

Kristallstrukturbestimmung

Fawzi Mohamed, Luca Ghiringhelli

FHI Max-Planck Gesellschaft



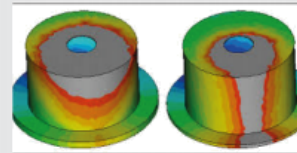
Big Data Applications



Information Marketplaces

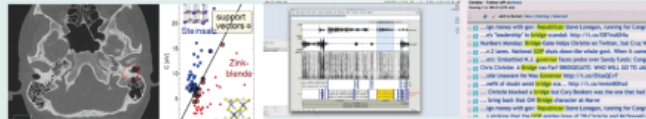


E-Health



Materials Science

Scalable
Machine
Learning



Statistical Analysis & Machine Learning

Video Annotation
Text Analytics
Material Characteristics

Emma (Declarative Specification)

ML

Graph

DataBag

Compiler / Optimizer

SystemML

R - Dialect

Compiler /
Optimizer



Apache Flink

Declarative,
Scalable
Data
Analytics

Framework • Testing • Development Tools • Benchmarking

Automatic
Optimization & Parallelization

XtreemFS

SDNs

Parameter Servers

Scalable Data
Management

Adaptive processing of data- & control flows • Optimization of storage distribution in modern file systems

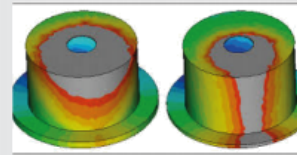
Big Data Applications



Information Marketplaces

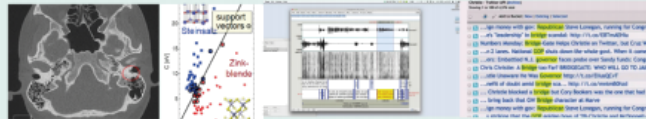


E-Health



Materials Science

Scalable
Machine
Learning



Statistical Analysis & Machine Learning

Video Annotation
Text Analytics
Material Characteristics

Emma (Declarative Specification)

ML

Graph

DataBag

Compiler / Optimizer

SystemML

R - Dialect

Compiler /
Optimizer



Apache Flink

Declarative,
Scalable
Data
Analytics

Framework • Testing • Development Tools • Benchmarking

Automatic
Optimization & Parallelization

XtreemFS

SDNs

Parameter Servers

Scalable Data
Management

Adaptive processing of data- & control flows • Optimization of storage distribution in modern file systems

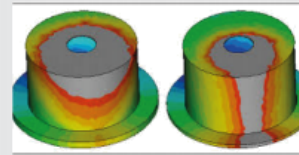
Big Data Applications



Information Marketplaces



E-Health



Materials Science

Scalable
Machine
Learning



Statistical Analysis & Machine Learning

Video Annotation
Text Analytics
Material Characteristics

Emma (Declarative Specification)

ML

Graph

DataBag

Compiler / Optimizer

SystemML

R - Dialect

Compiler /
Optimizer



Apache Flink

Declarative,
Scalable
Data
Analytics

Framework • Testing • Development Tools • Benchmarking

Automatic
Optimization & Parallelization

XtreemFS

SDNs

Parameter Servers

Scalable Data
Management

Adaptive processing of data- & control flows • Optimization of storage distribution in modern file systems



THE NOMAD LABORATORY

A EUROPEAN CENTRE OF EXCELLENCE

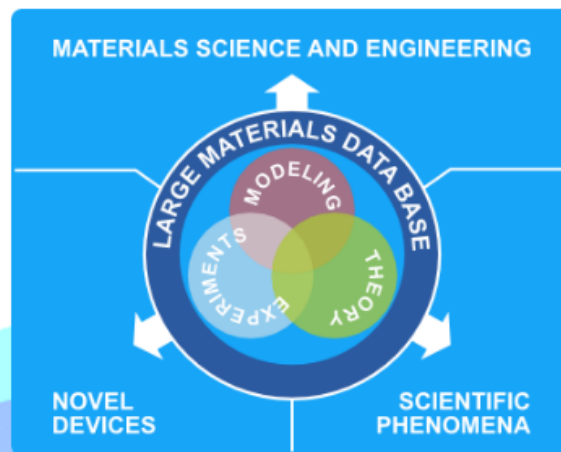
[HOME](#) [PROJECT](#) [INDUSTRY](#) [OUTREACH](#) [TEAM](#) [CODES](#) [NEWS](#) [PRESS KIT](#) [CONTACT US](#)

Enter Search...



The Novel Materials Discovery (NOMAD) Laboratory develops a *Materials Encyclopedia* and *Big-Data Analytics* and *Advanced Graphics Tools* for materials science and engineering.

Eight complementary computational materials science groups and four high-performance computing centers form the synergetic core of this Centre of Excellence.



Latest News

Nov 1, 2016

[Industry Interview - BIOVIA](#)

Oct 21, 2016

[NOMAD Year 1 Report](#)

Oct 7, 2016

[NOMAD Year 1 Meeting with Scientific Advisory Committee](#)

Sep 24, 2016

[7th School and Workshop on Time-Dependent Density-Functional Theory \(TDDFT\)](#)

Sep 19, 2016

[Industry Interview - Lockheed Martin](#)

<http://nomad-coe.eu>



MATERIALS ENCYCLOPEDIA



BIG-DATA ANALYTICS



ADVANCED GRAPHICS



HPC INFRASTRUCTURE

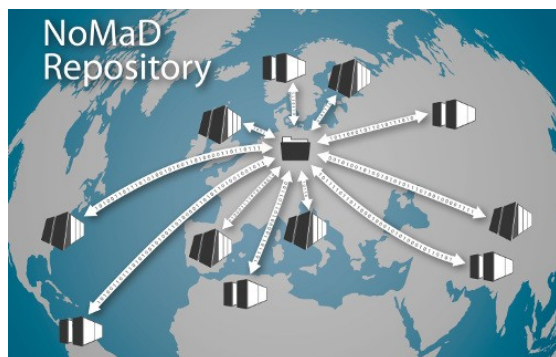


OUTREACH

© BBDC

Ziele:

- Korrelation und Struktur in Materialien Big Data
- Materialien für spezifische Applikationen
- Materialien zu erforschen in zukünftige Studien
- Den ersten Schritt ist die **Daten Verfügbar** für die Analyse **zu machen**
- Dann interaktive Anwendungen um Trends und Unregelmäßigkeit zu erforschen.
- BBDC Kenntnisse in Skalierbare Big Data Analyse bereichern die Expertise der NOMAD CoE



<http://nomad-repository.eu>

- Source of our data
- Established to host organize and share materials data
- Keeps data for at least 10 Years
- Open access and restricted data
 - >3M entries, >2.8M open access
- We use only open access data
- Joint effort by the groups of
 - Matthias Scheffler, FHI Berlin
 - Claudia Draxl, HU Berlin
 - Max Planck Computer & Data Facility (MPCDF), Garching, headed by Stefan Heinzl.



Claudia Draxl
HUB



Matthias Scheffler
FHI



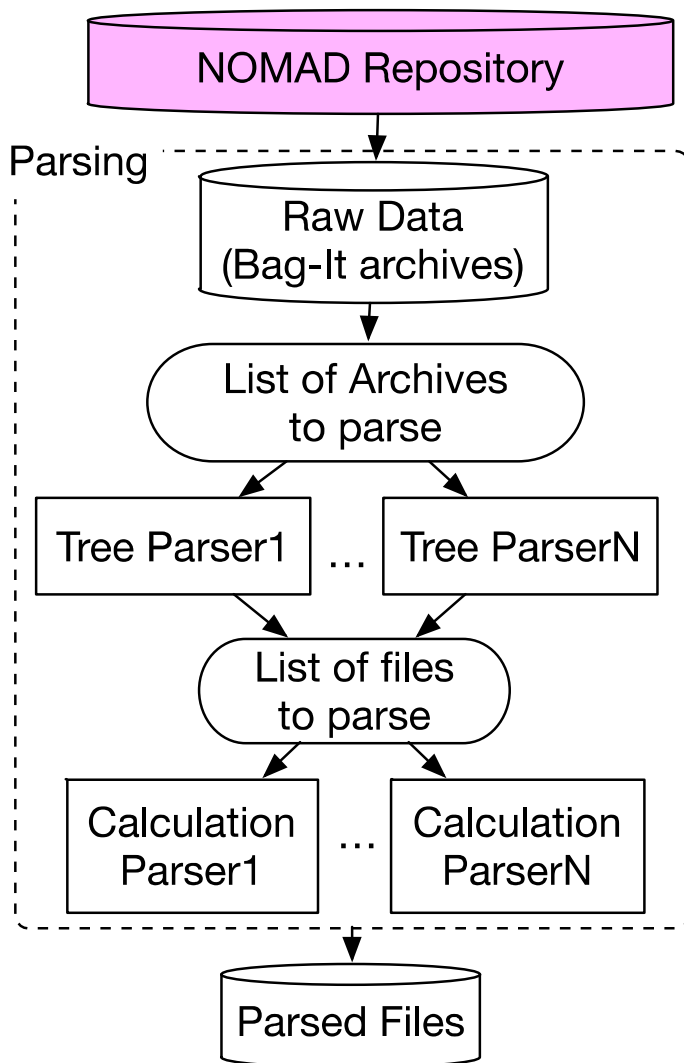
Jungho Shin
HUB



Lorenzo Pardini
HUB



Thomas Zastro
MPCDF



- Parsers **interpret** all calculation data
- **Organize** it according to the metadata structure
- Data not extracted is invisible
- Writing a parser cannot be automatized and requires a person with *scientific knowledge*
- **Parallel** execution
 - *Tree Parser* identifies the files
 - *Calculation Parser* performs the parsing and generates the parsed files
- Parsing is **pure**: the same version on the same data should give the same result

section_run

program_name FHI-aims

program_version 081912

section_system

simulation_cell $[[1.4e-9 \dots]]$

atom_positions $[[0.0, \dots] \dots]$

atom_labels ["Cu", ...]

section_method

basis_set fhi_aims_tight

XC_method DFT_GGA_PBE

section_single_configuration_calculation

section_scf_iteration

energy_total_scf_iteration $-1.326e-20$

section_scf_iteration

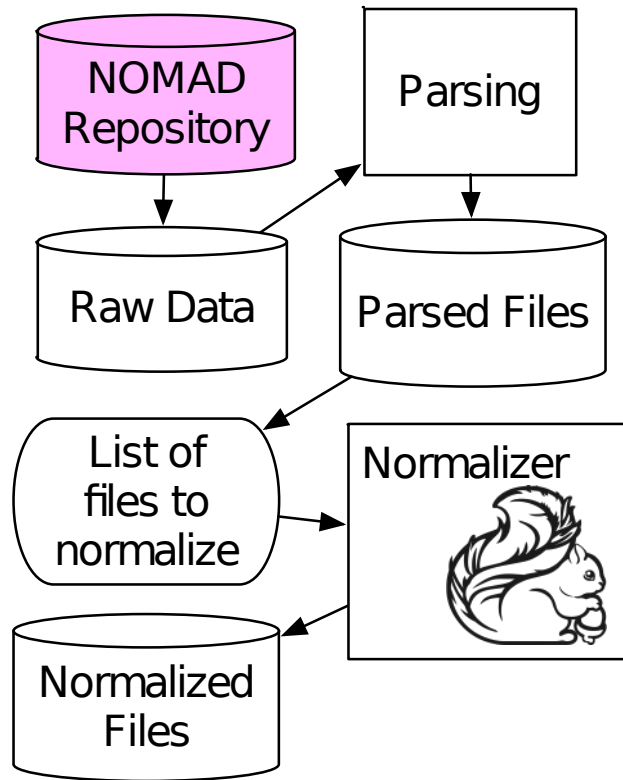
energy_total_scf_iteration $-1.344e-20$

energy_total $-1.344e-20$

Values: Data
Structure
and names: Metadata

SI Units:

- lengths: m
- energies: J
- ...

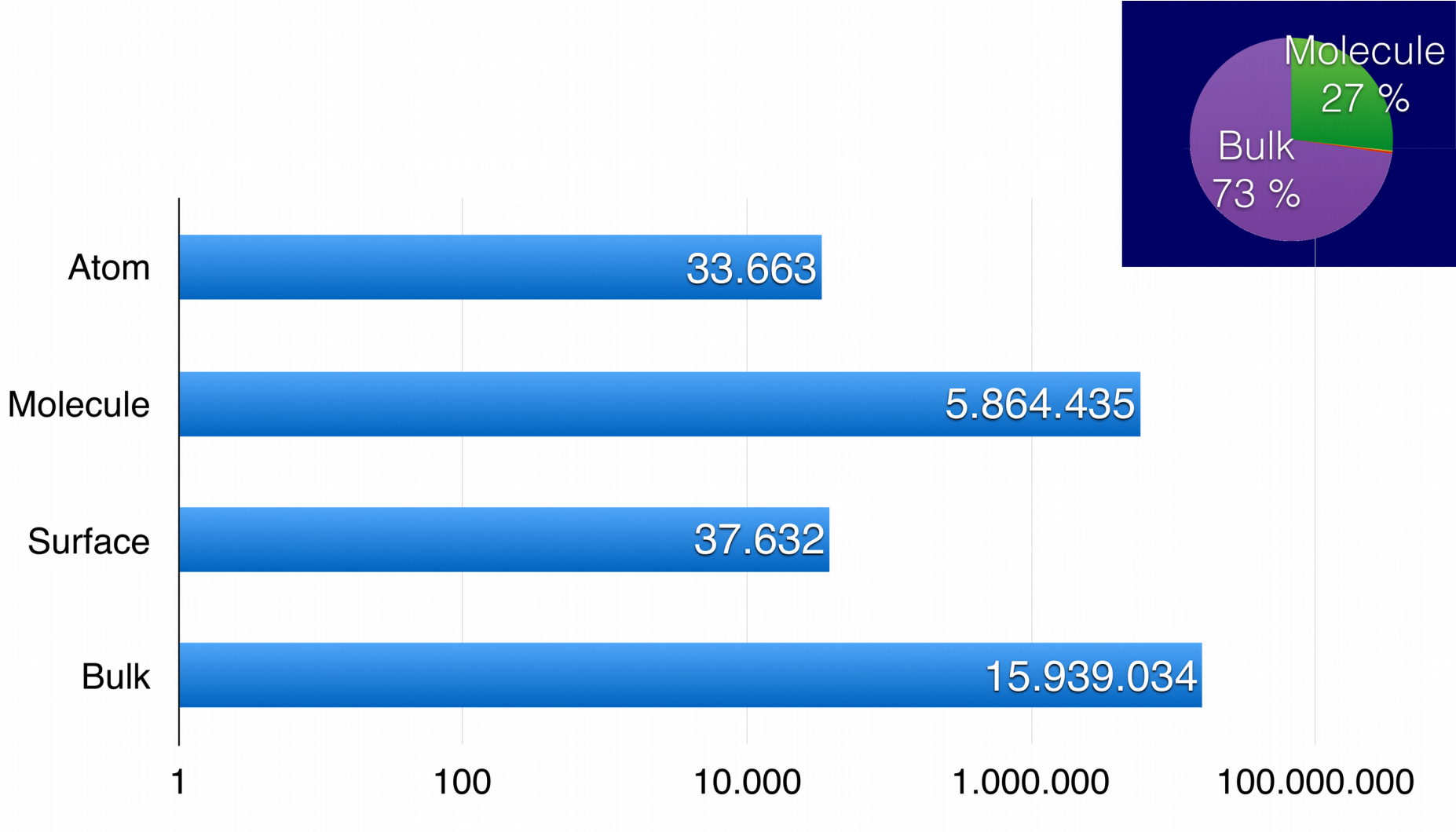


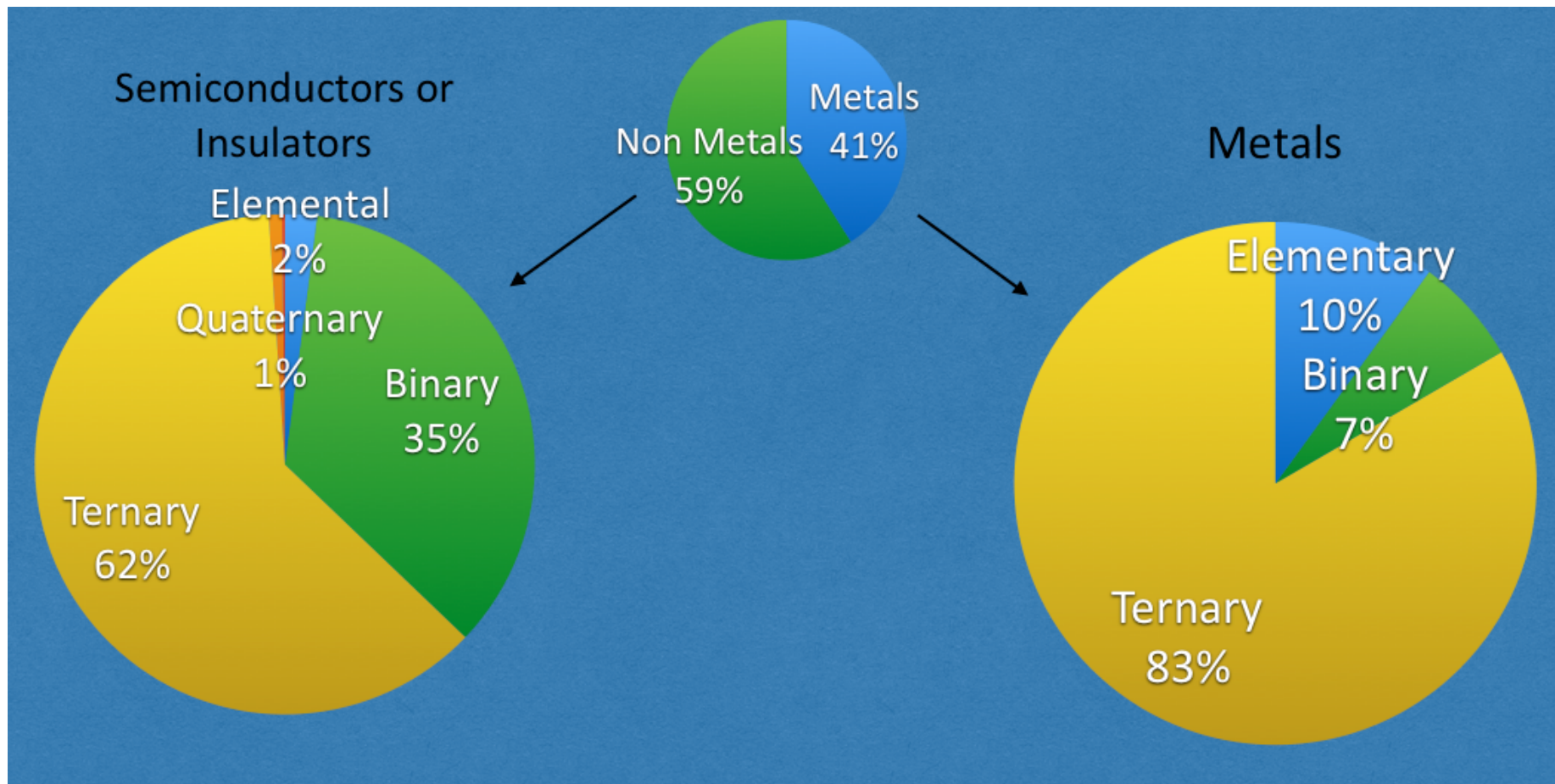
Standardization

- Parsers standardize the data,
- avoid the loss of information

Normalization

- The work packages define **derived quantities** (normalized representations,...)
- These can be generally useful for analysis or visualization
- Normalization is an infrastructure to apply automatically some transformations and store their result along with the parsed data



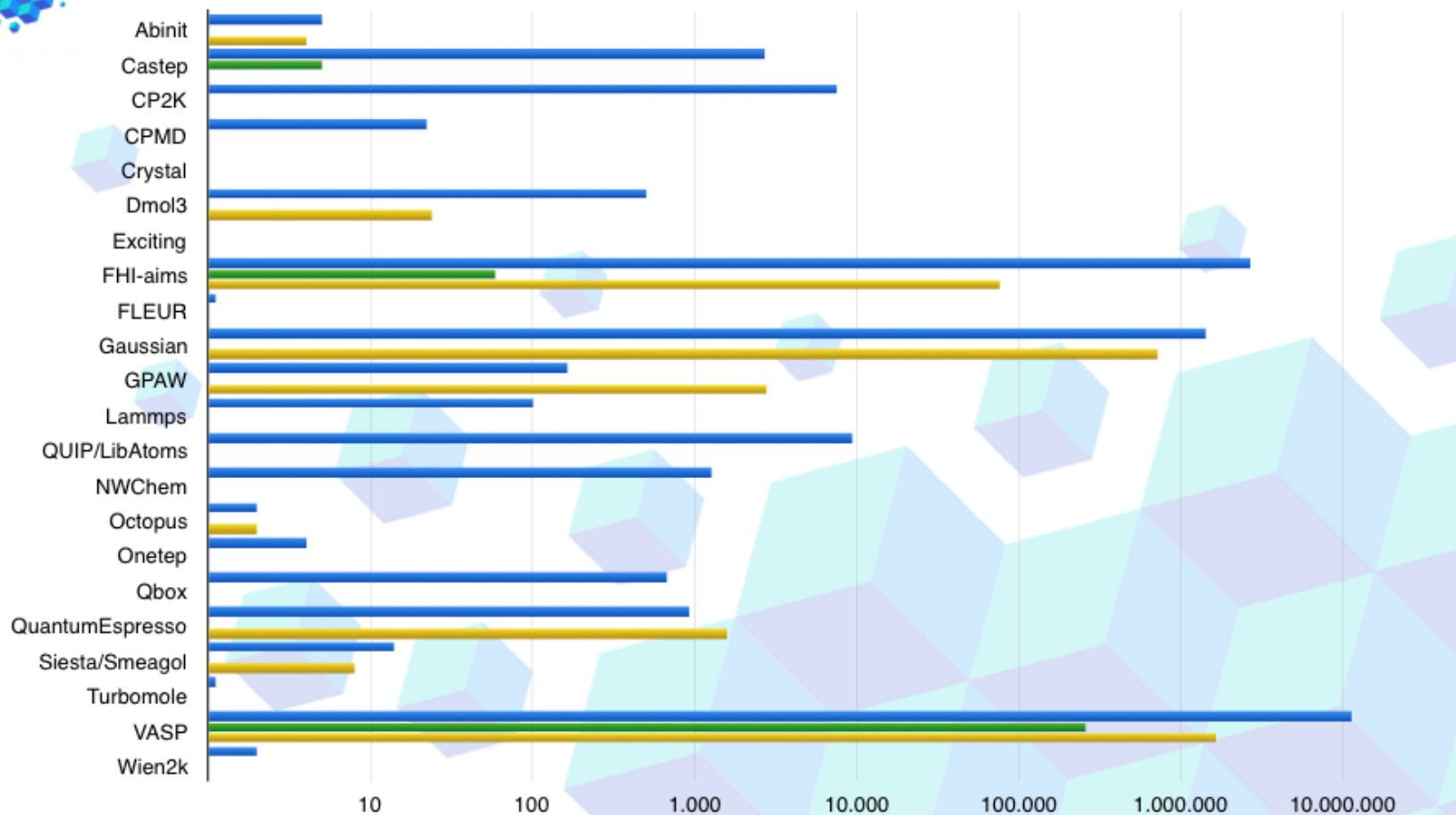




Unique Geometries

Bands

Eigenvalues



Kristallstrukturbestimmung

- Materialeigenschaften sind abhängig von Kristallstruktur
- Verschiedene Kristallstrukturen können geringe Energieunterschiede haben
- Man kann Kristallstrukturen Patentieren
- Kristallstruktur ist wichtig wenn mehrere Materialien kombiniert werden
- Wir versuchen klein dimensional Beschreibungen (descriptors) die robust sind zu finden (extrapolation)

NOMAD analytics toolkit

Tutorial example on Crystal prediction I: The case of octet-binary zincblende-vs-rocksalt semiconductor

developed by Angelo Ziletti, Ankit Karirya, Emre Ahmetcik, Fawzi Mohamed, Luca Ghiringhelli, and Matthias Scheffler. [Last update September 13, 2019]

Introduction and motivation

Machine learning method: Compressed sensing (LASSO performed on a tailor-made feature space, followed by L0-regularized minimization).

Click [here](#) for more info on the LASSO+L0 method.

Reference: "Big Data of Materials Science: Critical Role of the Descriptor"

L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, Phys. Rev. Lett. 114, 105503 (2015) ([Click here for the free access pdf](#))

Instructions

Primary features (hover the mouse pointer over the feature names to see a full description):

- ☒ IP^A
- ☒ EA^B
- ☒ E_{LUMO}^A
- ☒ r_s^B
- ☒ r_d^A
- ☐ d^{BB}
- ☐ E_{bb}^{AA}
- ☐ Z^B
- ☐ n_{period}^A
- ☐ r_{π}
- ☐ E_{bb}^{AB}

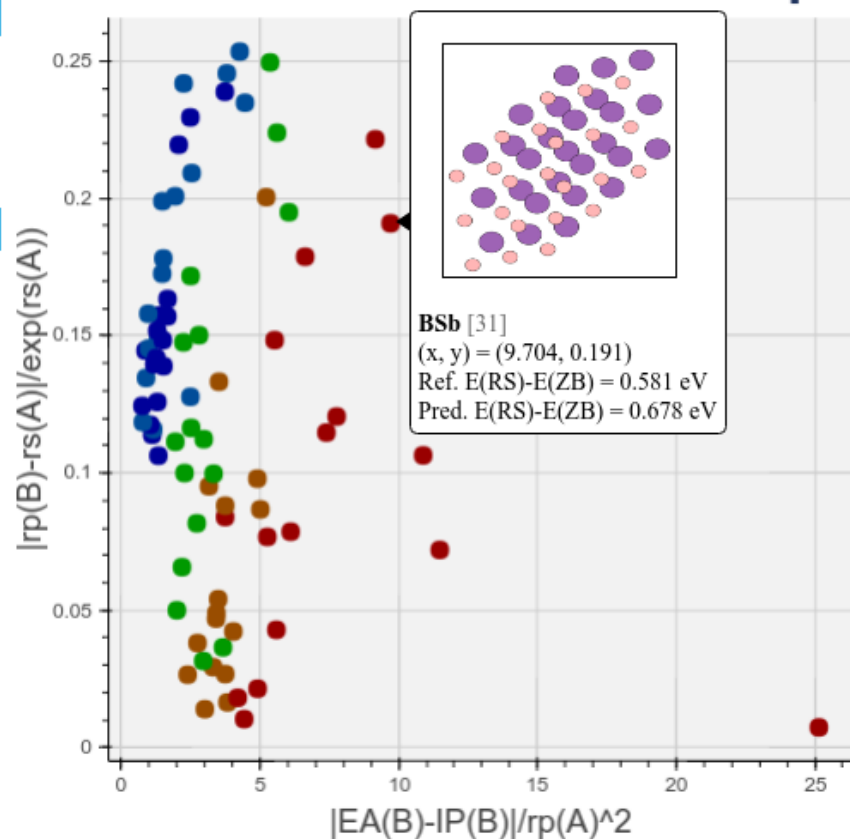
- ☒ IP^B
- ☒ E_{HOMO}^A
- ☒ E_{LUMO}^B
- ☒ r_p^A
- ☒ r_d^B
- ☐ ΔE_{HL}^{AA}
- ☐ E_{bb}^{BB}
- ☐ Z_{val}^A
- ☐ n_{period}^B
- ☐ d^{AB}

- ☒ EA^A
- ☒ E_{HOMO}^B
- ☒ r_s^A
- ☒ r_p^B
- ☐ d^{AA}
- ☐ ΔE_{HL}^{BB}
- ☐ Z^A
- ☐ Z_{val}^B
- ☐ r_{σ}
- ☐ ΔE_{HL}^{AB}

Allowed operations:

Given features x and y , apply these operations:

LASSO+L0 structure map

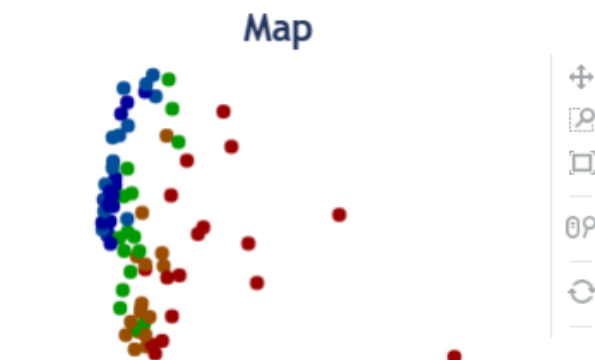
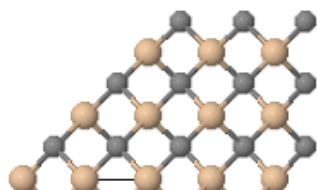


Reference $E(RS)-E(ZB)$

■ [-0.379eV, -0.164eV] ■ [-0.164eV, -0.058eV] ■ [-0.058eV, 0.11eV] ■ [0.11eV, 0.272eV] ■ [0.272eV, 2.629eV]

Name	Energy [eV]	Geometry File
SiC	-8903.9092	View

P 1 [P 1] #1
 $a=3.062\text{\AA}$
 $b=3.062\text{\AA}$
 $c=3.062\text{\AA}$
 $\alpha=60.000^\circ$
 $\beta=60.000^\circ$
 $\gamma=60.000^\circ$

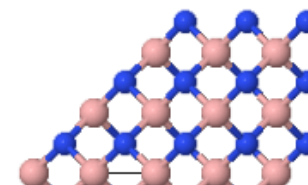


Selection		
Name	Reference $E(RS)-E(ZB)$ [eV]	Predicted $E(RS)-E(ZB)$ [eV]
BN	1.712	1.618
SiC	0.669	0.562

[Clear Selection](#)

Name	Energy [eV]	Geometry File
BN	-2153.491	View

P 1 [P 1] #1
 $a=2.534\text{\AA}$
 $b=2.534\text{\AA}$
 $c=2.534\text{\AA}$
 $\alpha=60.000^\circ$
 $\beta=60.000^\circ$
 $\gamma=60.000^\circ$



Andere Projekte

- Subgroup discovery
 - „spezielle“ Untergruppen Identifizieren
- Structural Similarity
 - Aktuelles Projekt
 - Kollaboration mit Shinichi Nakajima über ein Similaritätsmaß basierend auf die paarweise Korrelation

Danksagung

- Angelo Ziletti
- Ankit Kariryaa
- Emre Ahmetcik
- Matthias Scheffler
- Den Ganzen NOMAD Team
- BBDC

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

