SAND2021-2292C

Multi-fidelity ML/UQ and Bayesian Optimization for Materials Design: Application to Ternary Random Alloys

Anh Tran[2], Julien Tranchida[2], Aidan P. Thompson, Tim Wildey

Center for Computing Research, Sandia National Laboratories, Albuquerque, NM

[2] These two authors contributed equally to this work.

Main idea

This research uses multi-fidelity Gaussian process and Bayesian optimization to solve a multi-scale materials design inverse problem targeting composition-property relationship [TTWT20].

- Low-fidelity (LF): MD with machine learning potential.
- High-fidelity (HF): DFT.
- Input: chemical composition; output: bulk modulus.
- The global optimum is found at HF after 31 LF + 4 HF.

ML-IAP for MD: SNAP in LAMMPS [TST+15]

- A ML interatomic potential (ML-IAP) targeting energies, forces, and stress tensors.
- Describe local environment of each atom by a set of bispectrum components of the local neighbor density projected onto a basis of hyperspherical harmonics in 4D.
- ML-IAP implemented in LAMMPS.
- Trained using DFT & FitSnap.py wrapping DAKOTA [DEG+20].

DFT: Quantum ESPRESSO [GAB+17]

- Perdew–Burke-Ernzerhof (PBE) exchange-correlation energy.
- interaction between electrons and ions by projector augmented wave (PAW) pseudopotentials.
- Brillouin zone sampled using 2×2×2 k-point grid.
- N atoms and E elements: $\left(\binom{N}{E}\right) = \binom{N+E-1}{E-1}$ possibilities.
- Equation of State (EOS) computed through eight points, interpolated with a Birch-Murnaghan equation using a 3^{rd} order polynomial.

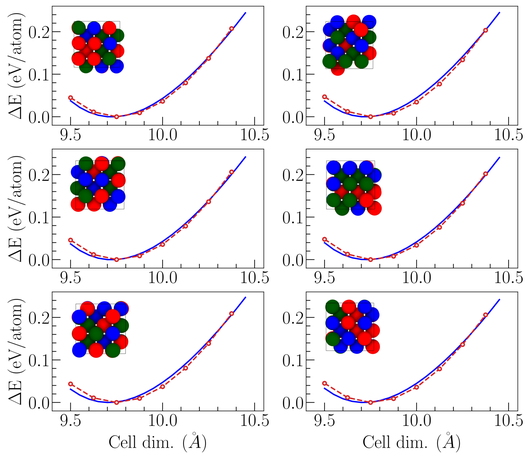


Figure: Equation of state calculations for six configurations. The red circles and blue lines display the high-fidelity DFT and low-fidelity SNAP results, respectively.

Multi-Fidelity Bayesian Optimization Theory [TWM20]

We assume auto-regressive model $f_H(\mathbf{x}) = \rho f_L(\mathbf{x}) + \delta(\mathbf{x})$. The posterior mean and posterior variance are given by

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

and

$$\sigma^2(\mathbf{x}) = \rho^2\sigma_L^2(\mathbf{x}) + \sigma_d^2(\mathbf{x}) - \tilde{\mathbf{k}}(\mathbf{x})(\tilde{\mathbf{K}} + \sigma^2\mathbf{I})^{-1}\tilde{\mathbf{k}}(\mathbf{x}),$$

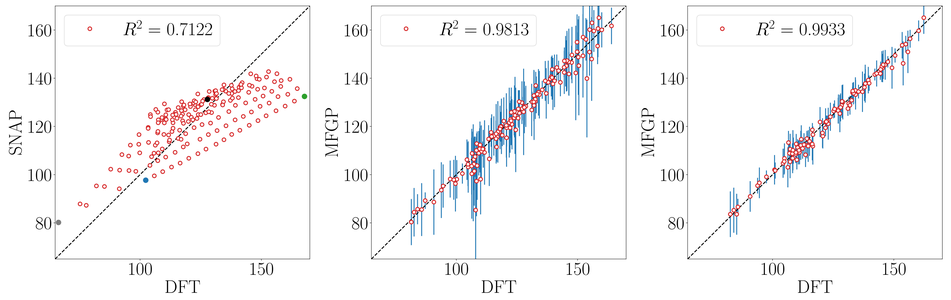
respectively, where the covariance matrix $\tilde{\mathbf{K}}$ is

$$\tilde{\mathbf{K}} = \begin{pmatrix} \sigma_L^2\mathbf{K}_L(\mathbf{x}_L, \mathbf{x}_L) & \rho\sigma_L^2\mathbf{K}_L(\mathbf{x}_L, \mathbf{x}_H) \\ \rho\sigma_L^2\mathbf{K}_L(\mathbf{x}_H, \mathbf{x}_L) & \rho^2\sigma_L^2\mathbf{K}_L(\mathbf{x}_H, \mathbf{x}_H) + \sigma_d^2\mathbf{K}_D(\mathbf{x}_H, \mathbf{x}_H) \end{pmatrix}.$$

The hyper-parameters in $\tilde{\theta} = (\theta_L, \theta_H)$ in $k_L(\cdot,\cdot)$ and $k_D(\cdot,\cdot)$ can be obtained by maximizing the log marginal likelihood as

$$\log p(\tilde{\mathbf{y}}|\mathbf{x}_{1:n_L}, \mathbf{x}_{1:n_H}, \tilde{\theta}) = -\frac{1}{2}(\tilde{\mathbf{y}} - \tilde{\mathbf{m}})^T(\tilde{\mathbf{K}}^{\tilde{\theta}} + \sigma^2\mathbf{I})^{-1}(\tilde{\mathbf{y}} - \tilde{\mathbf{m}}) - \frac{1}{2}\log|\tilde{\mathbf{K}}^{\tilde{\theta}} + \sigma^2\mathbf{I}| - \frac{n_H + n_L}{2}\log(2\pi).$$

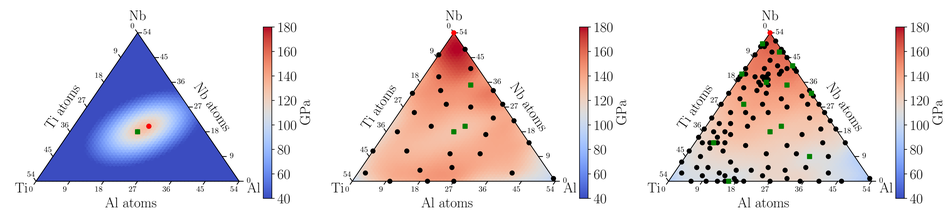
Multi-Fidelity Gaussian Process predictions



(a) DFT vs. SNAP (b) DFT vs. 57 LF+HF (c) DFT vs. 95 LF+HF

We compare the accuracy between MFGP predictions and DFT predictions. $R^2(19LF + HF) = 0.72$, $R^2(38LF + HF) = 0.72$, $R^2(57LF + HF) = 0.98$.

Multi-Fidelity Bayesian Optimization



(a) Iteration 4: 2 LF + 2 HF. (b) Iteration 35: 31 LF + 4 HF. (c) Iteration 130: 116 LF + 14 HF.

We use upper-confidence bound (UCB) $a_{\text{UCB}}(\mathbf{x}; \{\mathbf{x}_i, y_i\}_{i=1}^n, \theta) = \mu(\mathbf{x}; \{\mathbf{x}_i, y_i\}_{i=1}^n, \theta) + \kappa\sigma(\mathbf{x}; \{\mathbf{x}_i, y_i\}_{i=1}^n, \theta)$, where κ is tuned adaptively. Level t^* of fidelity is chosen as [TWM20] $t^* = \text{argmin}_t\left(C_t \int_{\mathcal{X}} \sigma^2(\mathbf{x})d\mathbf{x}\right)$, where C_t is the computational cost at level t. The optimal chemical composition for bulk modulus is found and verified at high-fidelity level after 31 LF + 4 HF iterations.

Speedup

- high-fidelity DFT: $\sim 1.44 \cdot 10^2$ core-hour. Reference.
- low-fidelity SNAP MD: $6.4 \cdot 10^{-1}$ core-hour. 2.25E2× faster.
- MFGP: $1.03 \cdot 10^{-3}$ core-hour. 1.4E5× faster.

Discussion/Conclusion

- Multi-fidelity is the right tool for multi-scale problems.
- Accuracy-speed trade-off is well interpreted in terms of fidelity:
  - high speed, low accuracy: low-fidelity → MD with ML-IAP: LAMMPS with SNAP
  - low speed, high accuracy: high-fidelity → DFT: Quantum ESPRESSO
- MFGP/MFBO is very accurate and efficient, with UQ naturally enabled.
- Extensible to high-entropy alloys.

References

[DEG+20] Keith Dalbey et al.
DAKOTA A Multilevel Parallel Object-Oriented Framework for Design Optimization Parameter Estimation Uncertainty Quantification and Sensitivity Analysis.
Technical report, Sandia National Lab, 2020.

[GAB+17] Paolo Giannozzi et al.
Advanced capabilities for materials modelling with Quantum ESPRESSO.
Journal of Physics: Condensed Matter, 29(46):465901, 2017.

[TST+15] Aidan P Thompson et al.
Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials.
Journal of Computational Physics, 285:316–330, 2015.

[TTWT20] Anh Tran, Julien Tranchida, Tim Wildey, and Aidan P Thompson.
Multi-fidelity machine-learning with uncertainty quantification and Bayesian optimization for materials design: Application to ternary random alloys.
The Journal of Chemical Physics, 153:074705, 2020.

[TWM20] Anh Tran et al.
sMF-BO-2CoGP: A sequential multi-fidelity constrained Bayesian optimization for design applications.
Journal of Computing and Information Science in Engineering, 20(3):1–15, 2020.

Acknowledgment

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. The research was partially supported by the U.S. Department of Energy, Office of Science, Early Career Research Program, under award 17020246.

anhtran@sandia.gov