Quantum Machine Learning

Parameters, Delta, Crystals

Anatole von Lilienfeld

"Machine Learning, Quantum Mechanics, and Chemical Compound Space"
By Ramakrishnan and von Lilienfeld
published in: *Reviews in Computational Chemistry*
edited by Abby L. Parrill and Kenny B. Lipkowitz
Volume **30**, Chapter 5, pages 225-256 (2017)


UNI BASEL
Research at Google
U.S. DEPARTMENT OF ENERGY
INCITE
LEADERSHIP COMPUTING
MARVEL
NATIONAL CENTRE OF COMPETENCE IN RESEARCH
Institute for Pure & Applied Mathematics
IPAM
University of California, Los Angeles
FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
UNITED STATES AIR FORCE

Conclusions

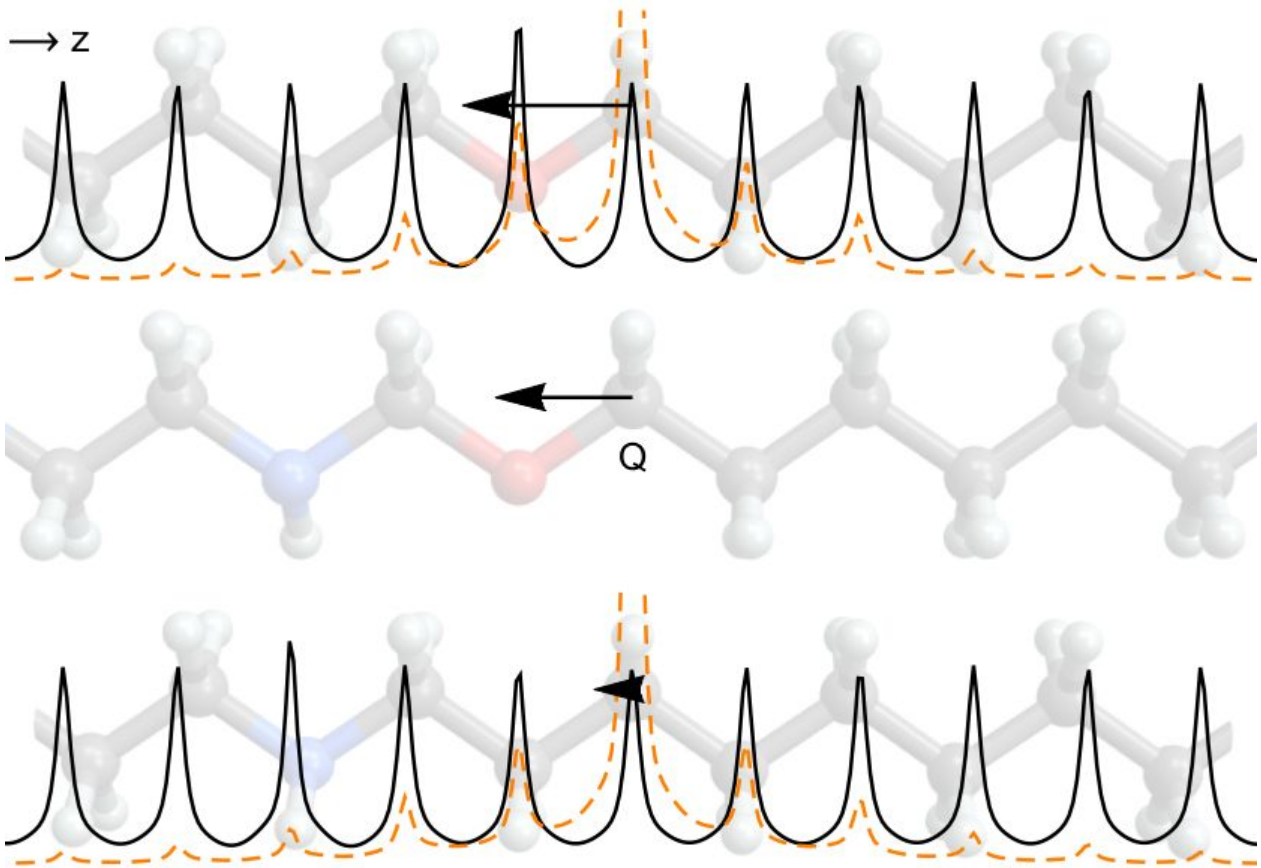
1. Parameter learning

2. Δ-ML (prior expert knowledge about property)

3. Crystals

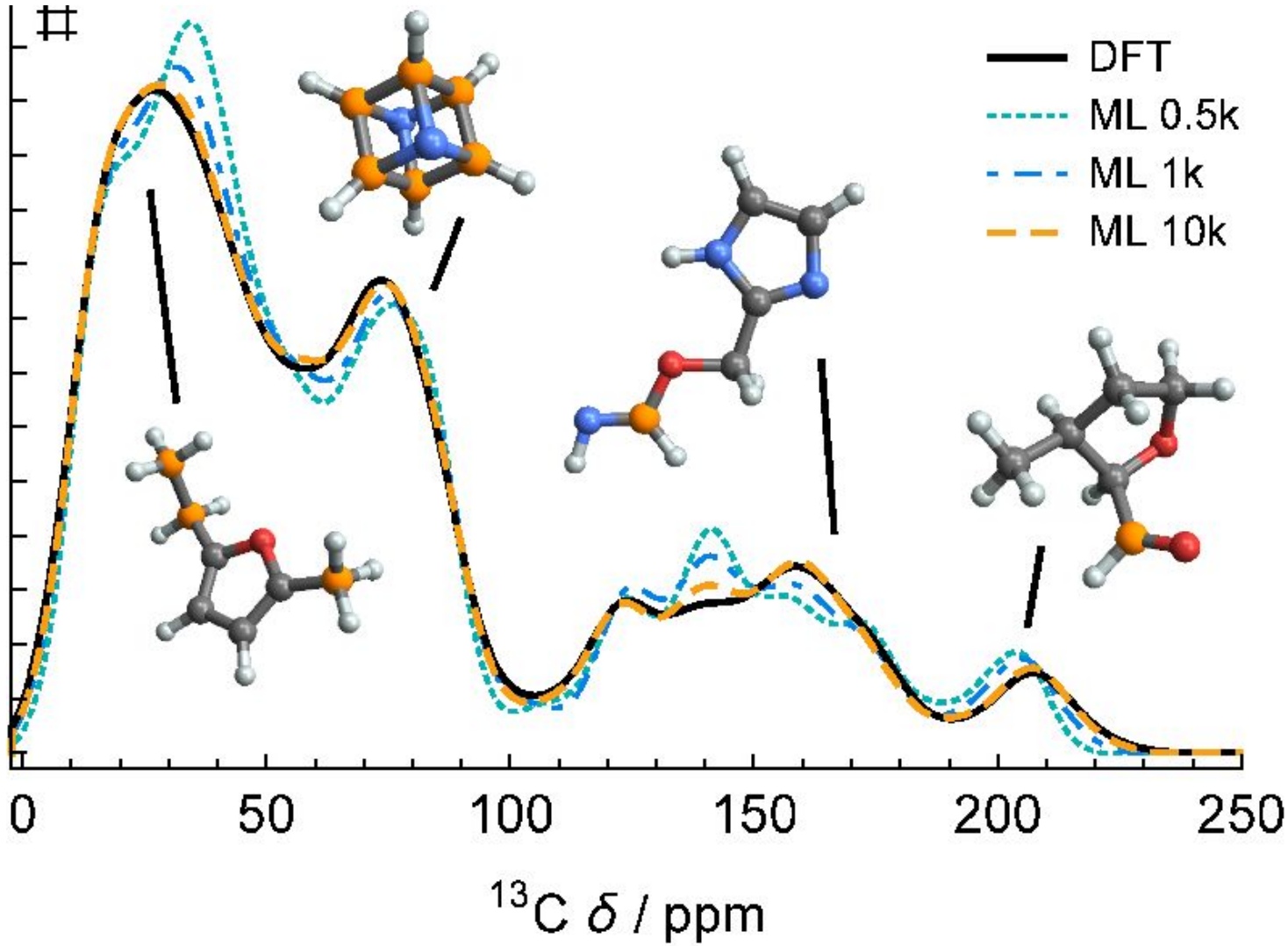
An Atom in Many Molecules

$$\langle \Psi | \partial_{\boldsymbol{R}_Q} \hat{H} | \Psi \rangle = \int \mathrm{d}\boldsymbol{r}\, (\boldsymbol{r} - \boldsymbol{R}_Q) Z_Q\, n(\boldsymbol{r}) / \|\boldsymbol{r} - \boldsymbol{R}_Q\|^3$$



M. Rupp, R. Ramakrishnan, OAvL, *JPCL* (2015), arXiv

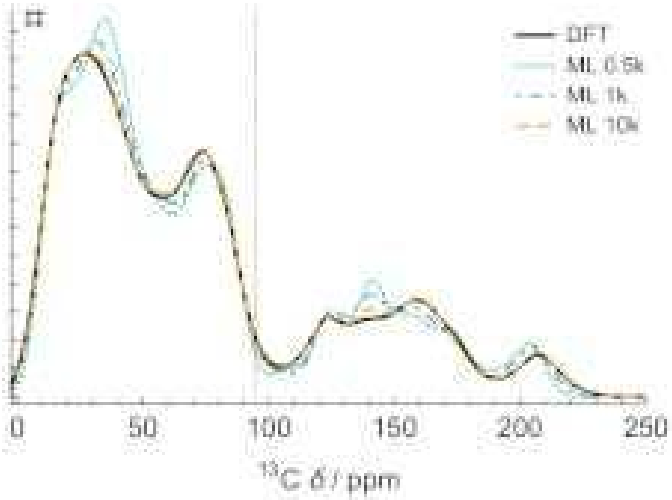
An Atom in Many Molecules

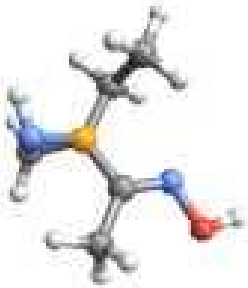


M. Rupp, R. Ramakrishnan, OAvL, *JPCL* (2015), arXiv

An Atom in Many Molecules

[redacted]

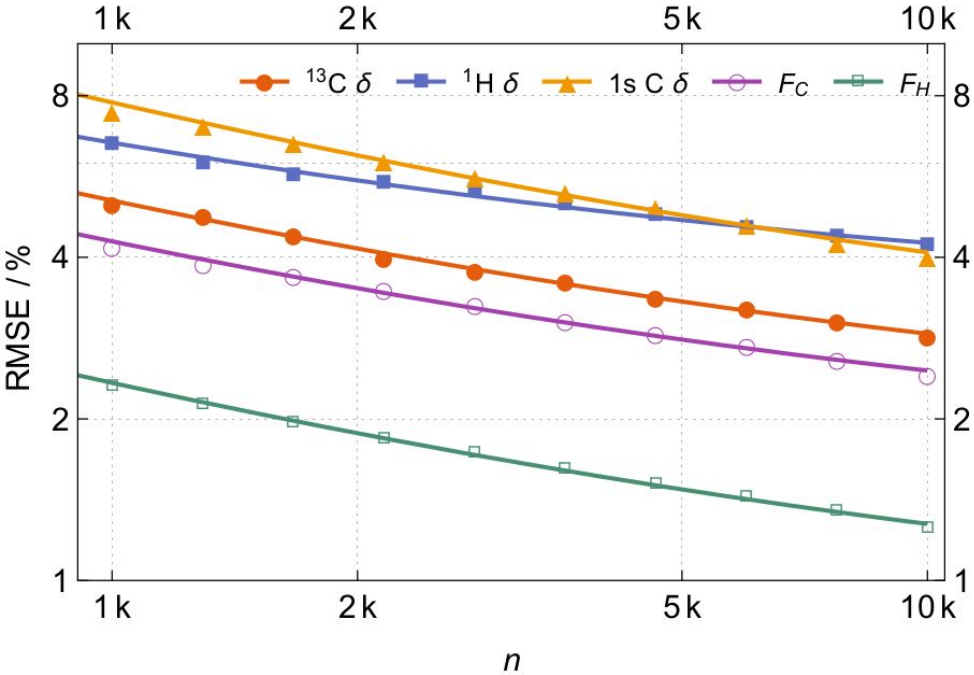




[redacted]

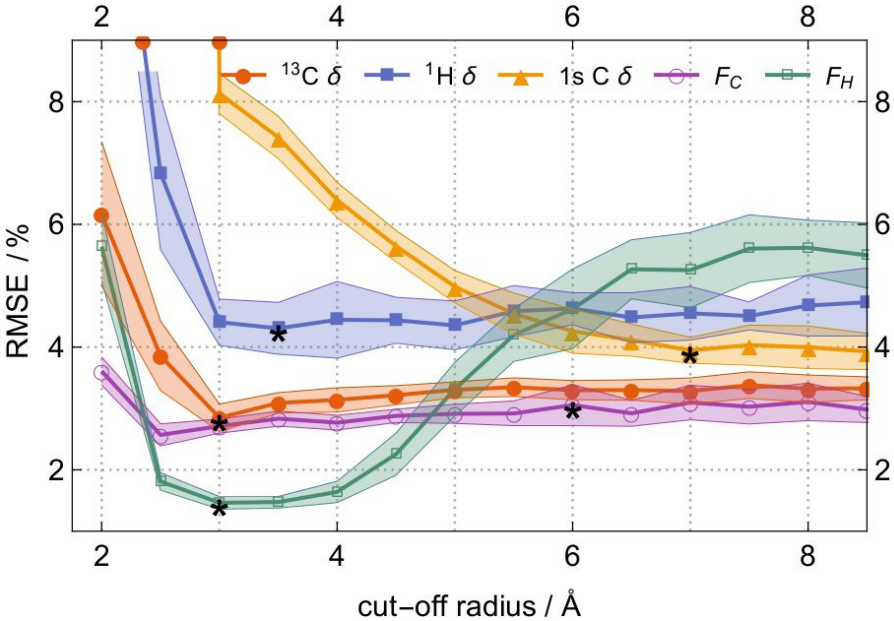
M. Rupp, R. Ramakrishnan, OAvL, *JPCL* (2015), arXiv

An Atom in Many Molecules



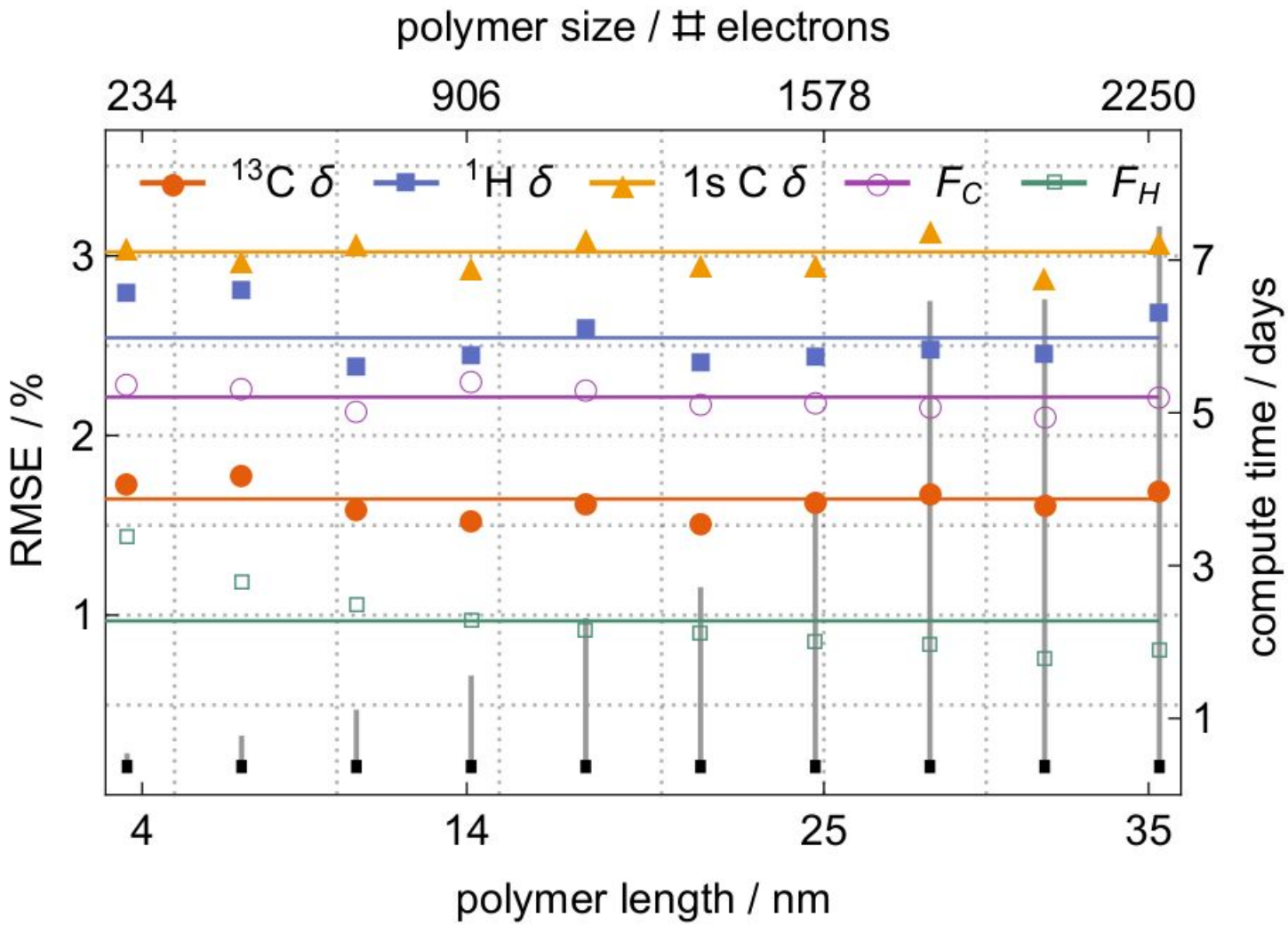
10k atoms from:
-16800 distortions from 168 $C_7H_{10}O_2$ isomers
-9k GDB molecules with 7 to 9 atoms CONF/molecule

For 10k models



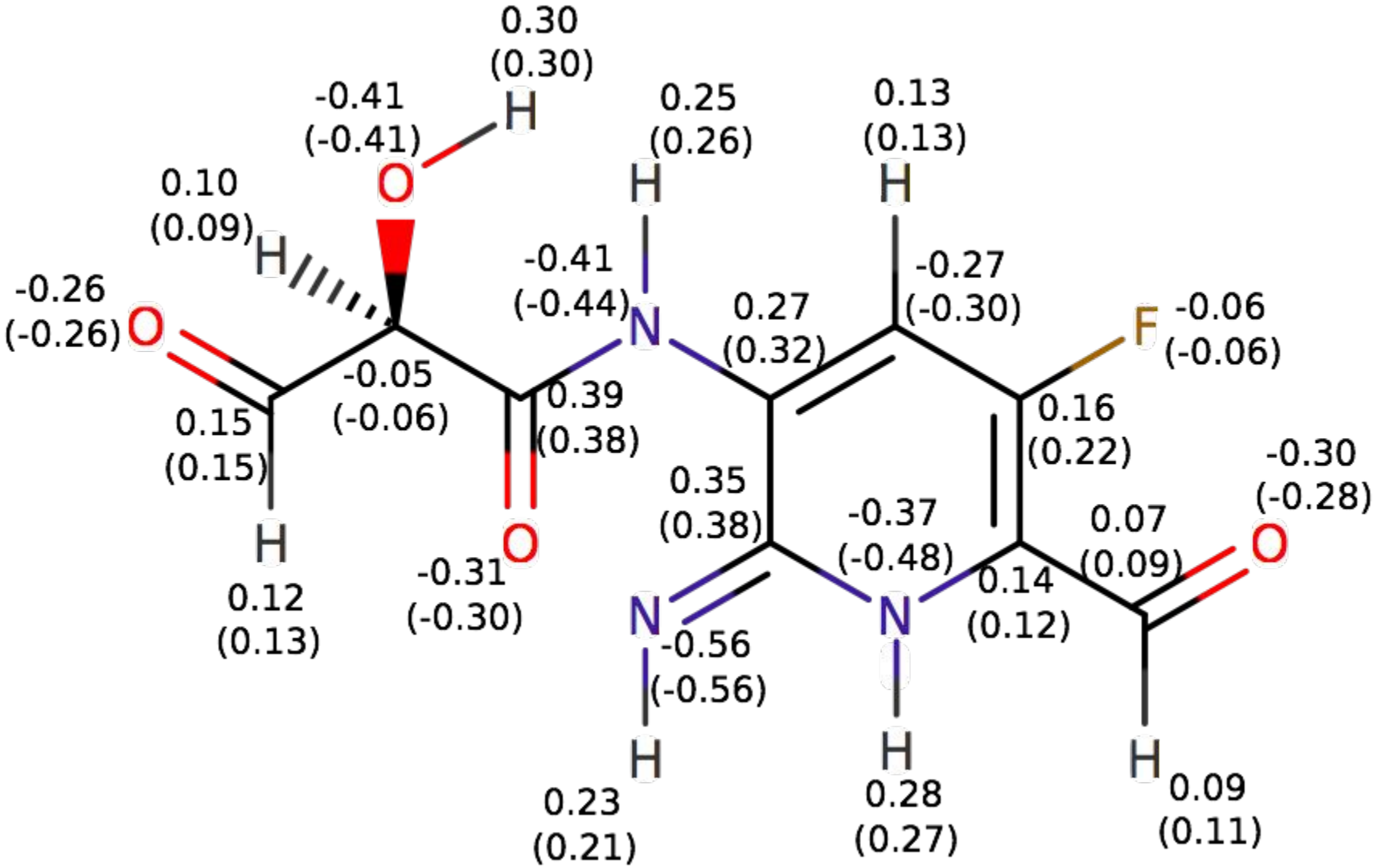
M. Rupp, R. Ramakrishnan, OAvL, *JPCL* (2015), arXiv

An Atom in Many Molecules

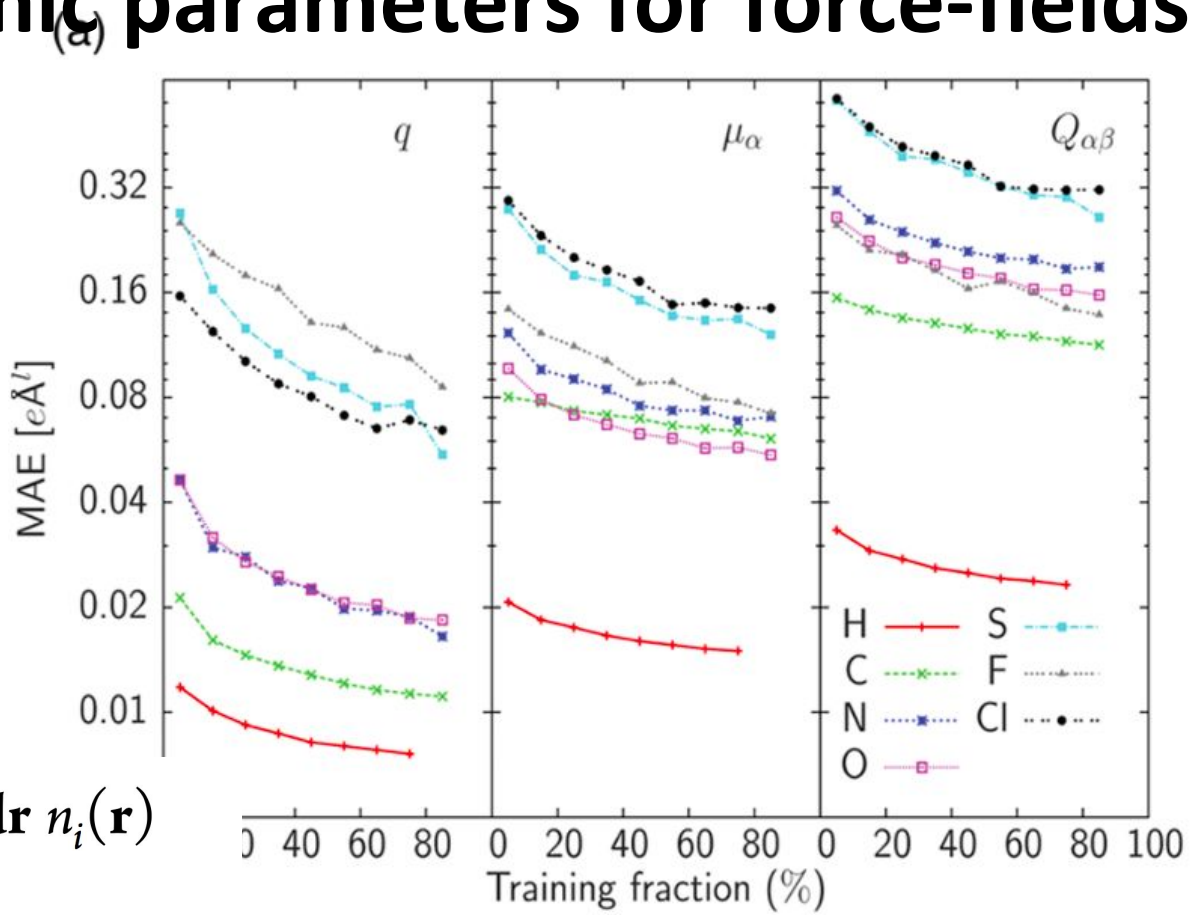


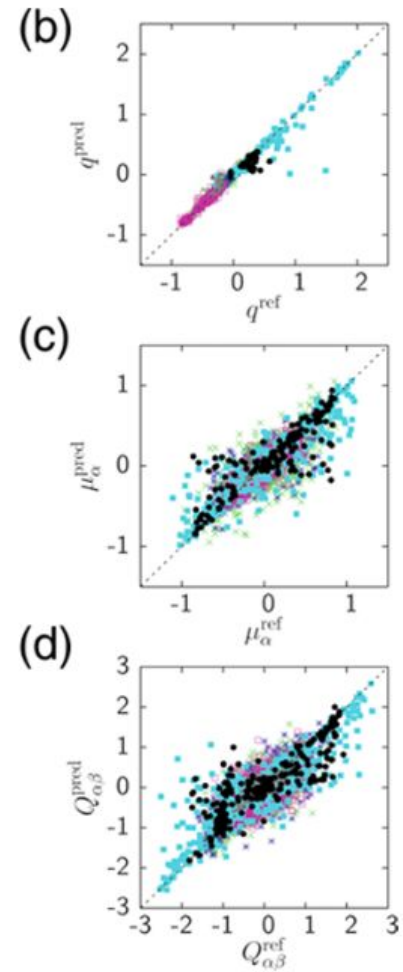
M. Rupp, R. Ramakrishnan, OAvL, *JPCL* (2015), arXiv

Atomic charges



Atomic parameters for force-fields

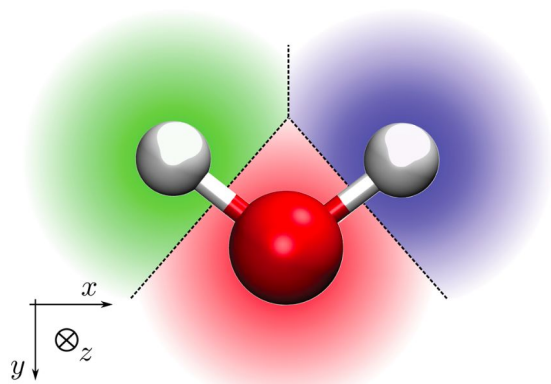




$$q^{(i)} = \int \mathrm{d}\mathbf{r}\, n_i(\mathbf{r})$$

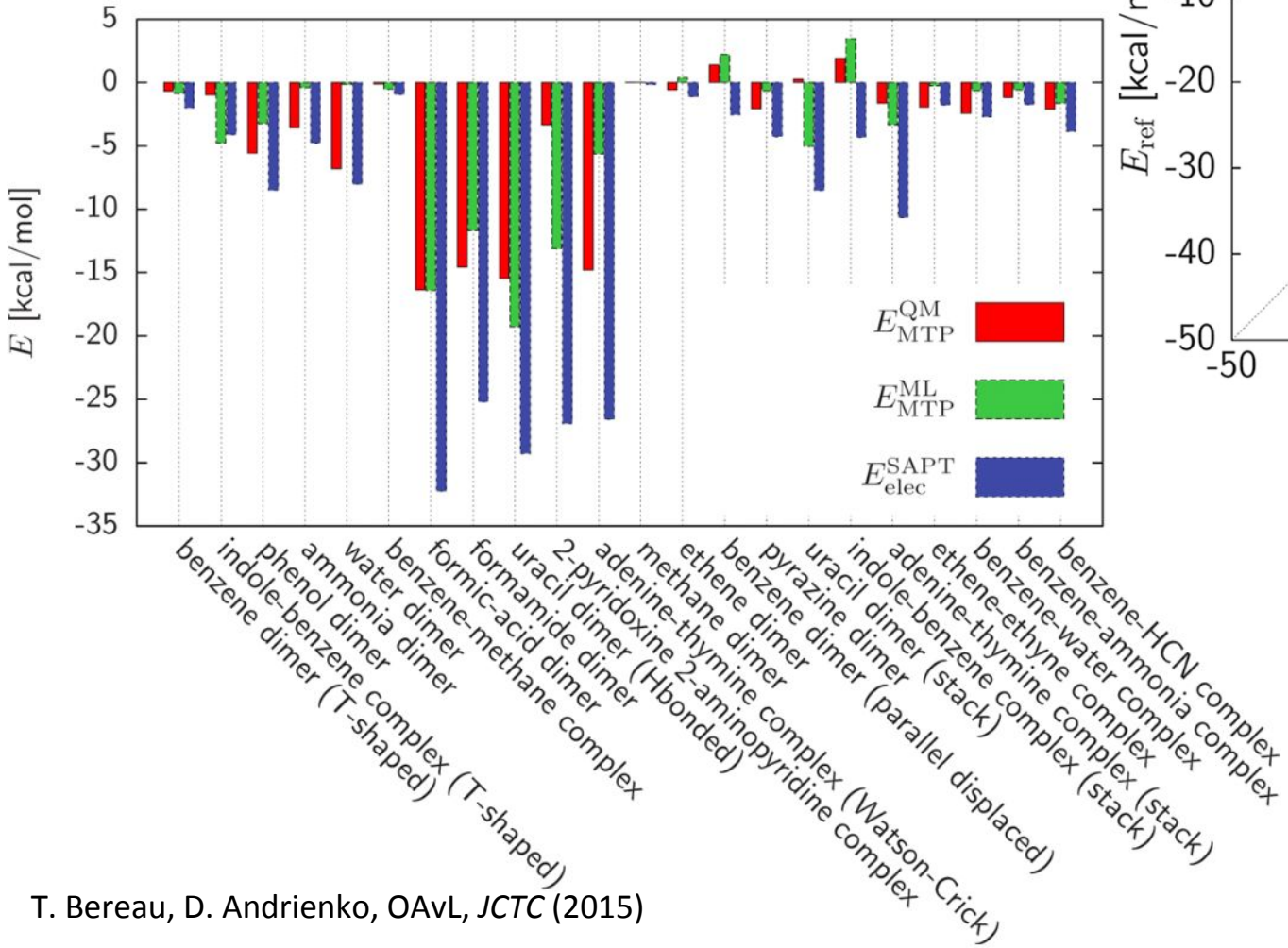
$$\mu_\alpha^{(i)} = \int \mathrm{d}\mathbf{r}\, n_i(\mathbf{r}) r_\alpha$$

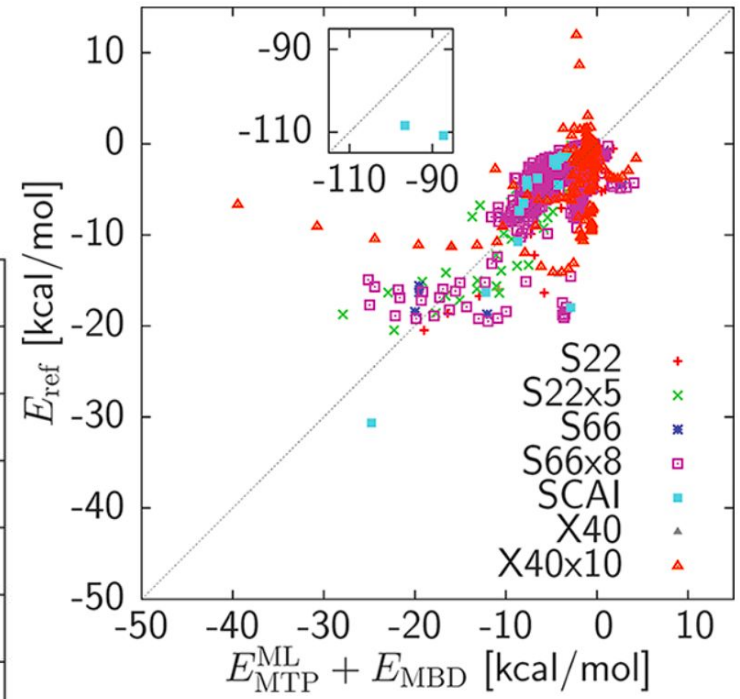
$$Q_{\alpha\beta}^{(i)} = \int \mathrm{d}\mathbf{r}\, n_i(\mathbf{r}) r_\alpha r_\beta$$



T. Bereau, D. Andrienko, OAvL, *JCTC* (2015)

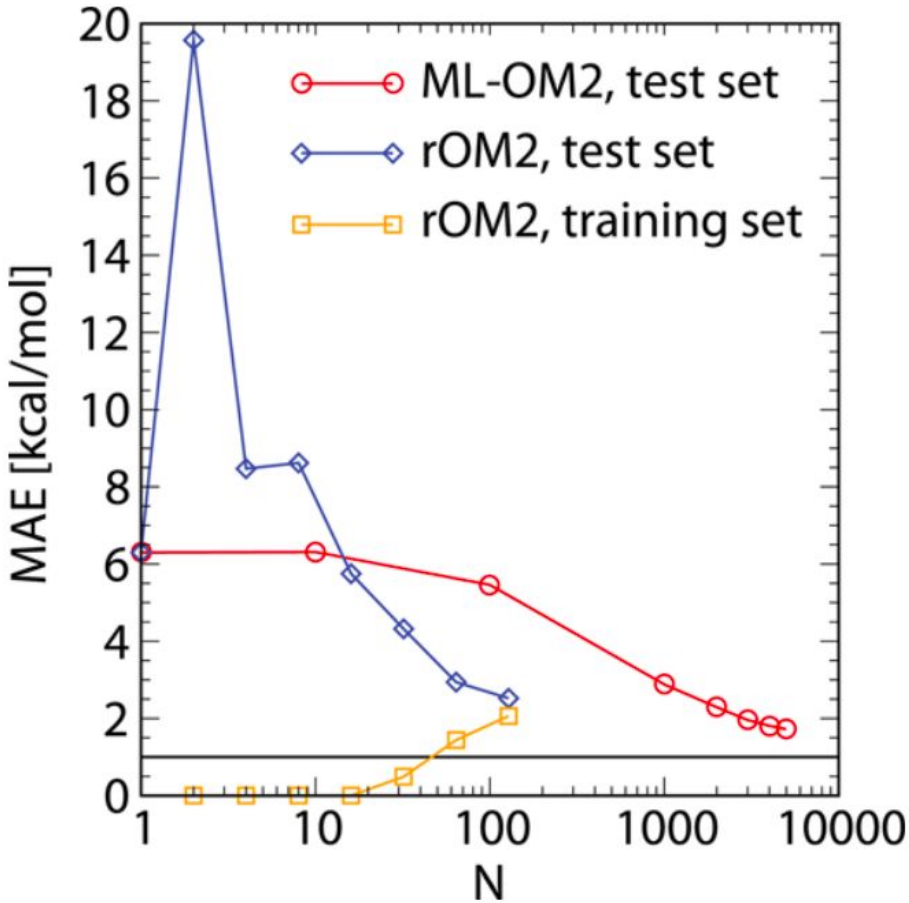
Atomic parameters for force-fields

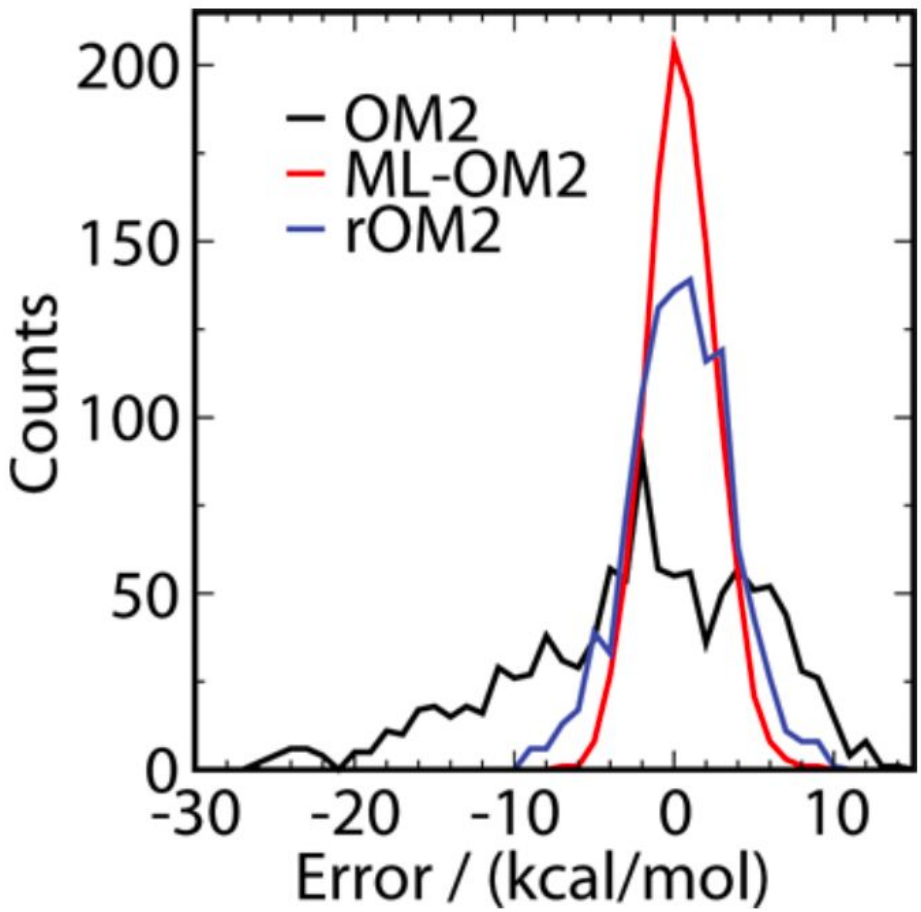




T. Bereau, D. Andrienko, OAvL, *JCTC* (2015)

Atomic parameters in OM2





$$\Delta P(M) = \sum_{i=1}^{N_{\text{train}}} \alpha_i \, \mathrm{e}^{-\|M-M_i\|_1/\sigma}$$

OM2 orbital exponent for Carbon ("screening") selected due to high sensitivity

P. Dral, OAvL, W. Thiel, *JCTC* (2015)

Conclusions

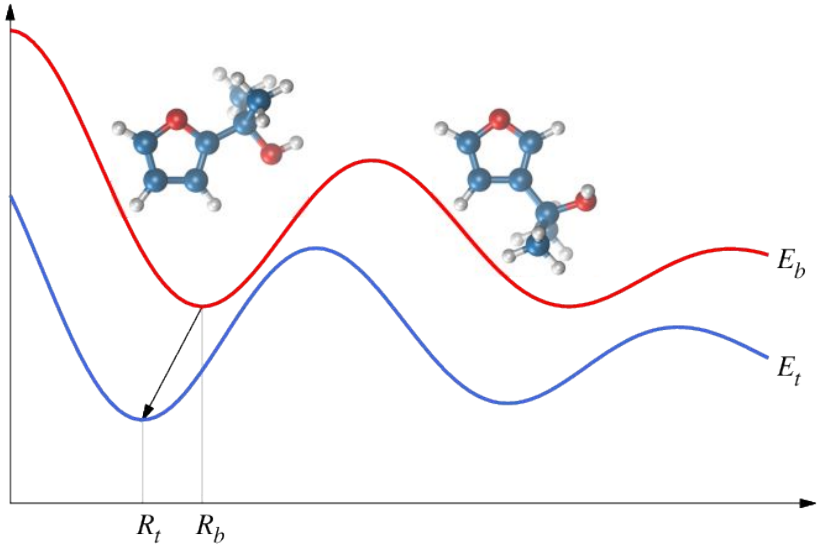
1. Parameter learning

2. Δ-ML (prior expert knowledge about property)

3. Crystals

Δ-ML

$$E_t(R_t) \approx E_b(R_b) + \Delta_b^t(R_b)$$



JCTC Journal of Chemical Theory and Computation
Article
pubs.acs.org/JCTC

Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach

Raghunathan Ramakrishnan,[†] Pavlo O. Dral,[¶,‡] Matthias Rupp,[†] and O. Anatole von Lilienfeld*[,†,§]

[†]Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials, Department of Chemistry, University of Basel, Klingelbergstraße 80, CH-4056 Basel, Switzerland

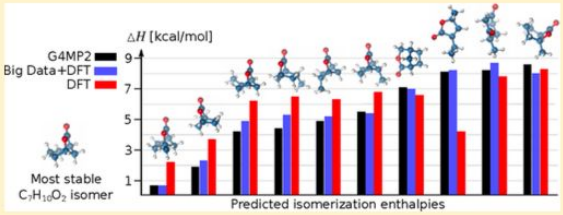
[¶]Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany

[‡]Computer-Chemie-Centrum and Interdisciplinary Center for Molecular Materials, Department Chemie und Pharmazie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstraße 25, 91052 Erlangen, Germany

[§]Argonne Leadership Computing Facility, Argonne National Laboratory, 9700 S. Cass Avenue, Lemont, Illinois 60439, United States

Supporting Information

ABSTRACT: Chemically accurate and comprehensive studies of the virtual space of all possible molecules are severely limited by the computational cost of quantum chemistry. We introduce a composite strategy that adds machine learning corrections to computationally inexpensive approximate legacy quantum methods. After training, highly accurate predictions of enthalpies, free energies, entropies, and electron correlation energies are possible, for significantly larger molecular sets than used for training. For thermochemical properties of up to 16k isomers of $C_7H_{10}O_2$ we present numerical evidence that chemical accuracy can be reached. We also predict electron correlation energy in post Hartree−Fock methods, at the computational cost of Hartree−Fock, and we establish a qualitative relationship between molecular entropy and electron correlation. The transferability of our approach is demonstrated, using

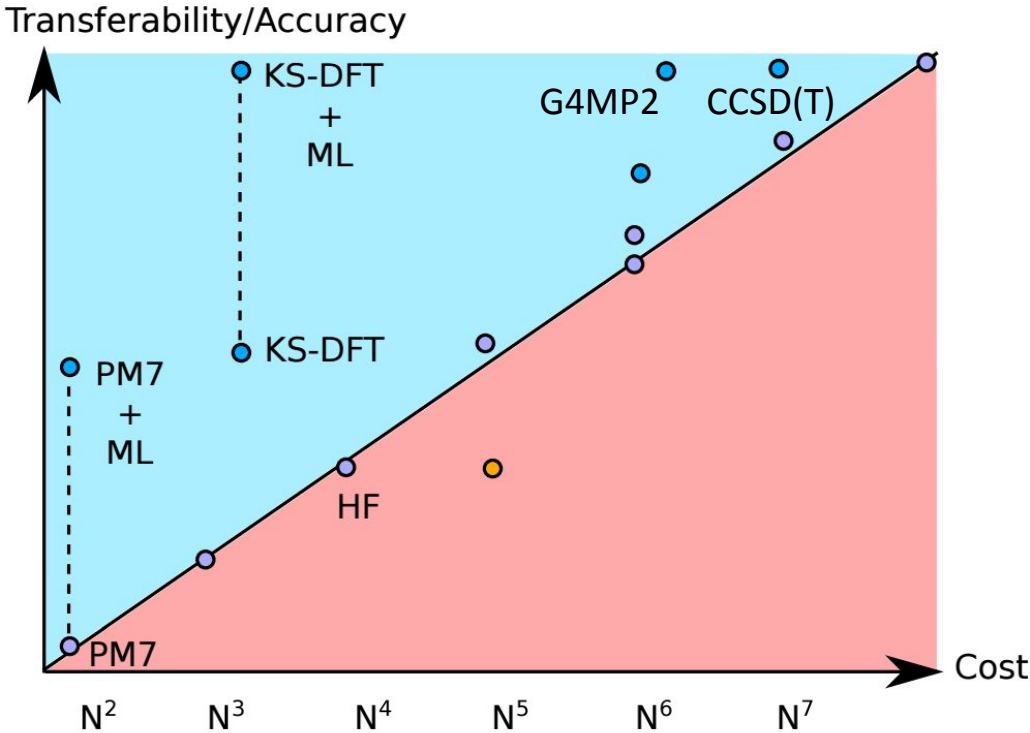


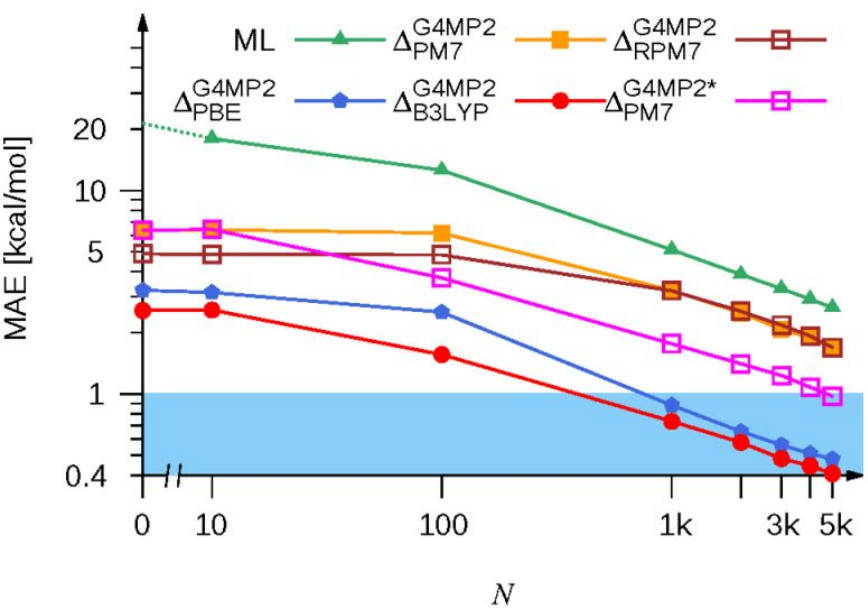
Ramakrishnan et al, *J Chem Theory Comput* (2015)

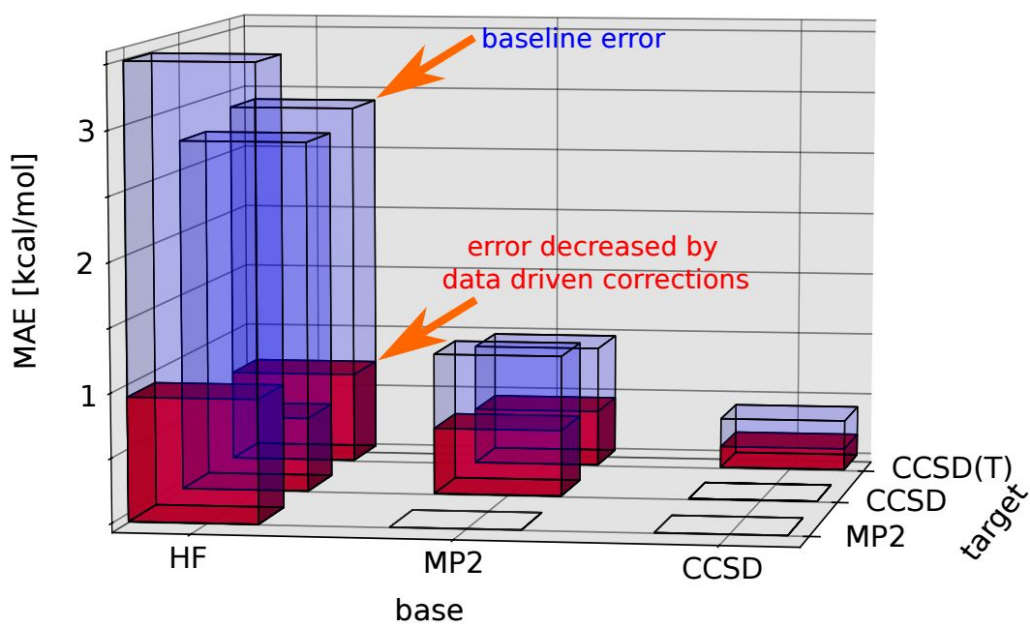
Δ-ML

$$E_t(R_t) \approx E_b(R_b) + \Delta_b^t(R_b)$$

6k constitutional isomers of $C_7O_2H_{10}$





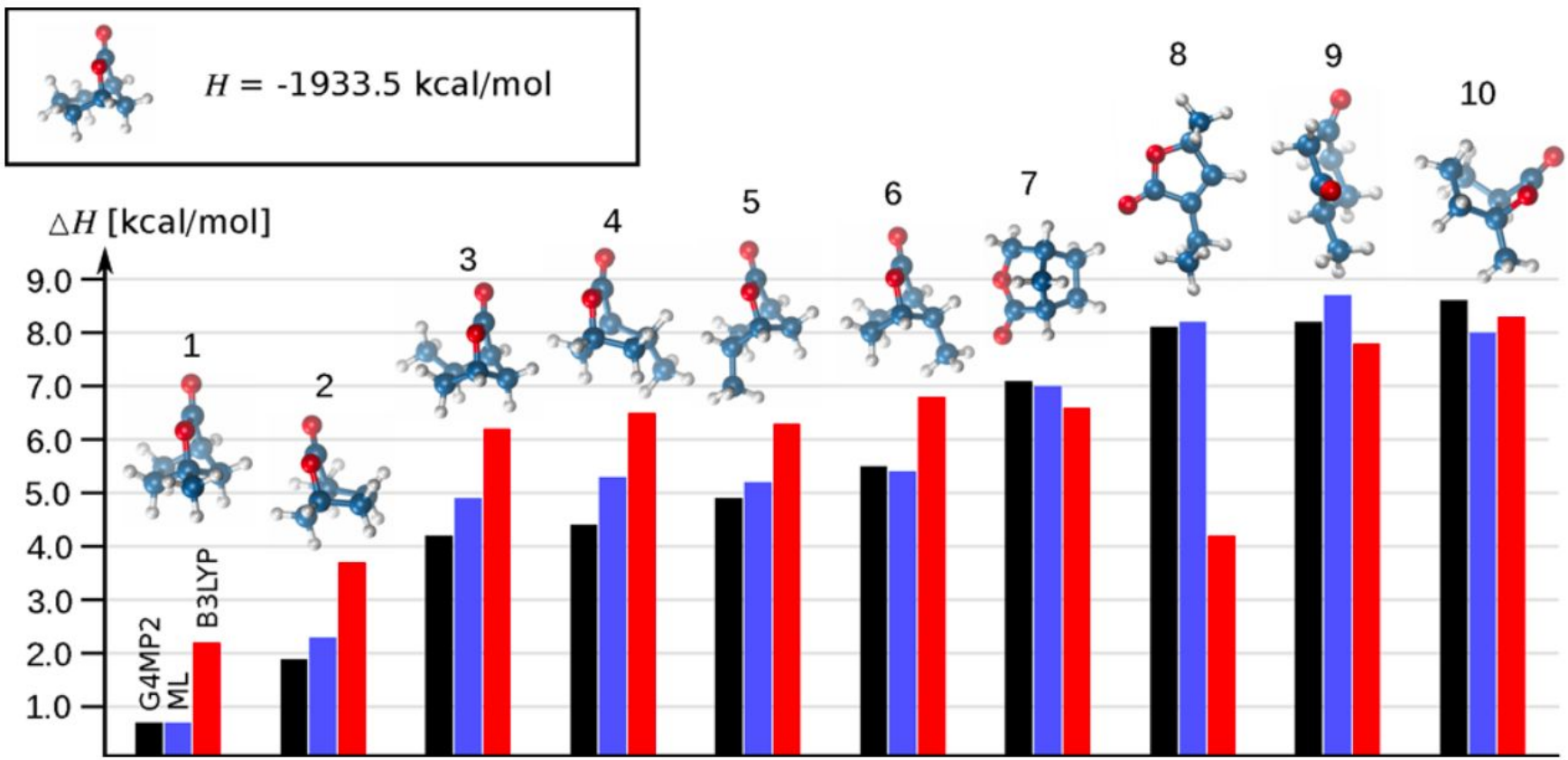


Ramakrishnan et al, *J Chem Theory Comput* (2015)

Δ-ML

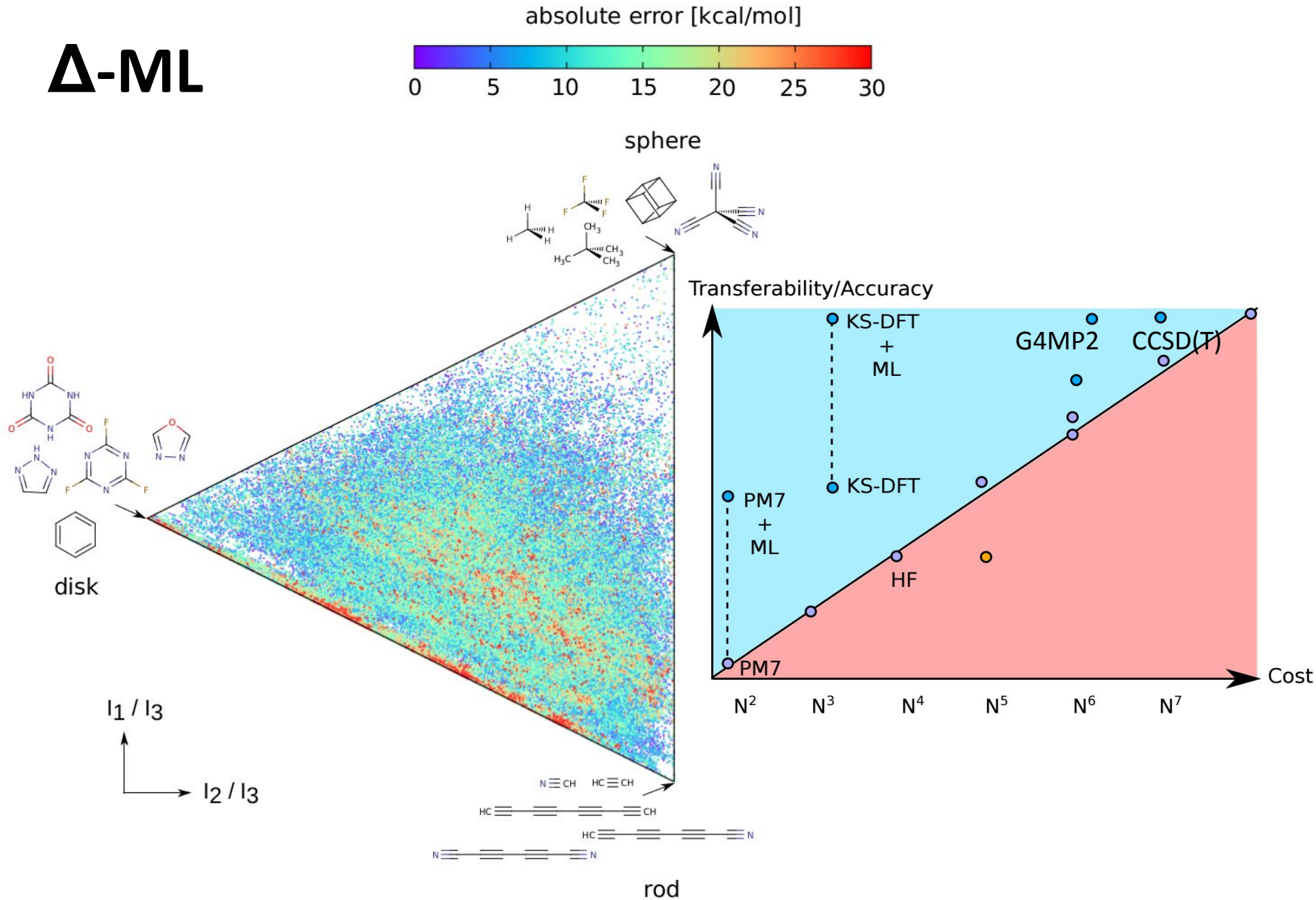
$$E_t(R_t) \approx E_b(R_b) + \Delta_b^t(R_b)$$

Ranking 10k diastereomers derived from 6k constitutional isomers of $C_7O_2H_{10}$
→ Global minimum, and its 10 closest isomers …



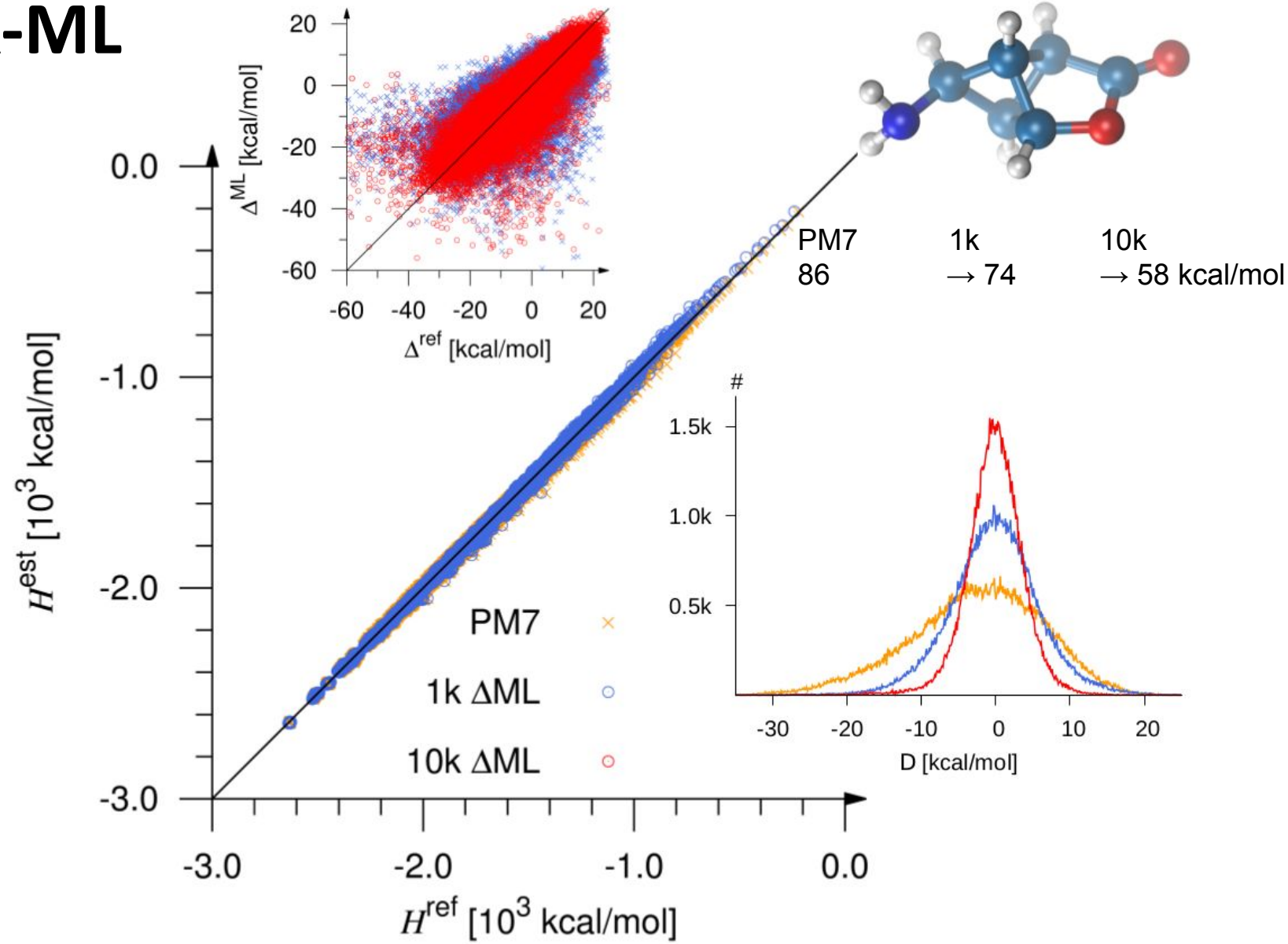
Ramakrishnan et al, *J Chem Theory Comput* (2015)

Δ-ML



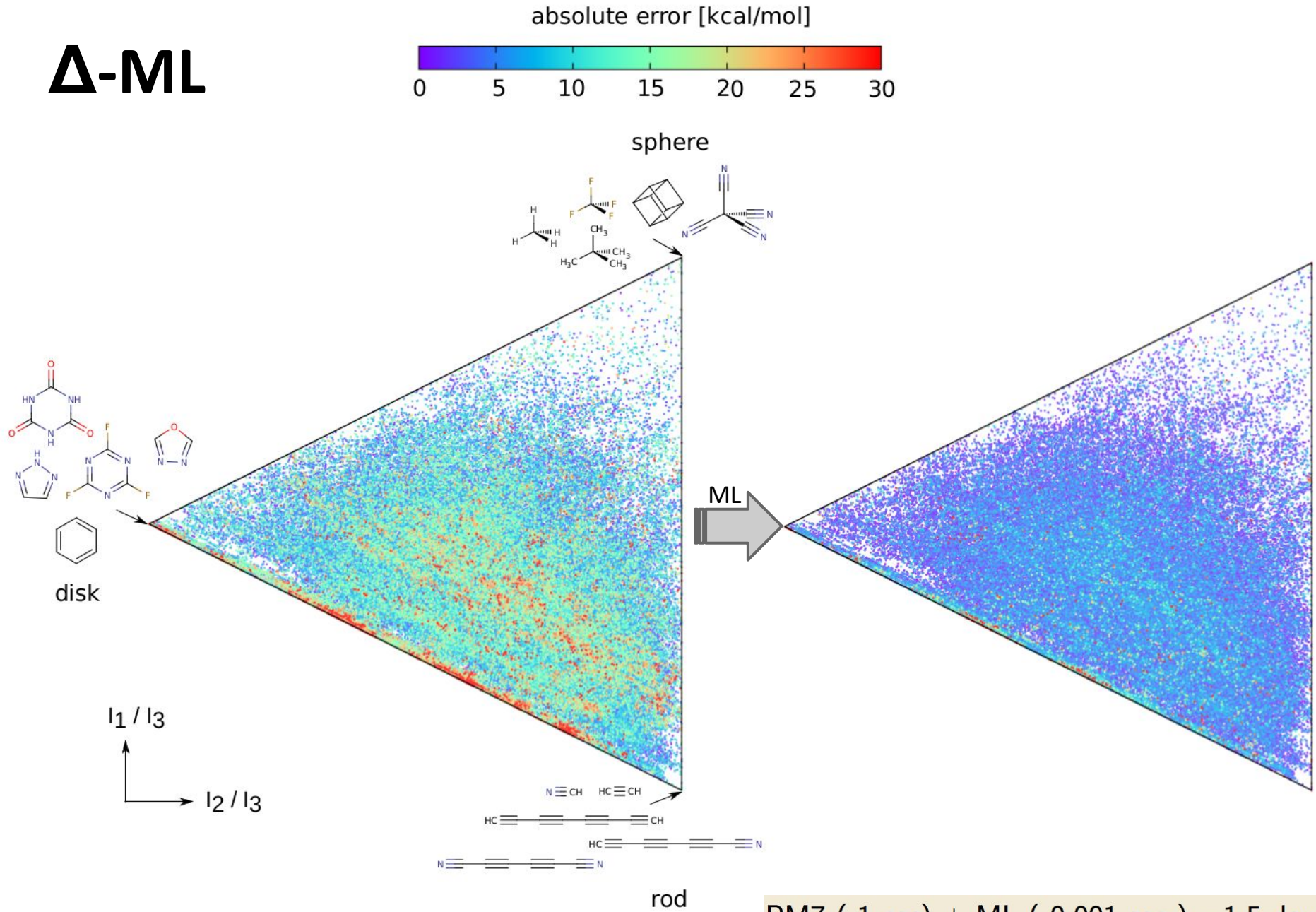
R Ramakrishnan et al *JCTC* (2015)

Δ-ML



R Ramakrishnan et al *JCTC* (2015)

Δ-ML



PM7 (1 sec) + ML (0.001 sec) – 1.5 days

DFT (30 min) – 8 yrs !

R Ramakrishnan et al *JCTC* (2015)

Conclusions

1. Parameter learning

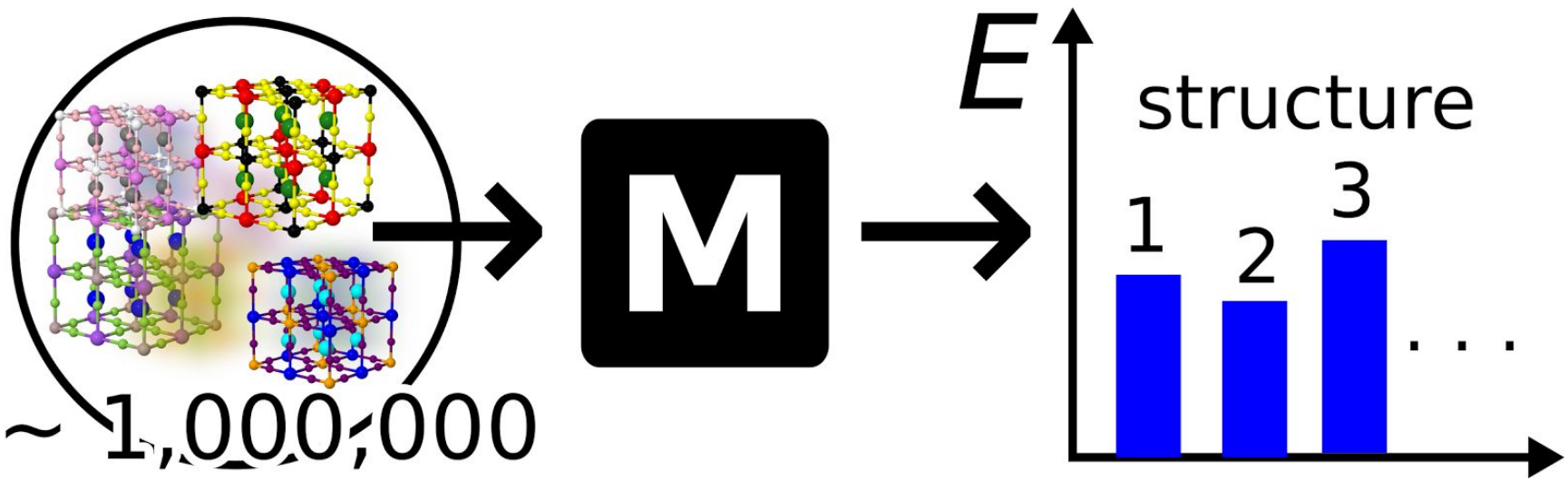
2. Δ-ML (prior expert knowledge about property)

3. Crystals

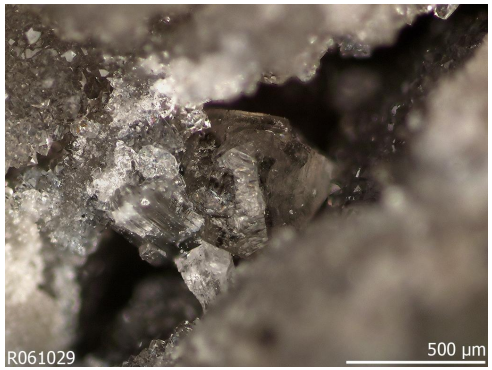
PRL 117, 135502 (2016) PHYSICAL REVIEW LETTERS week ending 23 SEPTEMBER 2016

Machine Learning Energies of 2 Million Elpasolite (ABC_2D_6) Crystals

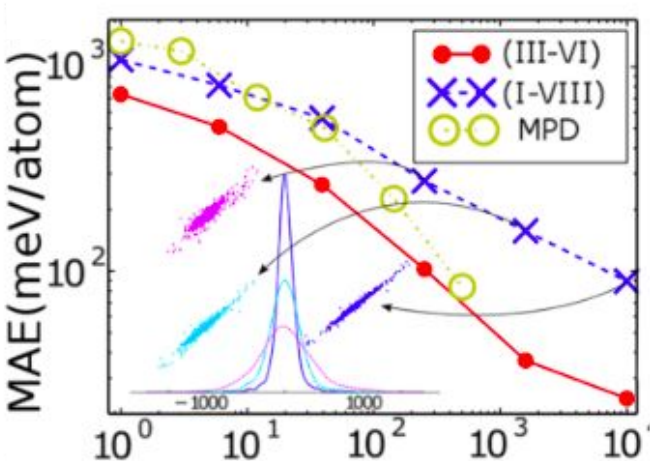
Felix A. Faber,[1] Alexander Lindmaa,[2] O. Anatole von Lilienfeld,[1,3,*] and Rickard Armiento[2,†]
[1]*Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials, Department of Chemistry, University of Basel, 4056 Basel, Switzerland*
[2]*Department of Physics, Chemistry and Biology, Linköping University, SE-581 83 Linköping, Sweden*
[3]*General Chemistry, Free University of Brussels, Pleinlaan 2, 1050 Brussels, Belgium*



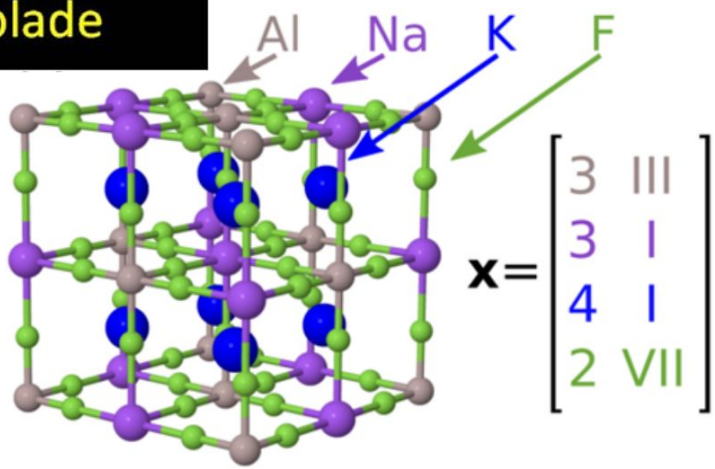
Crystals



- most abundant quaternary crystal structure in Inorganic Crystal Structure Database
- can emit light when exposed to ionic radiation (→ Scintillator candidates)
- transparent, glossy, colorless and soft crystal in Fm3m space group
- $AlNaK_2F_6$ found in Rocky Mountains, Virginia, or the Apennines

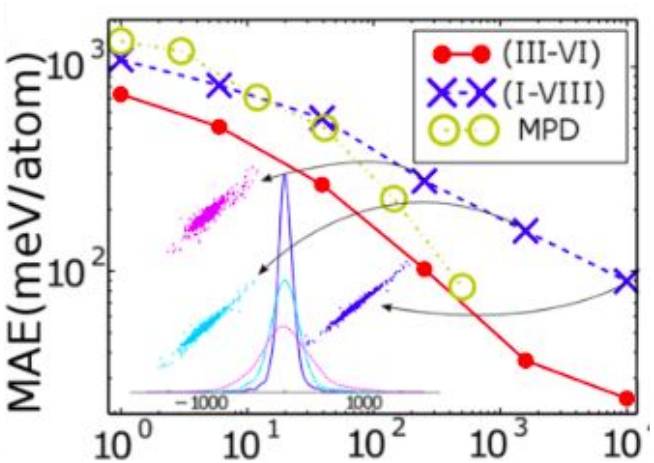


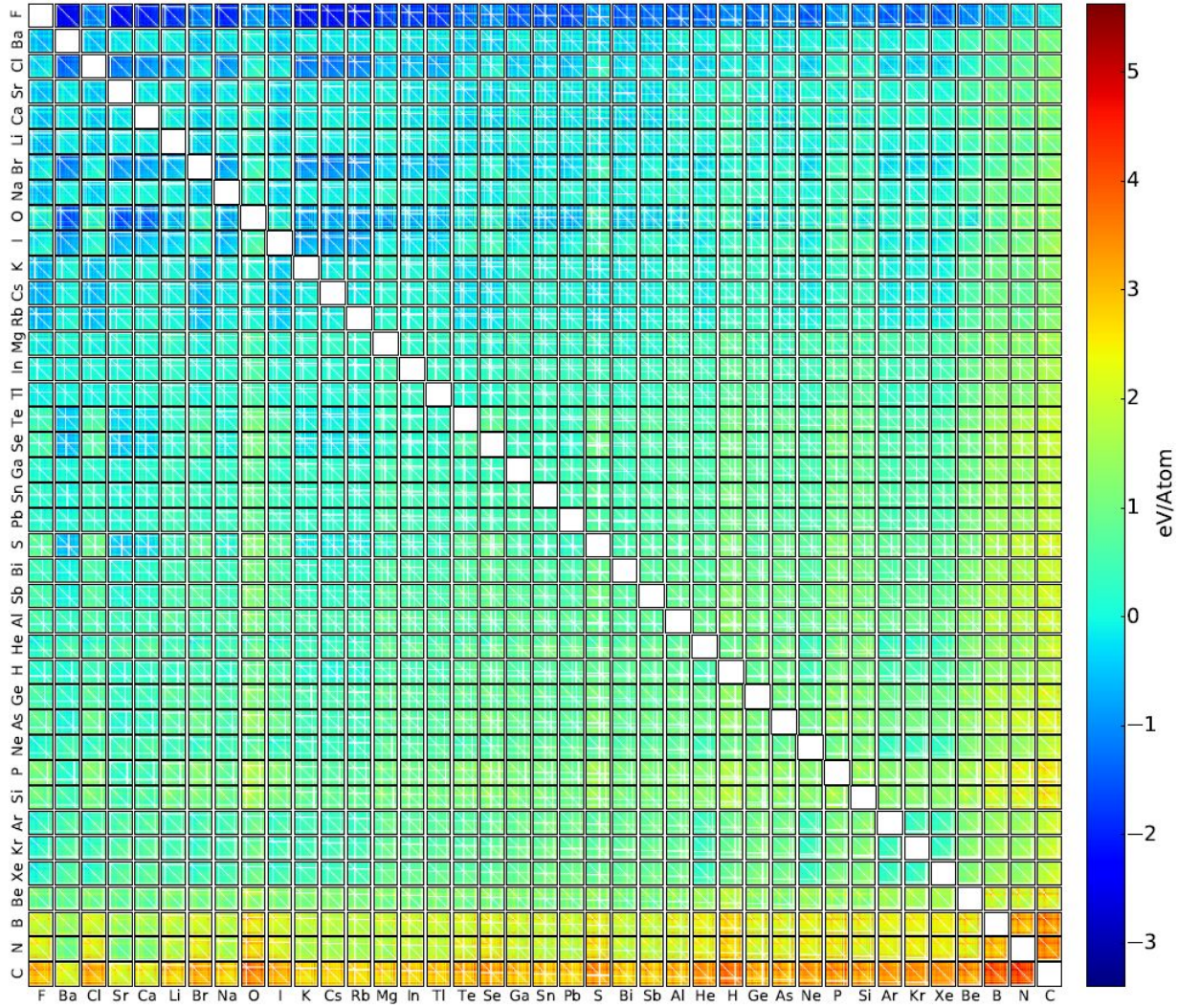
$$P^{\text{est}}(\mathbf{M}) = \sum_i \alpha_i k(\mathbf{M}, \mathbf{M}_i)$$
$$\vec{\alpha} = \mathbf{K}^{-1} \vec{P}^{\text{ref}}$$



Faber et al, *Phys Rev Lett* (2016) arxiv.org/abs/1508.05315

Crystals





Reduction in cost:
DFT: ~20 M CPU hours (optimized)
ML: ~20 CPU hours (not optimized)

Faber et al, *Phys Rev Lett* (2016) arxiv.org/abs/1508.05315

Crystals

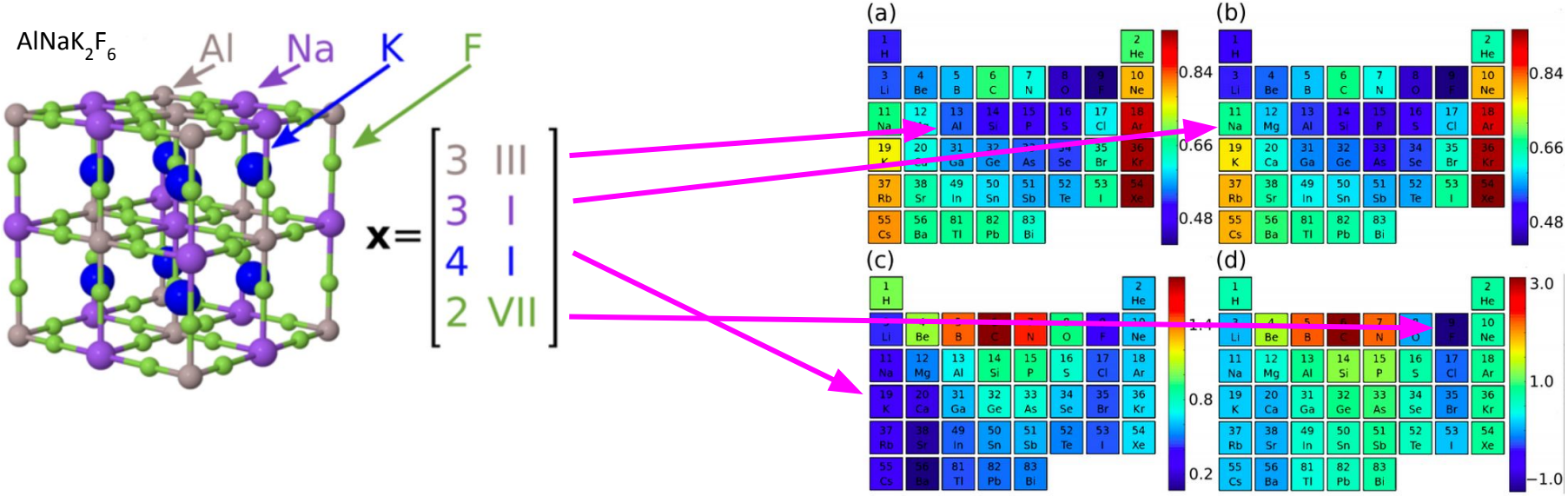
Discovery of ~90 new crystals!!!

TABLE VI – *Continued from previous page*

#	Elpasolite x_1	x_2	x_3	x_4	ΔE (eV/atom)	Type	$\Delta\epsilon$ (eV)	MP ID	Competing mixed phase in MP
50	Cl	N	Tl	Ba	−0.0312	C	0.00	mp-989542	$0.5\,Ba_6N_2 + 0.5\,Ba_1Cl_2 + 0.5\,Ba_2Tl_4 + 1.5\,Ba_1$
51	Tl	K	Cs	F	−0.0329	I	3.82	mp-989526	$0.6667\,Cs_3Tl_1F_6 + 0.0417\,K_{24}Tl_8F_{48}$
52	N	O	Sn	Sr	−0.0333	C	0.00	mp-989540	$Sr_2N_1 + Sr_3Sn_1O_1 + 0.5\,Sr_2Sn_2$
53	S	F	Cs	Cl	−0.0336	C	0.00	mp-989521	$0.1458\,Cl_{16} + 0.1667\,S_1F_6 + 0.1042\,S_8Cl_{16} + 2\,Cs_1Cl_1$
54	Se	Cl	Cs	F	−0.0349	C	0.03	mp-989544	$0.5\,Cs_1Cl_1 + 0.125\,Cl_4F_4 + 1.5\,Cs_1F_1 + 0.25\,Se_4F_{16}$
55	Ga	K	Cs	F	−0.0359	I	6.04	mp-989531	$0.5\,Cs_1F_1 + 0.0833\,Cs_{18}Ga_{12}F_{54} + K_1F_1$
56	Pb	Rb	Cs	F	−0.0370	C	0.00	mp-989525	$2\,Cs_1F_1 + 0.25\,Pb_4F_{12} + Rb_1F_1$
57	F	Br	Rb	Cl	−0.0382	S	0.92	mp-989573	$0.5\,Cl_4 + 0.5\,Br_2Cl_2 + 2\,Rb_1Cl_1 + 0.25\,Cl_4F_4$
58	N	Rb	Cs	F	−0.0384	S	2.96	mp-989519	$0.5\,Rb_2F_6 + 0.125\,N_8 + 2\,Cs_1F_1 + 0.25\,F_4$
59	N	Li	Na	F	−0.0396	S	2.72	mp-989504	$0.375\,F_8 + 0.5\,N_2 + Li_1F_1 + 2\,Na_1F_1$
60	Bi	Na	Rb	Cl	−0.0416	I	3.73	mp-989520	$0.25\,Bi_4Cl_{12} + 2\,Rb_1Cl_1 + Na_1Cl_1$
61	Na	Mg	Cs	F	−0.0449	C	0.00	mp-989568	$0.25\,F_4 + 0.6667\,Cs_1F_1 + 0.1667\,Cs_8Mg_6F_{20} + Na_1F_1$
62	Tl	Al	Rb	H	−0.0451	S	0.72	mp-989539	$H_1 + 0.5\,Al_2H_6 + Tl_1 + 2\,Rb_1H_1$
63	S	Br	Rb	Cl	−0.0452	C	0.00	mp-989518	$0.25\,Cl_4 + 0.5\,Br_2Cl_2 + 2\,Rb_1Cl_1 + 0.125\,S_8Cl_{16}$
64	N	F	Sn	Sr	−0.0454	C	0.00	mp-989592	$0.5\,Sr_8Sn_4 + 0.0625\,Sr_{32}N_{16}F_{16}$
65	As	Na	Rb	F	−0.0467	I	4.55	mp-989523	$2\,Rb_1F_1 + 0.25\,As_4F_{12} + Na_1F_1$
66	Pb	K	Cs	F	−0.0486	C	0.00	mp-989585	$2\,Cs_1F_1 + 0.25\,Pb_4F_{12} + K_1F_1$
67	In	Al	Cs	H	−0.0488	S	0.61	mp-989535	$0.5\,Al_2H_6 + 0.1111\,Cs_3In_9 + 1.6667\,Cs_1H_1 + 1.3333\,H_1$
68	Pb	Na	Cs	F	−0.0513	C	0.00	mp-989556	$2\,Cs_1F_1 + 0.25\,Pb_4F_{12} + Na_1F_1$
69	Ga	Na	Tl	F	−0.0514	I	4.39	mp-989561	$0.2\,Ga_2F_6 + 0.5\,Tl_4F_4 + 0.1\,Na_{10}Ga_6F_{28}$
70	In	Tl	Rb	Cl	−0.0528	S	2.40	mp-989550	$0.5\,In_2Cl_6 + Tl_1Cl_1 + 2\,Rb_1Cl_1$
71	In	Na	Rb	Cl	−0.0538	I	3.05	mp-989547	$0.3333\,In_2Cl_6 + 0.1667\,Na_6In_2Cl_{12} + 2\,Rb_1Cl_1$
72	In	Na	Cs	H	−0.0565	S	1.26	mp-989610	$3.3333\,H_1 + 0.1111\,Cs_3In_9 + 1.6667\,Cs_1H_1 + Na_1H_1$
73	Li	In	Rb	Cl	−0.0603	S	2.83	mp-989583	$0.0556\,Li_{18}In_6Cl_{36} + 2\,Rb_1Cl_1 + 0.3333\,In_2Cl_6$
74	In	Na	Tl	F	−0.0604	I	4.27	mp-989533	$Na_1F_1 + 0.5\,Tl_4F_4 + 0.5\,In_2F_6$
75	Br	F	Cs	Cl	−0.0610	S	0.95	mp-989543	$0.5\,Cl_4 + 2\,Cs_1Cl_1 + 0.5\,Br_2Cl_2 + 0.25\,Cl_4F_4$
76	In	Ga	Rb	F	−0.0616	I	3.27	mp-989566	$0.1667\,Ga_4 + 0.0833\,Rb_8In_{12}F_{44} + 0.1667\,Ga_2F_6 + 1.3333\,Rb_1F_1$
77	N	K	Cs	F	−0.0618	I	-	mp-989580	$0.25\,F_4 + 0.125\,K_8F_{24} + 0.0625\,N_{16} + 2\,Cs_1F_1$
78	Li	Na	Cs	F	−0.0634	C	0.00	mp-989559	$0.5\,F_4 + Cs_1F_1 + 0.25\,Cs_4Li_4F_8 + Na_1F_1$
79	Na	In	Rb	F	−0.0672	I	5.34	mp-989578	$0.0833\,Rb_8In_{12}F_{44} + Na_1F_1 + 1.3333\,Rb_1F_1$
80	O	N	Sn	Ca	−0.0686	C	0.00	mp-989584	$0.0625\,Ca_{24}N_{16} + Ca_3Sn_1O_1 + 0.0227\,Ca_{62}Sn_{40} + 0.0455\,Ca_2Sn_2$
81	N	F	Sn	Ca	−0.0738	S	0.14	mp-989590	$0.5\,Ca_4N_2F_2 + 0.5\,Ca_8Sn_4$
82	In	Rb	Cs	F	−0.0776	I	5.37	mp-989605	$0.0833\,Rb_8In_{12}F_{44} + 0.3333\,Rb_1F_1 + 2\,Cs_1F_1$
83	S	Br	Cs	Cl	−0.0848	C	0.00	mp-989517	$2\,Cs_1Cl_1 + 0.5\,Br_2Cl_2 + 0.25\,Cl_4 + 0.125\,S_8Cl_{16}$
84	In	K	Cs	F	−0.0875	I	5.46	mp-989639	$0.025\,K_{24}In_8F_{48} + 0.1\,K_4In_8F_{28} + 2\,Cs_1F_1$
85	Tl	Al	Cs	H	−0.0884	S	1.14	mp-989575	$2\,Cs_1H_1 + 0.5\,Al_2H_6 + Tl_1 + H_1$
86	Tl	Ga	Rb	F	−0.0945	I	4.40	mp-989565	$2\,Rb_1F_1 + 0.5\,Ga_2F_6 + 0.25\,Tl_4F_4$
87	Ga	Na	Rb	F	−0.1008	I	5.90	mp-989400	$2\,Rb_1F_1 + 0.1\,Na_{10}Ga_6F_{28} + 0.2\,Ga_2F_6$
88	Al	Na	Cs	H	−0.1019	S	2.14	mp-989642	$2\,Cs_1H_1 + 0.5\,Na_2Al_2H_8$
89	N	Na	Cs	F	−0.1064	S	2.80	mp-989527	$0.75\,F_4 + 0.5\,N_2 + Na_1F_1 + 2\,Cs_1F_1$

Faber et al, *Phys Rev Lett* (2016) arxiv.org/abs/1508.05315

Crystals

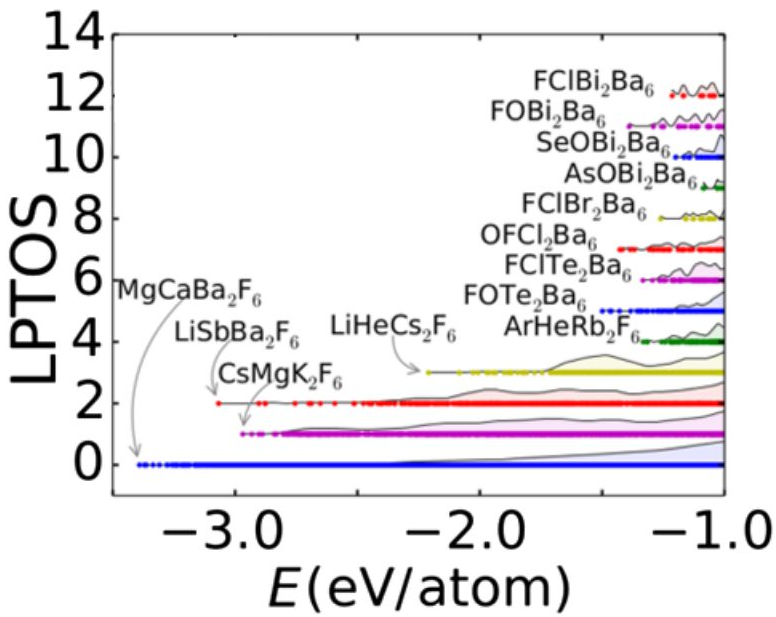


Faber et al, *Phys Rev Lett* (2016) arxiv.org/abs/1508.05315

Crystals

Element	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
H					✓		✓						
He						✓							
Li							✓						
Be							✓	✓					
B	✓						✓	✓	✓				
C		✓	✓	✓	✓	✓	✓	✓	✓	✓			
N	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓		
O				✓	✓		✓	✓					
F					✓								
Ne						✓							
Na					✓		✓						
Mg							✓	✓					
Al							✓	✓	✓				
Si		✓	✓	✓	✓		✓	✓	✓	✓			
P			✓	✓	✓		✓	✓	✓	✓	✓		
S				✓	✓		✓	✓	✓	✓	✓	✓	
Cl					✓		✓		✓	✓	✓	✓	✓
Ar						✓							
K					✓		✓						
Ca					✓		✓	✓					

Element	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8
Ga	✓	✓		✓			✓	✓	✓					
Ge		✓	✓	✓	✓	✓	✓	✓	✓	✓				
As			✓				✓	✓	✓		✓			
Se				✓			✓	✓		✓		✓		
Br					✓		✓		✓	✓	✓		✓	
Kr						✓	✓	✓						
Rb					✓		✓							
Sr							✓	✓						
In	✓						✓	✓	✓					
Sn		✓					✓	✓	✓	✓				
Sb			✓			✓	✓		✓		✓			
Te				✓		✓	✓	✓		✓	✓	✓		
I					✓		✓				✓		✓	
Xe							✓	✓		✓		✓		✓
Cs					✓		✓							
Ba								✓						
Ti	✓						✓	✓	✓					
Pb							✓	✓	✓	✓				
Bi							✓	✓	✓	✓	✓			



Lowest Possible Total Oxidation State (LPTOS)

Formula	LPTOS	E_{ML}	E_{DFT}	q_1	q_2	q_3	q_4
$MgSbBa_2F_6$	1	-2.88	-2.70	1.66	**0.42**	1.63	-0.89
$CaTeBa_2F_6$	1	-2.90	-2.68	1.58	**0.31**	1.67	-0.87
$TeCaBa_2F_6$	1	-2.83	-2.68	**0.31**	1.59	1.67	-0.87
$LiSbBa_2F_6$	2	-3.06	-2.62	0.89	**1.06**	1.62	-0.86
$CsMgRb_2F_6$	1	-2.93	-2.61	0.98	1.67	0.92	-0.75
$BeSbBa_2F_6$	2	-2.88	-2.60	1.68	**0.35**	1.62	-0.88
$CsMgK_2F_6$	1	-2.97	-2.58	1.01	1.68	0.92	-0.75
$SrSbBa_2F_6$	2	-2.90	-2.56	1.48	**0.60**	1.59	-0.88
$SrTeBa_2F_6$	2	-2.89	-2.55	1.70	**0.40**	1.66	-0.90

Faber et al, *Phys Rev Lett* (2016) arxiv.org/abs/1508.05315

Crystals

Calculated atomic charges in $NFAl_2\ Ca_6$ elpasolite using different methods (obtained using SIESTA[43]).

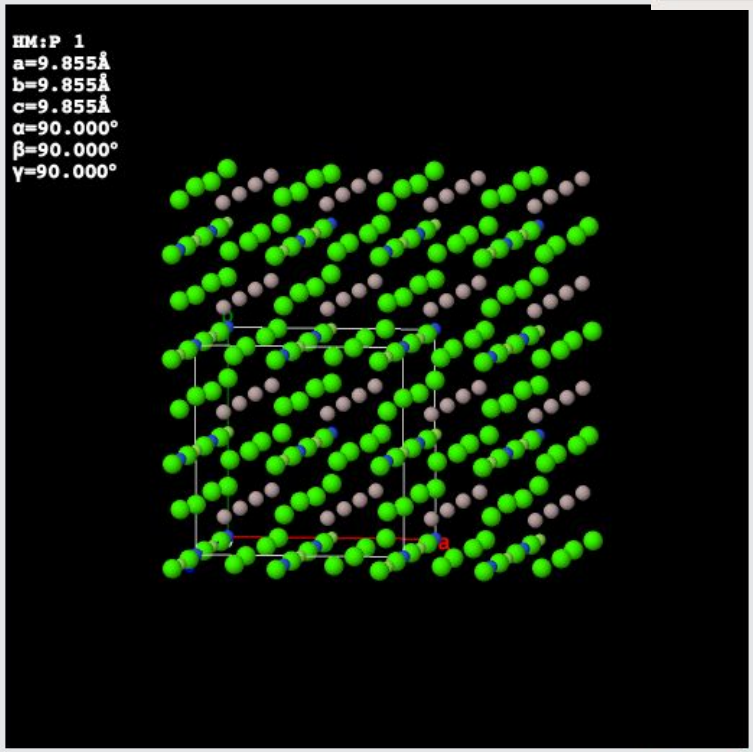
Method	N	F	Al	Ca
Bader	−2.00	−0.98	−2.13	1.20
Hirshfeld	−0.63	−0.36	−1.05	0.52
Voronoi deformation density	−0.81	−0.29	−1.13	0.56

https://materialsproject.org/materials/mp-989399/

Home

MATERIAL: Ca_6Al_2NF

ID: mp-989399



Structure Type: Conventional Standard | Primitive | Refined

CIF

Space Filling

Polyhedra

Material Details

Final Magnetic Moment: 0.882 μ_B

Magnetic Ordering: Non-magnetic

Formation Energy/Atom: -1.007 eV

Energy Above Hull: 0.000 eV/atom

Density: 2.27 g/cm^3

Decomposes To: Stable

Band Gap: 0.000 eV

Space Group

Lattice Parameters

a	6.969 Å	α	60.000°
b	6.969 Å	β	60.000°
c	6.969 Å	γ	60.000°

Volume 239.311 Å^3

Final Structure

Fractional Coordinates

CIF

Ca

a	b	c
0.2407	0.7593	0.7593
0.2407	0.7593	0.2407
0.2407	0.2407	0.7593
0.7593	0.2407	0.7593
0.7593	0.7593	0.2407

Al

Faber et al, *Phys Rev Lett* (2016) arxiv.org/abs/1508.05315

Quantum Machine: Elpasolite Crystal Energy Predictions

Input panel

Single Prediction

Group Prediction

Crystal Design

Elements in first position:

Elements in second position:

Elements in third position:

Elements in forth position:

Target formation energy:

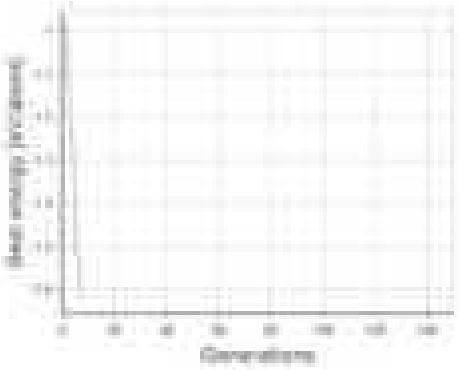
Iteration steps:

Population size:

Output report

Optimizing crystal with
target value: -5.00
max generations: 100
population size: 20

Optimization progress



Conclusions II

Scientific method - proper way to gain knowledge

Inductive (Data)

1. Assume a law
2. Metric
3. Examples
4. Infer
5. New combination

Fast (ms)
Arbitrary reference
Automatic improvement

Transferable?
Minimally condensed

Deductive (Laws)

1. Assume a law
2. Approximate
3. Solve
4. Predict
5. New regimes

Slow (depending on approx.)
Approximation dependent
Human improvement

Transferable?
Maximally condensed

-70.160669 96288144_10