Enabling Scale-up through Multi-fidelity Adaptive Computing

Juliane Mueller
Computational Science Center
March 2024

With contributions from Hilary Egan, Kevin Griffin

Scale-up of complex systems and associated risks

Scale-up: Extending systems and processes that were developed in the laboratory to function in the real world

Device and process scale-up comes with significant technical challenges and risk

Typical challenges:

- Data-driven models perform best when interpolating, **extrapolation is inherently uncertain,** and therefore **risky**
- Increasing ranges of scale (spatial, temporal) often lead to new/enriched physics
- High-fidelity physics-based models may capture new physics, but are typically too expensive for design/optimization work
- Operational regimes of existing experiments are limited, and new experiments are expensive

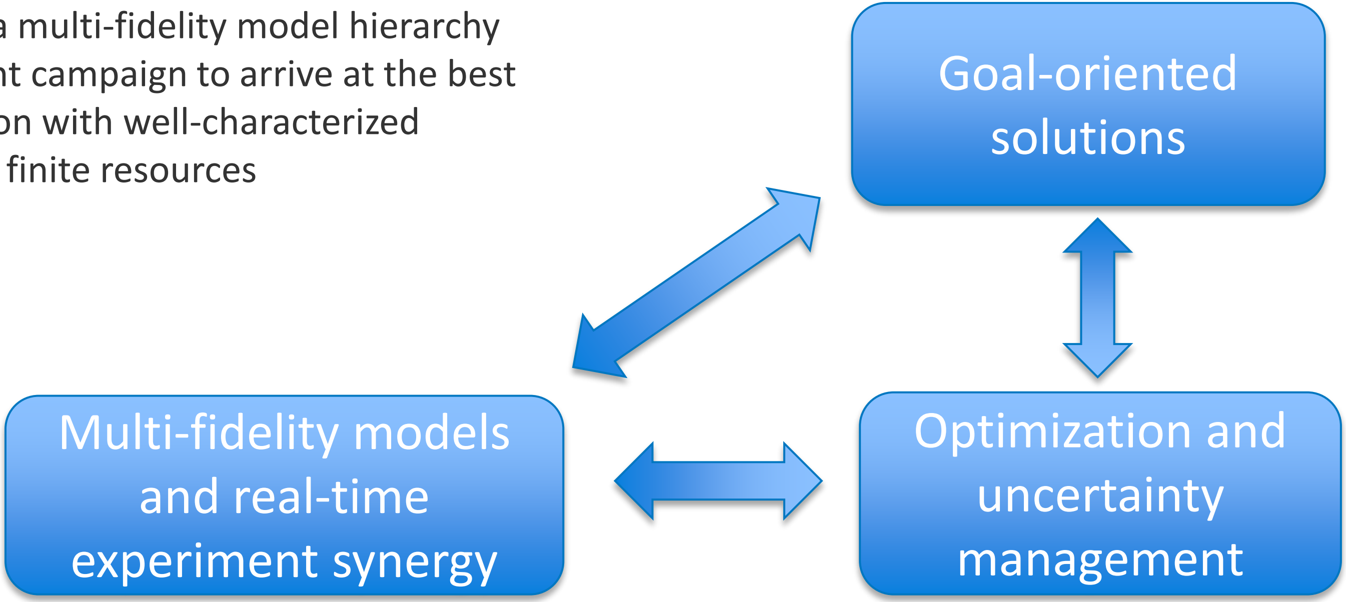
Image Credit: Dennis Schroeder, NREL 63958

Goal: reduce scale-up challenges by integrating multi-fidelity modeling and optimal compute resource use

Adaptive Computing

Orchestration of a multi-fidelity model hierarchy and/or experiment campaign to arrive at the best goal-based solution with well-characterized uncertainty given finite resources

Optimize the use of finite resources to achieve a specific science goal



Connect models with experiments to drive experiment design and data acquisition needs

Control of extrapolation uncertainty through targeted active learning

Key capability: multi-fidelity modeling

Most applications feature an assortment of models of widely varying fidelities, developed for different purposes:

- Experiment: “Truth”, but limited operational regime
- High-fidelity simulations: Physics-based (PDE/ODE), costly
- Lower fidelity levels: reduced physics, coarser meshes, less costly
- Data-driven surrogates: AI/ML, PINNs, Gaussian Processes (GPs), really cheap

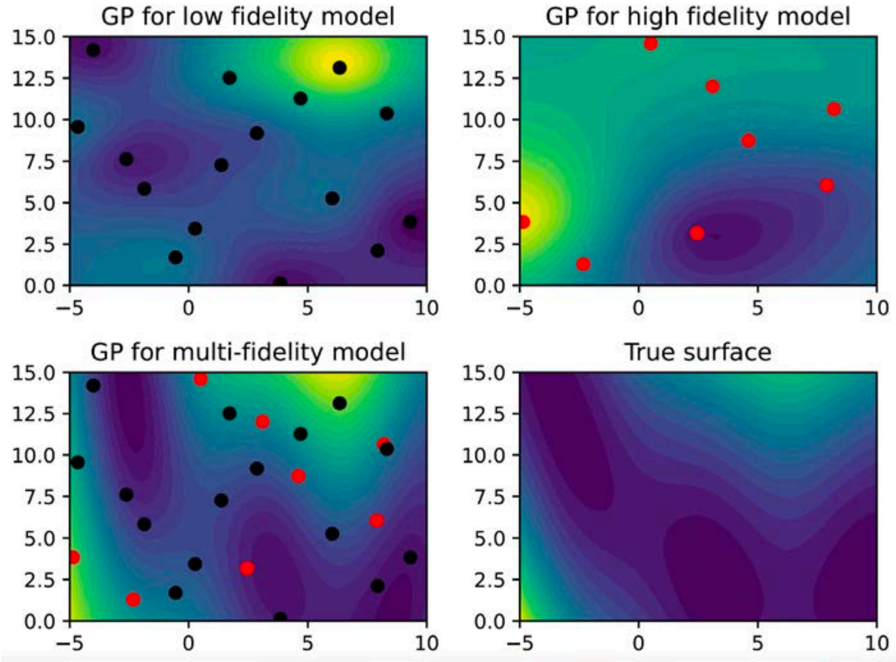
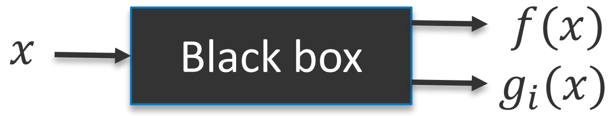
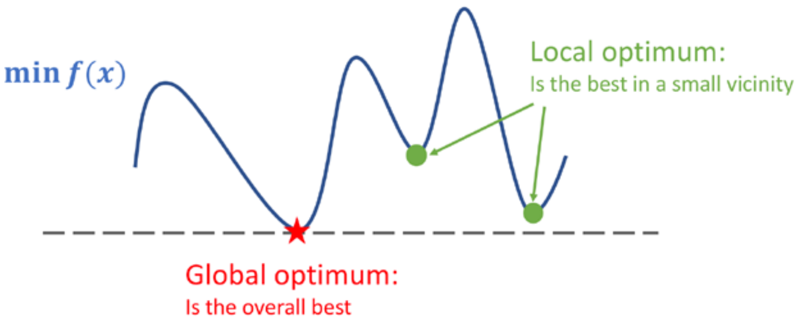


Fig: Exploiting information from multiple fidelity levels can increase surrogate accuracy

High-fidelity: Black-box expensive optimization

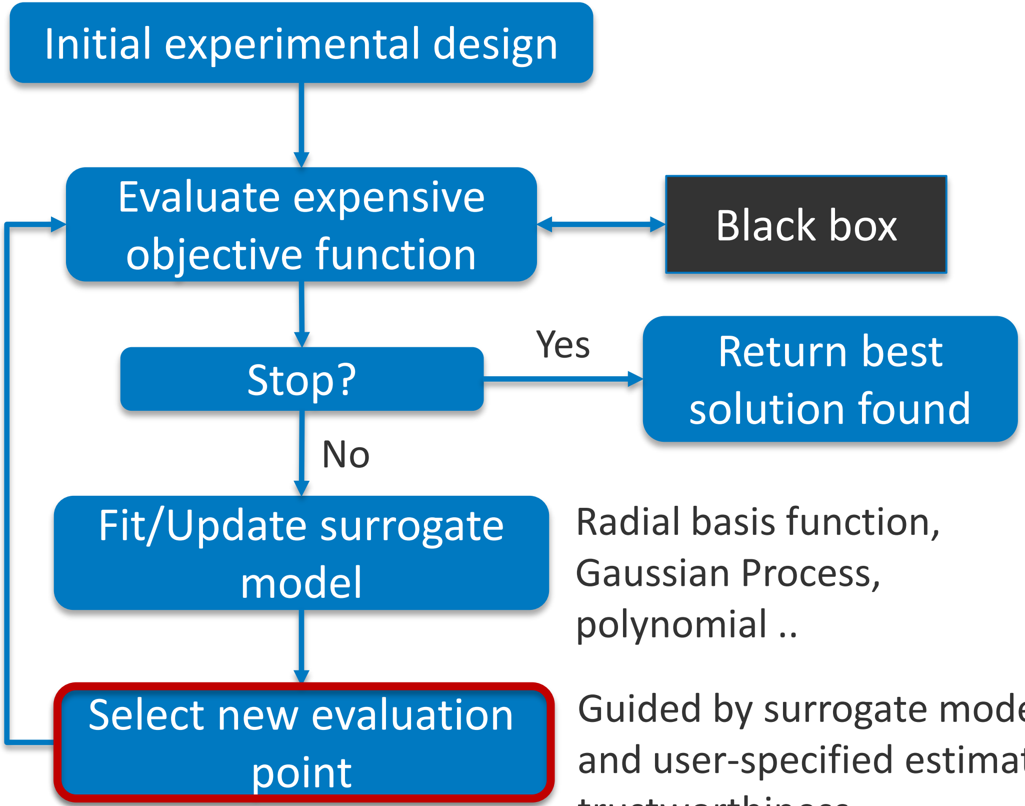
$$\min f(x)$$
$$s.t.\, g_i(x) \leq 0, i = 1, \dots, I$$
$$x \in \Omega$$





$$\cancel{\frac{df}{dx}}$$

Surrogate models steer the optimization loop



Stop when compute budget used up

Radial basis function, Gaussian Process, polynomial ..

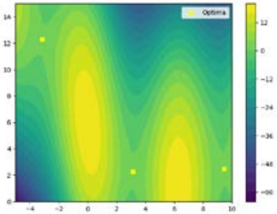
Guided by surrogate model and user-specified estimate of trustworthiness

Needs adaptation for problems with constraints, integers, failed evaluations, noise, multiple conflicting objective functions

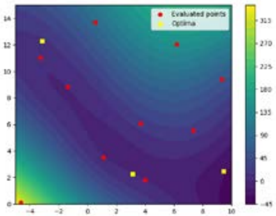
Correcting the low-fidelity model

- Multiplicative: $\hat{y}_{hf}(x) = \rho(x) * y_{lf}(x)$
- Additive: $\hat{y}_{hf}(x) = y_{lf}(x) + \delta(x)$
- Hybrid:
  - $\hat{y}_{hf}(x) = \rho(x) * y_{lf}(x) + \delta(x)$ (ρ const.)
  - $\hat{y}_{hf}(x) = w(x) * \rho(x) * y_{lf}(x) + \big(1 - w(x)\big) * (y_{lf}(x) + \delta(x))$, $w \in [0,1]$

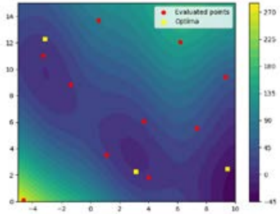
$y_{lf}(x)$

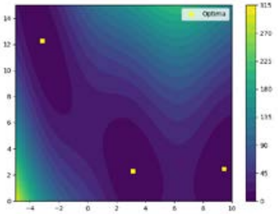


$\delta(x)$



$y_{lf}(x) + \delta(x)$



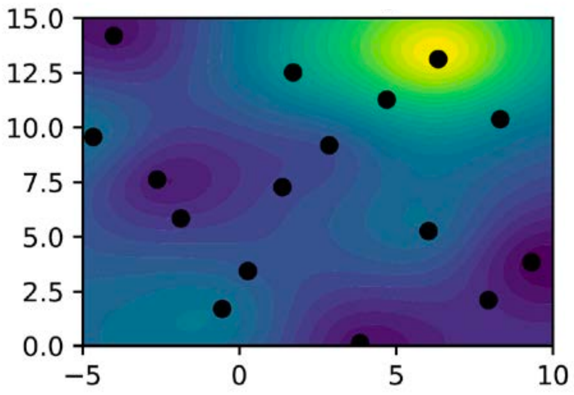


Ground truth

How do we make use of multiple fidelity levels during active learning?

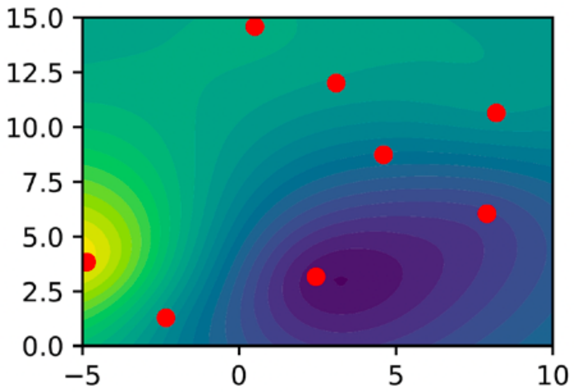
Exploiting multi-fidelity information

Build a surrogate model for the low(er) fidelity function



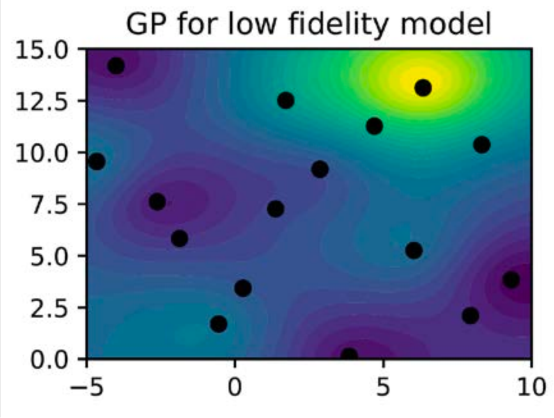
- Allow more samples than for high-fidelity function
- Use this surrogate to decide where to focus the search in the high-fidelity function
- *Low fidelity model is not necessarily accurate*

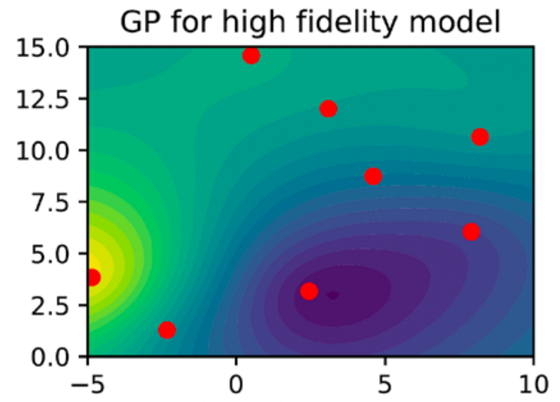
Build a surrogate model for the high-fidelity function



- Fewer samples are affordable
- Surrogate is less accurate (built on less data)
- Surrogate can be used to make (final) sample decisions

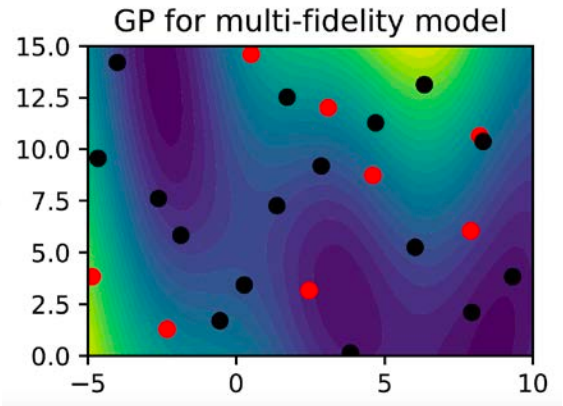
Gaussian Process: Using multiple fidelity information in one model

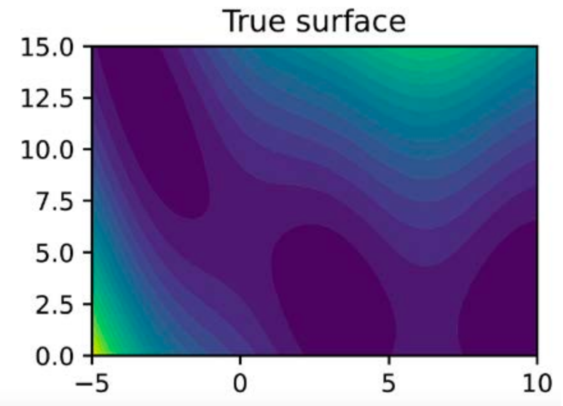




Red = high-fidelity evaluations
Black = Lower fidelity evaluations

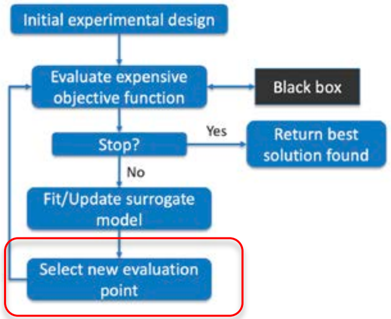
Python package surrogate modeling toolbox (SMT)



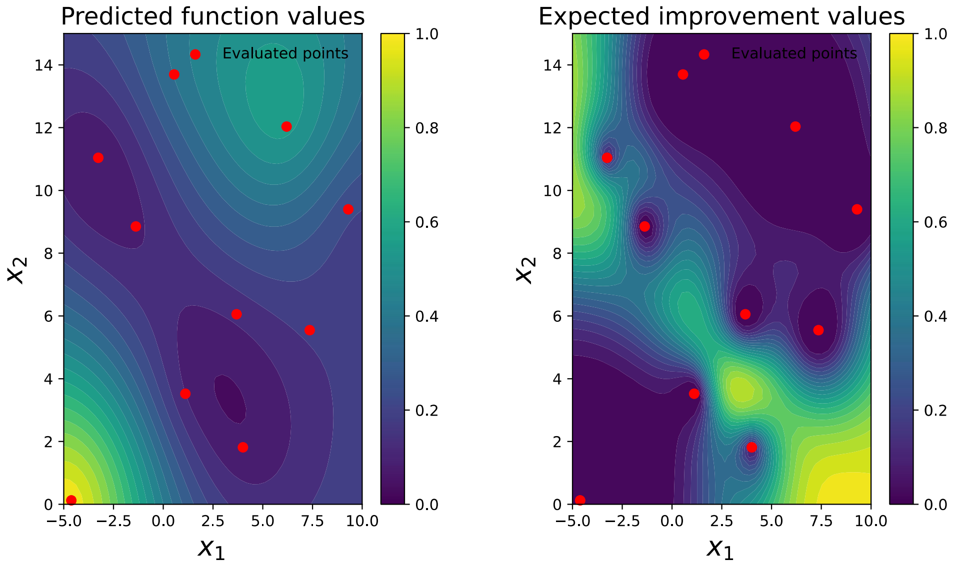


Combining high and lower fidelity information can lead to better approximation surface (compare to true contours)

Maximize the expected improvement to select a new point

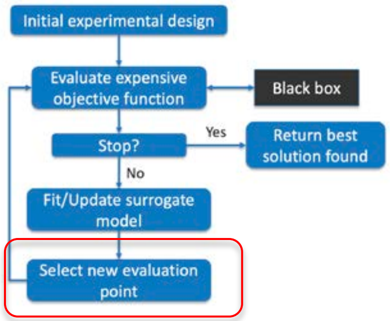


Expected improvement surface is multimodal and can become flat – making it difficult to find the global maximum…

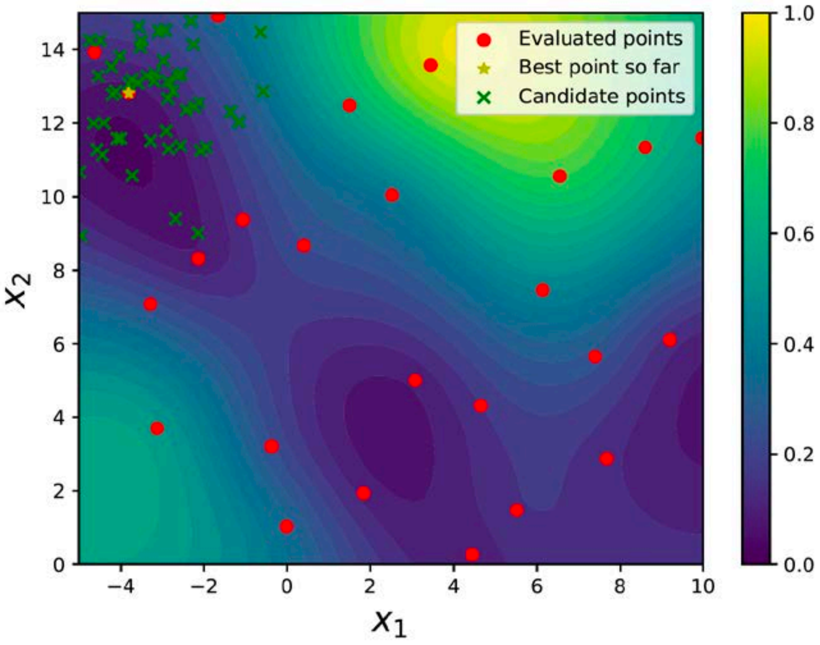


…requires development of other sampling strategies, guided by low-fidelity model

Sampling with candidate points



Add random perturbations to (select) variables of the best point(s) found so far



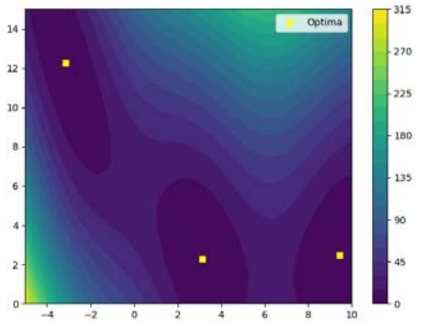
- Maximize a merit function that trades off predicted function value and distance to already evaluated points
  - Low function value -> local search
  - Large distance -> global search
- Select N new points for potential evaluation

Multi-fidelity sampling: when to ignore the low-fidelity model

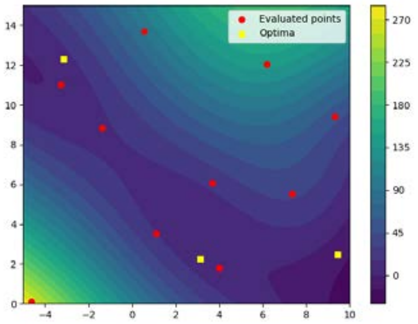
- Make use of as much information as is available
  - Surrogate of high-fidelity model
  - Low-fidelity (cheap) information – what if this one is very inaccurate/uncorrelated?
  - Surrogate of the difference as a selection constraint
  - User-specified estimates of local model trustworthiness

1. Define auxiliary function $a(x)$ using the surrogate model predictions
2. Optimize $a(x)$ to find x_{new}
3. If $-\delta \leq d(x_{new}) \leq \delta$ probe with low fidelity model first, otherwise ignore and evaluate high-fidelity model

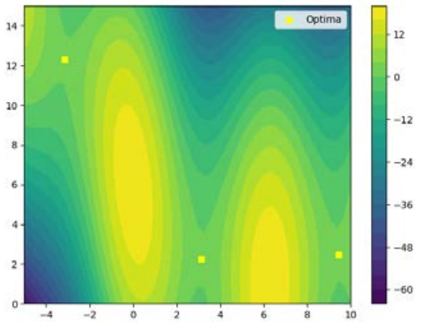
High-fidelity ground truth



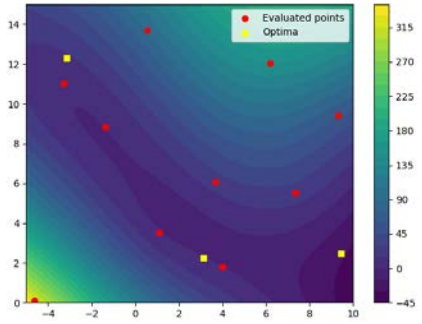
SM of high-fidelity model



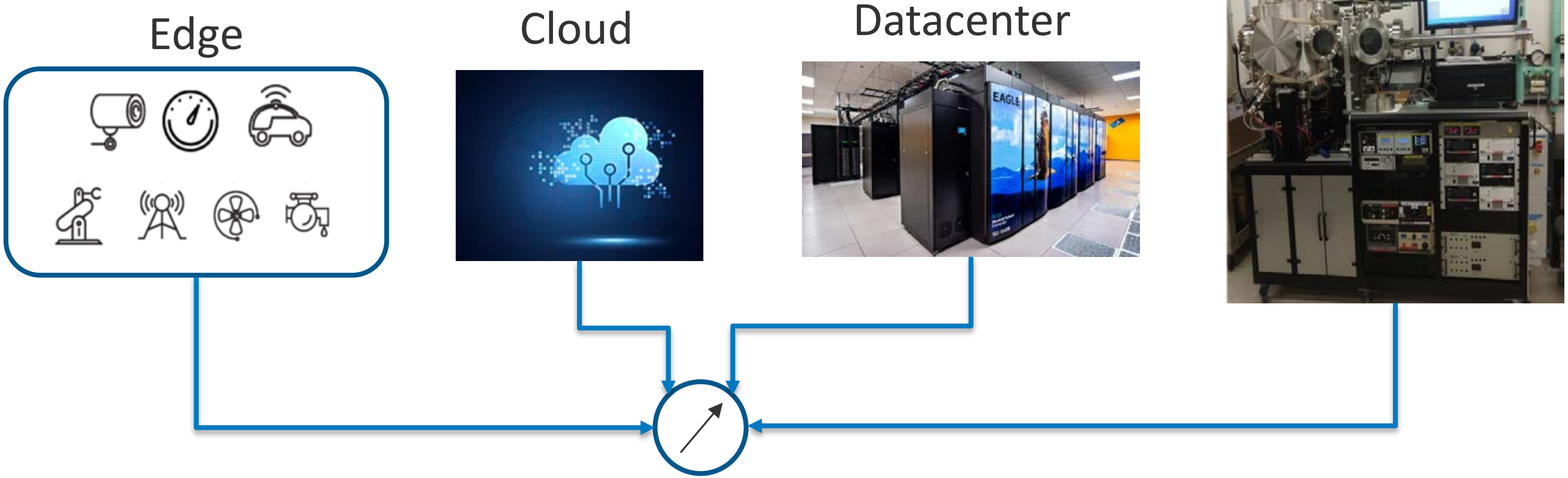
Low-fidelity model



Difference between high- and low-fidelity



Key capability: diverse compute resources



Resource manager:
how much and what kind of
resources do I have available?

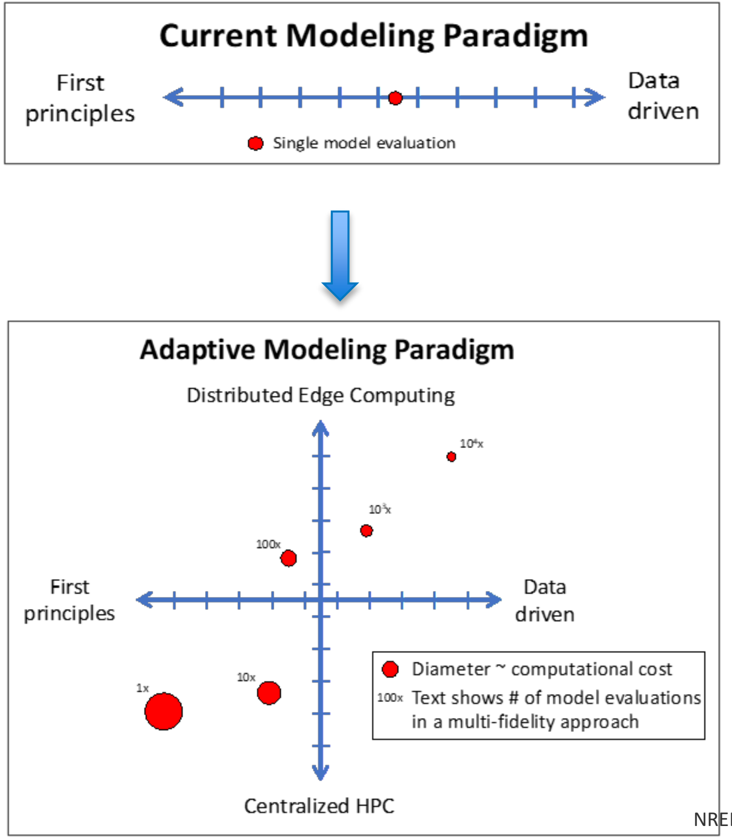
Solve stochastic discrete
optimization problem

Optimal computing strategy driven by specific output quantity of interest

Compute resource optimization problem

- What resources are available when?
- Formulate as optimization problems with stochasticity
- Implement solutions as constraints for multi-fidelity sampling
- Eventually must exploit asynchronous parallel computations

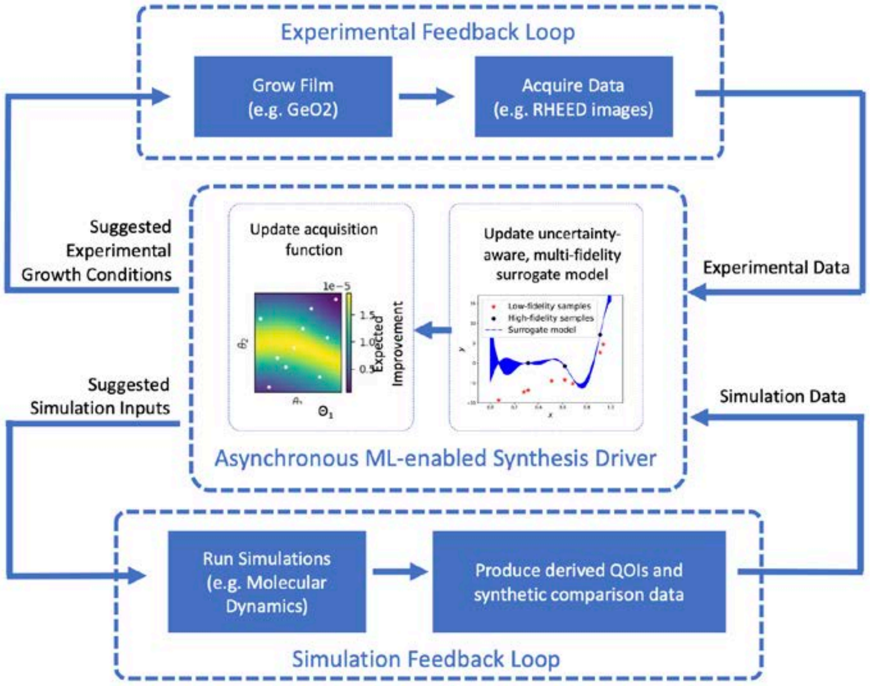
- Enumerate the user-defined simulation types (fidelity levels)
- Possible hardware configurations (# of CPUs, GPUs)
- Corresponding calculation duration
- Measurement noise estimate (aleatoric uncertainty)



Materials synthesis scale-up

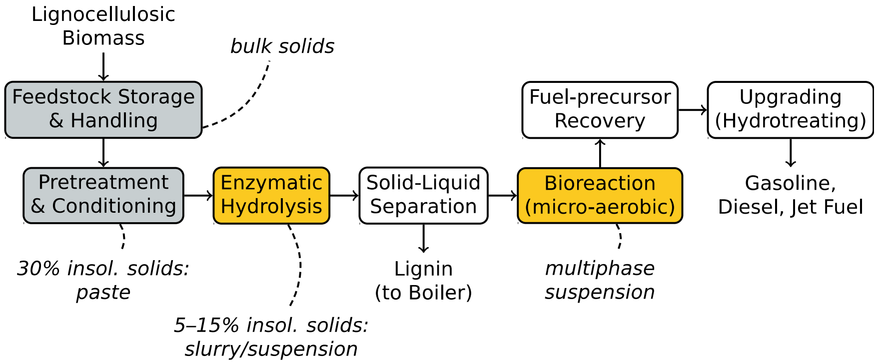
- **Big Picture Goal:** Accelerate the *deployment* of new materials and new capabilities of integrated electronics by avoiding materials growth and device integration barriers

- **Challenges:** Materials discovery has greatly outstripped reliable material synthesis and subsequent integration into devices. Each step currently requires a full experimental campaign for static, stove-piped optimization, and often suffers from reproducibility issues. This challenge is exacerbated by changing process conditions when moving from lab to pilot scale synthesis chambers.



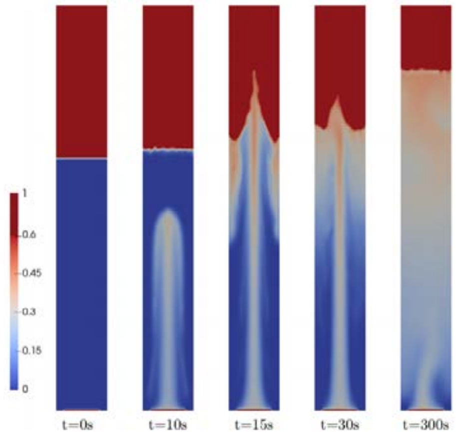
Example application: virtual engineering of biofuels

- Objective: maximize reactor-averaged oxygen uptake rate
- Inputs: O(10) chemical and processing design parameters



- Process lignocellulose-rich biomass into biofuel
- 3 step chemical processes
  - Pretreatment: fast simulation
  - Enzymatic hydrolysis: surrogate or CFD calculation
  - Bioreaction: surrogate or CFD calculation

- *Fidelity Level 1:* HF simulation (pretreatment, enzymatic hydrolysis, bioreactor): 32 CPU-cores @ 57 hours
- *Fidelity Level 2:* HF pretreatment, LF Lignocellulose model, and HF bioreactor: {72 CPU-cores @ 4 hours, or 32 CPU-cores @ 9 hours}
- Could add simulation type that varies the time to steady state/grid resolution



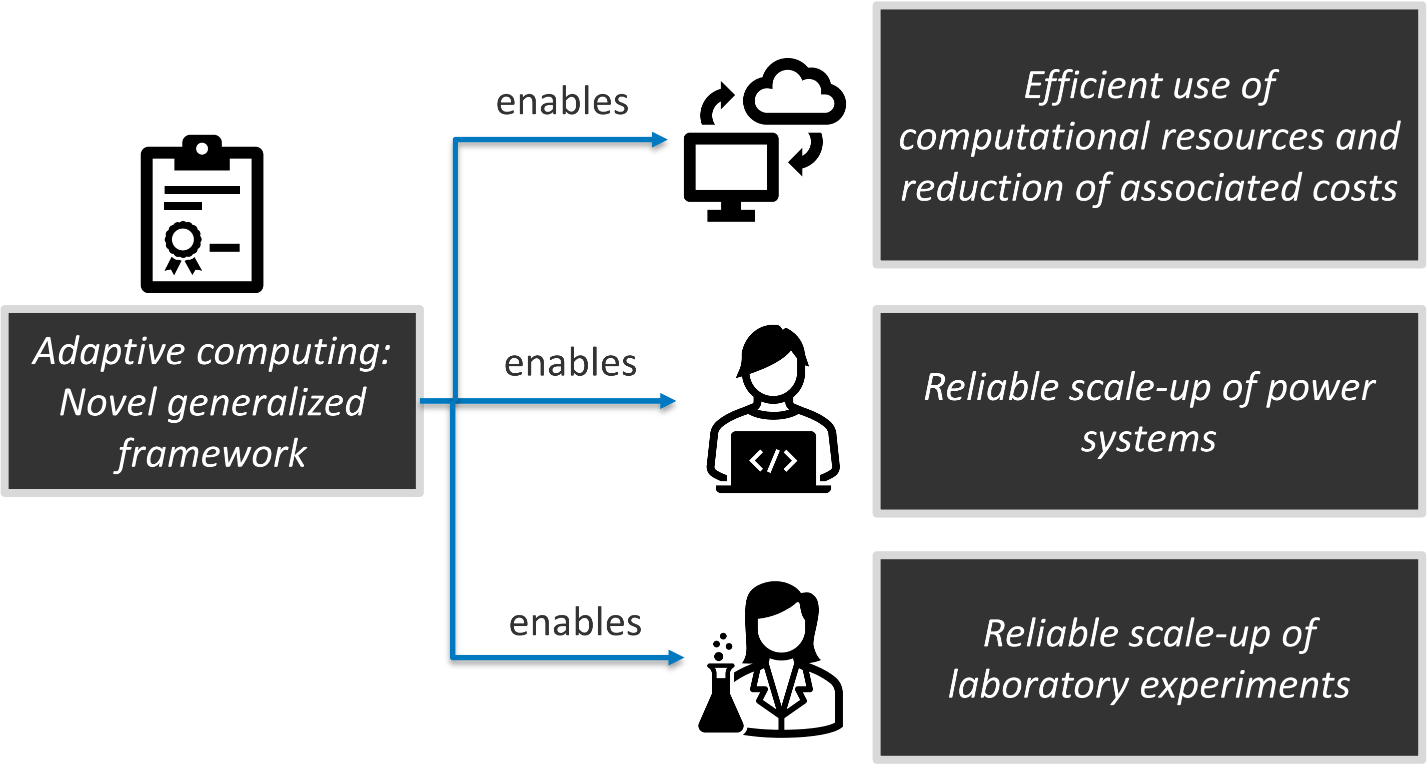
Preliminary tests on virtual engineering app

- 8 parameters, 9 random samples from LHS, 10 iterations
- Run for 5 minutes on 1 core, low fidelity models only
- Max Oxygen Uptake Rate = 0.06723323
- For some parameter settings, we obtained NaNs
  - The low fidelity models used may not be valid across the entire parameter space (hidden constraints – we know how to deal with these)

(No multi-fidelity business yet)

Parameter name	VE default	Min	Max	Final
Fraction of solids that is xylan	0.263	0	1	0.32
Fraction of solids that is glucan	0.4	0	1	0.29
Porous fraction of the biomass particles	0.8	0	1	0.64
Initial concentration of acid	1e-4	0	1e-3 (1)	1e-3
Steam temperature (C)	150	3.8	250.3	170
Fraction of insoluble solids	0.745	0	0.99 (1)	0.99
Enzymatic load	30	0	1000	57
FIS_0 target	0.05	0.005 (0)	1	0.005

Adaptive computing: optimizing the use of computational resources to target deficiencies and challenges related to scale-up



Thank You

www.nrel.gov

NREL/PR-2C00-89329

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. This work was supported by the Laboratory Directed Research and Development (LDRD) Program at NREL. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Photo from iStock-627281636

NREL
Transforming ENERGY