# Wannier90 v3.0 school, Virtual Edition 2020

## Speaker: Giovanni Pizzi

### 26 March 2020 - Questions and answers session

**QUESTIONS ON SCDM-AIIDA**

- **Is there a way to avoid WFs for the core energy levels? Is it possible to automatically remove the WFs for the core levels?**
    - Yes, one can simply use the 'exclude_bands' feature of Wannier90. It is possible to write a code to identify semi-core levels, even if there is no public code yet (one has to search for almost flat bands, disconnected from all the rest).
- **Since you shift \mu, you can probably decrease \sigma_fit as well? Like \sigma_fit/3 or something like this. Did you investigate this?**
    - Indeed, this could be done. We decided to keep the same sigma as this gives the physical energy bandwidth of the selected orbitals subspace. Other values might work as well (see slide on the validation of the parameters with tungsten (W) to check other possible values.
- **By SCDM+MLWF do you mean performing both disentanglement and maximal localization?**
    - No, actually we just do the Wannierisation procedure - in all our work (arXiv:1909.00433) we do not do disentanglement (except in the first part of the paper where we give a detailed introduction describing the effect of disentanglement or of not using it).
- **Have you tested this workflow for 2d materials?**
    - Yes. We did some preliminary tests. The workflow will get to the end, but unfortunately more than 50% of the systems we tried had wrong band interpolation as obtained from the SCDM method (we only tried ~10-20 common materials, though). We are investigating the reason.
- **With the goal in mind of reproducibility, does AiiDA store information about the compiler, libraries and hardware on which the calculations are run?**
    - AiiDA stores the information on which computer was used to run, and which code executable. By default, in order not to make things too complex for the user, only this is required. But the user can add metadata (called 'extras' in AiiDA), with any relevant information. At the moment we don't do it automatically because it's complex to get a robust command to obtain standardised information on e.g. the libraries that works on any supercomputer and operating system. But suggestions are welcome!

- **It is possible to see the plugins (codes) included in AiiDA?**
    - Yes. AiiDA comes by default with very few plugins, but installing them it's just 1 command. The full list, continuously growing, is on the AiiDA plugin registry: http://aiidateam.github.io/aiida-registry (in March 2020, we have 48 plugin packages registered, supporting over 96 code executables and providing almost 80 workflows).
- **By writing a workflow you implicitly define a standard for computing a certain property(ies). Do you give guidelines on how to create these standards/ontology?**
    - We are part of the OPTIMADE consortium (that include also most of the major materials databases existing today, and is open to new members) to have a standard to e.g. retrieve crystal structures, with the *same* API for all of them.
      For what concerns defining standard protocols, this is work in progress. In the scope of the MaX European Centre of Excellence, we are working on common interfaces to get a given result (e.g. relaxed crystal, electronic band structure, …) with the same API, for many codes (including VASP, Siesta, Quantum ESPRESSO, FLEUR, BigDFT, …). Each of the AiiDA plugin developers for these codes has been implementing turn-key solutions to get these results. Unifying the interface will make them very easy to be used by many people, and also facilitate verification and validation efforts.
- **How much overhead (storage etc) does using AiiDA add?**
    - There is some overhead, but this is typically acceptable. The overhead is very minimal on the files (they are just stored on disk as is). There is some overhead on parsed information and on the database. Anyway it's relatively small, for instance we have a DB of ~10'000'000 nodes (including over 100'000 DFT simulations): the files (inputs, outputs, …) on disk are ~500GB, and the DB is only a couple of GB. Moreover, by default AiiDA will store only the input file and the text/XML output file, trying to avoid duplication, and not storing in the long term big files like the wavefunctions (unless the user wants to).
      In terms of computation, the overhead of a calculation is of the order of 1/10 of a second. So, for DFT simulations, it's completely negligible. Of course, if you want to run very quick (millisecond) simulations, keeping the provenance of each of them in the DB can become a burden, but that's not the goal of AiiDA, and for this goal there are different tools. On the other hand, AiiDA has 2 focuses: 1) automatic tracking the provenance, and 2) not only it has a workflow manager, but this workflow manager can automatically connect to any number of remote computational resources (e.g. via SSH, and AiiDA can interact with the job schedulers), which make it unique in the world of workflow managers.
- **Do you think machine learning could be used to help researchers in creating standards for input parameters?**
    - Yes, even if, for machine learning, it's important to have a very clear control on the quality of the data (e.g. if you have a DB containing a lot of calculations where all parameters change randomly, including pseudopotentials, cutoffs, kpoints, …, it's very difficult for a machine-learning algorithm to learn something

meaningful). Combining ML with AiiDA can be very effective, because AiiDA has all provenance information, so an exact knowledge of the conditions under which the data was generated. See for instance this article: F. Pittino et al., *Prediction of Time-to-Solution in Material Science Simulations Using Deep Learning*, PASC 19 Proc., 10 (2019), where the authors took the data from the 2D materials section of the Materials Cloud (including all AiiDA provenance) and used it for a very different goal, a machine-learning prediction of the time needed for Quantum ESPRESSO to run the simulation, just knowing the input file.