Modeling large-scale structure in the golden age of cosmological surveys

Martin White (UCB/LBNL)

w/ Stephen Chen, Chirag Modi (UCB)
Emanuele Castorina, Zvonimir Vlah (CERN)

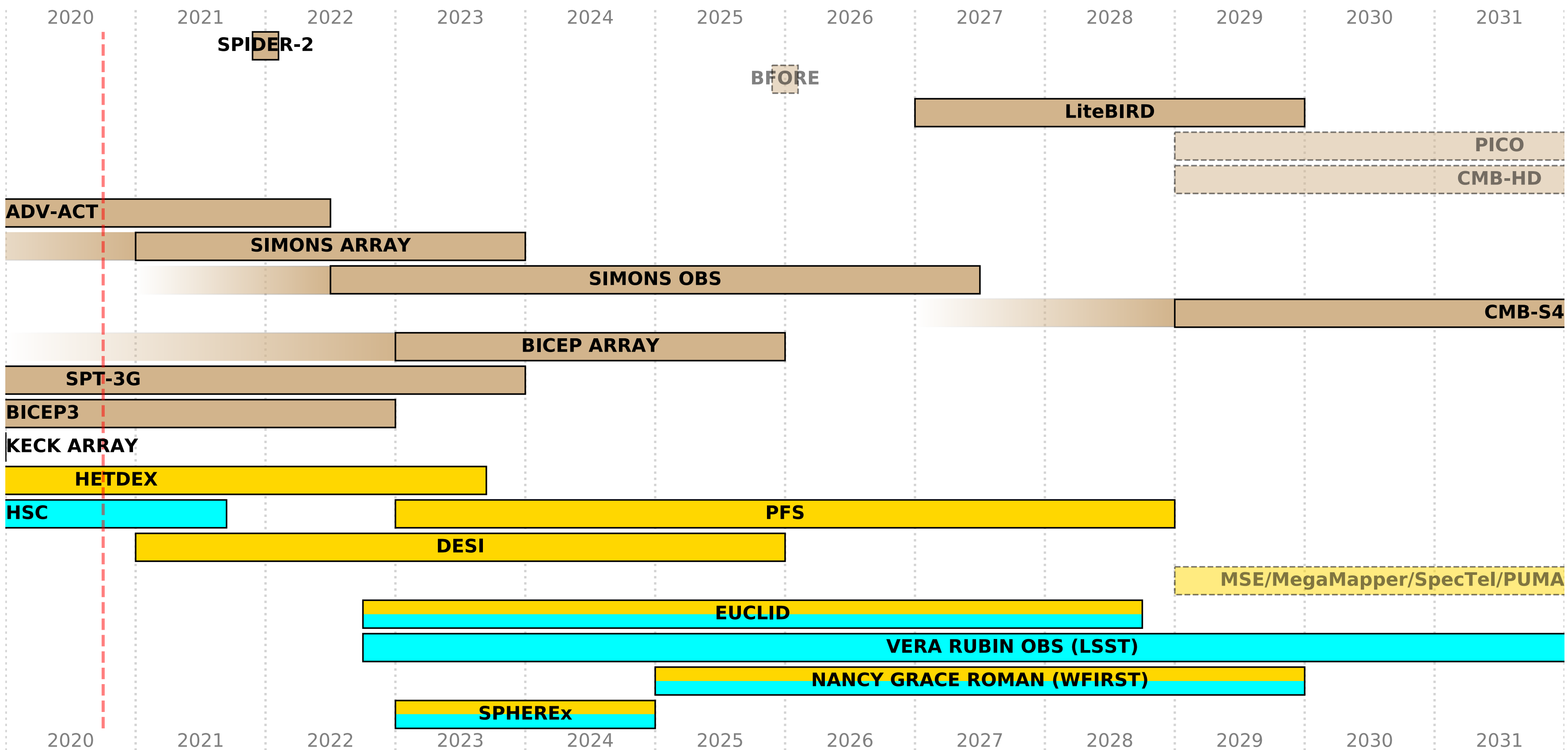
Harvard, 21 Sep 2020

Outline

- The golden age of cosmological surveys
  - We are gathering vast amounts of data that can inform us about many interesting questions in physics and astrophysics … given suitable models!
- Modeling the evolution of large-scale structure
- Concluding thoughts

Golden age of surveys

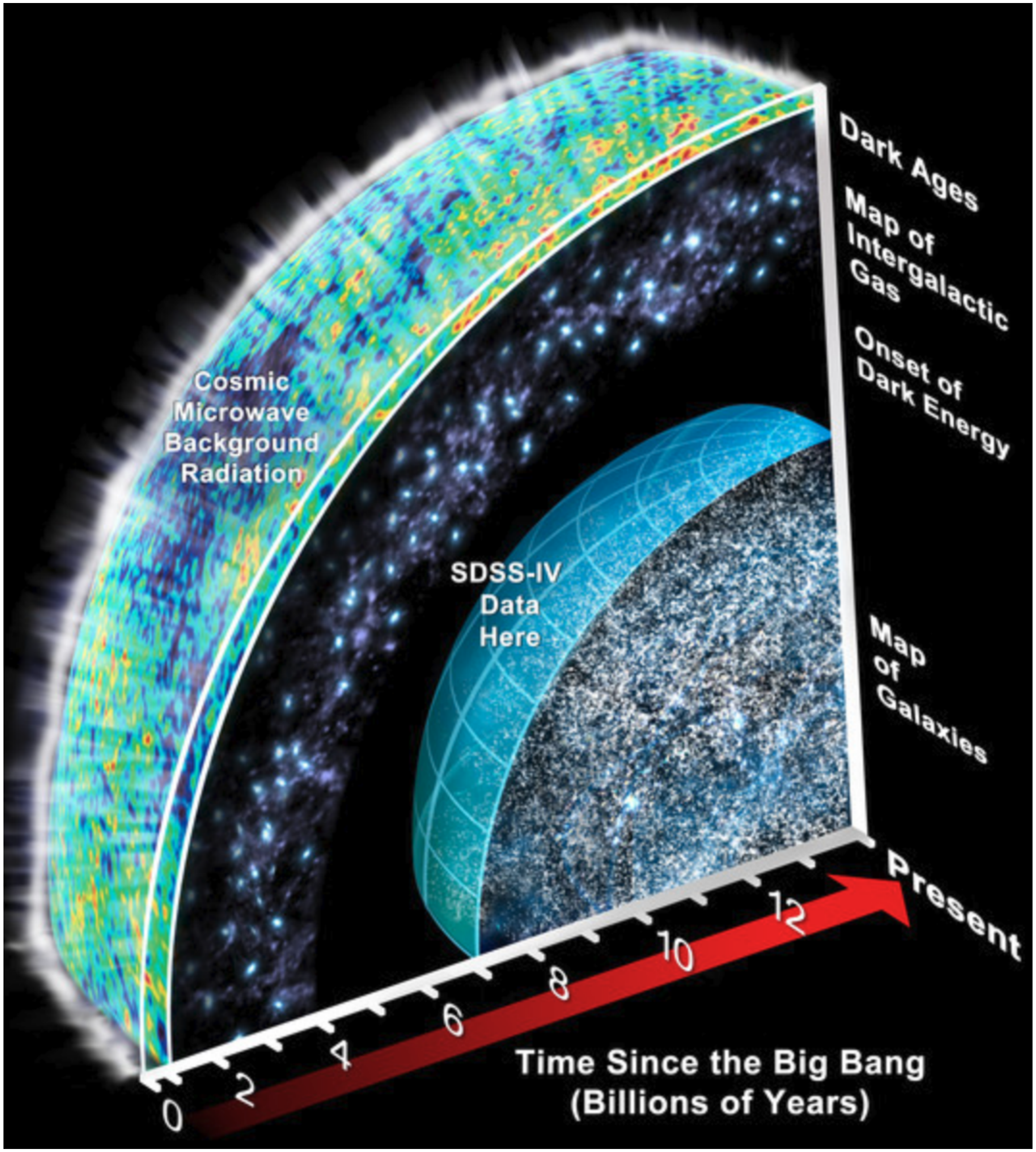
We are living in the "golden age" of cosmological surveys, with survey capabilities increasing exponentially ... (Moore's law)



D. Kirkby

Large-scale structure (LSS)

When we view the Universe today, we find structure on scales from the cosmological horizon to planetary systems.



Courtesy: SDSS/eBOSS

The story we tell ...

- Something like inflation turned quantum fluctuations (in the "inflaton" field) into classical perturbations in the density of all species at early times.
- Fluctuations grow over time through a process of gravitational instability to form all of the structure we see today.
- 14 Gyr of evolution shapes the spectrum.
  - Wide range of energy densities, temperatures, ...
  - We see these fluctuations at $z \simeq 10^3$ ($t \sim 400,000\,\mathrm{yr}$) as fluctuations in the temperature of the cosmic microwave background (CMB) radiation with $\Delta T/T \sim 10^{-4}$.
  - We see these fluctuations "today" with galaxy surveys.
- Growth is a competition between gravity and expansion
  - Depends on the laws of gravity (general relativity).
  - Depends upon the expansion of the Universe (metric).
  - Depends on constituents and their properties.

Probe metric, particle content and **both** epochs of accelerated expansion ... with high precision

Statistical inference

Since inhomogeneity arose from stochastic (QM) fluctuations, all inferences are statistical.

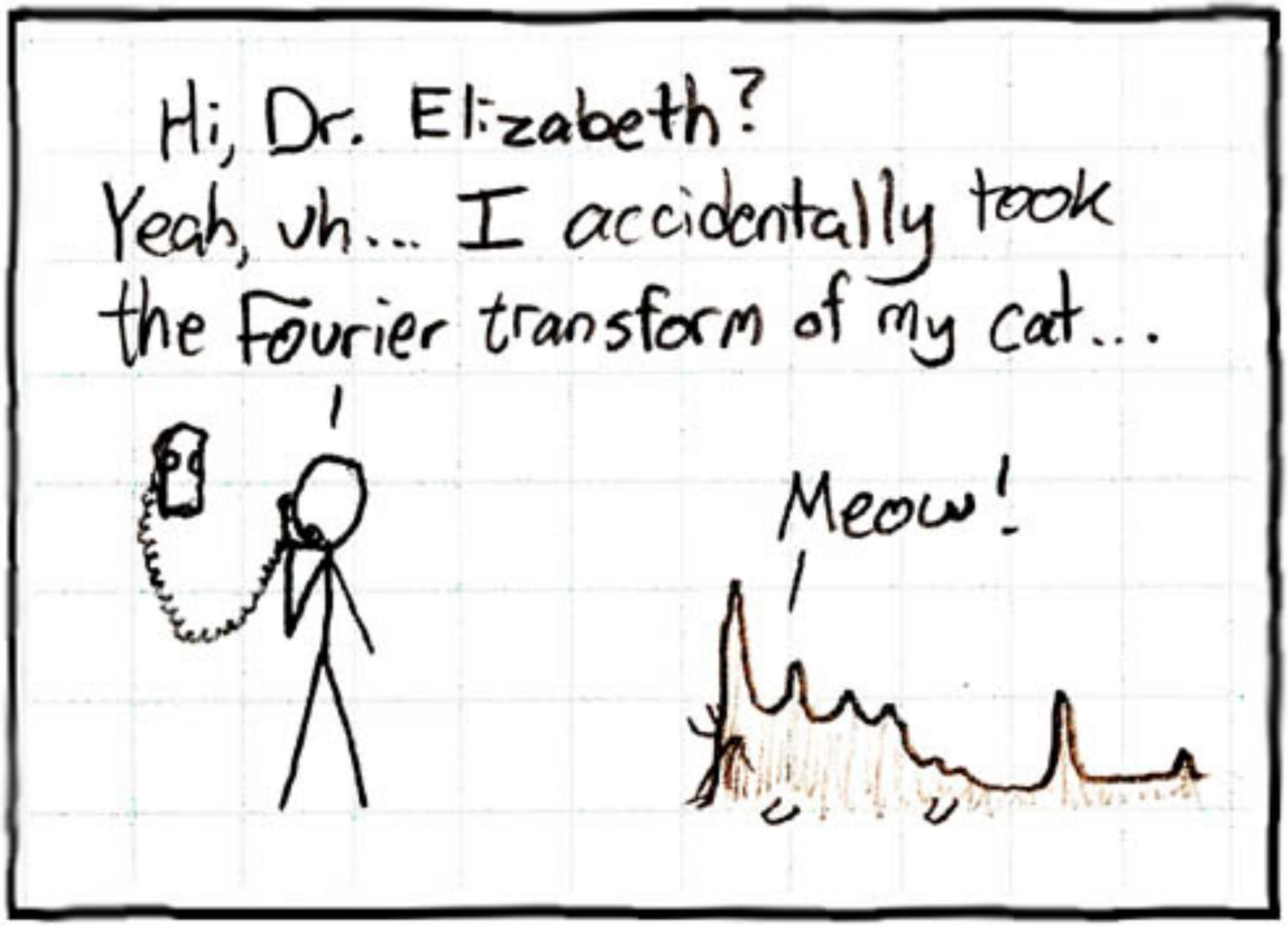
- The particular location of any given galaxy or star is not relevant … we look at correlations.
- Compare correlators of temperature, density, velocity, etc.
- For this talk my focus will be on 2-point functions, e.g. $\langle\rho(\mathbf{x}_1)\rho(\mathbf{x}_2)\rangle$.

An irreducible contribution to our uncertainty comes from the number of "independent samples" (modes) in our survey … want to push to larger volumes (i.e. earlier times)!

Statistical inference

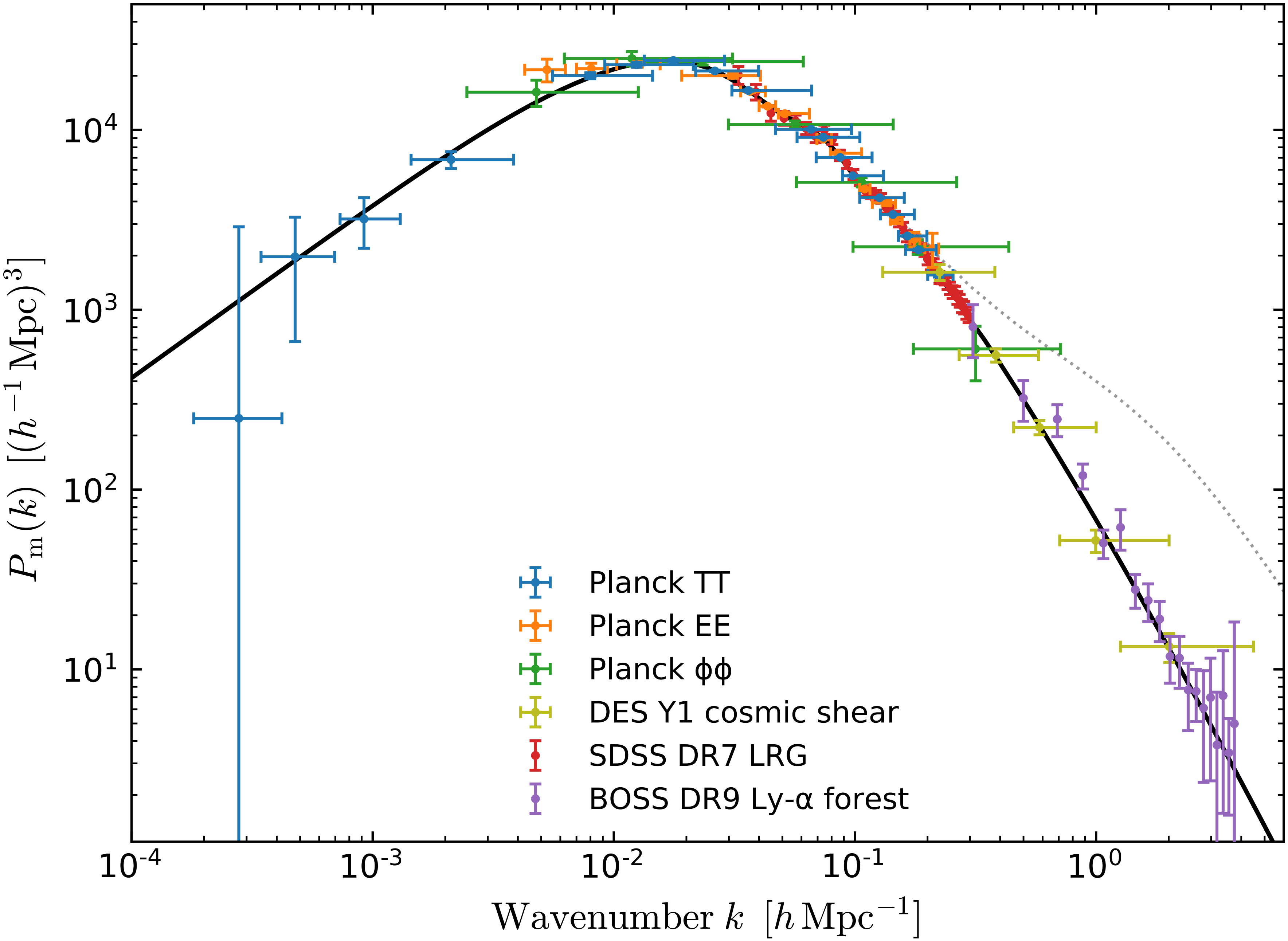
The underlying processes are translationally invariant, so we tend to work in Fourier space:

$$\rho(\mathbf{x}) = \bar{\rho}\,(1+\delta(\mathbf{x})) \quad , \quad \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\rangle = (2\pi)^3 \delta^{(D)}(\mathbf{k}_1+\mathbf{k}_2)\, P(k)$$



xkcd.com/26

The power spectrum: multiple tracers, late time



Planck+(2020)

Models of large-scale structure (LSS)

How do people model measurements of large-scale structure?

- There are two broad classes of approaches to modeling LSS: numerical and analytical.
- Numerical approaches (simulations)
  - Techniques for solving systems of pure dark matter are well developed; though the combination of volume and resolution required by next-gen surveys is very demanding.
  - The best way to deal with the complexities of galaxy formation, hydrodynamics and multiple species is still an open research problem.
- I will discuss analytic approaches based on perturbation theory (PT) – which have seen a renaissance in recent years.
- Most practitioners use some combination ...
  - All N-body codes use PT for initial conditions.
  - N-body can be used to test PT for fiducial models.
  - New ideas for combining the two: "best of both worlds".

Perturbation theory (PT) history

- Cosmology deals with relativistic gauge field theories, like many other sub-fields of physics.
- The equations of motion are both non-linear and non-local.
- PT developed starting in the 1960's, reached its "classical" form in the early 1990's.
  - At this point surveys could only probe well k-modes for which perturbation theory was barely applicable at all.
  - And in 1992 anisotropies in the CMB were discovered (queue Nobel Prize) for which linear theory is all that is needed.
  - The 1990's also saw the introduction of 'commodity' supercomputing and a huge advance in our ability to simulate large-scale structure.
- While it never fully died out, use of PT only really blossomed again within the last 5-10 years ...

A funny thing happened ...

- Cosmology is riding the Moore's law/big data revolution like many other fields.
- Even though computing/simulation is becoming a bigger component of the analysis toolkit, modern surveys are empowering theorists as never before ...
- We have the technology to survey very large volumes at larger distance (i.e. earlier times).
  - Large scales have undergone less processing, more correlated with ICs.
  - Fluctuations are linear, or quasi-linear ($\delta \lesssim 1$).
  - Such modes are under good "theoretical control" using PT.
  - We're now computing small corrections to "almost linear" quantities.
  - Bigger surveys demand higher precision.

Challenge I: non-linearity

- The large scale structure power spectrum is very blue: it has a lot of small-scale (UV) power.
  - Bottom-up structure formation.
  - Non-linear (length) scale grows with time.
- In the present day Universe non-linear scale is $\mathcal{O}(10\,\mathrm{Mpc})$ – on scales smaller than 10 Mpc $\delta = \delta\rho/\rho \gg 1$.
- Non-linearity has three disadvantages for our purposes:
  - It couples modes, reducing the amount of independent information.
  - It decorrelates the observed field from the initial conditions (primordial Universe).
  - It makes it harder to model! We need to go beyond "tree level", linear perturbation theory.

(This effect has been modeled out in the Planck figure!)

PT: two flavors

<u>Eulerian</u> (standard)
Treat cold dark matter as a pressureless (perfect) fluid obeying

$$\begin{aligned} \partial_\tau \delta + \nabla \cdot [(1+\delta)\mathbf{v}] &= 0 \\ \partial_\tau \mathbf{v} + \mathcal{H}\mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{v} &= -\nabla \Phi \end{aligned}$$

with the $\mathcal{H}v$ term being "Hubble drag" arising from the expansion of space.

<u>Lagrangian</u>
Treat cold dark matter as a collisionless system

$$\mathbf{x}(\mathbf{q}) = \mathbf{q} + \mathbf{\Psi}(\mathbf{q}, \tau)$$

with

$$\partial_\tau^2 \mathbf{\Psi} + \mathcal{H} \partial_\tau \mathbf{\Psi} = -\nabla \Phi \left(\mathbf{q} + \mathbf{\Psi}(\mathbf{q})\right)$$

then derive density from

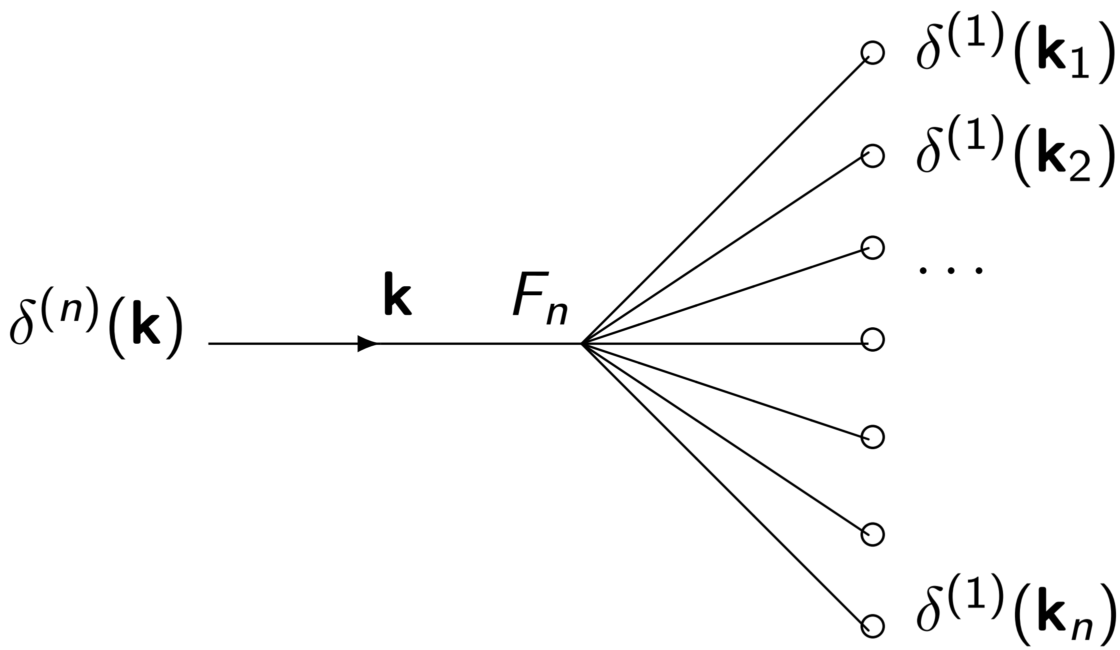
$$1+\delta(\mathbf{x}) = \int d^3 q \, \delta^{(D)} \left[\mathbf{x} - \mathbf{q} - \mathbf{\Psi}(\mathbf{q})\right]$$

(Both derivable from the Vlasov equation)

Method of solution

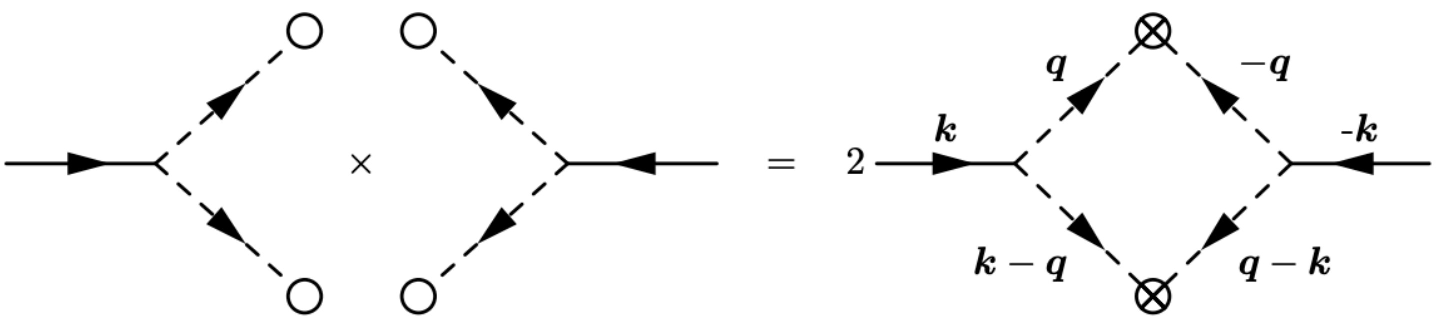
- Find Green's function for linear problem.
- Plug linear solution into non-linear terms and integrate against Green's function to get $\delta^{(2)}$ or $\mathbf{\Psi}^{(2)}$.
- Plug those in and integrate to get $\delta^{(3)}$ or $\mathbf{\Psi}^{(3)}$, etc.

$$\delta^{(n)}(\mathbf{k}) \propto \int \prod_{i=1}^{n} d^3k_i \, \delta^{(D)}\left(\sum \mathbf{k}_i - \mathbf{k}\right) F_n\left(\{\mathbf{k}_i\}\right) \delta^{(1)}(\mathbf{k}_1) \cdots \delta^{(1)}(\mathbf{k}_n)$$



Eulerian PT

Taking the expectation value "joins" pairs of $\delta^{(1)}$ together to form a power spectrum. So the propagator at 1-loop contains a contribution like e.g.:



$$= 2\int \frac{d^3q}{(2\pi)^3} F_2(\boldsymbol{q}, \boldsymbol{k}-\boldsymbol{q}) F_2(-\boldsymbol{q}, \boldsymbol{q}-\boldsymbol{k}) P_0(q) P_0(|\boldsymbol{k}-\boldsymbol{q}|)$$

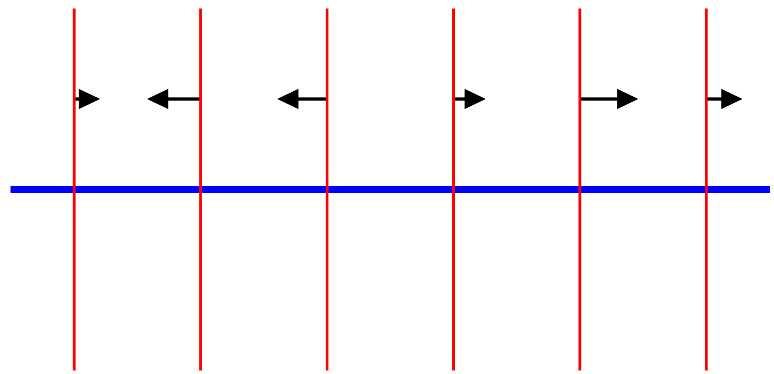
Many other rules look similar to particle or condensed matter physics – use path integrals, generating functions, cutoffs, EFT, RG flows, etc.

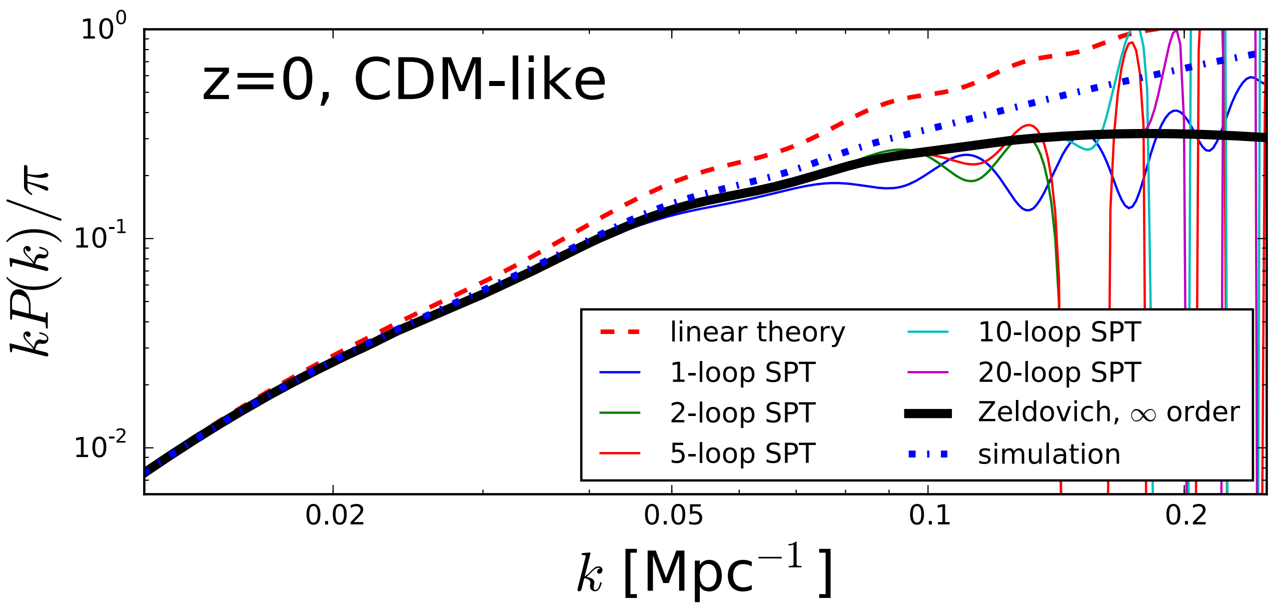
A problem emerges

- These two approaches give the same predictions, order by order in perturbation theory.
- **This is a problem!**
- Two frameworks for PT describe different systems:
  - pressureless fluid (Eulerian) and
  - collisionless fluid (Lagrangian),
- Shocks vs. caustics.

A toy model

Consider a collection of uniform, parallel, 2D sheets of matter moving normal to the sheets under gravity.





McQuinn & White (2016)

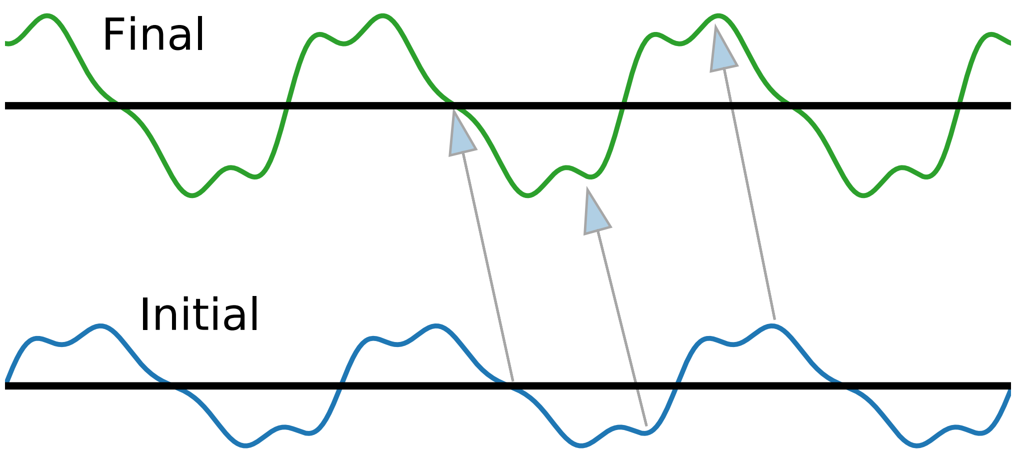
UV: effective field theory

This problem is well known in many areas of physics!

- EOM are non-linear, so have "composite" terms like $v\delta$.
- In Fourier space $\delta^{(2)}(k) \sim \int dk' \ \mathcal{K} \ \delta^{(1)}(k-k')\delta^{(1)}(k')$
- But $\delta^{(1)}(k')$ is not small for high k': PT breaks down.
- Need to regularize and introduce counter terms.
  - In Eulerian PT the lowest order counter term looks like a pressure force.
  - Lagrangian PT looks like a multipole expansion of extended objects – how they respond to low-k potentials and tides.
  - For the experts: it turns out that the lowest order counterterm also handles "peaks bias" and "baryonic effects".

IR resummation

- In addition to problems in the UV, there are issues in the IR.
- A lot of the difference between $\delta^{(\mathrm{non-lin})}(\mathbf{x})$ and $\delta^{(1)}(\mathbf{x})$ comes from displacement (advection).



- The displacement is driven by large-scale tidal fields.
- In "standard" perturbation theory this effect converges slowly.
- Need to "resum" the long-wavelength displacements
  - The Lagrangian formulation of PT is ideally suited to understanding "IR resummation".
  - Impacts topics like "reconstruction", primordial features, relative velocity effect, ...

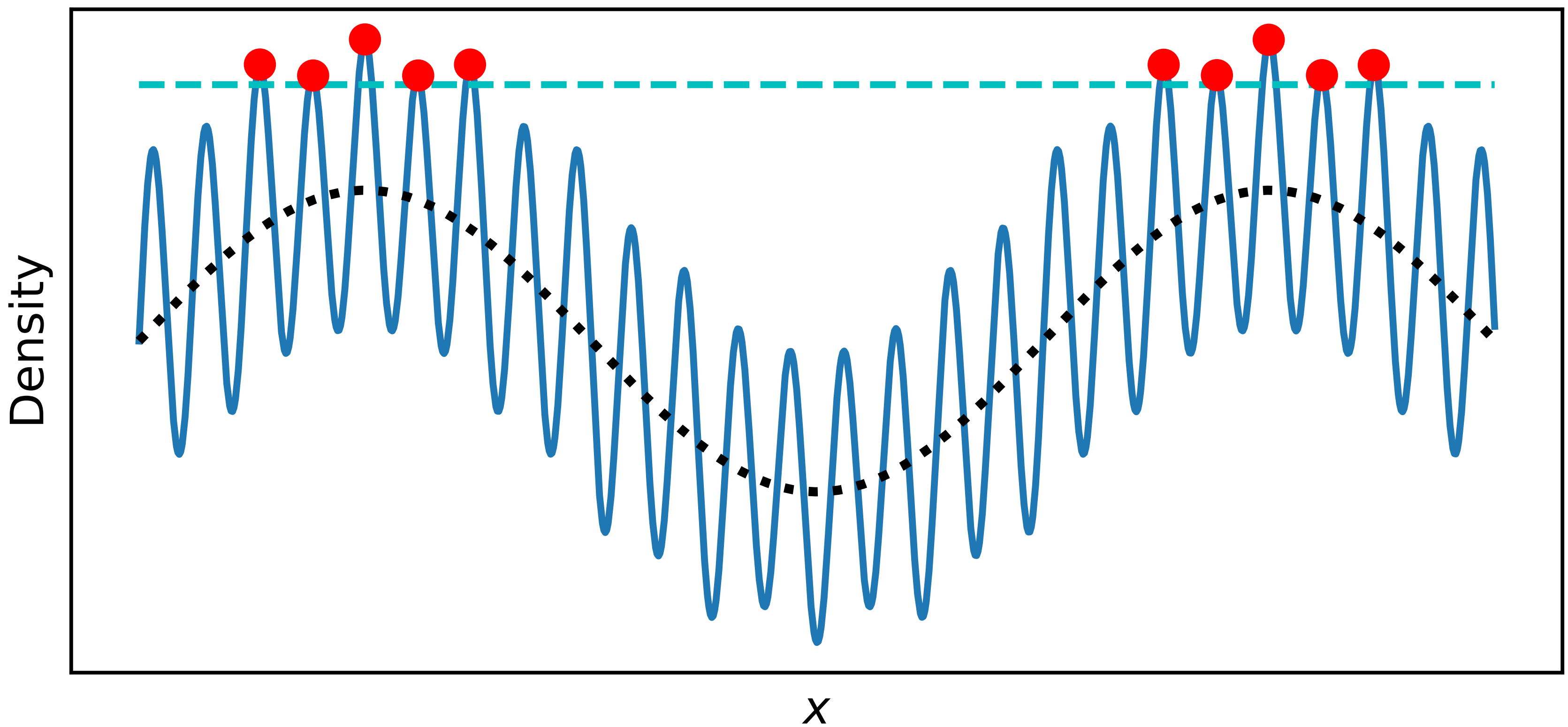
Challenge II: bias

- Different "tracers" of large-scale structure are related to the underlying perturbations in density and potential differently.
- The connection between how the tracer clusters and how the matter clusters is known as <u>bias</u>, and dealing with bias is one of the big challenges in modeling large-scale structure.
  - For example, more luminous galaxies tend to be more clustered than less luminous galaxies, even though both trace the same underlying density field.

(This effect has been modeled out in the Planck figure!)

Bias, peaks and EFT

- To make contact with galaxies, QSOs, 21 cm, Lyα, etc. we need to include bias.
- Simplest model: galaxies form at peaks in the initial density field:



Bias, peaks and EFT

- Write $\delta_{\rm gal}$ as a functional of the initial (long wavelength) density, velocity and potential fields: $\delta_{\rm gal}[\delta, \partial\mathbf{v}, \partial\partial\Phi, \cdots]$
- Coefficients of an expansion in e.g. δ are bias coefficients.

$$\delta_{\rm gal}(\mathbf{x}) = b_1\delta(\mathbf{x}) + b_2\delta^2(\mathbf{x}) + \cdots + \mathrm{stochastic} + \cdots$$

- Bias coefficients incorporate our uncertainty about complicated galaxy formation physics in addition to UV effects.
  - Dark matter halo formation, merger history, ...
  - Chemistry and gas cooling.
  - Star formation, SNe, AGN
  - Thermal and kinetic feedback
  - Background radiation

Bias, peaks and EFT

- While the process that form and shape galaxies and other objects are complex, all such objects arise from simple initial conditions acted upon by physical laws which obey well-known symmetries.
- For non-relativistic tracers these are
  - the equivalence principle
  - translational, rotational and
  - Galilean invariance.
- This highly restricts the kinds of terms that can arise in a bias expansion, no matter how complex the history.

Symmetry arguments are extremely powerful for bias since we really don't understand the small-scale physics of bias.

Aside: a new, hybrid technique

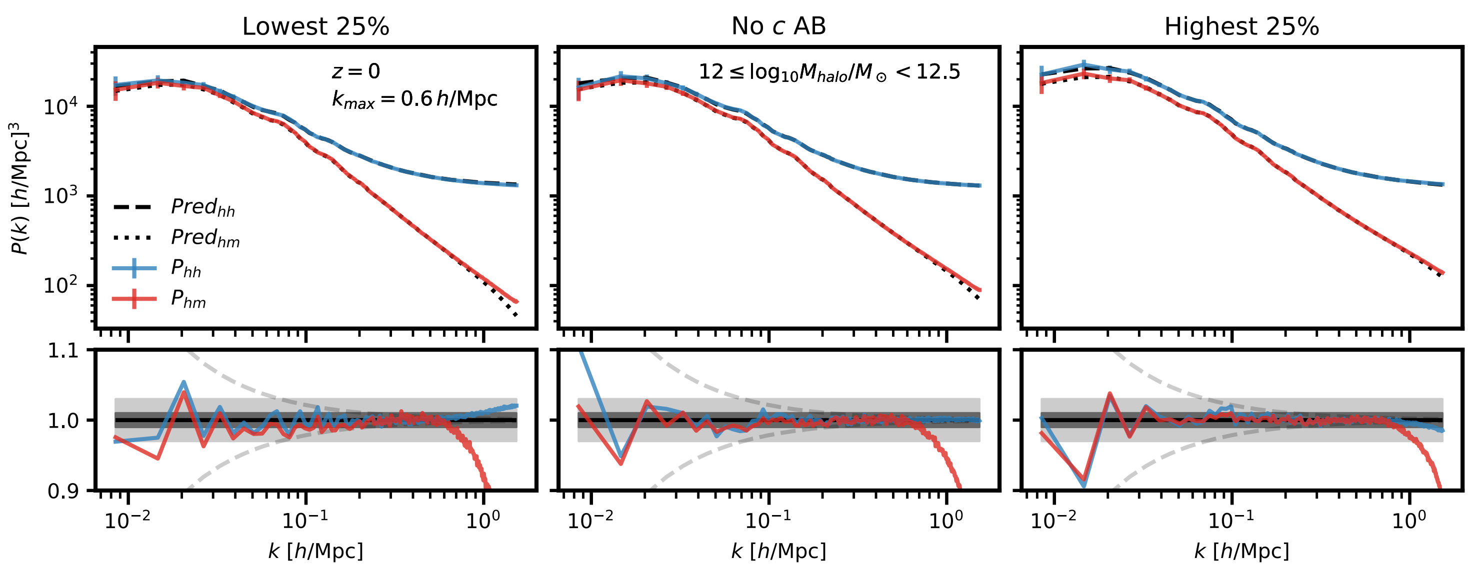
- Use dynamics from N-body simulations, but the symmetries-based bias technique from perturbation theory.
  - Generate initial conditions as before.
  - Measure density, tidal fields, etc. for each particle.
  - Weight each particle

$$\text{weight} = 1 + b_1\delta_{\rm ic} + b_2\left(\delta_{\rm ic}^2 - \langle\delta_{\rm ic}^2\rangle\right) + \cdots$$

  - Move the particles using an N-body code, and bin them to form a weighted density field: the “biased field”.
- This can be used to produce power spectra, as above, but it can also generate all of the polyspectra.
- Since it works at the field level, it can also be combined with new forward modeling techniques.

The Aemulus project

The combination of N-body displacements with symmetries-based bias expansion can be used as part of a "power spectrum emulator".
The emulator manages to fit mock catalog data for "3×2pt analyses" to 1-2% even for samples with assembly bias and other complex selections and even including hydrodynamics.

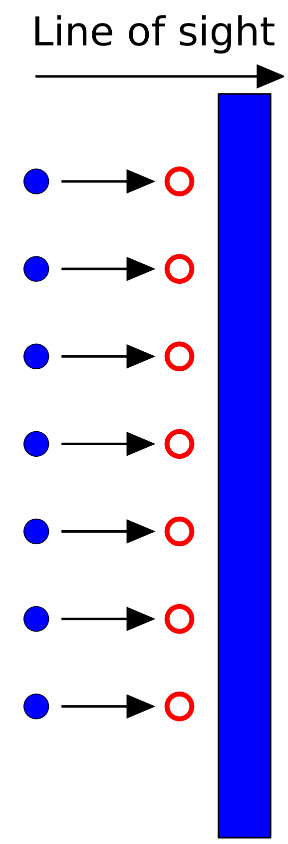


Kokron+20

Challenge III: redshift-space distortions (RSD)

We observe structure in redshift space: challenge & opportunity.

- $z_{\mathrm{obs}} \sim Hr + v_{\mathrm{pec}}^{\mathrm{los}}$.
- v_{pec} sourced by gravity, which is sourced by densities!
- Since $\delta = \rho/\bar{\rho} - 1$ grows by inflow of material, shifting by v_{pec} is like "looking into the future", but only in the line-of-sight direction!
- Comparison of clustering along and across the line-of-sight is a measure of growth rate.

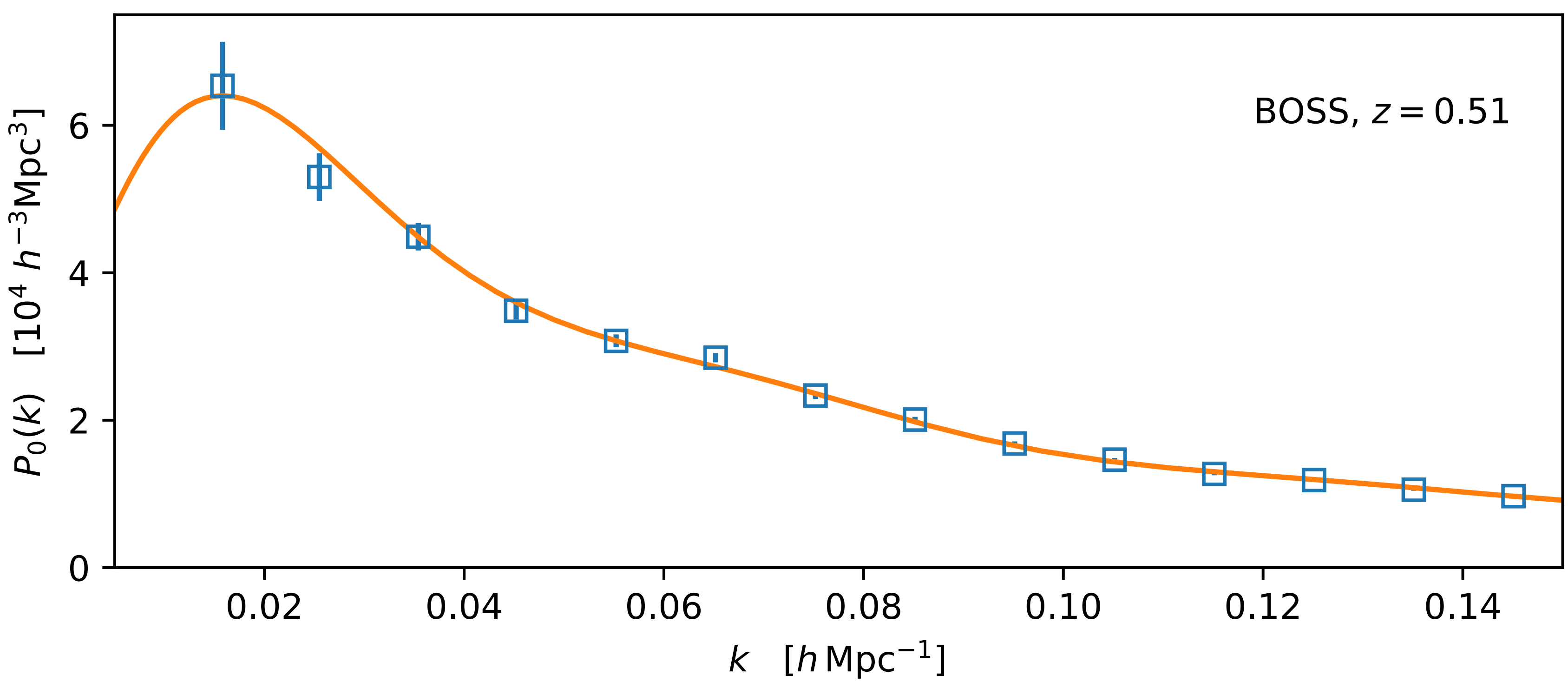


We constrain growth rate of structure by measuring the anisotropy in the clustering.

Up to the challenge

Key cosmological constraints from the latest generation of redshift surveys come from theories such as this!

- Constraints on cosmological parameters.
- Constraints on dark energy (distance scale).
- Constraints on neutrino masses and light relics.
- Constraints on primordial features and non-Gaussianity.



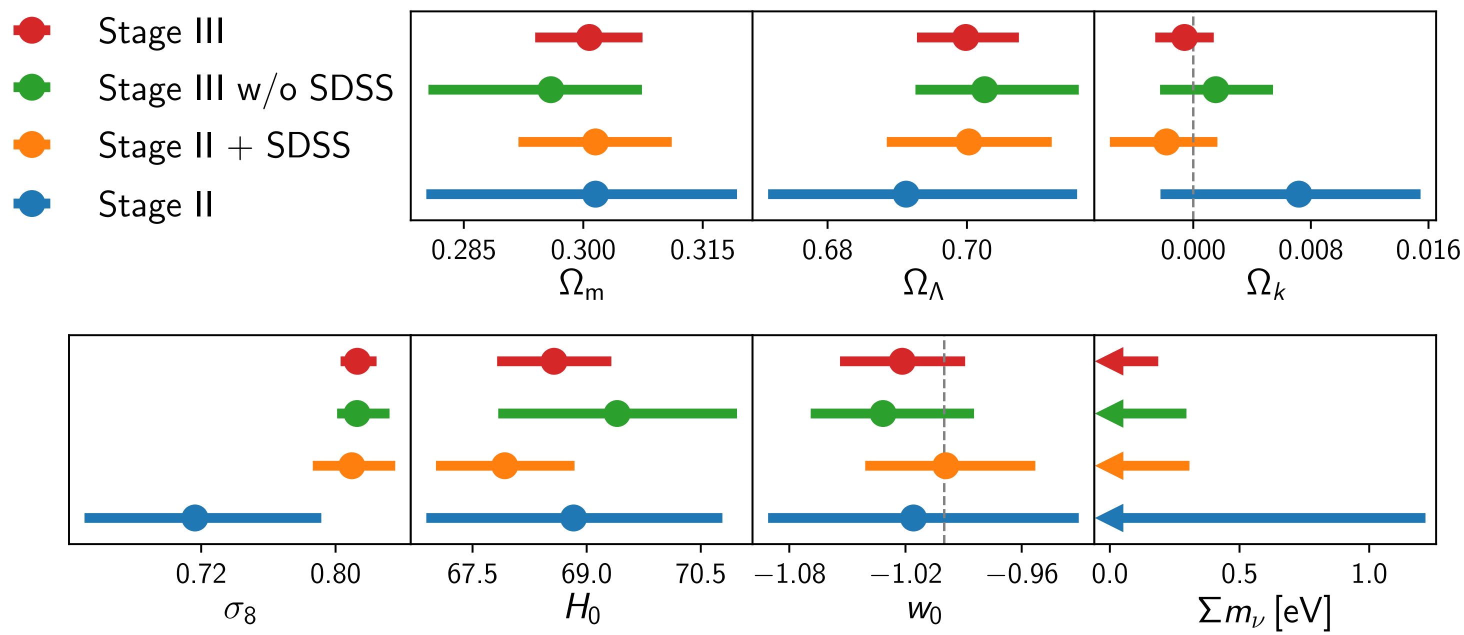
Velocileptors

- We have a public, Python package for these models.
- Being used in a number of surveys and data analyses now.
- Many ways to combine velocities and densities in power spectra: direct PT expansion, moment expansion, Gaussian streaming model, Fourier streaming model.
- Available in both LPT and EPT variants (allowing cross-checks!)
- Works in Fourier and configuration space.
- Fast and "easy to use".

http://github.com/sfschen/velocileptors

A decade of progress

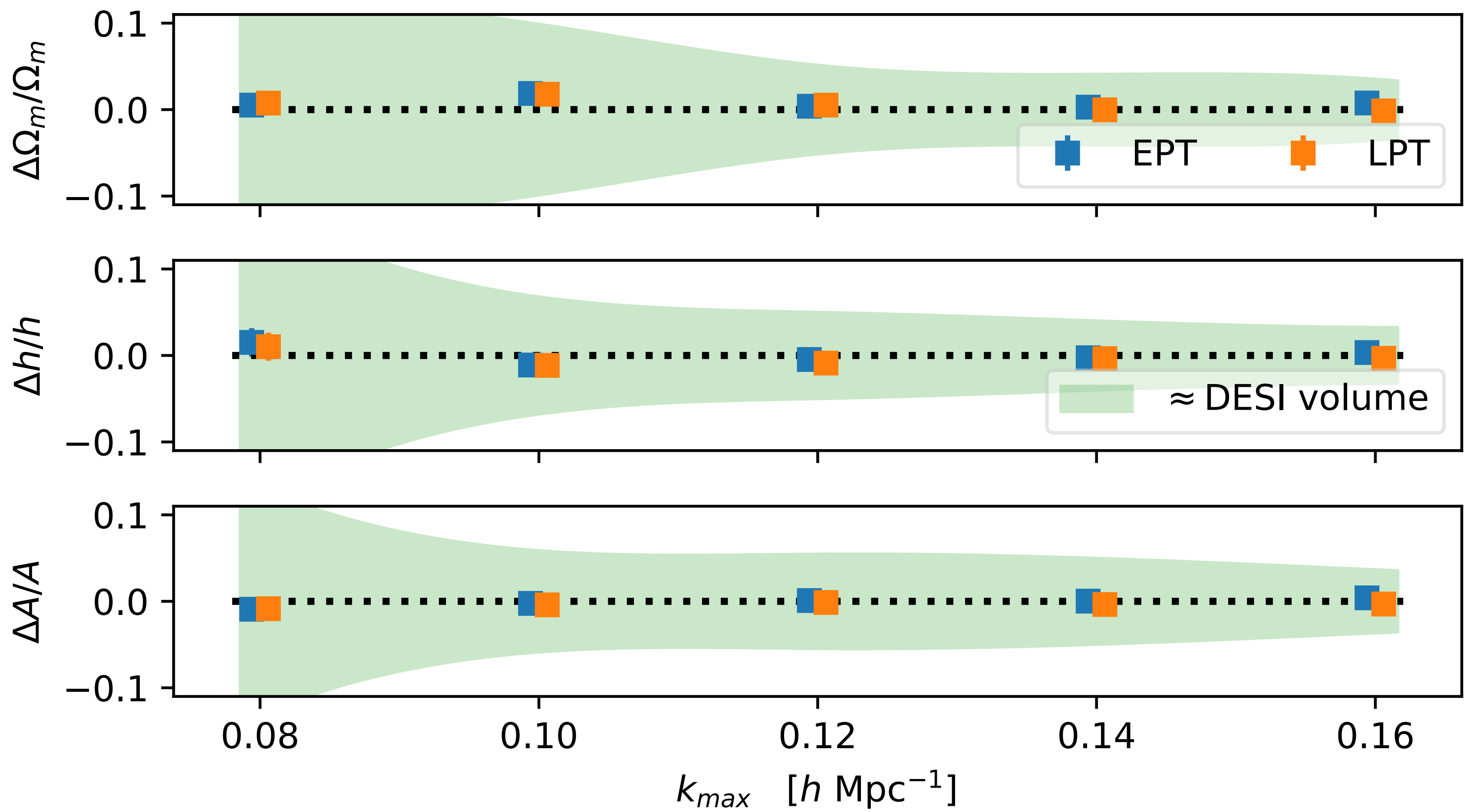
The SDSS surveys have used these kinds of perturbative models for their cosmology results.
This figure shows a decade of progress in using cosmology to constrain parameters:



eBOSS collaboration (2020)

PT blind challenge

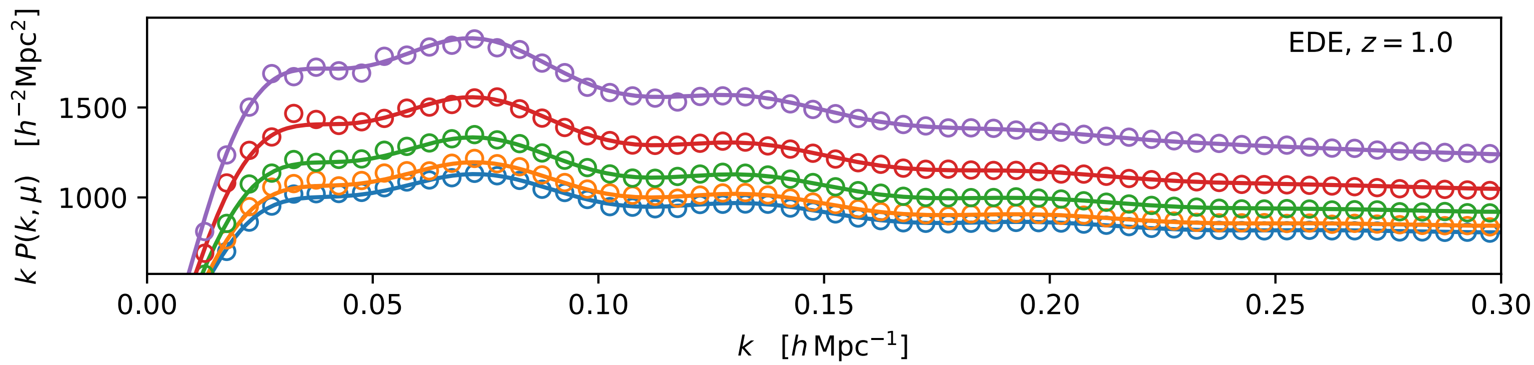
Inferring parameters from fits to mock survey data:



"PT blind challenge", Nishimichi+20

Features induced in $P(k)$ by expansion history

- Current theories achieve sub-percent accuracy on the scales most relevant to cosmological interpretations.



  - Future surveys could measure the effects of EDE or light degrees of freedom to percent level after the Universe was a few years old.

- Different schemes produce slightly different predictions – and the why is interesting – but agree well enough for next-generation experiments.
- Predictions for lensing and matter clustering can be even more precise, while involving fewer parameters.

Conclusions

- We are in the midst of the "golden age of cosmological surveys".
- Increasing survey power is driving a renaissance in analytic models of large-scale structure.
  - More perturbative modes at higher precision!
  - Form and techniques familiar from other areas of physics.
  - A few "cosmology" wrinkles.
- The models are well motivated and work well on current data.
  - Well motivated inference problem.
  - Allow us to forecast performance of future surveys reliably.
  - Survey optimization.
- Adding "beyond standard model" physics or new probes is an active area of research.
- Why the approaches differ in the way they do is still not fully understood... after 50 years we still don't understand structure formation as well as we'd like to!

The End!