Accurate global machine learning force fields for molecules with hundreds of atoms

IPAM Workshop: EMS2023
Force field reconstruction with ML
A unique set of challenges

- Learning on sets (of indistinguishable atoms)
- Invariant / equivariant outputs
- Stringent speed / accuracy requirements

Classical force fields

ML

*Ab initio* approximations to Schrödinger’s equation
Atomistic modeling with ML

Historical perspective

Feature engineering
- descriptor + linear models

Graph neural networks
- representation learning

Architecture engineering
- sophisticated filter parameterizations (e.g., YLM)
- “geometric deep learning”

2010
- ACSF
- GAP
- Rupp et al.
- SOAP
- Hirn et al.

2016
- Gilmer et al.
- SchNet
- DTNN

2018
- PaiNN
- 3D Steerable CNNs
- Tensor Field Networks
- SpookyNet
- Cormorant

Numerical analysis
- exploit model structure in training (e.g., correlations, PDE constraints)
- remove uncontrolled approximations

Behler 2011, Bartók et al. 2013, Hirn et al. 2013, Cohen-Welling 2016, Gilmer et al. 2017, Duvenaud et al. 2015, Schütt et al. 2017, Thomas-Smidt et al. 2018, Weiler et al. 2018, Anderson et al. 2019, Unke et al. 2021, Frank et al. 2021
All atoms interact in quantum many-body systems.

Atomistic modeling with ML
Efficient modeling of global force fields
Global atomic interactions
An emerging field without universal solution

TensorMol
Hybrid nearsighted NN potential with screened long-range electrostatic and van der Waals physics
Yao, K., Herr, J. E., Toth, D. W., Mckintyre, R. & Parkhill, J. The TensorMol-0.1 model chemistry: a neural network augmented with long-range physics. Chem. Sci. 9, 2261 (2018).

IPML
Physics-based mechanistic descriptions combined with environment-dependent ML corrections
Bereau, T., DiStasio Jr, R. A., Tkatchenko, A. & Von Lilienfeld, O. A. Non-covalent interactions across organic and biological subsets of chemical space: physics-based potentials parametrized from machine learning. J. Chem. Phys. 148, 241706 (2018).

Long-distance equivariant (LODE) representation + SOAP
Atom-density potential folds global structural and compositional information information into a local representation
Grisafi, A. & Ceriotti, M. Incorporating long-range physics in atomic-scale machine learning. J. Chem. Phys. 151, 204105 (2019).

Fourth-generation Behler-Parinello neural network (4G-BPNN)
ML-FF with non-local charge transfer correction using independent ML model
Ko, T. W., Finkler, J. A., Goedecker, S., & Behler, J. (2021). A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. Nat. Commun., 12(1), 1-11.

SpookyNet
ML force field augmented with physically motivated corrections for long-ranged electrostatic and dispersion interactions
Unke, O. T., Chmiela, S., Gastegger, M., Schütt, K. T., Sauceda, H. E., & Müller, K. R. (2021). SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects. Nat. Commun., 12(1), 1-14.
Structure and properties in MLFFs
Physical symmetries, data correlations, modeling choices

Symmetries / conservation laws

Physical

Frank et al. 2021, Schmitz et al. 2022, Chmiela et al. 2023, Blücher et al. 2023
Structure and properties in MLFFs

Physical symmetries, data correlations, modeling choices

Physically feasible subspace is much smaller.

e.g. Noether’s theorem, geometric priors etc.

Frank et al. 2021, Schmitz et al. 2022, Chmiela et al. 2023, Blücher et al. 2023
Structure and properties in MLFFs
Physical symmetries, data correlations, modeling choices

Symmetries / conservation laws

Correlation patterns in the reference data

Data driven interactions

Efficient training of constrained models

Physical · · · · · Statistical (ML) · · · · · Architectural (ML)

Frank et al. 2021, Schmitz et al. 2022, Chmiela et al. 2023, Blücher et al. 2023
Structure and properties in MLFFs
Physical symmetries, data correlations, modeling choices

Symmetries / conservation laws
Correlation patterns in the reference data
Physical

Data driven interactions
Statistical (ML)

Efficient training of constrained models
Architectural (ML)

Frank et al. 2021, Schmitz et al. 2022, Chmiela et al. 2023, Blücher et al. 2023
Correlation patterns in the reference data
Force field reconstruction with ML
Representing global interactions

All atoms interact:
quadtratic complexity (minimum)

Kernel function

Linear model
Pairwise atom-correlations: $\hat{f}(x) = K\alpha = \sum_{i}^{M} k(x, x_i)\alpha_i$

Reference data, but no models

Chmiela et al. 2023, Blücher et al. 2023
**Force field reconstruction with ML**

Representing global interactions

**All atoms interact:**
quadratic complexity (minimum)

**Linear model**
Pairwise atom-correlations:

\[
\hat{f}(x) = K\alpha = \sum_{i}^{M} k(x, x_i)\alpha_i
\]

Chmiela et al. 2023, Blücher et al. 2023
Force field reconstruction with ML
Atoms interact non-locally

“Donor-Bridge-Acceptor”
(E)-N,N-dimethyl-4-(4-nitrostyryl)aniline

All atoms need to interact to make accurate predictions.

Chmiela et al. 2023
Force field reconstruction with ML
Atoms interact non-locally

“Donor-Bridge-Acceptor”
(E)-N,N-dimethyl-4-(4-nitrostyryl)aniline

All atoms need to interact to make accurate predictions.
Force field reconstruction with ML

Atoms interact non-locally

"Donor-Bridge-Acceptor"
(E)-N,N-dimethyl-4-(4-nitrostyryl)aniline

All atoms need to interact to make accurate predictions.

Limit interaction length

Chmiela et al. 2023
Force field reconstruction with ML
Atoms interact non-locally

“Donor-Bridge-Acceptor”
(E)-N,N-dimethyl-4-(4-nitrostyryl)aniline

All atoms need to interact to make accurate predictions.
Force field reconstruction with ML
The most immediate application of ML in QM

All atoms interact:
quadratic memory complexity (minimum)

Linear model
Pairwise atom-correlations:
\[ \hat{f}(x) = K\alpha = \sum_{i}^{M} k(x, x_i)\alpha_i \]

Chmiela et al. 2023, Bücher et al. 2023
Correlation matrices
Numerical challenges in training

Involves matrix diagonalization: \( \alpha = (XX^T + \lambda \mathbb{I})^{-1} y = (LL^T)^{-1} y \)

Alternative approach: iterative optimization using matrix-vector products

- Gradient descent: \( \nabla_w L(\hat{f}(x), y) = (XX^T + \lambda \mathbb{I}) \alpha^{t-1} - y \)
  \[ \alpha^t = \alpha^{t-1} - \gamma \left[ (XX^T + \lambda \mathbb{I}) \alpha^{t-1} - y \right] \]
  \( \gamma \): learning rate

- Krylov subspace solver (conjugate gradients):
  Dynamic learning rate: \( \gamma_t = \frac{p_t^T y}{p_t^T K \lambda p_t} \)
  Conjugate optimization steps: \( p_t = r_t + \sum_{i < t} \frac{p_i^T K \lambda r_t}{p_i^T K \lambda p_i} p_i \)
  Rank-1 decomposition of \( K = \sum_{i} p_i p_i^T \) (N steps with perfect arithmetic!)

Complexity:
- Memory: \( \mathcal{O}(N^2) \)
- Time: \( \mathcal{O}(N^3) \)

\( X \): data points
\( K = XX^T \): symmetric, PSD

Cholesky decomposition
Correlation matrices

Strongly correlated reference geometries

Chmiela et al. 2023, Bücher et al. 2023
Correlation matrices

Strongly correlated reference geometries

Preconditioning captures dominant spectral components

Spectrum

Dominant eigenvectors

Chmiela et al. 2023, Bücher et al. 2023
Effective preconditioners for correlation matrices

Strongly correlated reference geometries

General idea: \( \mathbf{K} = \sum_{t=1}^{k} \mathbf{u}_t \mathbf{u}_t^\top + \sum_{t=k+1}^{n} \mathbf{p}_t \mathbf{p}_t^\top \)

Precond. \( \mathbf{K} = \mathbf{P}^{-1} \mathbf{K} \)

CG decomp. of \( \mathbf{K} \)

Computational cost / effectiveness trade-off: \( \mathbf{K} \approx \hat{\mathbf{K}} + \mathbb{I} \Lambda \)

- Nyström approximation
  \( \hat{\mathbf{K}} = \mathbf{K}_{mk} \mathbf{K}_{kk}^{-1} \mathbf{K}_{mk}^\top \)

- Incomplete Cholesky decomposition
  \( \hat{\mathbf{K}} = \mathbf{L}_{mk} \mathbf{L}_{mk}^\top \)

- Eigen-decomposition
  \( \hat{\mathbf{K}} = \mathbf{Q}_{mk} \Lambda_{kk} \mathbf{Q}_{mk}^\top \)

Chmiela et al. 2023, Blücher et al. 2023
Effective preconditioners for correlation matrices

Preconditioning methods (MD17 Aspirin)

Preconditioner dim.: $k = 50$, $k = 126$, $k = 313$

Eigenvalues

Construction of $\tilde{K}$: SVD, Incomplete Cholesky

Nyström approximation: Random sampling, Leverage score sampling

Chmiela et al. 2023, Blücher et al. 2023
Cost

Strongly correlated reference geometries

General idea: \( K = \sum_{t=1}^{k} u_t u_t^T + \sum_{t=k+1}^{n} p_t p_t^T \)

Precond.
\( \tilde{K} = P^{-1}K \)
CG decomp.

Computational cost / effectiveness trade-off: \( K \approx \tilde{K} + \mathbb{I} \lambda \)

- Nyström approximation
  \( \tilde{K} = K_{mk} K^{-1}_{kk} K_{mk}^T \)
- Incomplete Cholesky decomposition
  \( \tilde{K} = L_{mk} L_{mk}^T \)
- Eigen-decomposition
  \( \tilde{K} = Q_{mk} \Lambda_{kk} Q_{mk}^T \)

Woodbury identity:
\[
P^{-1} = \lambda^{-1} \left[ \mathbb{I} - K_{mk} \left( \lambda K_{kk} + K_{mk}^T K_{mk} \right)^{-1} K_{mk}^T \right]
\]

Stable decomposition into Cholesky factors,
indirectly using \( K_{mk}^T K_{mk} = R^T Q^T QR = QQ^T \)

Chmiela et al. 2023, Blücher et al. 2023
Global force fields with hundreds of atoms
A new frontier for machine learning

Examples: Conjugated system with delocalized electrons. Long timescale (PI)MD of supramolecular complexes

Chmiela, S., Vassilev-Galindo, V., Unke, O. T., Kabylda, A., Sauceda, H. E., Tkatchenko, A., & Müller, K. R. (2023). Accurate global machine learning force fields for molecules with hundreds of atoms. Science Advances, 9(2), eadf0873.
Efficient contraction of constrained models
Differential constraints in ML models with AD

Stronger inductive biases

Differential equations: $\mathcal{L}u(x) = f(x)$, where $\mathcal{L} = \sum_{|j| \leq n} a_j \mathcal{D}^j$ is a finite linear combination of differential operators $\mathcal{D}^j$ of order $n$.

- PDE's describe causal systems with interactions
- Learn from few examples and generalize

Challenges:
- Tedious manual implementation: “one new constraint per paper”
- Constraints often increase model complexity: e.g. $\mathcal{L} = \nabla_x$

Algorithmic differentiation: Express operators in terms of JVPs $J_u(x)v$ and VJPs $J_u^T(x)w$ at $O(C_u)$!

Example: Hessian-vector products in only $O(C_u)$!

Schmitz, N. F., Müller, K. R., & Chmiela, S. (2022). Algorithmic Differentiation for Automated Modeling of Machine Learned Force Fields. The Journal of Physical Chemistry Letters, 13(43), 10183-10189.
Differential constraints in ML models with AD

Gaussian process example

Constrained GP: \( f(x) \sim \mathcal{GP}(\mathcal{L}\mu(x), \mathcal{L}_x \otimes \mathcal{L}^T_x k(x, x')) \)

Example: \( \mathcal{L} = (1, \nabla, \nabla^2) \)

Constraints: function values, gradients & Hessians

\[
\mathcal{L}_x \otimes \mathcal{L}_x^T k = \begin{bmatrix}
    k & \nabla_x k & \nabla^2_x k \\
    \nabla_x k & \nabla_x \otimes \nabla_x^T k & \nabla^2_x \otimes \nabla_x^T k \\
    \nabla^2_x k & \nabla^2_x \otimes \nabla_x^T k & \nabla^2_x \otimes \nabla^2_x^T k
\end{bmatrix}
\]

large kernel matrix (but with structure)

tensor-product structure

Schmitz et al. 2022

Efficient contraction

Full instantiation of kernel can be avoided with AD:

\[
f(x) = \sum_i \mathcal{L}_x \left[ \mathcal{L}^T_x \alpha_i k(x, x_i) \right]
\]
Differential constraints in ML models with AD

Gaussian process example

Constrained GP: $\mathbf{f}(\mathbf{x}) \sim \mathcal{GP} \left( \mathcal{L}\mu(\mathbf{x}), \mathcal{L}_\mathbf{x} \otimes \mathcal{L}^T_{\mathbf{x}} k(\mathbf{x}, \mathbf{x}') \right)$

Schmitz, N. F., Müller, K. R., & Chmiela, S. (2022). Algorithmic Differentiation for Automated Modeling of Machine Learned Force Fields. The Journal of Physical Chemistry Letters, 13(43), 10183-10189.
Differential constraints in ML models with AD

Gaussian process example

Constrained GP: \( f(x) \sim \mathcal{GP} \left( \mathcal{L} \mu(x), \mathcal{L}_x \otimes \mathcal{L}_{x'}^\top k(x, x') \right) \)

| \( \mathcal{L}_x \) | \( \mathcal{L}_{x'} \) | \( \left[ \mathcal{L}_x \otimes \mathcal{L}_{x'}^\top k(x, x') \right] \) | cost | \( \mathcal{L}_x \left[ \mathcal{L}_{x'}^\top \alpha \left[ k(x, x') \right] \right] \) | cost |
|---|---|---|---|---|---|
| \( \nabla \) | \( \nabla \) | \( \text{Jac}_x(\text{Grad}_{x'}(k))^\top \) | \( NC_k \) | \( \text{Grad}_x(\text{Grad}_{x'}(k) \alpha) \) | \( C_k \) |
| \( \nabla \) | \( \nabla^2 \) | \( \text{Jac}_x(\text{Hess}_{x'}(k))^\top \) | \( N^2 C_k \) | \( \text{Grad}_x(\text{Hess}_{x'}(k) \alpha) \) | \( NC_k \) |
| \( \nabla^2 \) | \( \nabla^2 \) | \( \text{Hess}_x(\text{Hess}_{x'}(k))^\top \) | \( N^3 C_k \) | \( \text{Hess}_x(\text{Hess}_{x'}(k) \alpha) \) | \( N^2 C_k \) |

Gradient constraints in GPs:
Same complexity class as the unconstrained model!

---

Schmitz, N. F., Müller, K. R., & Chmiela, S. (2022). Algorithmic Differentiation for Automated Modeling of Machine Learned Force Fields. The Journal of Physical Chemistry Letters, 13(43), 10183-10189.
Differential constraints in ML models with AD

Gaussian process example

\[ f(\mathbf{x}) \sim \mathcal{GP} \left( \mathcal{L} \mu(\mathbf{x}), \mathcal{L}_x \otimes \mathcal{L}_x^\top k(\mathbf{x}, \mathbf{x}') \right) \]

2D Rosenbrock function: \[ u(x_1, x_2) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2 \]

Schmitz, N. F., Müller, K. R., & Chmiela, S. (2022). Algorithmic Differentiation for Automated Modeling of Machine Learned Force Fields. The Journal of Physical Chemistry Letters, 13(43), 10183-10189.
Differential constraints in ML models with AD

Combining differential equation constraints in GP’s

Solving Laplace's equation
\[ \Delta u(x) = 0 \text{ on the unit disk} \]

- Neumann boundary condition
  \[ \nabla u(x) \cdot \mathbf{n}(x) = \cos(5\phi) \]
- \( \phi \): radial angle of \( x \)
- \( \mathbf{n}(x) \): boundary normal vector

Schmitz, N. F., Müller, K. R., & Chmiela, S. (2022). Algorithmic Differentiation for Automated Modeling of Machine Learned Force Fields. The Journal of Physical Chemistry Letters, 13(43), 10183-10189.
Differential constraints in ML models with AD
Combining differential equation constraints in GP’s

1D wave equation $\Box u(x, t) = 0$, $\Box = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}$

Dirichlet boundary conditions:
$u(0,t) = u(1,t) = 0$, $u(x,0) = x(1 - x)$

Initial Neumann boundary condition:
$\partial_t u(x,0) = 0$

Schmitz, N. F., Müller, K. R., & Chmiela, S. (2022). Algorithmic Differentiation for Automated Modeling of Machine Learned Force Fields. The Journal of Physical Chemistry Letters, 13(43), 10183-10189.
Interactions from self-attention

\[ \alpha_{ij} = \langle v_i, k_j \rangle \]
Self-attention mechanism

Basic definition

\[ q_i = W_q x_i \quad k_i = W_k x_i \quad v_i = W_v x_i \]

“queries” “keys” “values”

\( x \): atom embeddings
\( W \): learnable weights

\[ A = \text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d}} \right) V \]

\( Q, K, V \): query, key and value embeddings for each atom
\( d \): embedding dimension

Frank, Thorben et al. "GeoPaTra: An Equivariant Transformer for Atomic Interactions on Arbitrary Length-Scales", 2021
Interactions from self-attention
Neural network architecture

Frank, Thorben et al. "GeoPaTra: An Equivariant Transformer for Atomic Interactions on Arbitrary Length-Scales", 2021
Emerging non-linear couplings between atoms
Attention coefficients vs. bond distances

Frank, Thorben et al. "GeoPaTra: An Equivariant Transformer for Atomic Interactions on Arbitrary Length-Scales", 2021
Want to know more?
Articles, datasets & code

**Ab initio accelerated.**

Accurate global machine learning force fields with hundreds of atoms.

Learn more  Get started

---

**MD22 benchmark dataset**

- four major classes of biomolecules and supramolecules
- up to 370 atoms

---

**Articles**

| Method | Software | Analysis | Applications | Reviews |
|--------|----------|----------|--------------|---------|
| Machine Learning of Accurate Energy-Conserving Molecular Force Fields |

Chmiela, S., Tkatchenko, A., Sauieda, H. E., Pottovsky, I., Schütt, K. T.,
Want to know more?
Articles, datasets & code

Parametrize your own datasets and use them as a force field.

OR

Upload your dataset and let us do the training.

www.sgdml.org