Enabling the fourth-paradigm of materials science through versatile Gaussian process & Bayesian optimization

Anh Tran

Sandia National Laboratories

October 9–12, 2022, Pittsburgh, PA
TMS Materials Science & Technology 2022
UQ in Data-Driven Materials and Process Design

Acknowledgment

Joint work with

- SNL: Tim Wildey, Mike Eldred, Bart G van Bloemen Waanders, Kathryn Maupin, John Mitchell, Laura Swiler, Julien Tranchida (now CEA), Aidan Thompson, Theron Rodgers, Hojun Lim, David Montes de Oca Zapiain
- ORNL: Hoang Tran
- Georgia Tech: Yan Wang, Dehao Liu (now Binghamton U), Stefano Travaglino, Wei Sun
- Others: Scott McCann (Xilinx), John Furlan (GIW), Krishnan Pagalthivarthi (GIW), Robert Visintainer (GIW)

Funded by

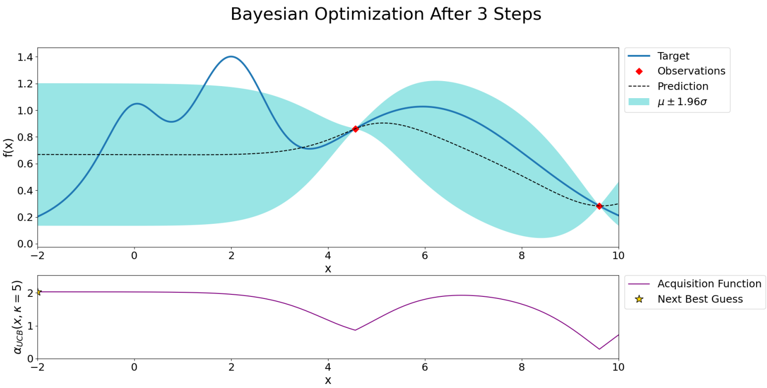
Sandia National Laboratories

NSF

The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

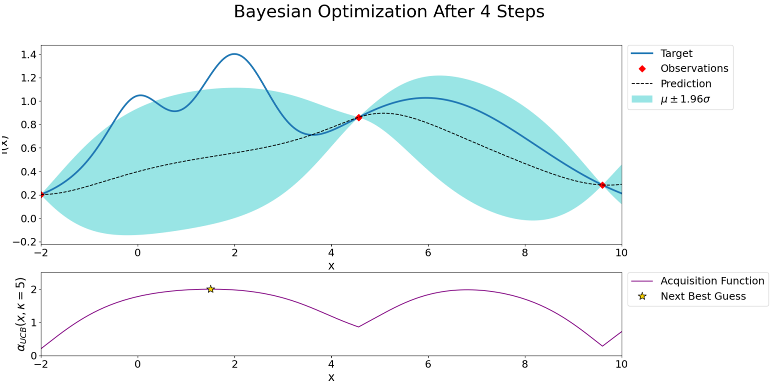
1. Gaussian process / Bayesian optimization
    - Demo
    - Introduction
    - Fundamentals
    - Acquisition
2. ICME applications
3. Conclusion

Bayesian optimization - Demo



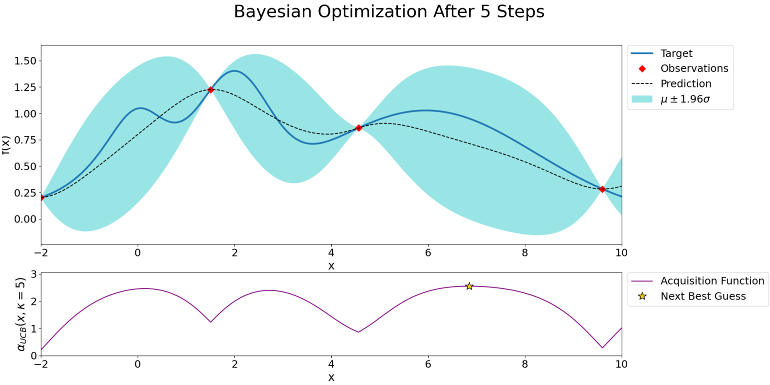
Bayesian optimization - Iteration 3

Bayesian optimization - Demo



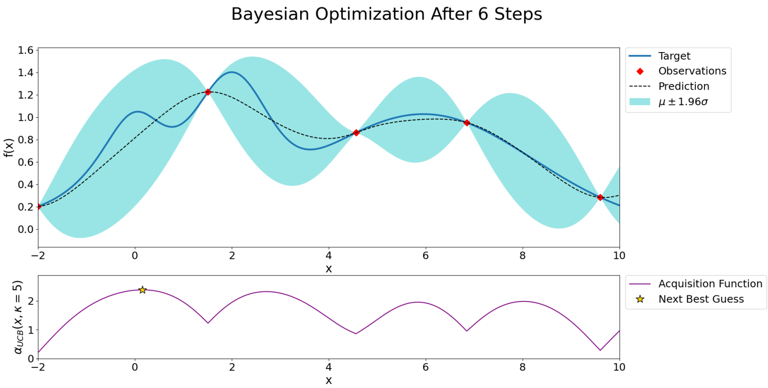
Bayesian optimization - Iteration 4

Bayesian optimization - Demo



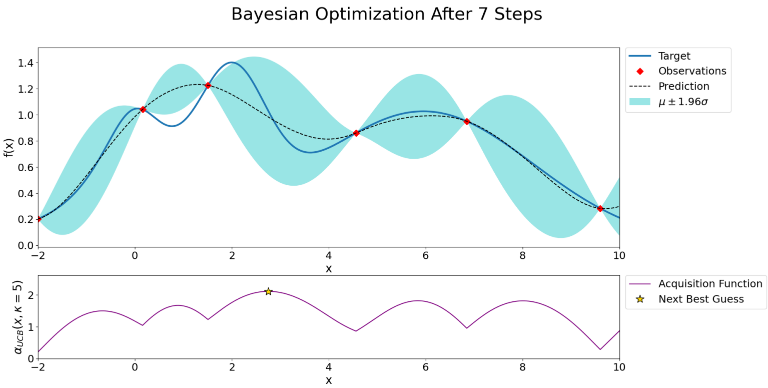
Bayesian optimization - Iteration 5

Bayesian optimization - Demo



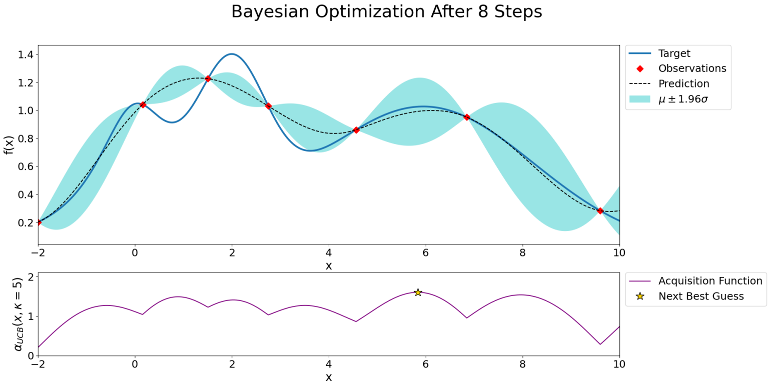
Bayesian optimization - Iteration 6

Bayesian optimization - Demo



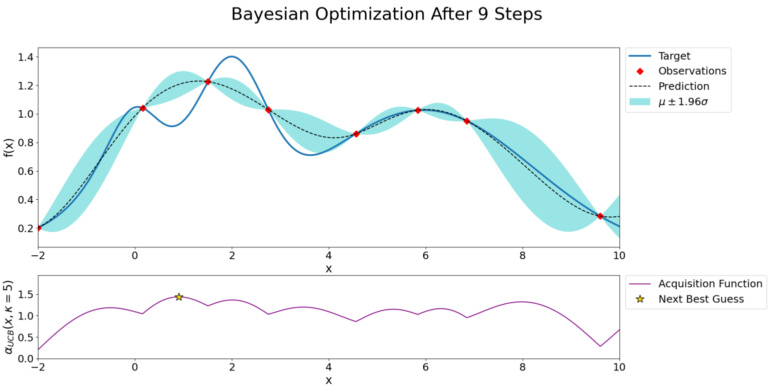
Bayesian optimization - Iteration 7

Bayesian optimization - Demo



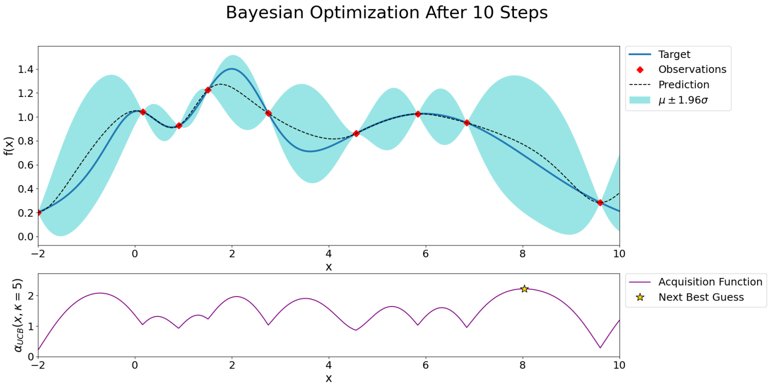
Bayesian optimization - Iteration 8

Bayesian optimization - Demo



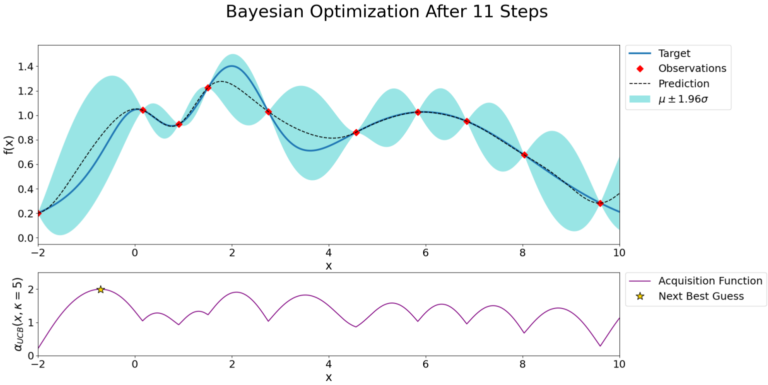
Bayesian optimization - Iteration 9

Bayesian optimization - Demo



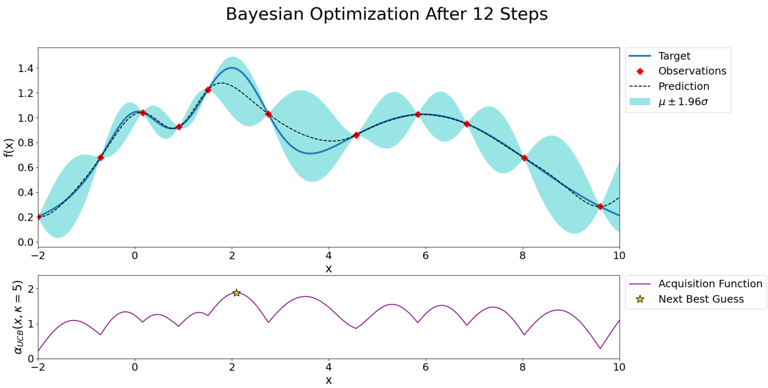
Bayesian optimization - Iteration 11

Bayesian optimization - Demo



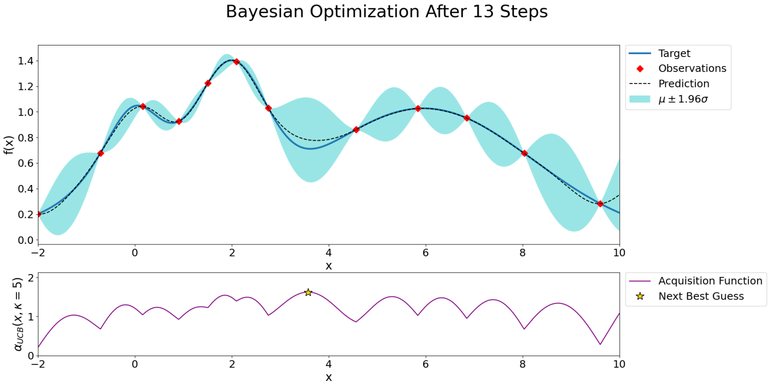
Bayesian optimization - Iteration 11

Bayesian optimization - Demo



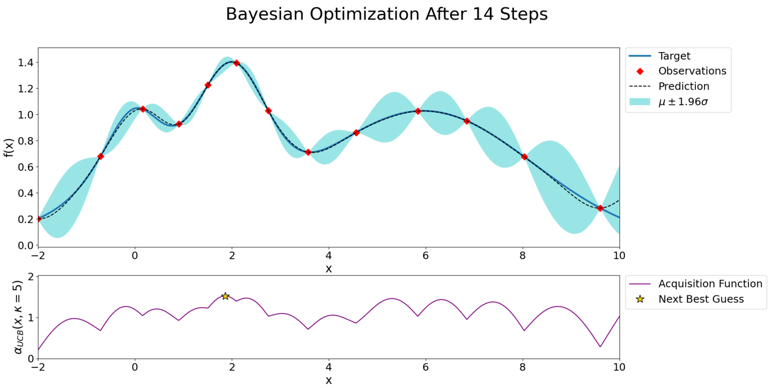
Bayesian optimization - Iteration 12

Bayesian optimization - Demo



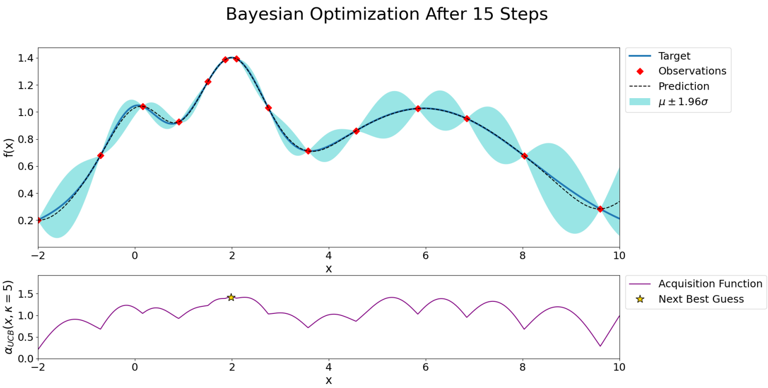
Bayesian optimization - Iteration 13

Bayesian optimization - Demo



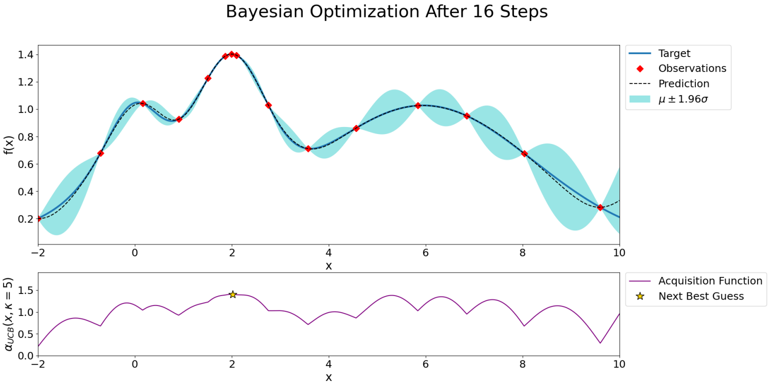
Bayesian optimization - Iteration 14

Bayesian optimization - Demo



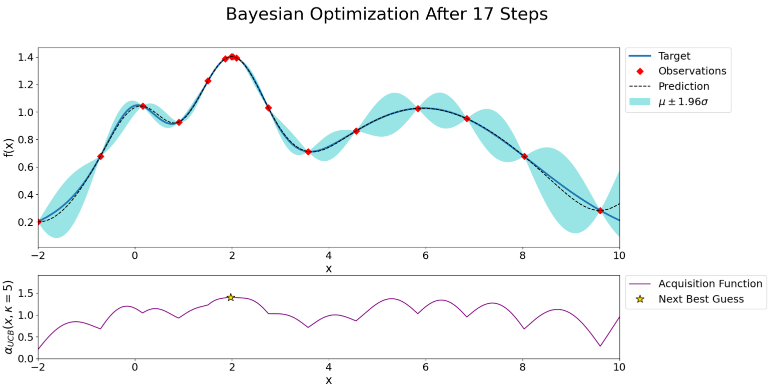
Bayesian optimization - Iteration 15

Bayesian optimization - Demo



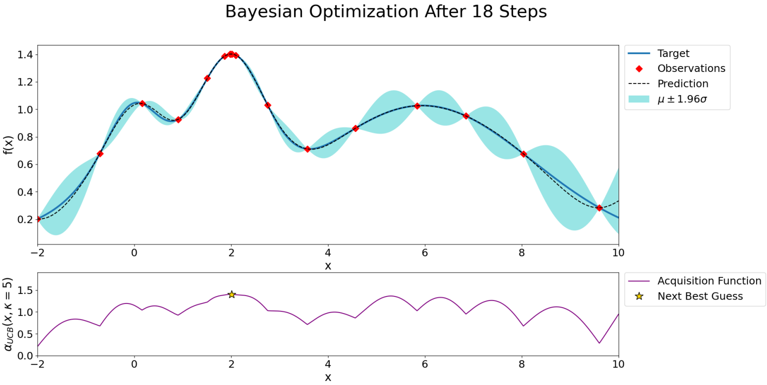
Bayesian optimization - Iteration 16

Bayesian optimization - Demo



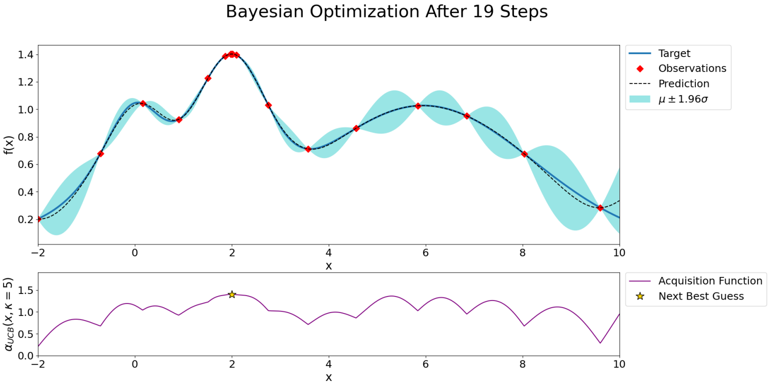
Bayesian optimization - Iteration 17

Bayesian optimization - Demo



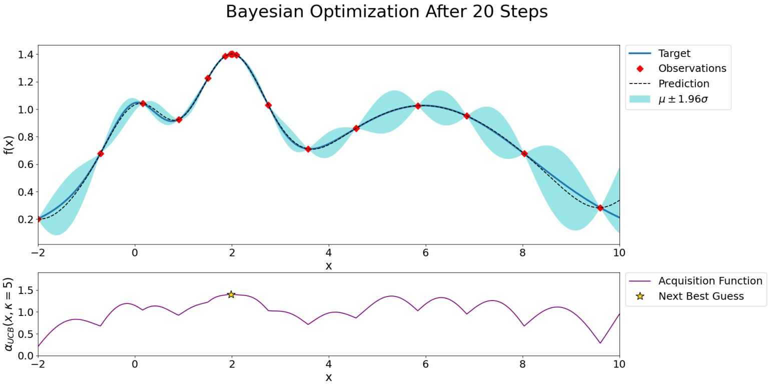
Bayesian optimization - Iteration 18

Bayesian optimization - Demo



Bayesian optimization - Iteration 19

Bayesian optimization - Demo



Bayesian optimization - Iteration 20

Advantages/Disadvantages

Bayesian optimization in a nutshell

Bayesian optimization = Gaussian process + sampling strategy

Advantages:

- optimize with uncertainty consideration (e.g. noisy observations)
- active machine learning (balance exploration-exploitation)
- derivative free (avoid computing Jacobian)
- global optimization (convergence in probability to global optimum)
- good convergence rate (provable asymptotic regret)

Disadvantages:

- high-dimensionality
- scalability: computational bottleneck $\mathcal{O}(n^3)$

Bayesian optimization features

very versatile (can solve a lot of practical problems)

- constrained on objectives (known + unknown constraints)
- constrained GP by physics (monotonicity, boundedness, convexity, symmetry)
- multi-task/multi-output
- multi-fidelity
- multi-objective optimization (Pareto frontier/optimal, domination)
- batch parallel optimization → asynchronous parallel optimization
- gradient-enhanced
- stochastic, heteroscedastic: non-Gaussian, student-t
- time-series (forecasting)
- mixed-integer, e.g. discrete/categorical
- scalable to Big Data
- latent variable model
- high-dimensional
- non-stationary

Classical GP: Fundamentals

Let $\mathcal{D}_n = \{\mathbf{x}_i, y_i\}_{i=1}^n$ denote the set of observations and $\mathbf{x}$ denote an arbitrary test points

$$\mu_n(\mathbf{x}) = \mu_0(\mathbf{x}) + \mathbf{k}(\mathbf{x})^T(\mathbf{K} + \sigma^2\mathbf{I})^{-1}(\mathbf{y} - \mathbf{m}) \tag{1}$$

$$\sigma_n^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}(\mathbf{x})^T(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{k}(\mathbf{x}) \tag{2}$$

where $\mathbf{k}(\mathbf{x})$ is a vector of covariance terms between $\mathbf{x}$ and $\mathbf{x}_{1:n}$.

Classical GP: Fundamentals

- assuming stationary kernel $\rightarrow k(\mathbf{x}, \mathbf{x}')$ only depends on $r = ||\mathbf{x} - \mathbf{x}'||$
- the covariance matrix: symmetric positive-semidefinite matrix made up of pairwise inner products

$$\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = k(\mathbf{x}_j, \mathbf{x}_i) = \mathbf{K}_{ji} \tag{3}$$

- kernel choice: assuming unknown function is smooth to some degree

Matérn kernels:

$$\mathbf{K}_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j) = \theta_0^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu} r)^\nu K_\nu(\sqrt{2\nu} r), \tag{4}$$

K_ν is a modified Bessel fuction of the second kind and order ν.
Common kernels:

- $\nu = 1/2 : k_{\text{Matérn}1}(\mathbf{x}, \mathbf{x}') = \theta_0^2 \exp(-r)$
- $\nu = 3/2 : k_{\text{Matérn}3}(\mathbf{x}, \mathbf{x}') = \theta_0^2 \exp(-\sqrt{3}r)(1 + \sqrt{3}r)$,
- $\nu = 5/2 : k_{\text{Matérn}5}(\mathbf{x}, \mathbf{x}') = \theta_0^2 \exp(-\sqrt{5}r)\left(1 + \sqrt{5}r + \frac{5}{3}r^2\right)$,
- $\nu \rightarrow \infty : k_{\text{sq-exp}}(\mathbf{x}, \mathbf{x}') = \theta_0^2 \exp\left(-\frac{r^2}{2}\right)$

Log (marginal) likelihood function:

$$\log p(\mathbf{y}|\mathbf{x}_{1:n}, \theta) = - \underbrace{\frac{n}{2}\log(2\pi)}_{\text{data likelihood} \downarrow \text{ as } n\uparrow} - \underbrace{\frac{1}{2}\log|\mathbf{K}^\theta + \sigma^2\mathbf{I}|}_{\substack{\text{“complexity” term} \\ \text{smoother covariance matrix}}} - \underbrace{\frac{1}{2}(\mathbf{y} - \mathbf{m}_\theta)^T(\mathbf{K}^\theta + \sigma^2\mathbf{I})^{-1}(\mathbf{y} - \mathbf{m}_\theta)}_{\substack{\text{“data-fit” term} \\ \text{how well model fits data}}} \tag{5}$$

Acquisition function: How to pick the next point(s)

- how to pick the next point: exploitation (if $\sigma_A^2 = \sigma_B^2$ but $\mu_A > \mu_B$ then choose A) or exploration (if $\mu_A = \mu_B$ but $\sigma_A^2 > \sigma_B^2$ then choose A). If
- different flavors:
  1. probability of improvement (PI) Mockus 1982
     $$\alpha_{\text{PI}}(\mathbf{x}) = \Phi(\gamma(\mathbf{x})), \tag{6}$$
     where
     $$\gamma(\mathbf{x}) = \frac{\mu(\mathbf{x}) - f(\mathbf{x}_{\text{best}})}{\sigma(\mathbf{x})}, \tag{7}$$
  2. expected improvement (EI) scheme Huang et al. 2006; Mockus 1975
     $$\alpha_{\text{EI}}(\mathbf{x}) = \sigma(\mathbf{x})[\gamma(\mathbf{x})\Phi(\gamma(\mathbf{x})) + \phi(\gamma(\mathbf{x}))] \tag{8}$$
  3. upper confidence bound (UCB) schemeSrinivas et al. 2009, 2012
     $$\alpha_{\text{UCB}}(\mathbf{x}) = \mu(\mathbf{x}) + \kappa\sigma(\mathbf{x}), \tag{9}$$
     where κ is a hyper-parameter describing the exploitation-exploration balance.
  4. pure exploration*:
     - ★ maximal MSE $\sigma^2(\mathbf{x}) \Leftrightarrow$ maximal entropy: $\frac{1}{2}\log\left[2\pi\sigma^2(\mathbf{x})\right] + \frac{1}{2}$
     - ★ maximal IMSE $\int_{\mathbf{x}\in\mathcal{X}} \sigma^2(\mathbf{x})$

Acquisition function: Reparameterization

Reparameterization in deep learning by Wilson, Hutter, and Deisenroth 2018.

Table: $1^{+/-}$: right-/left-continuous Heaviside step function; ReLU + sigmoid (σ) + softmax: activation function; $\Sigma = \mathbf{L}\mathbf{L}^\top$: Cholesky; $\gamma \sim \mathcal{N}(\mathbf{0}, \Sigma)$ residual

Abbr.	Acquisition function $\mathcal{L}$	Reparameterization
EI	$\mathbb{E}_{\mathbf{y}}[\max(\text{ReLU}(\mathbf{y} - \alpha))]$	$\mathbb{E}_{\mathbf{z}}[\max(\text{ReLU}(\mu + \mathbf{Lz} - \alpha))]$
PI	$\mathbb{E}_{\mathbf{y}}[\max(1^{-}(\mathbf{y} - \alpha))]$	$\mathbb{E}_{\mathbf{z}}[\max(\sigma(\frac{\mu + \mathbf{Lz} - \alpha}{\tau}))]$
SR	$\mathbb{E}_{\mathbf{y}}[\max(\mathbf{y})]$	$\mathbb{E}_{\mathbf{z}}[\max(\mu + \mathbf{Lz})]$
UCB	$\mathbb{E}_{\mathbf{y}}[\max(\mu + \sqrt{\frac{\beta\pi}{2}}\lvert\gamma\rvert)]$	$\mathbb{E}_{\mathbf{z}}[\max(\mu + \sqrt{\frac{\beta\pi}{2}}\lvert\mathbf{Lz}\rvert)]$
ES	$-\mathbb{E}_{\mathbf{y}_a}[\text{H}(\mathbb{E}_{\mathbf{y}_b \mid \mathbf{y}_a}[1^{+}(\mathbf{y}_b - \max(\mathbf{y}_b))])]$	$-\mathbb{E}_{\mathbf{z}_a}[\text{H}(\mathbb{E}_{\mathbf{z}_b}[\text{softmax}(\frac{\mu_{b\mid a} + \mathbf{L}_{b\mid a}\mathbf{z}_b}{\tau})])]$
KG	$\mathbb{E}_{\mathbf{y}_a}[\max(\mu_b + \Sigma_{b,a}\Sigma_{a,a}^{-1}(\mathbf{y}_a - \mu_a))]$	$\mathbb{E}_{\mathbf{z}_a}[\max(\mu_b + \Sigma_{b,a}\Sigma_{a,a}^{-1}\mathbf{L}_a\mathbf{z}_a)]$

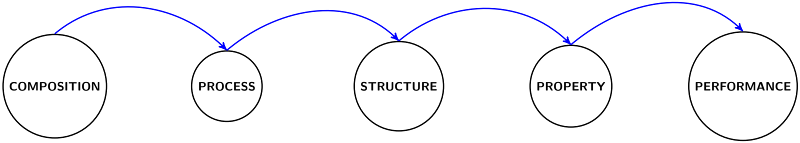
1. Gaussian process / Bayesian optimization

2. ICME applications
   - Overview – MGI with ICME
   - Inverse problems in process-structure (kinetic Monte Carlo)
   - Inverse problems in composition-property (DFT + MD)
   - Inverse problems in structure-property (CPFEM)

3. Conclusion

Process-structure-property relationship

- process: $\mathbf{x} + \delta, \delta \sim \mathcal{U}[\underline{\delta}, \overline{\delta}]$ – deterministic, controllable within a tolerance δ (in a manufacturing context)
- (micro)structure – random/stochastic, spatio-temporal, microstructure representations (physics-based vs. data-driven), image (i.e. high-dimensional), limited/scarce data
- property/performance: $y = f(\mathbf{x}) + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma^2)$ – deterministic with Gaussian noisy observations



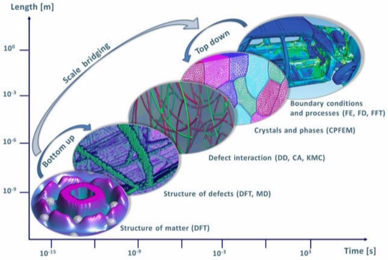
Process-structure-property relationship

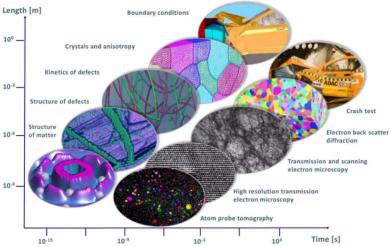
Nature of inverse problems

The nature of the input, i.e. deterministic or stochastic, determines the methodology for solving the inverse problem in PSPP.

- for deterministic variables in process $\rightarrow$ composition, process$\rightarrow$structure: Bayesian optimization (or any other optimization methods)
- for stochastic variables (typically affiliated with microstructure), such as grain size distribution, orientation distribution, in structure$\rightarrow$property: Bayesian inference is more appropriate to infer a distribution of features

Multi-scale ICME models as forward models



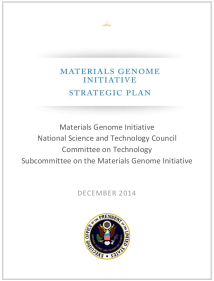


Courtesy of Prof. Dierk Raabe. https://www.dierk-raabe.com/multiscale-modeling/.

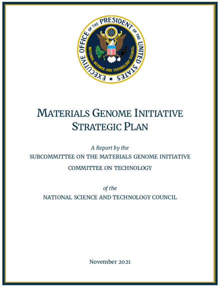
Common ICME models: DFT, MD, PF, CPFEM, kMC, CA, DDD, CFD
Multi-physics ICME models: PF+CPFEM, kMC+CPFEM, kMC+PF, DDD+CPFEM, MD+CPFEM, DFT+MD, etc.

Beyond forward ICME models: MGI as inverse problems

MGI - June 11 National Science and Technology Council (US) 2011

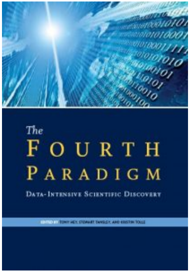


MGI - Dec 14 Holdren et al. 2014



MGI - Nov 21 Lander et al. 2021

Beyond forward ICME models: MGI as inverse problems



The Fourth Paradigm: Data-Intensive Scientific Discovery Hey, Tansley, Tolle, et al. 2009.

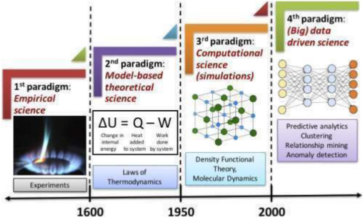


FIG. 1. The four paradigms of science: empirical, theoretical, computational, and data-driven.

The four paradigms of science: empirical, theoretical, computational, and data-driven. Agrawal and Choudhary 2016.

ICME is the 3rd block and ML/AI is the 4th block.

Beyond forward ICME models: MGI as inverse problems

Challenges:

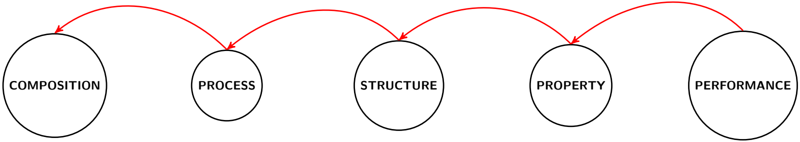
- optimization under (microstructure-induced) uncertainty
- small + noisy datasets
- high-dimensional
- high computational cost for ICME models → sample efficient

Goals:

- traditional approach: 20+ years
- accelerate materials design by "2× at a fraction of the cost"

Boosters:

- ICME developments: experimental[2] → computational[3]
- ML / AI: computational[3] → ML[4]



GP advantages

What are Gaussian process regression and Bayesian optimization useful for?

- (multiscale) design optimization, e.g. manufacturing, chemical processing, AM, materials discovery,
- ICME model calibration (e.g. CPFE, DDD, kMC, MD, DFT, PF)
- surrogate model for forward and inverse UQ.

Why are they useful?

- flexible – easy to modify to suit your needs,
- not too many hyper-parameters,
- optimization under (microstructure-induced) uncertainty,
- rigorous on mathematical ground: analytical convergence rate (both GP and BO),
- work well on small datasets with low- to intermediate-dimensional problems.

Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)

A formal problem statement:

- there exists a forward tool $f(\cdot)$ to predict microstructure, $u = f(x)$ (represented as images)
- given a target u^* (represented as images)
- task: find x^* such that $f(x^*) = u^* \approx u$

$\approx$ is defined in the sense of statistical equivalence for microstructures, $p_{\mathcal{D}}$ is the p.d.f. of statistical microstructure descriptors $\mathcal{D}$, i.e.

$$p_{\mathcal{D}} : \Omega \to L^1 : p_{\mathcal{D}}(u^*) \approx p_{\mathcal{D}}(u) \tag{10}$$

$$d\Big(p_{\mathcal{D}}(u^*), p_{\mathcal{D}}(u)\Big) \leq \text{TOL} \tag{11}$$

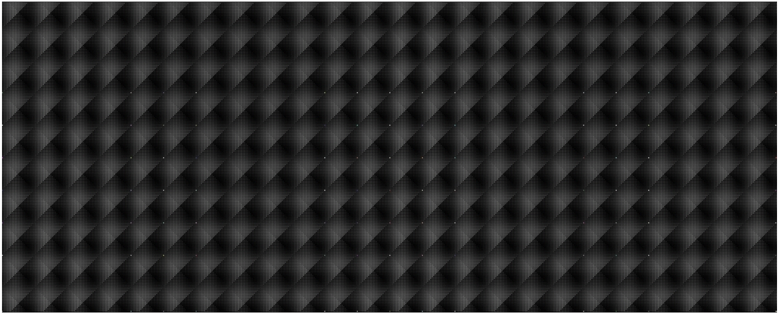
Hint: quantitatively differentiate microstructures using statistical microstructure descriptors

Reference

Anh Tran et al. (2020a). "An active-learning high-throughput microstructure calibration framework for process-structure linkage in materials informatics". In: *Acta Materialia* 194, pp. 80–92

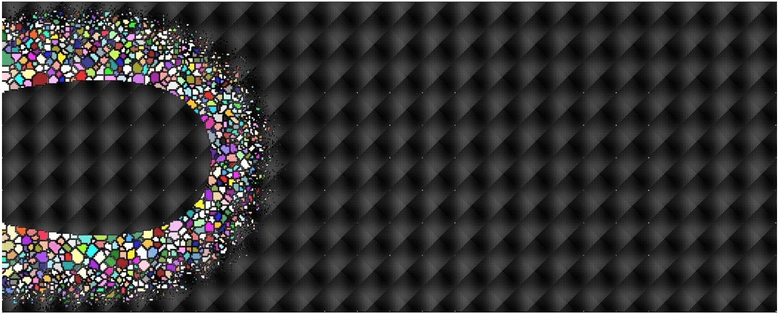
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



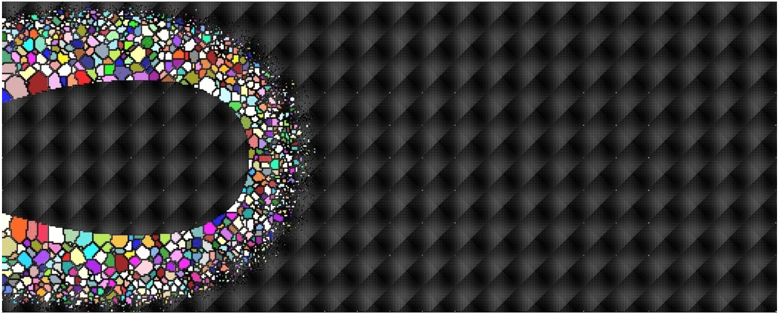
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)

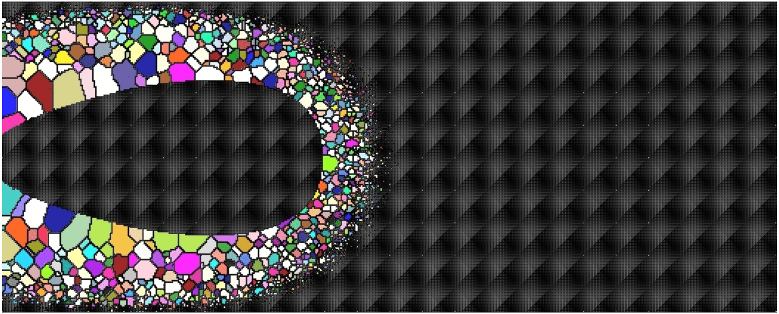


Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)

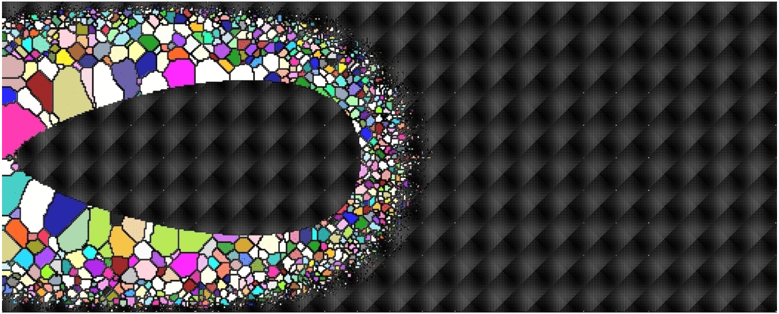
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



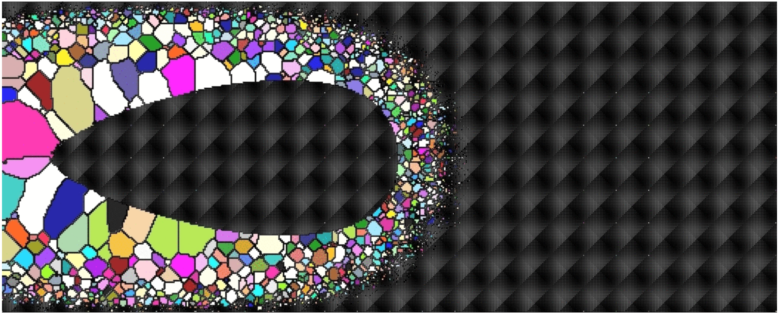
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



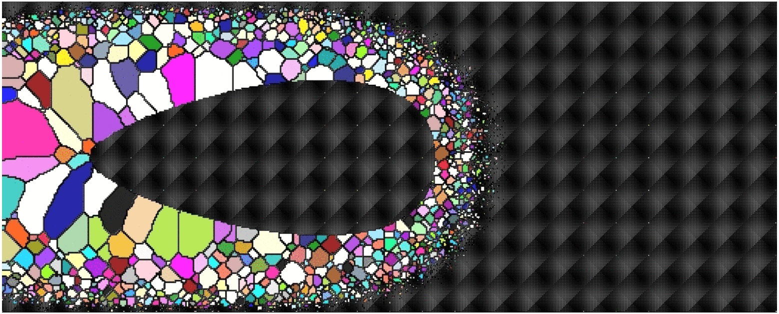
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



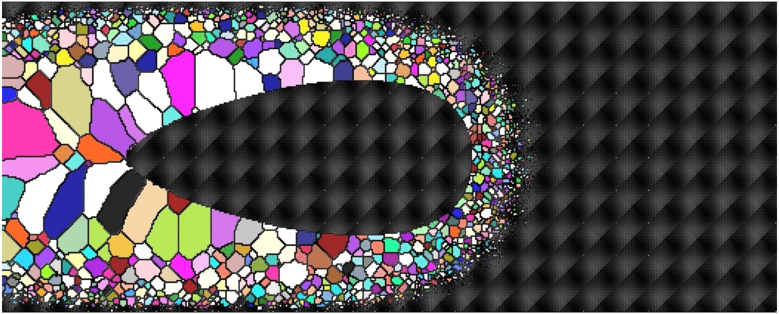
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



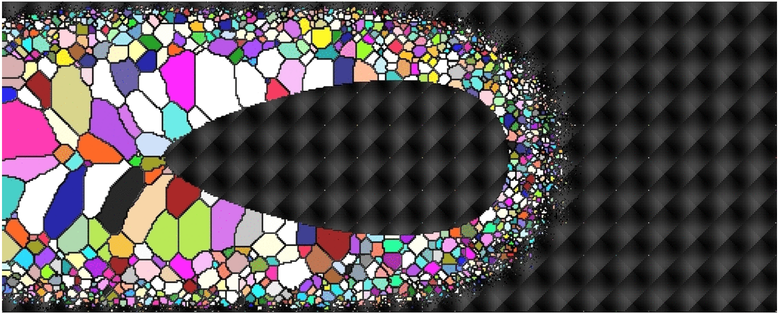
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



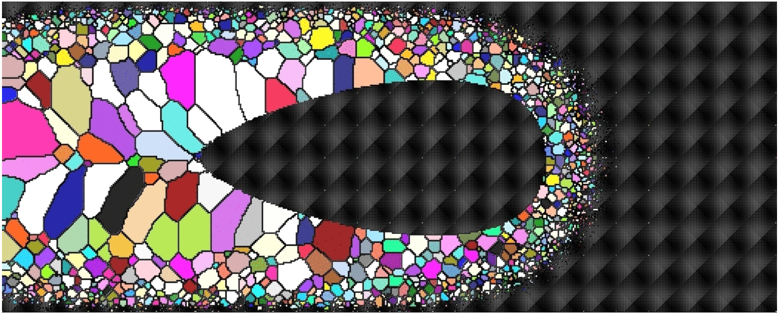
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



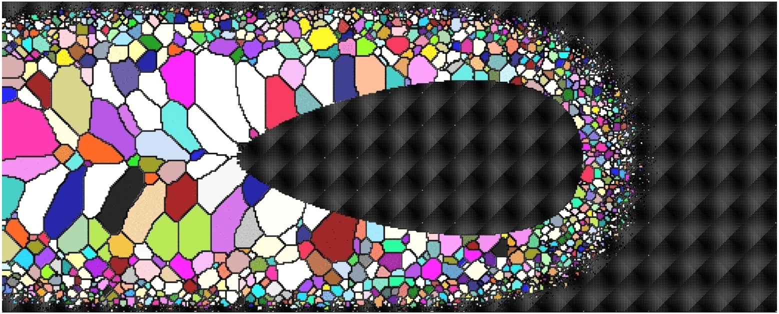
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



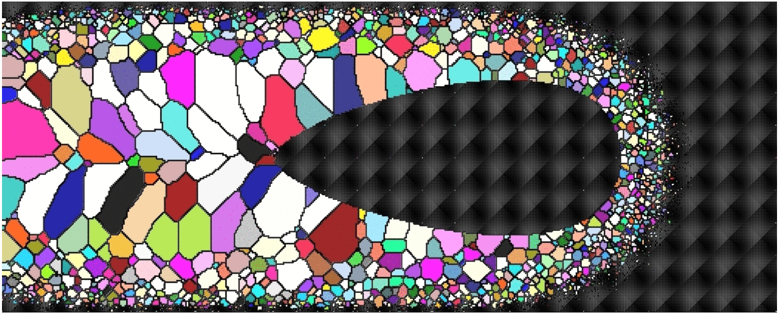
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



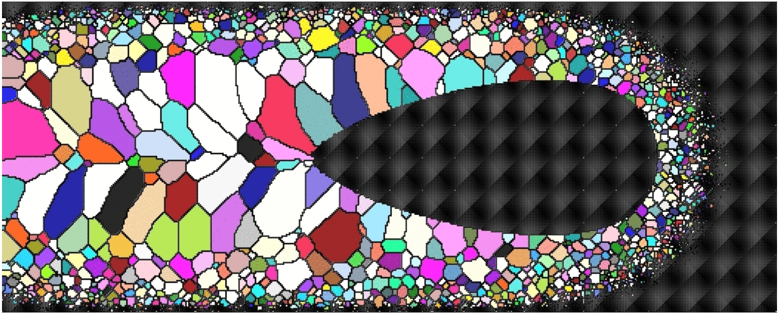
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



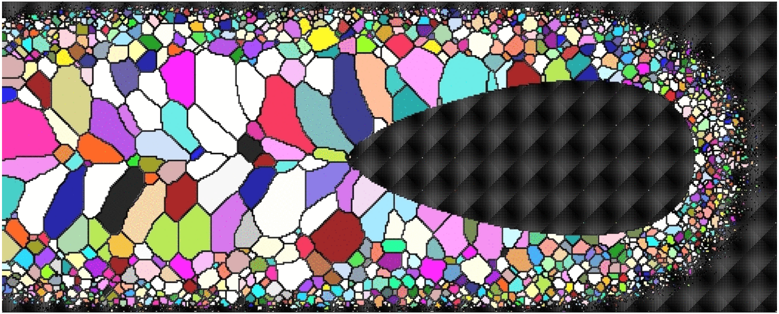
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



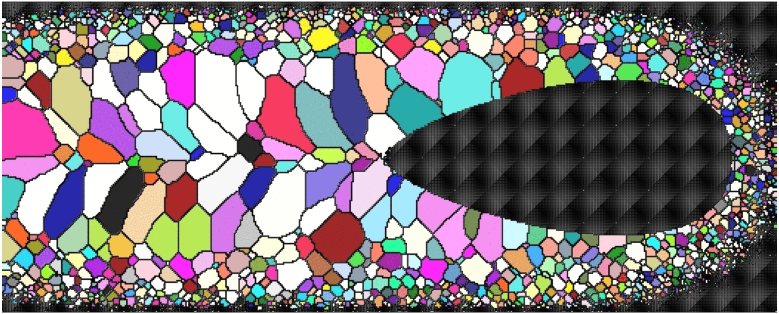
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



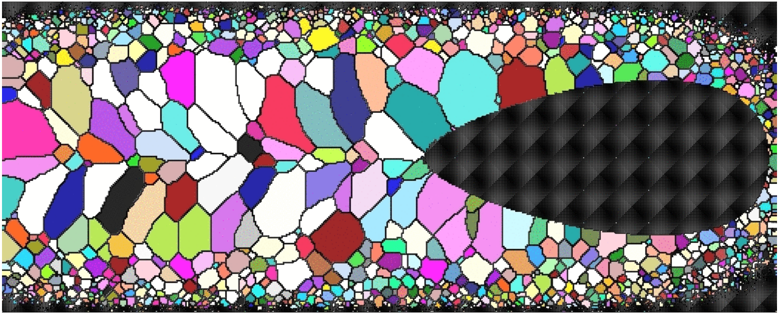
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



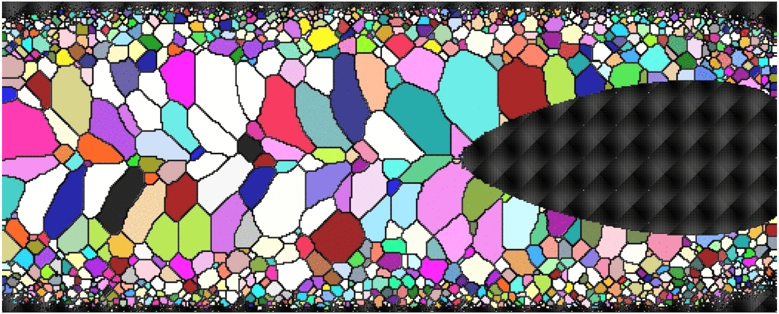
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



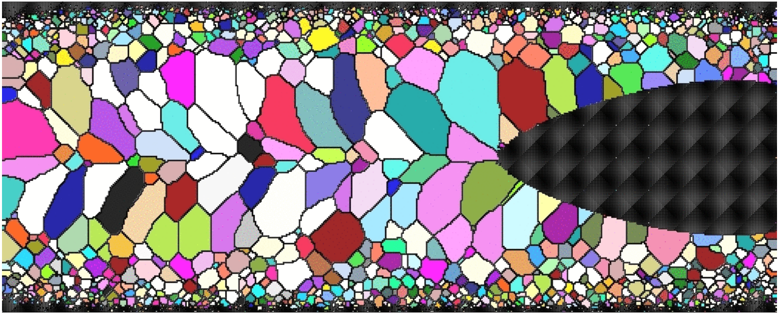
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



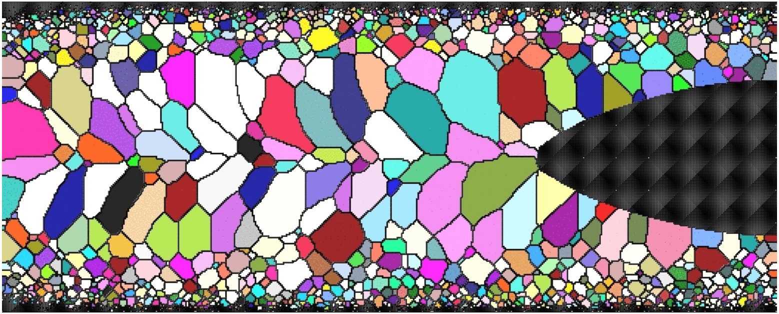
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



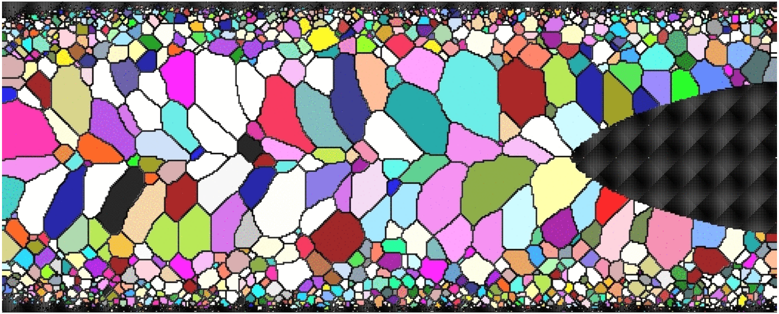
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



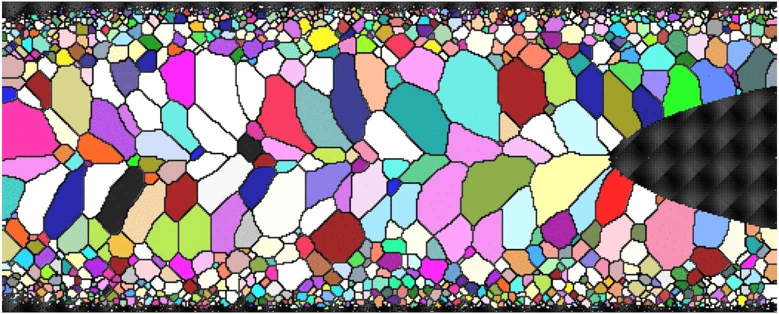
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



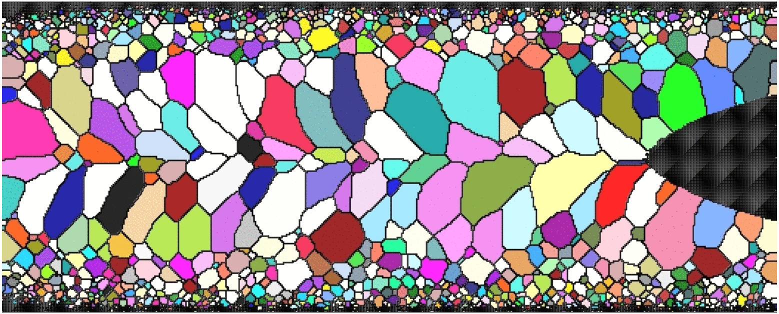
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



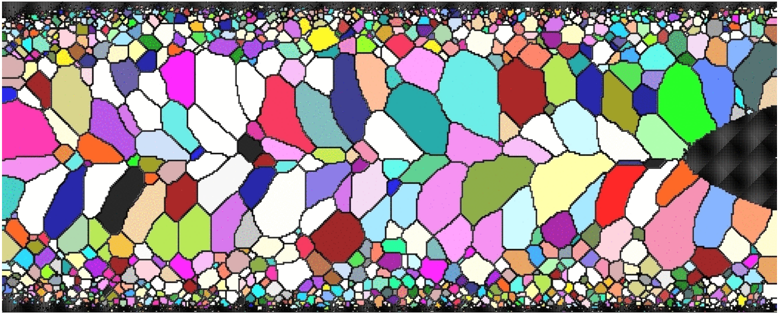
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



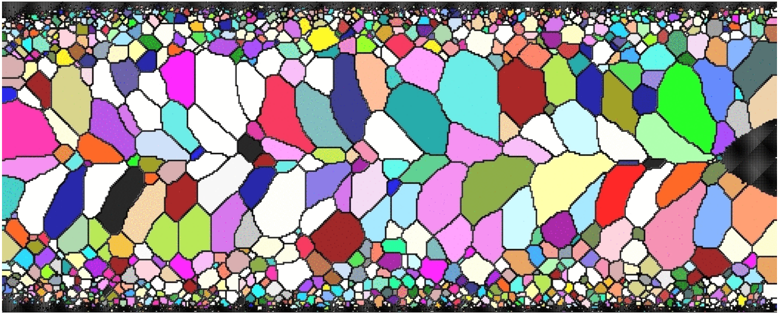
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



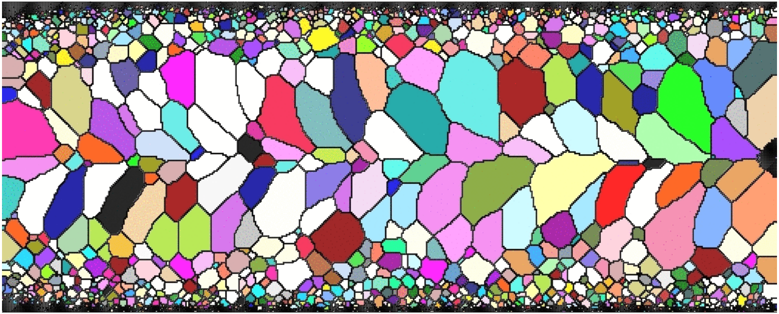
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



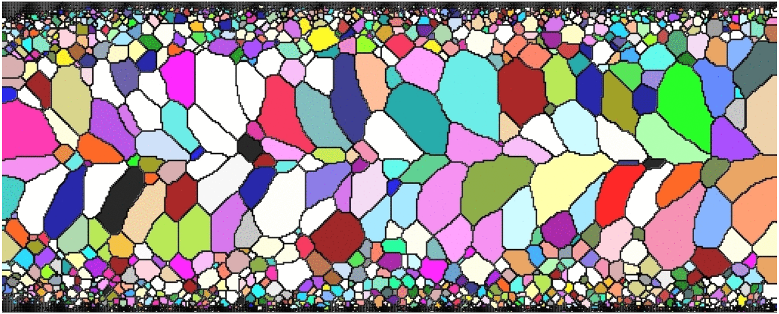
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



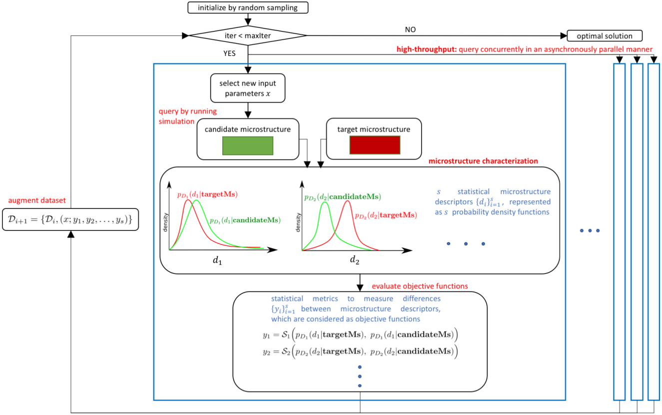
Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



Inverse problems in process-structure

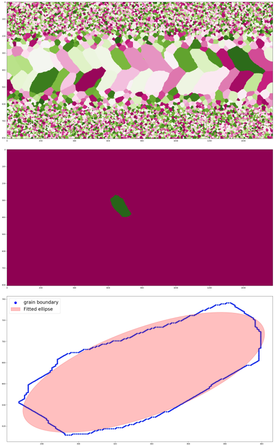
(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)

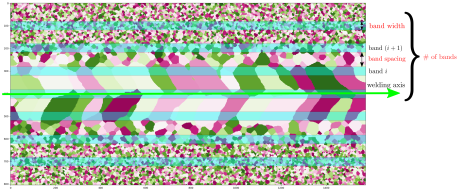


An asynchronous parallel Bayesian optimization workflow for inverse problems in process-structure linkage.

Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)

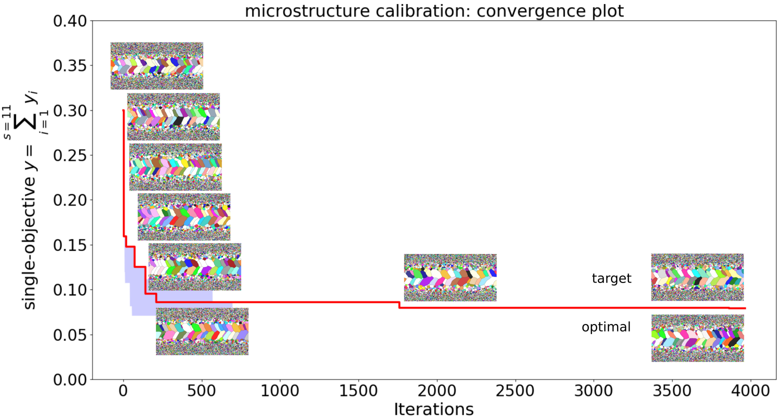




Collecting local + global statistical microstructure descriptors given a microstructure.

Inverse problems in process-structure

(joint work w/ Laura Swiler, John Mitchell, Tim Wildey, Theron Rodgers)



Reverse engineering an AM specimen through kinetic Monte Carlo (SPPARKS).

Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)

Reference

Active learning from chemical composition space to material property
Anh Tran et al. (2020b). "Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys". In: *The Journal of Chemical Physics* 153 (7), p. 074705.

Main ideas:

- Forward models:
  - MD-MLIAP: low-fidelity (low accuracy, low cost)
  - DFT: high-fidelity (high accuracy, high cost)
- Exploit correlation between low- and high-fidelity models
- Input: chemical composition
- Output/QoI: bulk modulus B_0
- What chemical composition would optimize the QoI?

Inverse problems in composition-property

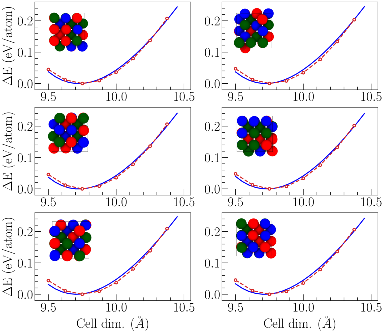
(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)

Ab-initio:

- DFT implemented in Quantum ESPRESSO
- high cost + high accuracy $\rightarrow$ high-fidelity

MD:

- MD with ML interatomic potential (SNAP)
- orders of magnitude faster
- low cost + low accuracy $\rightarrow$ low-fidelity



EOS calculations for 6 configs. red line: DFT; blue line: MD + SNAP

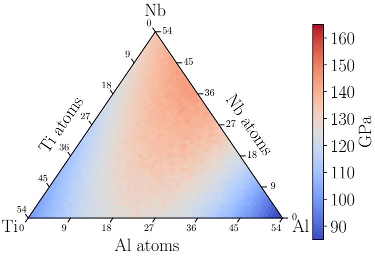
Birch-Murnaghan polynomials for B_0:

$$E(V) = E_0 + \frac{9 V_0 B_0}{16} \left\{ B_0' \left[\left(\frac{V_0}{V} \right)^{\frac{3}{2}} - 1 \right]^3 + \left[\left(\frac{V_0}{V} \right)^{\frac{3}{2}} - 1 \right]^2 \left[6 - 4 \left(\frac{V_0}{V} \right)^{\frac{3}{2}} \right] \right\} \quad (12)$$

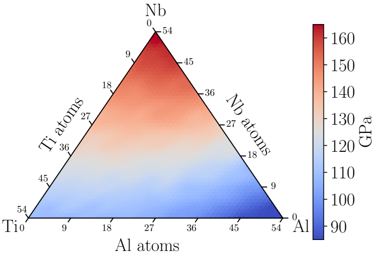
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)

$R^2 = 0.7122$: not exactly the same



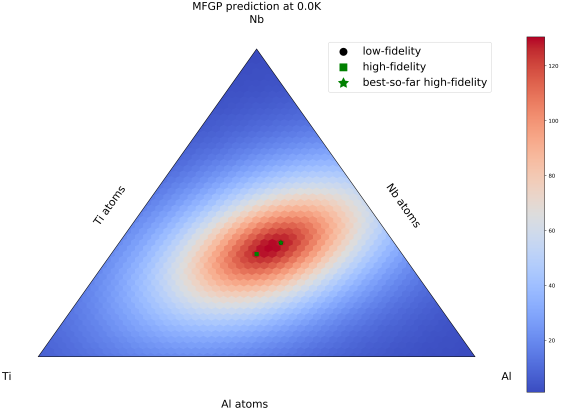
Low-fidelity: MD with SNAP potential.



Multi-fidelity GP ≈ high-fidelity: DFT.

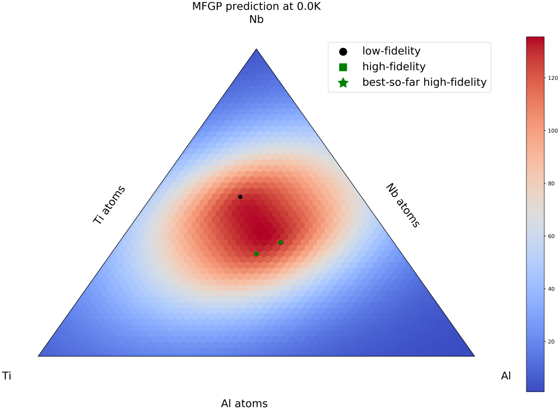
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



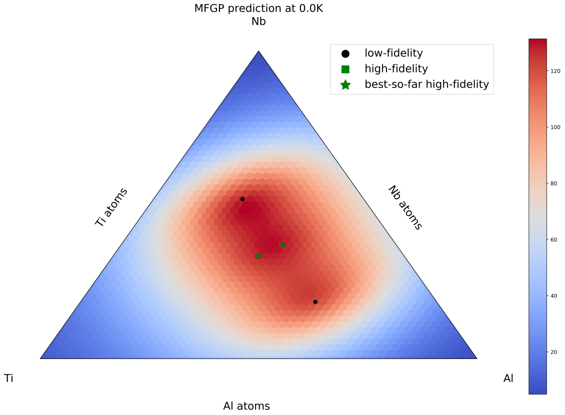
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



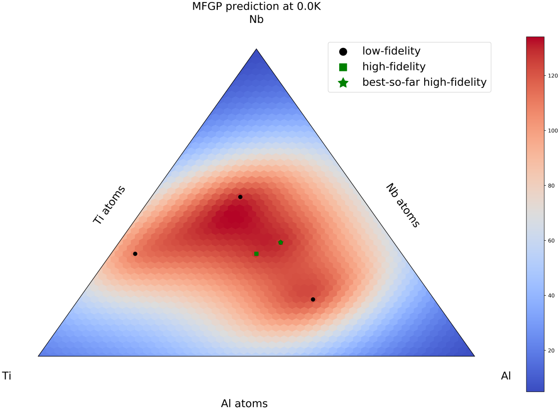
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



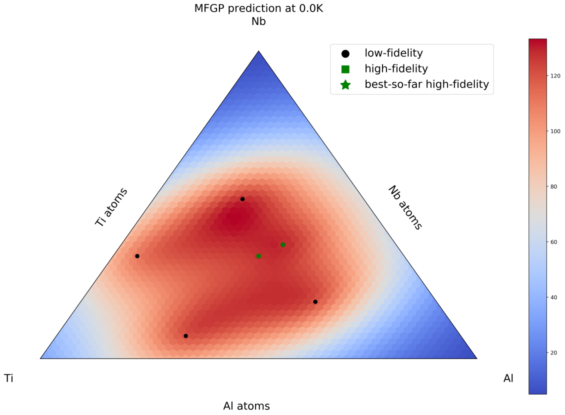
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



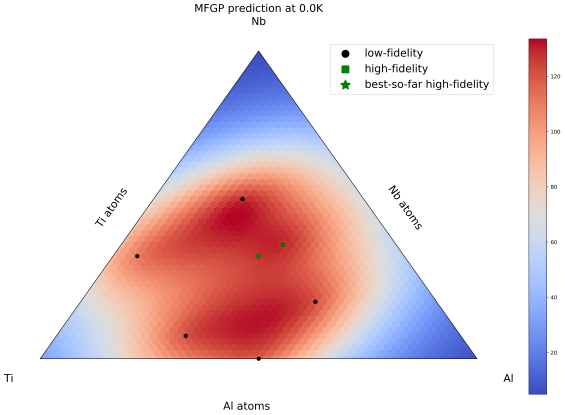
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



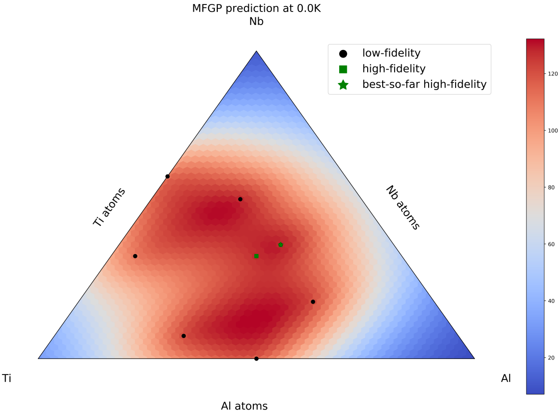
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



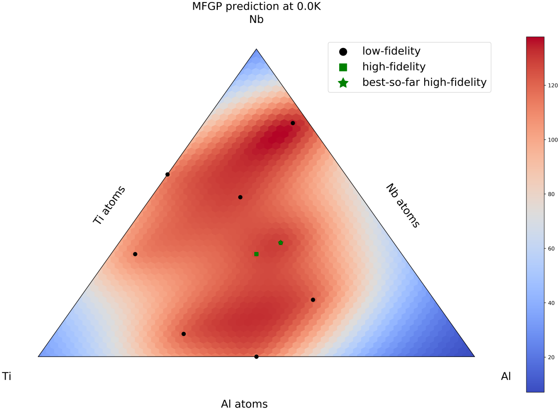
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



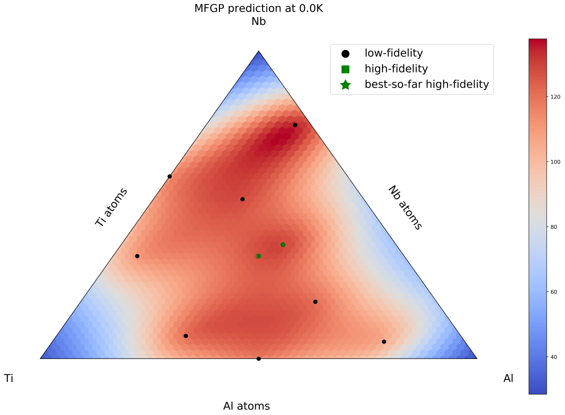
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



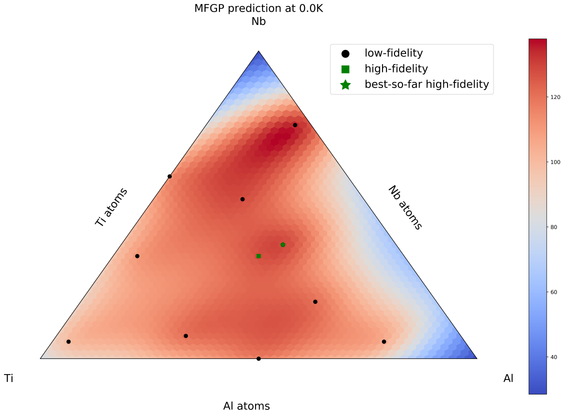
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



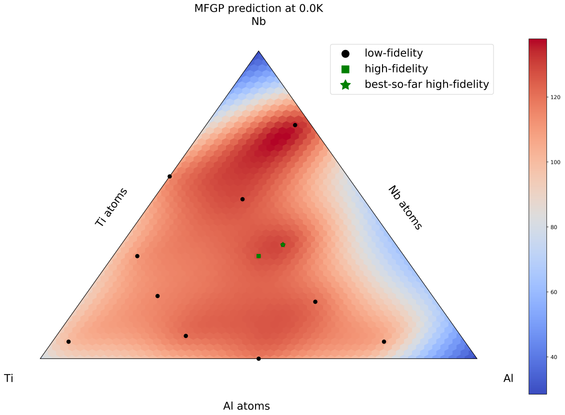
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



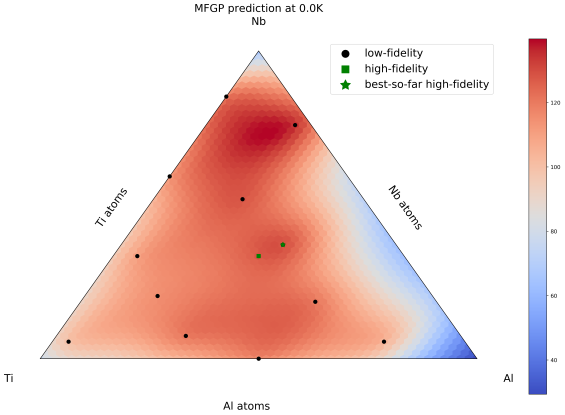
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



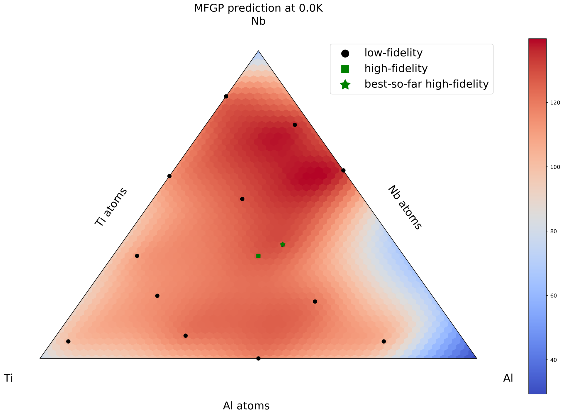
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



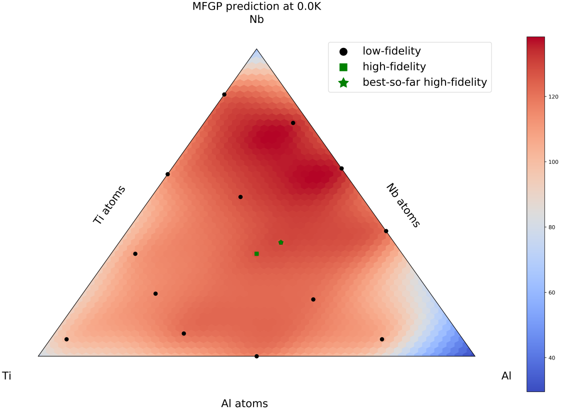
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



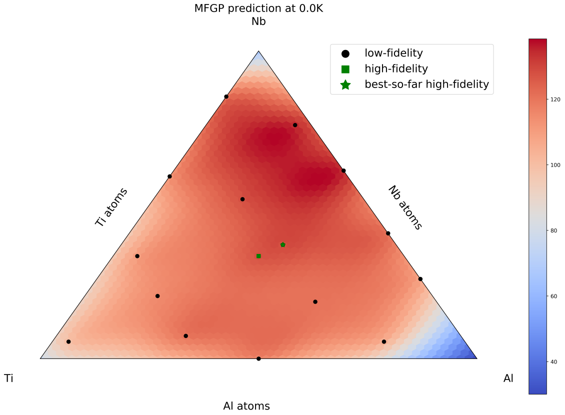
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



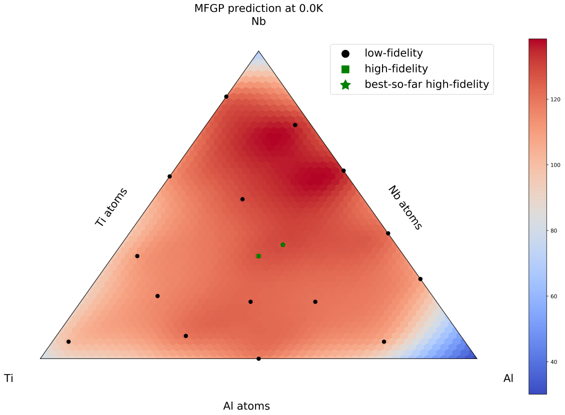
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



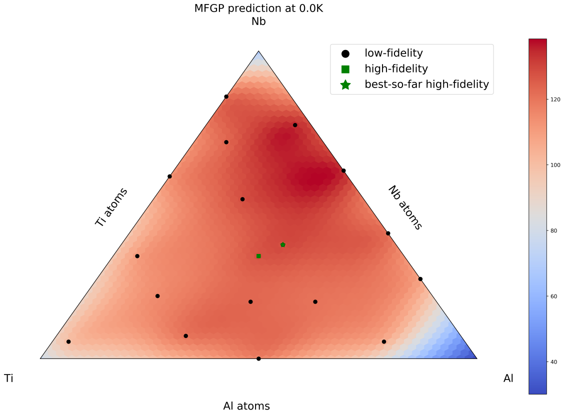
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



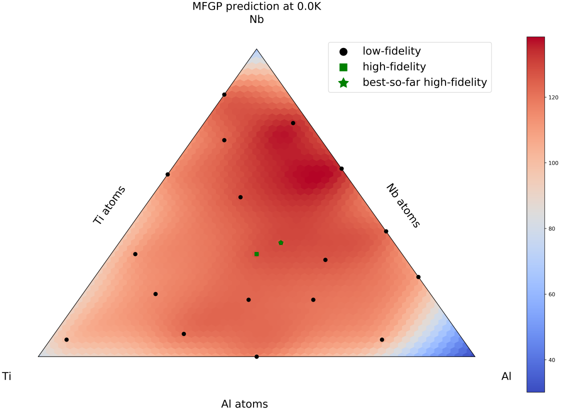
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



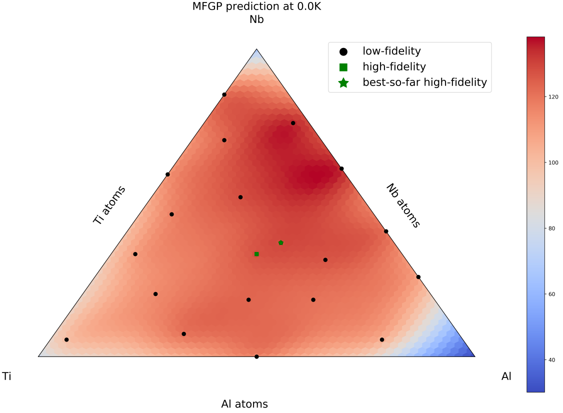
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



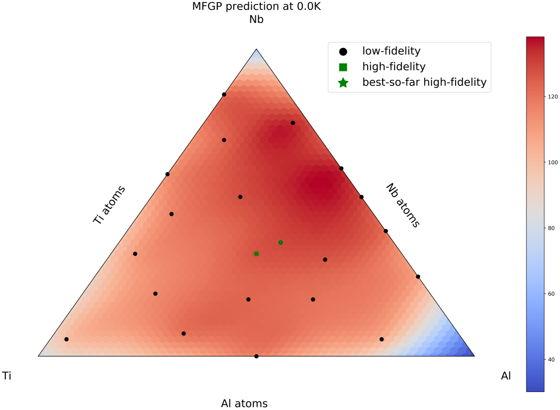
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



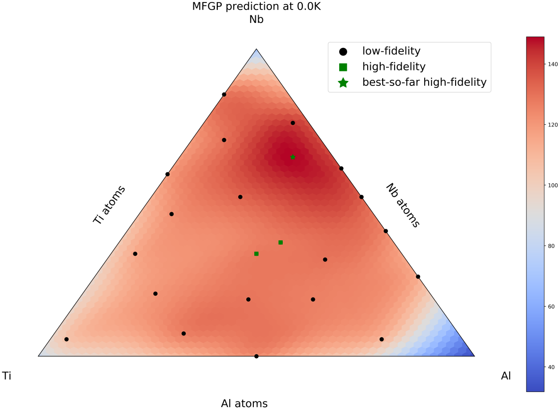
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



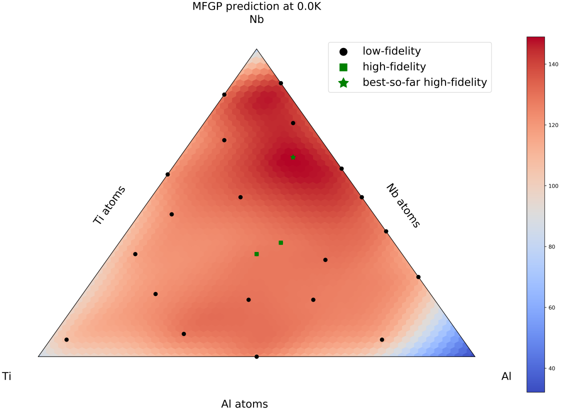
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



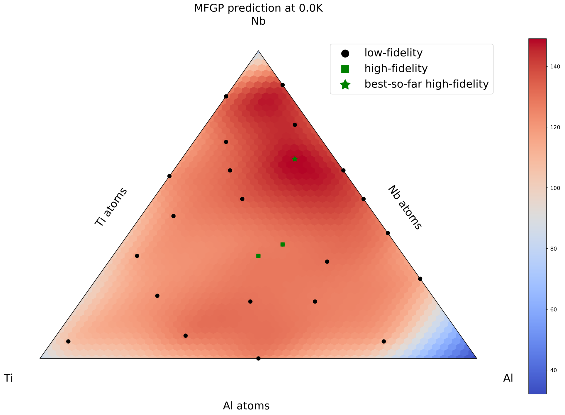
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



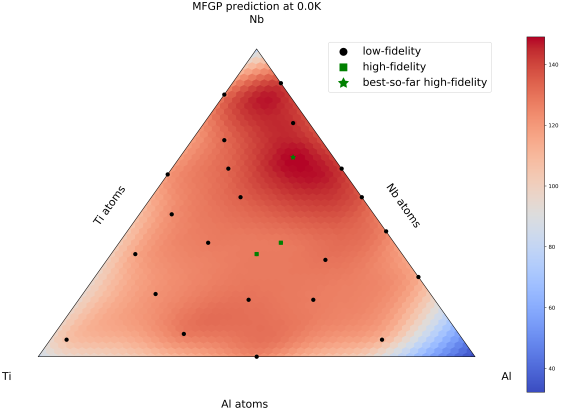
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



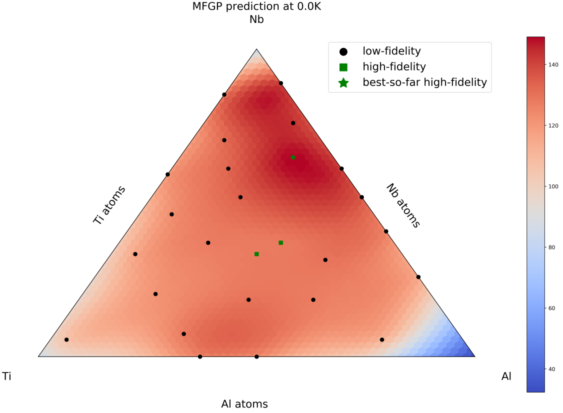
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



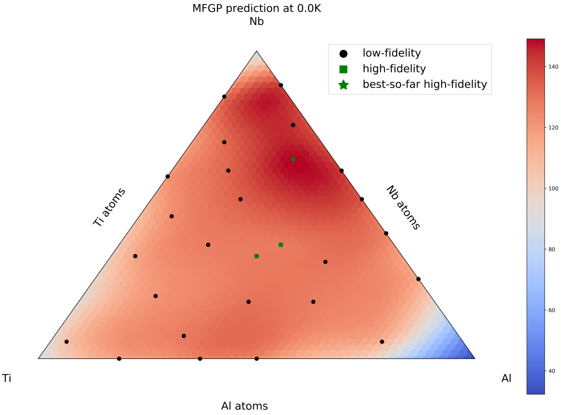
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



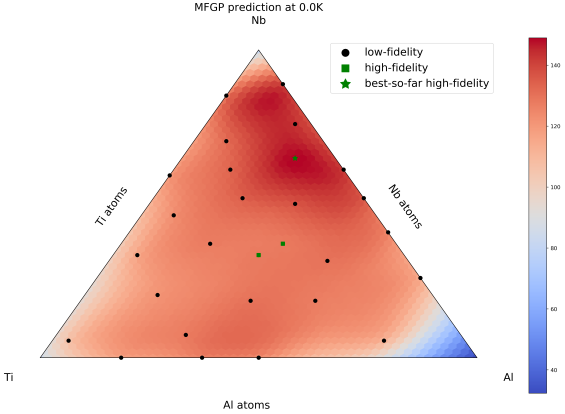
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



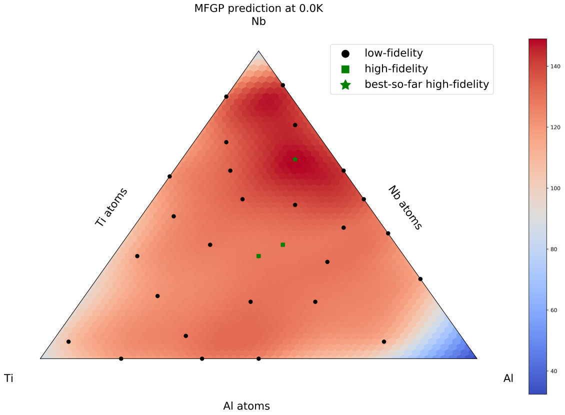
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



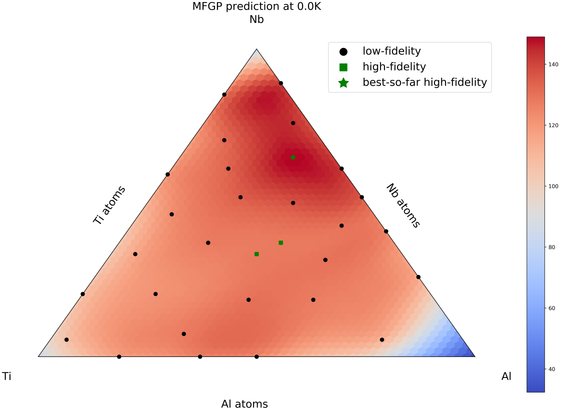
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



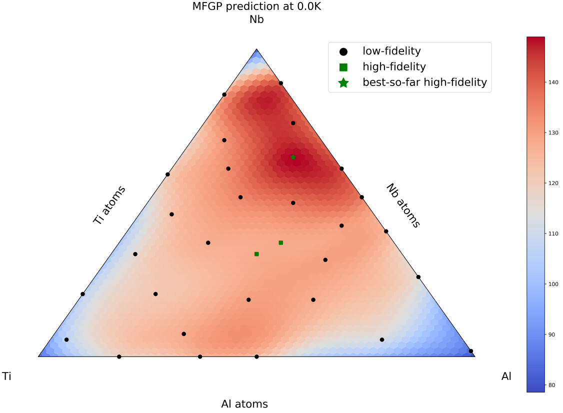
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



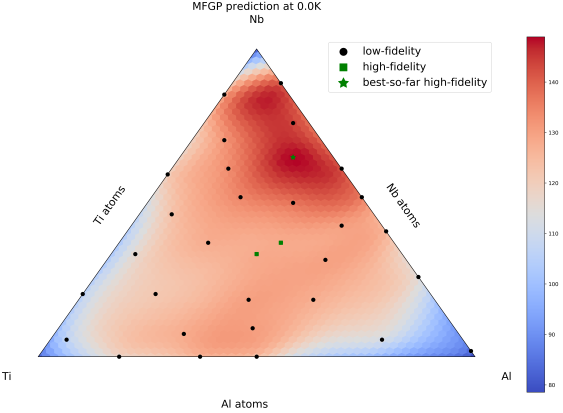
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



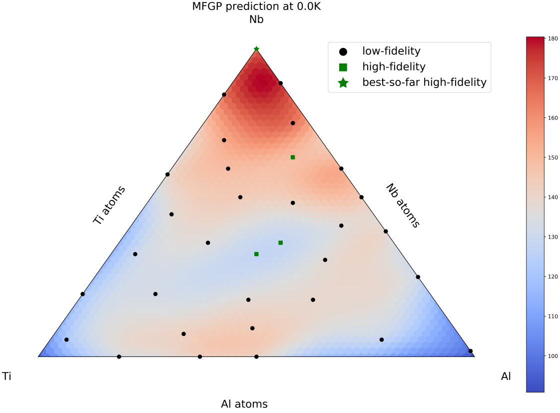
Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



Inverse problems in composition-property

(joint work w/ Julien Tranchida, Aidan Thompson, Tim Wildey)



Inverse problems in structure-property

(joint work w/ Tim Wildey)

Reference

Monday AM | October 10, 2022
310 | David L. Lawrence Convention Center
9:20 AM
Solving Stochastic Inverse Problems for Property–structure Linkages: Using Data-consistent Inversion and Machine Learning: Tim Wildey[1] ; Anh Tran[1] ; [1] Sandia National Labs

Anh Tran and Tim Wildey (2020). "Solving stochastic inverse problems for property-structure linkages using data-consistent inversion and machine learning". In: *JOM* 73, pp. 72–89

Conclusion

Takeaway message

Gaussian process is a versatile machine learning, uncertainty quantification, and optimization toolbox for ICME applications.

This talk: two parts

- a gentle tutorial to Gaussian process and Bayesian optimization
- multiscale ICME applications
  - density functional theory: Quantum ESPRESSO
  - molecular dynamics: LAMMPS
  - kinetic Monte Carlo: SPPARKS
  - crystal plasticity finite element: DREAM.3D + DAMASK

Thank you for your time and listening.

References

Methodology:

- Anh Tran et al. (2022). "aphBO-2GP-3B: a budgeted asynchronous parallel multi-acquisition functions for constrained Bayesian optimization on high-performing computing architecture". In: *Structural and Multidisciplinary Optimization* 65.4, pp. 1–45
- Anh Tran (Aug. 2021). "Scalable^3-BO: Big Data meets HPC - A scalable asynchronous parallel high-dimensional Bayesian optimization framework on supercomputers". In: *Proceedings of the ASME 2021 IDETC/CIE*. vol. Volume 1: 41th Computers and Information in Engineering Conference. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers
- Anh Tran et al. (Aug. 2020c). "srMO-BO-3GP: A sequential regularized multi-objective constrained Bayesian optimization for design applications". In: *Proceedings of the ASME 2020 IDETC/CIE*. vol. Volume 1: 40th Computers and Information in Engineering Conference. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers
- Anh Tran, Tim Wildey, and Scott McCann (2020). "sMF-BO-2CoGP: A sequential multi-fidelity constrained Bayesian optimization for design applications". In: *Journal of Computing and Information Science in Engineering* 20.3, pp. 1–15
- Anh Tran, Tim Wildey, and Scott McCann (Aug. 2019). "sBF-BO-2CoGP: A sequential bi-fidelity constrained Bayesian optimization for design applications". In: *Proceedings of the ASME 2019 IDETC/CIE*. vol. Volume 1: 39th Computers and Information in Engineering Conference. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. V001T02A073. American Society of Mechanical Engineers
- Anh Tran, Minh Tran, and Yan Wang (2019). "Constrained mixed-integer Gaussian mixture Bayesian optimization and its applications in designing fractal and auxetic metamaterials". In: *Structural and Multidisciplinary Optimization* 59 (6), pp. 2131–2154
- Anh Tran et al. (2019a). "pBO-2GP-3B: A batch parallel known/unknown constrained Bayesian optimization with feasibility classification and its applications in computational fluid dynamics". In: *Computer Methods in Applied Mechanics and Engineering* 347, pp. 827–852

Applications:

- Anh Tran and Tim Wildey (2020). "Solving stochastic inverse problems for property-structure linkages using data-consistent inversion and machine learning". In: *JOM* 73, pp. 72–89
- Anh Tran et al. (2020b). "Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys". In: *The Journal of Chemical Physics* 153 (7), p. 074705
- Anh Tran et al. (2020a). "An active-learning high-throughput microstructure calibration framework for process-structure linkage in materials informatics". In: *Acta Materialia* 194, pp. 80–92
- Stefano Travaglino et al. (2020). "Computational optimization study of transcatheter aortic valve leaflet design using porcine and bovine leaflets". In: *Journal of Biomechanical Engineering* 142 (1)
- Anh Tran et al. (2019b). "WearGP: A computationally efficient machine learning framework for local erosive wear predictions via nodal Gaussian processes". In: *Wear* 422, pp. 9–26
- Anh Tran, Lijuan He, and Yan Wang (2018). "An efficient first-principles saddle point searching method based on distributed kriging metamodels". In: *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering* 4.1, p. 011006

References I

Agrawal, Ankit and Alok Choudhary (2016). "Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science". In: *APL Materials* 4.5, p. 053208.

Bauer, Matthias, Mark van der Wilk, and Carl Edward Rasmussen (2016). "Understanding probabilistic sparse Gaussian process approximations". In: *Advances in neural information processing systems* 29, pp. 1533–1541.

Bonilla, Edwin V, Karl Krauth, and Amir Dezfouli (2019). "Generic inference in latent Gaussian process models.". In: *Journal of Machine Learning Research* 20.117, pp. 1–63.

Burt, David R., Carl Edward Rasmussen, and Mark van der Wilk (2020). "Convergence of Sparse Variational Inference in Gaussian Processes Regression". In: *Journal of Machine Learning Research* 21.131, pp. 1–63.

Hey, Tony, Stewart Tansley, Kristin M Tolle, et al. (2009). *The fourth paradigm: data-intensive scientific discovery*. Vol. 1. Microsoft research Redmond, WA.

Holdren, John P et al. (2014). "Materials genome initiative strategic plan (2014)". In: *National Science And Technology Council.*

References II

- Huang, Deng et al. (2006). "Global optimization of stochastic black-box systems via sequential kriging meta-models". In: *Journal of Global Optimization* 34.3, pp. 441–466.
- Lander, Eric et al. (2021). "Materials genome initiative strategic plan (2021)". In: *National Science And Technology Council*.
- Li, Mu, James Tin-Yau Kwok, and Baoliang Lü (2010). "Making large-scale Nyström approximation possible". In: *ICML 2010-Proceedings, 27th International Conference on Machine Learning*, p. 631.
- Martinsson, Per-Gunnar and Joel A Tropp (2020). "Randomized numerical linear algebra: Foundations and algorithms". In: *Acta Numerica* 29, pp. 403–572.
- Matthews, Alexander G de G et al. (2016). "On sparse variational methods and the Kullback-Leibler divergence between stochastic processes". In: *Artificial Intelligence and Statistics*. PMLR, pp. 231–239.
- Mockus, Jonas (1975). "On Bayesian methods for seeking the extremum". In: *Optimization Techniques IFIP Technical Conference*. Springer, pp. 400–404.

References III

- Mockus, Jonas (1982). "The Bayesian approach to global optimization". In: *System Modeling and Optimization*, pp. 473–481.
- National Science and Technology Council (US) (2011). *Materials Genome Initiative for global competitiveness*. Executive Office of the President, National Science and Technology Council.
- Quiñonero-Candela, Joaquin and Carl Edward Rasmussen (2005). "A unifying view of sparse approximate Gaussian process regression". In: *Journal of Machine Learning Research* 6.Dec, pp. 1939–1959.
- Quiñonero-Candela, Joaquin, Carl Edward Rasmussen, and Christopher KI Williams (2007). "Approximation methods for Gaussian process regression". In: *Large-scale kernel machines*, pp. 203–224.
- Srinivas, Niranjan et al. (2009). "Gaussian process optimization in the bandit setting: No regret and experimental design". In: *arXiv preprint arXiv:0912.3995*.
- Srinivas, Niranjan et al. (2012). "Information-theoretic regret bounds for Gaussian process optimization in the bandit setting". In: *IEEE Transactions on Information Theory* 58.5, pp. 3250–3265.

References IV

Titsias, Michalis (2009a). "Variational learning of inducing variables in sparse Gaussian processes". In: *Artificial Intelligence and Statistics*, pp. 567–574.

Titsias, Michalis K (2009b). "Variational model selection for sparse Gaussian process regression". In: *Report, University of Manchester, UK*.

Tran, Anh (Aug. 2021). "Scalable^3-BO: Big Data meets HPC - A scalable asynchronous parallel high-dimensional Bayesian optimization framework on supercomputers". In: *Proceedings of the ASME 2021 IDETC/CIE*. Vol. Volume 1: 41th Computers and Information in Engineering Conference. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers.

Tran, Anh, Lijuan He, and Yan Wang (2018). "An efficient first-principles saddle point searching method based on distributed kriging metamodels". In: *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering* 4.1, p. 011006.

Tran, Anh, Minh Tran, and Yan Wang (2019). "Constrained mixed-integer Gaussian mixture Bayesian optimization and its applications in designing fractal and auxetic metamaterials". In: *Structural and Multidisciplinary Optimization* 59 (6), pp. 2131–2154.

References V

Tran, Anh and Tim Wildey (2020). “Solving stochastic inverse problems for property-structure linkages using data-consistent inversion and machine learning”. In: *JOM* 73, pp. 72–89.

Tran, Anh, Tim Wildey, and Scott McCann (Aug. 2019). “sBF-BO-2CoGP: A sequential bi-fidelity constrained Bayesian optimization for design applications”. In: *Proceedings of the ASME 2019 IDETC/CIE*. Vol. Volume 1: 39th Computers and Information in Engineering Conference. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. V001T02A073. American Society of Mechanical Engineers.

— (2020). “sMF-BO-2CoGP: A sequential multi-fidelity constrained Bayesian optimization for design applications”. In: *Journal of Computing and Information Science in Engineering* 20.3, pp. 1–15.

Tran, Anh et al. (2019a). “pBO-2GP-3B: A batch parallel known/unknown constrained Bayesian optimization with feasibility classification and its applications in computational fluid dynamics”. In: *Computer Methods in Applied Mechanics and Engineering* 347, pp. 827–852.

References VI

- Tran, Anh et al. (2019b). "WearGP: A computationally efficient machine learning framework for local erosive wear predictions via nodal Gaussian processes". In: *Wear* 422, pp. 9–26.
- Tran, Anh et al. (2020a). "An active-learning high-throughput microstructure calibration framework for process-structure linkage in materials informatics". In: *Acta Materialia* 194, pp. 80–92.
- Tran, Anh et al. (2020b). "Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys". In: *The Journal of Chemical Physics* 153 (7), p. 074705.
- Tran, Anh et al. (Aug. 2020c). "srMO-BO-3GP: A sequential regularized multi-objective constrained Bayesian optimization for design applications". In: *Proceedings of the ASME 2020 IDETC/CIE*. Vol. Volume 1: 40th Computers and Information in Engineering Conference. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. American Society of Mechanical Engineers.

References VII

Tran, Anh et al. (2022). "aphBO-2GP-3B: a budgeted asynchronous parallel multi-acquisition functions for constrained Bayesian optimization on high-performing computing architecture". In: *Structural and Multidisciplinary Optimization* 65.4, pp. 1–45.

Travaglino, Stefano et al. (2020). "Computational optimization study of transcatheter aortic valve leaflet design using porcine and bovine leaflets". In: *Journal of Biomechanical Engineering* 142 (1).

Vanhatalo, Jarno et al. (2012). "Bayesian modeling with Gaussian processes using the GPstuff toolbox". In: *arXiv preprint arXiv:1206.5754*.

— (2013). "GPstuff: Bayesian modeling with Gaussian processes". In: *Journal of Machine Learning Research* 14.Apr, pp. 1175–1179.

Williams, Christopher and Matthias Seeger (2001). "Using the Nyström method to speed up kernel machines". In: *Advances in neural information processing systems* 13, pp. 682–688.

Wilson, James, Frank Hutter, and Marc Deisenroth (2018). "Maximizing acquisition functions for Bayesian optimization". In: *Advances in Neural Information Processing Systems* 31, pp. 9884–9895.

Big Data

- $p(\cdot)$: true pdf
- $q(\cdot)$: approximate pdf

Assume the fully independent training conditional (FITC) Quiñonero-Candela and Rasmussen 2005; Quiñonero-Candela, Rasmussen, and Williams 2007, augment the joint model $p(\mathbf{f}_*, \mathbf{f})$ as

$$p(\mathbf{f}_*, \mathbf{f}) = \int p(\mathbf{f}_*, \mathbf{f}, \mathbf{u}) d\mathbf{u} = \int p(\mathbf{f}_*, \mathbf{f}|\mathbf{u}) p(\mathbf{u}) d\mathbf{u}, \quad (13)$$

$\mathbf{u}$: inducing variables at m locations $\mathbf{X}_\mathbf{u}$. The training and testing conditionals are

$$p(\mathbf{f}|\mathbf{u}) = \mathcal{N}(\mathbf{m} + \mathbf{K}_{\mathbf{f},\mathbf{u}}\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}(\mathbf{u} - \mathbf{m}),\ \mathbf{K}_{\mathbf{f},\mathbf{f}} - \mathbf{Q}_{\mathbf{f},\mathbf{f}}), \quad (14)$$

and

$$p(\mathbf{f}_*|\mathbf{u}) = \mathcal{N}(\mathbf{m} + \mathbf{K}_{*,\mathbf{u}}\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}(\mathbf{u} - \mathbf{m}),\ \mathbf{K}_{*,*} - \mathbf{Q}_{*,*}), \quad (15)$$

where

$$\mathbf{Q}_{\mathbf{a},\mathbf{b}} := \mathbf{K}_{\mathbf{a},\mathbf{u}}\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{K}_{\mathbf{u},\mathbf{b}}. \quad (16)$$

The likelihood and inducing priors remain the same, i.e. $p(\mathbf{y}|\mathbf{f}) = \mathcal{N}(\mathbf{f}, \sigma^2\mathbf{I})$, and $p(\mathbf{u}) = \mathcal{N}(\mathbf{m}, \mathbf{K}_{\mathbf{u},\mathbf{u}})$.

Sparse variational GP

FITC training prior based on the inducing priors is modified as

$$q(\mathbf{f}|\mathbf{u}) = \prod_{i=1}^{n} p(\mathbf{f}_i|\mathbf{u}) = \mathcal{N}(\mathbf{m} + \mathbf{K}_{\mathbf{f},\mathbf{u}}\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}(\mathbf{u} - \mathbf{m}), \text{Diag}[\mathbf{K}_{\mathbf{f},\mathbf{f}} - \mathbf{Q}_{\mathbf{f},\mathbf{f}}]) \quad (17)$$

and keeping the testing prior the same

$$q(\mathbf{f}_*|\mathbf{u}) = p(\mathbf{f}_*|\mathbf{u}) = \mathcal{N}(\mathbf{m} + \mathbf{K}_{*,\mathbf{u}}\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}(\mathbf{u} - \mathbf{m}),\ \mathbf{K}_{*,*} - \mathbf{Q}_{*,*}), \quad (18)$$

the effective prior under the FITC assumption is

$$q(\mathbf{f}, \mathbf{f}_*) = \mathcal{N}\left(\begin{bmatrix} \mathbf{m} \\ \mathbf{m} \end{bmatrix}, \begin{bmatrix} \mathbf{Q}_{\mathbf{f},\mathbf{f}} - \text{Diag}[\mathbf{Q}_{\mathbf{f},\mathbf{f}} - \mathbf{K}_{\mathbf{f},\mathbf{f}}] & \mathbf{Q}_{\mathbf{f},*} \\ \mathbf{Q}_{*,\mathbf{f}} & \mathbf{K}_{*,*} \end{bmatrix} \right), \quad (19)$$

which implies the testing distribution as

$$\begin{aligned} q(\mathbf{f}_*|\mathbf{y}) &= \mathcal{N}(\mathbf{m} + \mathbf{Q}_{*,\mathbf{f}}(\mathbf{Q}_{\mathbf{f},\mathbf{f}} + \Lambda)^{-1}(\mathbf{y} - \mathbf{m}), \mathbf{K}_{*,*} - \mathbf{Q}_{*,\mathbf{f}}(\mathbf{Q}_{\mathbf{f},\mathbf{f}} + \Lambda)^{-1}\mathbf{Q}_{\mathbf{f},*}) \\ &= \mathcal{N}(\mathbf{m} + \mathbf{K}_{*,\mathbf{u}}\Sigma\mathbf{K}_{\mathbf{u},\mathbf{f}}\Lambda^{-1}(\mathbf{y} - \mathbf{m}), \mathbf{K}_{*,*} - \mathbf{Q}_{*,*} + \mathbf{K}_{*,\mathbf{u}}\Sigma\mathbf{K}_{\mathbf{u},*}) \end{aligned}, \quad (20)$$

where $\Sigma = [\mathbf{K}_{\mathbf{u},\mathbf{u}} + \mathbf{K}_{\mathbf{u},\mathbf{f}}\Lambda^{-1}\mathbf{K}_{\mathbf{f},\mathbf{u}}]^{-1}$ and $\Lambda = \text{Diag}[\mathbf{K}_{\mathbf{f},\mathbf{f}} - \mathbf{Q}_{\mathbf{f},\mathbf{f}} + \sigma^2\mathbf{I}]$.

Sparse variational GP

The marginal likelihood conditioned on the inducing inputs is therefore

$$q(\mathbf{y}|\mathbf{X_u}) = \int\int p(\mathbf{y}|\mathbf{f})q(\mathbf{f}|\mathbf{u})p(\mathbf{u}|\mathbf{X_u})d\mathbf{u}d\mathbf{f} = \int p(\mathbf{y}|\mathbf{f})q(\mathbf{f}|\mathbf{X_u})d\mathbf{f}, \qquad (21)$$

which implies the log marginal likelihood as

$$\log q(\mathbf{y}|\mathbf{X_u}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{Q_{f,f}} + \Lambda| - \frac{1}{2}(\mathbf{y}-\mathbf{m})^\top[\mathbf{Q_{f,f}} + \Lambda]^{-1}(\mathbf{y}-\mathbf{m}), \qquad (22)$$

where $\Lambda = \mathrm{Diag}[\mathbf{K_{f,f}} - \mathbf{Q_{f,f}}] + \sigma^2\mathbf{I}$.
Cost complexity: $\mathcal{O}(nm^2)$ Li, Kwok, and Lü 2010; Williams and Seeger 2001. (Note: do not multiply matrices directly – cf. Section 14.3 Martinsson and Tropp 2020).

Variational inference

Mostly follow Titsias 2009a; Titsias 2009b and Bonilla, Krauth, and Dezfouli 2019.
Definition of conditionally independent condition:

$$p(\mathbf{f}|\mathbf{u},\mathbf{y}) = p(\mathbf{f}|\mathbf{u}), \tag{23}$$

which implies $p(\mathbf{f},\mathbf{u}|\mathbf{y}) = p(\mathbf{f}|\mathbf{u},\mathbf{y})p(\mathbf{u}|\mathbf{y}) \approx q(\mathbf{f},\mathbf{u}) = p(\mathbf{f}|\mathbf{u})q(\mathbf{u})$, where $q(\mathbf{u})$ is the approximate variational posterior. Main tool: Jensen's inequality.

$$\begin{aligned}
\log q(\mathbf{y}|\mathbf{X_u}) &= \log \int\int p(\mathbf{y}|\mathbf{f})q(\mathbf{f}|\mathbf{u})p(\mathbf{u}|\mathbf{X_u}) \times \frac{q(\mathbf{u},\mathbf{f})}{q(\mathbf{u},\mathbf{f})} d\mathbf{u}d\mathbf{f} \\
&\geq \int\int q(\mathbf{u},\mathbf{f}) \log \frac{p(\mathbf{y}|\mathbf{f})q(\mathbf{f}|\mathbf{u})p(\mathbf{u}|\mathbf{X_u})}{q(\mathbf{u},\mathbf{f})} d\mathbf{u}d\mathbf{f} \\
&= \int\int p(\mathbf{f}|\mathbf{u})q(\mathbf{u}) \log \frac{p(\mathbf{y}|\mathbf{f})\cancel{q(\mathbf{f}|\mathbf{u})}p(\mathbf{u}|\mathbf{X_u})}{\cancel{p(\mathbf{f}|\mathbf{u})}q(\mathbf{u})} d\mathbf{u}d\mathbf{f} \\
&= \int q(\mathbf{u}) \left\{ \int p(\mathbf{f}|\mathbf{u}) \log p(\mathbf{y}|\mathbf{f}) d\mathbf{f} + \log \frac{p(\mathbf{u}|\mathbf{X_u})}{q(\mathbf{u})} \right\} d\mathbf{u} \\
&= \int q(\mathbf{u}) \left\{ \log G(\mathbf{u},\mathbf{y}) + \log \frac{p(\mathbf{u}|\mathbf{X_u})}{q(\mathbf{u})} \right\} d\mathbf{u} \\
&= \int q(\mathbf{u}) \left\{ \log \frac{G(\mathbf{u},\mathbf{y})p(\mathbf{u}|\mathbf{X_u})}{q(\mathbf{u})} \right\} d\mathbf{u} := \mathcal{F}_V(\mathbf{X_u},\mathbf{u}),
\end{aligned} \tag{24}$$

$$\begin{aligned}
\log G(\mathbf{u},\mathbf{y}) &= \int p(\mathbf{f}|\mathbf{u}) \log p(\mathbf{y}|\mathbf{f}) d\mathbf{f} \\
&= \int p(\mathbf{f}|\mathbf{u}) \left\{ -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \mathrm{Tr}\left[\mathbf{y}\mathbf{y}^\top - 2\mathbf{y}\mathbf{f}^\top + \mathbf{f}\mathbf{f}^\top\right] \right\} d\mathbf{f} \\
&= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \mathrm{Tr}\left[\mathbf{y}\mathbf{y}^\top - 2\mathbf{y}\alpha^\top + \alpha\alpha^\top + \mathbf{Q}_{\mathbf{f},\mathbf{f}} - \mathbf{K}_{\mathbf{f},\mathbf{f}}\right] \\
&= \mathcal{N}(\mathbf{y}|\alpha,\sigma^2\mathbf{I}) - \frac{1}{2\sigma^2}\mathrm{Tr}[\mathrm{Cov}(\alpha)],
\end{aligned} \tag{25}$$

Variational inference

where $\alpha = \mathbf{f}|\mathbf{u}$, with

$$\mathbb{E}[\alpha] = \mathbb{E}[\mathbf{f}|\mathbf{u}] = \mathbf{m} + \mathbf{K}_{\mathbf{f},\mathbf{u}}\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}(\mathbf{u} - \mathbf{m}) \tag{26}$$

and

$$\text{Cov}[\alpha] = \text{Cov}[\mathbf{f}|\mathbf{u}] = \mathbf{K}_{\mathbf{f},\mathbf{f}} - \mathbf{Q}_{\mathbf{f},\mathbf{f}} = \mathbf{K}_{\mathbf{f},\mathbf{f}} - \mathbf{K}_{\mathbf{f},\mathbf{u}}\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{K}_{\mathbf{u},\mathbf{f}}. \tag{27}$$

Reverse Jensen's inequality to maximize the variational evidence lower bound $\mathcal{F}_V(\mathbf{X}_\mathbf{u}, \mathbf{u})$ w.r.t. $q(\mathbf{u})$

$$\begin{aligned} \mathcal{F}_V(\mathbf{X}_\mathbf{u}, \mathbf{u}) &= \int q(\mathbf{u}) \left\{ \log \frac{G(\mathbf{u},\mathbf{y})p(\mathbf{u}|\mathbf{X}_\mathbf{u})}{q(\mathbf{u})} \right\} d\mathbf{u} \\ &\leq \int \log G(\mathbf{u},\mathbf{y}) p(\mathbf{u}|\mathbf{X}_\mathbf{u}) d\mathbf{u} \\ &= \log[\mathcal{N}(\mathbf{y}|\mathbf{m}, \sigma^2\mathbf{I} + \mathbf{Q}_{\mathbf{f},\mathbf{f}})] - \frac{1}{2\sigma^2}\text{Tr}\left[\mathbf{K}_{\mathbf{f},\mathbf{f}} - \mathbf{K}_{\mathbf{f},\mathbf{u}}\mathbf{K}_{\mathbf{u},\mathbf{u}}^{-1}\mathbf{K}_{\mathbf{u},\mathbf{f}}\right] =: \mathcal{F}_V(\mathbf{X}_\mathbf{u}) \end{aligned} \tag{28}$$

Train sparse GP by maximizing $\mathcal{F}_V(\mathbf{X}_\mathbf{u})$. See also Vanhatalo et al. 2012, 2013, Bauer, Wilk, and Rasmussen 2016; Burt, Rasmussen, and Wilk 2020, Matthews et al. 2016.