Preaching on first principles views on chemical compound space: Atom centered potentials and statistical learning

O. Anatole von Lilienfeld
Institute of Physical Chemistry, Department of Chemistry, University of Basel, Switzerland
Argonne Leadership Computing Facility, Argonne National Laboratory, Illinois, USA

``First principles view on chemical compound space: Gaining rigorous atomistic control of molecular properties'',
O. A. von Lilienfeld, Int J Quant Chem (2013), http://arxiv.org/abs/1209.5033

If, in some cataclysm, all scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words?

If, in some cataclysm, all scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words?

I believe it is the atomic hypothesis (or atomic fact, or whatever you wish to call it) that all things are made of atoms — little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another. In that one sentence you will see an enormous amount of information about the world, if just a little imagination and thinking are applied.

Feynman Lectures of Physics (1964)

``QMC is not a black-box''
(M. Foulkes)

``QMC is not a black-box''
(M. Foulkes)

…

``One material every 2 s by 2016''
(T. Mueller)

``QMC is not a black-box''
(M. Foulkes)

…

``One material every 2 s by 2016''
(T. Mueller)

``QMC can inform improvements of current DFTs (or other methods)''
(M. Gillan, A. Tkatchenko, ...)

``QMC is not a black-box''
(M. Foulkes)

…

``One material every 2 s by 2016''
(T. Mueller)

``QMC can inform improvements of current DFTs (or other methods)''
(M. Gillan, A. Tkatchenko, ...)

Increase DFT's transferability to properly account for

- spin states
- excited states
- van der Waals

``QMC is not a black-box''
(M. Foulkes)

…

``One material every 2 s by 2016''
(T. Mueller)

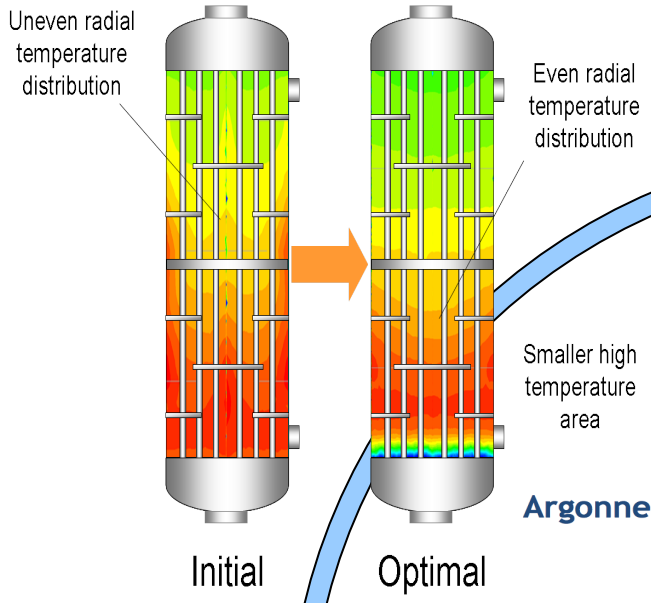
``QMC can inform improvements of current DFTs (or other methods)''
(M. Gillan, A. Tkatchenko, ...)

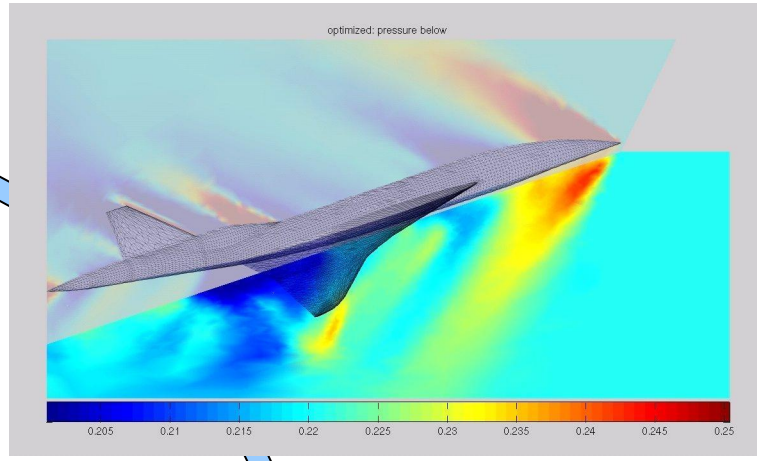
Increase DFT's transferability to properly account for

- spin states
- excited states
- van der Waals

Define transferability!

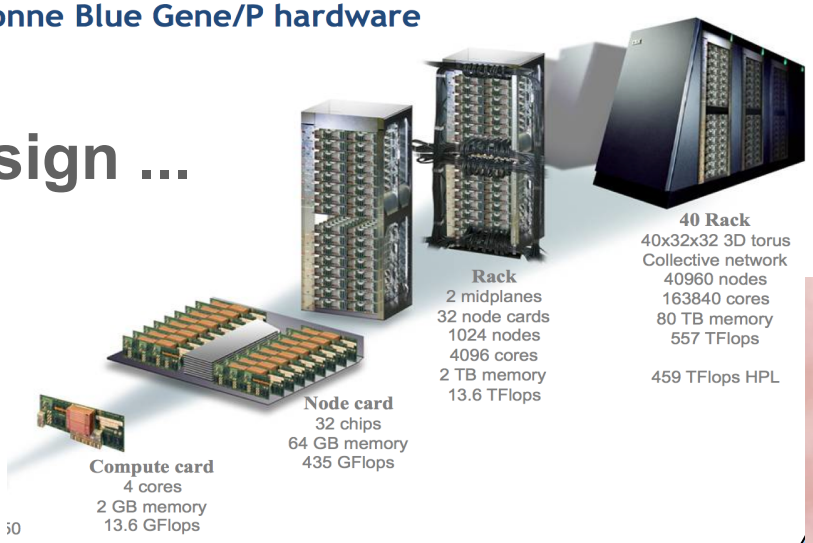
A method is called transferable if its error is invariant wrt changes in atomic configuration and composition == chemical compound space





Computational design ...

Argonne Blue Gene/P hardware



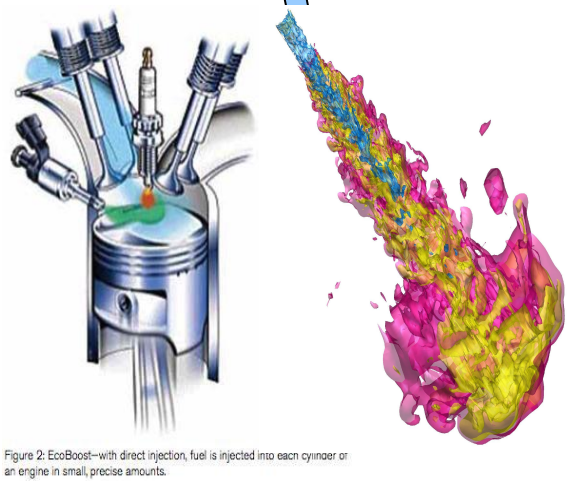
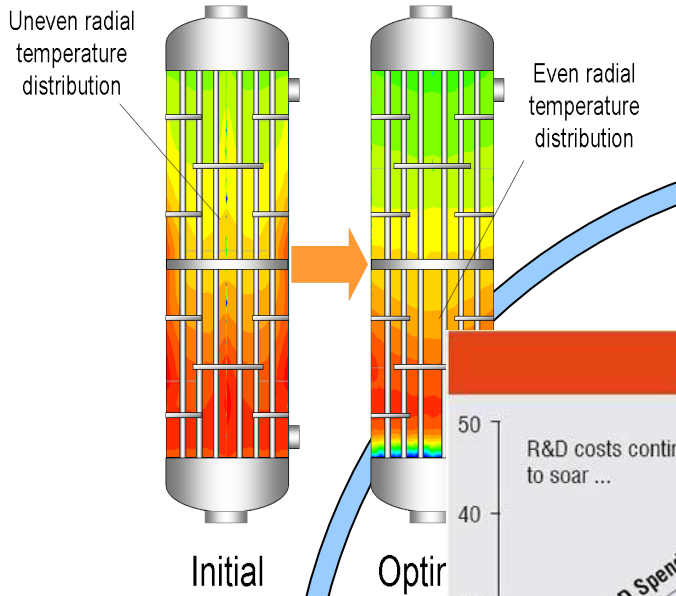
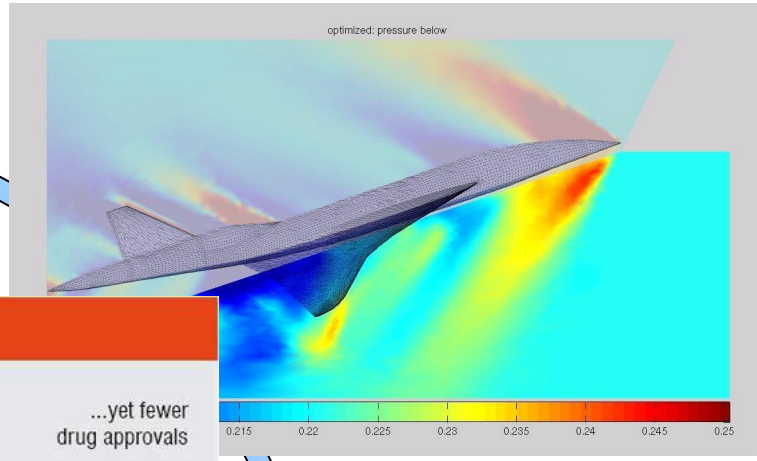
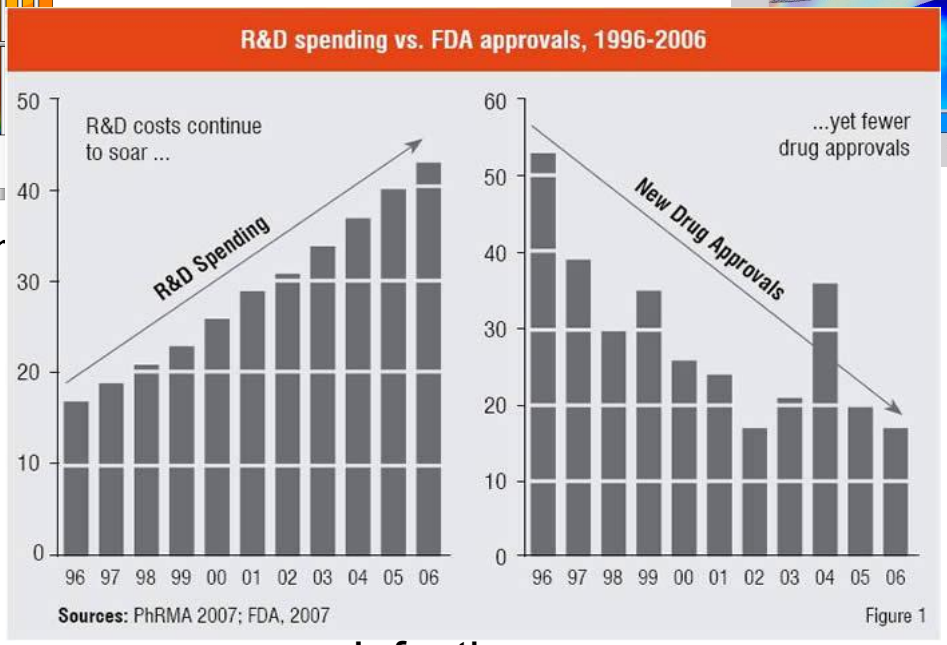
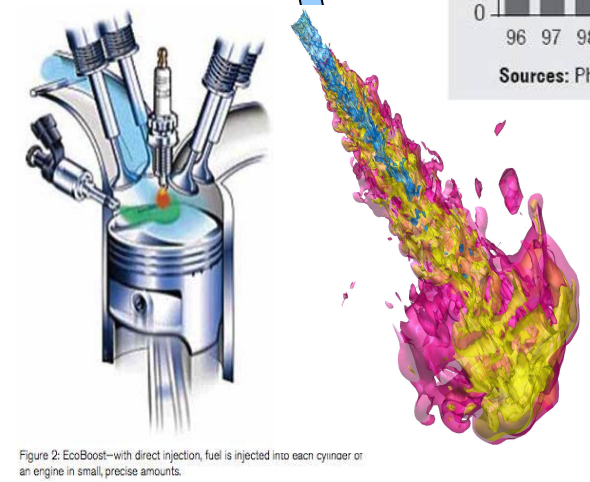


Figure 2: EcoBoost—with direct injection, fuel is injected into each cylinder of an engine in small, precise amounts.









- Infections
- Metabolic syndrome
- Aging
- Cancer

...

Why is this hard?

Combinatorial catastrophe

number of small organic molecules > 10^{60}

Nature Insight on chemical space (2004)

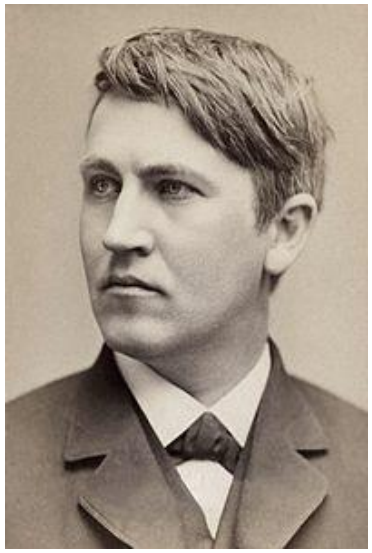
Assume 1 property evaluation ~ 1 s

→ exhaustive screening ~ 10^{52} yrs

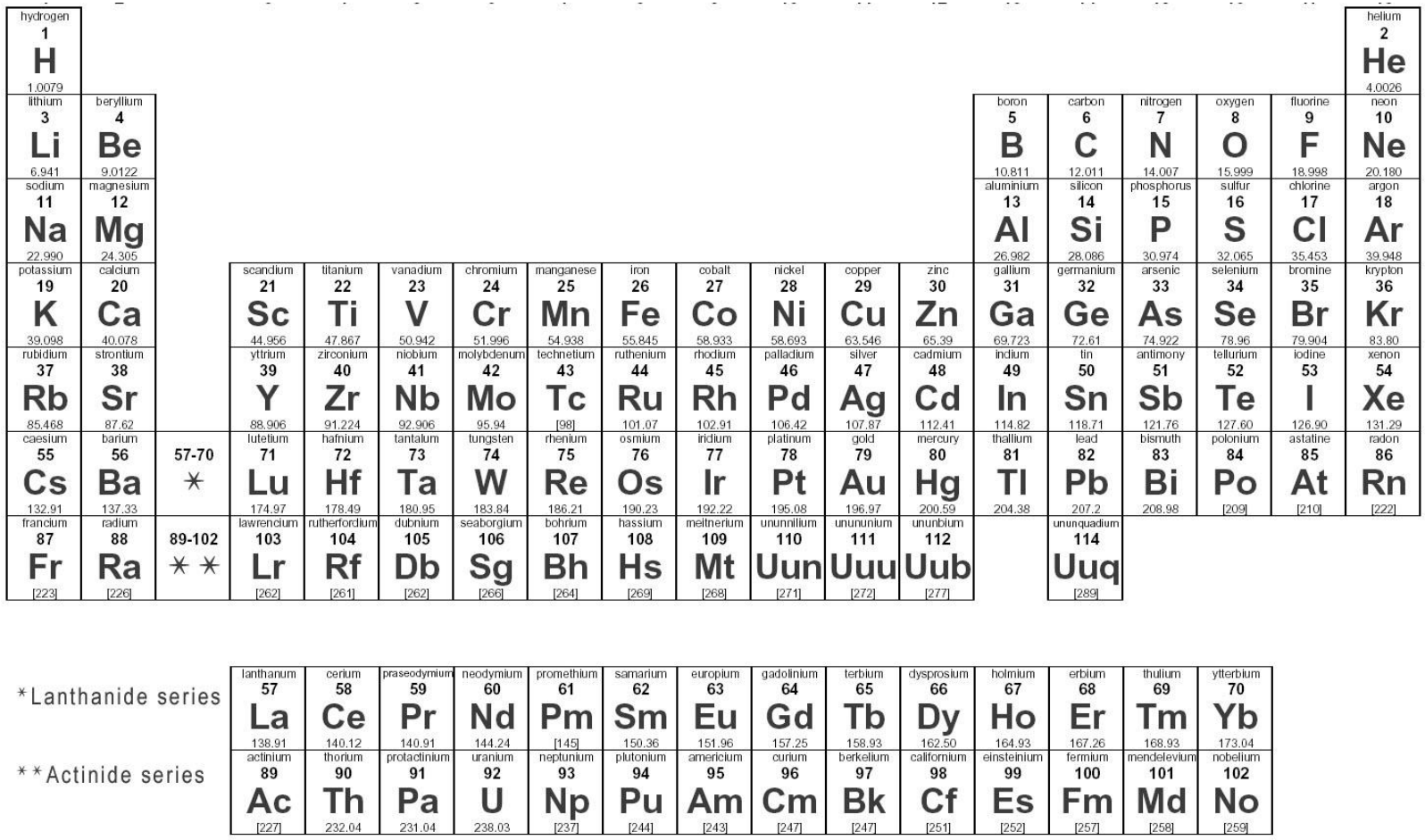
(age of universe ~10^{10} yrs)

Edisonian approach

1878



Combinatorial problem



Combinatorial problem

How many stoichiometries have N_p = 40 protons?

B_8

ScH_{19}

O_5

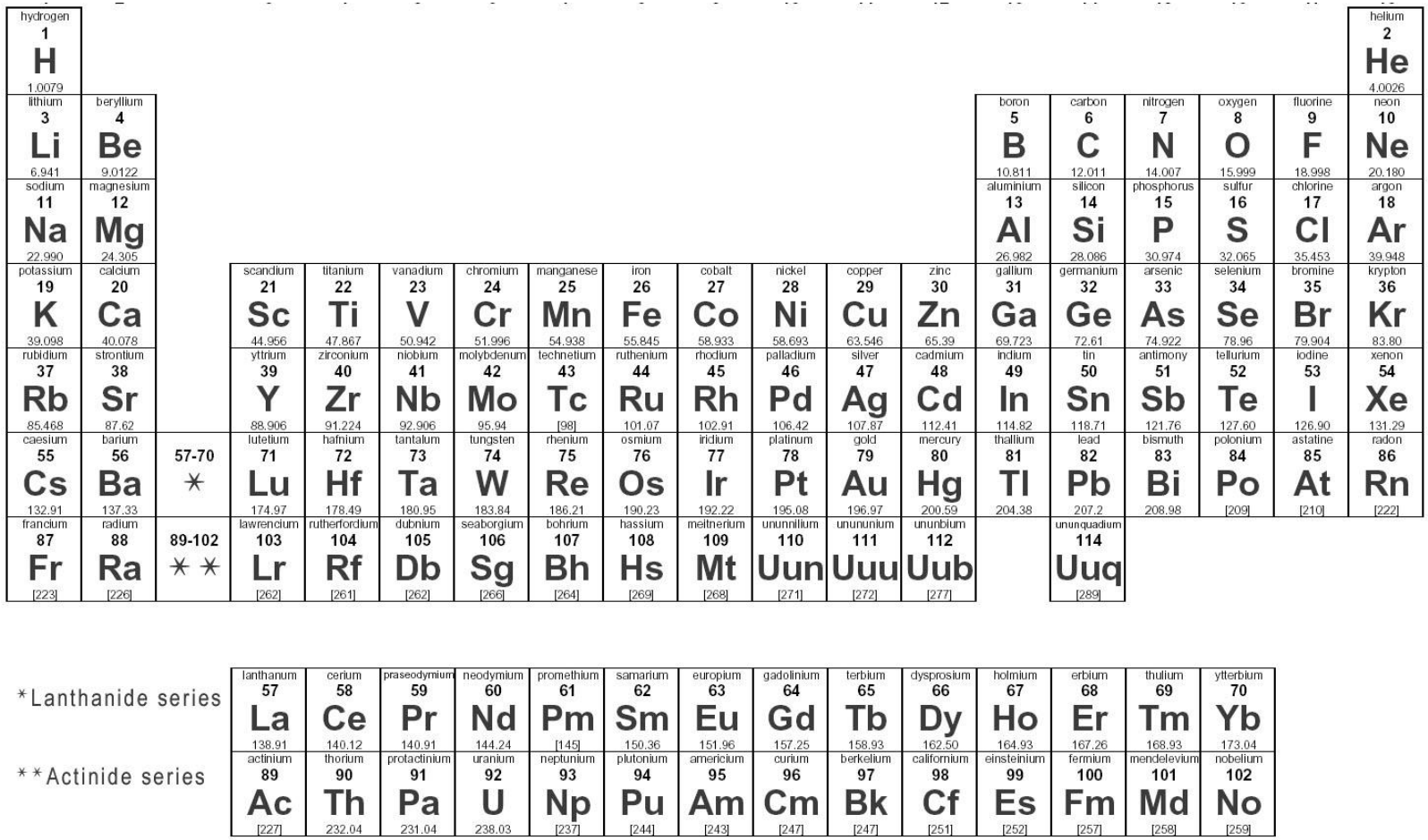
H_{40}

$ZnNe$

He_{20}

Ca_2

Zr



Combinatorial problem

How many stoichiometries have N_p = 40 protons?

B_8

ScH_{19}

O_5

H_{40}

$ZnNe$

He_{20}

Ca_2

Zr

Discrete number theory

- Integer partition of N_p
- Number of ways to write N_p as sum of positive integers
- Young-Ferrers diagrams

OAvL, *Int J Quant Chem*, (2013)

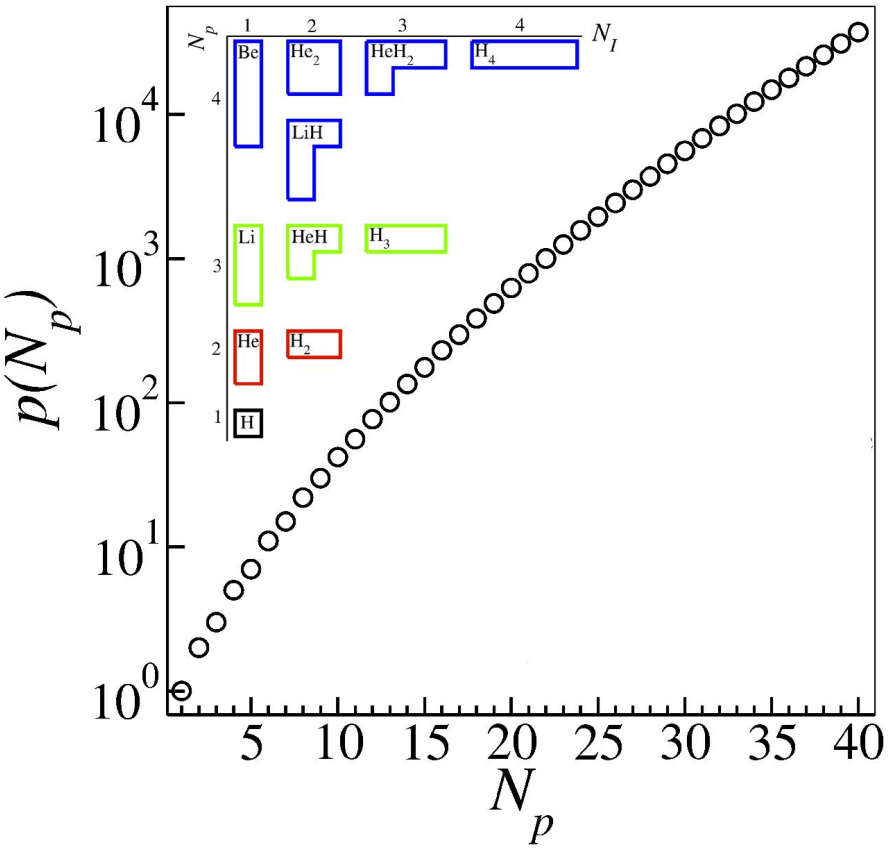
Combinatorial problem

How many stoichiometries have N_p = 40 protons?

B_8
ScH_{19}
O_5
H_{40}
$ZnNe$
He_{20}
Ca_2
Zr

Discrete number theory

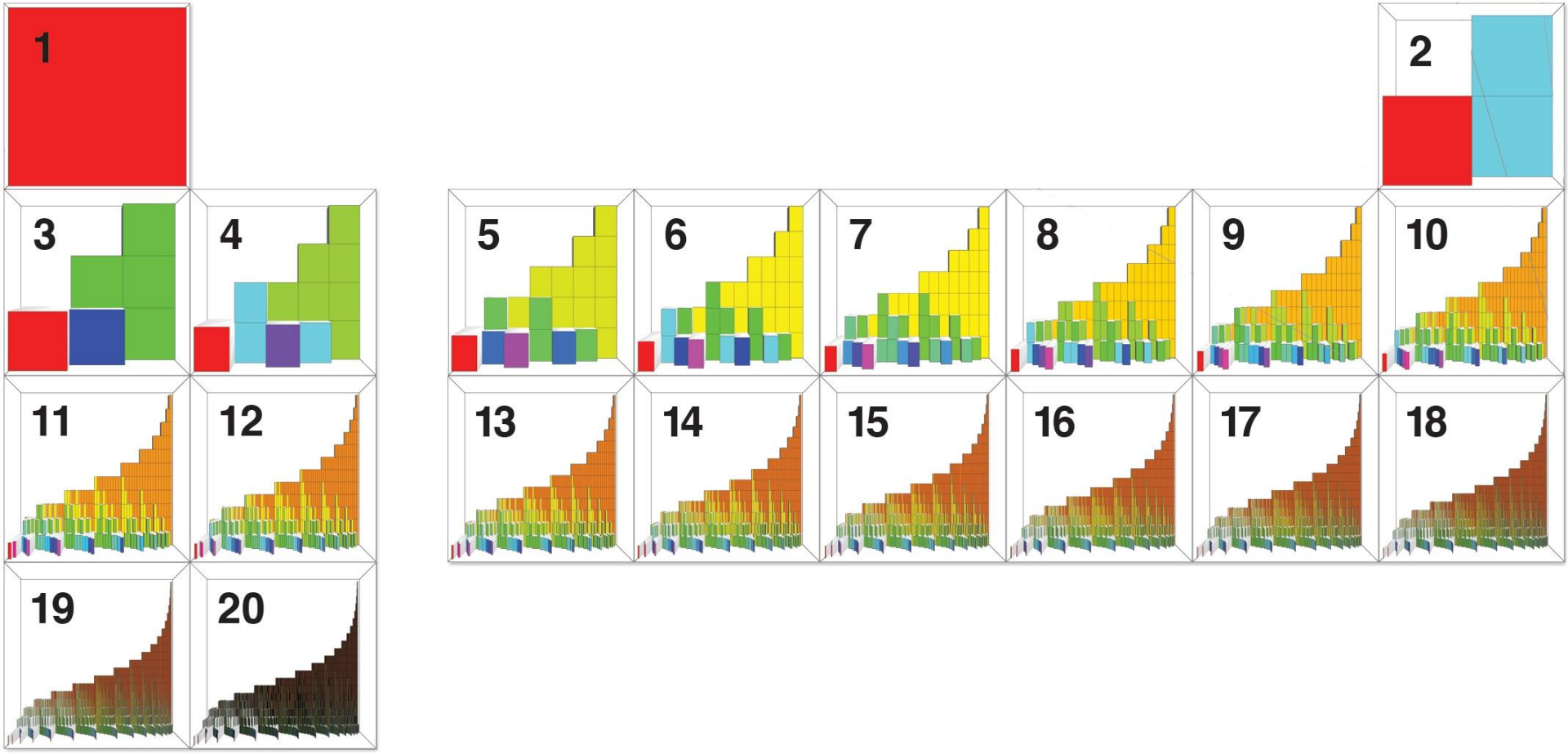
- Integer partition of N_p
- Number of ways to write N_p as sum of positive integers
- Young-Ferrers diagrams



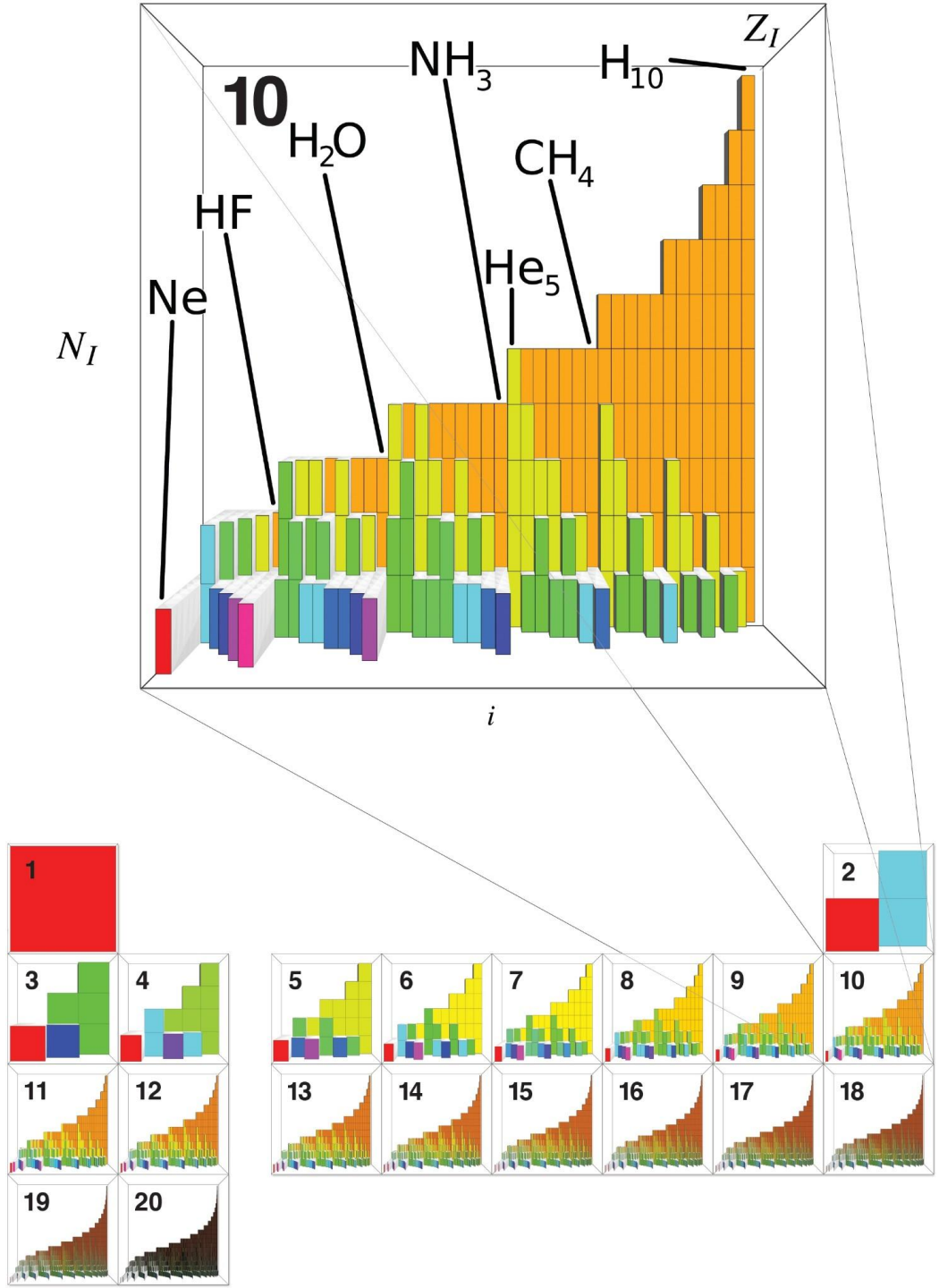
→ 40 protons yield > 37k stoichiometries

OAvL, *Int J Quant Chem*, (2013)

Combinatorial problem - availability heuristic?



OAvL, *Int J Quant Chem*, (2013)

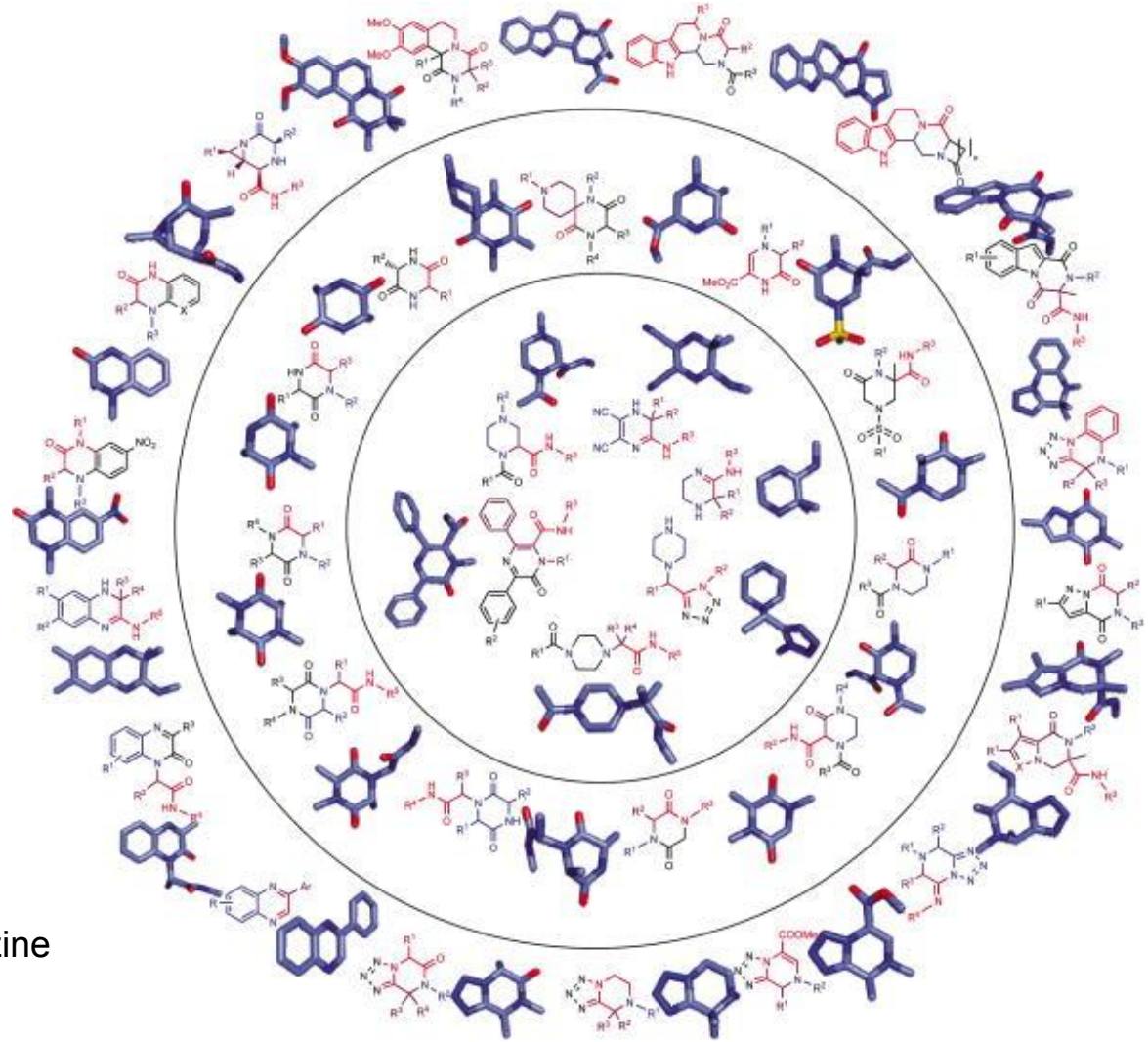

Z_I
NH_3
H_{10}
10
H_2O
CH_4
HF
He_5
Ne
N_I
i
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Why is this hard?

Combinatorial catastrophe

number of small organic molecules > 10^{60}

Nature Insight on chemical space (2004)



Piperazine

Why is this hard?

Combinatorial catastrophe

number of small organic molecules > 10^{60}

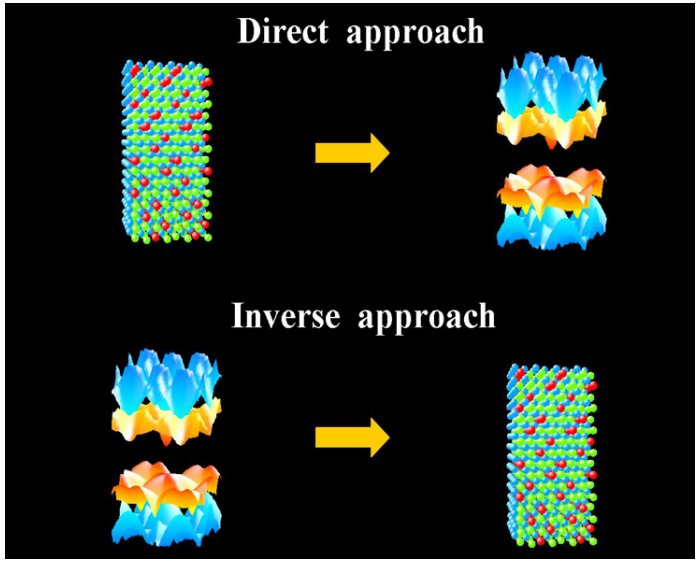
Nature Insight on chemical space (2004)

Assume 1 property evaluation ~ 1 s

→ exhaustive screening ~ 10^{52} yrs

(age of universe ~10^{10} yrs)

New



Franceschetti and Zunger, *Nature* (1999)

Why is this hard?

Combinatorial catastrophe

number of small organic molecules > 10^{60}

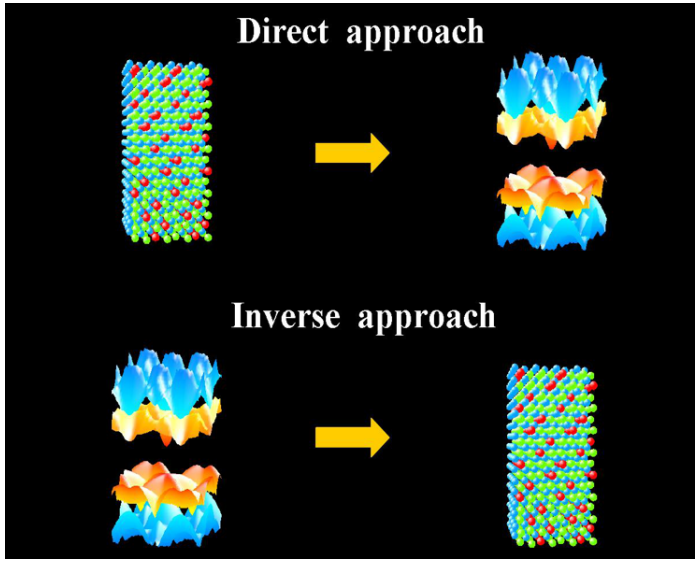
Nature Insight on chemical space (2004)

Assume 1 property evaluation ~ 1 s

→ exhaustive screening ~ 10^{52} yrs

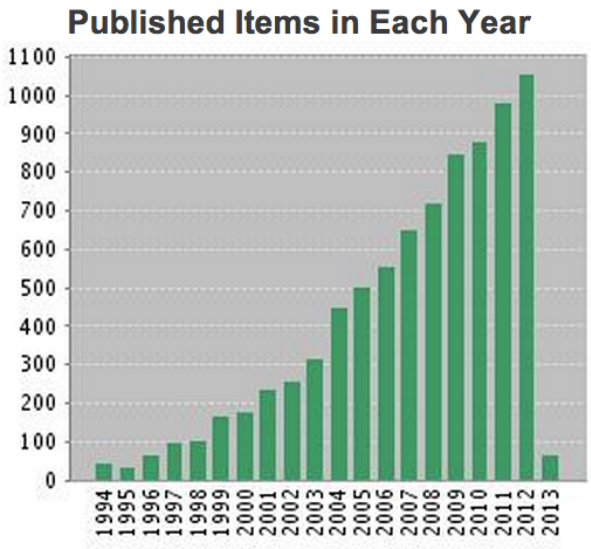
(age of universe ~10^{10} yrs)

New



Franceschetti and Zunger, *Nature* (1999)

DFT & Surface Adsorption



Why is this hard?

Combinatorial catastrophe

number of small organic molecules > 10^{60}

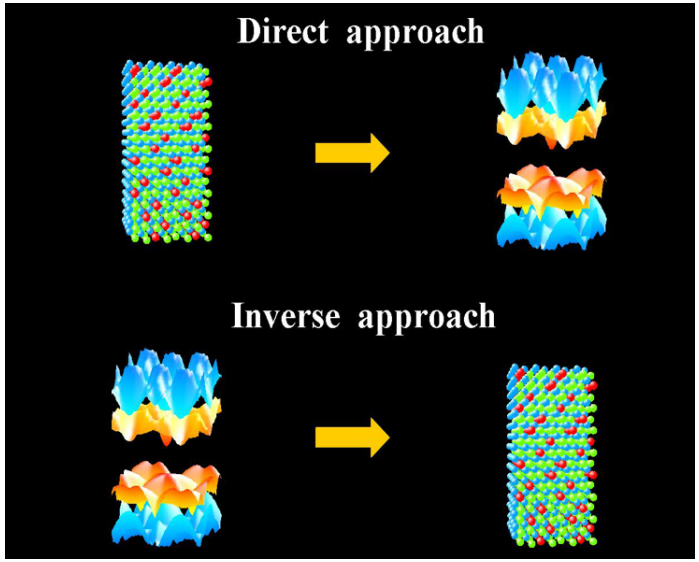
Nature Insight on chemical space (2004)

Assume 1 property evaluation ~ 1 s

→ exhaustive screening ~ 10^{52} yrs

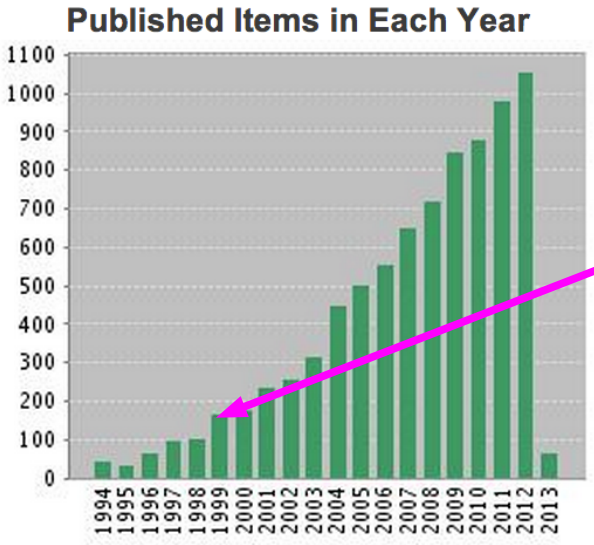
(age of universe ~10^{10} yrs)

New

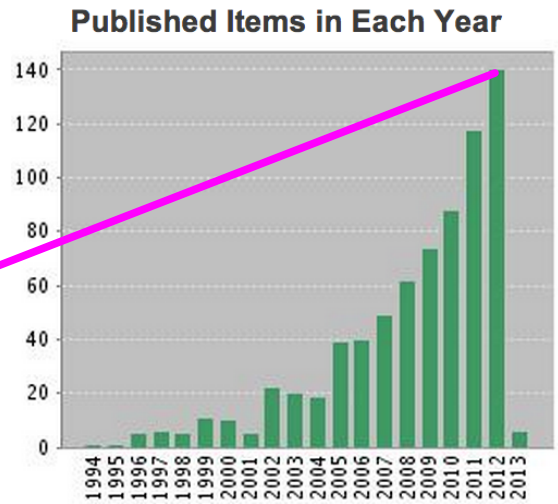


Franceschetti and Zunger, *Nature* (1999)

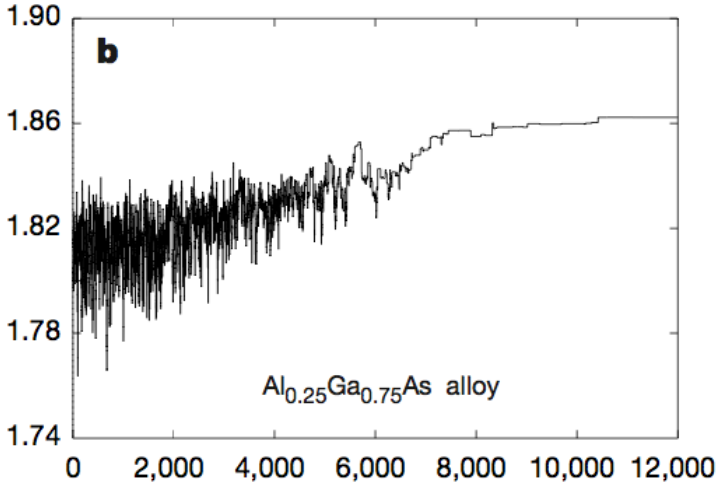
DFT & Surface Adsorption

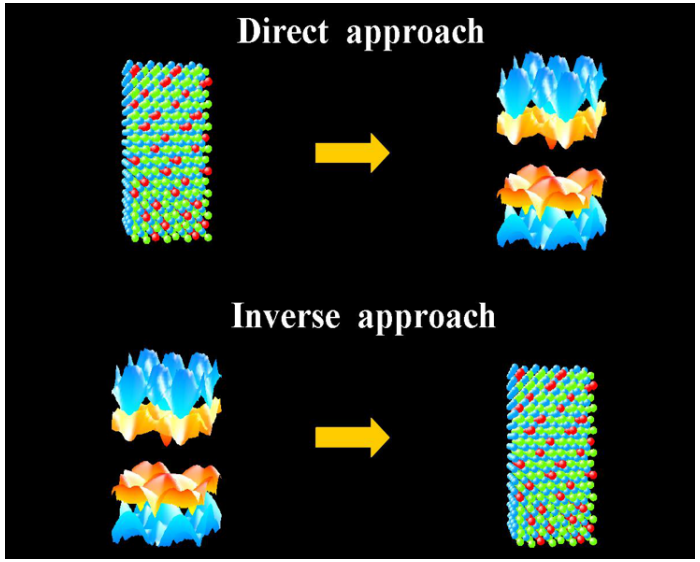


DFT & Surface Design



Right compound for right reason!

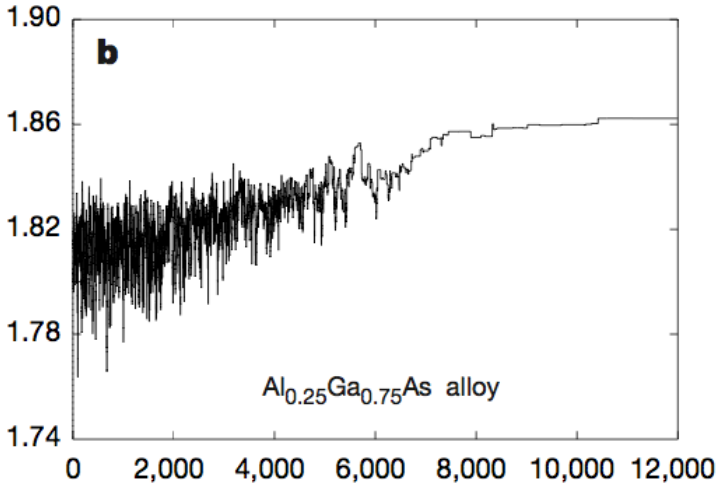


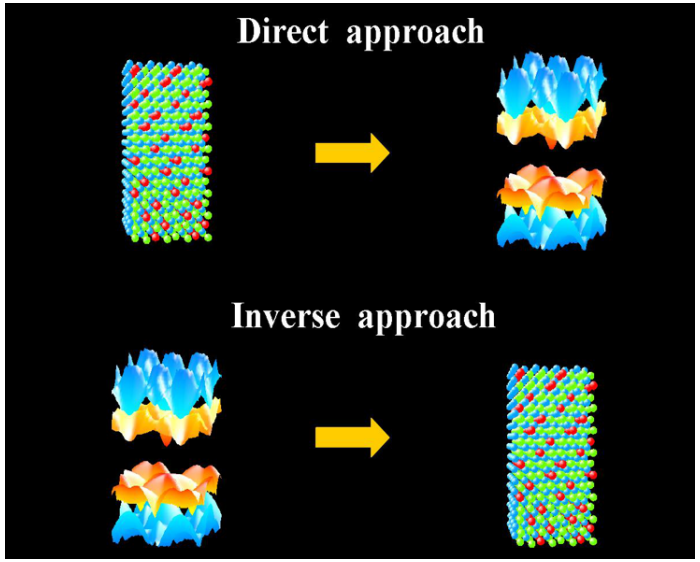


Franceschetti and Zunger, *Nature* (1999)

$$\min_{\{Z_I,\mathbf{R}_I\}} \sum_i \omega_i \left(P_i(\{Z_I,\mathbf{R}_I\}) - P_i^{\mathrm{ref}} \right)^2$$

Right compound for right reason!





Franceschetti and Zunger, *Nature* (1999)

- No analytic solution
- Ill-defined
- high dimensional
- expensive

→ Iterative minimization

$$\min_{\{Z_I,\mathbf{R}_I\}} \sum_i \omega_i \left(P_i(\{Z_I,\mathbf{R}_I\}) - P_i^{\mathrm{ref}} \right)^2$$

1) Transferable: First principles (QM, UFF)

2) Smart: Variational (dP/dX), Genetic, ...

3) Fast: Correlational (Machine Learning)

4) Muscle: Supercomputing & Data

First Principles

$$H(\{Z_I, \mathbf{R}_I\})\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

Schrödinger

First Principles

$$H(\{Z_I, \mathbf{R}_I\})\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

Schrödinger

variational (deductive)

Feynman

$$\frac{\partial E[H]}{\partial R_{Ix}} = \left\langle \Psi \left| \frac{\partial H}{\partial R_{Ix}} \right| \Psi \right\rangle$$

First Principles

$$H(\{Z_I, \mathbf{R}_I\})\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

Schrödinger

variational (deductive)

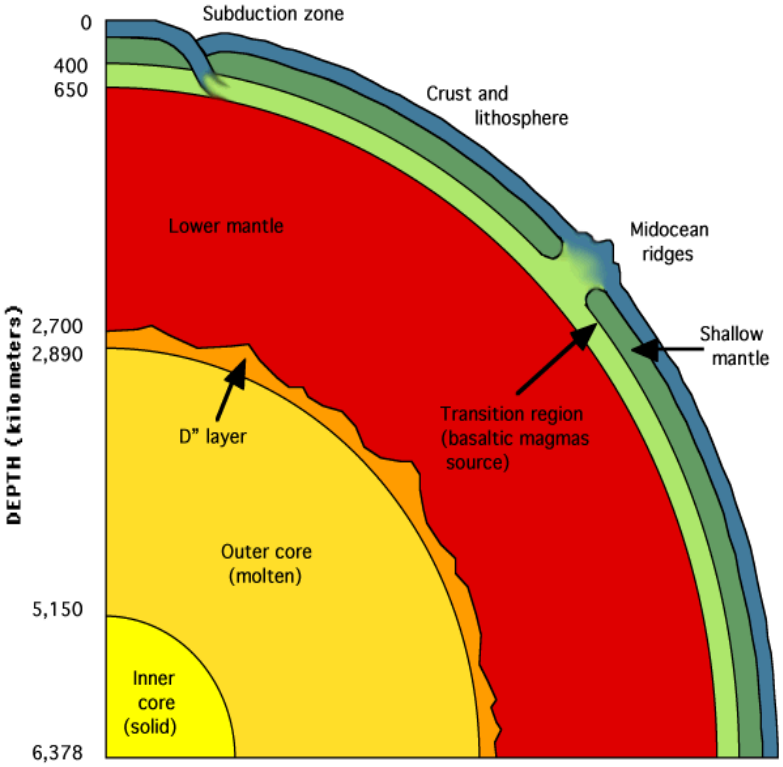
Alchemy

1. Free energies
2. Gradients to optimize

Feynman

$$\frac{\partial E[H]}{\partial R_{Ix}} = \left\langle \Psi \left| \frac{\partial H}{\partial R_{Ix}} \right| \Psi \right\rangle$$

$$\frac{\partial E[H]}{\partial Z_I} = \left\langle \Psi \left| \frac{\partial H}{\partial Z_I} \right| \Psi \right\rangle$$



Constraints on the composition of the Earth's core from *ab initio* calculations

D. Alfè*, M. J. Gillan† & G. D. Price*

* *Research School of Geological and Geophysical Sciences, Birkbeck College and University College London, Gower Street, London WC1E 6BT, UK*
† *Physics and Astronomy Department, University College London, Gower Street, London WC1E 6BT, UK*

Knowledge of the composition of the Earth's core[1–3] is important for understanding its melting point and therefore the temperature at the inner-core boundary and the temperature profile of the core and mantle. In addition, the partitioning of light elements between solid and liquid, as the outer core freezes at the inner-core boundary, is believed to drive compositional convection[4], which in turn generates the Earth's magnetic field. It is generally

Nature (2000)

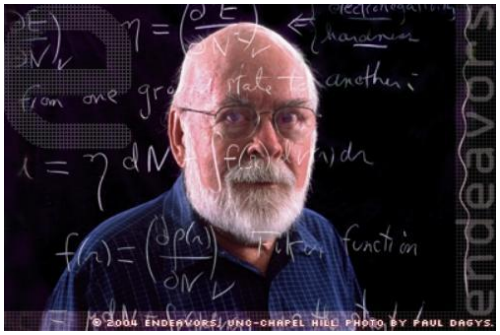
*

Variational: Gradients

Mermin

Fractional N_e

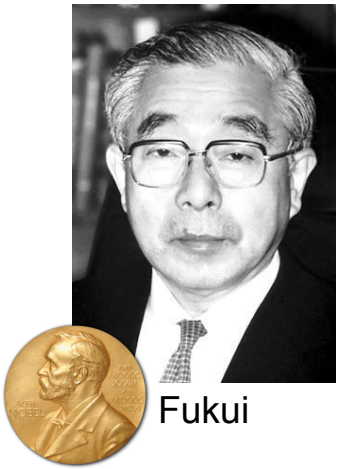
$$\frac{\partial E[H]}{\partial N_e} = \mu_e = \epsilon$$



Parr

Fukui function: Response of frontier orbitals to molecular changes

Conceptual DFT (Parr, Yang et al)



Fukui

Variational: Gradients

Fractional N_e

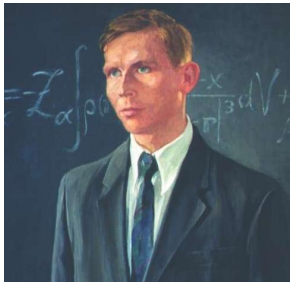
$$\frac{\partial E[H]}{\partial N_e} = \mu_e = \epsilon$$

OAvL et al, *Phys Rev Lett* (2005)

Fractional Z_I

$$\frac{\partial E[H]}{\partial Z_I} = \left\langle \Psi \left| \frac{\partial H}{\partial Z_I} \right| \Psi \right\rangle = \int d\mathbf{r} \frac{n(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_I|} - \sum_J \frac{Z_J}{|\mathbf{R}_J - \mathbf{R}_I|}$$

Hellmann



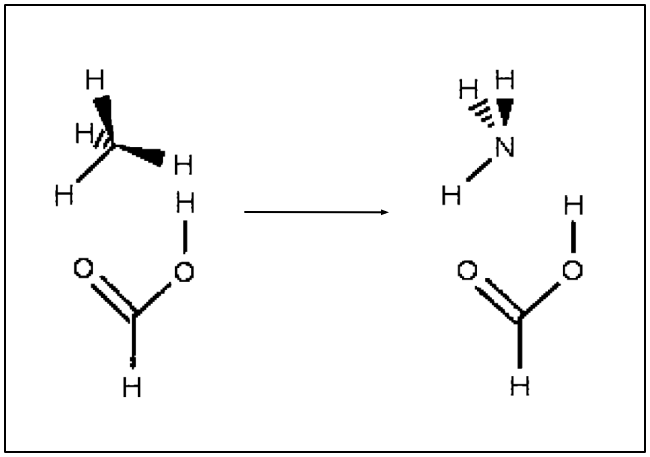
Feynman

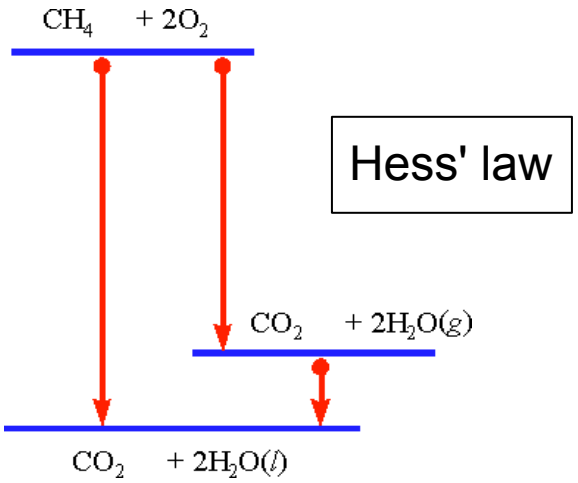
Weigend and Ahlrichs *J Chem Phys* (2004)

Yang & Beratan et al *JACS* (2006)

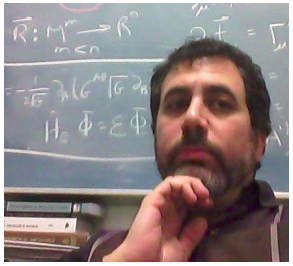
OAvL: *Phys Rev Lett* (2005), *J Chem Phys (2006, 2009)*, *J Chem Theory Comput* (2007)

Variational: Fractional nuclei





Hess' law



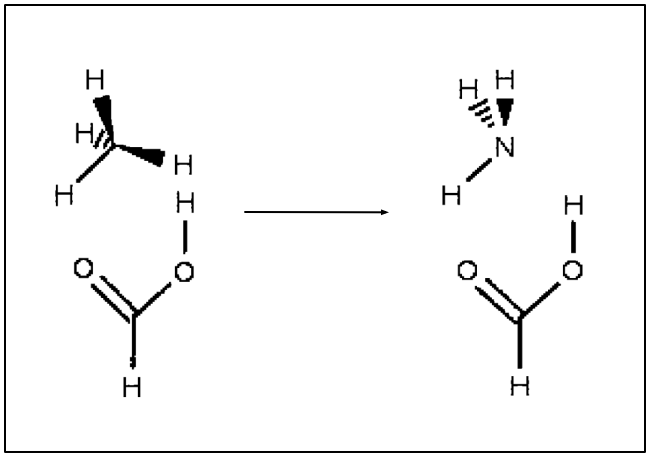
Tuckerman (NYU)

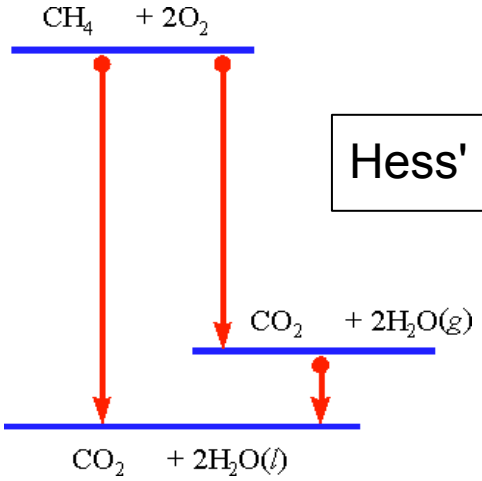
Weigend and Ahlrichs *J Chem Phys* (2004)

Yang & Beratan et al *JACS* (2006)

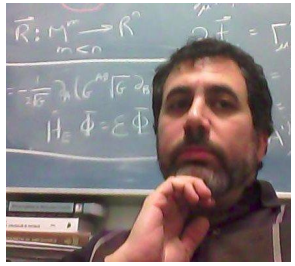
OAvL: *Phys Rev Lett* (2005), *J Chem Phys (2006, 2009)*, *J Chem Theory Comput* (2007)

Variational: Fractional nuclei

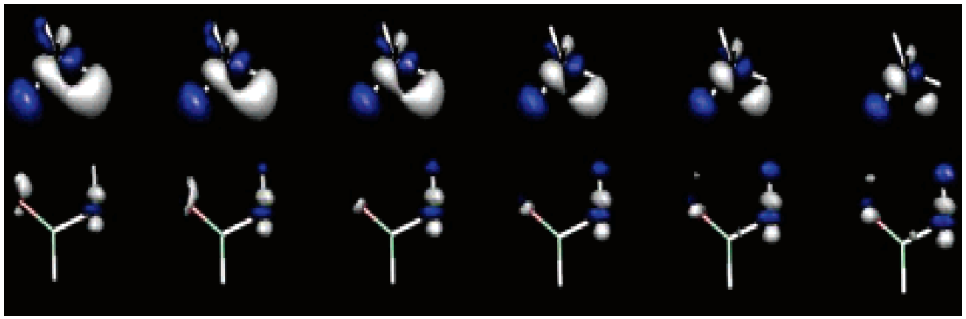


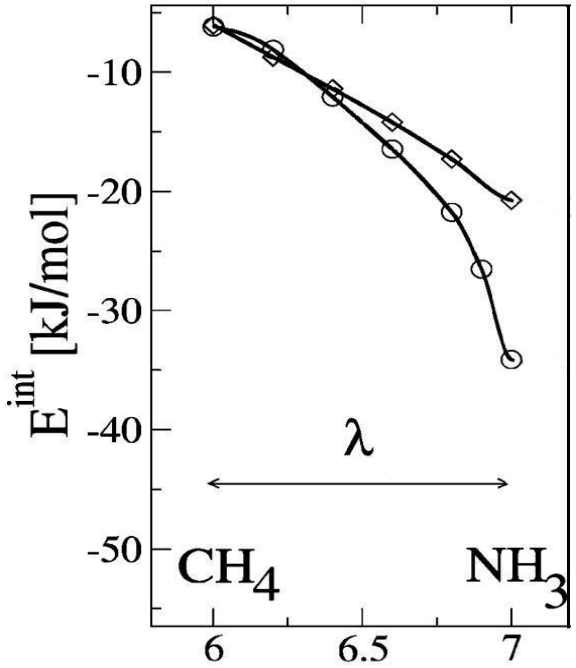


Hess' law



Tuckerman (NYU)





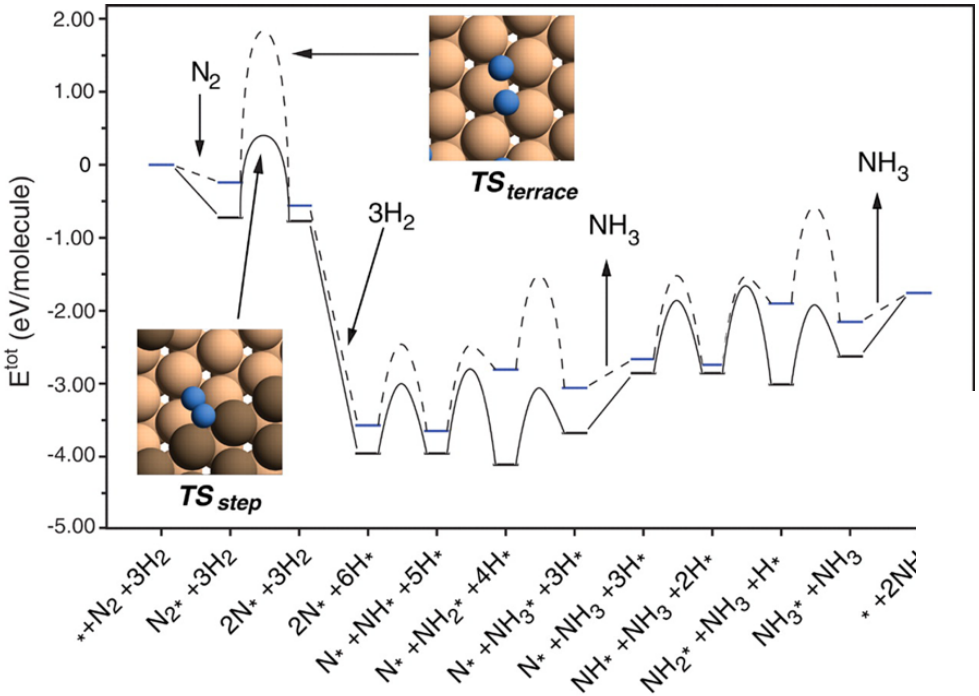
Weigend and Ahlrichs *J Chem Phys* (2004)

Yang & Beratan et al *JACS* (2006)

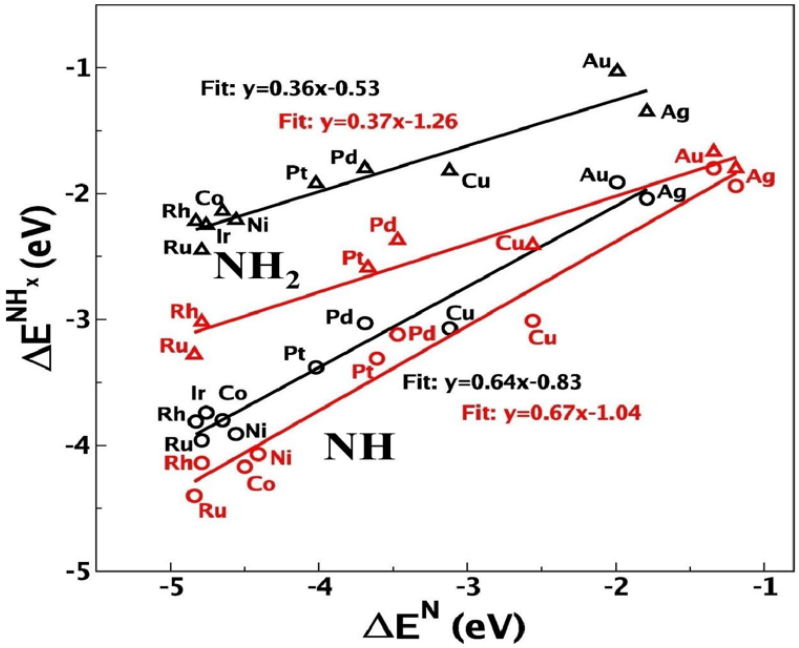
OAvL: *Phys Rev Lett* (2005), *J Chem Phys (2006, 2009)*, *J Chem Theory Comput* (2007)

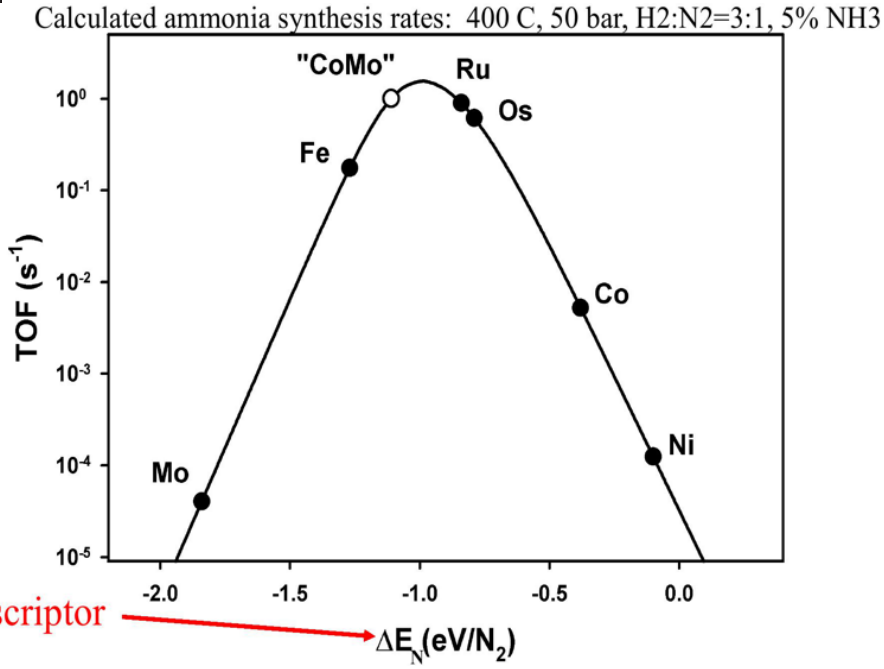
Adsorption

Haber-Bosch: $N_2 + 3\ H_2 \longrightarrow 2\ NH_3$



Sabatier's principle





Nørskov et al. *Nature Chemistry* (2009)
Nørskov et al. *Science* (2005)
Nørskov et al. *Phys Rev Lett* (2005)
Nørskov et al. *J Am Chem Soc.* (2001)

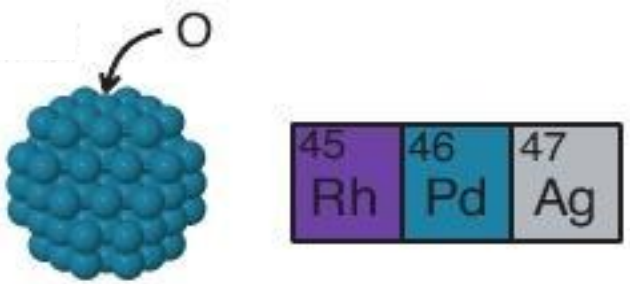
Adsorption

Volcano for oxygen reduction reaction: Oxygen binding

$$E^{\text{bind}} = E(\text{Pd}_{79}) - E(\text{Pd}_{79}\text{-O}) - 0.5\, E(\text{O}_2)$$

Henkelman (UT)

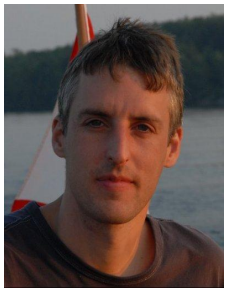
Sheppard (LANL)



How to dope?

D Sheppard, G Henkelman, OAvL, *J Chem Phys* (2010)

Adsorption

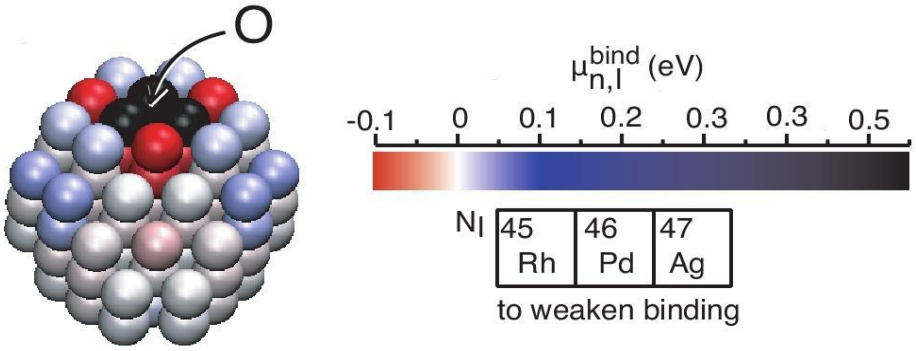



Henkelman (UT) Sheppard (LANL)

Volcano for oxygen reduction reaction: Oxygen binding

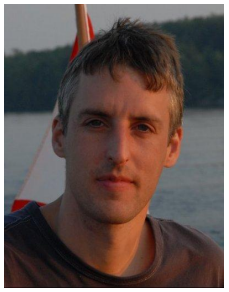
$$E^{\text{bind}} = E(\text{Pd}_{79}) - E(\text{Pd}_{79}\text{-O}) - 0.5\, E(\text{O}_2)$$

$$\mu_{n,I} = \partial E^{\text{bind}} / \partial Z_I$$



D Sheppard, G Henkelman, OAvL, *J Chem Phys* (2010)

Adsorption

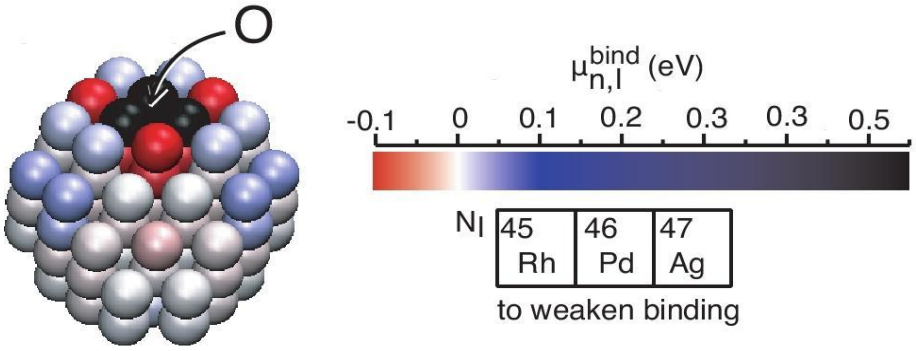

Henkelman (UT)


Sheppard (LANL)

Volcano for oxygen reduction reaction: Oxygen binding

$$E^{\text{bind}} = E(\text{Pd}_{79}) - E(\text{Pd}_{79}\text{-O}) - 0.5\, E(\text{O}_2)$$

$$\mu_{n,I} = \partial E^{\text{bind}} / \partial Z_I$$



1st order expansion for 10 doped mutants

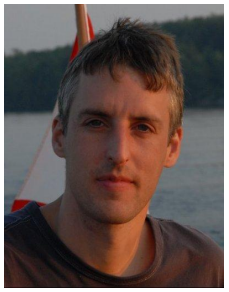
$$\partial_\lambda E^{\text{bind}} = \sum_I \mu^{\text{bind}}_{n,I} \partial_\lambda Z_I(\lambda)$$

D Sheppard, G Henkelman, OAvL, *J Chem Phys* (2010)

Adsorption

Volcano for oxygen reduction reaction: Oxygen binding

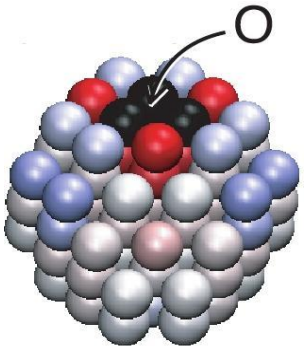
$$E^{\text{bind}} = E(\text{Pd}_{79}) - E(\text{Pd}_{79}\text{-O}) - 0.5\, E(\text{O}_2)$$

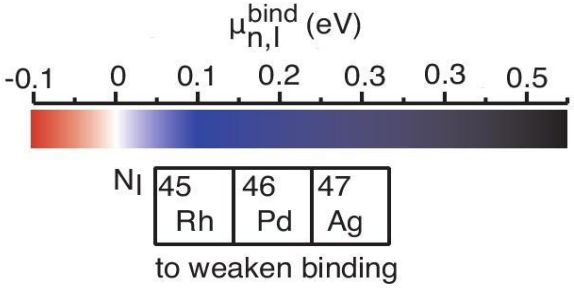


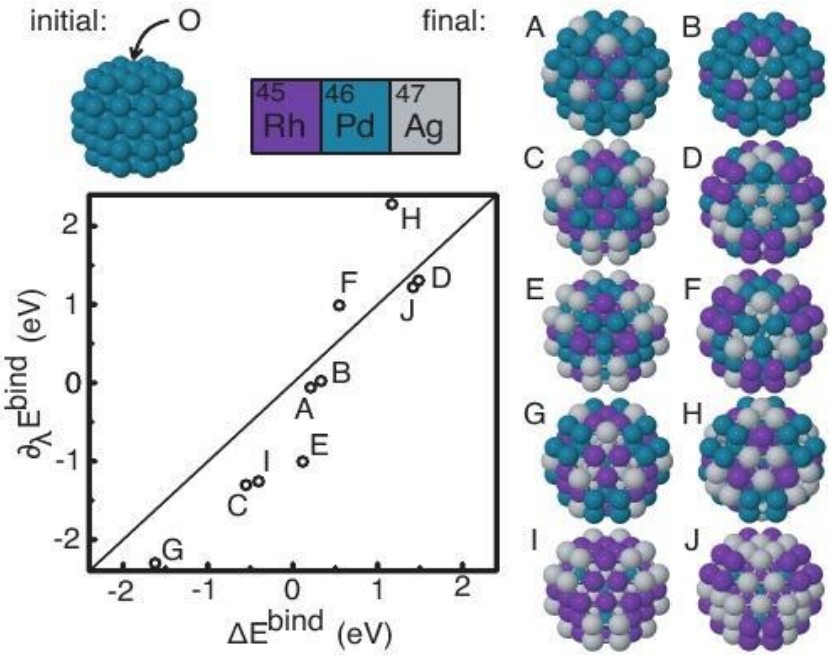
Henkelman (UT)

Sheppard (LANL)

$$\boxed{\mu_{n,I} = \partial E^{\text{bind}} / \partial Z_I}$$





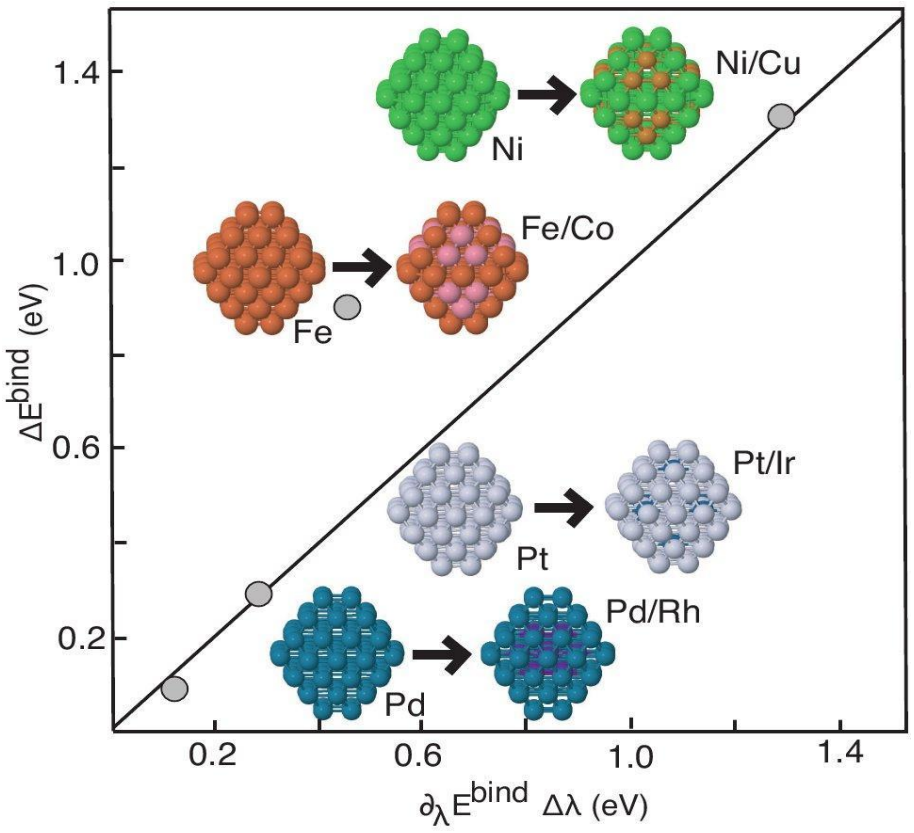


1st order expansion for 10 doped mutants

$$\partial_\lambda E^{\text{bind}} = \sum_I \mu^{\text{bind}}_{n,I} \partial_\lambda Z_I(\lambda)$$

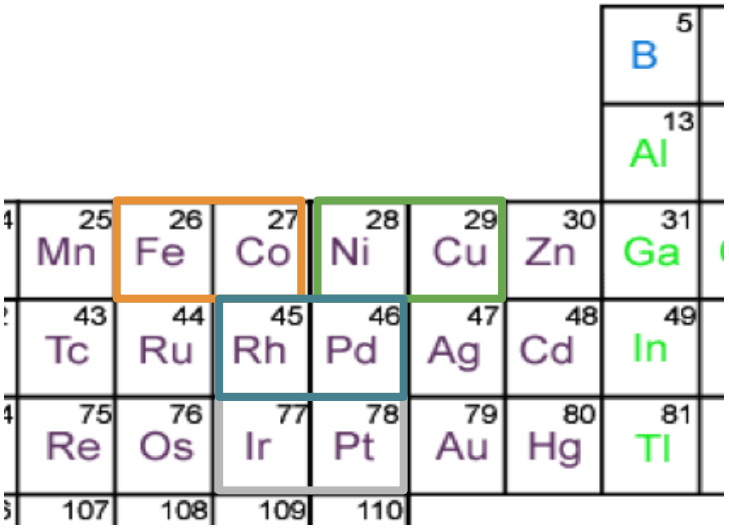
D Sheppard, G Henkelman, OAvL, *J Chem Phys* (2010)

Adsorption



Henkelman (UT)

Sheppard (LANL)

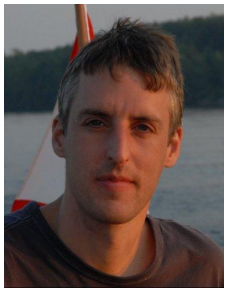


Target oxygen binding value: 1.65 eV

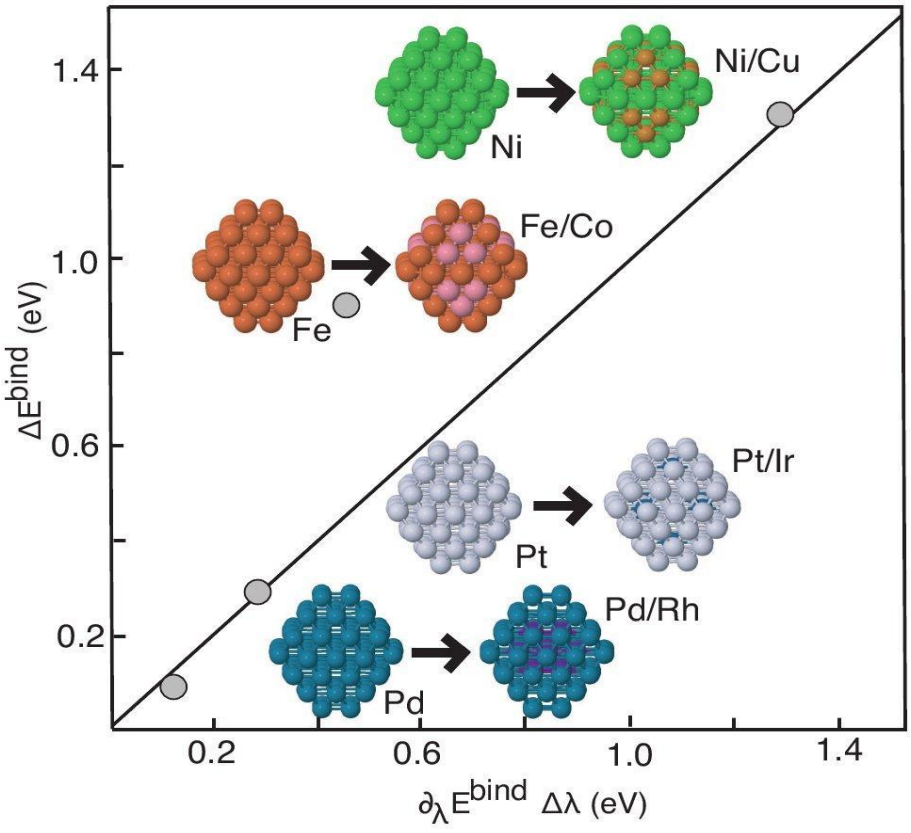
$$\min_{\{Z_I,\mathbf{R}_I\}} \left(P(\{Z_I,\mathbf{R}_I\}) - P^{\mathrm{ref}} \right)^2$$

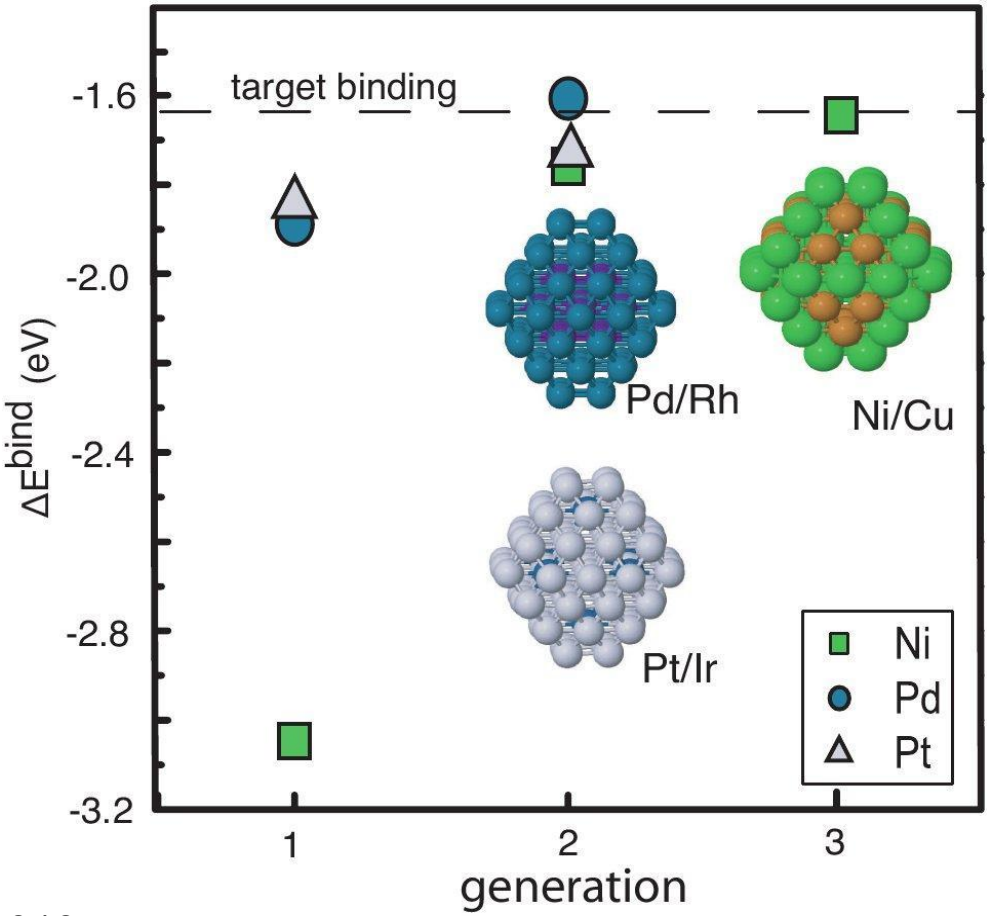
Dan Sheppard, PhD thesis, UT Austin 2010

Adsorption



Henkelman (UT) Sheppard (LANL)

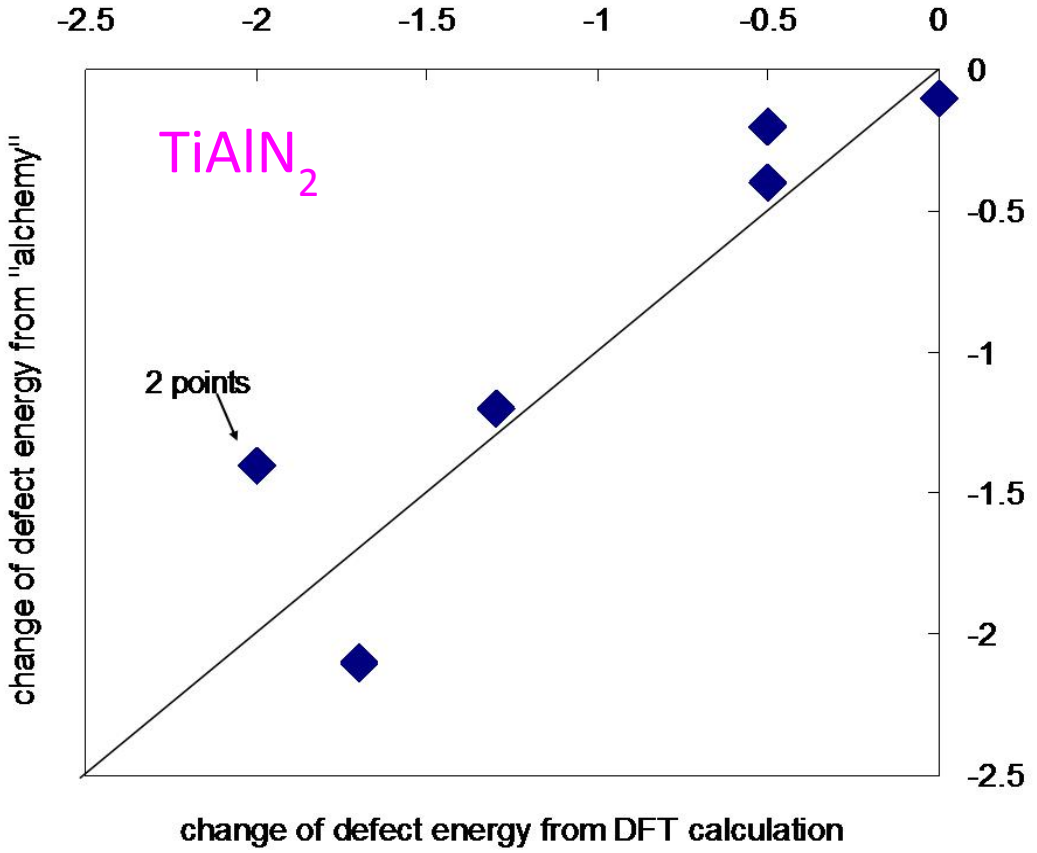




Dan Sheppard, PhD thesis, UT Austin 2010

Defects

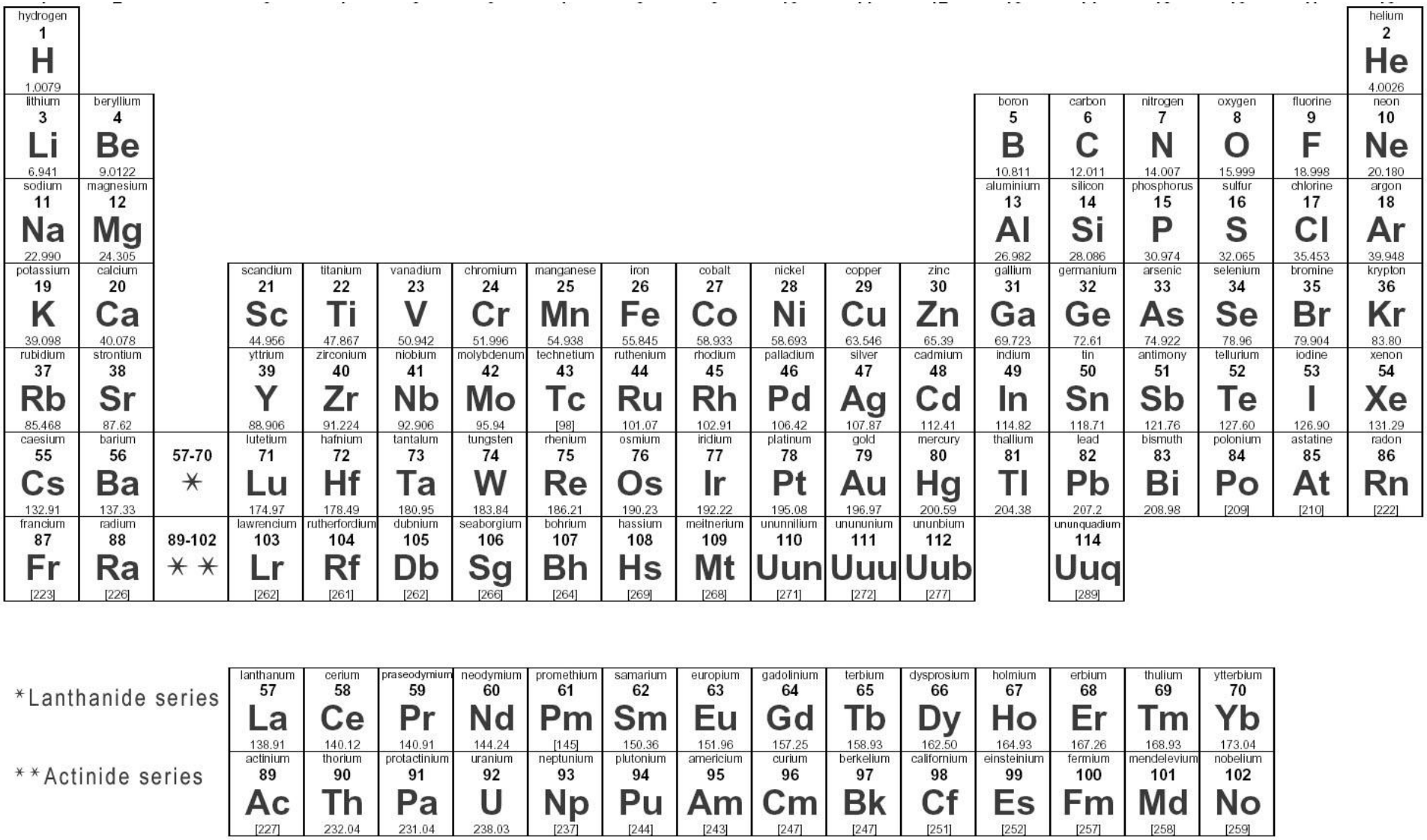
Al-vacancy energies [eV]



Predicted changes for various N to O mutations (out of 32)

Preliminary results from Moritz to Baben (group of Prof. Schneider, RWTH Aachen)

Is Z really a good variable?



Is Z really a good variable?

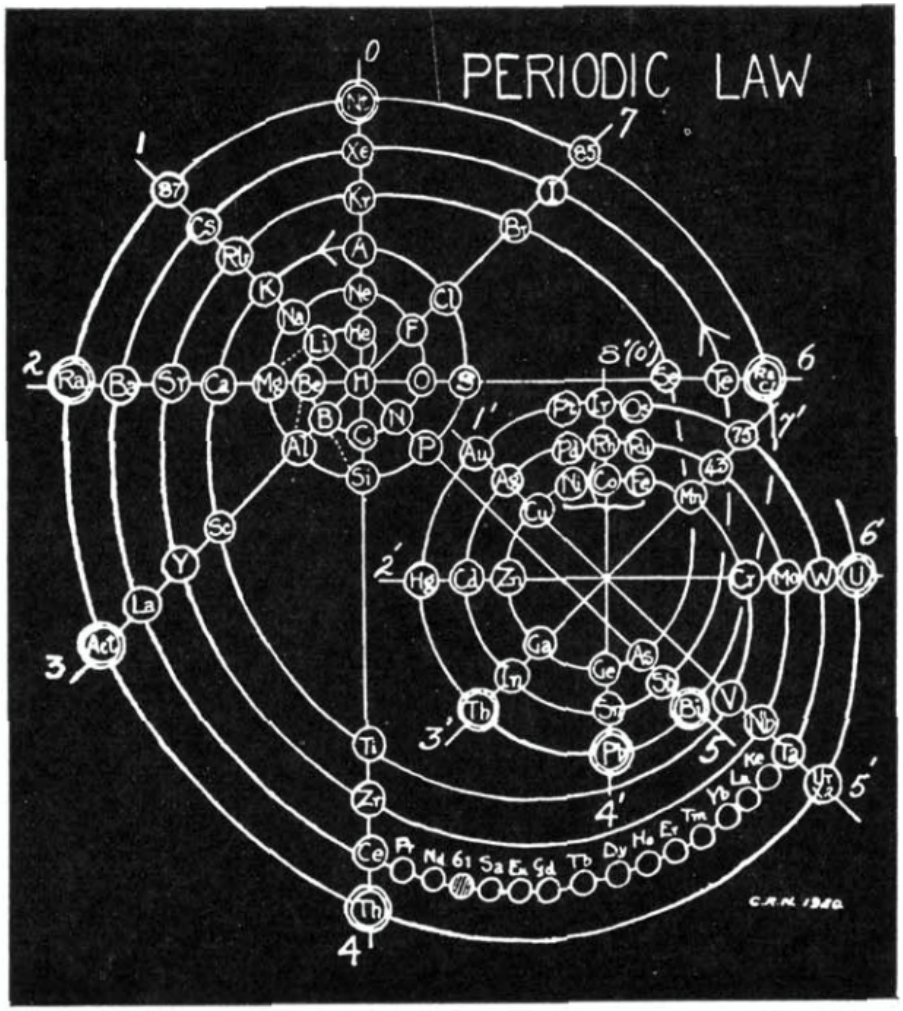


FIGURE 16.—NODDER'S PERIODIC TABLE

Quam & Quam, J Chem Educ (1934)

Is Z really a good variable?

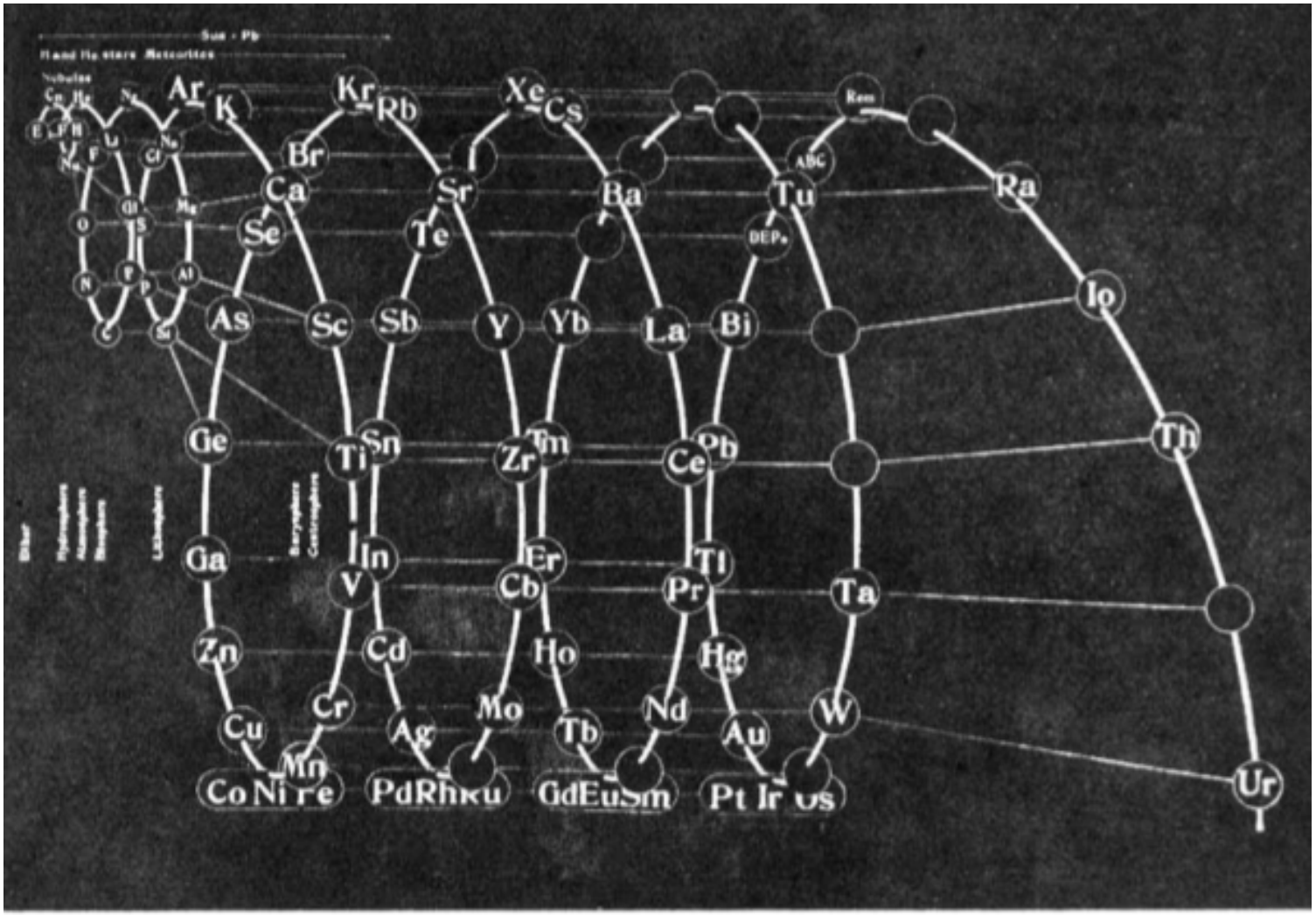
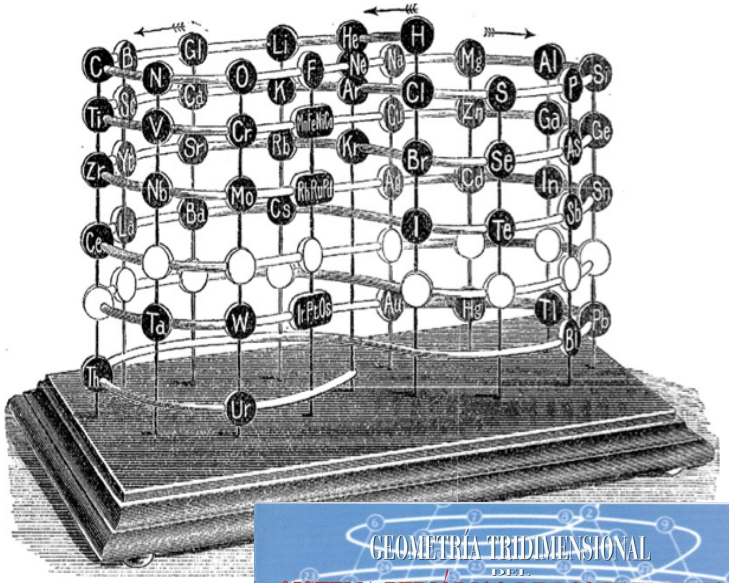


FIGURE 21.—EMERSON'S HELIX

Quam & Quam, J Chem Educ (1934)

Is Z really a good variable?



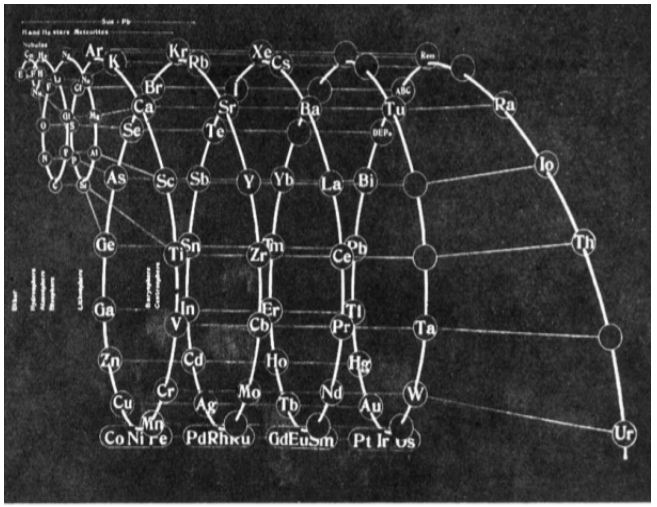
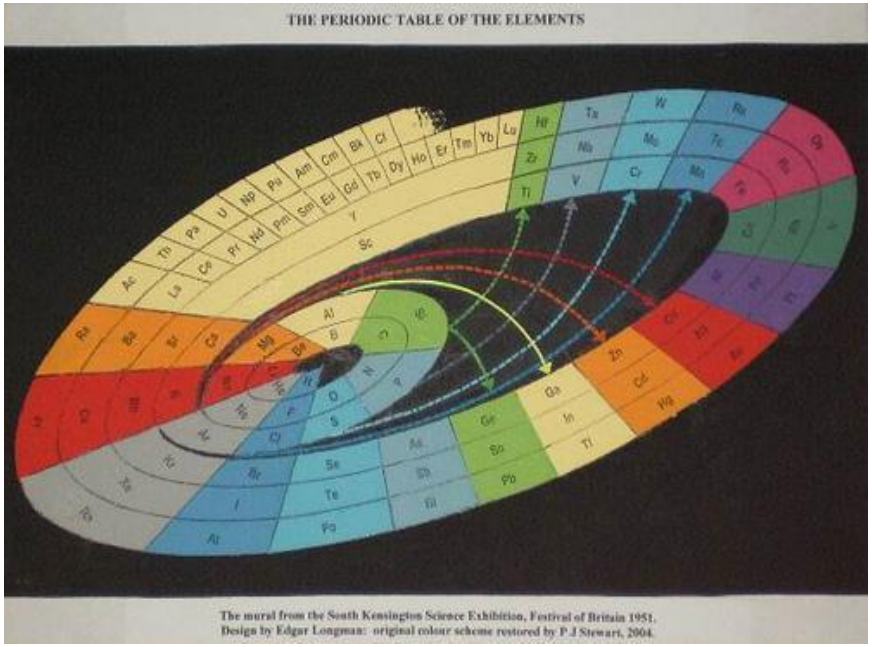
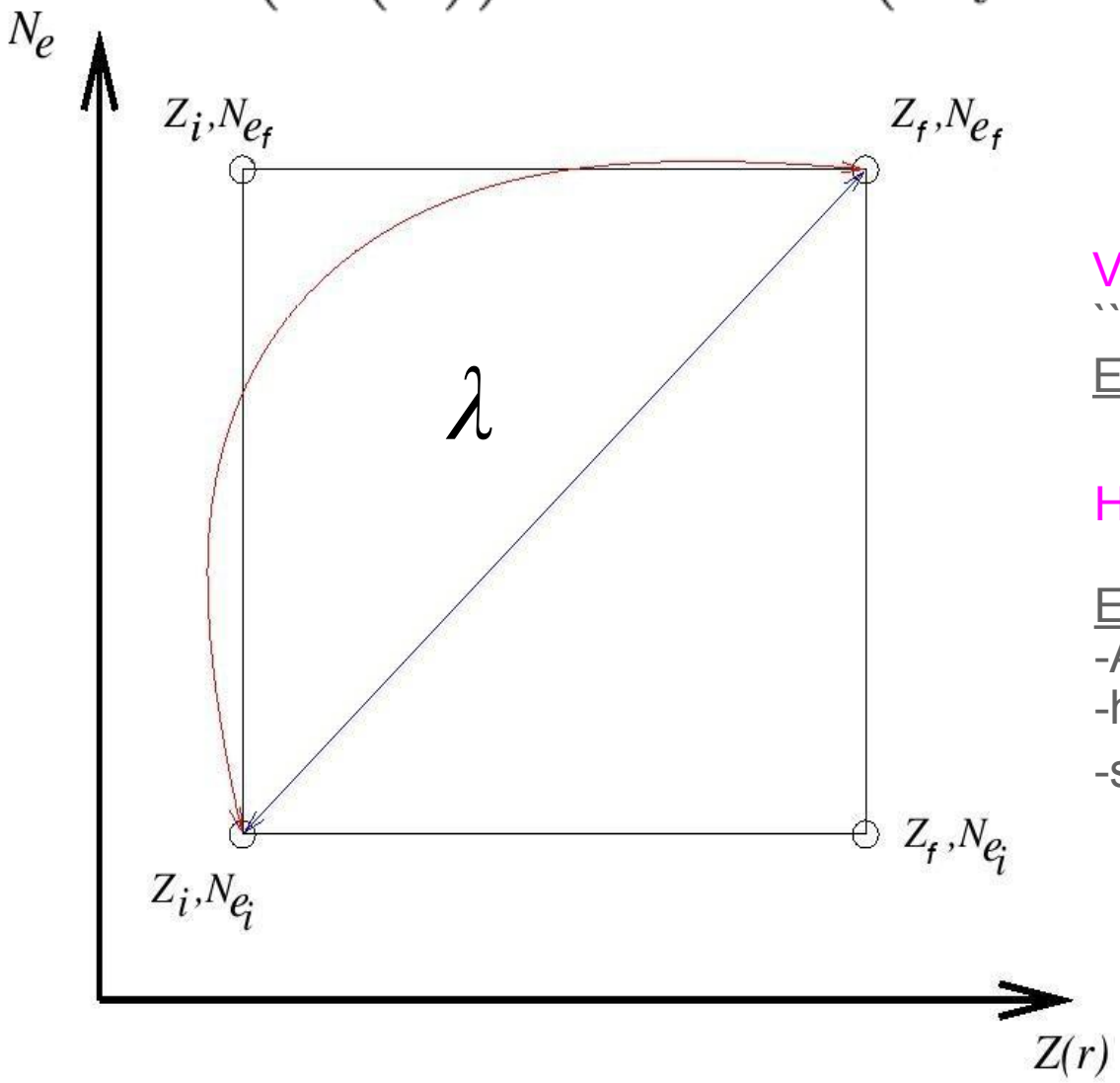


FIGURE 21.—EMERSON'S HELIX



Generalization

$$E(H(\lambda)) \quad = \quad E(H_i + \lambda(H_f - H_i))$$



Vertical changes:
``redox"
Example: $Li \rightarrow Li^+ + e^-$

Horizontal changes:
iso-electronic & ``alchemical"
Example:
-All constitutional isomers
-hydrazine (N_2H_4) $\rightarrow$ CH_3OH
-same number of valence electrons

First Principles

$$H(\{Z_I, \mathbf{R}_I\})\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

Schrödinger

variational (deductive)

Alchemy

1. Free energies
2. Gradients to optimize

Feynman

$$\frac{\partial E[H]}{\partial R_{Ix}} = \left\langle \Psi \left| \frac{\partial H}{\partial R_{Ix}} \right| \Psi \right\rangle$$

$$\frac{\partial E[H]}{\partial Z_I} = \left\langle \Psi \left| \frac{\partial H}{\partial Z_I} \right| \Psi \right\rangle$$

First Principles

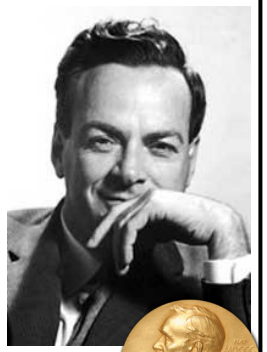
$$H(\{Z_I, \mathbf{R}_I\})\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

Schrödinger

variational (deductive)

Feynman



$$\frac{\partial E[H]}{\partial R_{Ix}} = \left\langle \Psi \left| \frac{\partial H}{\partial R_{Ix}} \right| \Psi \right\rangle$$

$$\frac{\partial E[H]}{\partial Z_I} = \left\langle \Psi \left| \frac{\partial H}{\partial Z_I} \right| \Psi \right\rangle$$

$$E(H(\lambda)) = E(H_i + \lambda(H_f - H_i))$$

$$\frac{\partial E[H]}{\partial \lambda} = \left\langle \Psi \left| \frac{\partial H(\lambda)}{\partial \lambda} \right| \Psi \right\rangle$$

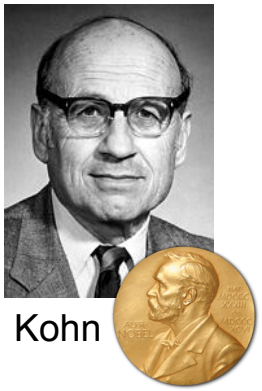
Alchemy

1. Free energies
2. Gradients to optimize

Generalization

$$
\begin{aligned}
E(H(\lambda)) &= E(H_i + \lambda(H_f - H_i)) \\
\frac{\partial E[H]}{\partial \lambda} &= \left\langle \Psi \left| \frac{\partial H(\lambda)}{\partial \lambda} \right| \Psi \right\rangle \\
&= \int d\mathbf{r}\; n_\lambda(\mathbf{r}) \times [v_j^{ext}(\mathbf{r}) - v_i^{ext}(\mathbf{r})]
\end{aligned}
$$

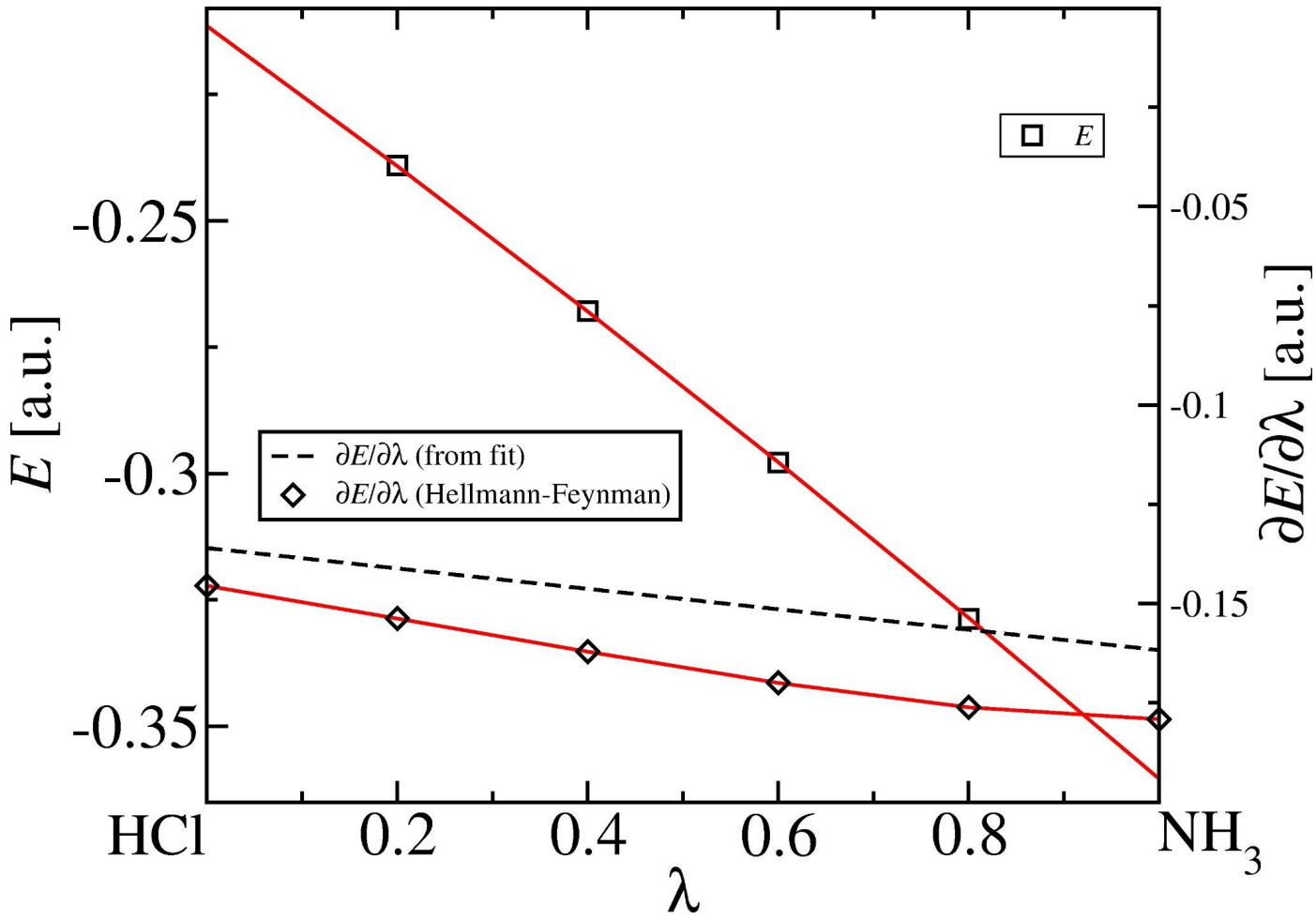
Hohenberg



Kohn

OAvL *J Chem Phys* (2009)

Generalization

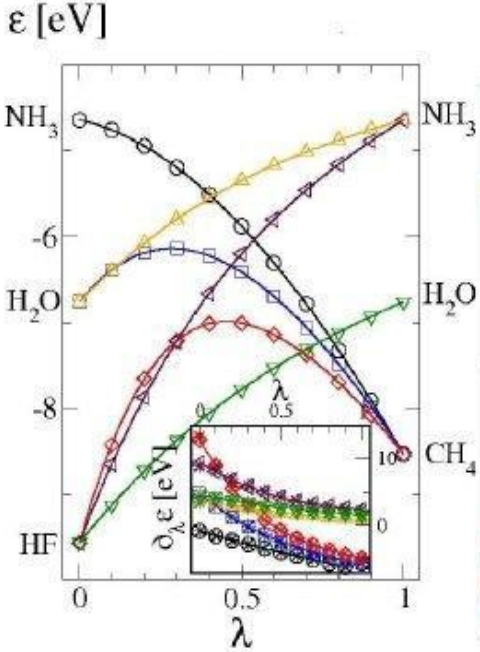


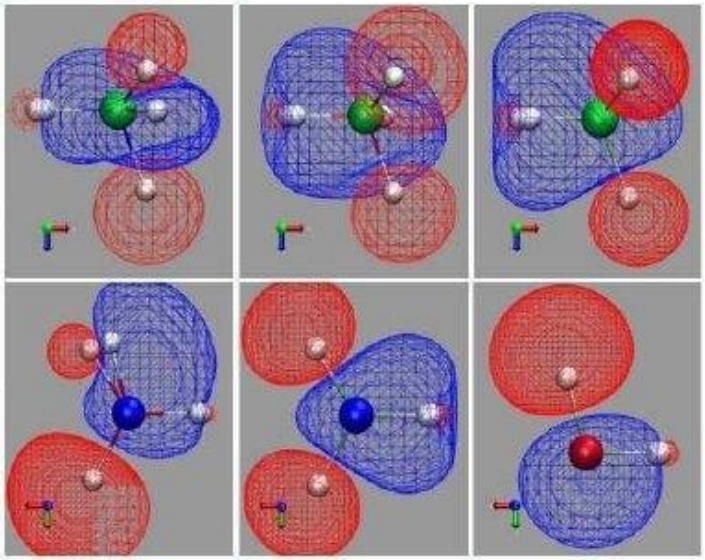
OAvL, *Int J Quant Chem*, (2013)

Example?

$$\epsilon(\lambda) = \frac{1}{\delta}\left(E(N_e,\lambda) - E(N_e-\delta,\lambda)\right)$$

$$\partial_\lambda E(\lambda) = \int d\mathbf{r}\, n_\lambda(\mathbf{r}) \times [v_j^{ext}(\mathbf{r}) - v_i^{ext}(\mathbf{r})]$$





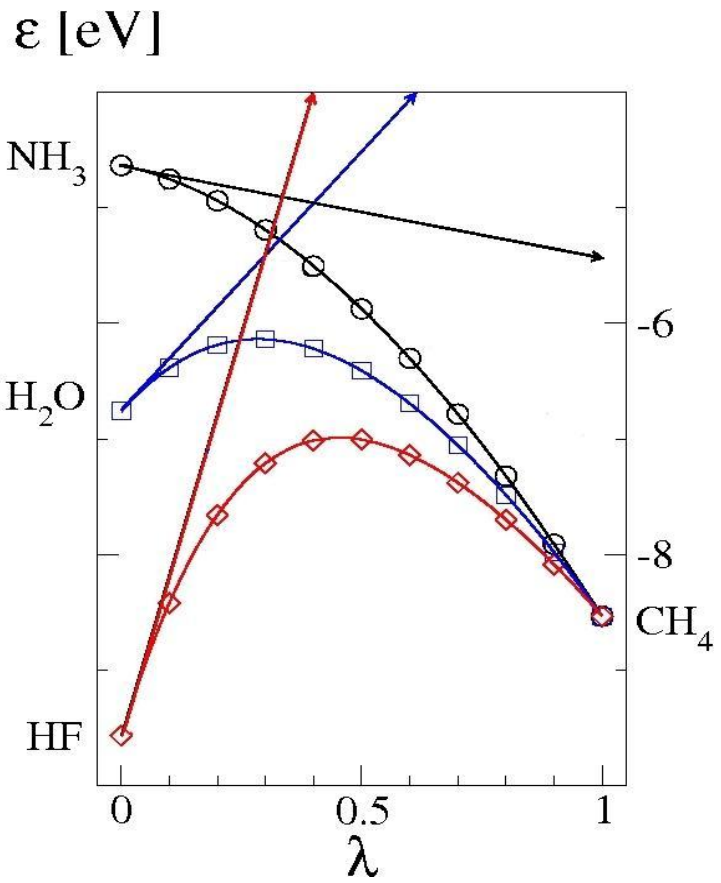
$$\partial_\lambda \epsilon(\lambda) = \frac{1}{\delta}\left(\int d\mathbf{r}[n_\lambda(\mathbf{r}) - n_\lambda^{+\delta}(\mathbf{r})] \times [v_j^{ext}(\mathbf{r}) - v_i^{ext}(\mathbf{r})]\right)$$

vs.

$$d\epsilon(\lambda)/d\lambda = \frac{1}{\delta}\left(\epsilon(\lambda+\delta) - \epsilon(\lambda)\right)$$

OAvL *JCP* (2009)

But what about prediction?

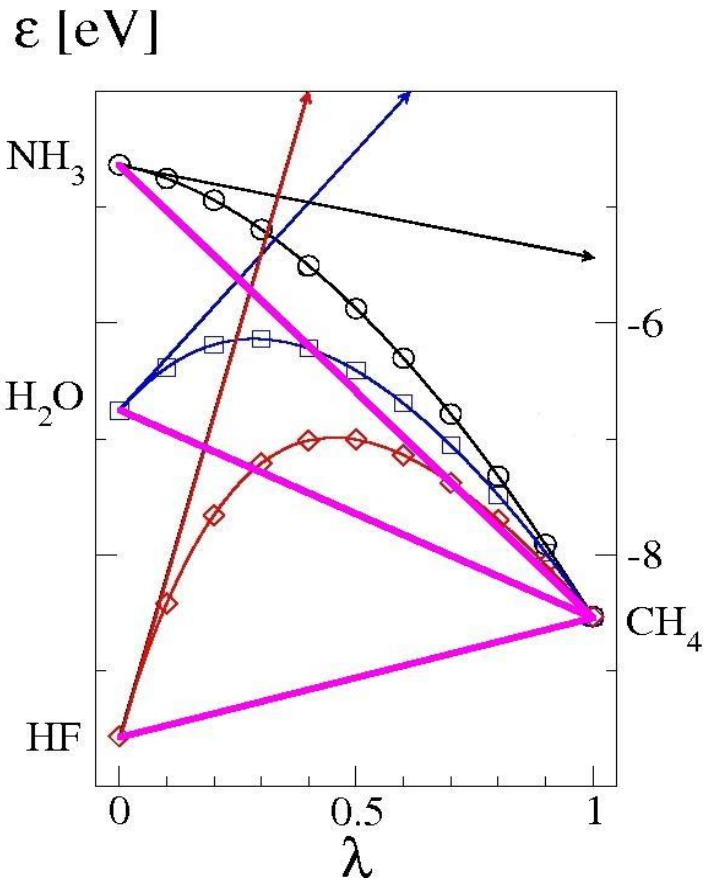


$$\varepsilon_{\lambda=1} \approx \varepsilon_{\lambda=0} + \left.\frac{\partial \varepsilon}{\partial \lambda}\right|_{\lambda=0} \Delta\lambda + H.O.T.$$

$$\Delta\lambda = 1$$

OAvL *JCP* (2009)

Prediction?

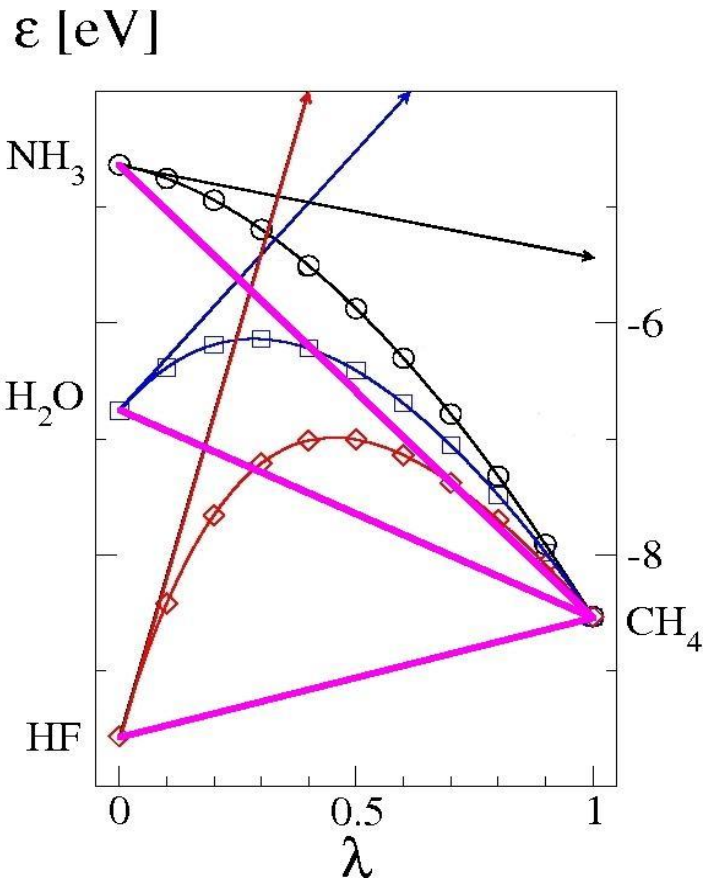


$$\varepsilon_{\lambda=1} \approx \varepsilon_{\lambda=0} + \left.\frac{\partial \varepsilon}{\partial \lambda}\right|_{\lambda=0} \Delta\lambda + H.O.T.$$

$$\Delta\lambda = 1$$

OAvL *JCP* (2009)

Prediction?



$$\varepsilon_{\lambda=1} \approx \varepsilon_{\lambda=0} + \left.\frac{\partial \varepsilon}{\partial \lambda}\right|_{\lambda=0} \Delta\lambda + H.O.T.$$

$$\Delta\lambda = 1$$

In analogy to:
Smith and van Gunsteren *JCP* (1994)

$$E^{lin} = E_i + \lambda \times (E_f - E_i)$$

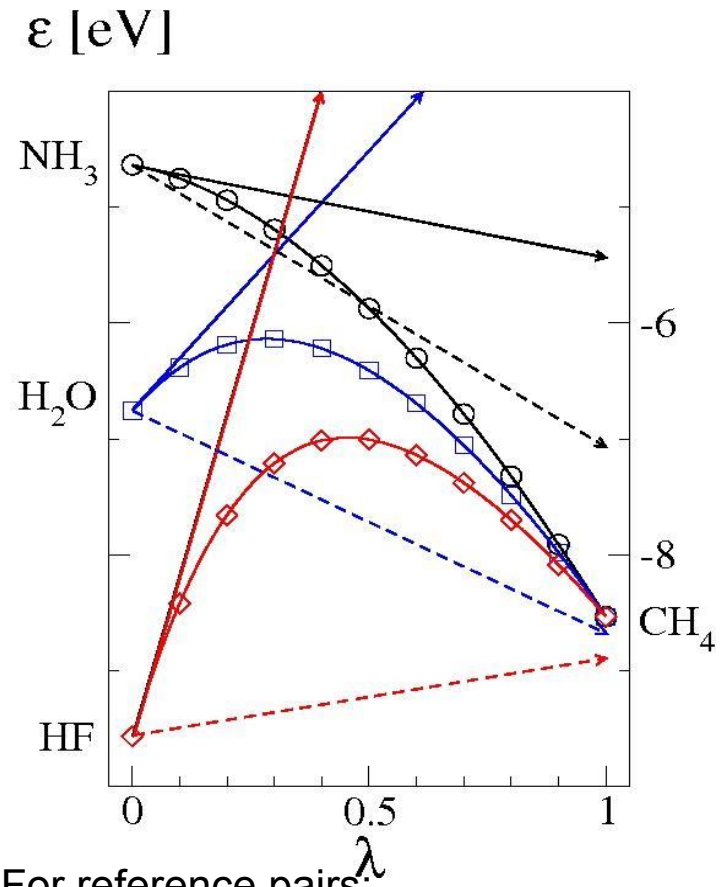
$$= \langle H_i + f_{if}(\lambda) \times (H_f - H_i)] \rangle_\lambda$$

$$f_{if}(\lambda) = \begin{cases} 0 & \text{if} \quad \lambda = 0 \\ 1 & \text{if} \quad \lambda = 1 \end{cases}$$

$$f_{if}(\lambda) = a_{if}(\lambda^2 - \lambda) + \lambda$$

OAvL *JCP* (2009)

Prediction?



$$\varepsilon_{\lambda=1} \approx \varepsilon_{\lambda=0} + \left.\frac{\partial \varepsilon}{\partial \lambda}\right|_{\lambda=0} \Delta\lambda + H.O.T.$$

$$\Delta\lambda = 1$$

In analogy to:
Smith and van Gunsteren *JCP* (1994)

$$E^{lin} = E_i + \lambda \times (E_f - E_i)$$

$$= \langle H_i + f_{if}(\lambda) \times (H_f - H_i)] \rangle_\lambda$$

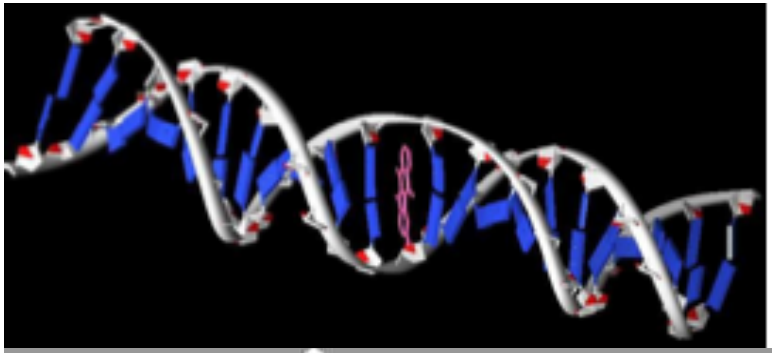
$$f_{if}(\lambda) = \begin{cases} 0 & \text{if} \quad \lambda = 0 \\ 1 & \text{if} \quad \lambda = 1 \end{cases}$$

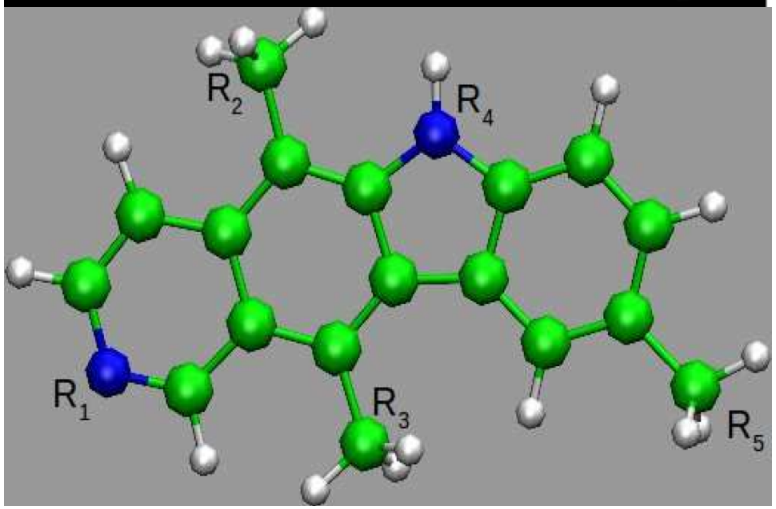
$$f_{if}(\lambda) = a_{if}(\lambda^2 - \lambda) + \lambda$$

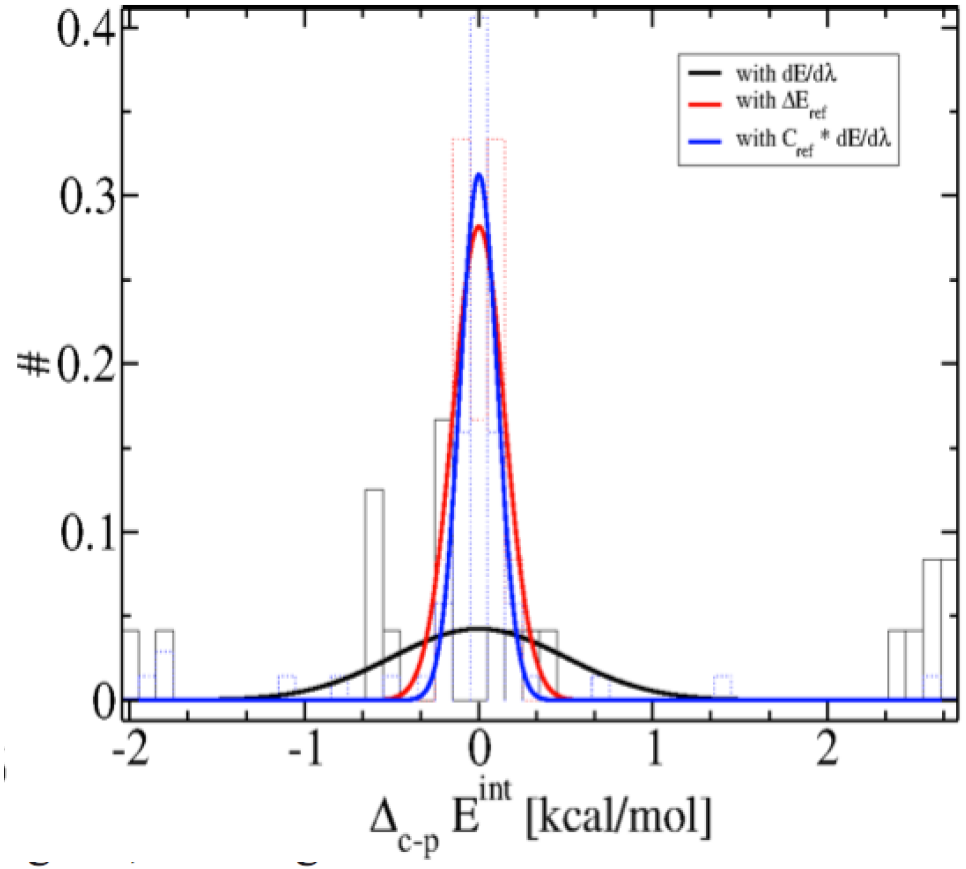
For reference pairs:
$CH_3NH_2 \rightarrow CH_3CH_3$
$CH_3OH \rightarrow CH_3CH_3$
$CH_3F \rightarrow CH_3CH_3$

OAvL *JCP* (2009)

Drug design?







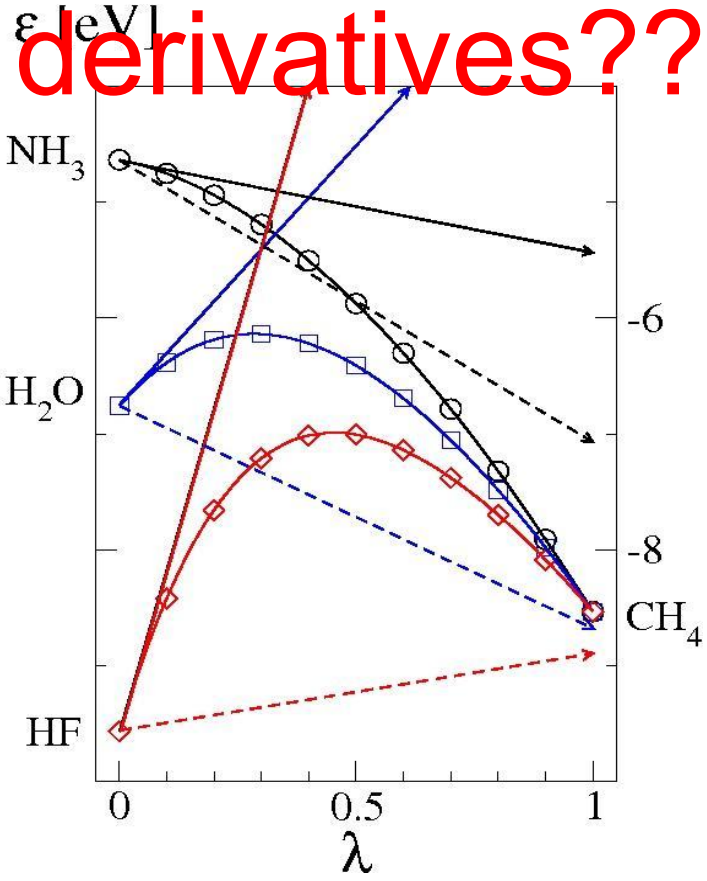
Ellipticine, intercalated between 2 Watson-Crick base-pairs w backbone, using vdW+DFT (GGA+DCACP)
J Phys Chem B (2007)

OAvL, Int J Quant Chem (2013)

site vs. group	1	2	3	4	5	6
R_1	CH	N	SiH	P	-	-
R_2	CH_3	NH_2	OH^{left}	OH^{right}	F	Cl
R_3	CH_3	NH_2	OH^{left}	OH^{right}	F	Cl
R_4	CH_2	NH	O	SiH_2	PH	S
R_5	CH_3	NH_2	OH^{left}	OH^{right}	F	Cl

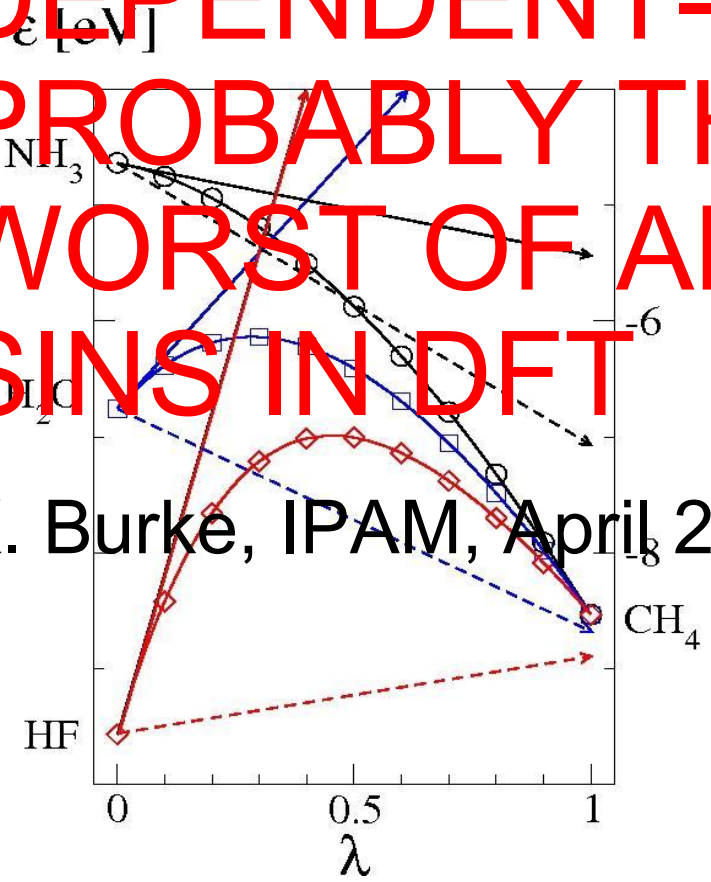
Prediction?

System dependent derivatives???

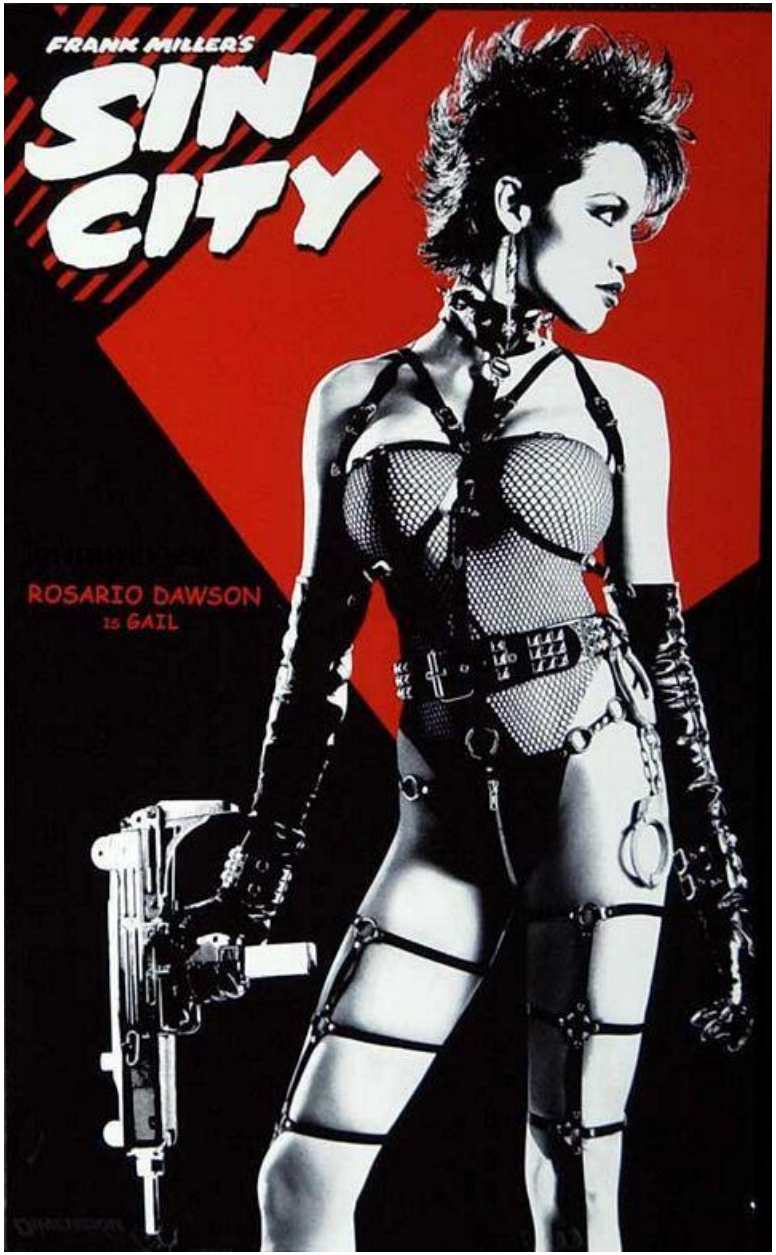


Prediction?

SYSTEM DEPENDENT---PROBABLY THE WORST OF ALL SINS IN DFT

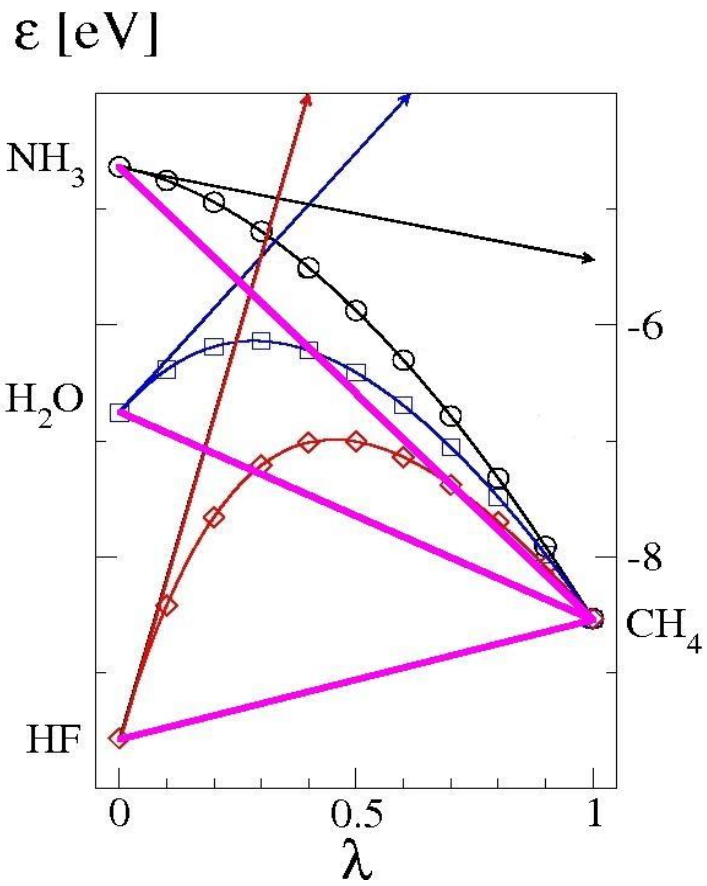


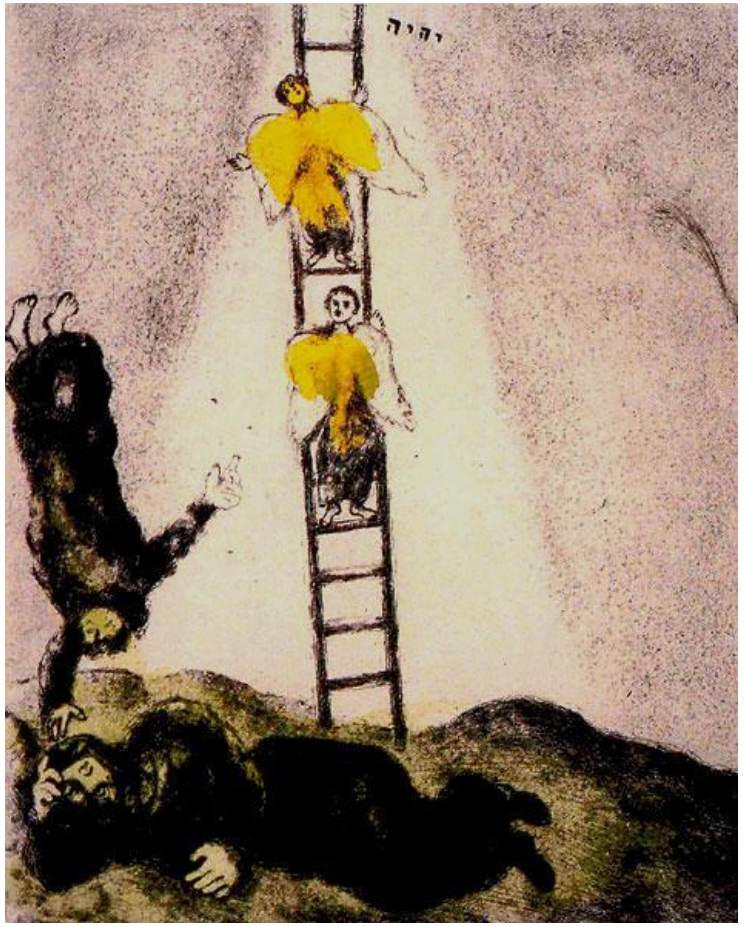
K. Burke, IPAM, April 2011



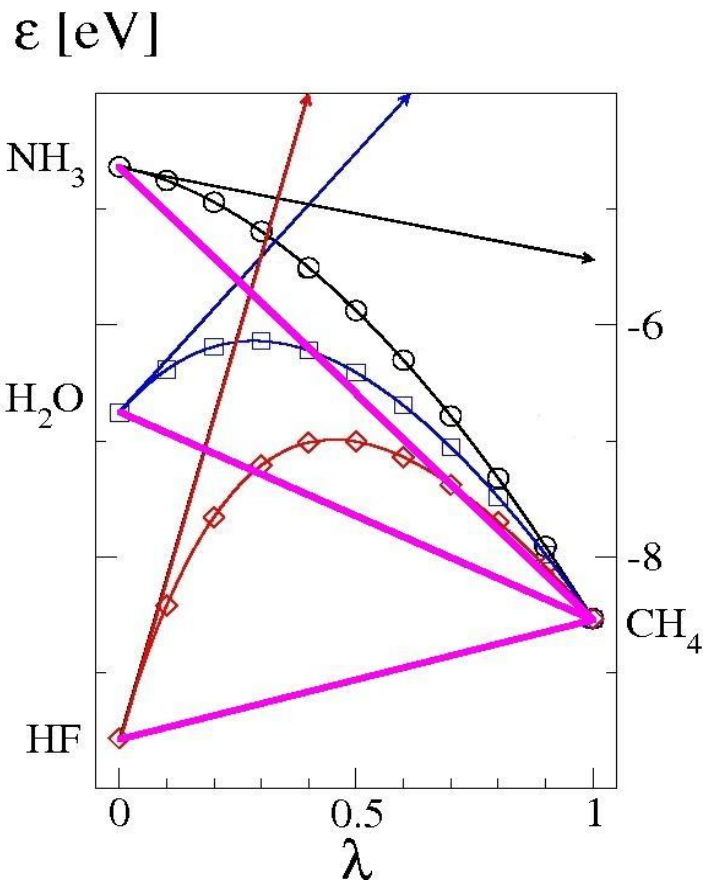
Help!

WANTED!
Jacob's Ladder for CCS





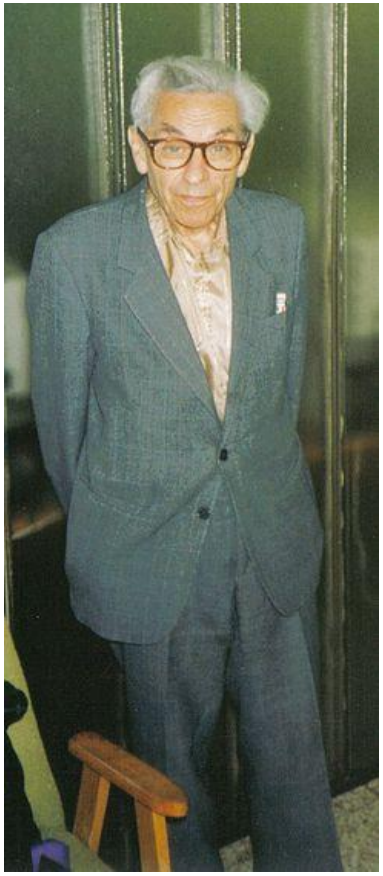
Help! → Swarm Intelligence



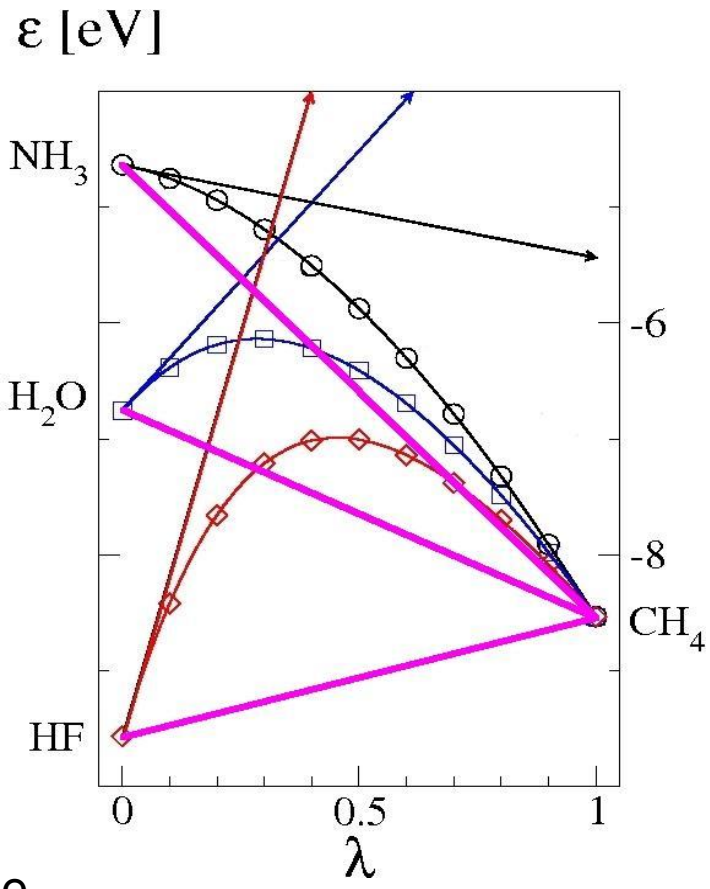
Erdős problems

Throughout his career, Erdős would offer US$ prizes for solutions to unresolved problems.

http://wikipedia.org



Win a prize!!!



See
www.alcf.anl.gov/~anatole
For more info

An ounce of Gold in the form of 100 shares in iShares Trust (IAU) --- currently worth a total of ~US$1.7k

for the first person who presents a solution to this problem:

Find---or show non-existence of---a system independent (i.e. valid for all of CCS as defined above) interpolating function *f* for which two differing (iso-)electronic Hamiltonians transform such that

$$\left.\frac{\partial E(\lambda)}{\partial \lambda}\right|_{\lambda=0} = \left\langle \frac{\partial H(f_{if}(\lambda))}{\partial \lambda} \right\rangle_{\lambda=0} = E_f - E_i$$

where

$$0 \le \lambda \le 1$$

$$E(\lambda=0) = \langle H(f(\lambda=0)) \rangle = \langle H_i \rangle = E_i$$

$$E(\lambda=1) = \langle H(f(\lambda=1)) \rangle = \langle H_f \rangle = E_f$$

First Principles

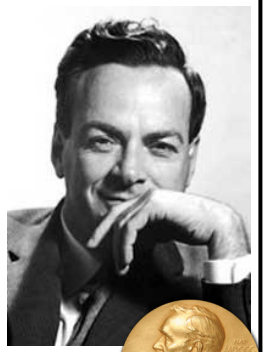
$$H(\{Z_I, \mathbf{R}_I\})\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

Schrödinger

variational (deductive)

Feynman



$$\frac{\partial E[H]}{\partial R_{Ix}} = \left\langle \Psi \left| \frac{\partial H}{\partial R_{Ix}} \right| \Psi \right\rangle$$

$$\frac{\partial E[H]}{\partial Z_I} = \left\langle \Psi \left| \frac{\partial H}{\partial Z_I} \right| \Psi \right\rangle$$

$$E(H(\lambda)) = E(H_i + \lambda(H_f - H_i))$$

$$\frac{\partial E[H]}{\partial \lambda} = \left\langle \Psi \left| \frac{\partial H(\lambda)}{\partial \lambda} \right| \Psi \right\rangle$$

Alchemy

1. Free energies
2. Gradients to optimize

First Principles

$$H(\{Z_I, \mathbf{R}_I\})\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

Schrödinger

correlational (inductive)

Vapnik

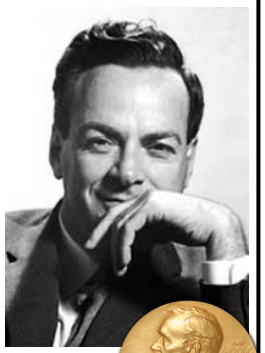
$$\{Z_I, \mathbf{R}_I\} \overset{H\Psi}{\longmapsto} E$$

supervised learning

$$\{Z_I, \mathbf{R}_I\} \overset{\mathrm{ML}}{\longmapsto} E$$

variational (deductive)

Feynman



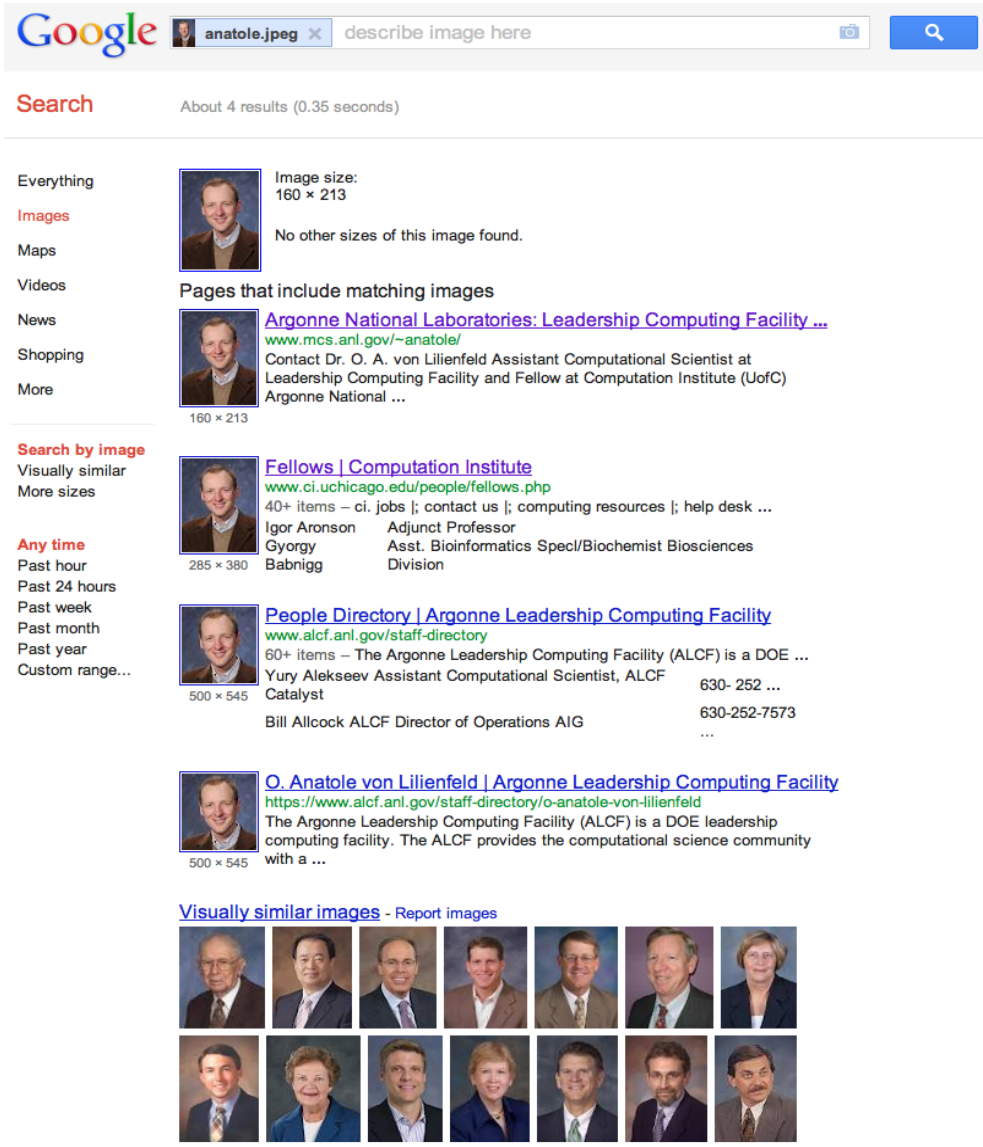
$$\frac{\partial E[H]}{\partial R_{Ix}} = \left\langle \Psi \left| \frac{\partial H}{\partial R_{Ix}} \right| \Psi \right\rangle$$

$$\frac{\partial E[H]}{\partial Z_I} = \left\langle \Psi \left| \frac{\partial H}{\partial Z_I} \right| \Psi \right\rangle$$

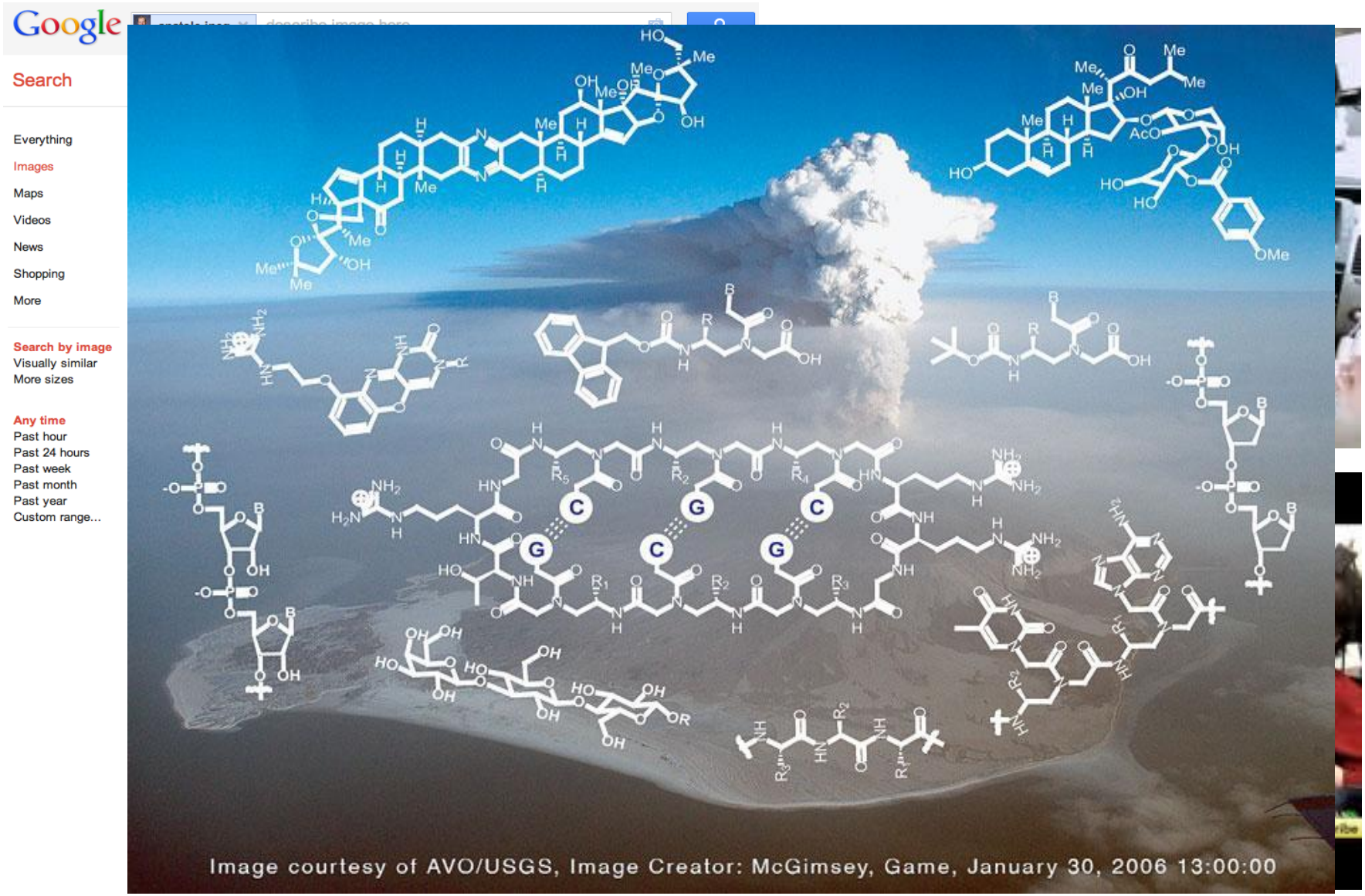
$$E(H(\lambda)) = E(H_i + \lambda(H_f - H_i))$$

$$\frac{\partial E[H]}{\partial \lambda} = \left\langle \Psi \left| \frac{\partial H(\lambda)}{\partial \lambda} \right| \Psi \right\rangle$$

Correlational: Machine Learning



Correlational: Machine Learning



First Principles

$$H(\{Z_I, \mathbf{R}_I\})\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

Schrödinger

correlational (inductive)

Vapnik

$$\{Z_I, \mathbf{R}_I\} \overset{H\Psi}{\longmapsto} E$$

supervised learning

$$\{Z_I, \mathbf{R}_I\} \overset{\text{ML}}{\longmapsto} E$$

Infer solution by comparison to previous examples

- Regression method?
- Function?
- Variables?
- Metric?
- Data?

$$\{Z_I, \mathbf{R}_I\} \xrightarrow{\text{ML}} E$$

Non-linear function

$$E^{est}(\mathbf{M}) = \sum_i \alpha_i e^{-\frac{d(\mathbf{M},\mathbf{M}_i)^2}{2\sigma^2}}$$

$$\{Z_I, \mathbf{R}_I\} \overset{\text{ML}}{\longmapsto} E$$

Non-linear function

$$E^{est}(\mathbf{M}) = \sum_i \alpha_i e^{-\frac{d(\mathbf{M},\mathbf{M}_i)^2}{2\sigma^2}}$$

Desirable descriptors are

- unique
- translation invariant
- rotation invariant
- symmetry invariant
- index invariant
- constant length

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

$$\{Z_I, \mathbf{R}_I\} \overset{\text{ML}}{\longmapsto} E$$

Non-linear function

$$E^{est}(\mathbf{M}) = \sum_i \alpha_i e^{-\frac{d(\mathbf{M},\mathbf{M}_i)^2}{2\sigma^2}}$$

Desirable descriptors are

- unique
- translation invariant
- rotation invariant
- symmetry invariant
- index invariant
- constant length

$$M_{IJ} = \begin{cases} 0.5 Z_I^{2.4} & \forall \ I = J, \\ \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} & \forall \ I \neq J. \end{cases}$$

Coulomb-matrix

- unique
- translation
- rotation
- symmetry
- sort/diagonalize
- fill up w zeros

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

$$\{Z_I, \mathbf{R}_I\} \overset{\text{ML}}{\longmapsto} E$$

Non-linear function

$$E^{est}(\mathbf{M}) = \sum_i \alpha_i e^{-\frac{d(\mathbf{M},\mathbf{M}_i)^2}{2\sigma^2}}$$

Desirable descriptors are

- unique
- translation invariant
- rotation invariant
- symmetry invariant
- index invariant
- constant length

$$M_{IJ} = \begin{cases} 0.5 Z_I^{2.4} & \forall \;\; I = J, \\ \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} & \forall \;\; I \neq J. \end{cases}$$

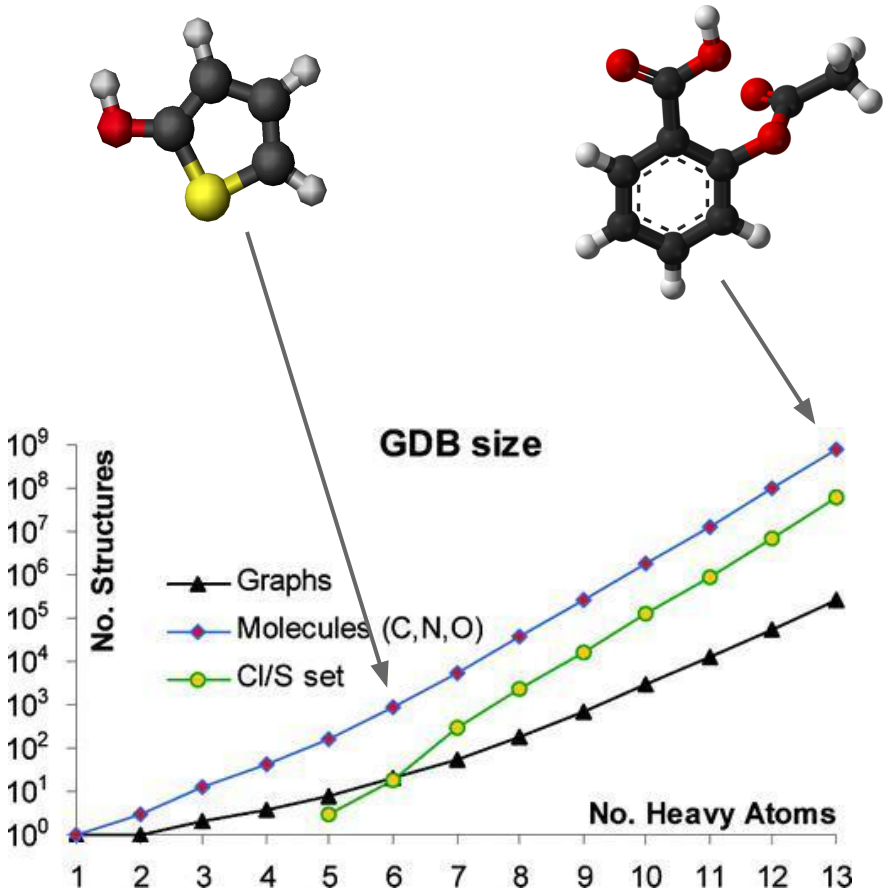
Coulomb-matrix

- unique
- translation
- rotation
- symmetry
- sort/diagonalize
- fill up w zeros

Euclidean distance

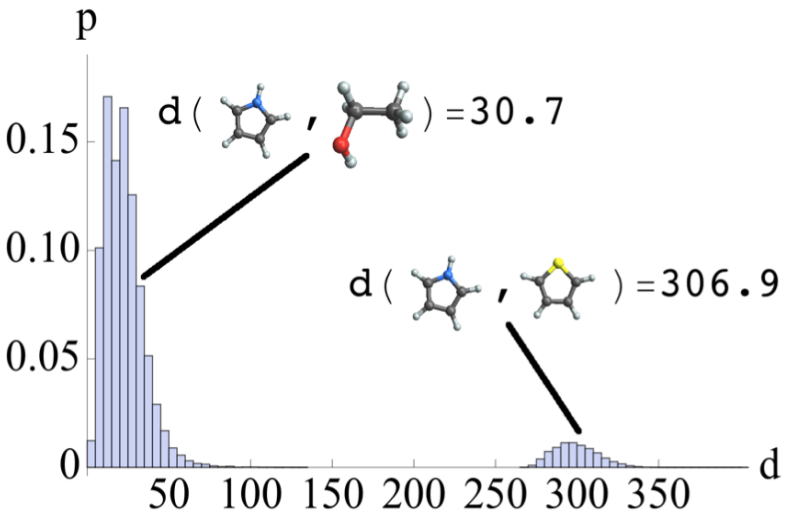
$$d(\mathbf{M}, \mathbf{M}_i) = \sqrt{\sum_{IJ} |M_{IJ} - M_{IJ}^{(i)}|^2}$$

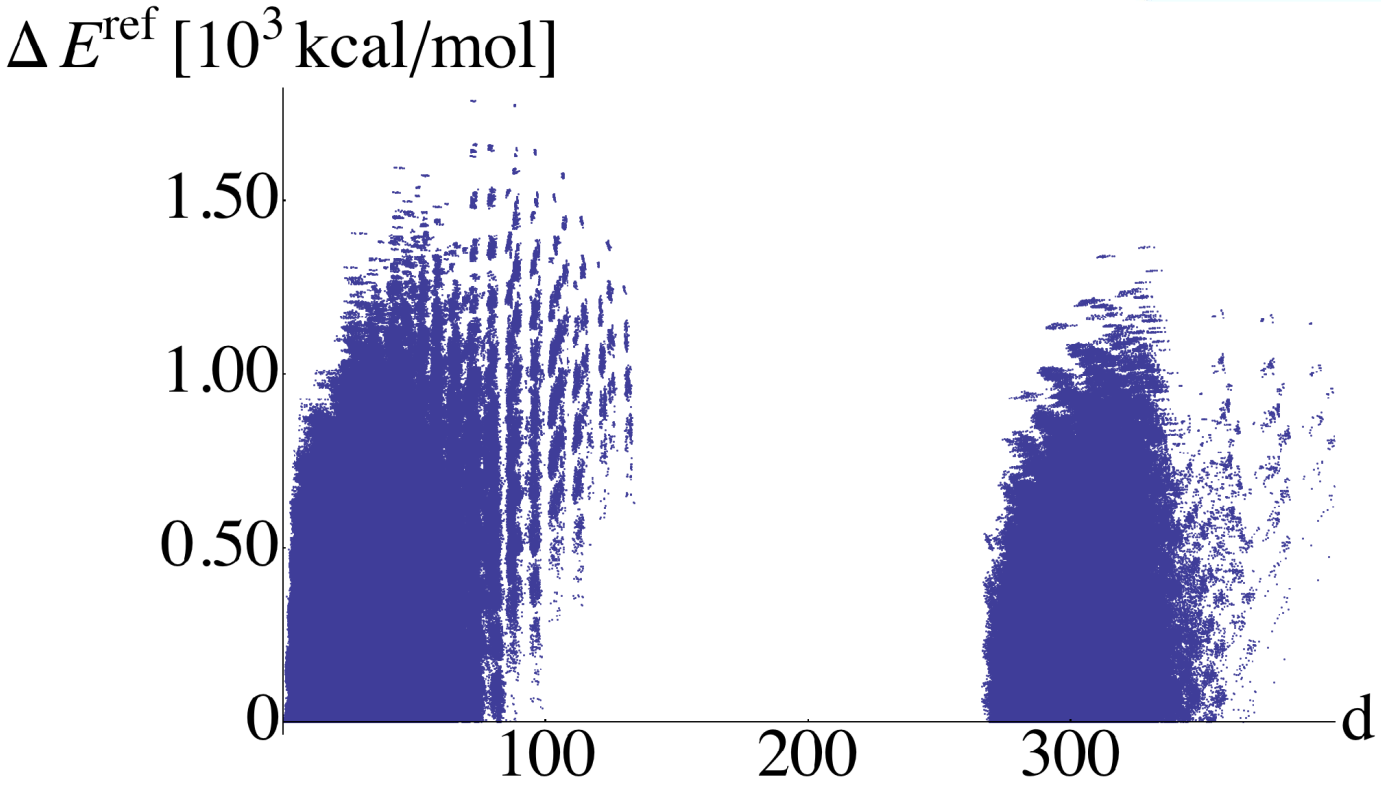
GDB: All organic molecules up to 13 atoms



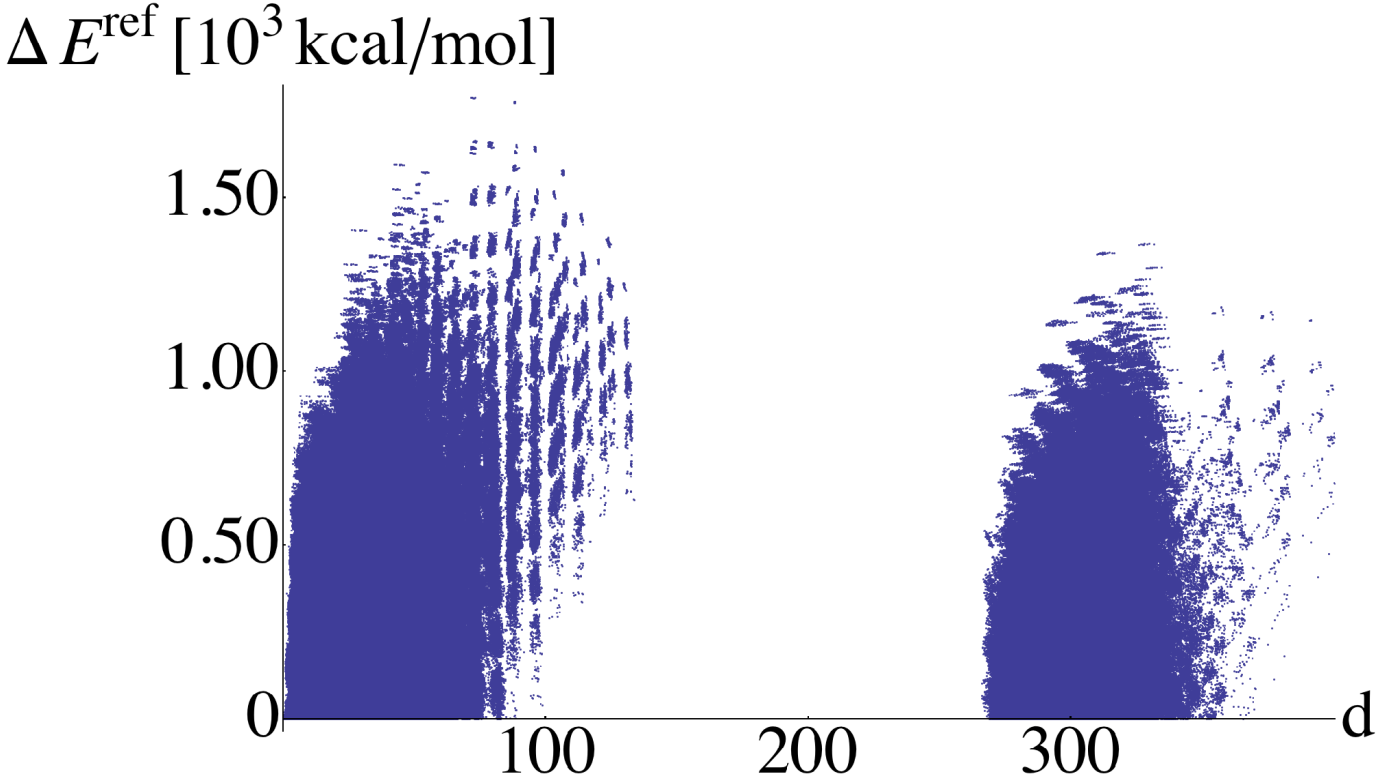
Fink, Bruggesser, Reymond *ACIE* (2005),
Blum, Reymond *JACS* (2009)

1. 7k compositional & constitutional isomers
2. Initial coordinates from universal force field [Goddard et al JACS (1992)]
3. Relax geometry with DFT
4. Calculate atomization energies





$$\min_{\alpha} \quad \sum_i \left(E^{est}(\mathbf{M}_i) - E_i^{ref}\right)^2 + \lambda \sum_i \alpha_i^2$$

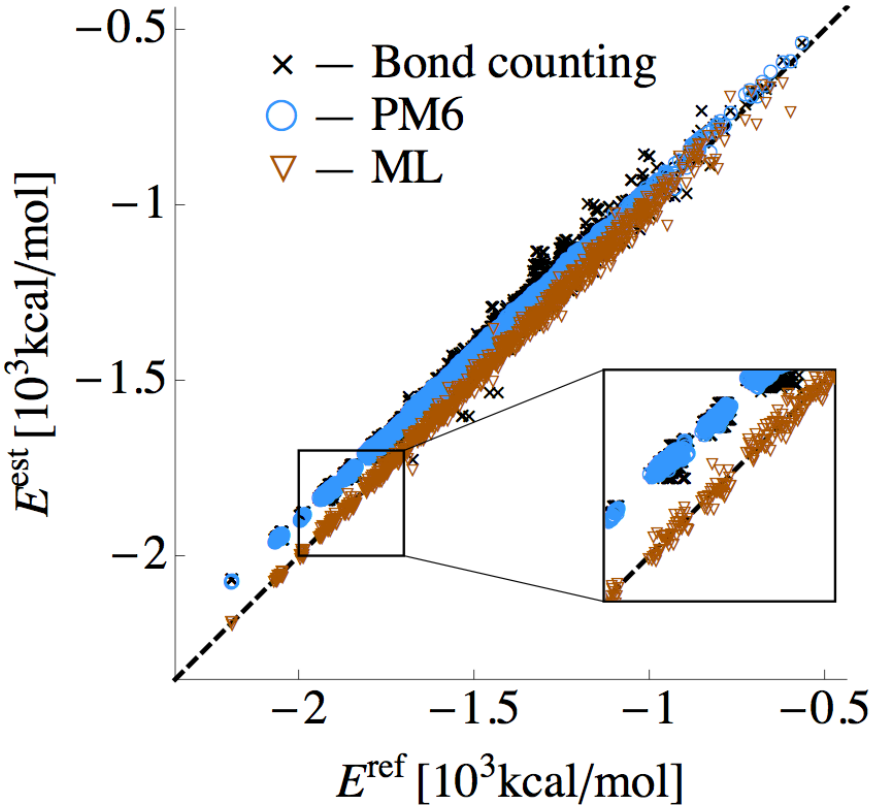


$$\min_{\alpha} \quad \sum_i \left(E^{est}(\mathbf{M}_i) - E_i^{ref}\right)^2 + \lambda \sum_i \alpha_i^2$$

$$\alpha \;=\; (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{E}^{ref}$$

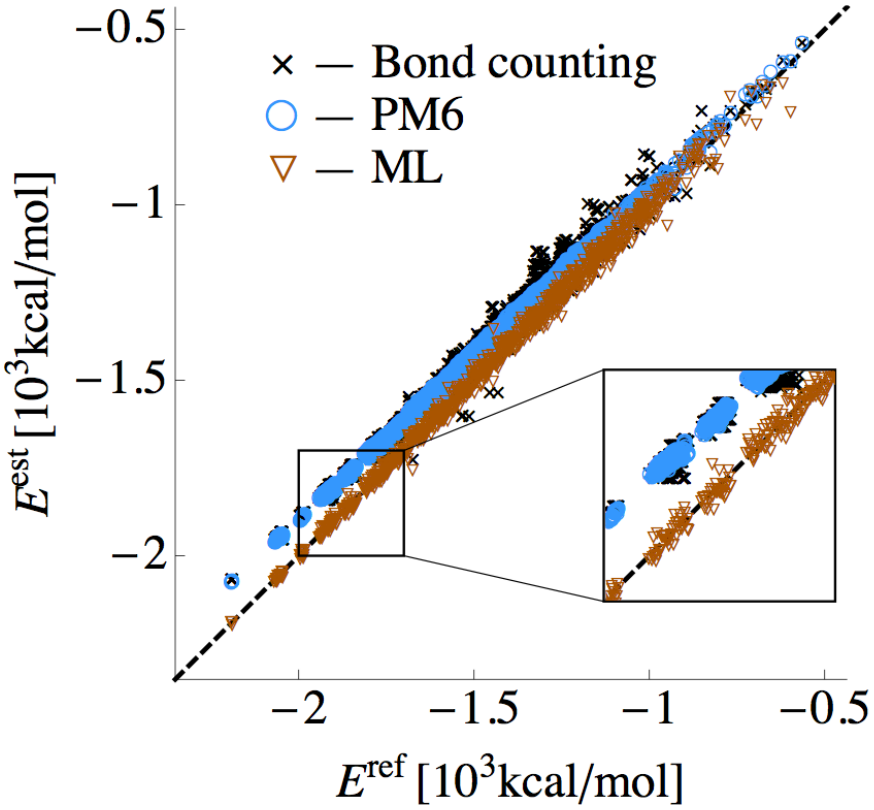
$$E^{est}(\mathbf{M}) \;=\; \sum_i \alpha_i e^{-\frac{d(\mathbf{M},\mathbf{M}_i)^2}{2\sigma^2}}$$

$$k(\mathbf{M},\mathbf{M}') \;=\; \exp\left(-\frac{d(\mathbf{M},\mathbf{M}')^2}{2\sigma^2}\right)$$



Training for N = 1000 molecules

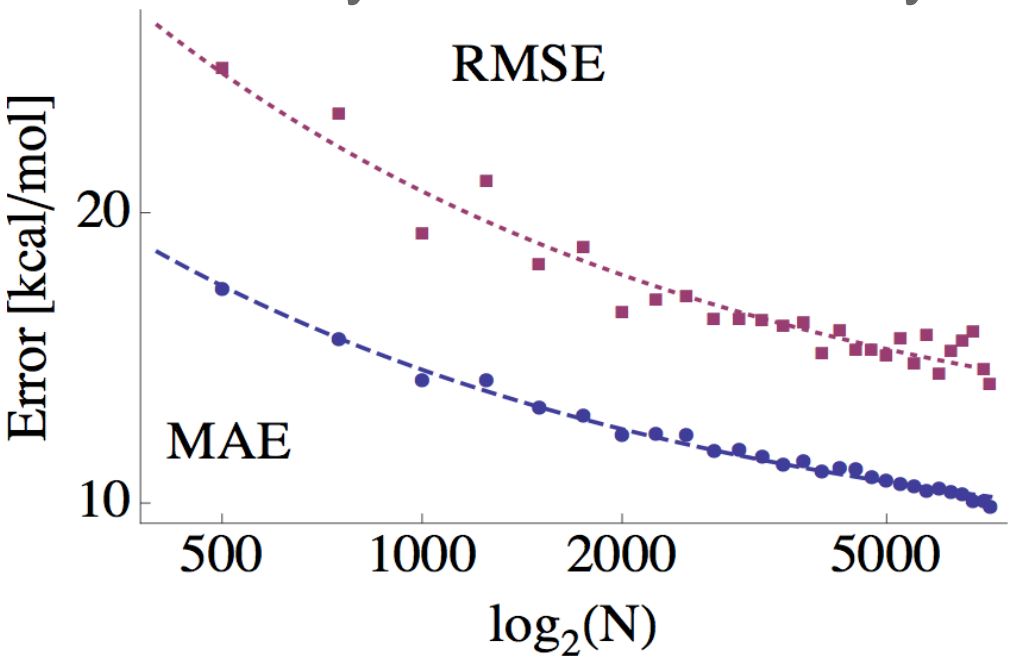
MAE ~15 kcal/mol

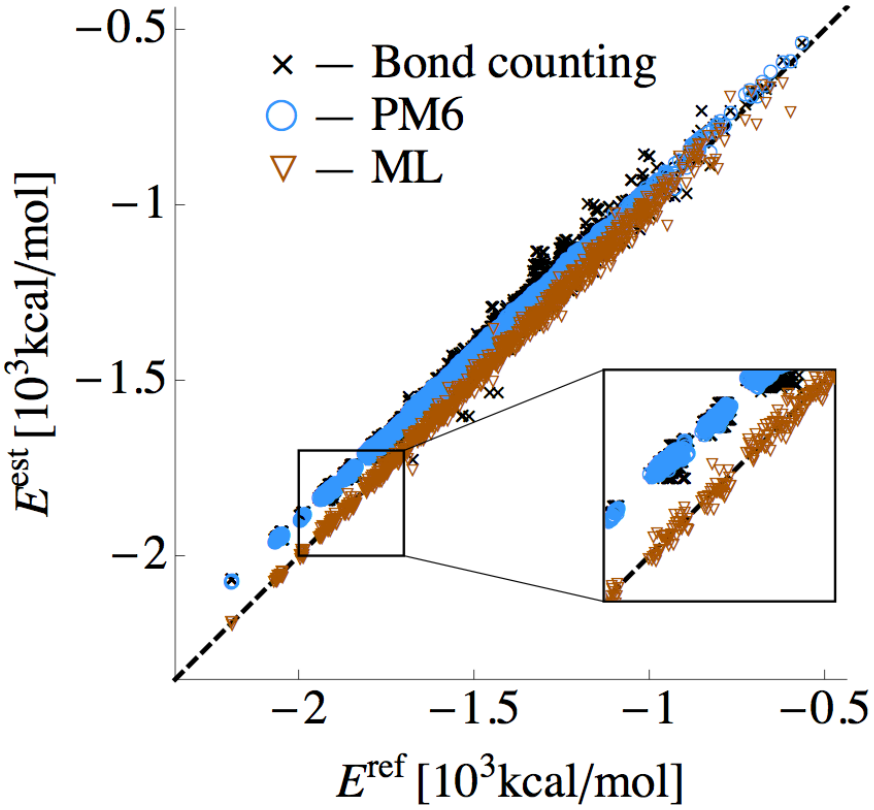


Training for N = 1000 molecules

MAE ~15 kcal/mol

Systematic eror decay

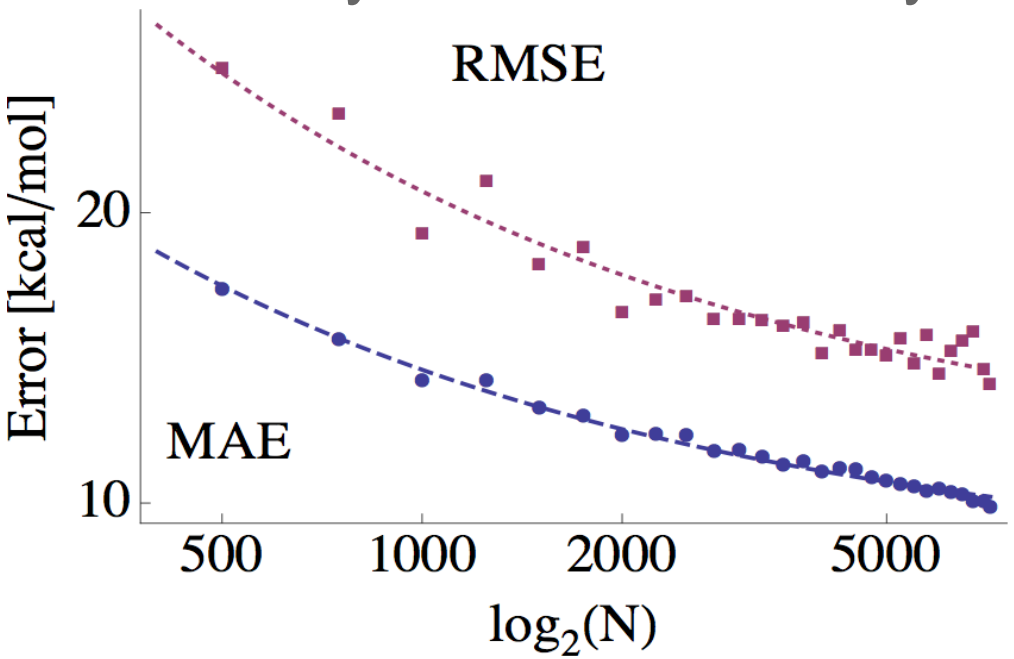




Training for N = 1000 molecules

MAE ~15 kcal/mol

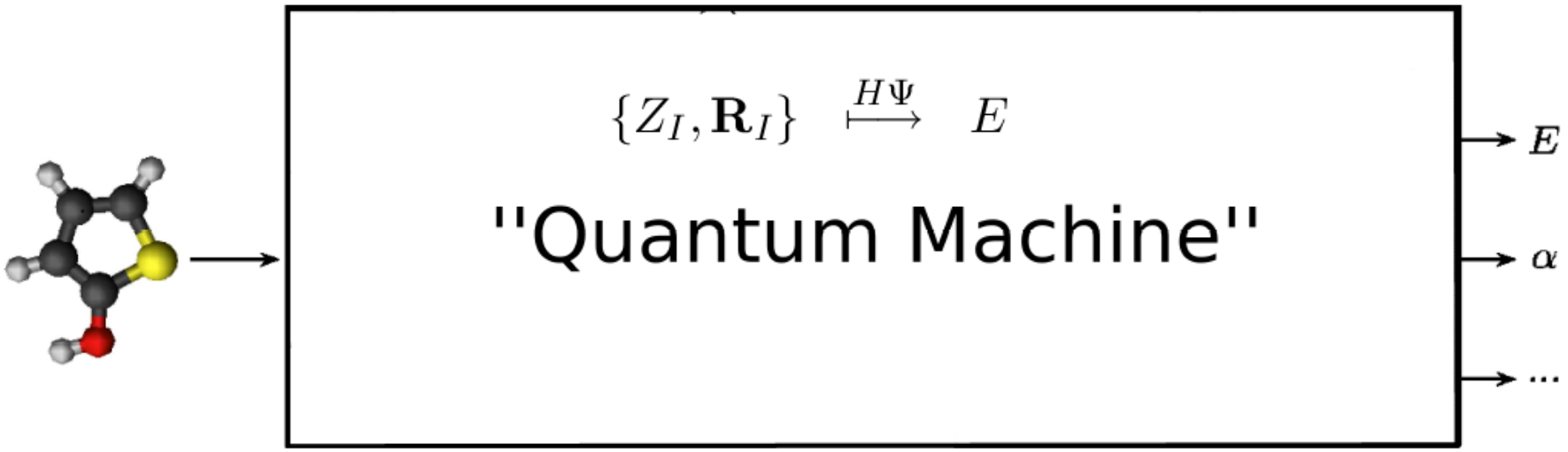
Systematic eror decay



PBE0: ~1000 seconds

ML: ~milli seconds

http://www.quantum-machine.org/



Tkatchenko (FHI)

Rupp (ETHZ)

Müller (TU Berlin)

M. Rupp, A. Tkatchenko, K.-R. Müller, OAvL, *Phys Rev Lett* (2012)

Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning

Matthias Rupp,[1,2] Alexandre Tkatchenko,[3,2] Klaus-Robert Müller,[1,2] and O. Anatole von Lilienfeld[4,2,*]

[1]*Machine Learning Group, Technical University of Berlin, Franklinstr 28/29, 10587 Berlin, Germany*

[2]*Institute of Pure and Applied Mathematics, University of California Los Angeles, Los Angeles, California 90095, USA*

[3]*Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin, Germany*

[4]*Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, Illinois 60439, USA*

(Received 15 June 2011; published 31 January 2012)

We introduce a machine learning model to predict atomization energies of a diverse set of organic molecules, based on nuclear charges and atomic positions only. The problem of solving the molecular Schrödinger equation is mapped onto a nonlinear statistical regression problem of reduced complexity. Regression models are trained on and compared to atomization energies computed with hybrid density-functional theory. Cross validation over more than seven thousand organic molecules yields a mean absolute error of ~10 kcal/mol. Applicability is demonstrated for the prediction of molecular atomization potential energy curves.

PRL 108, 058301 (2012) PHYSICAL REVIEW LETTERS week ending 3 FEBRUARY 2012

Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning

Matthias Rupp,[1,2] Alexandre Tkatchenko,[3,2] Klaus-Robert Müller,[1,2] and O. Anatole von Lilienfeld[4,2,*]

[1]Machine Learning Group, Technical University of Berlin, Franklinstr. 28/29, 10587 Berlin, Germany

PRL 109, 059801 (2012) PHYSICAL REVIEW LETTERS week ending 3 AUGUST 2012

Comment on "Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning"

In a recent Letter [1], the authors construct a machine learning (ML) model of molecular atomization energies, which they compare to bond counting (BC) and the PM6 semiempirical method [2]. However, their ML model was trained and tested on density functional theory (DFT)

Jonathan E. Moussa*
Sandia National Laboratories
Albuquerque, New Mexico 87185, USA

distributions

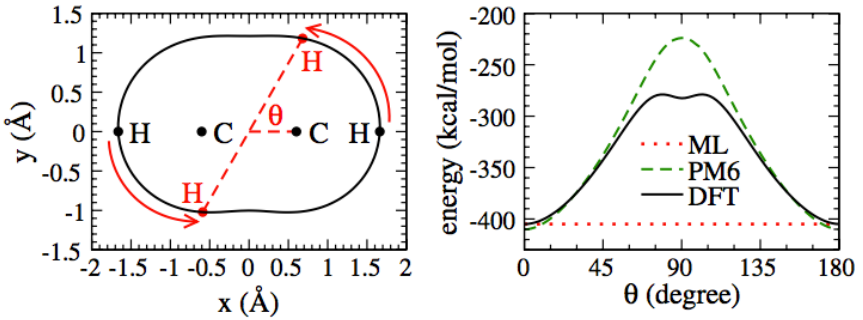


FIG. 2 (color online). A continuous deformation of acetylene. (left) Hydrogen atoms follow the closed curve with the line connecting them fixed to the origin. Carbon atoms remain near their equilibrium positions. (right) Atomization energy as a function of the H-origin-C angle.

$$M_{IJ} = \begin{cases} 0.5 Z_I^{2.4} & \forall \; I = J, \\ \dfrac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} & \forall \; I \neq J. \end{cases}$$

Coulomb-matrix

- unique
- translation
- rotation
- symmetry
- diagonalize
- fill up w zeros

???

N = 4
-> 3*N-6 = 6 degrees of freedom

Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning

Matthias Rupp,[1,2] Alexandre Tkatchenko,[3,2] Klaus-Robert Müller,[1,2] and O. Anatole von Lilienfeld[4,2,*]

[1]Machine Learning Group, Technical University of Berlin, Franklinstr. 28/29, 10587 Berlin, Germany

Comment on "Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning"

In a recent Letter [1], the authors construct a machine learning (ML) model of molecular atomization energies, which they compare to bond counting (BC) and the PM6 semiempirical method [2]. However, their ML model was trained and tested on density functional theory (DFT)

Jonathan E. Moussa*
Sandia National Laboratories
Albuquerque, New Mexico 87185, USA

distributions

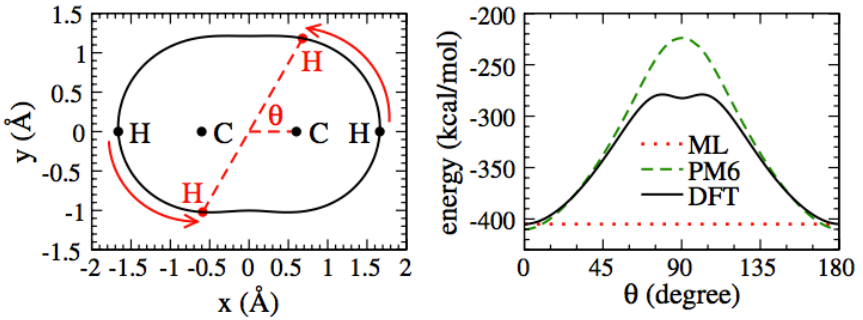


FIG. 2 (color online). A continuous deformation of acetylene. (left) Hydrogen atoms follow the closed curve with the line connecting them fixed to the origin. Carbon atoms remain near their equilibrium positions. (right) Atomization energy as a function of the H-origin-C angle.

$$M_{IJ} = \begin{cases} 0.5 Z_I^{2.4} & \forall \; I = J, \\ \dfrac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} & \forall \; I \neq J. \end{cases}$$

Coulomb-matrix

- unique
- translation
- rotation
- symmetry
- ~~diagonalize~~ sort
- fill up w zeros

N = 4
-> 3*N-6 = 6 degrees of freedom

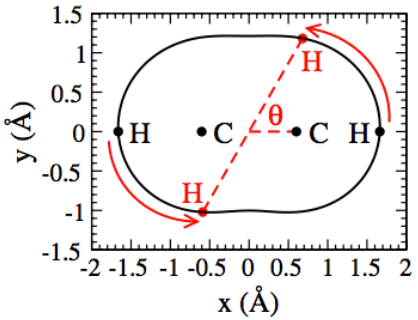
Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning

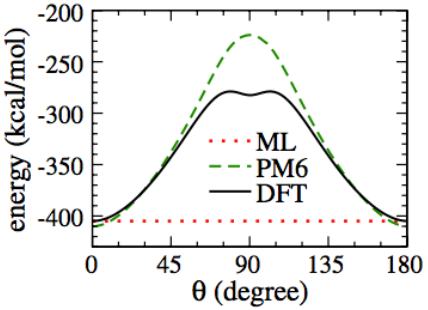
Matthias Rupp,[1,2] Alexandre Tkatchenko,[3,2] Klaus-Robert Müller,[1,2] and O. Anatole von Lilienfeld[4,2,*]

[1]Machine Learning Group, Technical University of Berlin, Franklinstr. 28/29, 10587 Berlin, Germany

Comment on "Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning"

In a recent Letter [1], the authors construct a machine learning (ML) model of molecular atomization energies, which they compare to bond counting (BC) and the PM6 semiempirical method [2]. However, their ML model was trained and tested on d





Jonathan E. Mous
Sandia National
Albuquerque, Ne

distributions

$$M_{IJ} =$$

N = 4
-> 3*N-6

Rupp *et al.* Reply: In his Comment [1], J. E. Moussa (JEM) raises concerns regarding the accuracy of our recently published Machine Learning (ML) model [2]. Our performance estimates, based on cross-validated Kernel Ridge Regression, amount to less than 10 kcal/mol mean absolute error (MAE) with respect to DFT-PBE0 [3,4] predictions of atomization energies, using a training set of more than 7000 small organic molecules from the GDB-13 data set [5]. As such, the ML model achieves an accuracy similar to generalized gradient DFT, and significantly exceeds that of Hartree-Fock or local density approximated DFT [6].

In our Letter we presented numerical evidence that ML models can be built using (i) sufficient examples and (ii) a molecular representation based on Cartesian coordinates and elemental composition *without* explicitly accounting for the electronic degrees of freedom. Therefore, performance of our

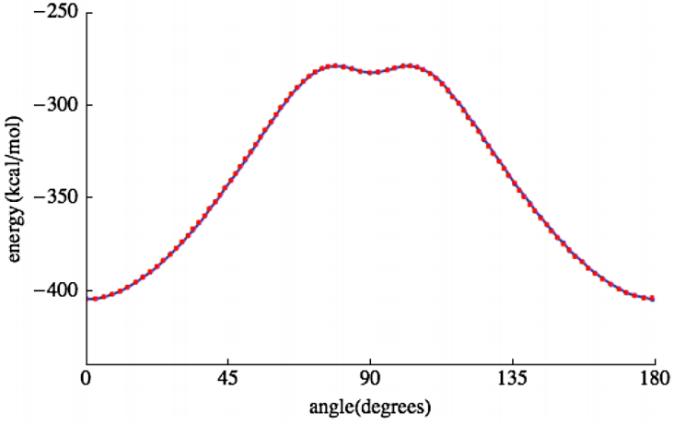


FIG. 1 (color online). Blue line: PBE0. Red dots: ML model using Frobenius norm of, and trained on, Coulomb matrices of geometries corresponding to JEM's example.

PRL **108,** 058301 (2012) PHYSICAL REVIEW LETTERS week ending 3 FEBRUARY 2012

Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning

Matthias Rupp,[1,2] Alexandre Tkatchenko,[3,2] Klaus-Robert Müller,[1,2] and O. Anatole von Lilienfeld[4,2,*]

[1]*Machine Learning Group, Technical University of Berlin, Franklinstr. 28/29, 10587 Berlin, Germany*

PRL **109,** 059801 (2012) PHYSICAL REVIEW LETTERS week ending 3 AUGUST 2012

Comment on "Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning"

In a recent Letter [1], the authors construct a machine learning (ML) model of molecular atomization energies, which they compare to bond counting (BC) and the PM6 semiempirical method [2]. However, their ML model was trained and tested on density functional theory (DFT)

Jonathan E. Moussa*
Sandia National Laboratories
Albuquerque, New Mexico 87185, USA

the ML mod
distributions

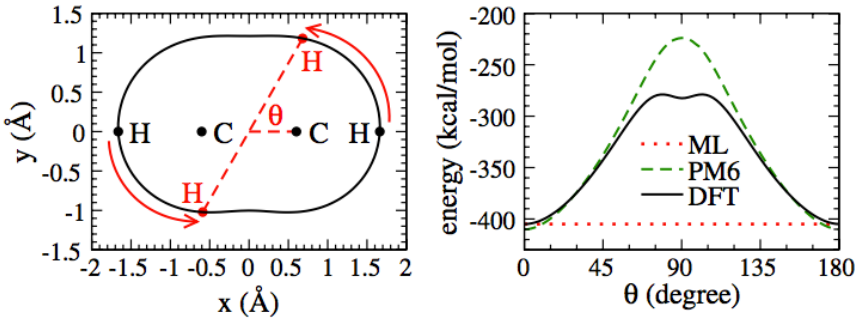


FIG. 2 (color online). A continuous deformation of acetylene. (left) Hydrogen atoms follow the closed curve with the line connecting them fixed to the origin. Carbon atoms remain near their equilibrium positions. (right) Atomization energy as a function of the H-origin-C angle.

$$M_{IJ} = \begin{cases} 0.5 Z_I^{2.4} & \forall \; I = J, \\ \dfrac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} & \forall \; I \neq J. \end{cases}$$

N = 4
-> 3*N-6 = 6 degrees of freedom

Coulomb-matrix

- unique
- translation
- rotation
- symmetry
- ~~diagonalize~~ ~~sort~~ mutants
- fill up w zeros

http://www.quantum-machine.org/

$$\{Z_I, \mathbf{R}_I\} \overset{H\Psi}{\longmapsto} E$$

"Quantum Machine"

E

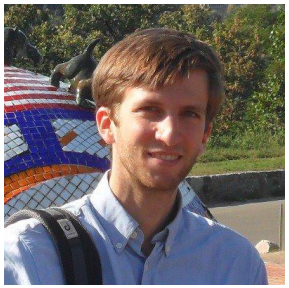
α

$\ldots$

Tkatchenko (FHI)

Gobre (FHI)

Hansen (FHI)



Montavon (TU Berlin)

Rupp (ETHZ)

Müller (TU Berlin)

Vazquez (ANL)

M. Rupp, A. Tkatchenko, K.-R. Müller, OAvL, *Phys Rev Lett* (2012);
G. Montavon, M. Rupp, V. Gobre, A. Vazquez, K. Hansen, A. Tkatchenko, K.-R. Müller, OAvL, *NJP* accepted (2013); Montavon et al *NIPS proceedings* (2013)

GDB: All organic molecules up to 13 atoms

Calculate 14 properties:

a. atomization energy (PBE0)

b. 2 x polarizability (PBE0/SCS)

c. 6 x HOMO/LUMO (GW/PBE0/ZINDO)

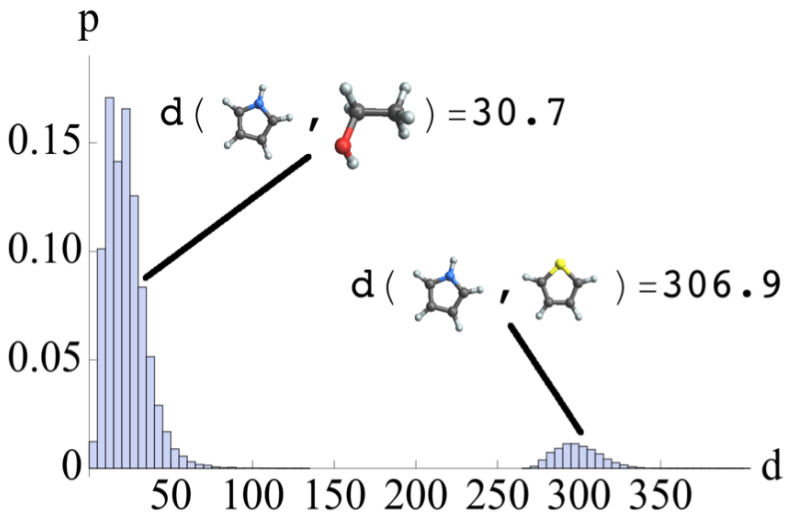
d. 2 x IP/EA (ZINDO)

e. 3 x Excitations (ZINDO)

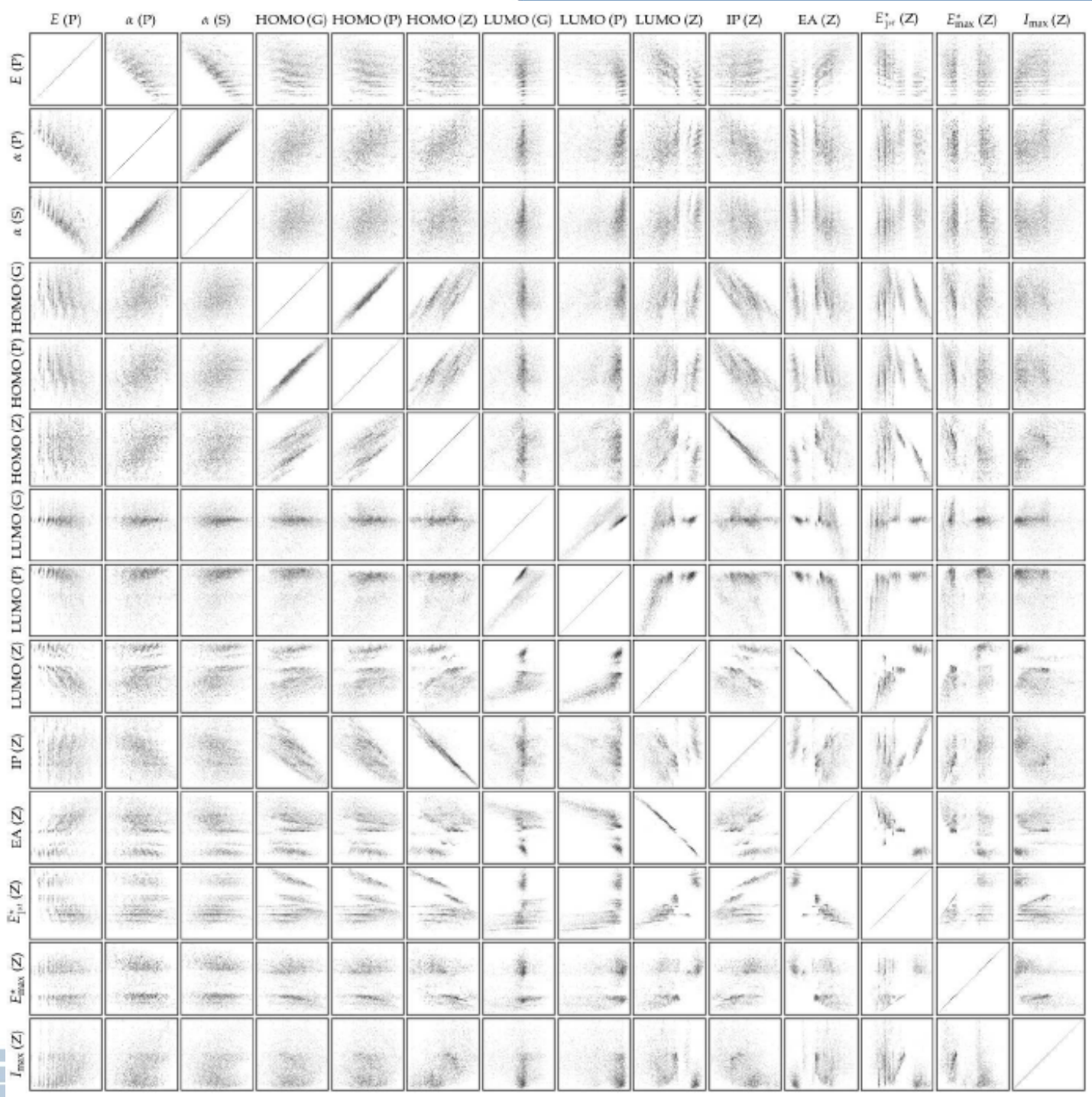
1. 7k compositional & constitutional isomers

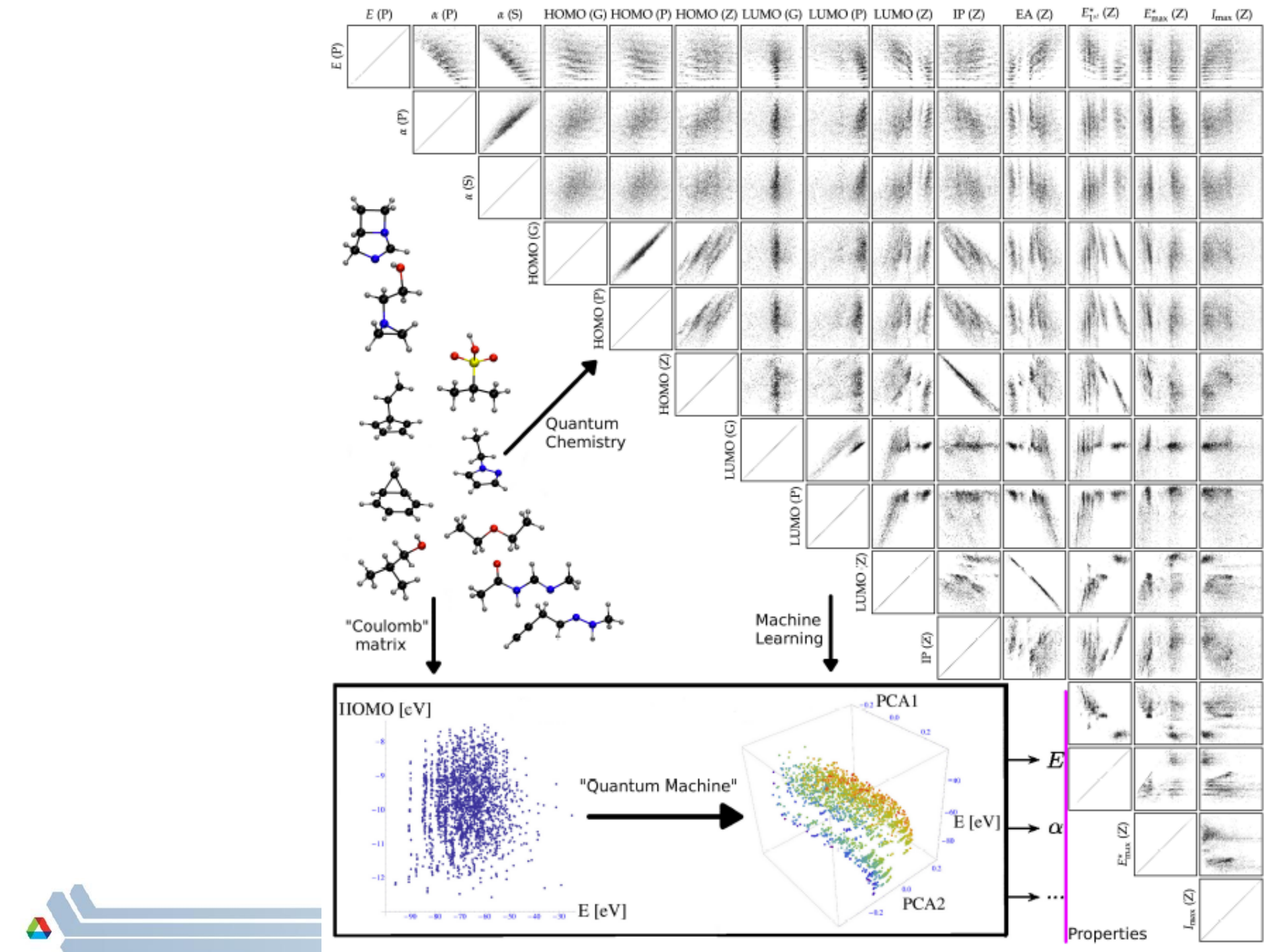
2. Initial coordinates from universal force field [Goddard et al JACS (1992)]

3. Relaxed geometry with DFT

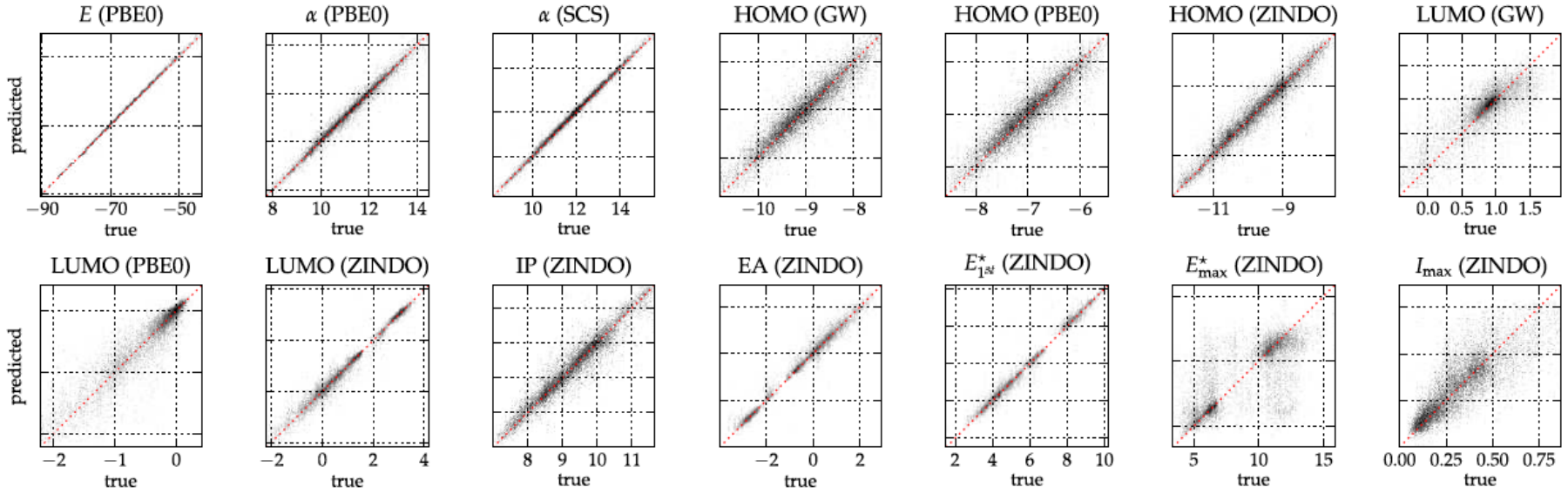


The data

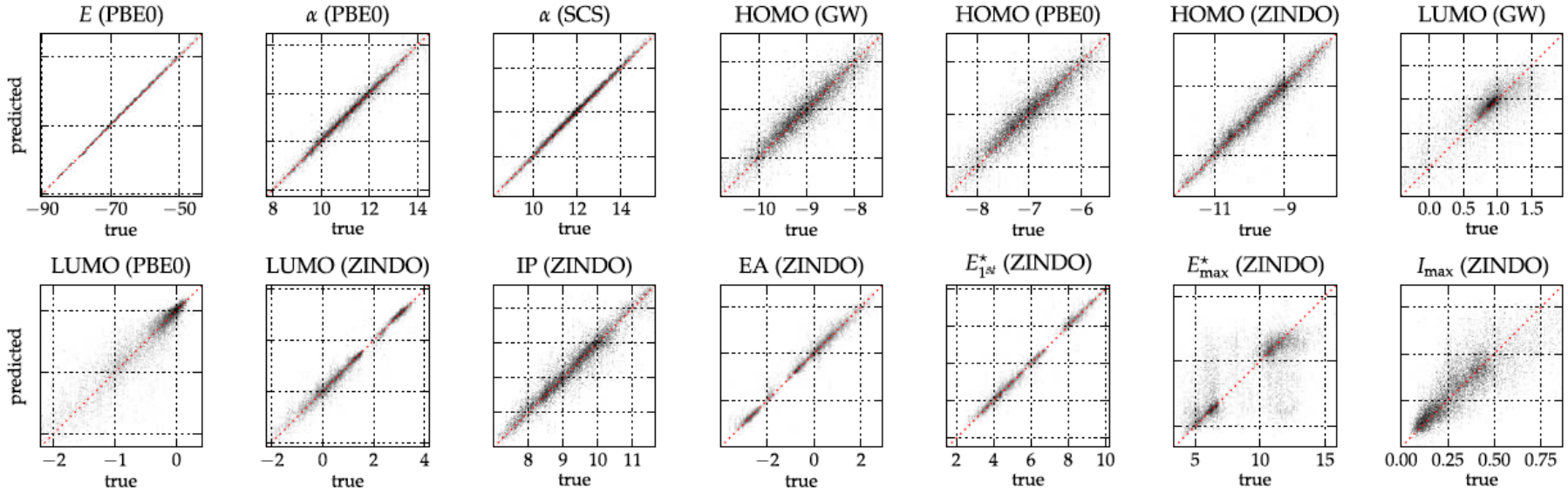

E (P)
α (P)
α (S)
HOMO (G)
HOMO (P)
HOMO (Z)
LUMO (G)
LUMO (P)
LUMO (Z)
IP (Z)
EA (Z)
E^*_{1st} (Z)
E^*_{max} (Z)
I_{max} (Z)

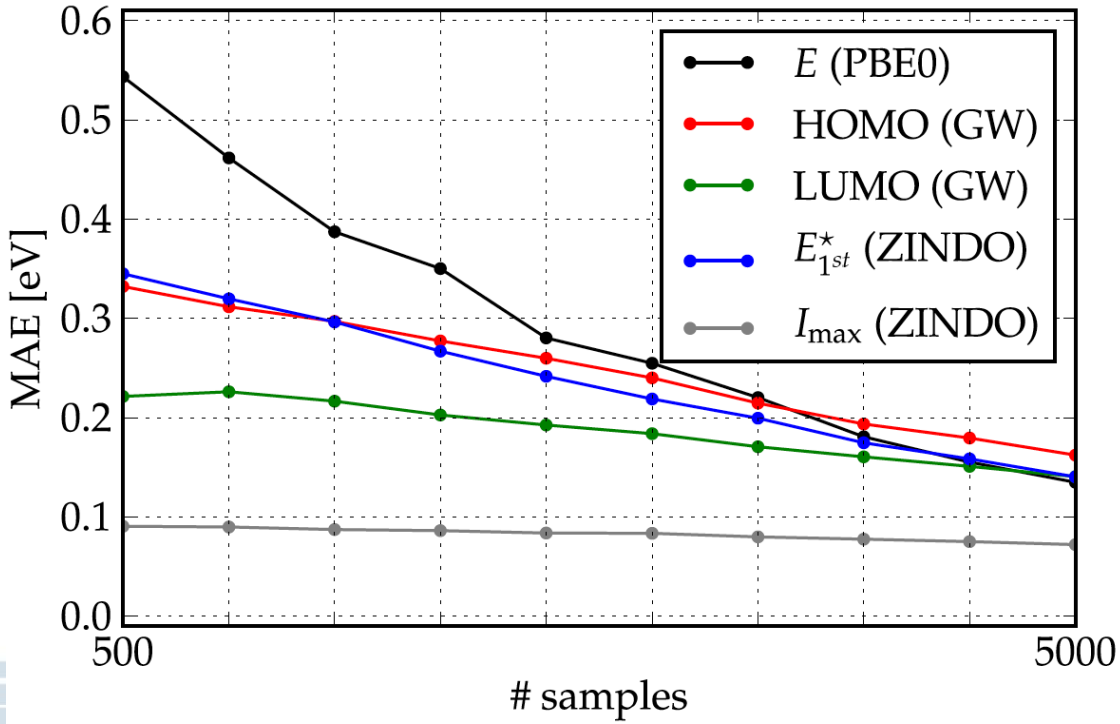

E (P)
α (P)
α (S)
HOMO (G)
HOMO (P)
HOMO (Z)
LUMO (G)
LUMO (P)
LUMO (Z)
IP (Z)
EA (Z)
$E^*_{1,0}$ (Z)
E^*_{max} (Z)
I_{max} (Z)
Quantum Chemistry
"Coulomb" matrix
Machine Learning
HOMO [eV]
E [eV]
"Quantum Machine"
PCA1
PCA2
E [eV]
E
α
...
Properties

http://www.quantum-machine.org/



http://www.quantum-machine.org/





Property [eV, A^3]	Mean	MAE		Reference MAE
E (PBE0)	-67.79	0.16		0.15[a], 0.23[b], 0.09 - 0.22[c]
α (PBE0)	11.11	0.11		0.05-0.27[d], 0.04-0.14[e]
α (SCS)	11.87	0.07		0.05-0.27[f], 0.04-0.14[g]
HOMO (GW)	-9.09	0.16		-
HOMO (PBE0)	-7.01	0.15		2.08[h]
HOMO (ZINDO)	-9.81	0.16		0.79[h]
LUMO (GW)	0.78	0.14		-
LUMO (PBE0)	-0.52	0.12		1.30[h]
LUMO (ZINDO)	1.05	0.11		0.93[h]
IP (ZINDO)	9.27	0.18		0.20[i], 0.15[j]
EA (ZINDO)	0.55	0.12		0.16[k], 0.11[l]
$E^*_{1^{st}}$ (ZINDO)	5.58	0.13		0.18[m], 0.21[n]
$E^*_{\max}$ (ZINDO)	8.82	1.07		-
$I_{\max}$ (ZINDO)	0.33	0.07		-

G. Montavon, M. Rupp, V. Gobre, A. Vazquez, K. Hansen, A. Tkatchenko, K.-R. Müller, OAvL, *NJP* accepted (2013)

http://www.quantum-machine.org/

$\{Z_I, \mathbf{R}_I\} \overset{H\Psi}{\longmapsto} E$

"Quantum Machine"

E

α

...

$\{Z_I, \mathbf{R}_I\} \overset{\mathrm{ML}}{\longmapsto} E$

Kernel Ridge Regression

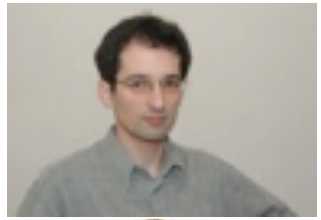
Artificial Neural Networks

Sumpter, Noid: Potential energy surfaces for macromolecules. A neural network technique *Chem Phys Lett (1992)*

Lorenz, Gross, Scheffler (FHI): Representing high-dimensional potential-energy surfaces for reactions at surfaces by neural networks *Chem Phys Lett (2004)*

Manzhos, T. Carrington (Montreal): Using neural networks to represent potential surfaces as sums of products, *J Chem Phys (2006)*

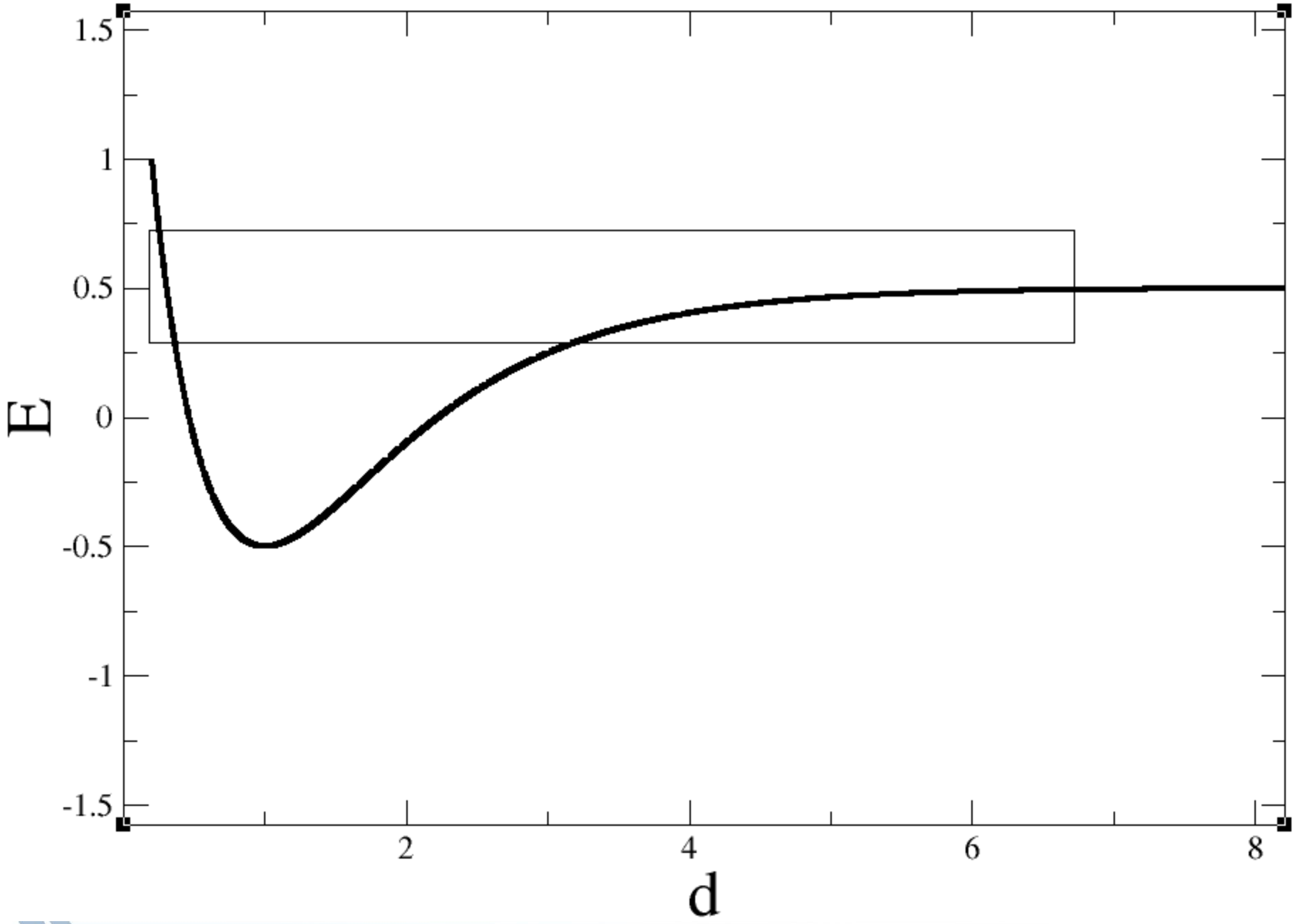
Parrinello, Behler (Bochum): Generalized neural-network representation of high-dimensional potential energy surfaces, *Phys Rev Lett (2010)*

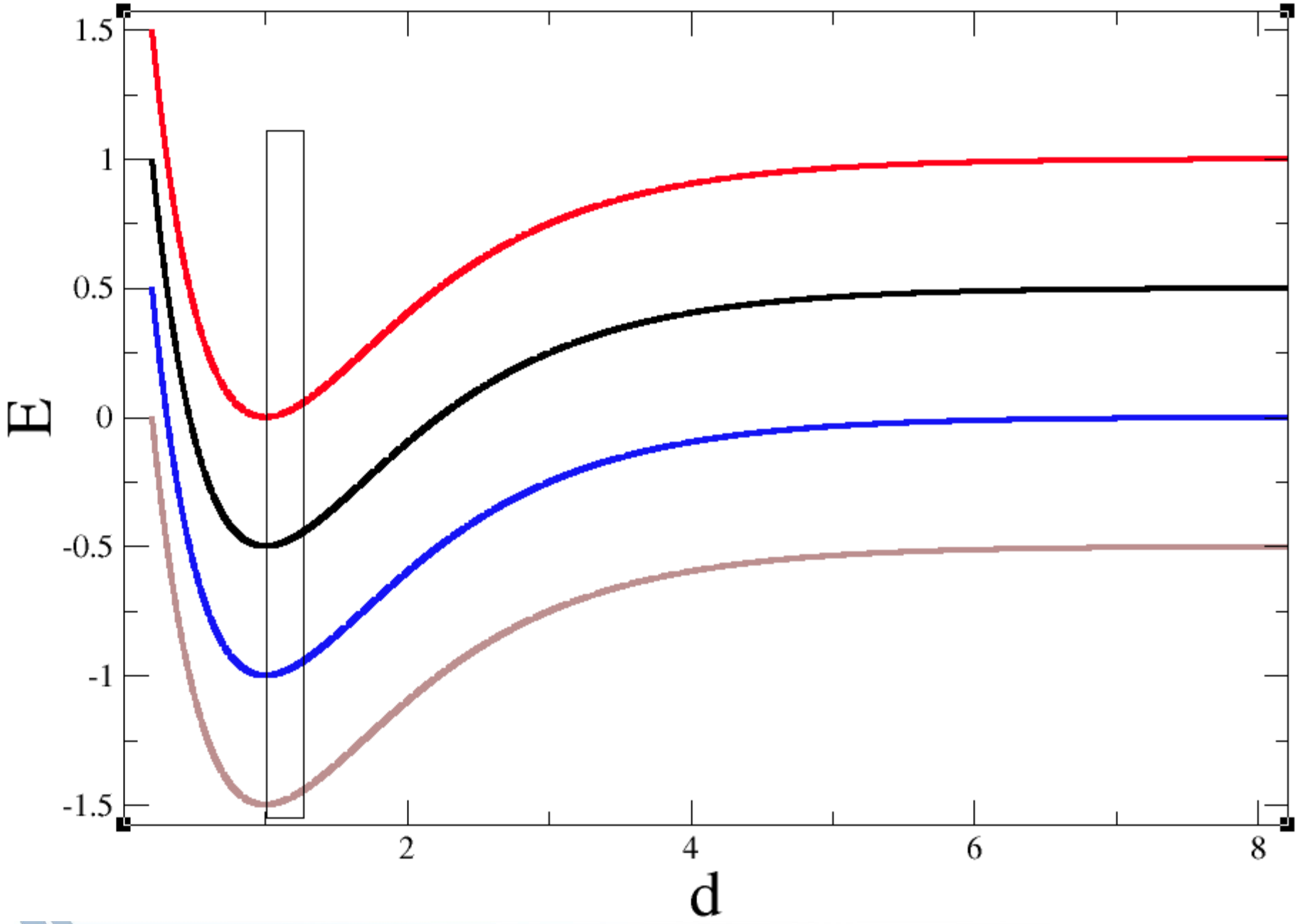


Csanyi (Cambridge): Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons, *Phys Rev Lett (2010)*

Henkelman (UT Austin): Optimizing transition states via kernel based machine learning, *J Chem Phys* (2012)

Burke (UC Irvine): Finding density functionals with machine learning, *Phys Rev Lett* (2012)


1.5
1
0.5
0
-0.5
-1
-1.5
E
2
4
6
8
d


1.5
1
0.5
0
-0.5
-1
-1.5
E
2
4
6
8
d

First Principles

$$H(\{Z_I, \mathbf{R}_I\})\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

Schrödinger

correlational (inductive)

Vapnik

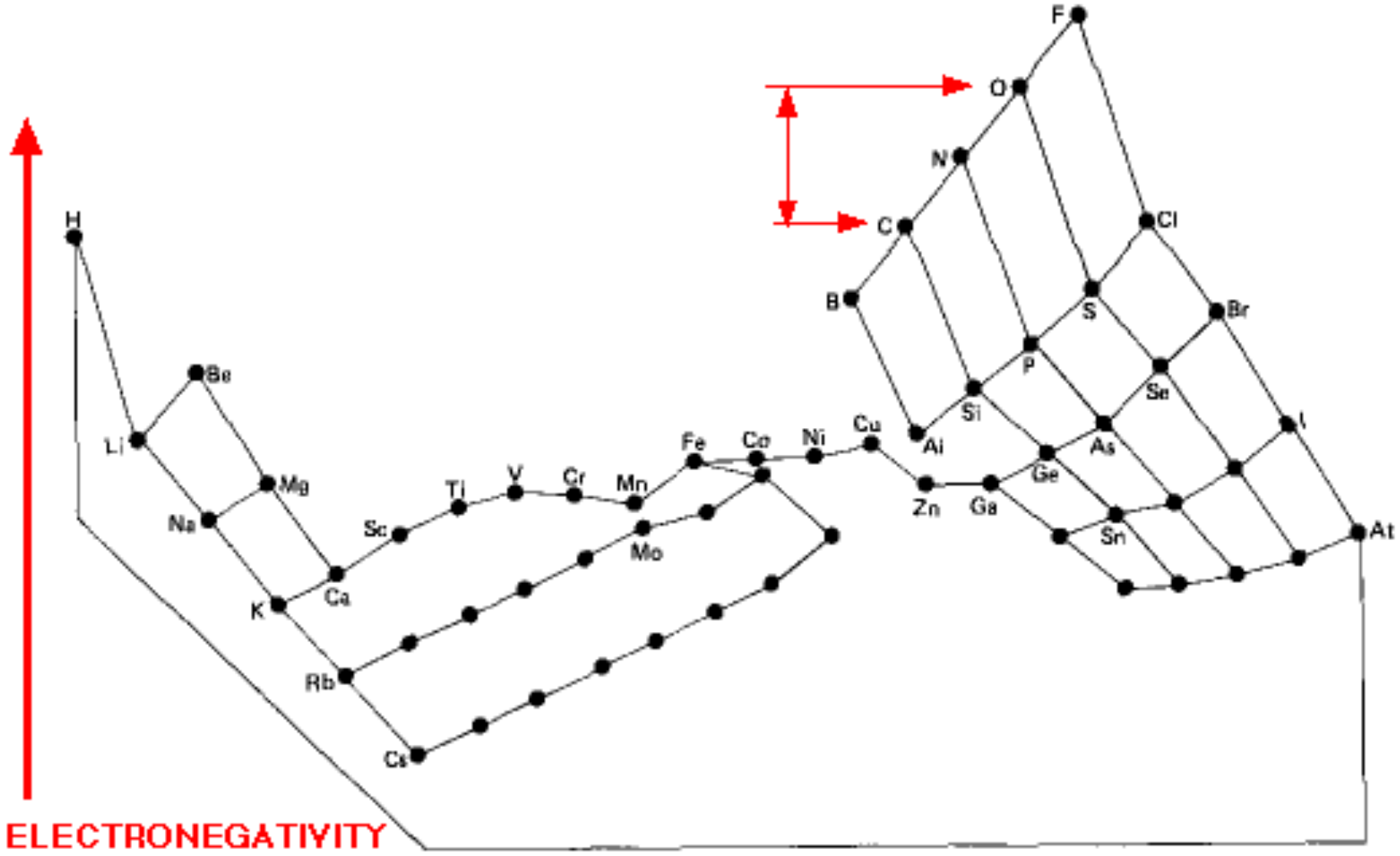
$$\{Z_I, \mathbf{R}_I\} \overset{H\Psi}{\longmapsto} E$$

supervised learning

$$\{Z_I, \mathbf{R}_I\} \overset{\text{ML}}{\longmapsto} E$$

Infer solution by comparison to previous examples

- Regression method?
- Function?
- **Variables?**
- Metric?
- Data?

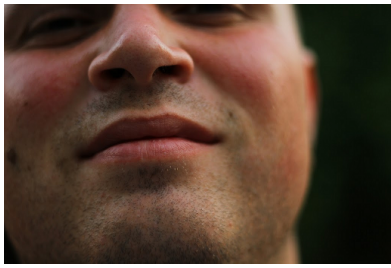

F
O
N
C
Cl
B
S
Br
P
Se
Si
I
Al
As
Ge
Ga
Sn
At
Zn
Cu
Ni
Co
Fe
Mn
Cr
V
Ti
Sc
Mo
H
Be
Li
Mg
Na
Ca
K
Rb
Cs
ELECTRONEGATIVITY

Property	Zxyz	CM	Eig(CM)
Unique	√	√	¬
First principles	√	√	√
Transl. invariant	¬	√	√
Rotat. invariant	¬	√	√
Permutat. invariant	¬	¬	√
Symmetry	¬	√	√
Size extensive	√	√	√
Complete/global	√	√	¬
Dimensionality	$4N$	$(N^2+N)/2$	N
Analytical	√	√	√
Differentiable	N.A.	√	√
Uniform length	¬	¬	¬
Variable ranges	√	√	√

Descriptor

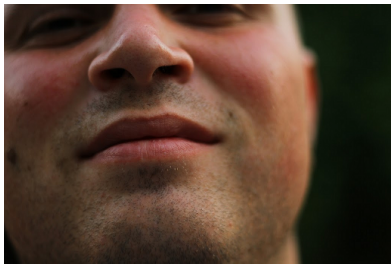
$$P(\mathbf{r}) = \sum_I Z_I e^{-a|\mathbf{r}-\mathbf{R}_I|^2}$$

$$\mathcal{F}(P) = \frac{1}{(2a)^{3/2}} e^{\frac{\omega^2}{4a}} \sum_I Z_I e^{i\omega^T \mathbf{R}_I}$$



Aaron Knoll
(TACC)

Descriptor



Aaron Knoll
(TACC)

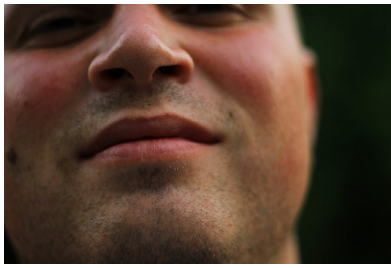
$$P(\mathbf{r}) = \sum_I Z_I e^{-a|\mathbf{r}-\mathbf{R}_I|^2}$$

$$\mathcal{F}(P) = \frac{1}{(2a)^{3/2}} e^{\frac{\omega^2}{4a}} \sum_I Z_I e^{i\omega^T \mathbf{R}_I}$$

$$\mathcal{F}\mathcal{F}^* = \frac{1}{(2a)^3} e^{\frac{\omega^2}{2a}} \sum_J \sum_I Z_I Z_J \cos[\omega^T(\mathbf{R}_I - \mathbf{R}_J)]$$

$$M_{IJ} = Z_I Z_J \cos[\omega^T(\mathbf{R}_I - \mathbf{R}_J)]$$

Descriptor



Aaron Knoll
(TACC)

$$P(\mathbf{r}) = \sum_I Z_I e^{-a|\mathbf{r}-\mathbf{R}_I|^2}$$

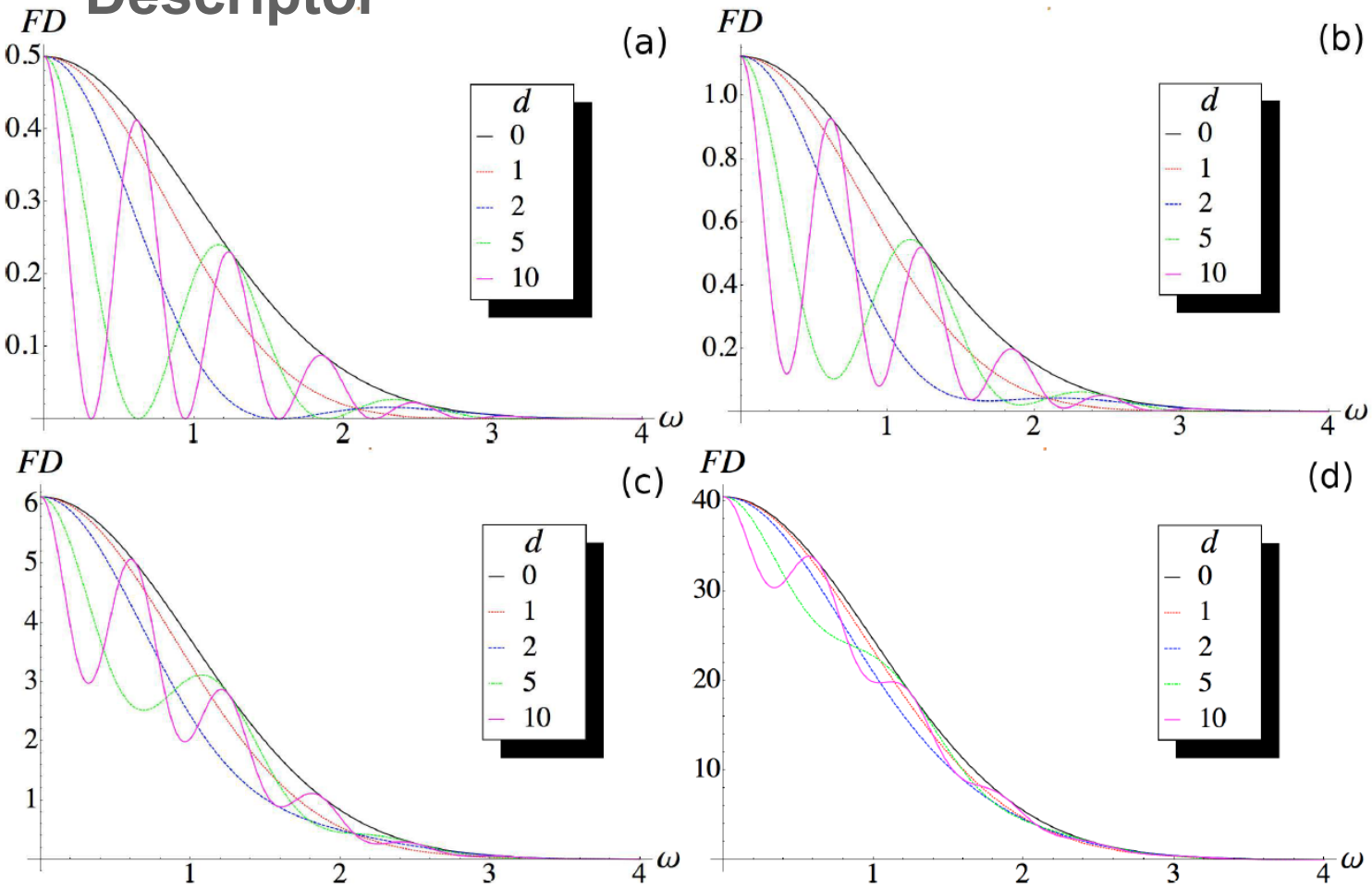
$$\mathcal{F}(P) = \frac{1}{(2a)^{3/2}} e^{\frac{\omega^2}{4a}} \sum_I Z_I e^{i\omega^T \mathbf{R}_I}$$

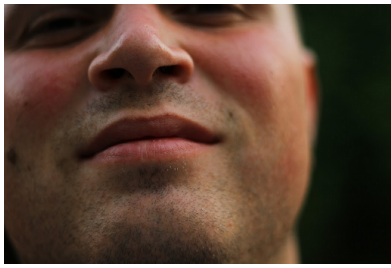
$$\mathcal{F}\mathcal{F}^* = \frac{1}{(2a)^3} e^{\frac{\omega^2}{2a}} \sum_J \sum_I Z_I Z_J \cos[\omega^T (\mathbf{R}_I - \mathbf{R}_J)]$$

$$M_{IJ} = Z_I Z_J \cos[\omega^T (\mathbf{R}_I - \mathbf{R}_J)]$$

$$FD = \frac{1}{(2a)^3} e^{\frac{\omega^2}{2a}} \sum_J \sum_I Z_I Z_J \cos[\omega \times d_{IJ}],$$

Descriptor



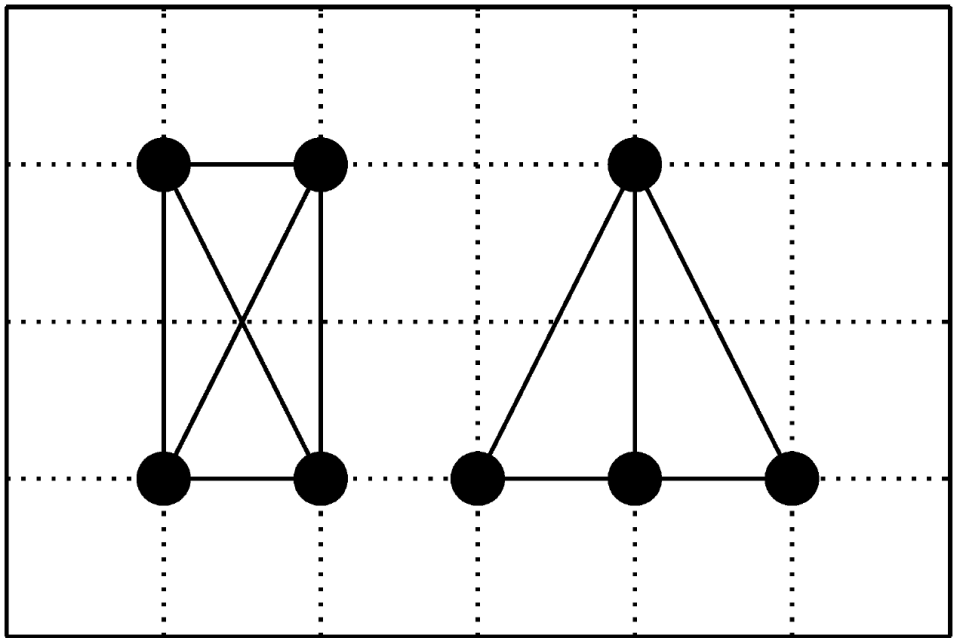


Aaron Knoll
(TACC)

Fourier descriptor FD of four diatomics, H_2 (a), HHe (b), HC (c), and HCl (d), for five interatomic distances d

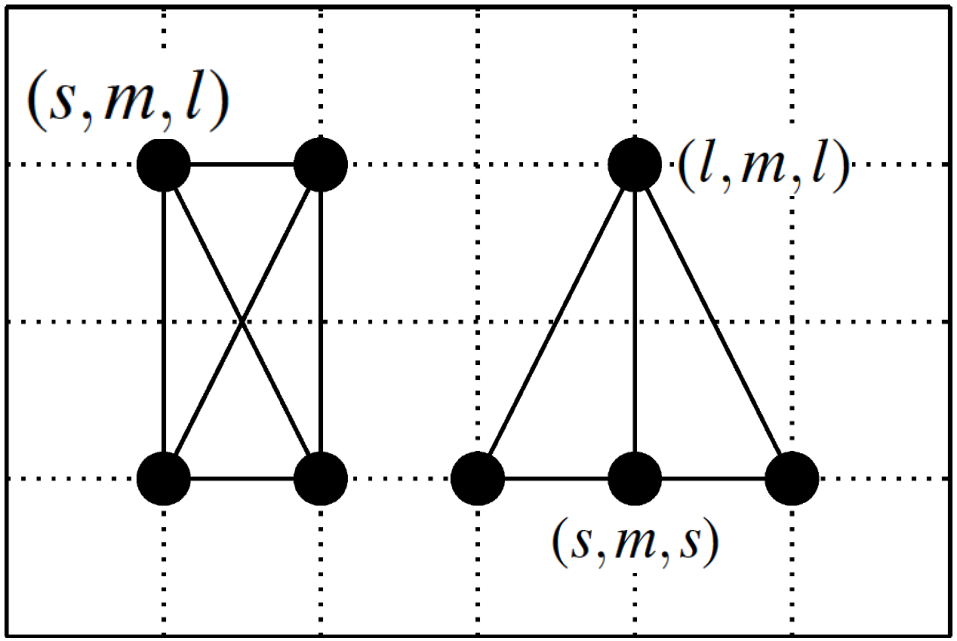
$$FD = \frac{1}{(2a)^3} e^{\frac{\omega^2}{2a}} \sum_J \sum_I Z_I Z_J \cos[\omega \times d_{IJ}],$$

$$FD = \frac{1}{(2a)^3} e^{\frac{\omega^2}{2a}} \sum_J \sum_I Z_I Z_J \cos[\omega \times d_{IJ}],$$



Homometric molecules?

$$FD = \frac{1}{(2a)^3} e^{\frac{\omega^2}{2a}} \sum_J \sum_I Z_I Z_J \cos[\omega \times d_{IJ}],$$

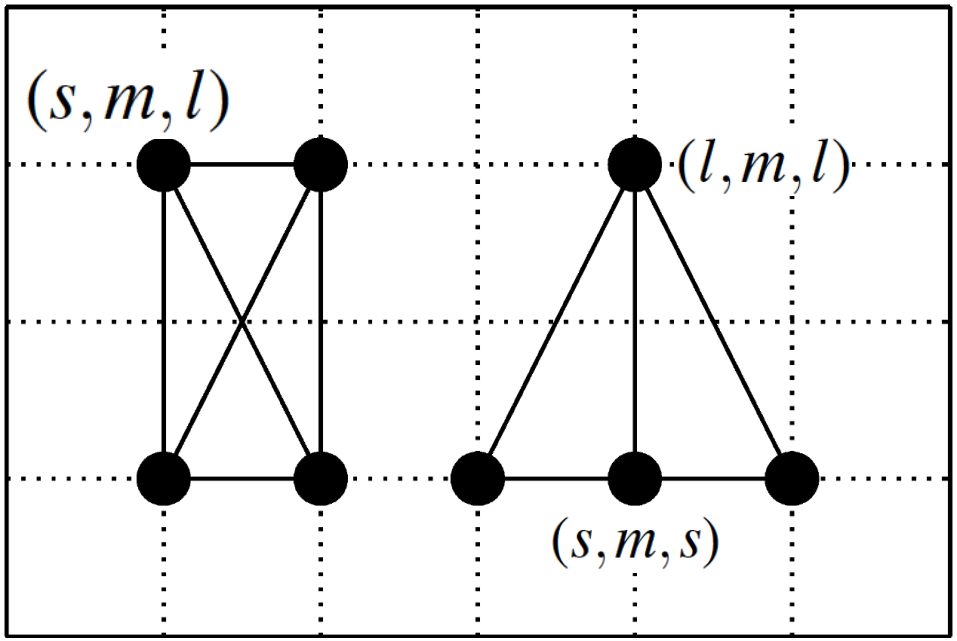


Homometric molecules?

$$\sum_J Z_J e^{-b(d-d_{IJ})^2}$$

OAvL, M. Rupp, A Knoll, *submitted* (2013), arxiv.org

$$FD = \frac{1}{(2a)^3} e^{\frac{\omega^2}{2a}} \sum_J \sum_I Z_I Z_J \cos[\omega \times d_{IJ}],$$

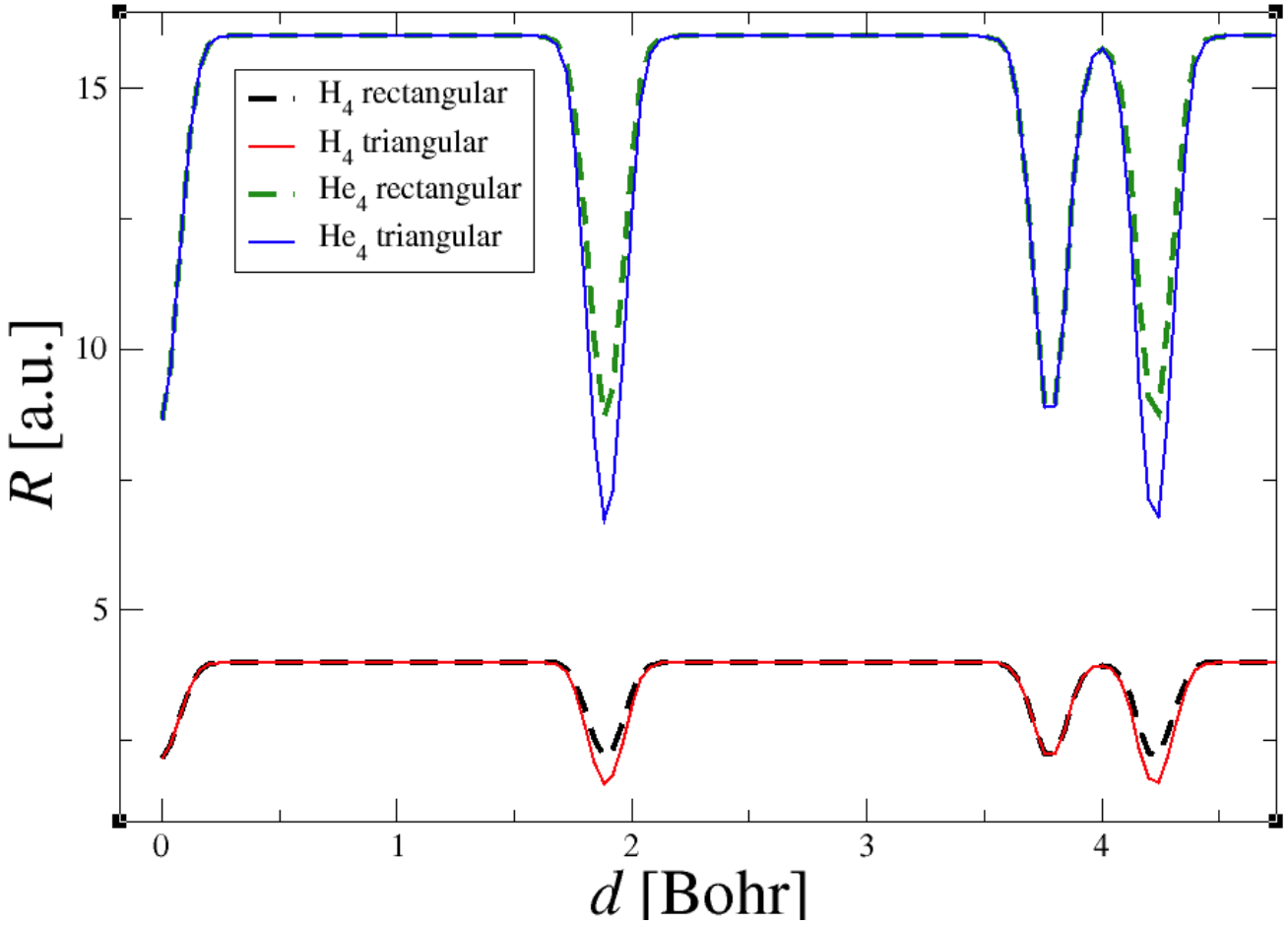


Homometric molecules?

$$\sum_J Z_J e^{-b(d-d_{IJ})^2}$$

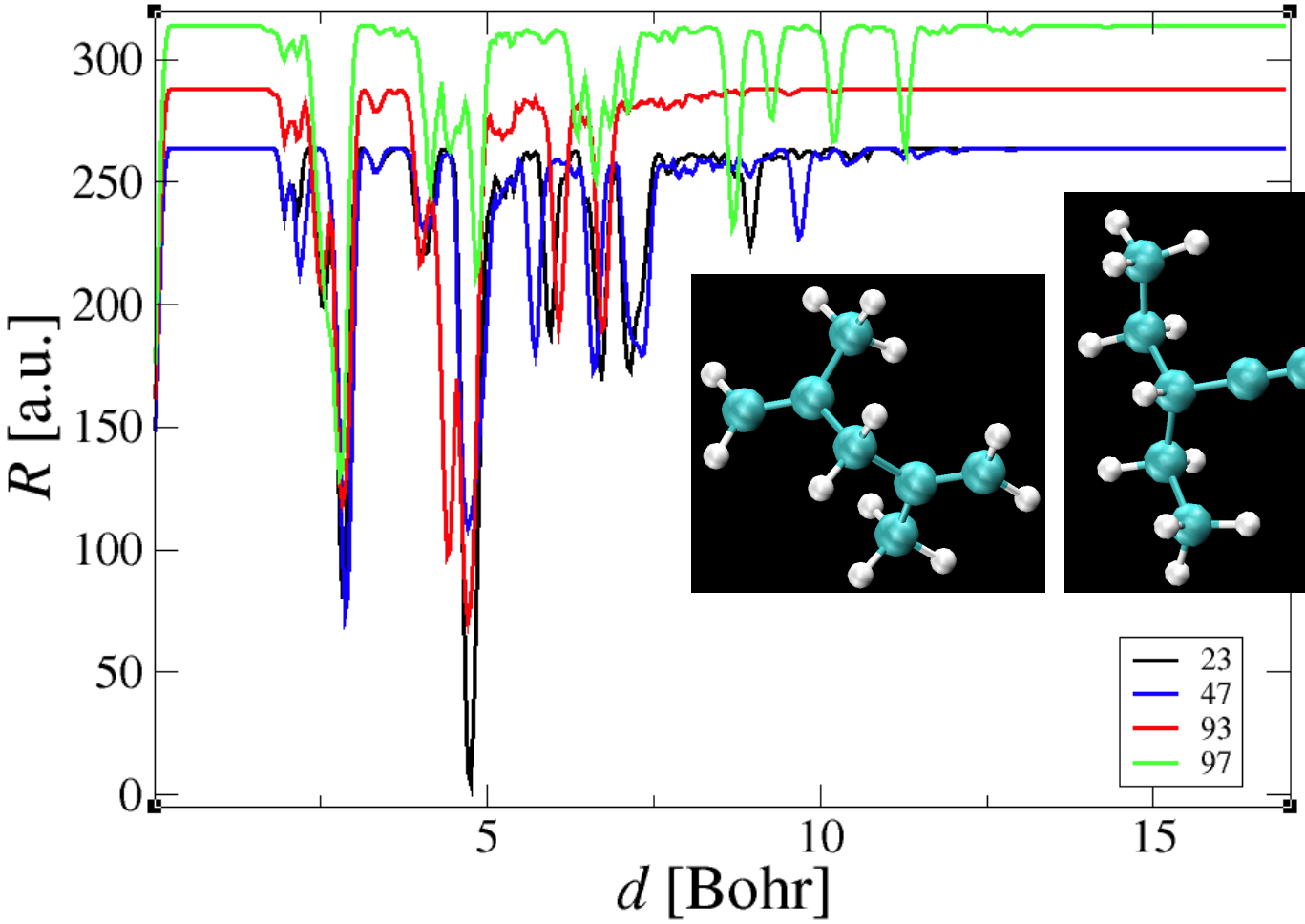
$$R(d) = \sum_I Z_I^2 \cos\left[\frac{1}{Z_I} \sum_J Z_J e^{-b(d-d_{IJ})^2}\right]$$

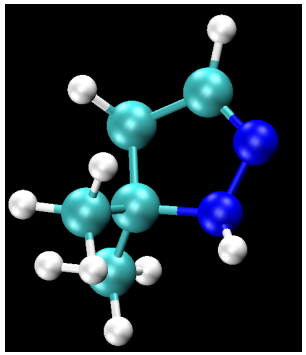
OAvL, M. Rupp, A Knoll, *submitted* (2013), arxiv.org

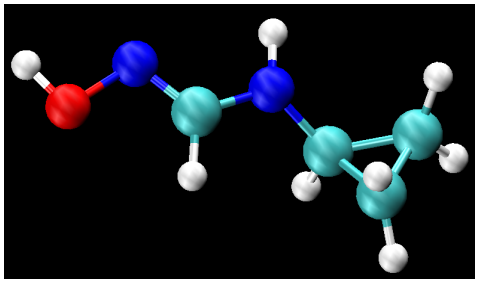


$$R(d) = \sum_I Z_I^2 \cos\left[\frac{1}{Z_I}\sum_J Z_J e^{-b(d-d_{IJ})^2}\right]$$

OAvL, M. Rupp, A Knoll, *submitted* (2013), arxiv.org



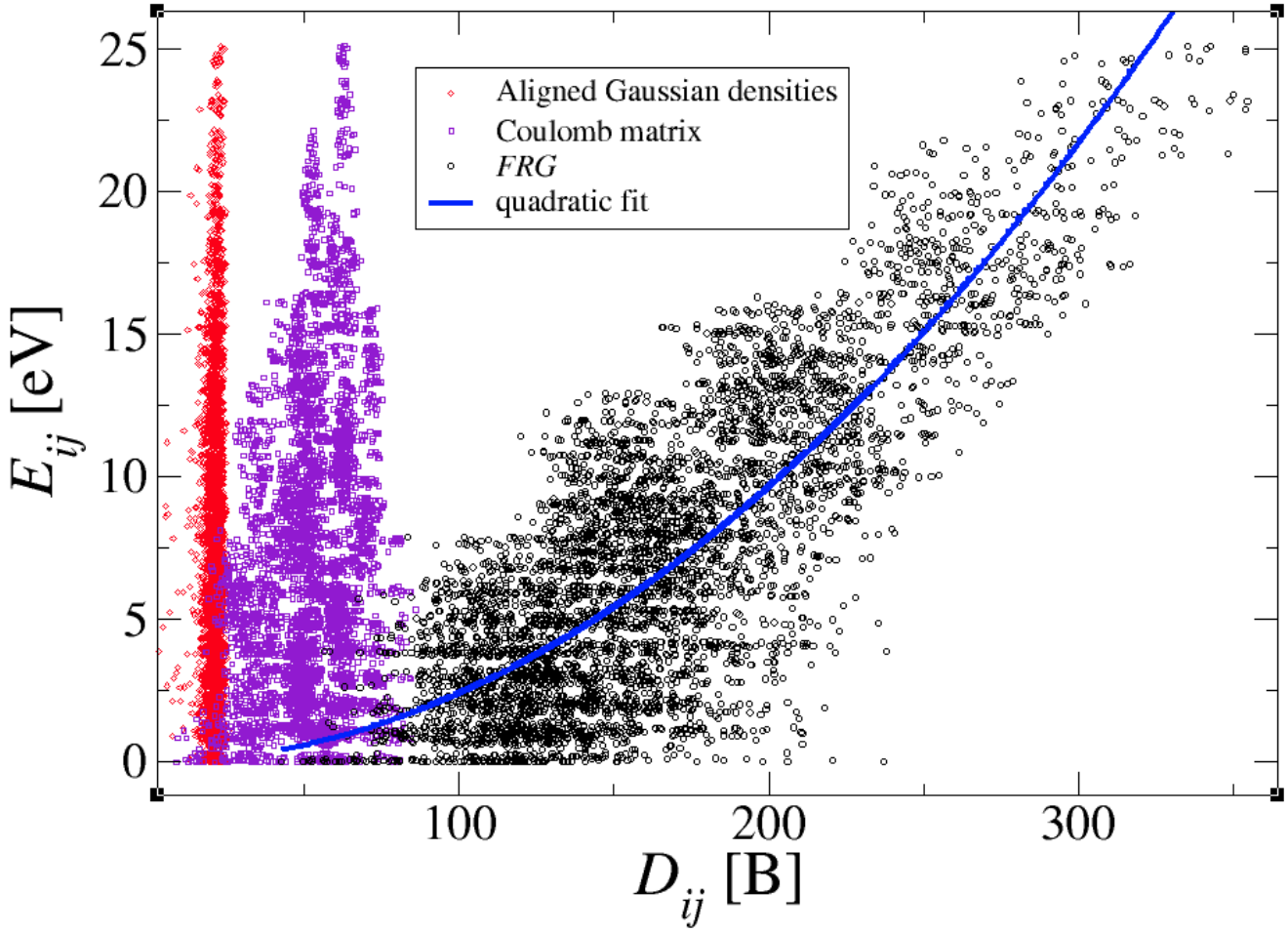




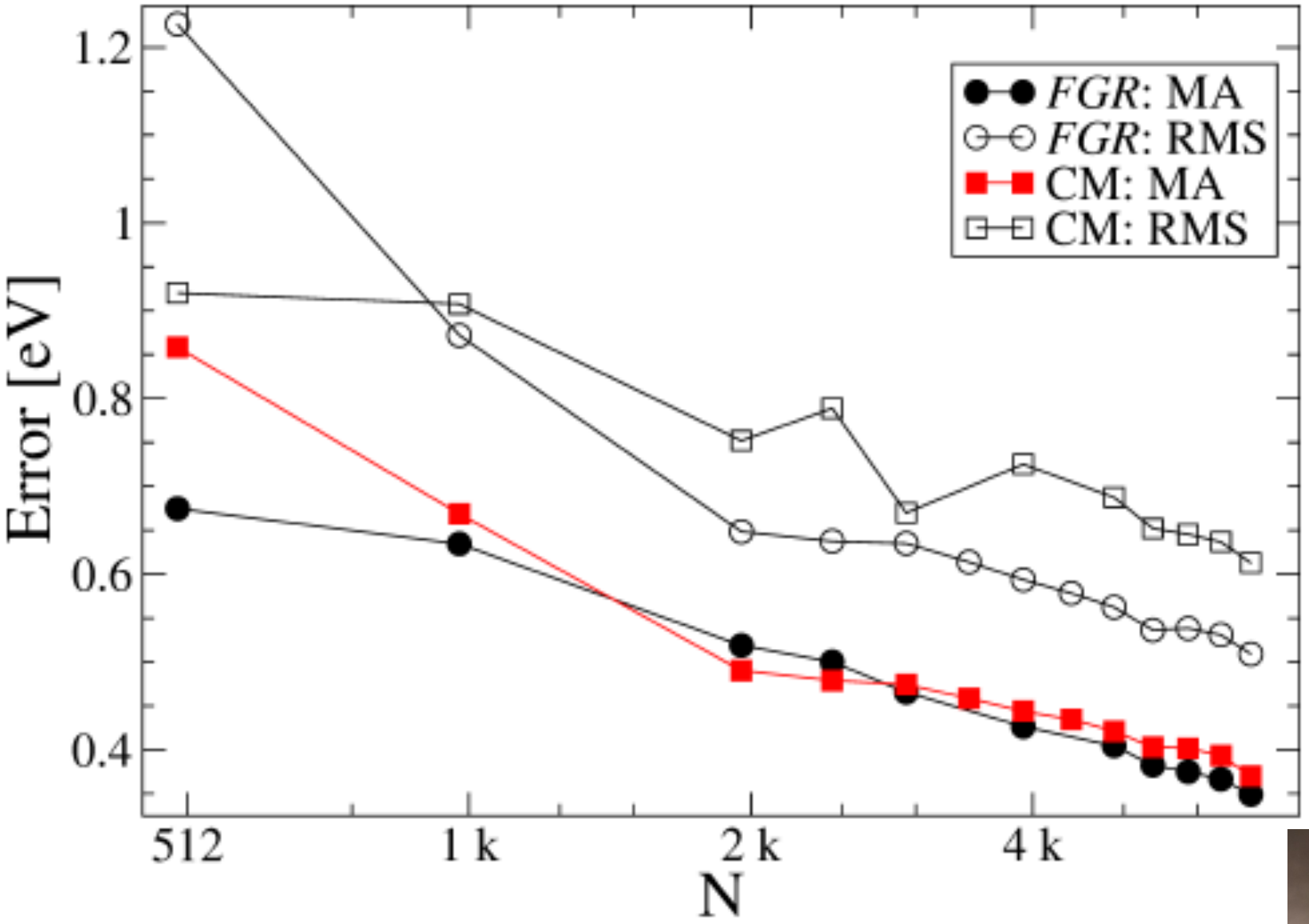
$$R(d) = \sum_I Z_I^2 \cos\left[\frac{1}{Z_I}\sum_J Z_J e^{-b(d-d_{IJ})^2}\right]$$

OAvL, M. Rupp, A Knoll, *submitted* (2013), arxiv.org

$$D(M_i, M_j) = \sqrt{\int_{d=0}^{d \geq d_{IJ}^{max}} dd \, (R_i(d) - R_j(d))^2}$$



OAvL, M. Rupp, A Knoll, *submitted* (2013), arxiv.org



OAvL, M. Rupp, A Knoll, *submitted* (2013), arxiv.org

Rupp (ETHZ)

First Principles

$$H(\{Z_I, \mathbf{R}_I\})\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

Schrödinger

correlational (inductive)

Vapnik

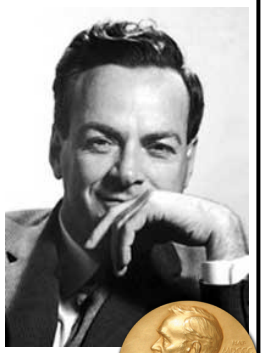
$$\{Z_I, \mathbf{R}_I\} \overset{H\Psi}{\longmapsto} E$$

supervised learning

$$\{Z_I, \mathbf{R}_I\} \overset{\mathrm{ML}}{\longmapsto} E$$

variational (deductive)

Feynman



$$\frac{\partial E[H]}{\partial R_{Ix}} = \left\langle \Psi \left| \frac{\partial H}{\partial R_{Ix}} \right| \Psi \right\rangle$$

$$\frac{\partial E[H]}{\partial Z_I} = \left\langle \Psi \left| \frac{\partial H}{\partial Z_I} \right| \Psi \right\rangle$$

$$E(H(\lambda)) = E(H_i + \lambda(H_f - H_i))$$

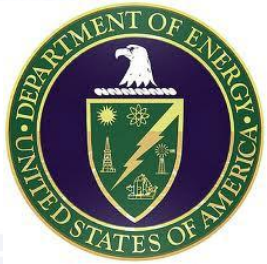
$$\frac{\partial E[H]}{\partial \lambda} = \left\langle \Psi \left| \frac{\partial H(\lambda)}{\partial \lambda} \right| \Psi \right\rangle$$

Thanks for your attention!

First principles view on chemical compound space: Gaining rigorous atomistic control of molecular properties
OAvL, Int J Quant Chem (2013), http://onlinelibrary.wiley.com/doi/10.1002/qua.24375/abstract

http://www.quantum-machine.org/

FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION



First Principles

$$H(\{Z_I, \mathbf{R}_I\})\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

Schrödinger

correlational (inductive)

Vapnik

$$\{Z_I, \mathbf{R}_I\} \overset{H\Psi}{\longmapsto} E$$

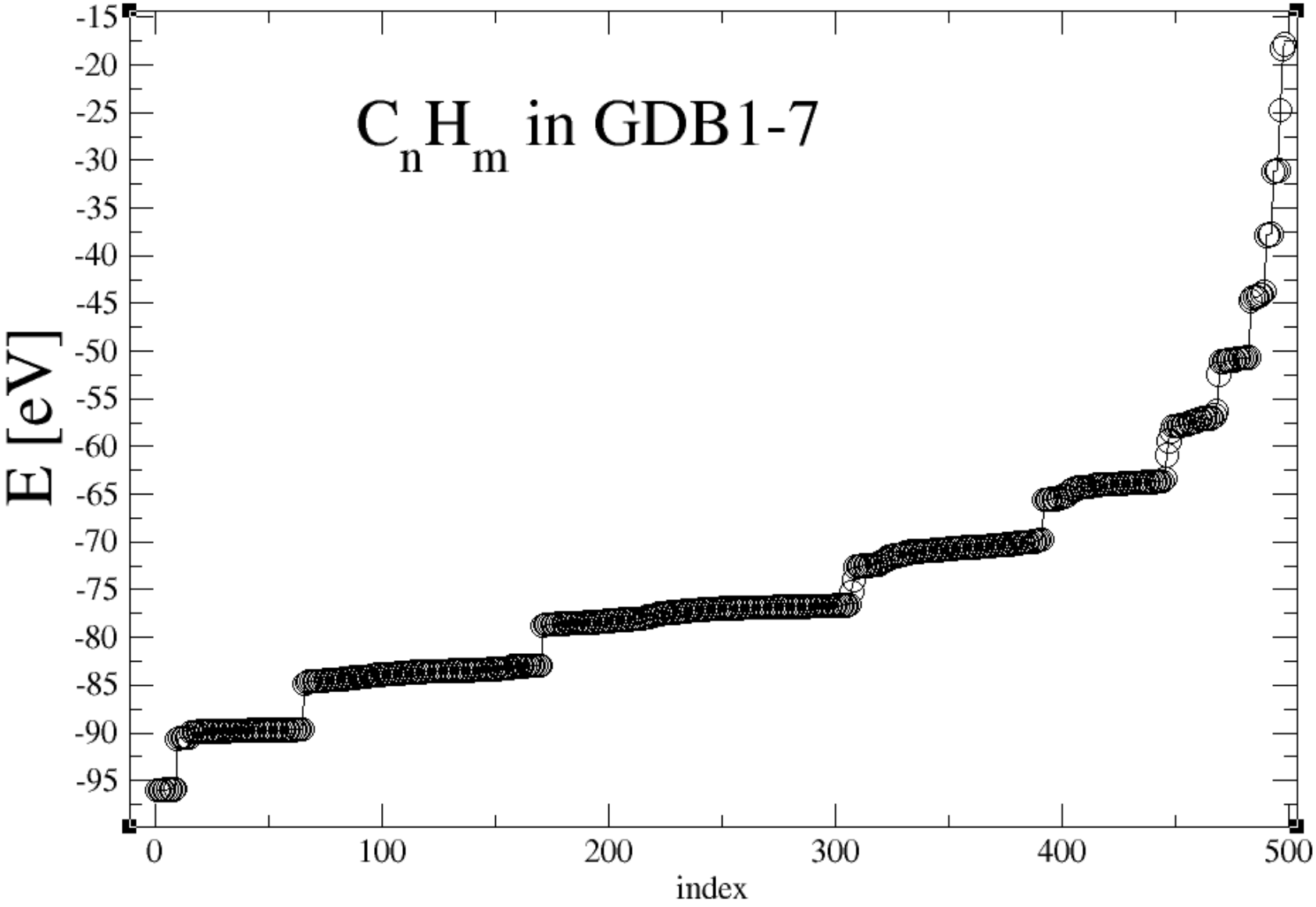
supervised learning

$$\{Z_I, \mathbf{R}_I\} \overset{\text{ML}}{\longmapsto} E$$

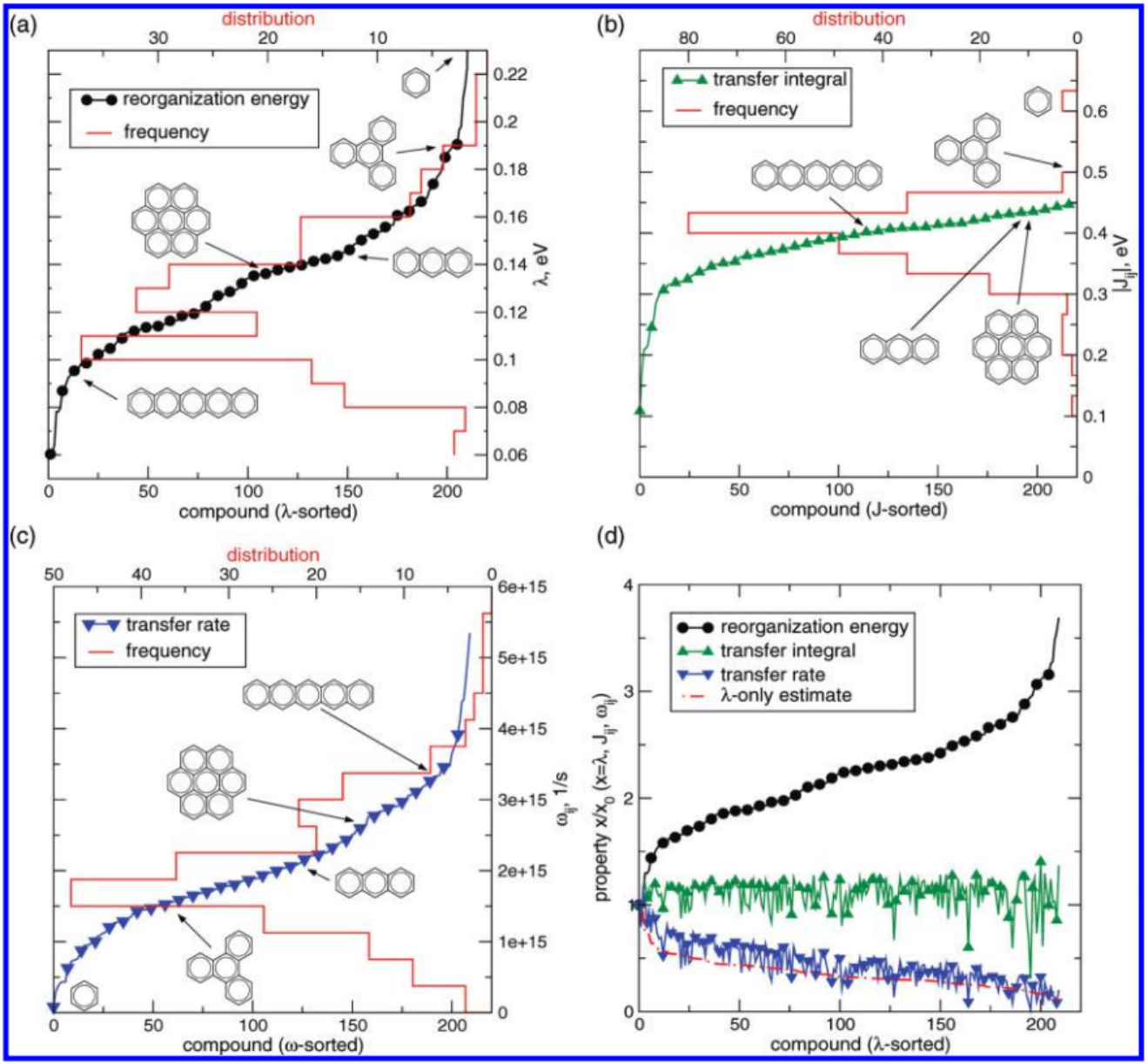
Infer solution by comparison to previous examples

- Regression method?
- Function?
- Variables?
- Metric?
- **Data?**

Data stratification



Data stratification



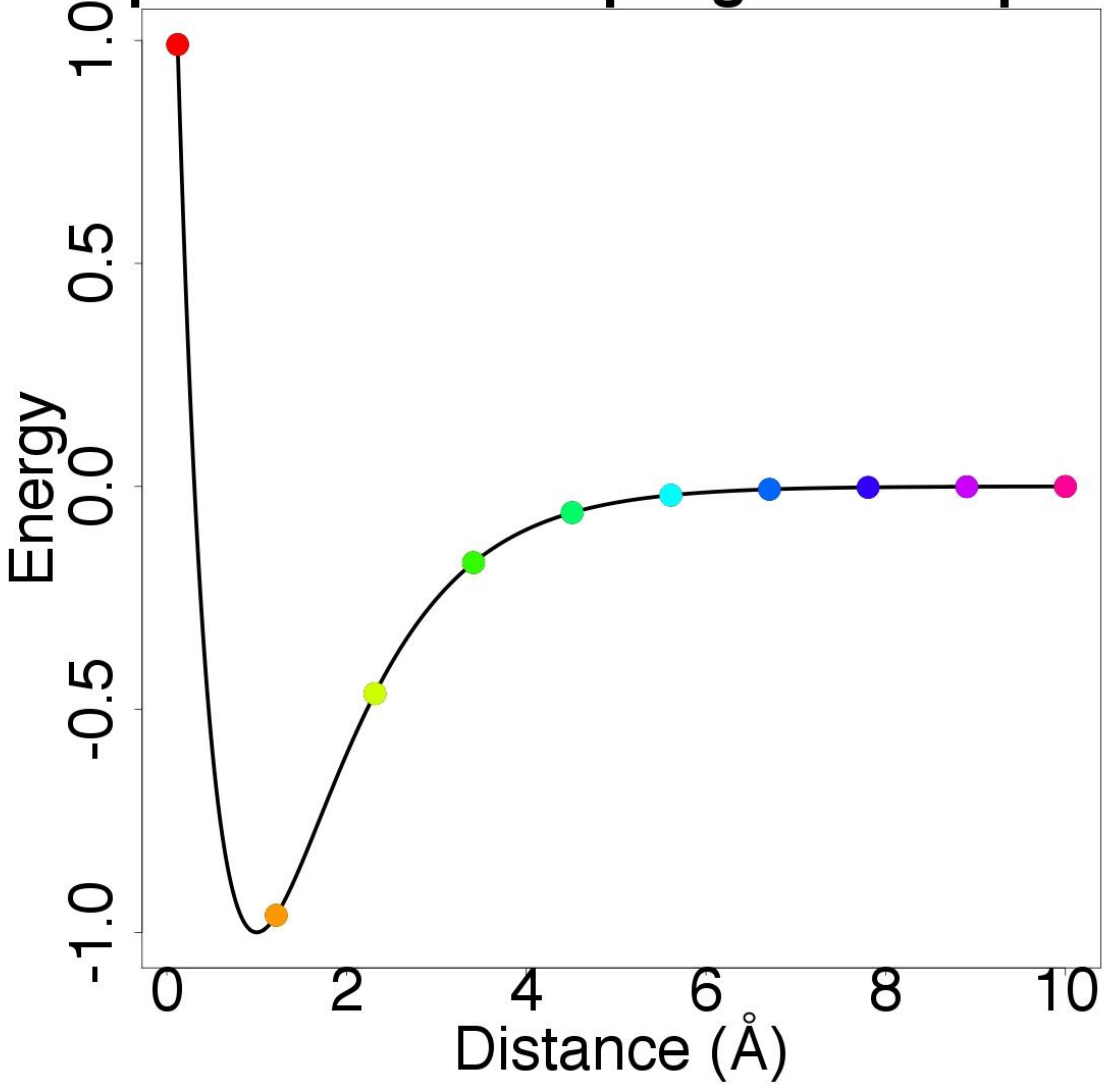
Misra et al JCTC (2011)

Outlook: Selection bias

Balaprakash (ANL)

Vazquez (ANL)

Equal interval sampling after 10 points



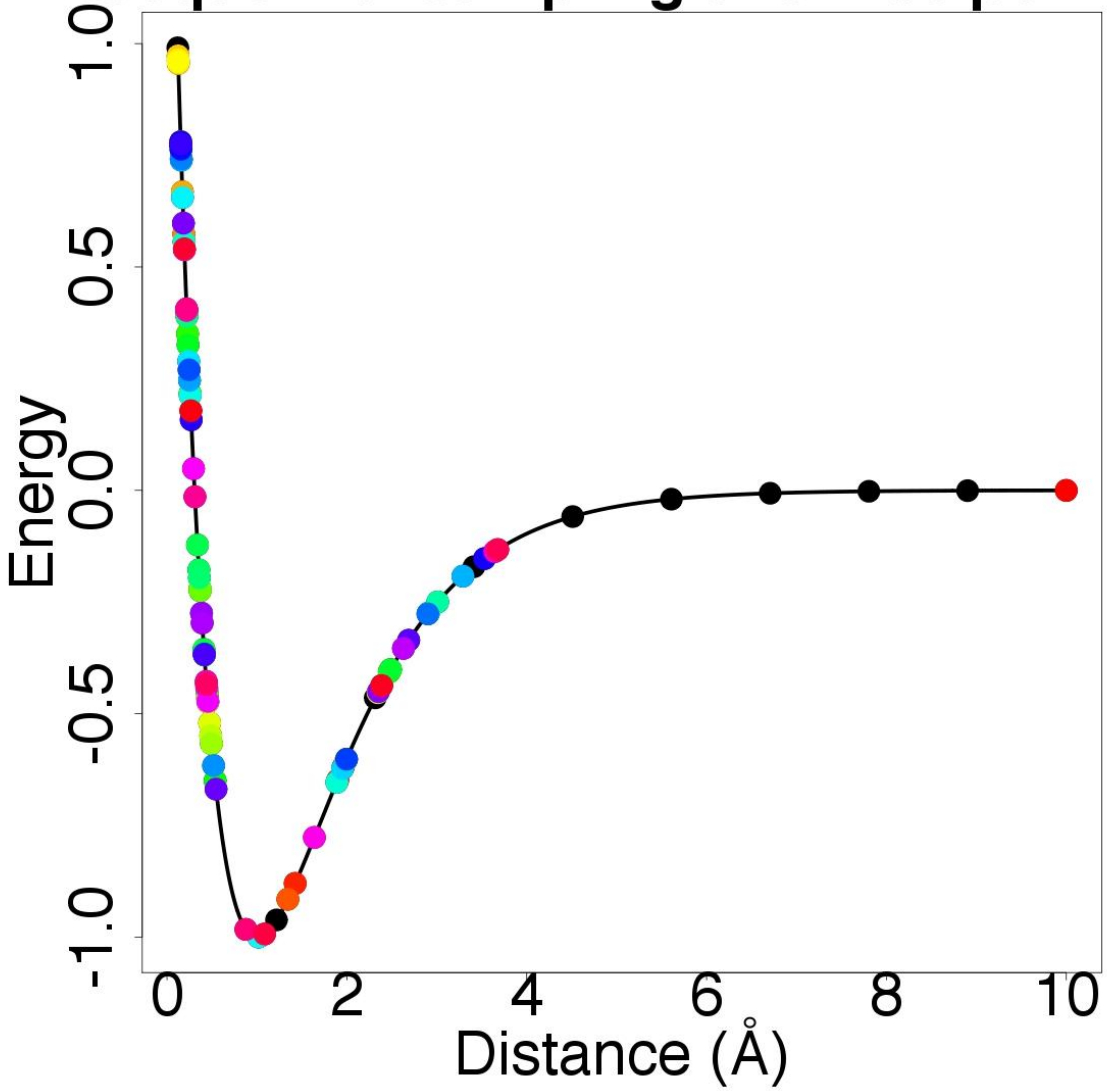
P. Balaprakash and A. Vazquez and OAvL, *in preparation* (2013)

Outlook: Selection bias

Balaprakash (ANL)

Vazquez (ANL)

Sequential sampling after 100 points



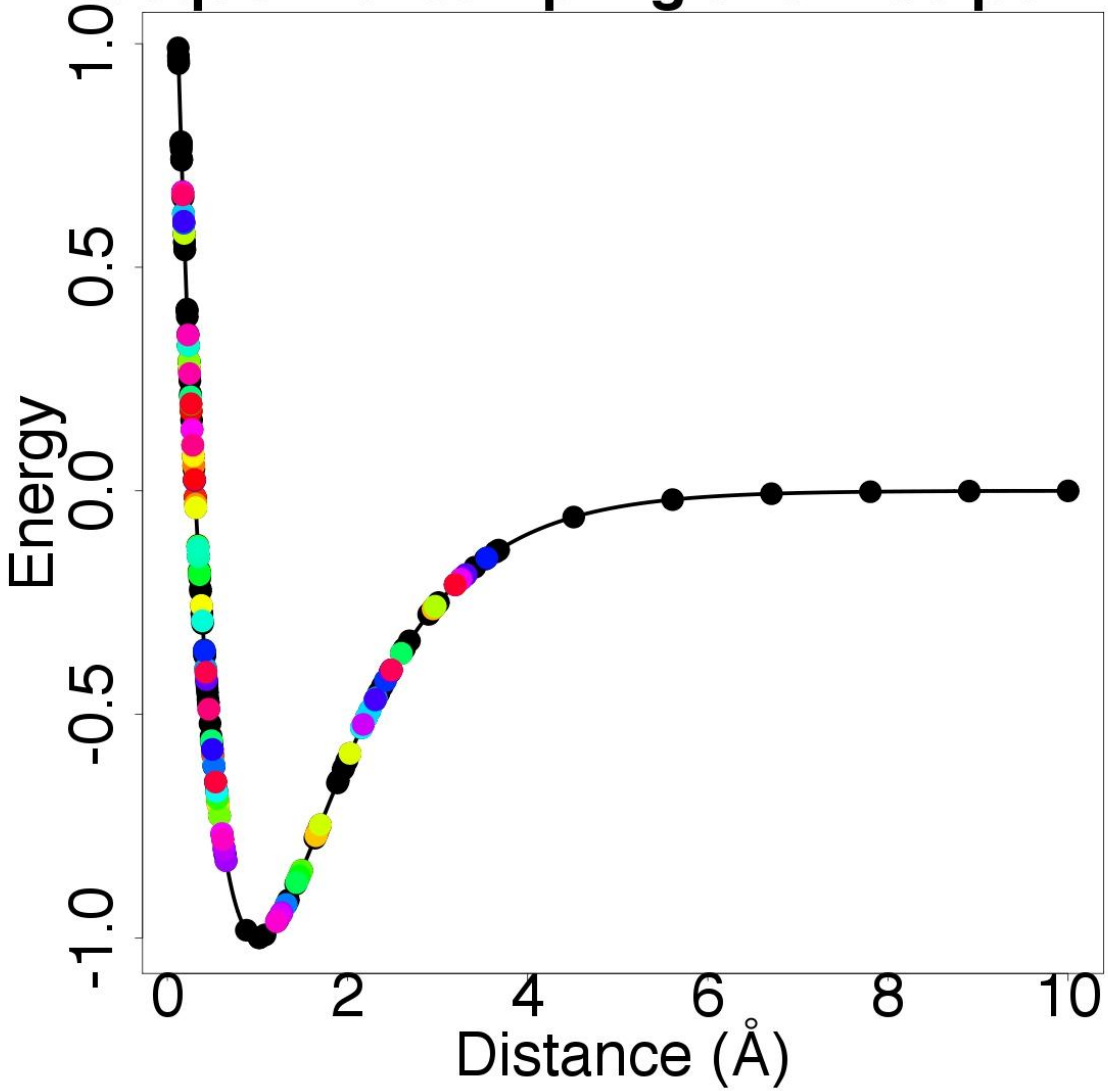
P. Balaprakash and A. Vazquez and OAvL, *in preparation* (2013)

Outlook: Selection bias

Balaprakash (ANL)

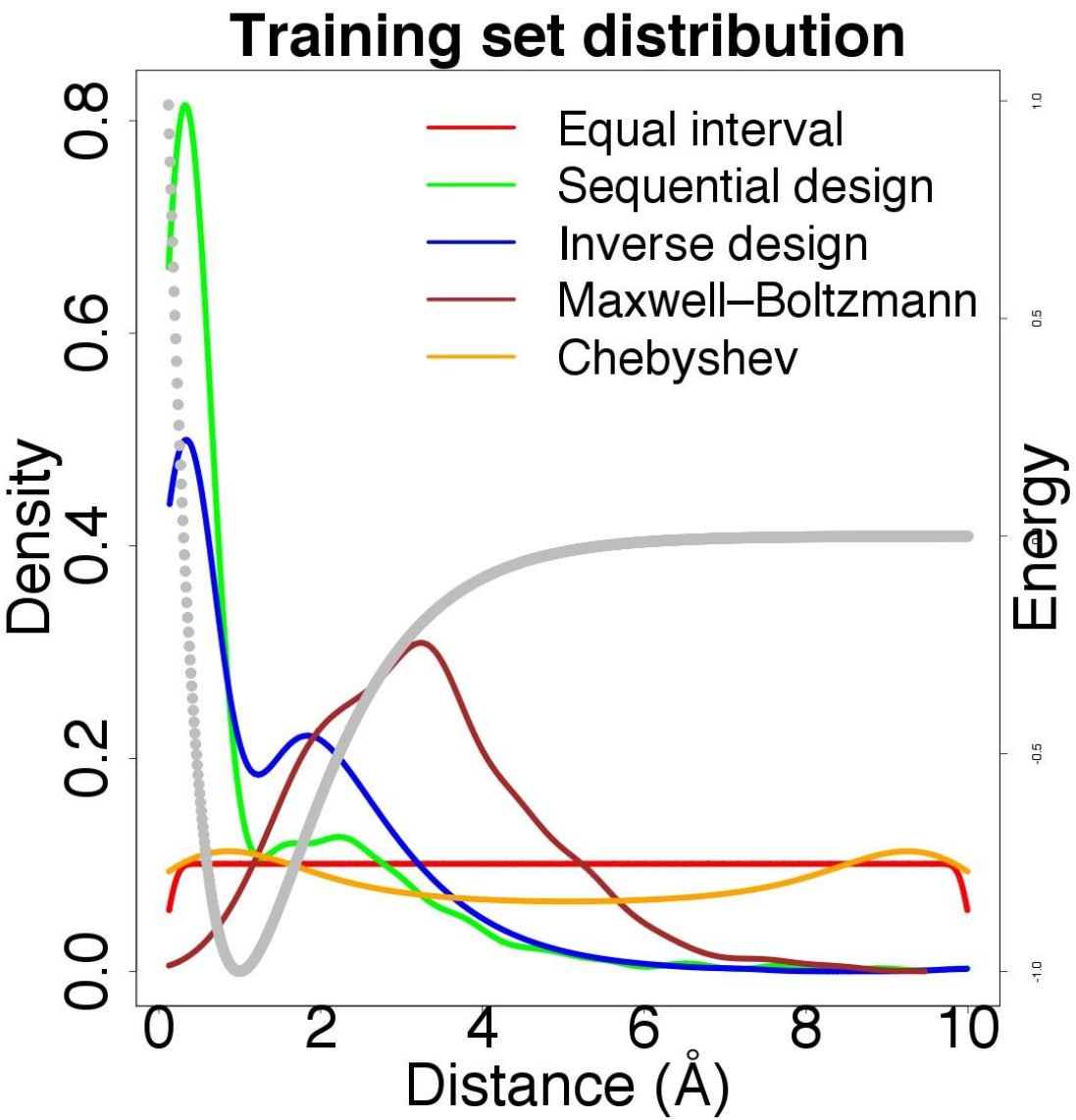
Vazquez (ANL)

Sequential sampling after 200 points



P. Balaprakash and A. Vazquez and OAvL, *in preparation* (2013)

Outlook: Selection bias



Balaprakash (ANL)

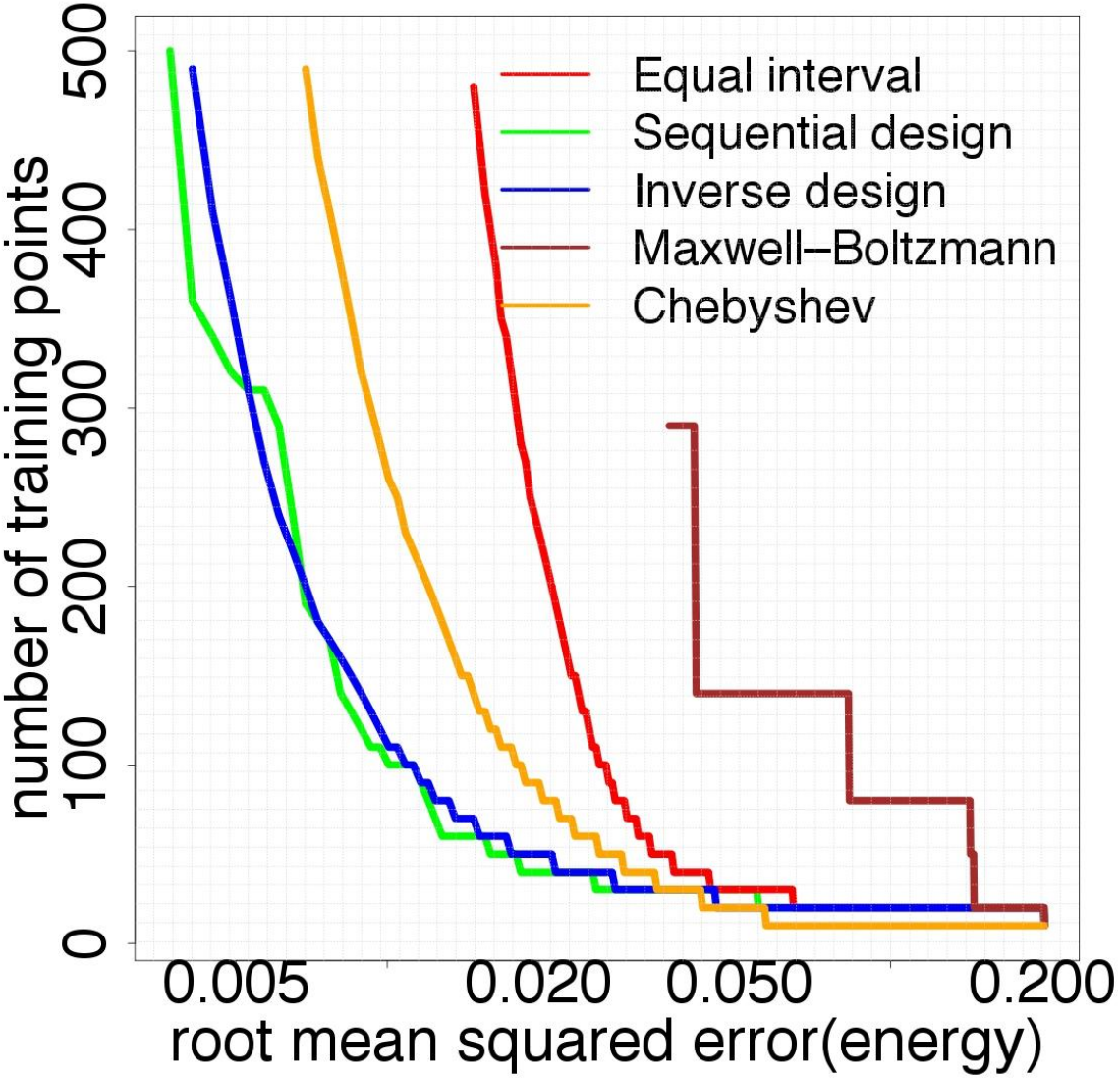
Vazquez (ANL)

P. Balaprakash and A. Vazquez and OAvL, *in preparation* (2013)

Outlook: Selection bias

Balaprakash (ANL)

Vazquez (ANL)



P. Balaprakash and A. Vazquez and OAvL, *in preparation* (2013)

First Principles

$$H(\{Z_I, \mathbf{R}_I\})\Psi(\mathbf{r}) = E\Psi(\mathbf{r})$$

$$H(\{Z_I, \mathbf{R}_I\}) = -\sum_i \nabla_i^2 - \sum_{I,i} \frac{Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}$$

Schrödinger

correlational (inductive)

Vapnik

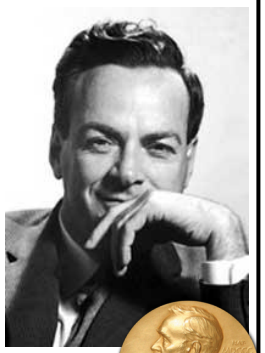
$$\{Z_I, \mathbf{R}_I\} \overset{H\Psi}{\longmapsto} E$$

supervised learning

$$\{Z_I, \mathbf{R}_I\} \overset{\mathrm{ML}}{\longmapsto} E$$

variational (deductive)

Feynman



$$\frac{\partial E[H]}{\partial R_{Ix}} = \left\langle \Psi \left| \frac{\partial H}{\partial R_{Ix}} \right| \Psi \right\rangle$$

$$\frac{\partial E[H]}{\partial Z_I} = \left\langle \Psi \left| \frac{\partial H}{\partial Z_I} \right| \Psi \right\rangle$$

$$E(H(\lambda)) = E(H_i + \lambda(H_f - H_i))$$

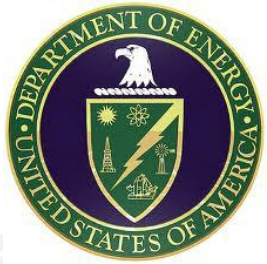
$$\frac{\partial E[H]}{\partial \lambda} = \left\langle \Psi \left| \frac{\partial H(\lambda)}{\partial \lambda} \right| \Psi \right\rangle$$

Thanks for your attention!

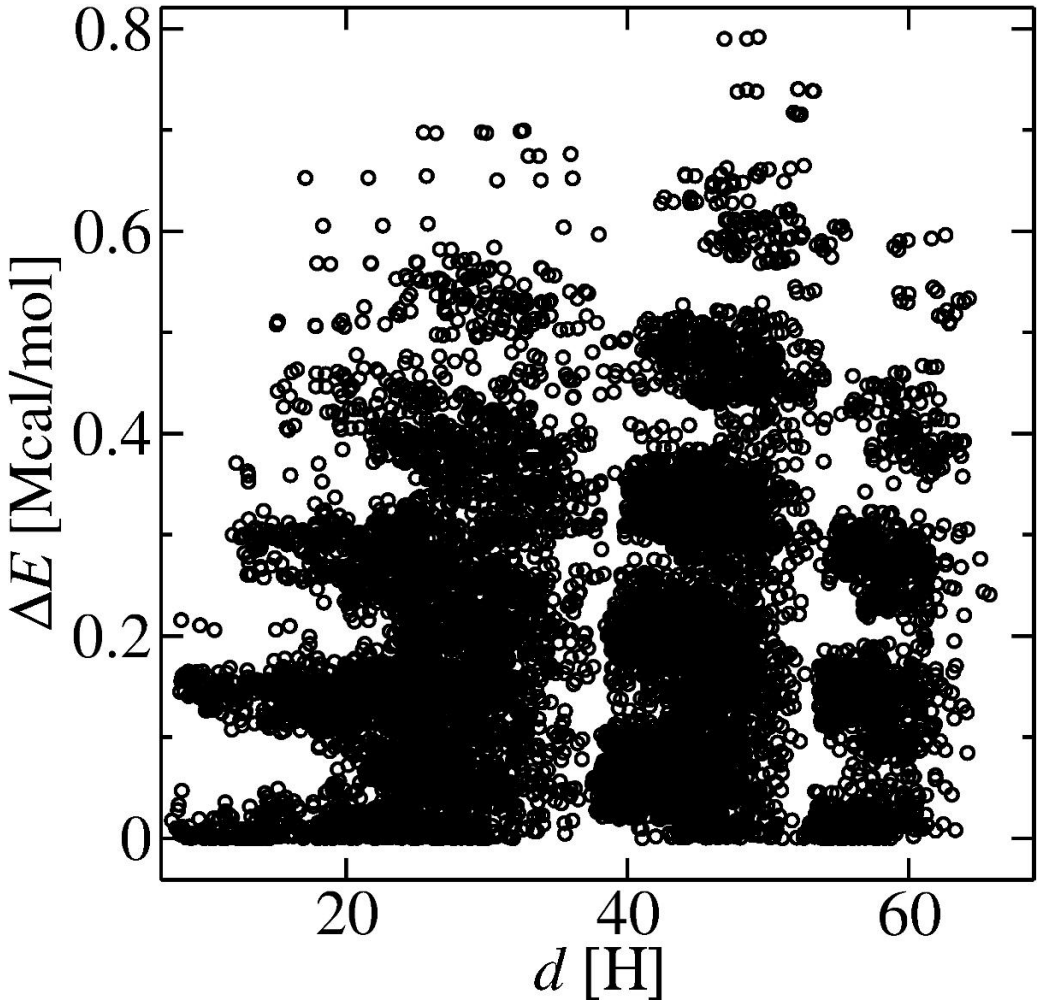
First principles view on chemical compound space: Gaining rigorous atomistic control of molecular properties
OAvL, Int J Quant Chem (2013), http://onlinelibrary.wiley.com/doi/10.1002/qua.24375/abstract

http://www.quantum-machine.org/

FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

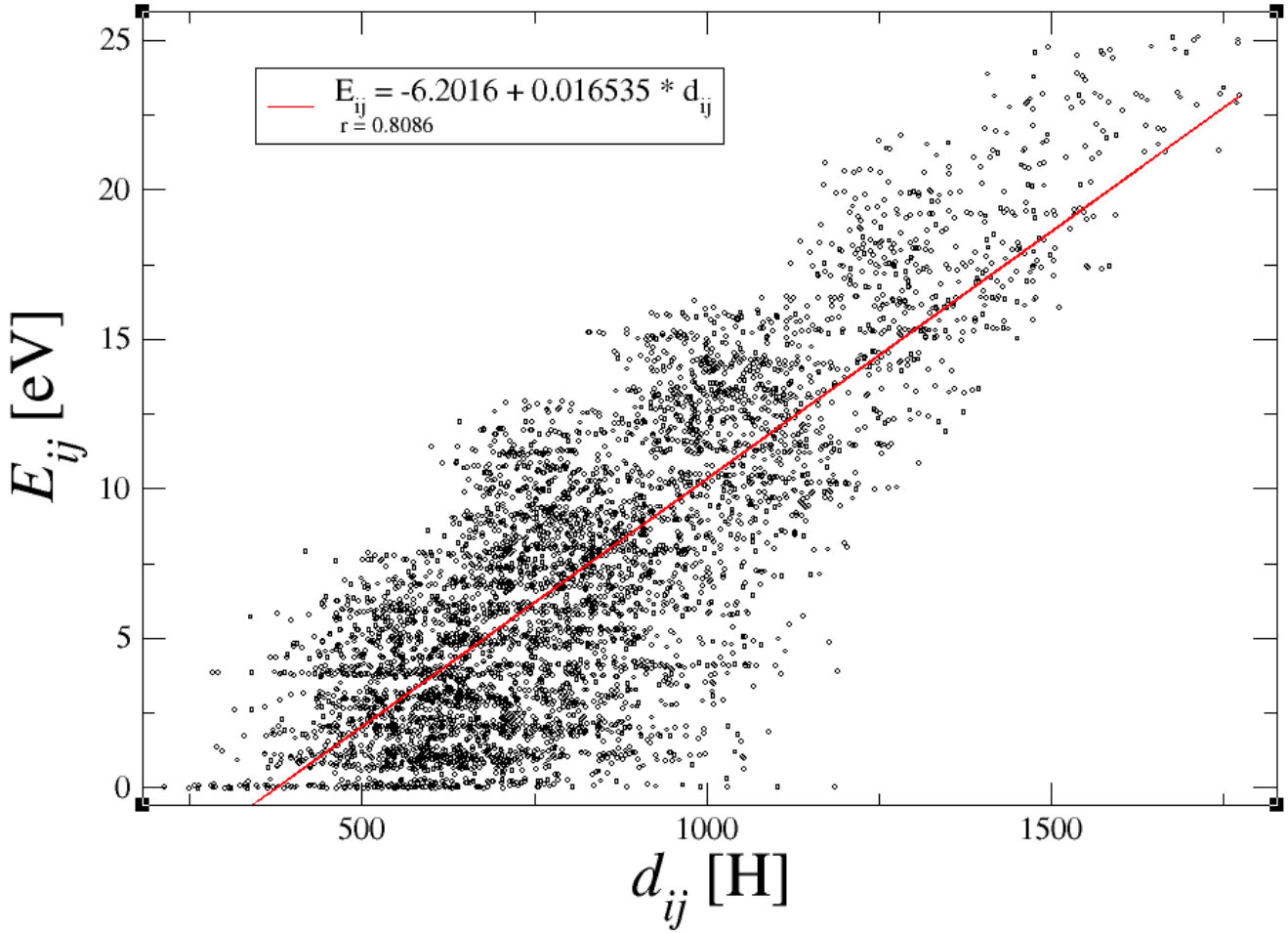


$$d(\mathbf{M}, \mathbf{M}_i) = \sqrt{\sum_{IJ} |M_{IJ} - M_{IJ}^{(i)}|^2}$$



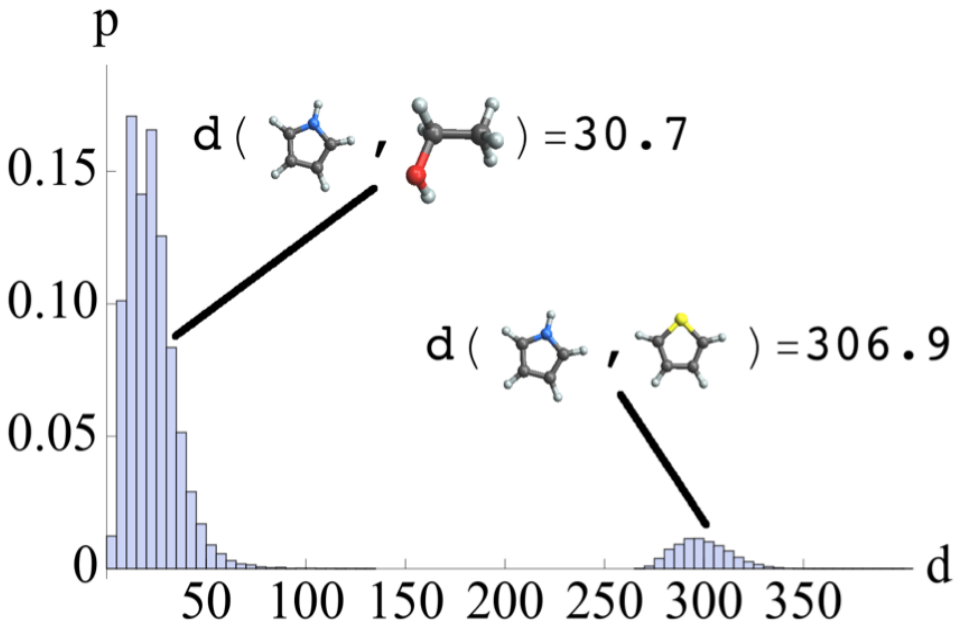
M. Rupp, A. Tkatchenko, K.-R. Müller, OAvL, *Phys Rev Lett* (2012)

$$D(M_i, M_j) = \sqrt{\int_{d=0}^{d \geq d_{IJ}^{max}} dd \, (R_i(d) - R_j(d))^2}$$



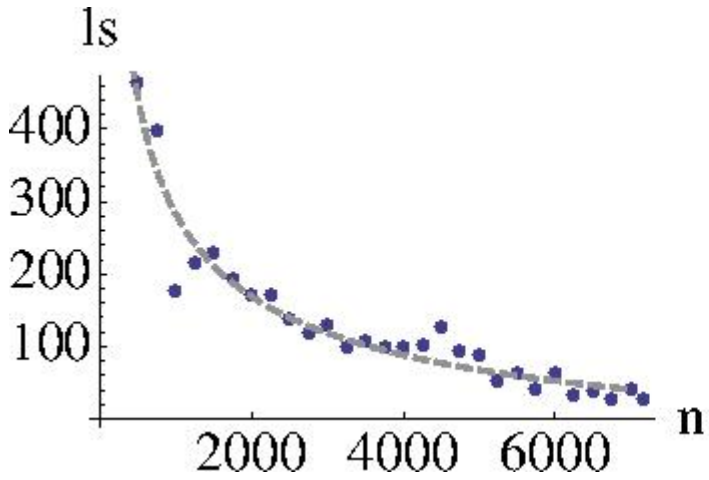
OAvL and A Knoll, *in preparation* (2013)

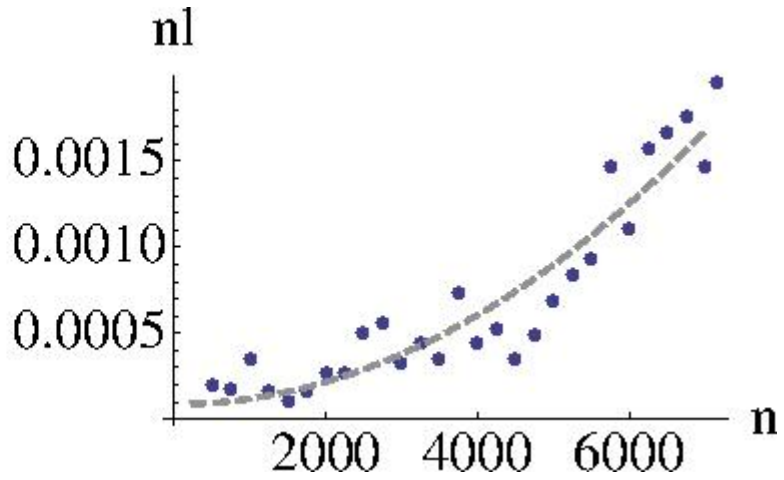
Locality



Model becomes local at ~5k molecules in training set

$$k(\mathbf{M}, \mathbf{M}') = \exp\left(-\frac{d(\mathbf{M}, \mathbf{M}')^2}{2\sigma^2}\right)$$

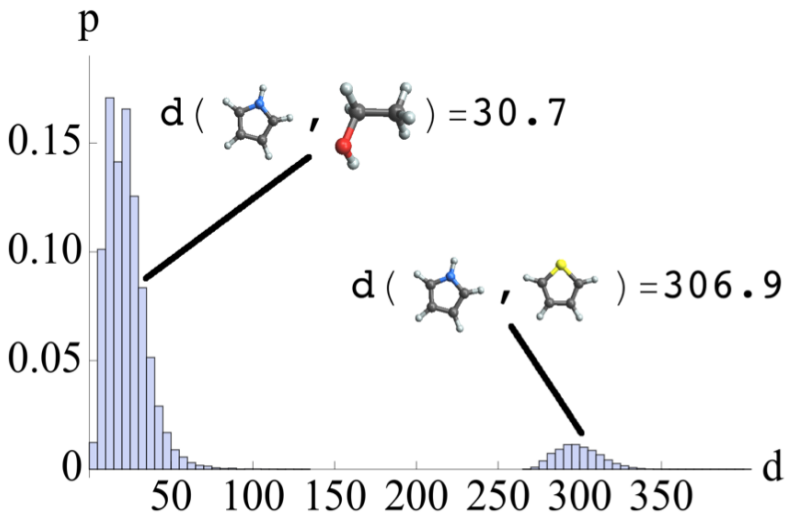




M. Rupp, A. Tkatchenko, K.-R. Müller, OAvL, *Phys Rev Lett* (2012)

Correlational: Kernel Ridge Regression

$$d(\mathbf{M}, \mathbf{M}') = \sqrt{\sum_{IJ} |M_{IJ} - M'_{IJ}|^2}$$



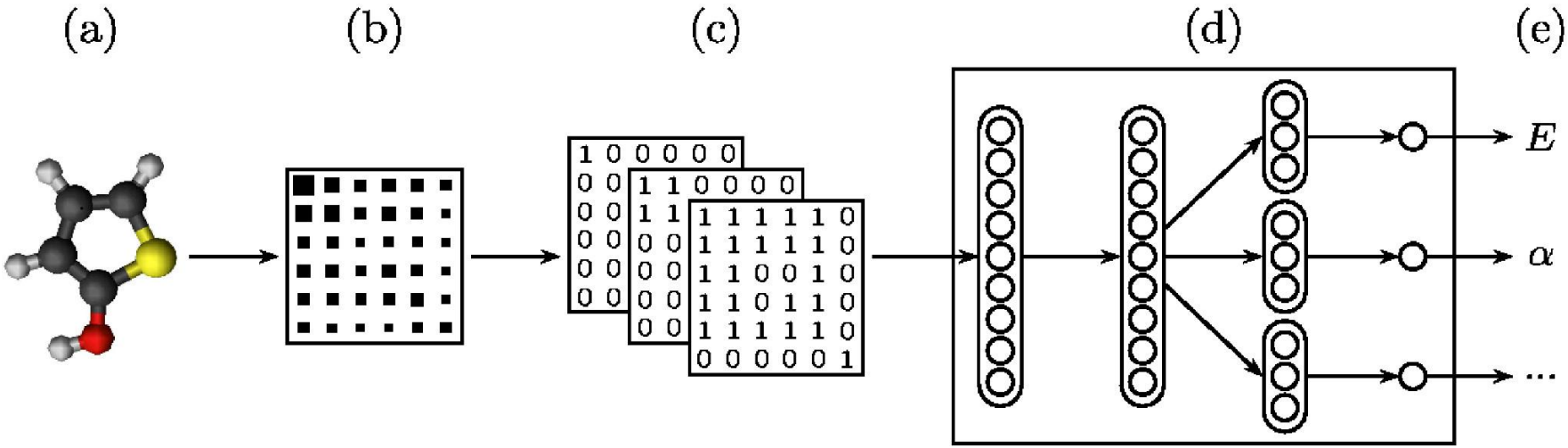
$$\min_{\alpha} \quad \sum_i \left(E^{est}(\mathbf{M}_i) - E_i^{ref}\right)^2 + \lambda \sum_i \alpha_i^2$$

$$\alpha = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{E}^{ref}$$

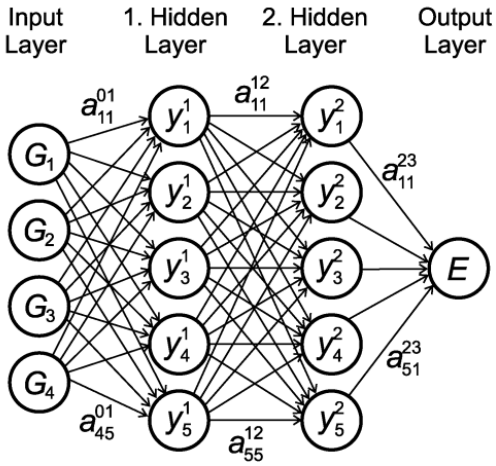
$$k(\mathbf{M}, \mathbf{M}') = \exp\left(-\frac{d(\mathbf{M}, \mathbf{M}')^2}{2\sigma^2}\right)$$

M. Rupp, A. Tkatchenko, K.-R. Müller, OAvL, *Phys Rev Lett* (2012)

Correlational: Regression 2



$$E = f_1^3\left(b_1^3 + \sum_{l=1}^{5} a_{l1}^{23} \cdot f_l^2\left(b_l^2 + \sum_{k=1}^{5} a_{kl}^{12} \cdot f_k^1\left(b_k^1 + \sum_{j=1}^{4} a_{jk}^{01} \cdot G_j\right)\right)\right)$$



T. B. Blank, S. D. Brown, A. W. Calhoun and D. J. Doren, *J. Chem. Phys.*, 1995, **103**, 4129.
J. Behler, *Phys Chem Chem Phys* (2011)

G. Montavon, M. Rupp, V. Gobre, A. Vazquez, K. Hansen, A. Tkatchenko, K.-R. Müller, OAvL, *submitted* (2012)

Transferability (no overfitting)

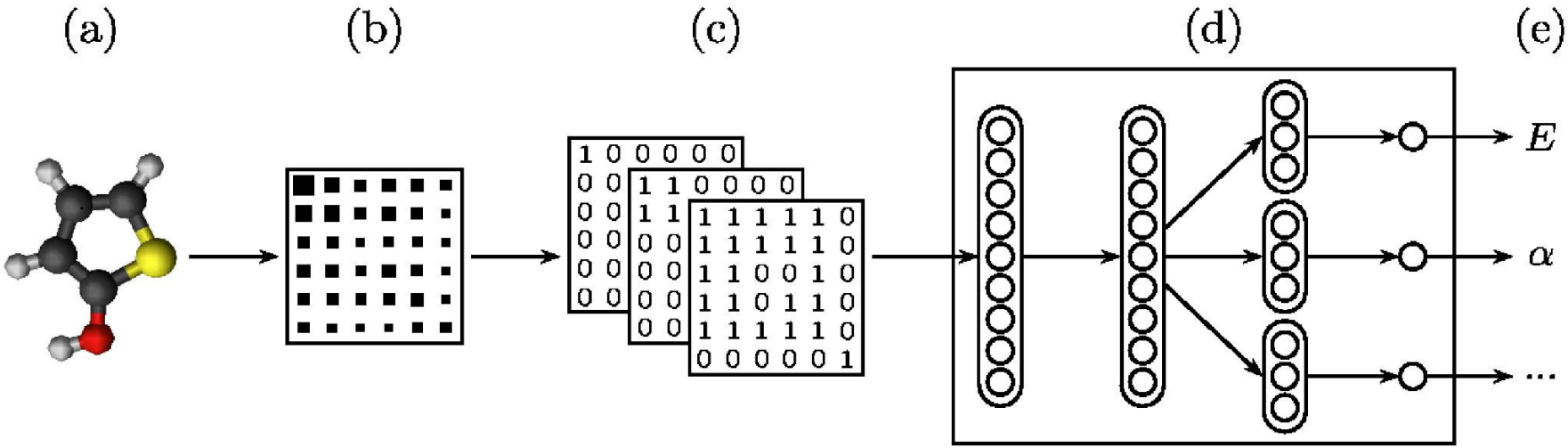
For training set: *k*-fold **cross-validation**

1. divide data into *k* blocks
2. predict each block with model trained on remaining blocks
3. average coefficients

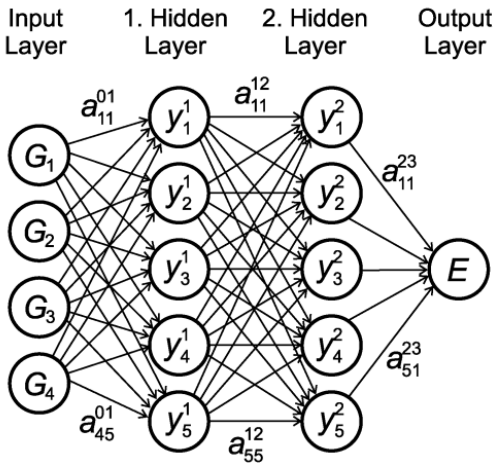
Two nested loops for training *and* hyper parameter optimization

Apply to test set to measure out-of-sample performance

Correlational: Regression 2



$$E = f_1^3\left(b_1^3 + \sum_{l=1}^{5} a_{l1}^{23} \cdot f_l^2\left(b_l^2 + \sum_{k=1}^{5} a_{kl}^{12} \cdot f_k^1\left(b_k^1 + \sum_{j=1}^{4} a_{jk}^{01} \cdot G_j\right)\right)\right)$$



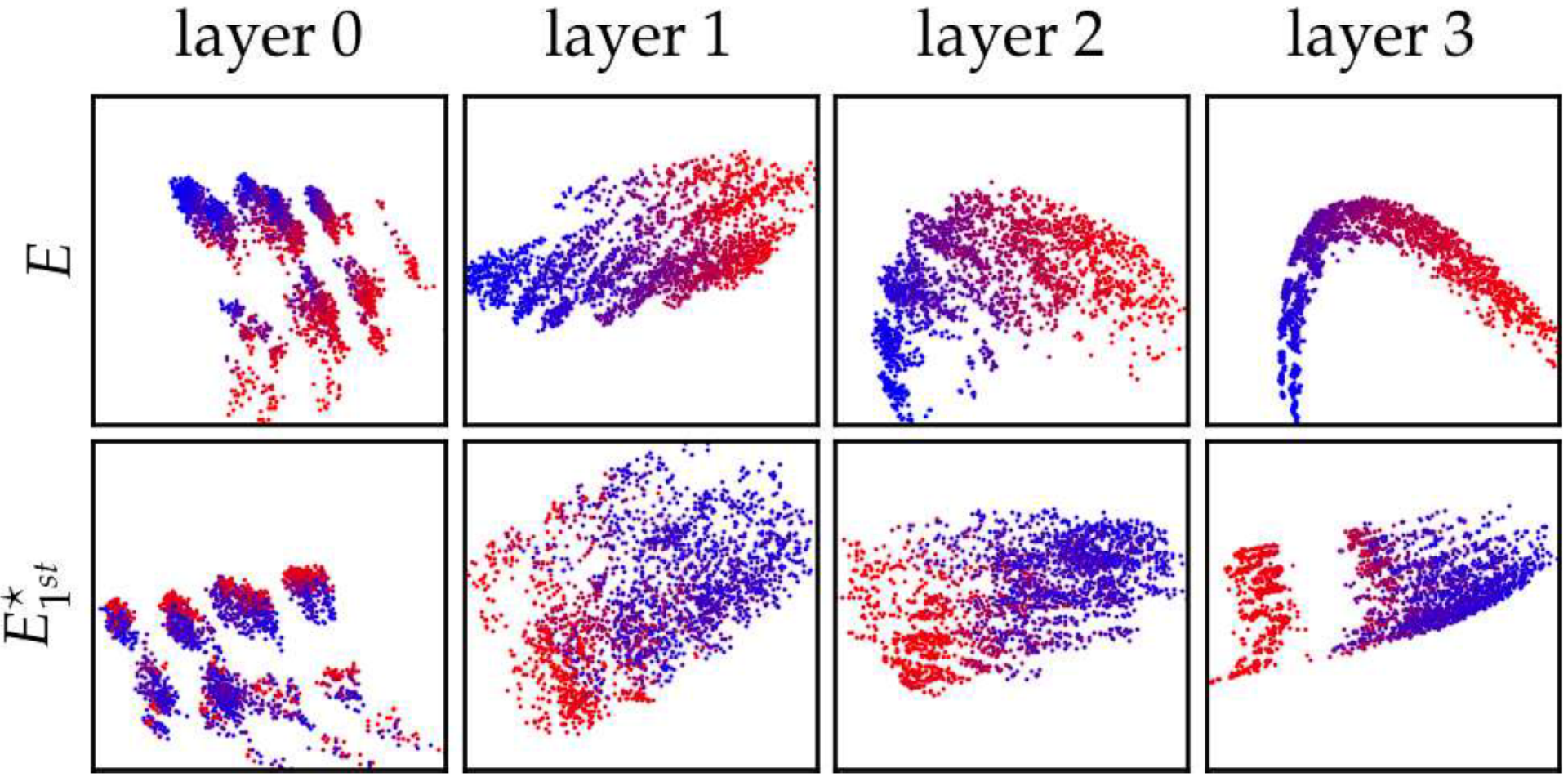
T. B. Blank, S. D. Brown, A. W. Calhoun and D. J. Doren, *J. Chem. Phys.*, 1995, **103**, 4129.

J. Behler, *Phys Chem Chem Phys* (2011)

G. Montavon, M. Rupp, V. Gobre, A. Vazquez, K. Hansen, A. Tkatchenko, K.-R. Müller, OAvL, *NJP* accepted (2013)

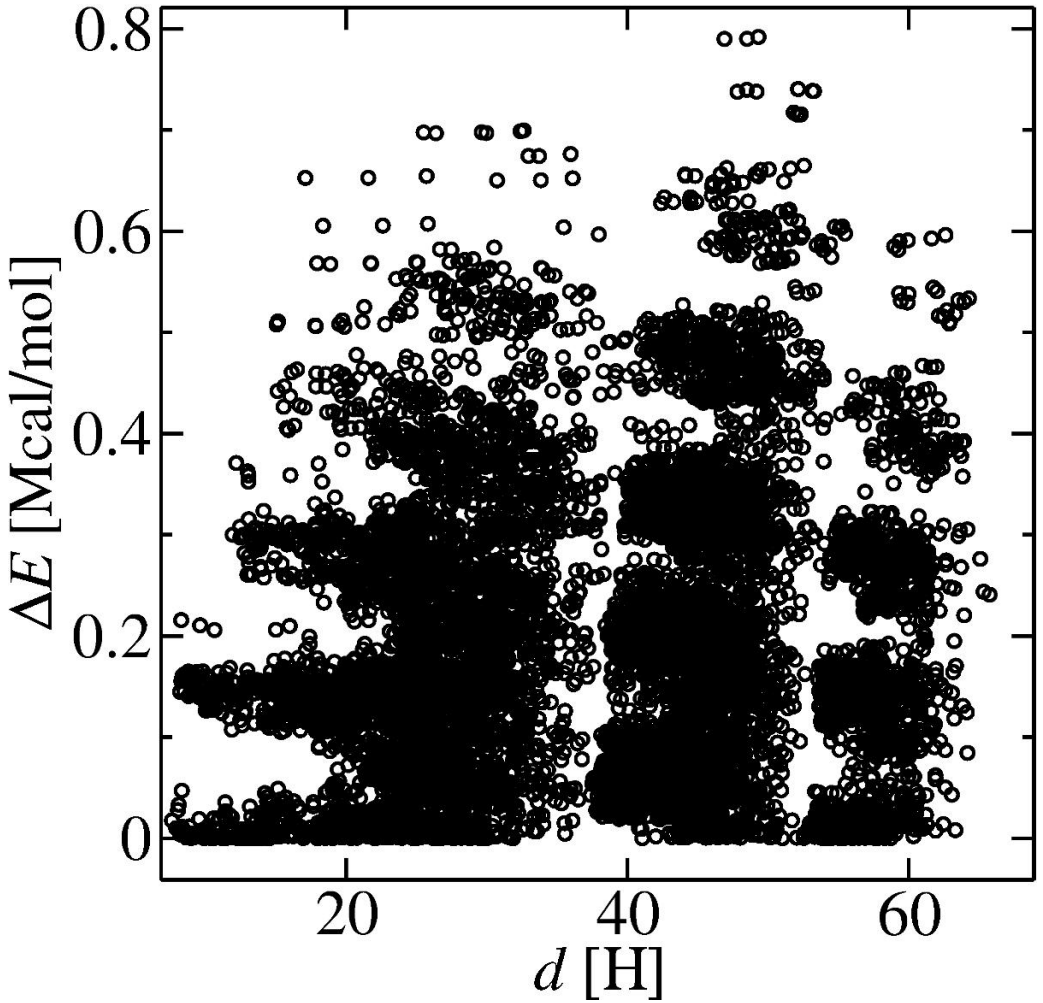
Deep Neural Networks

– PCA on properties for four layers



G. Montavon, M. Rupp, V. Gobre, A. Vazquez, K. Hansen, A. Tkatchenko, K.-R. Müller, OAvL, *NJP* accepted (2013)

$$d(\mathbf{M}, \mathbf{M}_i) = \sqrt{\sum_{IJ} |M_{IJ} - M_{IJ}^{(i)}|^2}$$

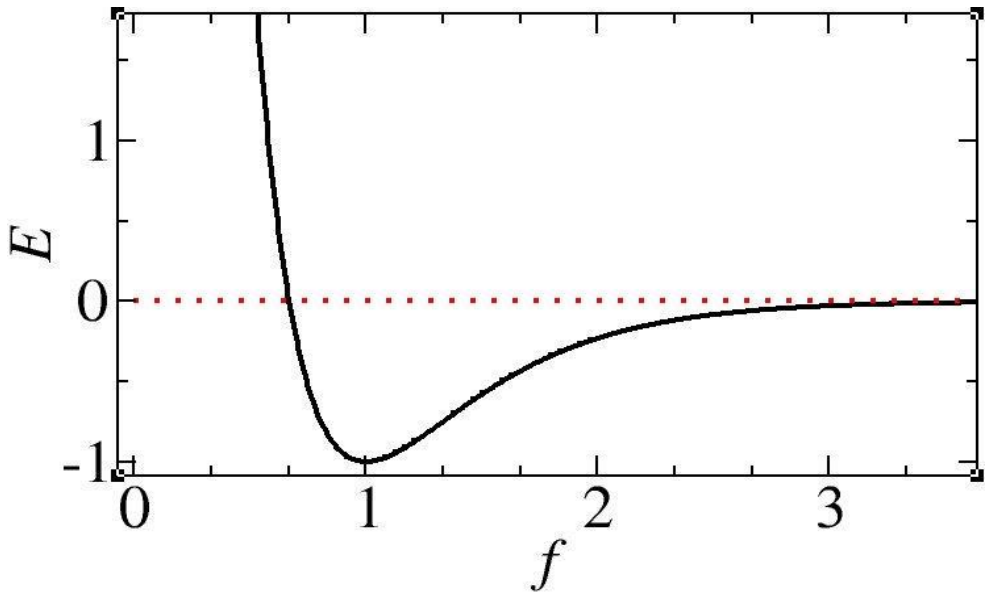


M. Rupp, A. Tkatchenko, K.-R. Müller, OAvL, *Phys Rev Lett* (2012)

Outlook: Forces

1. Interpolate binding:
   a. E(f=3) = 0
   b. E(f=1) = E(PBE0)
   c. dE/df(f=1) = 0
   d. E(f=2/3) = 0

Train on 1k molecules



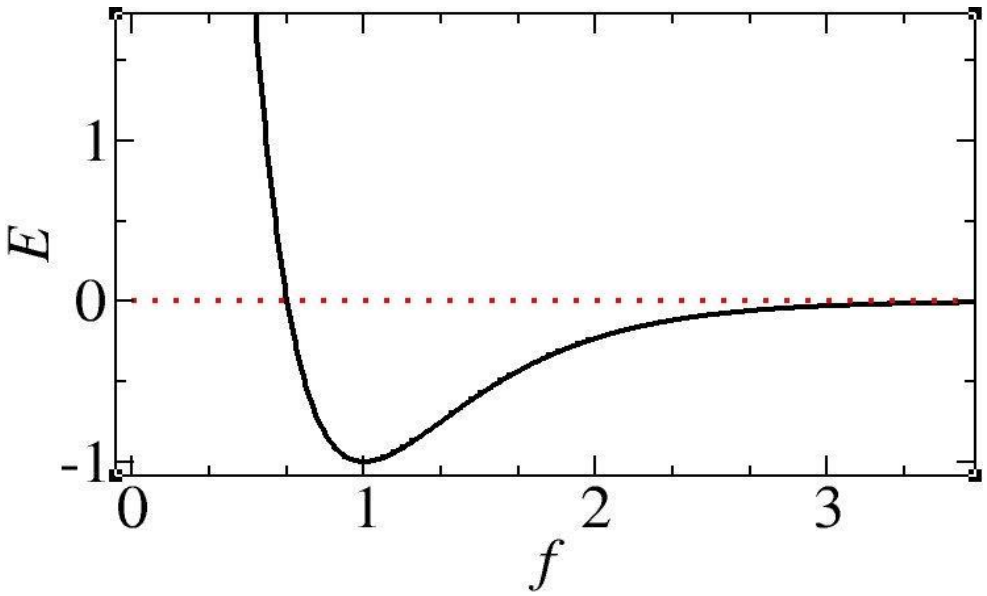
Outlook: Forces

1. Interpolate binding:
   a. E(f=3) = 0
   b. E(f=1) = E(PBE0)
   c. dE/df(f=1) = 0
   d. E(f=2/3) = 0

   Train on 1k molecules

2. Test on remaining 6k molecules

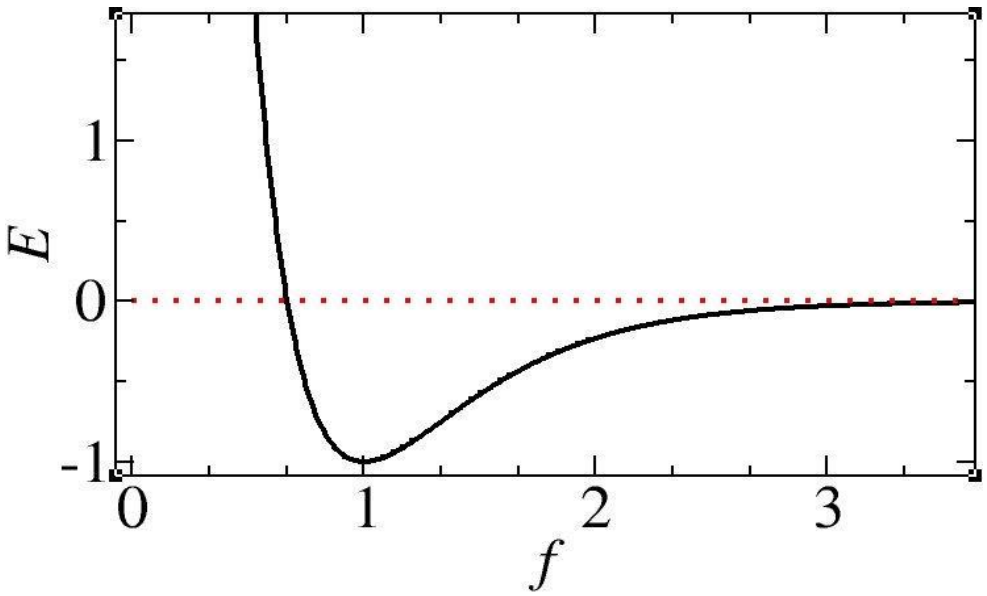
   MAE ~ 15kcal/mol



Outlook: Forces

1. Interpolate binding:
   a. E(f=3) = 0
   b. E(f=1) = E(PBE0)
   c. dE/df(f=1) = 0
   d. E(f=2/3) = 0

   Train on 1k molecules



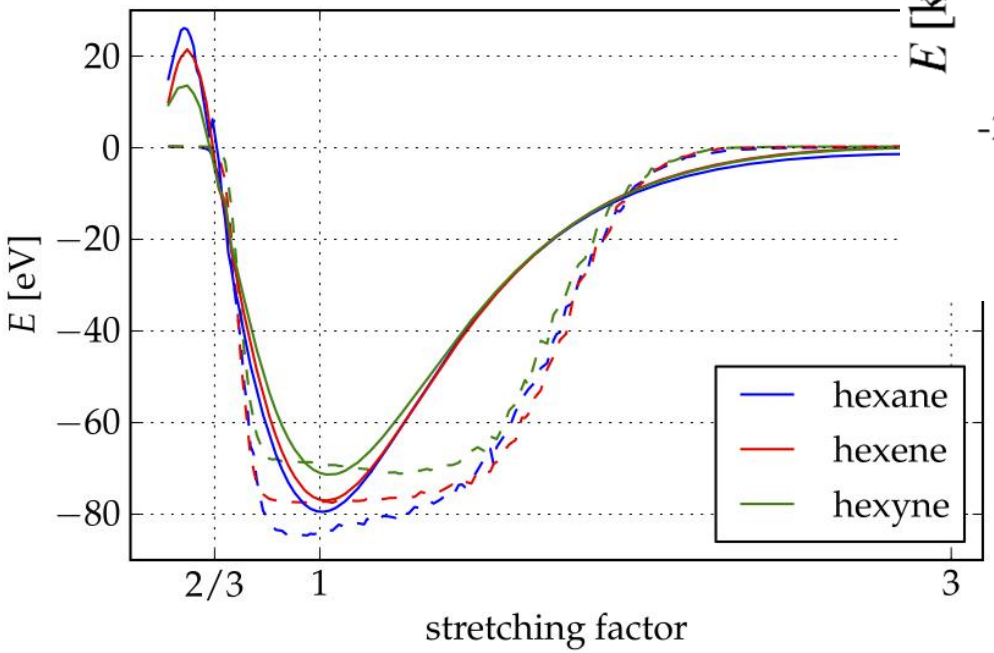
2. Test on remaining 6k molecules

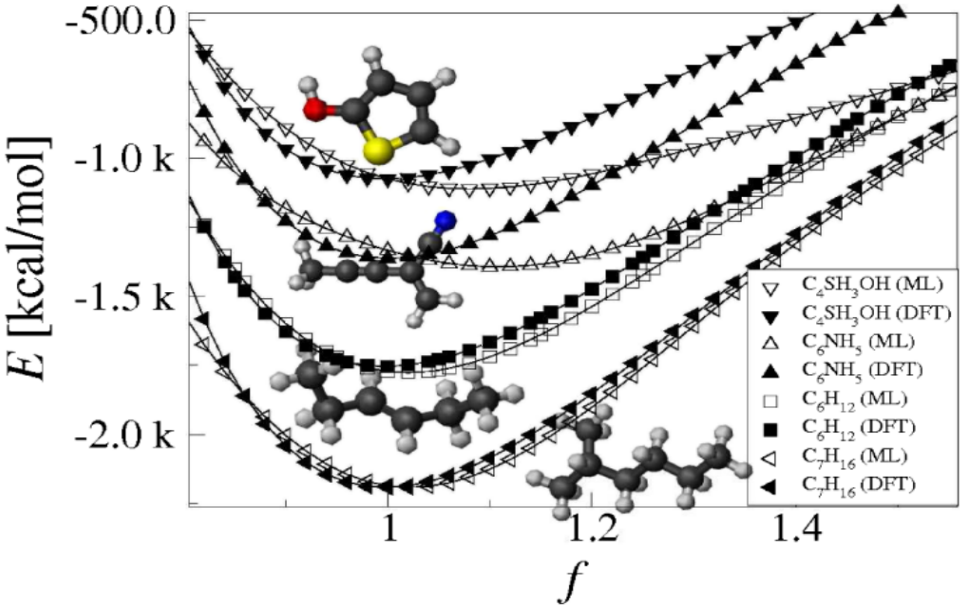
   MAE ~ 15kcal/mol

3. Experiment: Predict binding curve for some molecules

Outlook: Forces

Neural Networks





Kernel Ridge Regression

M. Rupp, A. Tkatchenko, K.-R. Müller, OAvL, *Phys Rev Lett* (2012)

G. Montavon, M. Rupp, V. Gobre, A. Vazquez, K. Hansen, A. Tkatchenko, K.-R. Müller, OAvL, *submitted* (2012)