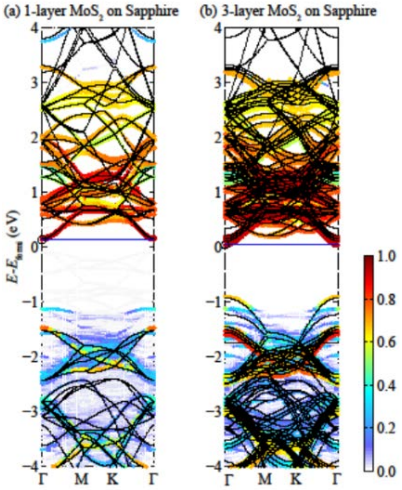
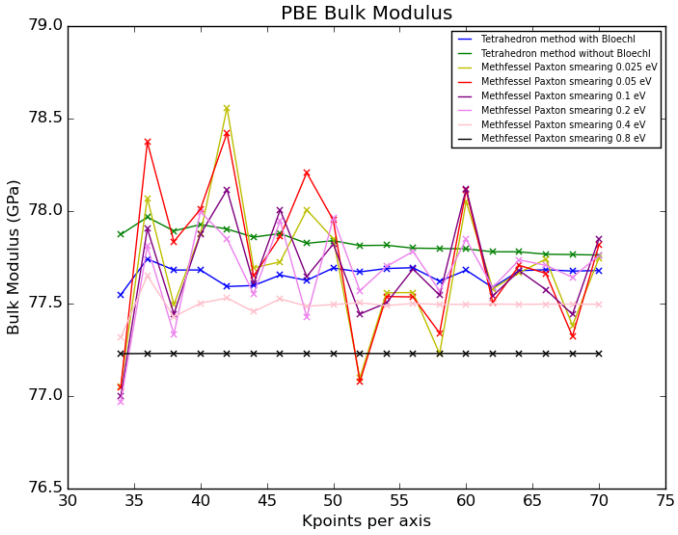
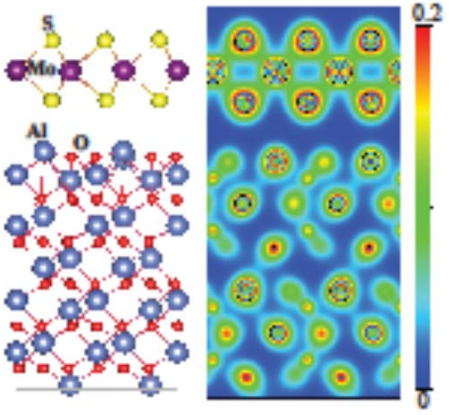
# A NIST effort towards the quantification and benchmarking of Density Functional Theory (DFT) uncertainty




***F. Tavazza (NIST),*** *T. Allison (NIST), Y. Congo (NIST), J. Gabriel (UFL), R. G. Hennig (UFL)*



# Motivation

- DFT values are often used in the literature as **reference values** (to compare to, to fit to, etc. )

**Table 5**
NiC (NaCl structure): physical constants from GGA and MEAM calculation.

| Materials | DFT | MEAM |
|---|---|---|
| Lattice constant (Å) | 4.08 | 4.08 |
| Cohesive energy (eV) | 10.54 | 9.89 |
| Bulk modulus (GPa) | 249 | 250 |
| $C_{11}$ (GPa) | 296 | 512 |
| $C_{12}$ (GPa) | 228 | 121 |
| $C_{44}$ (GPa) | 48 | 6.6 |

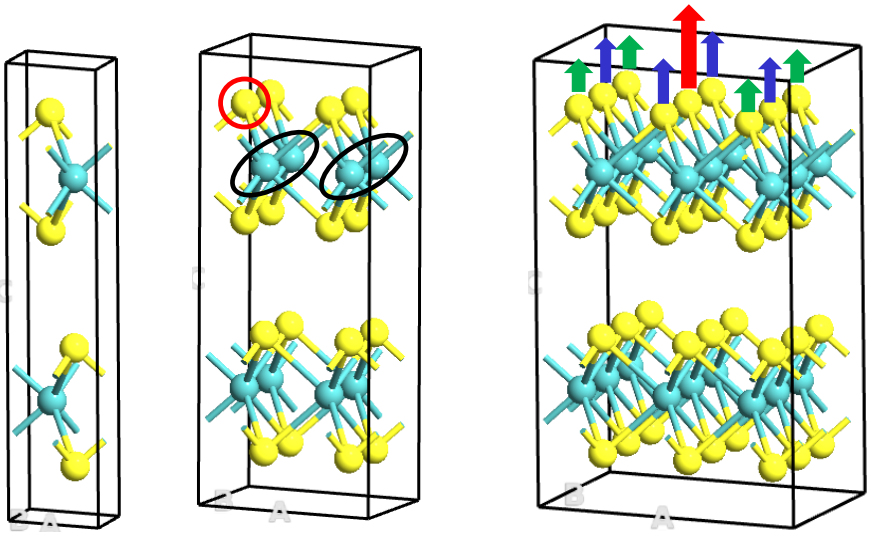
- They are always reported **without uncertainties**
- They are often reported with very **few technical specifications**

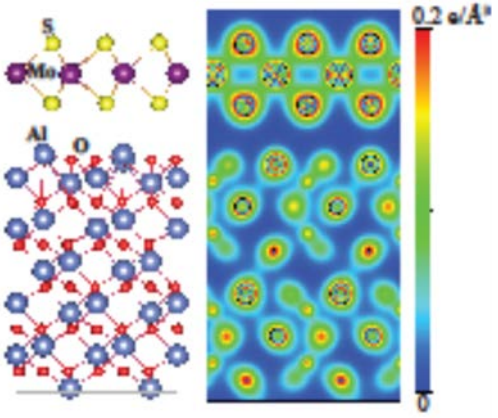
**however**

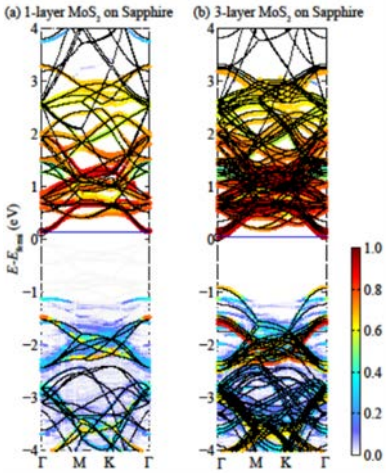
**Is DFT exact ?**

# Density Functional Theory (DFT)

- Atomistic modeling
- Based on quantum-mechanics (both ions and **electrons**)
- **Simulation cell**, often periodic boundary conditions





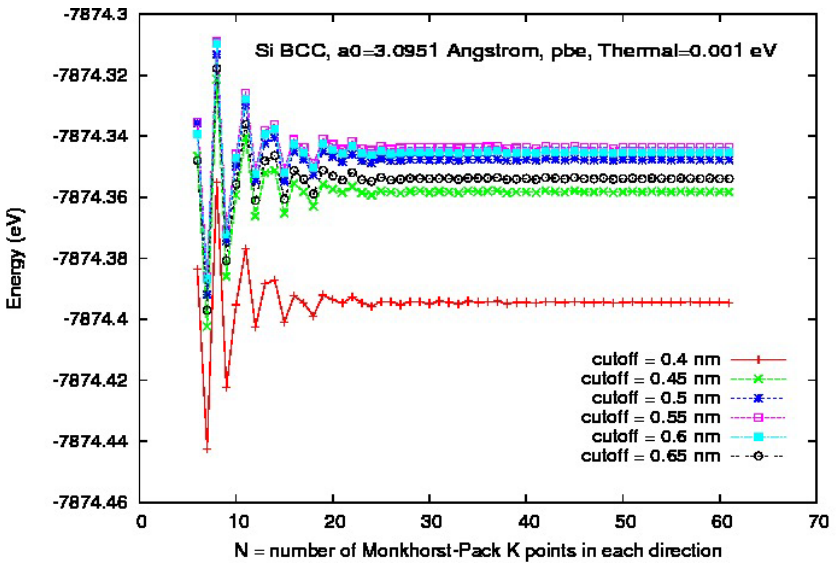




⇨
- Very few fitting parameters ➔ very transferable
- Very computationally costly:
  - small systems (**hundreds of atoms at most**)
  - Very short times (MD, semistatic deformations)

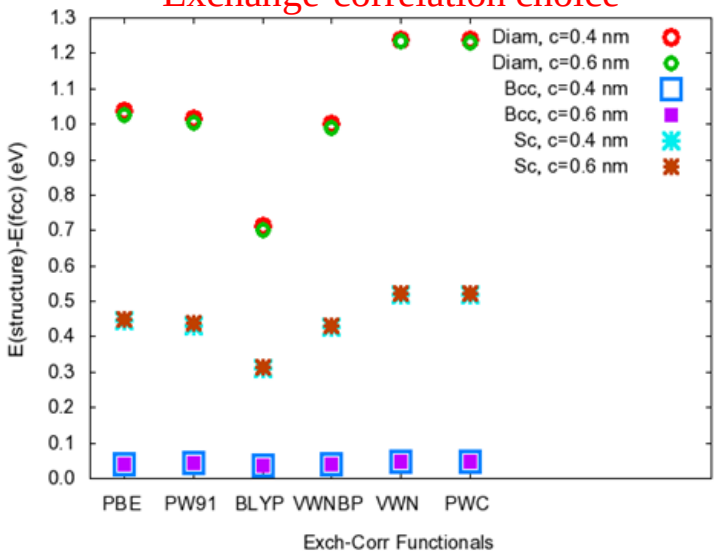
# Uncertainties in DFT

1) **controlled approximations**, whose errors can be made arbitrarily small through adjustable parameters typically at the expense of increased computational cost (Ex: k-points, real space or energy cutoff)

k-point convergence




Exchange-correlation choice




2) **uncontrolled approximations**, whose errors are unknown exactly and can't be reduced by increasing the computation
Ex: exchange, correlation, pseudopotential

# DFT Basics

❖ The **ground state (GS) energy** of a molecule/crystal can be determined from the **electron density** (3 d.o.f.) **instead of a wave function** (3N d.o.f., N= # of electrons)

❖ **Variational problem:**

the GS energy (E) and density (ρ) correspond to the **minimum of some universal functional E[ρ]** subject to the constrain that the density contains the correct number of electrons (Hohenburg-Kohn theorems )

**Because E= E[ρ] and ρ is unknown**, the above minimization is performed **self-consistently** (SCF)

Convergence parameters

❖**Basis sets**:

electronic structure methods rely on **expanding the unknown wave function** in terms of a set of basis functions

Some of possible **types:** - atom centered localized basis sets

"Parameters" for - plane waves.

basis set expansion

# Energy Functional E[ρ]

- Schrödinger eq. (+time independence, non relativistic, Born-Oppenheimer approximations) ➔ energy functional of a system of **interacting electrons**:

$$\mathbf{E\,[\rho]} = \text{energy functional} = T\,[\rho] + V_{ext}[\rho] + V_{ee}[\rho]$$

where: T=kinetic en. (unknown), $V_{ext}$ = electron-ion interaction, $V_{ee}$ = electron-electron interaction (unknown)

- Kohn- Sham introduced a fictitious system of N **non-interacting electrons** moving in an effective potential with density = to the true density ➔ the energy functional is:

$$E\,[\rho] = T_s[\rho] + V_{ext}[\rho] + V_{Coulomb}[\rho] + E_{xc}[\rho]$$

where **$E_{xc}$ = exchange –correlation functional** = error due to using a non-interacting kinetic energy + error due to treating the electron-electron interaction classically

**$E_{xc}$ not known exactly** and contains all the many-body quantum effects

"Parameter "= $E_{xc}$ functional form choice

# More questions

- What are the parameters that more strongly affect the result? **Do they depend on computed physical quantity**?
- How much do **different DFT codes** (periodic or molecular codes, local orbit basis or plane waves based codes or all-electron (augmented methods) codes) affect the results?
- How much does doing computations in a **non completely converged mode** affect the results?
- How can **uncertainties** be estimated?
- How systematic are they (by structure type, by material type, etc.)?
- How can they be propagated?
- How much is the **average user of DFT "numbers"** aware of DFT limitations/systematic problems etc.?

**DFT production codes**

- **Periodic codes (principally)**
  - Local orbital basis codes
    - QUEST; SeqQuest - gaussian basis pseudopotential code
    - SIESTA - numerical atom-centered basis pseudopotential code
    - CRYSTAL - CSE - gaussian basis all-electron code
    - AIMPRO
    - FPLO
    - OpenMX - GPL - numerical atom-centered basis PP code (Ozaki group)
  - All-electron (augmented methods) codes
    - ELK - GPL - FP-LAPW (one branch from the old EXCITING code)
    - EXCITING - FP-LAPW, focus on excited state properties (TDDFT, MBPT) [license not apparent on website, probably open source] (another branch from the old EXCITING code)
    - FLEUR - "freely available" - FLAPW code
    - RSPt - "Open Source" - FP-LMTO
    - WIEN2k - modest fee - full potential LAPW
  - Plane wave and related (real space, wavelet, etc.) methods
    - VASP - although check out its (trial?) spiffy new site
    - CASTEP and CETEP
    - CPMD
    - ABINIT - GPL
    - BigDFT - wavelets
    - Quantum-Espresso (formerly PWscf) - GPL
    - PEtot - GPL
    - DACAPO - GPL
    - Socorro - GPL
    - DFT++ - GPL
    - Octopus - GPL - real space TDDFT code
    - Paratec
    - DoD Planewave
    - PARSEC - GPL - real space, pseudopotential
    - CP2K - GPL (mixed basis DFT)
    - GPAW - GPL - real-space multigrid PAW code
    - SPHINX
    - QBOX - GPL - plane wave pseudopotential, large parallel

Return to Top

- **Molecular codes (principally)**
  - Gaussian.com (unless, of course, you have been "banned")
  - NWChem
  - DMol$^3$
  - Jaguar - Schrodinger, Inc.
  - GAMESS or GAMESS-UK
  - QCHEM
  - NRLMOL
  - MondoSCF (Matt Challacombe's Home Page)
  - ADF - SCM
  - deMon
  - CADPAC - The Cambridge Analytic Derivatives Package
  - PYQUANTE - GPL - python-based development toolset for DFT/HF
  - TURBOMOLE - DFT and HF for large molecular systems

Return to Top

- **DFT atomic pseudopotential codes**
  - fhi98PP pseudopotential program
    This is a well engineered, freely available package to generate pseudopotential

http://dft.sandia.gov/Quest/DFT_codes.html

# We are not alone …..

## The Kohn-Sham equation of state for elemental solids: a solved problem

Kurt Lejaeghere,[1] Chandler Becker,[2] Gustav Bihlmayer,[3] Torbjörn Björkman,[4] Peter Blaha,[5] Stefan Blügel,[3] Volker Blum,[6] Damien Caliste,[7] Ivano Eligio Castelli,[8] Stewart J. Clark,[9] Andrea Dal Corso,[10,11] Stefano de Gironcoli,[10,11] Thierry Deutsch,[7,12] Igor Di Marco,[13] Claudia Draxl,[14,15] Marcin Dułak,[16] Olle Eriksson,[13] Kevin F. Garrity,[2] Luigi Genovese,[7,12] Paolo Giannozzi,[11,17] Matteo Giantomassi,[18] Stefan Goedecker,[19] Xavier Gonze,[18] Oscar Grånäs,[13,20] Andris Gulans,[14,15] Donald R. Hamann,[21,22] Phil J. Hasnip,[23] Nathalie Holzwarth,[24] François Jollet,[25] Georg Kresse,[26] Klaus Koepernik,[27,28] Emine Küçükbenli,[10,11] Yaroslav O. Kvashnin,[13] Inka Locht,[13] Sven Lubeck,[14] Martijn Marsman,[26] Nicola Marzari,[8] Jens Jørgen Mortensen,[16] Taisuke Ozaki,[29] Lorenzo Paulatto,[30] Chris J. Pickard,[31] Ward Poelmans,[1] Matt I. J. Probert,[23] Keith Refson,[32,33] Manuel Richter,[27,28] Gian-Marco Rignanese,[18] Matthias Scheffler,[18] Francesca Tavazza,[2] Patrick Thunström,[34] Alexandre Tkatchenko,[15] Marc Torrent,[25] David Vanderbilt,[21] Michiel van Setten,[18] Veronique Van Speybroeck,[1] John M. Wills,[35] Jonathan R. Yates,[36] Guo-Xu Zhang,[37] and Stefaan Cottenier[1,38]

[1] *Center for Molecular Modeling, Ghent University, Technologiepark 903, BE-9052 Zwijnaarde, Belgium*

[2] *Material Measurement Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Stop 8553, Gaithersburg, MD 20899, USA*

[3] *Peter Grünberg Institut and Institute for Advanced Simulation, Forschungszentrum Jülich and JARA, D-52425 Jülich, Germany*

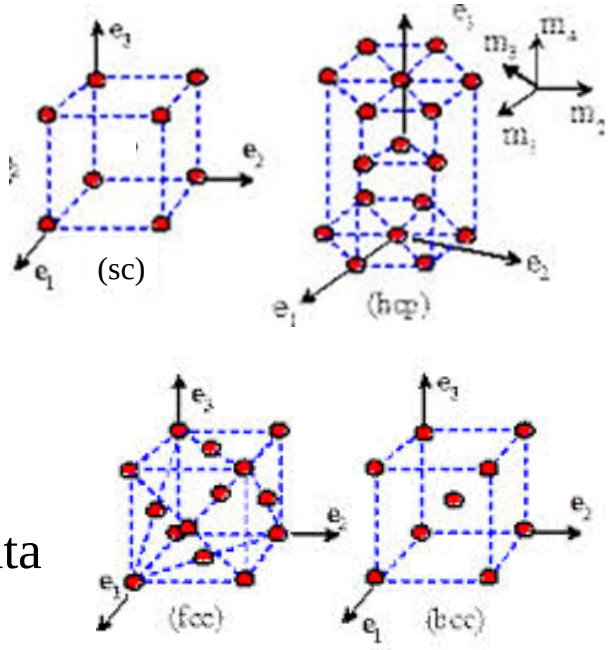
# NIST approach

❖ **Systematic** computation of

- basic **structural properties** (lattice and elastic constants, formation energies) of
- **single elements** (starting from Si, Fe, Ni, Cu, Al, C, Zr, W)
- **in stable and METASTABLE structures**

using as **many as possible** different

- Codes (➔ basis set expansions) ;
- Exchange-correlation choices;
- Thermal smearing, K-points convergence;
- ….




❖ Estimate of: - **uncertainties/trends** using these data

- uncertainties **propagation**

❖ Estimate of the effect of **not complete convergence**

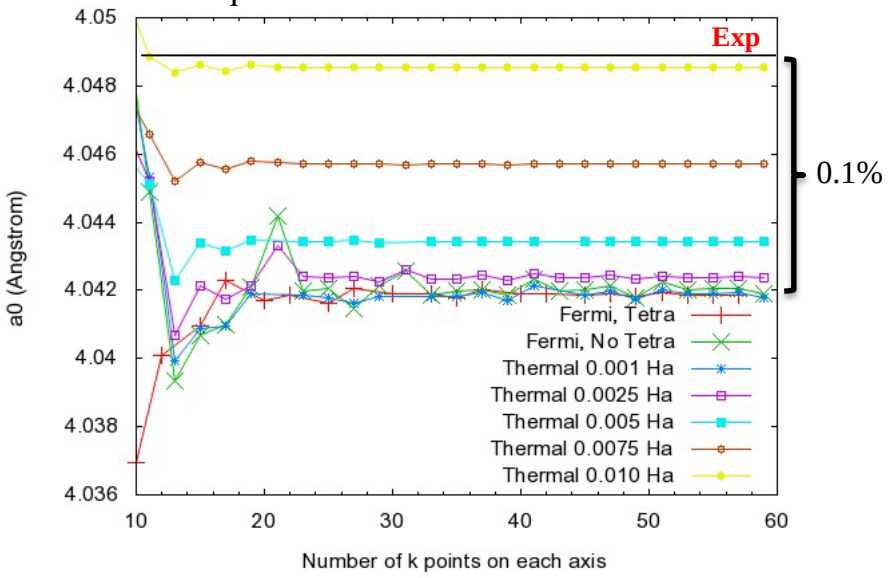
❖ To make these findings easily accessible to **non expert** users

# Physical properties versus **k point integration** and **smearing method**

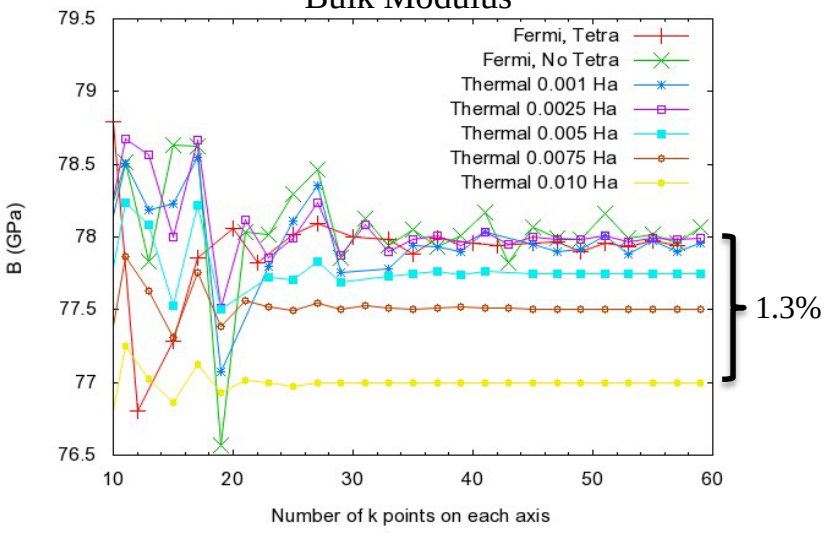
- Aluminum
- **PBE** exchange-correlation functional
- DMol3 code

Experimental values (T room):
- Lattice constant: 4.049 Å
- Bulk Modulus: 76 GPa

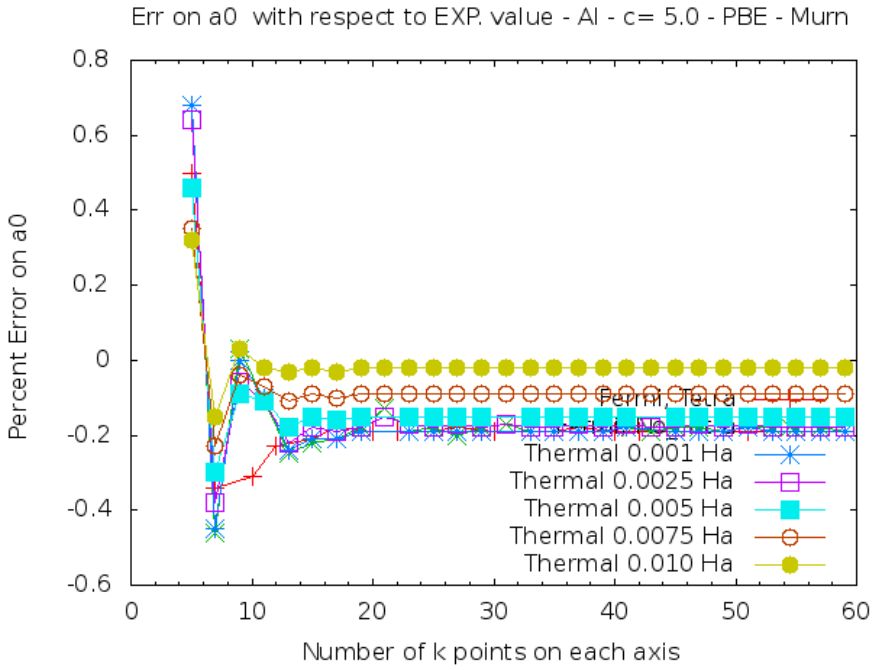
Equilibrium Lattice Constant

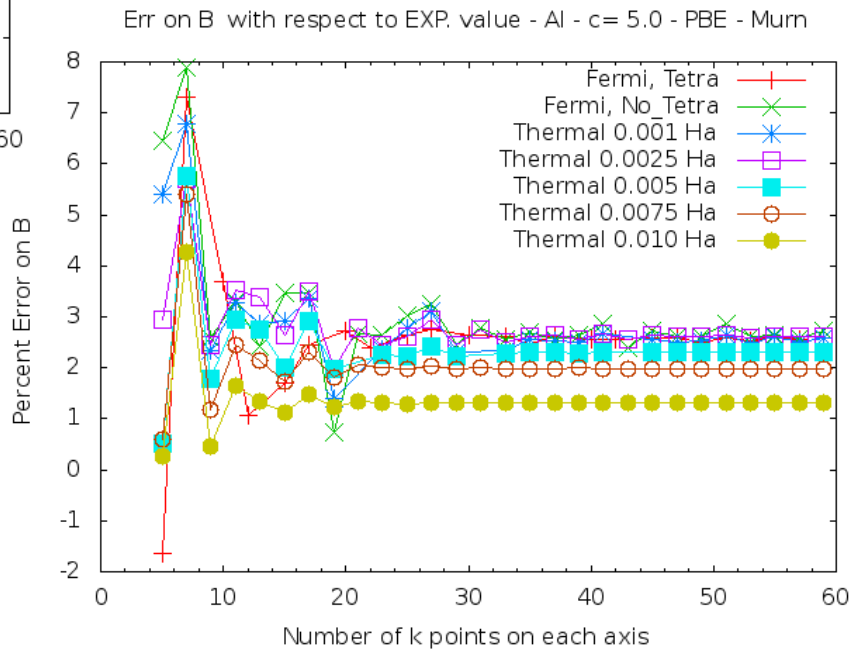



Bulk Modulus




- The onset of the k-point convergence depends on the physical quantity
- The smearing method controls: - the onset of k-point convergence (computation time)
  - the value of the physical quantity



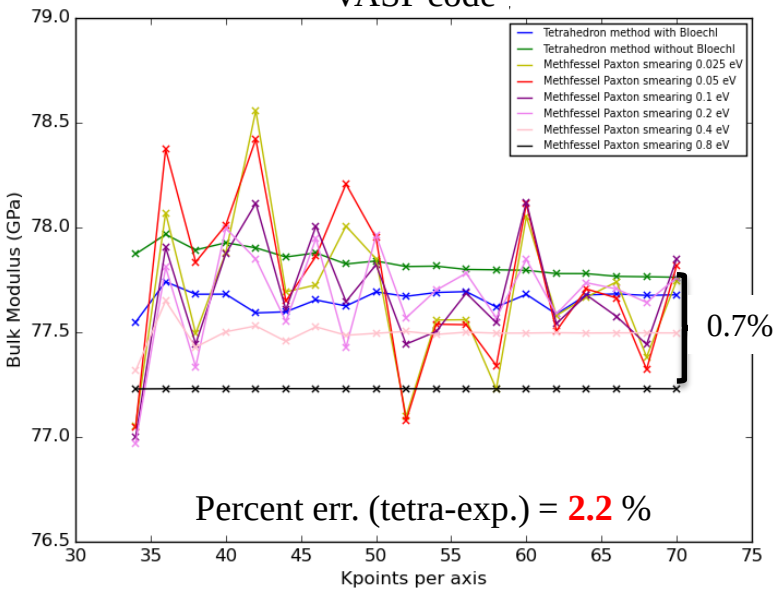



# Does it depend of the **code?**

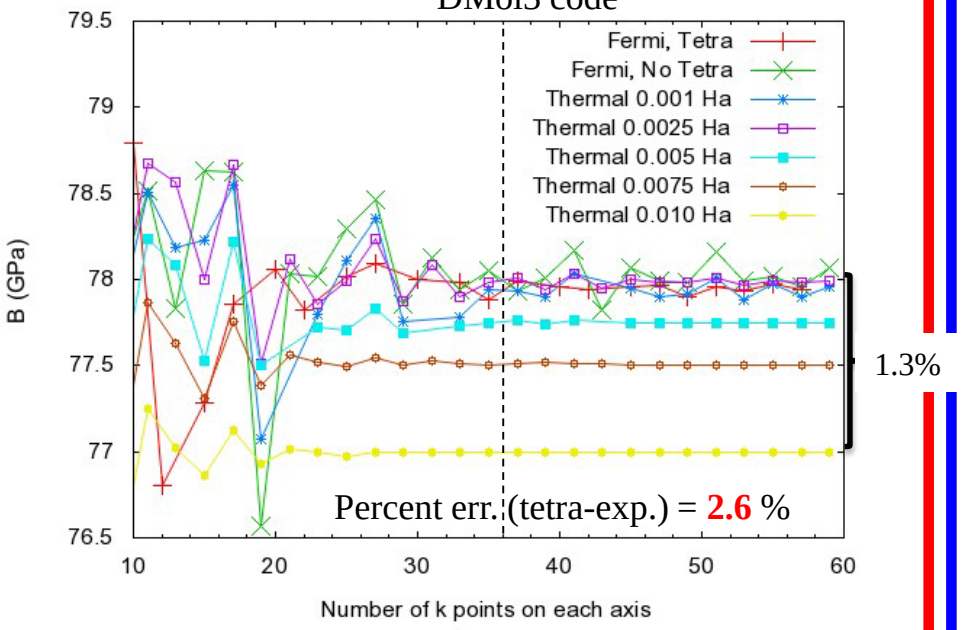
- **Bulk Modulus**
- Aluminum
- **PBE** exchange-correlation functional

Experimental value (T room): 76 GPa

VASP code



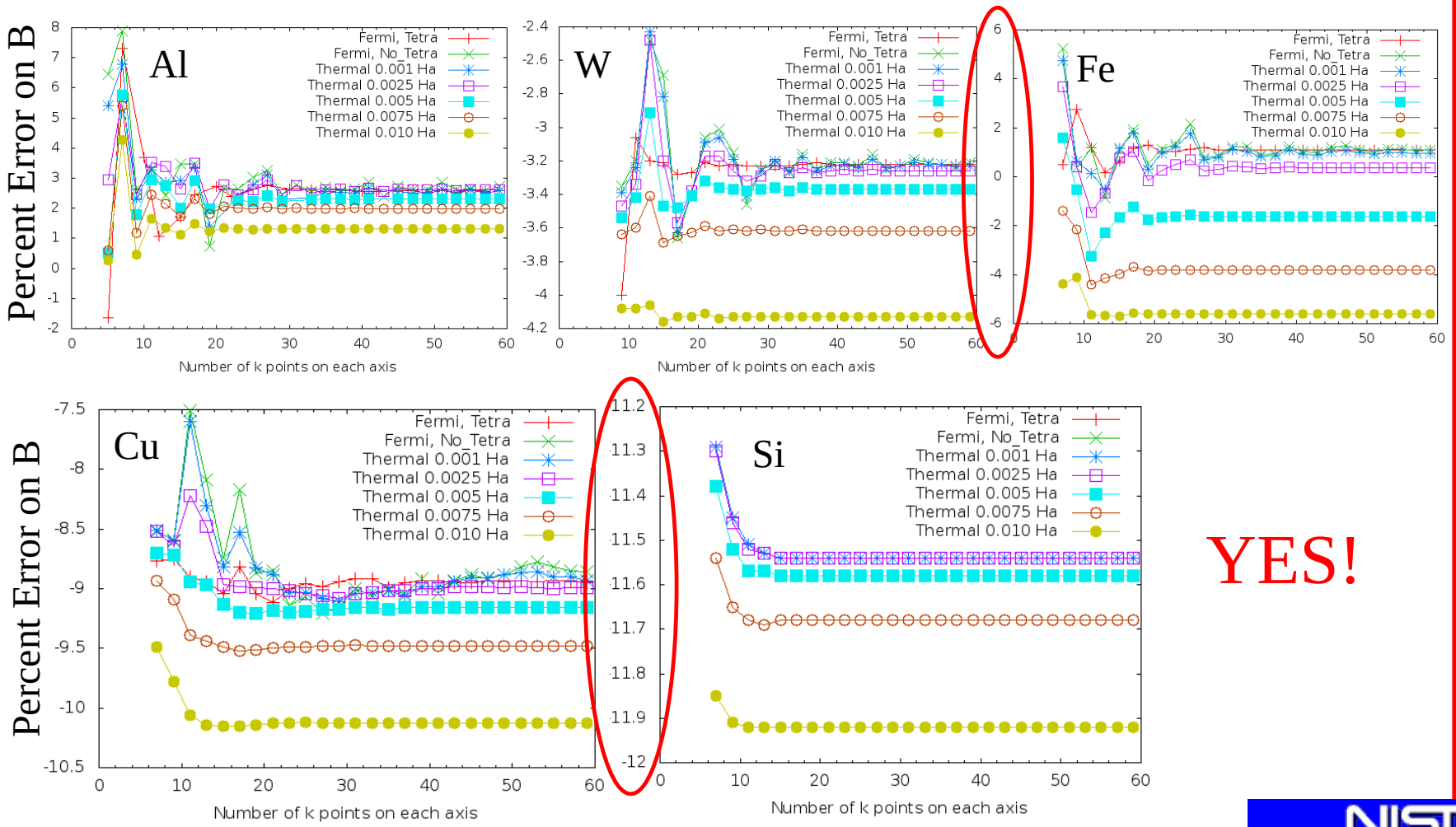

DMol3 code




NOT significantly

# Does it depend of the **element?**

- **Bulk Modulus, PBE** exchange-correlation functional, DMol3 code, c=5.0

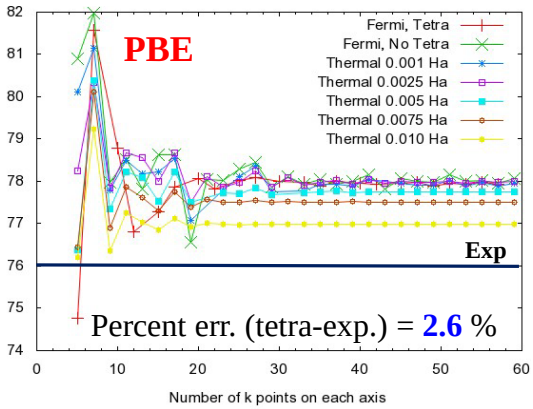


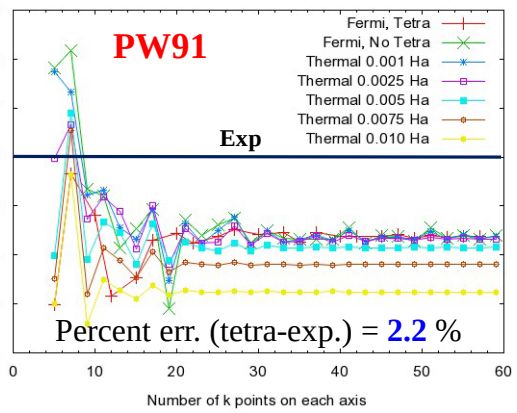

YES!

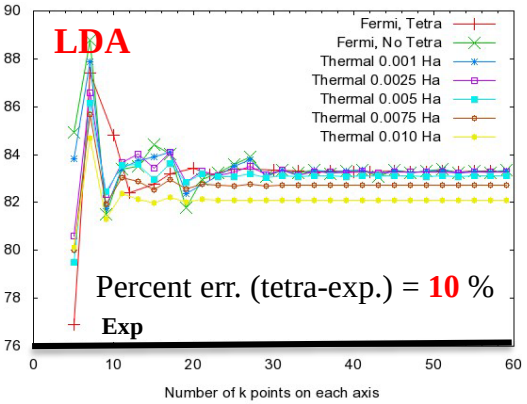
# Does it depend of the **exchange-correlation** functional?

Al Bulk Modulus Experimental value: 76 GPa

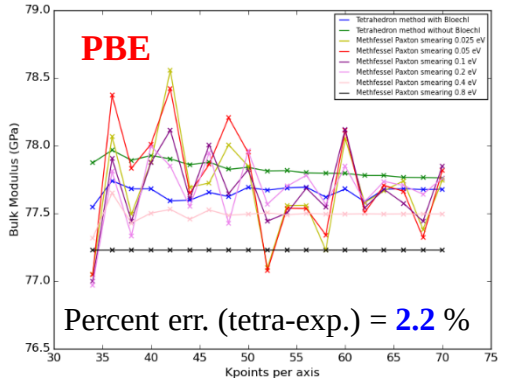
DMol3

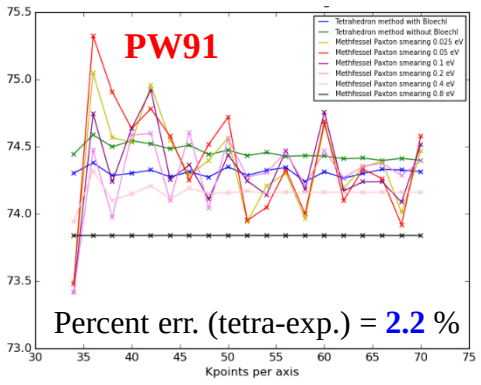


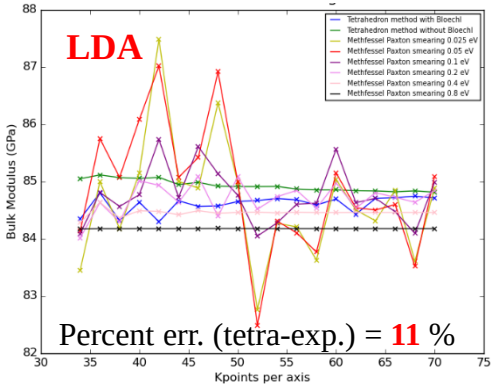




VASP



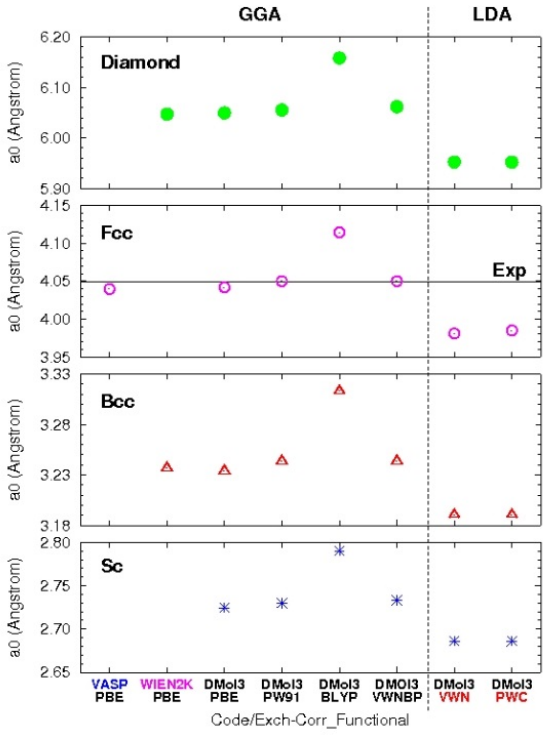



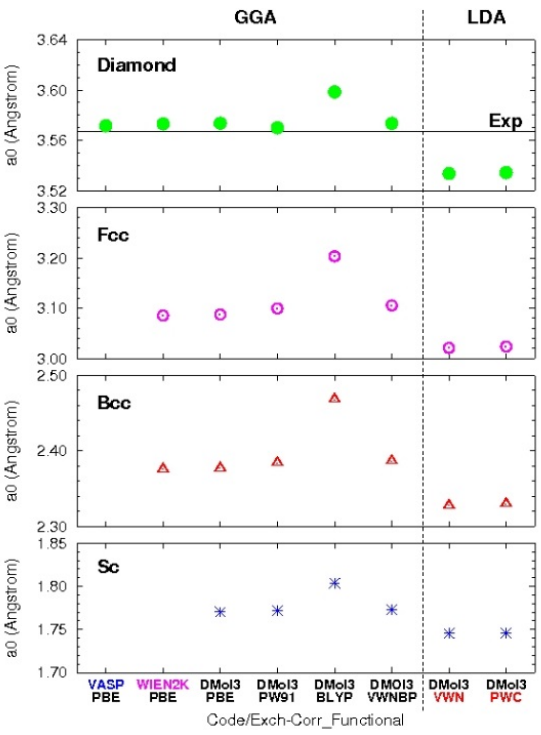

# YES!!!

# Lattice constant for **various elements vs $E_{xc}$** (at converged kp)

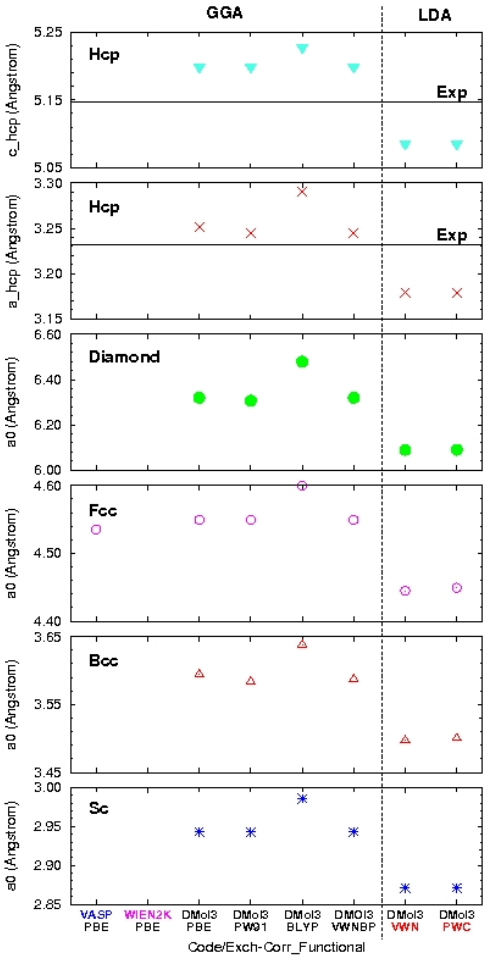
Aluminum




Carbon



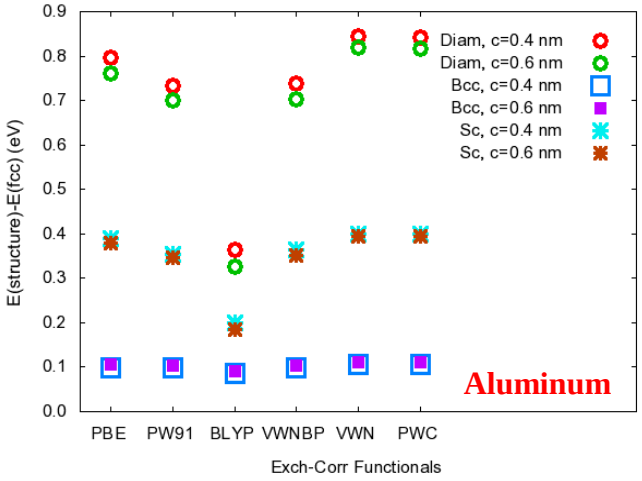

Zirconium

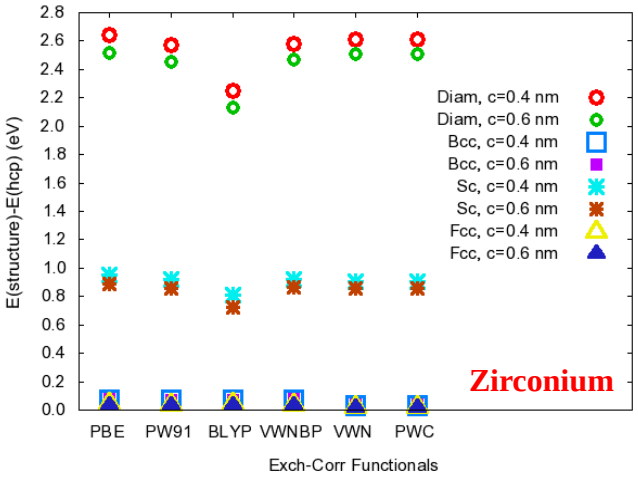


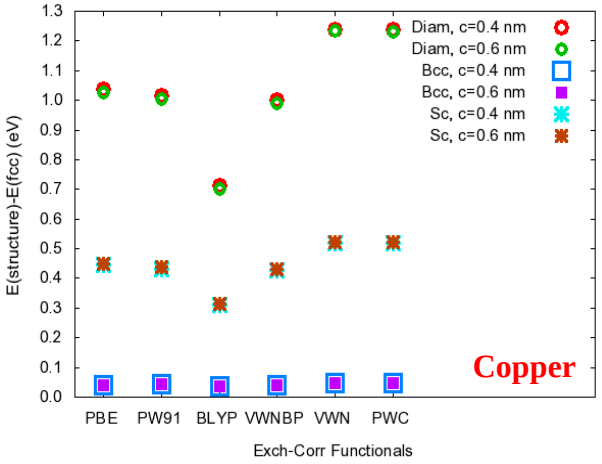

Systematic dependence on $E_{xc}$ functional choice

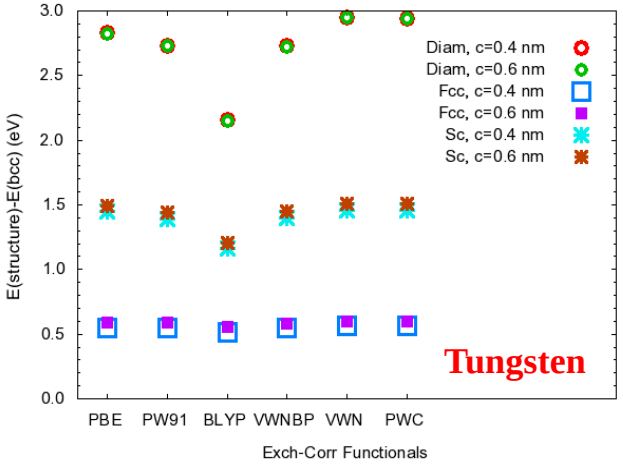
➔ possibility to estimate **"transferable" uncertainties**

# **Energy** for various elements (metals)



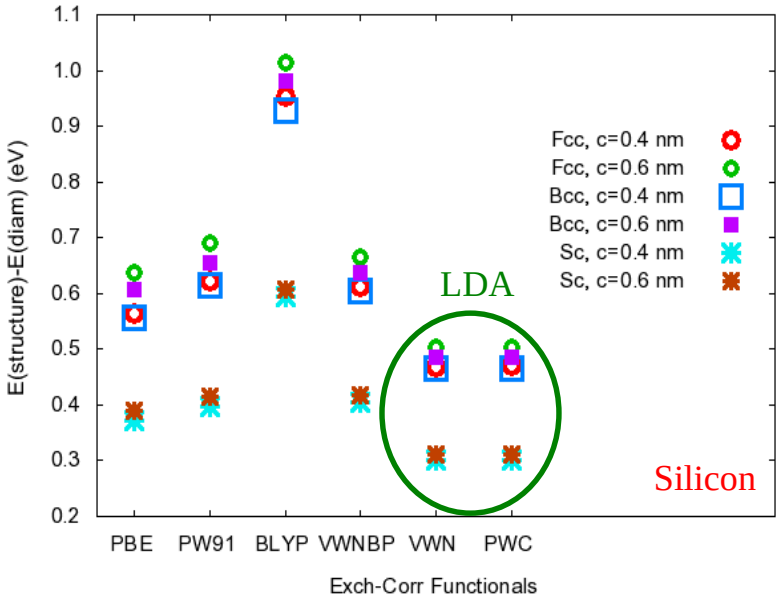



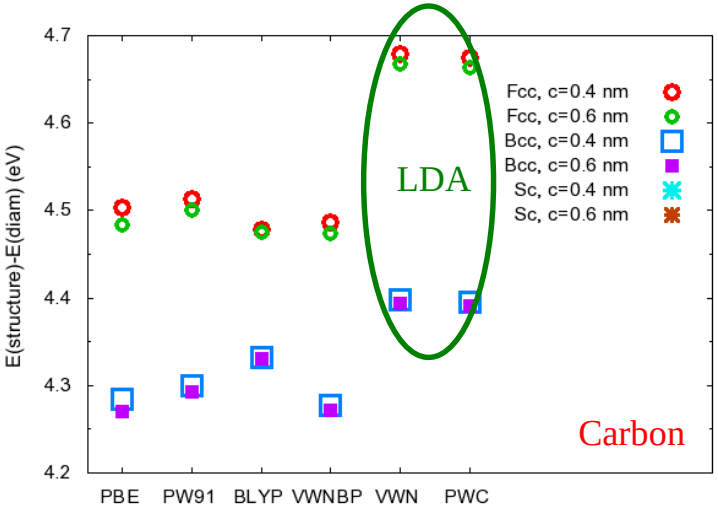



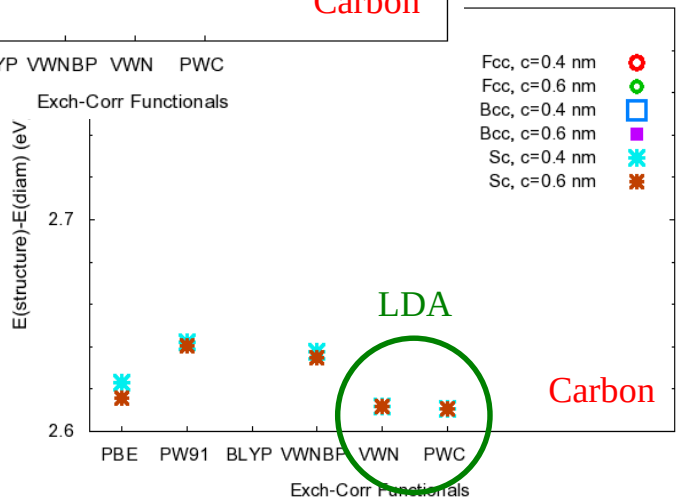

- Always the same ordering!!
- Semi-systematic dependence on $E_{xc}$

All DMol3 calculations

# **Energy** for various non-metallic elements




- Always the same ordering!!
- NOT systematic dependence on $E_{xc}$

**$E_{xc}$ effect is property dependent!!!!**

All DMol3 calculations

# DFT Benchmarking WEB-INTERFACE

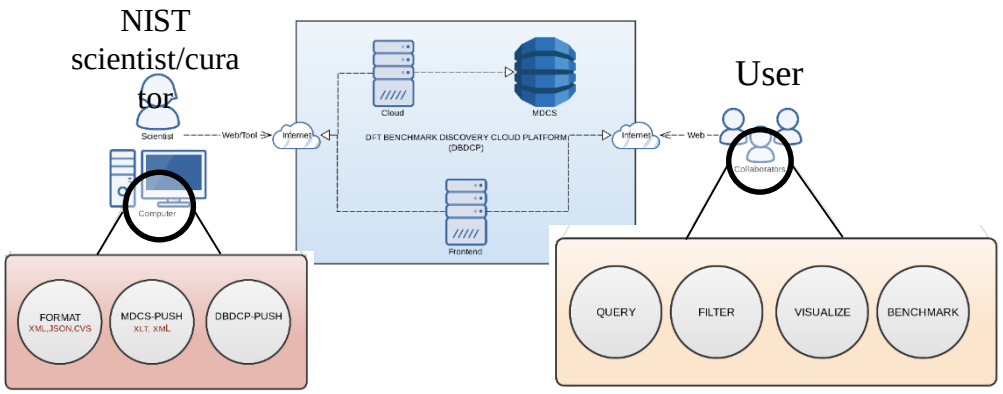
**NIST Scientist(s):**

1. Computing data
2. Posting data/curating database:
   - **Access** using the web browser to custom manage users: data, meta-data and a discussions forum with users.
   - **Upload** new data using a tool that will synchronize the scientist's local repository with the curator and, therefore, updates the platform in real time
   - **Create** custom access for collaborators if needed.

**Users:**

Users will be able to:

- access the platform only from a web browser.
- **query, filter, visualize and access** the benchmark data.



- Large MULTIDIMENTIONAL space:

  Using the platform the user can virtual cross-filter/plot/download any combination of parameters ➔ **huge number of combinations** and plots

- Search options:

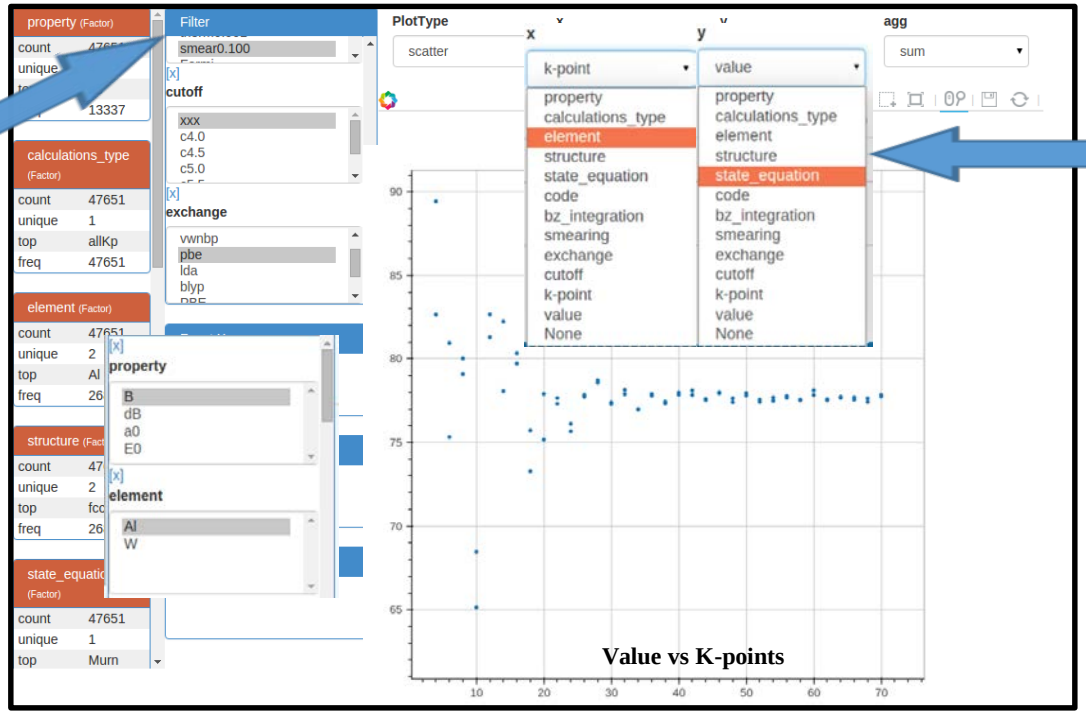
**1)** **defaults parameter sets** are provided for non-expert users or quicker investigations;

2) it is **possible to specify every single** computational **parameter**, if so desired;

3) default configuration can be overwritten by the user (combination of 1) and 2));

4) parameter combinations that don't make sense for the chosen domain are restrained

Multiple possible **FILTERs/ defaults** available




Multiple options for **PLOTTING variables**

# Conclusions

- **DFT needs uncertainties**
- To estimate/predict such uncertainties is definitely not trivial:
  - Large multidimensional parameter-space needs to be explored
  - Certain “parameters” have a much larger effect than others
  - Material dependent

- Future: to extend the scope from:
  - single elements to (at least) binary compounds
  - mechanical to electronic properties

- We are collecting data
- We are showing the need for UQ in DFT
- We **need help** with the next steps:
  - to compute the uncertainties
  - how to propagate uncertainties