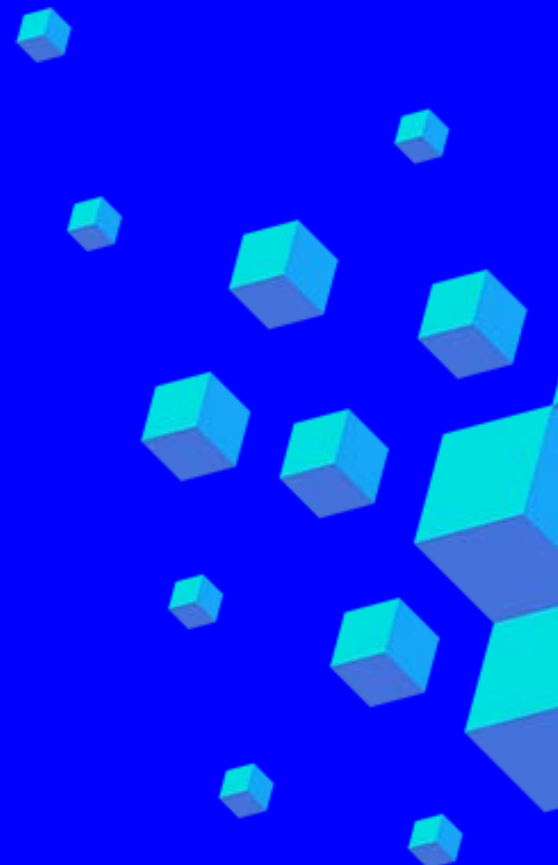


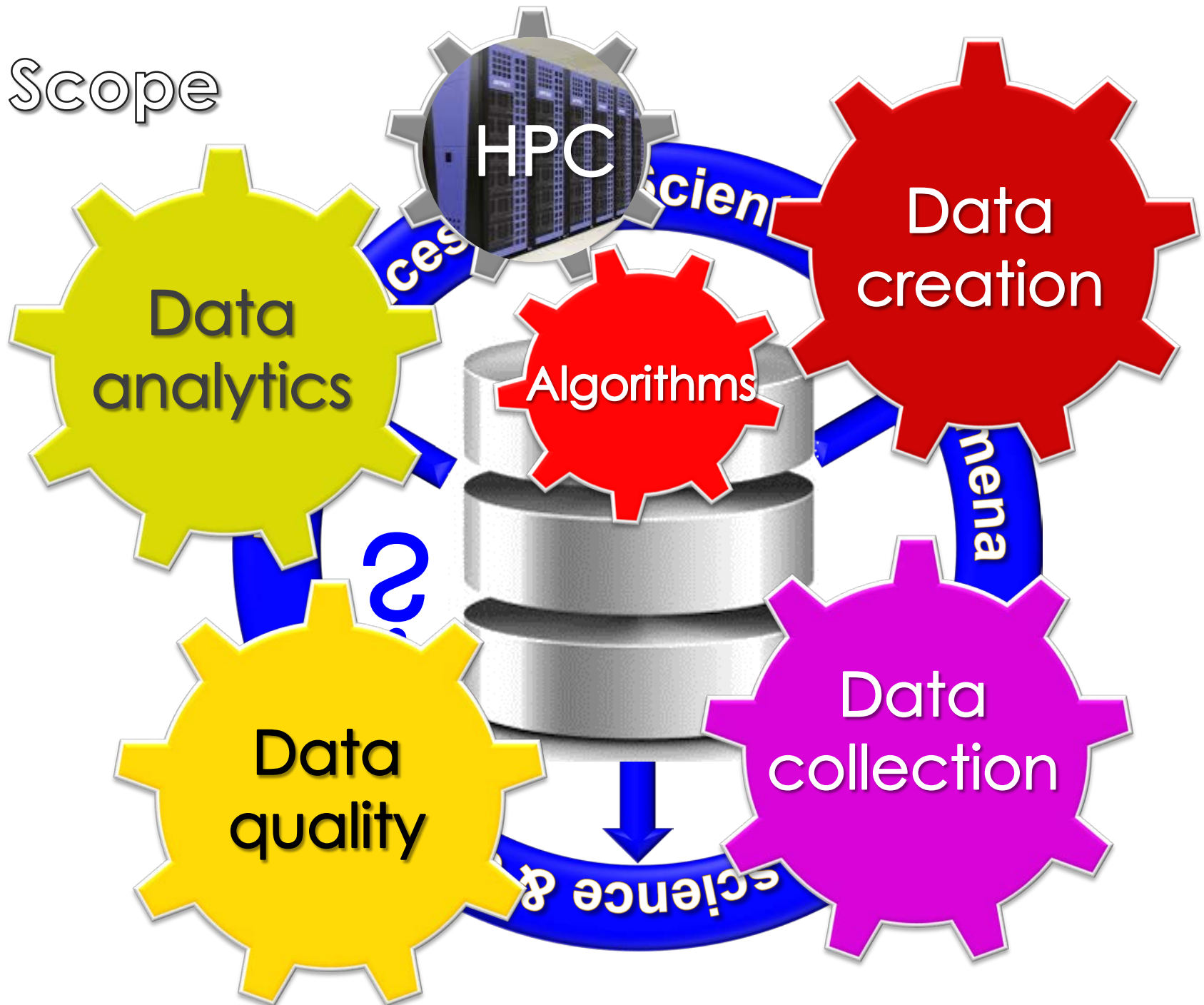
From high-performance computing to big-data analytics of materials science and back



Claudia Draxl



Scope



Materials data & their structure

Level	Properties	Methods	Size
I	Atomic positions and nuclear charges, properties of free atoms, symmetry, temperature, pressure	Input: definition of material <i>gene</i>	10 kB - 10 MB
II			10 MB - 10 TB
III	optical spectra, electrical conductivity, phonon spectra, thermal conductivity, etc.	DF perturbation theory, <i>ab initio</i> MD	1 GB - 1 TB
IV	Efficiency of solar cell, thermoelectric figure of merit, turn-over frequency of catalyst, etc. as a function of temperature and pressure	Modeling, output derived from levels I-III <i>phenotype</i>	10 kB - 1 MB

The amount of materials data produced on workstations, compute clusters, and supercomputers is growing exponentially. Most of it is thrown away

Methodology

Bottleneck: matrix diagonalization

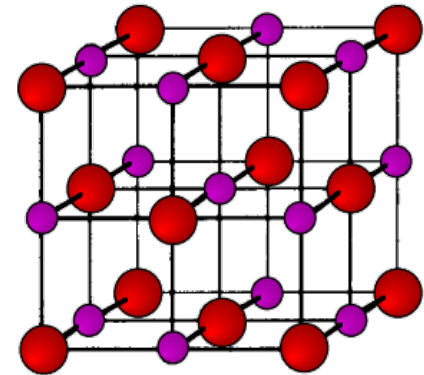
Generalized eigenvalue problem

Scaling: N^3

N ... number of atoms in the unit cell

N ... 1 - > 1000

>100 basis functions / atom



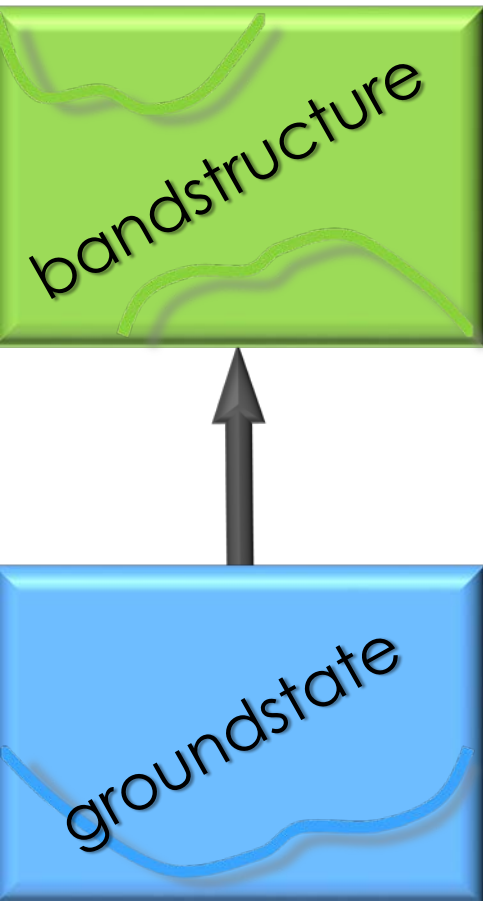
Density-functional theory

Kohn-Sham equation

$$\left[T + V_{\text{ext}}(\mathbf{r}) + V_H(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}) \right] \psi_i^{\text{KS}}(\mathbf{r}) = \epsilon_i^{\text{KS}} \psi_i^{\text{KS}}(\mathbf{r})$$



Methodology



Density-functional theory

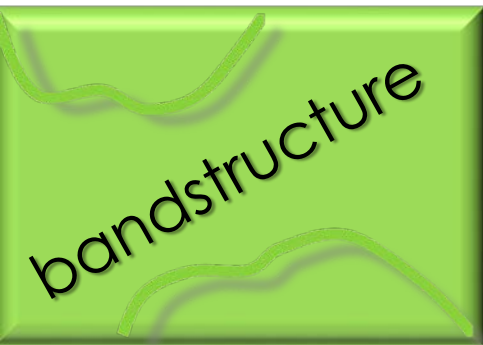
Kohn-Sham equation

$$\left[T + V_{\text{ext}}(\mathbf{r}) + V_H(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}) \right] \psi_i^{\text{KS}}(\mathbf{r}) = \epsilon_i^{\text{KS}} \psi_i^{\text{KS}}(\mathbf{r})$$

Methodology

Bottleneck: non-local operators

Scaling: $\sim N^4$



Many-body perturbation theory

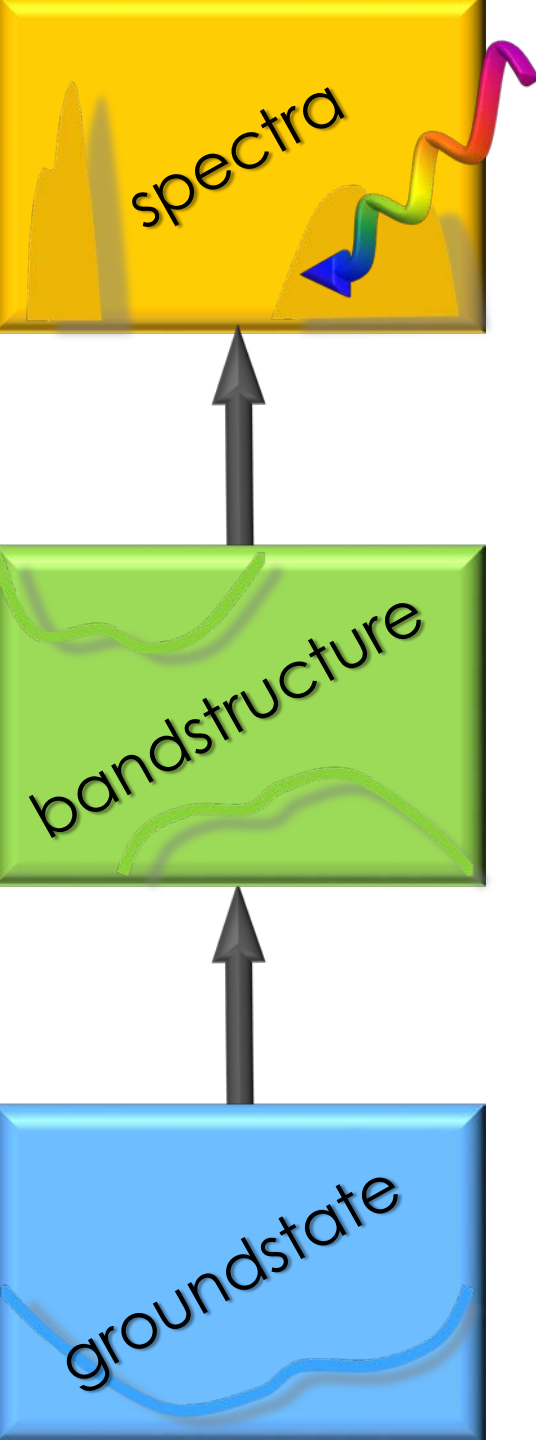
G_0W_0 approximation

$$\epsilon_{nk}^{QP} = \epsilon_{nk}^{KS} + \langle nk | \Sigma - V_{xc}^{KS} | nk \rangle$$

Density-functional theory

Kohn-Sham equation

$$\left[T + V_{\text{ext}}(\mathbf{r}) + V_H(\mathbf{r}) + V_{xc}(\mathbf{r}) \right] \psi_i^{KS}(\mathbf{r}) = \epsilon_i^{KS} \psi_i^{KS}(\mathbf{r})$$



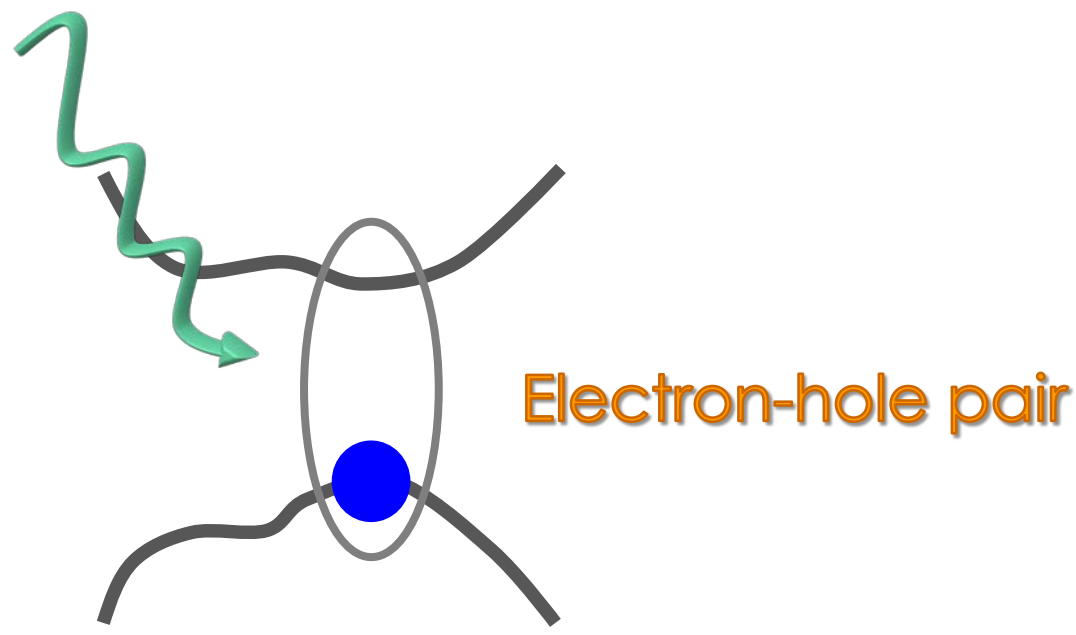
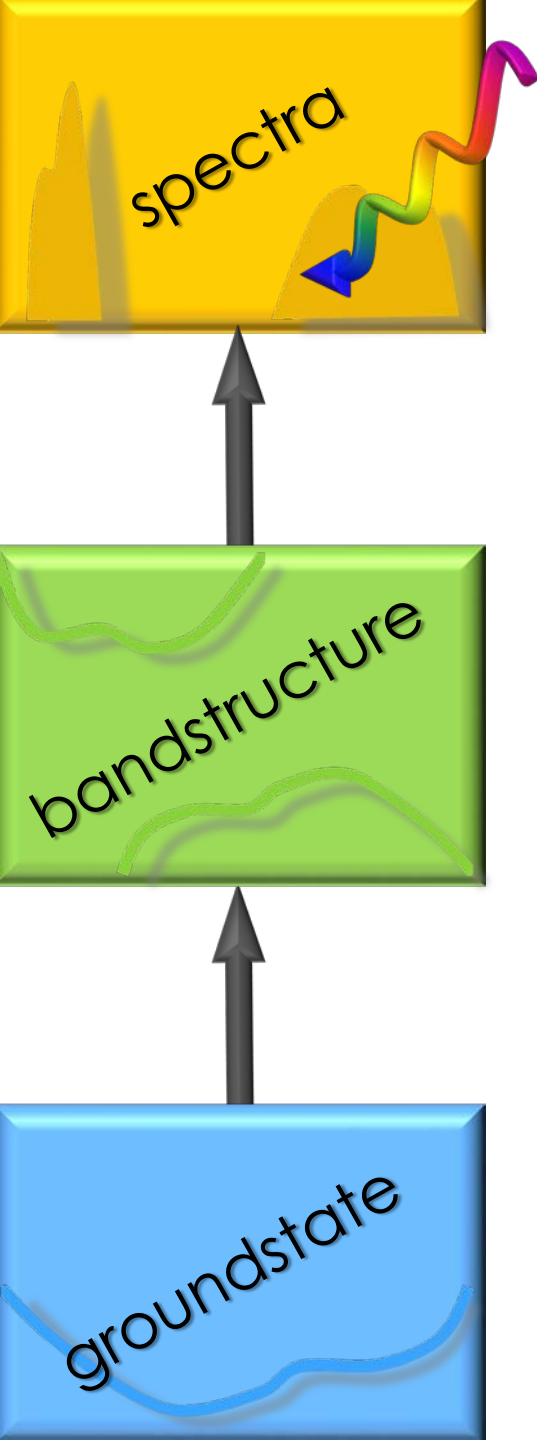
G_0W_0 approximation

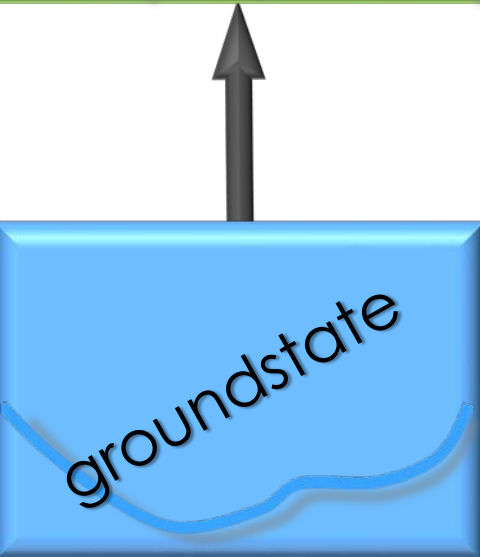
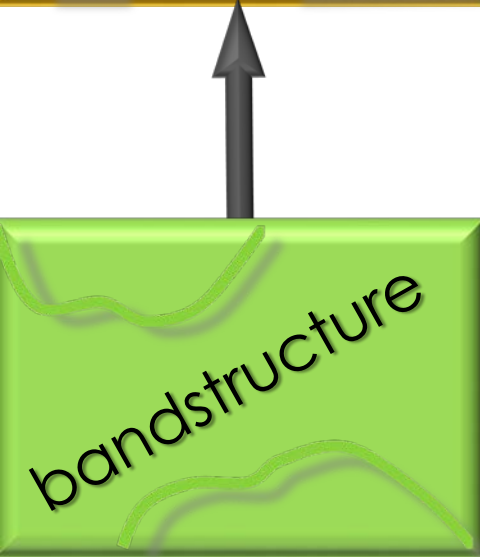
$$\epsilon_{nk}^{QP} = \epsilon_{nk}^{KS} + \langle nk | \Sigma - V_{xc}^{KS} | nk \rangle$$

Density-functional theory

Kohn-Sham equation

$$\left[T + V_{\text{ext}}(\mathbf{r}) + V_H(\mathbf{r}) + V_{xc}(\mathbf{r}) \right] \psi_i^{KS}(\mathbf{r}) = \epsilon_i^{KS} \psi_i^{KS}(\mathbf{r})$$





Bethe-Salpeter equation

$$\left[H_{el} + H_{hole} + H_{el-hole} \right] A_{\lambda} = E_{\lambda} A_{\lambda}$$

Many-body perturbation theory

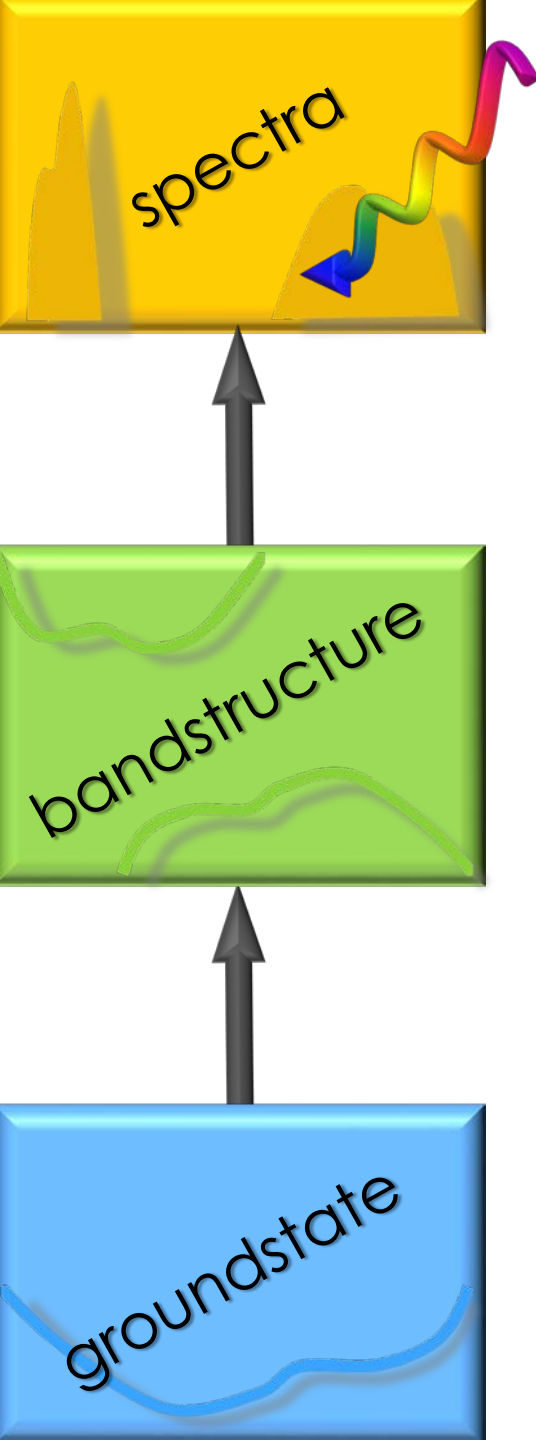
G_0W_0 approximation

$$\epsilon_{nk}^{QP} = \epsilon_{nk}^{KS} + \left\langle nk \left| \Sigma - V_{xc}^{KS} \right| nk \right\rangle$$

Density-functional theory

Kohn-Sham equation

$$\left[T + V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) + V_{xc}(\mathbf{r}) \right] \psi_i^{KS}(\mathbf{r}) = \epsilon_i^{KS} \psi_i^{KS}(\mathbf{r})$$



Bethe-Salpeter equation

$$\left[H_{el} + H_{hole} + H_{el-hole} \right] A_{\lambda} = E_{\lambda} A_{\lambda}$$

Many-body perturbation theory

$$\sum_{v'c'k'} H_{vck,v'c'k'}^{e-h} A_{v'c'k'}^{\lambda} = E_{\lambda} A_{vck}^{\lambda}$$

Spectroscopy

Bottleneck: matrix size

10 000 - 100 000 for small systems

Non-local operators

Thermoelectrics

Direct conversion between
difference and electric

End



100 MW

Thermoelectrics



Waste-heat recovery

What makes a good thermoelectric?

Figure of merit ZT with

$$Z = \frac{\sigma S^2}{\kappa_{el}^0 + \kappa_{ph}}$$

S Seebeck coefficient
 σ electronic conductivity
 κ thermal conductivity

Current values: $Z = 0.6 - 1.5$

Profitable applications: $Z > 2$

Problem:

High electrical conductivity σ and low thermal conductivity κ is required at the same time

Materials data & their structure

Level	Properties	Methods	Size
I	Atomic positions and nuclear charges, properties of free atoms, symmetry, temperature, pressure	Input: definition of material <i>gene</i>	10 kB - 1 MB
II	Total energy, electron density, potential, wavefunctions, atomic forces, optimized geometry, elastic constants, etc.	Density-functional theory (DFT) and <i>ab initio</i> molecular dynamics (MD)	10 MB - 10 TB
III	Excitation energies, dielectric screening, matrix elements of Coulomb interaction, etc. optical spectra, electrical conductivity, phonon spectra, thermal conductivity, etc.	Many-body perturbation theory (MBPT), DF perturbation theory, <i>ab initio</i> MD	1 GB - 1 TB
IV	Efficiency of solar cell, thermoelectric figure of merit, turn-over frequency of catalyst, etc. as a function of temperature and pressure	Modeling, output derived from levels I-III <i>phenotype</i>	10 kB - 1 MB

Novel Materials Discovery

<http://nomad-repository.eu>

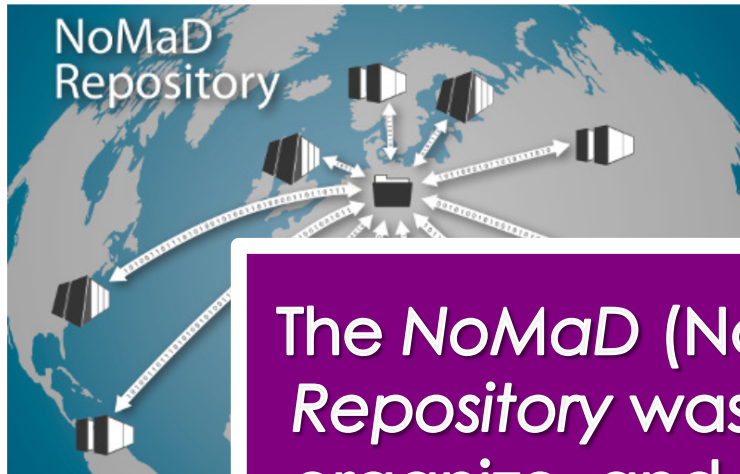


The *NoMaD* Repository

Insight by sharing

[Home](#) [NoMaD Team](#) [Why sharing?](#) [DOIs](#) [Terms](#) [FAQ](#) [Other repositories](#) [Upload your files](#) [Search and download](#) [Contact us](#)

Welcome to the NoMaD Repository



The *NoMaD* Repository
[youtube](#) to see our

The *NoMaD* (Novel Materials Discovery) Repository was established to host, organize, and share materials data.

NoMaD copes with the increasing demand and requirement of storing scientific data and making them available for longer periods. Rules of good scientific

The *NoMaD* (Novel Materials Discovery) Repository was established to host, organize, and share materials data.

News

Currently, the NoMaD Repository contains

3 314 840

entries.

[Upload to NoMaD from MedaA application ... more](#)

[or related conferences workshops.](#)

[al Support](#)



Kristian Thygesen
DTU Lyngby



Ciaran Clissman
Pintail Dublin



Arndt Bode
LRZ Munich



Jose Maria Cela
BSC Barcelona



Alessandro De Vita
Kings College London



Matthias Scheffler
FHI Berlin



Claudia Draxl
HU Berlin



Angel Rubio
MPSD Hamburg



Risto Nieminen
Aalto Univ. Helsinki



Kimmo Koski
CSC Helsinki



Francesc Illas
Univ. Barcelona

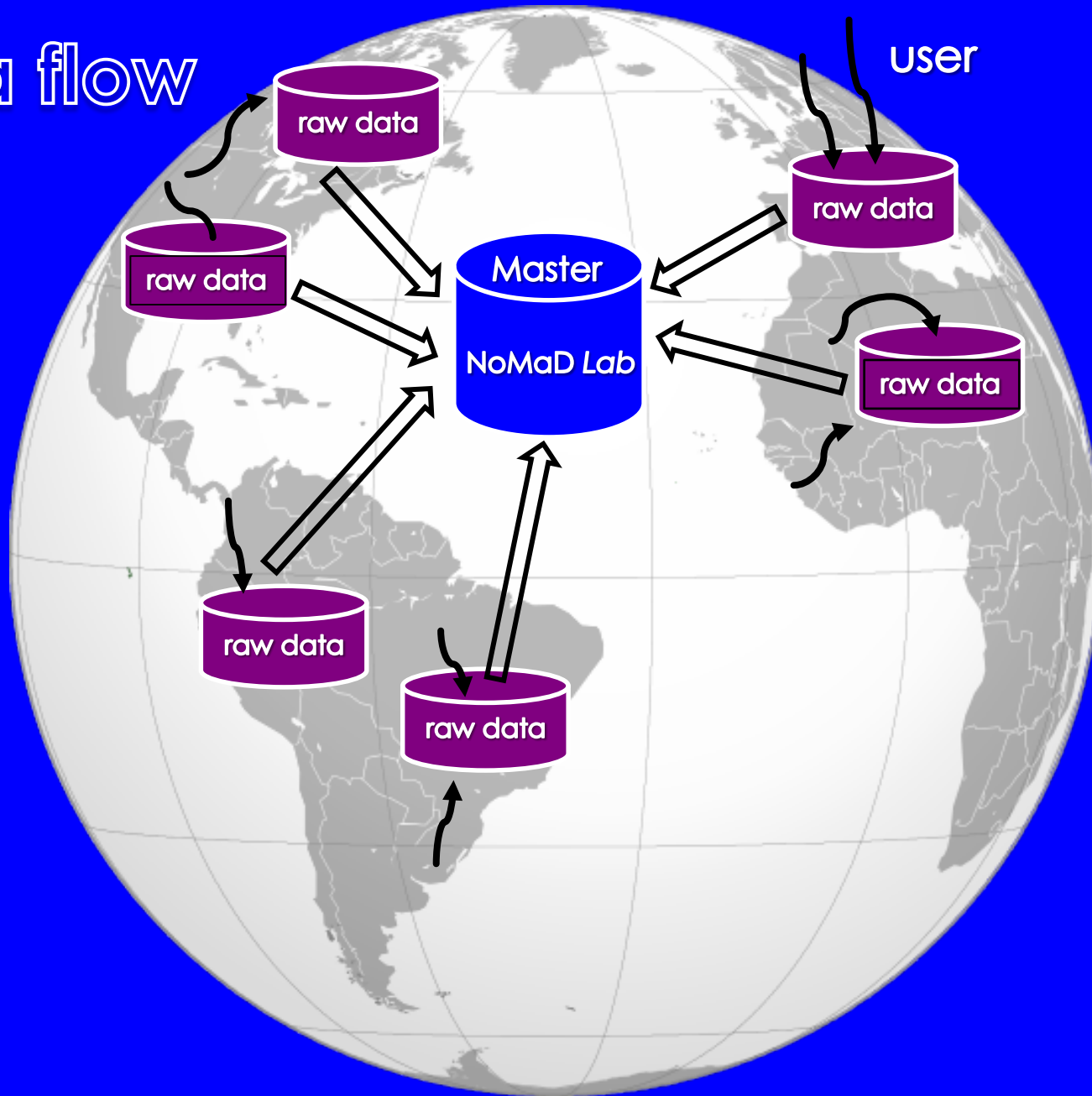


Stefan Heinzel
MPSCD Garching



Daan Frenkel
Univ. Cambridge

Data flow



NOMAD Laboratory



Existing resources

Code-dependent data



Data conversion

Give access to the vast
amount of materials data
computed worldwide

Big-data
analytics

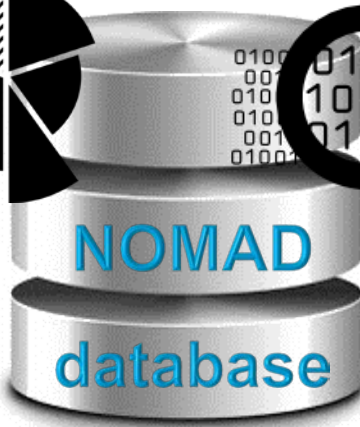


Materials
encyclopedia

0100 0100 11001
001 0010
010 001
010 10010 1001
001 01001 0010
01001 01001



NOMAD
database



HPC expertise
& hardware



Visualization

Data conversion

How to make data comparable?

NOMAD supports ~40 different computer codes

Common representation for various quantities

E.g., pseudopotentials vs all-electron methods

Evaluate error bars

Different functionals, force fields, ...

Metadata

Generic and code-specific

<https://nomad-coe.eu/index.php?page=nomad-meta-info>

Delta factors

RESEARCH ARTICLE

DFT METHODS

Reproducibility in density functional theory calculations of solids

Kurt Lejaeghere,^{1*} Gustav Bihlmayer,² Torbjörn Björkman,^{3,4} Peter Blaha,⁵ Stefan Blügel,² Volker Blum,⁶ Damien Caliste,^{7,8} Ivano E. Castelli,⁹ Stewart J. Clark,¹⁰ Andrea Dal Corso,¹¹ Stefano de Gironcoli,¹¹ Thierry Deutsch,^{7,8} John Kay Dewhurst,¹² Igor Di Marco,¹³ Claudia Draxl,^{14,15} Marcin Dułak,¹⁶ Olle Eriksson,¹³ José A. Flores-Livas,¹² Kevin F. Garrity,¹⁷ Luigi Genovese,^{7,8} Paolo Giannozzi,¹⁸ Matteo Giantomassi,¹⁹ Stefan Goedecker,²⁰ Xavier Gonze,¹⁹ Oscar Grånäs,^{13,21} E. K. U. Gross,¹² Andris Gulans,^{14,15} François Gygi,²² D. R. Hamann,^{23,24} Phil J. Hasnip,²⁵ N. A. W. Holzwarth,²⁶ Diana Iuşan,¹³ Dominik B. Jochym,²⁷ François Jollet,²⁸ Daniel Jones,²⁹ Georg Kresse,³⁰ Klaus Koepernik,^{31,32} Emine Küçükbenli,^{9,11} Yaroslav O. Kyashnin,¹³ Inka L. M. Locht,^{13,33} Sven Lubeck,¹⁴ Martijn Marsman,³⁰ Nicola Marzari,⁹ Ulrike Nitzsche,³¹ Lars Nordström,¹³ Taisuke Ozaki,³⁴ Lorenzo Paulatto,³⁵ Chris J. Pickard,³⁶ Ward Poelmans,^{1,37} Matt J. L. Probert,²⁵ Keith Refson,^{38,39} Manuel Richter,^{31,32} Gian-Marco Rignanese,¹⁹



Delta factors

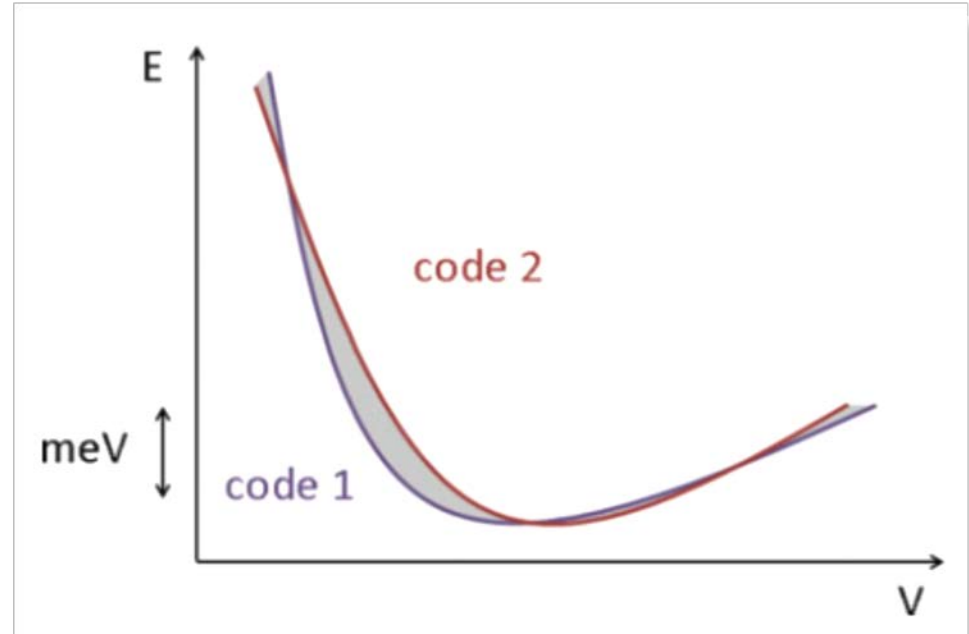
Compute $E(V)$ using PBE

Fit to the Birch-Murnaghan equation of state

Compare with other codes / method

Quality factor Δ

$$\Delta = \left\langle \sqrt{\frac{\int \Delta E^2(V) dV}{\Delta V}} \right\rangle$$






RESEARCH ARTICLE SUMMARY

DFT METHODS






Reproducibility in density functional theory calculations of solids

K. Lejaeghere et al.,
Science **351**, aad3000 (2016).

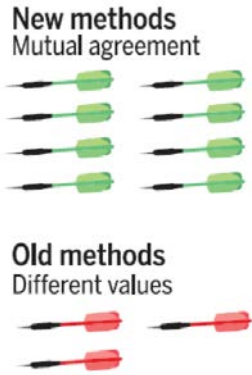
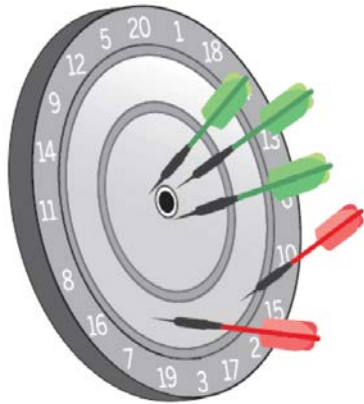
Delta factors

Code	Version	Basis	Electron treatment	Δ -value	Authors
Exciting	development version	LAPW+xlo	all-electron	0 meV/atom	Exciting [10,16] 
FHI-aims	081213	tier2 numerical orbitals	all-electron (relativistic atomic_zora scalar)	0.1 meV/atom	ASE [2,16] 
WIEN2k	13.1	LAPW/APW+lo	all-electron	0.2 meV/atom	S. Cottenier [16] 
FHI-aims	081213	tier2 numerical orbitals	all-electron (relativistic zora scalar 1e-12)	0.3 meV/atom	ASE [2] 
Quantum ESPRESSO	5.1	plane waves	SSSP Accuracy (mixed NC/US/PAW potential library)	0.3 meV/atom	QuantumESPRESSO [12,16] 

Delta factors

Code	Version	Basis	Electron treatment	Δ -value	Authors
Exciting	development version	LAPW+xlo	all-electron	0 meV/atom	Exciting [10,16] 
FHI-aims	081213	tier2 numerical orbitals	all-electron (relativistic atomic_zora scalar)	0.1 meV/atom	ASE [2,16] 
WIEN2k	13.1	LAPW/APW+lo	all-electron	0.2 meV/atom	S. Cottenier [16] 
FHI-aims	081213	tier2 numerical orbitals	all-electron (relativistic zora scalar 1e-12)	0.3 meV/atom	ASE [2] 
Quantum ESPRESSO	5.1	plane waves	SSSP Accuracy (mixed NC/US/PAW potential library)	0.3 meV/atom	QuantumESPRESSO [12,16] 

Delta factors

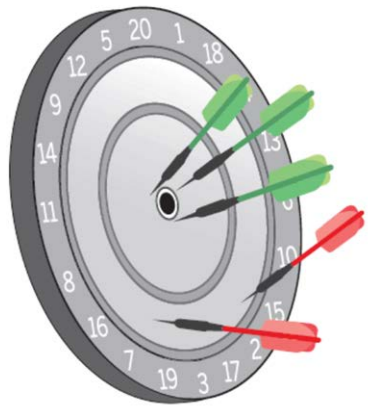


exciting
FHI-AIMS/tier2
FLEUR
FPLO/T+F+S
RSpt
WIEN2k/acc
VASP2011/VASP

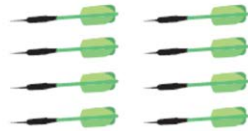
O₂

exciting		0.1	0.4	1.8	0.7	0.6	7.7
FHI-AIMS/tier2	0.1		0.3	1.7	0.6	0.5	7.6
FLEUR	0.4	0.3		1.4	0.3	0.3	7.2
FPLO/T+F+s	1.8	1.7	1.4		1.1	1.2	5.8
RSpt	0.7	0.6	0.3	1.1		0.1	6.9
WIEN2k/acc	0.6	0.5	0.3	1.2	0.1		7.1
VASP2011/VASP	7.7	7.6	7.2	5.8	6.9	7.1	

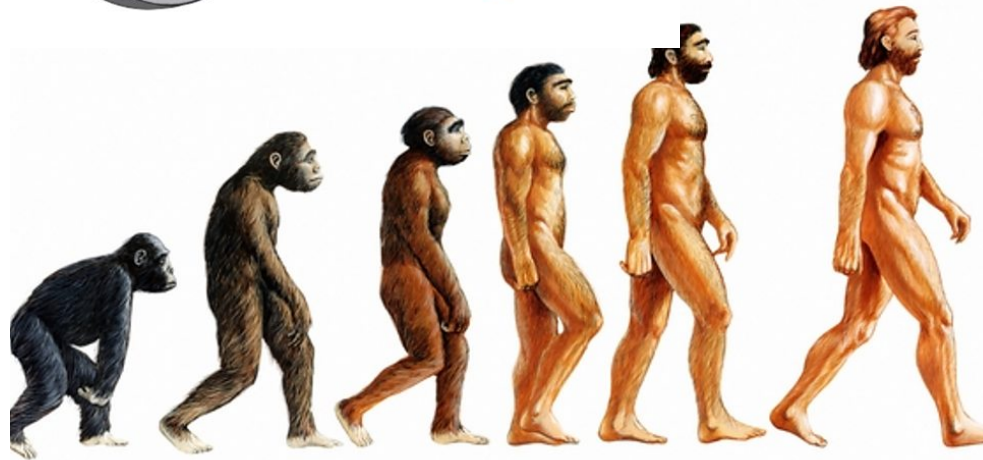
Delta factors



New methods
Mutual agreement



Old methods
Different values



	year	$\langle \Delta \rangle$ vs AE
JTH01/ABINIT	2013	1.1
JTH02/ABINIT	2014	0.6
Vdb/CASTEP	1998	6.5
OTFG7/CASTEP	2013	2.6
OTFG9/CASTEP	2015	0.7
GPAW06/GPAW	2010	3.6
GPAW09/GPAW	2012	1.6
PSlib031/QE	2013	1.7
PSlib100/QE	2013	1.0
VASP2007/VASP	2007	2.0
VASP2012/VASP	2012	0.8
VASPGW2015/VASP	2015	0.6

This is all great

... but just the beginning

What about other systems?

Surfaces, defects, molecules, ...

What about other quantities?

Band gaps, barriers, spectra ...



Can we reach
ultimate precision?

Total energies of atoms
compared to MADNESS
using multiresolution analysis



Andris Gulans



Same for molecules

Yes we can!

	Ha	μ Ha
	$E^{\text{LAPW+lo}}$	ΔE^{MRA}
H	-0.4787107	0.5
He	-2.8344552	0.2
Li	-7.3432843	1.6
Be	-14.4464735	0.6
B	-24.3548568	1.3
C	-37.4685404	0.9
N	-54.1343867	0.7
O	-74.5286993	0.2
F	-99.1118530	-0.3
Ne	-128.2299171	0.3
Na	-161.4436320	2.1
Mg	-199.1352882	0.7
Al	-241.3178300	2.3
Si	-288.2171655	1.6
P	-340.0000526	1.8
S	-396.7390648	0.7
Cl	-458.6643433	0.1
Ar	-525.9397933	0.8

<http://exciting-code.org>

A. Gulans, S. Kontur, C. Meisenbichler, D. Nabok, P. Pavone, S. Rigamonti, S. Sagmeister, U. Werner, and C. Draxl

exciting: a full-potential all-electron package implementing density-functional theory and many-body perturbation theory
J. Phys: Condens. Matter 26, 363202 (2014).



Home

[Documentation](#)
[Tutorials](#)
[Input Reference](#)
[Template Market](#)

Downloads

[exciting](#)
[Other Packs](#)

How to reach us

[Contact](#)
[Forum](#)

Development

[Developers Space](#)
[Manage Wiki](#)
[Internal](#)

[edit this panel](#) | [recent changes](#)

The exciting Code

Download exciting

exciting is a full-potential all-electron density-functional-theory package implementing the families of linearized augmented planewave methods. It can be applied to all kinds of materials, irrespective of the atomic species involved, and also allows for exploring the physics of core electrons. A particular focus are excited states within many-body perturbation theory.

A. Gulans, S. Kontur, C. Meisenbichler, D. Nabok, P. Pavone, S. Rigamonti, S. Sagmeister, U. Werner, and C. Draxl, "exciting — a full-potential all-electron package implementing density-functional theory and many-body perturbation theory", J. Phys.: Condens. Matter **26**, 363202 (2014)



Documentation
How exciting! Hands-on w

Developers Team

- [exciting hub at the Humboldt Uni](#)
- [Current developers](#)

Events

Dual basis for WF, density, potential, ...

Atomic spheres α , β

Atomic-like basis functions

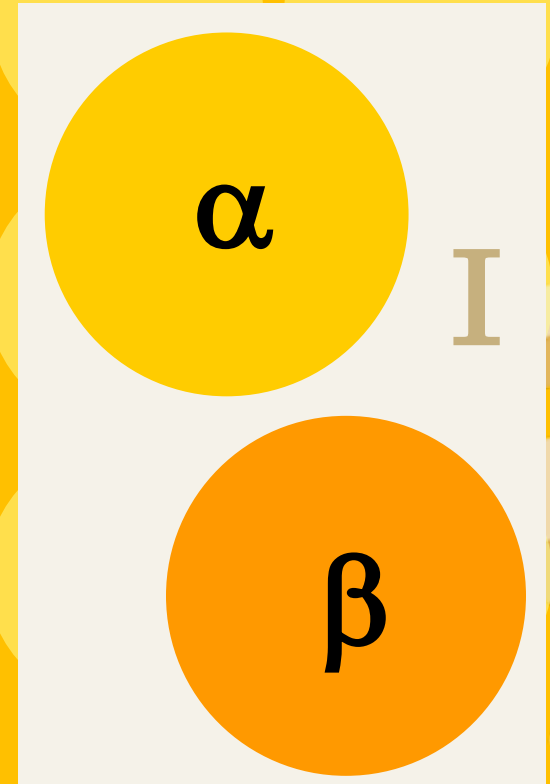
Interstitial

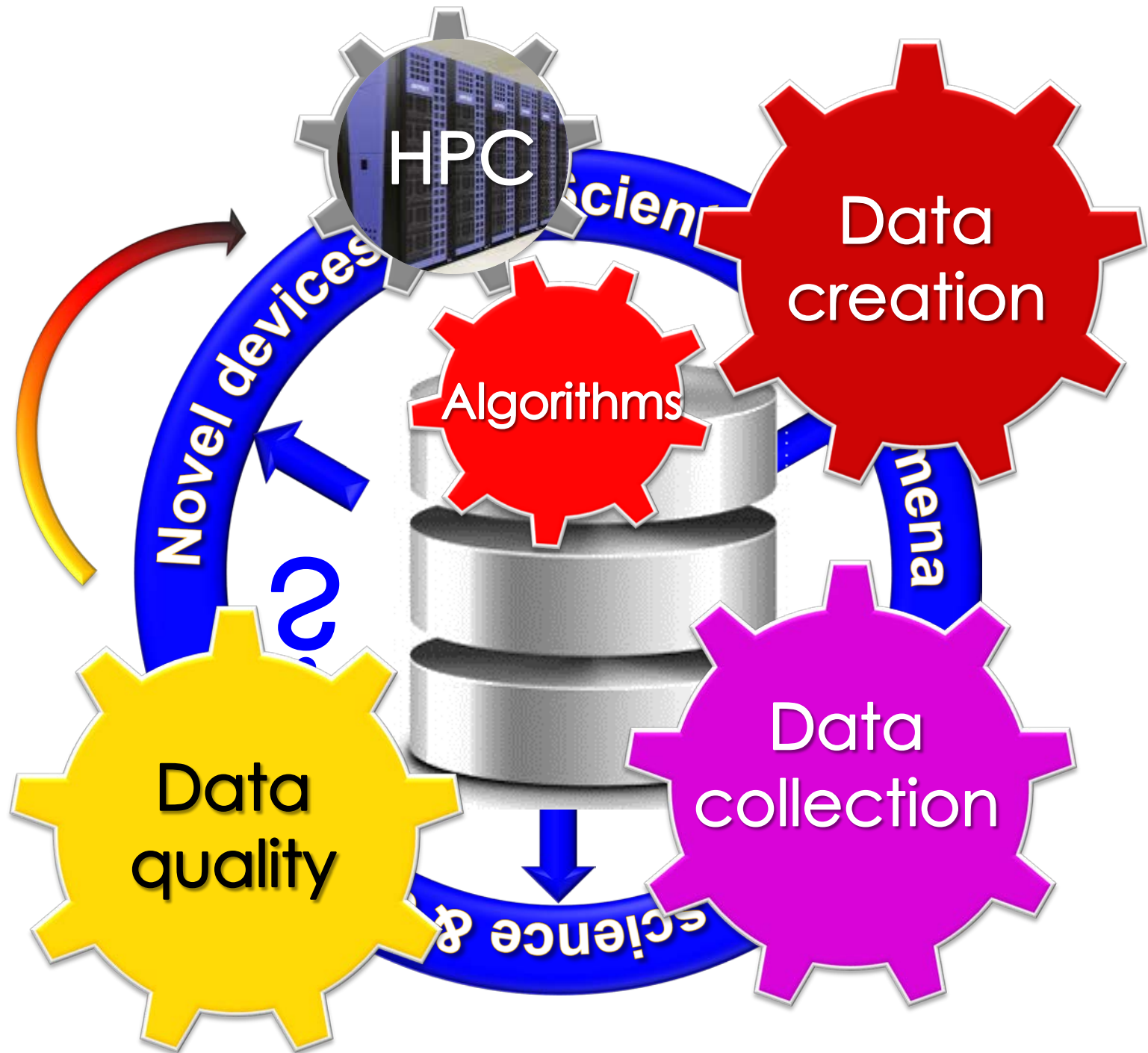
Planewave basis

All-electron method

Can handle strong variations

Can explore the core region





Algorithms

Matrix size

Tested up to 8 mio.

Needed up to 1 mio.

after implementing a Coulomb cutoff

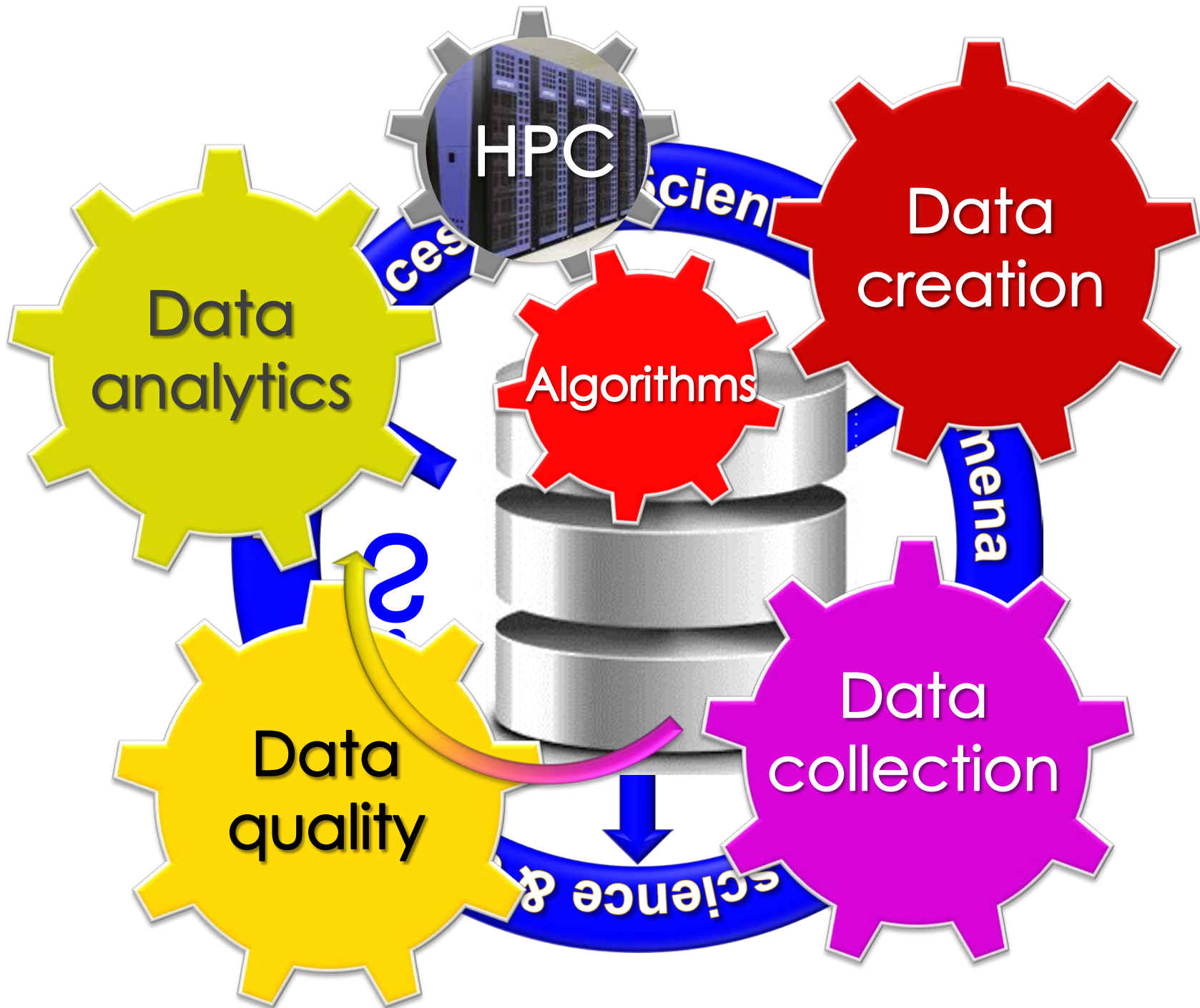
Block-Davidson

Special implementation to be used in connection with LAPW method

Scaling $O(V \cdot \log(V))$

V ...volume of unit cell

Will domain scientists meet exascale challenges?



Big-data Analytics

Identifying correlations and structure in big data of materials will enable scientists and engineers to decide which materials are useful for specific applications or which new materials should be the focus of future studies.



Big-Data Analytics

One example ...



Classification of materials

Can we predict the crystal structure from the nuclear charges Z_A and Z_B ?

Classical example: Phillips – Van Vechten Problem

J. A. Van Vechten, PRB 182 , 891 (1969).
J. C. Phillips, Rev. Mod. Phys. 42, 317 (1970).

A. Zunger, PRB 22, 5839 (1980).
D. G. Pettifor, Solid State Commun. 51, 31 (1984).

Y. Saad, D. Gao, T. Ngo, S. Bobbitt, J. R. Chelikowsky, and W. Andreoni, PRB 85, 104104 (2012).



L. Ghiringhelli



J. Vybiral

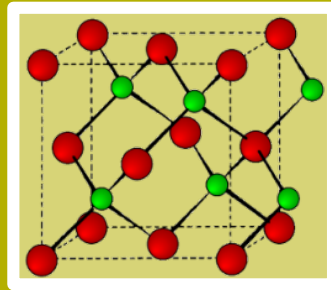
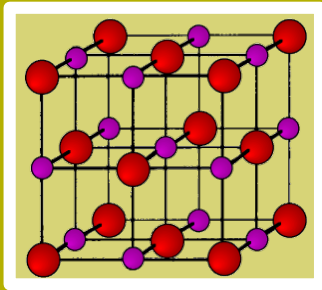


S. Levchenko

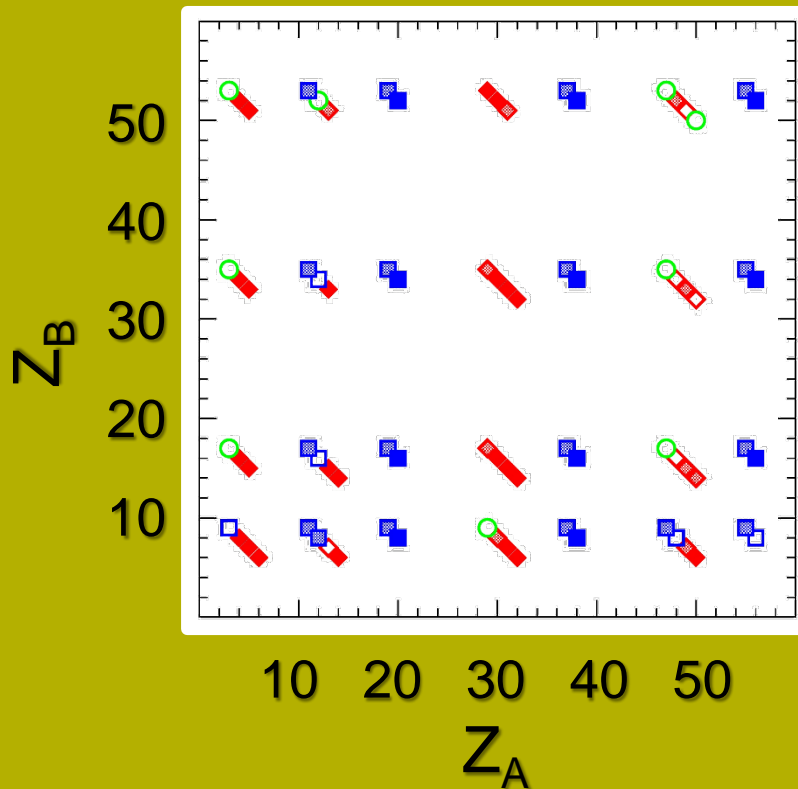


M. Scheffler

Rocksalt vs zincblende



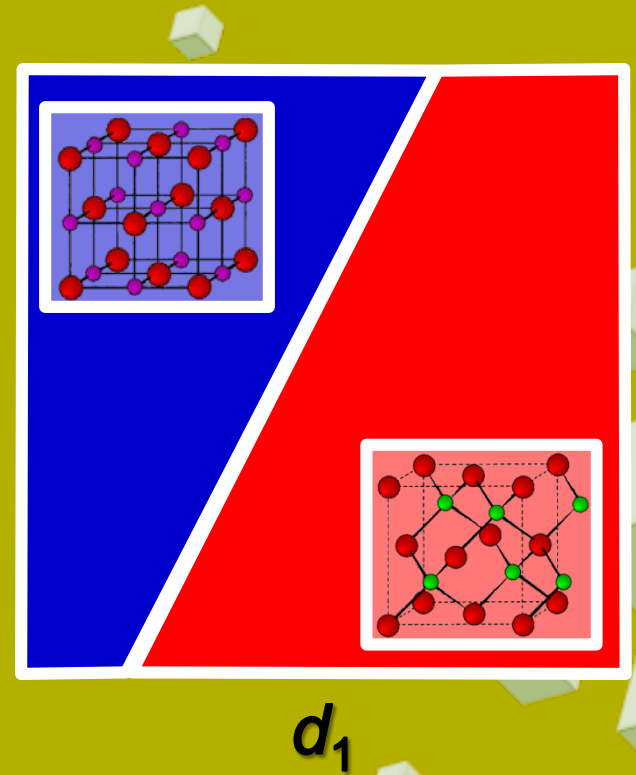
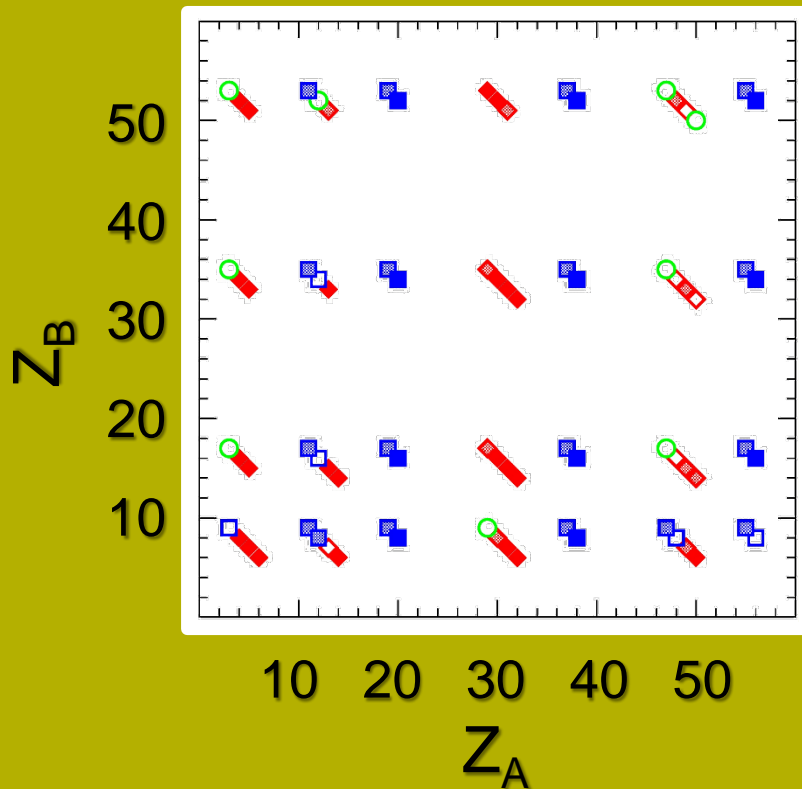
$$\Delta = E(rs) - E(zb) \text{ [eV]}$$



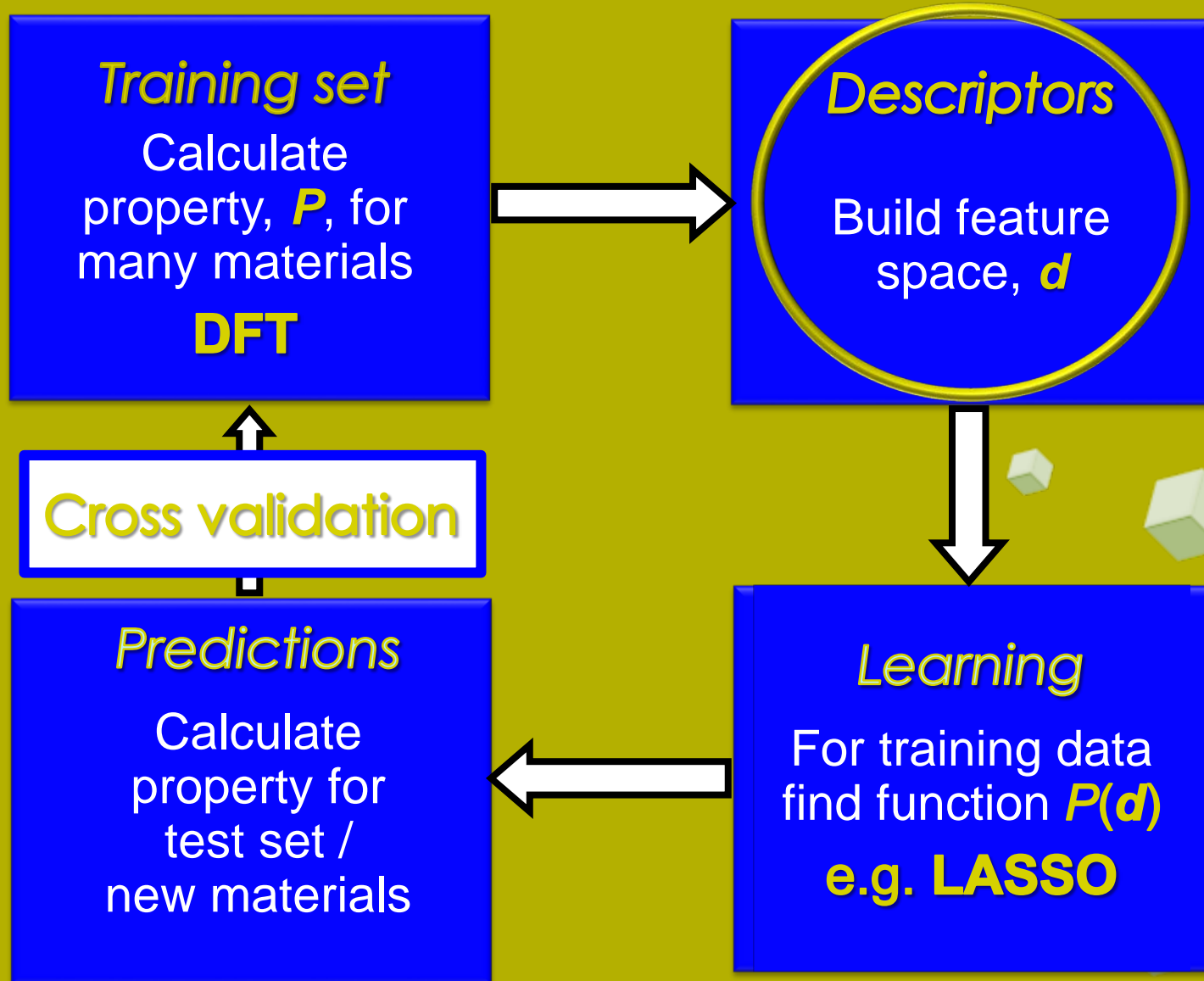
Rocksalt vs zincblende

Classification problem

J. A. Van Vechten, PRB 182 , 891 (1969).
J. C. Phillips, RMP 42, 317 (1970).



How to proceed?



Building descriptors

Primary features

Free atoms

$IP(A), IP(B)$

Ionization potential

$EA(A), EA(B)$

Electron affinity

$H(A), H(B)$

Highest occupied Kohn-Sham level

$L(A), L(B)$

Lowest unoccupied Kohn-Sham level

$r_s(A), r_s(B)$

Radius at max. of s-like wavefunction

$r_p(A), r_p(B)$

Radius at max. of p-like wavefunction

$r_d(A), r_d(B)$

Radius at max. of d-like wavefunction

Dimers

$HL(AA), HL(BB), HL(AB)$

HOMO-LUMO KS gap

$E_b(AA), E_b(BB), E_b(AB)$

Binding energy

$d(AA), d(BB), d(AB)$

Equilibrium distance

Building descriptors

Full feature space

10 000 **nonlinear** combinations of primary features

+, -, *, /, 2, 3, $\sqrt{\quad}$, exp., ...

Linear relationship

$$P(d) = c d$$

Let the machine choose most relevant descriptors

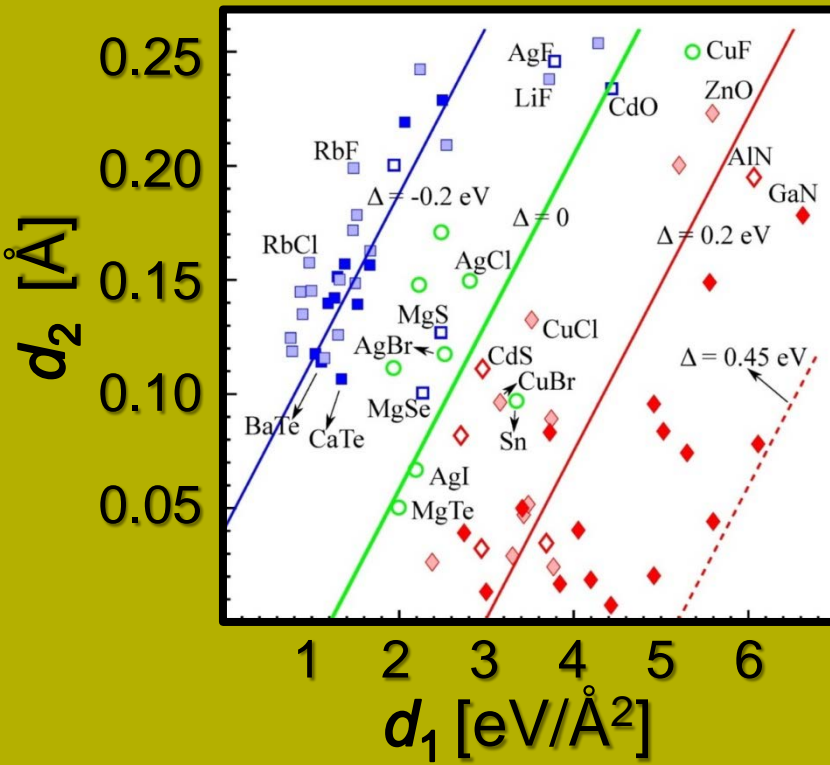
$$\min_{c \in R^M} \|P - c d\|^2 + \lambda \|c\|_1$$

Descriptors selected by LASSO

Least Absolute Shrinkage and Selection Operator

2D representation

Most relevant descriptors



$$d_1 = \frac{IP(A) - EA(B)}{r_p(A)^2}$$

$$d_2 = \frac{|r_s(A) - r_p(B)|}{\exp[rp(A)]}$$

$$d_3 = \frac{|r_p(B) - r_s(B)|}{\exp[rd(A) + r_s(B)]}$$

L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, C. Draxl, and M. Scheffler, PRL 114, 105503 (2015).
L.M. Ghiringhelli, et al., New J. Phys., in print.

Novel Materials Discovery

<http://nomad-repository.eu>

Currently 57 mio. files

Amount rapidly increasing

Several replica (BSC, China 2, Korea ...)

Working on the entire DB

on the search for unusual phenomena

Additional dedicated high-throughput
calculations needed

HPC resources required

... by sharing

Home NoMaD Team

Welcome to

NoMaD
Repository

NoMaD Repository

3 314 840

NoMaD from MedA
... more

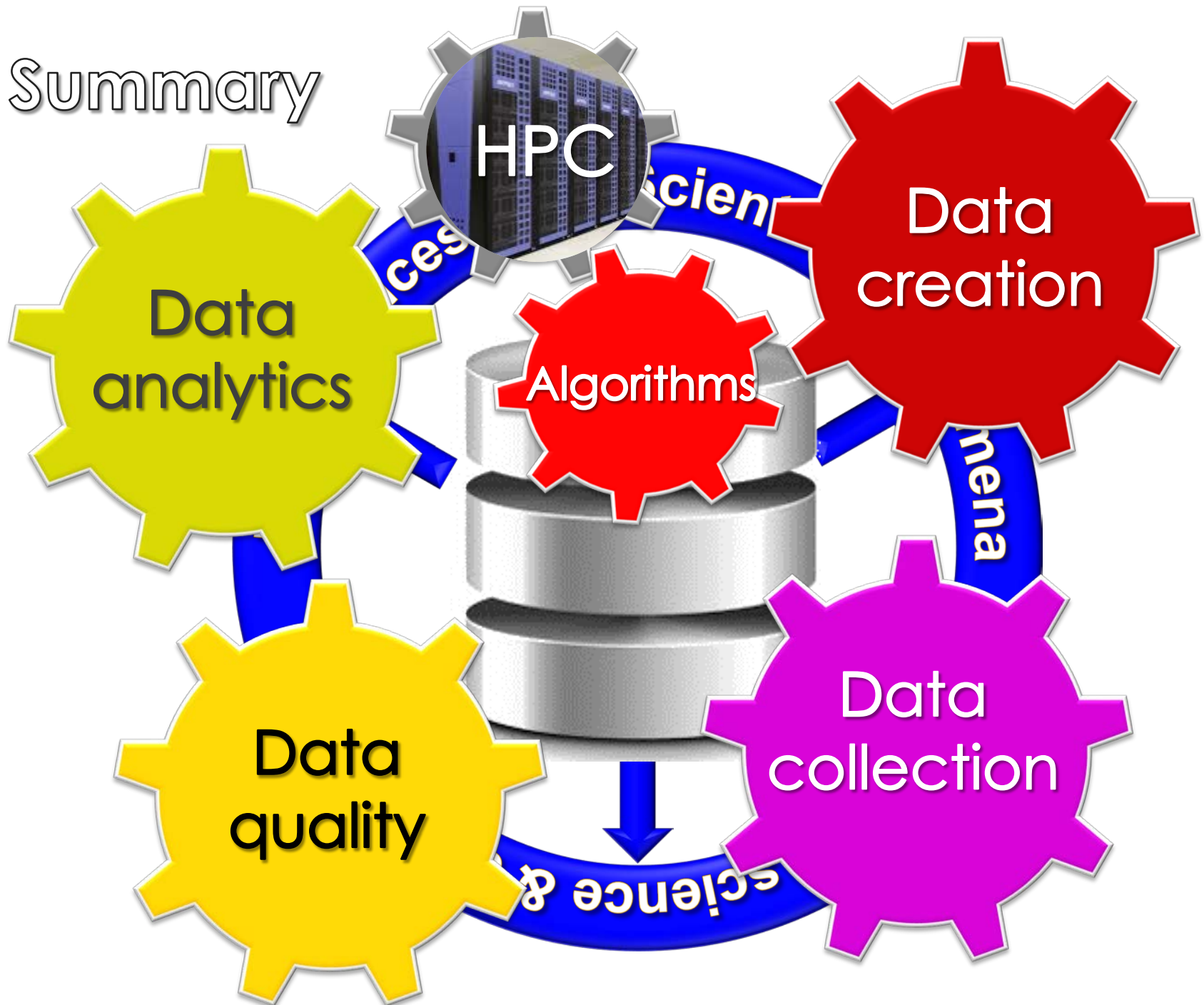
related conferences
ops.

Financial Support

exchange their results, inside a single group or between two or more, and to recall what was actually done some years ago.

The **NoMaD Repository** enables the confirmatory analysis of materials data, their reuse, and repurposing. Have a look at [youtube](#) to see our [movies on the concept](#) and a [basic tutorial](#).

Summary



This project has received funding from the European Union's Horizon 2020 research and innovation programme, grant agreement No 676580.

Thank you!

