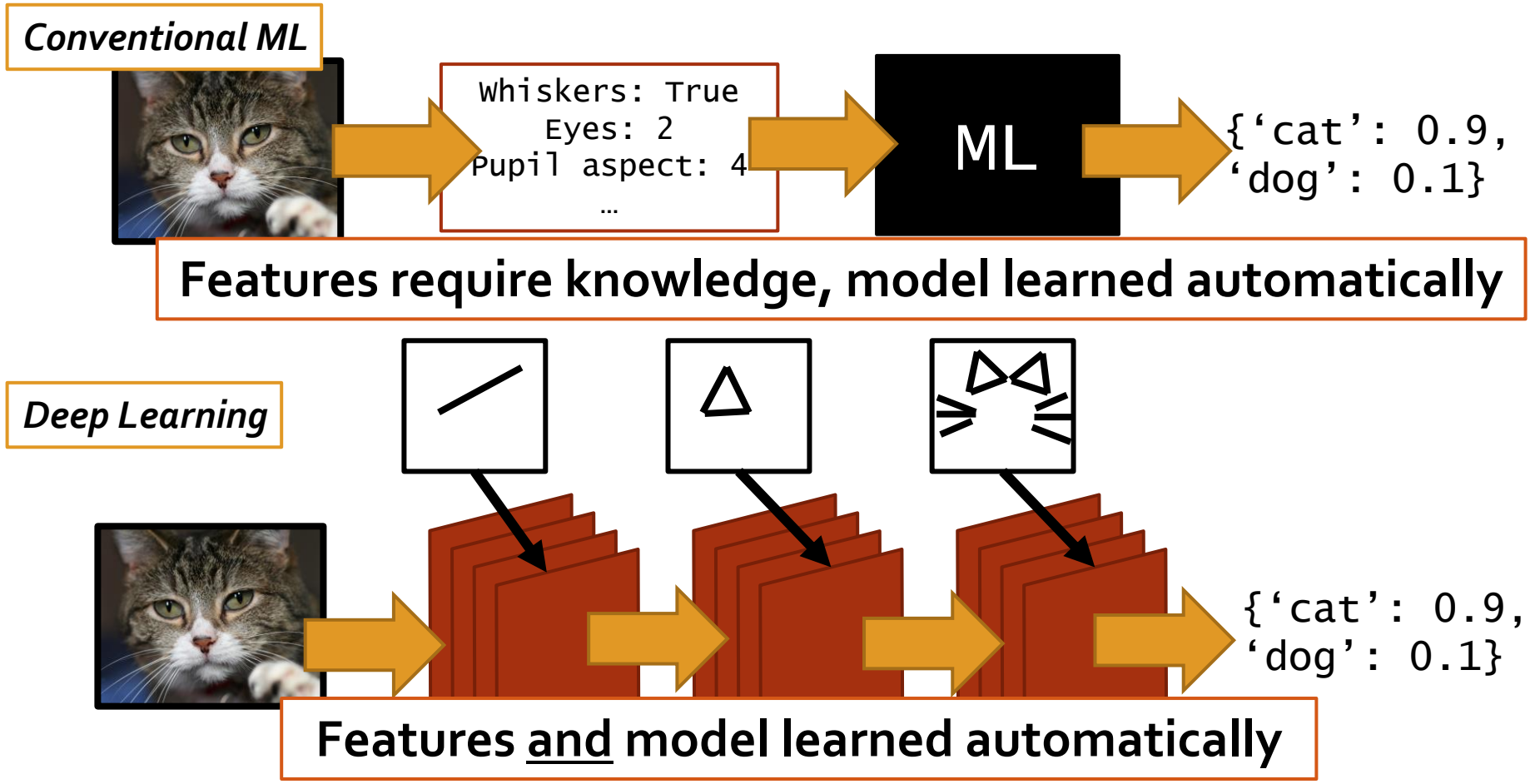
DEEP LEARNING FOR MOLECULAR DATA

Logan Ward
Asst. Computational Scientist
Argonne National Laboratory

26 January 2021

What is different about deep learning?



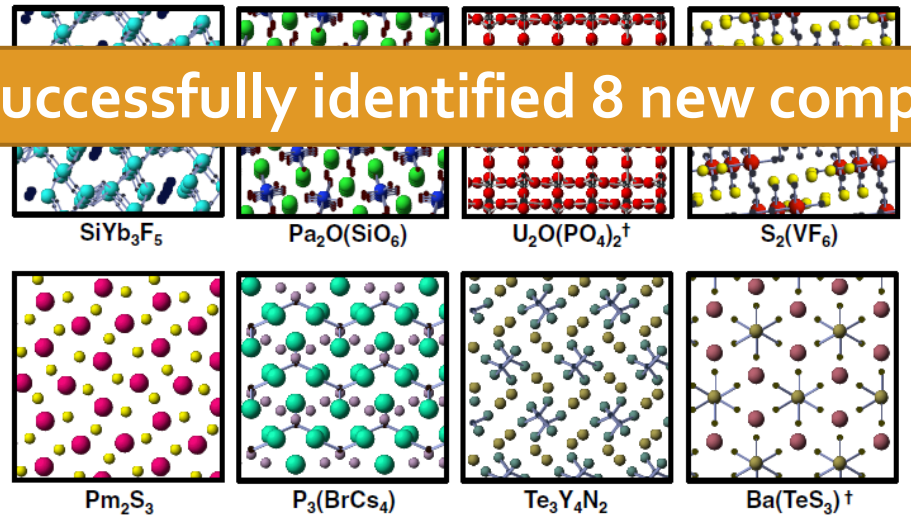
FIRST, WHAT CONVINCED ME DEEP LEARNING WAS WORTHWHILE

Predicting Crystalline Materials

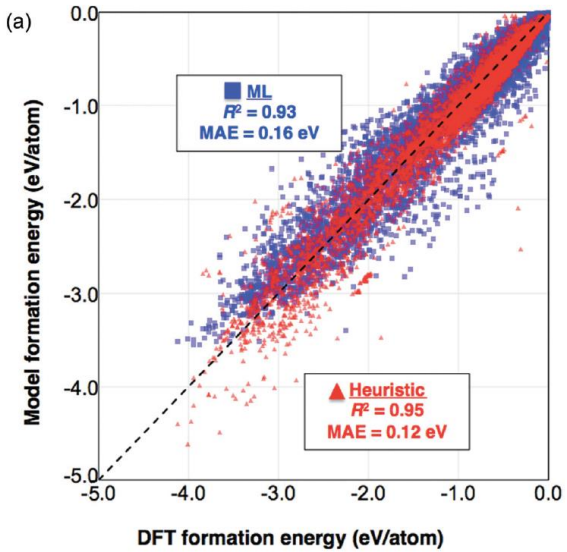
Goal: Identify new crystalline materials

Model: Given, composition predict ΔH_f

Initial attempt: 2014



Successfully identified 8 new compounds!



Error still ~10% typical values

Can we do even better with deep neural networks?

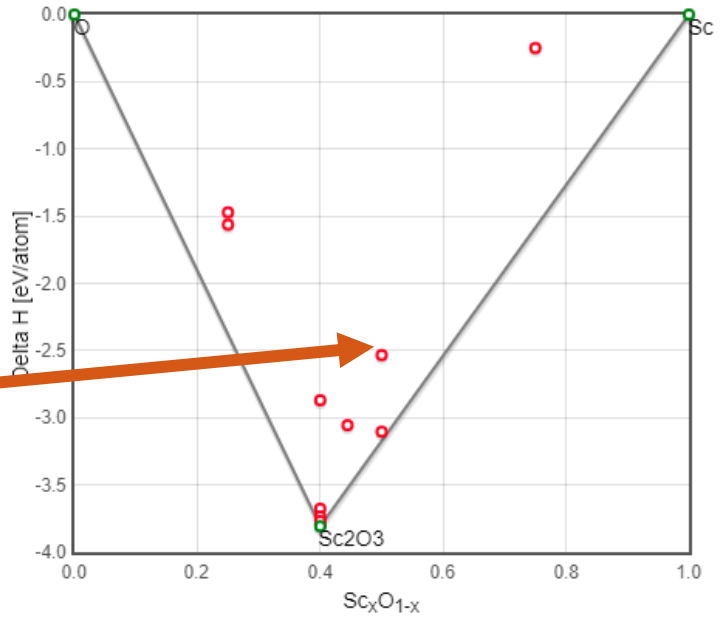
Ref: Meredig et al. PRB. (2014), 094104. doi: 10.1103/PhysRevB.89.094104

Training Data

Data Source: Open Quantum Materials Database
~470k DFT Calculations

Why OQMD?

- ✔ Open ✔ Large
- ✔ Contains unstable materials



Training Set:

Subset: Lowest energy structure

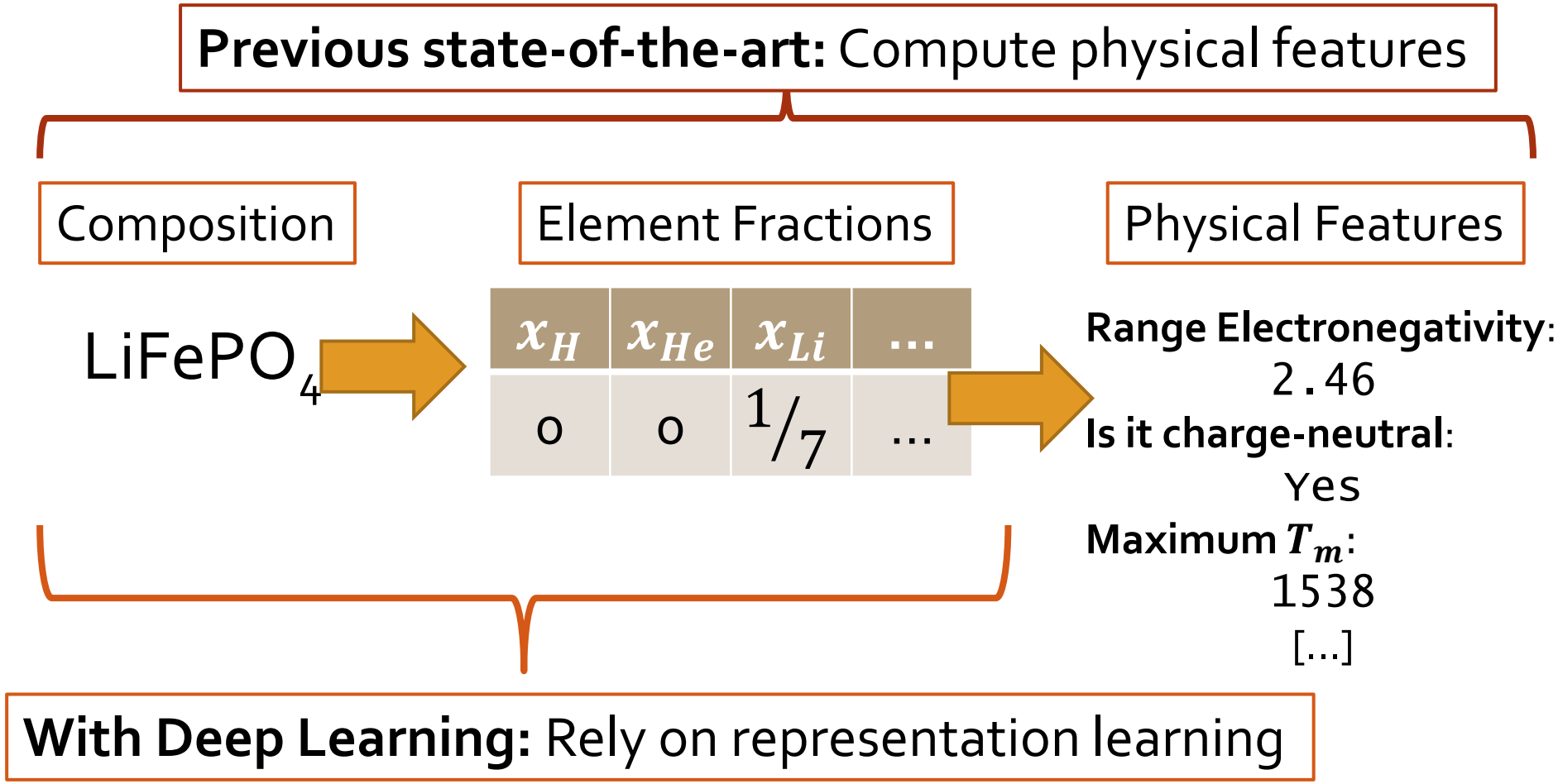
Size: 230336

Input: Composition

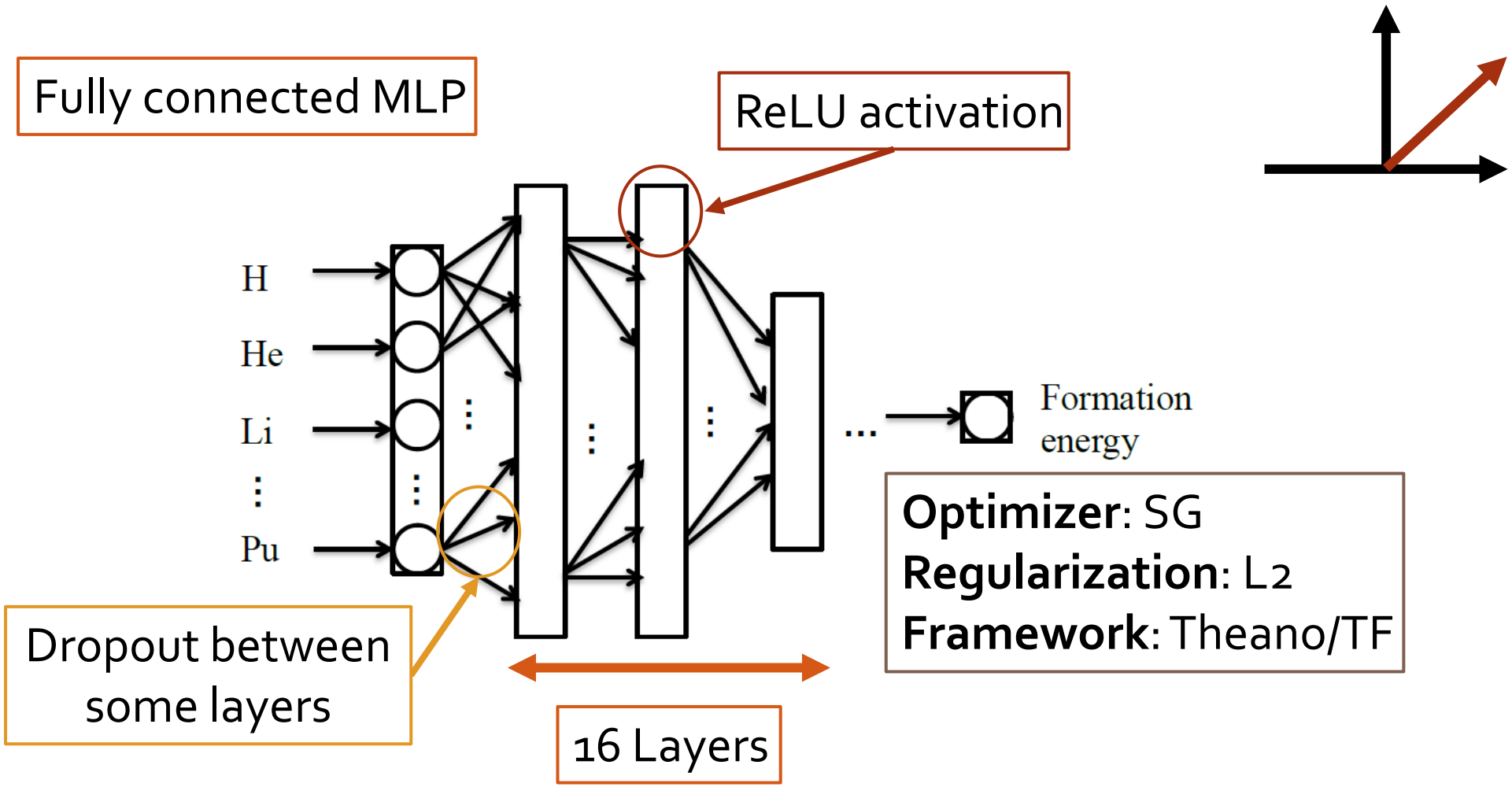
Output: Formation Enthalpy ΔH_f

Ref: Jha et al., Sci. Rep. (2020) 17593

Input Features

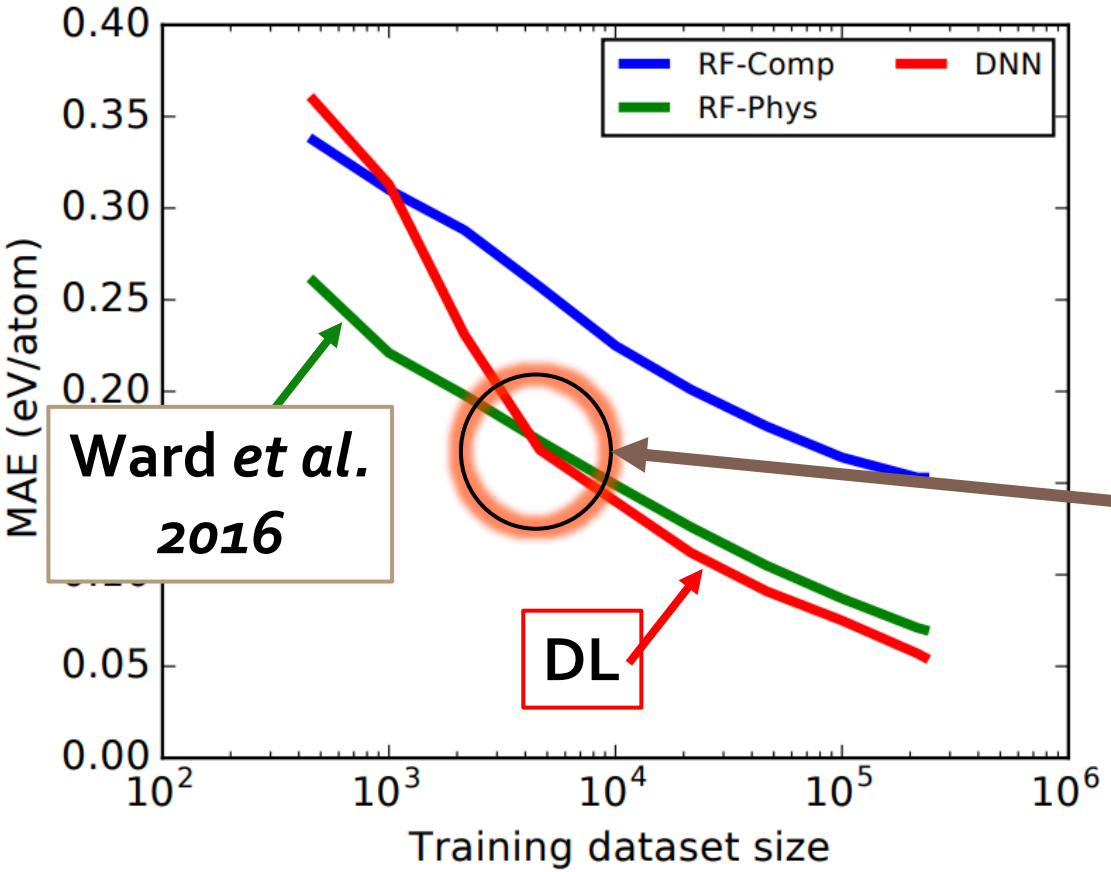


ElemNet Architecture: Nothing Fancy



Ref: Jha et al., Sci. Rep. (2020) 17593

Better than conventional learning?



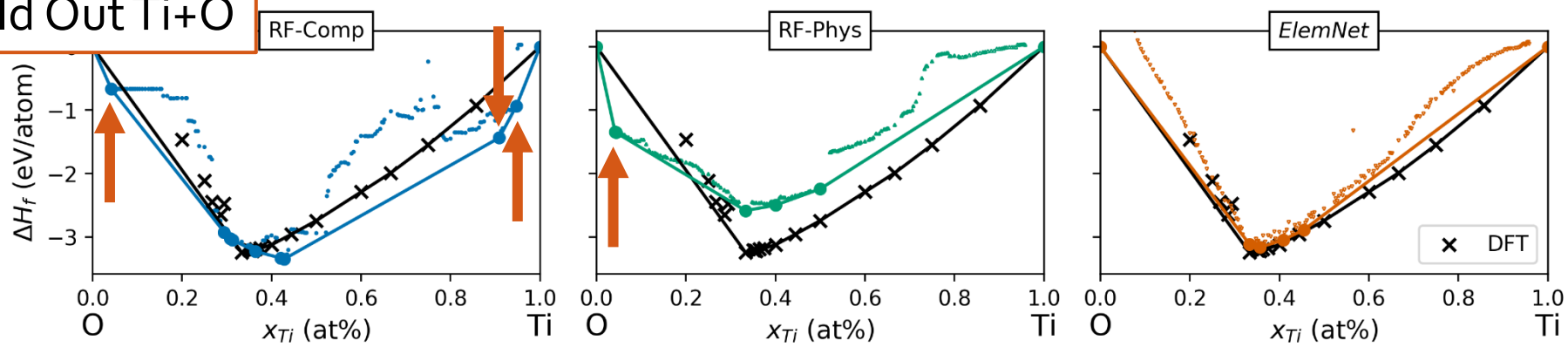
(ML) Meredig 2014: 0.092 eV/atom
(ML) Ward 2016: 0.070 eV/atom
(DL) Jha 2017: 0.055 eV/atom

Beats state-of-the-art with
no input from materials scientist,
and only 4000 data points!

Ref: Jha et al., Sci. Rep. (2020) 17593

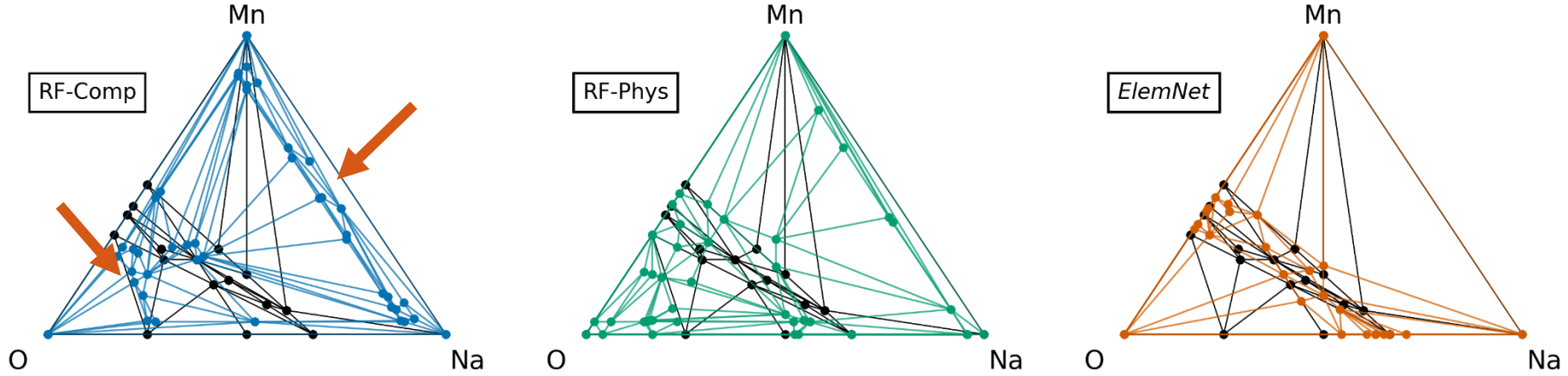
Can DL interpolate between elements?

Hold Out Ti+O



Deep Learning Yields Fewer Spurious Predictions

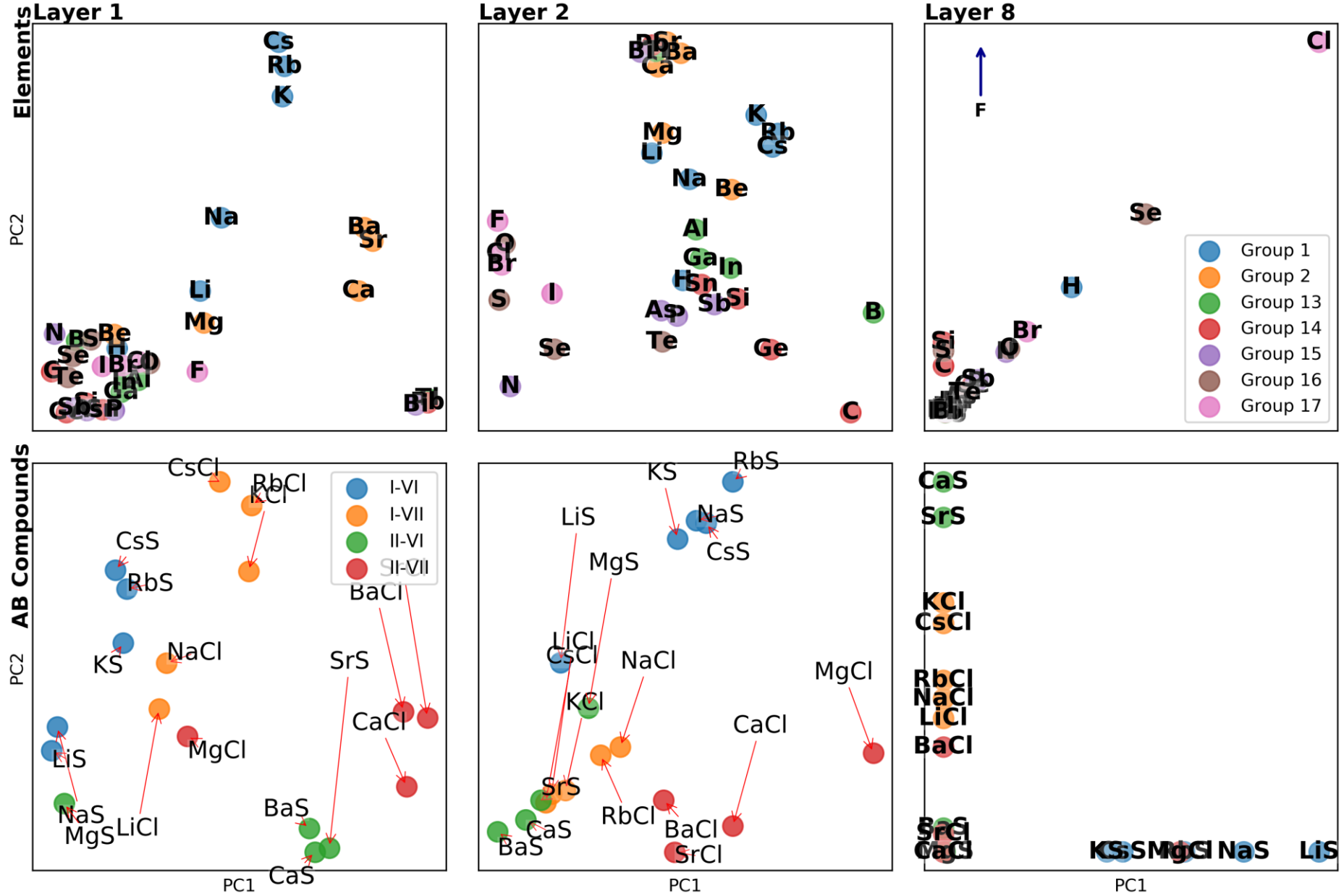
Hold Out Na-Fe-Mn-O



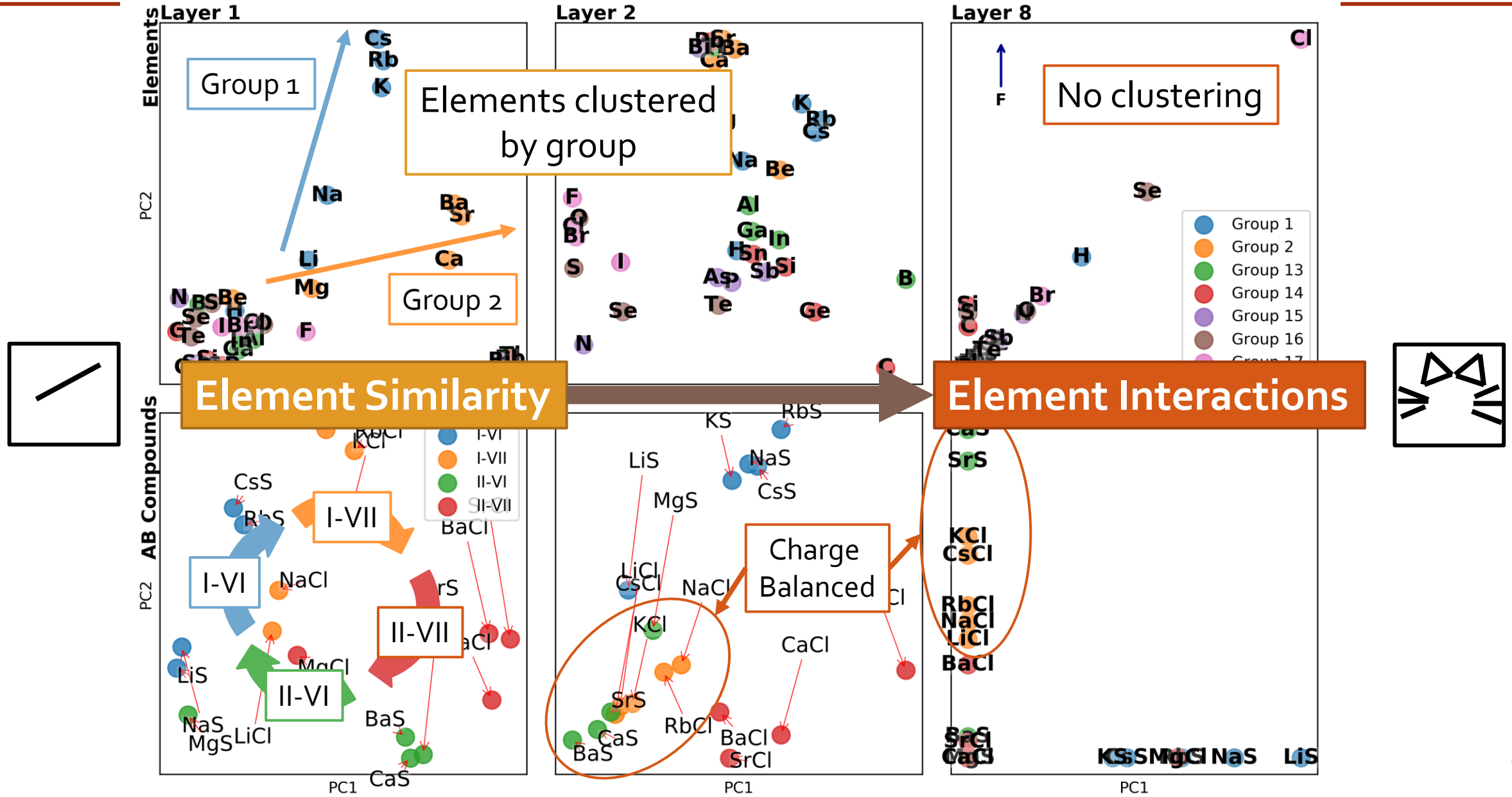
How is this possible? Model hasn't seen Ti-O bonds

Ref: Jha et al., Sci. Rep. (2020) 17593

How is it working so well?



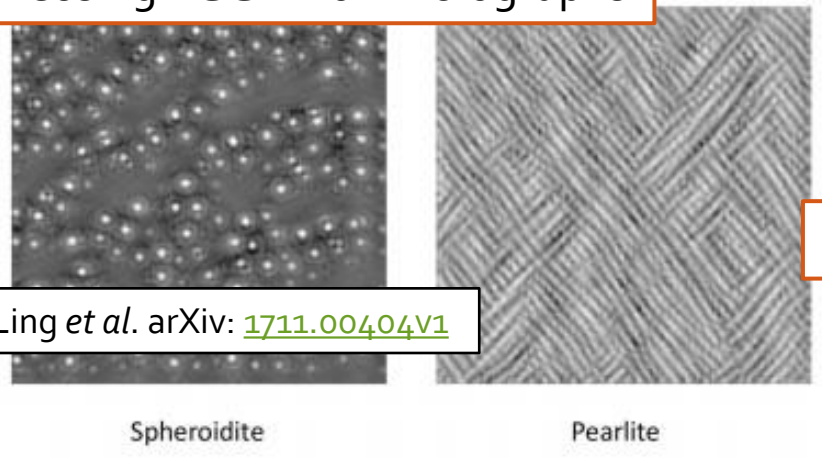
How is it working so well?



Big Opportunity: Transfer Learning

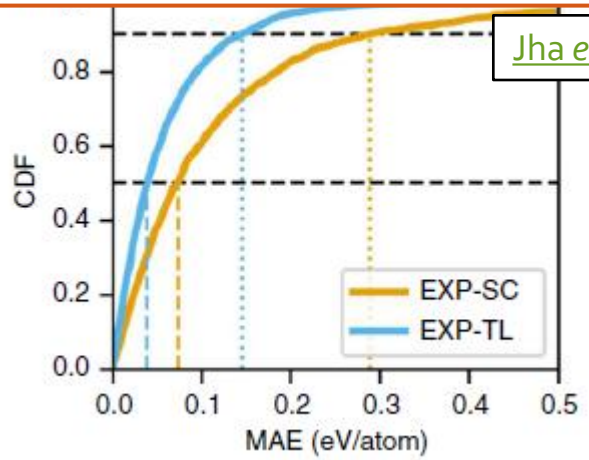
Features learned on one dataset applicable to others

Reusing VGG16 for micrographs



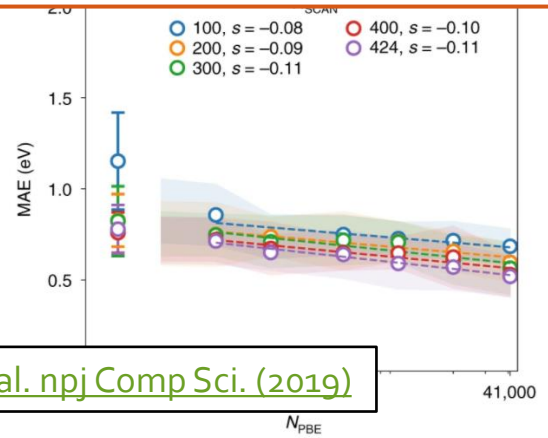
Ling *et al*. arXiv: 1711.00404v1

Reusing ElemNet for experimental data



Jha *et al*. Nat Comm. (2020)

Learning from multiple fidelities



Chen et al. npj Comp Sci. (2019)

Deep Learning is not Cure-All

Table 3: Summary of performances(test subset): conventional methods versus graph-based methods. Graph-based models outperform conventional methods on 11/17 datasets.

Category	Dataset	Metric	Best performances - conventional methods	Best performances - graph-based methods
Quantum Mechanics	QM7	MAE	KRR(CM): 10.22	**DTNN: 8.75**
	QM7b	MAE	KRR(CM): 1.05	**DTNN: 1.77***
			Multitask: 0.0150	**MPNN: 0.0143**
			Multitask(CM): 4.35	**DTNN: 2.35**
		E	XGBoost: 0.99	**MPNN: 0.58**
		E	XGBoost: 1.74	**MPNN: 1.15**
		E	XGBoost: 0.799	**GC: 0.655**
		PRC	Logreg: 0.129	**GC: 0.136**
Biophysics	MUV	AUC-PRC	**Multitask: 0.184**	Weave: 0.109
	HIV	AUC-ROC	**KernelSVM: 0.792**	GC: 0.763
	BACE	AUC-ROC	**RF: 0.867**	Weave: 0.806
	PDBbind(full)	RMSE	**RF(grid): 1.25**	GC: 1.44
Physiology	BBBP	AUC-ROC	**KernelSVM: 0.729**	GC: 0.690
	Tox21	AUC-ROC	KernelSVM: 0.822	**GC: 0.829**
	ToxCast	AUC-ROC	Multitask: 0.702	**Weave: 0.742**
	SIDER	AUC-ROC	**RF: 0.684**	GC: 0.638
	ClinTox	AUC-ROC	Bypass: 0.827	**Weave: 0.832**

* As discussed in section 4.4, DTNN outperforms KRR(CM) on 14/16 tasks in QM7b while the mean-MAE is skewed due to different magnitudes of labels.

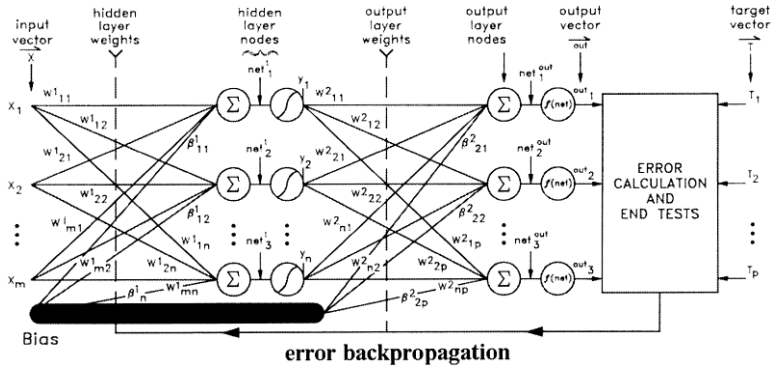
For benchmark problems of learning from molecular data, conventional ML better for **6/17** cases

Ref: Wu *et al.* Chem Sci. (2017) 10.1039/C7SC02664A

MANY APPROACHES FOR DEEP LEARNING WITH MOLECULAR DATA

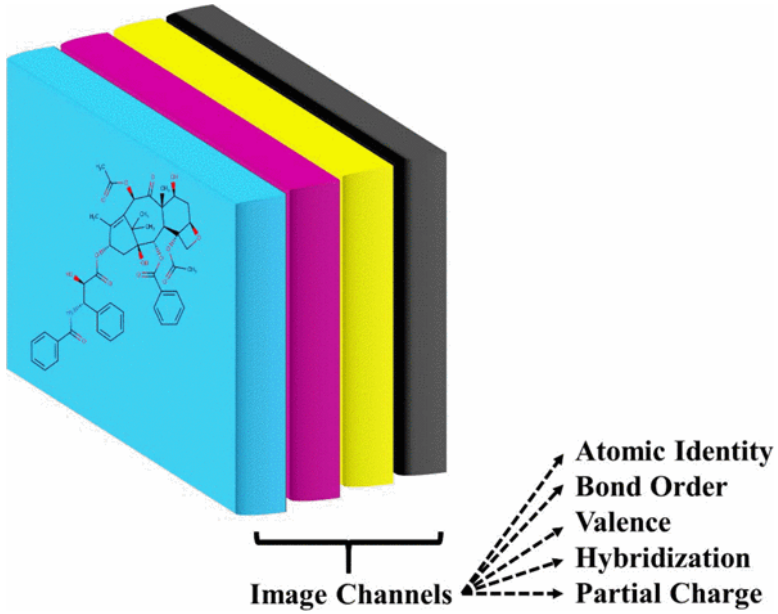
How do we do deep learning with molecules?

As a normal supervised learner



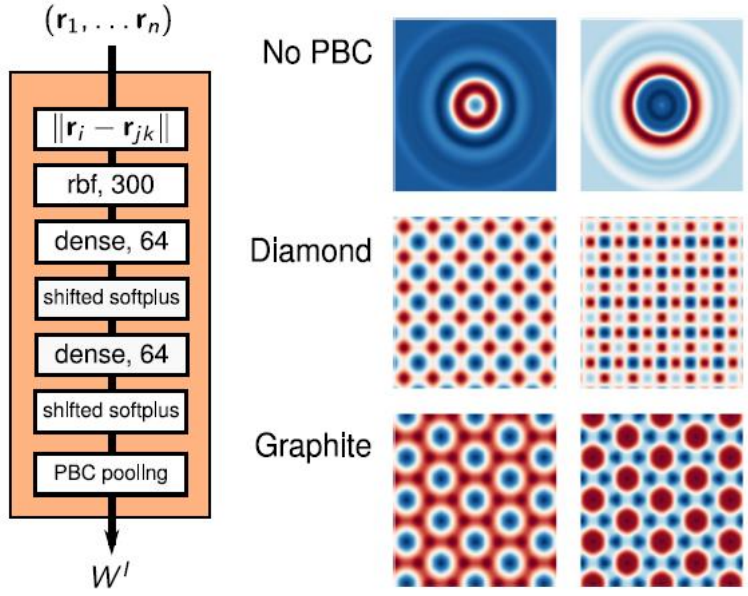
Ex: Ulmer et al. CTPS. (1998), 311.

Repurposing Image Models



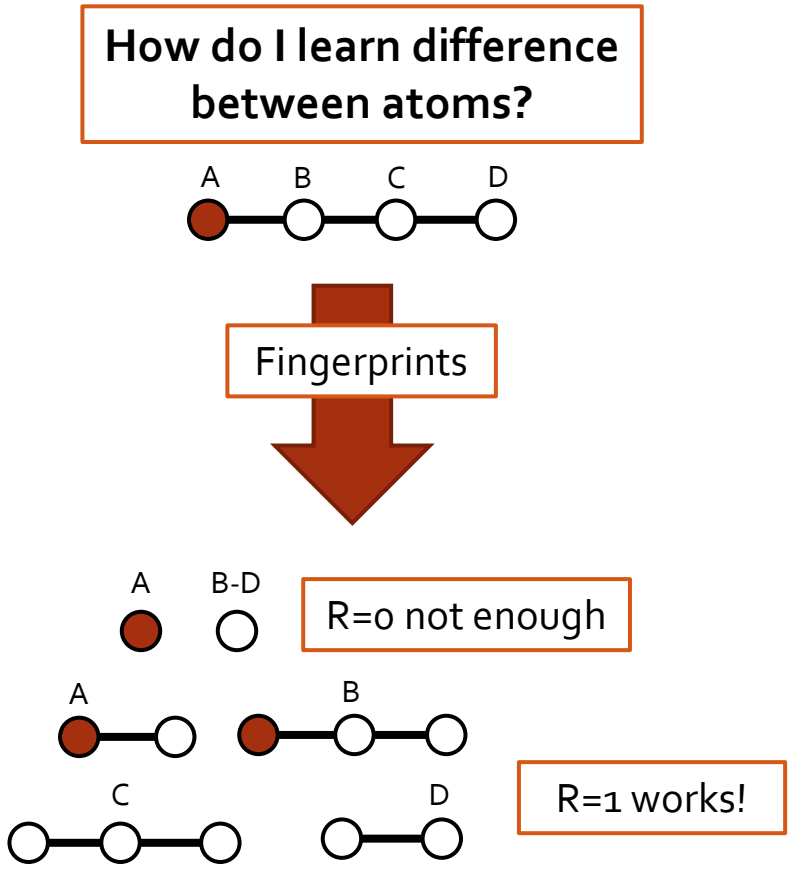
Ex: Goh et al. IEEE WACV. (2018)

Message-Passing Networks



Ex: Schütt et al. JCP, (2018)

Our focus: Message-Passing Networks

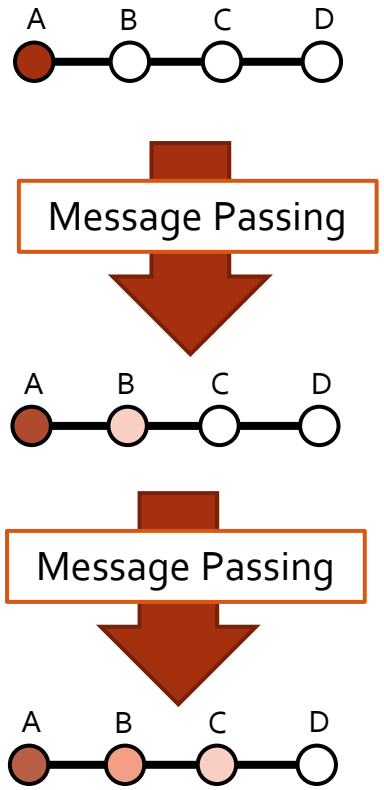


Problems:

1. Combinatorics: Many features!
2. No similarity between groups (is A:B > B:C)

Our focus: Message-Passing Networks

How do I learn difference between atoms?



Atomic fingerprints are learned by "convolution"

Message Passing is Two Steps

Step 1: Message generation

$$m(B) = f_M(\text{B}, \text{A}, \text{—}) + f_M(\text{B}, \text{C}, \text{—}) =$$

Step 2: Update

$$B^{\{n+1\}} = f_U(\text{B}, \text{m}) =$$

Solution: Message passing gives continuous features

~~1. Combinatorics: Many features!~~ Fixed feature count

~~2. No similarity between groups (is A:B > B:C)~~ Distances

New Challenge: Learning update and message functions

Formalizing Message Passing Neural Networks

A generalized form of neural networks for graph data, introduced by Gilmer et al. (Google)

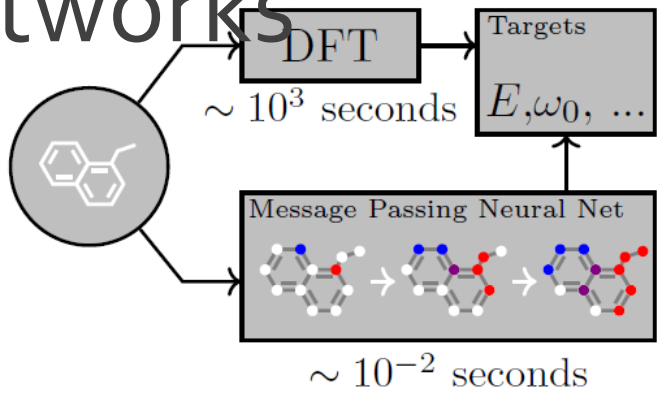


Figure 1. A Message Passing Neural Network predicts quantum properties of an organic molecule by modeling a computationally expensive DFT calculation.

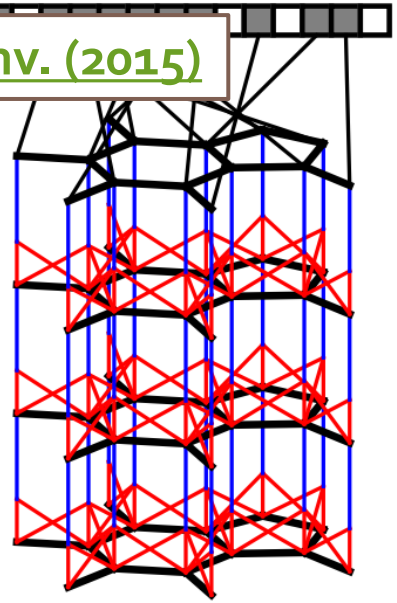
Existing strategies mostly variants of

- $m_v^{t+1} = \sum_{w \in N(v)} \boldsymbol{M_t}(h_v^t, h_w^t, e_{vw})$ 1. **Gather** messages from neighboring nodes
- $h_v^{t+1} = \boldsymbol{U_t}(h_v^t, m_v^{t+1})$ 2. **Update** node state given messages
- $\hat{y} = \boldsymbol{R}(\{h_v^T | v \in G\})$ 3. **Readout** graph properties given node states

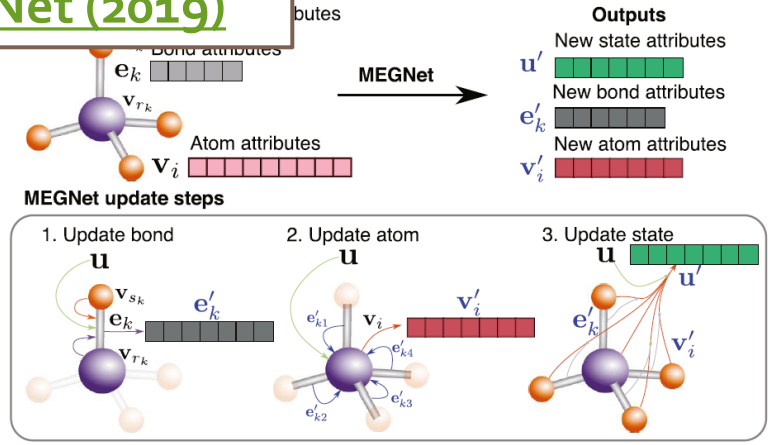
Ref: Gilmer et al. (2017). arXiv:1704.01212v2

Many variations of MPNNs

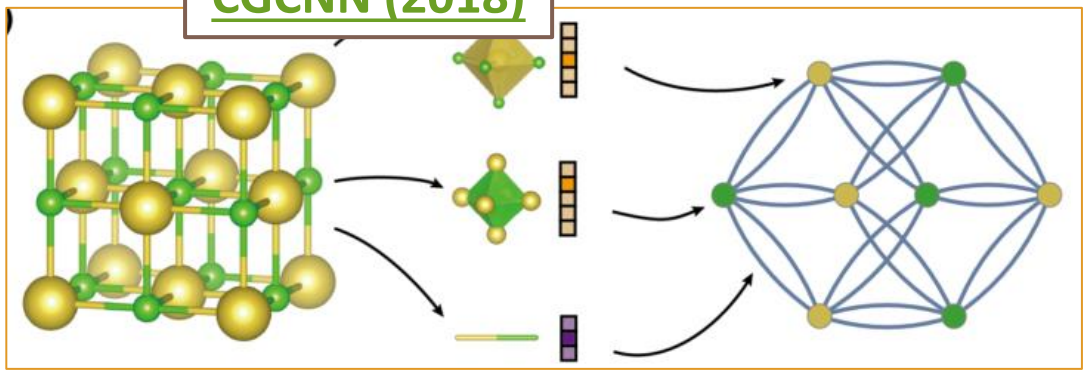
Graph Conv. (2015)



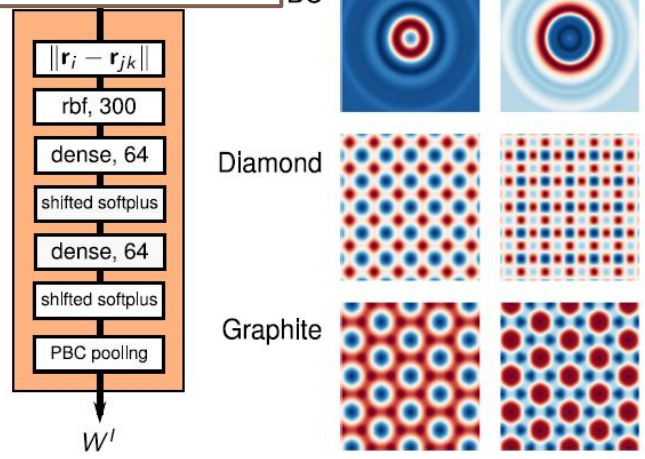
MEGNet (2019)



CGCNN (2018)

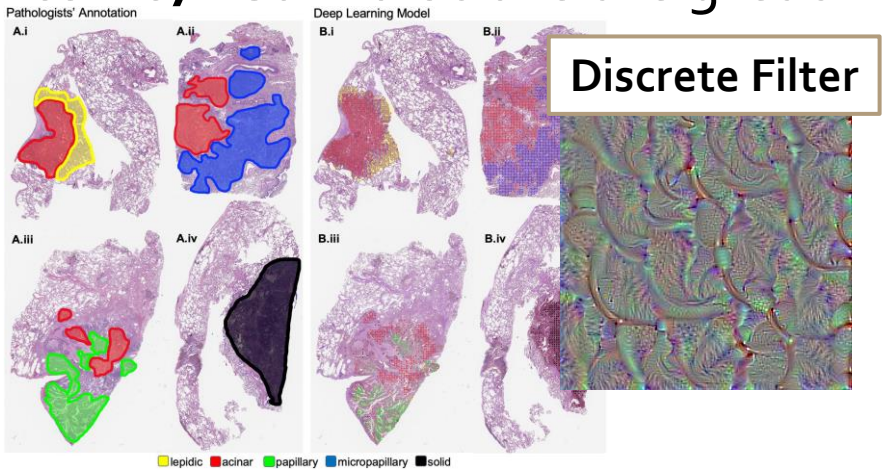


SchNet (2018)



SchNet: Continuous Convolutions

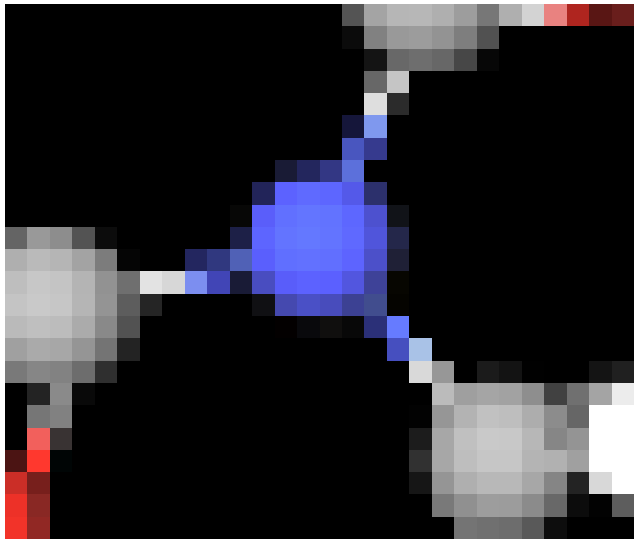
Opportunity: Convolutions are great!



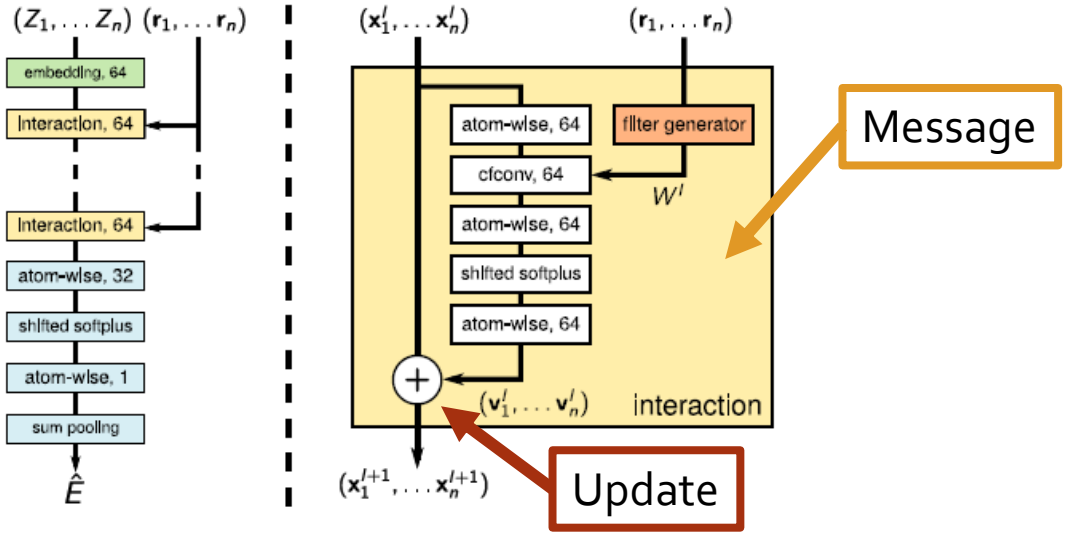
Wei *et al.* Sci Rep. (2019), 3358

Discrete Filter

Problem: Atoms are not pixel

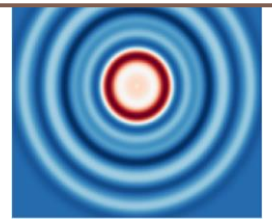


Key Innovation: "Continuous Convolutions"



$$\text{cfconv}(x_i) = \sum_{j}^{n_{atoms}} x_j \circ W(r_j - r_i)$$

Continuous Filter



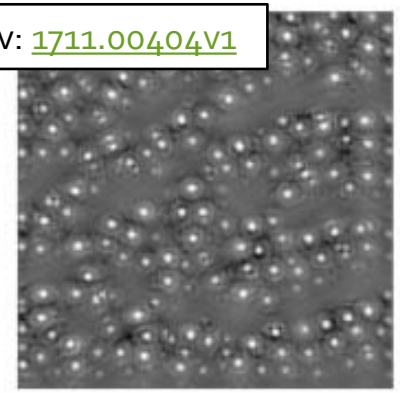
Ref: Schütt et al. JCP, (2018)

Take-home Points

Deep Learning has many advantages…

- Learning from sparse features
- High model complexity -> High accuracy
- Transfer learning

Ling *et al.* arXiv: 1711.00404v1

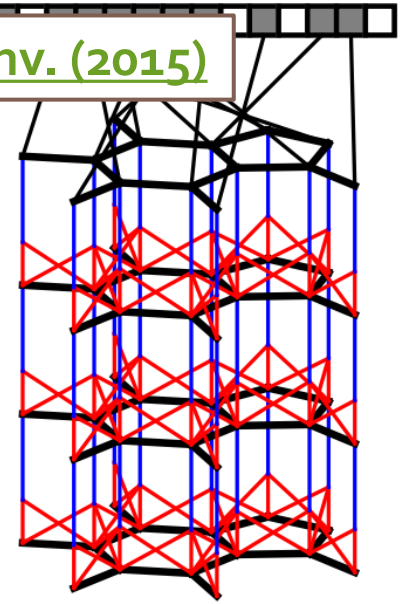


Spheroidite

Major method for molecules: Message Passing Networks

- Automatically learning fingerprints
- Combination of "message," "update" and "readout"
- Many variations in the literature

Graph Conv. (2015)



Practical Exercises: Learn how to implement these in TensorFlow