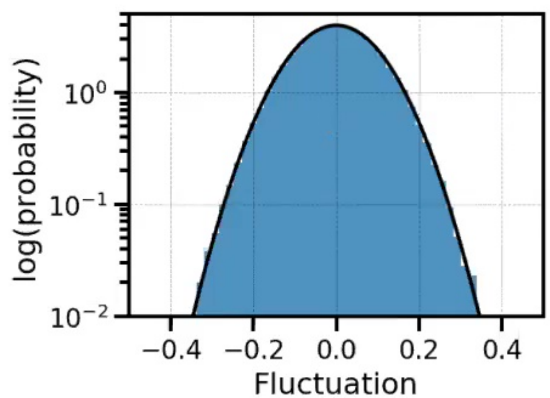

Multi-field inflation
Quantum fluctuations
not Gaussian
Cosmic Microwave Background radiation
Galaxies
not Gaussian
Peaks of non-Gaussian field

Non-Gaussian fluctuations from inflation

Cosmic inflation

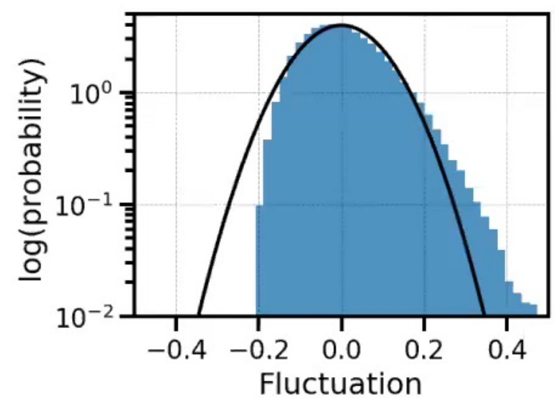
Single field

Gaussian fluctuations
Skewness $f_{\mathrm{NL}} \ll 1$

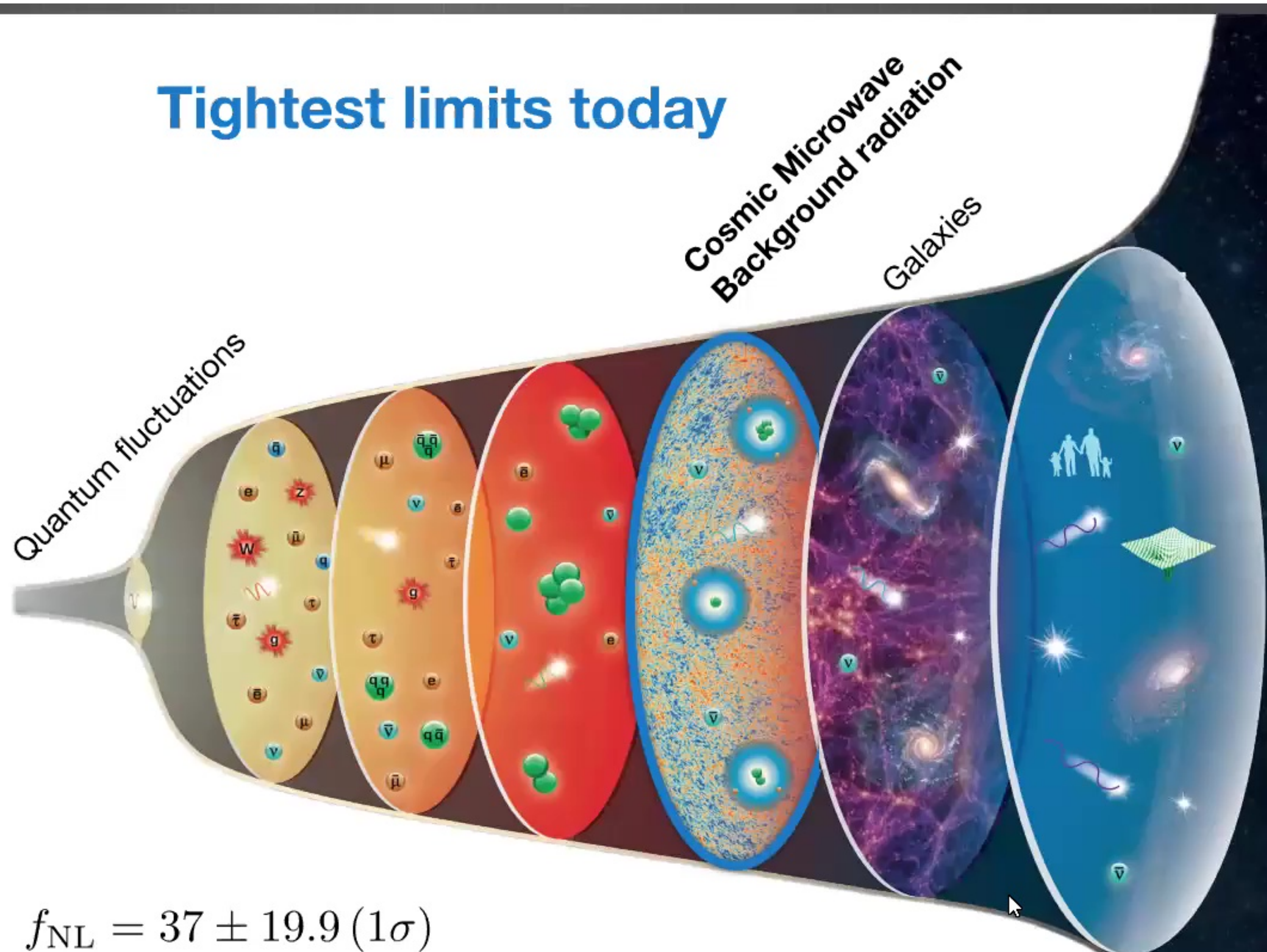


Multi-field

Non-Gaussian fluctuations
Skewness $f_{\mathrm{NL}} \gtrsim 1$



Tightest limits today



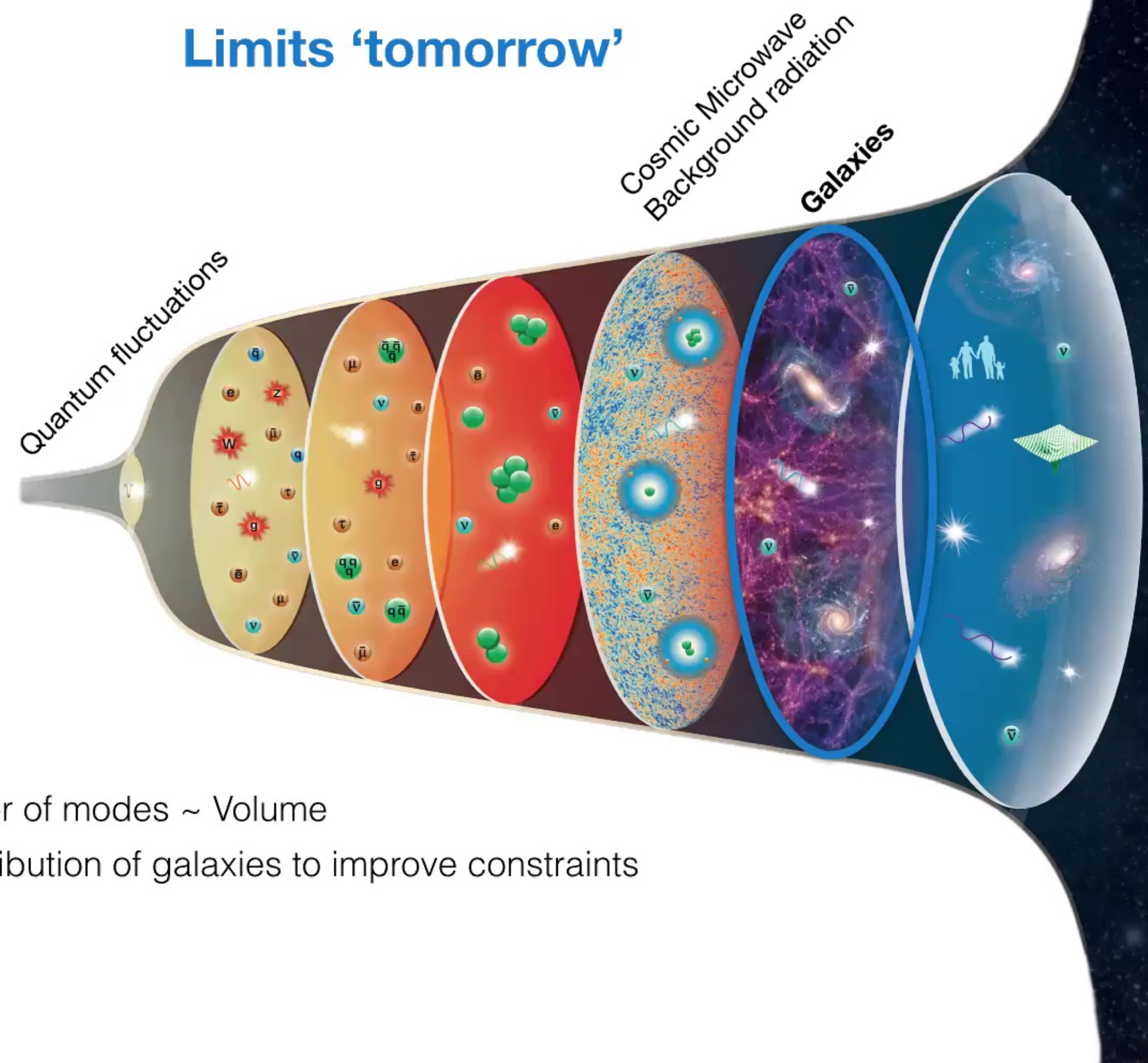
WMAP satellite: $f_{\rm NL} = 37 \pm 19.9\,(1\sigma)$

Planck satellite: $f_{\rm NL} = -0.9 \pm 5.1\,(1\sigma)$

Both consistent with zero (2σ)

WMAP Collaboration, Bennett et al. (2013)
Planck Collaboration, Akrami et al. (1905.05697)

Limits 'tomorrow'



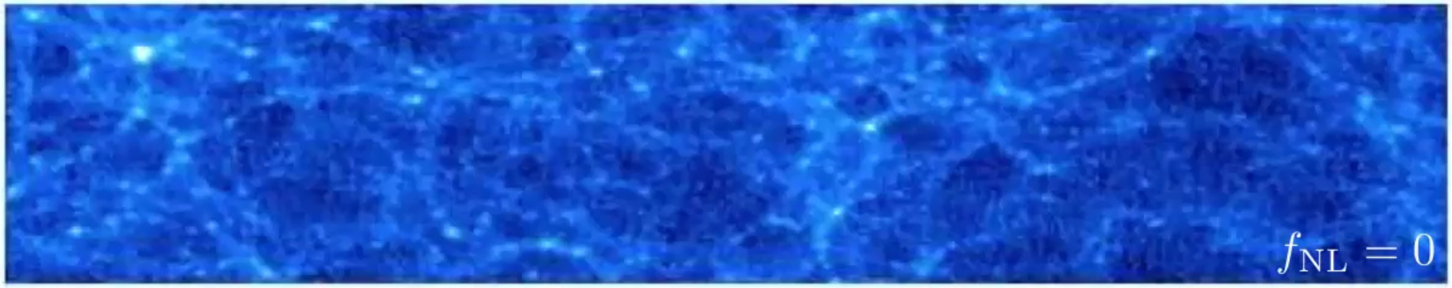
SNR² ~ Number of modes ~ Volume

⇒ Use 3D distribution of galaxies to improve constraints

The distribution of galaxies

Single-field inflation

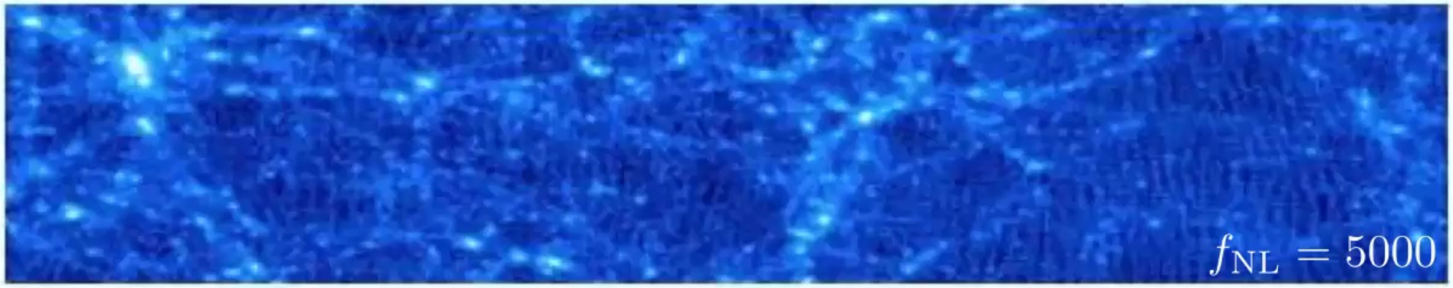
Galaxies evolve from normally distributed initial conditions



Multi-field inflation

Galaxies evolve from non-Gaussian initial conditions with enhanced peaks

We are looking for a 5000x smaller signal



Dalal et al. (2007)

How to measure this signal?

Count #peaks

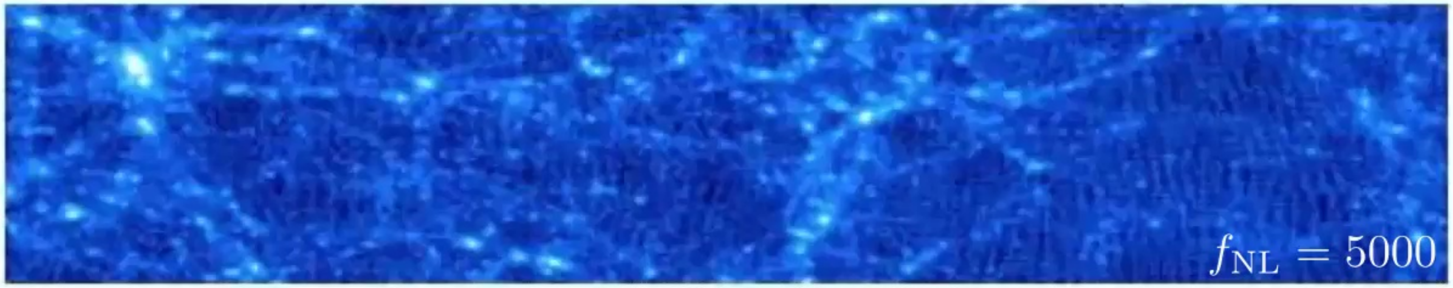
Count #voids

Measure histogram of the galaxy number density

Measure skewness, kurtosis, etc

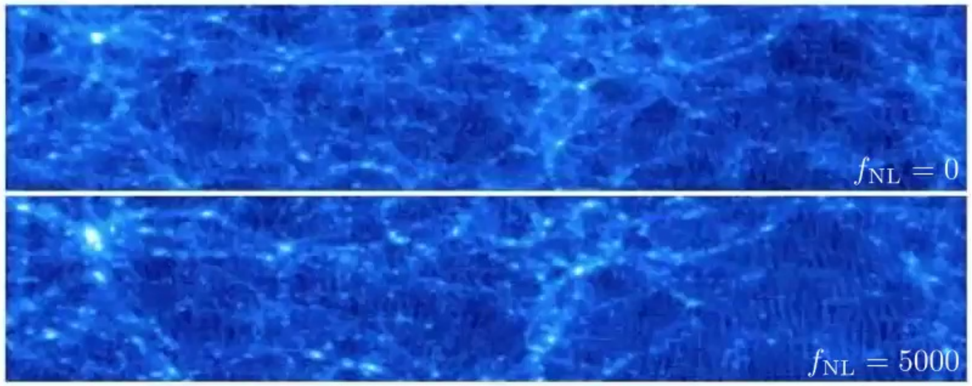
Measure skewness of all 3D Fourier modes

+ Many more ideas proposed in the literature



Some challenges

Signal is tiny



Data is complicated, nonlinear function of initial conditions

⇒ *Not easy to model the data*

Data is not normally distributed

⇒ *What is the optimal data analysis method?*

Other questions suffer from similar challenges

Cosmic inflation
What particle physics model led to the rapid expansion?
How did our Universe begin?

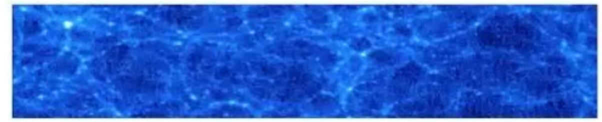
Dark energy
What causes this 2nd epoch of rapid expansion?
Is it a cosmological constant? Is General Relativity broken?

Dark matter
What particle(s) is it made of?

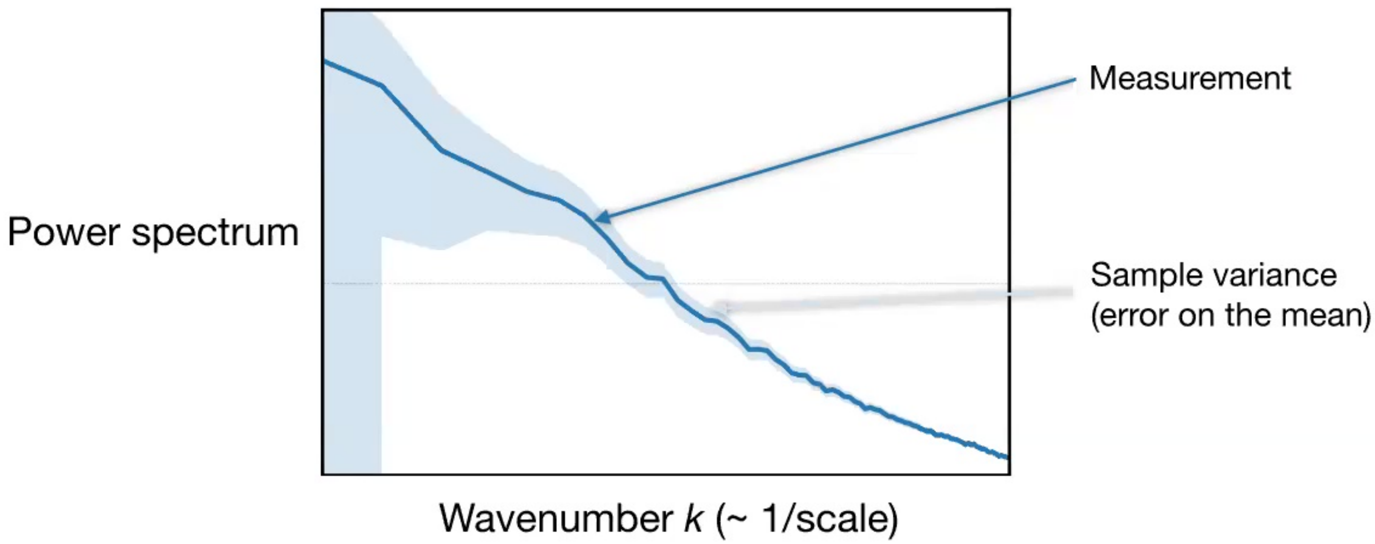
Relativistic particles
What is the mass (hierarchy) of neutrinos?
Are there additional light particles?

Power spectrum as a summary statistic

Power spectrum $P(k) = \frac{1}{N(k)} \sum_{\mathbf{k}, |\mathbf{k}|=k} |\delta(\mathbf{k})|^2$



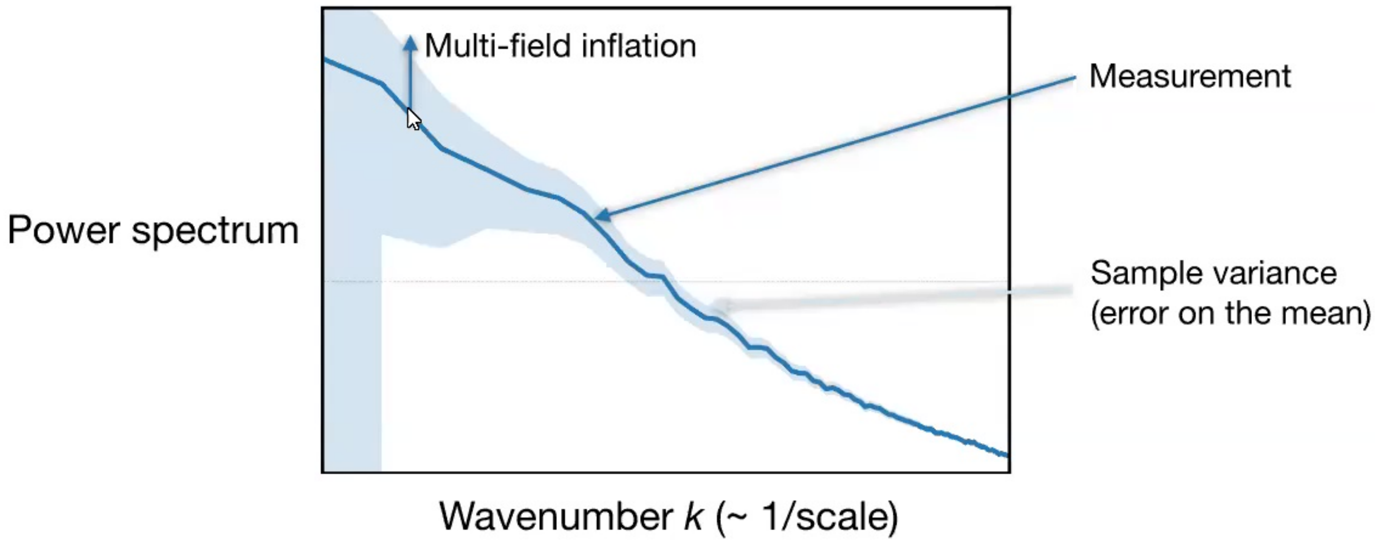
Squared size of fluctuations



Power spectrum as a summary statistic

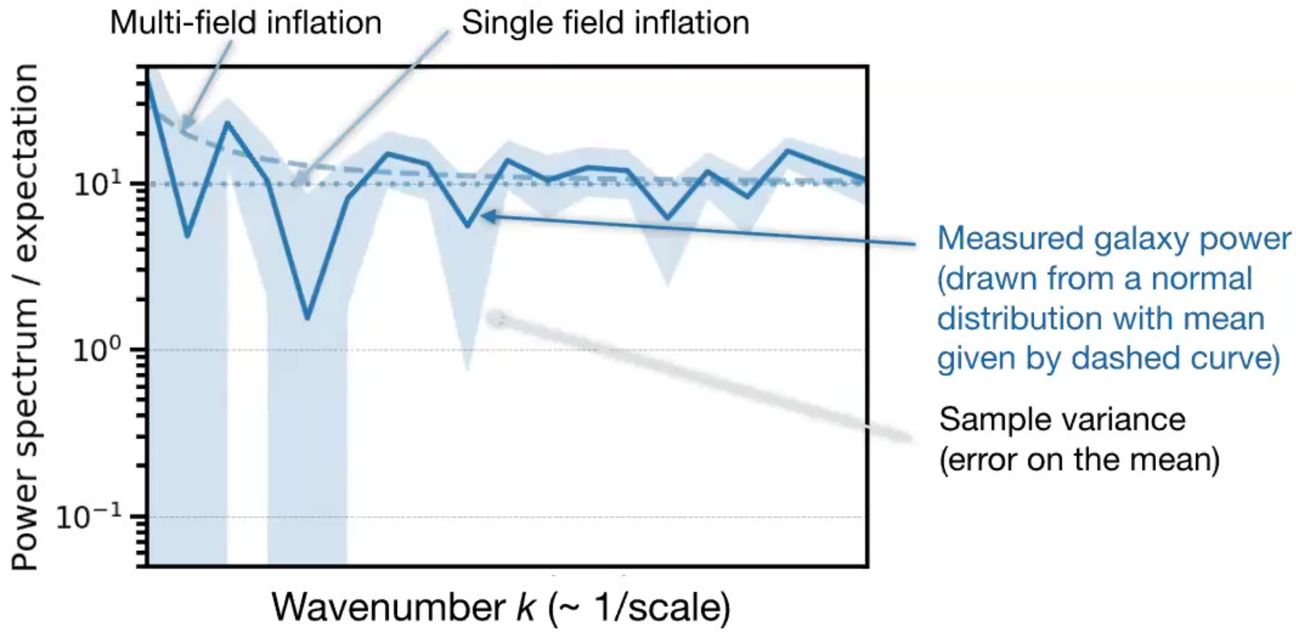
Power spectrum $P(k) = \frac{1}{N(k)} \sum_{\mathbf{k}, |\mathbf{k}|=k} |\delta(\mathbf{k})|^2$

Squared size of fluctuations

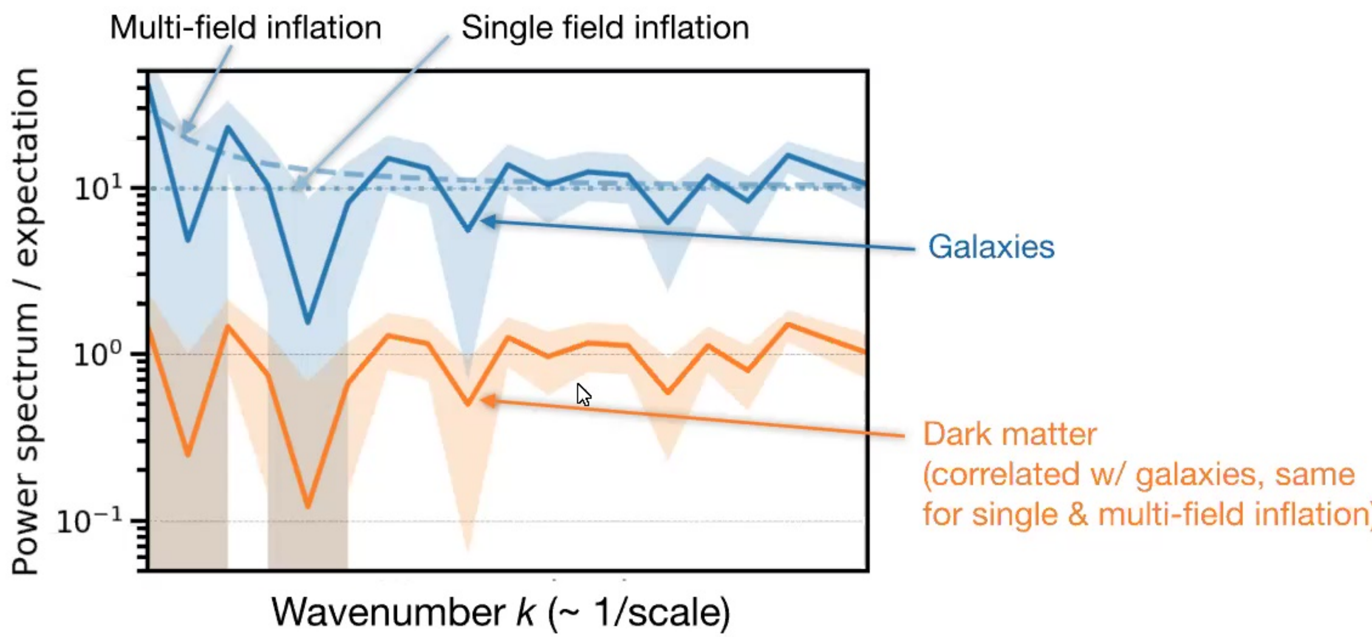


Multi-field inflation couples peaks to grav. potential, enhancing power at low k

Divide by expectation for single-field inflation

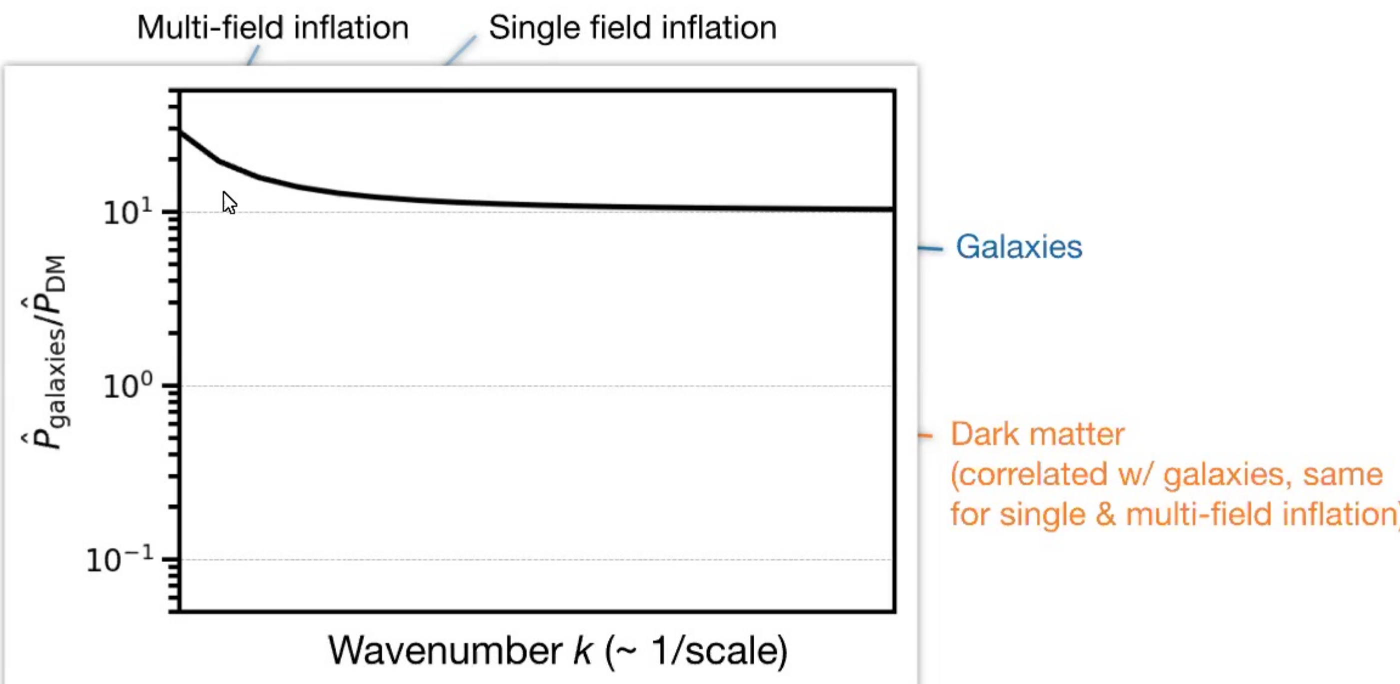


Make a correlated measurement, insensitive to signal



Seljak (2009), McDonald & Seljak (2009), MS & Seljak (2018)

Make another, correlated measurement



Sample variance cancels in the ratio, so can detect multi-field inflation

Seljak (2009), McDonald & Seljak (2009), MS & Seljak (2018)

How did this work?

Imagine you come up with a new image compression algorithm

Is it better than JPEG?

Method 1

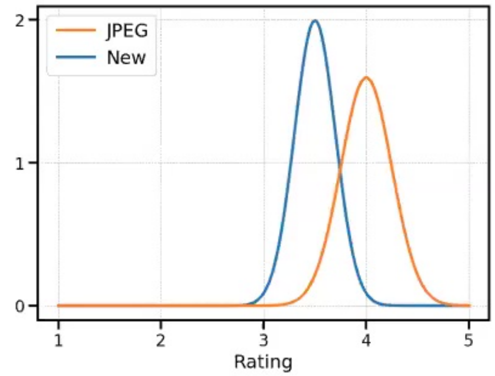
a. Ask people to rate JPEG-compressed images

...

b. Also ask to rate *other* images compressed with new algorithm

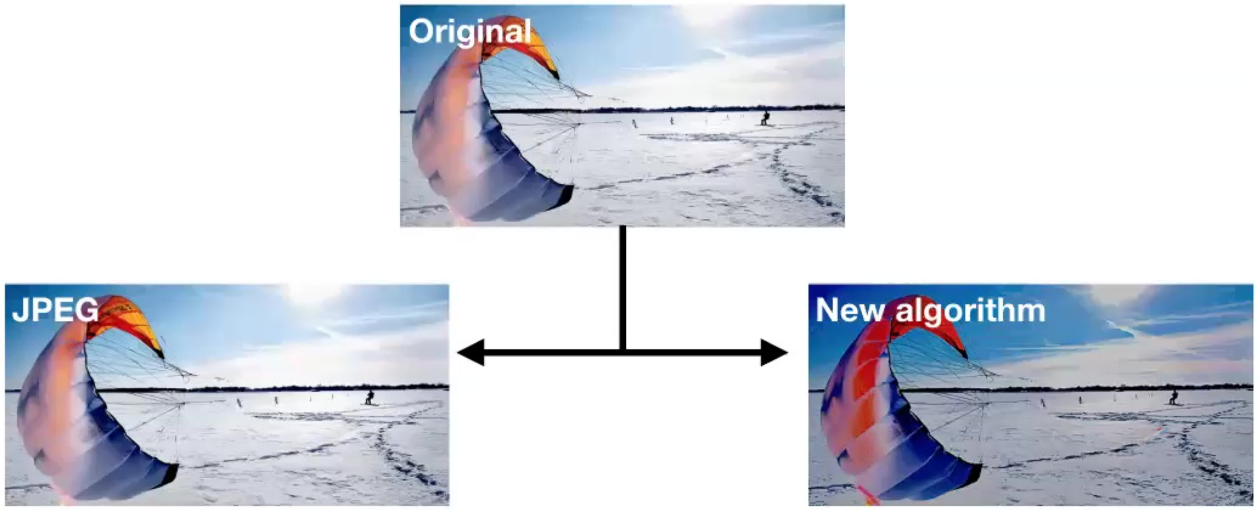
...

c. Compare ratings to find winner



Subject to sample variance (error of the mean)

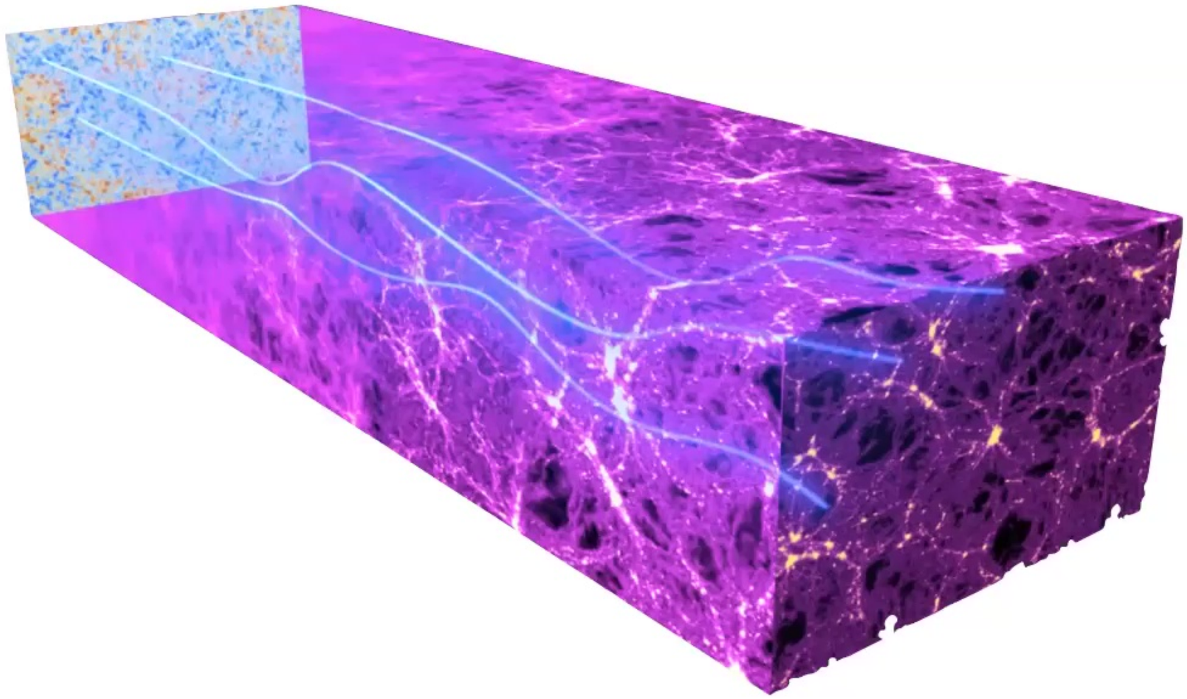
Method 2



a. Ask people to rate same image compressed with JPEG & new algorithm
b. Compare ratings 1-by-1 for each image

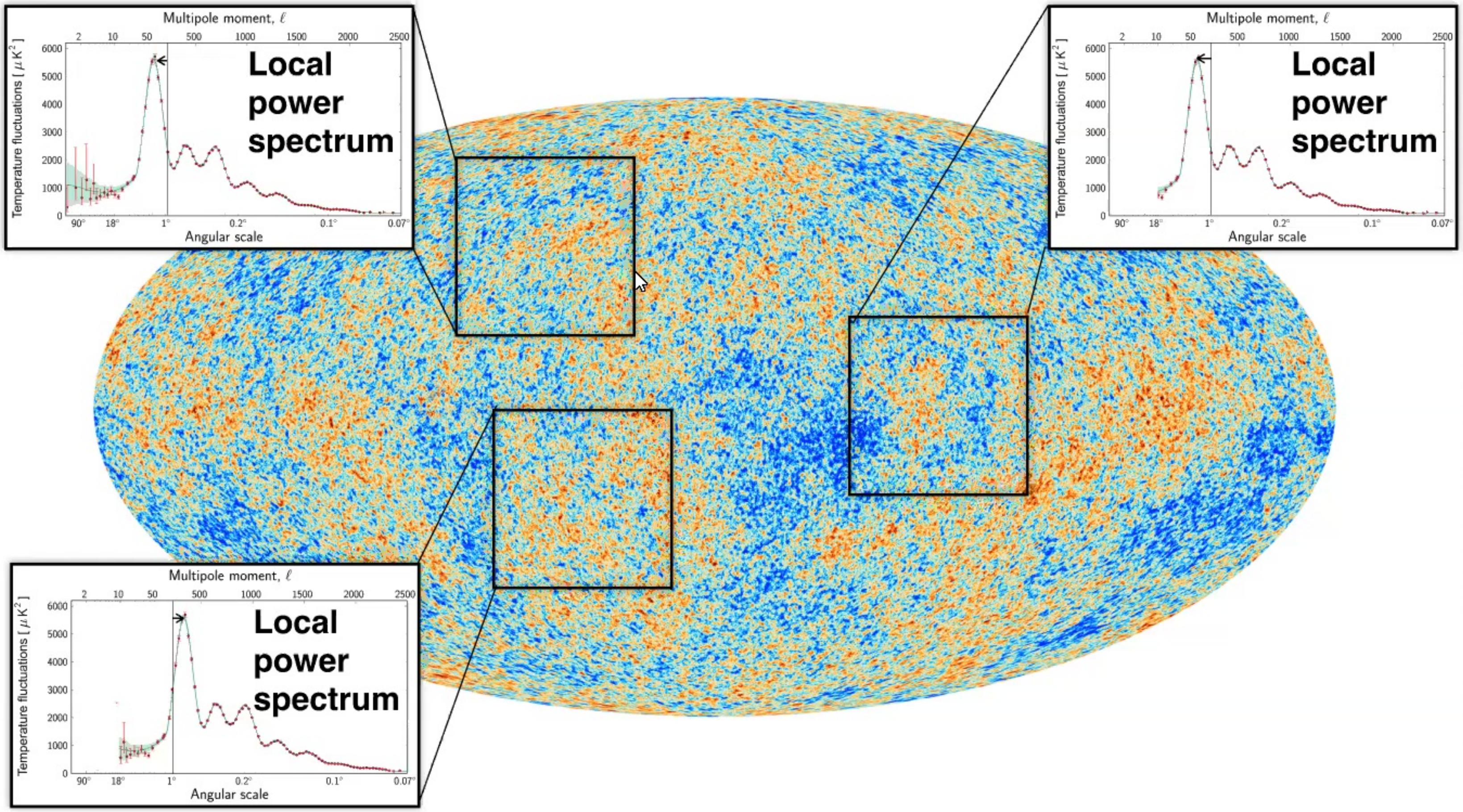
Less sample variance (can tell winner with 1 image)

How to measure the distribution of dark matter?



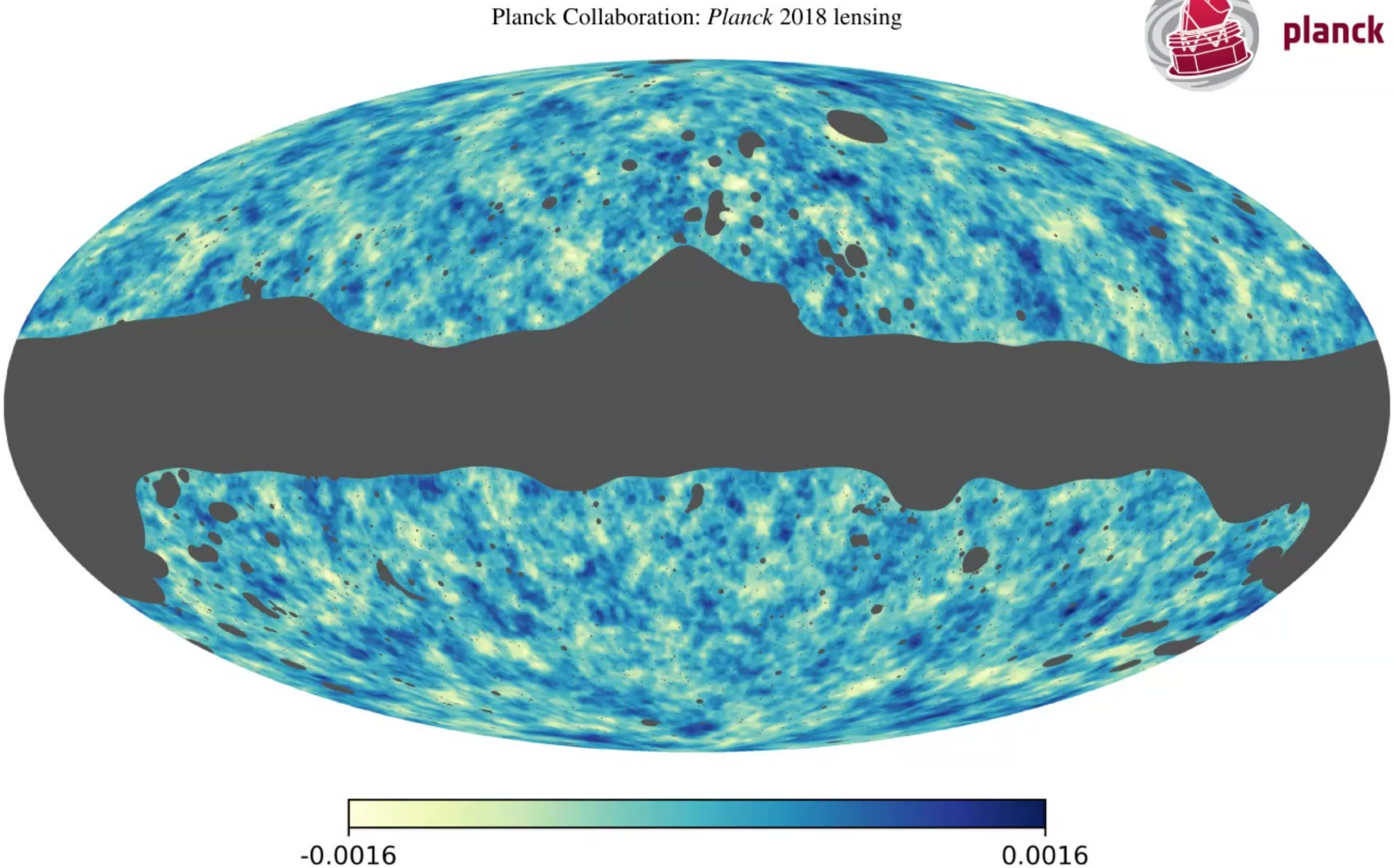
Measure gravitational lensing of Microwave Background radiation

Lensed Cosmic Microwave Background radiation



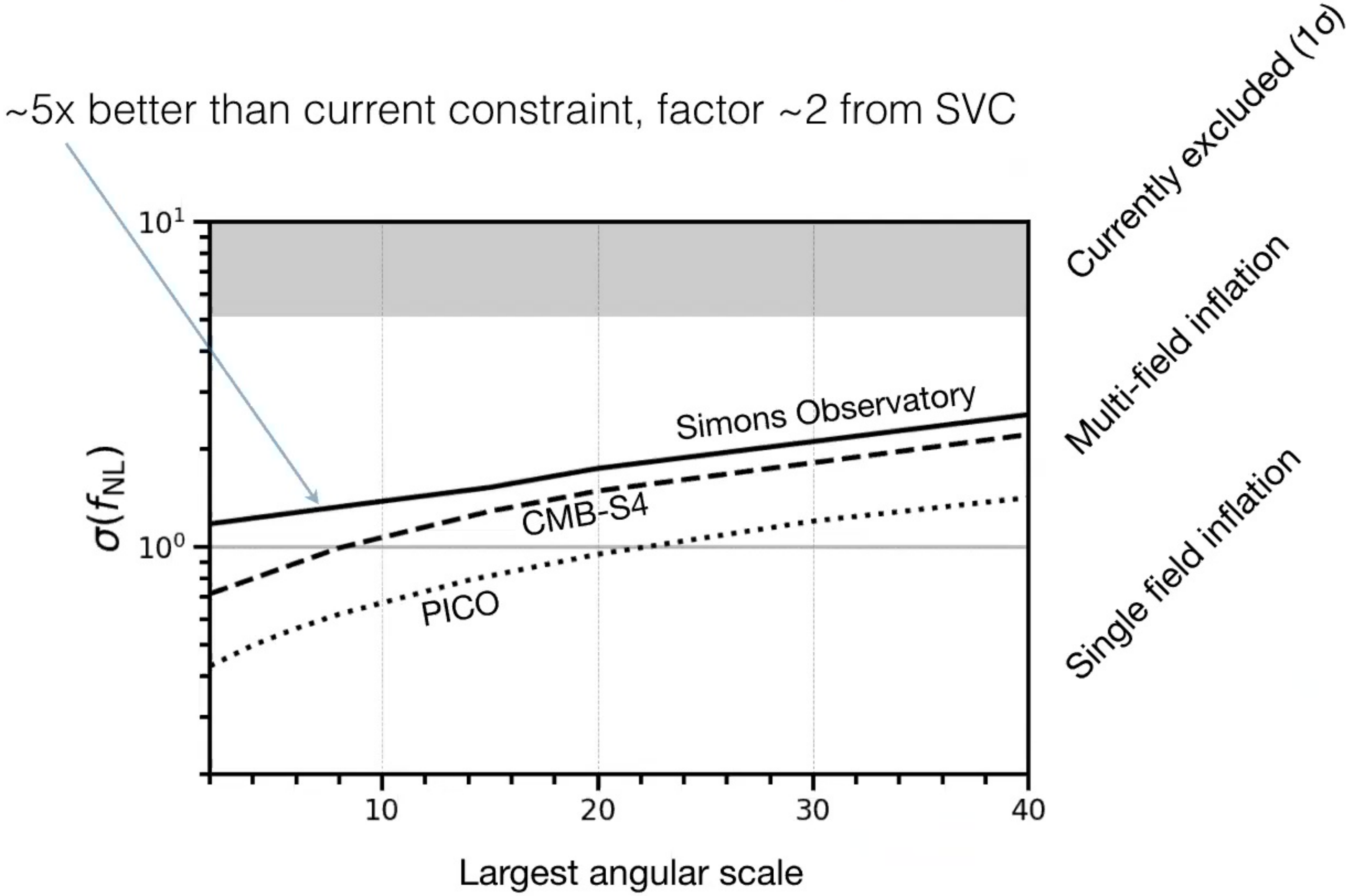
Estimate local magnification / demagnification

Measured magnification map



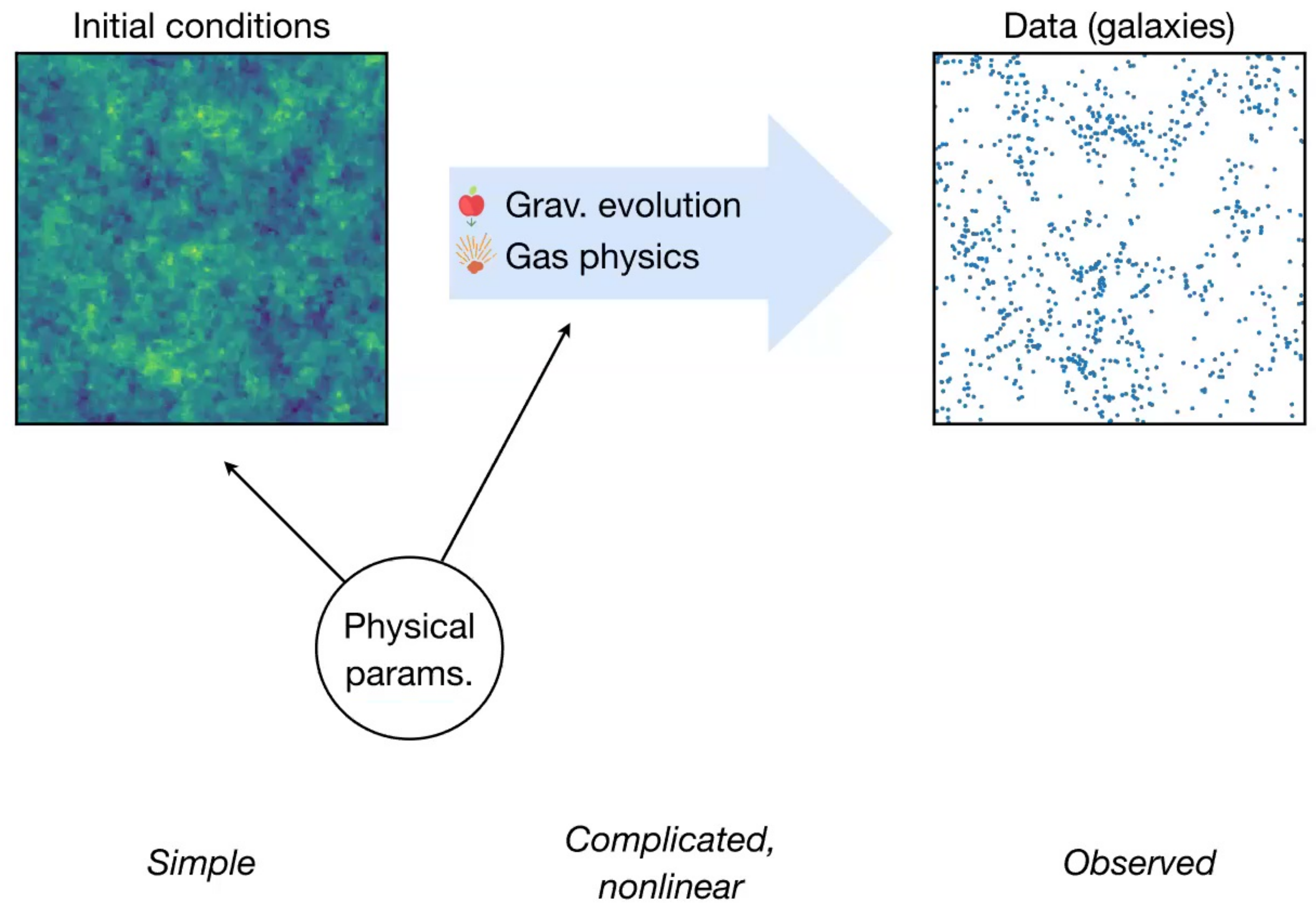
Forecasts for future experiments

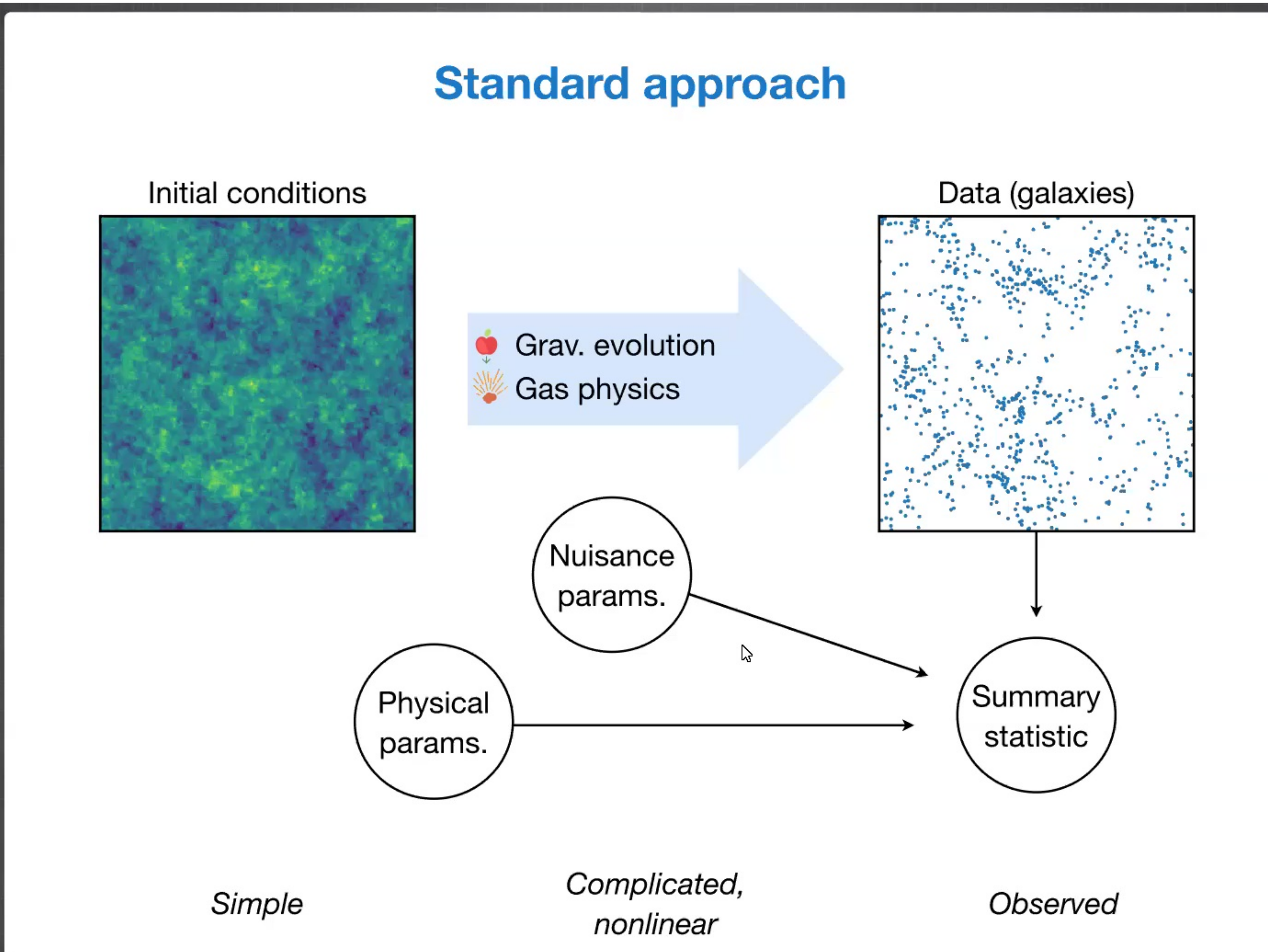
~5x better than current constraint, factor ~2 from SVC



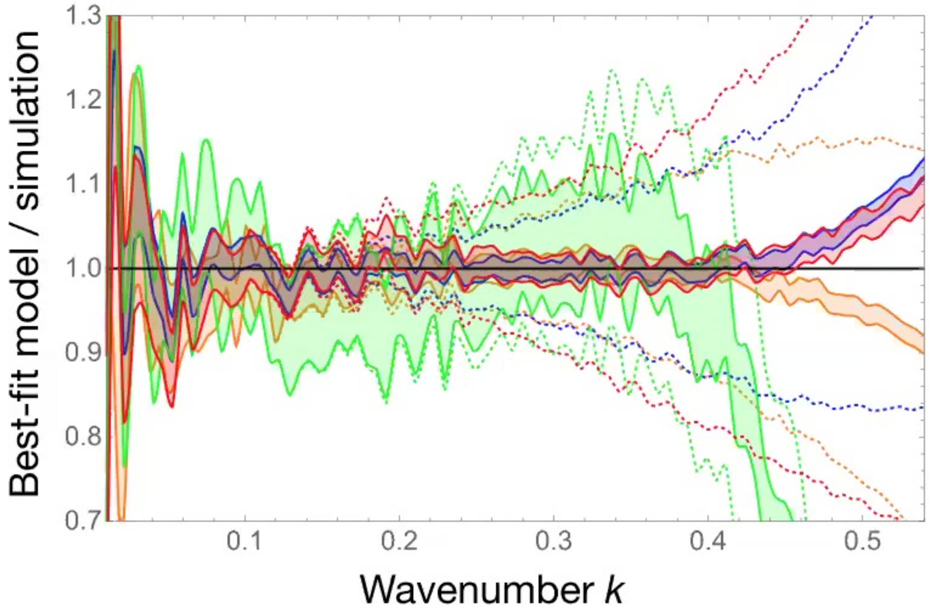
MS & Seljak (2018); arXiv: 1808.07445, 1907.08284; 1907.04473, 1908.01062; 1902.10541, 1908.07495

Challenge


Initial conditions
Data (galaxies)
Grav. evolution
Gas physics
Physical params.
Simple
Complicated, nonlinear
Observed


Standard approach
Initial conditions
Data (galaxies)
Grav. evolution
Gas physics
Nuisance params.
Physical params.
Summary statistic
Simple
Complicated, nonlinear
Observed

Power spectrum as a summary statistic



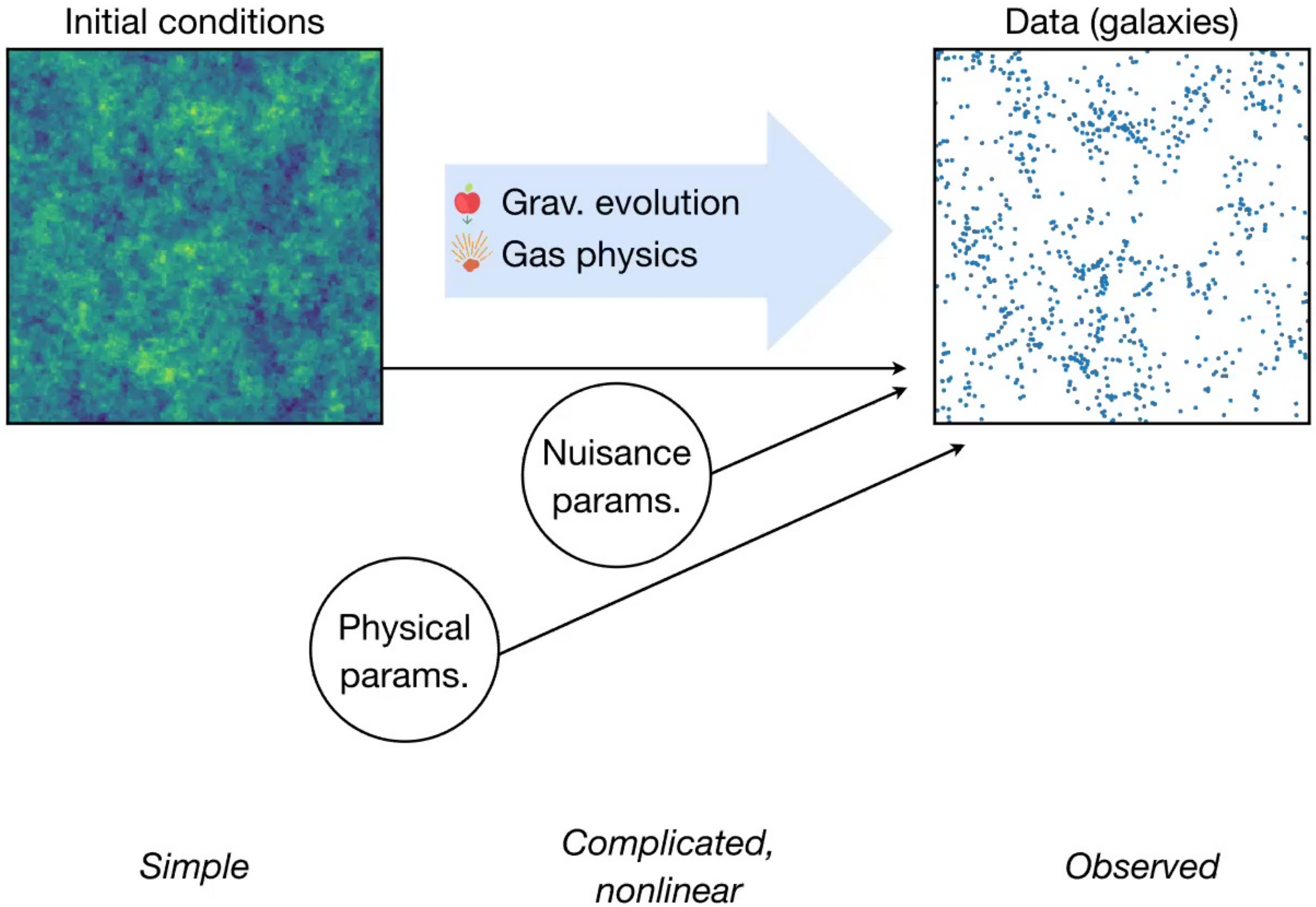
Fitted parameters:

$$
\begin{aligned}
b_1 &= 0.98 \pm 0.01 \\
b_2 &= 0.01 \pm 2.73 \\
b_3 &= -0.62 \pm 1.43 \\
b_4 &= 0.58 \pm 2.33 \\
c_{\rm ct}^{(\delta_h)} &= (5.3 \pm 4.7)\left(\frac{k_{\rm NL}}{h\ {\rm Mpc}^{-1}}\right)^2 \\
\tilde{c}_{r,1} &= (-14 \pm 5)\left(\frac{k_{\rm M}}{h\ {\rm Mpc}^{-1}}\right)^2 \\
\tilde{c}_{r,2} &= (-0.69 \pm 1.67)\left(\frac{k_{\rm M}}{h\ {\rm Mpc}^{-1}}\right)^2 \\
c_{\epsilon,1} &= (0.76 \pm 14.74) \\
c_{\epsilon,2} &= (8.9 \pm 3.4)\left(\frac{k_{\rm M}}{h\ {\rm Mpc}^{-1}}\right)^2 \\
c_{\epsilon,3} &= (8.0 \pm 7.8)\left(\frac{k_{\rm M}}{h\ {\rm Mpc}^{-1}}\right)^2 .
\end{aligned}
$$

Beware of overfitting: 50 data points, 10 free parameters

Perko, Senatore, Jennings & Wechsler (arXiv:1610.09321)

Alternative: Predict 3D data given initial conditions



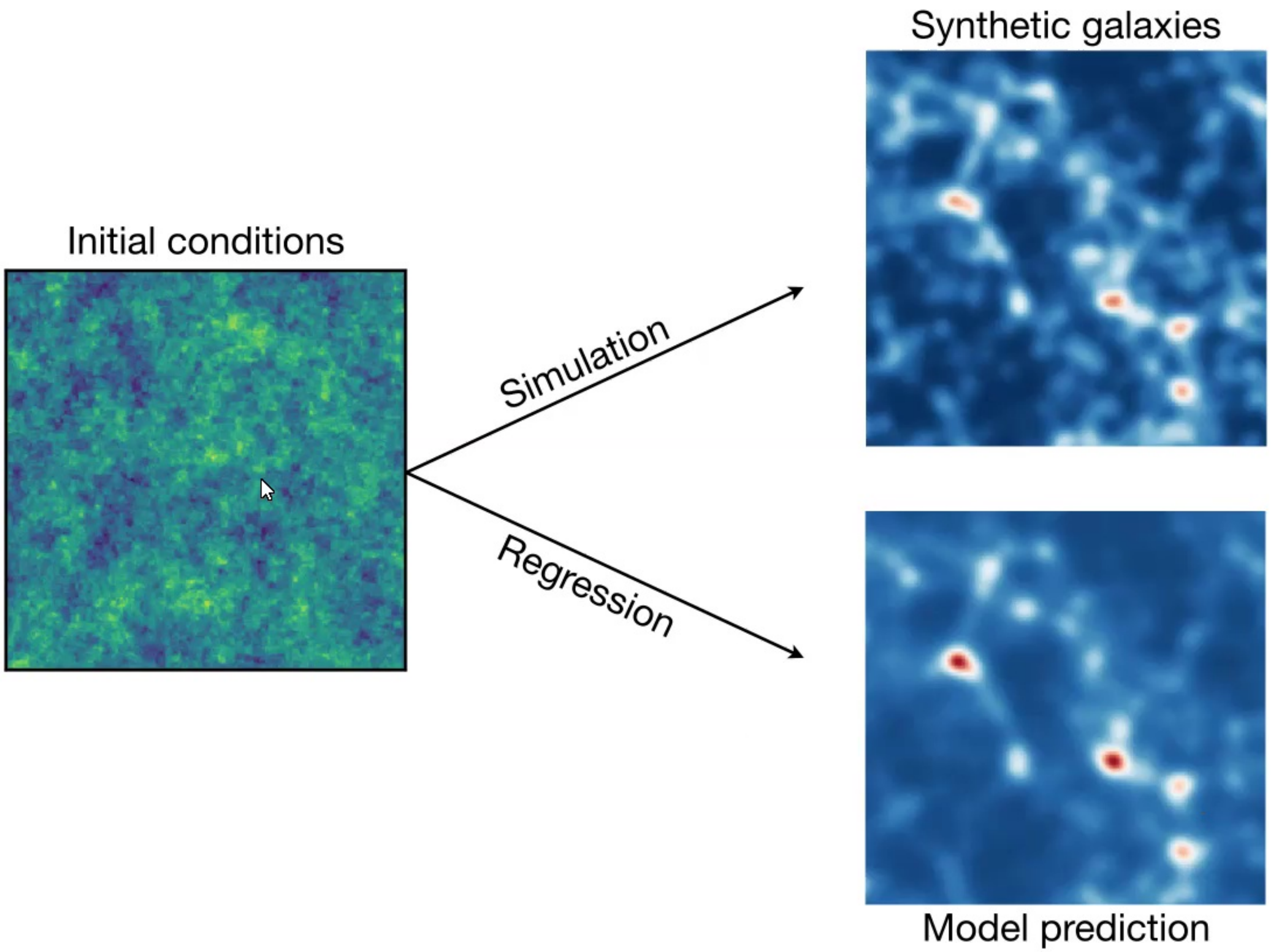
Benefits

Benefits of using 3D fields rather than summary statistics

+ No overfitting (6 parameters describe >1 million galaxy positions)
+ No sample variance, can use small volumes with high resolution
+ 'All' *n*-point functions measured simultaneously
+ Easy to isolate mistakes of the model
+ Useful for field-level likelihood and initial condition reconstruction

MS, Simonović, Assassi & Zaldarriaga (2019)

Setup

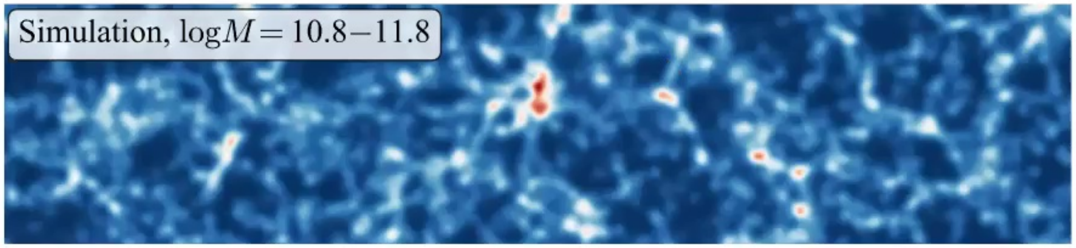


MS, Simonović, Assassi & Zaldarriaga (2019)

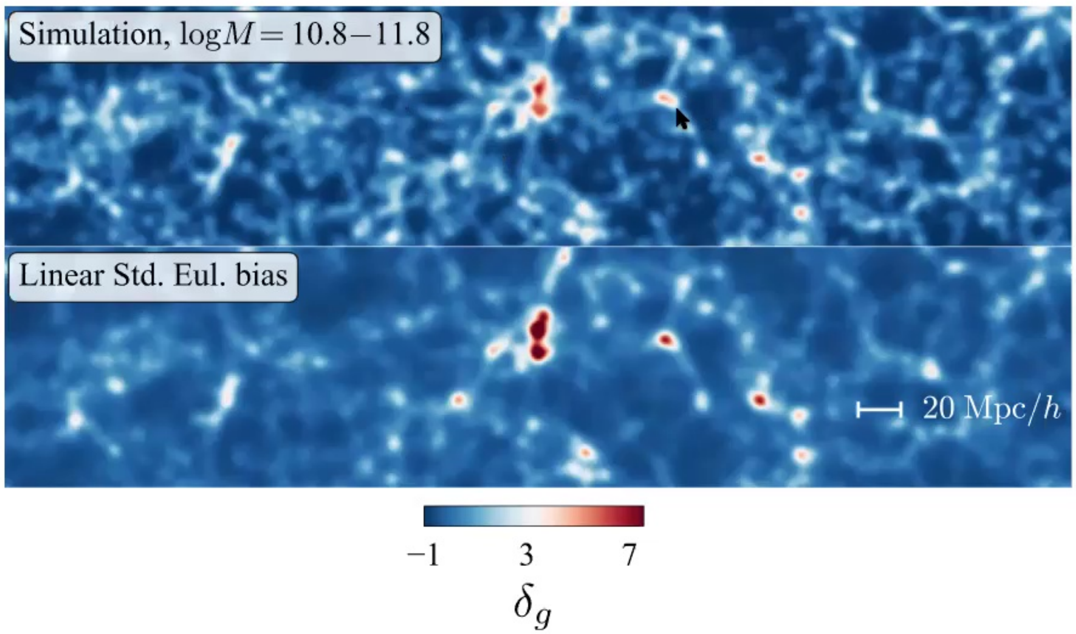
Simulation

1536^3 = 3.6B particles in a 3D cubic box
3072^3 = 29B grid points for long-range force computation
4000 time steps
5 realizations

1M CPU hours on a local cluster (~2000 CPUs) using MP-Gadget N-body code



Comparison with linear regression model



Simulation

Linear model
$b_1 \delta_m(\mathbf{x})$

Reasonable prediction on large scales

Missing structure on small scales

MS, Simonović, Assassi & Zaldarriaga (2019)

How to improve?

Include all terms allowed by symmetries (effective field theory)

$$\delta_g(\mathbf{x}) = b_1\delta_m(\mathbf{x}) + b_2\delta_m^2(\mathbf{x}) + \text{tidal term} + b_3\delta_m^3(\mathbf{x}) + \cdots$$

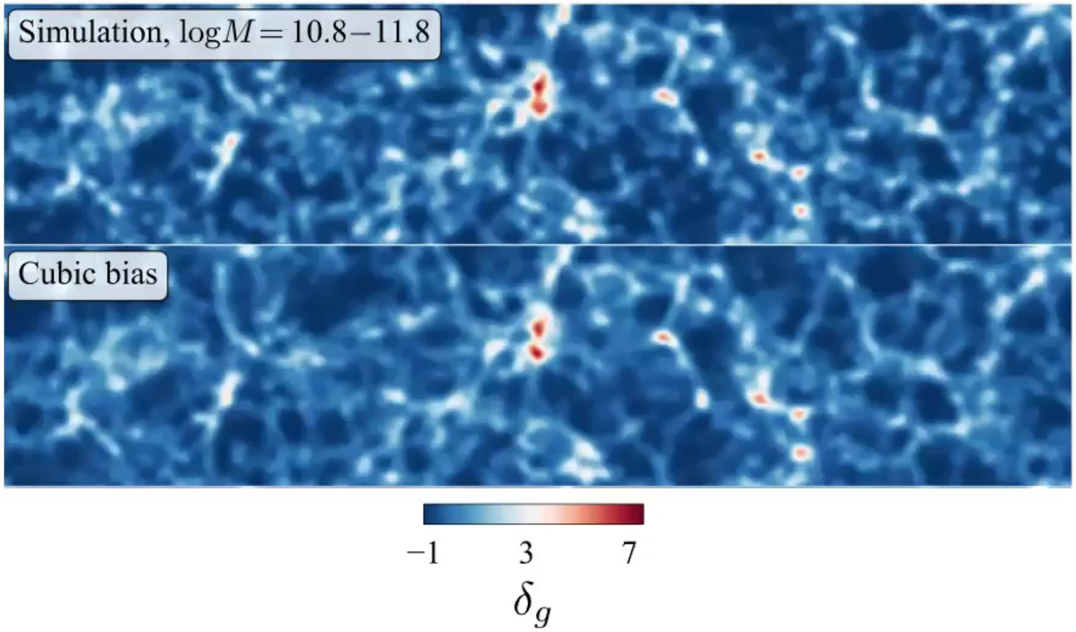
Desjacques, Jeong & Schmidt: Review of Large-Scale Galaxy Bias (2018)

Fit coefficients b_i using least-squares regression

MS, Simonović, Assassi, Zaldarriaga (2019)

In practice:
- Run independent regression for each Fourier mode shell
- Fit resulting regression coefficients $b_i(k)$ with 6-parameter model
- Orthogonalize operators for robust numerics and interpretation
- Include large bulk flows nonperturbatively (see later)

Comparison with nonlinear model



Simulation

Nonlinear regression model

Much better agreement than linear model

MS, Simonović, Assassi, Zaldarriaga (2019)

How to improve?

Include all terms allowed by symmetries (effective field theory)

$$\delta_g(\mathbf{x}) = b_1\delta_m(\mathbf{x}) + b_2\delta_m^2(\mathbf{x}) + \text{tidal term} + b_3\delta_m^3(\mathbf{x}) + \cdots$$

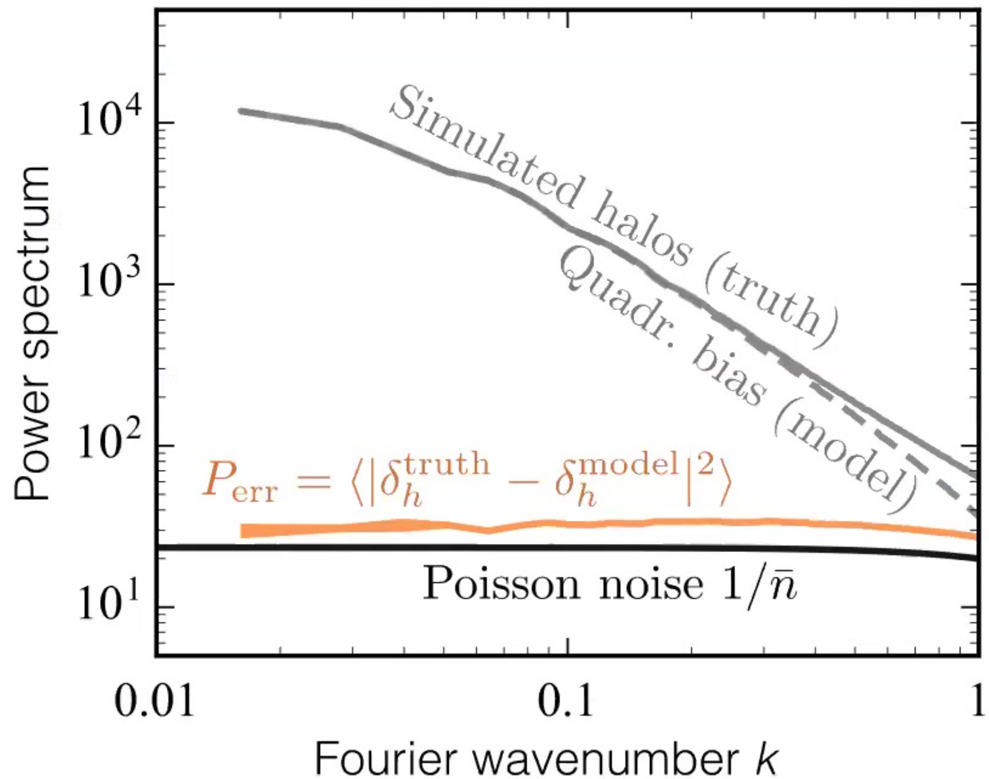
Desjacques, Jeong & Schmidt: Review of Large-Scale Galaxy Bias (2018)

Fit coefficients b_i using least-squares regression

MS, Simonović, Assassi, Zaldarriaga (2019)

In practice:
- Run independent regression for each Fourier mode shell
- Fit resulting regression coefficients $b_i(k)$ with 6-parameter model
- Orthogonalize operators for robust numerics and interpretation
- Include large bulk flows nonperturbatively (see later)

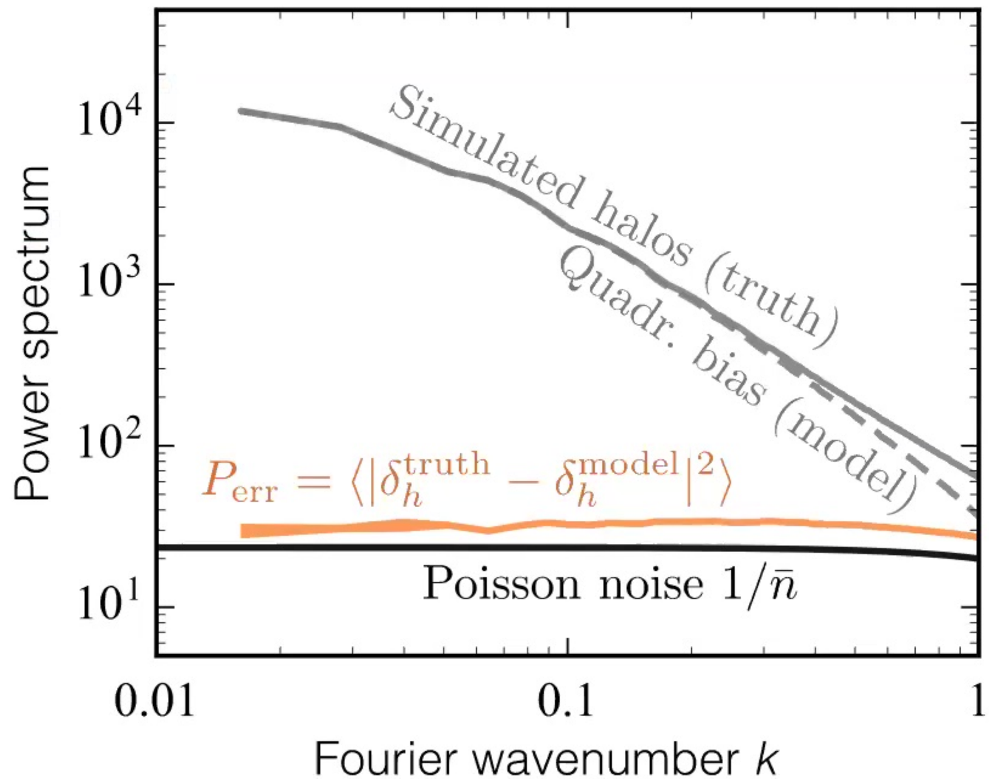
Mean-squared model error



This white noise is crucial to avoid biasing physical parameters

MS, Simonović, Assassi, Zaldarriaga (2019)

Mean-squared model error



This white noise is crucial to avoid biasing physical parameters

MS, Simonović, Assassi, Zaldarriaga (2019)

How to improve?

Include all terms allowed by symmetries (effective field theory)

$$\delta_g(\mathbf{x}) = b_1\delta_m(\mathbf{x}) + b_2\delta_m^2(\mathbf{x}) + \text{tidal term} + b_3\delta_m^3(\mathbf{x}) + \cdots$$

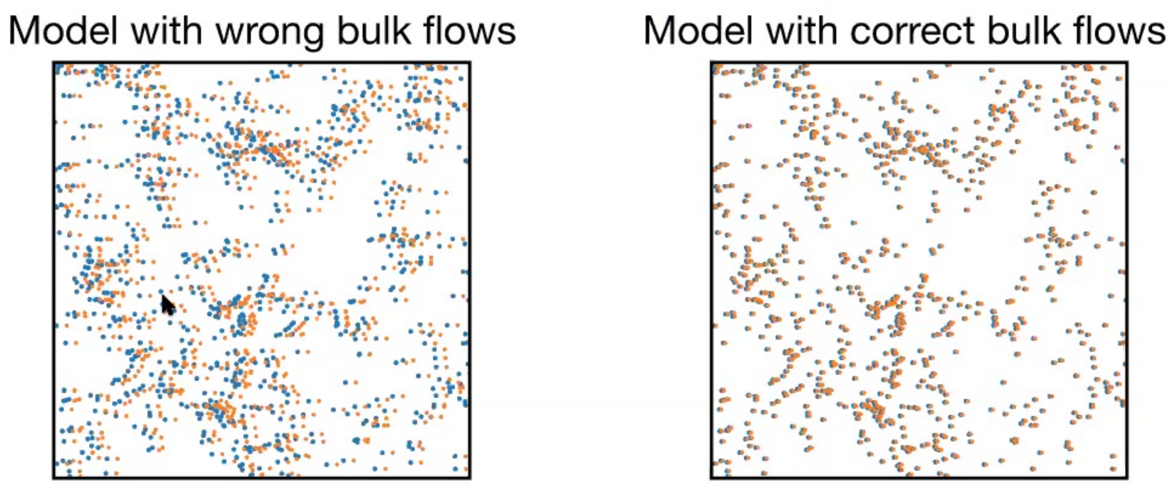
Desjacques, Jeong & Schmidt: Review of Large-Scale Galaxy Bias (2018)

Fit coefficients b_i using least-squares regression

MS, Simonović, Assassi, Zaldarriaga (2019)

In practice:
- Run independent regression for each Fourier mode shell
- Fit resulting regression coefficients $b_i(k)$ with 6-parameter model
- Orthogonalize operators for robust numerics and interpretation
- Include large bulk flows nonperturbatively (see later)

Bulk flows



Correct bulk flows are crucial for small pixel-level residual

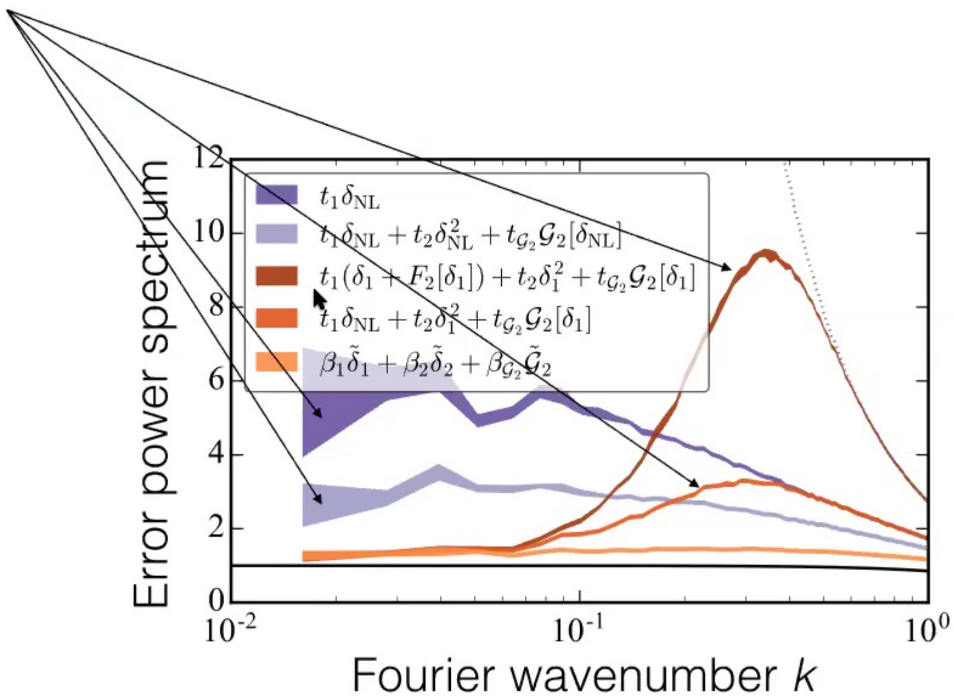
Cannot Taylor expand because bulk flows are large

$$\delta(x+\psi) \neq \delta(x) + \psi \nabla \delta(x)$$

MS, Simonović, Assassi, Zaldarriaga (2019)

Tried many other nonlinear models

3-6x larger model error



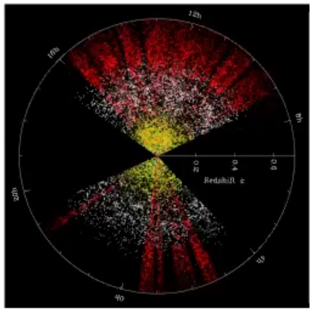
Our model allows using small-scale data (~10x larger volume)
Main reason: Bulk flows included nonperturbatively

MS, Simonović, Assassi, Zaldarriaga (2019)

Real data

Same model as recent EFT analyses of BOSS power spectrum

D'Amico, Gleyzes, Kokron et al. (1909.05271)
Ivanov, Simonović & Zaldarriaga (1909.05277)
Tröster, Sanchez, Asgari et al. (2020)

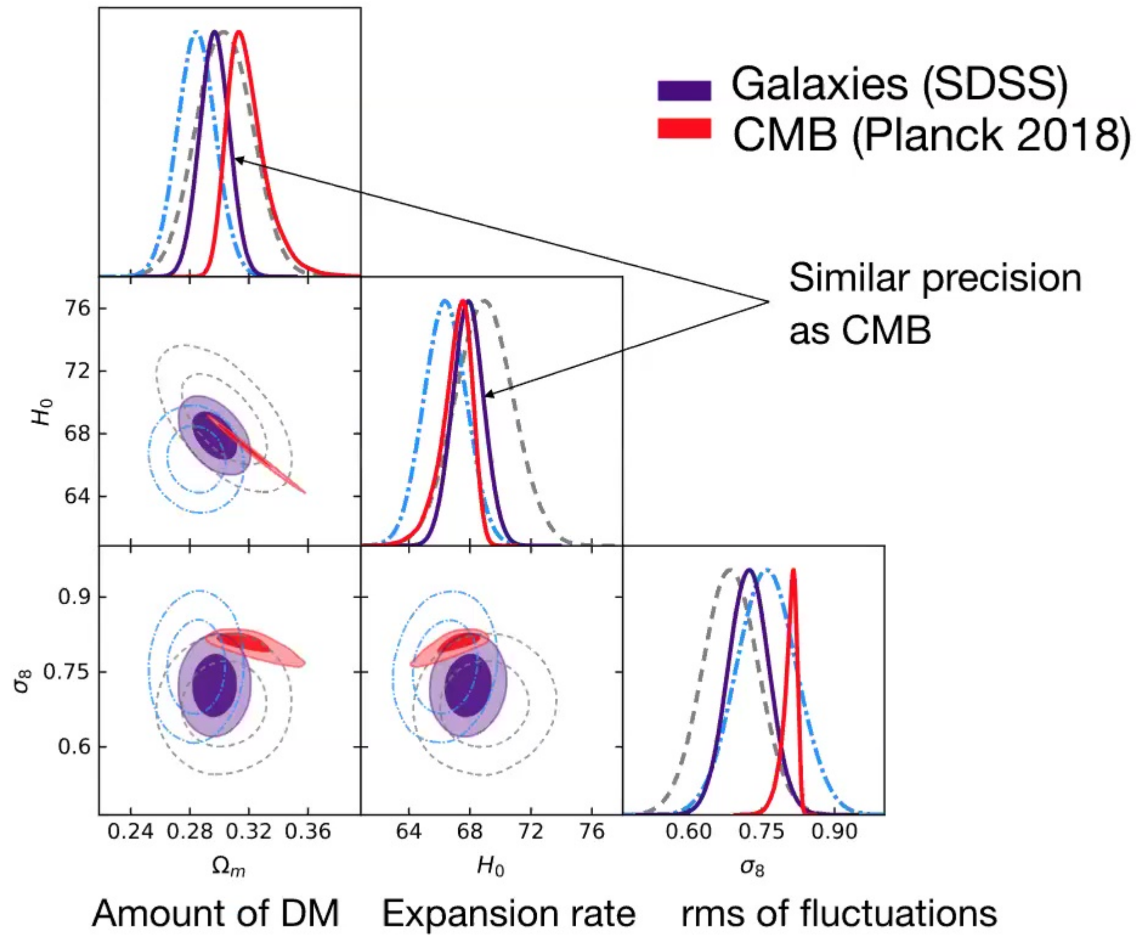


SDSS

Model power spectrum is evaluated using FFTLog trick to speed up MCMC chains (reduce 2D loop integrals to 1D FFTs)

Hamilton (2000)
MS, Vlah & McDonald (2016)
McEwen, Fang, Hirata & Blazek (2016)
Cataneo, Foreman & Senatore (2017)
Simonović, Baldauf, Zaldarriaga et al. (2018)

Competitive for some parameters



Ivanov et al. (arXiv:1909.05277)

Redshift space

Distance to galaxy inferred from redshift = true distance + velocity

Model velocity field in Lagrangian perturbation theory to get a field level model for galaxies in redshift space

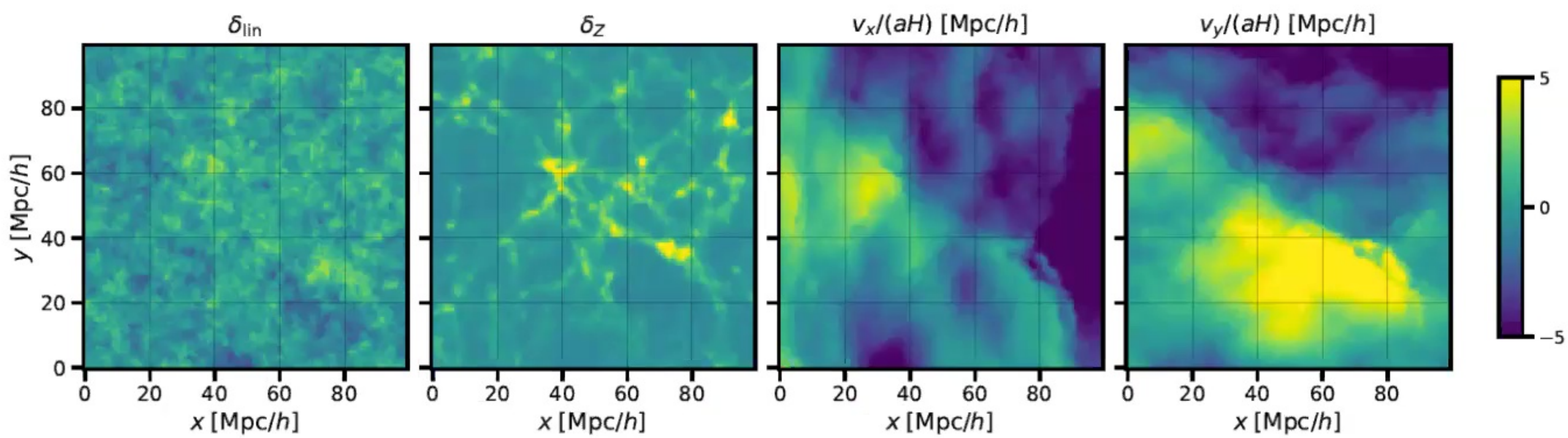
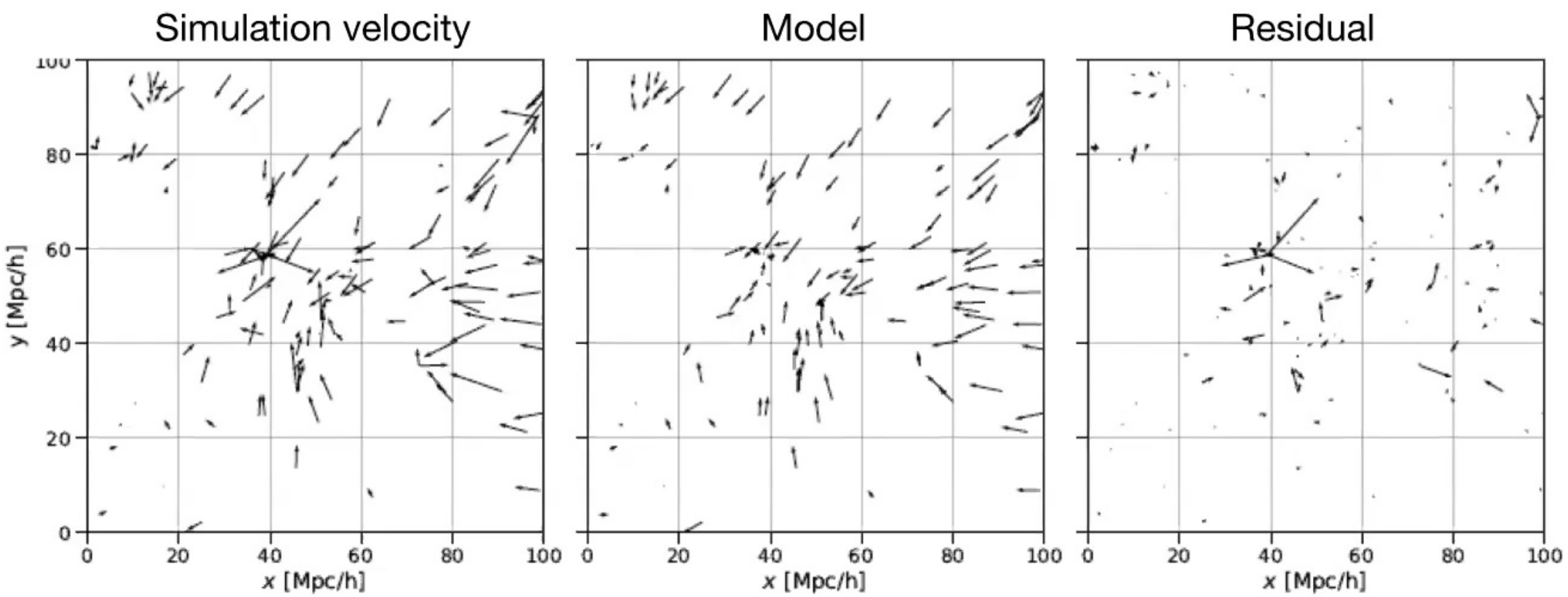


Figure 1: 2D slice of the linear density, Zel'dovich density, and x- and y-component of the continuous velocity field predicted by Eq. (2.12) for $n_{\max} = 1$. The predicted velocity field is coherent over tens of Megaparsecs, with most regions flowing towards the cluster and filament in the center of the slice. To generate the Zel'dovich density and the velocity prediction, 1536^3 particles in a Lagrangian space box with $L = 500\ h^{-1}$Mpc were shifted by the first-order displacement. All fields are evaluated at redshift $z = 0.6$.

MS, Simonović, Ivanov, Philcox & Zaldarriaga (arXiv:2012.03334)

Comparison with simulations

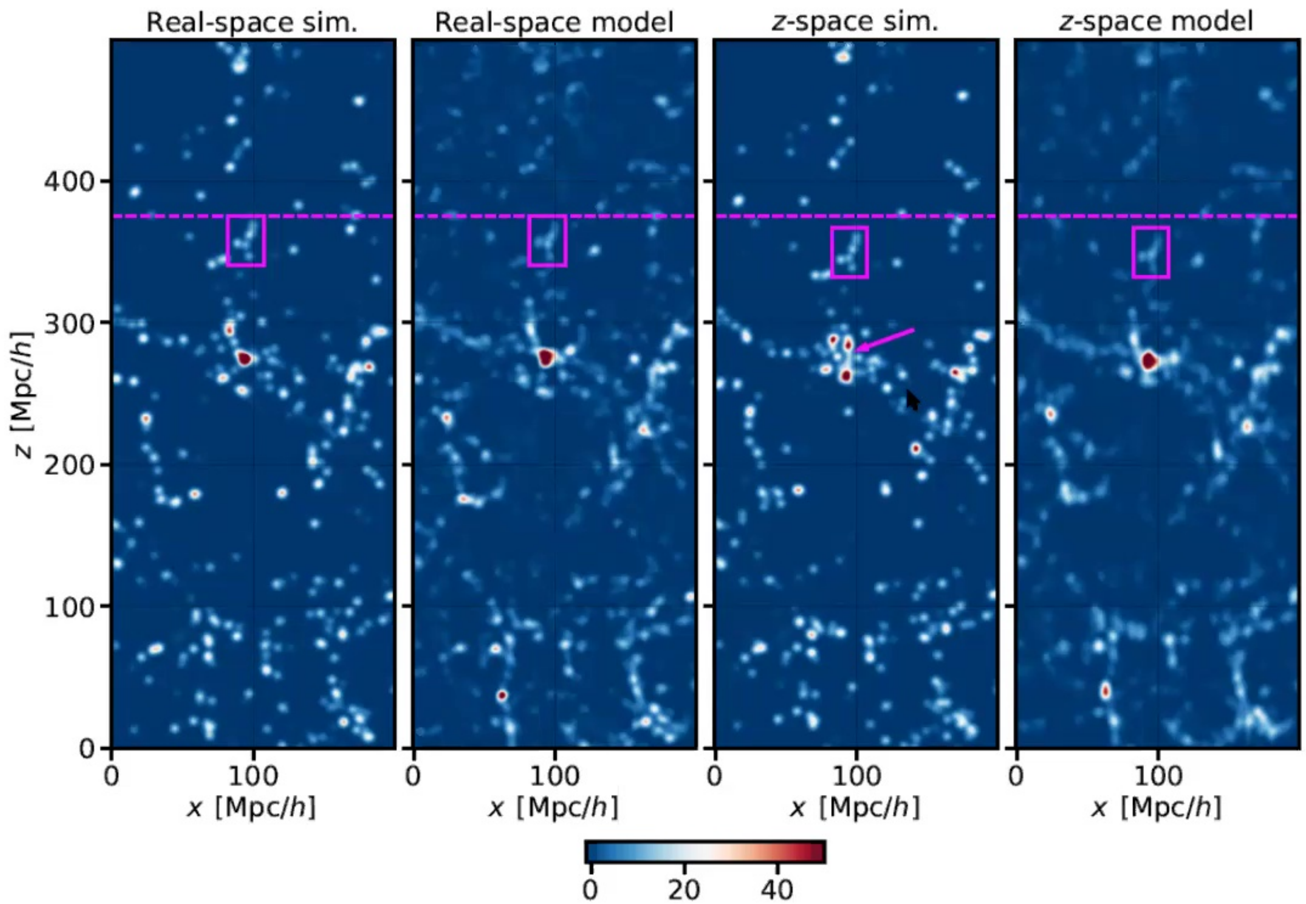
Model describes bulk flows in simulations well, but does not capture large velocities in highly clustered regions (satellites):



MS, Simonović, Ivanov, Philcox & Zaldarriaga (arXiv:2012.03334)

Comparison with simulations

Model captures large-scale flows, but not Fingers of God



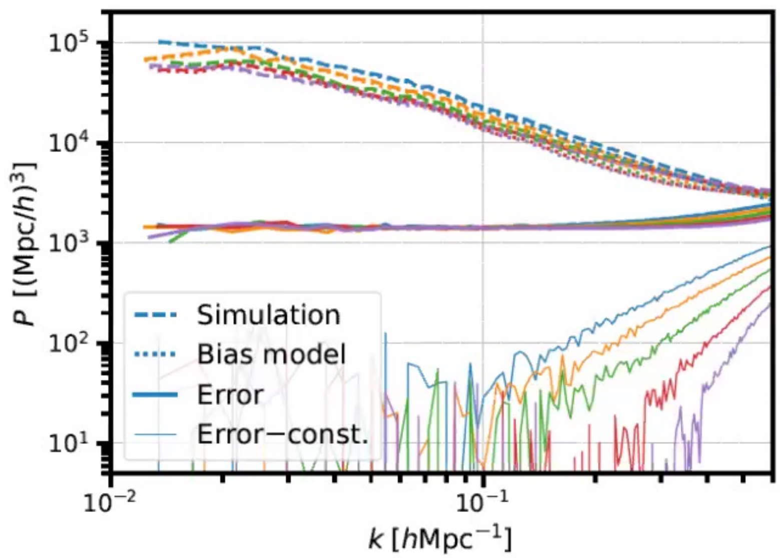
Comparison with simulations

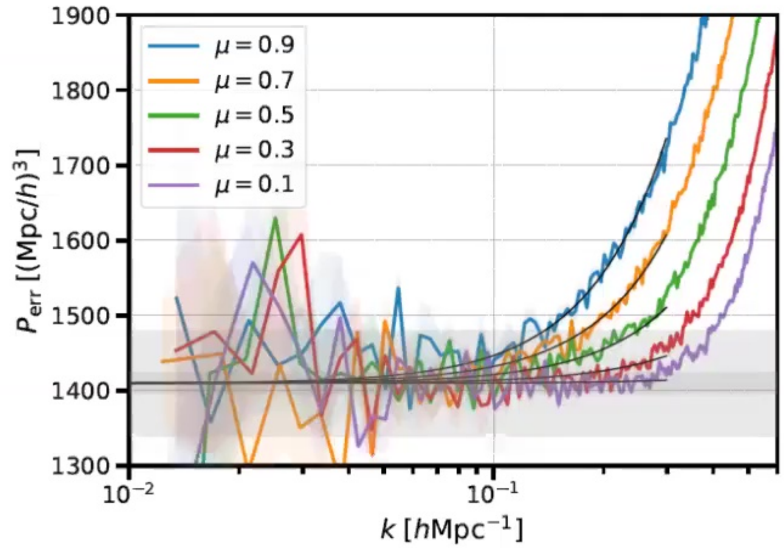
Functional form of model error power spectrum:

$$P_{\rm err}(k,\mu) = \frac{1}{\bar{n}_g}\left(c_{\epsilon,1} + c_{\epsilon,3} f \mu^2 \left(\frac{k}{k_{\rm M}}\right)^2\right)$$

$$c_{\epsilon,1} = 0.599$$

$$c_{\epsilon,3} = 2.45 \left(\frac{k_{\rm M}}{1\, h{\rm Mpc}^{-1}}\right)^2$$





Perko, Senatore, Jennings & Wechsler (arXiv:1610.09321)
MS, Simonović, Ivanov, Philcox & Zaldarriaga (arXiv:2012.03334)

https://github.com/mschmittfull/perr

Additional thoughts from 3D tests of models

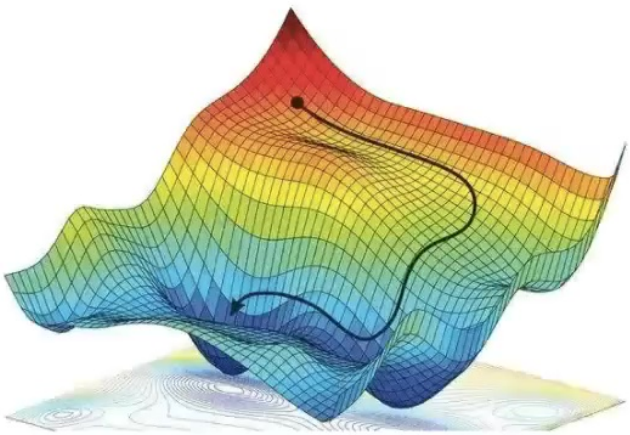
- Weighting galaxies by their mass can reduce model error 10x
- Removing 13% of galaxies gives 2x smaller rms RSD displacement, enabling higher model k_{max}

Additional thoughts from 3D tests of models

- Weighting galaxies by their mass can reduce model error 10x
- Removing 13% of galaxies gives 2x smaller rms RSD displacement, enabling higher model $k_{\max}$

How to find weights/detect outliers?

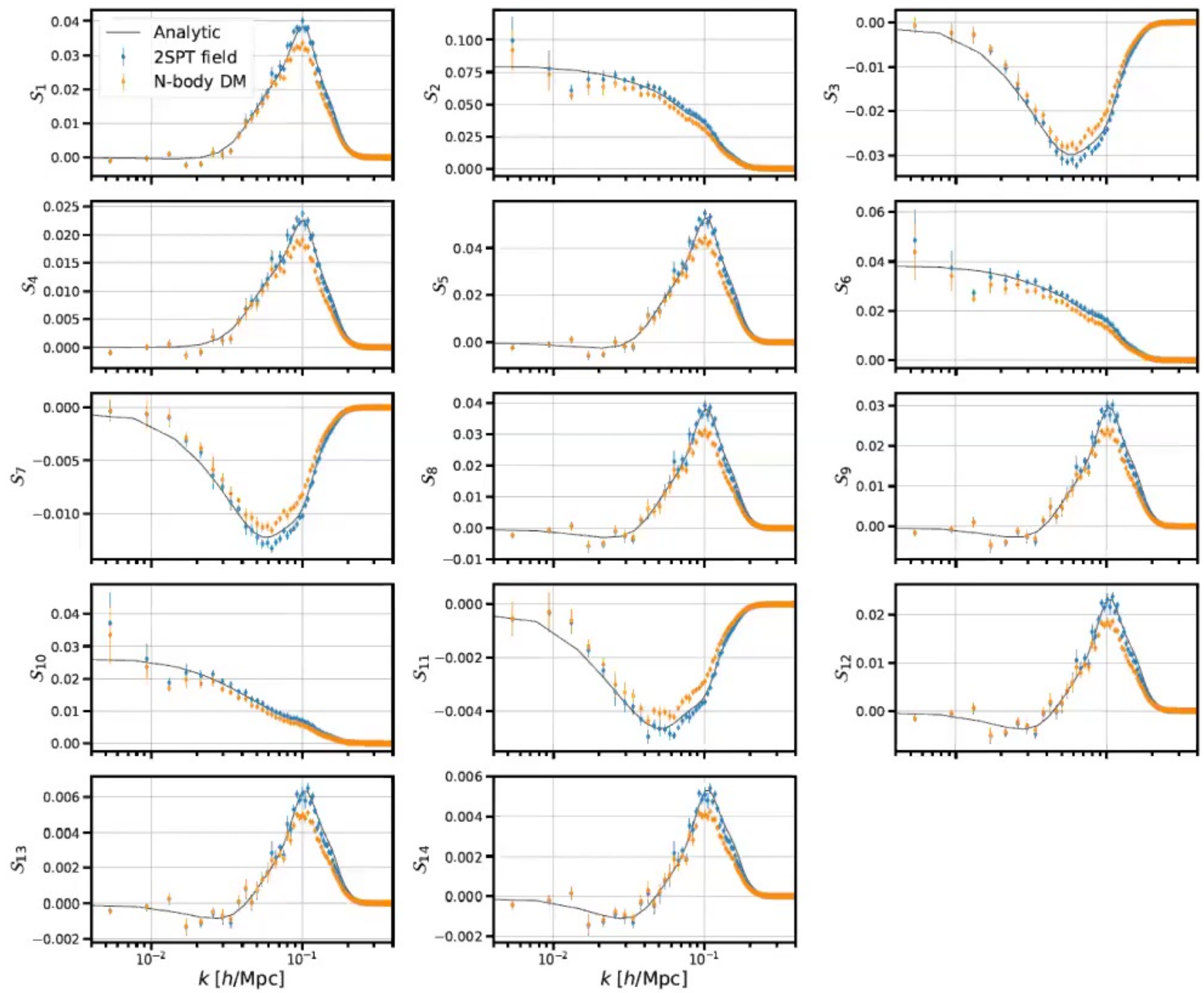
Solve as optimization problem with custom-made objective (e.g. maximize quadrupole/monopole ratio at high k to suppress FoG and shot noise)



Note: We can weight galaxies by any function of local observables and still use the same galaxy bias model

Inspired by Obuljen, Percival & Dalal (2020)

Test on synthetic dark matter data



MS & Moradinezhad Dizgah (2021)

https://github.com/mschmittfull/skewspec

Galaxy skew-spectra in redshift space

Cross-spectra between 14 quadratic fields $\mathcal{S}_n$ and galaxy density

$$b_1^3: \quad \mathcal{S}_1 = F_2[\delta, \delta]$$
$$b_1^2 b_2: \quad \mathcal{S}_2 = \delta^2$$
$$b_1^2 b_{\mathcal{G}_2}: \quad \mathcal{S}_3 = S^2[\delta, \delta]$$
$$b_1^3 f: \quad \mathcal{S}_4 = \hat{z}_i \hat{z}_j \, \partial_i \left(\delta \frac{\partial_j}{\nabla^2} \delta \right)$$
$$b_1^2 f: \quad \mathcal{S}_5 = 2F_2[\delta^{\|}, \delta] + G_2^{\|}[\delta, \delta]$$
$$b_1 b_2 f: \quad \mathcal{S}_6 = \delta \delta^{\|}$$
$$b_1 b_{\mathcal{G}_2} f: \quad \mathcal{S}_7 = S^2[\delta, \delta^{\|}]$$
$$b_1^2 f^2: \quad \mathcal{S}_8 = \hat{z}_i \hat{z}_j \partial_i \left(\delta \frac{\partial_j}{\nabla^2} \delta^{\|} + 2\delta^{\|} \frac{\partial_j}{\nabla^2} \delta \right)$$
$$b_1 f^2: \quad \mathcal{S}_9 = F_2[\delta^{\|}, \delta^{\|}] + 2G_2^{\|}[\delta^{\|}, \delta]$$
$$b_2 f^2: \quad \mathcal{S}_{10} = (\delta^{\|})^2$$
$$b_{\mathcal{G}_2} f^2: \quad \mathcal{S}_{11} = S^2(\delta^{\|}, \delta^{\|})$$
$$b_1 f^3: \quad \mathcal{S}_{12} = \hat{z}_i \hat{z}_j \partial_i \left(\delta^{\|\|} \frac{\partial_j}{\nabla^2} \delta + 2\delta^{\|} \frac{\partial_j}{\nabla^2} \delta^{\|} \right)$$
$$f^3: \quad \mathcal{S}_{13} = G_2^{\|}[\delta^{\|}, \delta^{\|}]$$
$$f^4: \quad \mathcal{S}_{14} = \hat{z}_i \hat{z}_j \partial_i \left(\delta^{\|\|} \frac{\partial_j}{\nabla^2} \delta^{\|} \right) .$$

MS & Moradinezhad Dizgah (2021)

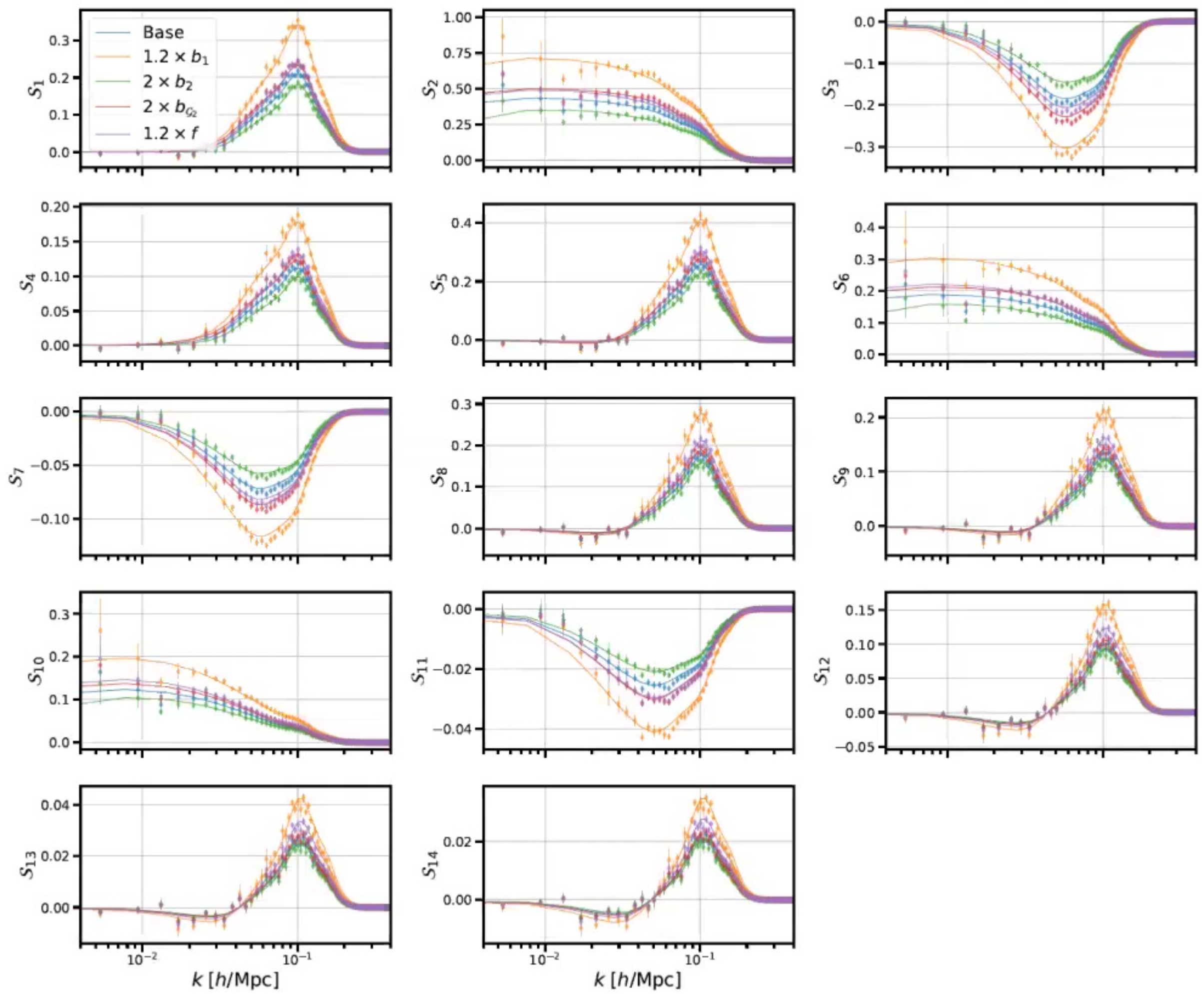
Galaxy skew-spectra in redshift space

Cross-spectra between 14 quadratic fields $\mathcal{S}_n$ and galaxy density

$$b_1^3: \quad \mathcal{S}_1 = F_2[\delta, \delta]$$
$$b_1^2 b_2: \quad \mathcal{S}_2 = \delta^2$$
$$b_1^2 b_{\mathcal{G}_2}: \quad \mathcal{S}_3 = S^2[\delta, \delta]$$
$$b_1^3 f: \quad \mathcal{S}_4 = \hat{z}_i \hat{z}_j \, \partial_i \left(\delta \frac{\partial_j}{\nabla^2} \delta \right)$$
$$b_1^2 f: \quad \mathcal{S}_5 = 2F_2[\delta^{\|}, \delta] + G_2^{\|}[\delta, \delta]$$
$$b_1 b_2 f: \quad \mathcal{S}_6 = \delta \delta^{\|}$$
$$b_1 b_{\mathcal{G}_2} f: \quad \mathcal{S}_7 = S^2[\delta, \delta^{\|}]$$
$$b_1^2 f^2: \quad \mathcal{S}_8 = \hat{z}_i \hat{z}_j \partial_i \left(\delta \frac{\partial_j}{\nabla^2} \delta^{\|} + 2\delta^{\|} \frac{\partial_j}{\nabla^2} \delta \right)$$
$$b_1 f^2: \quad \mathcal{S}_9 = F_2[\delta^{\|}, \delta^{\|}] + 2G_2^{\|}[\delta^{\|}, \delta]$$
$$b_2 f^2: \quad \mathcal{S}_{10} = (\delta^{\|})^2$$
$$b_{\mathcal{G}_2} f^2: \quad \mathcal{S}_{11} = S^2(\delta^{\|}, \delta^{\|})$$
$$b_1 f^3: \quad \mathcal{S}_{12} = \hat{z}_i \hat{z}_j \partial_i \left(\delta^{\|\|} \frac{\partial_j}{\nabla^2} \delta + 2\delta^{\|} \frac{\partial_j}{\nabla^2} \delta^{\|} \right)$$
$$f^3: \quad \mathcal{S}_{13} = G_2^{\|}[\delta^{\|}, \delta^{\|}]$$
$$f^4: \quad \mathcal{S}_{14} = \hat{z}_i \hat{z}_j \partial_i \left(\delta^{\|\|} \frac{\partial_j}{\nabla^2} \delta^{\|} \right) .$$

MS & Moradinezhad Dizgah (2021)

Test on synthetic galaxy data



MS & Moradinezhad Dizgah (2021)

https://github.com/mschmittfull/skewspec

Code & Simulations

- Field level model — https://github.com/mschmittfull/perr
- Skew-spectra — https://github.com/mschmittfull/skewspec
- Iterative reconstruction — https://github.com/mschmittfull/iterrec

All based on nbodykit — https://github.com/bccp/nbodykit

For questions email mschmittfull@gmail.com