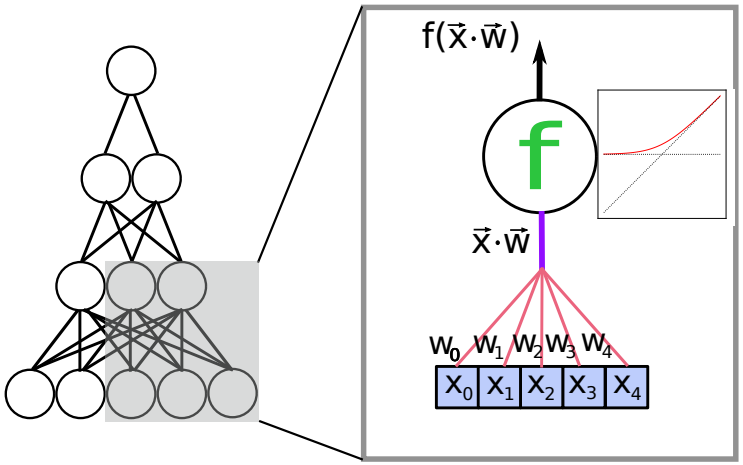
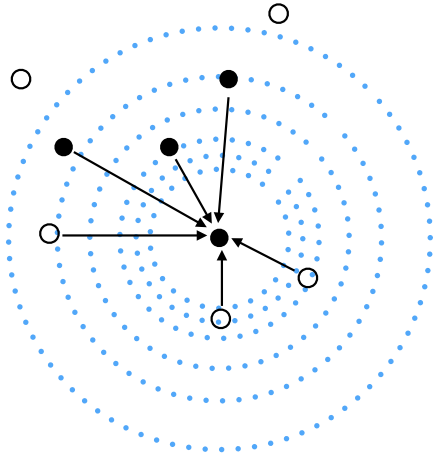
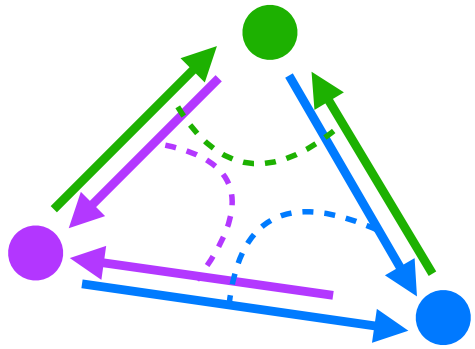
Neural Architectures for Atomistic Machine Learning

2025 Los Alamos Computational Condensed Matter Summer School







Nicholas Lubbers
Information Sciences Group
Computer, Computational, and Statistical Sciences Division
Los Alamos National Laboratory

LA-UR-25-26171

Acknowledgements

- Funding through the years supplied by LANL-LDRD, LANL-CNLS, and DOE-BES.
- Amazing and large team at LANL, cannot list all…

Self

Galen Craven

Emily Shinkle

Justin S Smith (Nvidia)

Alice Allen (Max Planck)

Yulia Pimonova

Brendan Gifford

Sergei Tretiak

Jared Averitt

Aleks Pachalieva

Ben Nebgen

Sakib Matin

Riti Bahl (Emory)

Michael Chigaev (UC Berkeley)

Roxana Bujack

Kipton Barros

Richard Messerly (ORNL)

Ying Wai Li

Outline

- Quick background on molecular dynamics, classical models, and the blessing and curse of electronic structure
- A primer on neural networks, neurons, and convolutional networks
- An overview of neurons for atomistic machine learning
- A dive into our latest work in this area

Outline

- **Quick background on molecular dynamics, classical models, and the blessing and curse of electronic structure**
- A primer on neural networks, neurons, and convolutional networks
- An overview of neurons for atomistic machine learning
- A dive into our latest work in this area

Atomistic Simulation

- Uses Born-Oppenheimer approximation, based on electron mass

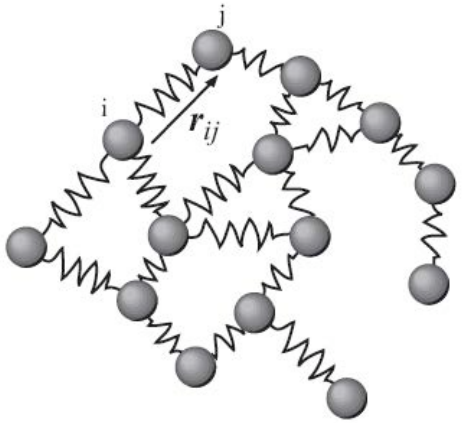
- Vast amounts of molecular dynamics are computed with Newton's laws applied to the atom positions:

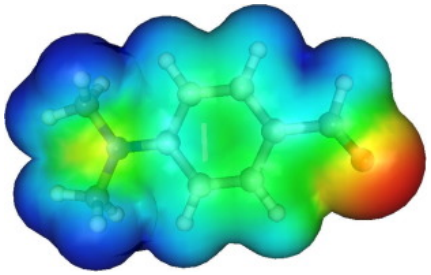
$$F = ma \qquad a = \frac{d^2x}{dt^2}$$

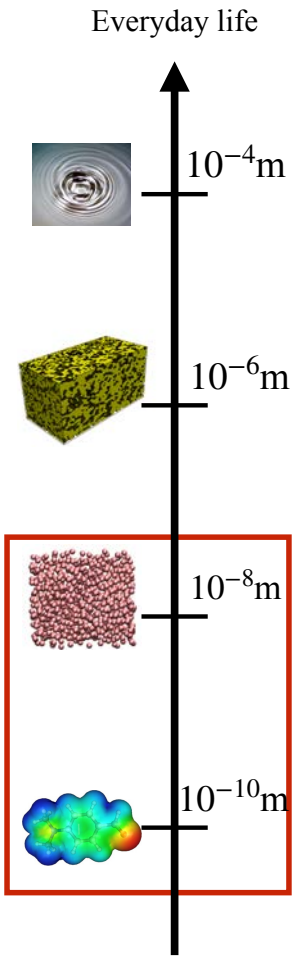
$$F = \frac{\partial E(x)}{\partial x}$$

- But the underlying energies derive from electrons, obeying the Schrödinger equation:

$$\hat{H}\psi = E\psi$$





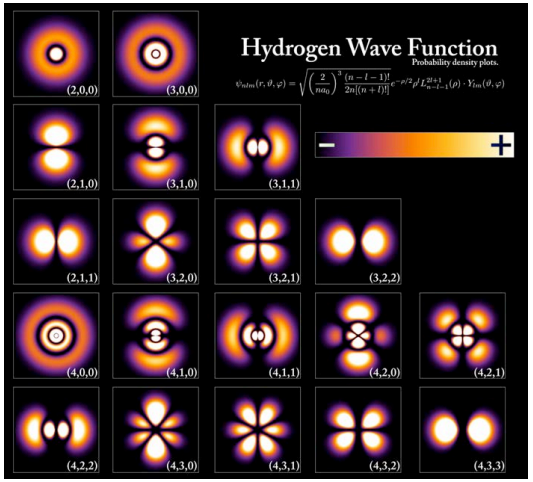


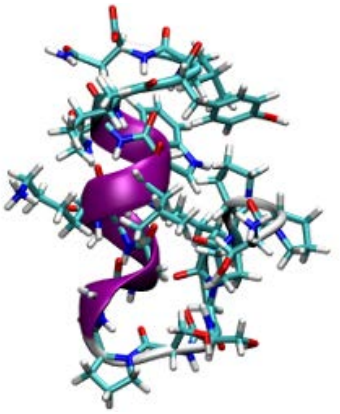
M. Rocha, A. D. Santo *et al.*, "Ab-initio and DFT calculations on molecular structure, NBO, HOMO–LUMO study and a new vibrational analysis of 4-(Dimethylamino) Benzaldehyde," *Spectrochim. Acta, Part A*, vol. 136, pp. 635 – 643, .2015.

The size of the problem

$$\hat{H}\psi = E\psi$$

- Physically, all electrons mutually interact with each other. Only very simple cases (e.g. hydrogen atom) are solvable.
- As a result, the energy is highly many-body, and most QM methods scale as N^3 in system size.
- A simulation may need $O(10^3)$ atoms to determine if a drug binds to a receptor, and $O(10^7)$ time steps.
- More complex simulations sometimes require $O(10^6)$ atoms or more.
- Can we build a neural network to model the many-body energy with an $O(n)$ scaling?

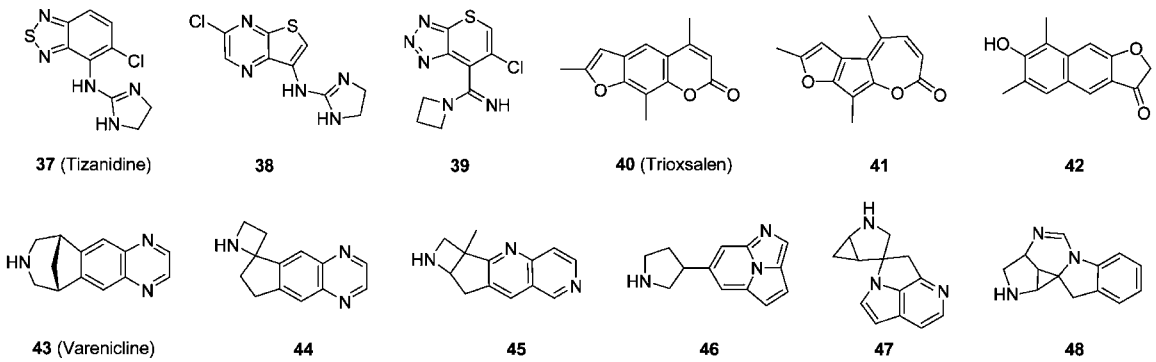




1 hour per time step?

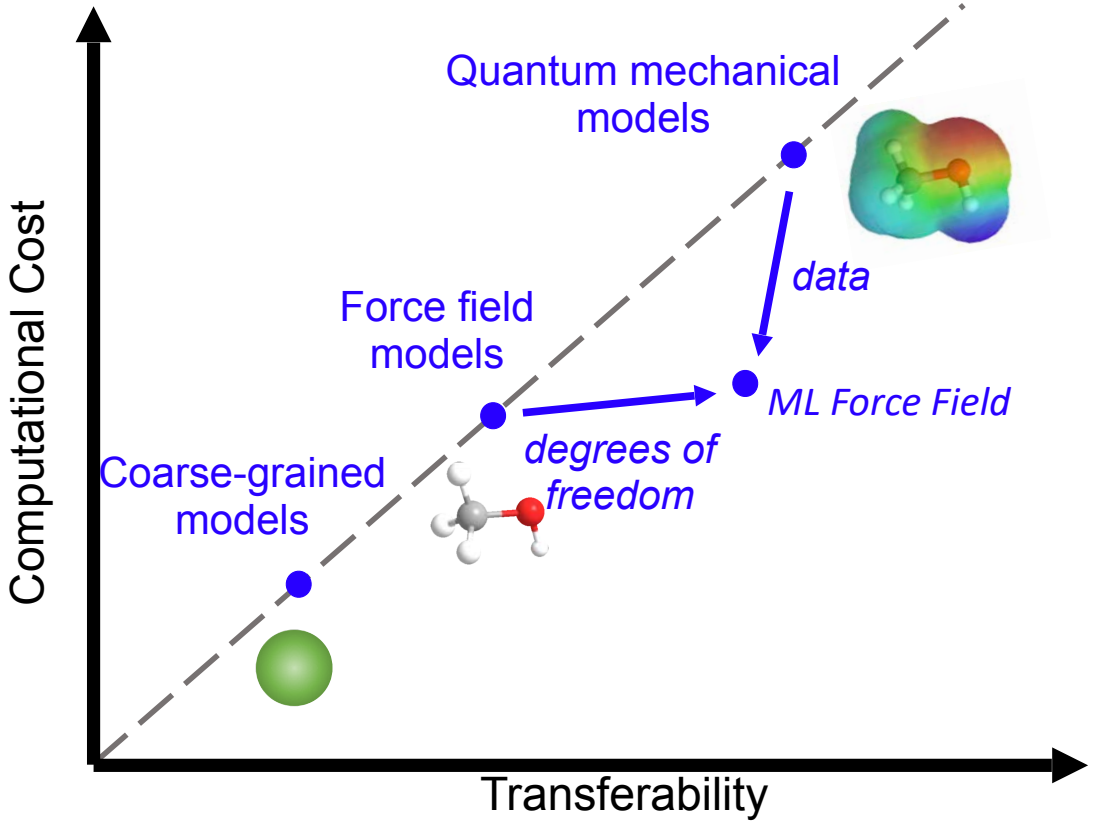
The data space

- Key fact: There are 166 billions+ *small* organic molecules.
- There are needles in this haystack!
- Condensed materials are not any easier to find



Ruddigkeit, Lars, et al. "Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17." *Journal of chemical information and modeling* 52.11 (2012): 2864-2875.

A data-driven goal



Can we build models that are:

- O(n) Linear scaling computational complexity
- Capture complex many-body effects
- Transferable across many systems

Main Goal: Predict energy

$$E(\{Z, R\})$$

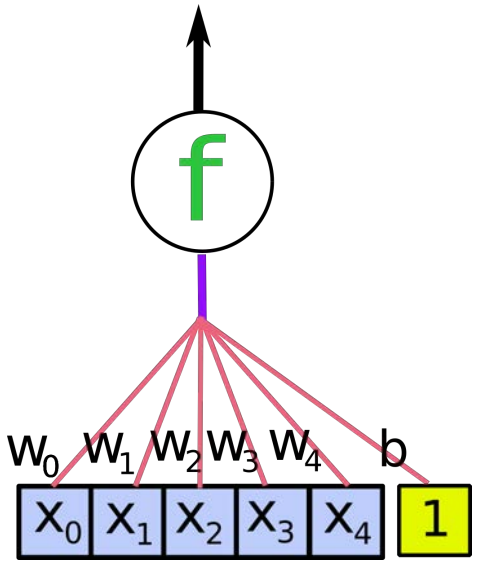
Outline

- Quick background on molecular dynamics, classical models, and the blessing and curse of electronic structure
- **A primer on neural networks, neurons, and convolutional networks**
- An overview of neurons for atomistic machine learning
- A dive into our latest work in this area

The Perceptron

$$a = f\left(\vec{w} \cdot \vec{x} + b\right)$$

- Invented in 1958 by Frank Rosenblatt.
- Learns binary output with threshold function f(z):
  - 0 if z < 0,
  - 1 if z > 0
- Threshold applied to input *features*
- Implemented with hardware and hand-tunable weights.



Controversially, *The New York Times* reported the perceptron to be "the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence."

https://news.cornell.edu/stories/2019/09/professors-perceptron-paved-way-ai-60-years-too-soon

Deep Learning

- Not very long after, in 2006, Geoffrey Hinton makes Deep Learning mainstream
- Builds many-layer networks by stacking one-layer networks that are built up one-at-a-time
- Able to classify and separate news articles of varying types based on the word-stem counts.

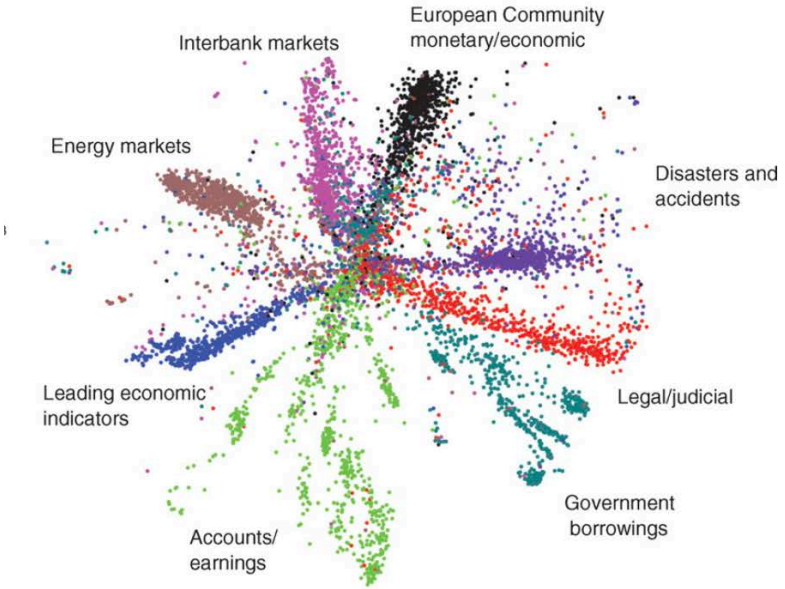
> It has been obvious since the 1980s that backpropagation through deep autoencoders would be very effective for nonlinear dimensionality reduction, provided that computers were fast enough, data sets were big enough, and the initial weights were close enough to a good solution. All three conditions are now satisfied.

Reducing the Dimensionality of Data with Neural Networks

G. E. HINTON AND R. R. SALAKHUTDINOV Authors Info & Affiliations

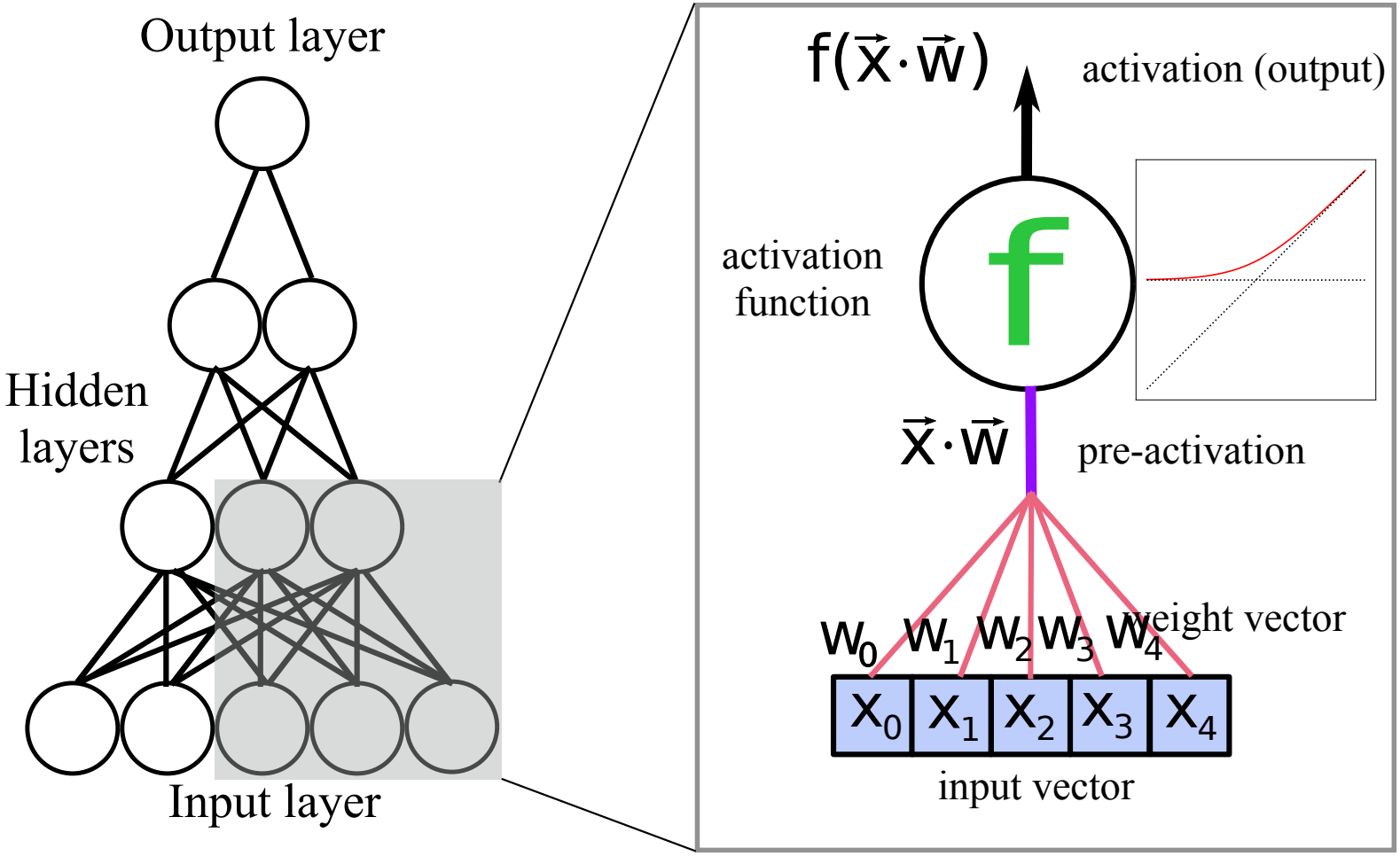
SCIENCE · 28 Jul 2006 · Vol 313, Issue 5786 · pp. 504-507

DOI: 10.1126/science.1127647



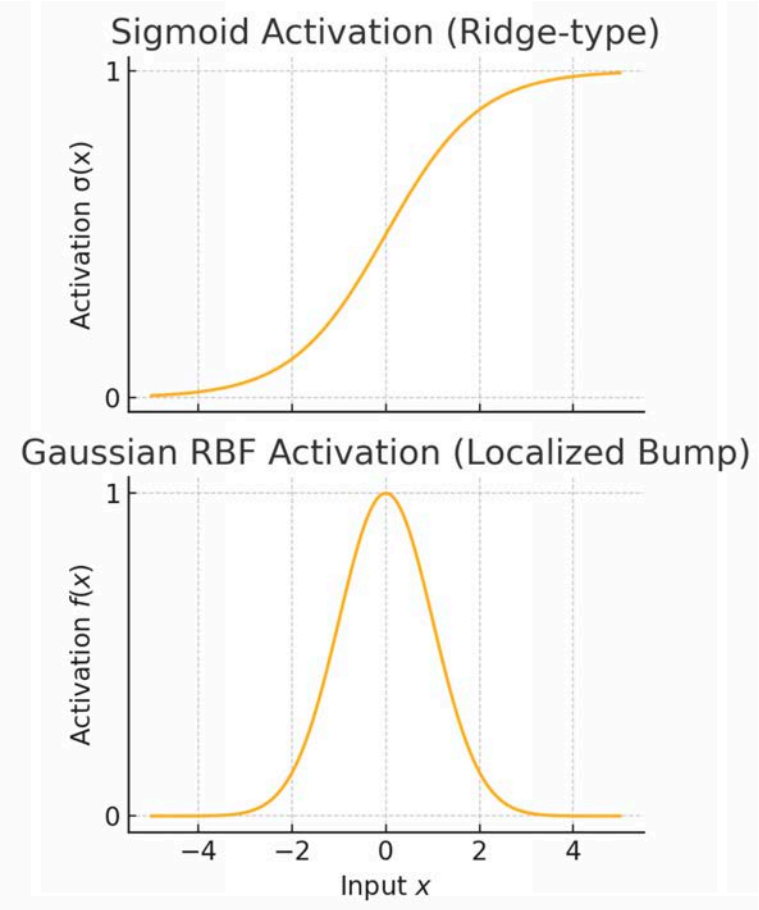
Deep Neural Networks

(in a nutshell)



Universal Approximation

- In 1989, it was proven (George Cybenko) that if you have one hidden layer and an unbounded width, you can approximate any function y=f(x)
- Based on sigmoid neurons, or generally “ridge functions”, which can identify arbitrary half-spaces.
- Straightforward alternatives using gaussian neurons (RBF) to interpolate a function.

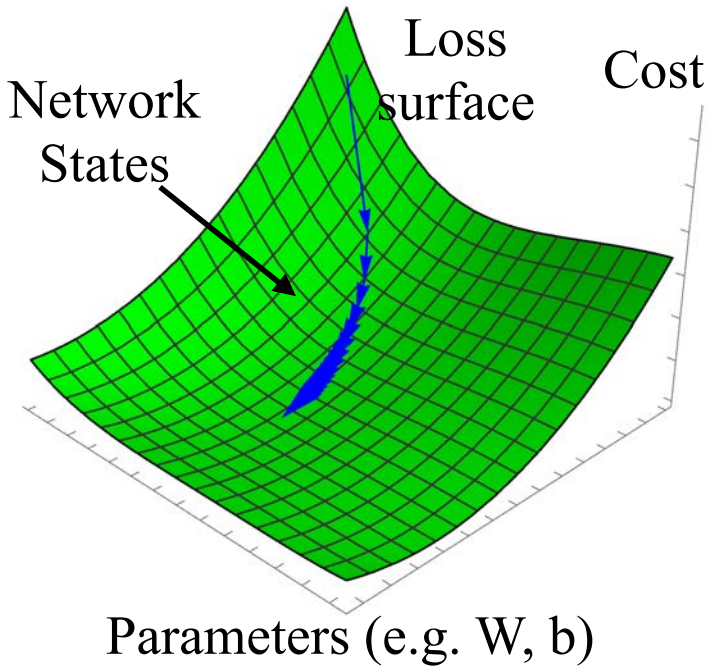


Training with Gradient Descent

- Cost function that measures error in a differentiable way — lower is better.
- Train network by finding minimum of the cost function
- Update with a *learning rate*

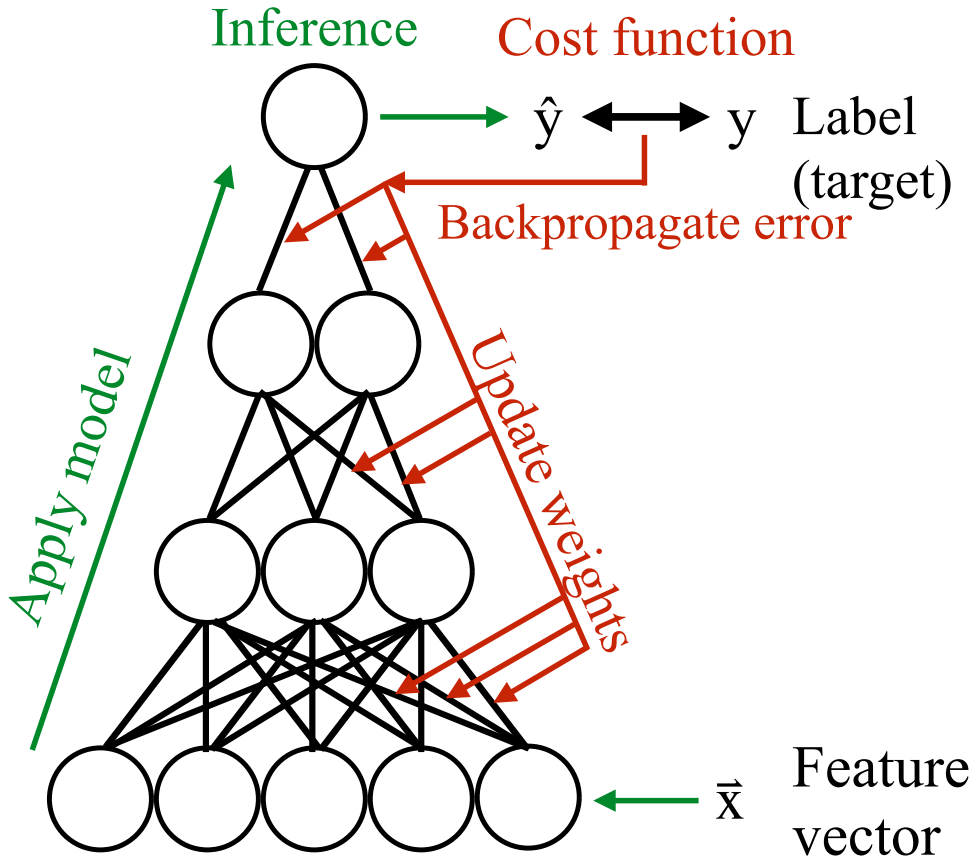
$$\theta_{n+1} = \theta_n - \alpha \nabla_\theta \mathscr{L}$$

$$\nabla_\theta \mathscr{L} = \frac{\partial \mathscr{L}}{\partial \theta}$$



Backpropagation

- Popularized by Hinton (1986)
- Allows us to compute derivative of the cost with respect to the parameters
- Computing derivative by saving key intermediate variables, and then working backwards with the chain rule
- Speed of gradient is same complexity class as speed of forward pass



2012: Deep Convolutional Neural Networks

- Krizhevsky et al. breakthrough in image classification competition on *ImageNet*

  1. Custom GPU programming for extreme computational throughput

  2. End-to-end training rather than feature creation + learning

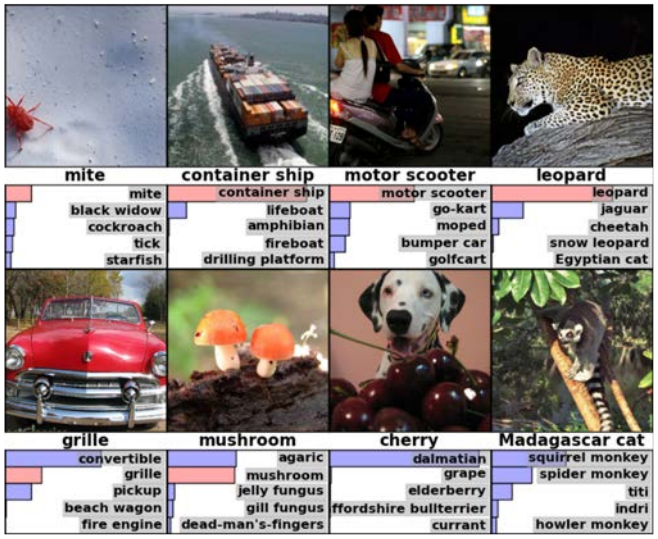
  3. Convolutional network architecture that captures translational symmetry using co-moving approach, learning image processing filters.

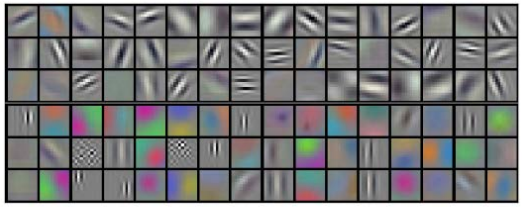
ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

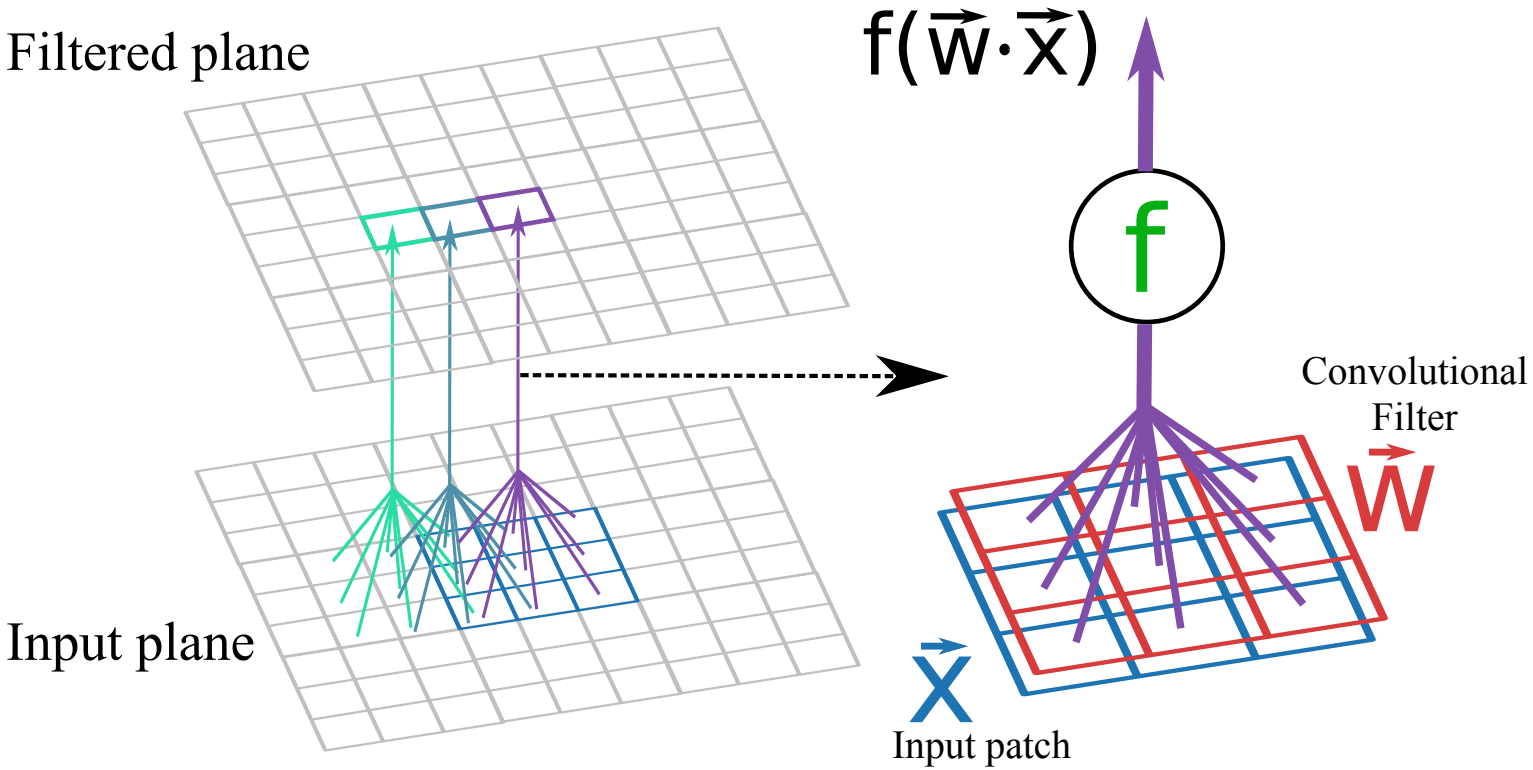
Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca





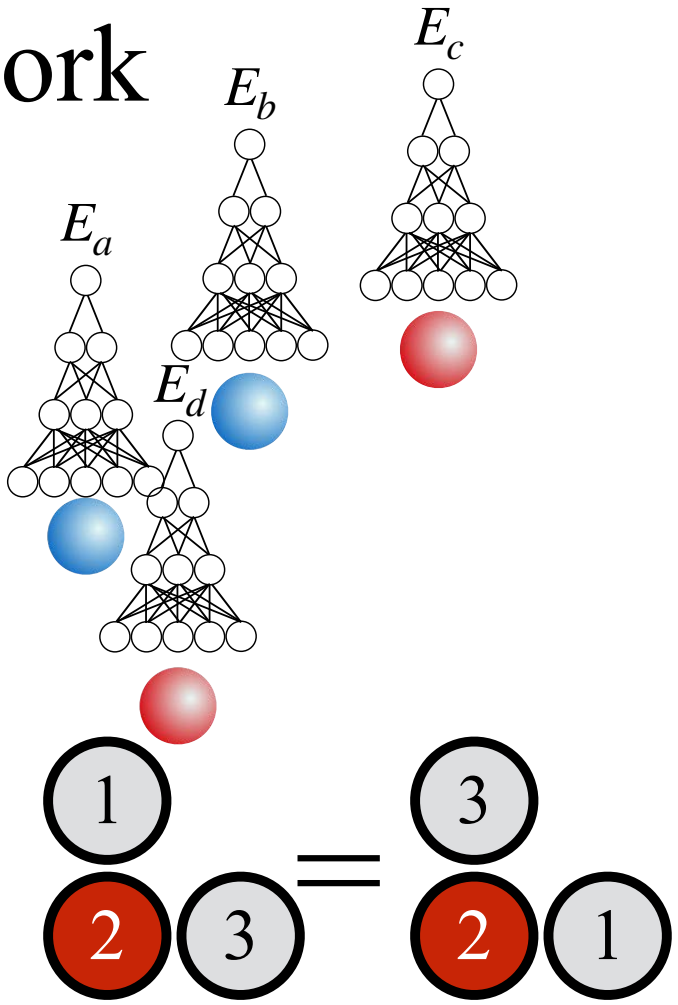
Convolutional Neural Networks: Image Processing



Convolutional neurons apply the same weights and activation to each local patch in an image, reducing the number of parameters and allowing pattern detection regardless of location.

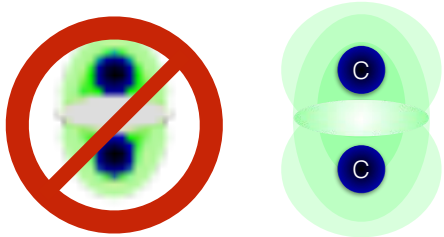
What we need — an atomistic neural network

- Associate a sub-network for each atom; each atom has its own activations
- Atom ⟷ Pixel
- Each sub-net has the same structure (weights and biases), covers permutation symmetry
- Require some analogy to convolutional networks to capture *local atomic environments*



Physics-informed Considerations

- Extensibility & locality
- Smoothness
- Permutation/Rotation/Translation Invariance



Translation Covariance and Invariance

- Convolution corresponds to cross-correlation with kernel
- Covariant (moves with) shifts of the input domain
- Invariance can be achieved by summation
- CNNs get these right for our problem!

$$a(x) = \int k(x - x')z(x')\mathrm{d}x'$$

$$x \to x + c$$
$$z(x) \to z(x + c)$$
$$a(x) \to a(x + c)$$

Translation Covariance

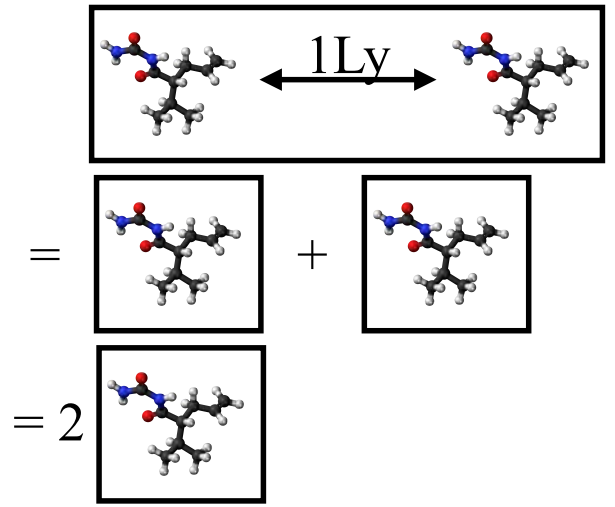
$$E = \int e(x)\mathrm{d}x$$

$$x \to x + c$$
$$E \to E$$

Translation Invariance

Extensivity and Locality

- The energy of two very separated systems should be the sum of the energy of each system.
- If energy is local, this is guaranteed
- Finite range of CNN Kernel takes care of this



Outline

- Quick background on molecular dynamics, classical models, and the blessing and curse of electronic structure
- A primer on neural networks, neurons, and convolutional networks
- **An overview of neurons for atomistic machine learning**
- A dive into our latest work in this area

1992: NNs predict coefficients for PESs

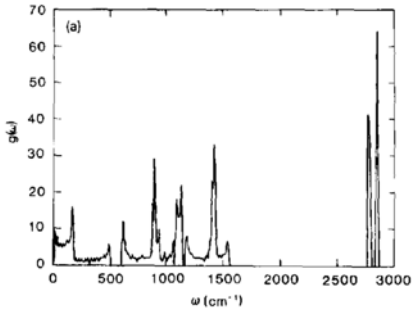
- Earliest example I know of relating NNs and Potential Energy Surfaces
- Consider polyethylene-like force field model
- Build random parameters into the force-field, which uses a "classical" form, and simulate the spectrum of vibrations for a large molecule
- Train a neural network to go from spectrum back to the force-field parameters; "*deconvolute the spectrum*"

For the present application we have determined an optimal network architecture as: a fully connected network with 426 input nodes, one hidden layer of 7 nodes, and an output layer of 18 nodes.

Potential energy surfaces for macromolecules. A neural network technique

Bobby G. Sumpter and Donald W. Noid
Chemistry Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6182, USA and Department of Chemistry, The University of Tennessee, Knoxville, TN 37996-1600, USA

Received 19 November 1991; in final form 29 January 1992



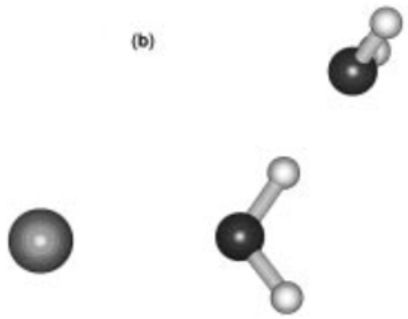
$$V(\text{bonded}) = \sum V_{CC} + \sum V_{CH} + \sum V_{CCC} + \sum V_{HCH} + \sum V_{HCC} + \sum V_{CCCC} \quad (2)$$

and

$$V(\text{nonbonded}) = \sum V_{CC} + \sum V_{CH} + \sum V_{HH} . \quad (3)$$

$$V(\theta) = \sum \tfrac{1}{2} K_{\theta ijk} (\theta_{ijk} - \theta^{e}_{ijk})^2, \quad \theta = \theta_{CCC}, \theta_{HCC}, \text{ or } \theta_{HCH} ; \quad (5)$$

$$V_{CCCC}(\tau) = \sum -\alpha' \cos(\tau_{ijkl}) + \beta' \cos^3(\tau_{ijkl}) ; \quad (6)$$



1998: NNs make PESs

Representation of Intermolecular Potential Functions by Neural Networks

Helmut Gassner,† Michael Probst,*,† Albert Lauenstein,‡ and Kersti Hermansson‡

Institute of General and Inorganic Chemistry, Innsbruck University, Innrain 52a, A-6020 Innsbruck, Austria, and Inorganic Chemistry, The Ångström Laboratory, Uppsala University, Box 538, S-75121 Uppsala, Sweden

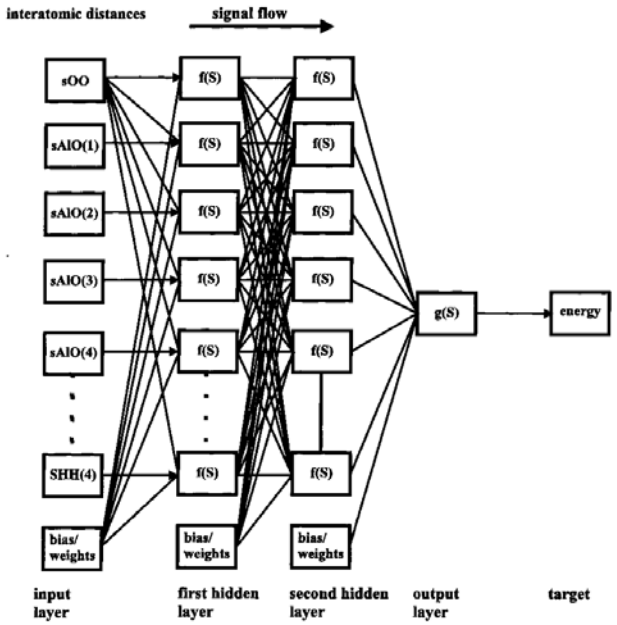
Received: July 8, 1997; In Final Form: March 17, 1998

- Identified difficulty of fitting *ab initio* with analytical formulas - Al3+ H_2O 3-body interactions
- Feed scaled atomic distances into network
- Account for permutation symmetry of interactions using polynomials

> "The second reason that the interatomic distances themselves cannot be used is the requirement to ensure the correct symmetry of the interactions."

$$r^{(1)}_{\mathrm{Al-O}} = |r_{\mathrm{Al-O_1}} + r_{\mathrm{Al-O_2}}| \quad \text{and} \quad r^{(2)}_{\mathrm{Al-O}} = |r_{\mathrm{Al-O_1}} - r_{\mathrm{Al-O_2}}|$$

- Spiritual successors continue to the present day, but not the most common approach



15 features 5 Neurons 5 Neurons

2007: BP Nets: Symmetry Functions and Extensivity

PRL 98, 146401 (2007) PHYSICAL REVIEW LETTERS week ending 6 APRIL 2007

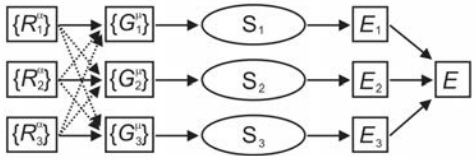
Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces

Jörg Behler and Michele Parrinello

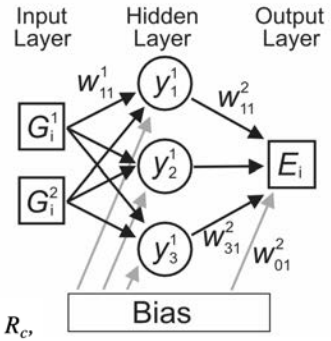
Department of Chemistry and Applied Biosciences, ETH Zurich, USI-Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland

(Received 27 September 2006; published 2 April 2007)

- Behler and Parrinello describe a model form and apply it to Bulk Silicon
- Recognize that the entire potential can be constructed with ML
- Extensive energy formulation assigning energy to each atom
- Each atom has associated *symmetry functions* (radial and angular) which describe its environment using a *cutoff function*



$$E = \sum_i E_i.$$

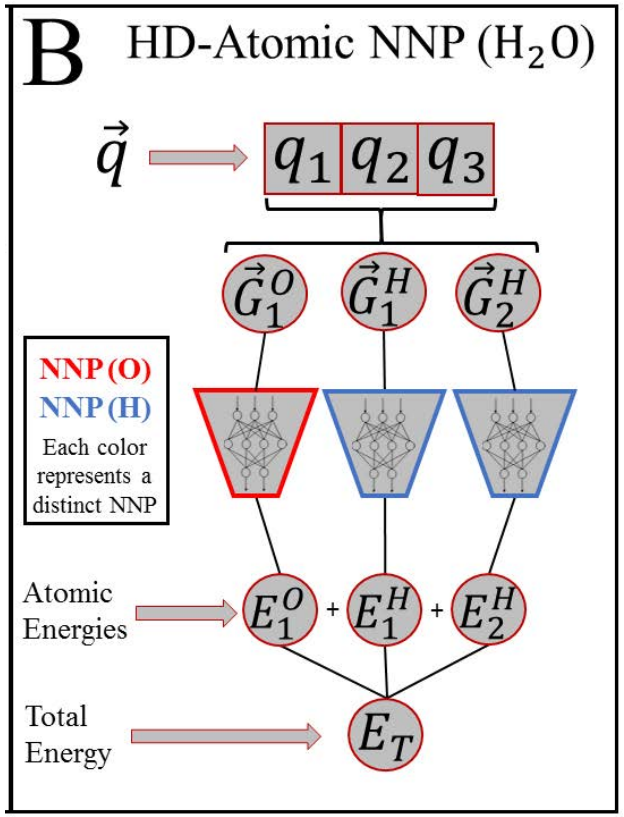


$$f_c(R_{ij}) = \begin{cases} 0.5 \times \left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1\right] & \text{for } R_{ij} \le R_c, \\ 0 & \text{for } R_{ij} > R_c. \end{cases}$$

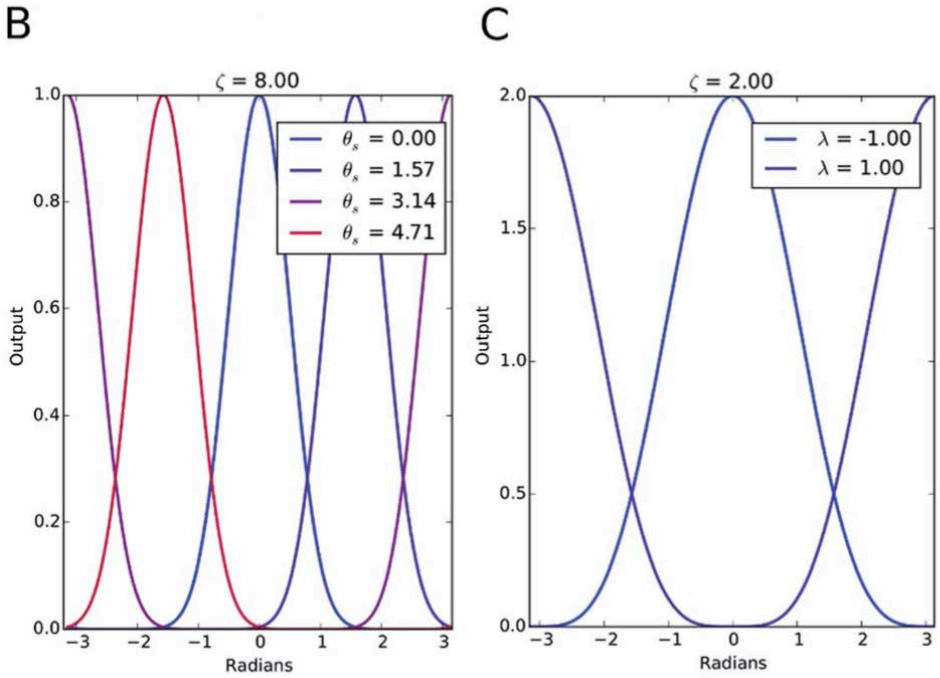
$$G_i^1 = \sum_{j \neq i}^{\text{all}} e^{-\eta(R_{ij} - R_s)^2} f_c(R_{ij}).$$

$$G_i^2 = 2^{1-\zeta} \sum_{j,k \neq i}^{\text{all}} (1 + \lambda \cos\theta_{ijk})^{\zeta} \times e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}),$$

Atomic Environment Vectors



Smith et al 2017



Angular *Symmetry Functions*
(Probe functions; basis sets)

2012: Deep Convolutional Neural Networks

- Krizhevsky et al. breakthrough in image classification competition on *ImageNet*

  1. Custom GPU programming for extreme computational throughput

  2. End-to-end training rather than feature creation + learning

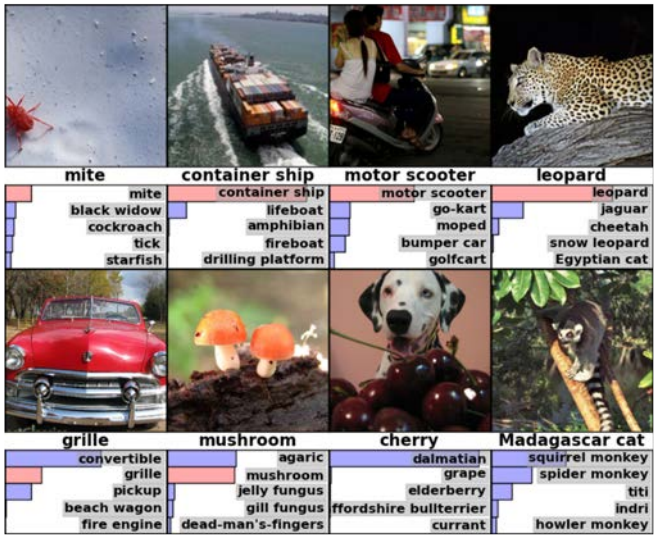
  3. Convolutional network architecture that captures translational symmetry using co-moving approach, learning image processing filters.

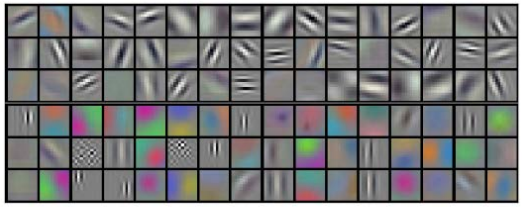
ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca



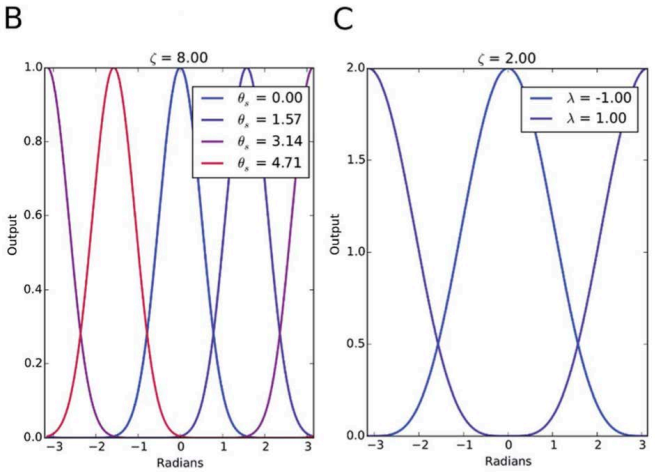


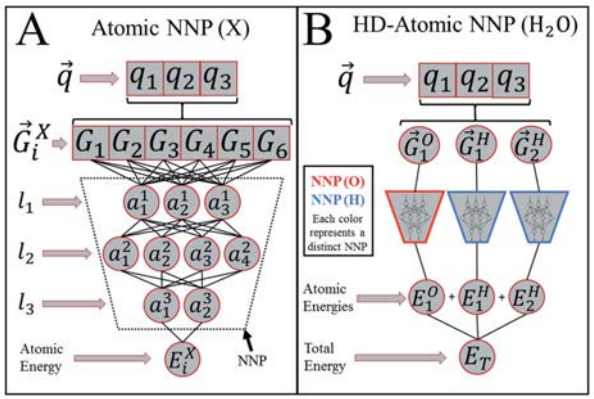
2017: ANI-1, a universal potential

- Krizhevsky success attributed to ultra-fast custom GPU implementation
- Take Behler-Parrinello framework, modify some parts, and build an ultra-fast custom GPU implementation
- Trained to 20M DFT Calculations: 200x more than anyone else at the time was even trying
- Targeted not only PES, but also chemical degrees of freedom
- Able to be applied to larger chemicals and get reasonable behavior - better than AM1, PM6, and DFTB

ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost†

J. S. Smith,[a] O. Isayev*[b] and A. E. Roitberg*[a]



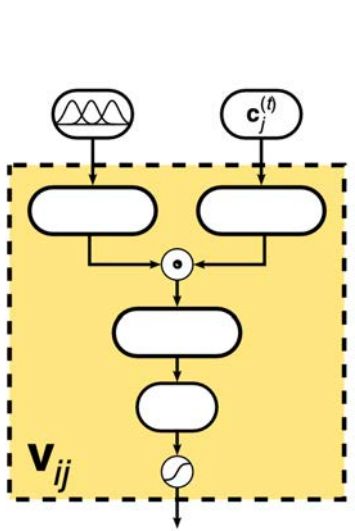


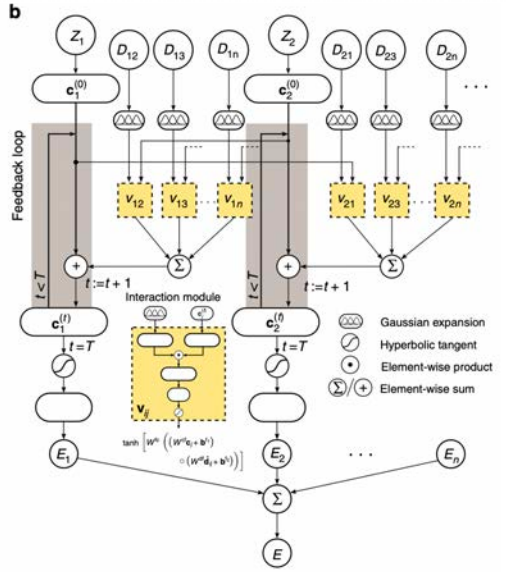
2017: Deep Tensor Neural Networks

- Krizhevsky success attributed to end-to-end learning scheme, so stack layers of updates to atomic variables.
- Distance represented with g*aussian distance embedding* (similar to Behler radial symmetry functions)
- Related to MPNNs (Message Passing Neural Networks)
- Used tensor factorization to reduce number of possible parameters

Quantum-chemical insights from deep tensor neural networks

Kristof T. Schütt[1], Farhad Arbabzadah[1], Stefan Chmiela[1], Klaus R. Müller[1,2] & Alexandre Tkatchenko[3,4]





Full Dependence $$v_{ijk} = \tanh\left(\mathbf{c}_j^{(t)} V_k \hat{\mathbf{d}}_{ij} + \left(W^{c}\mathbf{c}_j^{(t)}\right)_k + \left(W^{d}\hat{\mathbf{d}}_{ij}\right)_k + b_k\right),$$

Low-rank Factorized $$\mathbf{v}_{ij} = \tanh\left[W^{fc}\left(\left(W^{cf}\mathbf{c}_j + \mathbf{b}^{f_1}\right) \circ \left(W^{df}\hat{\mathbf{d}}_{ij} + \mathbf{b}^{f_2}\right)\right)\right],$$

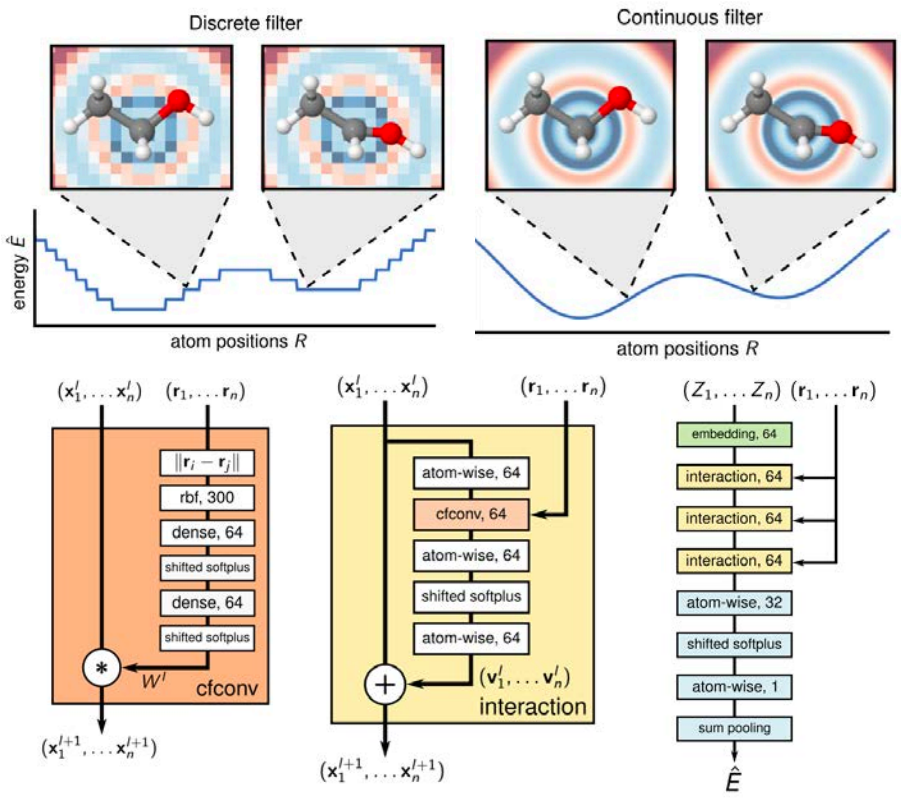
2017: SchNet - Continuous Filter Convolutions

- Krizhevsky success attributed to convolutional filters
- Analagous to LeNet, the first convolutional image model, construct filters as a function of distance between atoms
- Each filter acts independently on a feature channel, thus necessitating atom-wise layers to 'rotate' features between interactions.

$$\mathbf{x}_i^{l+1} = (X^l * W^l)_i = \sum_j \mathbf{x}_j^l \circ W^l(\mathbf{r}_i - \mathbf{r}_j),$$

SchNet: A continuous-filter convolutional neural network for modeling quantum interactions

K. T. Schütt[1]*, P.-J. Kindermans[1], H. E. Sauceda[2], S. Chmiela[1]
A. Tkatchenko[3], K.-R. Müller[1,4,5]†



2018: HIP-NN - Full-rank convolutions

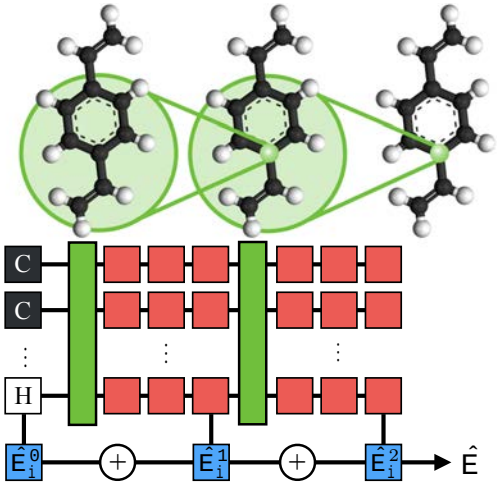
- Hierarchical Interacting Particle Neural Network
- Per-layer more expressive than SchNet (but more parameters)
- Messages do not include terms that depend only on species, nor only on distance
- Re-introduce cutoff distance to end-to-end approaches in anticipation of MD
- Add other physics-inspired architecture and training routines

RESEARCH ARTICLE | MARCH 19 2018

Hierarchical modeling of molecular energies using a deep neural network

Special Collection: Data-Enabled Theoretical Chemistry

Nicholas Lubbers; Justin S. Smith; Kipton Barros



HIP-NN Interaction:

$$\tilde{z}_{i,a}^{\ell+1} \underset{\text{inter.}}{=} f\left(\sum_{j,b} v_{ab}^{\ell}(r_{ij}) z_{j,b}^{\ell} + \sum_{b} W_{ab}^{\ell} z_{i,b}^{\ell} + B_a^{\ell}\right),$$

Full Dependence

$$v_{ijk} = \tanh\left(\mathbf{c}_j^{(t)} V_k \hat{\mathbf{d}}_{ij} + \left(W^{c}\mathbf{c}_j^{(t)}\right)_k + \left(W^{d}\hat{\mathbf{d}}_{ij}\right)_k + b_k\right),$$

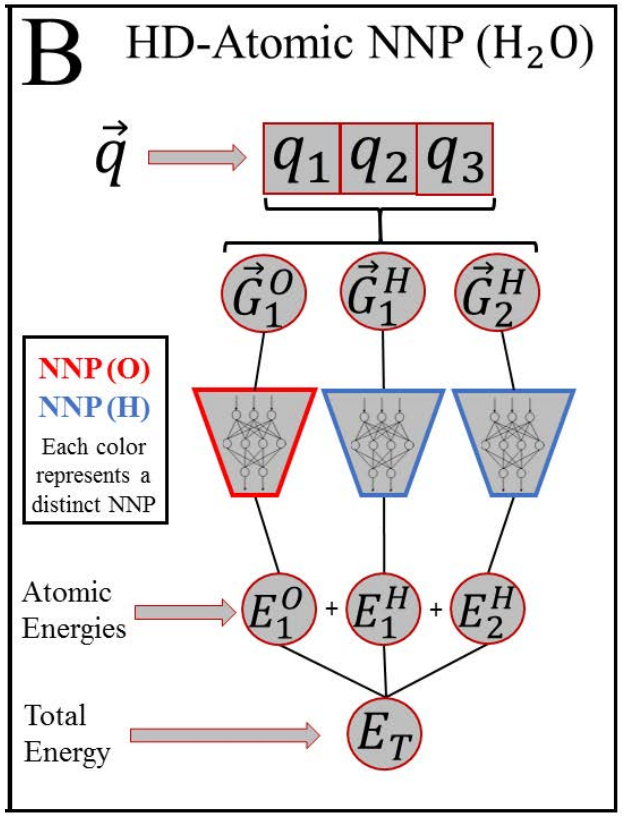
Atomic Environment Picture

$$\sum_{\nu,j,b} V_{ab}^{\nu} s^{\nu}(r_{ij}) z_{j,b}$$

$$G_b^{\nu} = \sum_{j} s^{\nu}(r_{ij}) z_{j,b}$$

$$\sum_{\nu,b} V_{ab}^{\nu} G_b^{\nu}$$

So this form of message passing is very similar to the *radial* part of BP Symmetry functions!



Smith et al 2017

Message Passing Picture

Magnitude r_{ij} is what ensures rotation and translation invariance

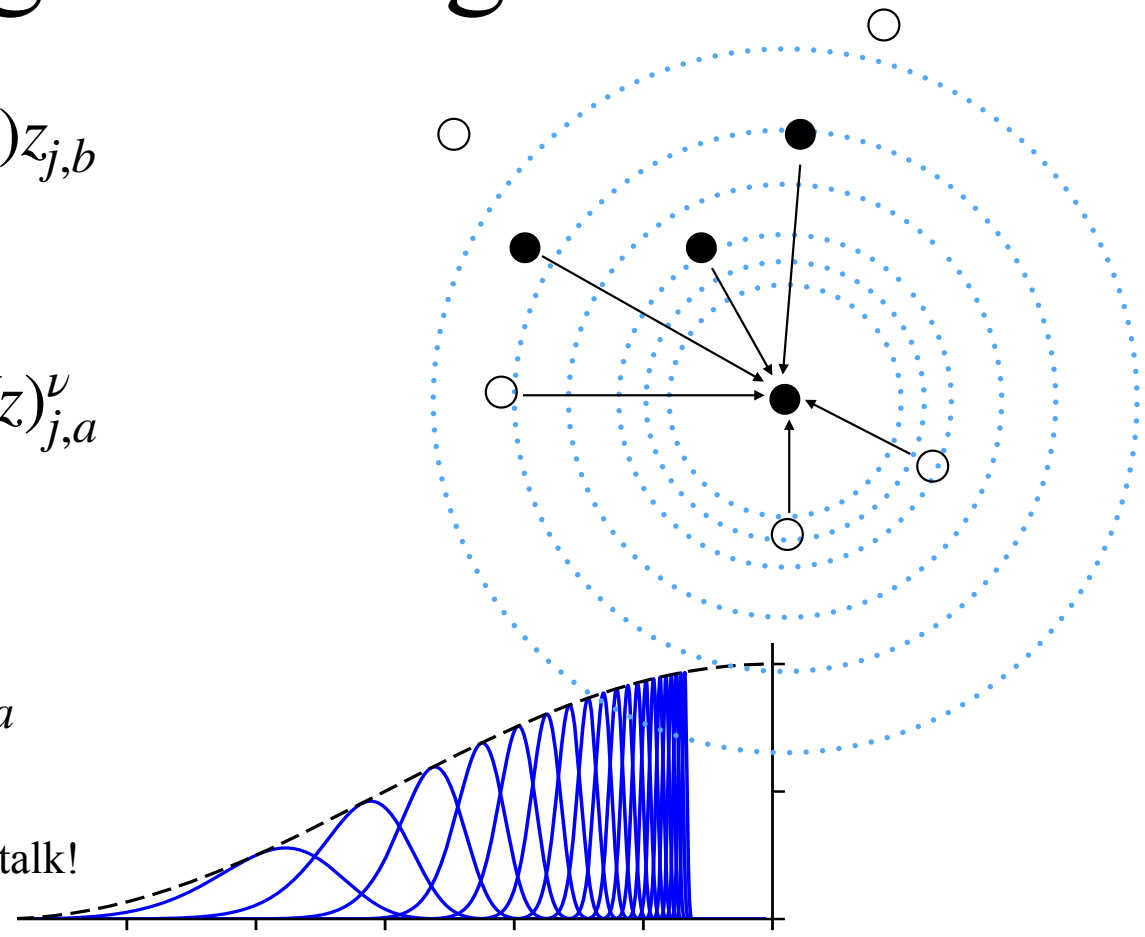
$$\sum_{\nu,j,b} V^{\nu}_{ab} s^{\nu}(r_{ij}) z_{j,b}$$

$$\sum_{\nu,j} s^{\nu}(r_{ij})(Vz)^{\nu}_{j,a}$$

"Message"

$$\sum_{j} m_{ij,a} = \sum_{\nu,j} s^{\nu}(r_{ij}) \zeta^{\nu}_{j,a}$$

This form will come back later in the talk!



Variable Weight Picture

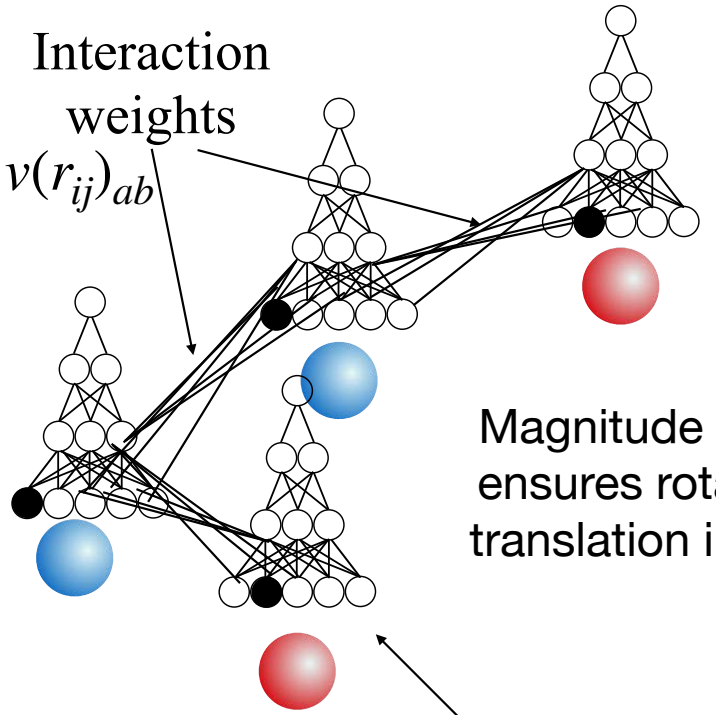
$$\sum_{\nu,j,b} V^{\nu}_{ab} s^{\nu}(r_{ij}) z_{j,b}$$

$$\sum_{j,b} (VS)(r_{ij})_{ab} z_{j,b}$$

"Interaction weights"

$$\sum_{j,b} v(r_{ij})_{ab} z_{j,b}$$

This is strictly *more* expressive than Schnet-style, which uses a variable weight which is *diagonal* in feature space (point-wise product)



Magnitude r_{ij} is what ensures rotation and translation invariance

2018: HIP-NN - a few more details

- *Sensitivity functions* (distance embedding, basis set) give more compact representation of distance than previous models, while retaining short-range flexibility
- Used energy terms from each level of interactions, and loss penalty to make them hierarchical
- Showed uncertainty quantification based on hierarchically for a prediction

Hierarchical Decomposition:

$$\hat{E}_i = \sum_{n=0}^{N_{\text{interaction}}} \hat{E}_i^n.$$

Regularization Term:

$$R = \sum_{n=1}^{N_{\text{interaction}}} \sum_{i=1}^{N_{\text{atom}}} \frac{(\hat{E}_i^n)^2}{(\hat{E}_i^n)^2 + (\hat{E}_i^{n-1})^2}.$$

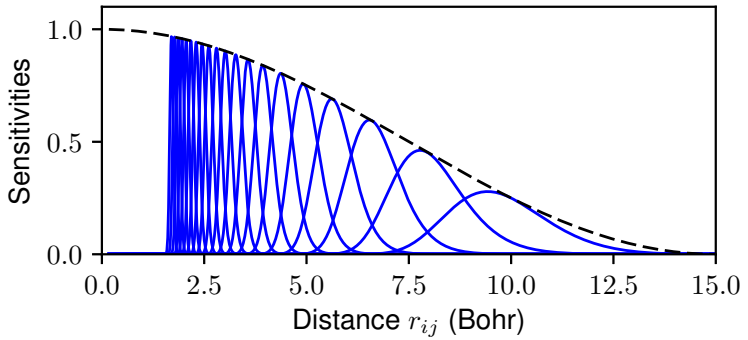
RESEARCH ARTICLE | MARCH 19 2018

Hierarchical modeling of molecular energies using a deep neural network

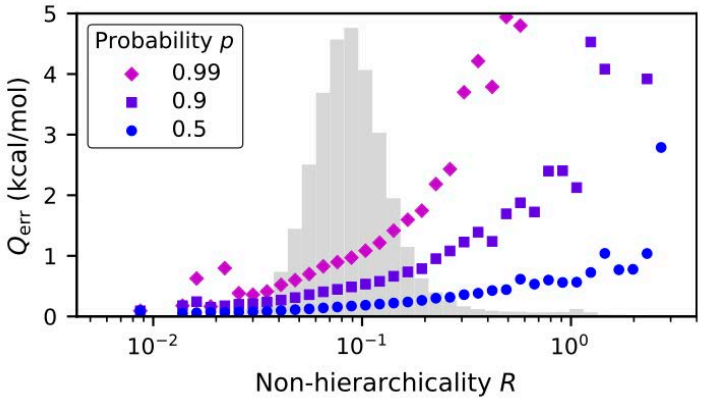
Special Collection: Data-Enabled Theoretical Chemistry

Nicholas Lubbers; Justin S. Smith; Kipton Barros

Expanded: $z'_{i,a} = f\left(\sum_{\nu,j,b} V^{\nu}_{ab} s^{\nu}(r_{ij}) z_{j,b} + \sum_{b} W_{ab} z_{i,b} + B_a\right)$



Sensitivity functions *s* using 1/r transformed Gaussians

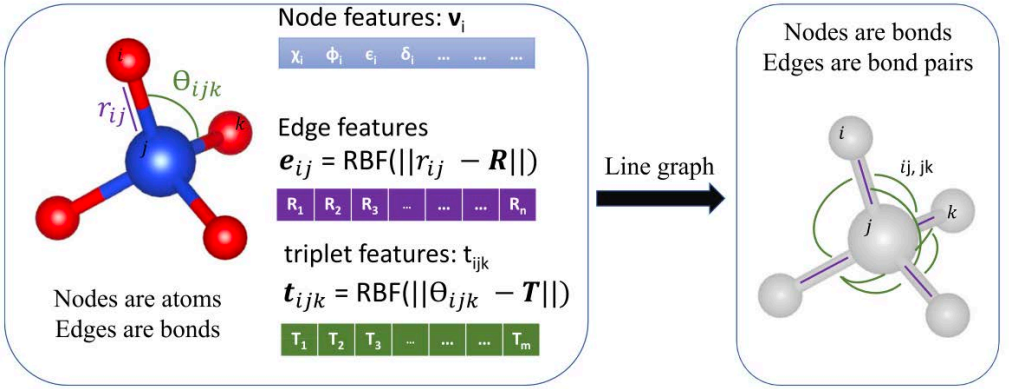


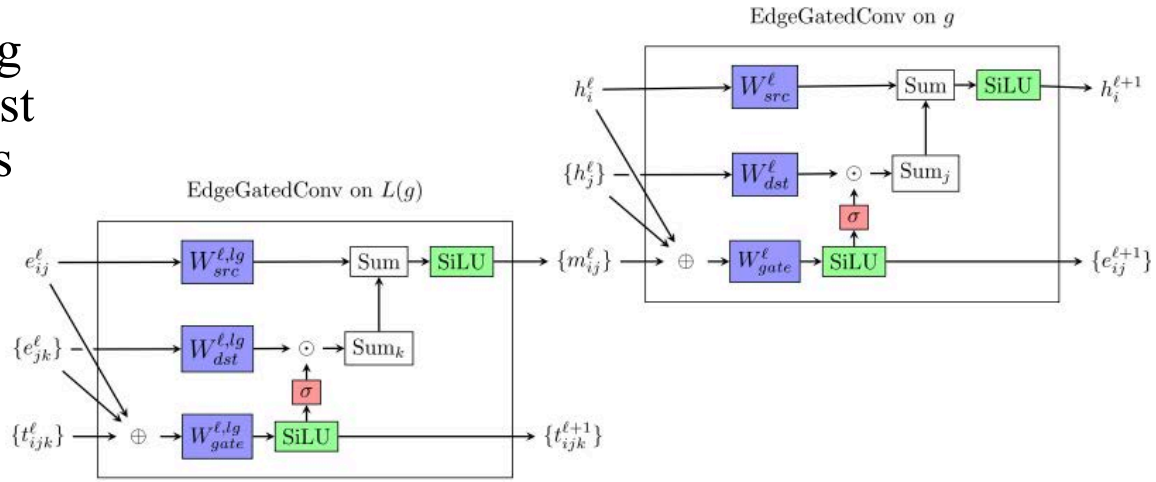
2022: ALIGNN

- Some earlier networks updated nodes and edges - ALIGNN introduces the idea of passing messages with angles as higher-order edges. (Edges for edges)
- This is called the *line graph*
- Showed very strong performance going after the *Materials Project* dataset - first major step towards universal potentials beyond organics

Atomistic Line Graph Neural Network for improved materials property predictions

Kamal Choudhary [1,2,3,4]✉ and Brian DeCost [1,4]





Is invariance the best we can do?

- Scalar approaches like HIP-NN or ALIGNN seem artificially limited when you think back to the fact that it only looks at distances and angles.
- Arbitrary dependence on input coordinates is dangerous, because at the end of the day we will want to have a completely invariant prediction. (Noether's theorem implies that we must obey the symmetry in order to get the corresponding conservation law)
- Instead of invariance, could we allow our network to look at functions in a way that changes *with* the rotational symmetry?

2018: Tensor Field Networks

- Huge limitation of prior works was limitation to scalar processing
- Augmented Schnet with *tensor product* and *equivariant neurons*
- Point out that non-invariant neurons can be fine as long as we know how to make invariants from them in the end
- Use group theory and *Clebsh-Gordan decomposition* to define layers using tensor product
- One down-side: expensive to compute all terms

Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds

Authors Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, Patrick Riley

Layer: $\mathcal{L}(\vec{r}_a, x_a) = (\vec{r}_a, y_a)$

Translation Equivariance: $\mathcal{L} \circ \mathcal{T}_{\vec{t}} = \mathcal{T}_{\vec{t}} \circ \mathcal{L}$

Rotation Equivariance:

$$\mathcal{L} \circ \left[\mathcal{R}(g) \oplus D^{\mathcal{X}}(g)\right] = \left[\mathcal{R}(g) \oplus D^{\mathcal{Y}}(g)\right] \circ \mathcal{L}$$

Tensor Product:

$$(u \otimes v)^{(l)}_m = \sum_{m_1=-l_1}^{l_1} \sum_{m_2=-l_2}^{l_2} C^{(l,m)}_{(l_1,m_1)(l_2,m_2)} u^{(l_1)}_{m_1} v^{(l_2)}_{m_2}$$

Layer definition:

$$\mathcal{L}^{(l_o)}_{acm_o}(\vec{r}_a, V^{(l_i)}_{acm_i}) := \sum_{m_f,m_i} C^{(l_o,m_o)}_{(l_f,m_f)(l_i,m_i)} \sum_{b \in S} F^{(l_f,l_i)}_{cm_f}(\vec{r}_{ab}) V^{(l_i)}_{bcm_i}$$

Equivariance and group representations

- Rotation is formally described by a Lie Group group structure.
  SO(3) → *special orthogonal group in 3 dimensions*

- Some properties are *equivariant* with rotation. When the input is rotated, the output obeys a transformation rule as well.

$$x \xrightarrow{g} x' = R(g)x$$
$$f(x) \xrightarrow{g} f(x') = U(g)f(x)$$

Coordinates x
Group element g
Rotation Matrix R
Function f
Representation matrix U

- If $U(g) = 1$, f is an invariant function or *scalar.*

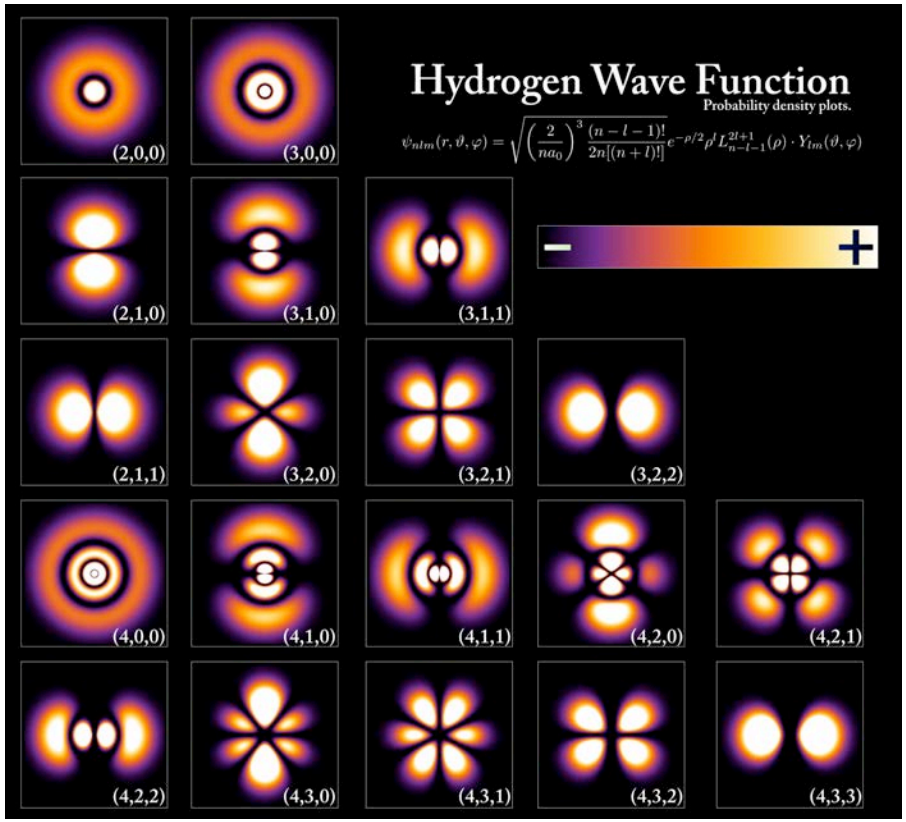
- For instance, if $U(g)=R(g)$, then f is a *vector* function. (Sometimes called covariant, although this word has subtly different meanings in different contexts)

A basis set for the spherical functions

- A general function $f(\theta,\varphi)$ may transform arbitrarily, but could we break it down into equivariant pieces?
- Remember the hydrogen atom solutions to the Schrödinger equation
- The spherical harmonics form a basis set for functions that separates the action of the rotation group into *irreducible representations*.

$$f(\theta,\phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c[f]_m^l Y_m^l(\theta,\phi)$$

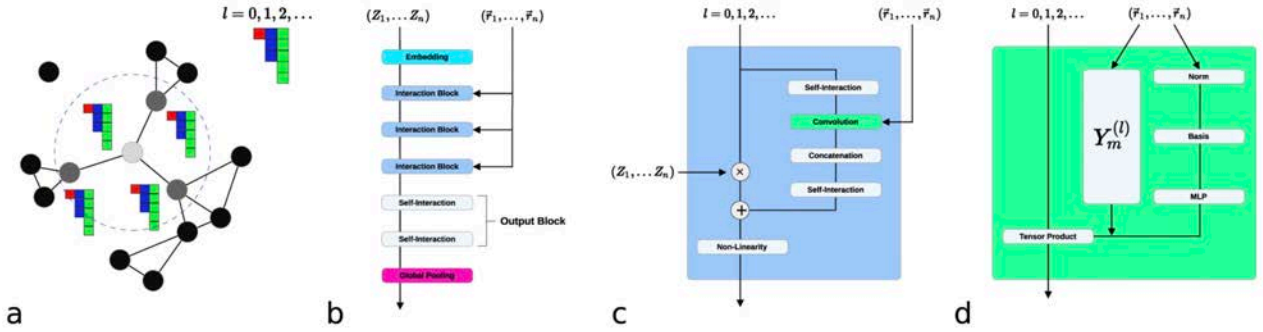
$$Y_m^l(\theta,\phi) \xrightarrow{g} Y_m^l(\theta',\phi') = \sum_{m'} U_{mm'}^l(g) Y_{m'}^l(\theta,\phi)$$



Spherical tensor approaches to atomistic ML

- The modern approach is to generalize from sensitivity functions that express radial density into a tensor expansion of density for multiple orders.

$$S_m^{(l)}(\vec{r}_{ij}) = R(r_{ij})Y_m^{(l)}(\hat{r}_{ij})$$



- In many approaches the *tensor product* is the key operation, based on the multiplication identity for spherical harmonics

Article | Open access | Published: 04 May 2022

E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt & Boris Kozinsky

Nature Communications **13**, Article number: 2453 (2022) | Cite this article

$$Y_{a,\alpha}(\theta,\varphi)\, Y_{b,\beta}(\theta,\varphi) = \sqrt{\frac{(2a+1)(2b+1)}{4\pi}} \sum_{c=0}^{\infty} \sum_{\gamma=-c}^{c} (-1)^{\gamma} \sqrt{2c+1} \begin{pmatrix} a & b & c \\ \alpha & \beta & -\gamma \end{pmatrix} \begin{pmatrix} a & b & c \\ 0 & 0 & 0 \end{pmatrix} Y_{c,\gamma}(\theta,\varphi)$$

2022: MACE: Deep, equivariant, many-bodied modeling

- Message-passing ACE, based on a non-deep version called the *Atomic Cluster Expansion* (ACE)
- Messages are formed from tensor products of multiple neighbors at once
- Can be shown to be complete in the sense that arbitrary functions can be created out of one layer operations

Single-particle environment tensors:

$$A^{(t)}_{i,kl_3m_3} = \sum_{l_1m_1,l_2m_2} C^{l_3m_3}_{l_1m_1,l_2m_2} \sum_{j\in\mathcal{N}(i)} R^{(t)}_{kl_1l_2l_3}(r_{ji}) Y^{m_1}_{l_1}(\hat{\boldsymbol{r}}_{ji}) \sum_{\tilde{k}} W^{(t)}_{k\tilde{k}l_2} h^{(t)}_{j,\tilde{k}l_2m_2},$$

Multi-particle environment tensors

$$\boldsymbol{B}^{(t)}_{i,\eta_\nu kLM} = \sum_{\boldsymbol{lm}} \mathcal{C}^{LM}_{\eta_\nu,\boldsymbol{lm}} \prod_{\xi=1}^{\nu} \sum_{\tilde{k}} w^{(t)}_{k\tilde{k}l_\xi} A^{(t)}_{i,\tilde{k}l_\xi m_\xi}, \quad \boldsymbol{lm} = (l_1m_1,\ldots,l_\nu m_\nu)$$

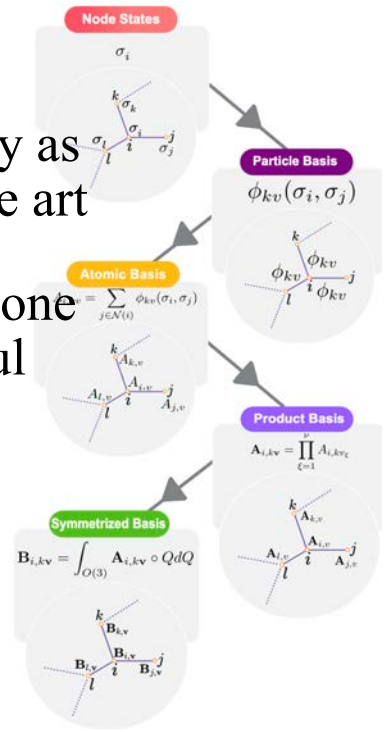
Under mild conditions on the two-body bases $\boldsymbol{A}^{\nu'}_i$, the higher order features $\boldsymbol{B}^{\nu'}_{i,\eta_\nu kLM}$ can be interpreted as a *complete basis* of many-body interactions [17], which can be computed at a cost comparable to pairwise interactions. Because of this, the expansion (11) is *systematic*. It can in principle be converged to represent any smooth $(\nu+1)$-body equivariant mapping in the limit of infinitely many features (proof in [17]).

MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields

Authors

Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, Gabor Csanyi

- Regarded by many as current state of the art
- Much effort has gone into training useful models

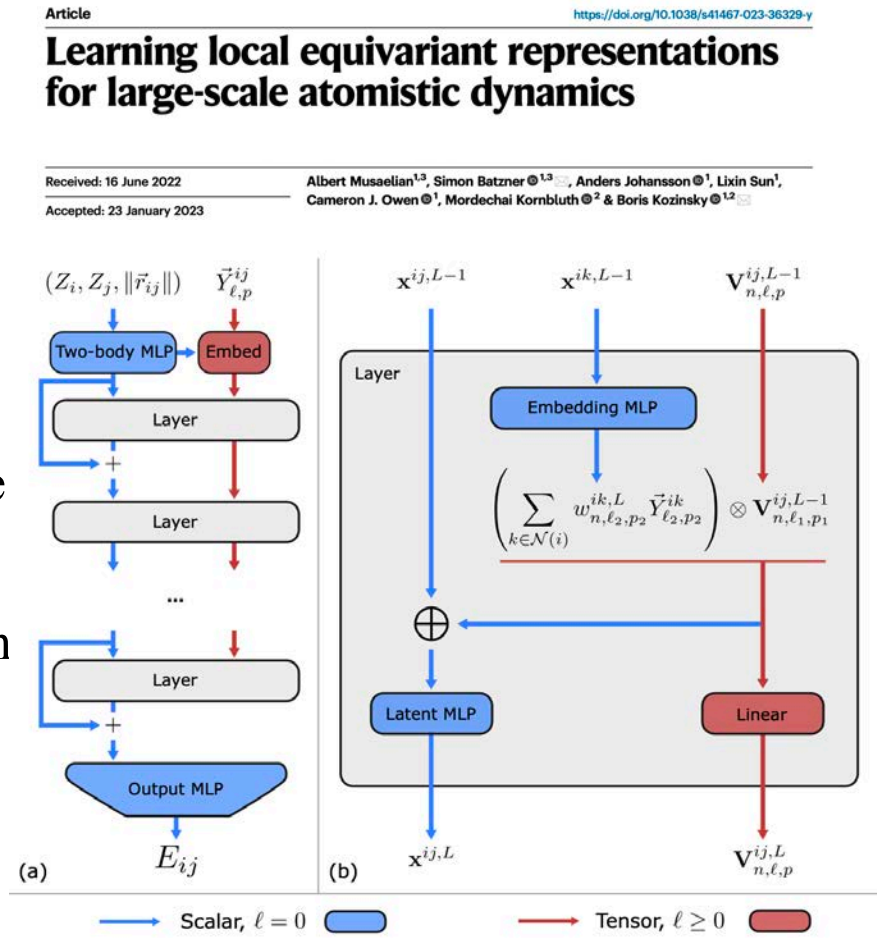


The Design Space of E(3)-Equivariant Atom-Centered Interatomic Potentials

Ilyes Batatia,[1,2,*] Simon Batzner,[3,*] Dávid Péter Kovács,[1] Albert Musaelian,[3] Gregor N. C. Simm,[1] Ralf Drautz,[4] Christoph Ortner,[5] Boris Kozinsky,[3,6] and Gábor Csányi[1]

2023: Allegro: *local* depth

- Highly novel architecture which has fixed range even with many interaction layers; great for molecular dynamics
- Exploits tensor product for rich geometric variance
- Edge variables but no atom variables, uses a directional-bond version of the line graph approach
- Aims and achieves high speed, accuracy, and parallelizability



2023: Allegro - *local* depth

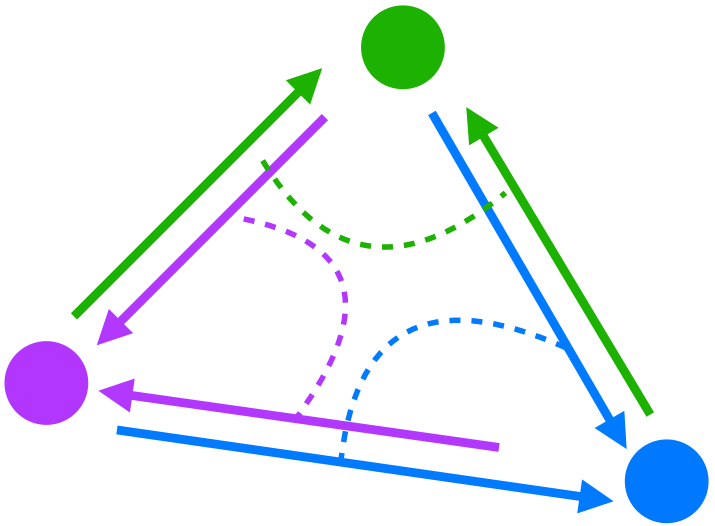
- Highly novel architecture which has fixed range even with many interaction layers; great for molecular dynamics
- Exploits tensor product for rich geometric variance
- Edge variables but no atom variables, uses a directional-bond version of the line graph approach
- Aims and achieves high speed, accuracy, and parallelizability


Article https://doi.org/10.1038/s41467-023-36329-y

Learning local equivariant representations for large-scale atomistic dynamics

Received: 16 June 2022
Accepted: 23 January 2023

Albert Musaelian[1,3], Simon Batzner[1,3], Anders Johansson[1], Lixin Sun[1], Cameron J. Owen[1], Mordechai Kornbluth[2] & Boris Kozinsky[1,2]



Allegro uses *directed,* edge-centered variables, and passes information through the line graph. Thus, all edges connected to a given atom form a closed set which do not communicate

Outline

- Quick background on molecular dynamics, classical models, and the blessing and curse of electronic structure
- A primer on neural networks, neurons, and convolutional networks
- An overview of neurons for atomistic machine learning
- **A dive into our latest work in this area**

The Cartesian tensor basis

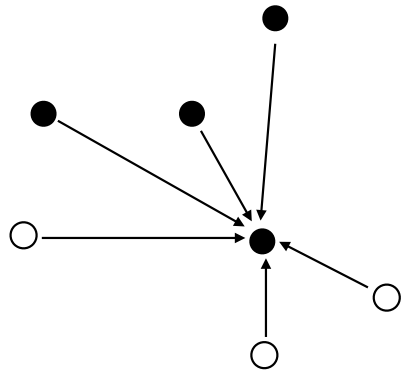
- There is a nice basis for describing the functions of angles in terms of *cartesian components* instead.
- These are called irreducible tensors, and they may be familiar from electrostatics in the *multipole expansion*
- They are 1-to-1 convertible with the spherical harmonics for a given value of *l*
- By projecting the environmental message distribution on to them, we get a set of symmetric traceless tensor describing the distribution of messages from the environment.

Charge	$T^{(0)}(\mathbf{r}) = 1$
Dipole	$T^{(1)}_{\alpha}(\mathbf{r}) = r_\alpha$
Quadrupole	$T^{(2)}_{\alpha\beta}(\mathbf{r}) = r_\alpha r_\beta - \frac{1}{3}\delta_{\alpha\beta} r^2$
Octopole	$T^{(3)}_{\alpha\beta\gamma}(\mathbf{r}) = r_\alpha r_\beta r_\gamma - \frac{1}{5}(\delta_{\alpha\beta} r_\gamma + \delta_{\alpha\gamma} r_\beta + \delta_{\beta\gamma} r_\alpha) r^2$

Each tensor is totally symmetric and traceless.

$$\mathcal{E}_{i,a}(\hat{\mathbf{r}}) = \sum_j \delta(\hat{\mathbf{r}} - \hat{\mathbf{r}}_{ij}) m_{ij,a},$$

$$\boldsymbol{\mathcal{E}}^{(\ell)}_{i,a} = \int \mathbf{T}^{(\ell)}(\hat{\mathbf{r}}) \mathcal{E}_{i,a}(\hat{\mathbf{r}})\, d^2\hat{\mathbf{r}}.$$

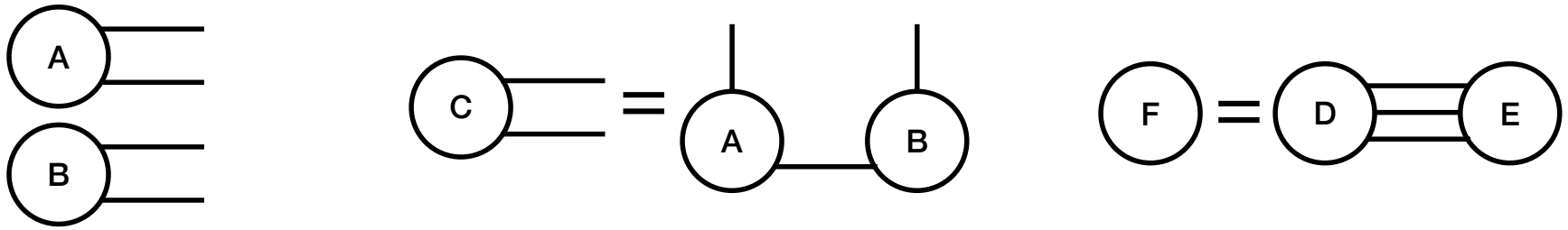


Tensor contractions as graphs

- Just like vectors, we can make new tensors by contracting indices (dot product)

$$C_{ab} = \sum_{c} A_{ac} B_{bc} \qquad\qquad F = \sum_{abc} D_{abc} E_{abc}$$

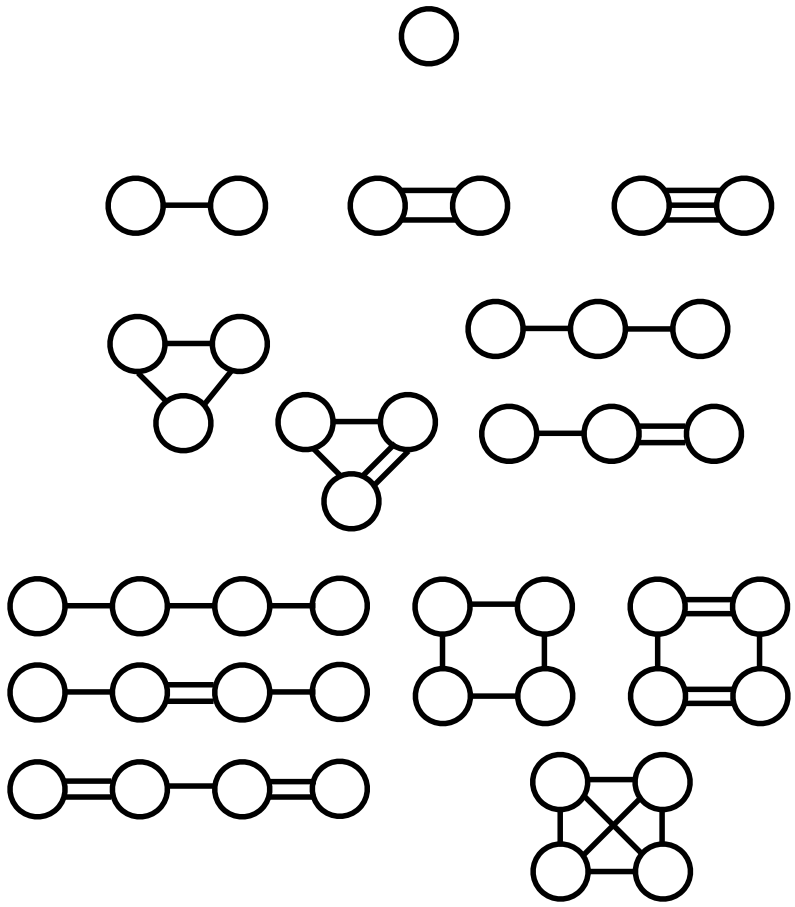
- This is far simpler to think about than the spherical tensor product rule.
- Because our tensors are symmetric, they can be represented without order or name associated to the indices using a Penrose graphical notation:



- Because the tensors are traceless, a self-loop gives a value of zero.

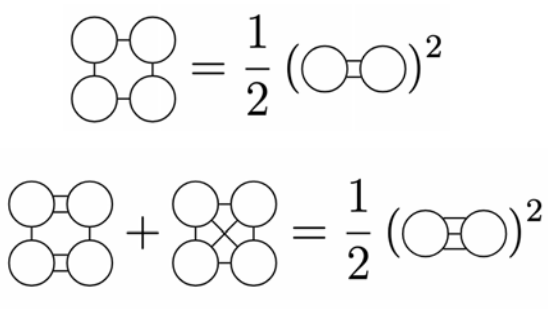
ACE, MACE, and Beyond

- Other recent approaches have created features such as Atomic Cluster Expansion (ACE) and Message-passing ACE (MACE) which (in some sense) use every graph available up to a certain order.
- This approach is expressive but can be expensive, as the number of graphs grows combinatorially with the order of the expansion.

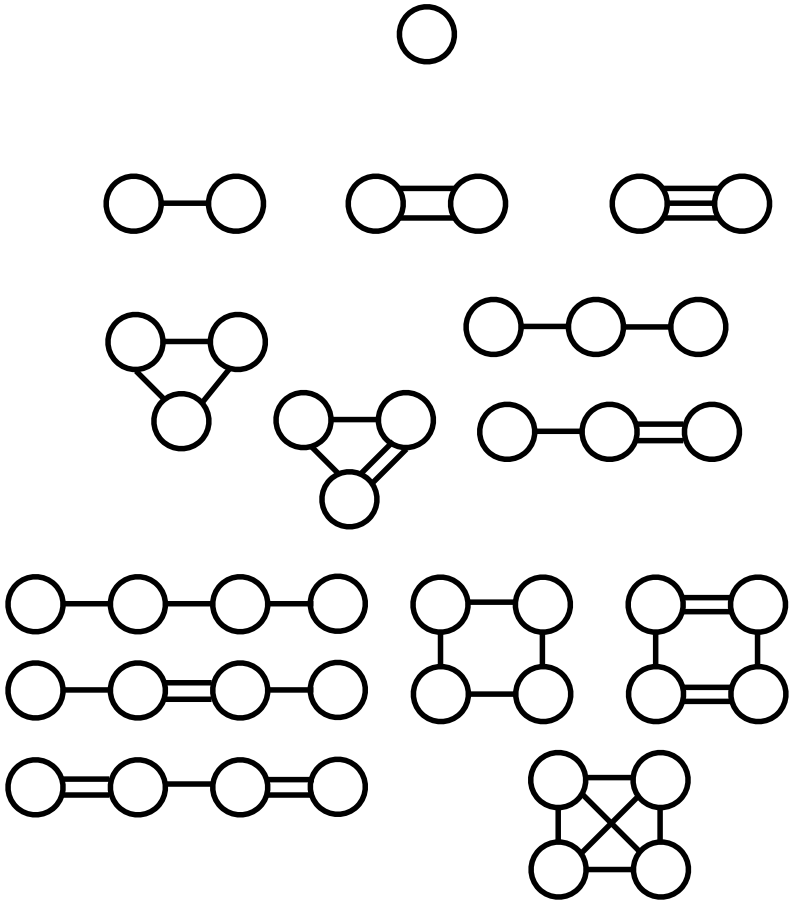


Taming the explosion

- Central question: Are there any graphs which are not useful?

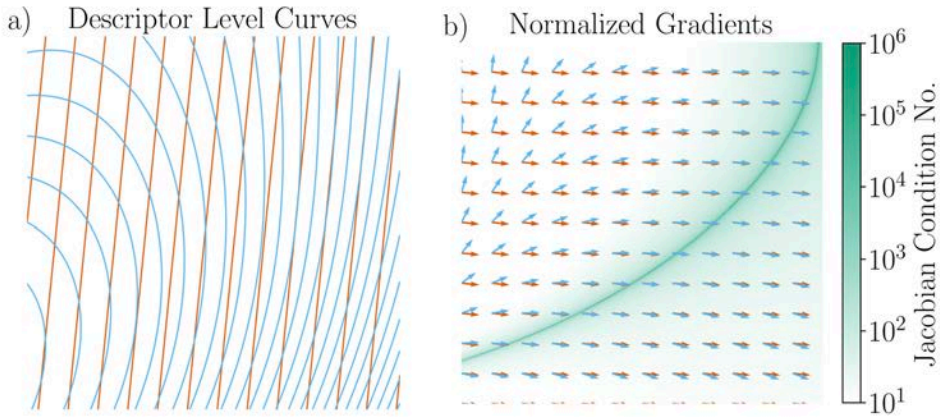


- We'd be happy to have a set which can localize a pattern: within some neighborhood of the values of the set of graphs, could we guarantee being in a neighborhood of some configuration, modulo symmetry?



Finding complete and flexible bases

- Although the complete set is good almost everywhere, it can break down on sub-manifolds
- A *flexible* set of graphs ensures that a degeneracy does not happen
- We have analyzed the set of graphs to account for the source of degeneracies: they occur when one of the moments around an atom are zero (local symmetry).
- Knowing this allows us to construct a flexible set.



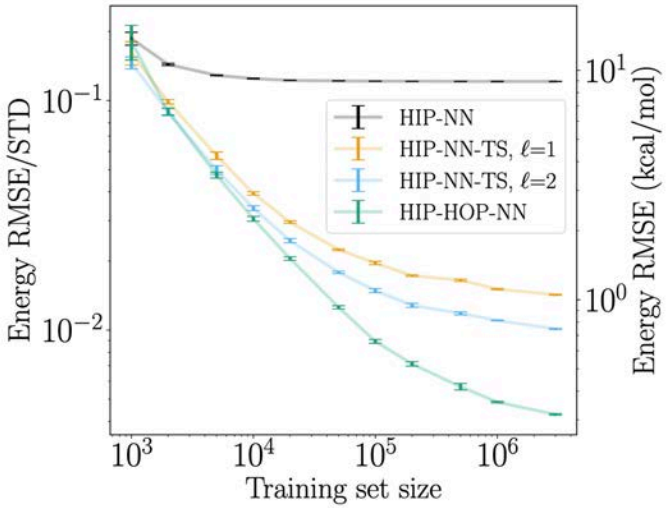
$$I_n(r) = f(I_1(r), ..., I_{n-1}(r)).$$

$$\frac{\partial I_n(r)}{\partial r} = \sum_{i=1}^{n_1} \frac{\partial f}{\partial I_i} \frac{\partial I_i}{\partial r}$$

Max Langbein and Hans Hagen, "A generalization of moment invariants on 2D vector fields to tensor fields of arbitrary order and dimension," in *International Symposium on Visual Computing* (Springer, 2009) pp. 1151–1160.

2025: HIP-HOP-NN

- The resulting model architecture we call HIP-NN with **H**igher **O**rder **P**olynomials, or HIP-HOP-NN.
- First complete five-body neurons
- Competitive accuracy and speed



Methane Completeness Benchmark

Accuracy on the COMP6 Benchmark Set:

		ANI-1x [44]	GM-NN [47]	MACE 96-1 [46]	MACE 192-2 [46]	NewtonNet [5]	TensorNet 2L [48]	HIP-HOP
COMP6	E	1.93	2.03	0.76	0.48	1.45	-	0.64
	F	3.09	1.85	0.77	0.52	1.79	-	0.80

Molecule	N	TensorNet 2L	MACE-OFF small	MACE-OFF medium	MACE-OFF large	HIP-HOP-NN 256
Alanine Dipeptide	*22*	26.5	15.5	19.2	23.0	14.1
Chignolin	*166*	26.9	15.5	24.4	68.7	13.7
DHFR	*2489*	106.7	68.2	272.2	1025.6	95.6
Factor IX	*5807*	248.6	143.1	610.2	-	210.7

Timings (ms/step)

Conclusions

Shameless plug: our code is available open-source on GitHub in a library called *hippynn*

- Ideas of NNs for PES's go back to the early 1990s at least. Symmetry has been recognized from the start.
- Behler's early successes with *Atomic Environment Vectors*, leading eventually to ANI towards universal potentials
- End-to-end NN approaches based on atom-centered message passing with *scalars,* such as distance and angle
- Tensor field networks and descendants involving *equivariant* message passing
- MACE uses both tensors and *higher-order messages*
- Our latest research, HIP-HOP-NN, uses flexible subsets invariants to retain universal approximation without combinatoric explosions

Incomplete References

- Chigaev, M., Smith, J.S., Anaya, S., Nebgen, B., Bettencourt, M., Barros, K. and Lubbers, N., 2023. Lightweight and effective tensor sensitivity for atomistic neural networks. The Journal of Chemical Physics, 158(18).
- Shinkle, E., Pachalieva, A., Bahl, R., Matin, S., Gifford, B., Craven, G. T., & Lubbers, N. (2024). Thermodynamic Transferability in Coarse-Grained Force Fields using Graph Neural Networks. arXiv preprint arXiv:2406.12112.
- Lubbers, N., Smith, J.S. and Barros, K., 2018. Hierarchical modeling of molecular energies using a deep neural network. The Journal of chemical physics, 148(24).
- Smith, J.S., Isayev, O. and Roitberg, A.E., 2017. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. Chemical science, 8(4), pp.3192-3203.
- Kulichenko, M., Smith, J.S., Nebgen, B., Li, Y.W., Fedik, N., Boldyrev, A.I., Lubbers, N., Barros, K. and Tretiak, S., 2021. The rise of neural networks for materials and chemical dynamics. The Journal of Physical Chemistry Letters, 12(26), pp.6227-6243.
- Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K. and Riley, P., 2018. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. arXiv preprint arXiv:1802.08219.
- Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J.P., Kornbluth, M., Molinari, N., Smidt, T.E. and Kozinsky, B., 2022. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. Nature communications, 13(1), p.2453.
- Drautz, R., 2019. Atomic cluster expansion for accurate and transferable interatomic potentials. Physical Review B, 99(1), p.014104.
- Batatia, I., Kovacs, D.P., Simm, G., Ortner, C. and Csányi, G., 2022. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. Advances in Neural Information Processing Systems, 35, pp.11423-11436.