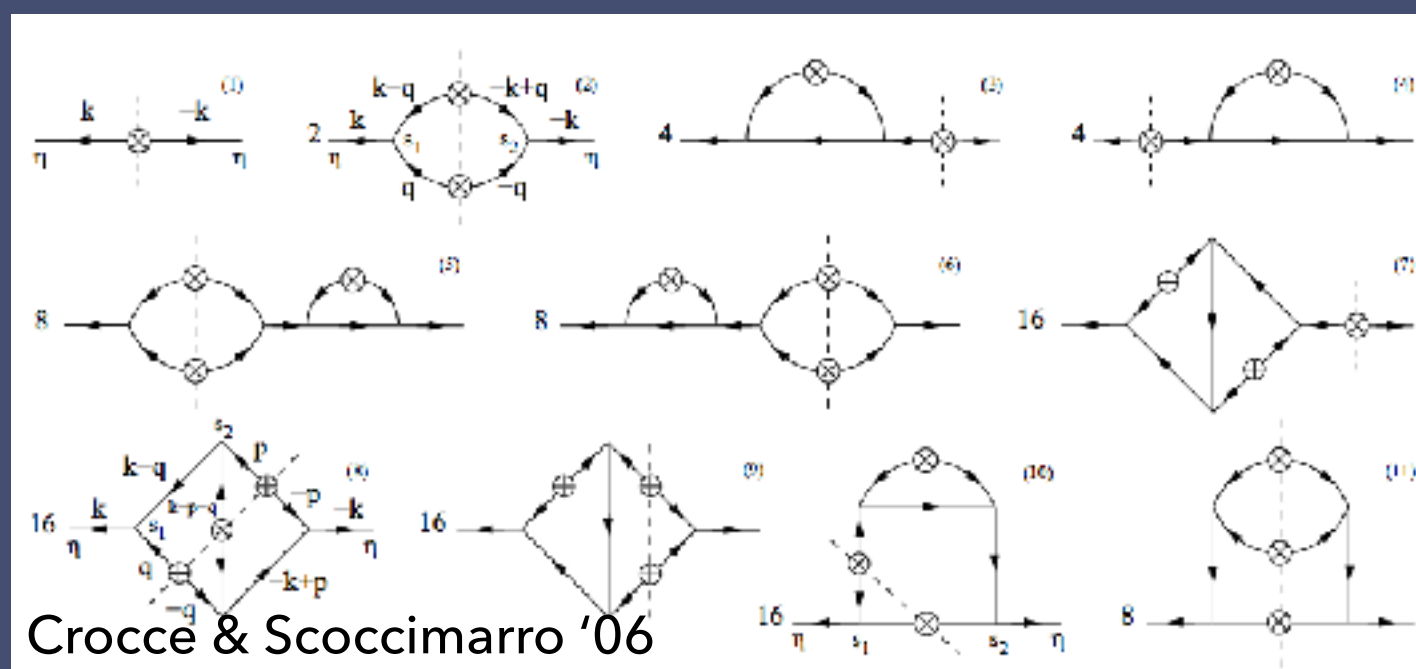


宇宙大規模構造を用いた宇宙論の データ科学的側面

西道 啓博 (京都大学 基礎物理学研究所/Kavli IPMU)

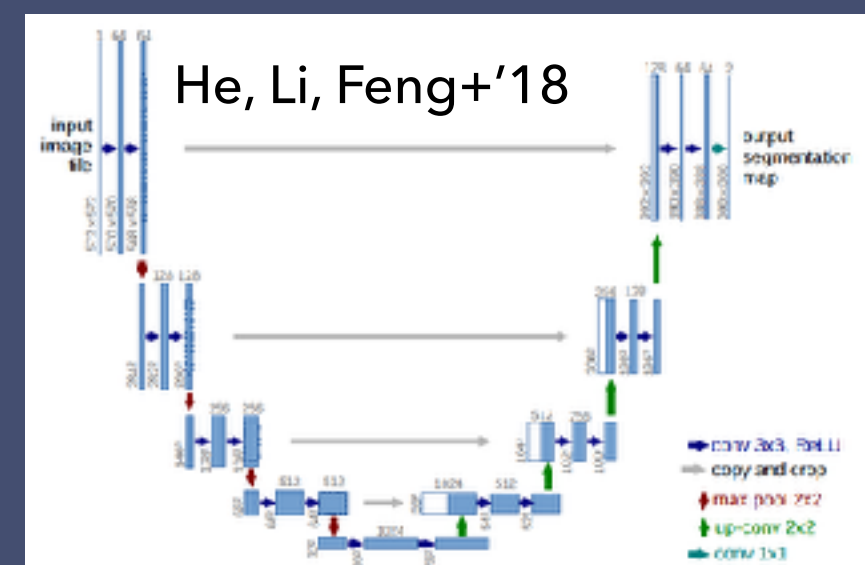
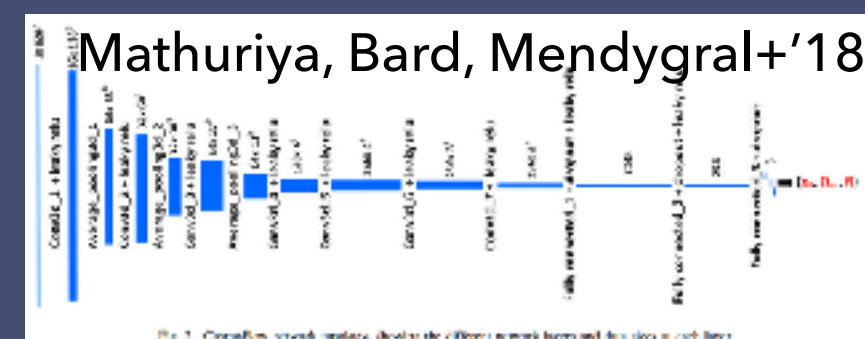


Crocce & Scoccimarro '06

$$P_{2b}(k) = \text{[Diagram 1]} + 2 \text{[Diagram 2]} + 6 \text{[Diagram 3]}$$

Diagram illustrating the two-loop power spectrum $P_{2b}(k)$ using Feynman diagrams. The diagrams are labeled (1) through (11). The diagrams show various loop topologies and external momenta k , q , p , and s_1, s_2 . The diagrams are arranged in three rows: the first row has (1), (2), (3), and (4); the second row has (5), (6), and (7); the third row has (8), (9), (10), and (11). The diagrams are connected by plus signs, indicating they are summed to form the total power spectrum.

Bernardeau, Crocce & Scoccimarro '08



観測的宇宙論：基本的な考え方

多次元モデルパラメタ空間

Model A

順モデル

observable X

observation

distance, time,...

- モデル選択 / パラメタ推定の問題
- 宇宙を記述する本質的な物理量（パラメタ）をデータから読み解く
- 統計的な手法で
- 思い込みを排除して

問題設定と典型的な大きさ



生データ：画像、ペタバイト級



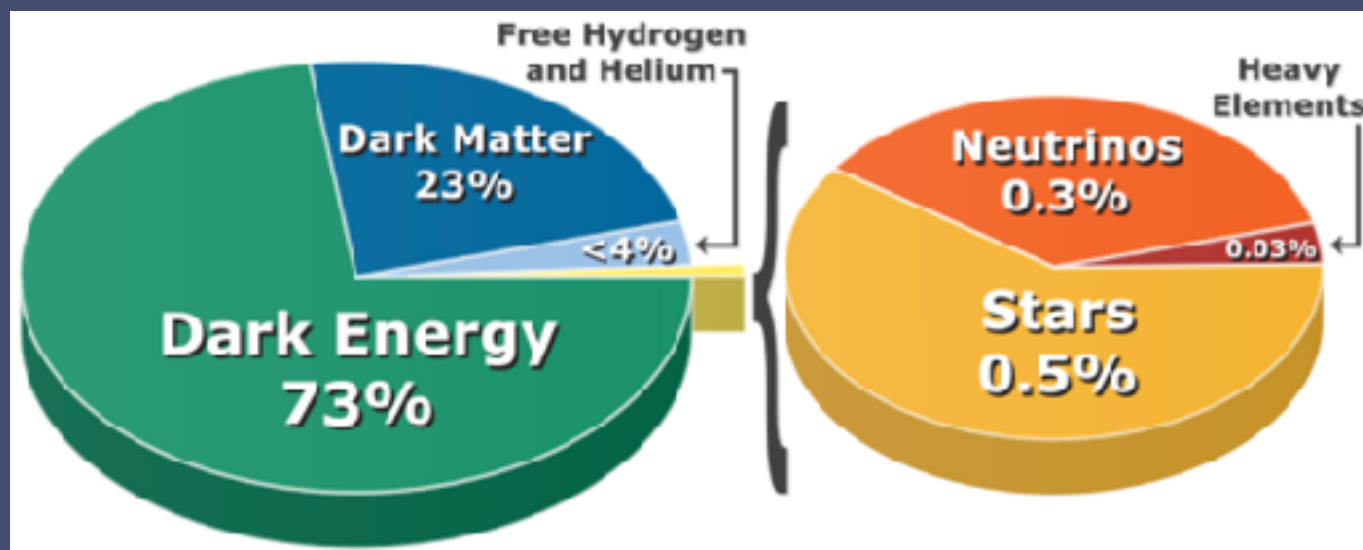
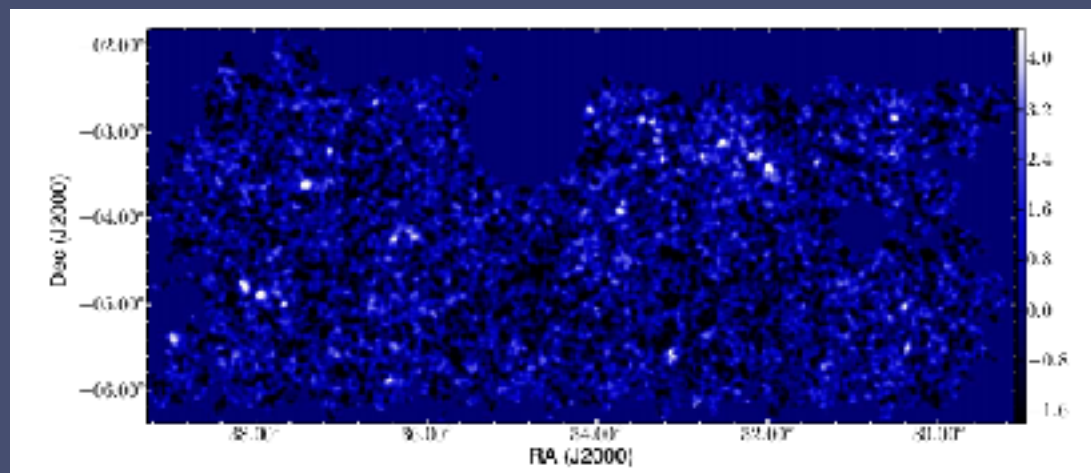
サイエンスデータ：銀河の位置と形状、ギガバイト級



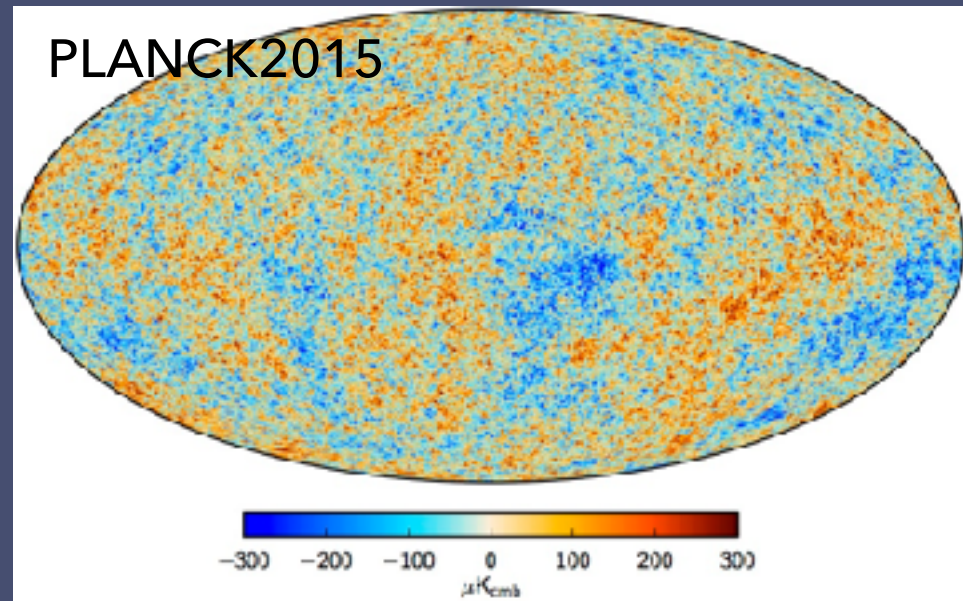
要約統計量：～100個程度の互いに相関した確率変数



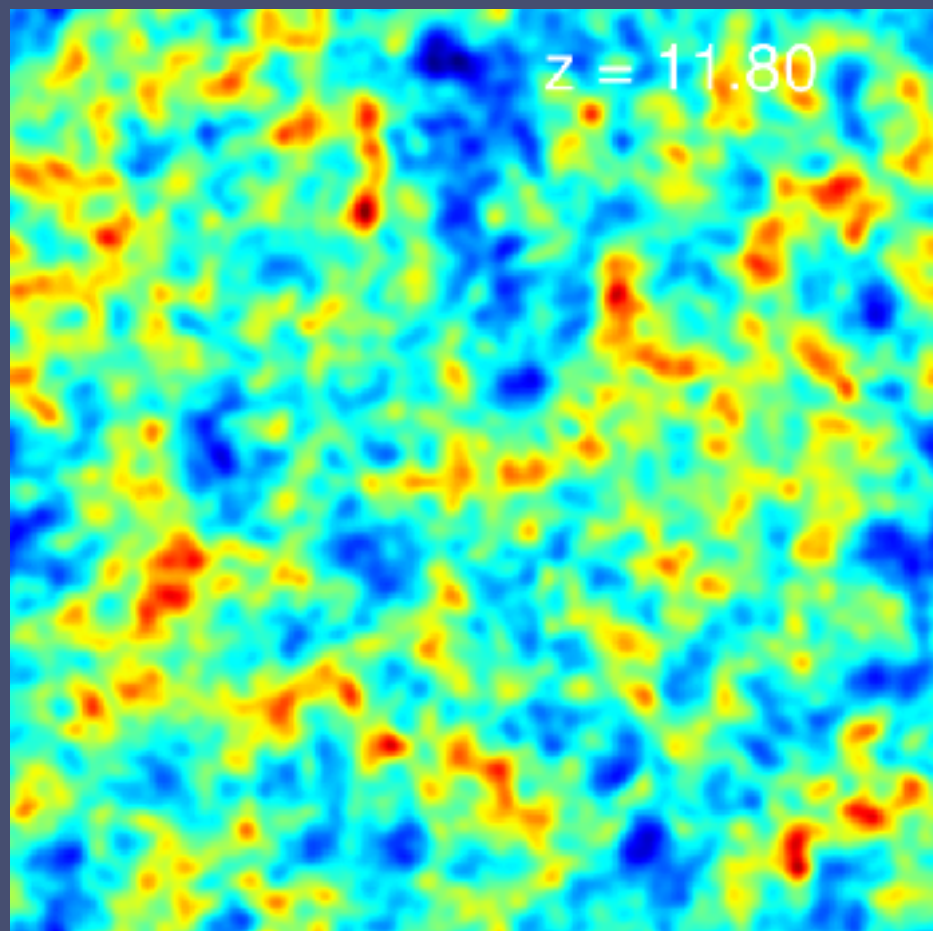
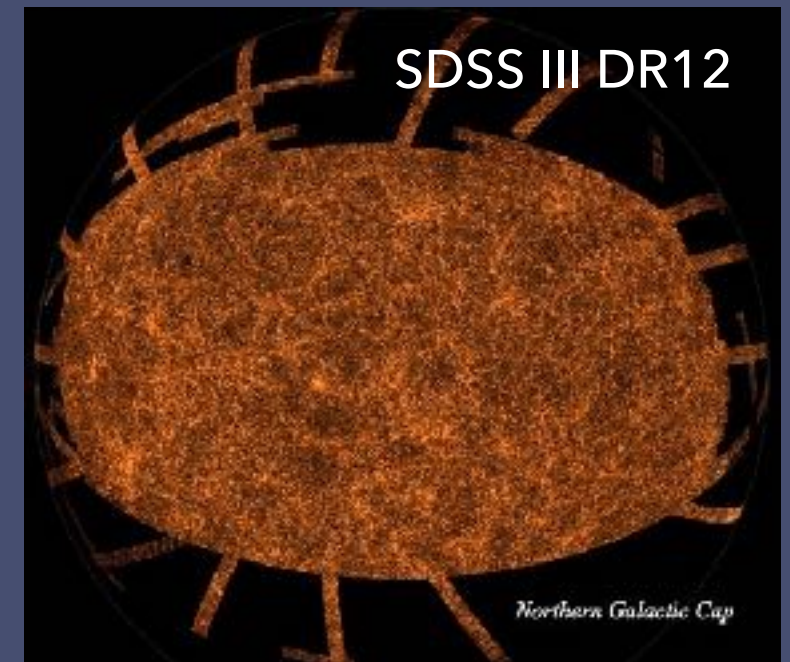
宇宙論パラメータ：数個～10個程度まで



大規模構造を用いた宇宙論



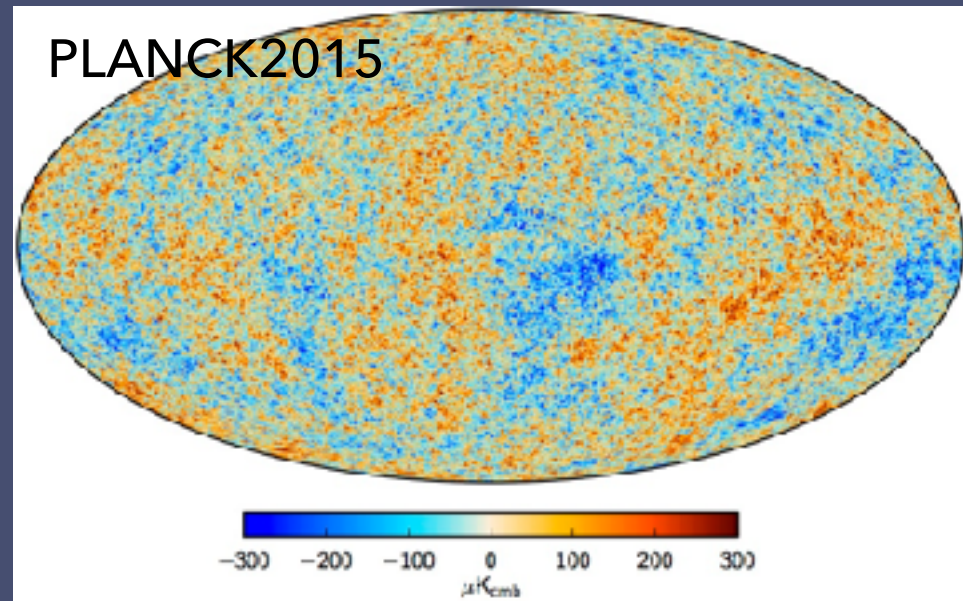
Gravity
→
Ingredients



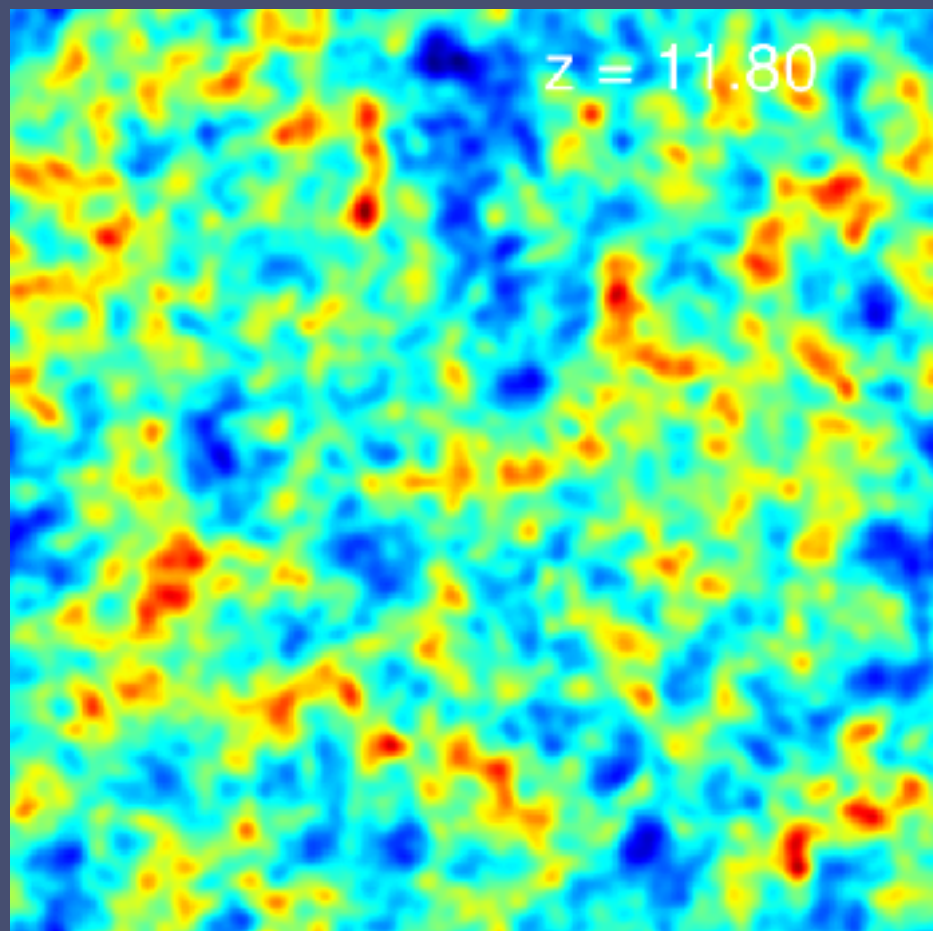
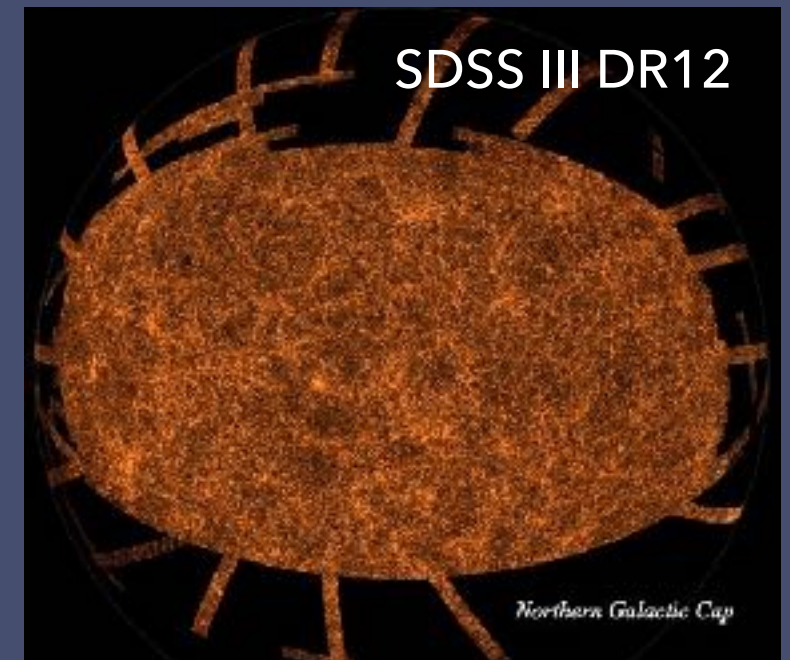
- **CMBを初期条件＋理論模型**
 - 定量的な理論予言を与える
- 近傍宇宙観測との比較により「答え合わせ」が可能
 - ダークエネルギーの正体
 - 物質（ダークマター、ニュートリノ）の性質
 - 重力模型の修正可能性

どこに情報があって、どの程度解釈できているのか？

大規模構造を用いた宇宙論



Gravity
→
Ingredients



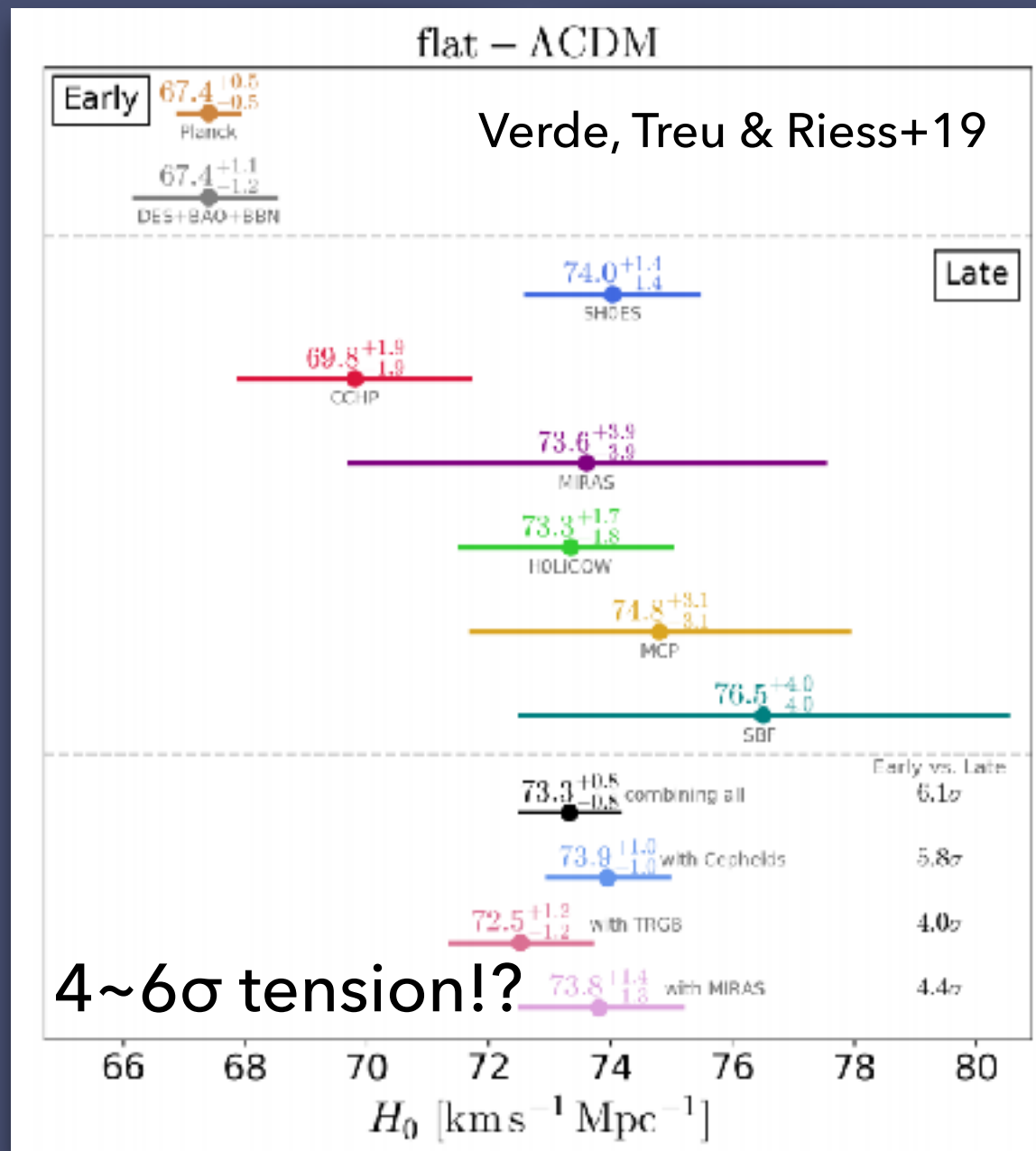
- **CMBを初期条件＋理論模型**
 - 定量的な理論予言を与える
- 近傍宇宙観測との比較により「答え合わせ」が可能
 - ダークエネルギーの正体
 - 物質（ダークマター、ニュートリノ）の性質
 - 重力模型の修正可能性

どこに情報があって、どの程度解釈できているのか？

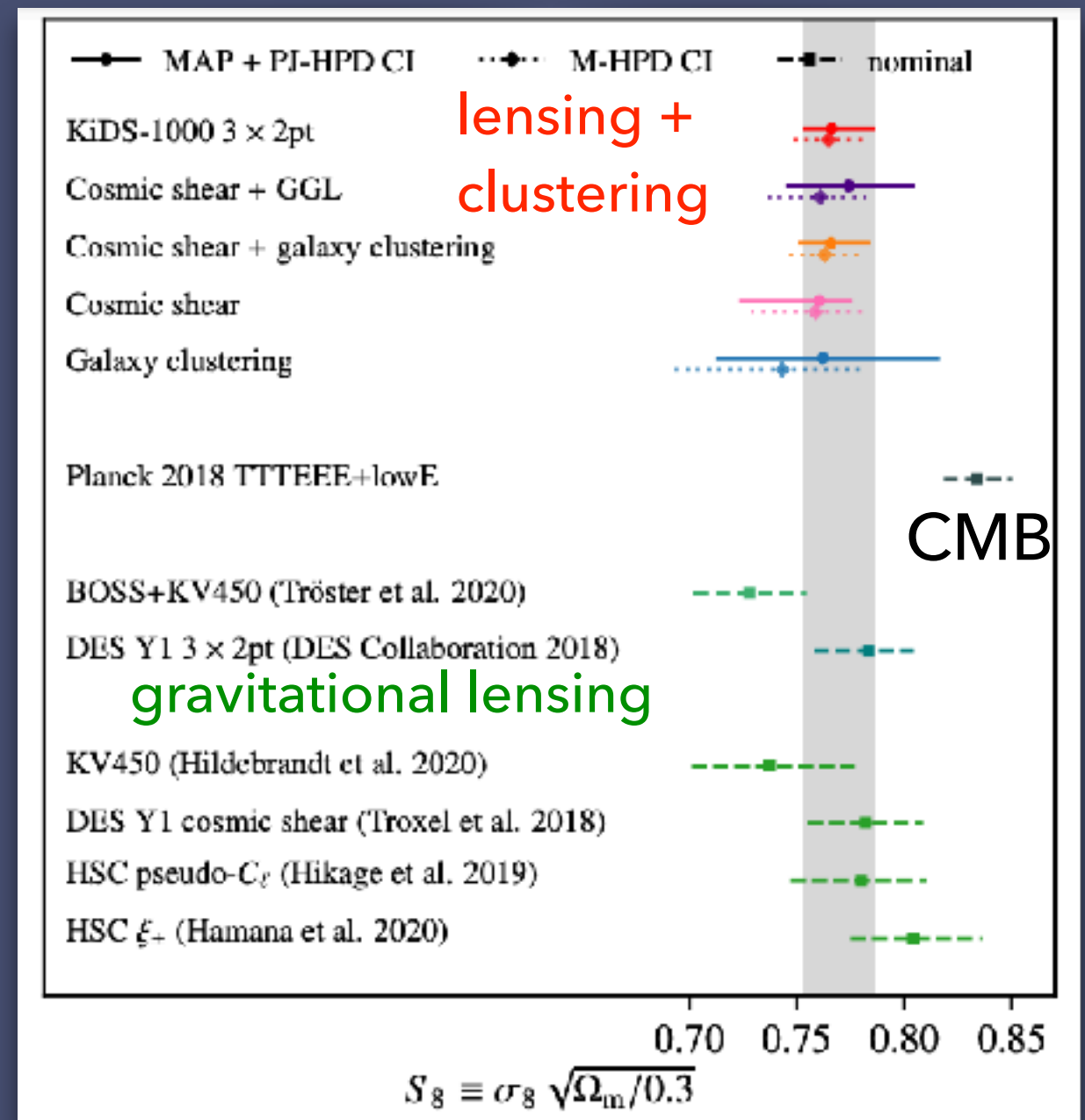
Λ CDMの綻び？

- CMB + Λ CDM vs 近傍観測
- 統計的に無視できないレベルの食い違い？

KiDS-1000 result
(arXiv:2007.15632)



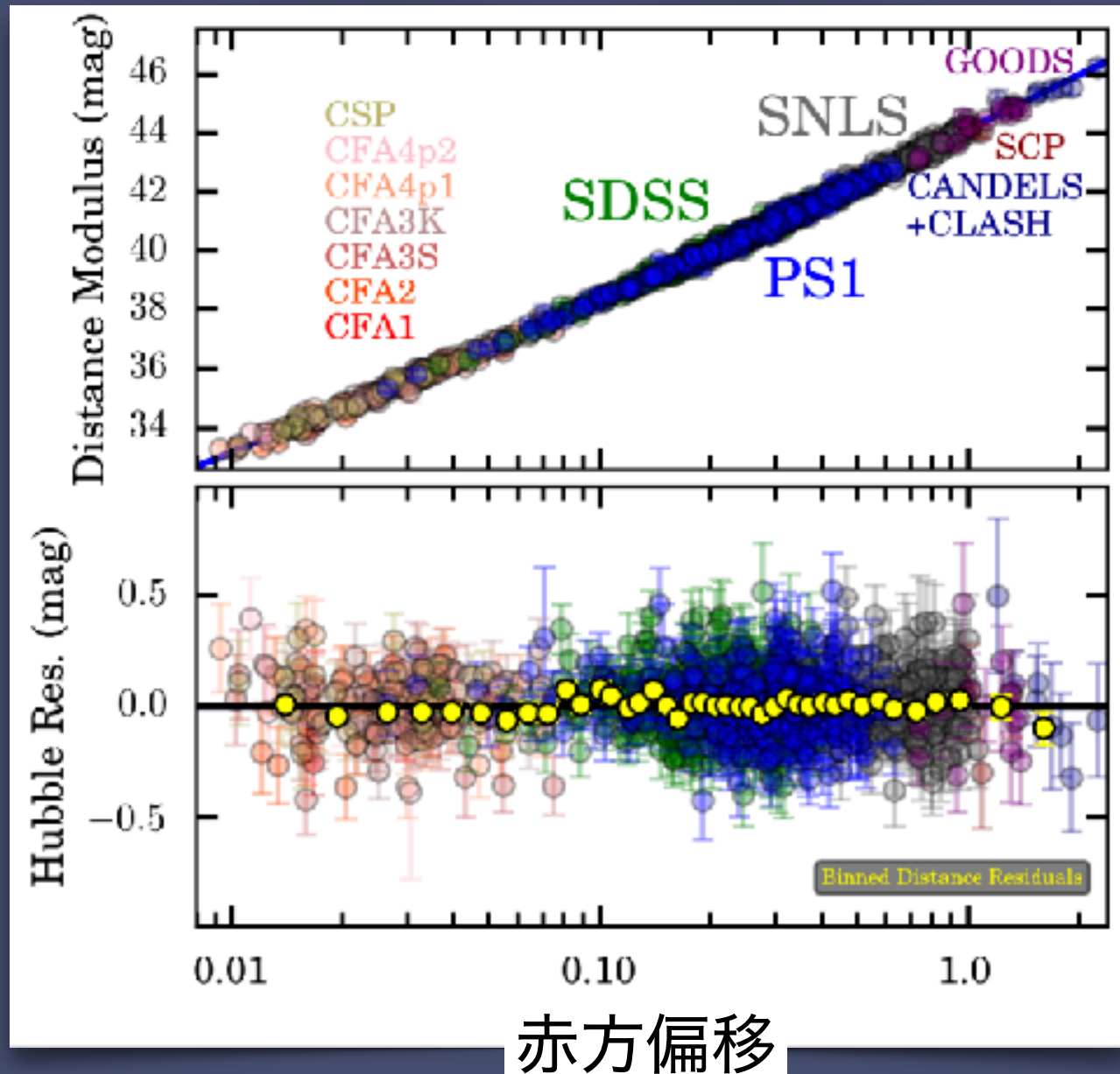
Hubble 膨張率 (km/s/Mpc)



揺らぎの振幅 $S_8 = \sigma_8 (\Omega_m/0.3)^{0.5}$

バックグラウンド vs 揺らぎ

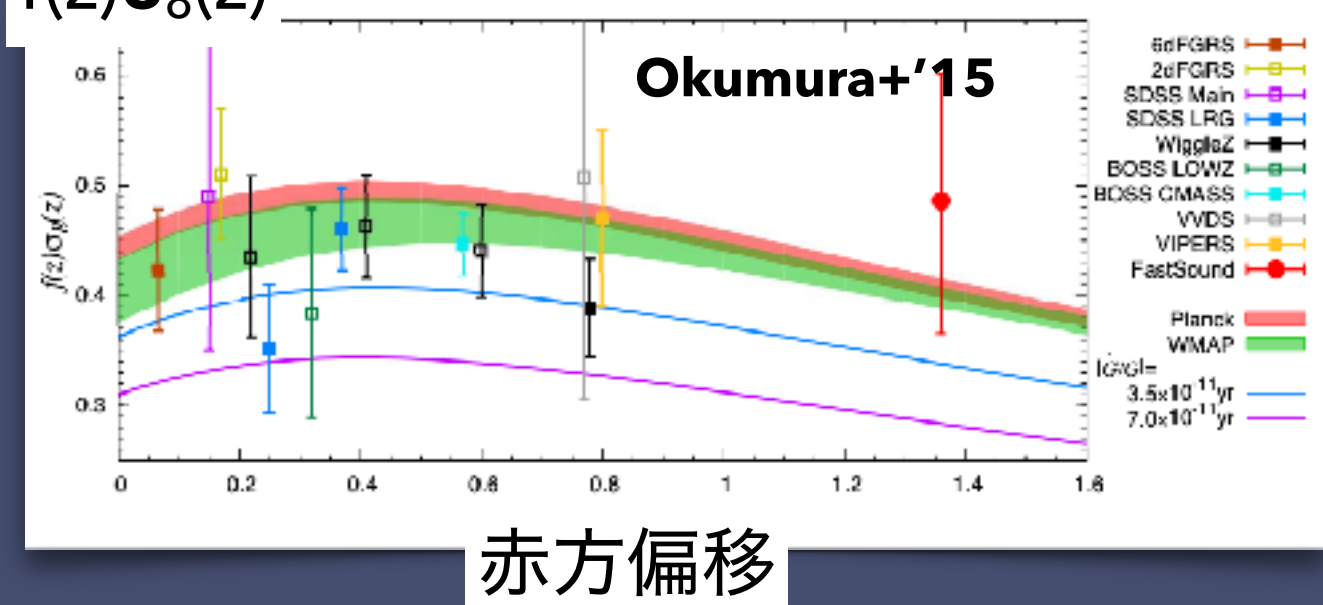
A compilation of type Ia SNe



Scolnic + '18

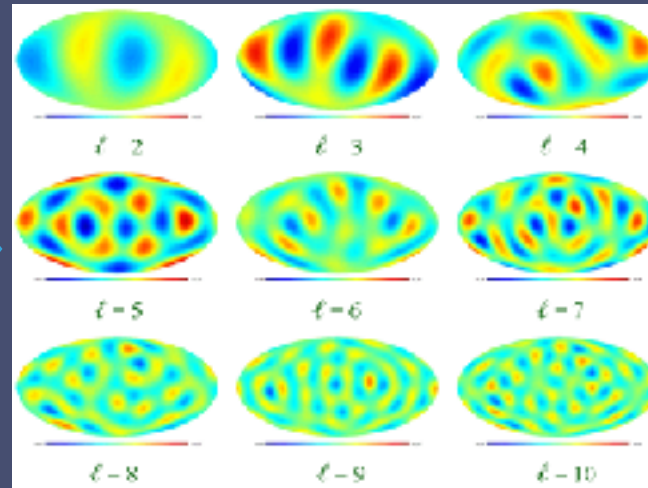
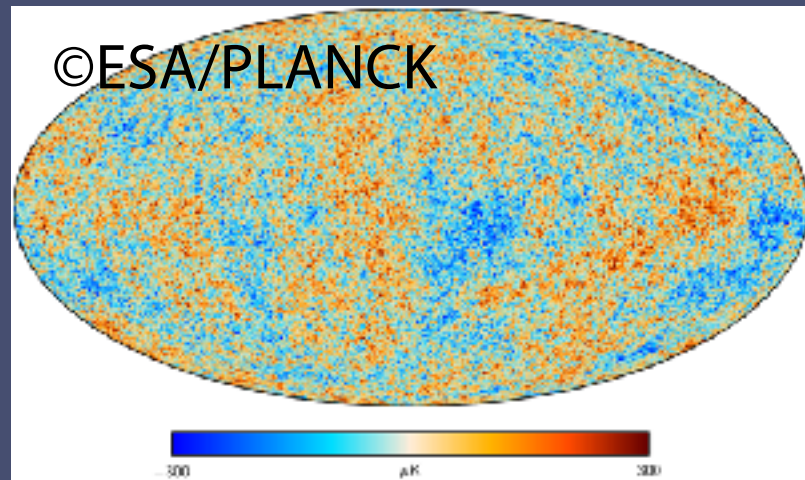
Redshift-space distortion

$f(z)\sigma_8(z)$



- ダークエネルギー vs 修正重力
- (重力レンズもそうだが) 揺らぎの成長に関わる特別なパラメタの組み合わせのみ測定可能
- 概して膨張則よりも揺らぎの振幅の方が測るのが難しい

CMBの持つ情報



spherical
harmonics

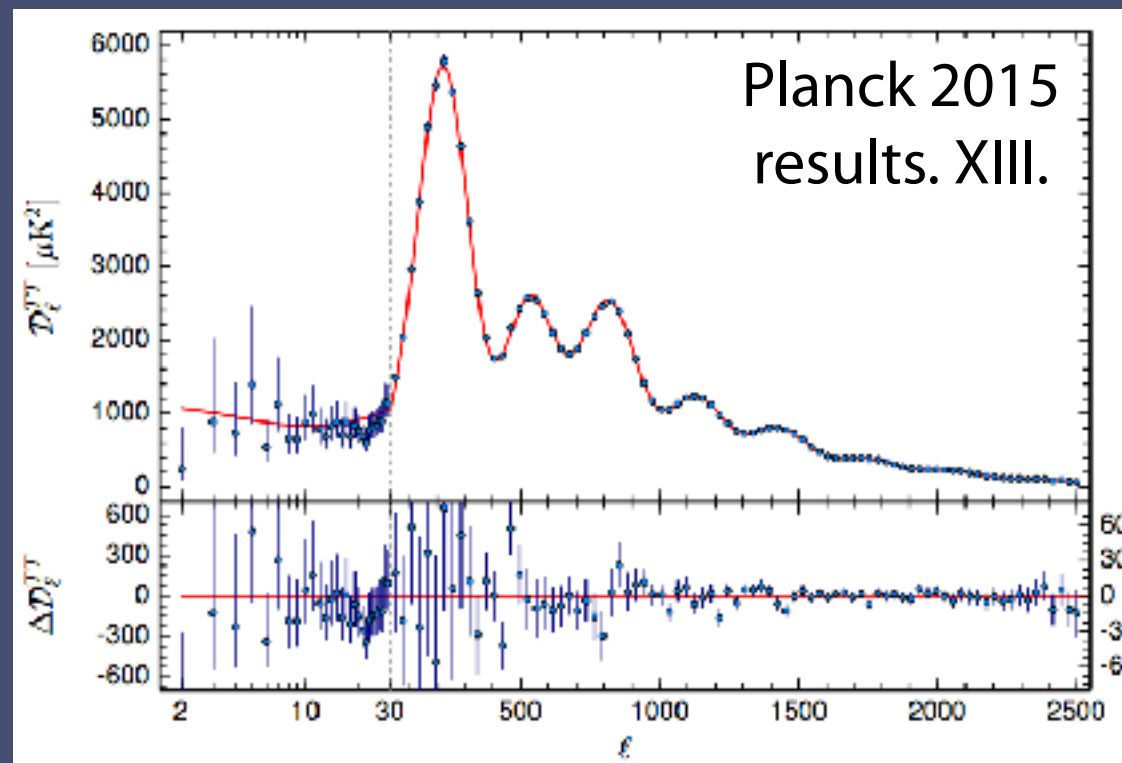
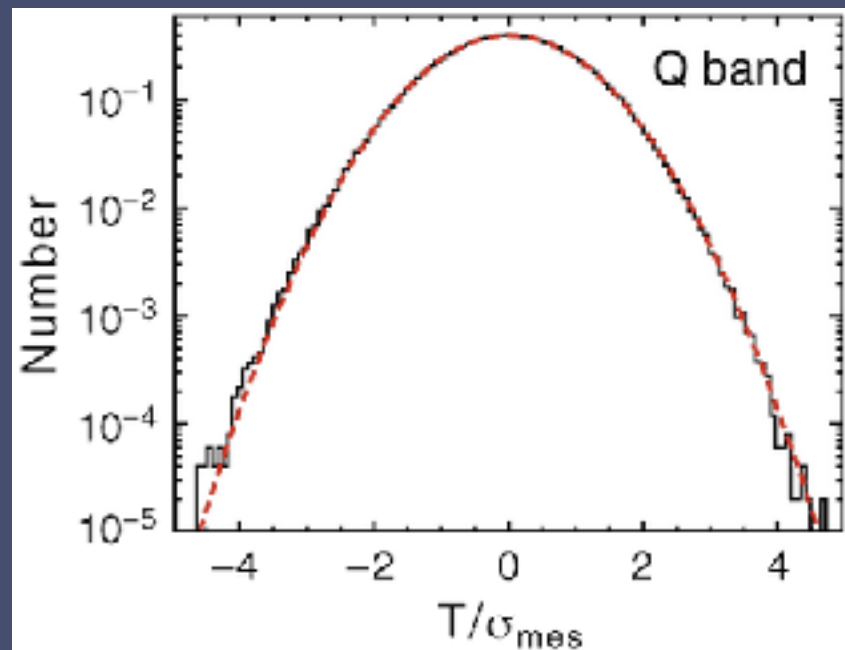
$$\Theta(\hat{p}) = \sum_{\ell, m} a_{\ell m} Y_{\ell m}(\hat{p})$$

“power spectrum”

$$\langle a_{\ell m} a_{\ell' m'}^* \rangle = \delta_{\ell \ell'} \delta_{m m'} C_{\ell}$$

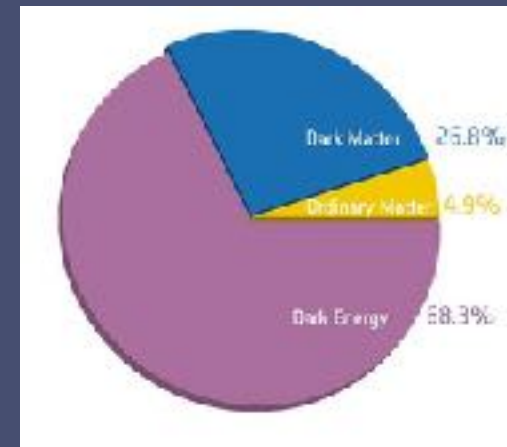
十分統計量

1pt probability density



ほぼ正規分布 (非ガウス性<0.1%) ell_max = 2,500まで解析

→ 2,500^2 ~ 数百万 モード



CMBの理論予言

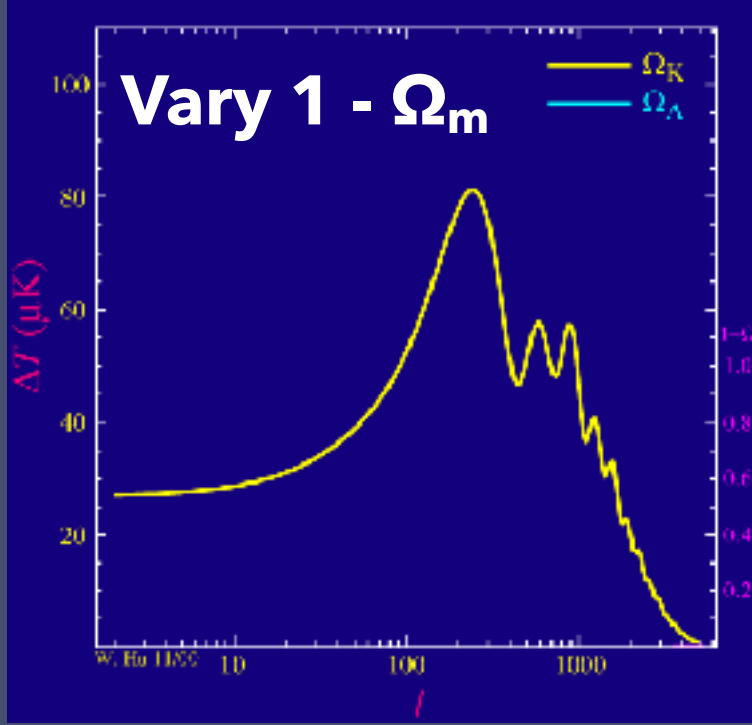
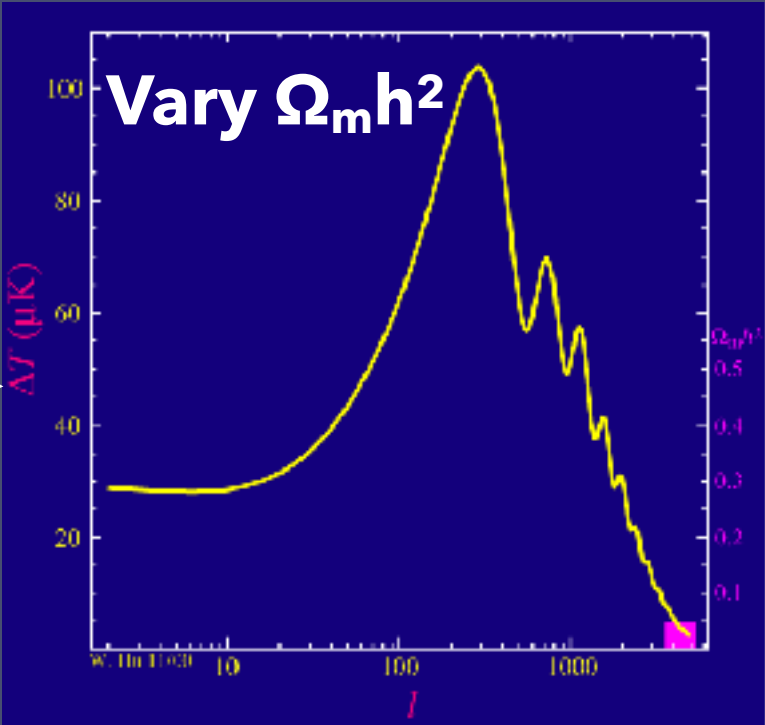
Bond & Efstathiou '84

TABLE 1
COLD DARK MATTER FLUCTUATION PARAMETERS

Ω (1)	Ω_B (2)	h_0 (3)	$a(\text{Mpc})$ (4)	$b(\text{Mpc})$ (5)	$c(\text{Mpc})$ (6)	ν (7)	θ_c (arc min) (8)	β (9)	$C(0)^{1/2} \times 10^5$ (10)	$\Delta T/T \times 10^5$ (11)	$J_3(x_0)$ ($h^{-3} \text{Mpc}^3$) (12) (13)		z_{nl} (14)	z_{reh} (15)	Ω_{BC} (16)	
1.0	0.03	0.75	11.3	5.29	3.10	1.13	8.4	0.94	1.2	0.28	210	340	30	117	1.9×10^{-6}	
1.0	0.03	0.50	23.1	11.4	6.48	1.25	8.8	1.25	2.8	0.42	190	380	20	151	6.9×10^{-15}	
0.4	0.03	0.75	23.5	17.3	7.15	1.07	5.2	0.90	3.0	1.1	180	450	25	86	1.2×10^{-5}	
0.3	0.03	0.75	37.1	21.1	10.8	1.12	4.5	0.88	4.4	1.8	170	470	20	79	4.2×10^{-6}	
0.2	0.03	0.75	48.2	31.6	15.6	1.30	3.7	0.94	8.0	3.6	160	520	19	70	1.9×10^{-5}	
0.2	0.10	0.75	76.2	71.1	27.7	2.15	5.2	1.82	22.0	7.9	150	580	15	32	2.6×10^{-3}	
0.2	0.03	1.00	37.9	18.4	10.1	1.21	3.9	0.75	5.4	2.3	170	490	23	58	2.7×10^{-4}	
0.2	0.03	0.50	207.0	47.9	38.5	1.73	3.6	1.22	15.6	7.2	160	530	24	91	1.1×10^{-5}	
CfA Survey												180	550			
$x_0 h \text{ Mpc}$												7.5	15			

CMBfast
CAMB
CLASS

...



CMBの理論予言

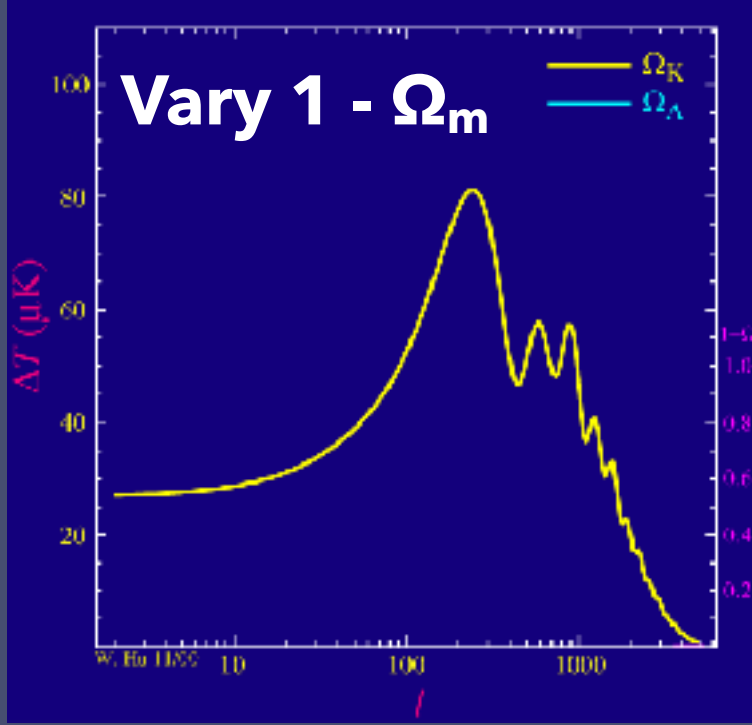
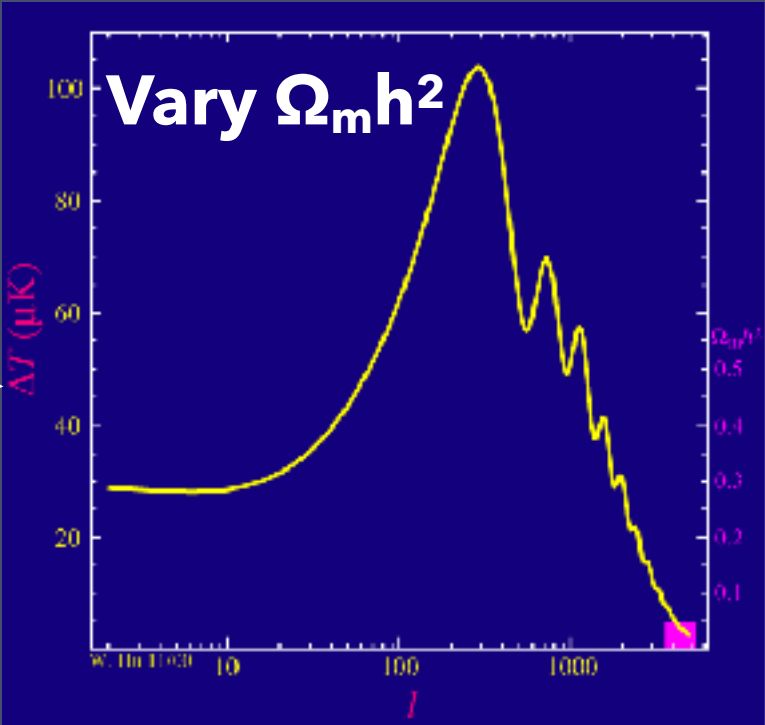
Bond & Efstathiou '84

TABLE 1
COLD DARK MATTER FLUCTUATION PARAMETERS

Ω (1)	Ω_B (2)	h_0 (3)	$a(\text{Mpc})$ (4)	$b(\text{Mpc})$ (5)	$c(\text{Mpc})$ (6)	ν (7)	θ_c (arc min) (8)	β (9)	$C(0)^{1/2} \times 10^5$ (10)	$\Delta T/T \times 10^5$ (11)	$J_3(x_0)$ ($h^{-3} \text{Mpc}^3$) (12)	(13)	z_{nl} (14)	z_{reh} (15)	Ω_{BC} (16)	
1.0	0.03	0.75	11.3	5.29	3.10	1.13	8.4	0.94	1.2	0.28	210	340	30	117	1.9×10^{-6}	
1.0	0.03	0.50	23.1	11.4	6.48	1.25	8.8	1.25	2.8	0.42	190	380	20	151	6.9×10^{-15}	
0.4	0.03	0.75	23.5	17.3	7.15	1.07	5.2	0.90	3.0	1.1	180	450	25	86	1.2×10^{-5}	
0.3	0.03	0.75	37.1	21.1	10.8	1.12	4.5	0.88	4.4	1.8	170	470	20	79	4.2×10^{-6}	
0.2	0.03	0.75	48.2	31.6	15.6	1.30	3.7	0.94	8.0	3.6	160	520	19	70	1.9×10^{-5}	
0.2	0.10	0.75	76.2	71.1	27.7	2.15	5.2	1.82	22.0	7.9	150	580	15	32	2.6×10^{-3}	
0.2	0.03	1.00	37.9	18.4	10.1	1.21	3.9	0.75	5.4	2.3	170	490	23	58	2.7×10^{-4}	
0.2	0.03	0.50	207.0	47.9	38.5	1.73	3.6	1.22	15.6	7.2	160	530	24	91	1.1×10^{-5}	
CfA Survey												180	550			
$x_0 h \text{ Mpc}$												7.5	15			

CMBfast
CAMB
CLASS

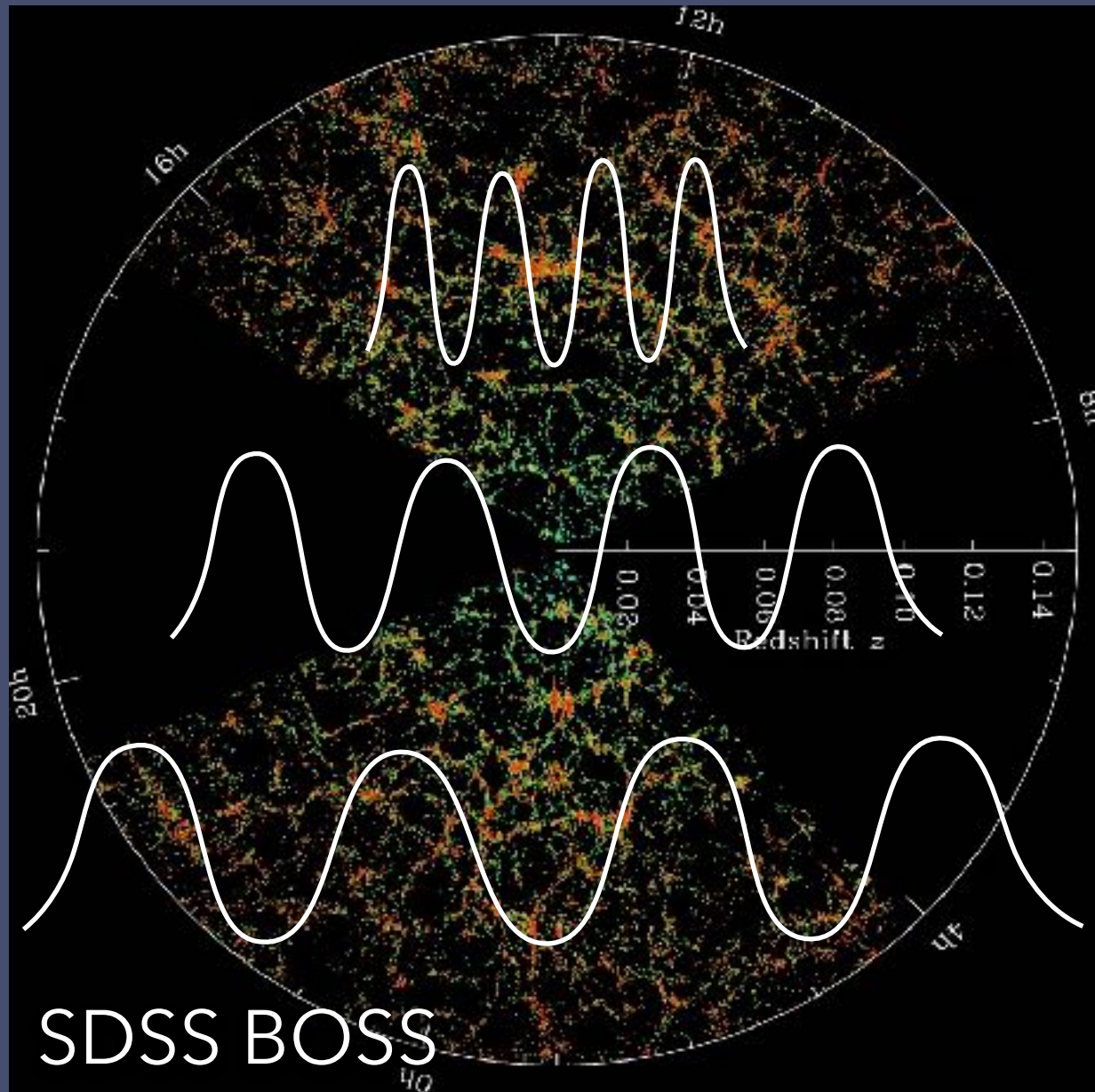
...



LSSの持つ情報

Galaxy clustering

3次元的点過程

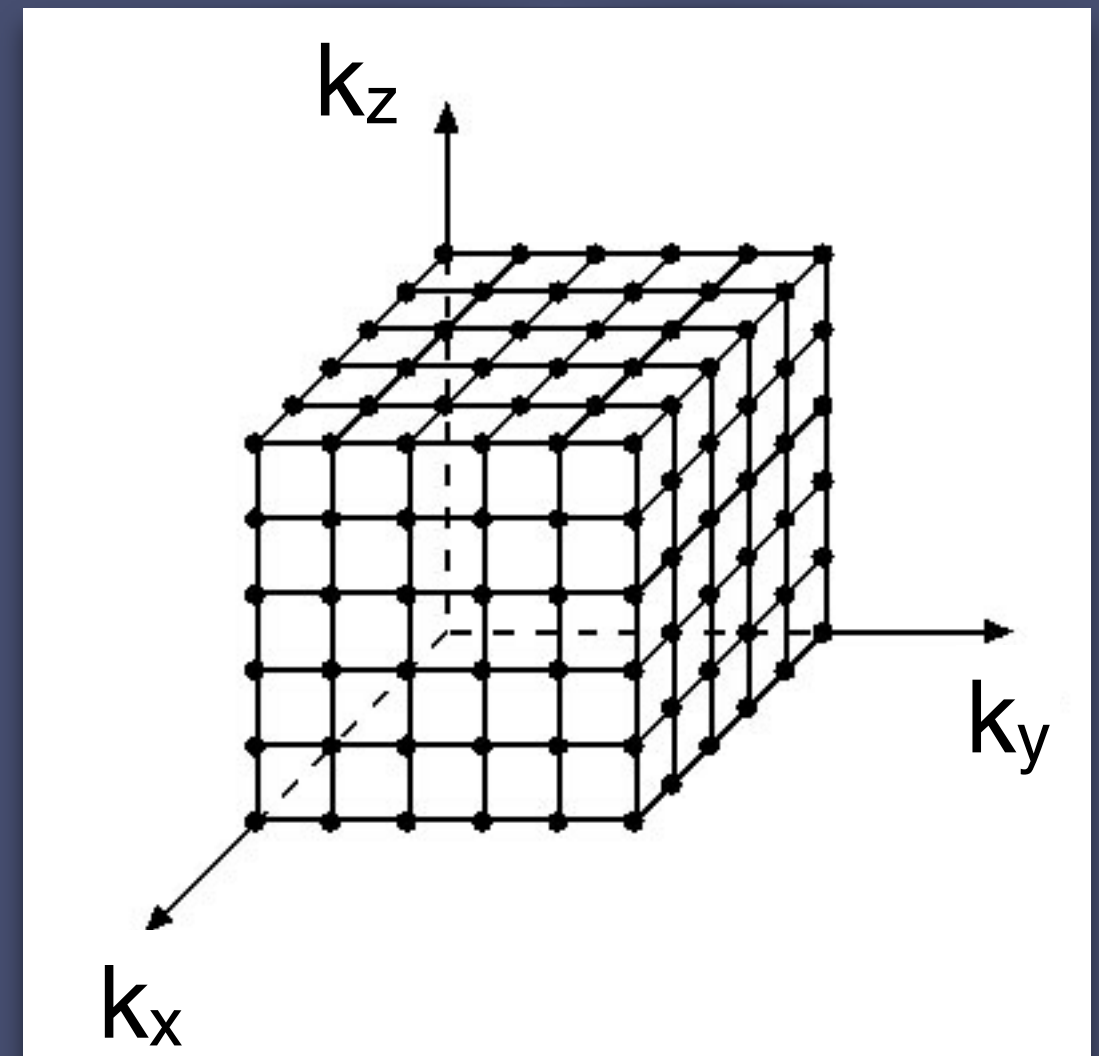
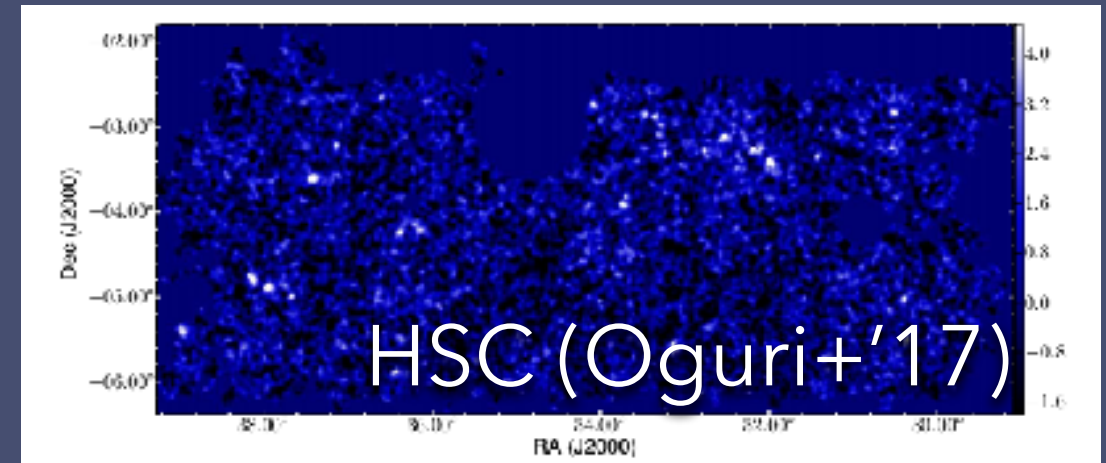


$2\pi/V^{1/3}$ を単位とした離散的なフーリエ
モードの数がナイーブな情報量

西道 啓博 (基研)

Weak gravitational lensing

射影された密度マップ



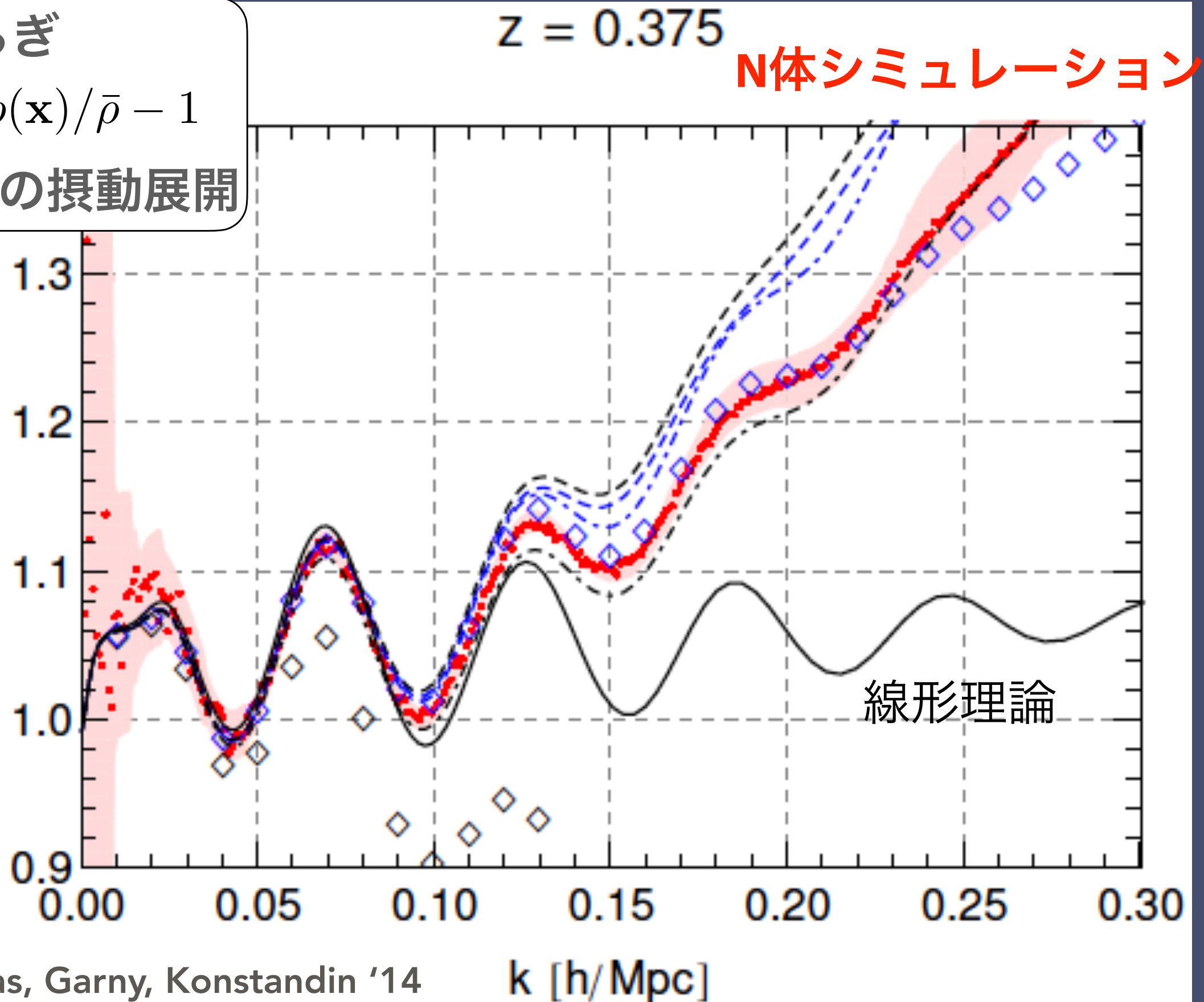
非線型性をモデリングできるか？

密度揺らぎ

$$\delta(\mathbf{x}) = \rho(\mathbf{x})/\bar{\rho} - 1$$

についての摂動展開

規格化されたパワースペクトル



Blas, Garny, Konstandin '14

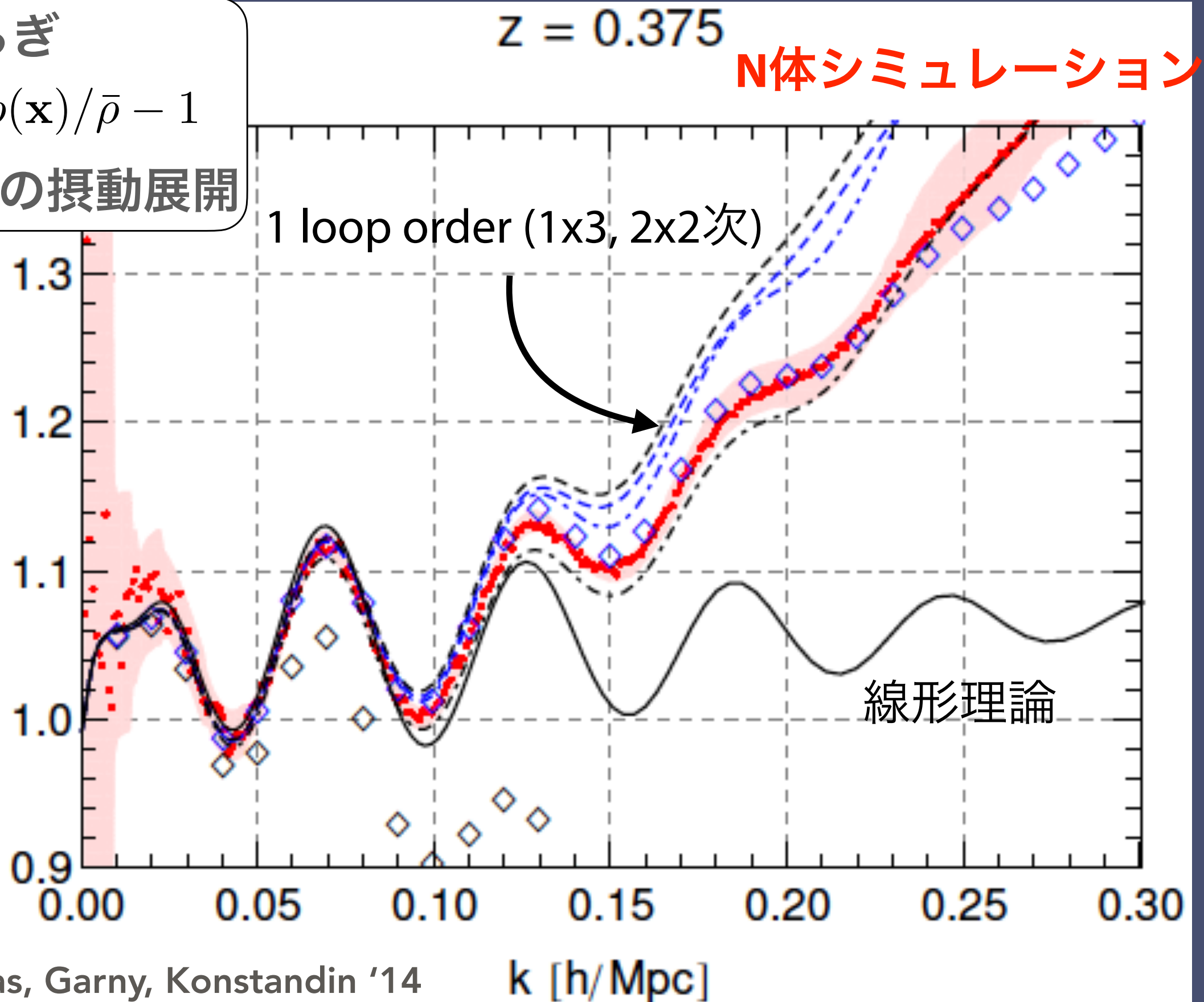
非線型性をモデリングできるか？

密度揺らぎ

$$\delta(\mathbf{x}) = \rho(\mathbf{x})/\bar{\rho} - 1$$

についての摂動展開

規格化されたパワースペクトル



Blas, Garny, Konstandin '14

k [h/Mpc]

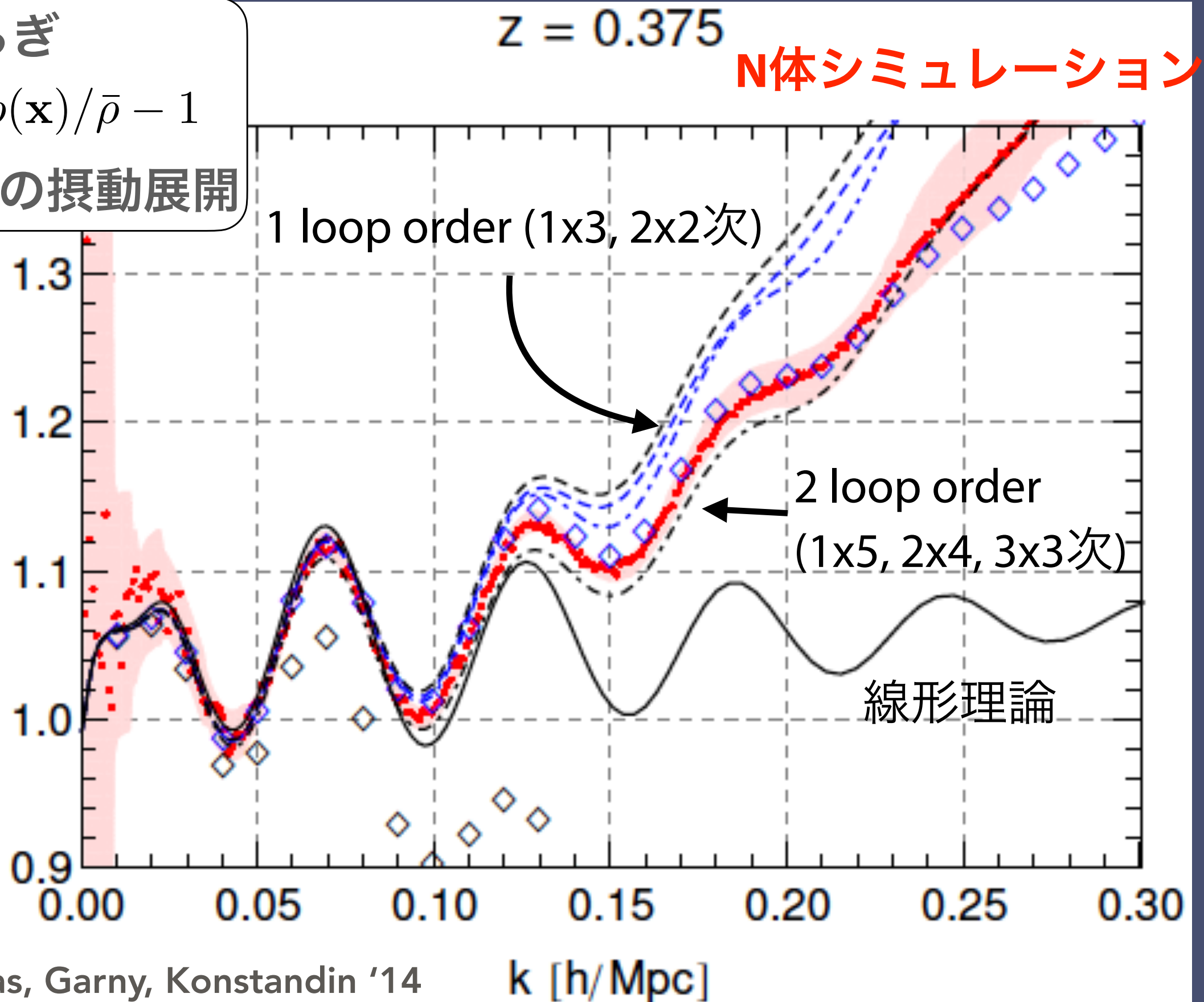
非線型性をモデリングできるか？

密度揺らぎ

$$\delta(\mathbf{x}) = \rho(\mathbf{x})/\bar{\rho} - 1$$

についての摂動展開

規格化されたパワースペクトル



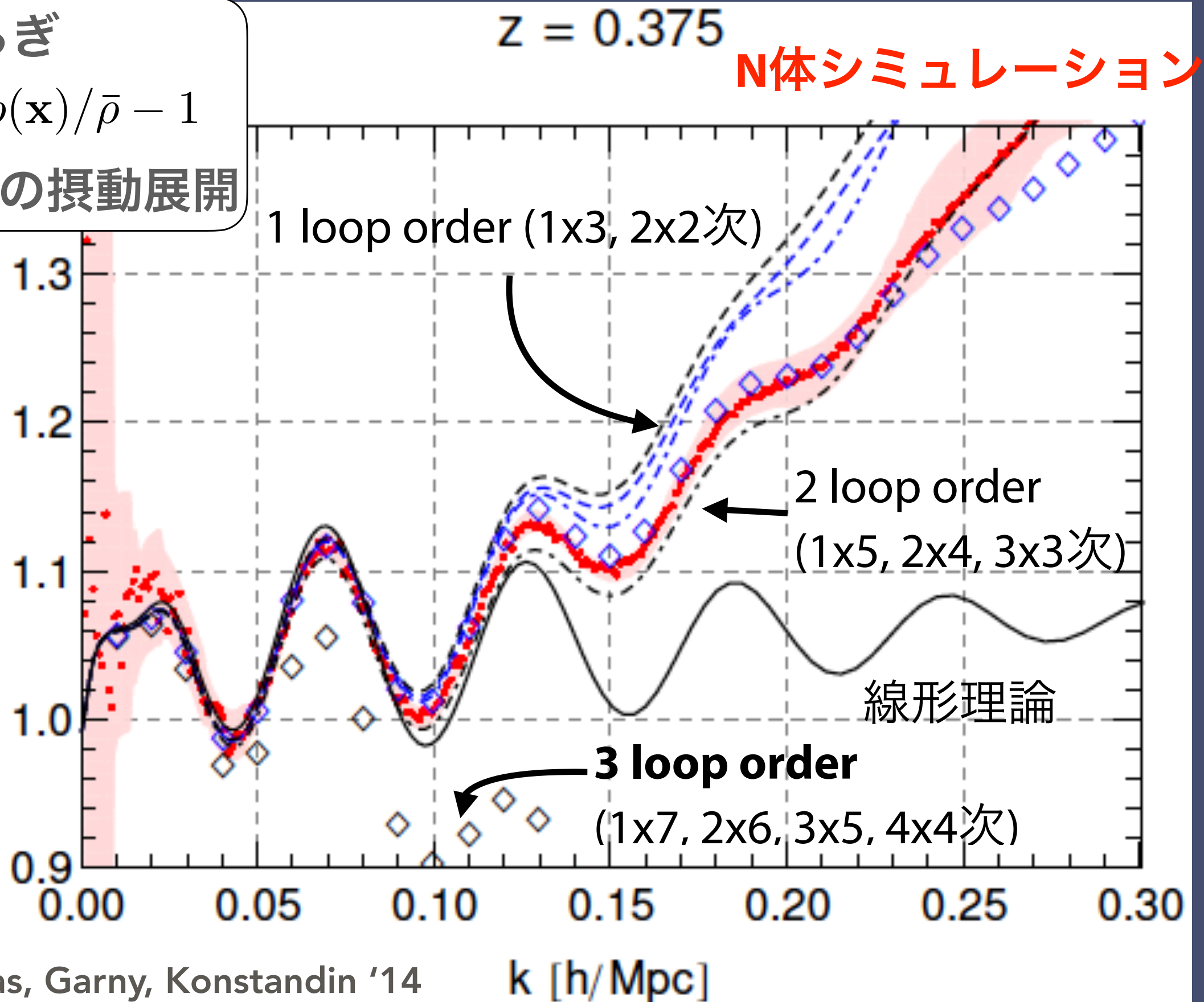
非線型性をモデリングできるか？

密度揺らぎ

$$\delta(\mathbf{x}) = \rho(\mathbf{x})/\bar{\rho} - 1$$

についての摂動展開

規格化されたパワースペクトル



Blas, Garny, Konstandin '14

非線型性をモデリングできるか？

密度揺らぎ

$$\delta(\mathbf{x}) = \rho(\mathbf{x})/\bar{\rho} - 1$$

についての摂動展開

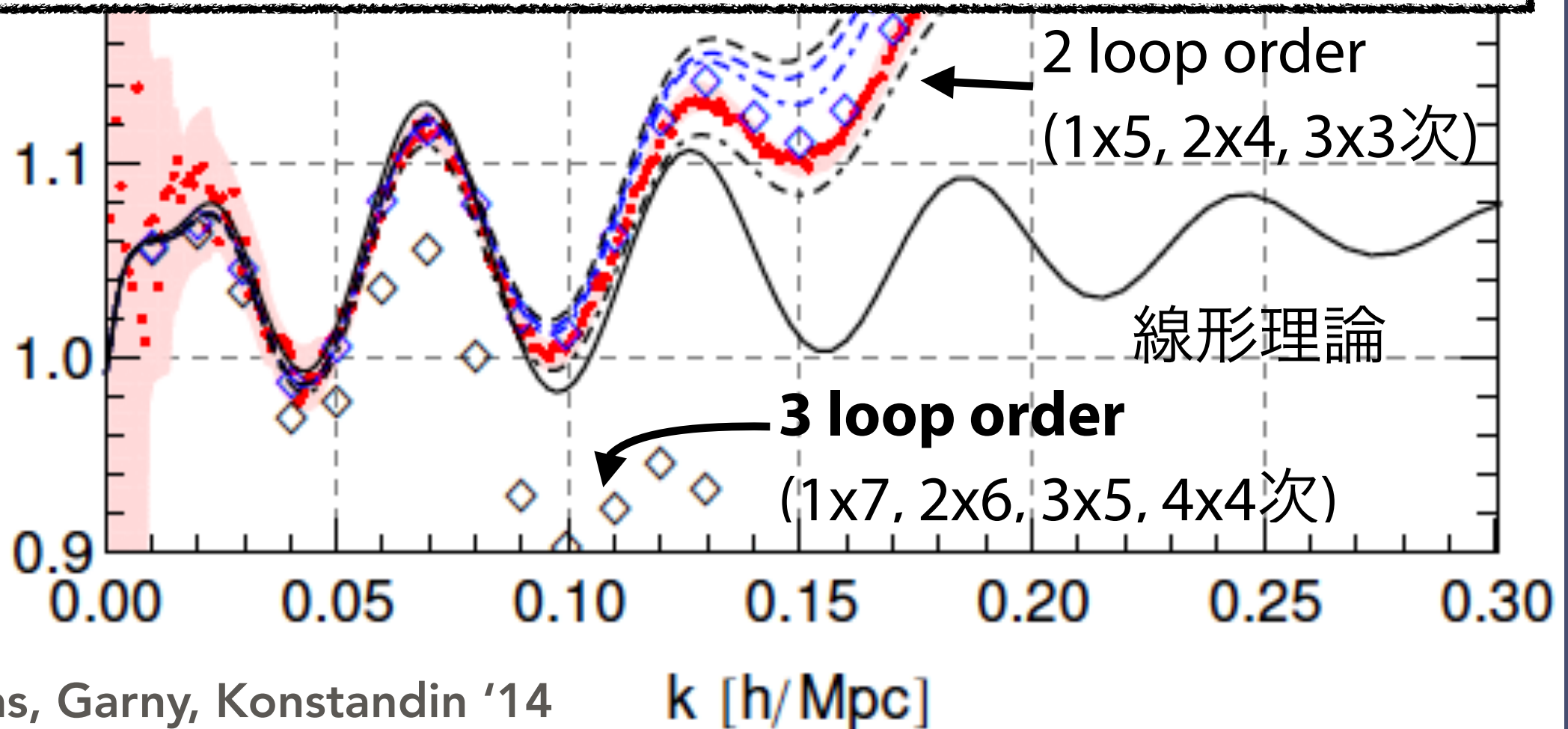
$z = 0.375$

N体シミュレーション

1 loop order (1x3, 2x2次)

2 loopオーダーの成功は偶然か？

規格化されたパワースペクトル



Blas, Garny, Konstandin '14

システムレベルでの「応答関数」

TN, Bernardeau, Taruya '16, 17, Halle, TN, Taruya Colombi, Bernardeau '20

$$K(k, q) = q \frac{\delta P_{nl}(k)}{\delta P_{lin}(q)}$$

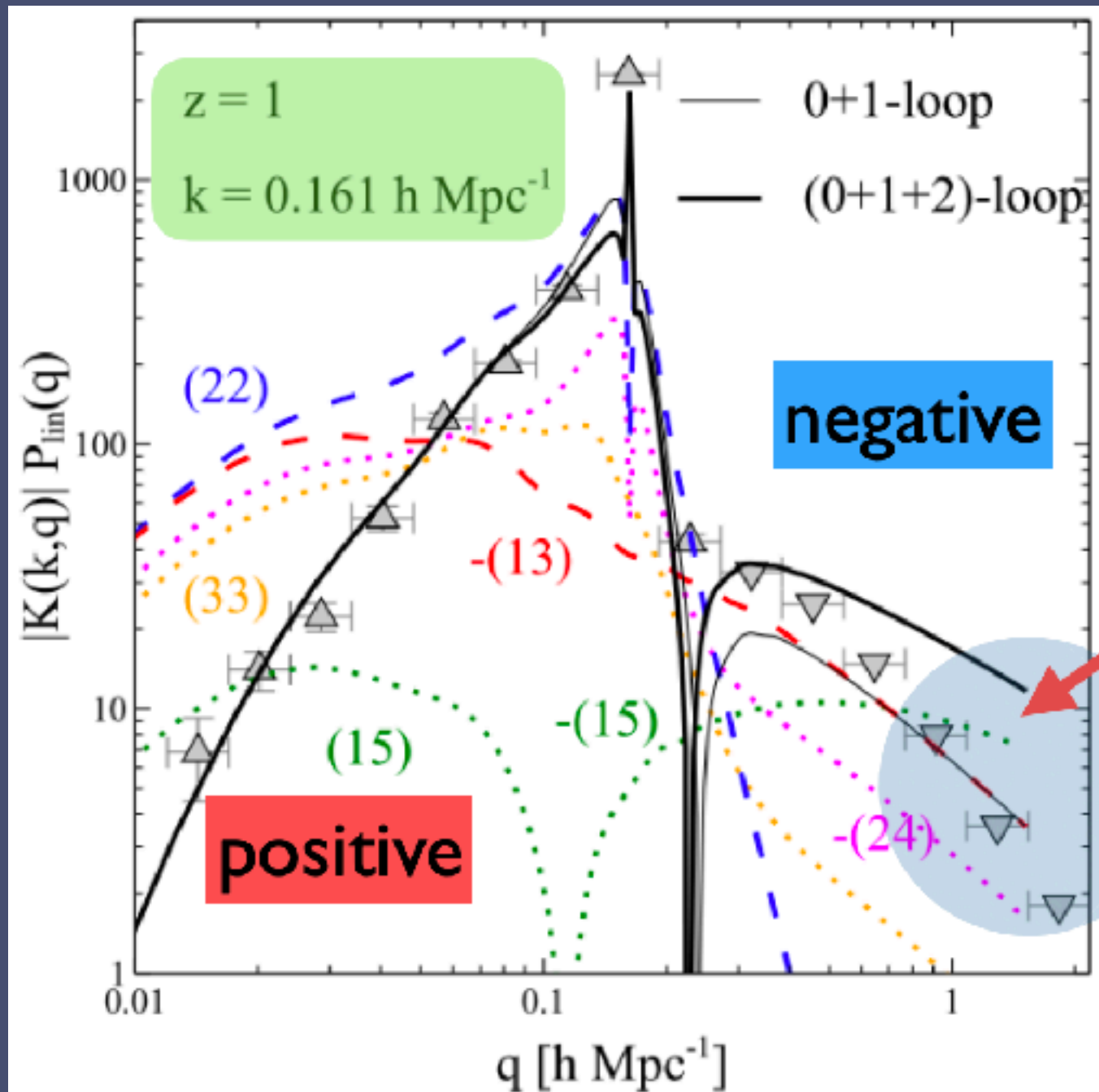
I want to study this mode at some late time t



応答関数: シミュレーション vs 摂動論

TN, Bernardeau, Taruya '16, 17, Halle, TN, Taruya Colombi, Bernardeau '20

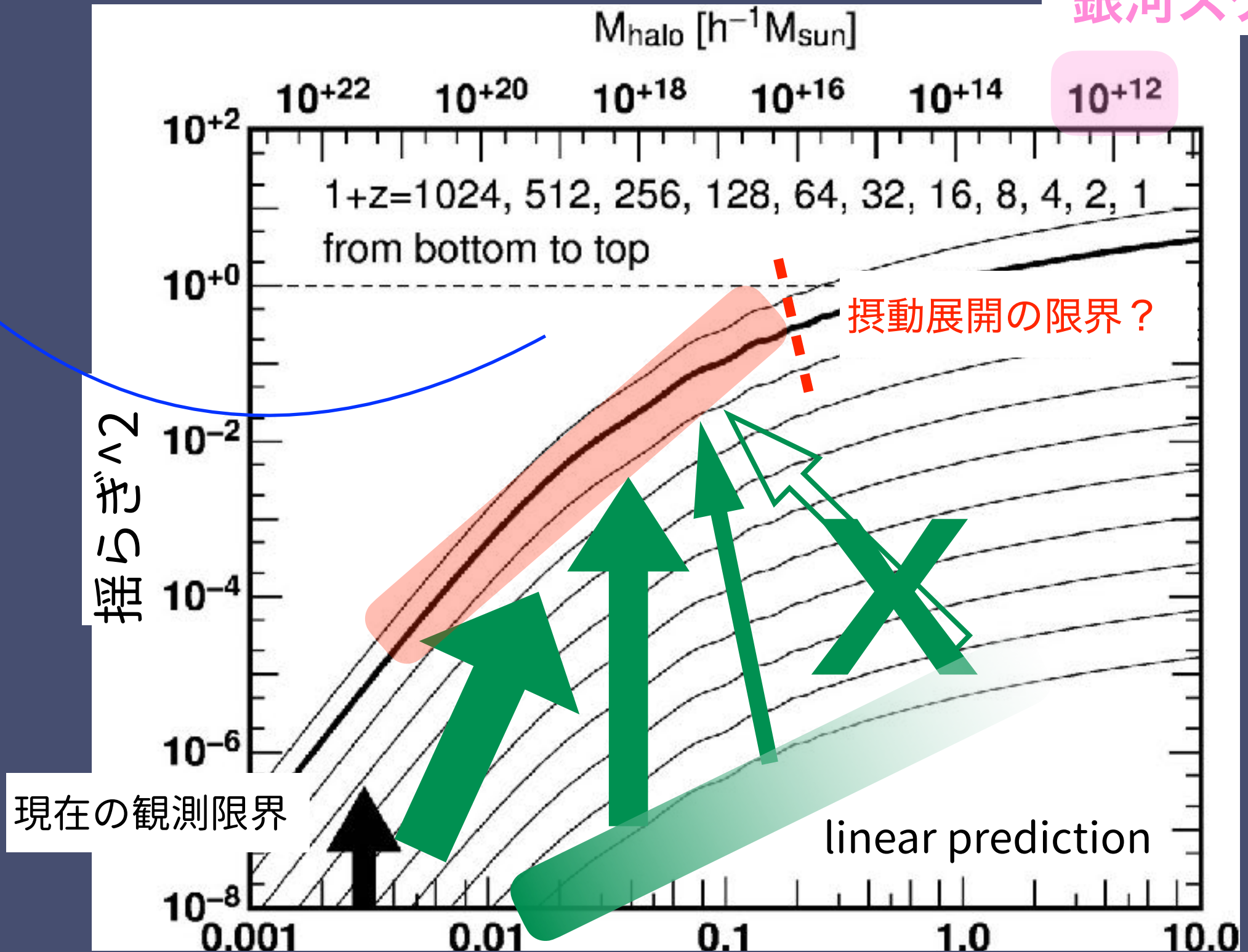
$$K(k, q) = q \frac{\delta P_{nl}(k)}{\delta P_{lin}(q)}$$



- 大→小スケールへの伝搬
- 全体的な傾向は摂動論が良くとらえている
- 摂動論 >> N体 @ high q
- ここで摂動展開が破れる
 - 小スケール(UV)の揺らぎは大スケール(IR)に伝播しない
- **multi-streaming領域の物理**
 - 何らかの正則化が必要

摂動論の限界

銀河スケール



長波長

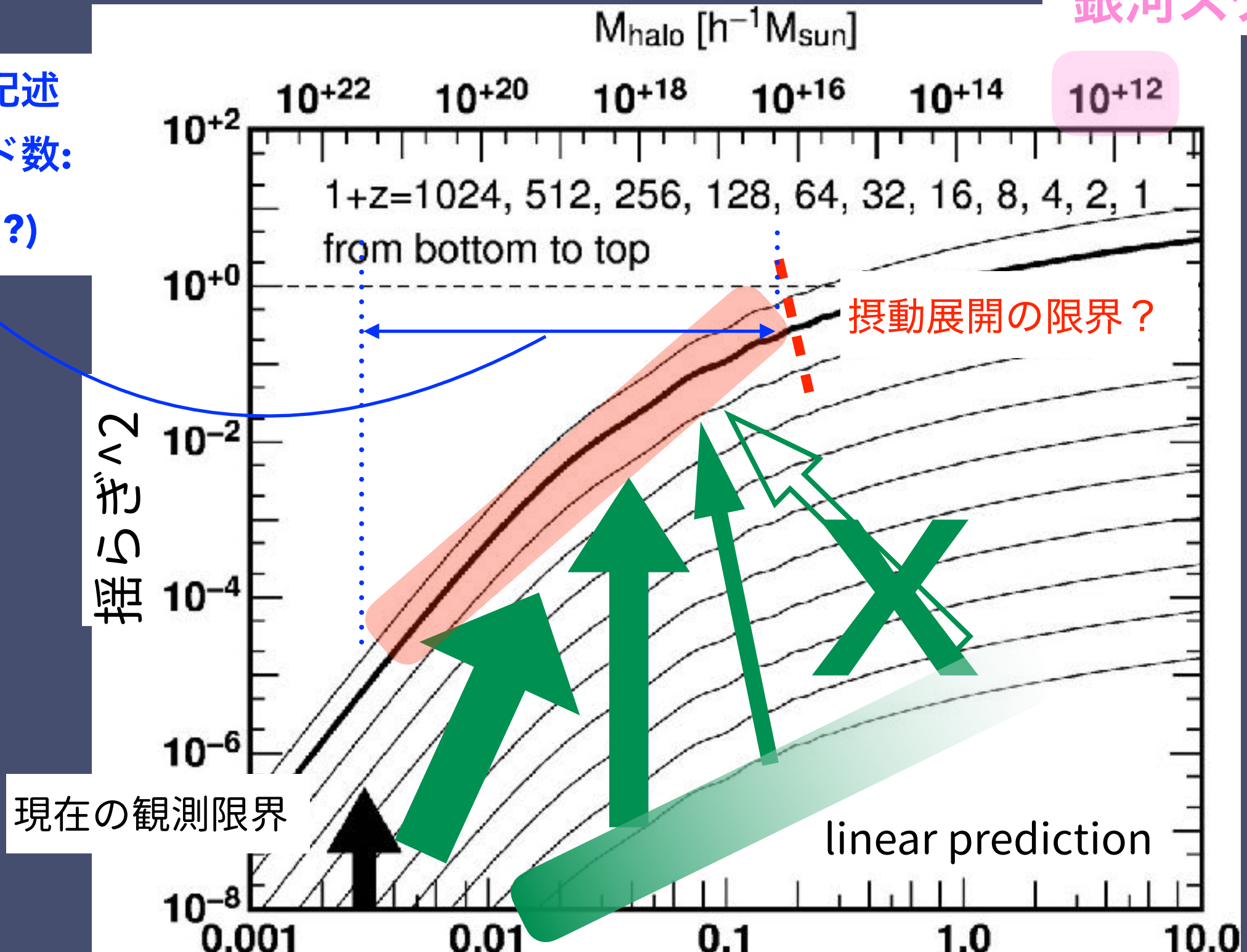
波数 $[h/\text{Mpc}]$

短波長

摂動論の限界

銀河スケール

ロバストに記述
できるモード数:
 $\sim 10^{(5-6)}$ (?)



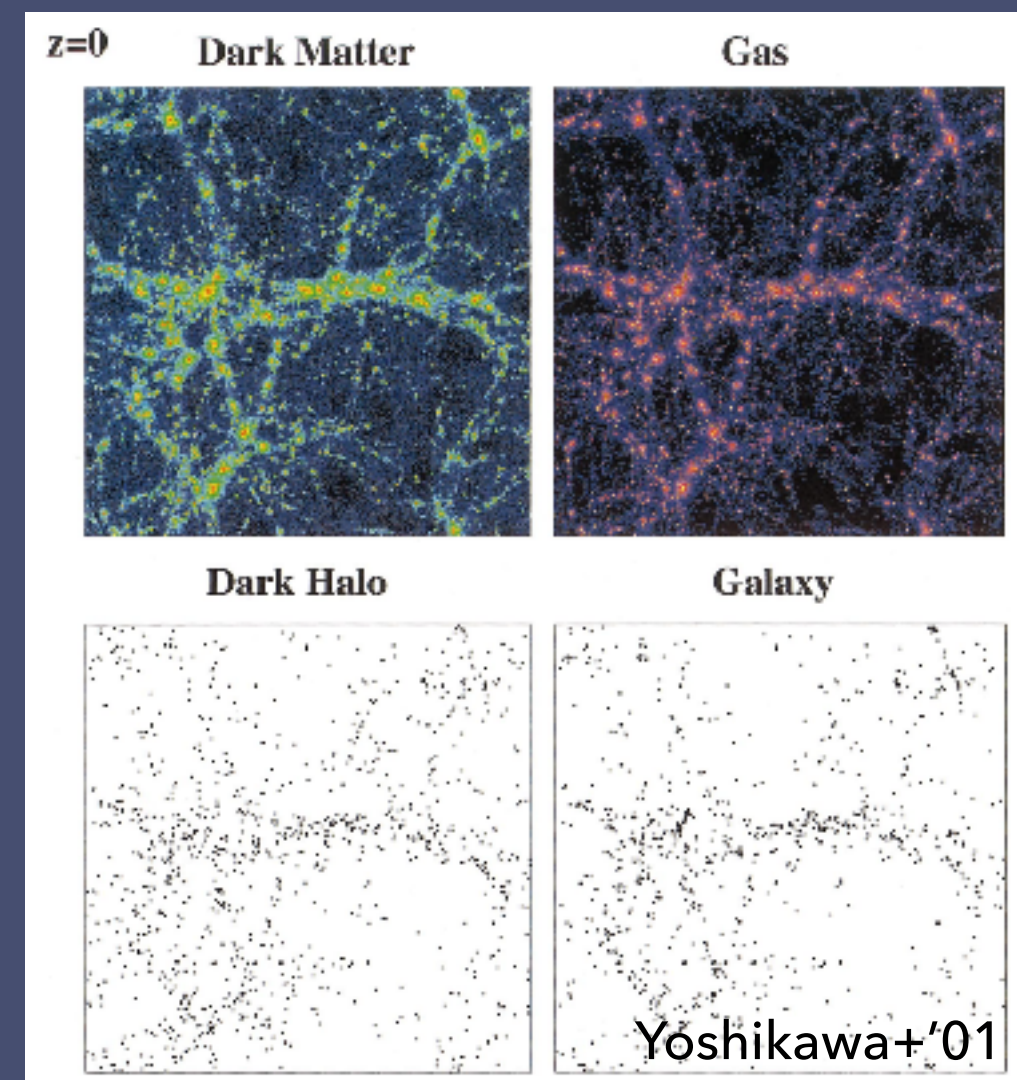
長波長

短波長

波数 $[h/\text{Mpc}]$

銀河バイアス

- 光を放つ「天体」は背後の密度場の「**トレーサー**」に過ぎない (Kaiser '84)
- 銀河の分布を第一原理的に予言するのは難しい (but hydro sims)
- 簡便化した何らかのレシピを入れて多数の攪乱パラメタを入れるしかない？



The Galaxy Power Spectrum and Bispectrum in Redshift Space

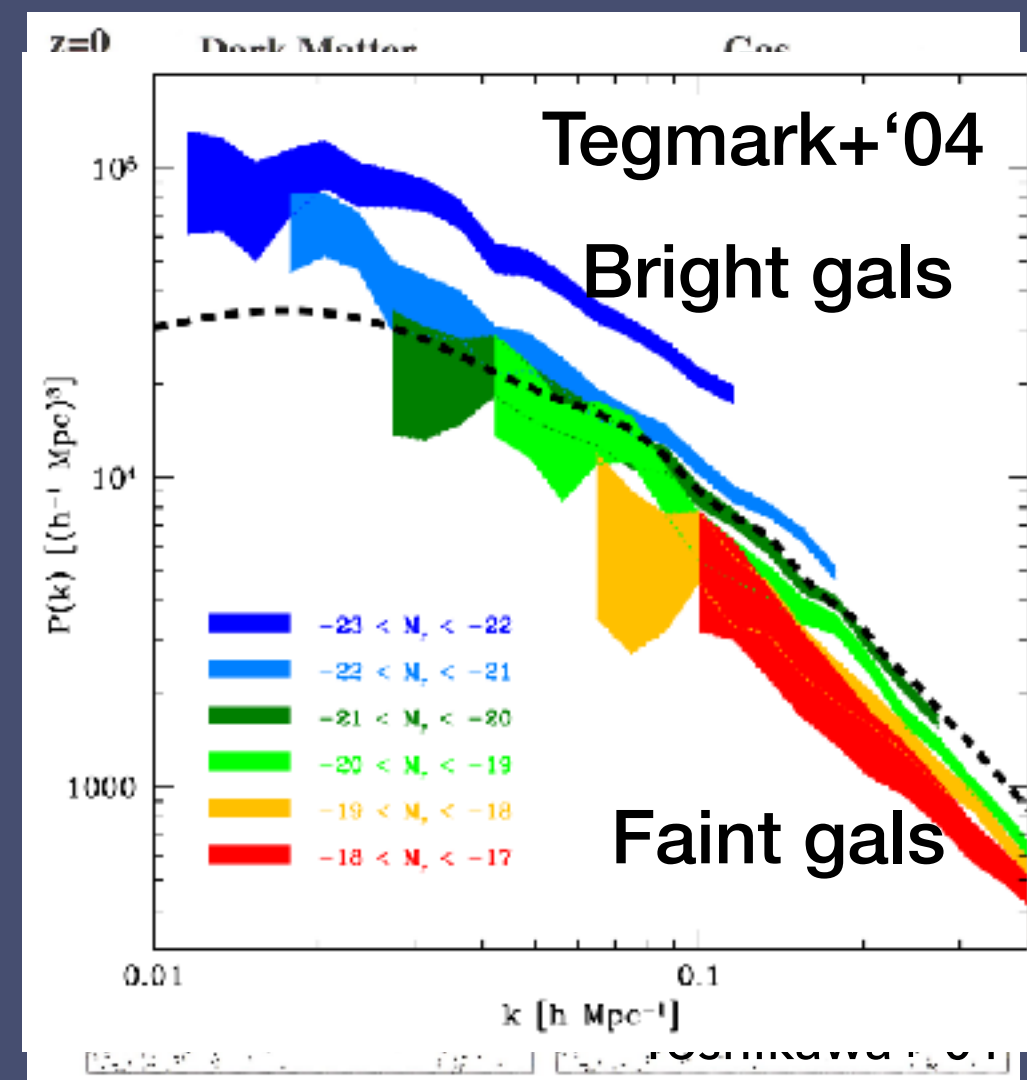
Vincent Desjacques, Donghui Jeong, Fabian Schmidt

(Submitted on 11 Jun 2018)

We present the complete expression for the next-to-leading (1-loop) order galaxy power spectrum and the leading-order galaxy bispectrum in redshift space in the general bias expansion, or equivalently the effective field theory of biased tracers. We consistently include selection effects. These are degenerate with many, but not all, of the redshift-space distortion contributions, and have not been included before. Moreover, we show that, in the framework of effective field theory, a consistent bias expansion in redshift space requires selection contributions. Physical arguments about the tracer sample considered and its observational selection have to be used to constrain the selection contributions. In summary, the next-to-leading order galaxy power spectrum and leading-order galaxy bispectrum in redshift space are described by 22 parameters, which reduces to 11 parameters if selection effects can be neglected. All contributions to the power spectrum and bispectrum are expressed in terms of 28 independent loop integrals.

銀河バイアス

- 光を放つ「天体」は背後の密度場の「**トレーサー**」に過ぎない (Kaiser '84)
- 銀河の分布を第一原理的に予言するのは難しい (but hydro sims)
- 簡便化した何らかのレシピを入れて多数の攪乱パラメタを入れるしかない？



The Galaxy Power Spectrum and Bispectrum in Redshift Space

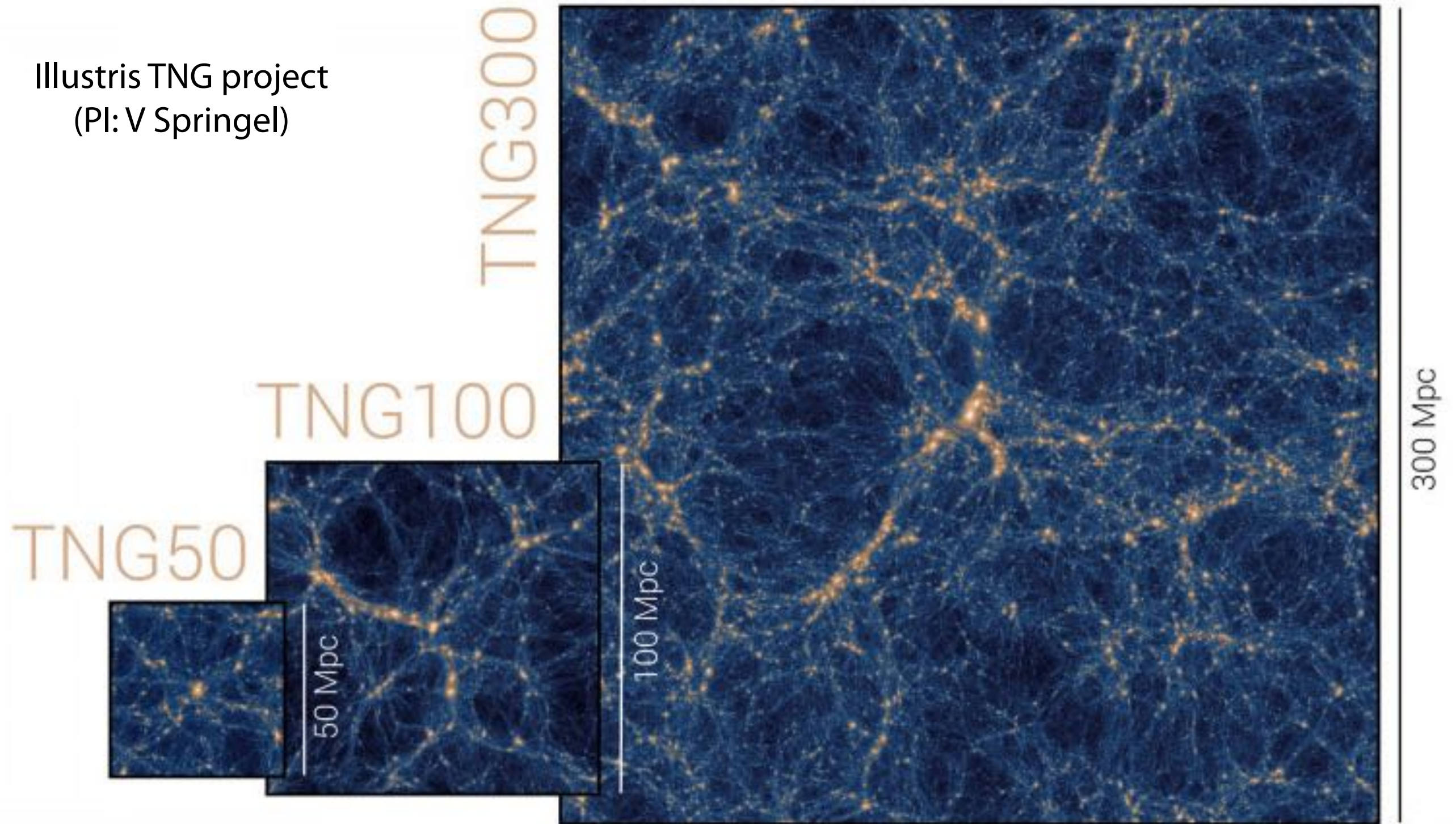
Vincent Desjacques, Donghui Jeong, Fabian Schmidt

(Submitted on 11 Jun 2018)

We present the complete expression for the next-to-leading (1-loop) order galaxy power spectrum and the leading-order galaxy bispectrum in redshift space in the general bias expansion, or equivalently the effective field theory of biased tracers. We consistently include selection effects. These are degenerate with many, but not all, of the redshift-space distortion contributions, and have not been included before. Moreover, we show that, in the framework of effective field theory, a consistent bias expansion in redshift space must include selection contributions. Physical arguments about the tracer sample considered and its observational selection have to be used to fix the selection contributions. In summary, the next-to-leading order galaxy power spectrum and leading-order galaxy bispectrum in redshift space are described by 22 parameters, which reduces to 11 parameters if selection effects can be neglected. All contributions to the power spectrum and bispectrum are expressed in terms of 28 independent loop integrals.

宇宙論的流体計算のstate of the art

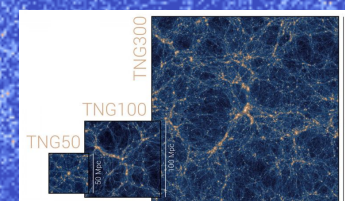
Illustris TNG project
(PI: V Springel)



宇宙

About the size of the Subaru PFS survey

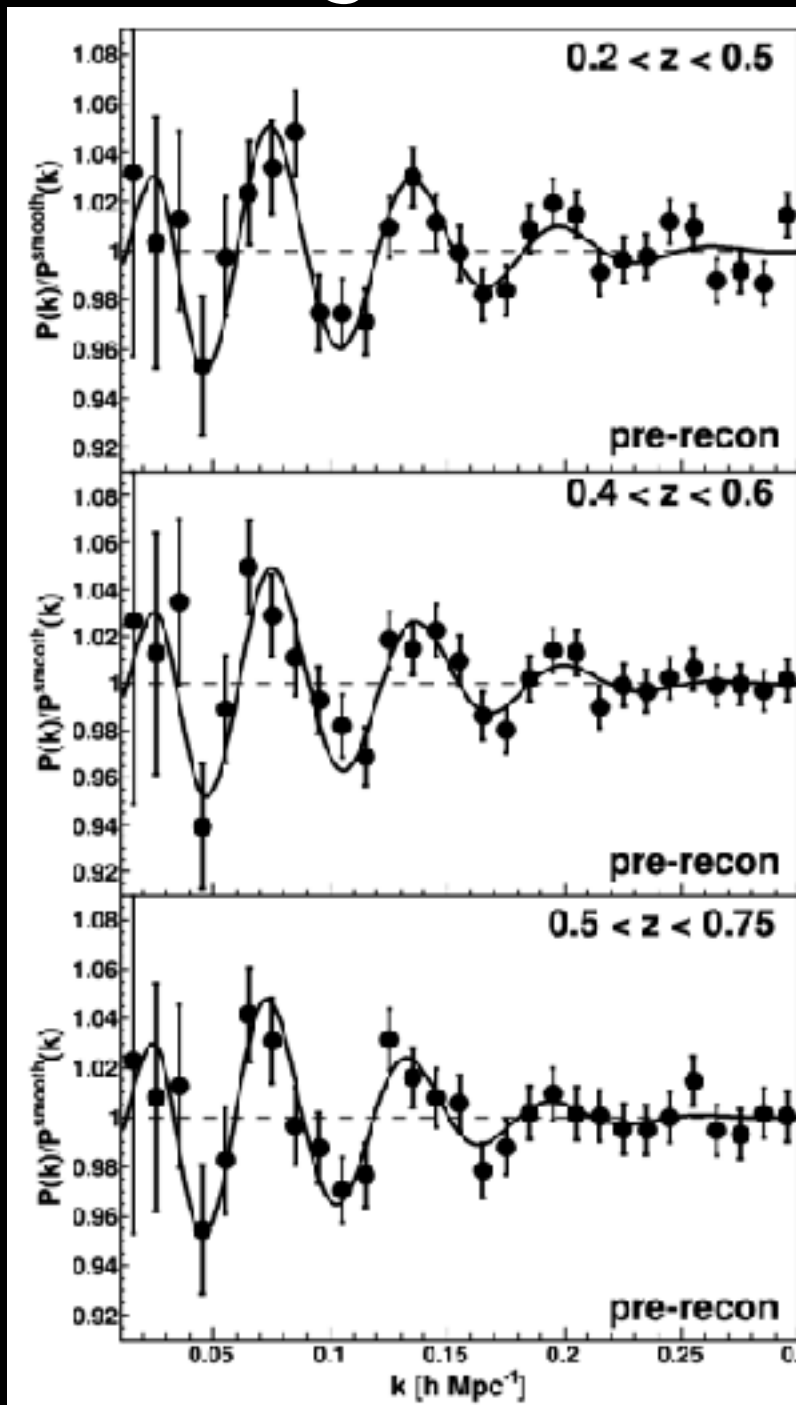
art



Baryon Acoustic Oscillation

SDSS III BOSS; Beutler+'17

Power spectrum of
~1.2M galaxies



Imprint in the nearby
Universe

CMB acoustic scale
Photon vs baryon
(pressure and gravity)

Robust probe of cosmic expansion

©Gen Chiaki, Atsushi Taruya

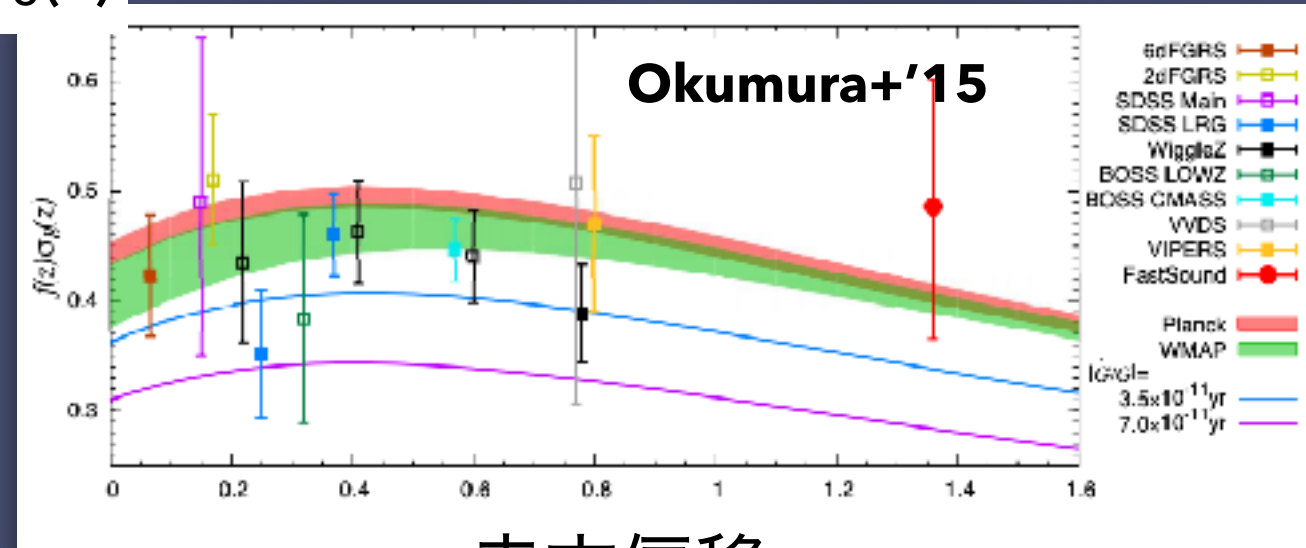
赤方偏移歪み

z-space r-space

$$\mathbf{z} = \mathbf{r} + \frac{u_z}{aH} \hat{\mathbf{z}}$$

peculiar velocity

$$f(z)\sigma_8(z)$$



赤方偏移

Large-scales



- 視線方向が特別な方向
- 視線方向との成す角のコサインに対する2次及び4次の依存性
- ほんの一部の情報しか解釈できない

$$P(k, \mu) = P_{gg}(k) + 2f\mu^2 P_{g\theta}(k) + f^2\mu^4 P_{\theta\theta}(k)$$

$$\propto (b\sigma_8)^2$$

$$\propto (b\sigma_8)(f\sigma_8)$$

$$\propto (f\sigma_8)^2$$

EFT-of-LSS

Baumann+'12, Carrasco, Herzberg, Senatore'12 ...

continuity + Euler + Poisson eqs.

$$\begin{aligned} \frac{\partial \delta}{\partial t} + \frac{1}{a} \nabla \cdot [(1 + \delta) \mathbf{v}] &= 0, \\ \frac{\partial \mathbf{v}}{\partial t} + H \mathbf{v} + \frac{1}{a} (\mathbf{v} \cdot \nabla) \mathbf{v} &= -\frac{1}{a} \nabla \phi - \frac{1}{\rho_m} \frac{1}{a} \nabla \tau_{ij} \\ \nabla^2 \phi &= 4\pi G \rho a^2 \delta. \end{aligned}$$

- 標準摂動論は渦なし圧力なし完全流体を仮定

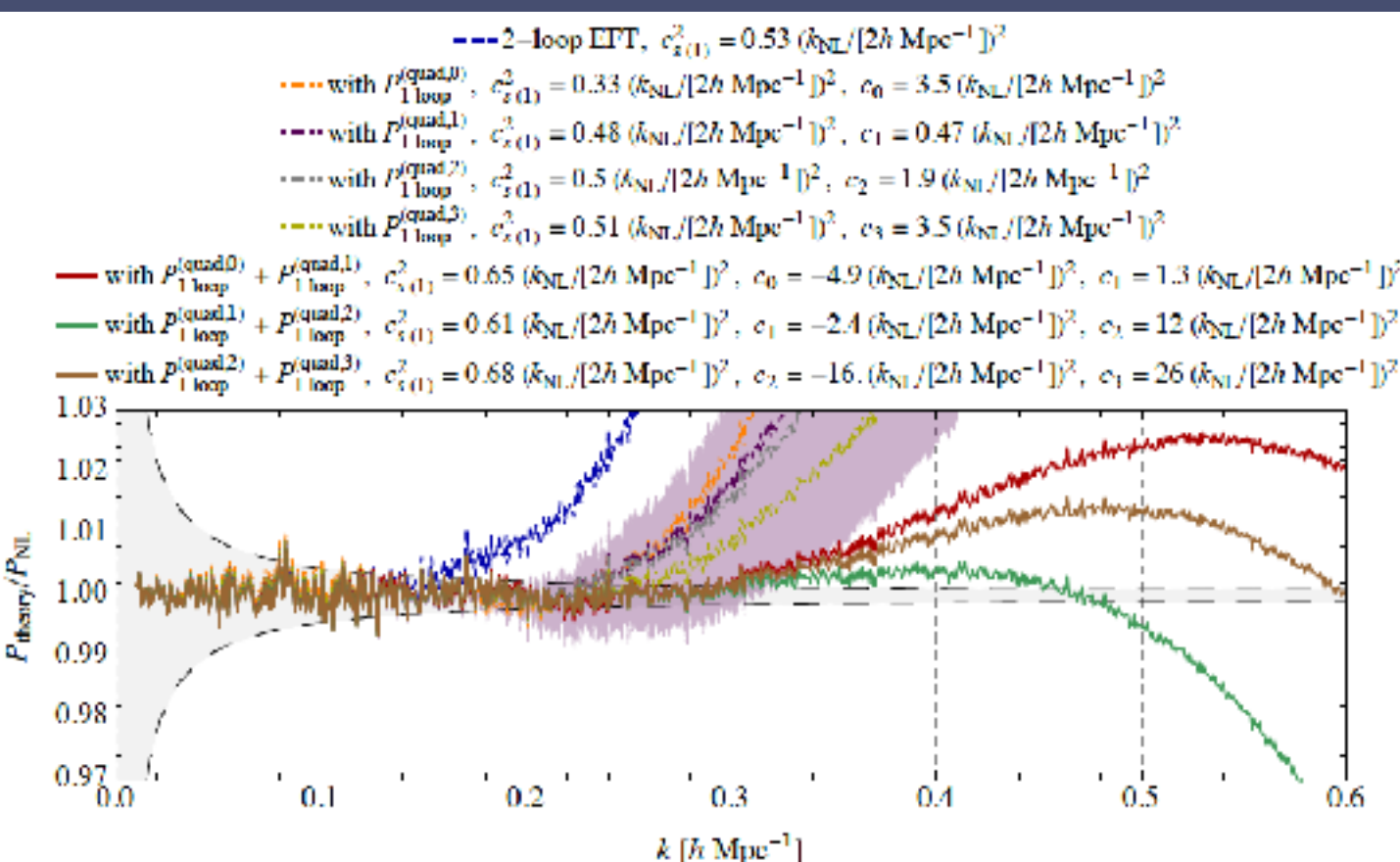
- CDMが支配する宇宙では少なくともhigh-zでは良い近似

- シェルクロッシング？

- 応力テンソル由来の counterterm を展開次数毎に系統的に導入

- 自由係数 (実効音速等) はシミュレーションか観測データからfix

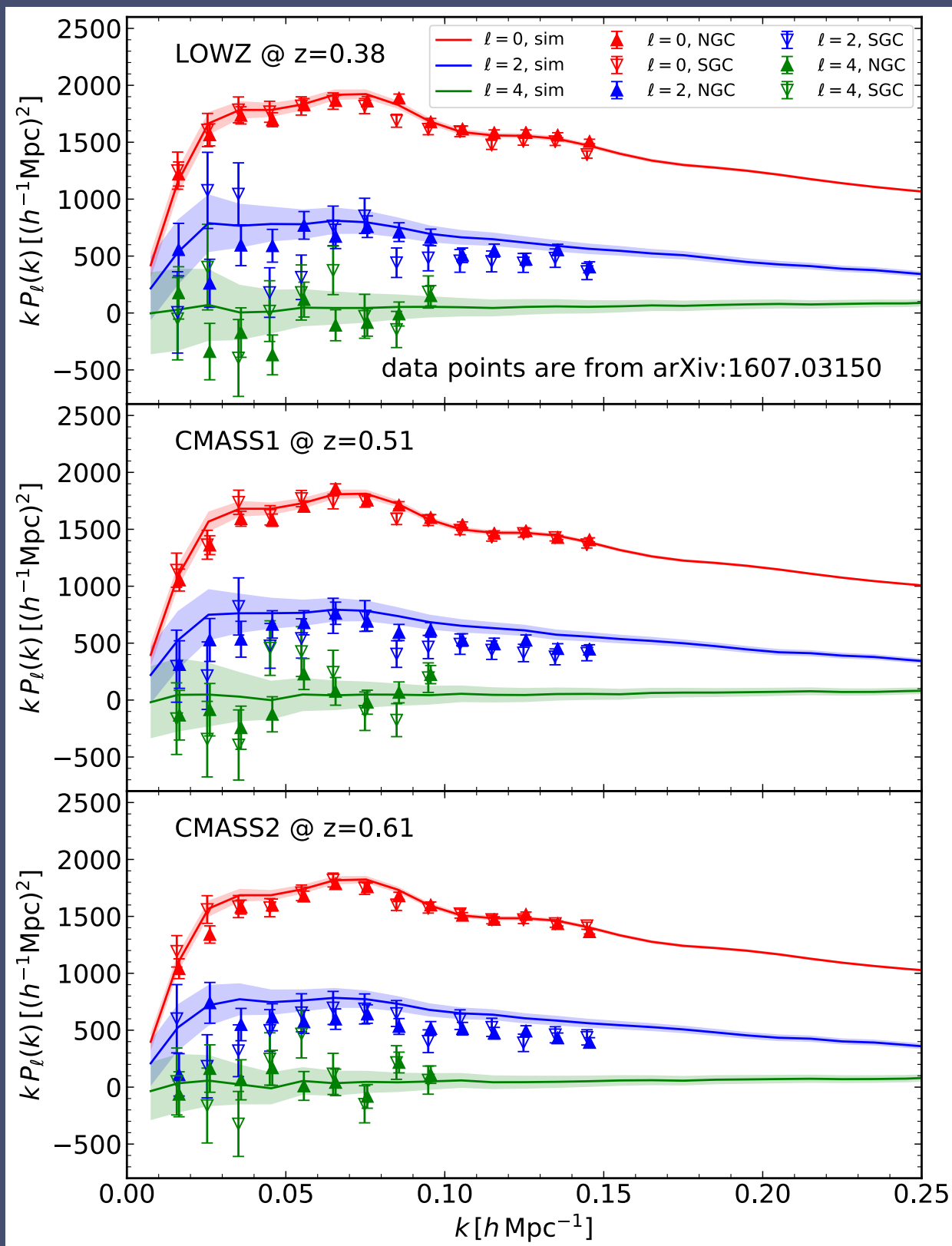
- 銀河のバイアス展開はある意味既にEFT



Foreman, Perrier, Senatore'16

Blinded cosmology challenge

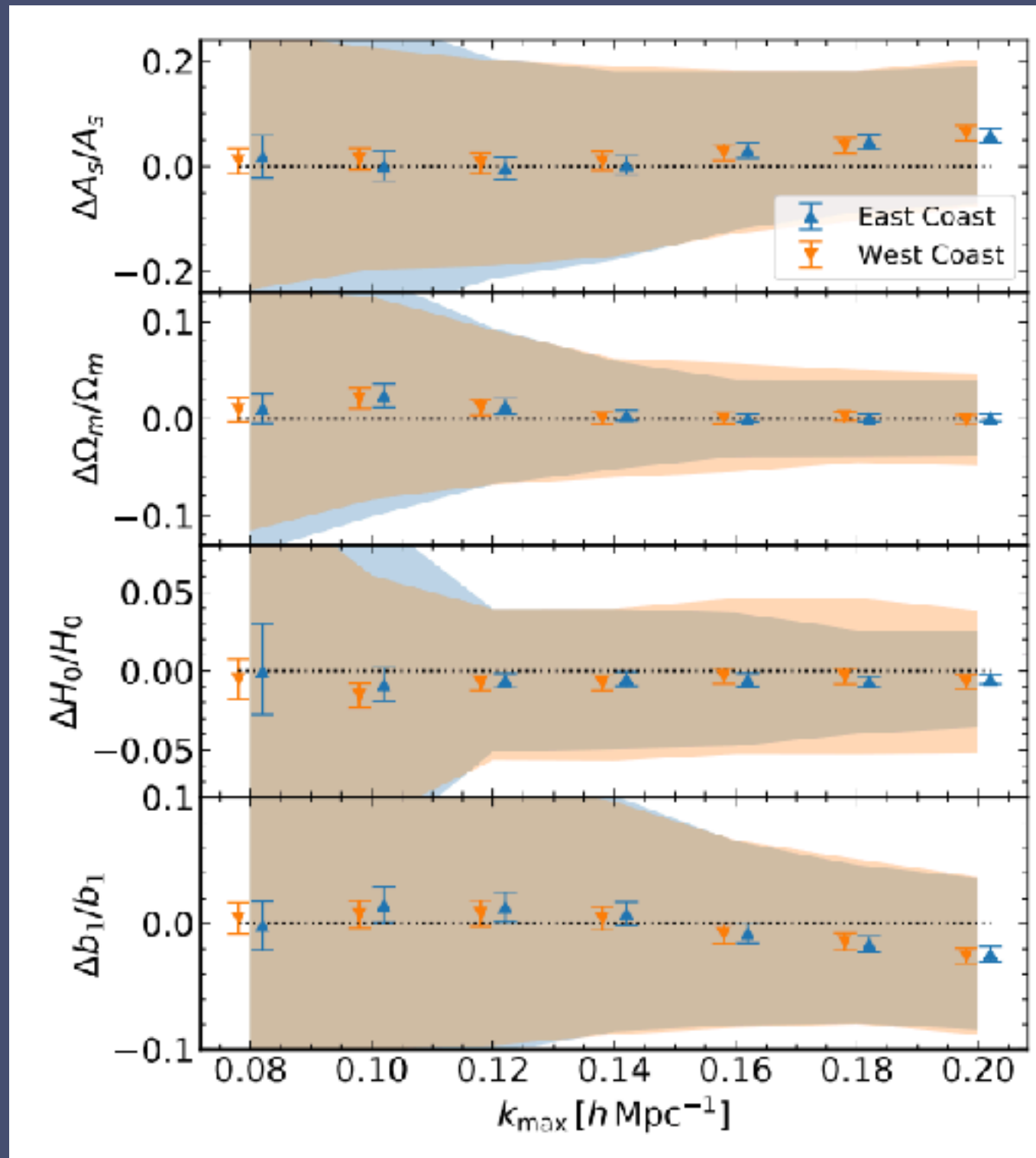
TN, et al. (arXiv:2003.08277, w/ Senatore, Zaldarriaga++)



- 全部入りの模擬データを作成
 - $10 \times (3,840 \text{ Mpc}/h)^3$ の体積を用意し、統計誤差を極限まで低減
 - 概ねSDSS BOSSのCMASS銀河と同様なセレクションで銀河を同定（やり方の詳細は伏せる）
 - Λ CDMを仮定（これは教える）
 - 3つの宇宙論パラメタを当てられるか（ A_s , Ω_m , H_0 を伏せ、他は教える）
 - <https://www2.yukawa.kyoto-u.ac.jp/~takahiro.nishimichi/data/PTchallenge/>

Blinded cosmology challenge

TN, et al. (arXiv:2003.08277, w/ Senatore, Zaldarriaga++)



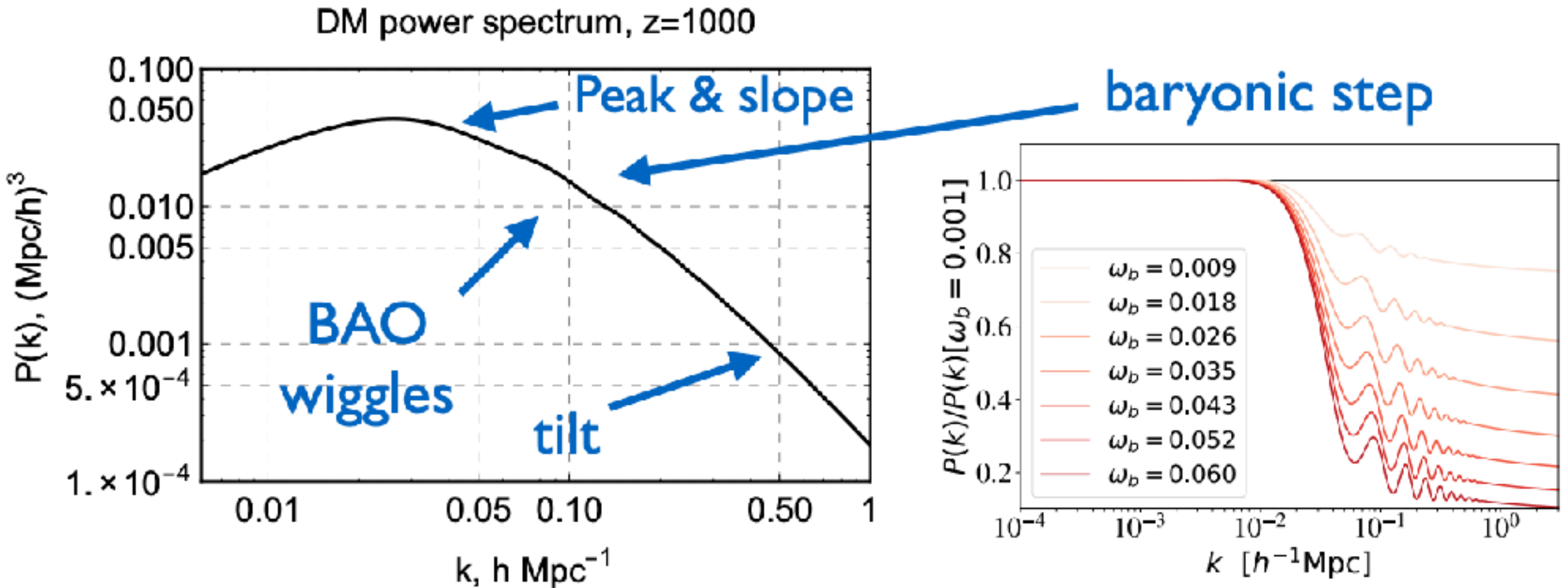
シンボル: シミュレーションデータが持つエラーレベルを仮定

shade: 現在のSDSS BOSSのエラーレベルへスケール

2 チームとも $k_{\text{max}}=0.12 \text{ Mpc}/h$ の結果を最終的なものとして提出

(少なくともとあるレシピで撒いた銀河サンプルに対しては) EFTはうまく行っているようだ

EFT view of information content



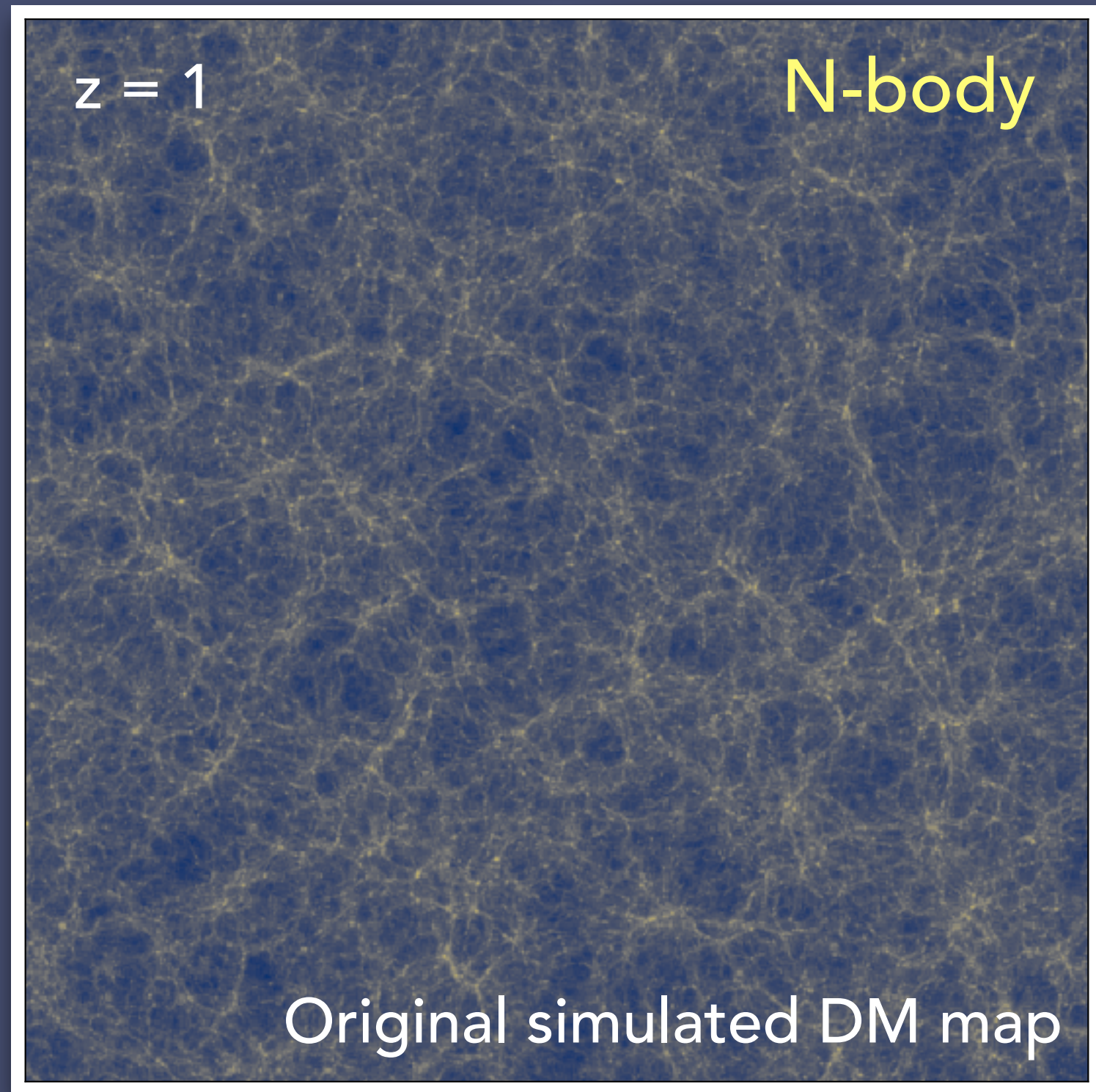
- ✓ Shape $(\omega_b, \omega_{cdm}, n_s)$
- ✓ Distance $D_V \rightarrow H_0$
- ✓ Amplitude $f\sigma_8 \rightarrow \sigma_8$

✗ AP effect $D_A, H \rightarrow$ nothing
✗ TN comment: assuming Λ CDM

Numerical evaluation of equation (7) shows that $\Delta z/(z \Delta \theta)$ is not a powerful estimator of q_0 in the $\Lambda = 0$ case—there is only 11% variation of $\Delta z/(z \Delta \theta)$ between $q_0 = 0$ and $q_0 = 1$ at $z = 2$.

From M. Ivanov's slide

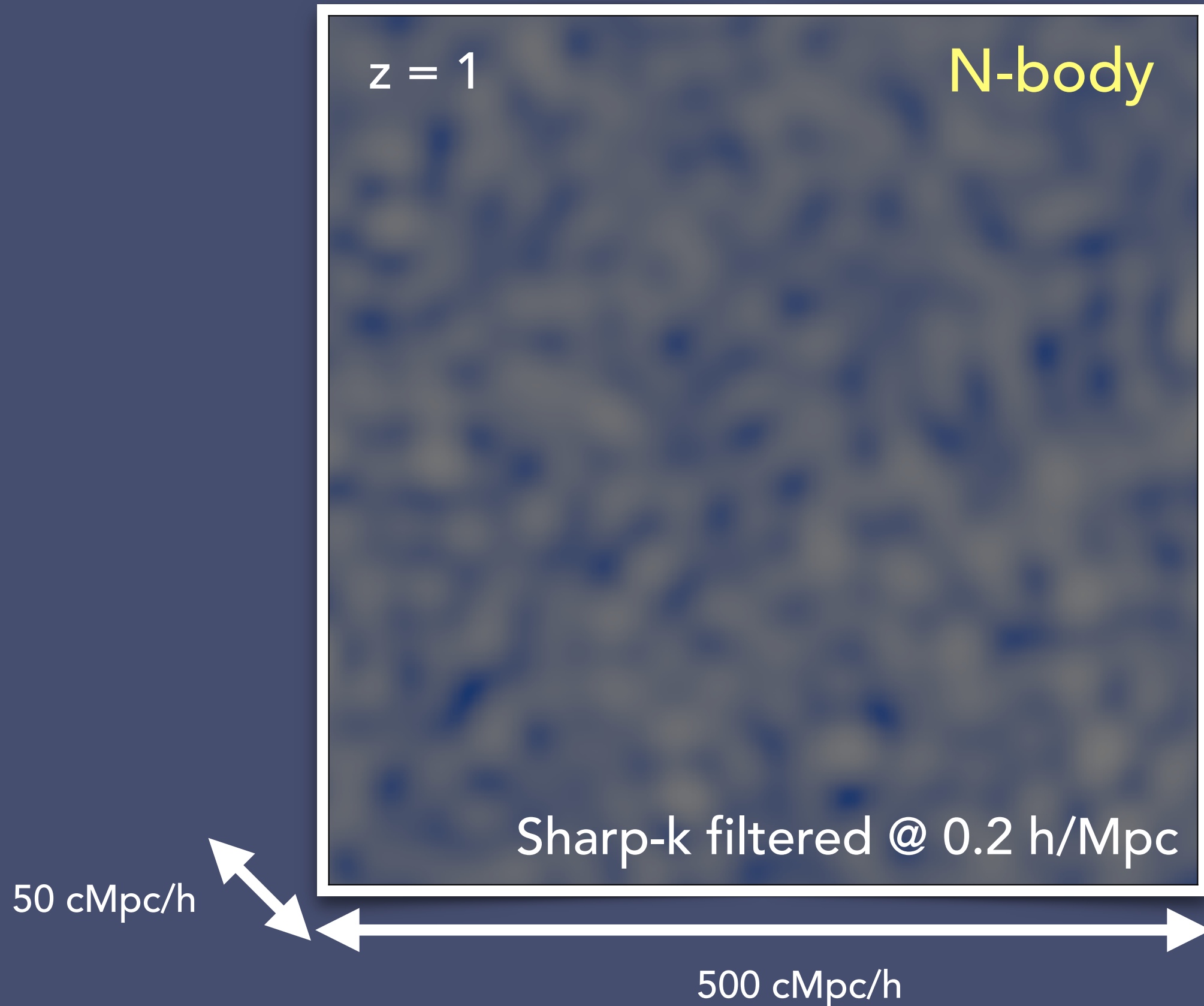
現状で解釈できている情報



50 cMpc/h

500 cMpc/h

現状で解釈できている情報



Simulation-based inference?

従来の解析パイプライン

新たな解析パイプライン

観測量

測定

要約統計量

順モデル:
各点での近
似理論計算

高精度シミュレーション
(100パラメタセットに対
してデータベース化)

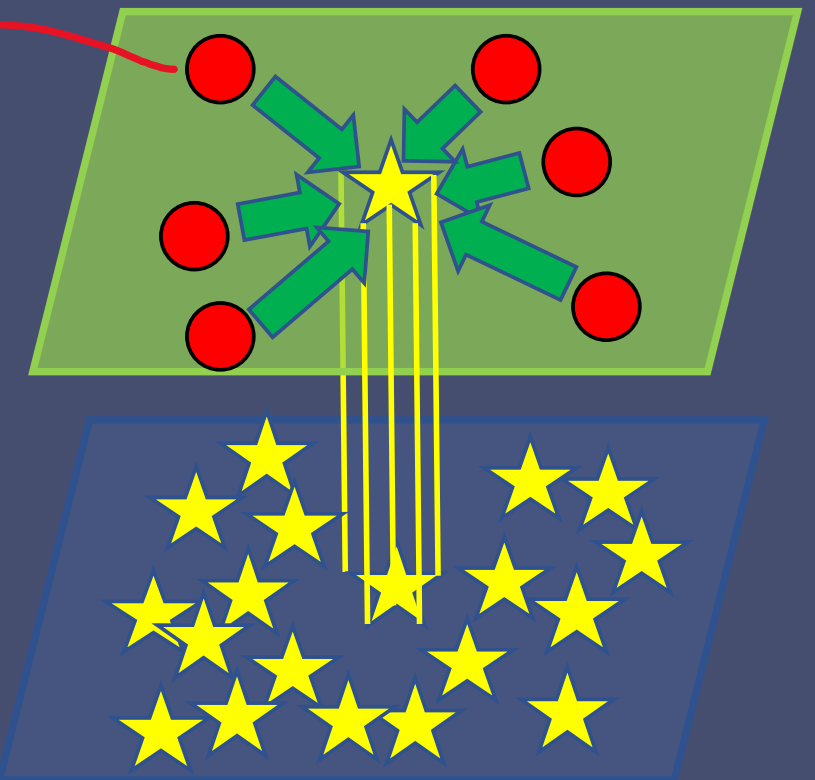
エミュレータ
(ベイズ推定)

MCMCサンプラー
多次元パラメタ空間から約
100万パラメタセットを探索

観測量

測定

要約統計量



計算コストの高い数値シミュレーションをベイズの枠
組みで利用可能に！

DARK QUEST PROJECT

TN+ arXiv:1811.09504

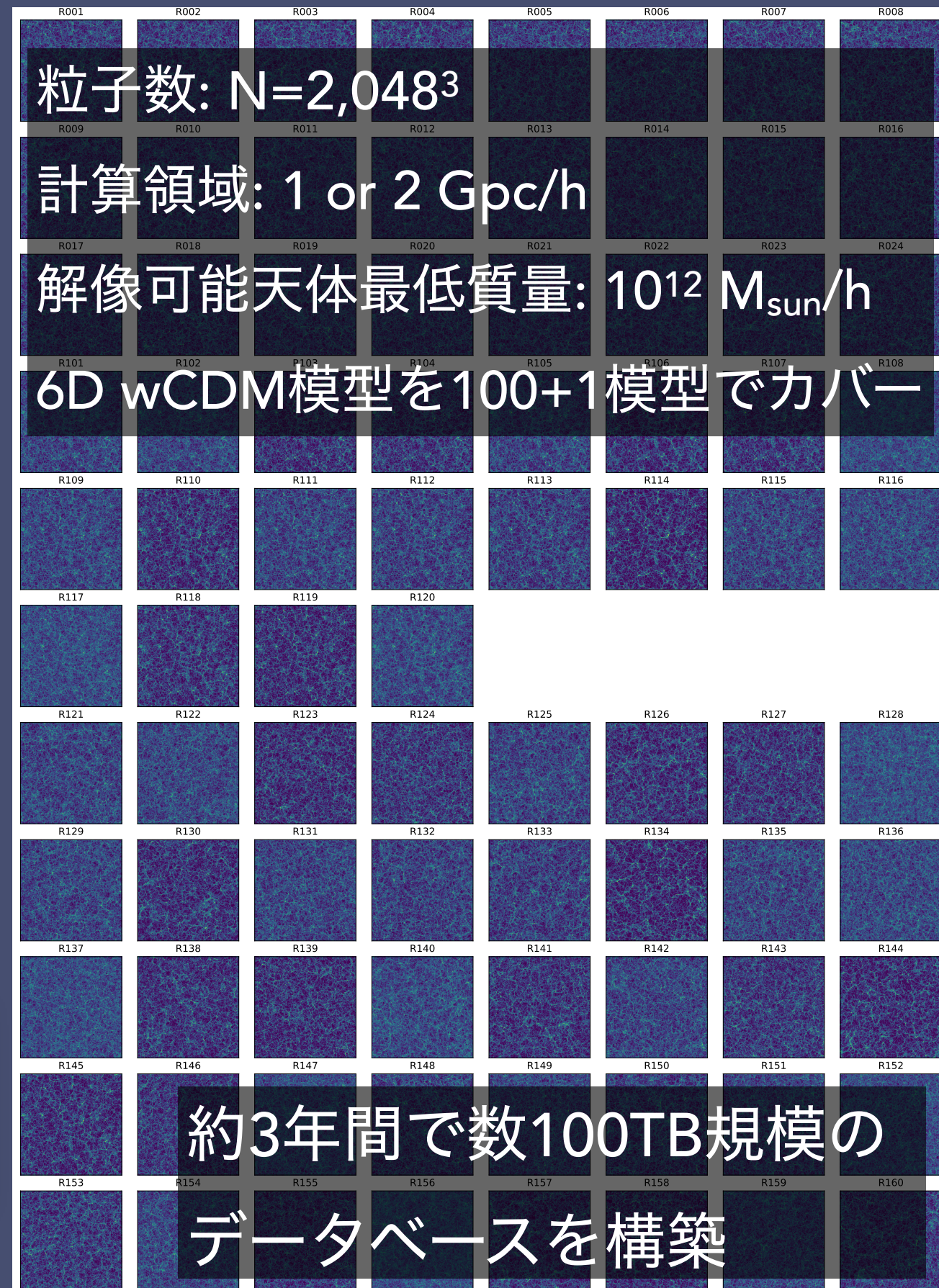
ハローの統計をN体シミュレーションで調べ尽くす

ハイブリッド統計モデリング

- 階層的ラテン超方格サンプリング
- ガウス過程回帰 (nonparametric Bayes)
- 重み付き主成分分析 (データ圧縮)
- パラメトリック銀河・ハロー結合模型
- 訓練・検証データスプリットによる厳密な精度評価

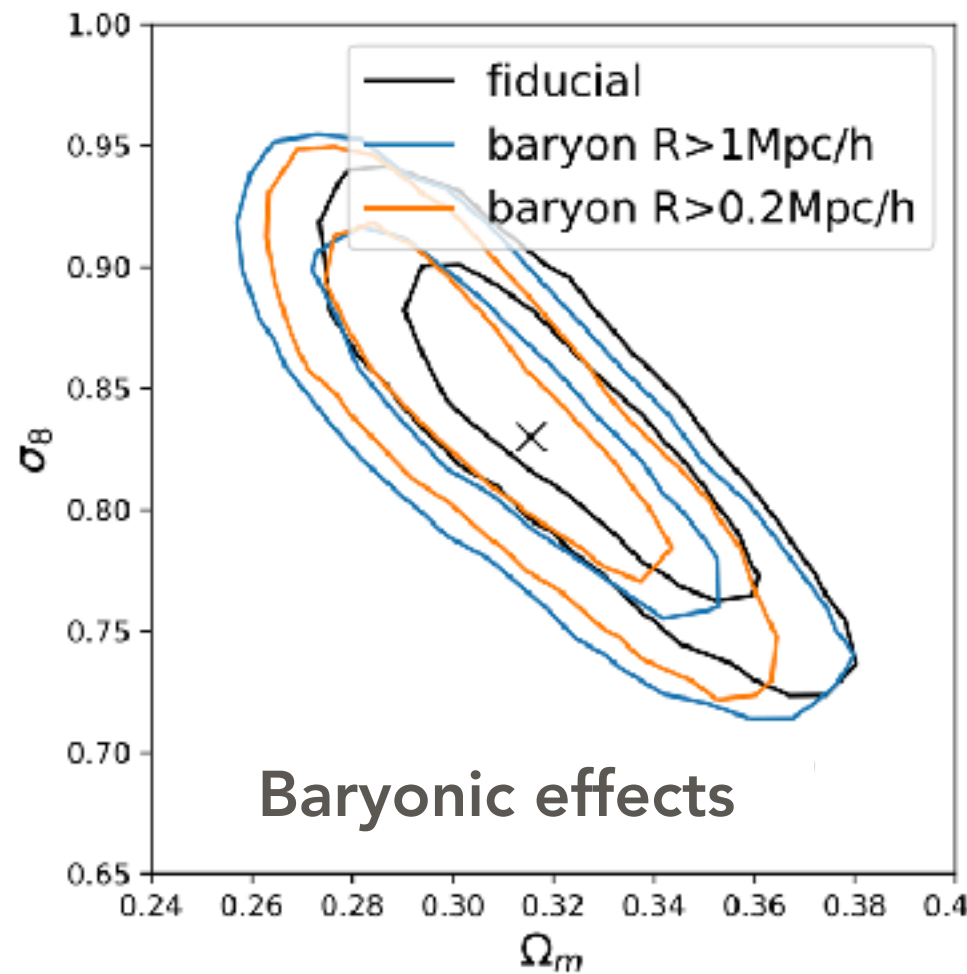
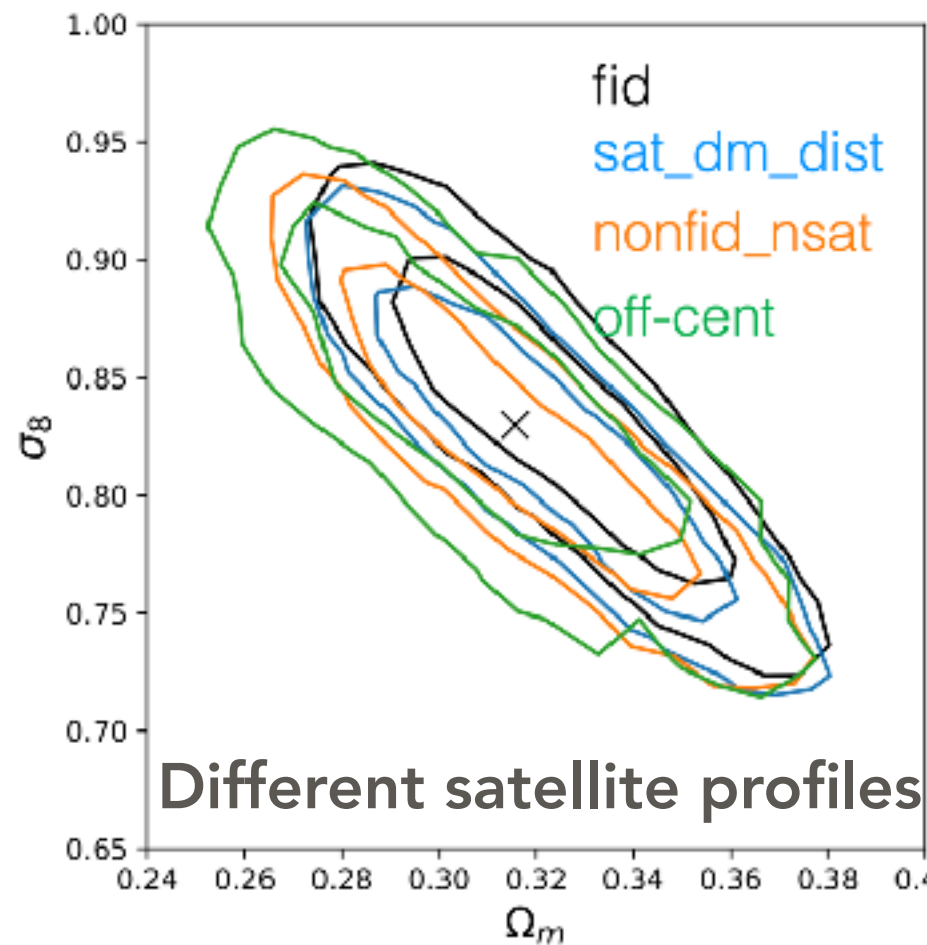
Emulate!

スパコンで2日をラップ
トップで数秒に短縮!



模擬宇宙 vs Emulator

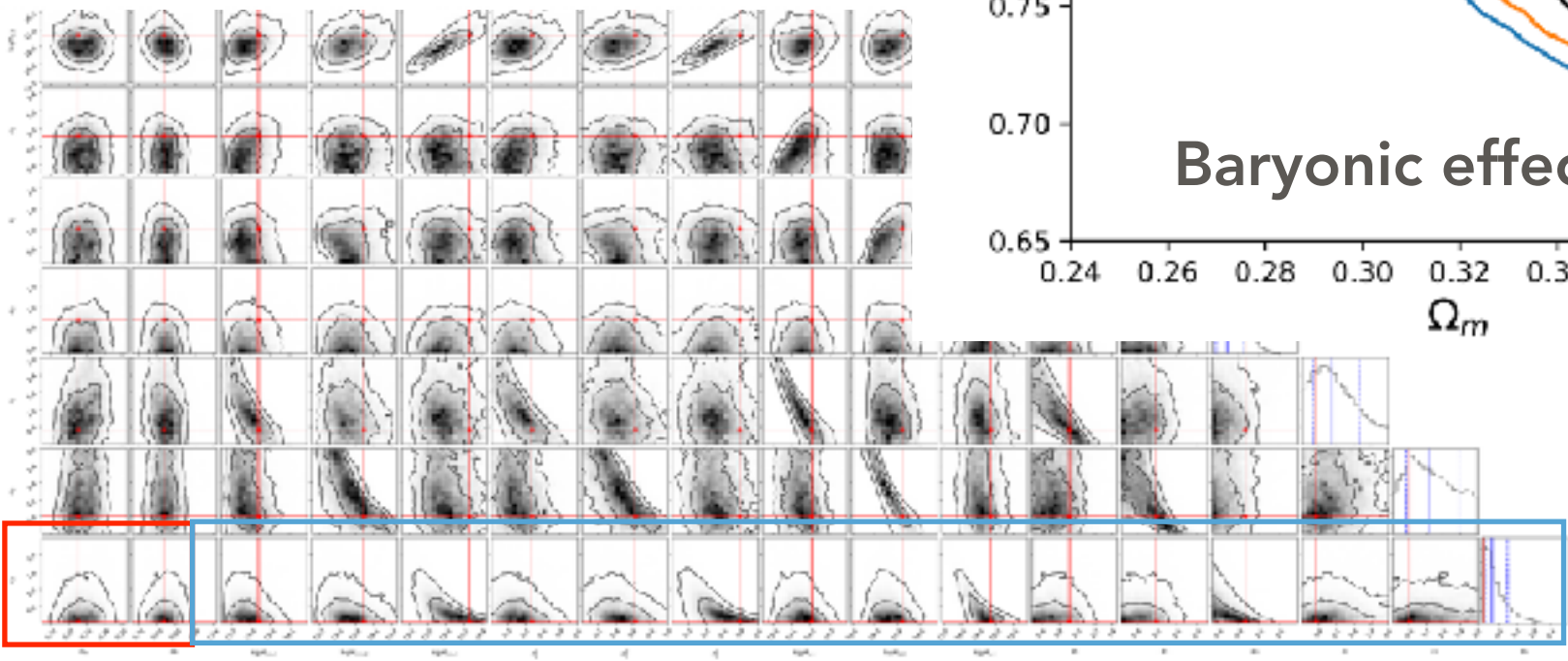
“数値計算 x AI”が観測データを分析する舞台がようやく整った！！



HSC x SDSSで得られるパラメタ制限の予測

fiducial: $r_{\text{lens}} = [0.2, 30] \text{ Mpc}/h$, $r_{\text{cl}} = [0.2, 30] \text{ Mpc}/h$
sat_subhalo: $r_{\text{lens}} = [2, 30] \text{ Mpc}/h$, $r_{\text{cl}} = [1, 30] \text{ Mpc}/h$

Substructures instead of HOD



COSMO

GALAXY (HOD + extra)

宇宙論の将来？

Ravanbakhsh, Oliva, Fromenteau+'17

Mathuriya, Bard, Mendygral+'18

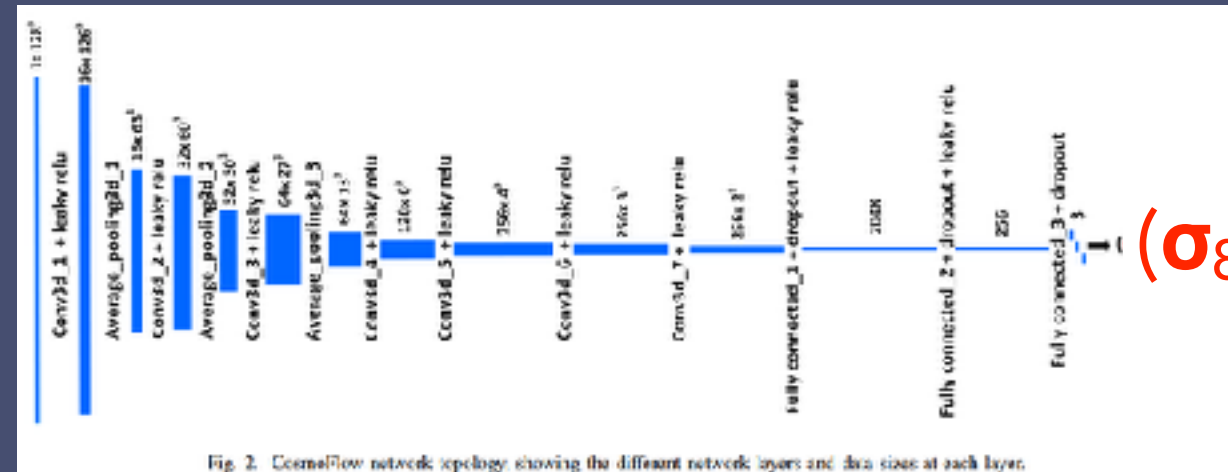
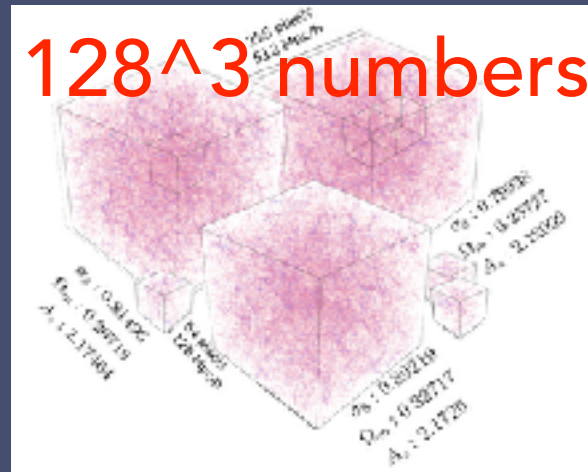
>7,000,000 parameters

69.33 Gflop / mimi-batch (mini-batch size = 1)

観測データ

$10^{(6-7)}$
numbers

128³ numbers

 $(\sigma_8, \Omega_m, n_s)$

Inverse problem

- 512Mpc/h, 粒子数 512^3
- 12,632 COLA simulations (高速簡易版N体)
- 8つの小領域に分け 128^3 グリッドの密度データを生成 (計101,056の学習・検証・テストデータ)

6 or so
numbers

宇宙論パラメタ

西道 啓博 (基研)

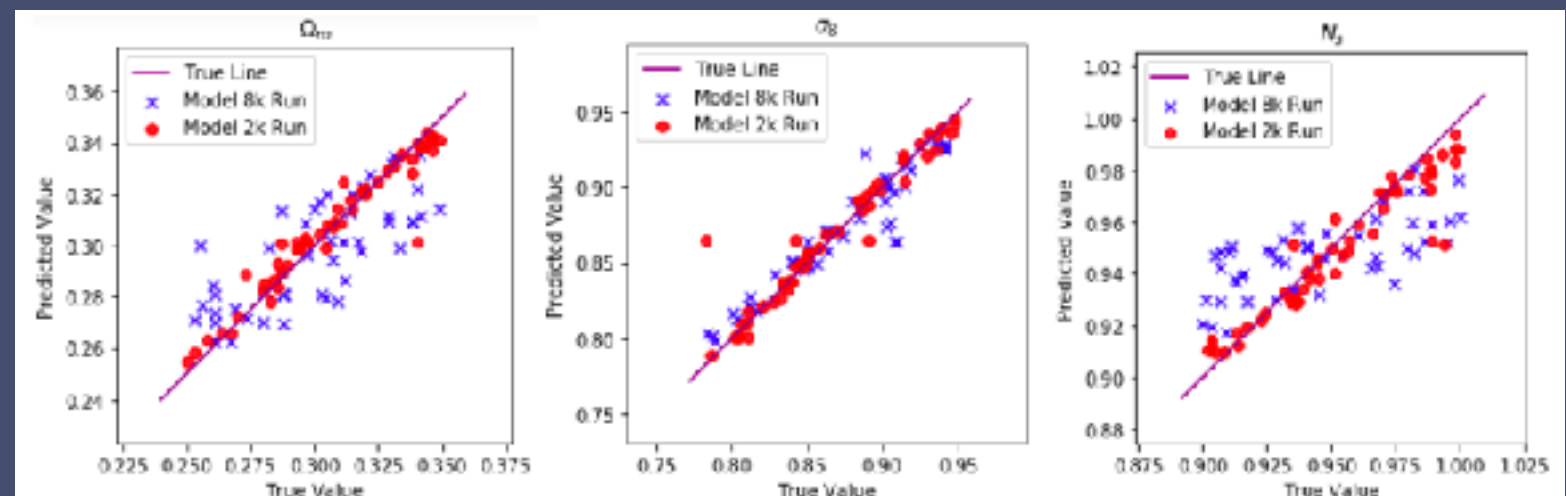
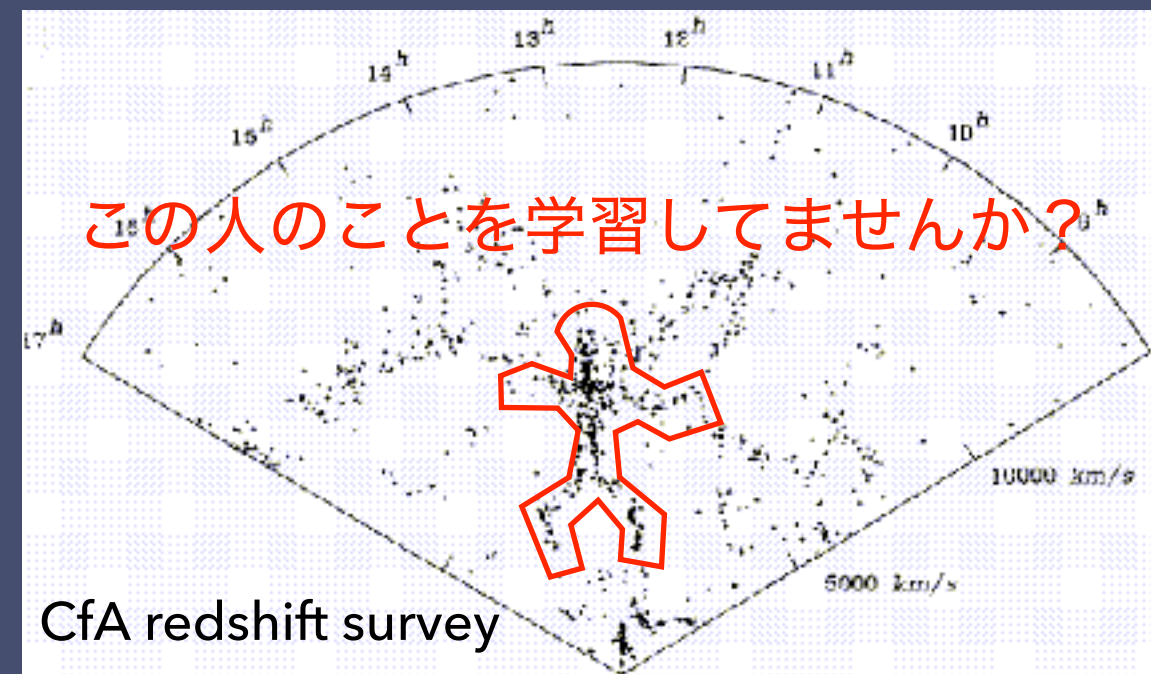
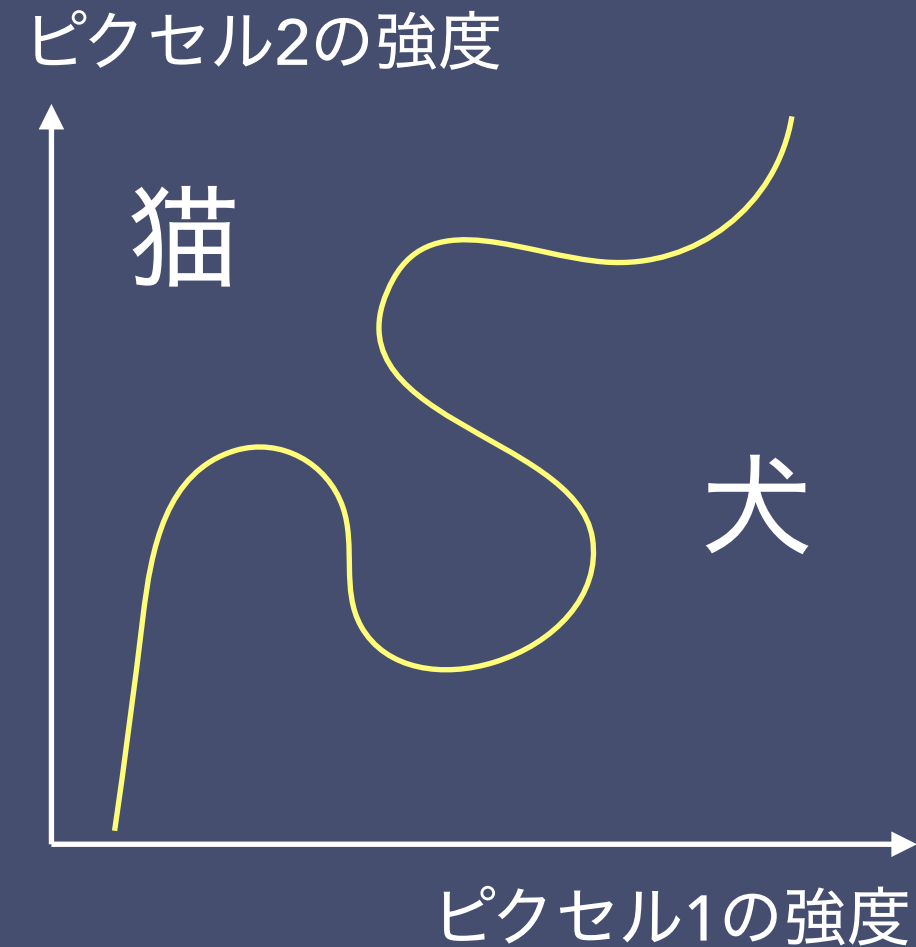
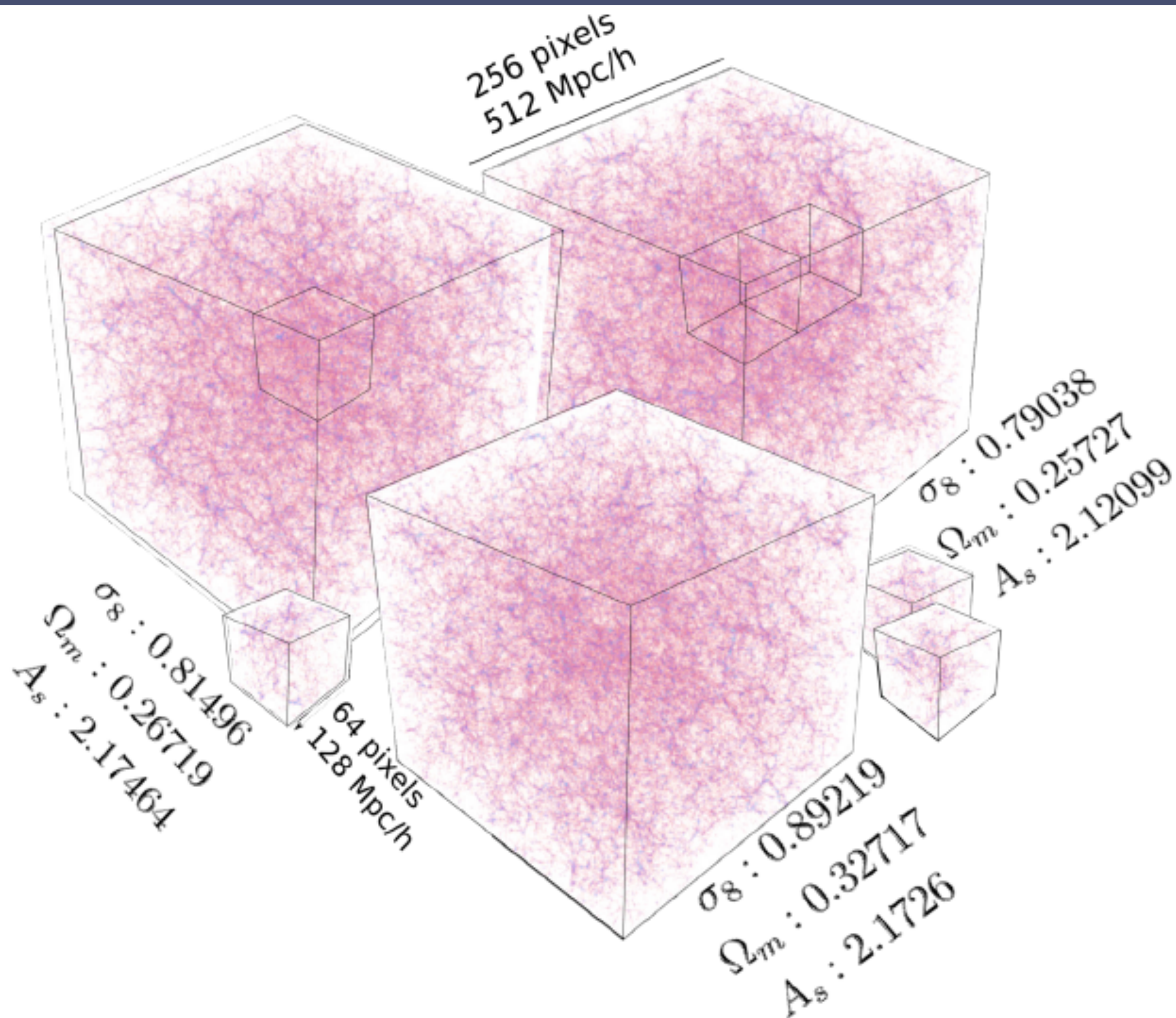


Fig. 6. Estimates of Ω_M , σ_8 and n_s from the 2048- and 8192-node runs.

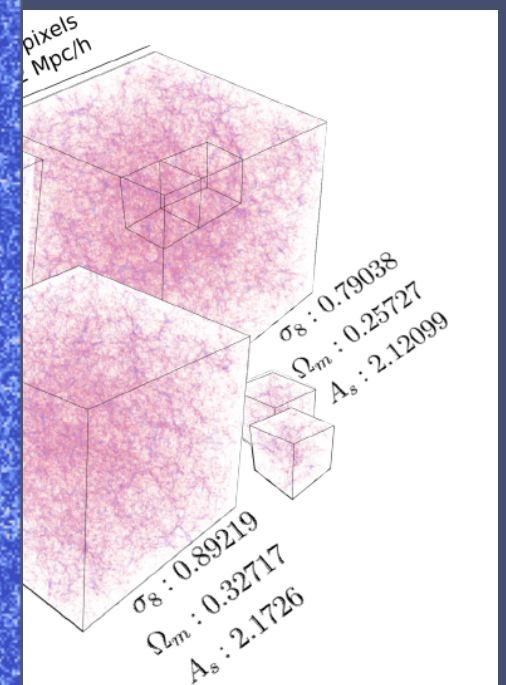
私見：本当にこれでいいのか？

- 近年の深層学習の発展の主な舞台の一つに画像認識がある
 - 非線形性が極端に強く
 - 情報が局在化されている
 - ほとんどの情報は無駄
 - Ex. ReLU活性化関数、max pooling ...
- 観測データの膨大な自由度を調査し切れるか？
 - 扱っているものがそもそも確率場
 - 過学習を防ぐ機構は十分？
 - 長距離相関を拾える？





About the size of the Subaru PFS survey



Super resolutionとその亜種

観測データ

$10^{(6-7)}$
numbers

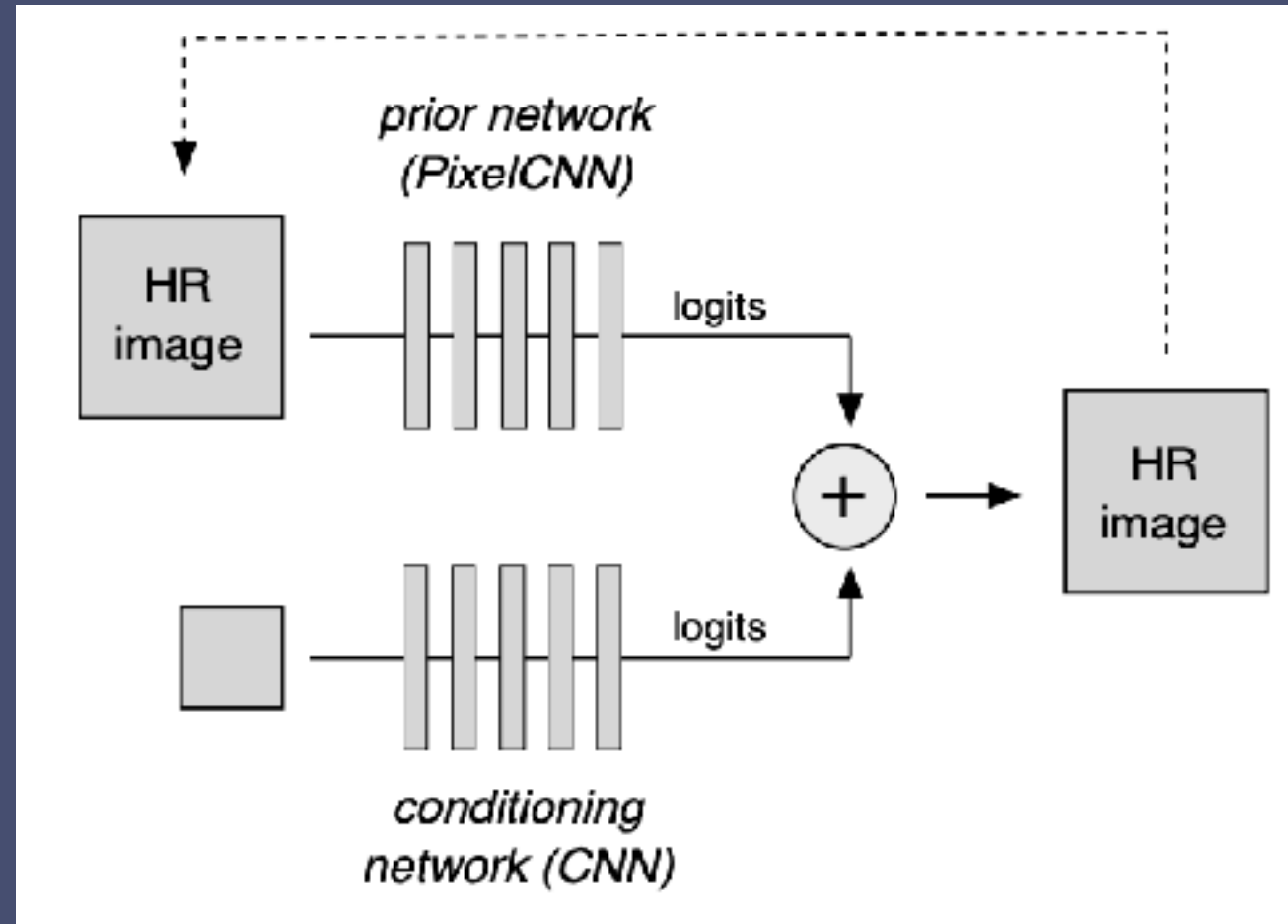
高精度模型

機械学習

近似模型

6 or so
numbers

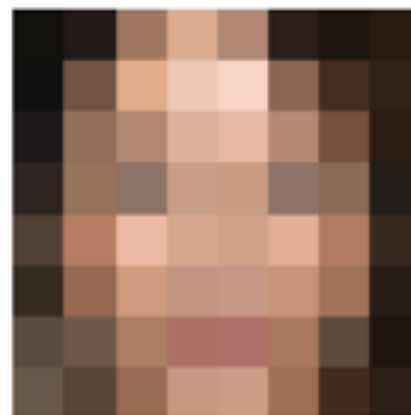
宇宙論パラメタ



8×8 input

32×32 samples

ground truth



Super resolution, arXiv:1702.00783, by Google Brain

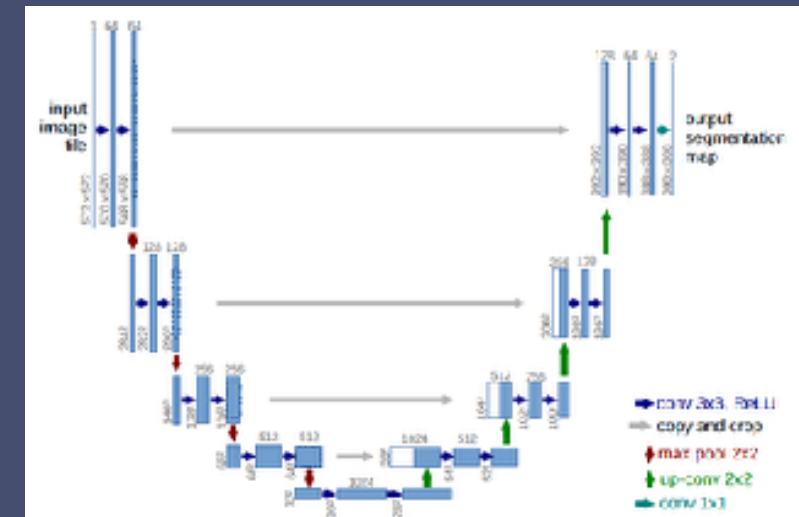
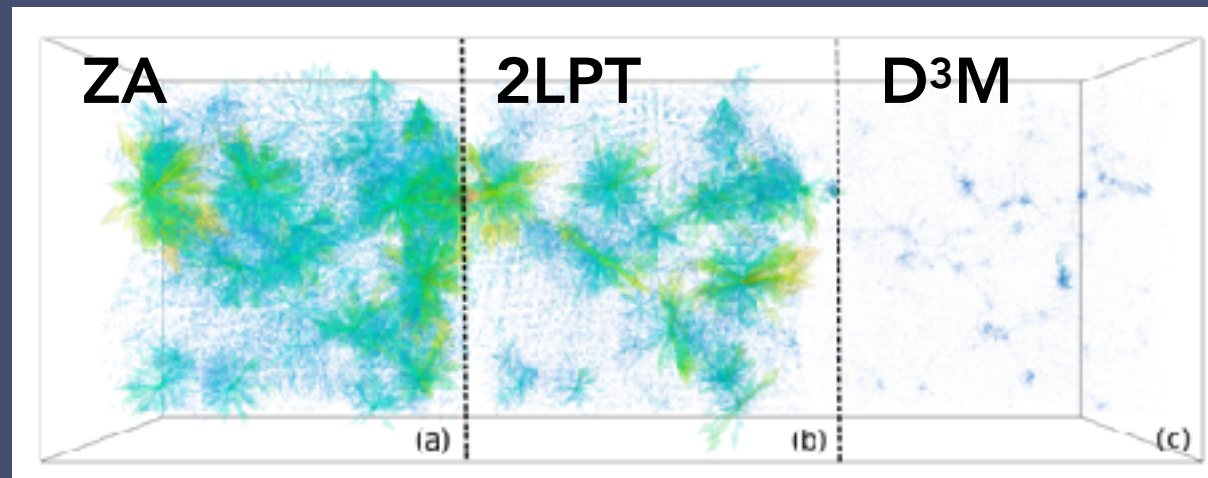
その一例

He, Li, Feng+'18

観測データ

フルN体からの残差 (displacement)

$10^{(6-7)}$
numbers



高精度模型

- 学習させるもの：
 - インプット: ゼルドビッチ近似
 - アウトプット: N体
- 10,000ペア作成
 - 128 Mpc/h, 粒子数 32^3
- 機械はダイナミクスを学ぶので、学習に使ったものとは別の宇宙論パラメタでもある程度ワークすると主張

機械学習

近似模型

6 or so
numbers

宇宙論パラメタ

まとめ

- 大規模構造を用いた宇宙論
 - 大スケールのピュアな情報にロバストにアクセスする手段が確立されつつある
 - 言い換えると情報のほんの一部しか使えていない！！($k_{\text{max}} \sim 0.2h/\text{Mpc}$)
- N体シミュレーション + 機械学習
 - エミュレータ
 - フィッティング公式の代用
 - HSCのデータ解析へ実用段階に
 - もっと機械学習に頼る方向性
 - 2点統計以上の情報を取り出す可能性
 - まだまだproof-of-concept段階だが、将来無視できなくなる？