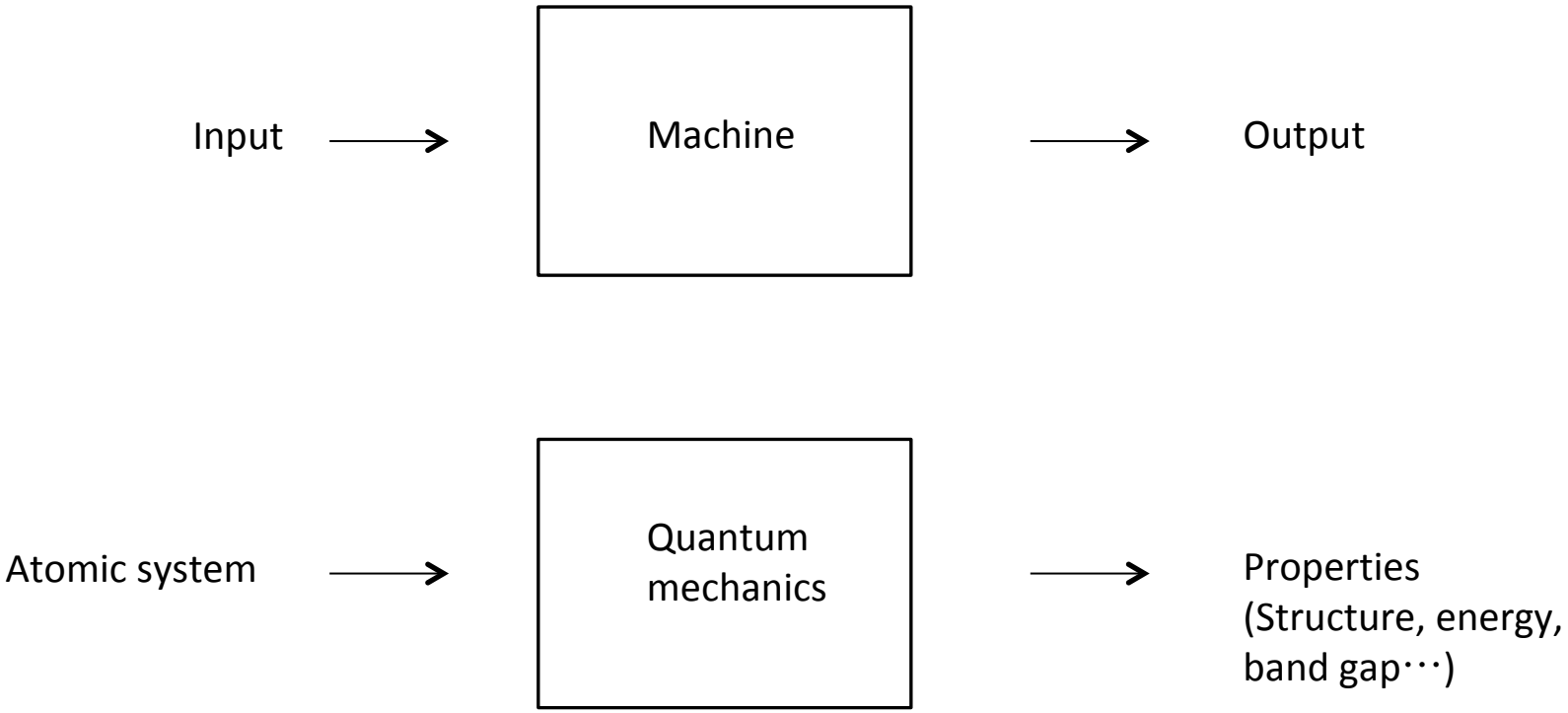
Machine Learning and the BEEF functionals

Karsten W. Jacobsen

CAMD, Dept. of Physics

Technical University of Denmark

Machine learning – why?



The machine is potentially much faster than quantum mechanics.

Maybe we can also understand something new from the machine?

Overview

- Computational screening
  - Descriptors
  - How to search?
    - Databases
    - Brute force calculations
    - **Machine learning**
      - Fitting a function
      - Bayesian inference
      - Gaussian processes and kernel regression
- Machine learned exchange-correlation functionals

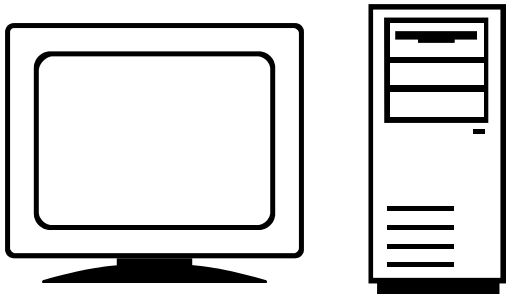
Computational Materials Design Descriptors

"Real material"

What do we want:

- Battery: High power, rechargeability, long lasting…
- Chemical reactor: High Turn-Over-Frequency…
- Structural material: High strength, ductility…
- Solar cell: High solar to electrical energy conversion efficiency
- Photoelectrochemical cell: High solar to fuel conversion efficiency

Identifying key parameters
Multiscale modeling

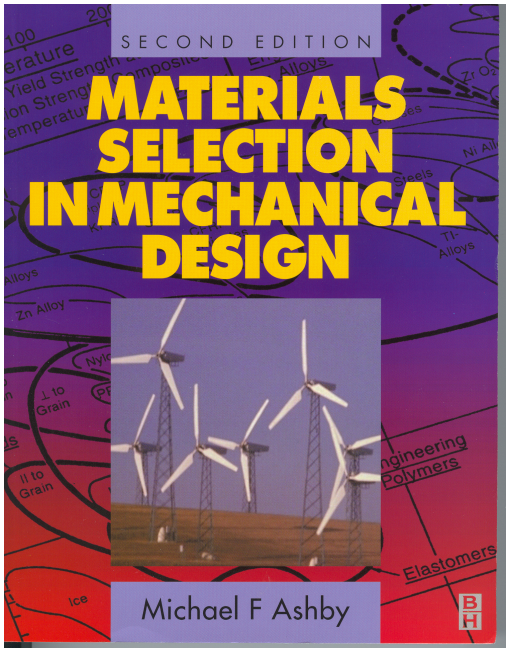


Computing

What can we compute at the electronic/atomic level?
"Descriptors"!

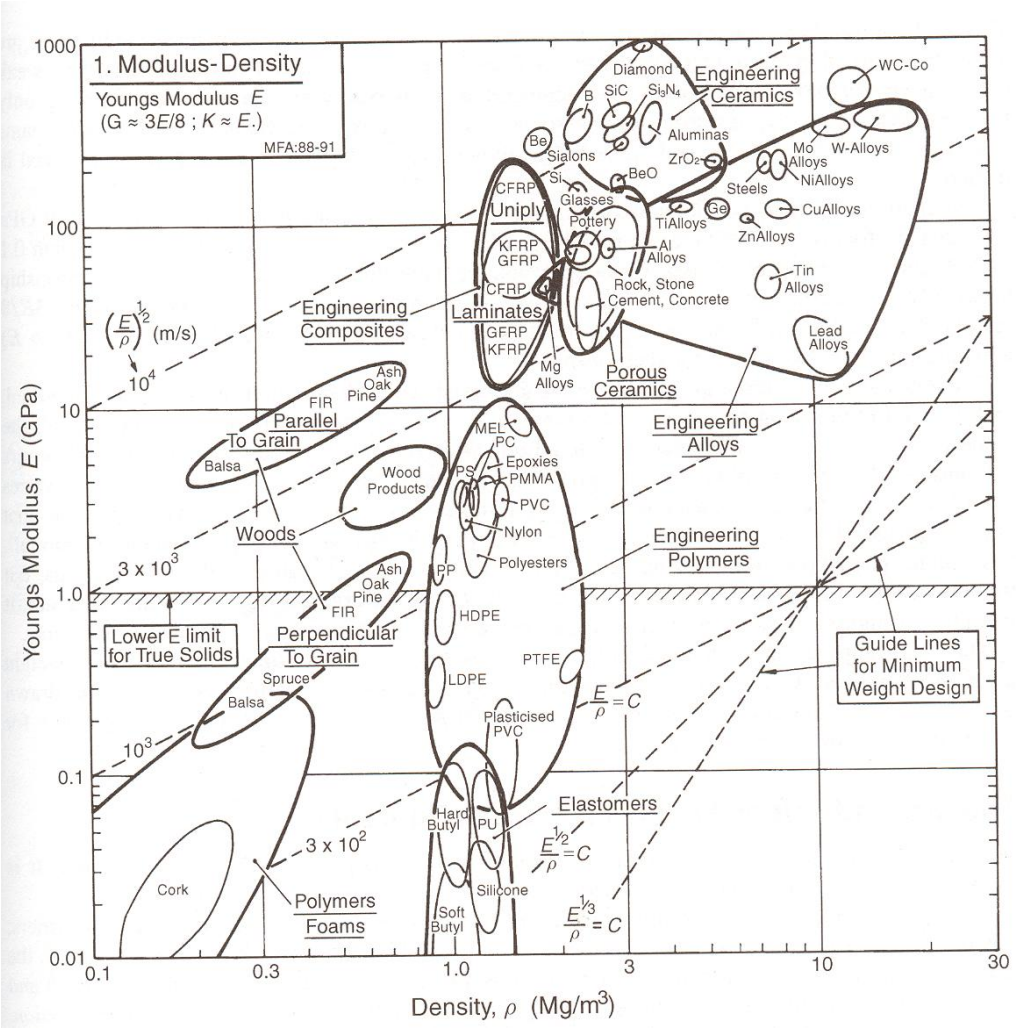
Ashby diagrams

"Materials Selector"



Great variety of materials
Multi-dimensionality
Known materials

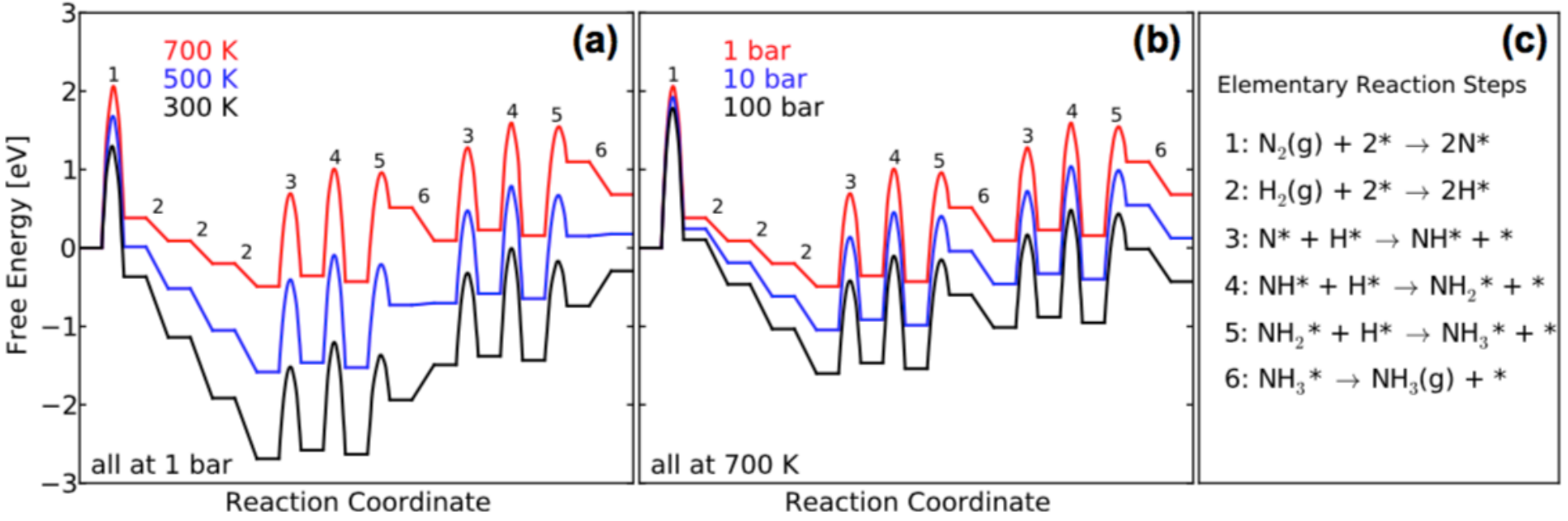
(1992)



Multiscale modeling: Microkinetics
Ammonia synthesis

$$N_2 + 3H_2 \rightarrow 2NH_3$$

Descriptors: Adsorption energies and reaction barriers
Multiscale modeling: Microkinetics



Number of descriptors can be reduced using scaling relations: the energy barrier scales with the binding energy

Vojvodic, Medford, Studt, Abild-Pedersen, Khan, Bligaard, and Nørskov, *Chemical Physics Letters*, **598**, 108 (2014)

Combined Electronic Structure and Evolutionary Search Approach to Materials Design

Superalloys

- Mechanical strength
- Resistance to thermal creep
- Surface stability
- Resistance to corrosion and oxidation

Descriptor

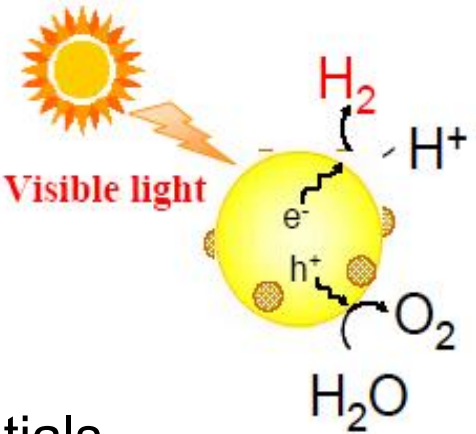
- Try alloy heat of formation

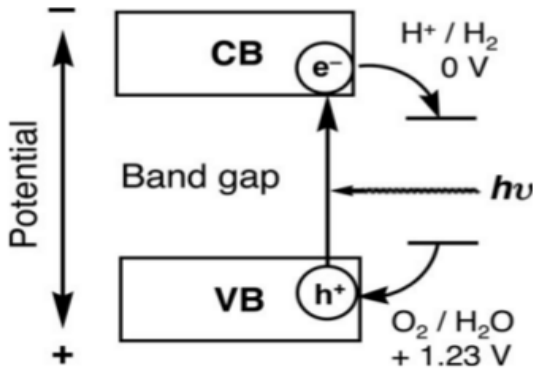
G. Johannesson, Thomas Bligaard, A. Ruban, H. L. Skriver, K. W. Jacobsen, and J. K. Nørskov, Phys. Rev. Lett. **88**, 255506 (2002)
T. Bligaard, G. Johannesson, A. V. Ruban, H. L. Skriver, K. W. Jacobsen, and J. K. Nørskov, Appl. Phys. Lett. (2003)

$AlNi_3$	-0.49
Ni_3Ti	-0.46
$HfNi_3$	-0.44
Al_2Ti_2	-0.43
Al_3Sc	-0.43
Al_2Zr_2	-0.42
Al_2ZnZr	-0.42
Al_2Sc_2	-0.41
Ni_3Sc	-0.41
Al_3Zr	-0.40
Al_2TiZn	-0.39
Al_2ScZn	-0.38
Al_3Ti	-0.38
Co_3Ti	-0.38
Ni_3Zr	-0.36
Al_2NbTi	-0.36
Al_2CuTi	-0.35
Al_2HfZn	-0.34
Al_2CuZr	-0.34
Al_3Lu	-0.34

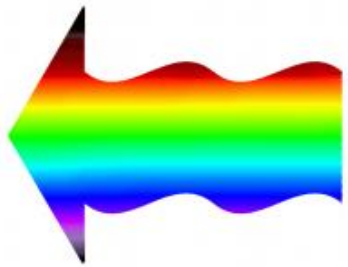
Light-induced water splitting

- **Stability of material**
  - Heat of formation
- **Good light absorption**
  - Bandgap in the visible range
- **Photogenerated charges at right potentials**
  - Band edges straddle the water redox potentials





Principle of water splitting using semiconductor photocatalysts.

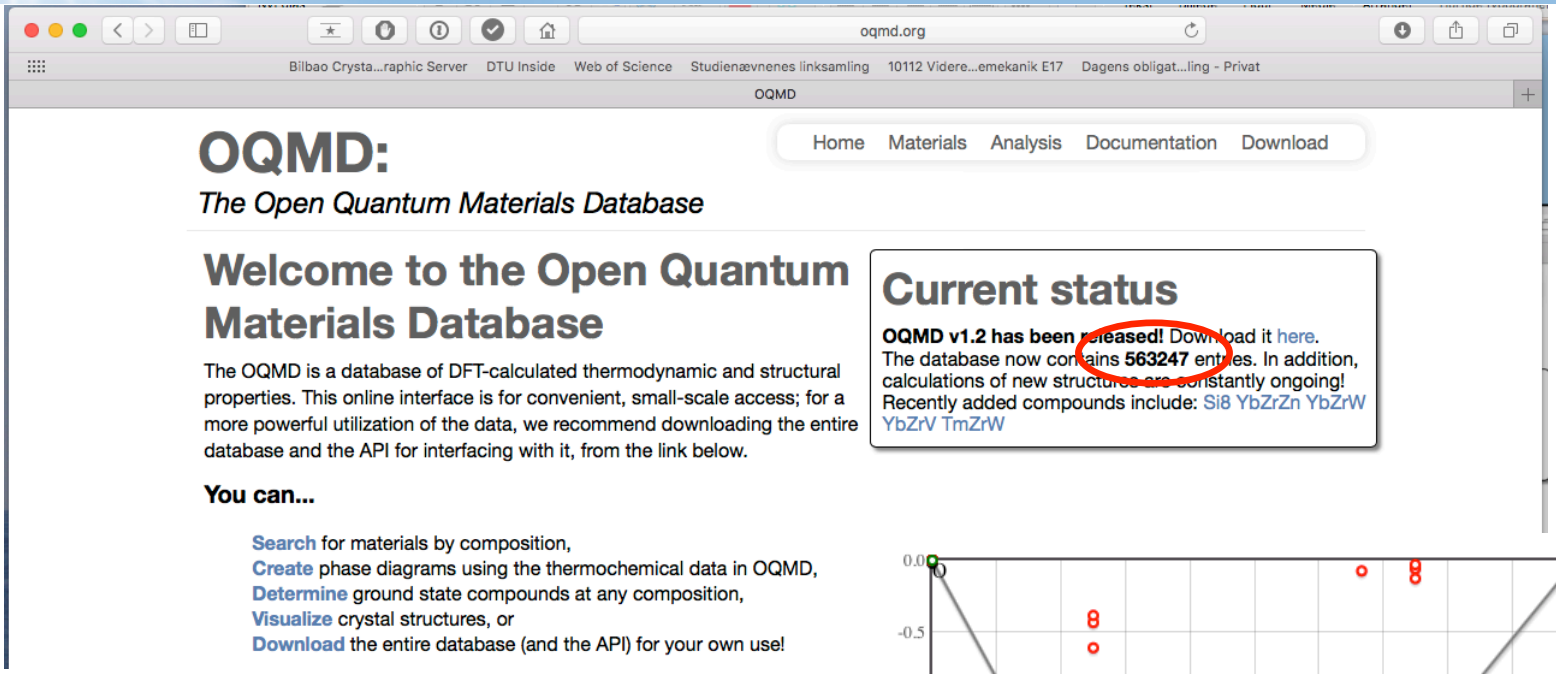


(Fujishima and Honda, Nature 1972)

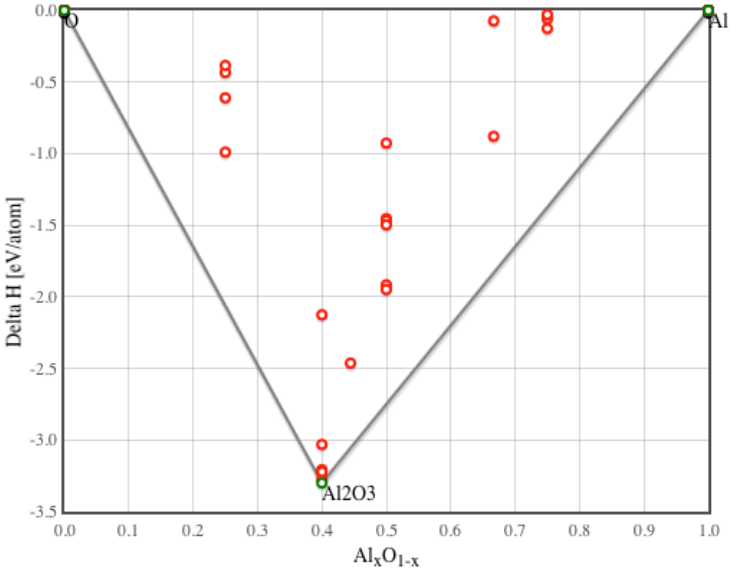
Overview

- Computational screening
  - Descriptors
  - How to search?
    - Databases
    - Brute force calculations
    - **Machine learning**
      - Fitting a function
      - Bayesian inference
      - Gaussian processes and kernel regression
      - Neural nets
- Machine learned exchange-correlation functionals

“Standard” material quantities now available in public computational databases



Al_xO_{1-x} heats of formation



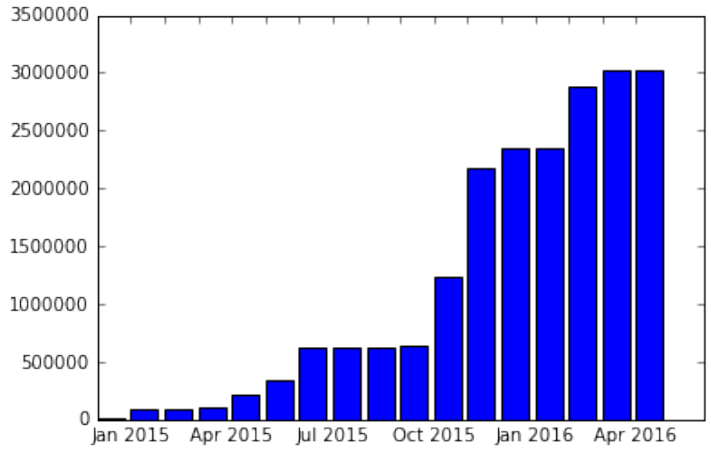
Stevanovic, Lany, Zhang, Zunger, *Phys. Rev. B*, **85**, 115104 (2012)

J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, *JOM*, vol. 65, no. 1, pp. 1501–1509, Nov. 2013.

Computational Databases

- OQMD – ICSD + specific structures
- Materials Project – ICSD, project specific
- AFLOWLIB – ICSD, project specific
- NOMAD Repository – store everything
- AiiDA
- CatApp/CatMap
- Computational Materials Repository (CMR) …
- …

Number of calc. in NOMAD



Now more than
49 million calculations
in NOMAD

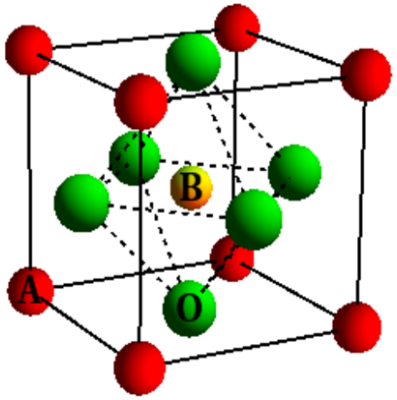
- *Several experimental databases*
  - Inorganic Crystal Structure Database (ICSD)
  - …

Brute force screening: Water splitting

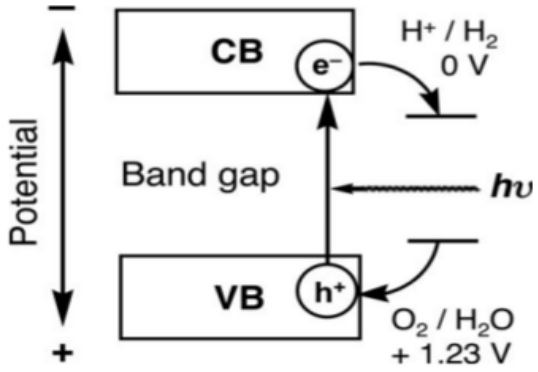
- **Stability**
  - Heat of formation
- **Good light absorption**
  - Bandgap in the visible range
- **Photogenerated charges at right potentials**
  - Band edges straddle the water redox potentials

Perovskites

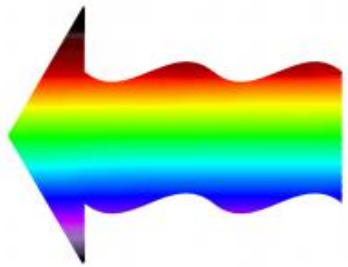
ABO_3



5 atom unit cell



Principle of water splitting using semiconductor photocatalysts.



Oxides, oxynitrides, oxysulfides, oxyfluorides, oxyfluornitrides

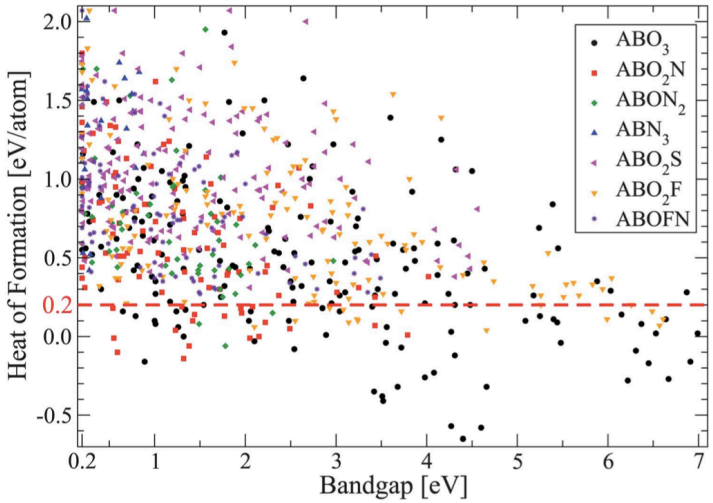
Materials candidates:

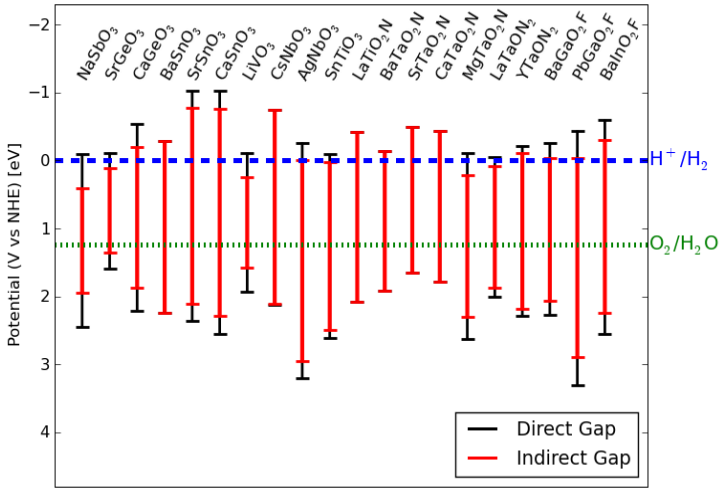
- ABO_3 10 4 known, 6 unknown
- ABO_2N 5 4 known, 1 unknown
- $ABON_2$ 2 $LaTaON_2$ (known) $YTaON_2$ (unknown)
- ABN_3 0
- ABO_2S 0
- ABO_2F 3
- ABOFN 0

20 candidate materials

About half are known

~19000 materials





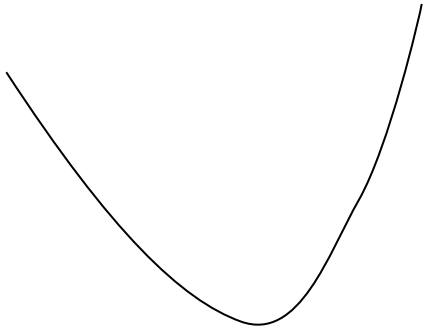
(Castelli, Landis, Thygesen, Dahl, Chorkendorff, Jaramillo, Jacobsen, Energy Environ Sci **5**, 9034 (2012))

More “intelligent” searches

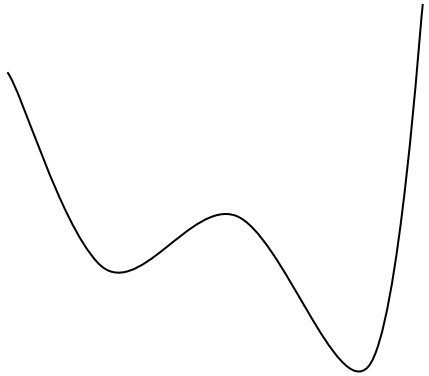
Correlation is the key

Find the minimum value

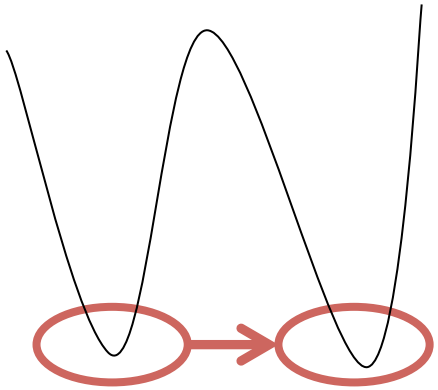
Easy (why?)



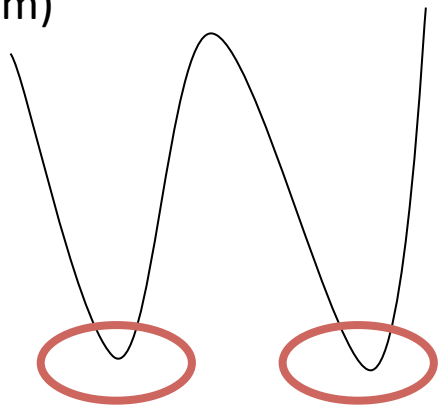
More difficult



Really difficult



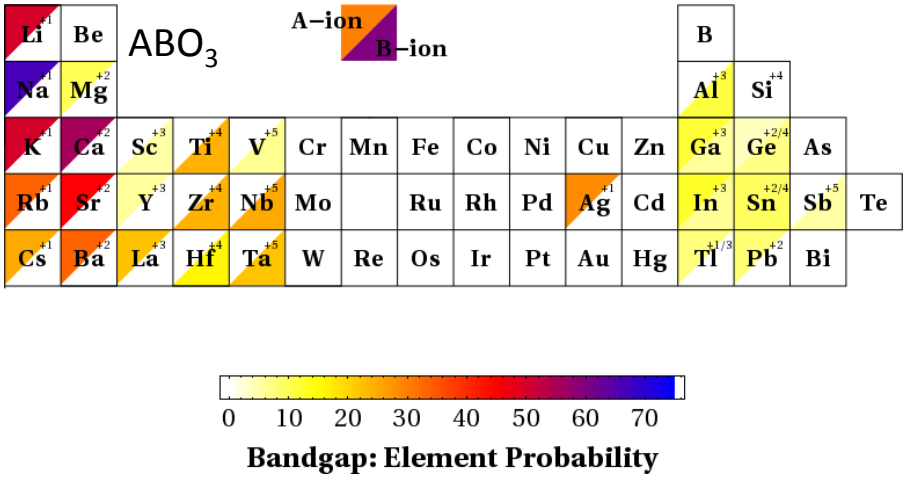
Impossible (random)



The minima resemble each other somehow
Smart moves from one well to another (genetic algorithm)

No similarity

Information transfer: From oxides to oxynitrides



Probability for an element to generate a stable semiconductor for the ABO_3 stoichiometry

Probabilities and rules transferable to oxynitrides:

Sum of valences = 0

$$\tilde{P}(\mathrm{ABO_2N}) = P_A(\mathrm{A})P_B(\mathrm{B})P_{\mathrm{rules}}(\mathrm{ABO_2N})$$

Castelli and Jacobsen,
Modelling Simul. Mater. Sci. Eng. **22**, 055007 (2014)

Ranking of oxynitrides

14 out of 16 compounds quickly identified

A-ion	B-ion	Stable?	Gap [eV]
Ca	Ta	✓	2.2
Sr	Ta	✓	2.1
Ca	Nb	✓	1.4
Sr	Nb	✓	1.4
Ba	Ta	✓	2.0
Ba	Nb	✓	1.3
La	Zr	✓	3.4
La	Hf	✓	3.8
La	Ti	✓	2.5
Sn	Ta	✓	1.2
Sn	Nb	✓	0.5
Pb	Ta	✓	2.0
Pb	Nb	✓	1.3
Sr	Sb		0
Ca	Sb		0
Sr	V		0
Ca	V		0
In	Zr		0
Mg	Ta	✓	2.1
Mg	Nb		1.5
Ba	Sb		0
Ba	V		0

A-ion	B-ion	Stable?	Gap [eV]
(cont.)			
In	Hf		0
La	Sn		1.8
In	Ti		0
La	Ge		0
Y	Zr		3.3
Ge	Ta		1.8
Ge	Nb		1.1
Y	Hf		3.4
In	Sn		0
Y	Ti		2.4
Sn	Sb		0
Pb	Sb		0
Sn	V		0
Pb	V		0
In	Ge		0
Mg	Sb		0
Mg	V		0
Y	Sn		2.7
Y	Ge		1.3
Ge	Sb		0
Ge	V		0

Machine learning
Kernel regression

Fitting a function *f(x)* based on data points $y_i = f(x_i)$

Drop a Gaussian on each data point:

$$k(x, x_i) = \exp(-|x - x_i|^2/2\rho^2)$$

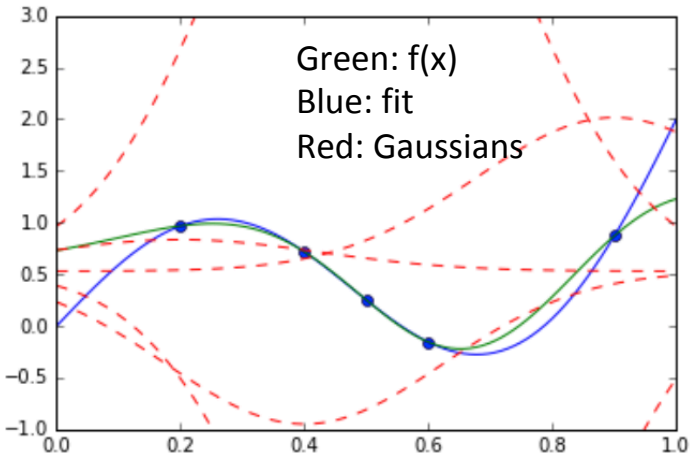
Interpolation: $y(x) = \sum_i k(x, x_i)\alpha_i$

Coefficients determined by data points:

$$y_j = \sum_i k(x_j, x_i)\alpha_i = \sum_i K_{ji}\alpha_i \rightarrow \mathbf{y} = \mathbf{K}\boldsymbol{\alpha} \rightarrow \boldsymbol{\alpha} = \mathbf{K}^{-1}\mathbf{y}$$

Interpolation: $y(x) = \mathbf{k}^T\mathbf{K}^{-1}\mathbf{y}$

with $k_i = k(x, x_i)$



Machine learning
Kernel ridge regression

Fitting a function *f(x)* based on data points $y_i = f(x_i)$

Drop a Gaussian on each data point:

$$k(x, x_i) = \exp(-|x - x_i|^2/2\rho^2)$$

Interpolation: $y(x) = \sum_i k(x, x_i)\alpha_i$

Coefficients determined by data points:

$$y_j = \sum_i k(x_j, x_i)\alpha_i = \sum_i K_{ji}\alpha_i \rightarrow \mathbf{y} = \mathbf{K}\boldsymbol{\alpha} \rightarrow \boldsymbol{\alpha} = \mathbf{K}^{-1}\mathbf{y}$$

Noise in data or problem inverting **K**:

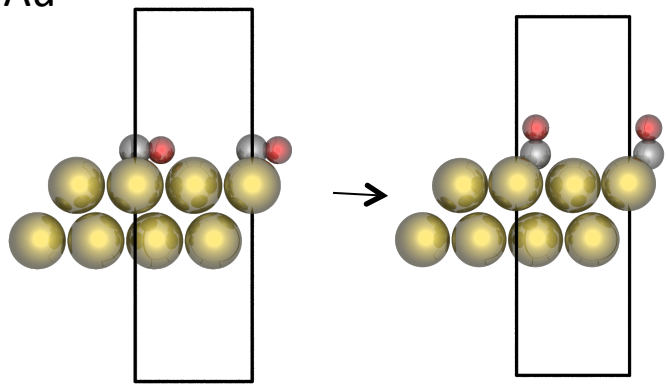
$$\mathbf{K} \rightarrow \mathbf{K} + \lambda\mathbf{I}$$

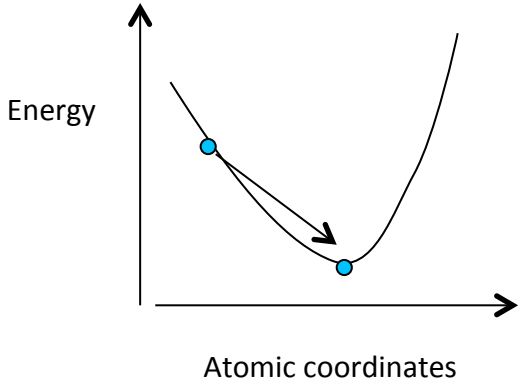
Small regularization parameter

Example: Local structure optimization of atomic systems

- Multidimensional local optimization
- A number of well-developed techniques are available: Conjugate Gradients, BFGS, …
- Takes up a large fraction of CPU hours on supercomputers performing electronic structure calculations

CO/Au





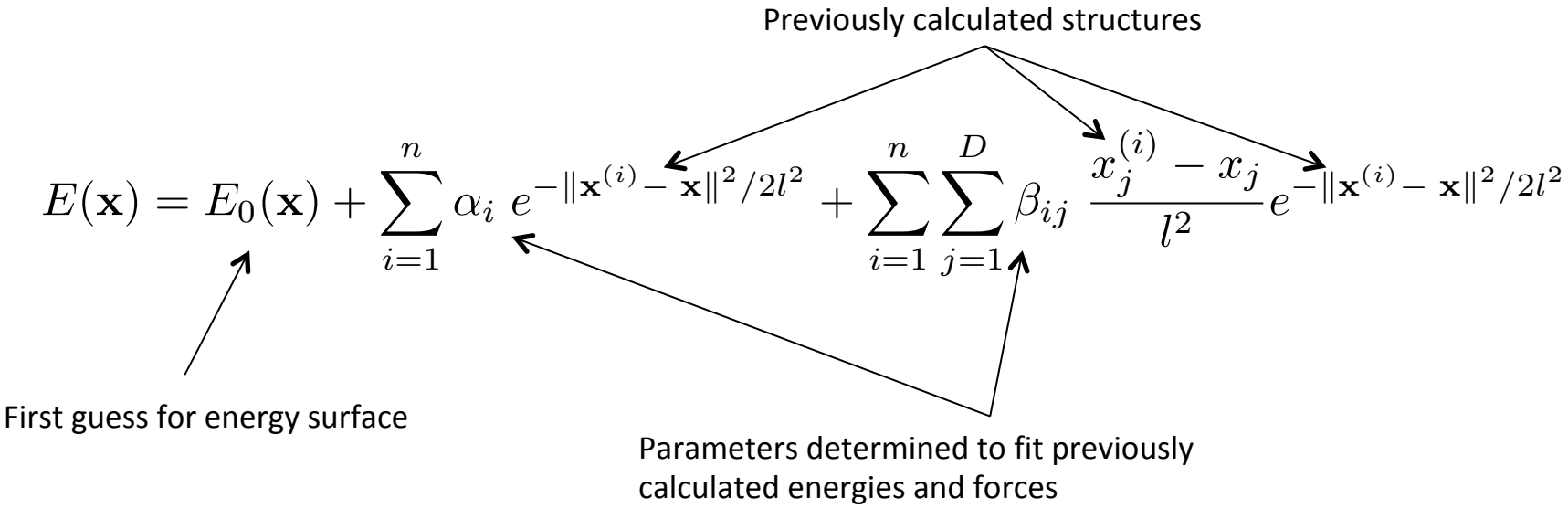
Use kernel regression on energy surface!

(Estefanía Garcia del Río, Jens Jørgen Mortensen, KWJ, arXiv:1808.08588 [physics.comp-ph])

Kernel regression with gradient information

Atomic coordinates: $\mathbf{x} = (x_1, x_2, \ldots, x_D)$

Prediction of potential energy surface:

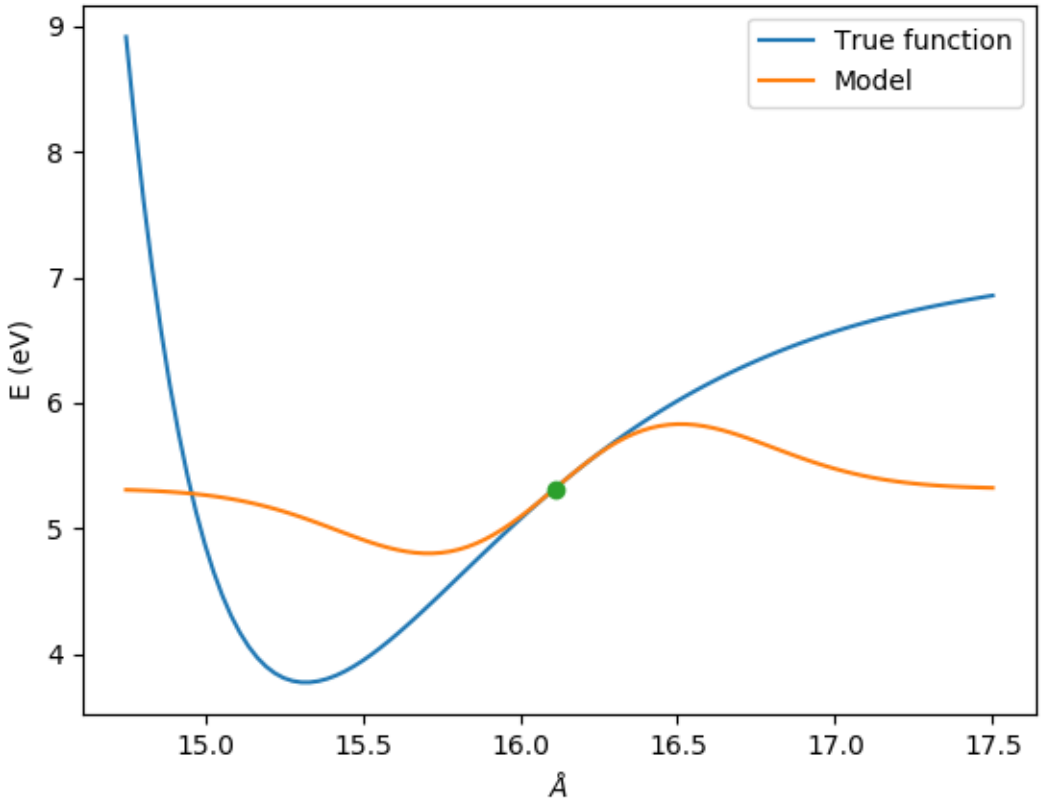


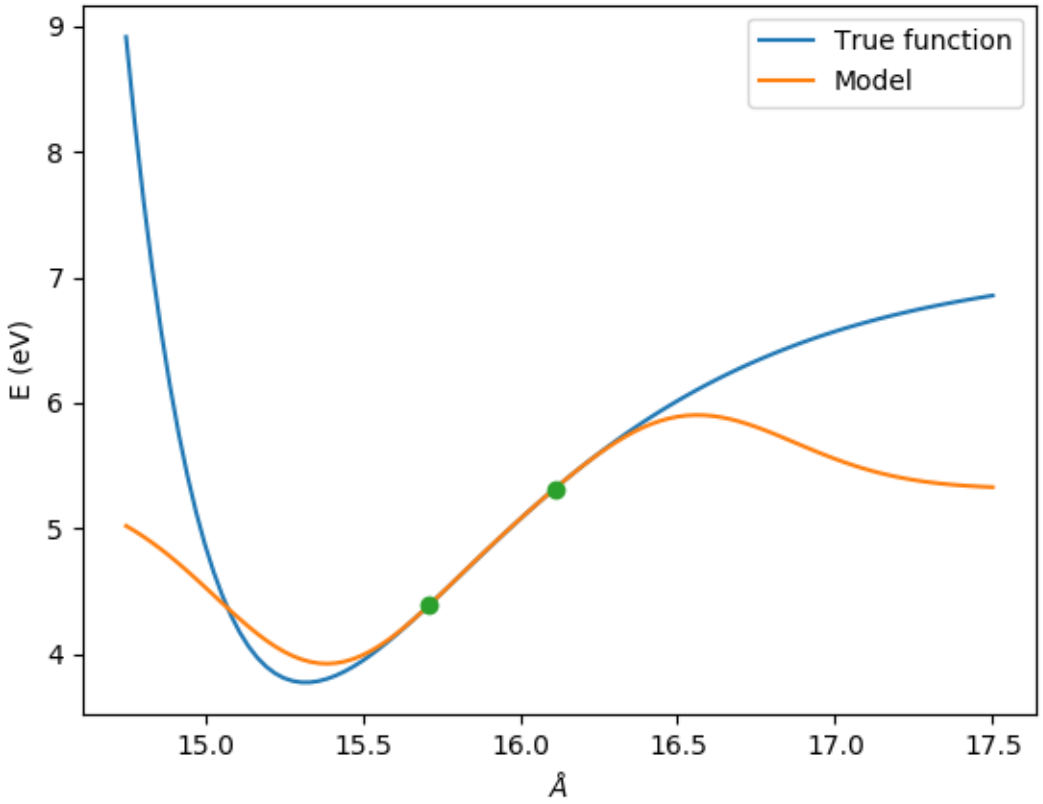
J. Wu, M. Poloczek, A. G. Wilson, and P. I. Frazier. arXiv:1703.04389

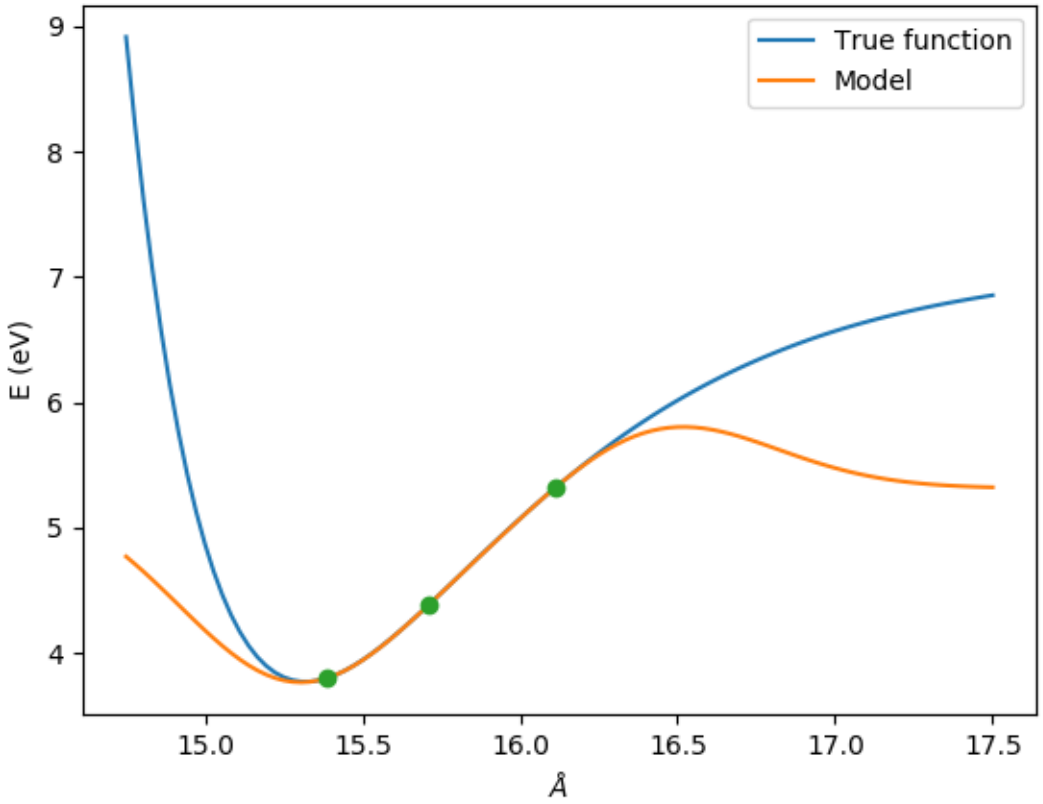
Applied to Nudged Elastic Band simulations:
O.-P. Koistinen, F. B. Dagbjartsdóttir, V. Ásgeirsson, A. Vehtari, and H. Jónsson, *J Chem Phys*, **147,** 152720 (2017)

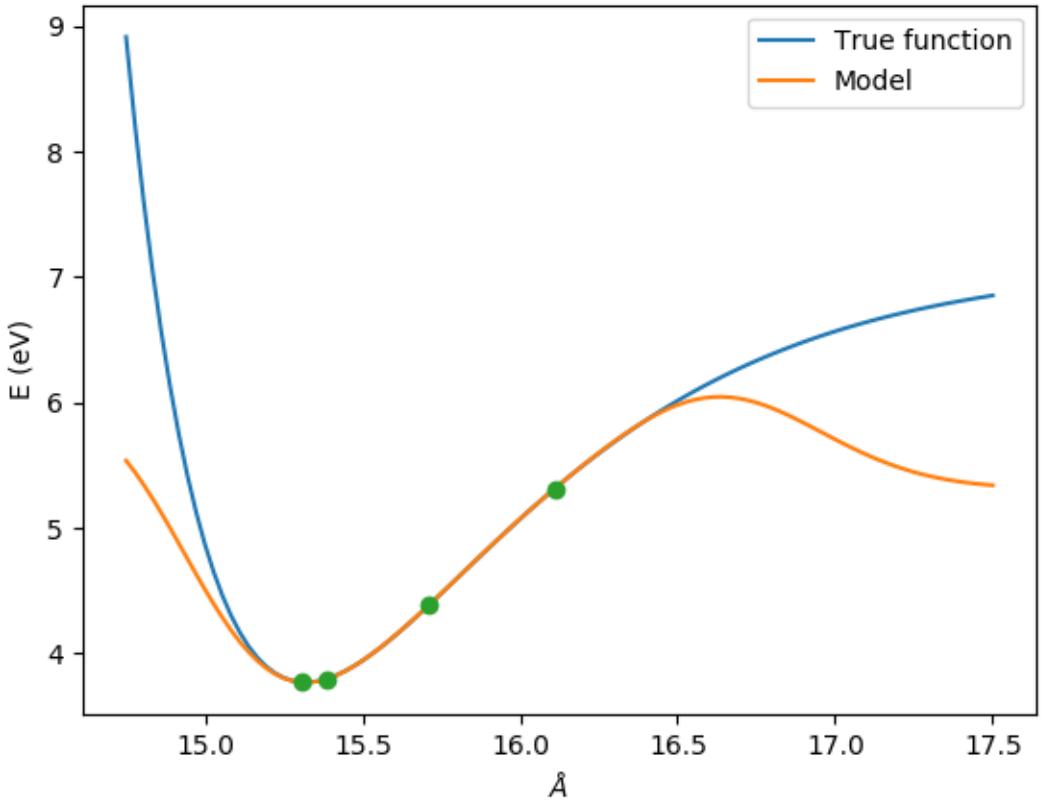
Optimization algorithm

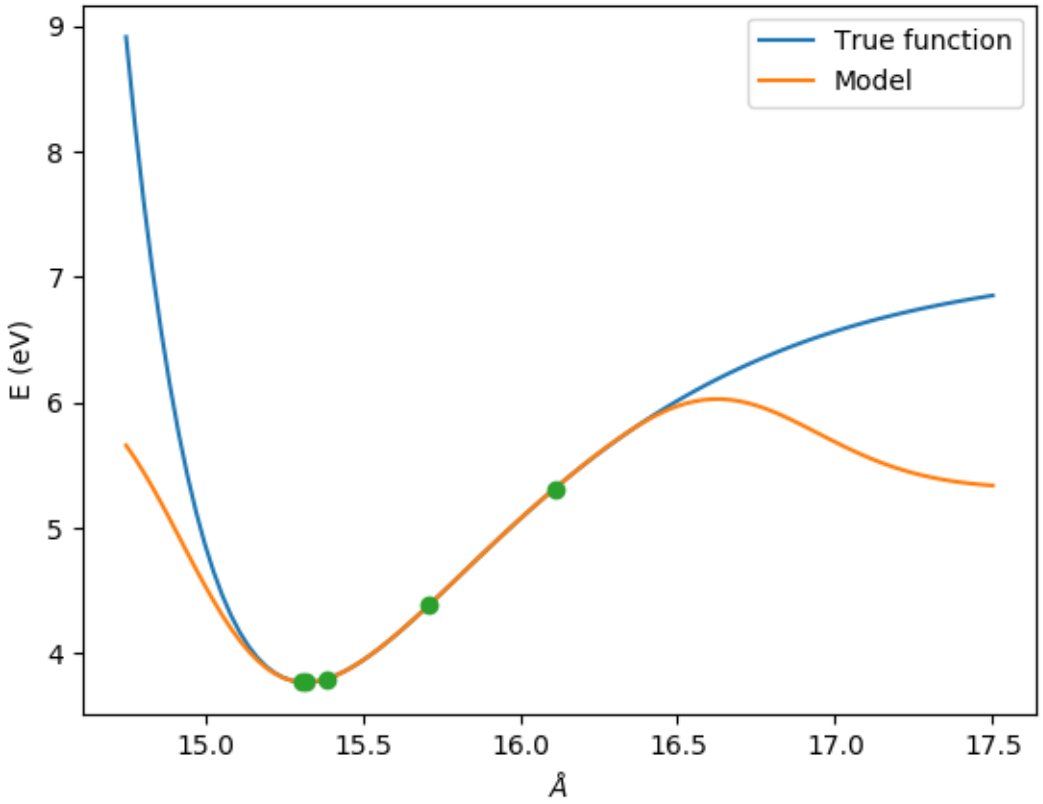
- Choose initial structure
- Predict energy surface based on previous structures
- Find minimum structure on predicted energy surface using BFGS (a standard optimizer)
- Calculate energy and forces at the predicted minimum point with DFT


True function
Model
E (eV)
9
8
7
6
5
4
15.0
15.5
16.0
16.5
17.0
17.5
Å


True function
Model
E (eV)
Å
9
8
7
6
5
4
15.0
15.5
16.0
16.5
17.0
17.5


True function
Model
E (eV)
9
8
7
6
5
4
15.0
15.5
16.0
16.5
17.0
17.5
Å

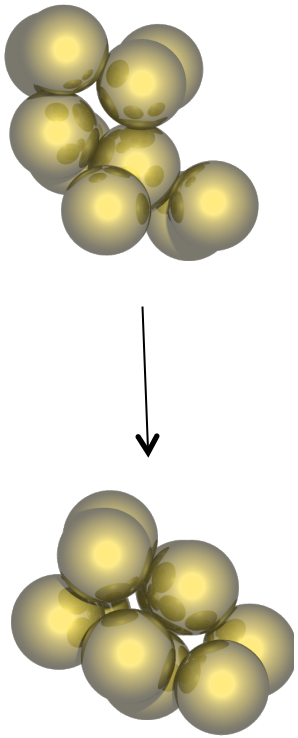

True function
Model
E (eV)
9
8
7
6
5
4
15.0
15.5
16.0
16.5
17.0
17.5
Å

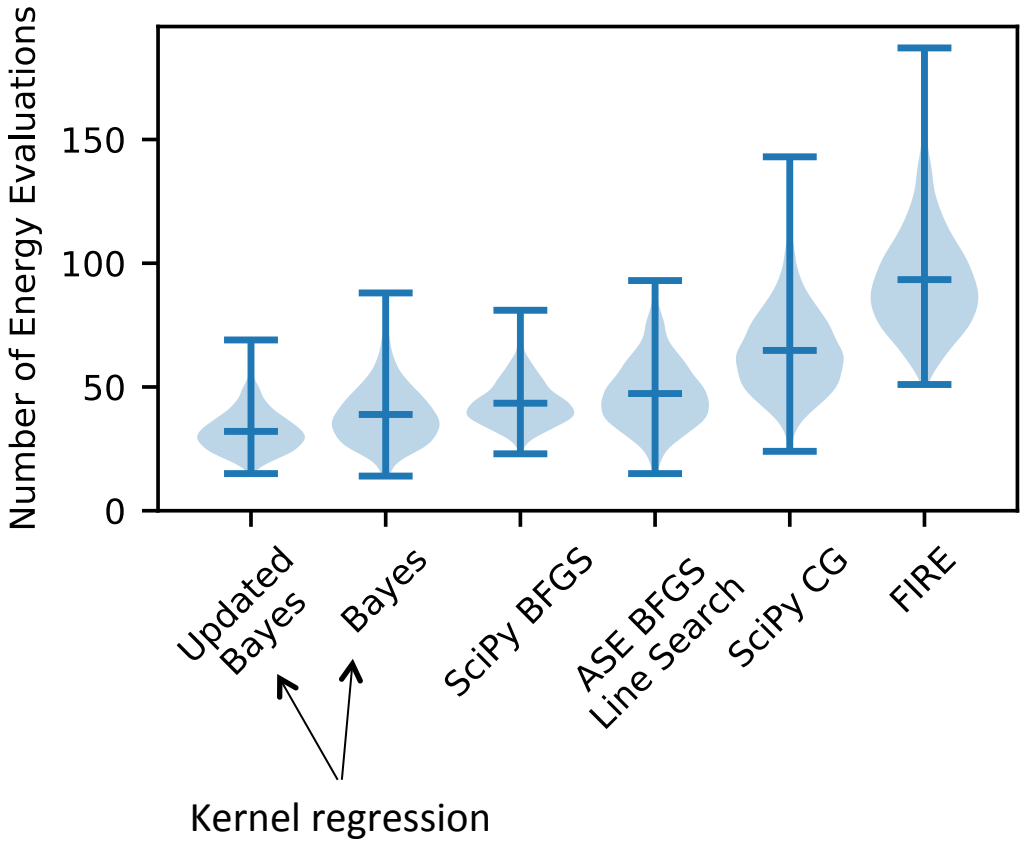

True function
Model
E (eV)
9
8
7
6
5
4
15.0
15.5
16.0
16.5
17.0
17.5
Å

Test case: 10 atom Au cluster

10 atom Au cluster with Effective Medium Theory interatomic potential.

1000 energy minimizations with different initial conditions.

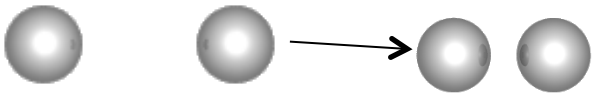




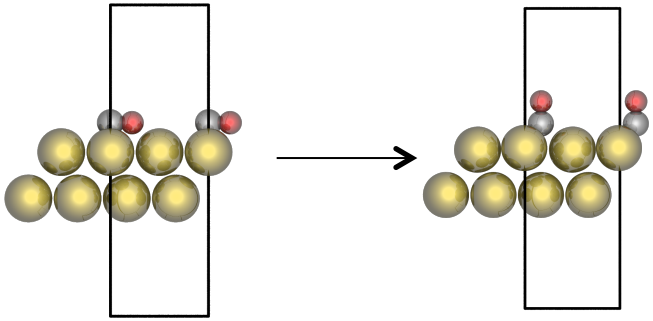
Comparison with other optimizers in the Atomic Simulation Environment (ASE).

“Standard” test systems from ASE

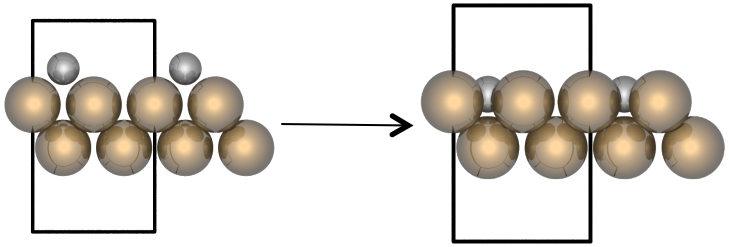
H_2



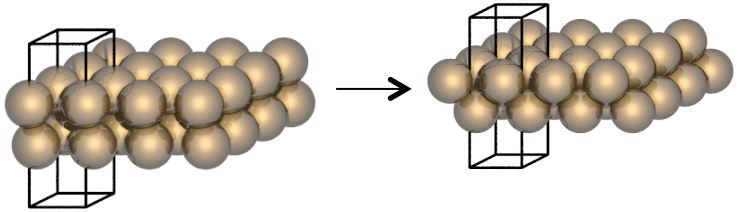
CO/Ag



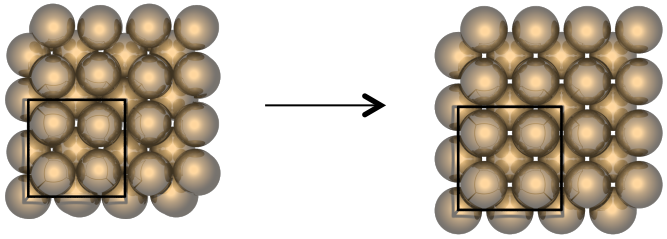
C/Cu



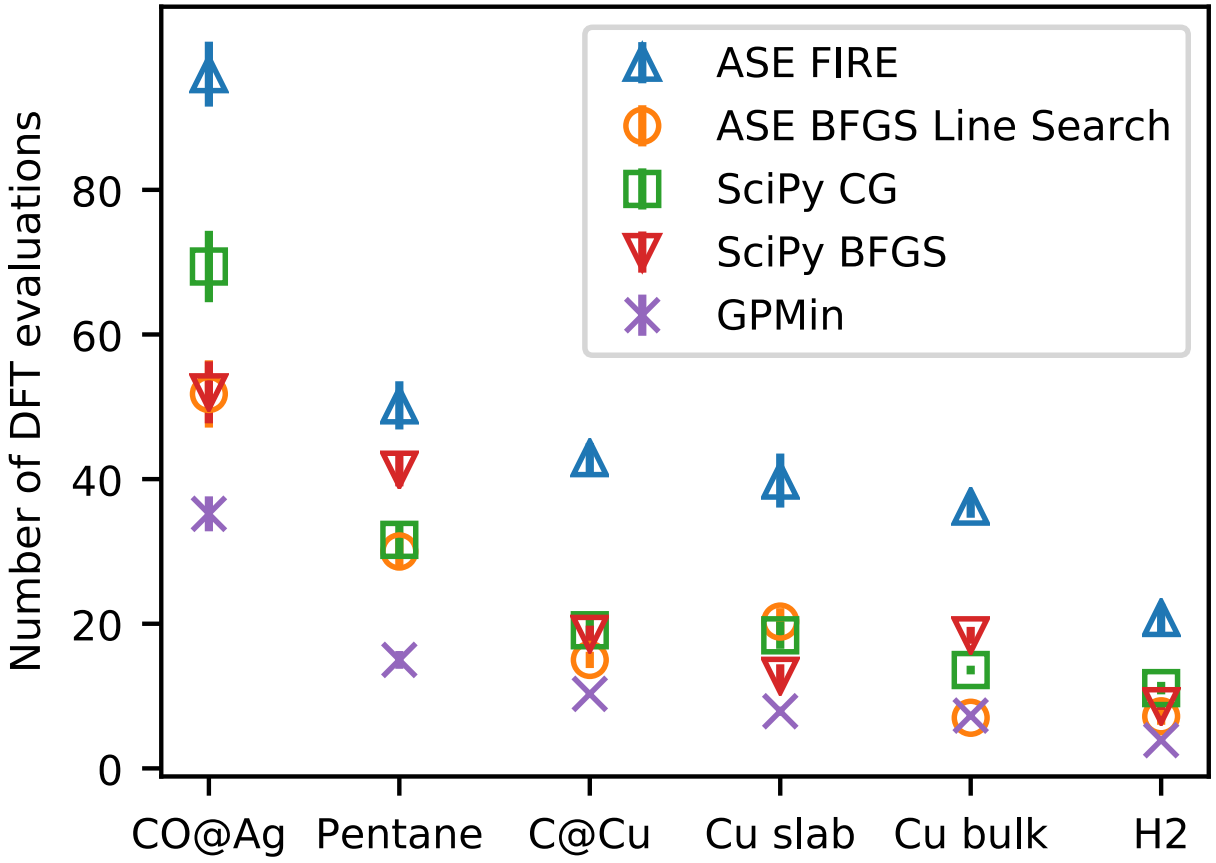
Cu slab



Cu bulk



Test results for optimizers in ASE with standard settings



10 minimizations for each system

GPAW calculations using LCAO basis set

Available in ASE now!

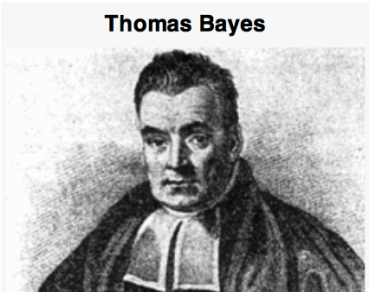
Bayes' theorem

Probability theory: $P(A, B) = P(A|B)P(B) = P(B|A)P(A)$

Bayes theorem: $P(A|B) = \frac{1}{P(B)} P(B|A)P(A)$

Inference:

$$P(\text{cause}|\text{effect}) = \frac{1}{P(\text{effect})} P(\text{effect}|\text{cause})P(\text{cause})$$



(1701-1761)

Laplace
Turing

Example: Disease (D) and test (T)

1/1000 of the population has the disease: $P(D) = 0.001, \ P(\neg D) = 0.999$

The test is 99% good: $P(T|D) = P(\neg T|\neg D) = 0.99$

$$P(T|\neg D) = P(\neg T|D) = 0.01$$

You get a positive test. What is the risk you have the disease P(D|T)?

1%
10%
50%
90%
99%

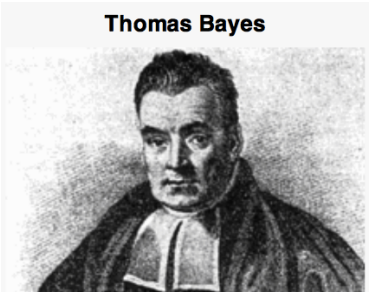
Bayes' theorem

Probability theory: $P(A,B) = P(A|B)P(B) = P(B|A)P(A)$

Bayes theorem: $P(A|B) = \frac{1}{P(B)} P(B|A)P(A)$

Inference:

$$P(\text{cause}|\text{effect}) = \frac{1}{P(\text{effect})} P(\text{effect}|\text{cause})P(\text{cause})$$



(1701-1761)

Laplace
Turing

Example: Disease (D) and test (T)

1/1000 of the population has the disease: $P(D) = 0.001, \ P(\neg D) = 0.999$

The test is 99% good: $P(T|D) = P(\neg T|\neg D) = 0.99$

$$P(T|\neg D) = P(\neg T|D) = 0.01$$

You get a positive test. What is the risk you have the disease P(D|T)?

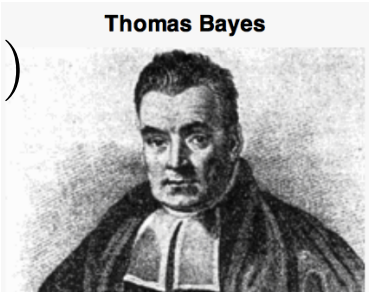
$$P(D|T) = \frac{P(T|D)P(D)}{P(T)} = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\neg D)P(\neg D)}$$

$$= \frac{0.99 * 0.001}{0.99 * 0.001 + 0.01 * 0.999} = 0.09 = 9\%$$

Bayesian inference

Probability theory: $P(A, B) = P(A|B)P(B) = P(B|A)P(A)$

Bayes theorem: $$P(A|B) = \frac{1}{P(B)} P(B|A)P(A)$$



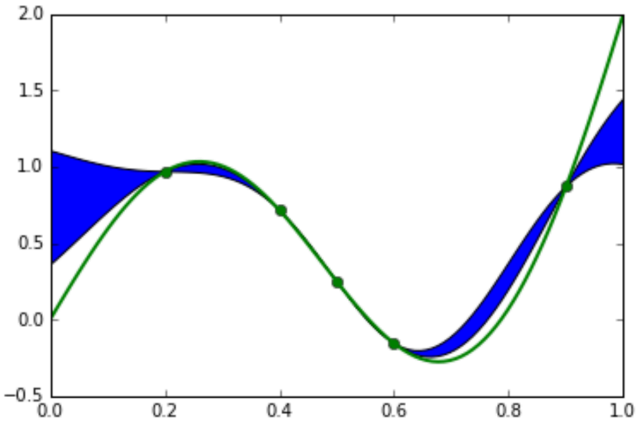
(1701-1761)

Laplace
Turing

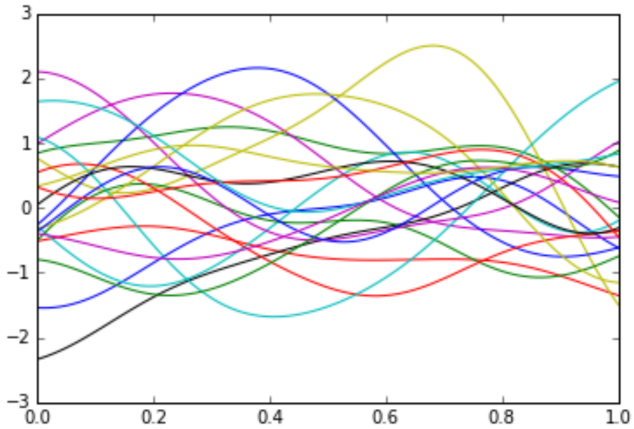
$$P(\text{Model}|\text{Data}) = \frac{1}{P(\text{Data})} P(\text{Data}|\text{Model})P(\text{Model})$$

Likelihood

Prior probability



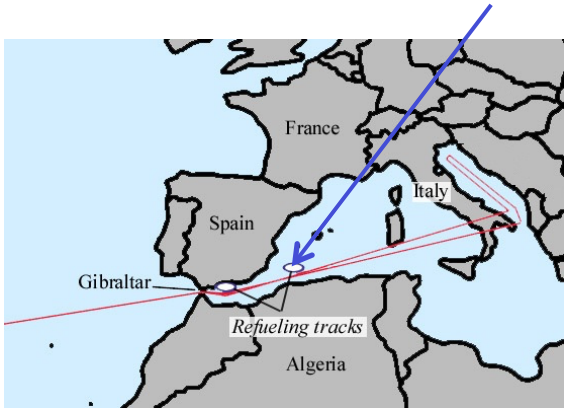
Update of model because of data



Bayesian Search Theory in Practice
The 1966 Palomares B-52 crash

B-52G collided with KC-135 tanker when fueling

Refueling



4 H-bombs dropped
3 on land
1 in the Mediterranean Sea

Bayesian search theory applied:
Assign probabilities to different areas of the sea
based on available information
(a local fisherman saw the bomb dropping)
Update your probability depending on your search.

Bomb recovered

Example: heads and tails

Model: Probability p for heads

Data: h heads and t tails

$$P(\text{Model}|\text{Data}) = \frac{1}{P(\text{Data})} P(\text{Data}|\text{Model}) P(\text{Model})$$

$$P(p|h,t) \propto P(h,t|p)P(p) = \frac{(h+t+1)!}{h!t!} p^h (1-p)^t$$

Average value of *p*: $\langle p \rangle = \frac{h+1}{(h+1)+(t+1)}$

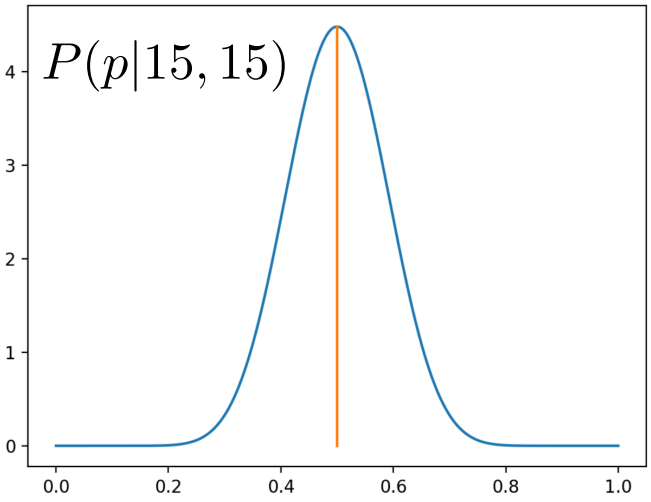
Binomial distribution of *h* and *t* given parameter *p*

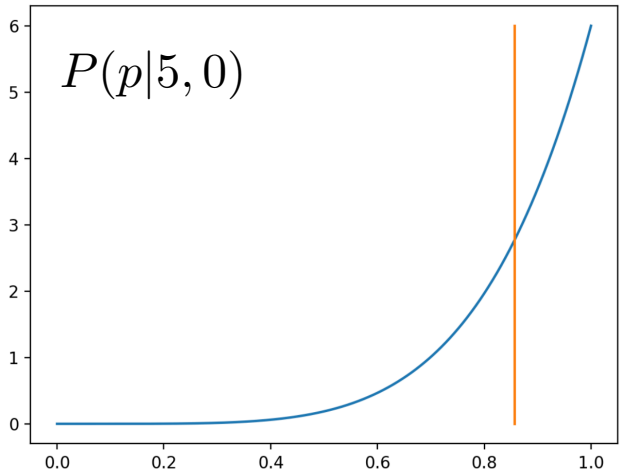
Conjugate distribution: Beta distribution of *p* given parameters *h* and *t*

$h = 15$

$t = 15$

$\langle p \rangle = \frac{1}{2}$





$h = 5$

$t = 0$

$\langle p \rangle = \frac{6}{7}$

From Bayes to least squares regression

$$P(\text{Model}|\text{Data}) = \frac{1}{P(\text{Data})} P(\text{Data}|\text{Model}) P_0(\text{Model})$$

Data: $\{y_i\}$

Model parameters: a

Predictions by model: $y_i(a)$

Likelihood with Gaussian noise in the data: $P(\{y_i\}|a) \propto \exp\left(-\sum_i (y_i - y_i(a))^2/2\sigma^2\right)$

Prior: $P_0(a) \propto \exp\left(-R(a)\right) \propto \exp\left(-\lambda|a|^{1 \text{ or } 2}\right)$

Posterior: $P(a|\{y_i\}) \propto \exp\left(-\sum_i (y_i - y_i(a))^2/2\sigma^2 - R(a)\right) = \exp(-C(a))$

Maximize posterior probability -> minimize cost function

Regularization

$$C(a) = \sum_i (y_i - y_i(a))^2/2\sigma^2 + R(a)$$

Least squares regression!

Gaussian Processes

Consider just two data points: y_1, y_2

Prior probability:

The two points are Gaussian distributed with zero mean and some correlation between them:

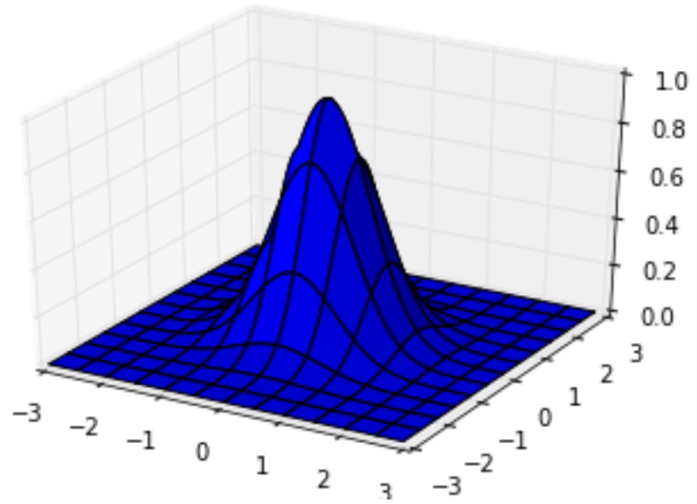
$$\langle y_1^2 \rangle = \sigma^2, \ \langle y_2^2 \rangle = \sigma^2, \ \langle y_1 y_2 \rangle = \tau^2 \ or \ \langle \mathbf{y}\mathbf{y}^{\mathbf{T}} \rangle = \begin{pmatrix} \sigma^2 & \tau^2 \\ \tau^2 & \sigma^2 \end{pmatrix} \equiv \mathbf{K}$$

This corresponds to the probability distribution:

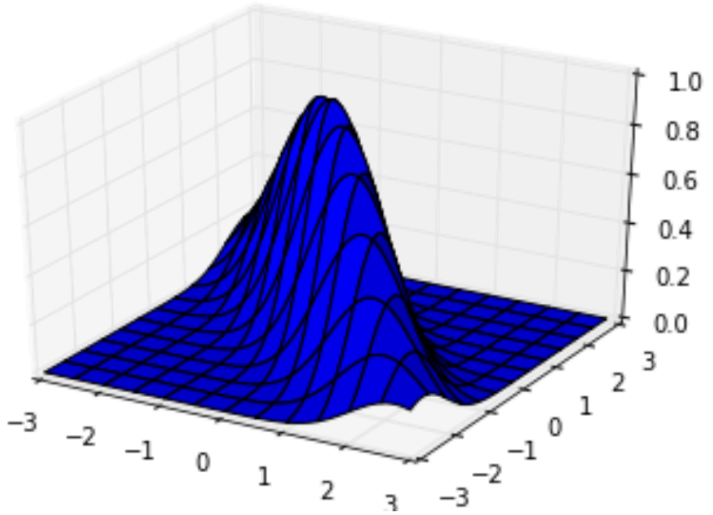
$$P_0(y_1, y_2) = \frac{1}{\sqrt{2\pi det(\mathbf{K})}} \exp\left(-\frac{1}{2}\mathbf{y}^{\mathbf{T}}\mathbf{K}^{-\mathbf{1}}\mathbf{y}\right)$$

$P(y_1, y_2)$

No correlation



Strong correlation



Gaussian Processes

Now we get the information that y_1 actually has the value y_1^0

What is then the probability distribution for y_2?

$$P(y_2) \propto \int dy_1 \delta(y_1 - y_1^0) P_0(y_1, y_2) = P_0(y_1^0, y_2) \propto \exp\left[-\frac{1}{2(\sigma^2(1-(\tau/\sigma)^4))} \left(y_2 - \left(\frac{\tau}{\sigma}\right)^2 y_1^0 \right)^2 \right]$$

A new Gaussian!

Mean:

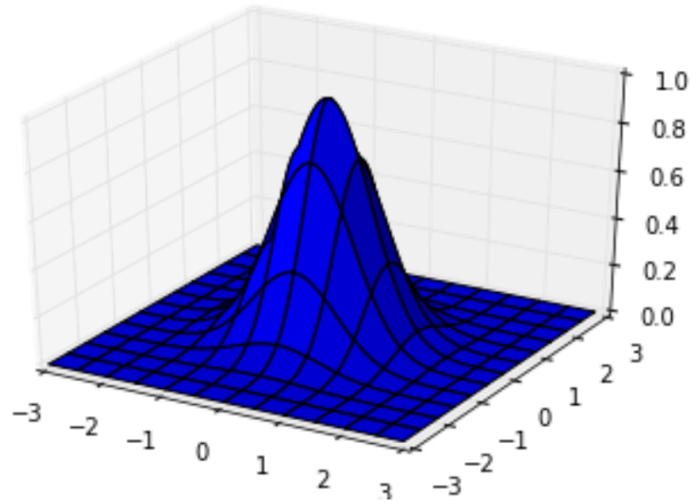
$$\langle y_2 \rangle = \left(\frac{\tau}{\sigma}\right)^2 y_1^0$$

Width:

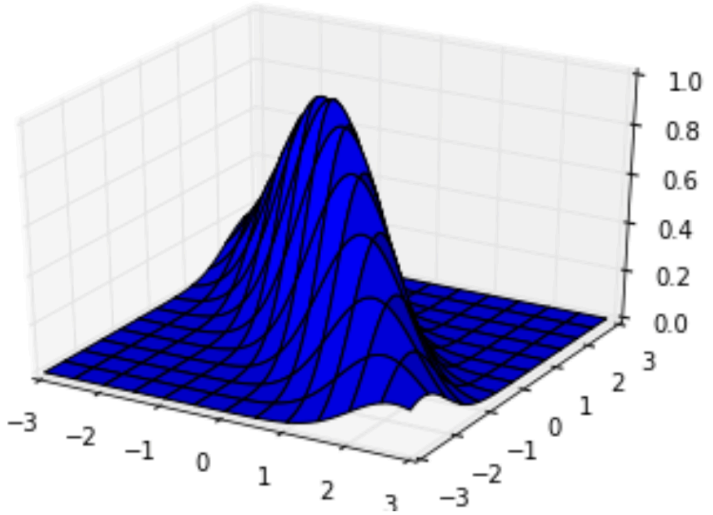
$$\langle (y_2 - \langle y_2 \rangle)^2 \rangle = \sigma^2(1 - (\tau/\sigma)^4)$$

$P(y_1, y_2)$

No correlation



Strong correlation



Gaussian Processes

Now consider N data points: $\mathbf{y}^T = (y_1, y_2, \ldots, y_N)$

Prior probability: Gaussian with zero mean

$$\langle \mathbf{y} \rangle = 0, \;\; \langle \mathbf{y}\mathbf{y}^T \rangle \equiv \mathbf{K} \quad i.e. \quad P_0(\mathbf{y}) = \frac{1}{\sqrt{2\pi det(\mathbf{K})}} \exp\left(-\frac{1}{2}\mathbf{y^T K^{-1} y}\right)$$

Now we come with a new data point $y_{N+1}, \quad \mathbf{y}_{N+1}^T = (y_1, y_2, \ldots, y_N, y_{N+1})$

And again we assume a Gaussian distribution: $\langle \mathbf{y_{N+1}\ y_{N+1}^T} \rangle = \begin{pmatrix} \mathbf{K} & \mathbf{k} \\ \mathbf{k} & c \end{pmatrix}$

Given that the first N data points take the values $\mathbf{y}^0$ we can now update our probability distribution and determine, that the last data point is Gaussian distributed with

$$\langle y_{N+1} \rangle = \mathbf{k}^T \mathbf{K}^{-1} \mathbf{y}^0 \qquad \langle (y_{N+1} - \langle y_{N+1} \rangle)^2 \rangle = c - \mathbf{k}^T \mathbf{K}^{-1} \mathbf{k}$$

The same formula for the mean as the simple fitting with superposed Gaussians!
New meaning to the kernel K: A measure of correlations.
Data which are “similar” or “close” have high correlation.

Fitting a function
The prior distribution

Functions on the interval [0,1]:

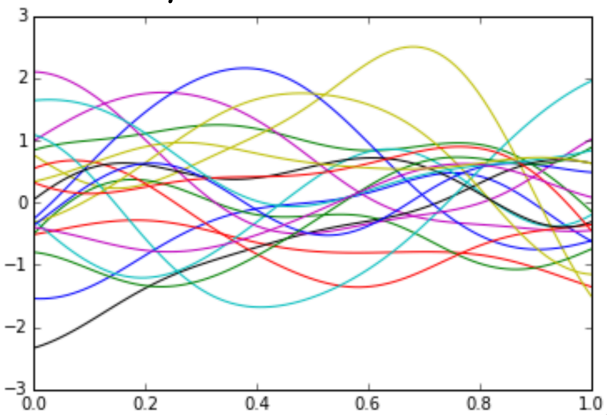
Kernel function

$$K_{ij} = \langle y(x_i)y(x_j)\rangle = k(x_i, x_j) = \exp(-|x_i - x_j|^2/2\rho^2)$$

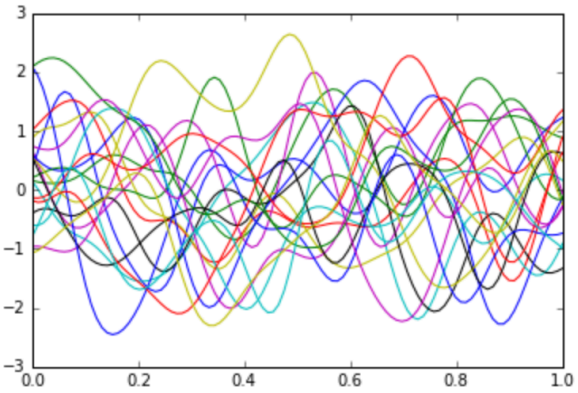
Prior distribution:

$$P_0(\mathbf{y}) = \frac{1}{\sqrt{2\pi det(\mathbf{K})}} \exp\left(-\frac{1}{2}\mathbf{y^T K^{-1} y}\right), \ \mathbf{y}^T = (y(x_1), y(x_2), \ldots, y(x_N))$$

$\rho^2 = 0.1$



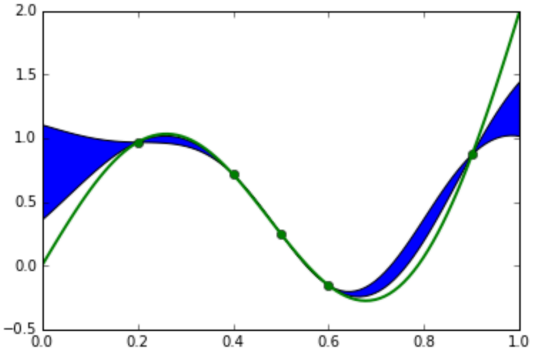
$\rho^2 = 0.01$



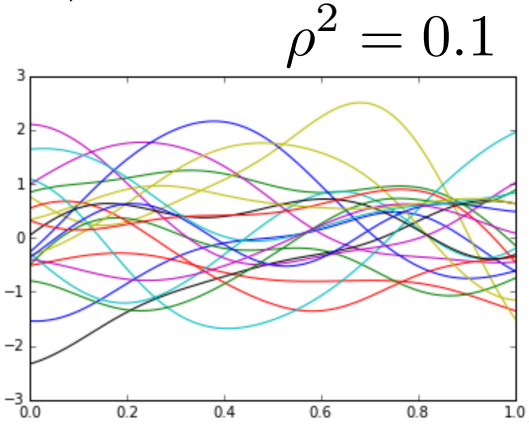
Fitting a function

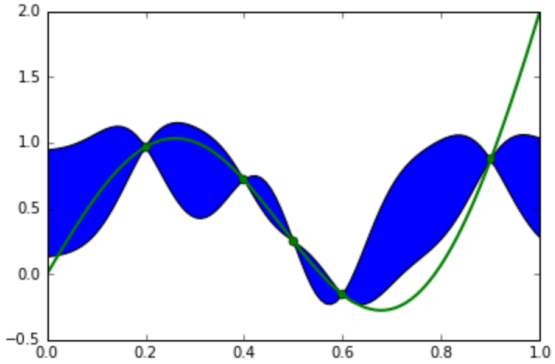
Fitting a function f(x) based on data points $y_i=f(x_i)$

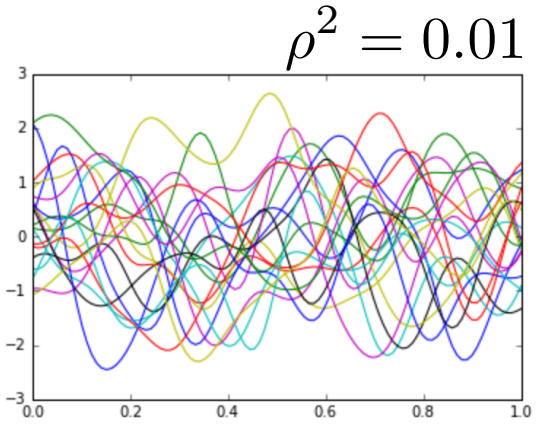
Kernel: $k(x_i, x_j) = \exp(-|x_i - x_j|^2/2\rho^2)$



Update of model because of data



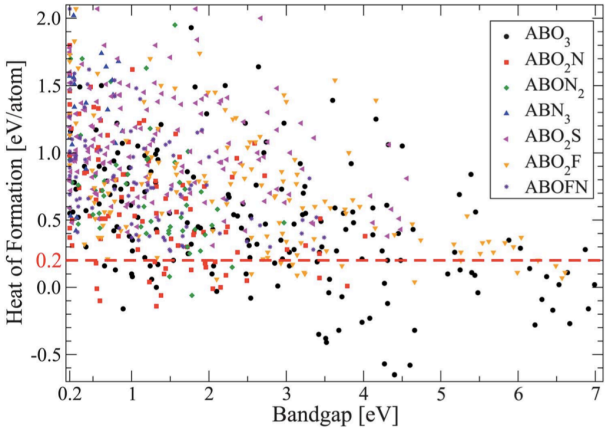




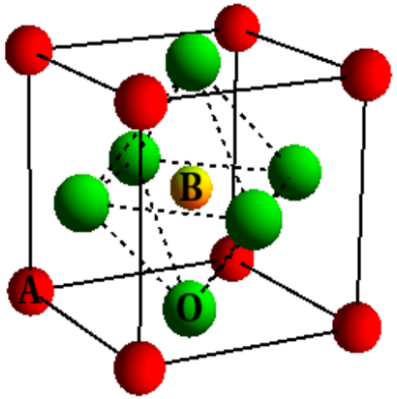
The value of ρ can be addressed by so-called cross validation

Back to water splitting with machine learning

About 19000 cubic perovskites oxides, oxynitrides, oxysulfides, oxyfluorides, oxyfluornitrides



ABO_3, $ABON_2$, ..



Fingerprint (x-vector):

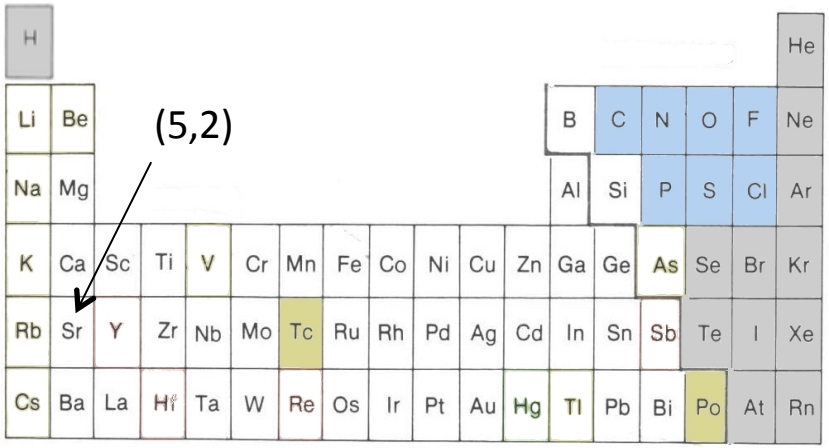
$$x(\mathrm{SrTaO_2N}) = (5, 2, 6, 5, 2, 1, 0, 0)$$

O, N, S, F

Sr "coordinates"

Kernel function:

$$k(x_i, x_j) = \exp(-|x_i - x_j|^2/2\rho^2)$$



Water splitting
Gaussian process

Training on 500 perovskites (~2.6 % of the total dataset).

Kernel: $K_{ij} = k(x_i, x_j) = \exp(-\|x_i - x_j\|^2/2)$

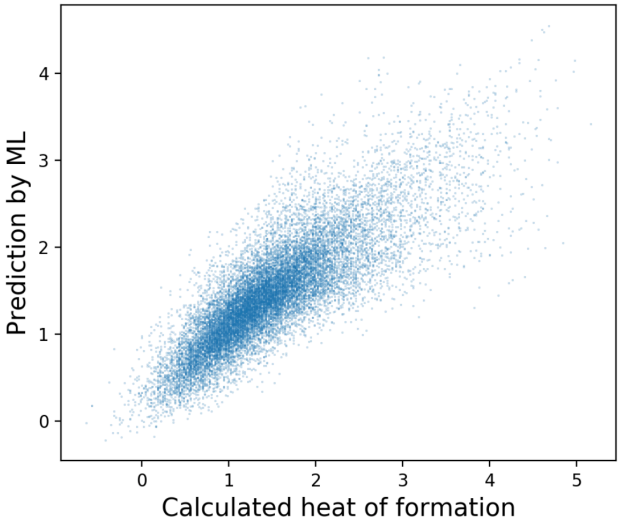
Prediction: $y(x) = \mathbf{k}^T \mathbf{K}^{-1} \mathbf{y}$ with $k_i = k(x, x_i)$

Only determined by "metric" (not by data)

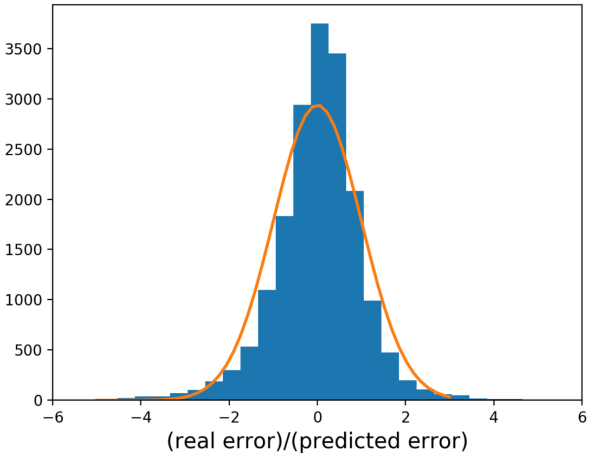
Data

+ error prediction

Example: Heat of formation



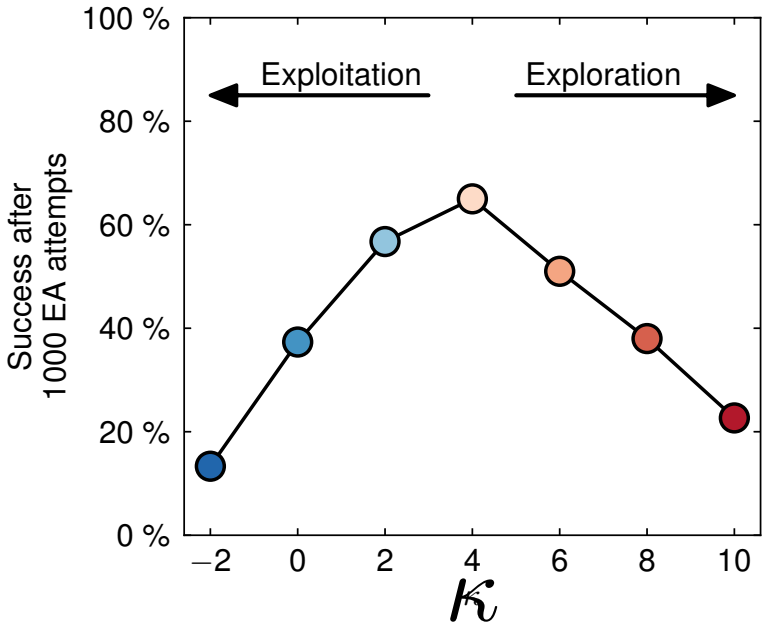
Mean Absolute Error: 0.28 eV
Mean Absolute Predicted Error: 0.38 eV

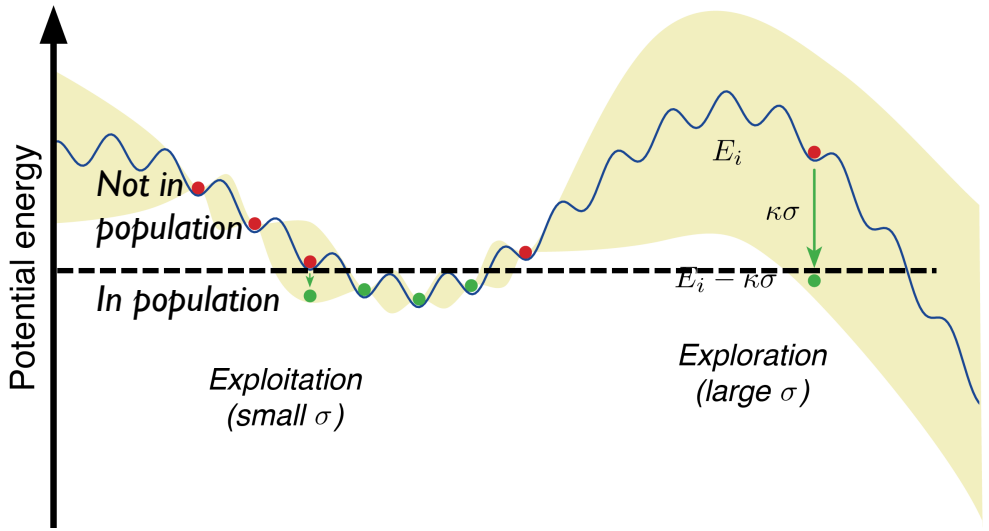


Global optimization using Bayesian estimation

Acquisition function:

$$A_i = E_i - \kappa\sigma_i$$





M. S. Jørgensen, U. F. Larsen, K. W. Jacobsen, and B. Hammer, The Journal of Physical Chemistry A 122, 1504 (2018).

Bayesian Error Estimation Functionals (BEEF)

- Density Functional Theory
  - Predictive power but many different approximations to the xc-functional.
  - Reliability evaluated based on experience from previous investigations or maybe apply several different functionals. Need for systematic approach.
  - The reliability depends on both functional and investigated property.
    - PBEsol better than RPBE for lattice constants
    - RPBE better than PBEsol for chemisorption energies

Bayesian Error Estimation Functionals (BEEF)

- Exchange-correlation functionals fitted to experimental and high-quality computational data
- Different levels: BEEF(GGA), BEEFvdW (GGA+vdW), mBEEF, mBEEF-vdW
- Provides probability distribution of functionals – ensemble
- Error bars obtained from ensemble

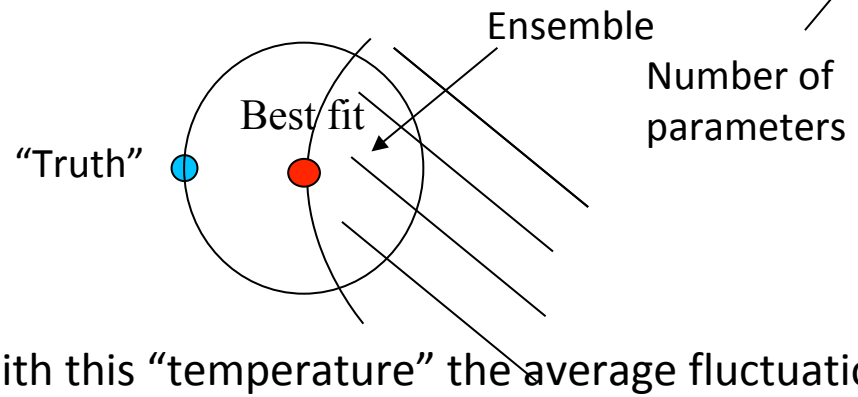
Approach for insufficient models i.e. models which cannot fit the data points

$$P(a|D) \propto \exp(-C(a)/T), \quad C(a) = \sum_n (y_n - y_n(a))^2$$

Effective temperature given by the best-fit cost function:

$$T = \frac{2C_0}{N_p}$$

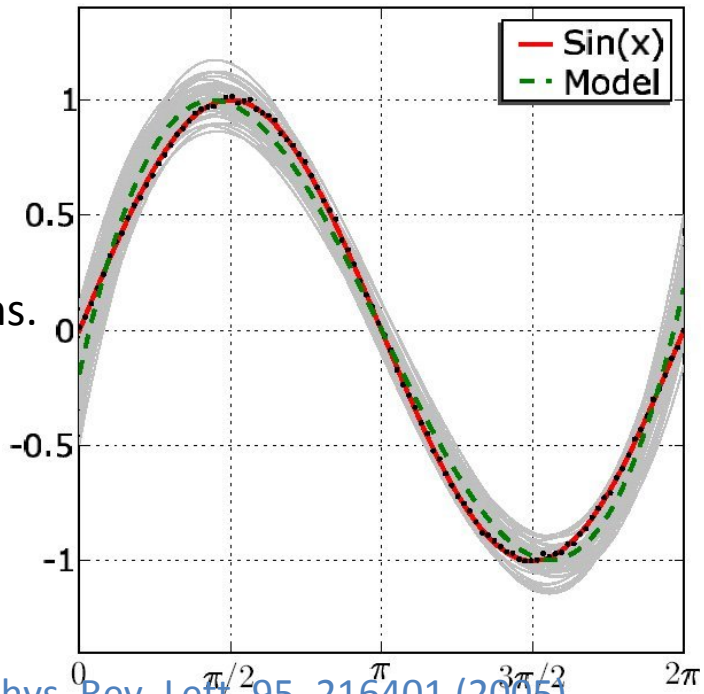
Number of parameters



With this "temperature" the average fluctuations for predicted points in the database equal the actual deviations. Derivable from max-entropy principle.

$$\left\langle \sum_n \delta y_n^2 \right\rangle = \sum_n (y_n - y_n(a_{\text{best-fit}}))^2$$

Insufficient model: 3rd order polynomial fit to a sine function



Frederiksen, Jacobsen, Brown, Sethna PRL 93, 165501 (2004)
Mortensen, Kaasbjerg, Frederiksen, Nørskov, Sethna, Jacobsen, Phys. Rev. Lett. 95, 216401 (2005).

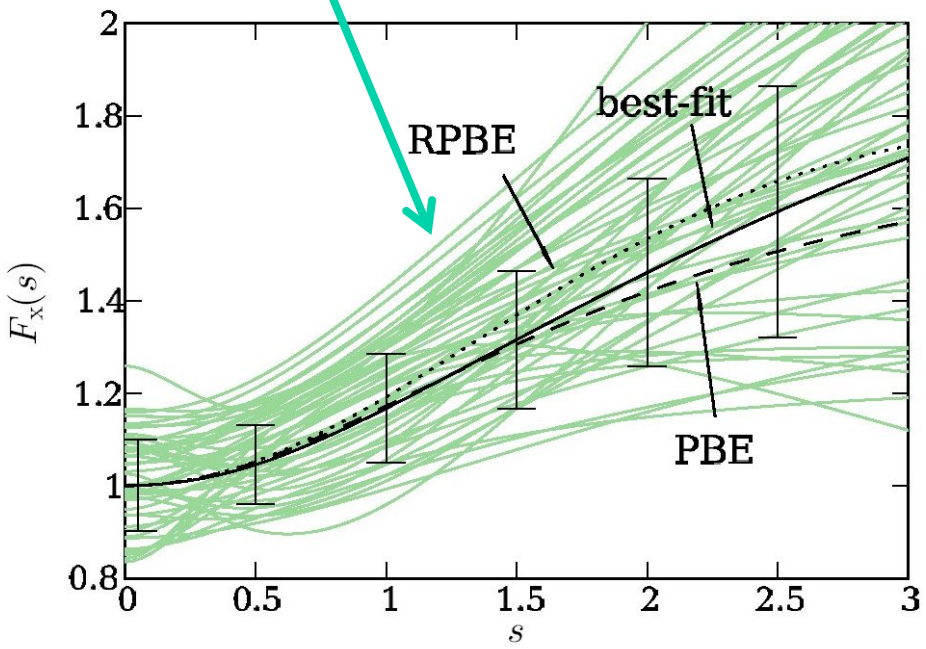
GGA: ensemble of enhancement factors

$$E_x[n] = \int d\mathbf{r}\; n(\mathbf{r})\epsilon_x^{\mathrm{unif}}(n)F_x(s)$$

Fitted to molecules and solids

Error estimation:

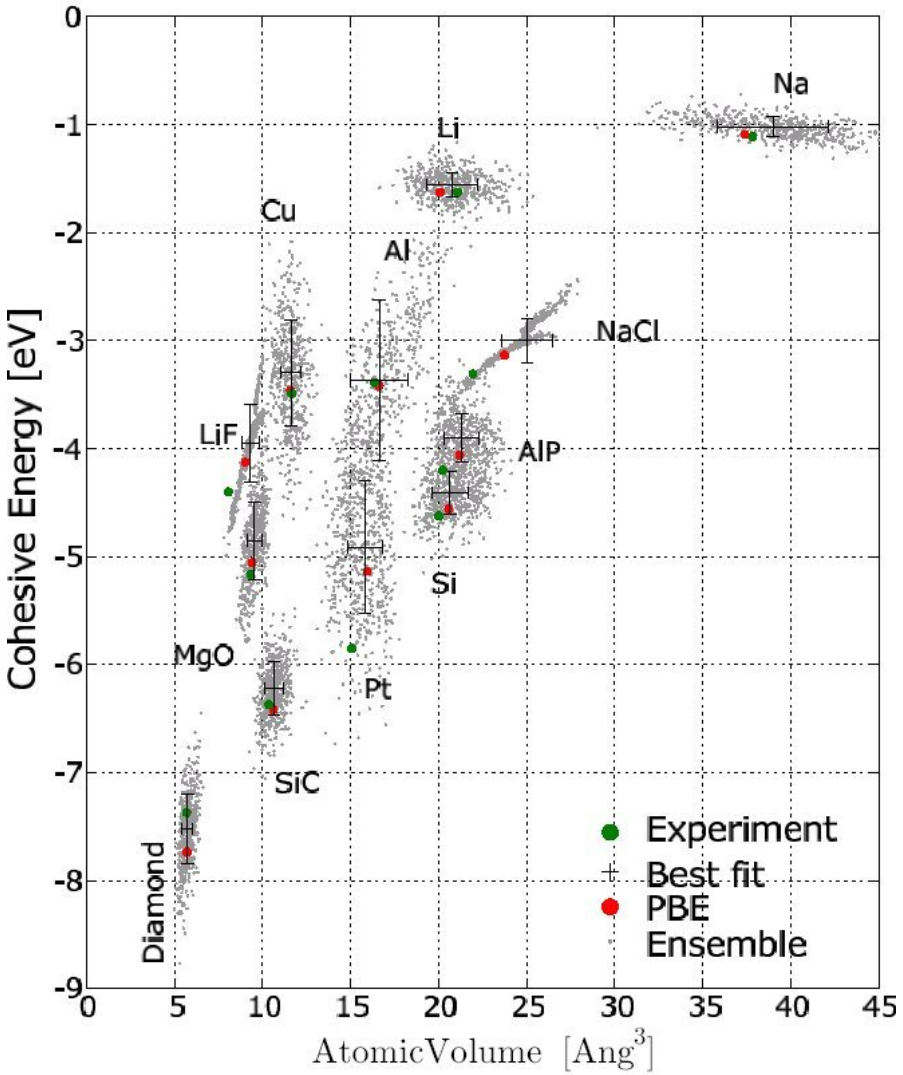
$$\sigma_{\mathrm{BEE}}(O) = \sqrt{\frac{1}{N}\sum_{\mu=1}^{N}\left(O(a^\mu) - O_{\text{best-fit}}\right)^2}$$

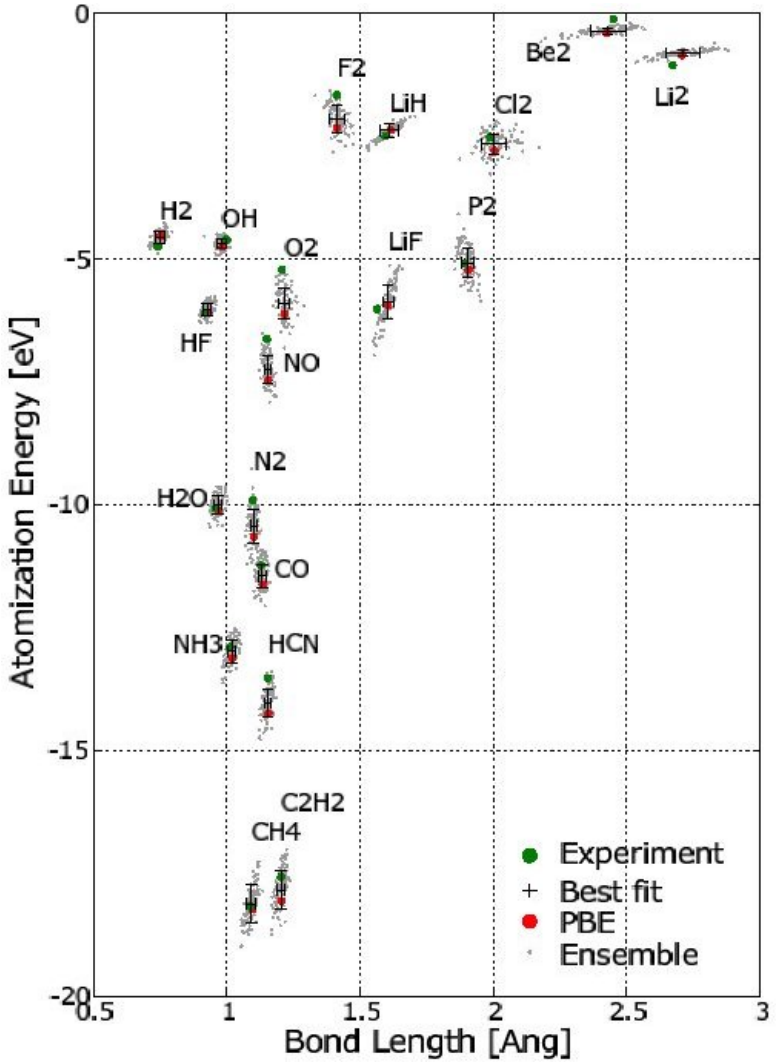


Essentially no additional computational cost (no self-consistency)

Mortensen, Kaasbjerg, Frederiksen, Nørskov, Sethna, Jacobsen, Phys. Rev. Lett. 95, 216401 (2005).

Error estimation





Mortensen, Kaasbjerg, Frederiksen, Nørskov, Sethna, Jacobsen, Phys. Rev. Lett. 95, 216401 (2005).

Figure courtesy of Kristen Kaasbjerg

Stepping up the ladder: Larger databases and more functional forms

BEEF-vdW

$$E_{xc} = \sum_m a_m E_m^{GGA-x} + \alpha_c E^{LDA-c} + (1-\alpha_c) E^{PBE-c} + E_{vdW-DF2}^{nl-c}$$

Expansion of enhancement factor (density gradient)

mBEEF:

$$E_{xc} = \sum_{m,n} a_{mn} E_{mn}^{mGGA-x} + E^{PBEsol-c}$$

Expansion of enhancement factor (density gradient, kinetic energy density)

mBEEF-vdW

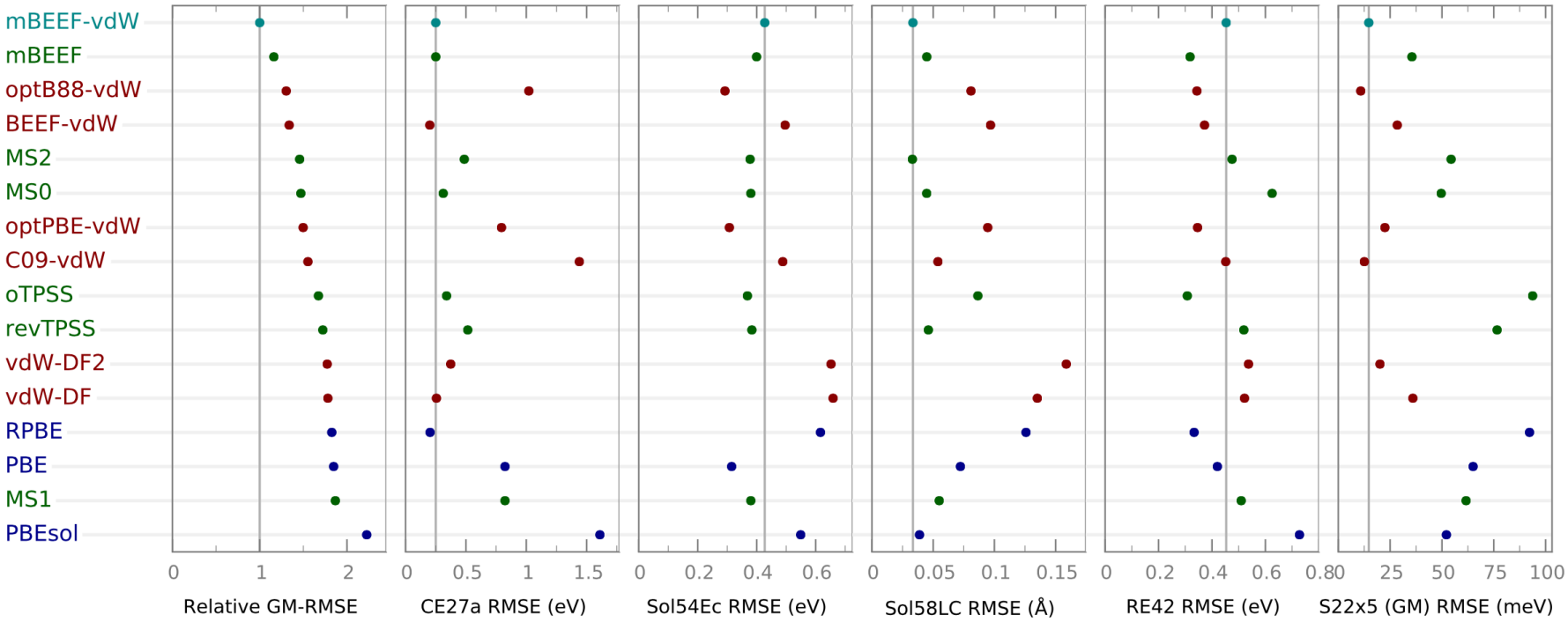
$$E_{xc} = \sum_{m,n} a_{mn} E_{mn}^{mGGA-x} + \alpha_c E^{LDA-c} + \alpha_{PBEsol} E^{PBE-c} + \alpha_{nl-c} E_{vdW-DF2}^{nl-c}$$

BEEF-vdW:
Wellendorff, Lundgaard, Møgelhøj, Petzold, Landis, Nørskov, Bligaard, Jacobsen, Phys. Rev. B **85**, 235149 (2012).

mBEEF:
Wellendorf, Lundgaard, Jacobsen, Bligaard, J. Chem. Phys. **140**, 144107 (2014)

mBEEF-vdW:
K. T. Lundgaard, J. Wellendorff, J. Voss, K. W. Jacobsen, and T. Bligaard, *Phys. Rev. B*. **93**, 235162 (2016).
Petzold, Bligaard, and Jacobsen, Top Catal 55, 402 (2012).

Functional comparison on mBEEF-vdW **training** sets



Overall geometric mean	Chemisorption energies on TM surfaces	Solid cohesive energies	Solid lattice constants	Gas phase reaction energies	Non-covalent bonding

(mBEEF-vdW: K. T. Lundgaard, J. Wellendorff, J. Voss, K. W. Jacobsen, and T. Bligaard, *Phys. Rev. B*. **93**, 235162 (2016).)

Examples of error estimation

- Copper cohesive energy vs structural energy difference
  - Energy errors can be very different
- The CO puzzle
  - When error bars show we don't know
- Compound formation energies
  - Comparing models
- Water splitting ABS_3 compounds
  - Error bars in practice
- Ammonia synthesis
  - When error bars show we know more than one could expect

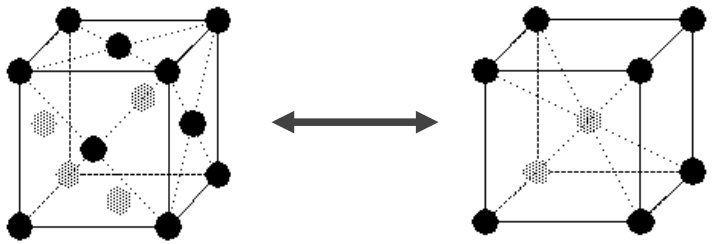
Fcc-bcc structural energy difference

Cu

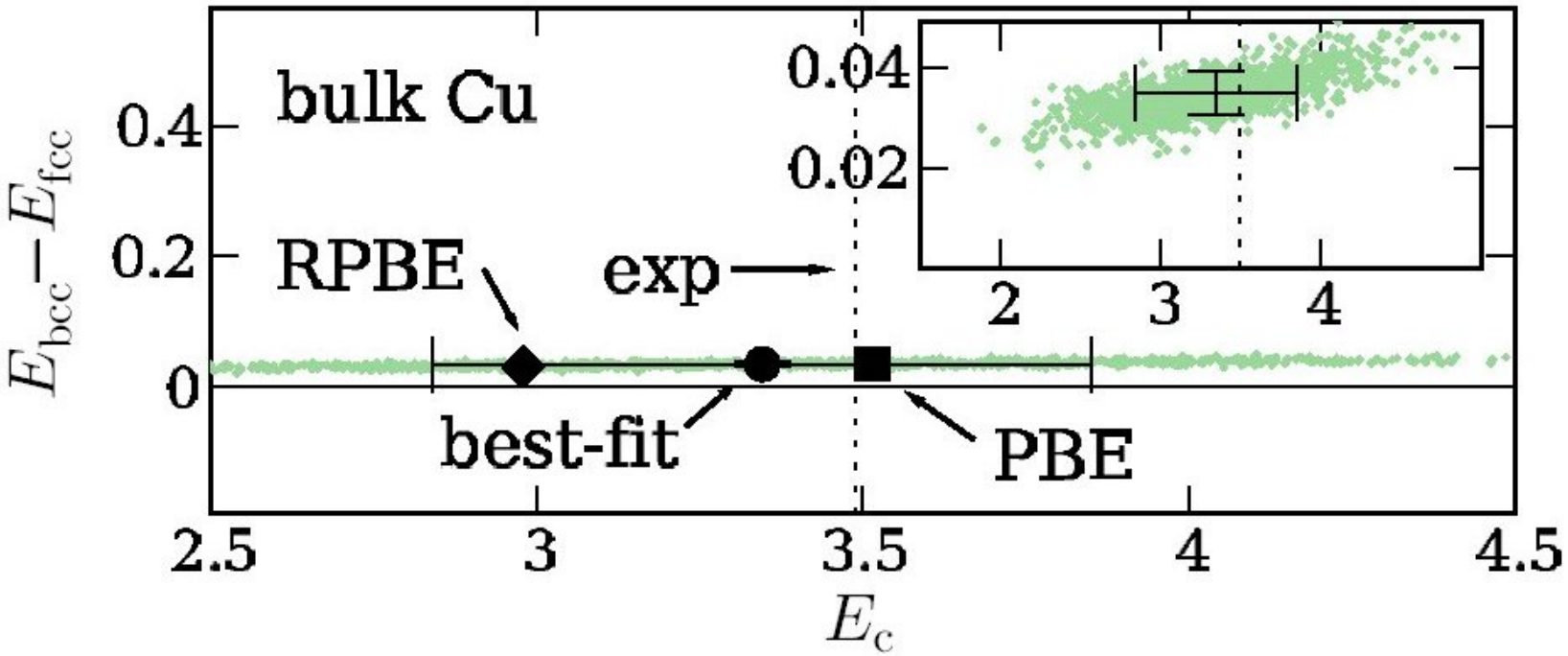
Best fit plus BEE:

E_c = 3.3 ± 0.5 eV

$E_{bcc} - E_{fcc}$ = 35 ± 4 meV

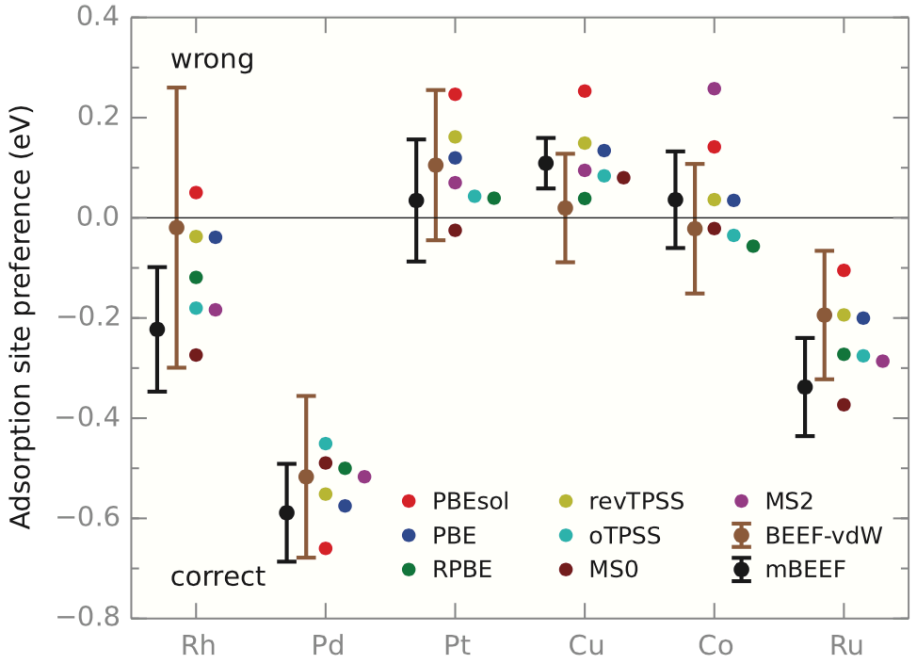


fcc bcc



Mortensen, Kaasbjerg, Frederiksen, Nørskov, Sethna, Jacobsen, Phys. Rev. Lett. 95, 216401 (2005).

The CO puzzle

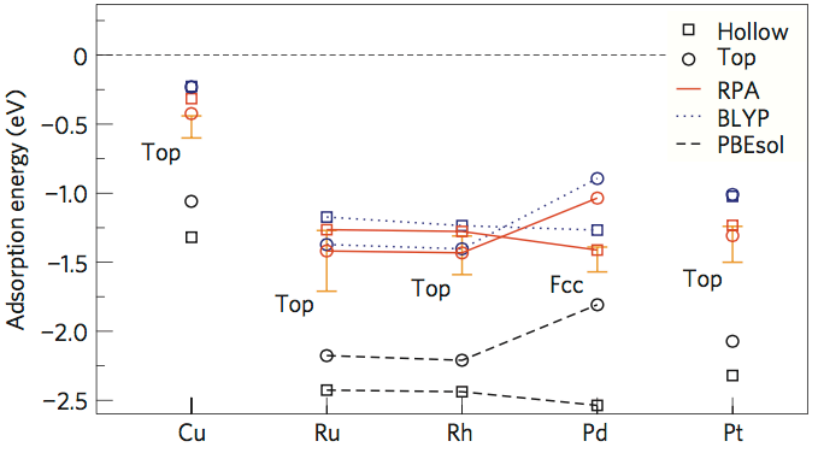


Hexagonal surfaces at 25% coverage

Binding energy at hollow site vs atop site

Expt: Pd(111) hollow site; the others atop

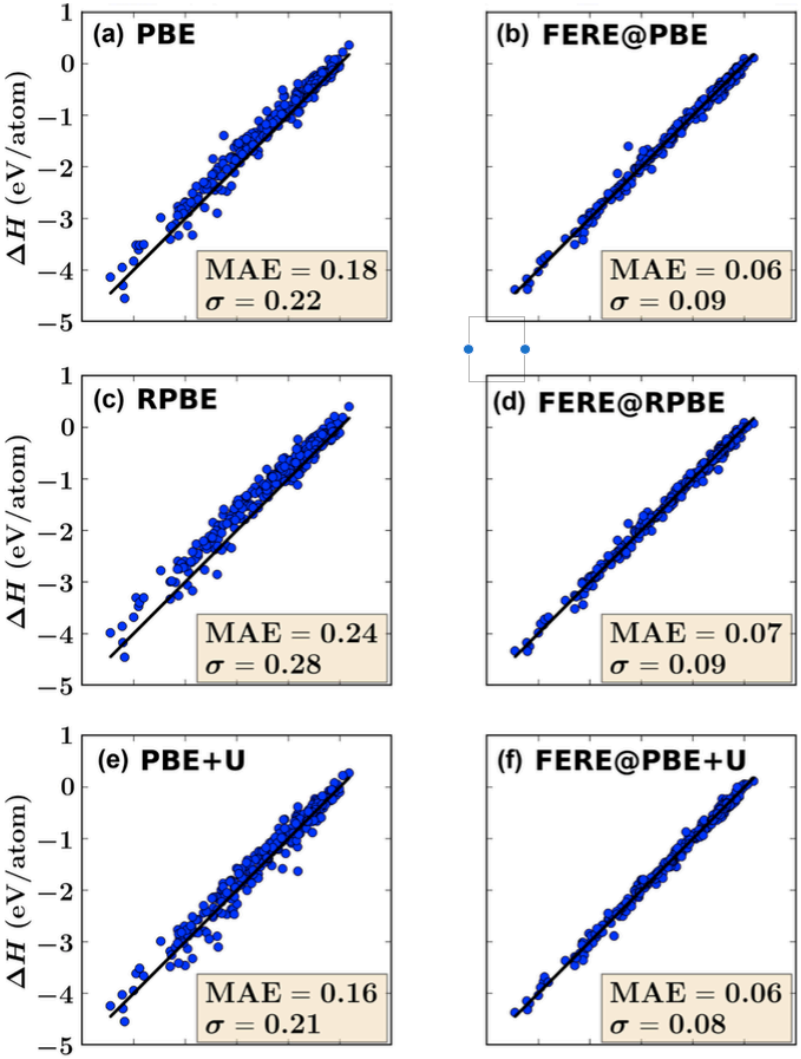
mBEEF:
Wellendorf, Lundgaard, Jacobsen, Bligaard,
J. Chem. Phys. 140, 144107 (2014)



RPA calcuations by
Schimka et al. Nature Materials **9**, 741 (2010).

Compound formation energies
Errors and error estimation

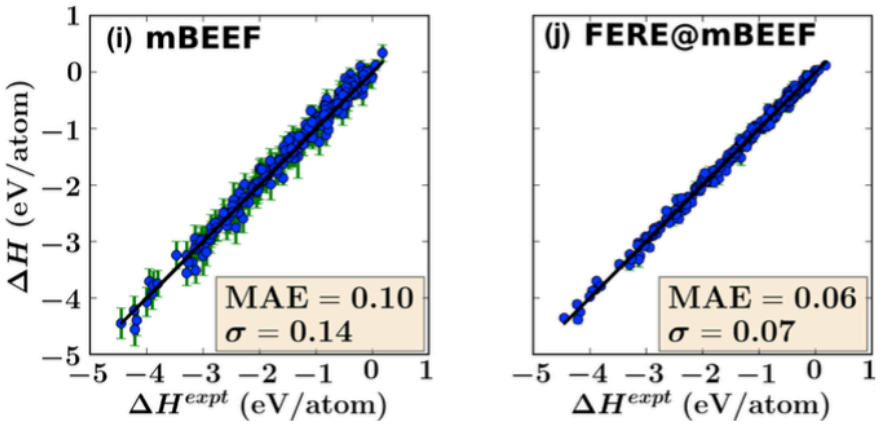
257 binary compounds with known heats of formation



Fitting of **E**lemental **R**eference **E**nergies:

$$\Delta H_f(A_{n_1}B_{n_2}\ldots) = E_{tot}(A_{n_1}B_{n_2}\ldots) - \sum_i n_i \mu_i^0$$

Fitting parameters



Test set of 24 binary and ternary compounds:

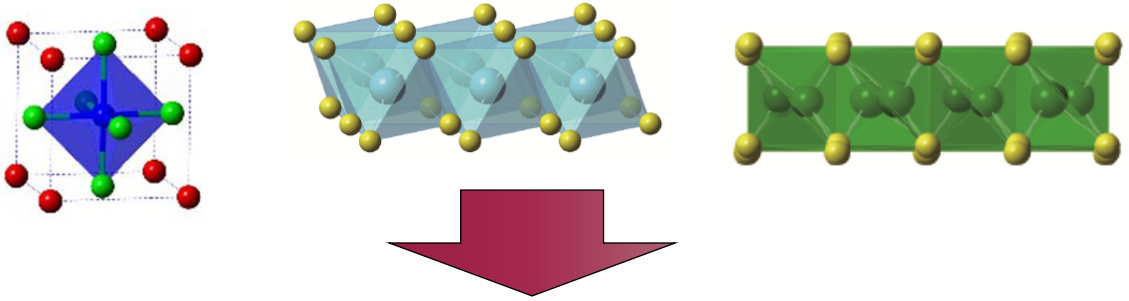
	PBE	PBE-FERE	mBEEF	mBEEF-FERE
MAE	0.24 eV	0.12 eV	0.12 eV	0.09 eV

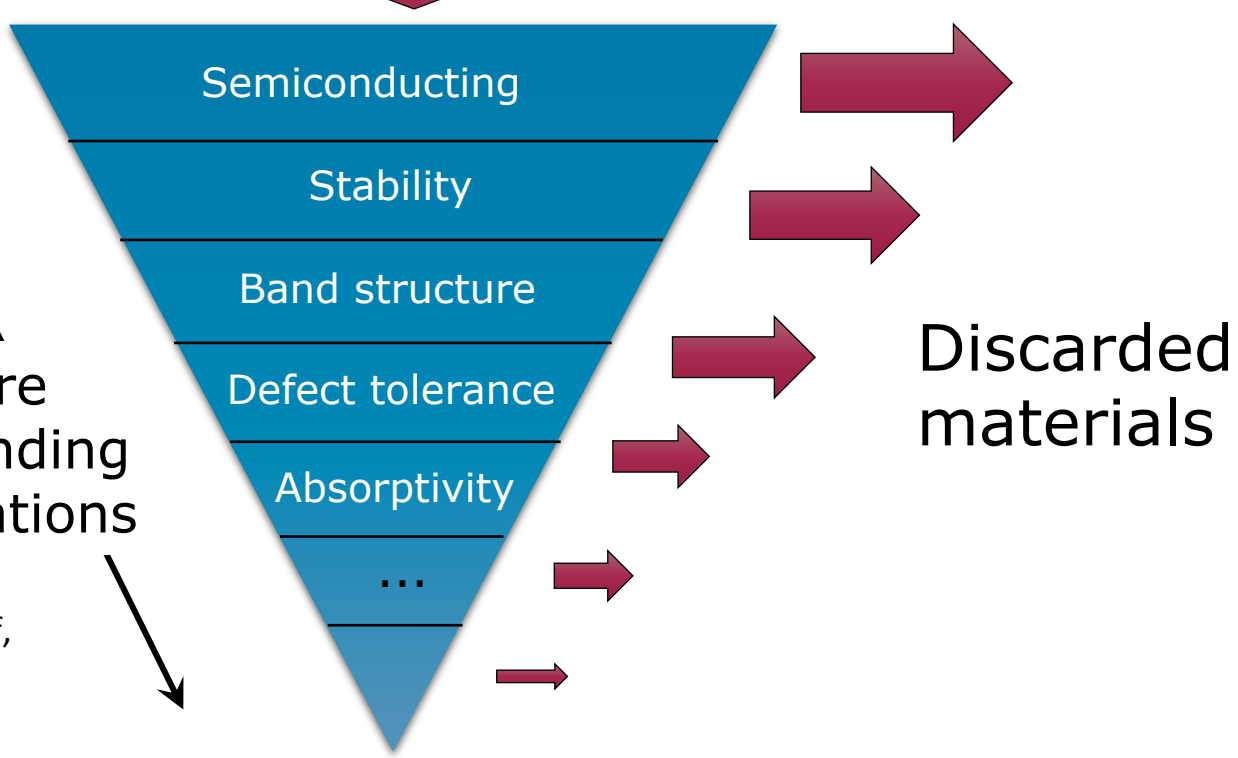
Pandey, Jacobsen, *Phys. Rev. B*, **91**, 235201 (2015)

FERE: Stevanovic, Lany, Zhang, Zunger, *Phys. Rev. B*, **85**, 115104 (2012)

Water splitting sulfide perovskites
Screening funnel

Initial structures:
(Experimental databases or hypothetical)





Kuhar, Crovetto, Pandey, Thygesen, Seger, Vesborg, Hansen, Chorkendorff, Jacobsen,
Energy Environ. Sci., **10**, 2579 (2017)

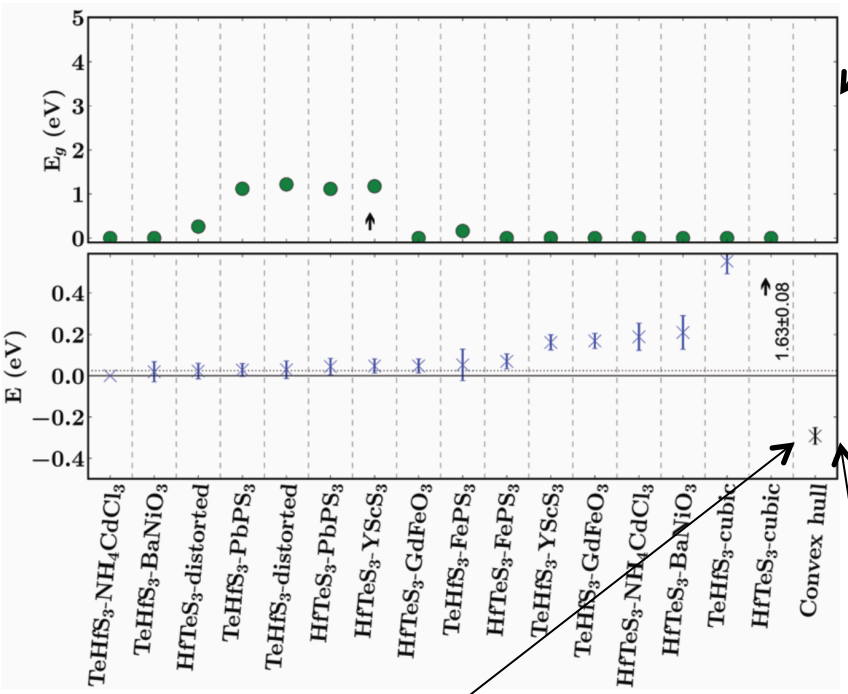
Candidates for experimental investigation

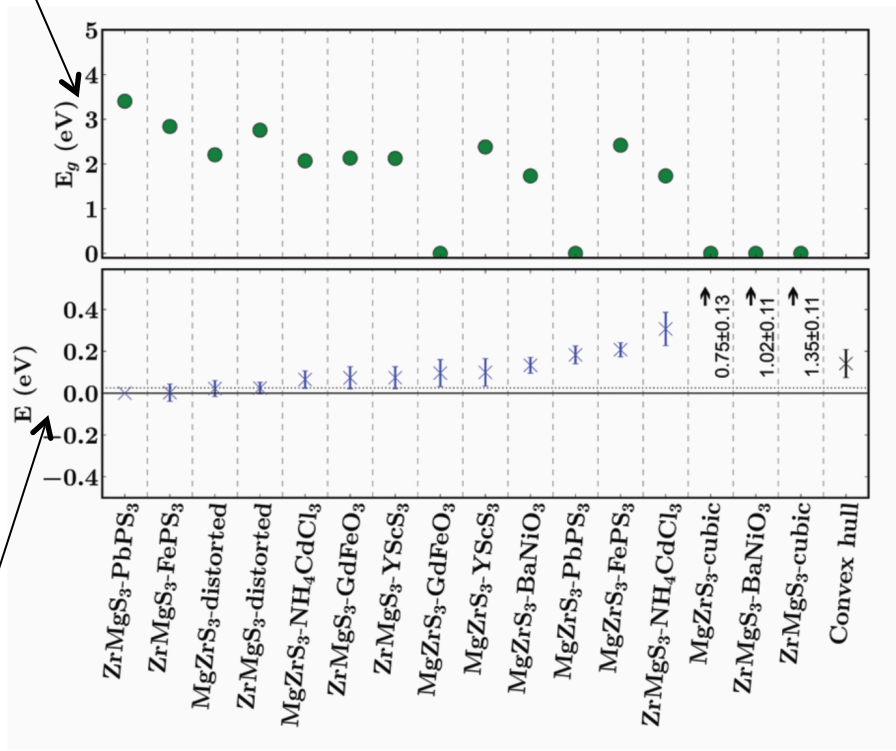
Sulfide perovskites
Stability

$TeHfS_3$

Band gap

$ZrMgS_3$





Energy of formation

Unstable relative to decomposition

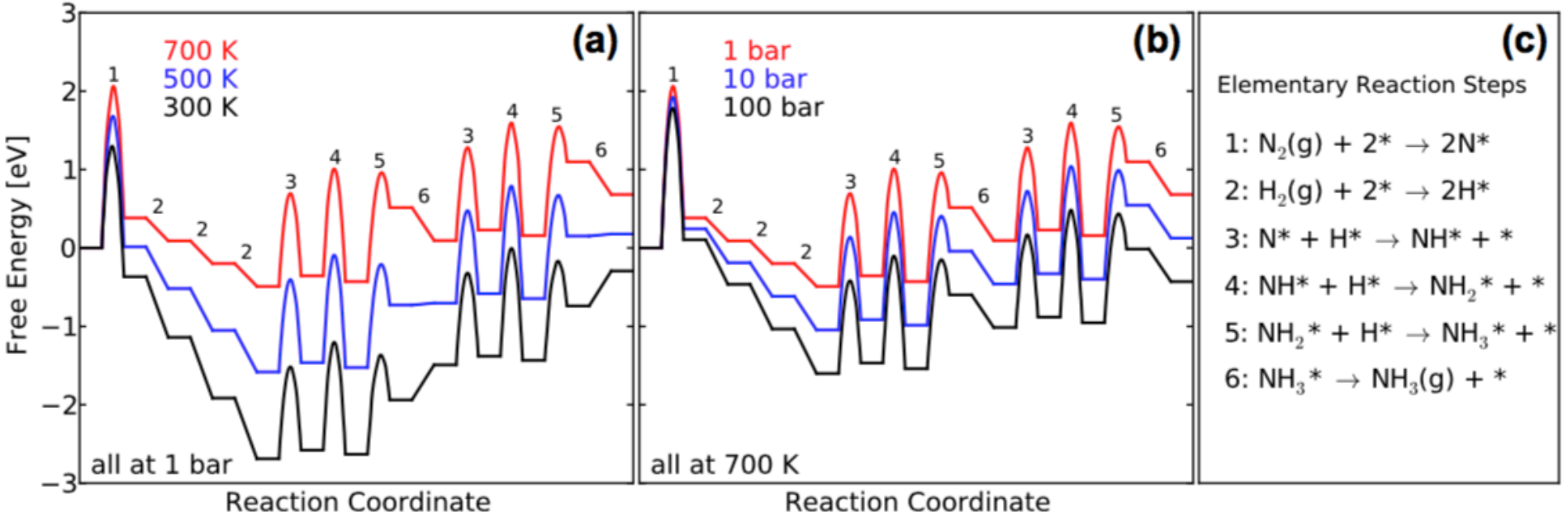
Use mBEEF uncertainties on energy of formation to improve predictions.

Kuhar, Crovetto, Pandey, Thygesen, Seger, Vesborg, Hansen, Chorkendorff, Jacobsen,
Energy Environ. Sci., **10**, 2579 (2017)

Ammonia synthesis

$$N_2 + 3H_2 \rightarrow 2NH_3$$

Descriptors: Adsorption energies and reaction barriers
Rates depend exponentially on barriers. **Can we predict anything at all?**

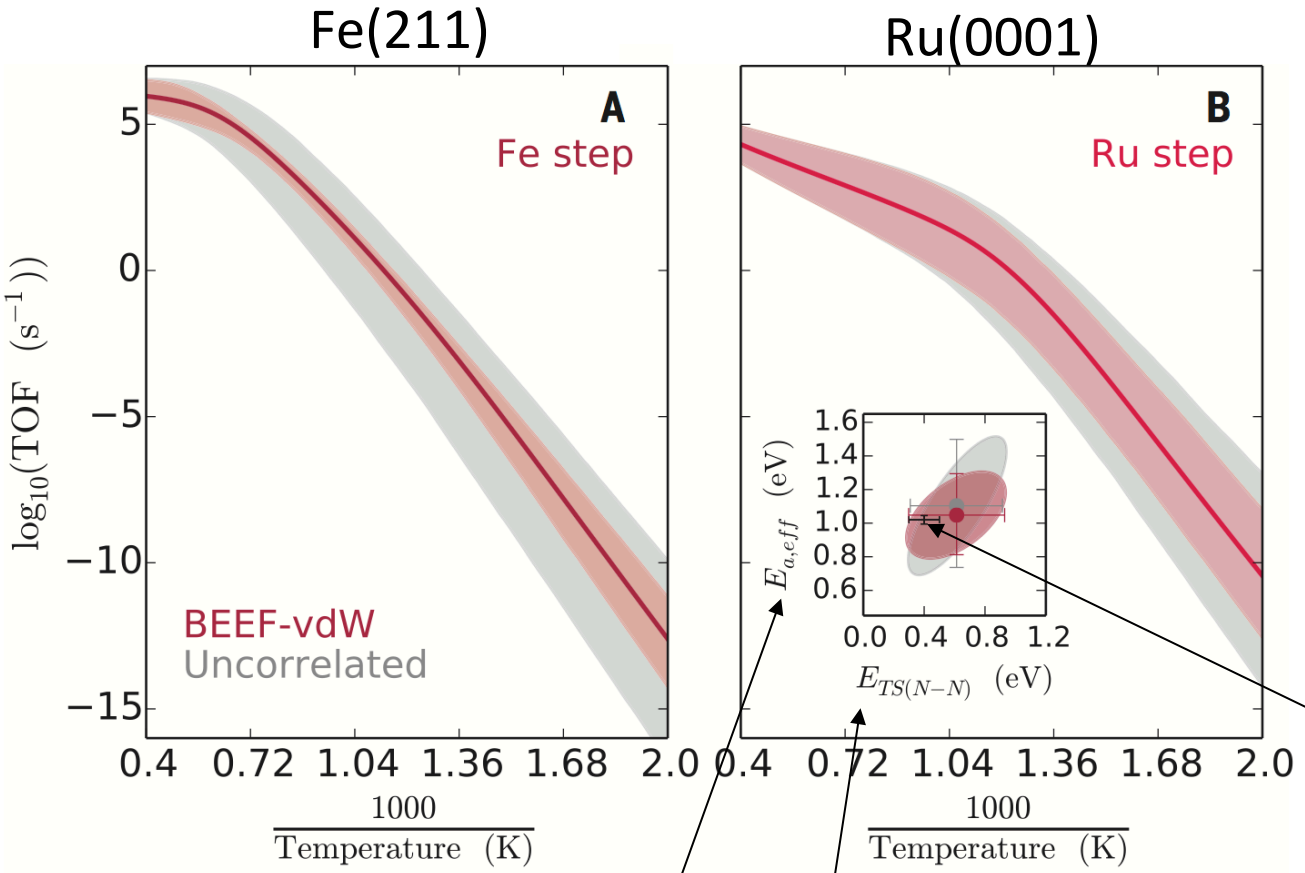


Vojvodic, Medford, Studt, Abild-Pedersen, Khan, Bligaard, and Nørskov, *Chemical Physics Letters*, **598**, 108 (2014)

Ammonia synthesis: Turn-over-frequency

Calculations including error bars.

BEEF-vdW



Industrial conditions: 100 bar, 50% approach to equilibrium

Correlations reduce error bars

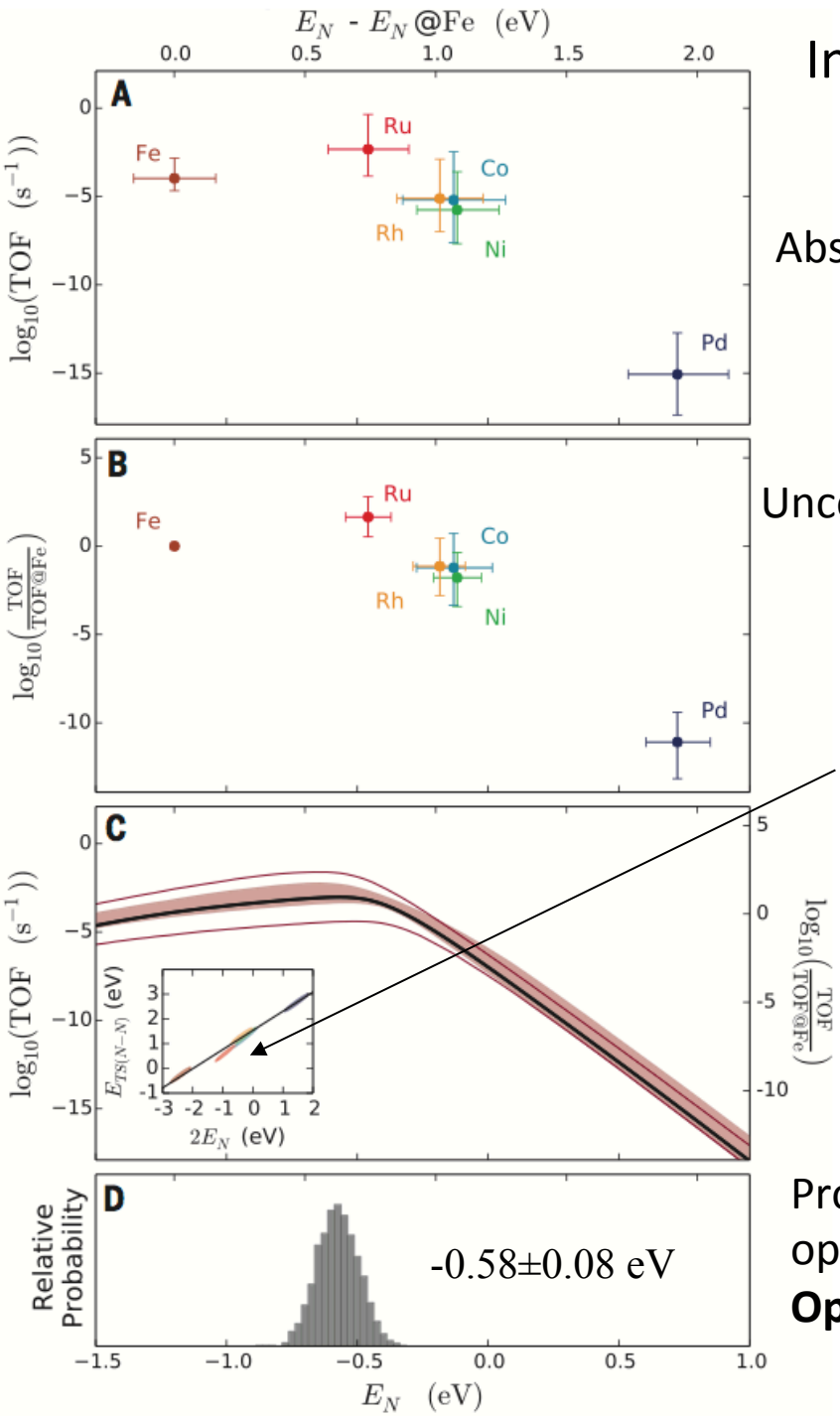
Experiment

Medford, Wellendorf, Vojvodic, Studt, Abild-Pedersen, Jacobsen, Bligaard, Nørskov, *Science* **345,** 197 (2014)

Effective activation energy

Rate limiting step: N-N dissociation

Investigating other metals



Absolute uncertainties

Uncertainties relative to Fe

Ensembles collapse along scaling relations (Brønsted-Evans-Polanyi)

Red curves: absolute uncertainties

Shaded area: relative to Fe

Probability distribution for optimum of volcano

Optimum is well determined!

Medford, Wellendorf, Vojvodic, Studt, Abild-Pedersen, Jacobsen, Bligaard, Nørskov, *Science* **345,** 197 (2014)

Limitations

- Functional and error estimates depend on model (GGA) and database (solids and molecules)
- Problems if
  - a property is not represented in the database
  - model is unable to describe the property at all
    - vd-Waals interactions with GGA

The End!