Multi-Fidelity Emulation for matter power spectrum and Ly*α* flux power spectrum

Cosmology from Home 2022

Ming-Feng Ho
4th year PhD student
UC Riverside
NASA FINESST FI

Multi-fidelity emulation for matter power spectrum using Gaussian processes
Ming-Feng Ho[1,3], Simeon Bird[1], Christian R. Shelton[2] (2022 MNRAS)
arXiv:2105.01081

A multi-fidelity emulator for Lyα forest
Martin Fernandez[1,4], Ming-Feng Ho[1,3], Simeon Bird[1] (in prep)

Multi-fidelity emulation using simulation from different boxsizes
Ming-Feng Ho[1,3], Simeon Bird[1], Christian R. Shelton[2] (in prep)

[1]Physics and Astronomy, University of California, Riverside
[2]Computer Science & Engineering, University of California, Riverside
[3]NASA FINESST FI
[4]NSF GRFP

Ming-Feng Ho: jibancat.github.io
Martin Fernandez: mafern.github.io
Simeon Bird: sbird.github.io

PHYSICS & ASTRONOMY

July 2022 Virtual

Outline

- What's the problem? Why multi-fidelity
- What's emulation?
- What's multi-fidelity?
- Example 1: matter power spectrum
- Example 2: emulation for different simulation box sizes
- Example 3: Lyα 1-D flux power

Why multi-fidelity emulation?
Problem of current emulation

- **Problem:** As the simulations become increasingly more realistic & expensive, the simulation data needed to train an accurate emulator can be difficult to generate.
  - EuclidEmulator2 (2020): **needs 200+ simulations (3000^3)** in 8D even though using Latin hypercube
  - And we need to emulate beyond ΛCDM in the future … (more dimensions!)

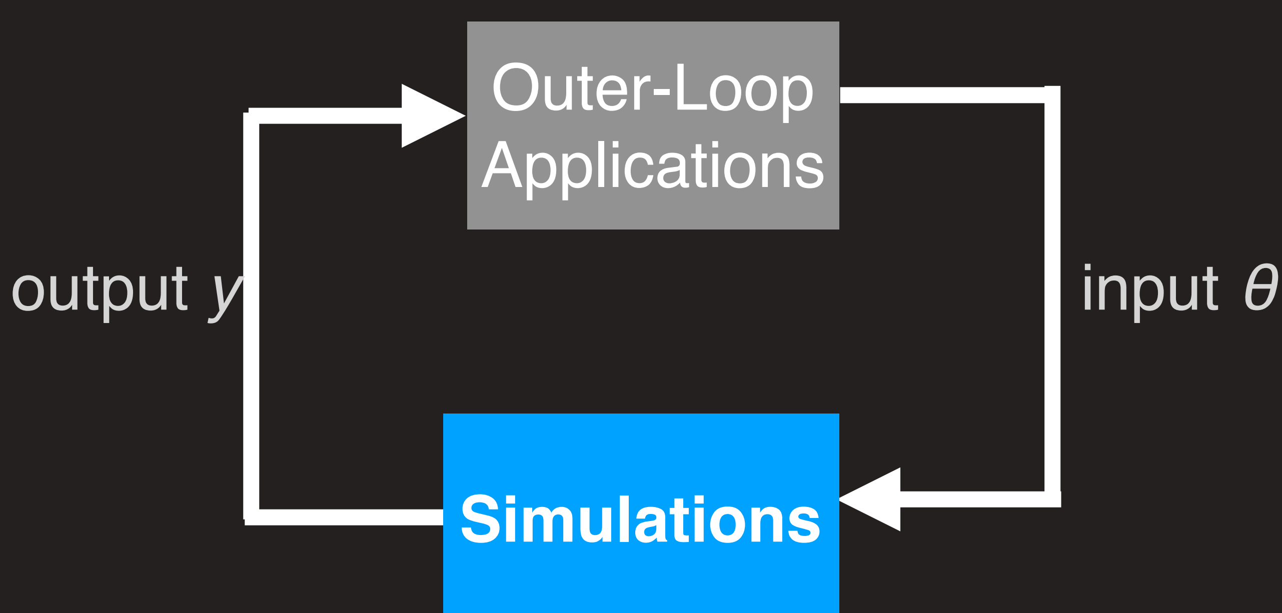
*Emulator: A model learns the mapping relationship from cosmological parameters θ to summary statistics (e.g., matter power spectrum) through a statistical learning process

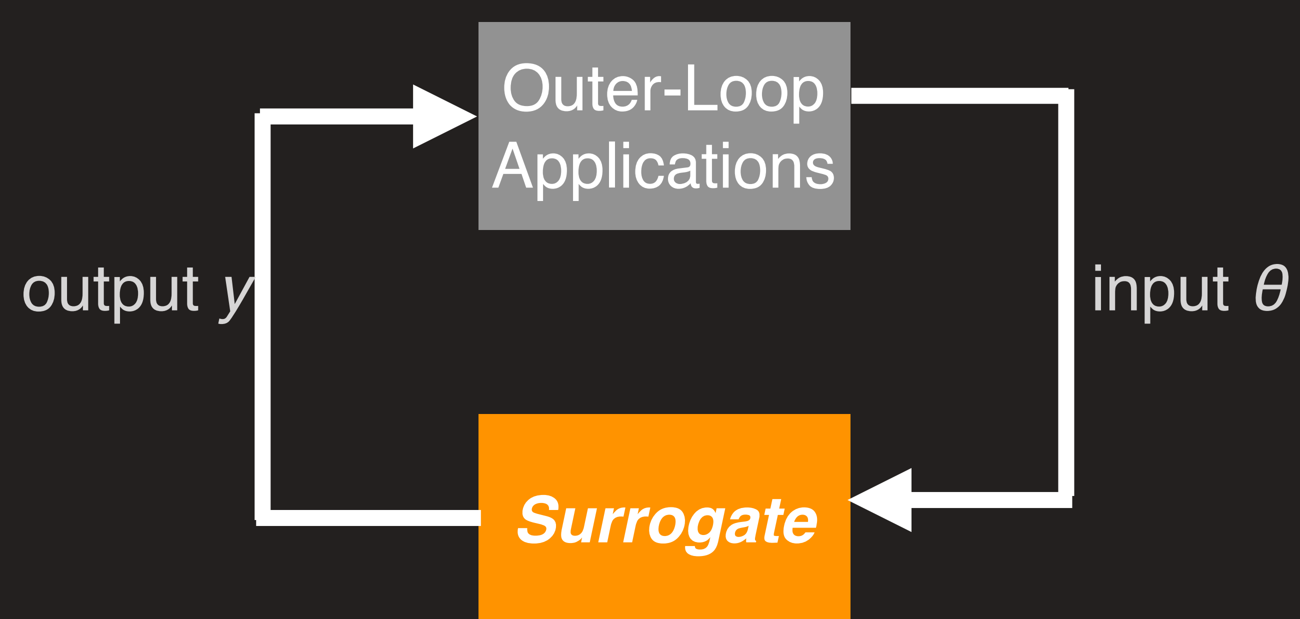
What’s emulation?
Cosmic calibration

- History: analytical calculation is not enough anymore for future surveys, we have to rely on numerical simulations to give accurate theoretical predictions …

- Key ingredients for emulation:
  - Experimental design: space filling
  - Surrogate modelling: interpolation

- Outer-loop applications:
  - Uncertainty quantification
  - Inference
  - Optimization





Perherstorfer et al (2018)
Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization

What's emulation? ~~Cosmic calibration~~

Bayesian calibration for computer experiments

- History: analytical calculation is not enough anymore for future surveys, we have to rely on numerical simulations to give accurate theoretical predictions …

- Key ingredients for ~~emulation~~ *Bayesian* modelling:
  - Experimental design: ~~space filling~~ ***Data***
  - Surrogate modelling: ~~interpolation~~ ***Prior***

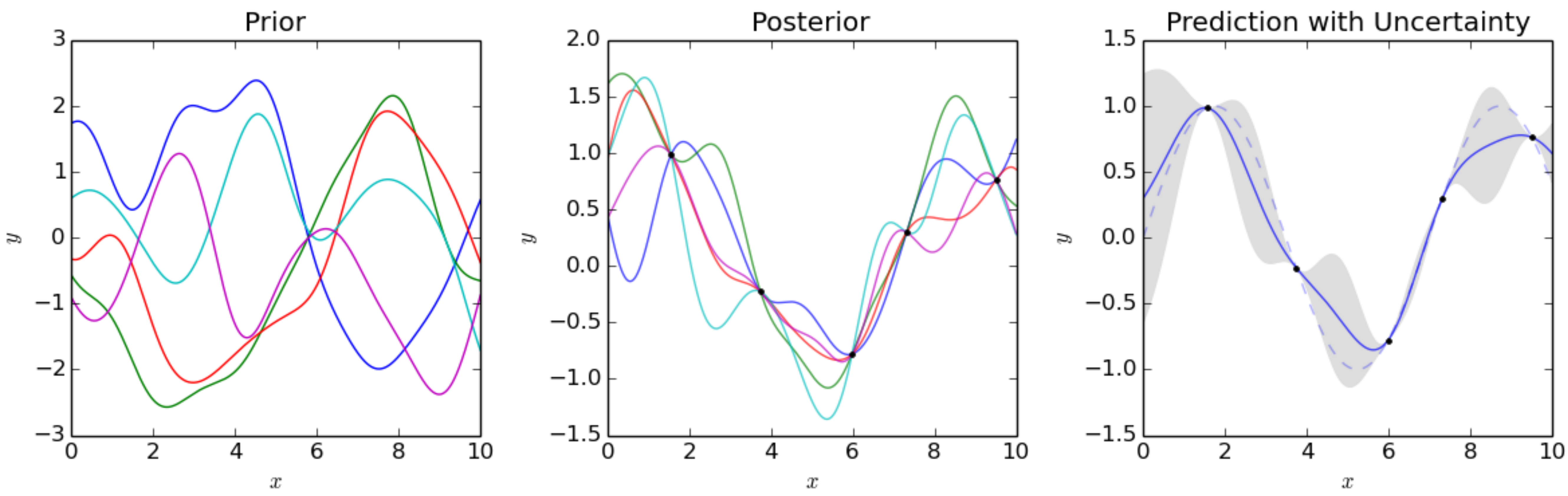
Simulations ↑

Emulation = *Posterior* predictions given *Prior* and *Data*

↓ Simulations you haven't run

↓ A distribution over smooth functions

Gaussian process: Bayesian function prediction



- Gaussian process prior: Smoothness and monotonicity features of y(x) before data are collected.
- Bayesian approach: Choose a flexible prior allowing many shapes of y(x), and let the Bayesian machinery to direct the details of the predictions.*

image credit: wikipedia

* This Bayesian attitude is mentioned in Santner (2003), The Design and Analysis of Computer Experiments

What's emulation?
Roadmap of building an emulator

Bayesian calibration of computer experiments

Simulations

- **N-body simulations** — Most of current applications are N-Body
- **Baryonification** — BACCO: Arico (2020), Giri (2021)
- **Hydro simulations** — CAMELS: F Villaescusa-Navarro et al, Lyman-α: Bird (2019), Rogers (2019)

… *More*!

Surrogates

- **Gaussian process**
- **Neural Network**

… *More*!

Applications

- **Matter power** — Heitmann (2006), Habib (2007), etc. Coyote Universe: 37 highRes; Mira-Titan: 36 highRes; EuclidEmulator: 250 highRes
- **Halo mass functions** — Aemulus: T McClintock (2018); Mira-Titan: Bocquent (2020)
- **Cosmic shear** — Harnois-De´raps (2019), Giblin (2019)
- **Field level** — KiDs: Yiu (2021); Neerav (2021), CAMELS papers
- **Lyα flux power** — Bird (2019), Rogers (2019), Rogers (2021), Pedersen (2021)
- **21cm power** — Sudipta (2022)
- **Gravitational wave** — Damon H. T. (2022)
- **Exoplanet** — Rogers (2021), they use Random Forest

… *More*!

What's emulation?
Roadmap of building an emulator

Simulations

First cosmic emulator- Heitmann (2006)

Bayesian calibration of computer experiments

N-body simulations
Most of current applications are N-Body

Baryonification
BACCO: Arico (2020), Giri (2021)

Hydro simulations
CAMELS: F Villaescusa-Navarro et al, Lyman-α: Bird (2019), Rogers (2019)

… *More*!

Surrogates

Gaussian process

Neural Network

… *More*!

Applications

Matter power
Heitmann (2006), Habib (2007), etc.
Coyote Universe: 37 highRes
Mira-Titan: 36 highRes
EuclidEmulator: 250 highRes

Halo mass functions
Aemulus: T McClintock (2018)
Mira-Titan: Bocquent (2020)

Cosmic shear
Harnois-De´raps (2019), Giblin (2019)

Field level
KiDs: Yiu (2021)
Neerav (2021),
CAMELS papers

Lyα flux power
Bird (2019), Rogers (2019), Rogers (2021), Pedersen (2021)

21cm power
Sudipta (2022)

Gravitational wave
Damon H. T. (2022)

Exoplanet
Rogers (2021), they use Random Forest

… *More*!

What's emulation?
Roadmap of building an emulator

Bayesian calibration of computer experiments

Simulations

CAMELS

N-body simulations

Most of current applications are N-Body

Baryonification

BACCO: Arico (2020),
Giri (2021)

Hydro simulations

CAMELS: F Villaescusa-Navarro et al,
Lyman-α: Bird (2019), Rogers (2019)

… *More*!

Surrogates

Gaussian process

Neural Network

… *More*!

Applications

Matter power

Heitmann (2006), Habib (2007), etc.
Coyote Universe: 37 highRes
Mira-Titan: 36 highRes
EuclidEmulator: 250 highRes

Halo mass functions

Aemulus: T McClintock (2018)
Mira-Titan: Bocquent (2020)

Cosmic shear

Harnois-De´raps (2019),
Giblin (2019)

Field level

KiDs: Yiu (2021)
Neerav (2021),
CAMELS papers

Lyα flux power

Bird (2019), Rogers (2019),
Rogers (2021), Pedersen (2021)

21cm power

Sudipta (2022)

Gravitational wave

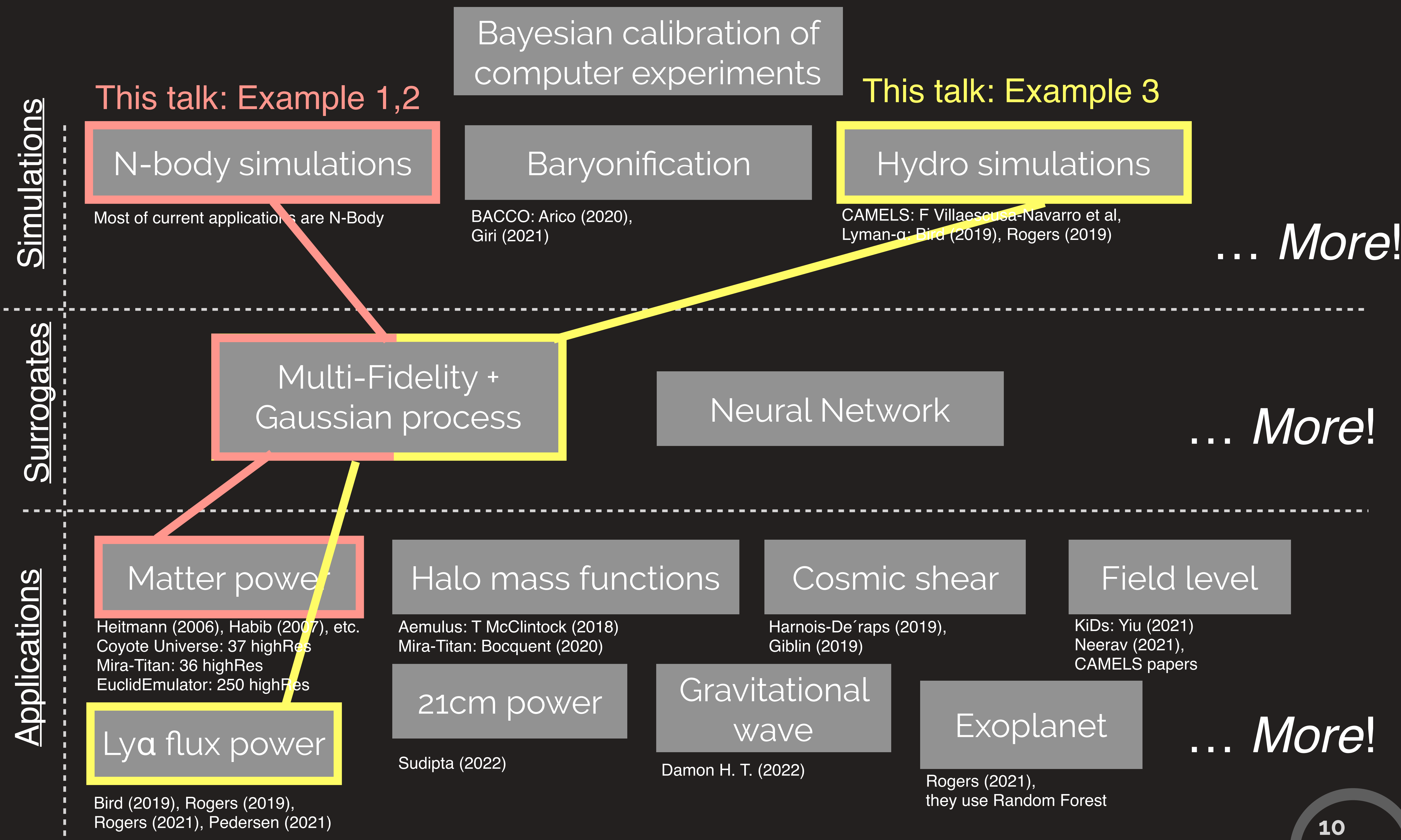
Damon H. T. (2022)

Exoplanet

Rogers (2021),
they use Random Forest

… *More*!

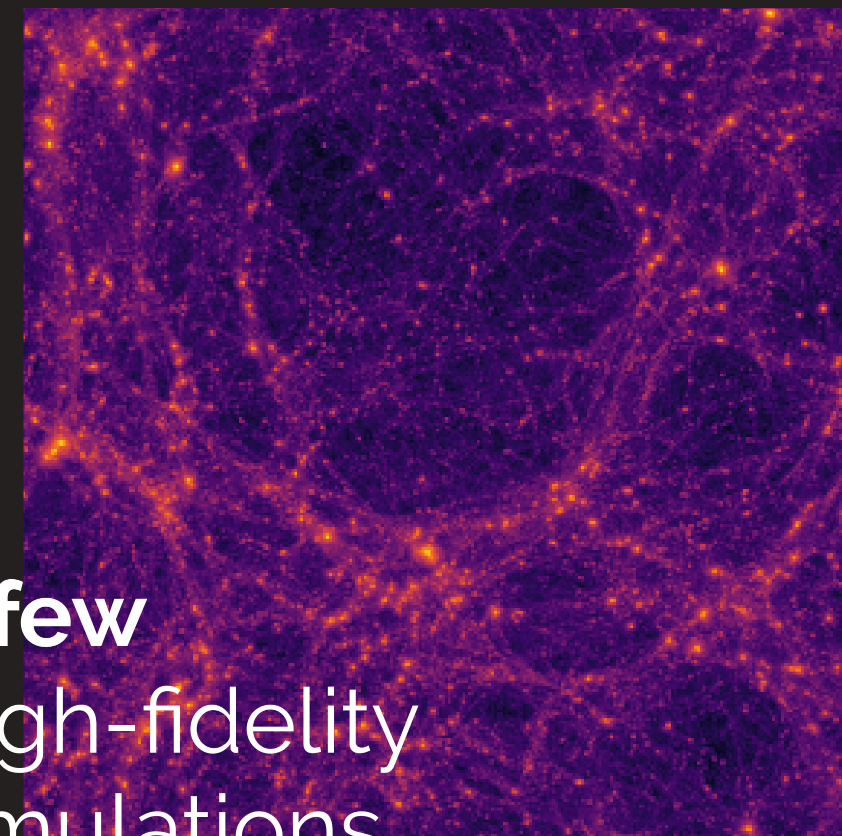
What's emulation?
Roadmap of building an emulator



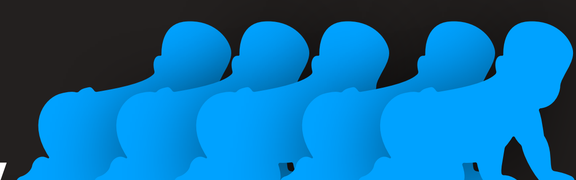
What’s multi-fidelity?
An analogy

Many
Low-fidelity
simulations

+



A few
High-fidelity
simulations



Many
Grad Student Researchers

+

A few
Professors

= efficiency + accuracy

What's multi-fidelity?
Statistical modelling - regression view

- KO method (AR1): linear autoregressive GP

$$f_{\mathrm{HF}}(x) = \rho f_{\mathrm{LF}}(x) + \delta(x).$$

Bias function, a GP

Low-fidelity output, a GP

Scale, a value

- Nonlinear Autoregressive Gaussian Process (NARGP): An improved KO method using a deep GP

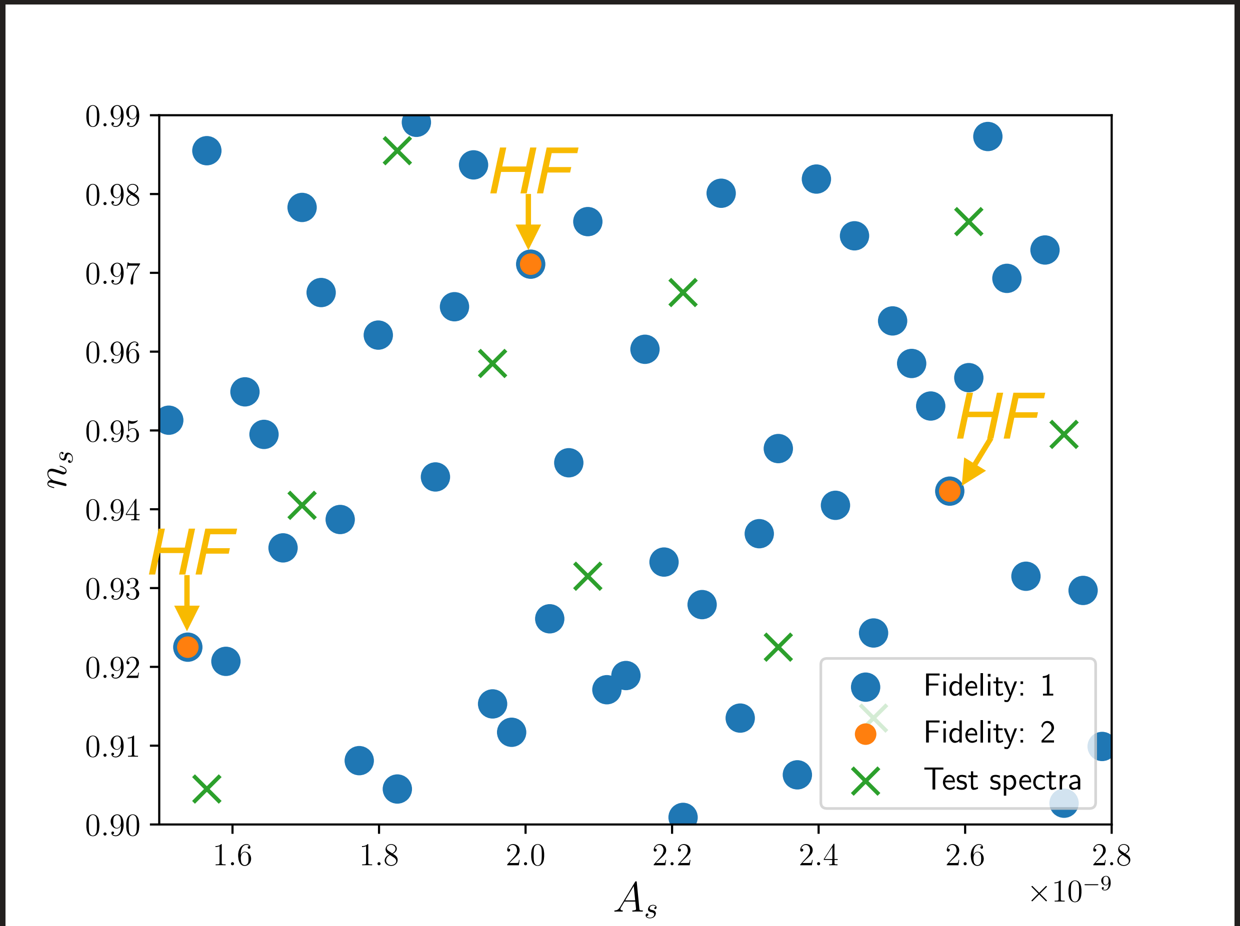
$$f_{\mathrm{HF}}(x) = \rho(x, f_{\mathrm{LF}}(x)) + \delta(x).$$

Scale function, a GP

KO: Kennedy & O'Hagan (2000)
NARGP: Perdikaris et al. (2017)

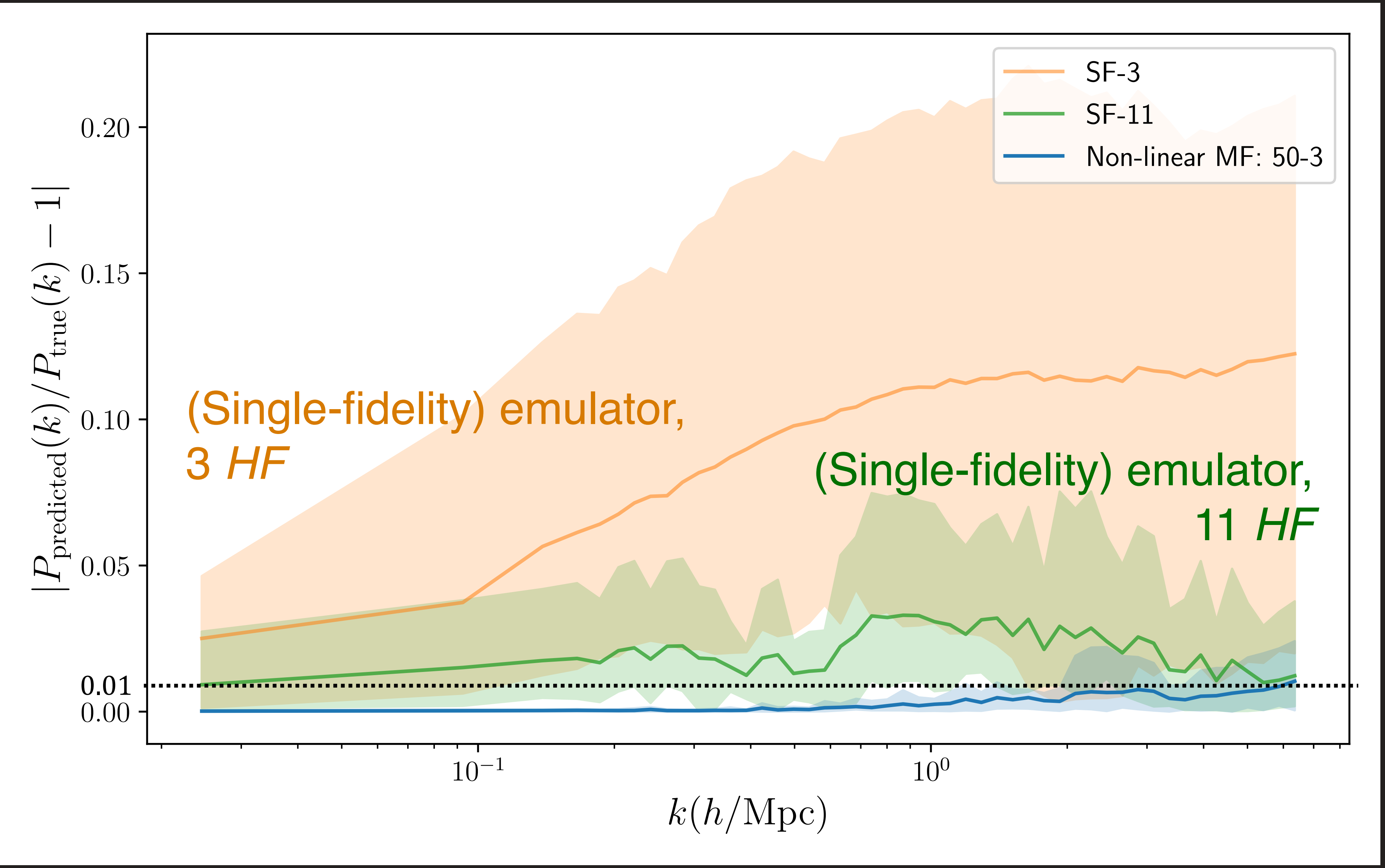
Example 1: matter power spectrum
Experimental design

- Parameters: $(h, \Omega_0, \Omega_b, A_s, n_s)$
- *Low-fidelity*: space-filling strategy (Latin hypercube)
  - $128^3, 256\,\mathrm{Mpc\,h^{-1}}$
- *High-fidelity*: a subset of low-fidelity runs
  - $512^3, 256\,\mathrm{Mpc\,h^{-1}}$
- *HF* choices were optimized using *LF* simulations



Ho, Bird, Shelton (2022)

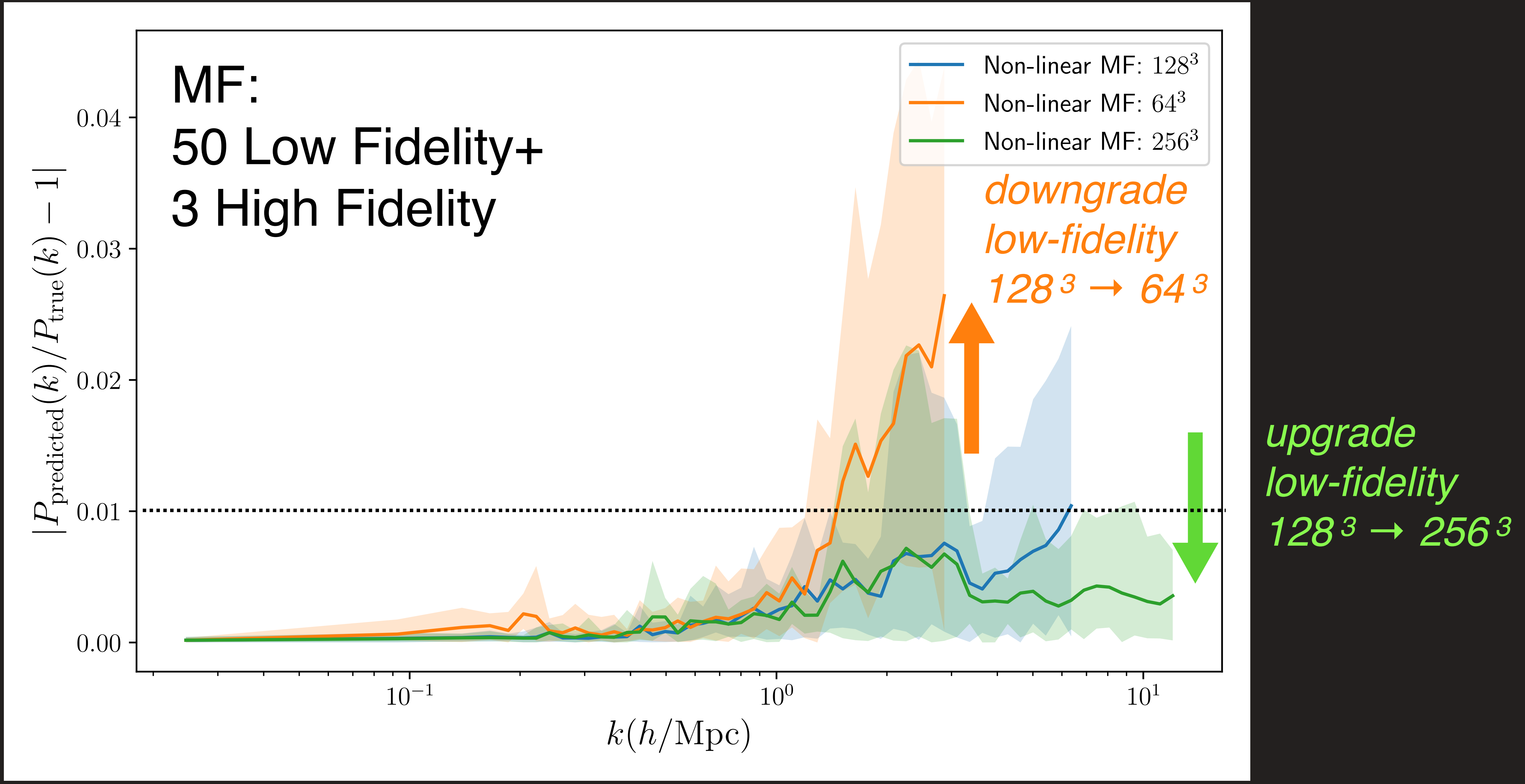
Example 1: matter power spectrum
Emulation error (z = 0)



Multi-fidelity, 50 *LF* + 3 *HF*

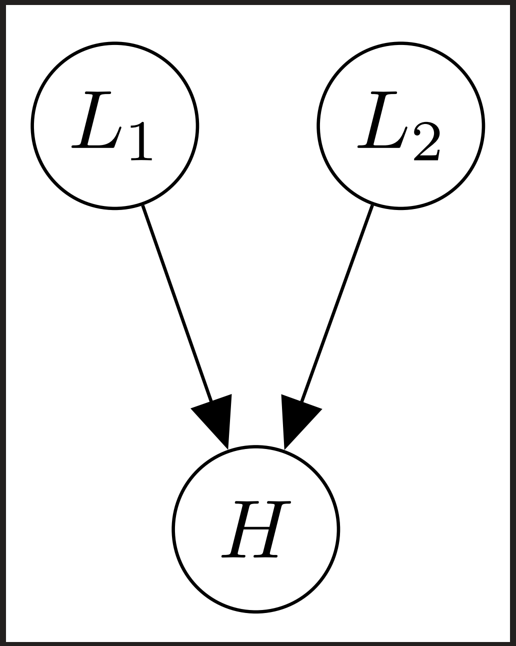
- The cost of 50 LF + 3 HF simulations is ~ 4 HF simulations.
- Still, the accuracy of MFEmulator is even better than using 11 HF.

Example 1: matter power spectrum
Accuracy increases with a better LF training set



- The quality of LF simulations affects the accuracy of MF emulation
- Small scales emulation can be improved with a better quality of LF simulation suite. → *Question: Can we use small box LF to improve emulation?*

Example 2: matter power spectrum
Extending to using **boxsize** as a fidelity



- L_1: 128^3, 256 Mpc/h
- L_2: 128^3, 100 Mpc/h
- H: 512^3, 256 Mpc/h

- Number of particles is not the only fidelity variable, ***boxsize*** is also a fidelity variable
- A smaller boxsize, ***better*** resolution at ***small scales***
- We can combine both ***large box (L1)*** and ***small box (L2)*** information through a graphical model construction.
- A graphical GP (Ji et al., 2021) allows us to do so.

Deep Graphical Multi-fidelity GP (dGMGP)

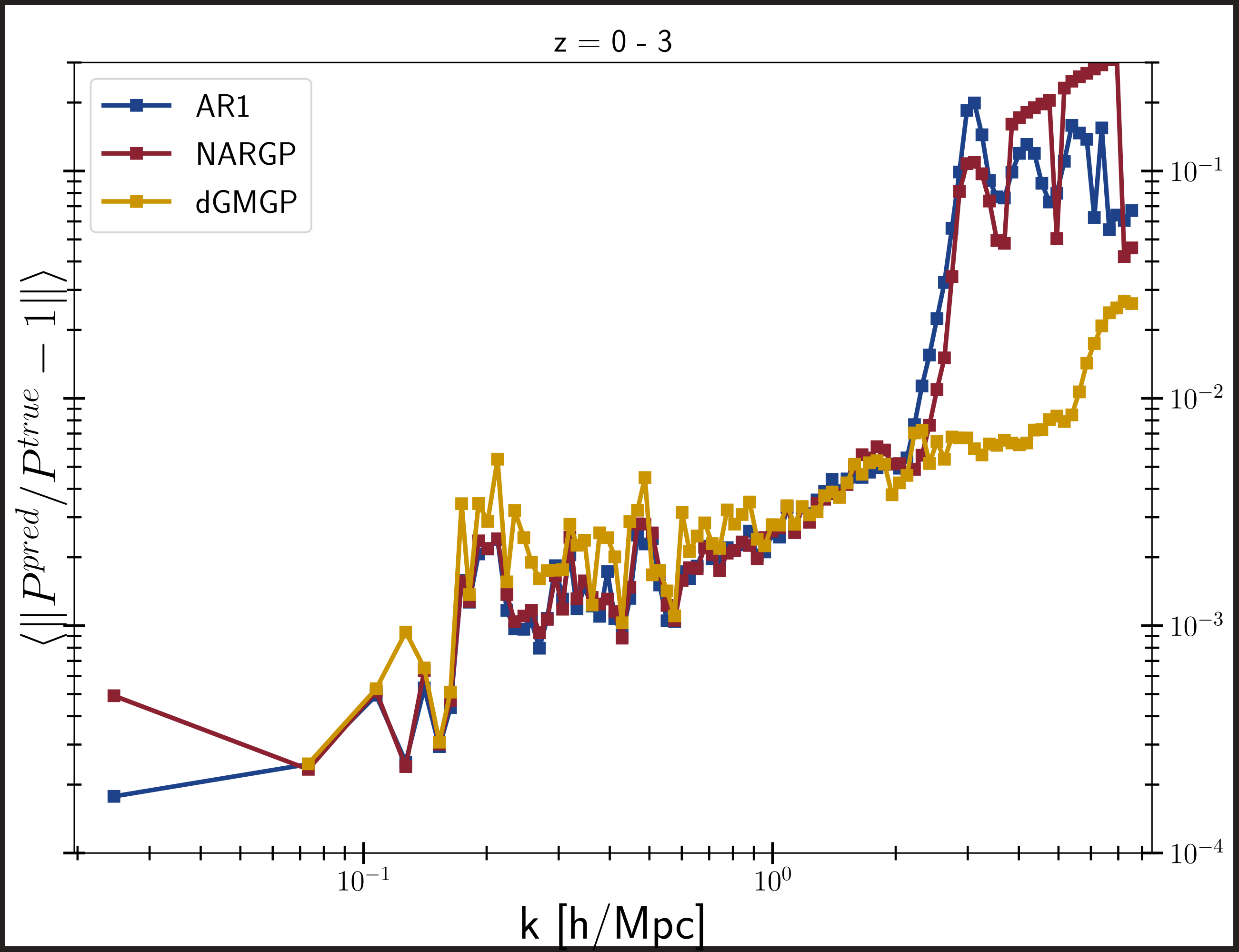
$$f_{\mathrm{HF}}(x) = \rho(\{f_{\mathrm{LF},1}, f_{\mathrm{LF},2}\} \cup x) + \delta(x)$$

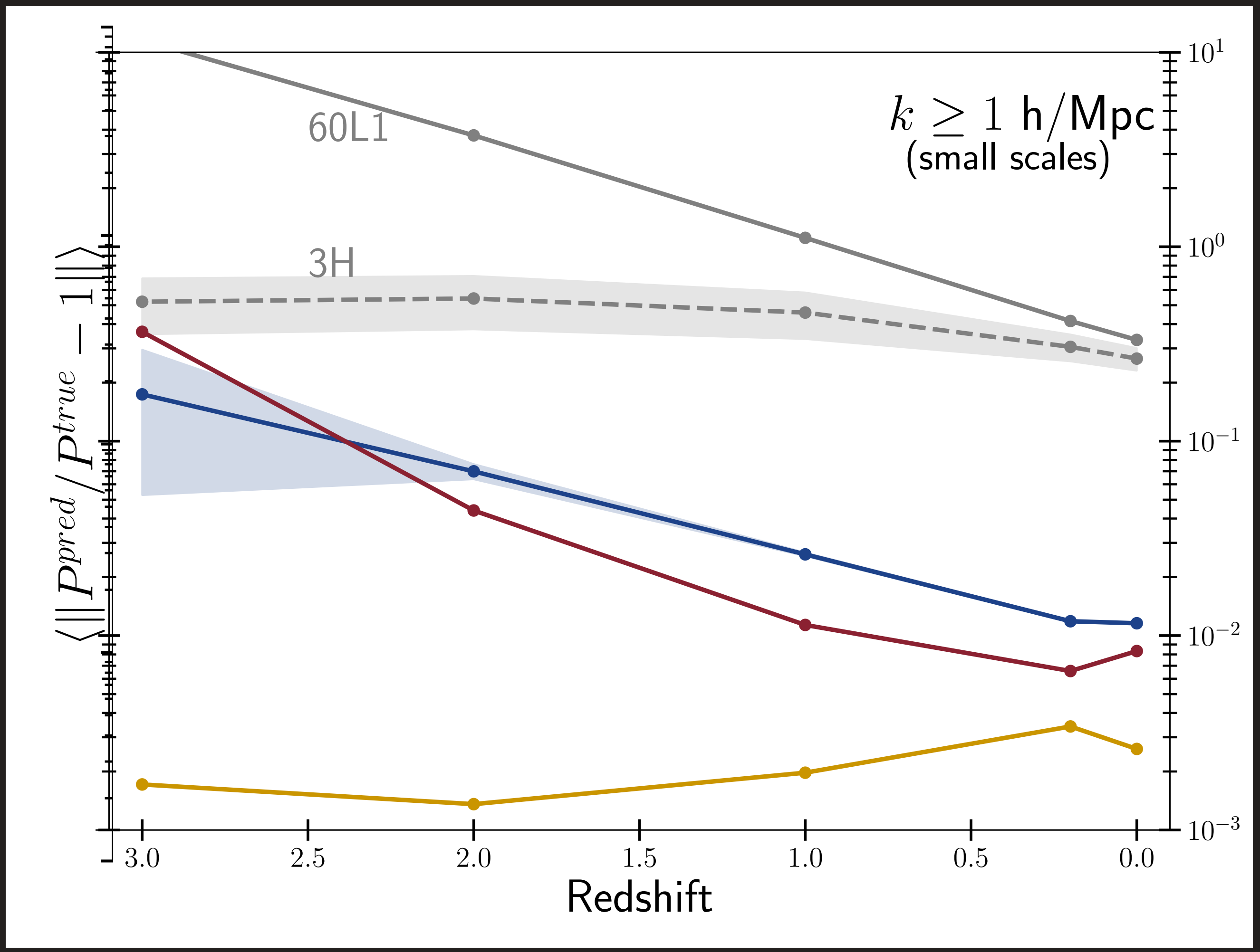
$$K([x, f_{\mathrm{LF}}], [x', f'_{\mathrm{LF}}]) = K_{\mathrm{SE}}(x, x')[K_{\mathrm{LIN}}(f_{\mathrm{LF}}, f'_{\mathrm{LF}}) + K_{\mathrm{SE}}(f_{\mathrm{LF}}, f'_{\mathrm{LF}})] + K_{\mathrm{SE}}(x, x')$$

K_{SE} : Squared-exponential kernel, guarantees smooth functions
K_{LIN} : Linear kernel, doing Bayesian linear regression

GMGP: Ji (2021)
A graphical multi-fidelity Gaussian process model, with application to emulation of expensive computer simulations

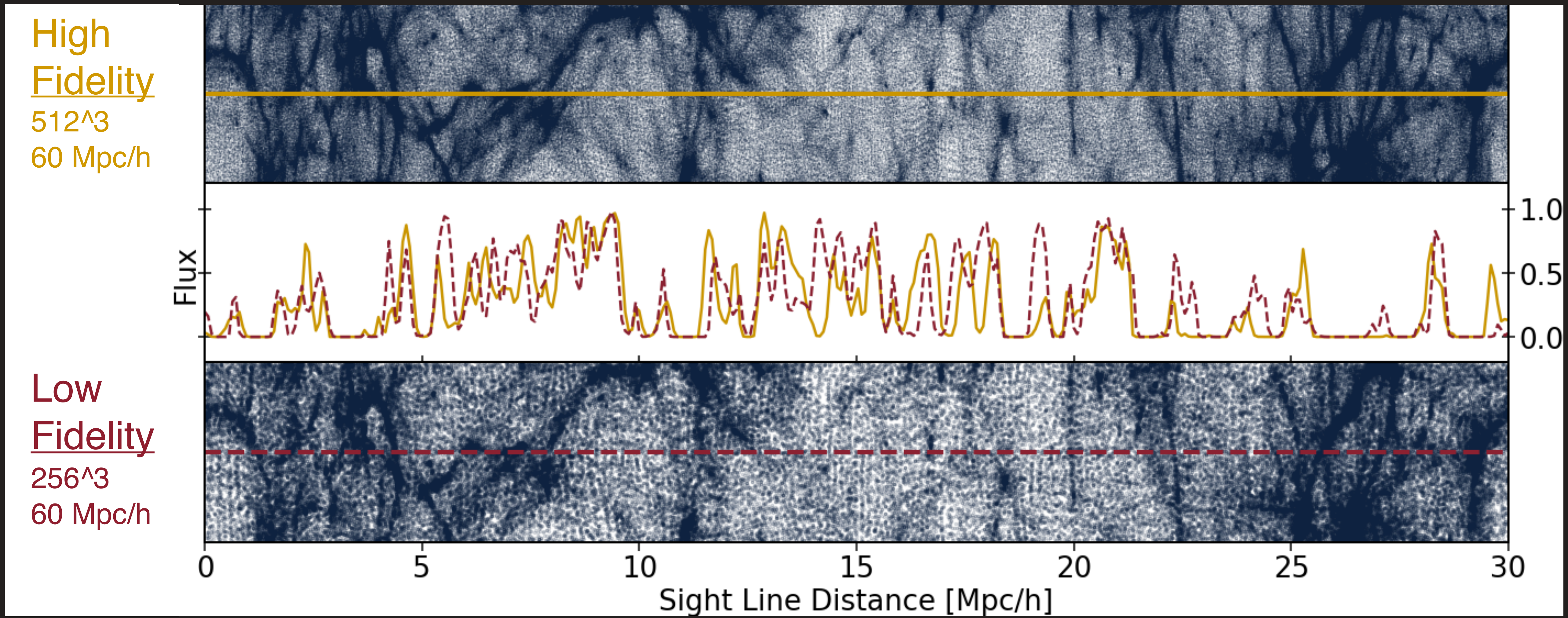
Example 2: matter power spectrum
Extending to using **boxsize** as a fidelity





- GMGP uses smaller boxsize simulations to enhance the emulation accuracy at small scales.
- GMGP outperforms AR1 and NARGP at z ≧ 1.
- It is possible to emulate HF using LF from various box sizes.

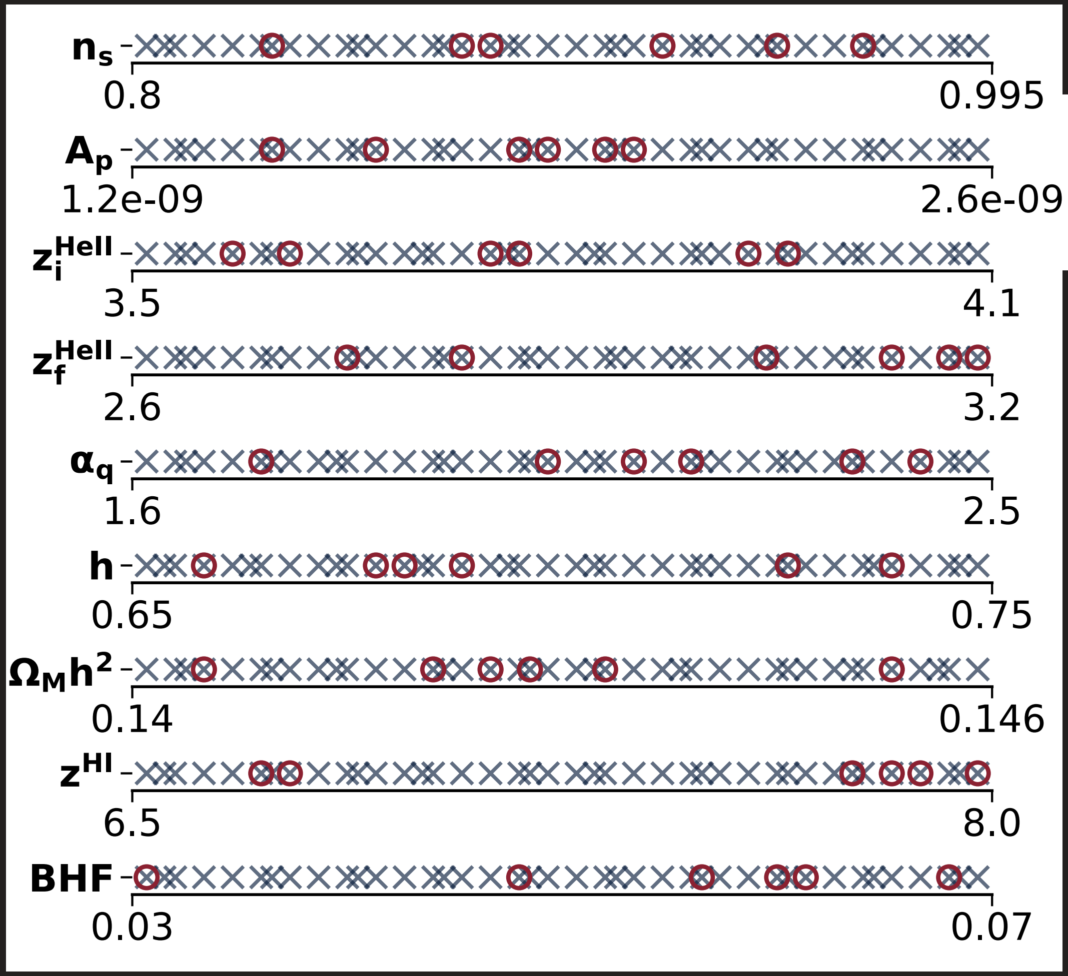
Example 3: Lya flux power spectrum
Simulated spectra generation



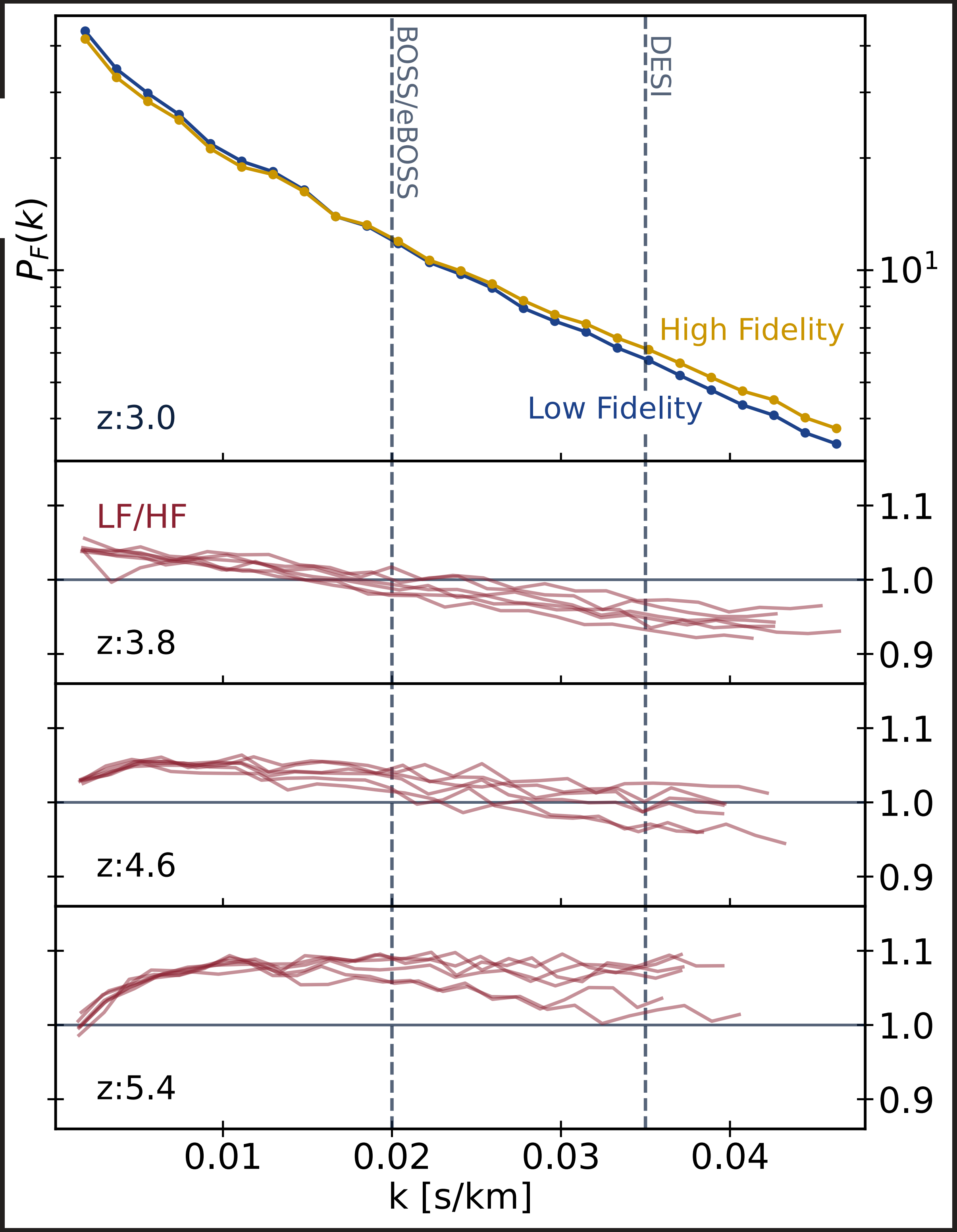
- Fake spectra: 32,000 (seeded) random skewers per snapshot using Bird (2017).
- Lya flux power spectrum: Measure correlation between neutral hydrogen within a slightline

figure credit: Martin Fernandez

Example 3: Lya flux power spectrum
Experimental design in 9 dimensions



x: LF
o: HF

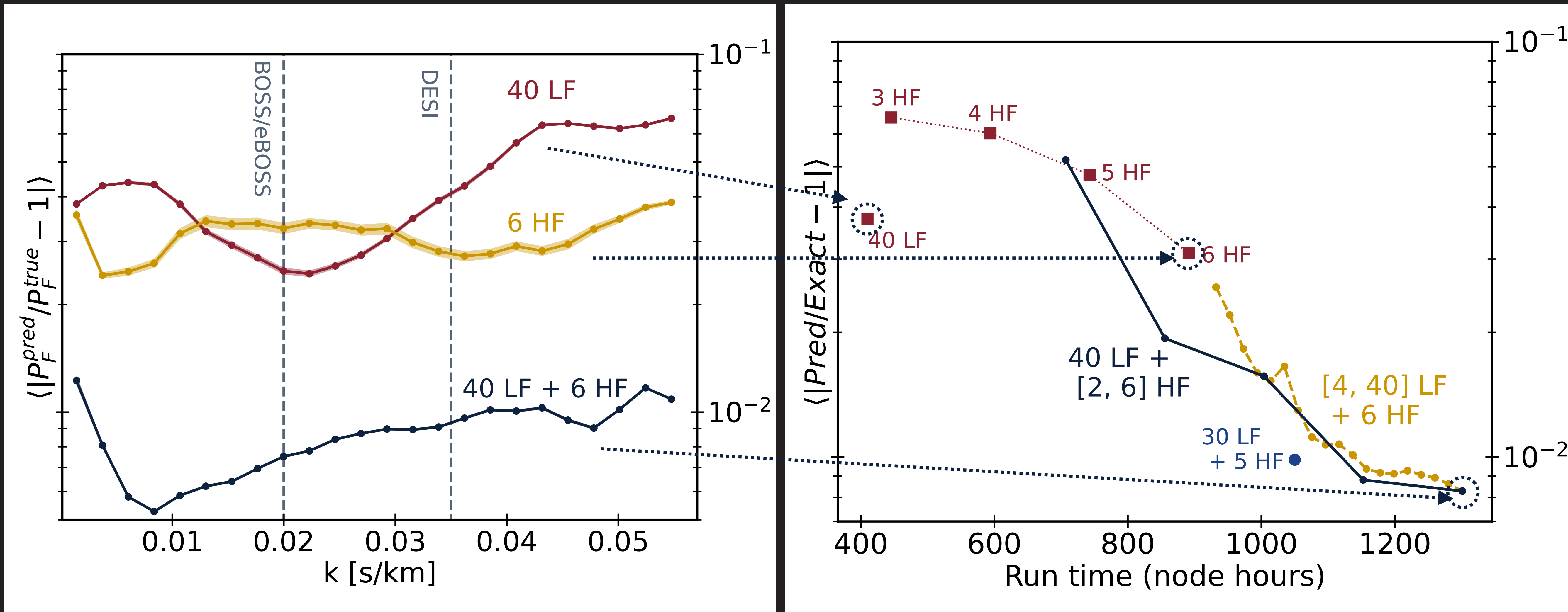


- 9 parameters (z = 2 - 5.4), including reionization parameters and black hole feedback
- The discrepancy between LF/HF appears across scales (k), varies with redshifts

figure credit: Martin Fernandez

Fernandez, Ho, Bird (2022 in prep)

Example 3: Lya flux power spectrum
Emulation errors and computational cost



- MFEmulator with 40 LF + 6 HF has ≈ 1% accuracy across all scales.
- Effectively increase the emulation accuracy by combining existing low and high-fidelity simulations.

figure credit: Martin Fernandez

Conclusion

- Save computation budget with multi-fidelity experimental designs.
- Multi-fidelity emulation can be directly applied to existing simulation suite.
- Cheaply increasing dynamic range of simulations.
- Example 1: matter power spectrum
- Example 2: matter power spectrum, emulating with different box sizes
- Example 3: Lya flux power spectrum
- Future applications: Halo mass function, cosmic shear power spectrum

arXiv:2105.01081
github.com/jibanCat/matter_multi_fidelity_emu
Papers for example 2 and 3 are expected to be submitted later this year.

We thanks Yi Ji (Duke, Stat) and Simon Mak (Duke, Stat) for kindly providing the GMGP code in Python.