Multi-Fidelity Machine Learning for Extending the Range of High-Fidelity Molecular Dynamics Data

Lucas J. Stanek[1], Shaunak D. Bopardikar[2], Michael S. Murillo[1]

[1] Dept. of Computational Mathematics, Science, and Engineering, Michigan State University
[2] Dept. of Electrical and Computer Engineering, Michigan State University

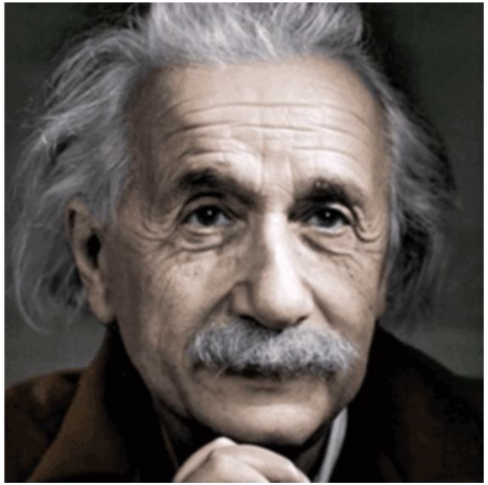
Multi-Fidelity Machine Learning for Extending the Range of High-Fidelity Molecular Dynamics Data

Lucas J. Stanek[1], Shaunak D. Bopardikar[2], Michael S. Murillo[1]

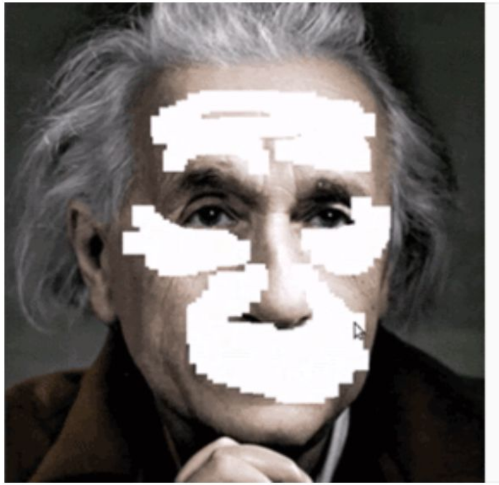
[1] Dept. of Computational Mathematics, Science, and Engineering, Michigan State University
[2] Dept. of Electrical and Computer Engineering, Michigan State University

CMSE

Image Inpainting: Filling in the gaps with Machine Learning



Original Image

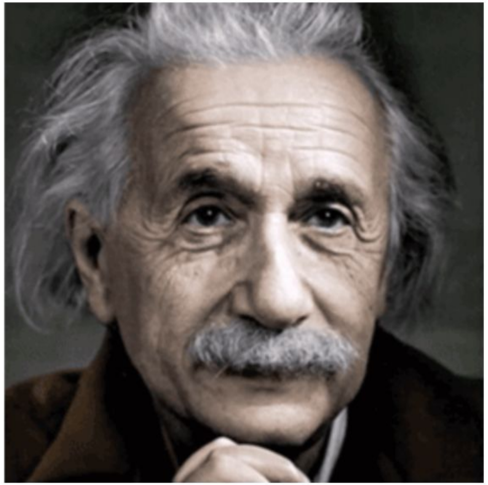


Masking

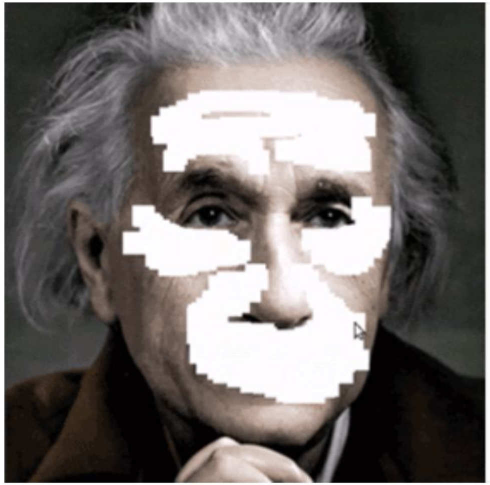
→ Utilize machine learning to make a prediction

Image: nvidia

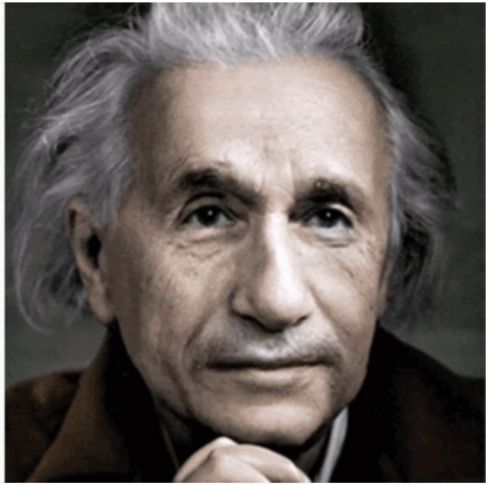
Image Inpainting: Filling in the gaps with Machine Learning



Original Image



Masking

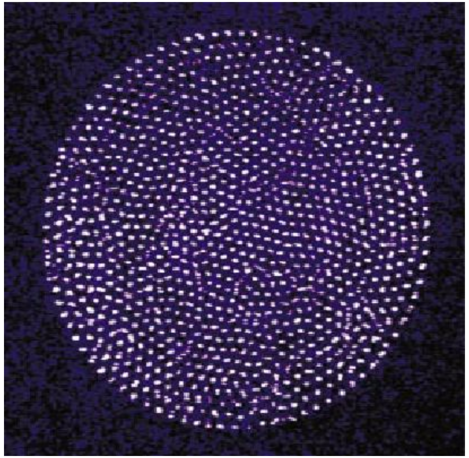


Predicted image

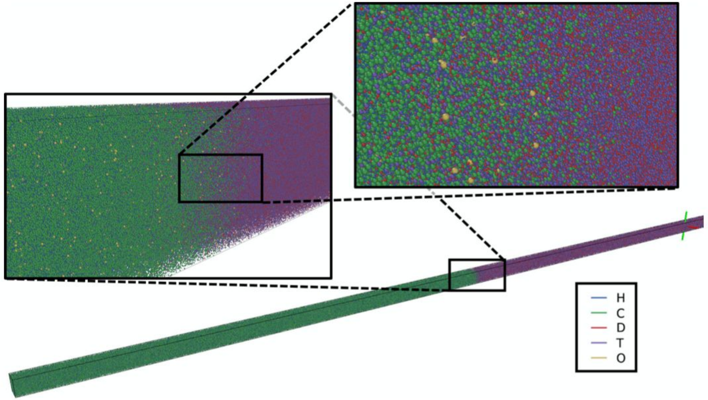
Image: nvidia

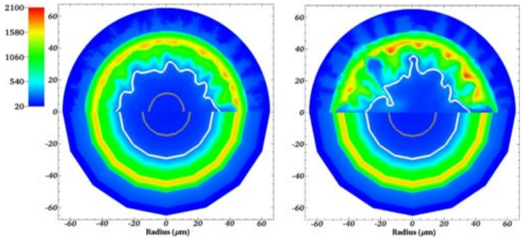
Atomic Transport in Disparate Regimes

Self-diffusion in magnetized dusty plasmas



Atomic-scale mixing in nuclear fusion simulations and experiments





1) Murillo, Physics of Plasmas 11:5 2964-2971 (2004)
2) Stanton et. al, Phys. Rev. X 8(2) (2018)
3) Rana et. al Phys. Rev. E 95(1) (2017)

Physical Models are Often Asymptotically Accurate

Γ Coulomb Coupling Parameter

$\Gamma \gg 1$ $\qquad$ $\frac{1}{\Gamma} \gg 1$

Model A $\longleftrightarrow$ Model B

Model ?

An interpolation problem

Ultracold Neutral Plasma Experiment

Kohn-Sham Molecular Dynamics

Molecular Dynamics

Kinetic Simulation

⋮

Inertial Confinement Fusion Experiment

Path Integral Monte Carlo

Analytic Theory

Hydrodynamic Simulation

⋮

Interpolation with Gaussian Process Regression (GPR)

Radial Basis Function kernel

$$k(x_i, x_j) = a \exp\left(-\frac{\|x_i - x_j\|^2}{b}\right) \longrightarrow K(X, X)$$

A measure of similarity between data points

a,b are optimizable hyper-parameters

Kernel matrix

Prediction domain

Training domain

$$f^* = K(X_*, X)K(X,X)^{-1}f$$

m x n   n x n   n x 1

Prediction

$$\sigma^2 = K(X_*, X_*) - K(X_*, X)K(X,X)^{-1}K(X, X_*)$$

m x m   m x n   n x n   n x m

Covariance matrix → Uncertainty

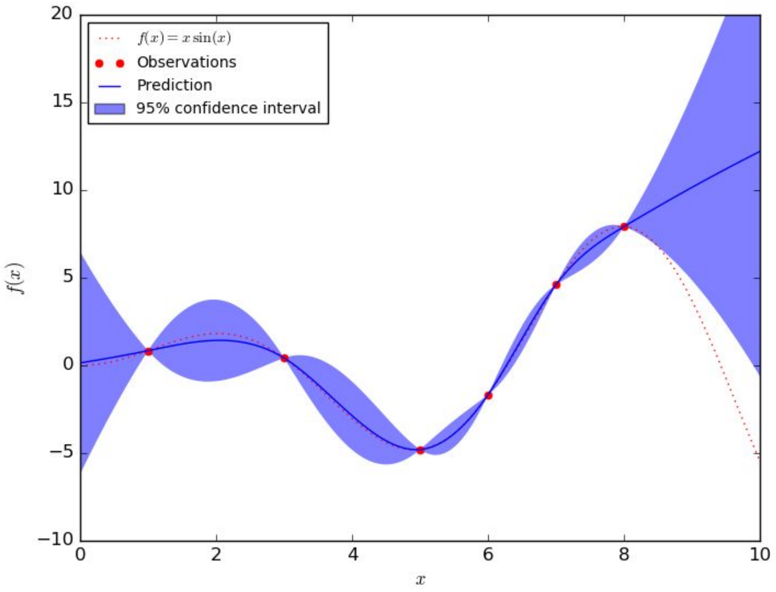


Image: sklearn GPR documentation

Linear Autoregressive Mult-Fidelity (MF) Model

$$f_{HF}(x) = \rho f_{LF}(x) + \delta_{HF}(x)$$

HF prediction

Correlation between model outputs at both fidelities

LF prediction using Gaussian process regression

Bias of HF data

Informing High-Fidelity Fits with Low-Fidelity Data

Model A: high-fidelity (HF) - expensive to generate data, very accurate

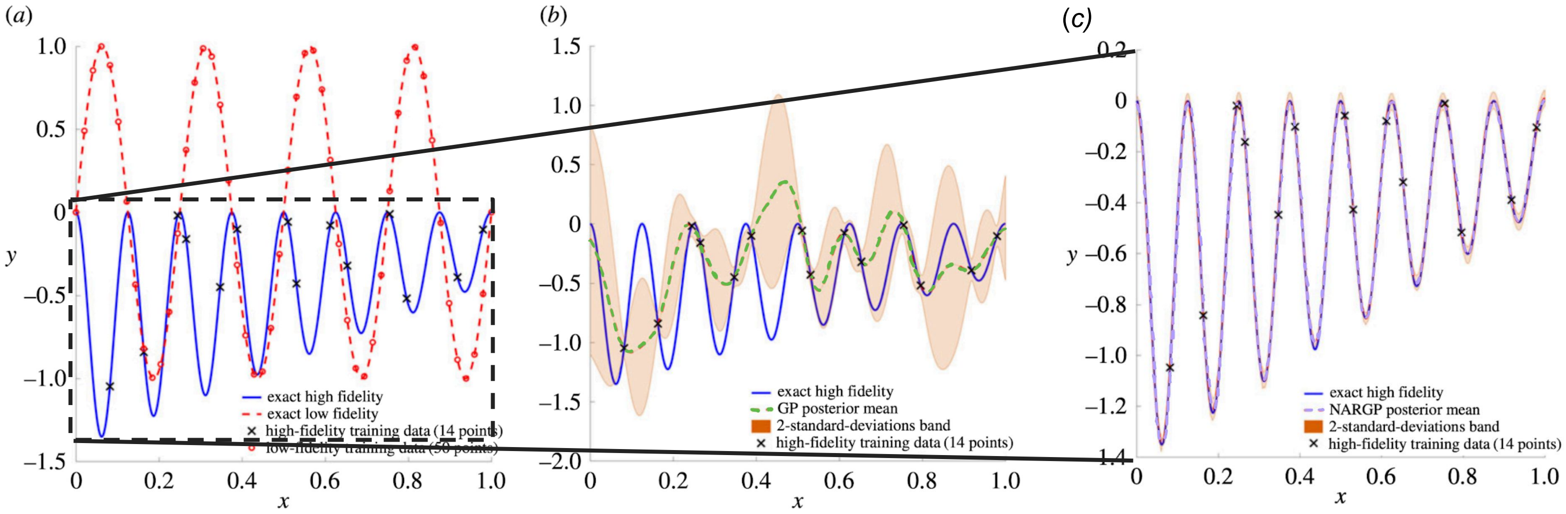
Model B: low-fidelity (LF) - cheap to generate data, low accuracy

GPR on HF data

Informing HF fit with LF data

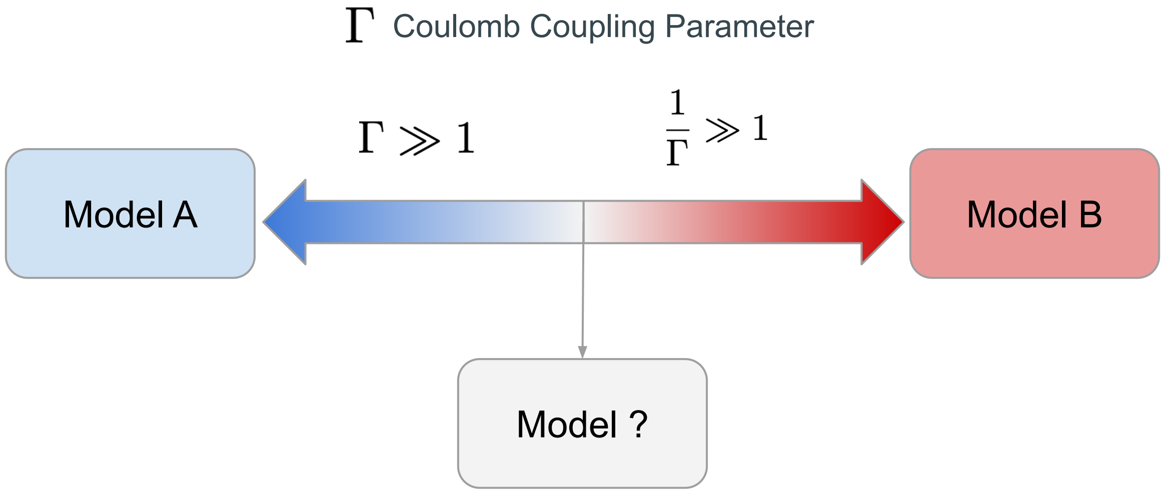
HF: 14 points

LF: 50 points

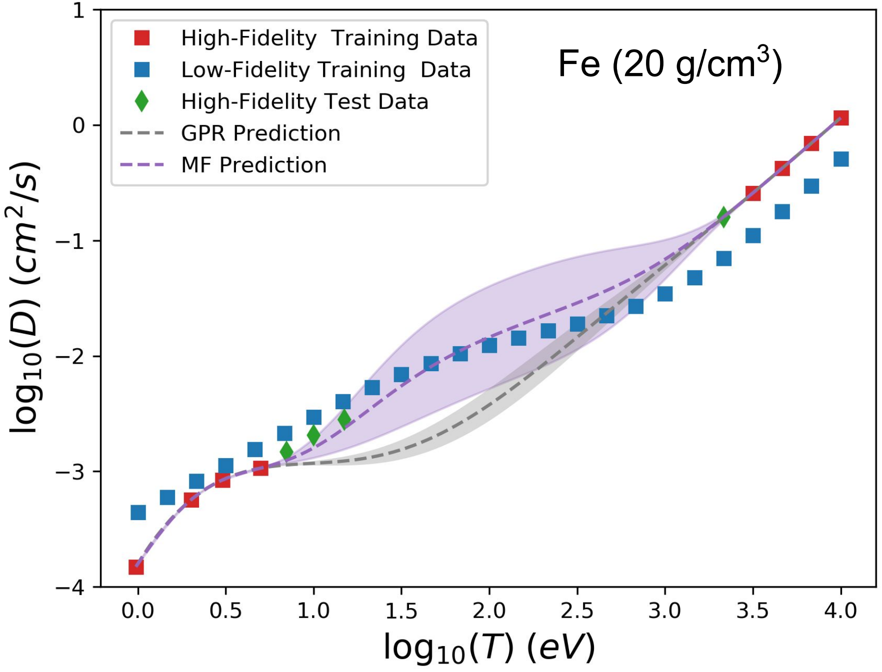


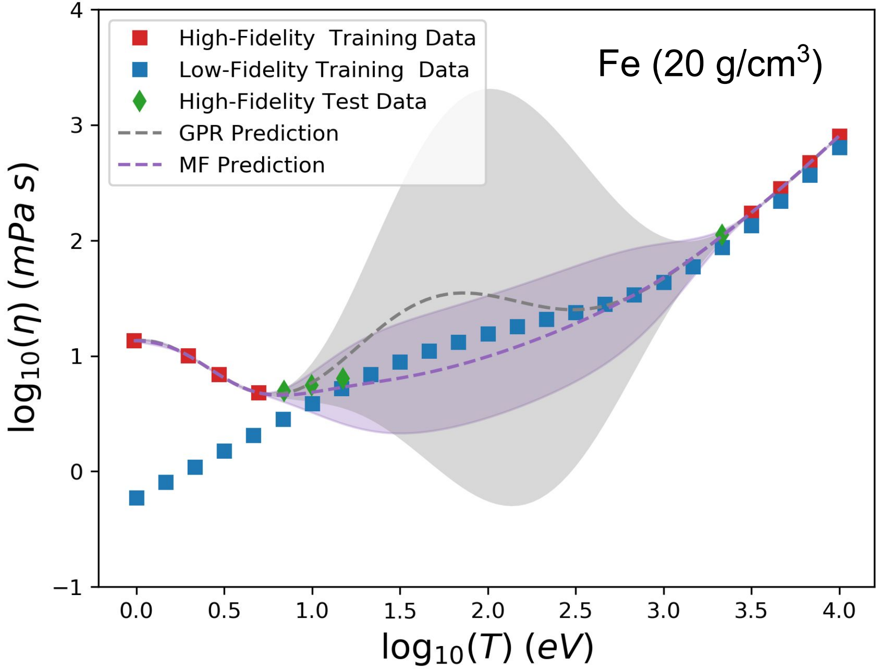
Perdikaris et. al, *Proc. R. Soc. A.*47320160751 (2017)

Physical Models are Often Asymptotically Accurate



Predicting Transport Coefficients with Multiple Models



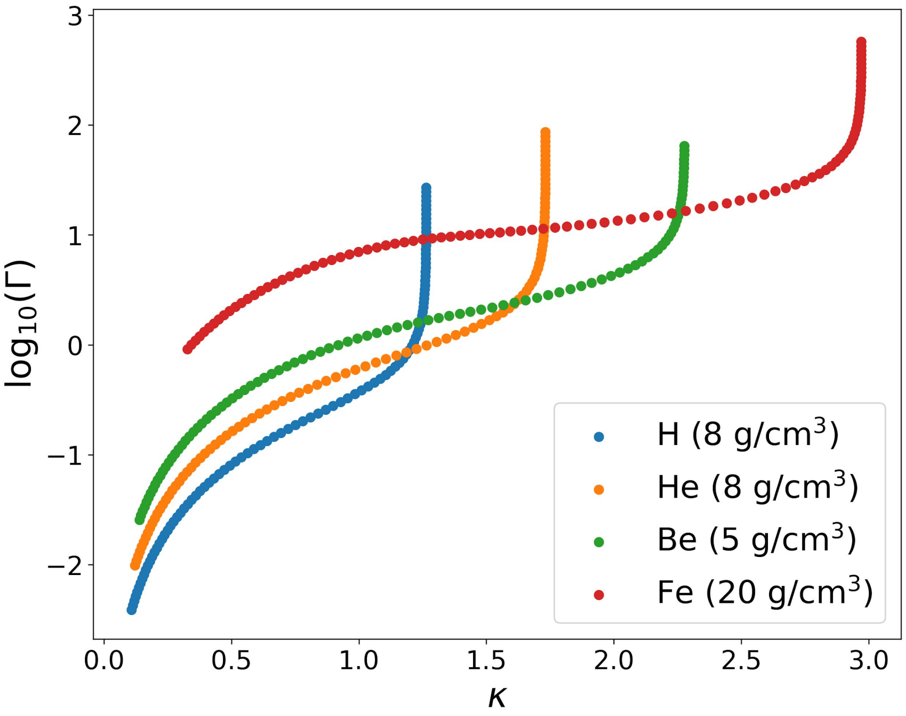


Projecting $\{Z, n_i, T\}$ to 2-Dimensions

$$\kappa = \frac{a_i}{\lambda_s}$$ Inverse electron screening length

$$\Gamma = \frac{\langle Z \rangle^2 e^2}{a_i T}$$ Coulomb coupling parameter

A (κ, Γ) pair defines a Yukawa system.



Multi-Fidelity Modeling in 3D

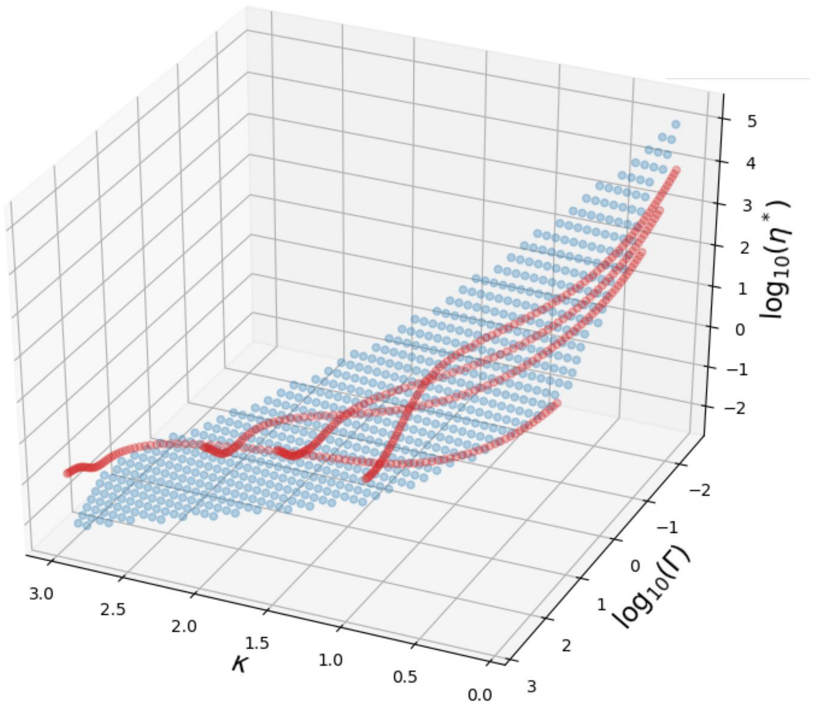
κ x-axis

Γ y-axis (log-scale)

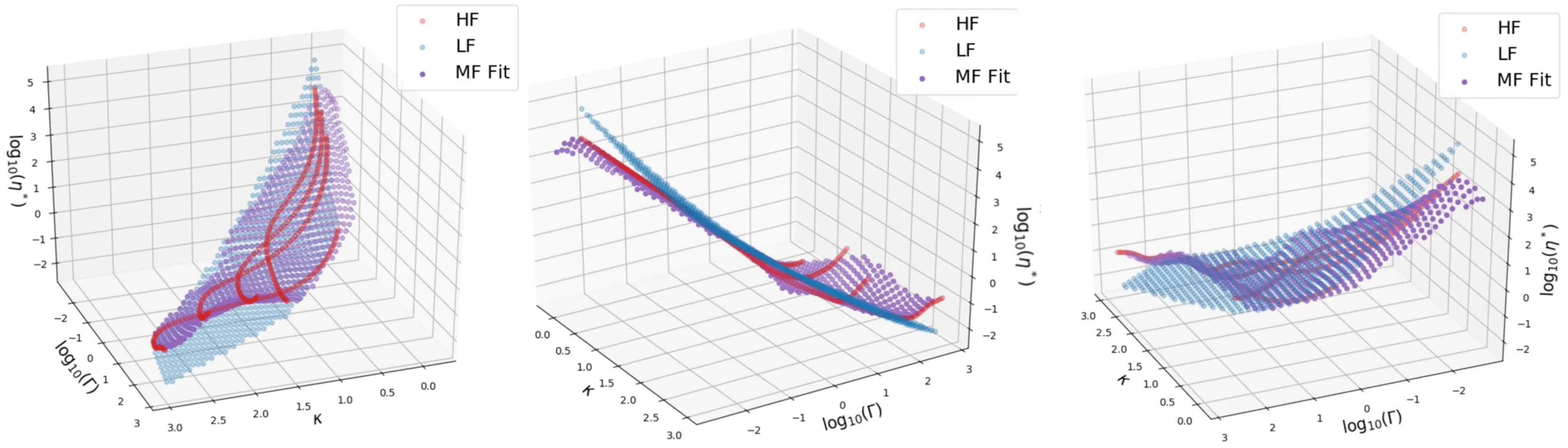
$\eta^* = \eta/(\rho_i \; a_i \; \omega_p^2)$

z-axis (log-scale)

- Low-Fidelity Model
- High-Fidelity Model



Multi-Fidelity Modeling in 3d Backup, Backup



Contributions

- Multi-fidelity models span large scales and preserve asymptotics
- Table interpolation improved when curse of dimensionality becomes problematic
- Confidence intervals aid in experimental design