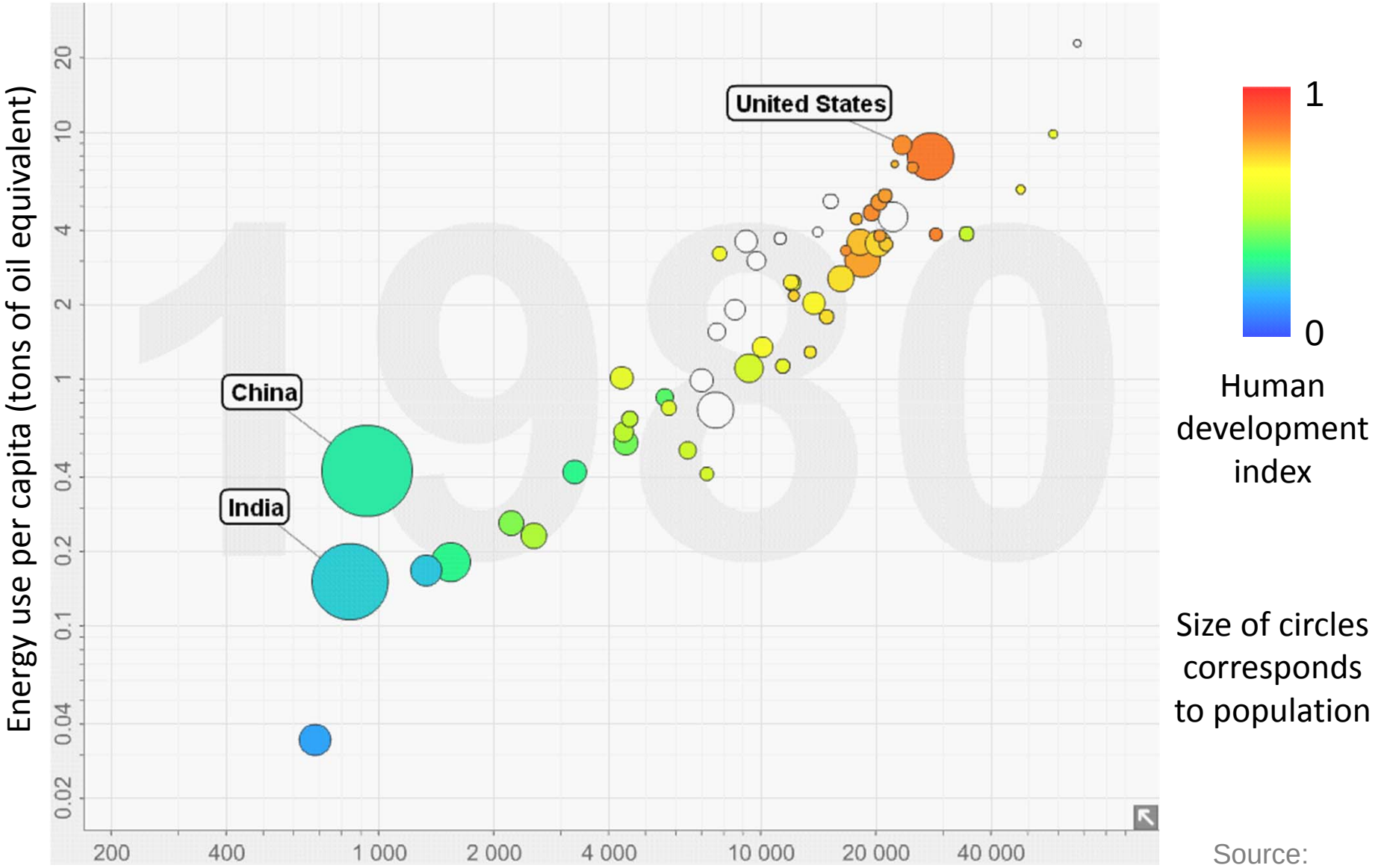
Quantum Monte Carlo for Materials Design

Tim Mueller

Johns Hopkins University

Supported by INCITE

Increasing demand for resources



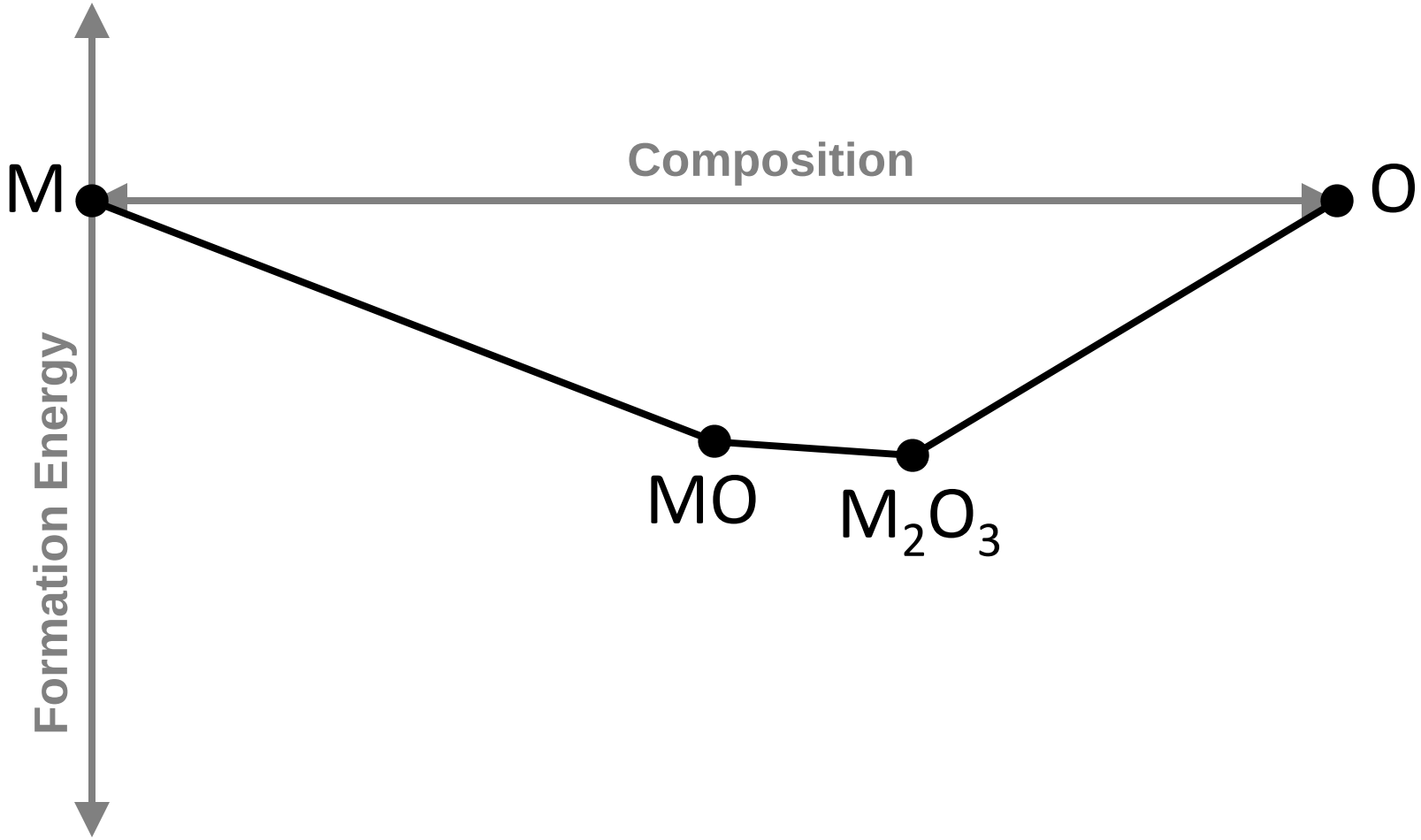
Source: www.gapminder.org

The Materials Genome Initiative

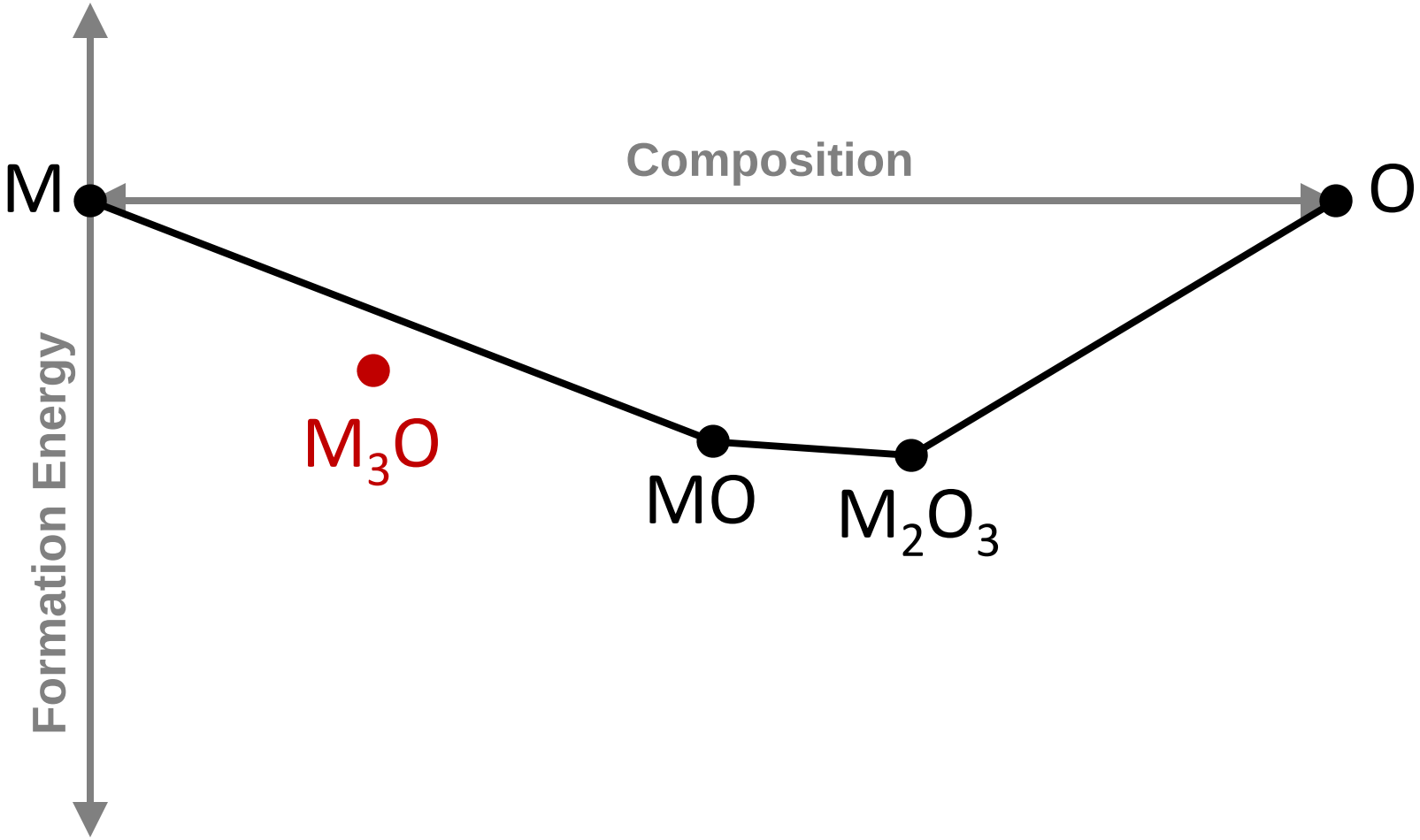
Use information technology to accelerate the development of advanced materials

- Develop database of properties for all materials
- Make available data and tools for data analysis

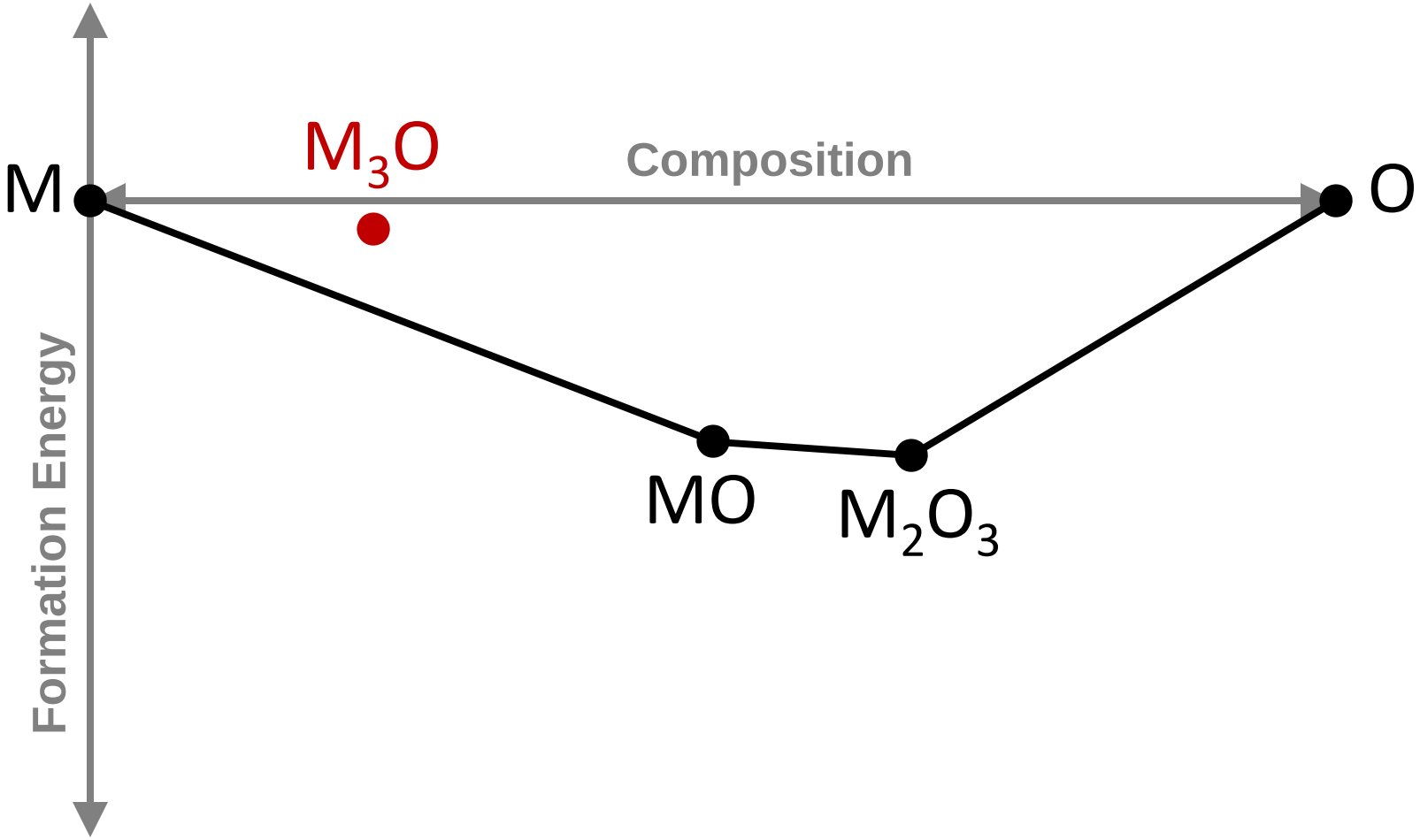
Formation energies predict stability and reactivity



Formation energies predict stability and reactivity



Formation energies predict stability and reactivity



Formation energies predict stability and reactivity

A database of material formation energies (and eventually formation free energies) would allow researchers to predict whether a hypothetical new material will be thermodynamically stable.

The Long-Term Goal

1. Material designer dreams up new material for some technology. Calculated material properties look promising.

The Long-Term Goal

1. Material designer dreams up new material for some technology. Calculated material properties look promising.

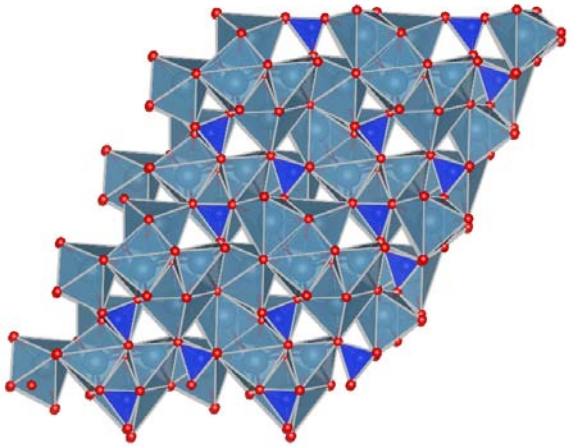
2. Designer runs QMC on hypothetical material and checks stability against database of known free energies.

The Long-Term Goal

1. Material designer dreams up new material for some technology. Calculated material properties look promising.

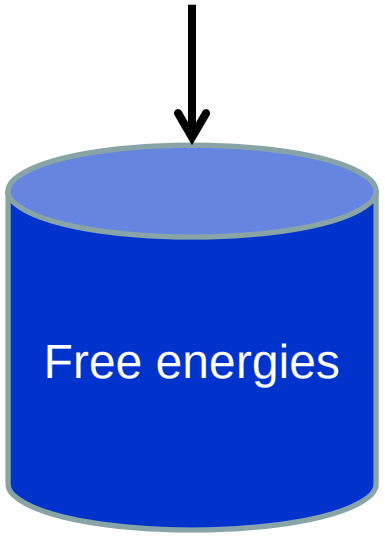
2. Designer runs QMC on hypothetical material and checks stability against database of known free energies.

3. Designer determines whether it is worth synthesizing the material, and under what conditions it might be synthesized.

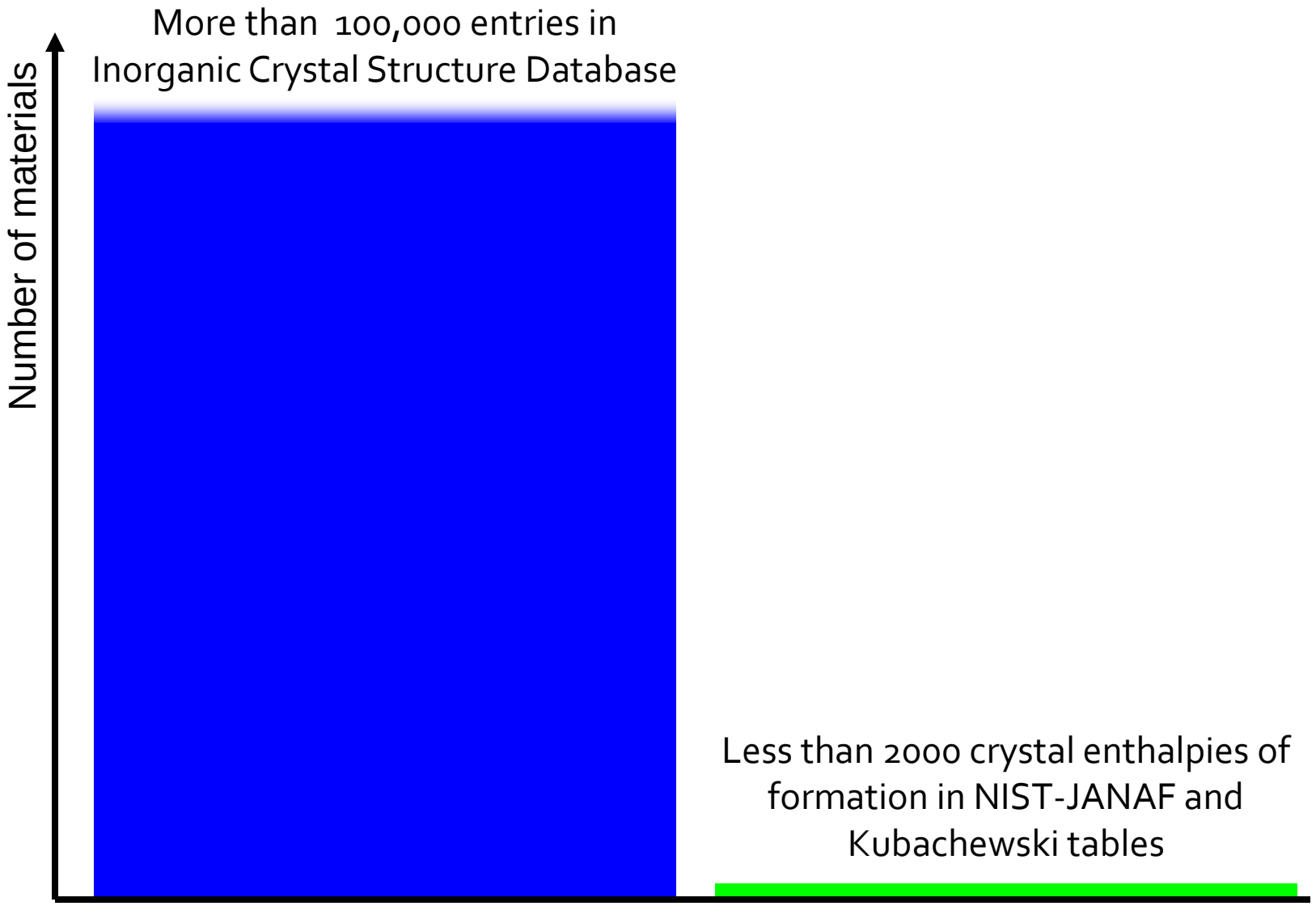


The Long-Term Goal

The values in this database only need to be calculated once, and they can be re-used for a very long time.



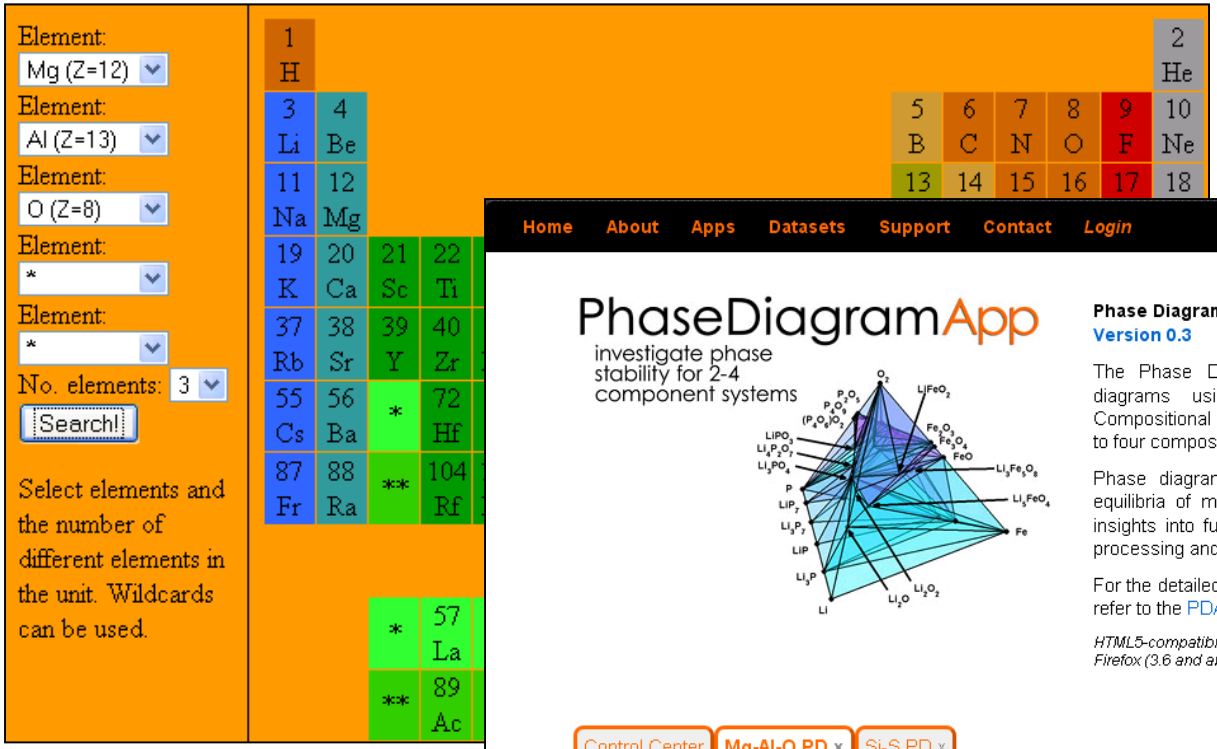
We lack experimental data on many materials

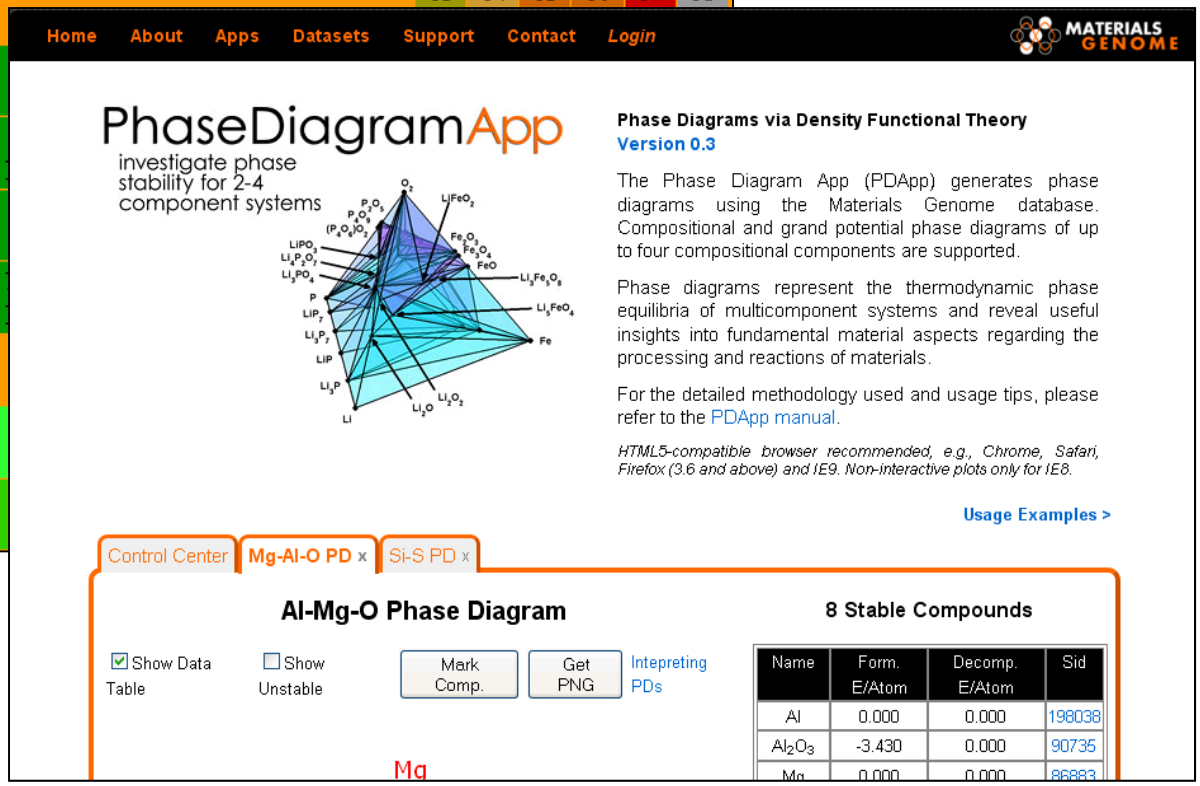


Existing online data and tools – typically created using DFT

http://gurka.fysik.uu.se/ESP/

C. Ortiz, O. Eriksson, and M. Klintenberg, **Computational Materials Science** 44 (2009) 1042

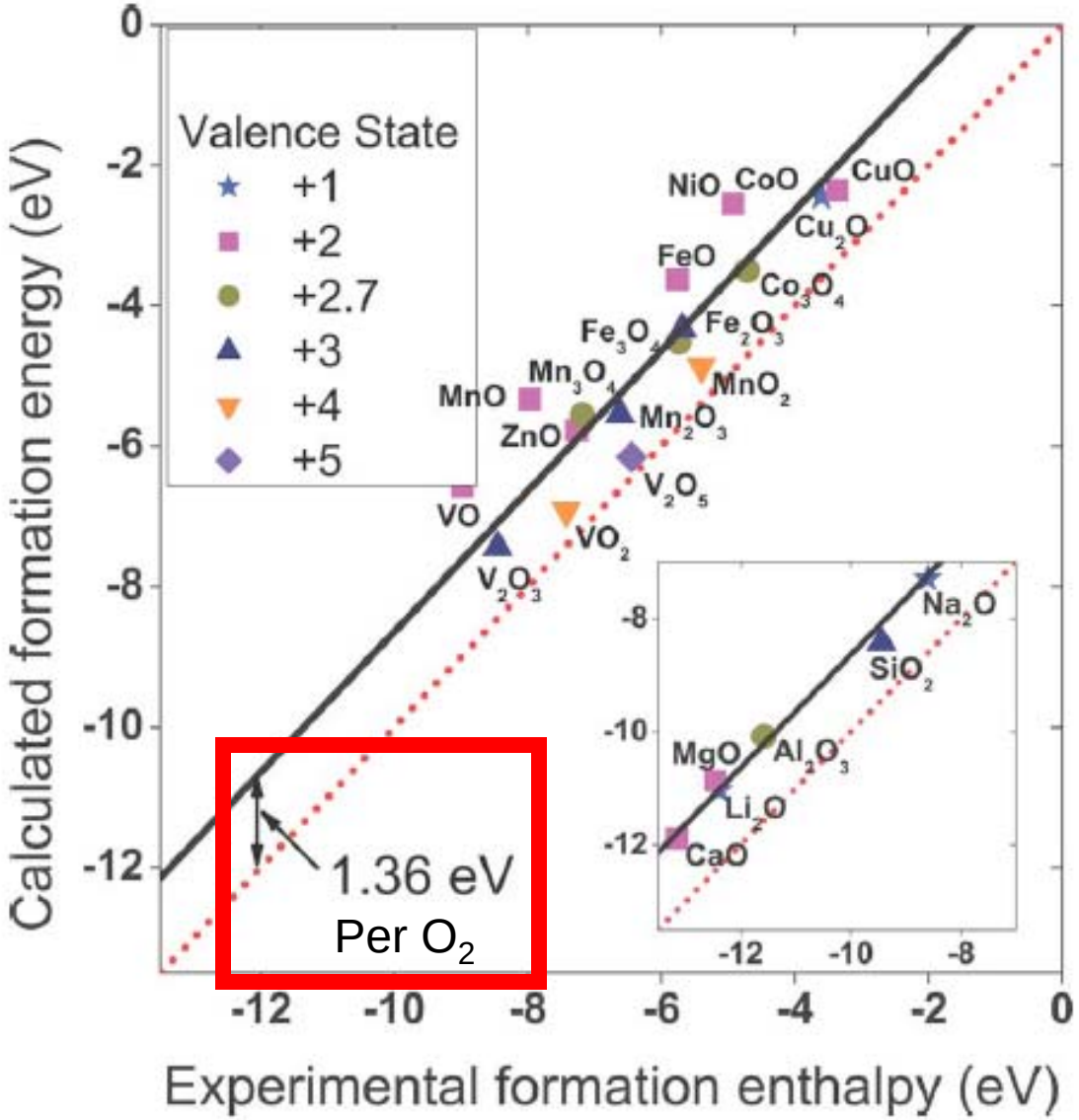




A. Jain et al., **Computational Materials Science** 50 (2011) 2295

http://www.materialsproject.org

DFT gets many energies wrong, including oxidation energies



L. Wang et al., **Phys. Rev. B.** 73 (2006) 195107

Correcting for oxygen exposes other problems

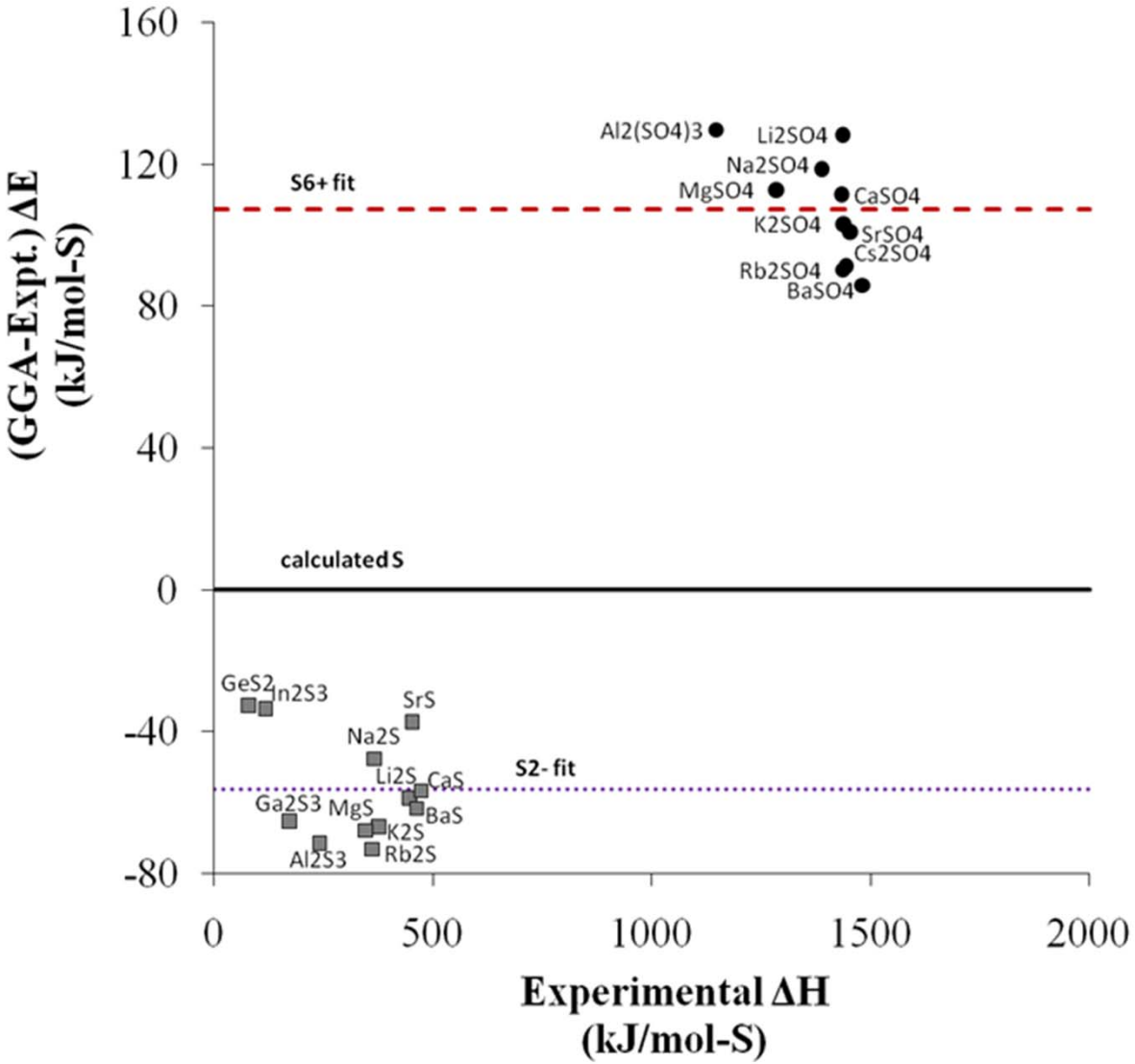


Image courtesy of Anubhav Jain

No experimental data on regions of chemical space

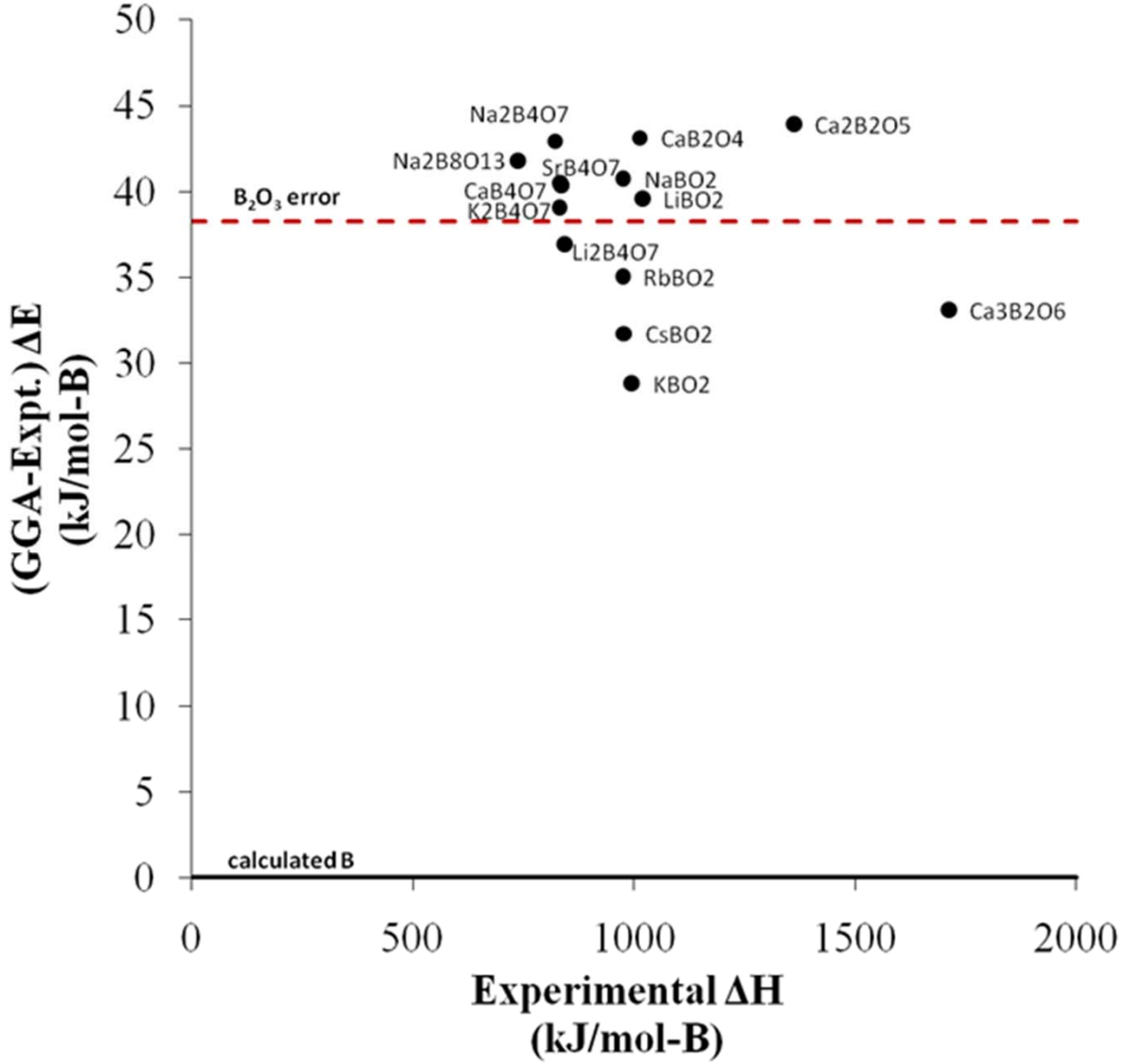
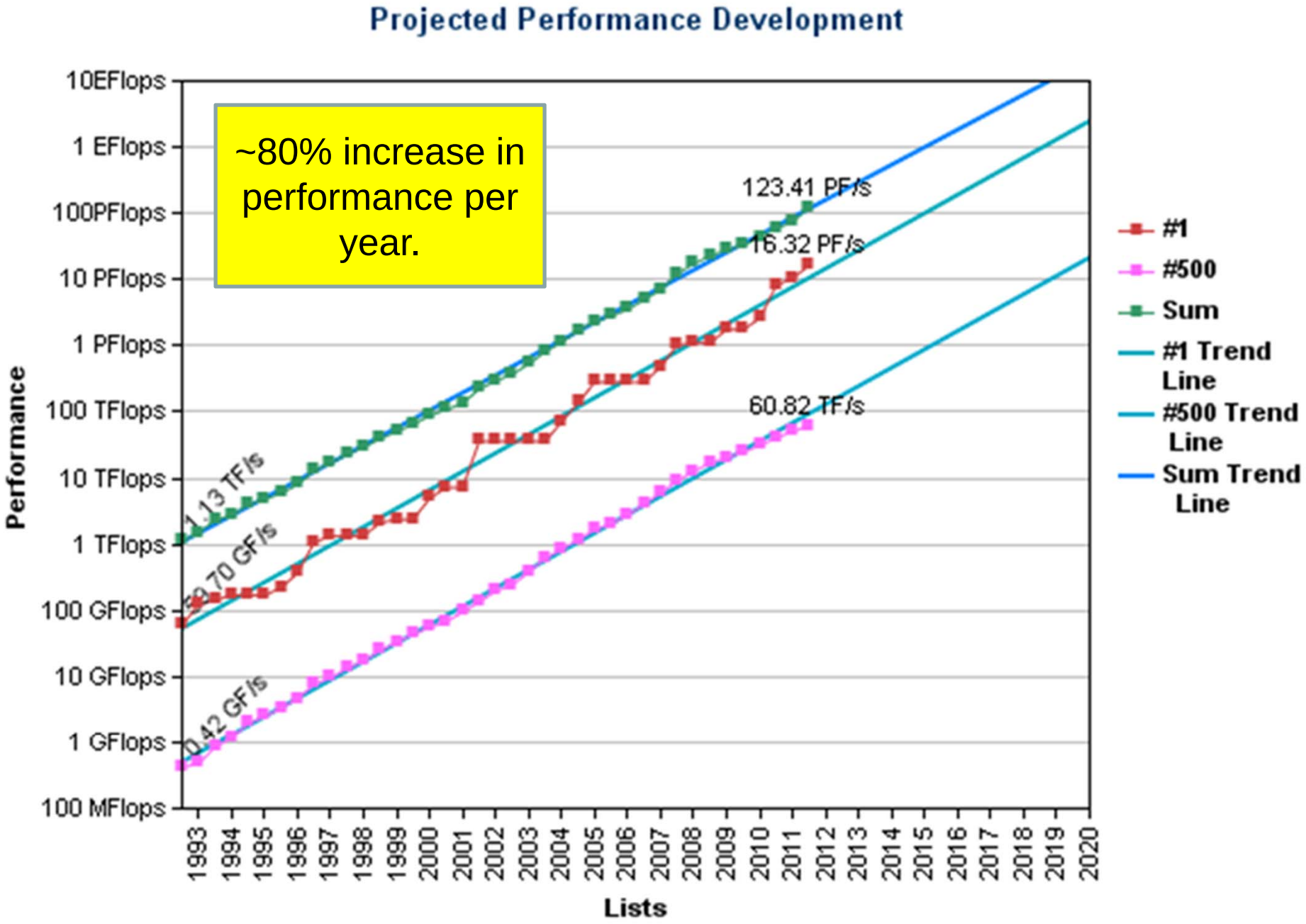
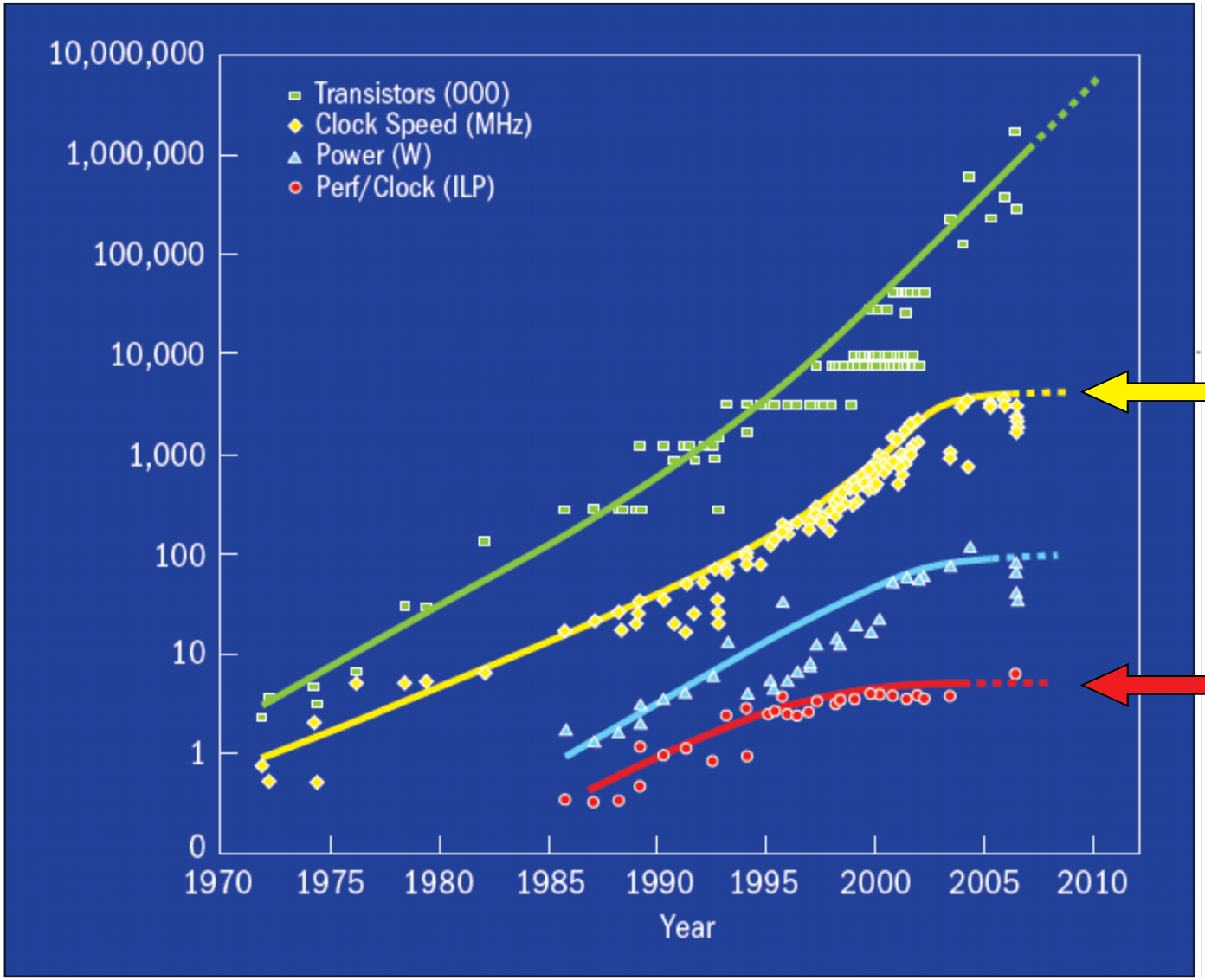


Image courtesy of Anubhav Jain


Projected Performance Development
~80% increase in performance per year.
123.41 PF/s
16.32 PF/s
60.82 TF/s
1.13 TF/s
59.70 GF/s
0.42 GF/s
10EFlops
1 EFlops
100PFlops
10 PFlops
1 PFlops
100 TFlops
10 TFlops
1 TFlops
100 GFlops
10 GFlops
1 GFlops
100 MFlops
Performance
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Lists
#1
#500
Sum
#1 Trend Line
#500 Trend Line
Sum Trend Line

Performance increases will come from parallelism



Individual processor speeds are reaching a plateau

John Shalf et al., **SciDAC Review**, Fall 2009

QMC is the best available method for building a database of formation energies

- Scales well with number of processors

QMC is the best available method for building a database of formation energies

- Scales well with number of processors
- Scales well with system size – O(N) when calculating formation energies per atom

QMC is the best available method for building a database of formation energies

- Scales well with number of processors
- Scales well with system size – O(N) when calculating formation energies per atom
- Works for everything (molecules, metals, insulators, semiconductors…)

QMC is the best available method for building a database of formation energies

- Scales well with number of processors
- Scales well with system size – O(N) when calculating formation energies per atom
- Works for everything (molecules, metals, insulators, semiconductors…)
- Accurate energies

QMC energies of solids

Literature

Compound	QMC	Year	Experiment
Li	1.09 ± 0.05	1989	1.65
	1.57 ± 0.01	1996	
Na	1.14 ± 0.01	2008	1.11
	1.0221 ± 0.0003	2003	
Mg	1.51 ± 0.01	2008	1.52
Al	3.23 ± 0.08	2002	3.43
MgH_2	6.84 ± 0.01	2008	6.83
BN	12.85 ± 0.09	1997	12.9
C (diamond)	7.346 ± 0.006	2003	7.37
Si	4.62 ± 0.01	2004	4.62
Ge	3.85 ± 0.10	1995	3.86
GaAs	4.9 ± 0.2	1996	6.7
MnO	9.29 ± 0.04	2010	9.5
FeO	9.66 ± 0.04	2008	9.7
NiO	9.442 ± 0.002	2003	9.5
$BaTiO_3$	31.2 ± 0.3	2007	31.57

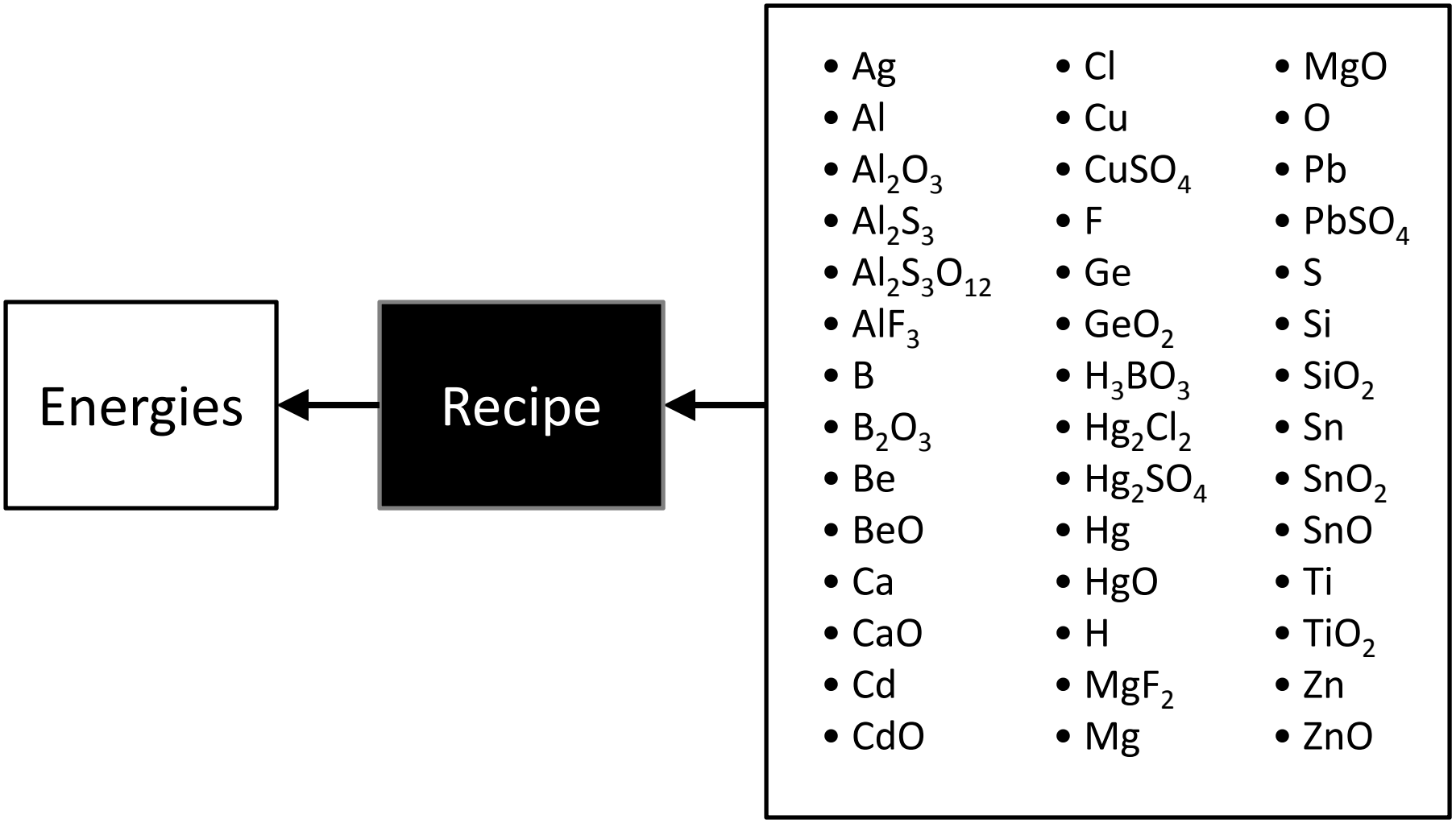
Jindřich Kolorenč and Lubos Mitas, **Rep. Prog. Phys** 74 (2011) 025602

This work

- Ag
- Al
- Al_2O_3
- Al_2S_3
- $Al_2S_3O_{12}$
- AlF_3
- B
- B_2O_3
- Be
- BeO
- Ca
- CaO
- Cd
- CdO
- Cl
- Cu
- $CuSO_4$
- F
- Ge
- GeO_2
- H_3BO_3
- Hg_2Cl_2
- Hg_2SO_4
- Hg
- HgO
- H
- MgF_2
- Mg
- MgO
- O
- Pb
- $PbSO_4$
- S
- Si
- SiO_2
- Sn
- SnO_2
- SnO
- Ti
- TiO_2
- Zn
- ZnO

. . .

For high-throughput QMC, need a "recipe"



The recipe

The recipe

Basis set

Plane waves	
Pros	• Natural basis for periodic systems • Used by many software packages • Single-parameter convergence
Cons	• Non-local, which can be expensive in real space.

The recipe

Basis set

Plane waves

Pros	• Natural basis for periodic systems • Used by many software packages • Single-parameter convergence
Cons	• Non-local, which can be expensive in real space.

Blips (splines)

Pros	• Can represent any basis • Local
Cons	• Require a lot of memory and disk space

The recipe

Basis set

Pseudopotentials

Norm-conserving pseudopotentials

ABINIT (FHI)	• Covers most of periodic table • Mostly large-core (soft)

The recipe

Basis set

Pseudopotentials

Norm-conserving pseudopotentials

ABINIT (FHI)	• Covers most of periodic table • Mostly large-core (soft)
Rappe	• An alternative to ABINIT • Only available for select elements

The recipe

Basis set

Pseudopotentials

Eric Shirley and Richard Martin, **Phys. Rev. B** 57 (1993) 15413

Core-polarization correction	
Material	Oxidation energy improvement*
CaO	48 kJ / mol O
SnO	-1 kJ / mol O
SnO_2	9 kJ / mol O
GeO_2	3 kJ / mol O

*After correcting for O_2 energy error

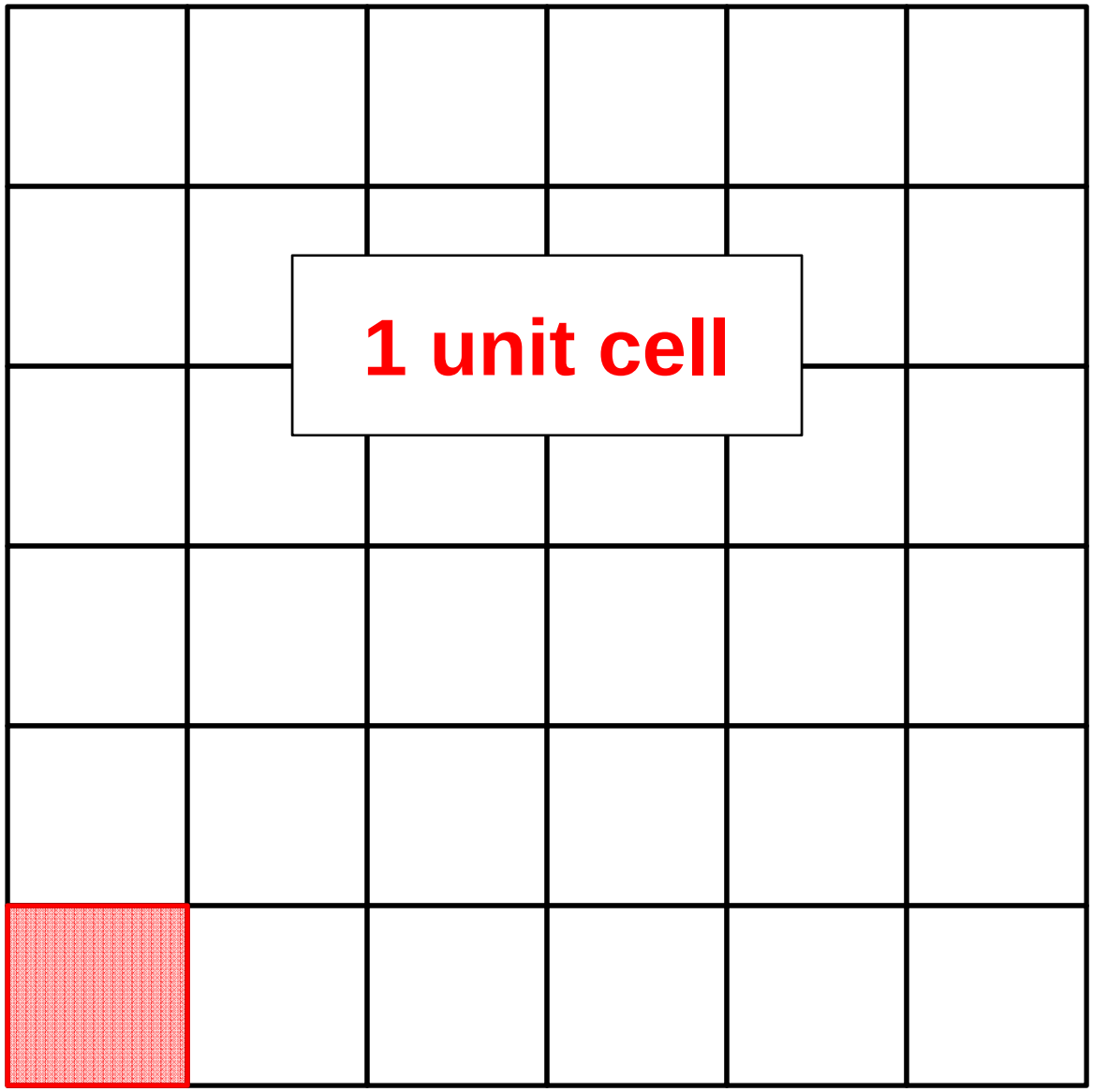
The recipe

Basis set

Pseudopotentials

Finite-size effects

Extrapolate to infinitely sized supercells



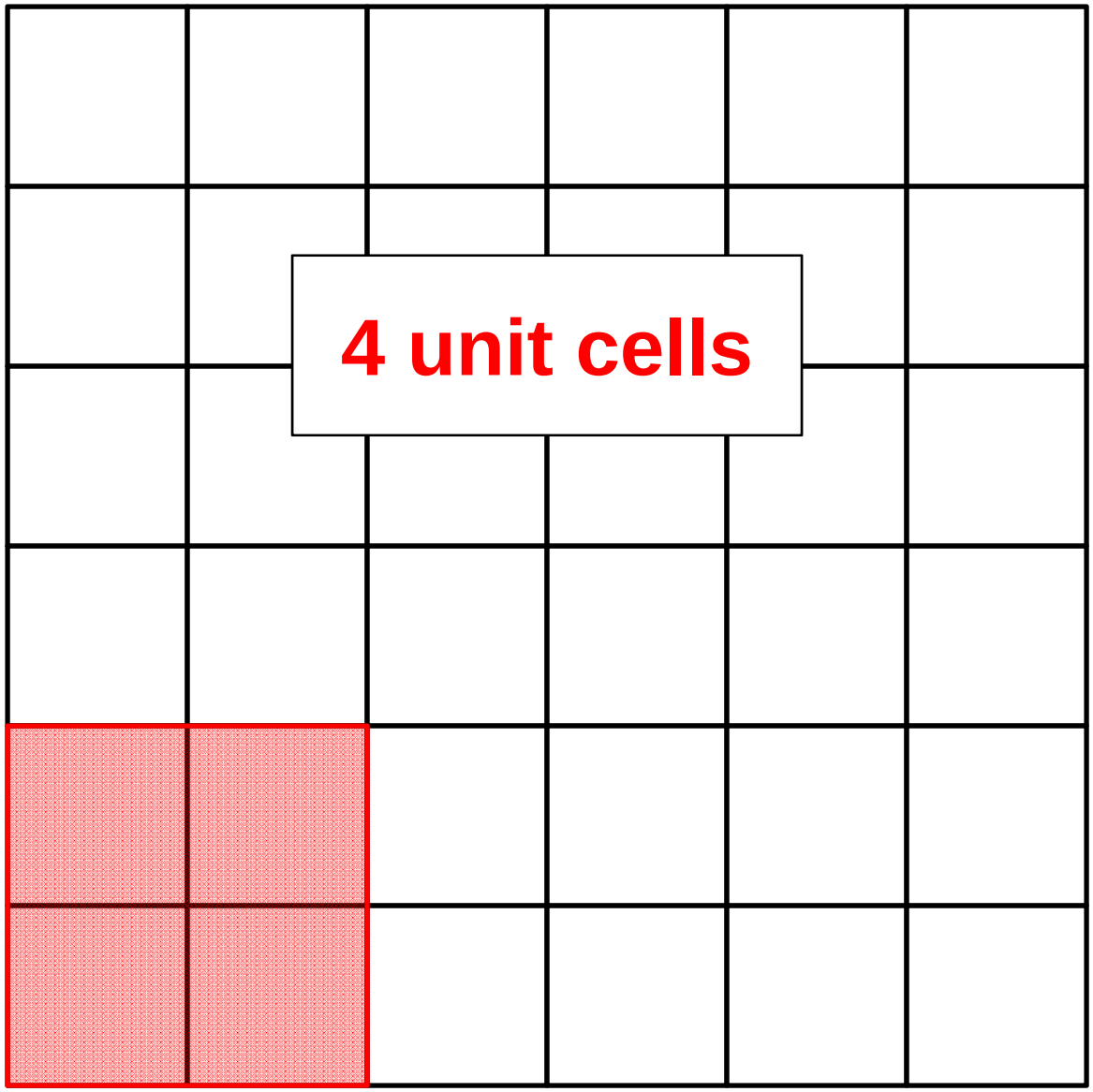
The recipe

Basis set

Pseudopotentials

Finite-size effects

Extrapolate to infinitely sized supercells



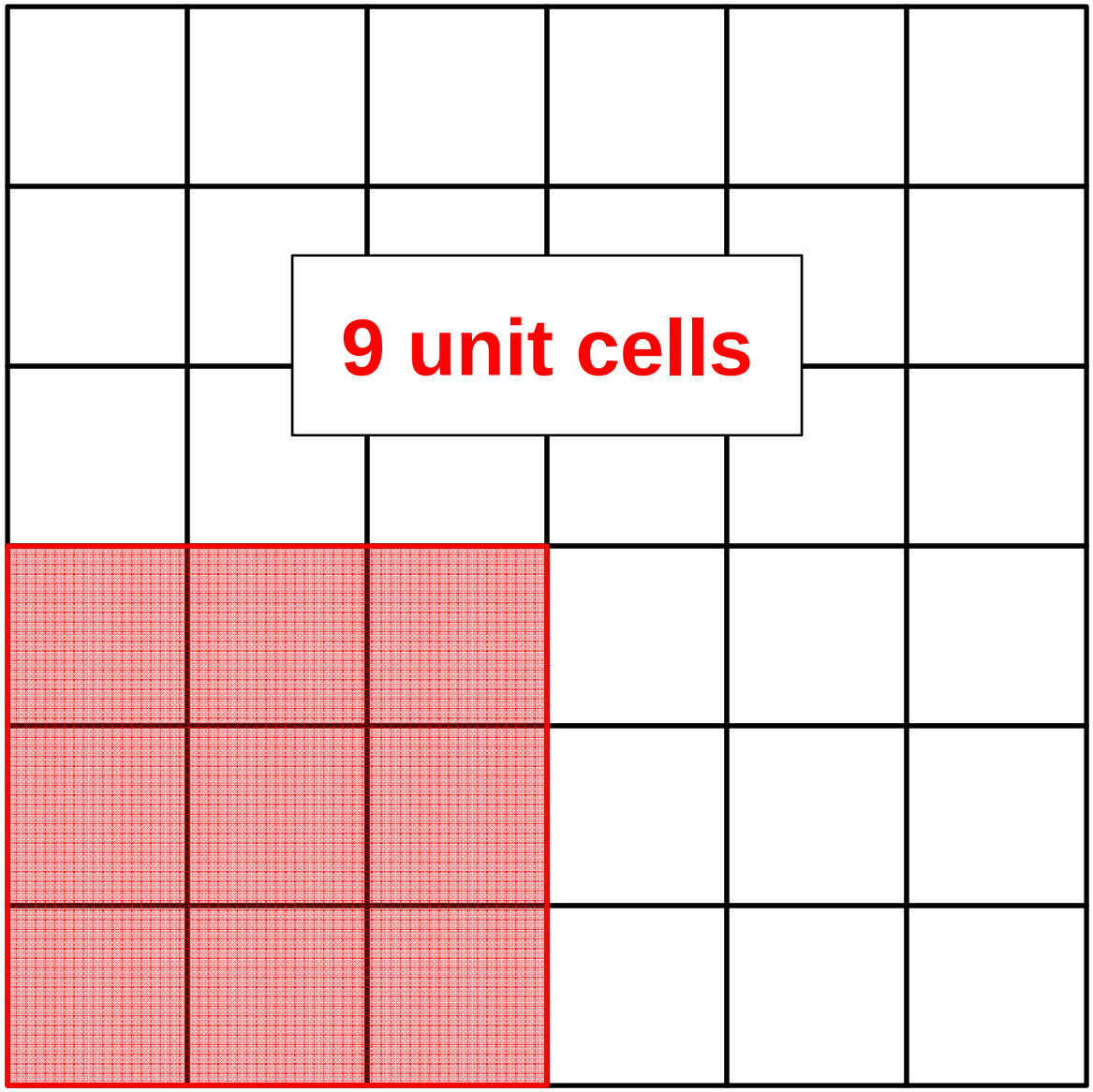
The recipe

Basis set

Pseudopotentials

Finite-size effects

Extrapolate to infinitely sized supercells



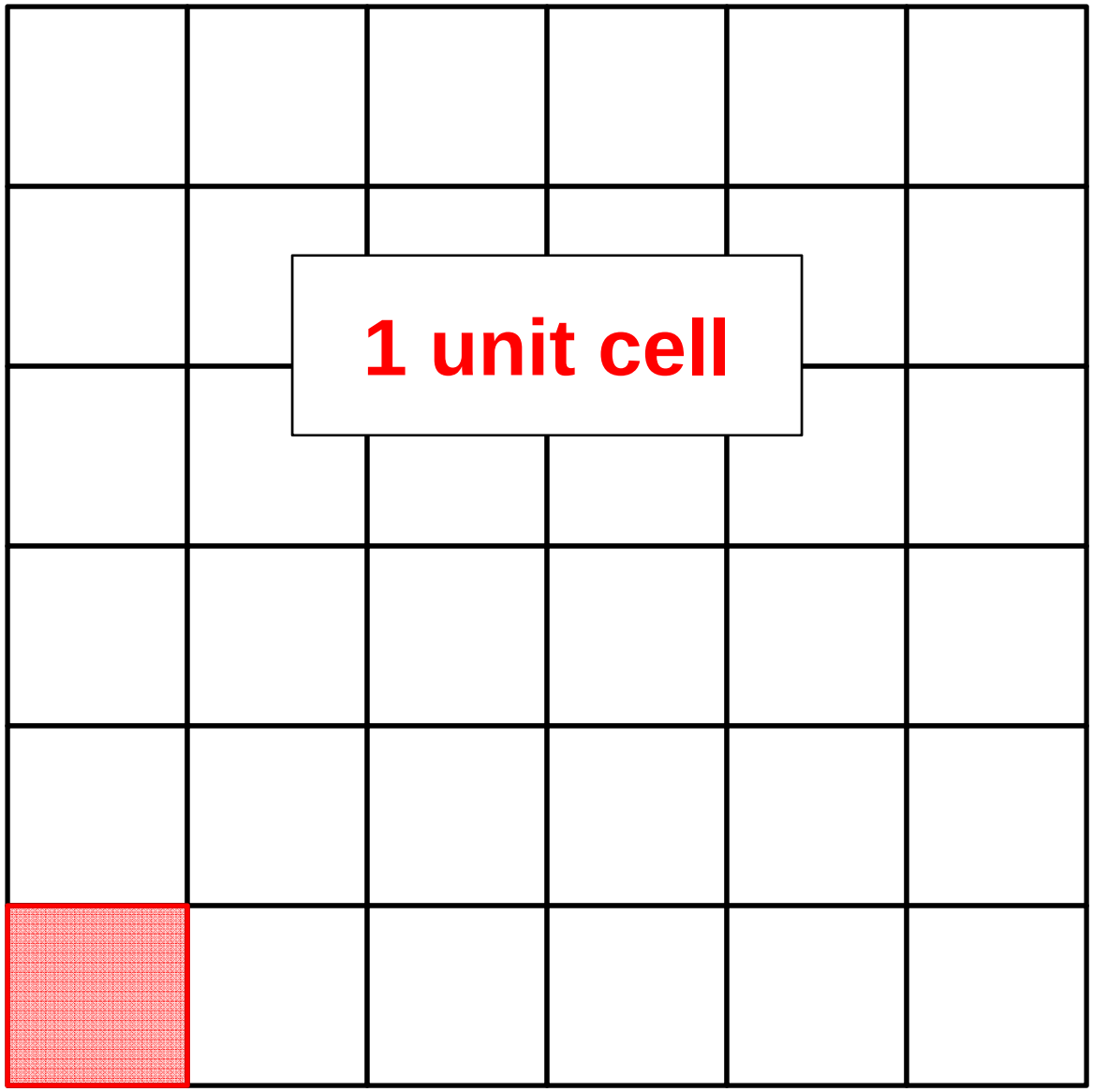
The recipe

Basis set

Pseudopotentials

Finite-size effects

Extrapolate to infinitely sized supercells



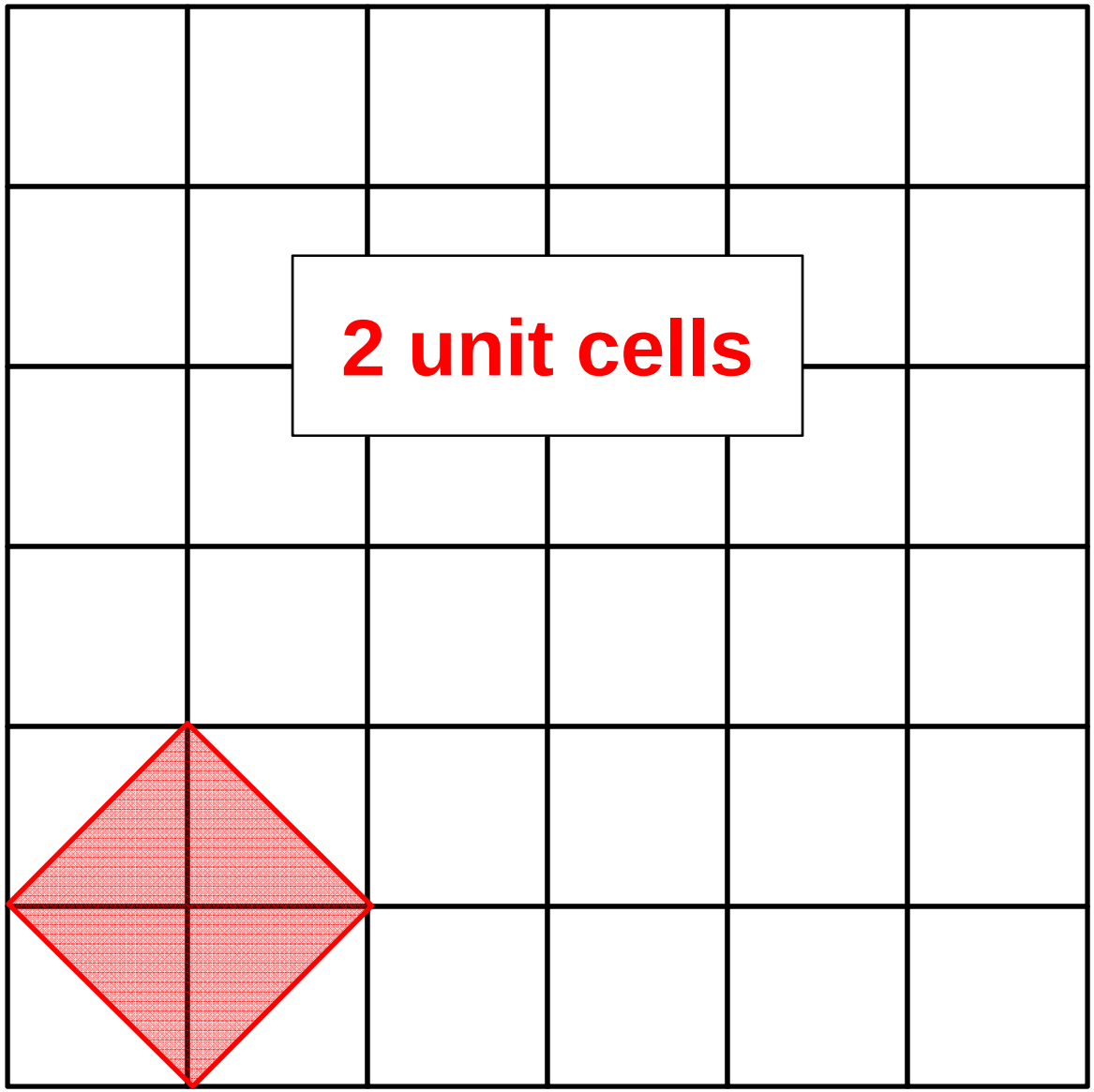
The recipe

Basis set

Pseudopotentials

Finite-size effects

Extrapolate to infinitely sized supercells



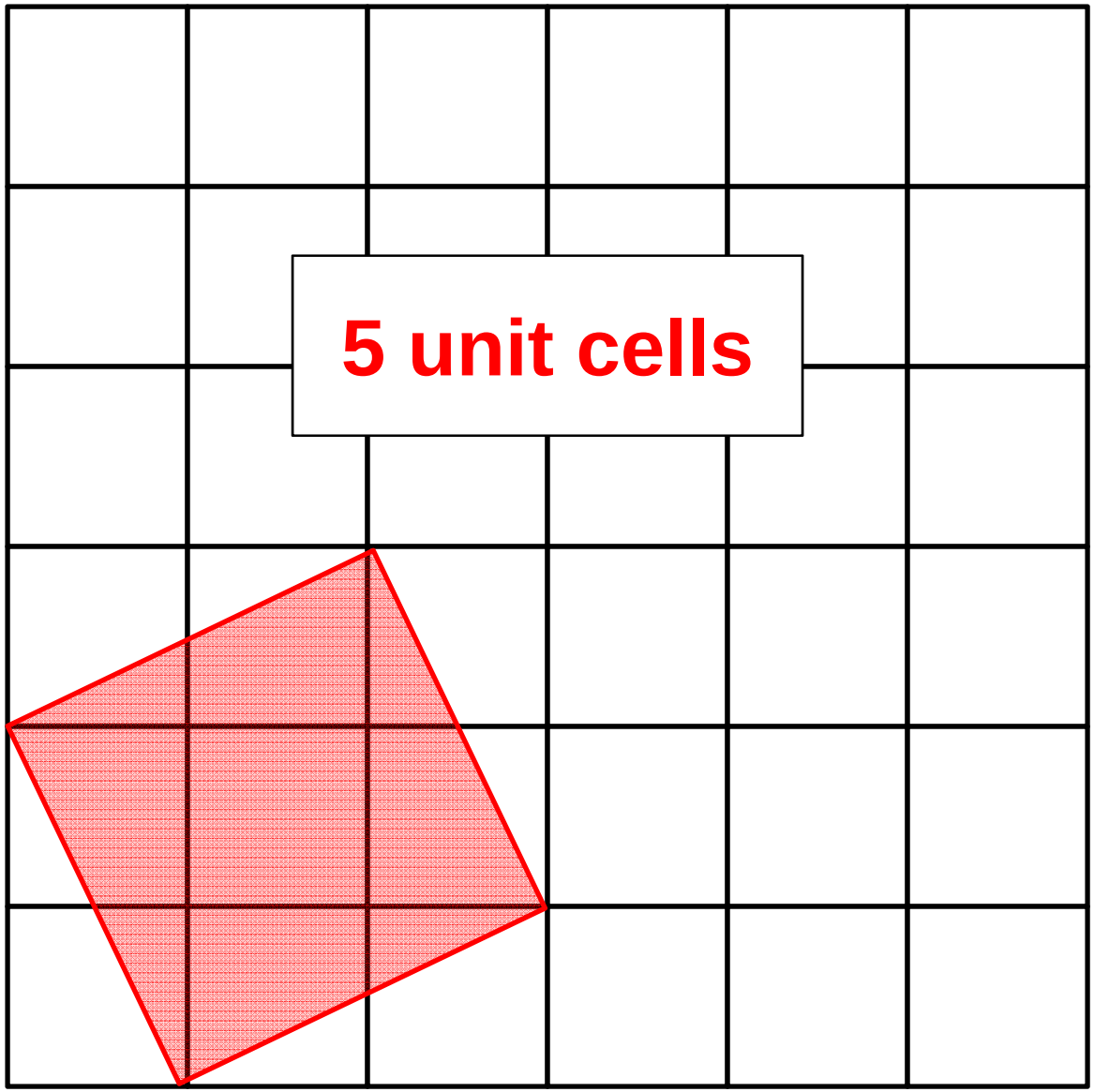
The recipe

Basis set

Pseudopotentials

Finite-size effects

Extrapolate to infinitely sized supercells



The recipe

Basis set

Pseudopotentials

Finite-size effects

Twist averaging

2x2x2 grid

- 2,500 statistics accumulation samples at each point .
- Real arithmetic can be done at these points.
- Costs only a little more than collecting 20,000 samples at the Gamma point (due to equilibration).

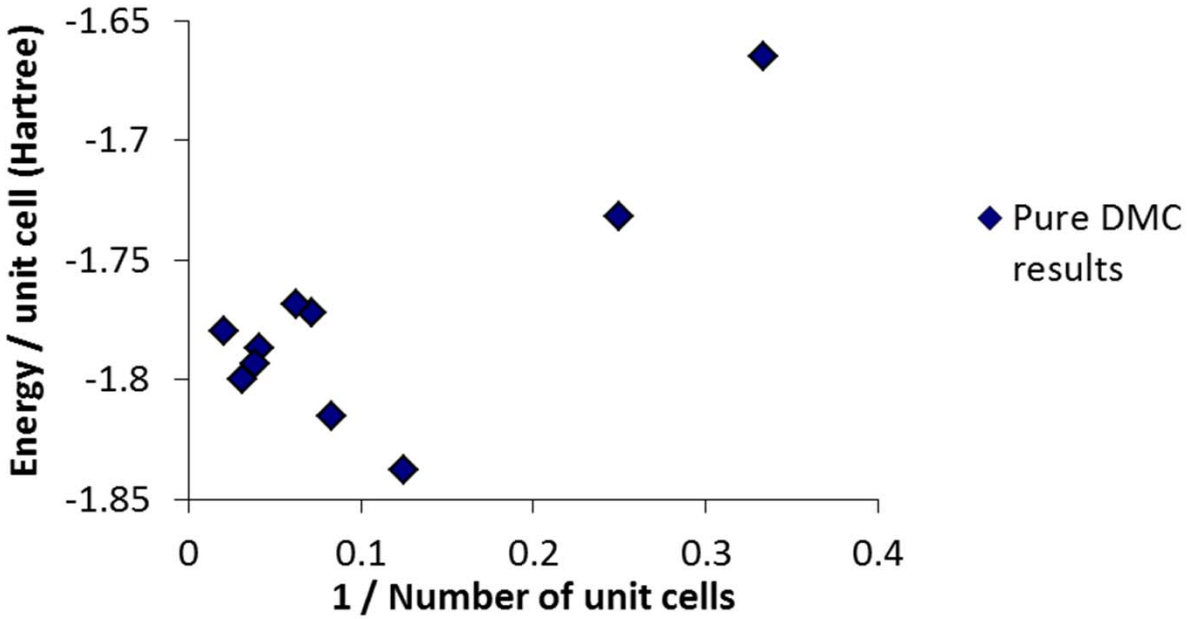
The recipe

Basis set

Pseudopotentials

Finite-size effects

The trend in pure QMC data is not clear (Γ-point data below)



The recipe

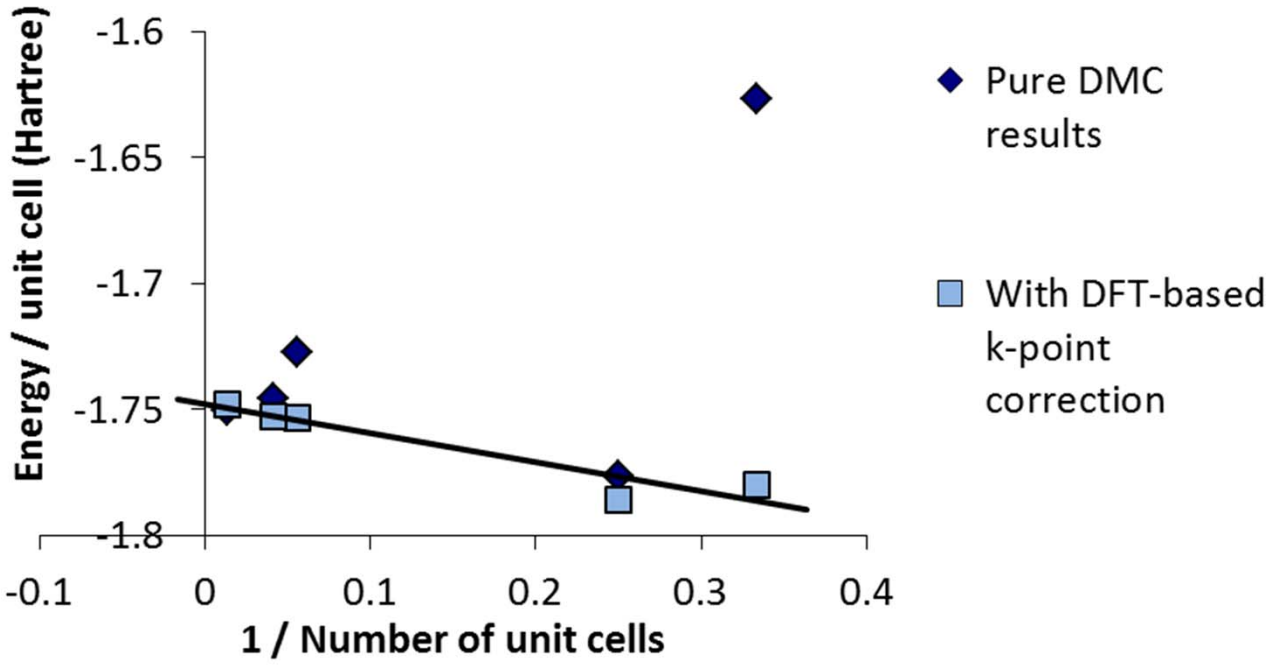
Basis set

Pseudopotentials

Finite-size effects

We adjust the QMC energies using:

$$E_{adjusted} = E_{(QMC,KPTS)} + [E_{(DFT,\infty)} - E_{(DFT,KPTS)}]$$



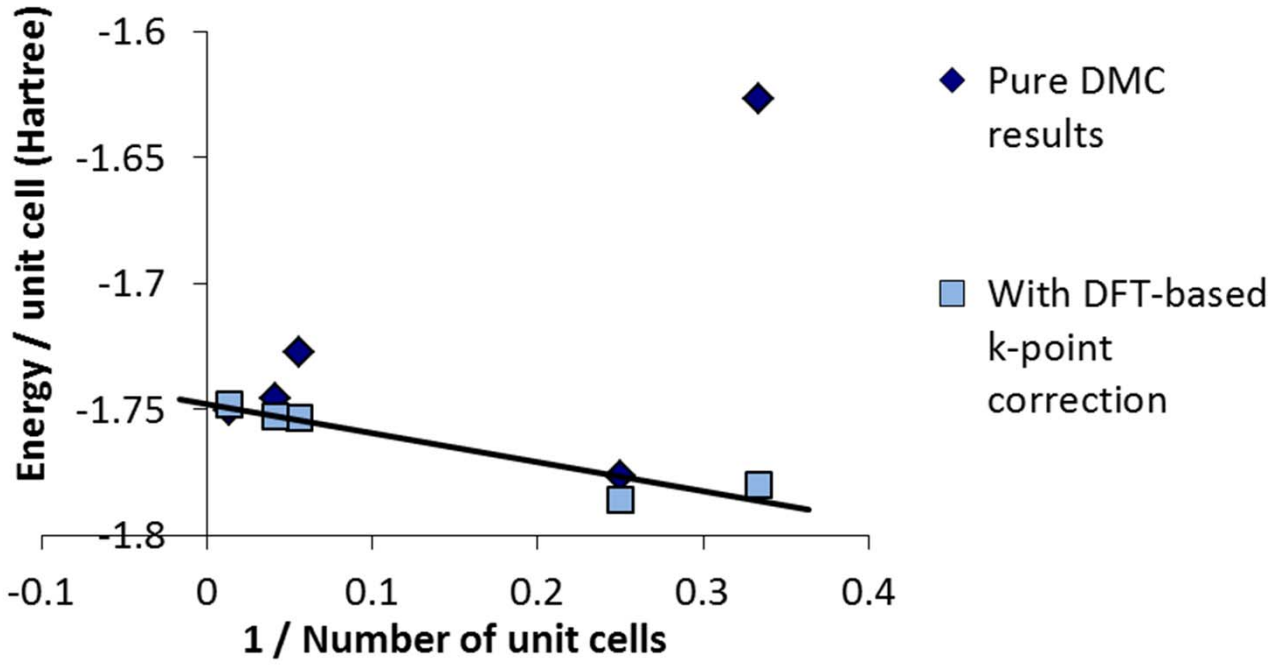
The recipe

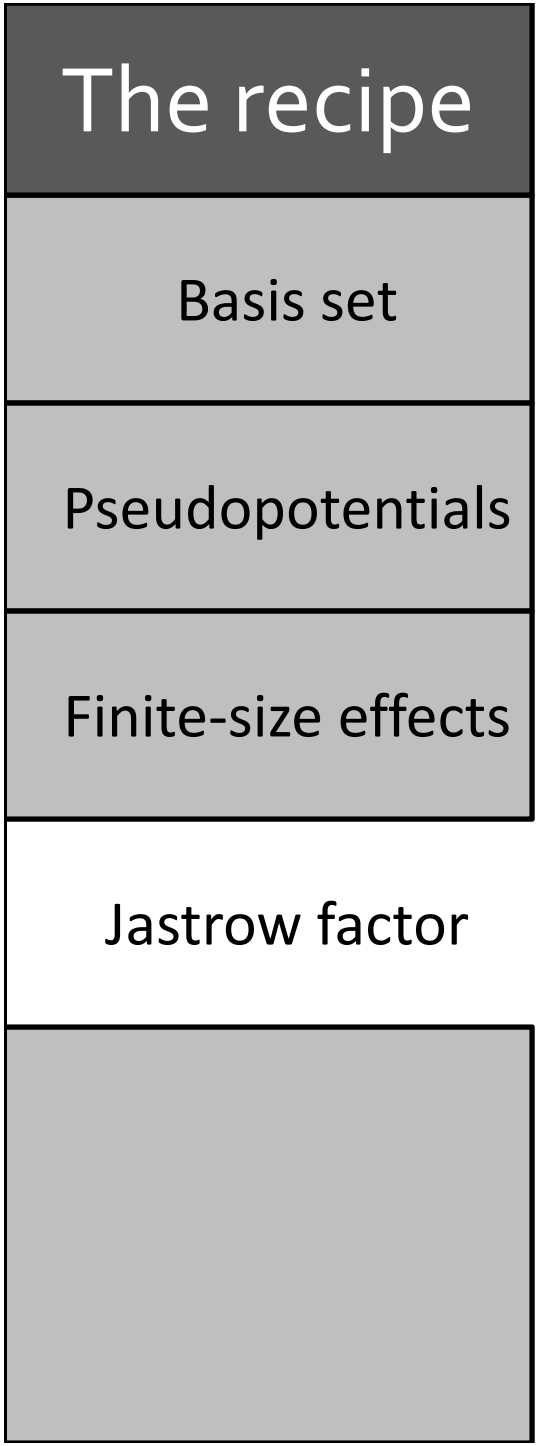
Basis set

Pseudopotentials

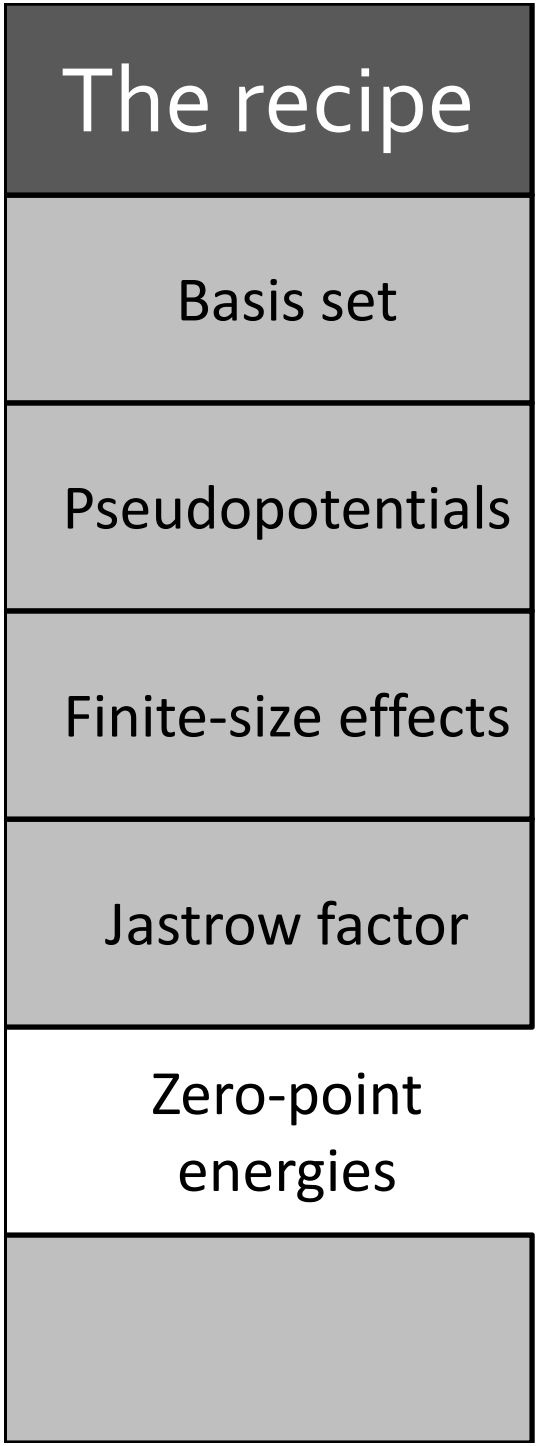
Finite-size effects

This only improves the results by about 1 kJ / mol atoms on average – possibly because we only fit through largest three cells, so adjustment size is small.





- Initialize parameters with converged parameters from smaller unit cell
- All k-points use same Jastrow factor.
- Three-body terms not included – they only help a little, and made VMC parameter optimization more difficult.

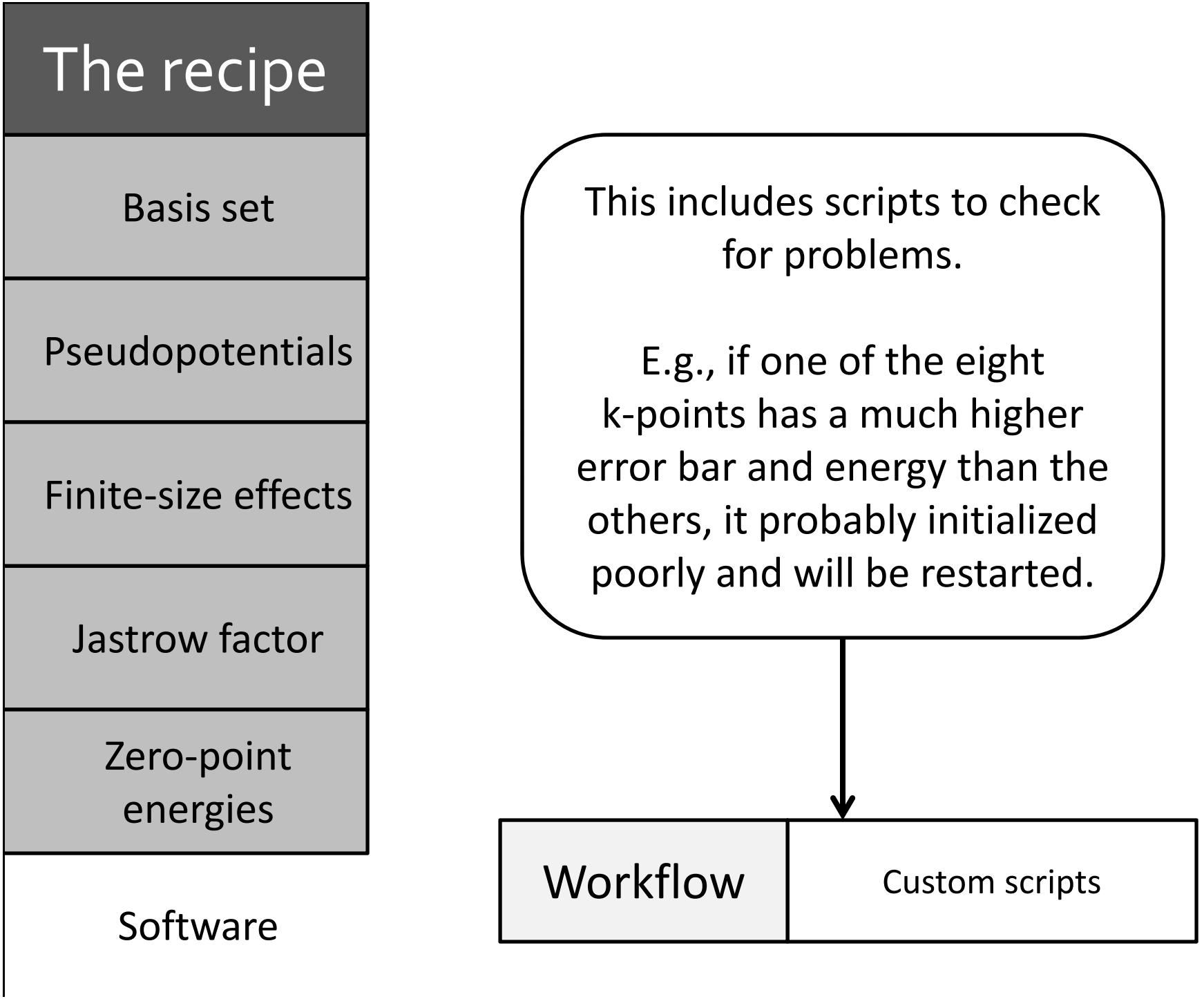


- Use DFT (PAW / VASP) to calculate zero-point energies.

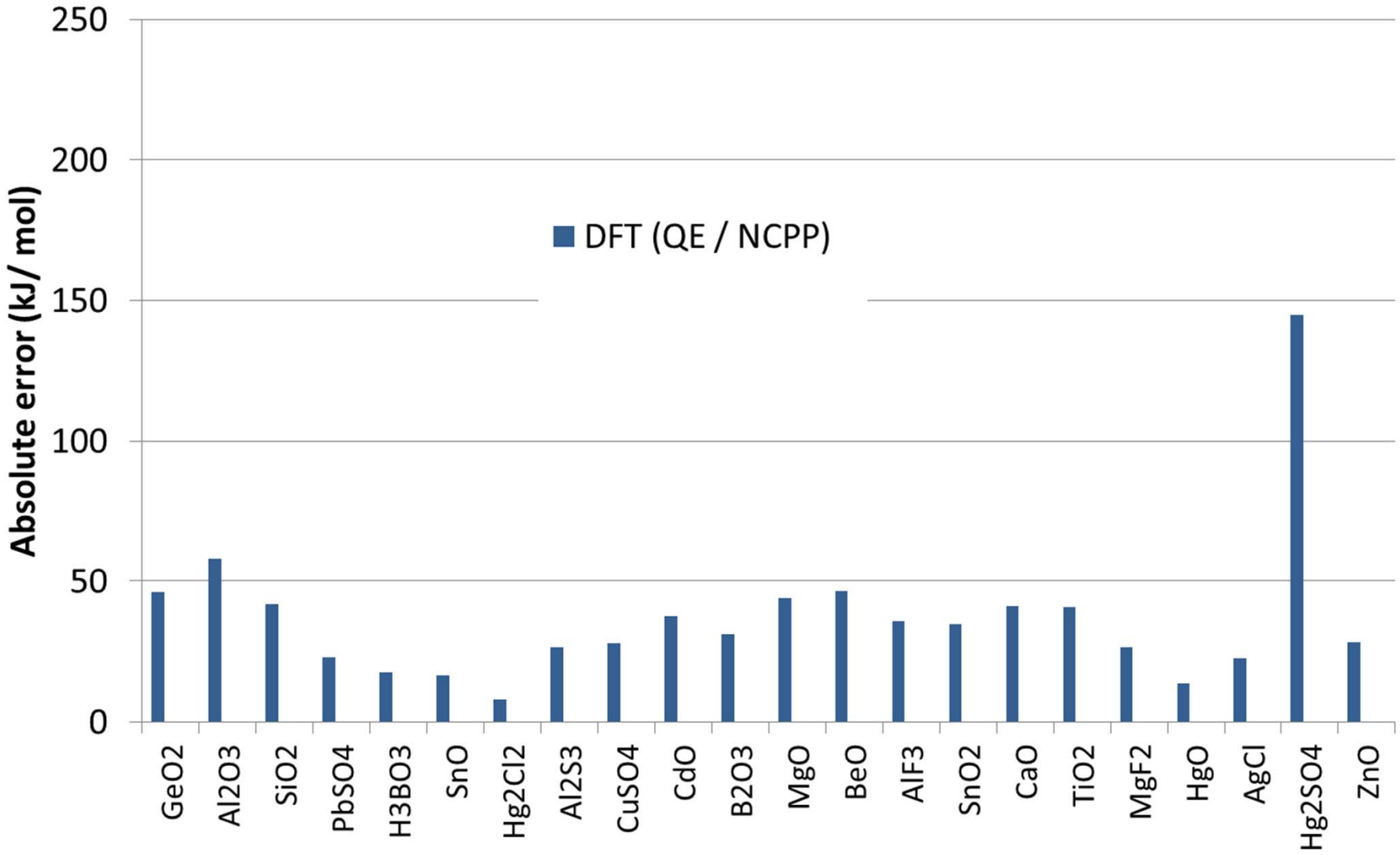
The recipe

- Basis set
- Pseudopotentials
- Finite-size effects
- Jastrow factor
- Zero-point energies
- Software

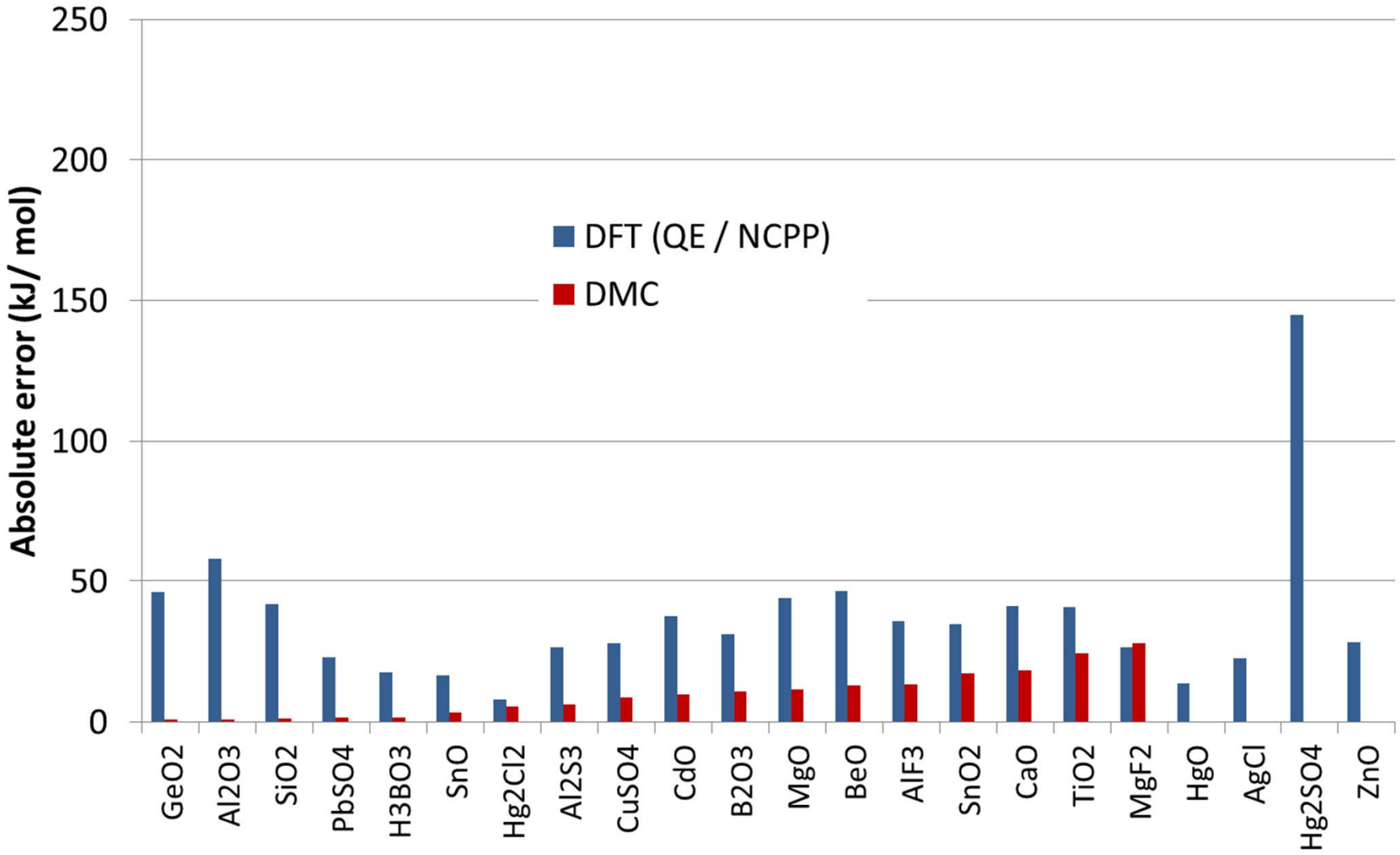
Software	
DFT	Quantum ESPRESSO
QMC	CASINO
Phonons	VASP PHONOPY
Workflow	Custom scripts


The recipe
Basis set
Pseudopotentials
Finite-size effects
Jastrow factor
Zero-point energies
Software
This includes scripts to check for problems.
E.g., if one of the eight k-points has a much higher error bar and energy than the others, it probably initialized poorly and will be restarted.
Workflow
Custom scripts

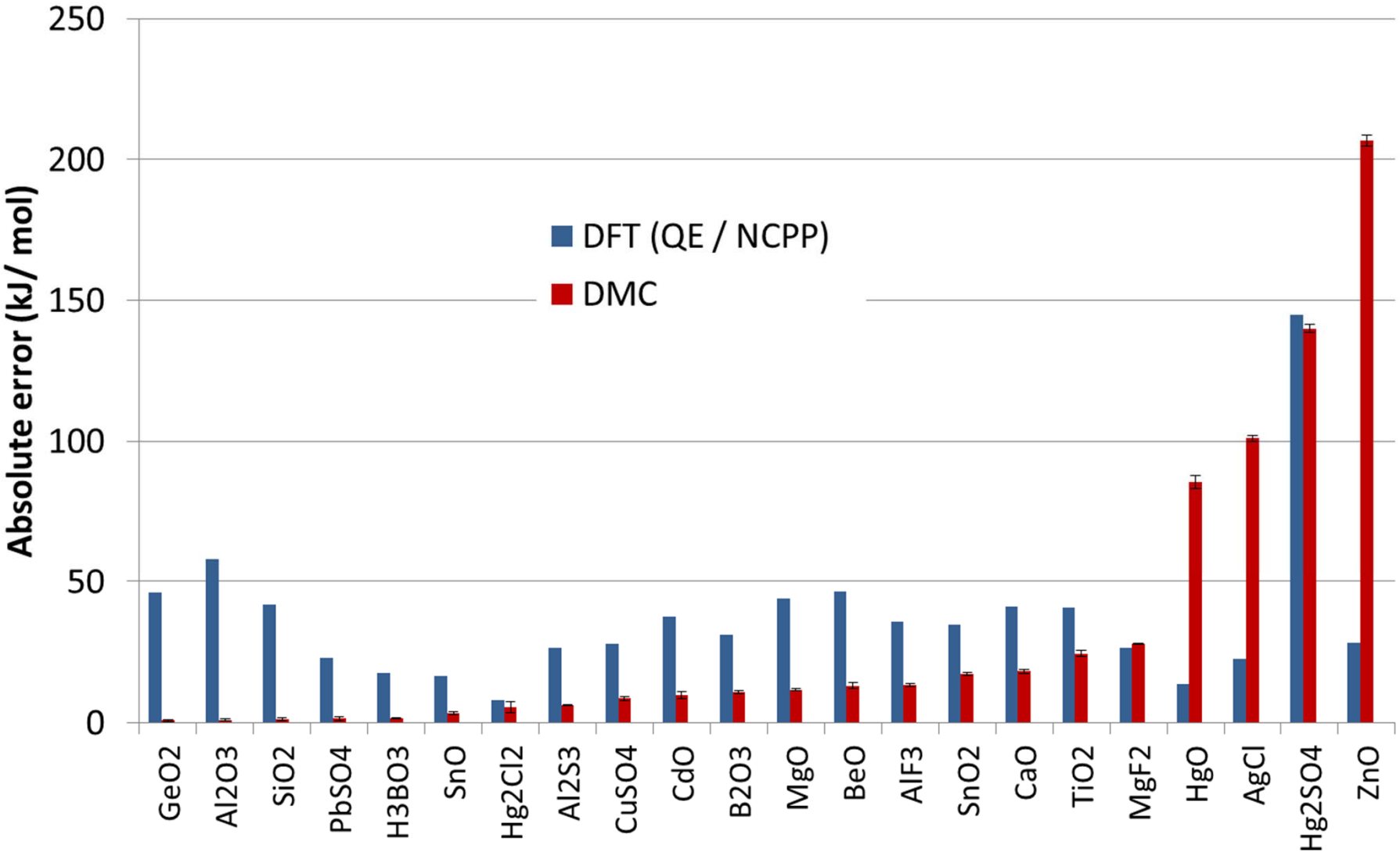
Results: DFT Errors



DMC Results


250
200
150
100
50
0
Absolute error (kJ/ mol)
DFT (QE / NCPP)
DMC
GeO2
Al2O3
SiO2
PbSO4
H3BO3
SnO
Hg2Cl2
Al2S3
CuSO4
CdO
B2O3
MgO
BeO
AlF3
SnO2
CaO
TiO2
MgF2
HgO
AgCl
Hg2SO4
ZnO

DMC Results

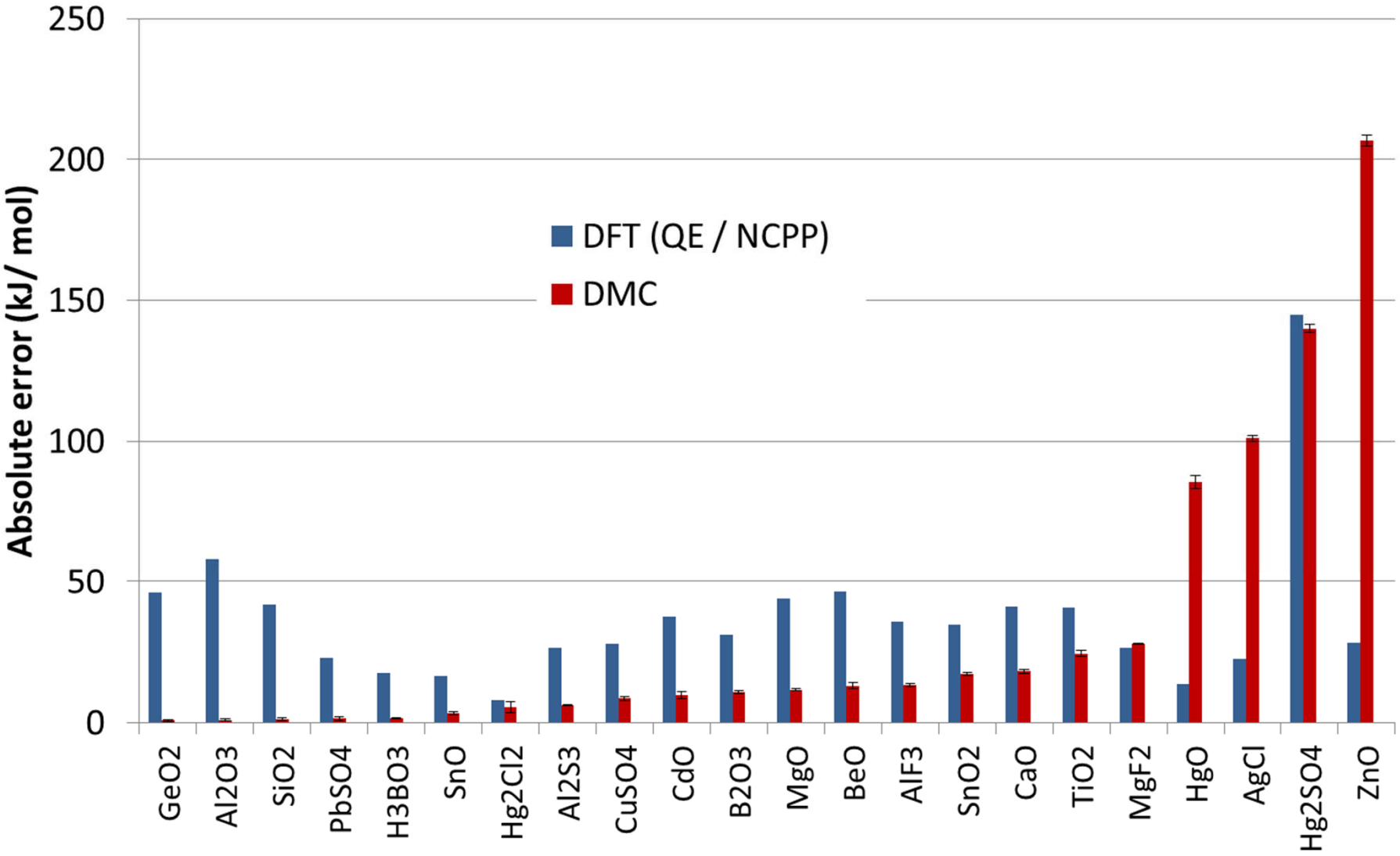

250
200
150
100
50
0
Absolute error (kJ/ mol)
DFT (QE / NCPP)
DMC
GeO2
Al2O3
SiO2
PbSO4
H3BO3
SnO
Hg2Cl2
Al2S3
CuSO4
CdO
B2O3
MgO
BeO
AlF3
SnO2
CaO
TiO2
MgF2
HgO
AgCl
Hg2SO4
ZnO

Where the problems are

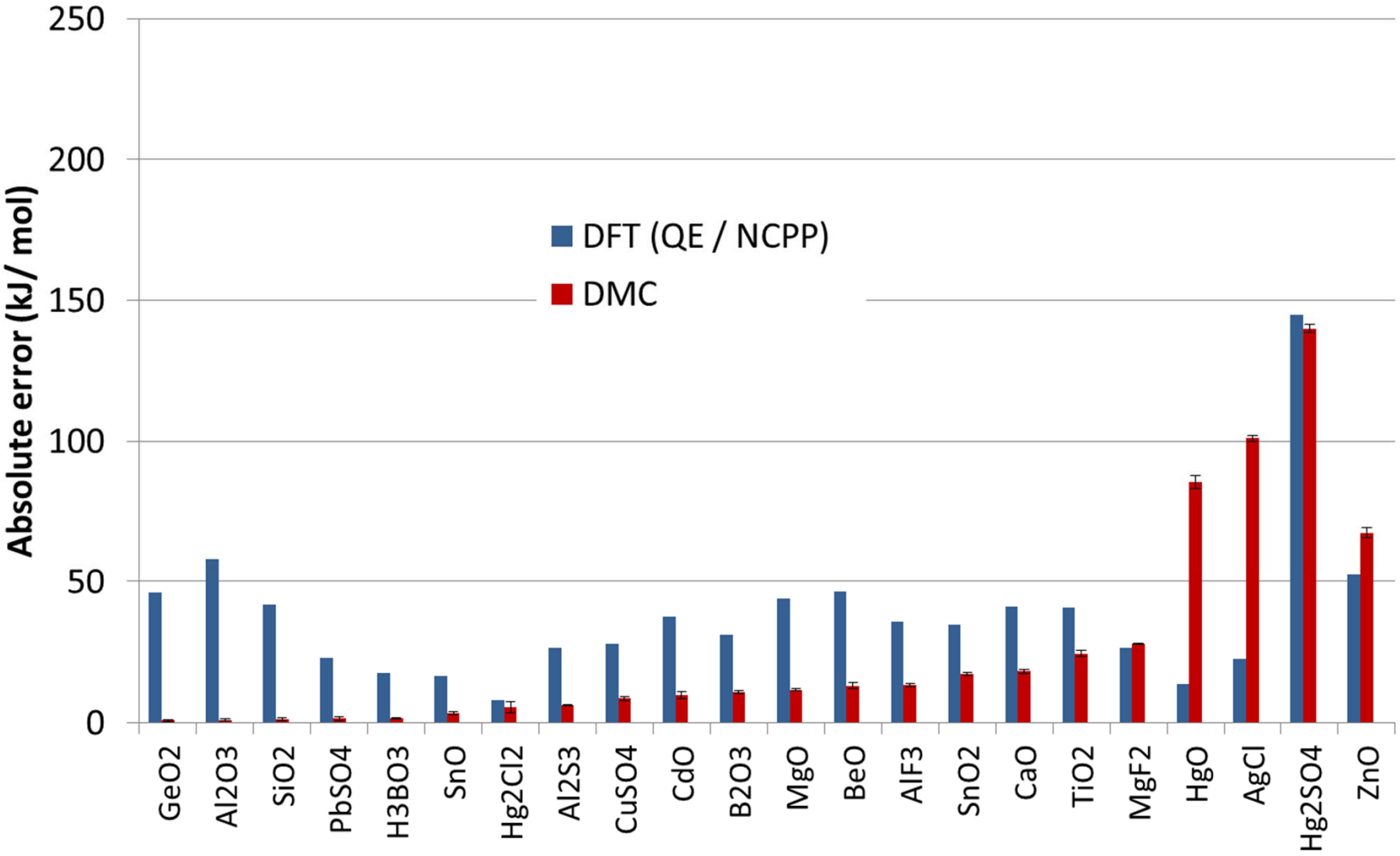
	IA																VIIIB
1	H	IIA										IIIB	IVB	VB	VIB	VIIB	He
2	Li	Be										B	C	N	O	F	Ne
3	Na	Mg	IIIA	IVA	VA	VIA	VIIA	[-------	--VIIIA--	-------]	IB	IIB	Al	Si	P	S	Cl	Ar
4	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
5	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
6	Cs	Ba		Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
7	Fr	Ra		Rf	Db	Sg	Bh	Hs	Mt	Uun				Uuq				

(6)	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
(7)	Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr

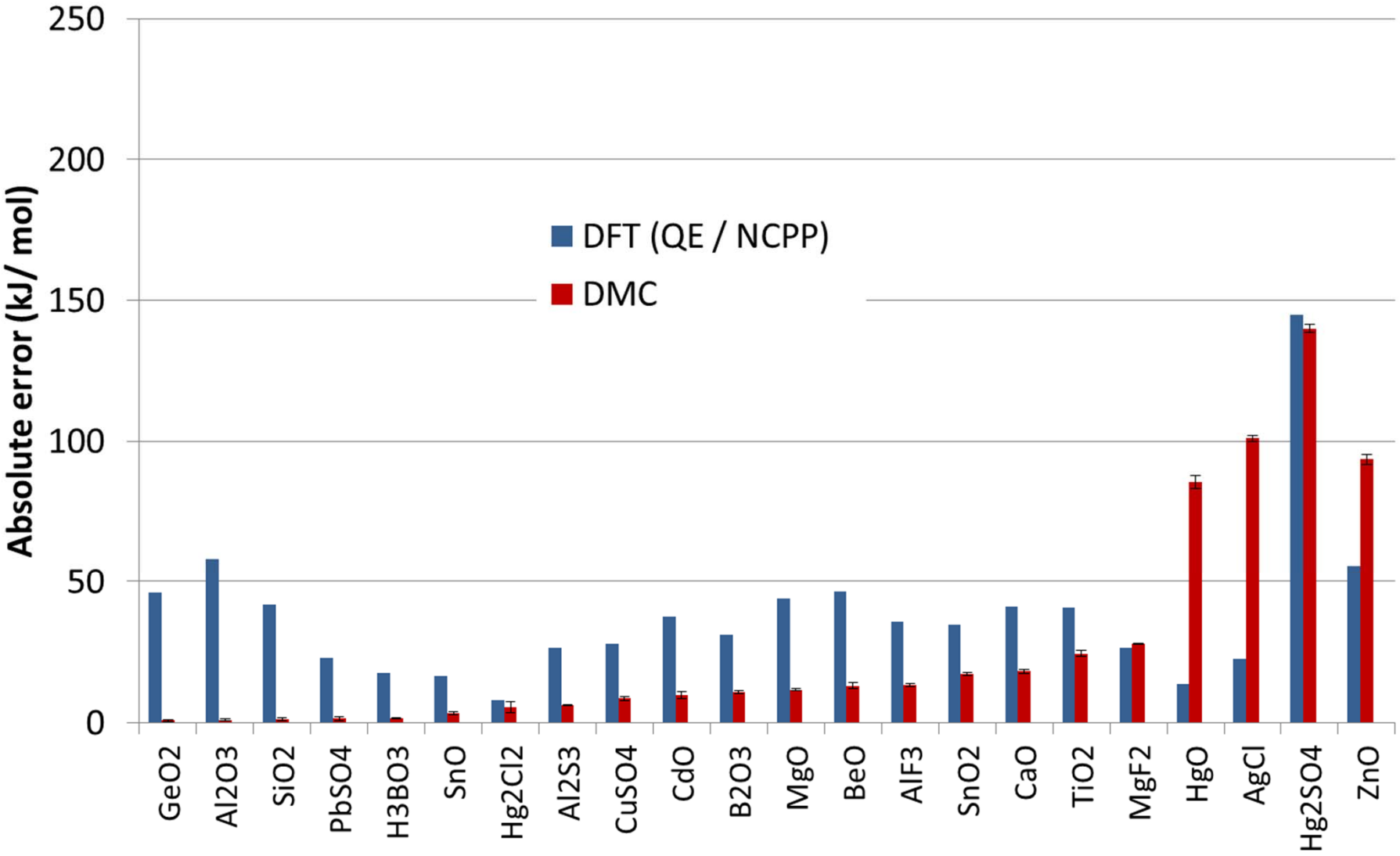
DMC Results


250
200
150
100
50
0
Absolute error (kJ/ mol)
DFT (QE / NCPP)
DMC
GeO2
Al2O3
SiO2
PbSO4
H3BO3
SnO
Hg2Cl2
Al2S3
CuSO4
CdO
B2O3
MgO
BeO
AlF3
SnO2
CaO
TiO2
MgF2
HgO
AgCl
Hg2SO4
ZnO

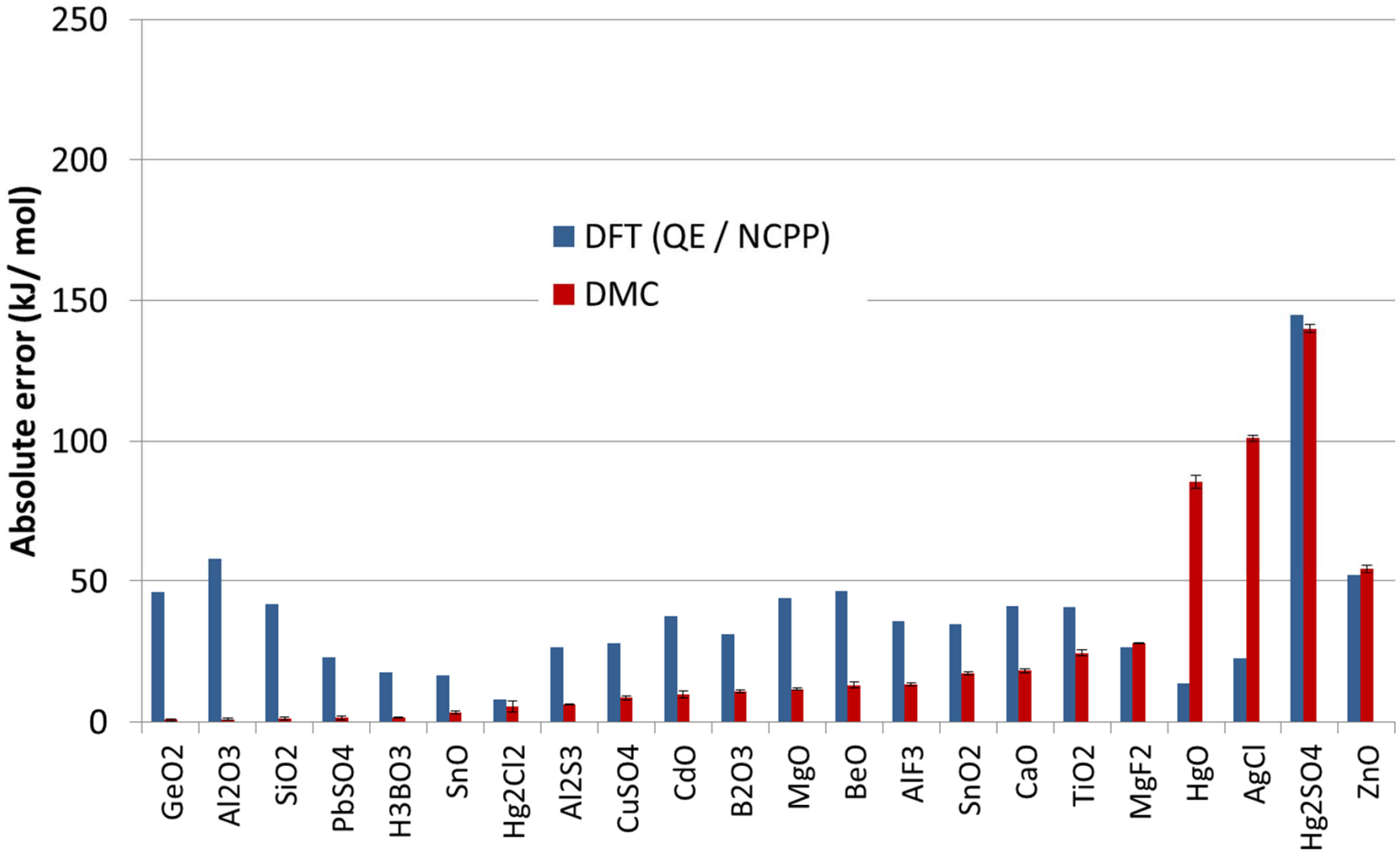
A new Zn pseudopotential based on Cd pseudopotential



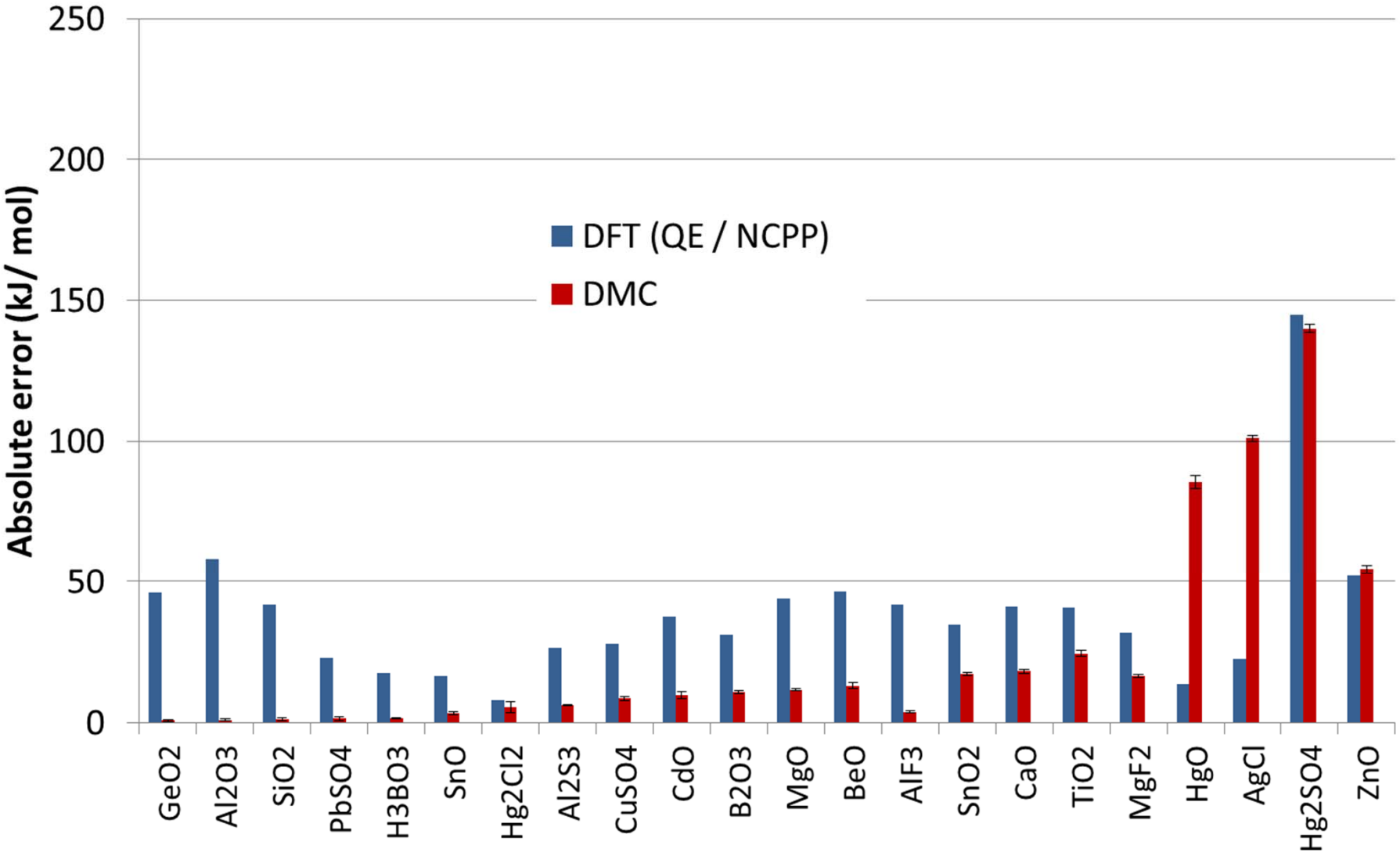
Small-core Zn pseudopotential



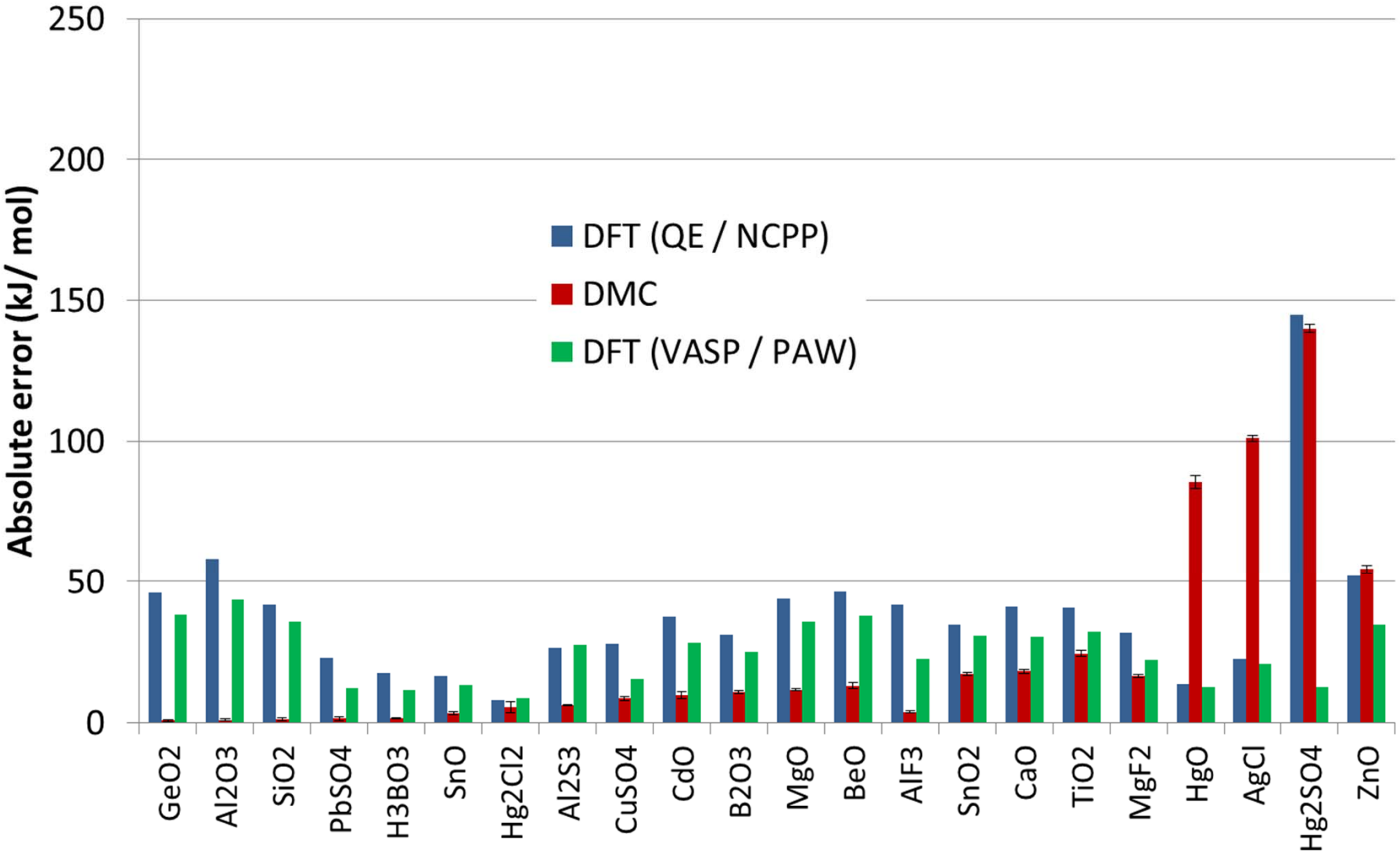
Best for Zn: ABINIT (FHI) pseudopotential



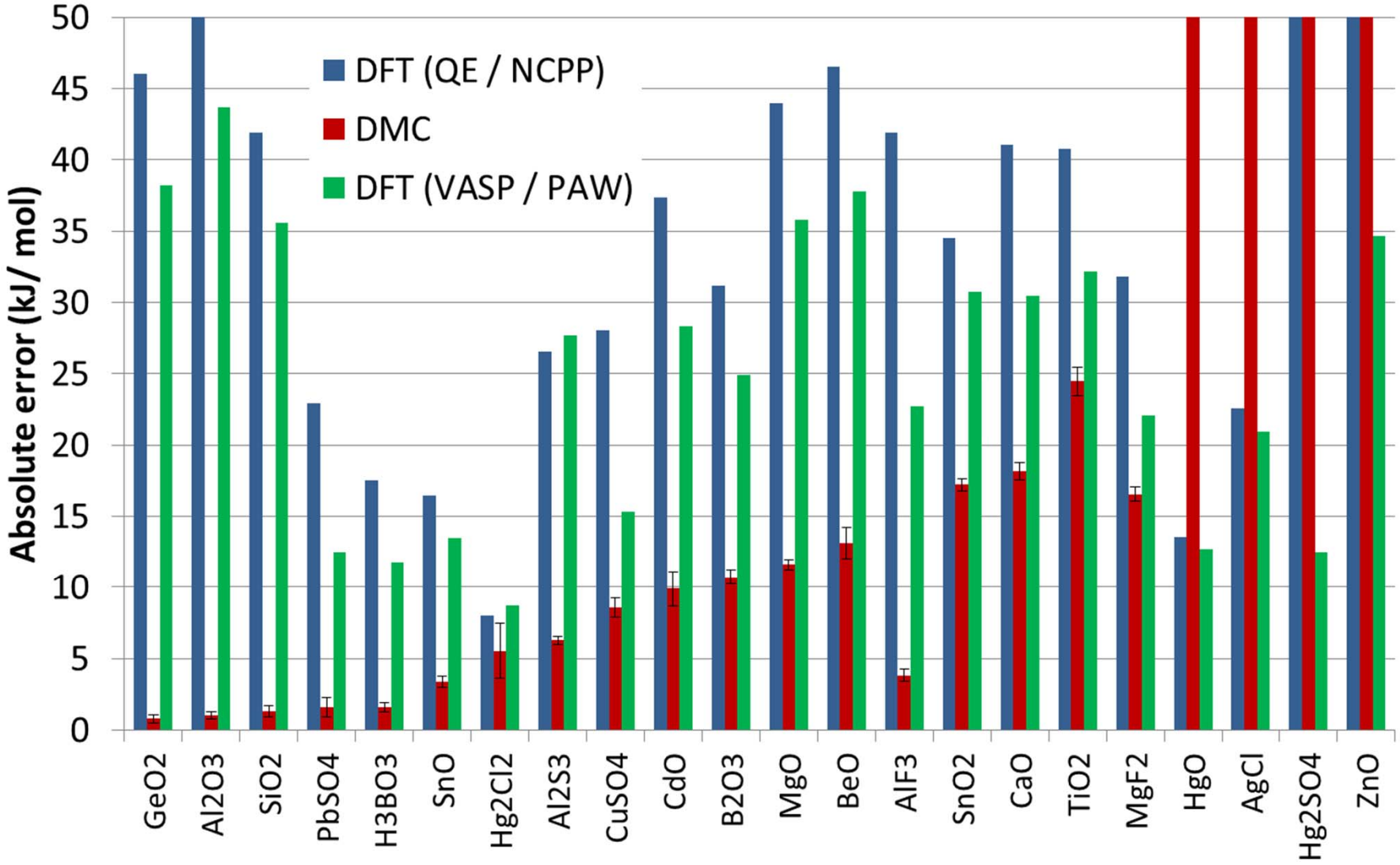
ABINIT F pseudopotential worked better as well



VASP / PAW is better than QE / NCPP



Close-up view of errors



The plan

1) Develop “recipe” for reliable QMC formation energies.

2) Generate benchmark data for important materials (e.g. each element in each common oxidation state).

3) Long term: Calculate them all.

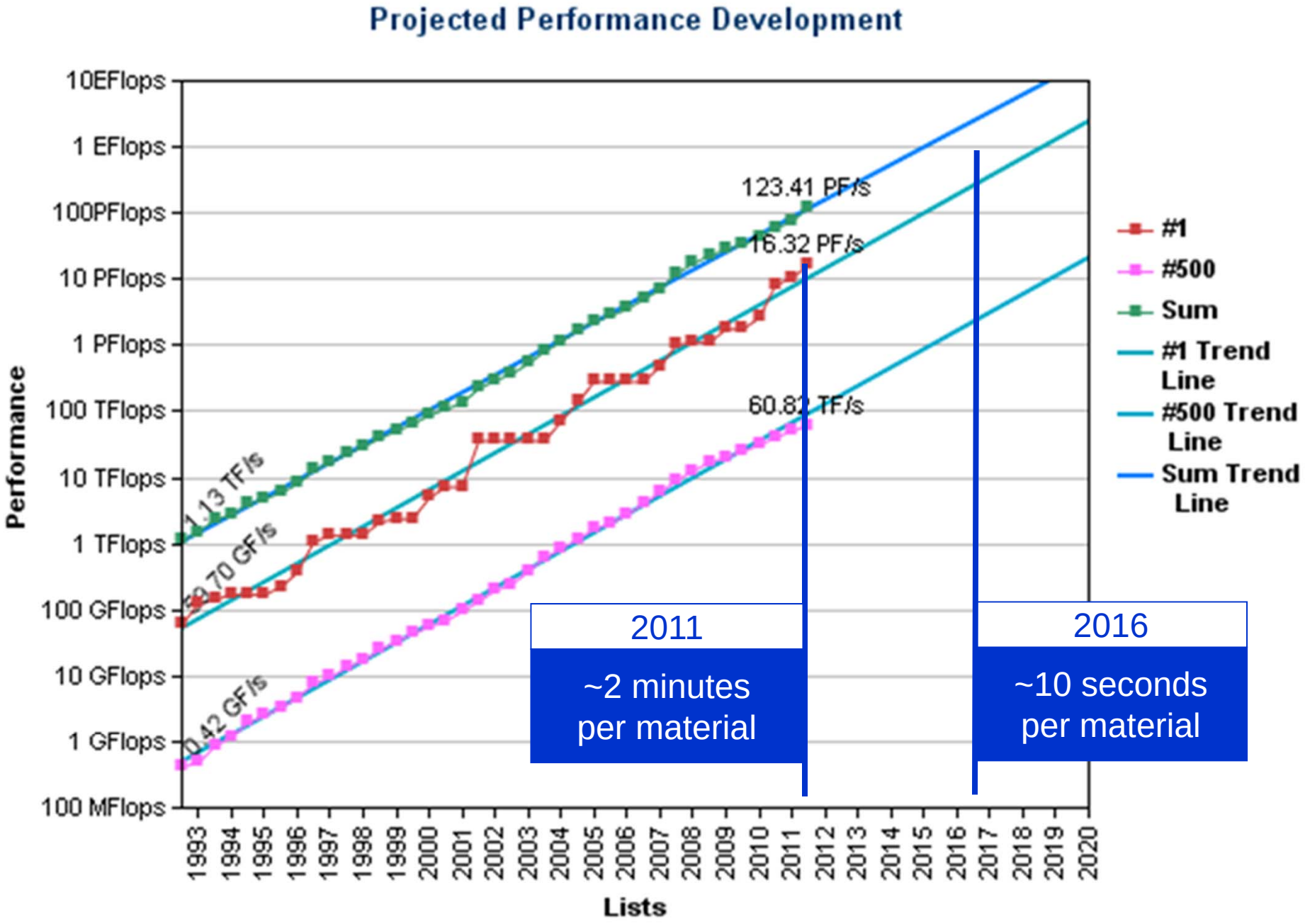
The plan

Can this be done?

3) Long term: Calculate them all.

The plan

The energies we calculated took an average of about 50,000 CPU-hours per material.


Projected Performance Development
10EFlops
1 EFlops
100PFlops
10 PFlops
1 PFlops
100 TFlops
10 TFlops
1 TFlops
100 GFlops
10 GFlops
1 GFlops
100 MFlops
Performance
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
Lists
123.41 PF/s
16.32 PF/s
60.82 TF/s
1.13 TF/s
59.70 GF/s
0.42 GF/s
#1
#500
Sum
#1 Trend Line
#500 Trend Line
Sum Trend Line
2011
~2 minutes per material
2016
~10 seconds per material

The plan

By 2016-ish, we should be able to calculate QMC energies for every known inorganic material on a single supercomputer in about a week (roughly).

Getting QMC ready for high-throughput

- Better pseudopotentials
- Simplified workflow.
- User-friendly features (e.g. stop if convergence criterion is met, even if all steps haven't been run).
- Accuracy improvements.

Discussion