Updates on high-throughput DFPT

Guido Petretto, Matteo Giantomassi, Henrique P. C. Miranda, Michiel J. van Setten, David Waroquiers, Shyam Dwaraknath, Donald Winston, Xavier Gonze, Kristin A. Persson, Geoffroy Hautier, Gian-Marco Rignanese

MODL, Institute of Condensed Matter and Nanosciences, Université catholique de Louvain, 1348 Louvain-la-neuve, Belgium

22/05/2019

Outline

Outline

Aim of the project

Diffusion of large databases based on DFT calculations

$\Downarrow$

High-throughput workflows for Abinit

$\Downarrow$

DFPT phonon band structures at Materials Project

Where were we?

ABIDEV 2017:

- Infrastructure to run high-throughput calculations with Abinit
  - dependencies on different python frameworks
  - high-throughput framework: Abiflows
- Preliminary results:
  - Convergence study
  - Workflows at NERSC

Where were we?

ABIDEV 2017:

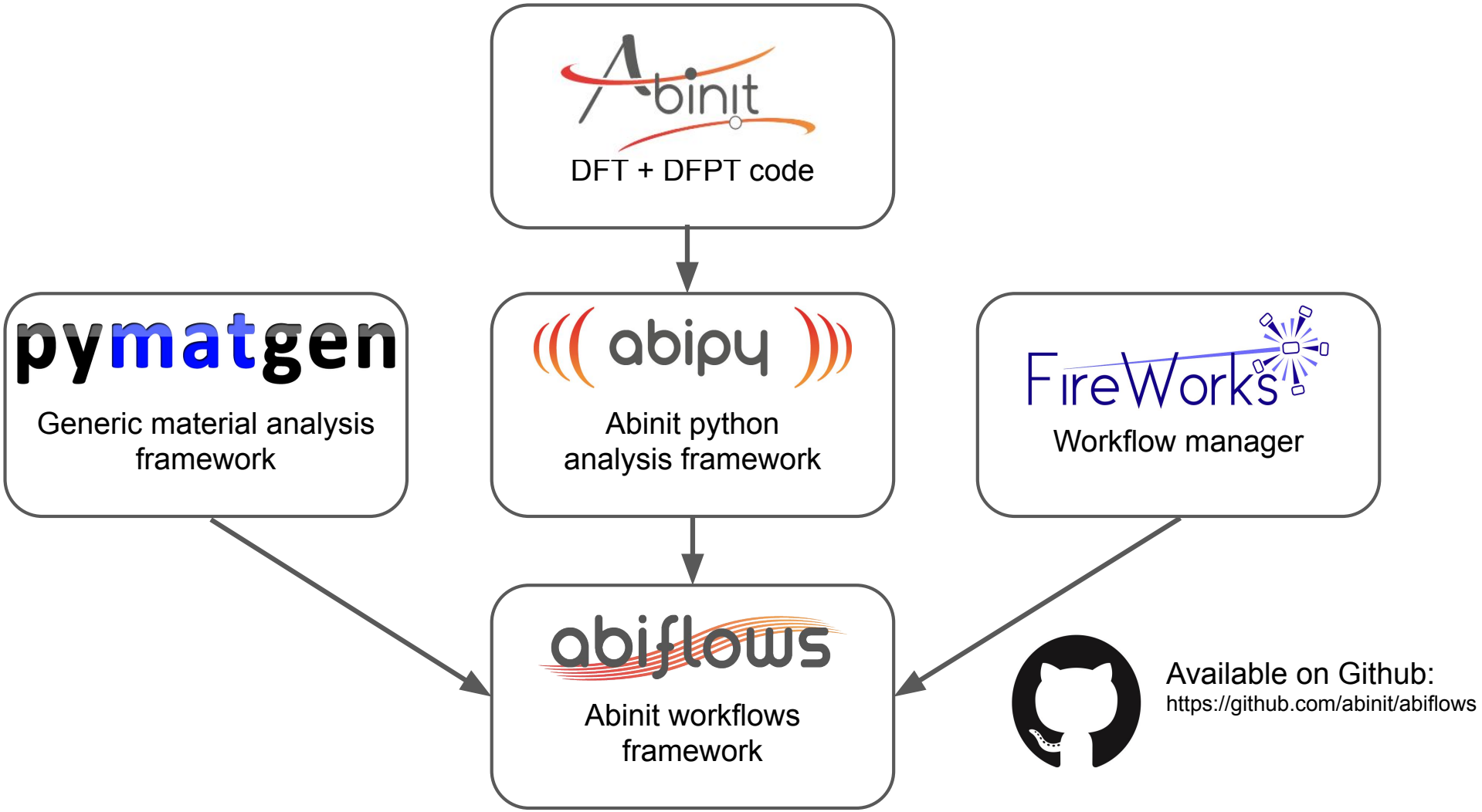
- Infrastructure to run high-throughput calculations with Abinit
  - dependencies on different python frameworks
  - high-throughput framework: Abiflows
- Preliminary results:
  - Convergence study
  - Workflows at NERSC

ABIDEV 2019:

- Results
  - Materials project
- Problems encountered
- Next steps

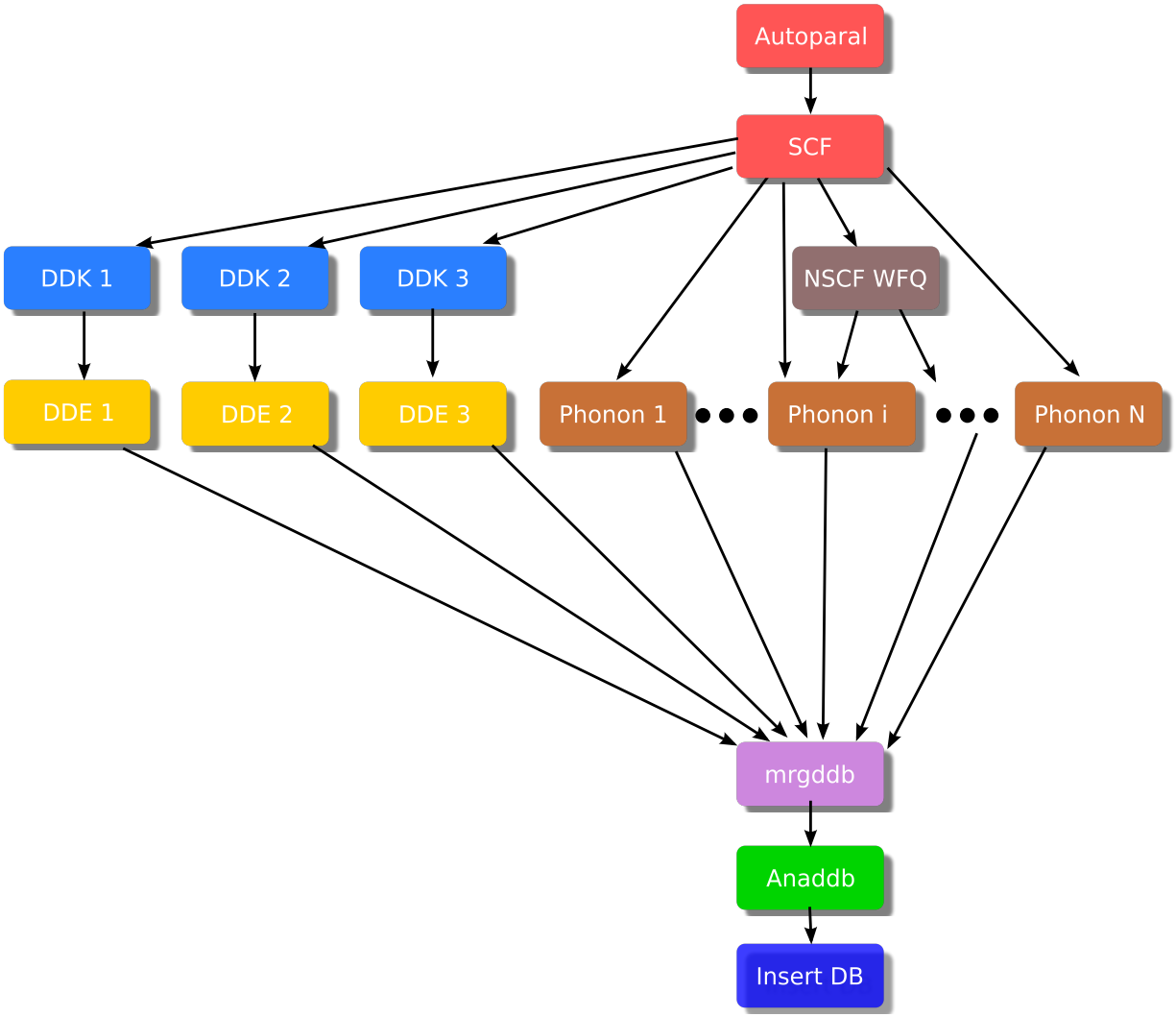
Outline

High-throughput framework



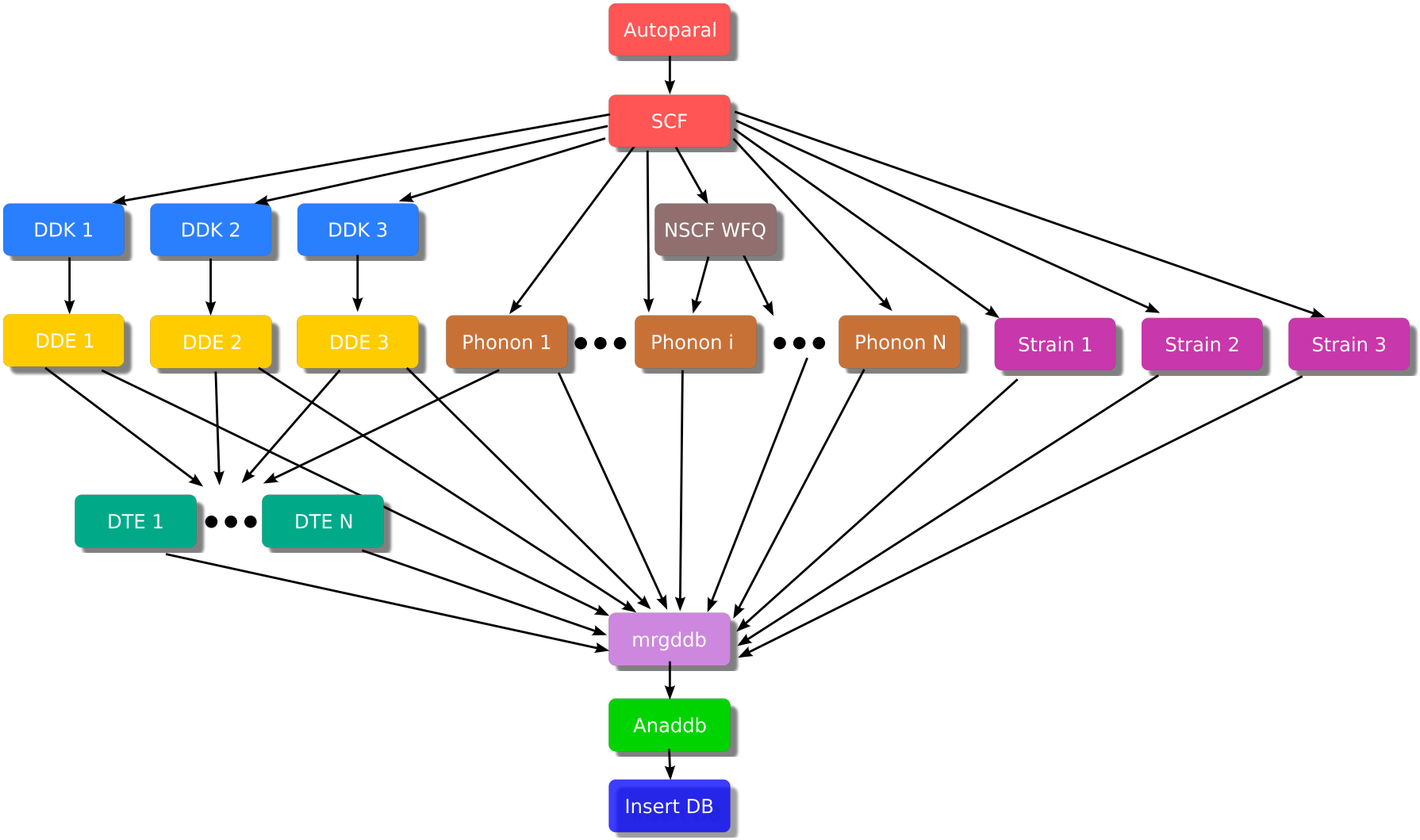
DFPT workflow

Extended workflow to cover all possible DFPT calculation available



DFPT workflow

Extended workflow to cover all possible DFPT calculation available



Convergence study

Find optimal k-points and q-points sampling for high-throughput

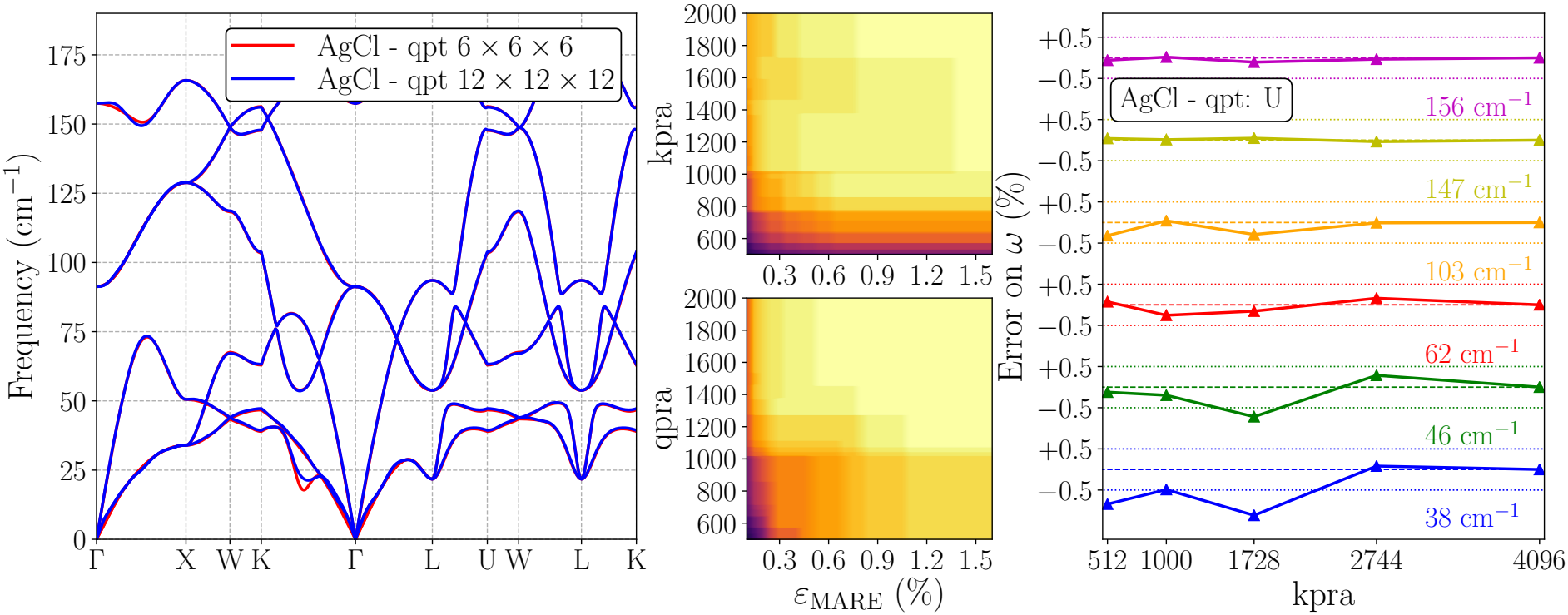
Petretto, Gonze, Hautier, Rignanese, *Comp. Mat. Sci.*, 144, 331 (2018)

- Set of 48 semiconductors
  - Various sizes, crystal symmetries, gaps
- Several K and Q grids
- Statistic on error with respect to dense grids:
  - relative and absolute error
  - mean and maximum error

NaLi2Sb	Ca(CdP)2	CdS	SrLiP	InS	GaN	RbYO2
SiO2	BP	AlSb	LiZnP	MgCO3	ScF3	ZnGeN2
LiMgAs	P2Ir	Si	Li3Sb	K2O	Ga3Os	Be3P2
ZnSe	MgO	AgCl	SiC	YWN3	SrO	PbF2
MgSiP2	SiO2	GaP	Be2C	SnHgF6	MgMoN2	ZnO
ZrSiO4	Ba(MgP)2	Ba(MgAs)2	Ca(MgAs)2	C	RbI	FeS2

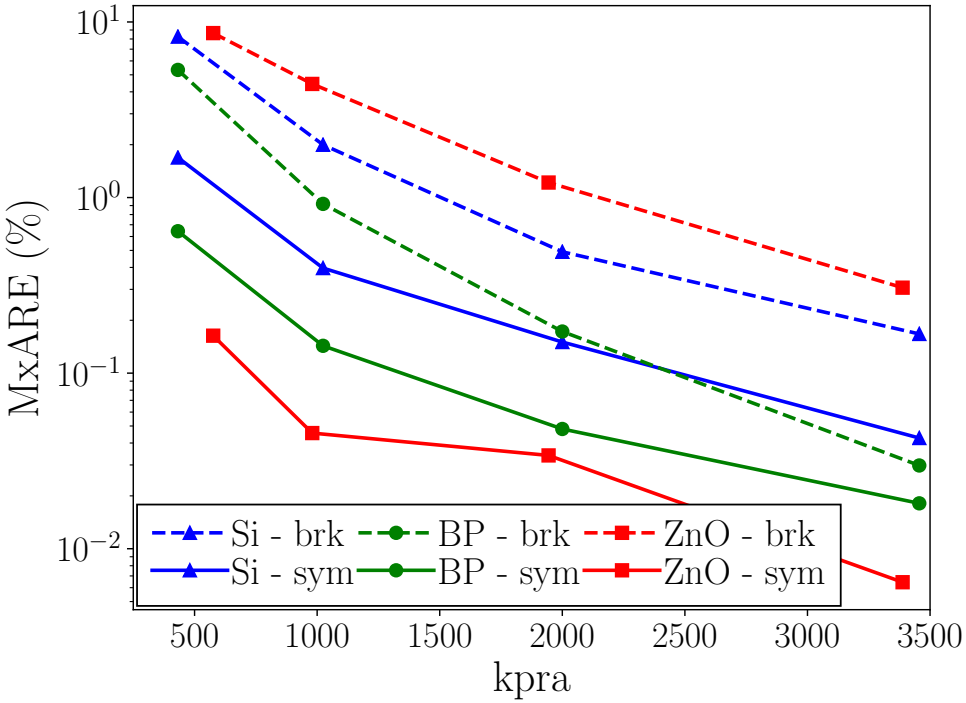
Convergence study: grids density

- absolute and relative errors on ω, E_{at}, Z^* and ϵ
- 1500 points per reciprocal atom $\Rightarrow$ $\mathrm{N_{kpt}}$*$\mathrm{N_{atoms}} \backsimeq 1500$
  - $\Rightarrow$ All materials converged with 0.5 cm^{-1} MAE and 0.6% MARE
- Better using a Q-grid commensurable with K-grid
  - $\Rightarrow$ Smoother close to Γ



Convergence study: symmetry of the grid

Convergence rate of phonon frequencies ω for symmetric versus non-symmetric grids



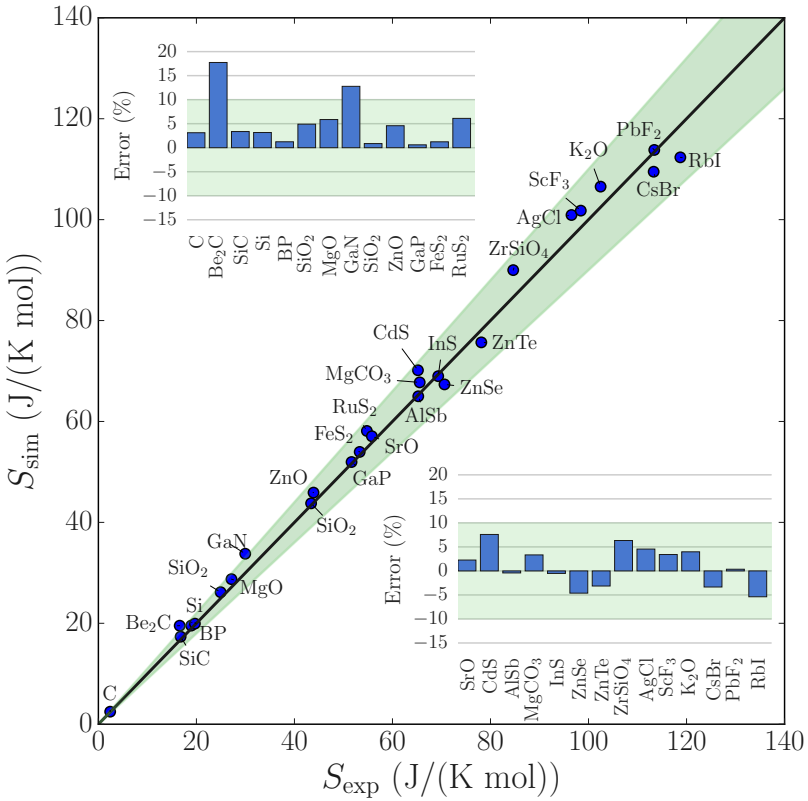
$\Downarrow$

Grids should preserve the symmetries of the system

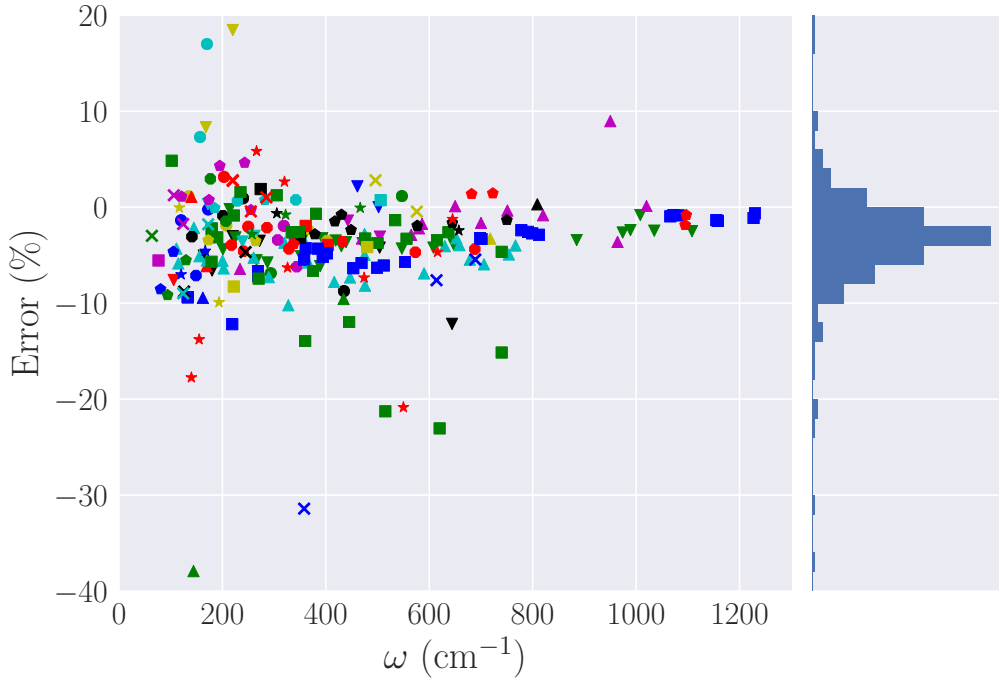
Results validation

Validation versus experimental data

Vibrational entropy at 300K



Γ phonon frequencies



Outline

Materials Project phonons database

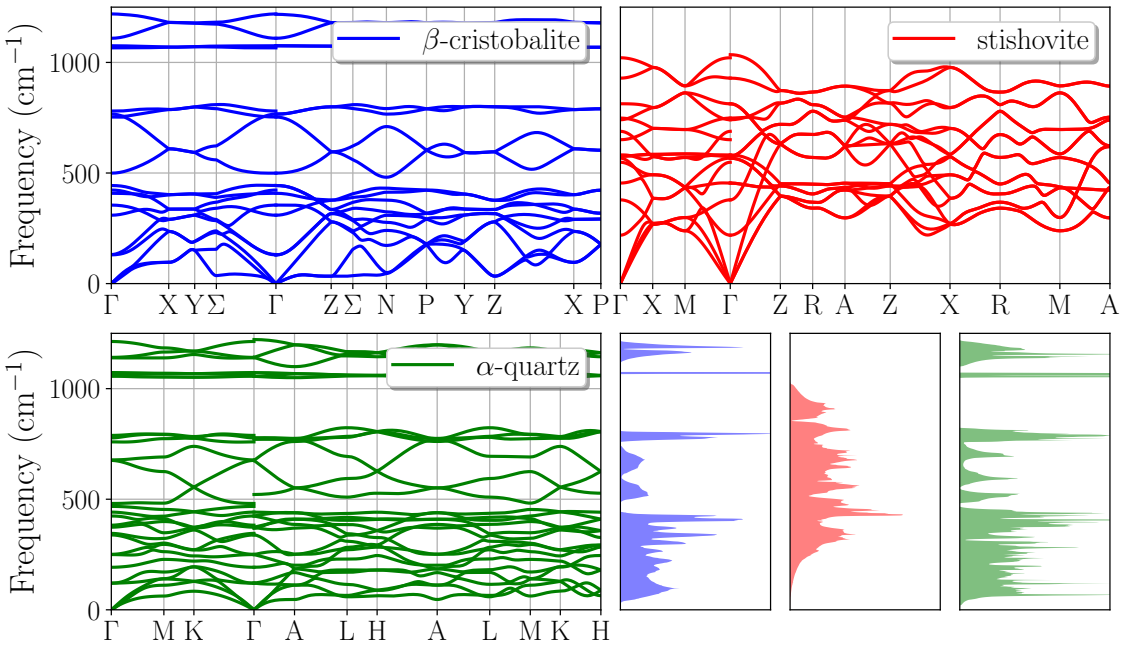
Open access database: 1521 semiconductor materials (and growing...)
1508 of those materials with less than 13 atomic sites
$\sim$ 5M CPU-hours

Materials Project phonons database

Open access database: 1521 semiconductor materials (and growing...)
1508 of those materials with less than 13 atomic sites
$\sim$ 5M CPU-hours

SCIENTIFIC DATA

- Interatomic force constants (DDB files)
- Phonon dispersion
- Born effective charges
- Dielectric tensor
- Thermodynamic prop.:
  - ΔF, ΔE_{ph}, C_v, S



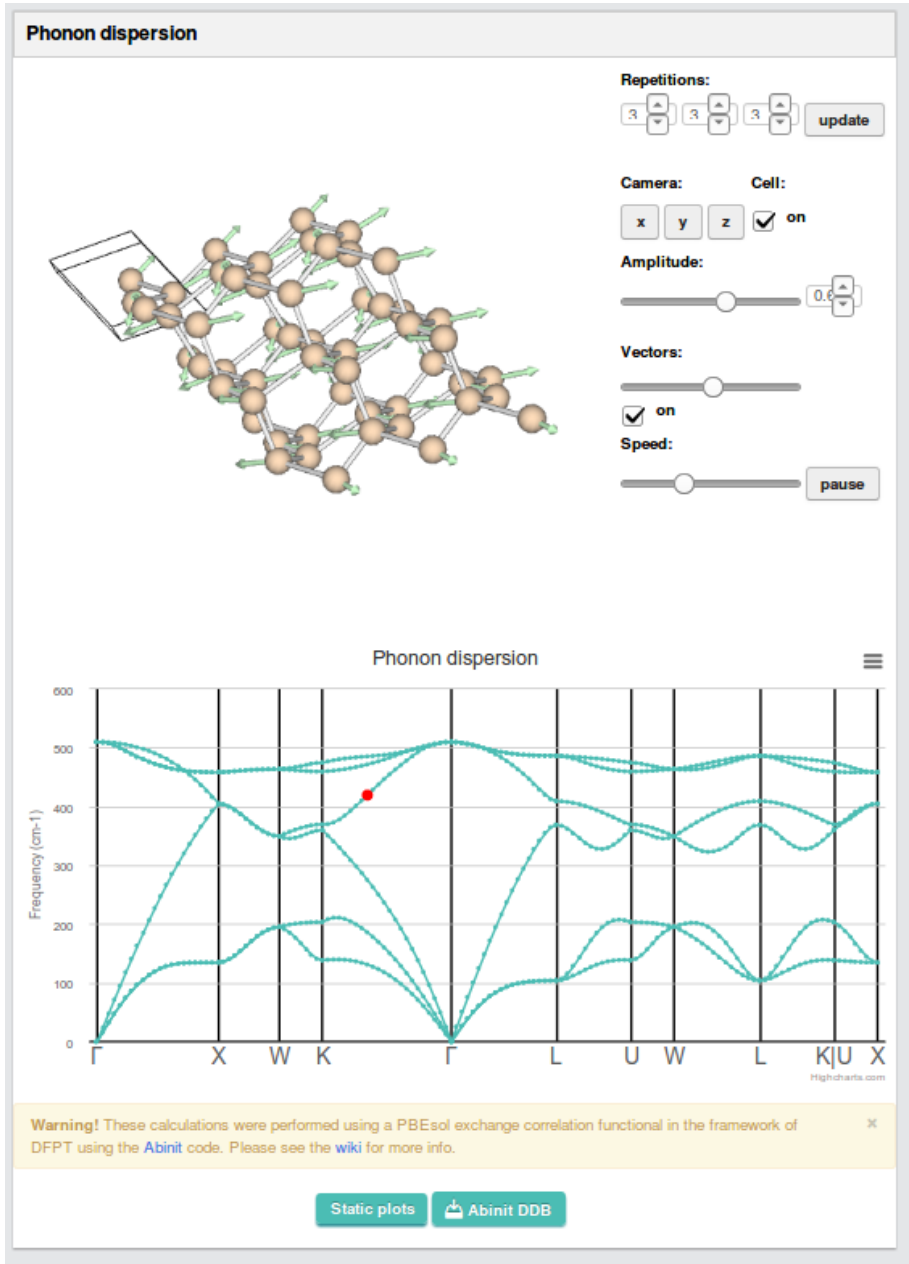
Petretto, Dwaraknath, Miranda, Winston, Giantomassi, Van Setten, Gonze, Persson, Hautier, Rignanese, Scientific Data (2018)

Materials Project phonons database

All the data available on the website and through REST service.

Interactive visualization of the phonon dispersion using the phononwebsite

`http://henriquemiranda.github.io/phononwebsite/`



Outline

Reliability: workflows

For the 1521 phonon band structures:

- Relax workflow
- Phonon workflow (+ anaddb)

Out of all the submitted workflows only ~ 30 did not complete successfully

Reliability: workflows

For the 1521 phonon band structures:

- Relax workflow
- Phonon workflow (+ anaddb)

Out of all the submitted workflows only ~ 30 did not complete successfully

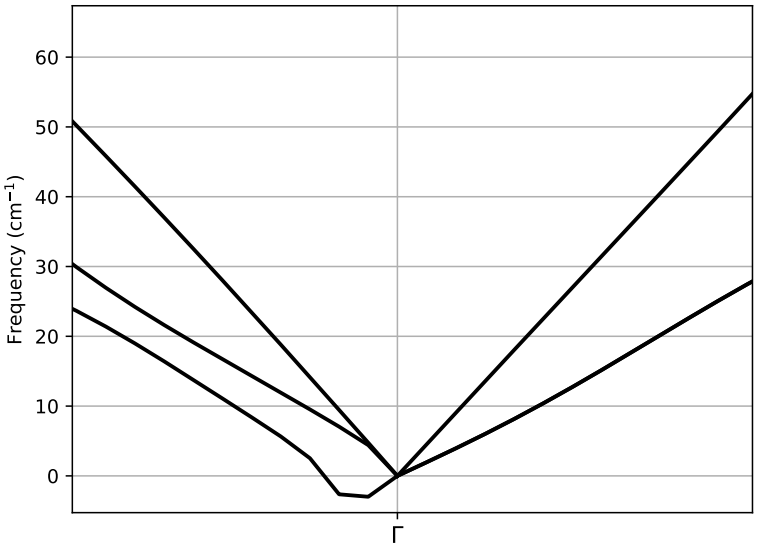
Main reasons:

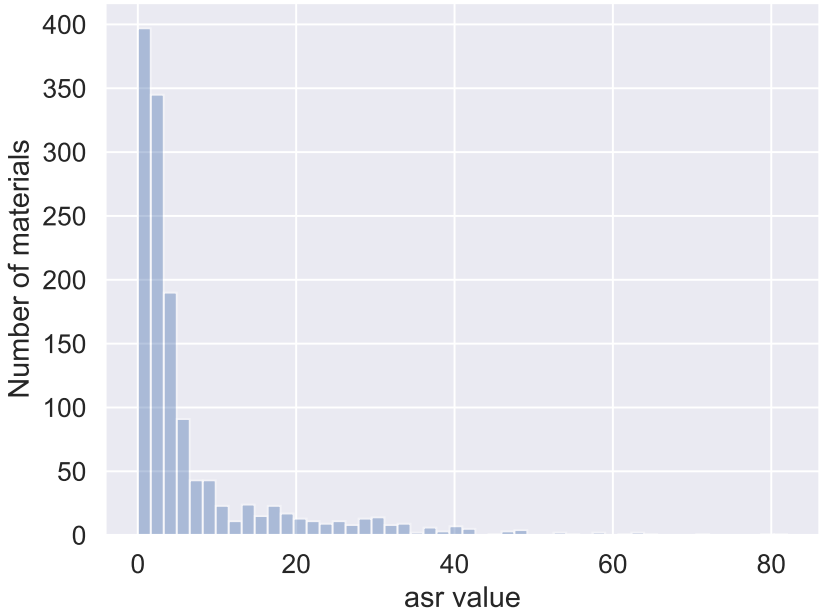
- Too slow relaxation/relaxation did not converge
- Too small gap (switched to metallic)
- Presence of La
- Poor choice of materials

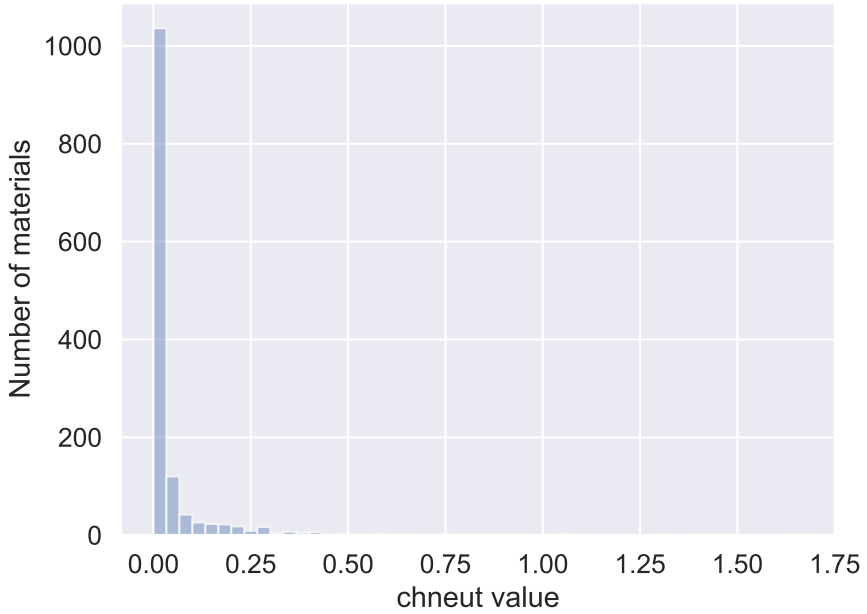
Reliability: results

Warnings available in the database:

- Negative ω close to Γ: 24 materials
- ASR break $>$30 cm^{-1}: 72 materials
- CNSR break $>0.2e$: 92 materials







Problems encountered

- Current MP cluster: KNL nodes
  - not optimized
  - reserve full node
  - relatively poor performances
  - difficult to fine tune parallelization at high-throughput level
- Relax - *ionmov* 22: seems faster but may fail at small *tolmxf* ($1e^{-6}$)

  $\Rightarrow$ switch *ionmov* at python level.
- Autoparal for DFPT
  - always gives the maximum number of preocesses allowed
  - often parallelizing over just the k-points is advantegeous

  $\Rightarrow$ could be improved?
- Memory
  - moving to larger materials already caused a few jobs to fail due to memory issues

  $\Rightarrow$ might be needed to rely on estimation of total memory

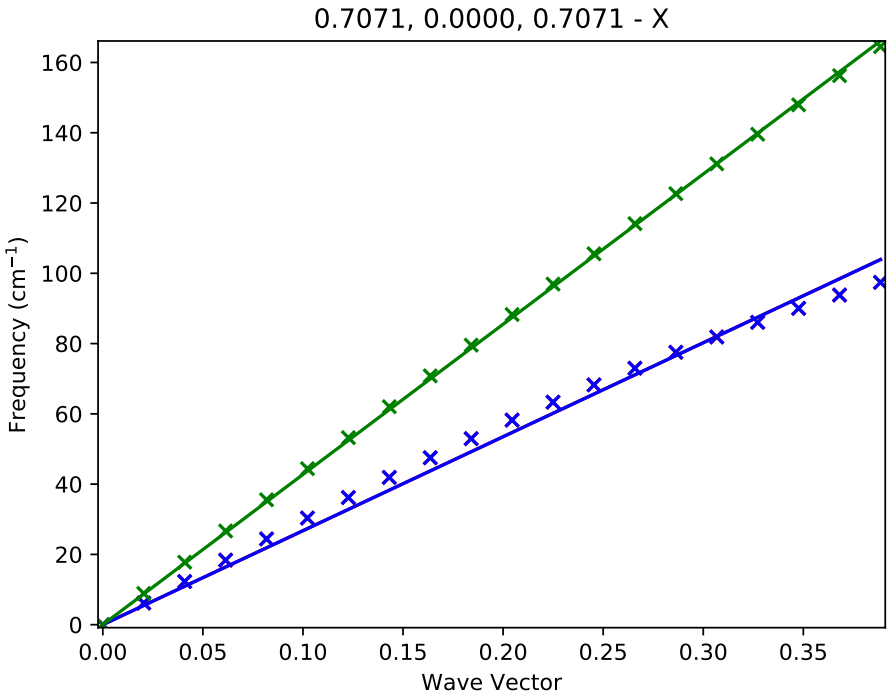
Outline

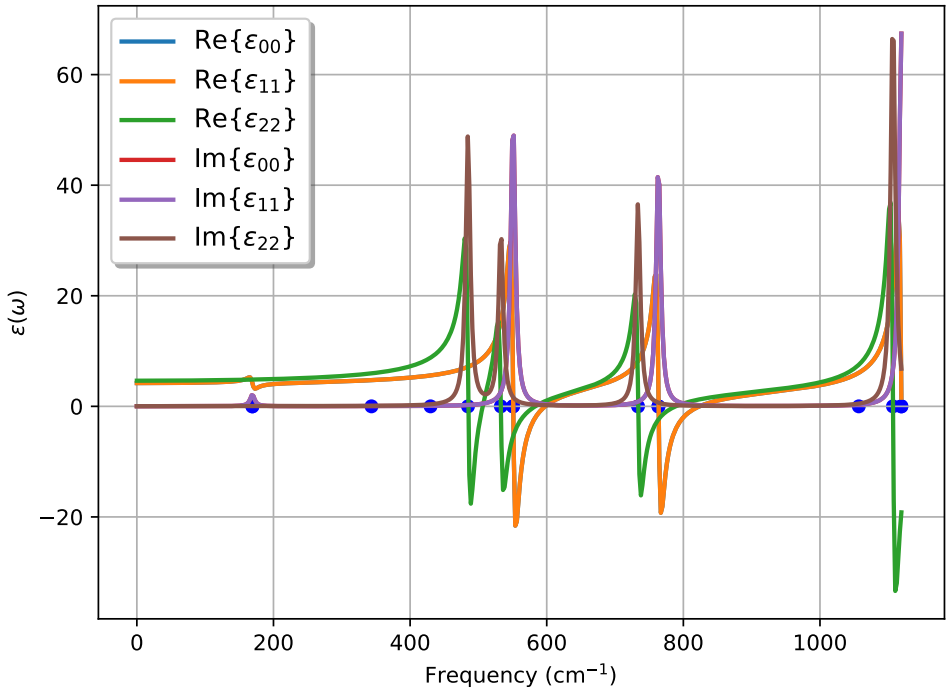
More data on the MP database

Calculations have proceed: almost 500 more materials

More physical quantities will be extracted from the phonon data:

- Sound velocity as slope of acoustic modes
- Low-frequency dielectric permittivity tensor $\epsilon_{\alpha\beta}(\omega)$
- Thermal displacement ellipsoids (Debye-Waller)





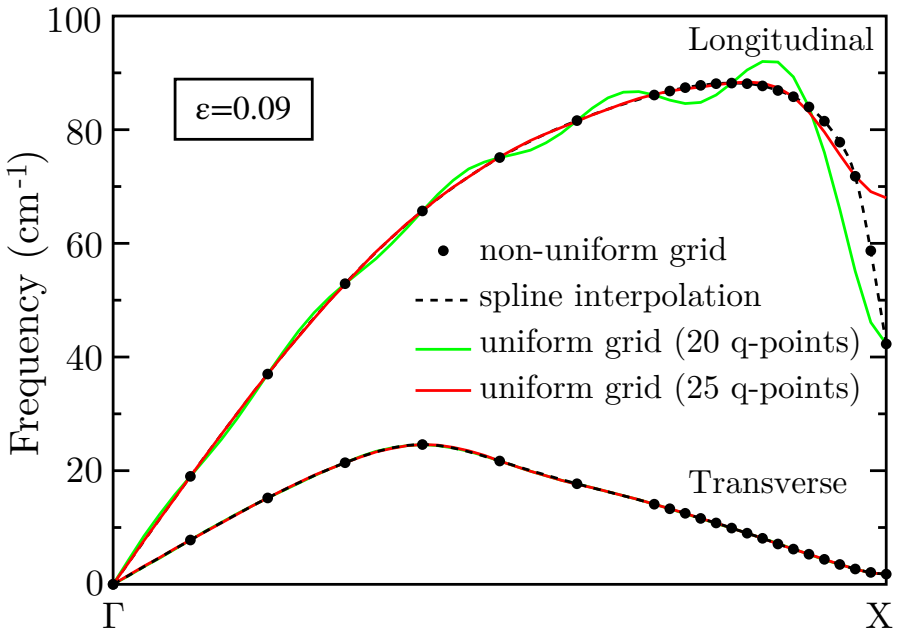
Phonons for metals

Extend the calculation to metals as well. Materials project:

- 24356 metals
- 8242 with less than 6 atoms

Potential issues:

- Denser k-point grids
- Q-point grids?
- Smearing
- Kohn anomalies
  - Fourier interpolation



He, Liu, Li, Rignanese, Zhou (2019)

⇓

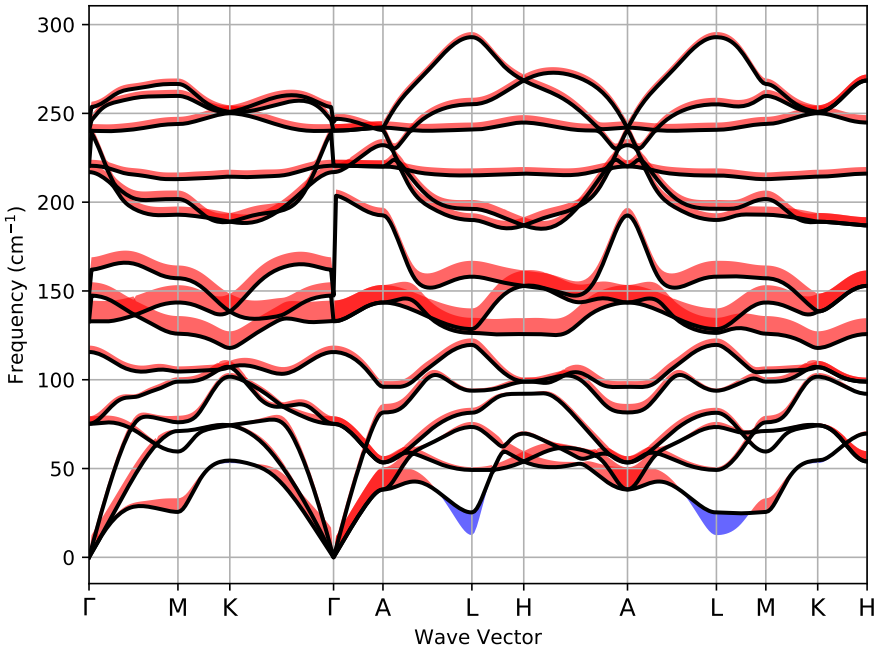
New convergence study required

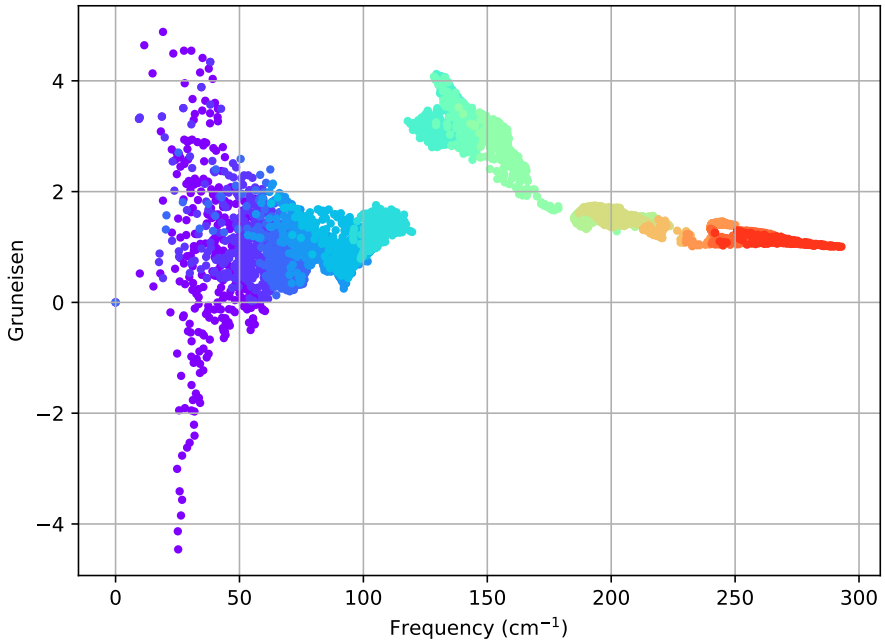
Volume: Grüneisen parameters

Phonons at different volumes (e.g. $\pm 2\%$) $\Rightarrow$ Grüneisen

- Tools already available in Abinit (netcdf) and Abipy
- Preferable to have separate workflows

```
g = GrunsNcFile.from_ddb_list(["-2_DDB", "+0_DDB", "+2_DDB"])
g.plot_phbands_with_gruns(with_doses=None)
g.plot_gruns_scatter()
phbst.plot_phbands(units="cm-1")
```





Volume: Quasi-Harmonic Approximation

Tools for QHA implemented in Abipy:

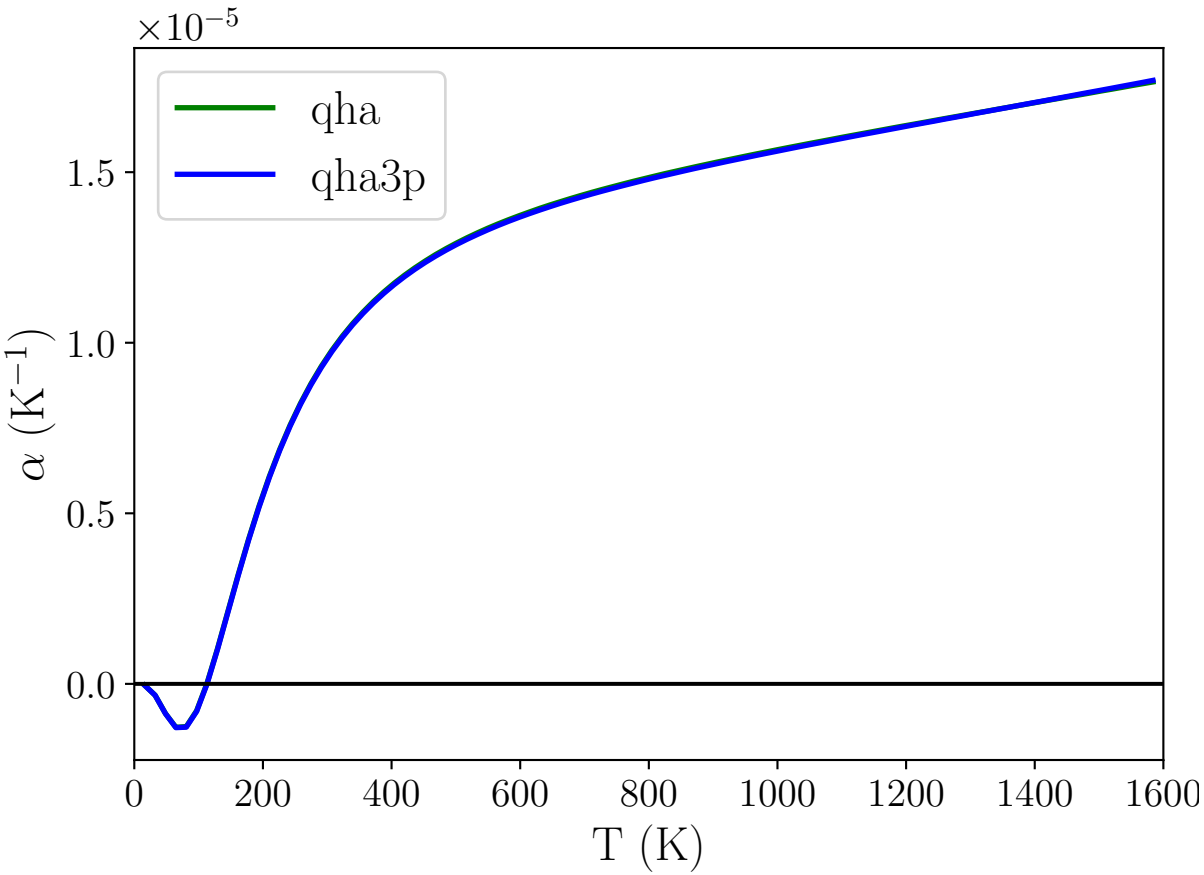
- Standard QHA object
  - generated from GSR and PHDOS netcdf files
  - Fittings
  - Thermal expansion coefficient
  - Interface to Phonopy for further functionalities
- Cheaper QHA: QHA-3P (Nath *et al.* arXiv:1807.04669)
  - several electronic energies at different V and 3 phonon calculations
  - extrapolate phonon contribution at other V
  - Satisfactory results

  $\Rightarrow$ Interesting for high-throughput

Volume: Quasi-Harmonic Approximation

Example QHA-3P for Si

```
qha = QHA.from_files(gsr_paths, dos_paths)
qha3p = QHA3P.from_files(gsr_paths, gruns_path, ind_doses=[1,2,3])
fig = qha.plot_thermal_expansion_coeff()
qha3p.plot_thermal_expansion_coeff(ax=fig.axes[0])
```



Thank you for your attention

High-throughput framework

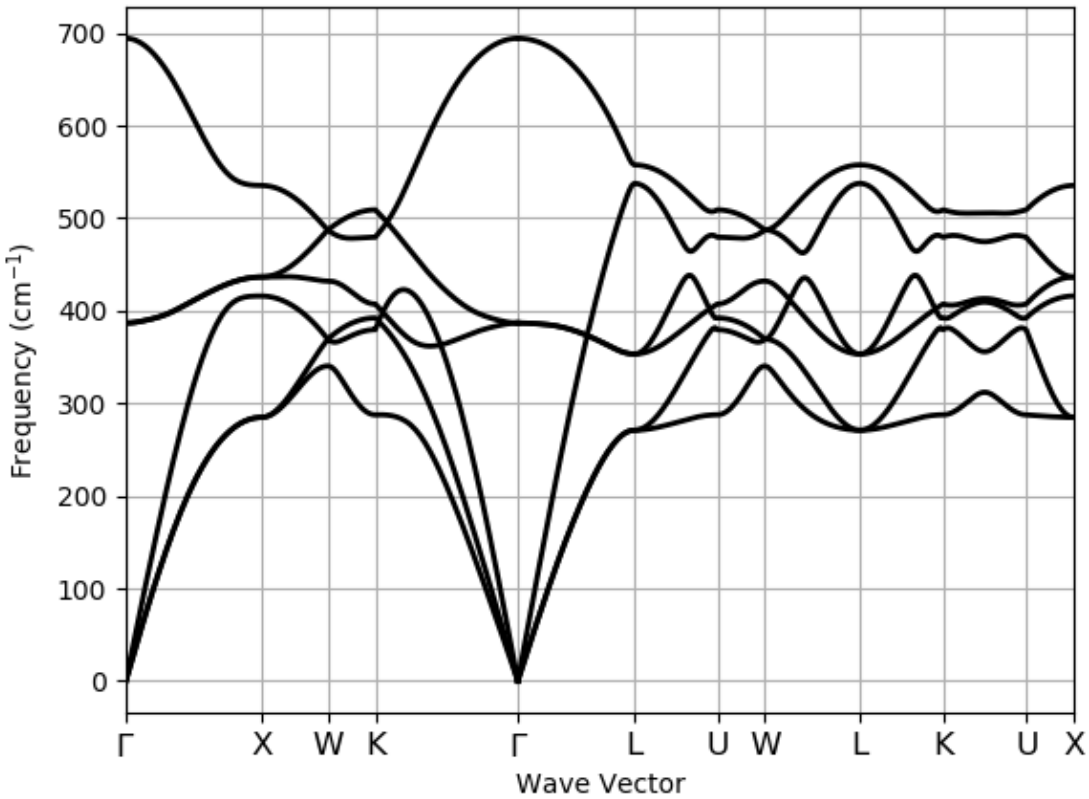
What do we need for high-throughput with abinit?

- Python interface to DFT codes pymatgen
- Inputs
  - pseudopotentials and cutoffs Pseudo Dojo
  - automatic generation abipy
- Workflow management FireWorks
- Database interface mongoengine
- Workflows abiflows
- Error handling and data analysis abipy

Materials Project phonons database: rester

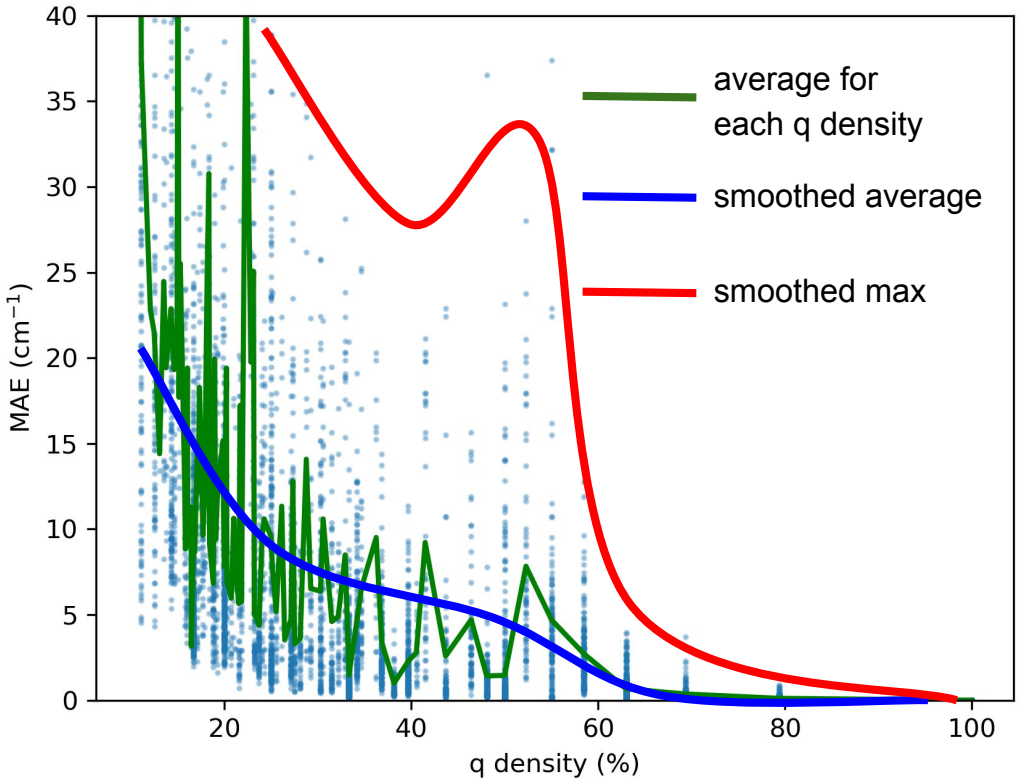
Fetching DDB files from MP and analyze results with Abipy

```
ddb = abilab.DdbFile.from_mpid("mp-1265")  # MgO
phbst, phdos = ddb.anaget_phbst_and_phdos_files(ndivsm=20, nqsmall=20,
                                                lo_to_splitting=True)
phbst.plot_phbands(units="cm-1")
```



q-points convergence

Use subgrids of the q-point grid to check the convergence w.r.t. qpt



- Material dependent
- Suggest good convergence level at 1500 qppa
- reducing by a factor 2 may lead to sizeable errors on average