Tutorial on how to run the automated workflow to obtain interpolated bands using Wannier90, the SCDM method and the protocol described in the paper Automated High-throughput Wannierisation by Vitale *et al.*

All data, virtual machine, etc. can be found on the Materials Cloud Archive entry with doi **https://doi.org/10.24435/materialscloud:2019.0044/v2**

This document assumes that you have already properly obtained, installed and configured the virtual machine – if this not the case, please follow first the instructions in the file `README-virtual-machine.txt`

Acknowledgements
The virtual machine has been generated by modified versions of the MaX/MARVEL "Quantum Mobile" Virtual machine, whose original version can be found here: https://www.materialscloud.org/work/quantum-mobile

This version is adapted to contain only the codes relevant to this project, and to adapt to the exact code versions used in the work [1].

For maximal reproducibility, we also attach the ansible roles that allow to rebuild from scratch the Virtual Machine, using the Ansible and Vagrant softwares.

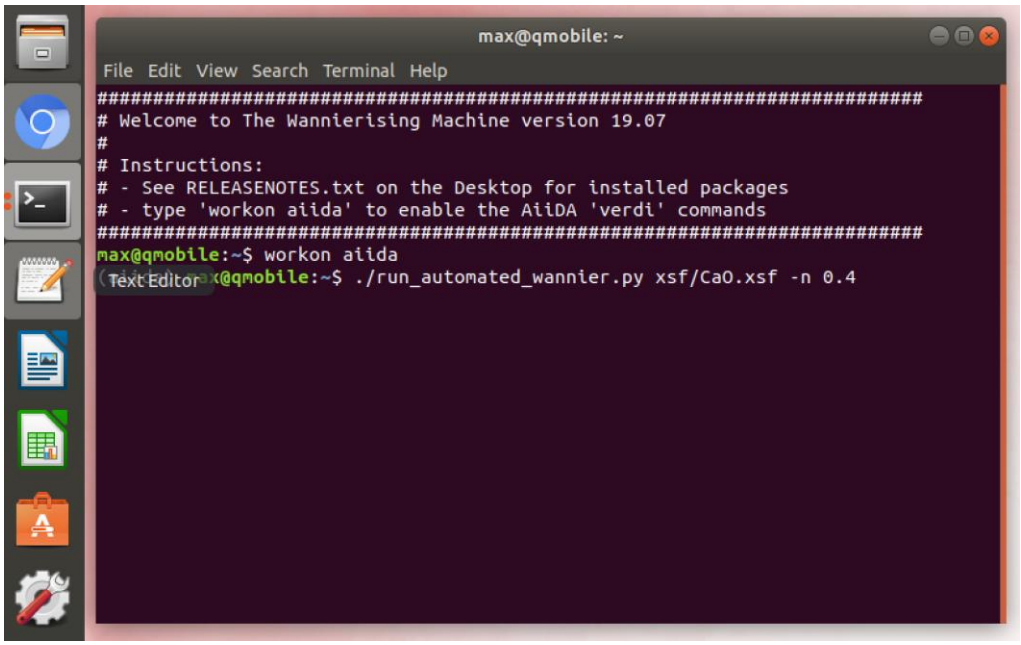
This Virtual Machine reproduces the work of the following paper [1].
Please cite this work if you use the automated workflows presented here.

[1] Valerio Vitale, Giovanni Pizzi, Antimo Marrazzo, Jonathan R. Yates, Nicola Marzari, Arash A. Mostofi, Automated high-throughput Wannierisation, preprint: arXiv:1909.00433 (2019), https://arxiv.org/abs/1909.00433

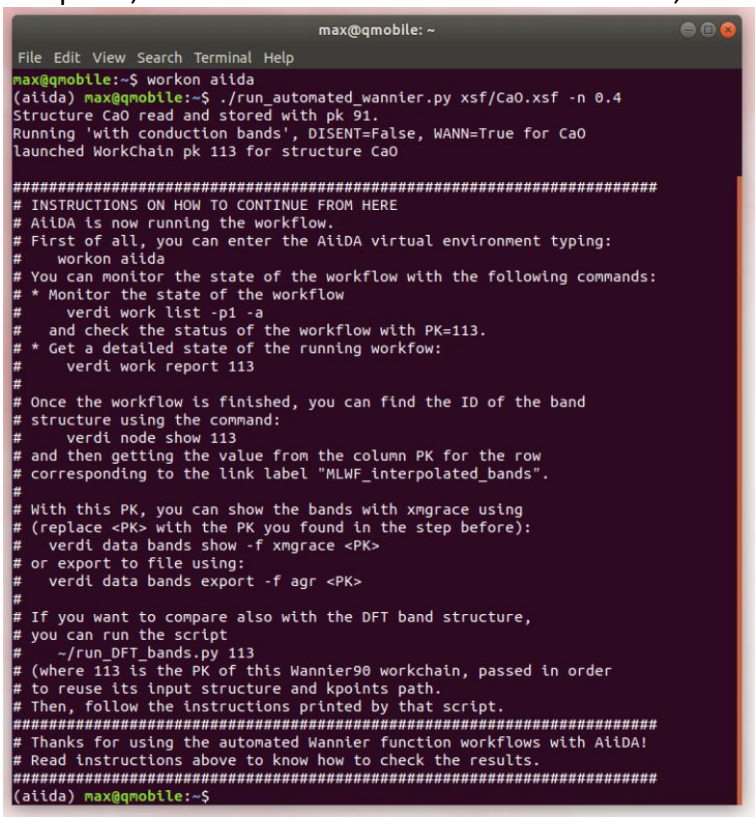
Running a bands interpolation for CaO

1. Open the terminal (icon on the left bar, represented by a black screen with a `>_` symbol inside) and type `workon aiida` to enter the virtual environment where AiiDA is installed.
   Then execute, in the home folder, the command
   `./run_automated_wannier.py xsf/CaO.xsf -n 0.4`
   to run the AiiDA workflows (using Quantum ESPRESSO and Wannier90, with the SCDM method) for CaO, using a linear density of k-points of 0.4 $Å^{-1}$. This density, coarser than the one suggested in the paper, is used to speed up a bit the simulation, as you will be running with 2 CPUs on your workstation, and not on a supercomputer!

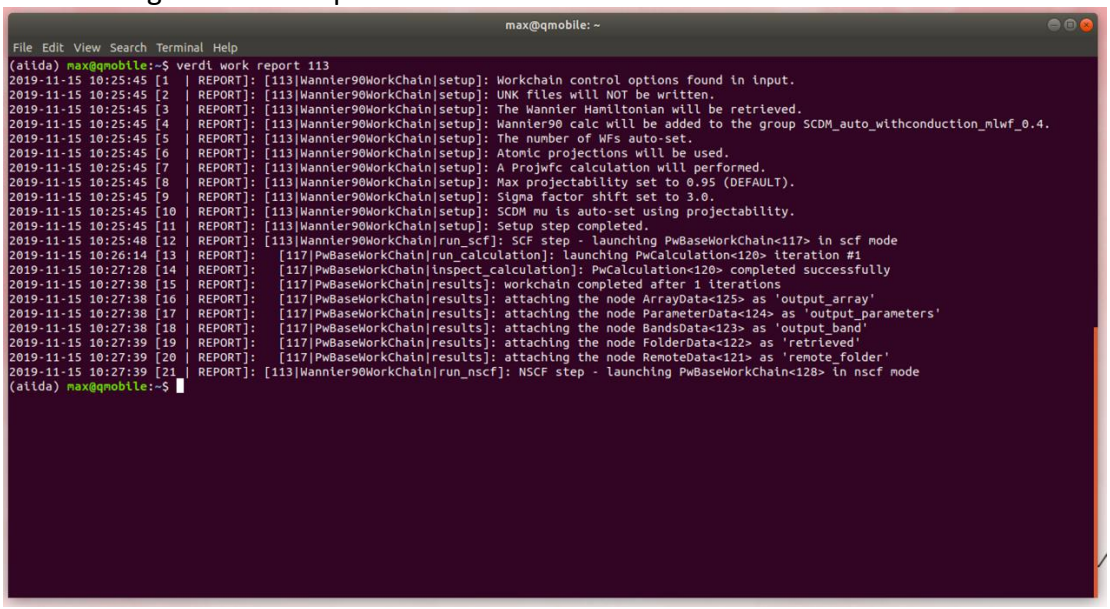
   Notes:
   - In the `xsf` subfolder you can also find all other materials used in the corresponding scientific paper.
   - You can also run `./run_automated_wannier.py -h` to inspect other possible variables to change the behavior of the script (e.g. to choose a different density of k-points).



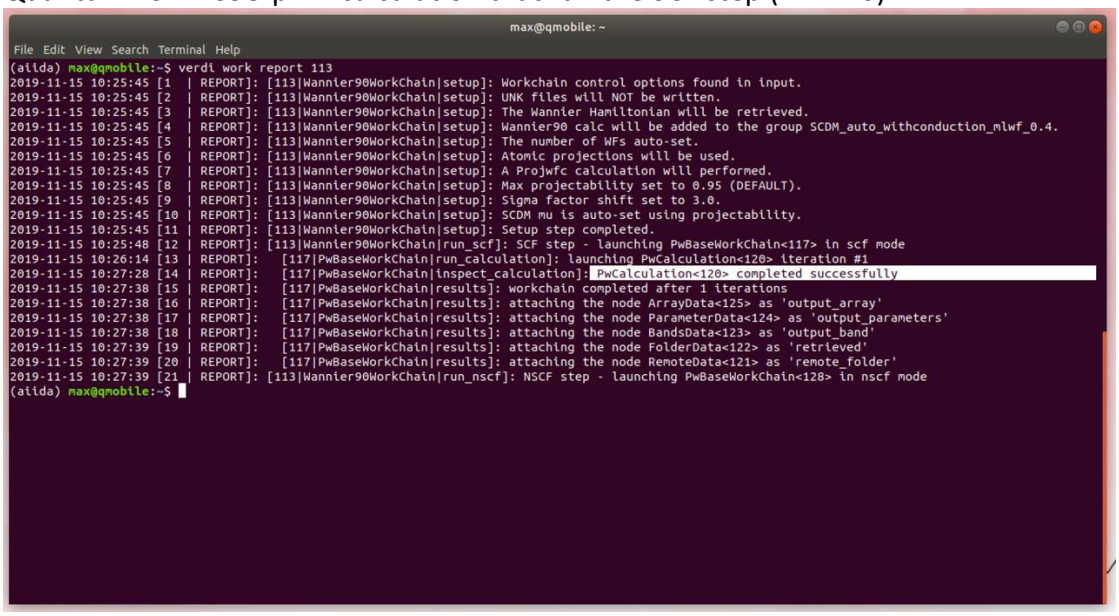
2. When you run the script, it prints a long explanation of how to continue. The important information is the ID of the workflow (called WorkChain PK, in the screenshot below it is 113). Note that, depending on the performance of your computer, the workflow will take some time to run, at least a few minutes.



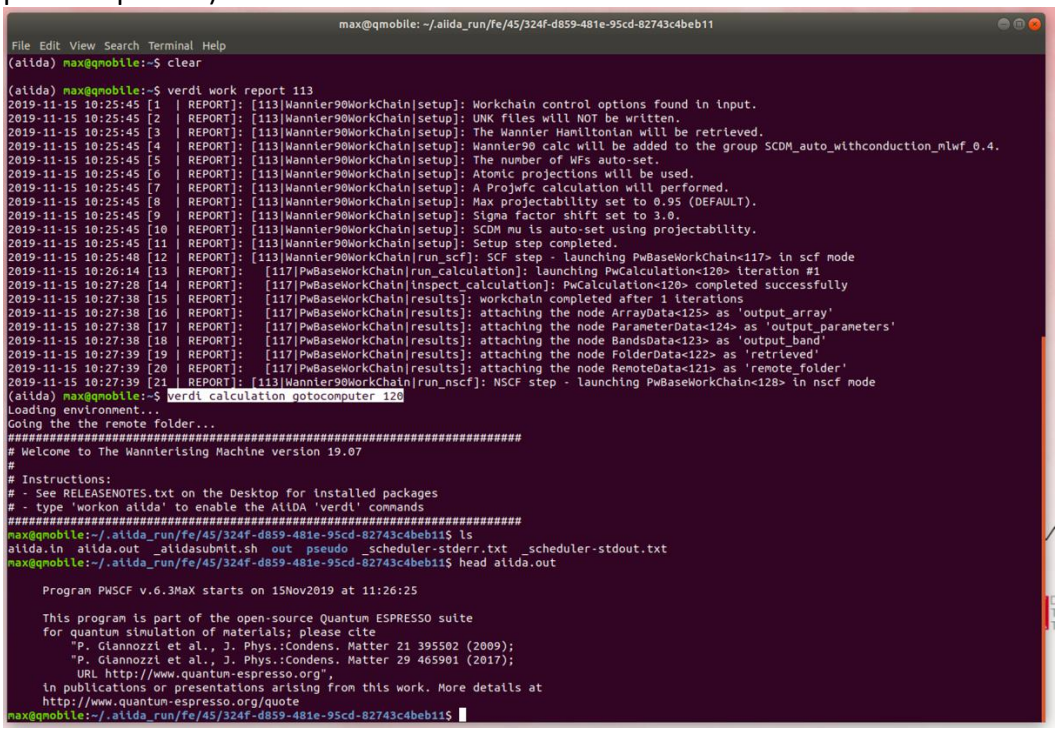
3. You can check the list and status of recently run workflows with `verdi work list -p1 -a`
4. If you want to inspect the workflow status, you can use the command `verdi work report 113`
   (replace 113 with the WorkChain PK obtained in the previous steps). The screenshot below, for instance, was taken while ther workflow had completed the SCF step and was running the NSCF step.



5. The report is also useful to go and inspect the independent calculations that were run by AiiDA. For instance, in the screenshot below, we identify the PK of the Quantum ESPRESSO pw.x calculation that run the SCF step (PK=120).

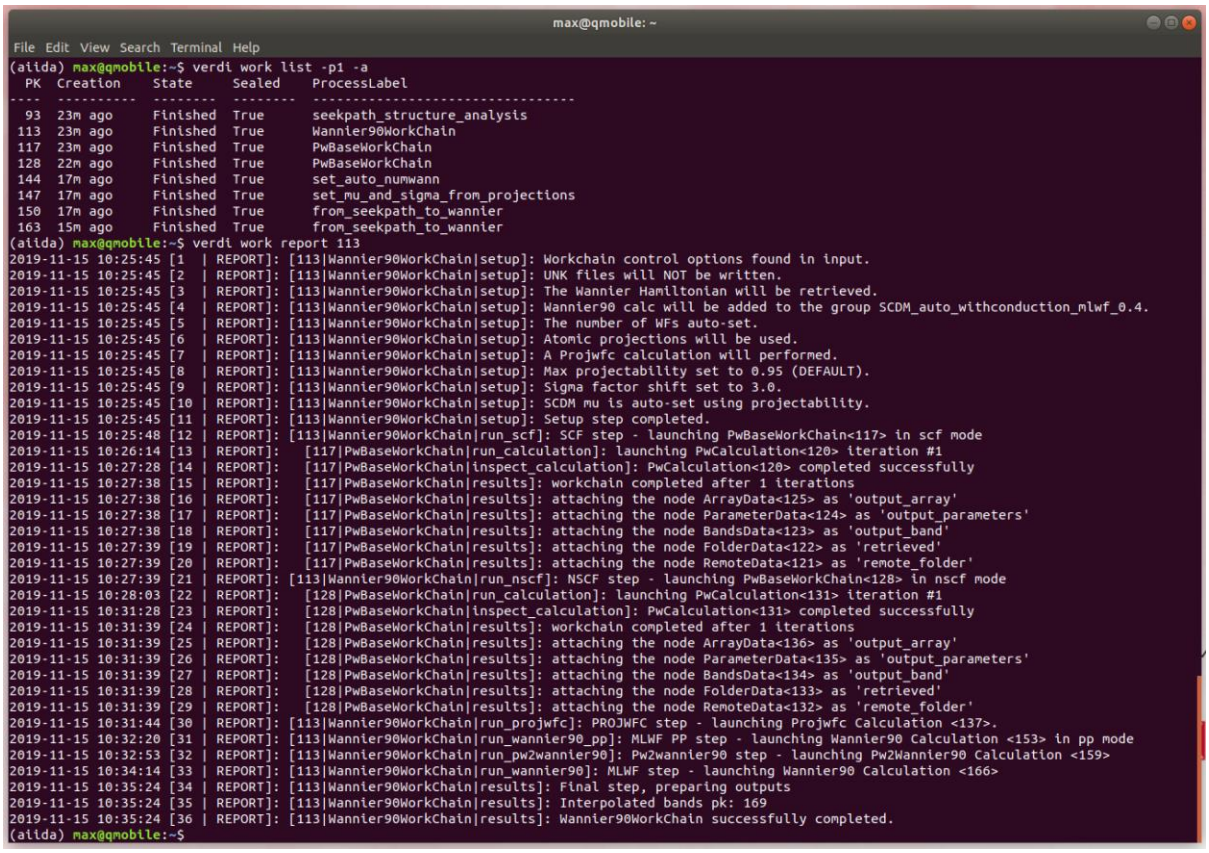


6. With this PK, we can check the results of the Quantum ESPRESSO calculation itself. AiiDA provides different ways of inspecting the results (raw and parsed). Here we show how to access the raw results, since they are most useful for users not experienced with AiiDA.
   In particular (see screenshot below), you can open a terminal directly in the folder where the simulation run with the command
   `verdi calculation gotocomputer 120`
   (replace 120 with the PK of the PwCalculation). As shown below, you will find the various files (for instance, `aiida.in` is the pw.x input file, and `aiida.out` is the pw.x output file).



7. Once you have finished inspecting the files, you can type `exit` (only once) to go back to the terminal with the AiiDA virtual environment.
8. When the workflow completes, `verdi work list -p1 -a` will list all workflows and subworkflows as `Finished`.
   Additionally, the last line of `verdi work report 113` will state that the Wannier90WorkChain successfully completed.
   In the report, you will also find the PKs of all calculations that were run. In the following screenshot, for instance, these are:
   - PwCalculation (Quantum ESPRESSO pw.x, SCF step): 120
   - PwCalculation (Quantum ESPRESSO pw.x, NSCF step): 131
   - ProjwfcCalculation (Quantum ESPRESSO projwfc.x): 137
   - Wannier90Calculation (preprocess mode): 153
   - Pw2wannier90Calculation (Quantum ESPRESSO pw2wannier90.x): 159
   - Wannier90Calculation (main run): 166

   As before, you can use the command `verdi calculation gotocomputer PK` (replacing PK with the correct PK) to go to the folder where the simulation has run and inspect the raw inputs and outputs from Quantum ESPRESSO or wannier90).
   Additionally, also the PK of the final interpolated band structure is reported: 169

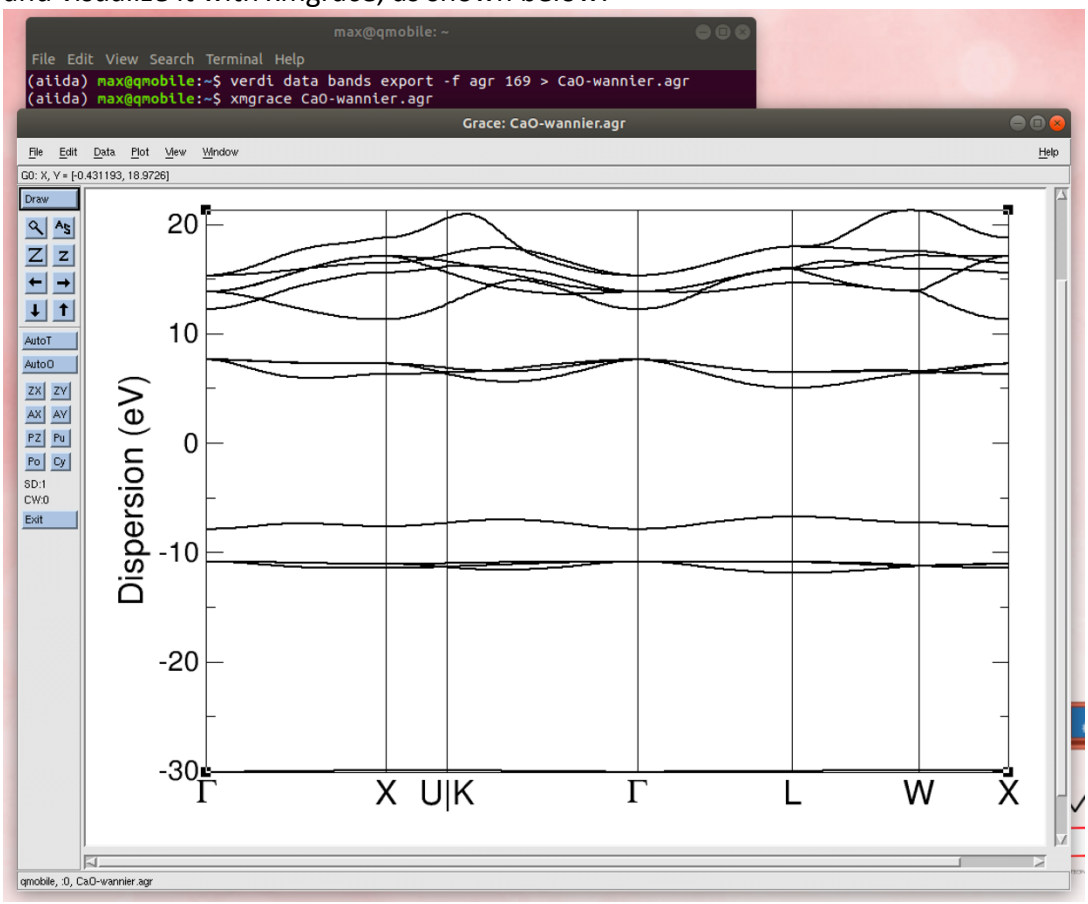


9. Equivalently, the PK of the interpolated band structure can be obtained by inspecting the output nodes of the initial workflow that we submitted (it was PK=113 in our example). This can be done by running the command `verdi node show 113` and checking the PK of the output node with link label `MLWF_interpolated_bands`. Indeed, as shown in the screenshot below, the PK

is 169, the same we obtained from the output of `verdi work report 113`.

```
max@qmobile: ~
File  Edit  View  Search  Terminal  Help
(aiida) max@qmobile:~$ verdi node show 113 | tail -n 28
##### OUTPUTS:
Link label                       PK  Type
-------------------------------  ----  ----------------------
CALL                            144  WorkCalculation
CALL                            128  WorkCalculation
CREATE                          127  ParameterData
CREATE                          114  ParameterData
mlwf_output_parameters          170  ParameterData
overlap_matrices_local_folder   161  FolderData
CREATE                          126  ParameterData
nscf_output_parameters          135  ParameterData
CALL                            117  WorkCalculation
CREATE                          115  Bool
CREATE                          116  Int
CALL                            166  Wannier90Calculation
CREATE                          146  ParameterData
CALL                            137  ProjwfcCalculation
CALL                            150  WorkCalculation
overlap_matrices_remote_folder  160  RemoteData
CALL                            159  Pw2wannier90Calculation
CALL                            153  Wannier90Calculation
scf_output_parameters           124  ParameterData
MLWF_interpolated_bands         169  BandsData
CALL                            147  WorkCalculation
CALL                            163  WorkCalculation
##### LOGS:
There are 20 log messages for this calculation
Run 'verdi work report 113' to see them
(aiida) max@qmobile:~$
```

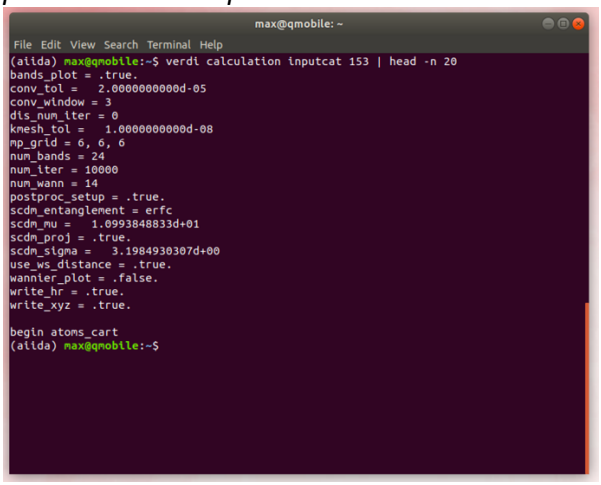
10. Finally, you can export the band structure (e.g. in xmgrace `.agr` format) with the command
`verdi data bgands export -f -agr 169 > CaO-wannier.agr`
and visualize it with xmgrace, as shown below:



Extract the SCDM parameters

We now discuss how to extract the SDCM μ and σ parameters from the automatic workflow, and how to obtain them in general also without the AiiDA workflows described in the paper.

1. To obtain the parametrs used in the run, use the command verdi calculation inputcat 166 (where 166 is the PK of the Wannier90 calculation, obtained above). Among the parameters you will find the values, in this case (see also screenshot below)
   `scdm_mu =  1.0993848833d+01`
   `scdm_sigma =  3.1984930307d+00`
   An important note: in more recent versions of Quantum ESPRESSO and Wannier90, these values are not present anymore in the Wannier90 input, but directly in the pw2wannier90 input.

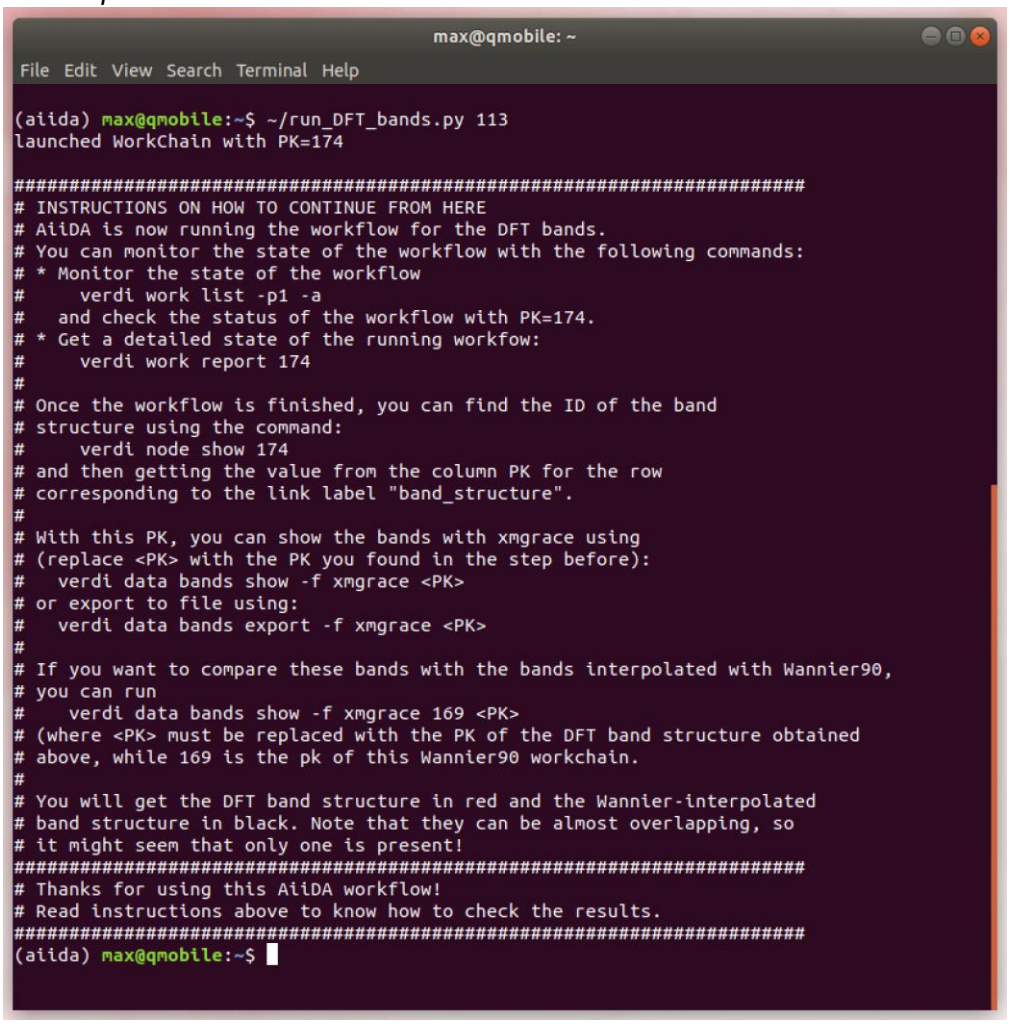


2. If instead you want to compute the values yourself, without using the automated workflow, you will need to reproduce the fit of Fig. 8 of the Vitale *et al.* paper. To this aim, we describe below the algorithm to be followed:
   a. Get the squared modulus of the projection of a given wavefunction (n, **k**) of each of the atomic orbitals defined in the pseudopotential files. This is computed by the projwfc.x code. If you want to inspect the output, you can see it `with verdi calculation gotocomputer 137` in our example.
   b. Sum these components for each given (n, **k**) for all orbitals, to obtain the projectability of the band (n, **k**).
   c. Plot these projectability values (on the y axis) vs. the energy of the corresponding band $E_{n\mathbf{k}}$, obtaining a plot similar to Fig. 8 of the paper. You might want to exclude all n values associated to excluded bands, if you are excluding bands in the Wannier90 calculation.
   d. Fit the resulting points with the following function:
      $f(x) = 0.5\ erfc(\ (x - \mu_{fit})/\sigma_{fit}\ )$
   e. Obtain the SCDM parameters as $\mu_{SCDM} = \mu_{fit} - 3\sigma_{fit}$, $\sigma_{SCDM} = \sigma_{fit}$
3. *Note*: this algorithm is implemented in the AiiDA parser (for the extraction of the data from the output of projwfc.x) and in the function `set_mu_and_sigma_from_projections` at the bottom of the workflow file that you can find in the virtual machine at the path:
   `~/codes/aiida-wannier90-theosworkflows/aiida_wannier90_theosworkflows/workflows`

Compare the interpolated bands with the DFT bands

We now discuss how to run also the bands with DFT and compare them with the interpolated bands.

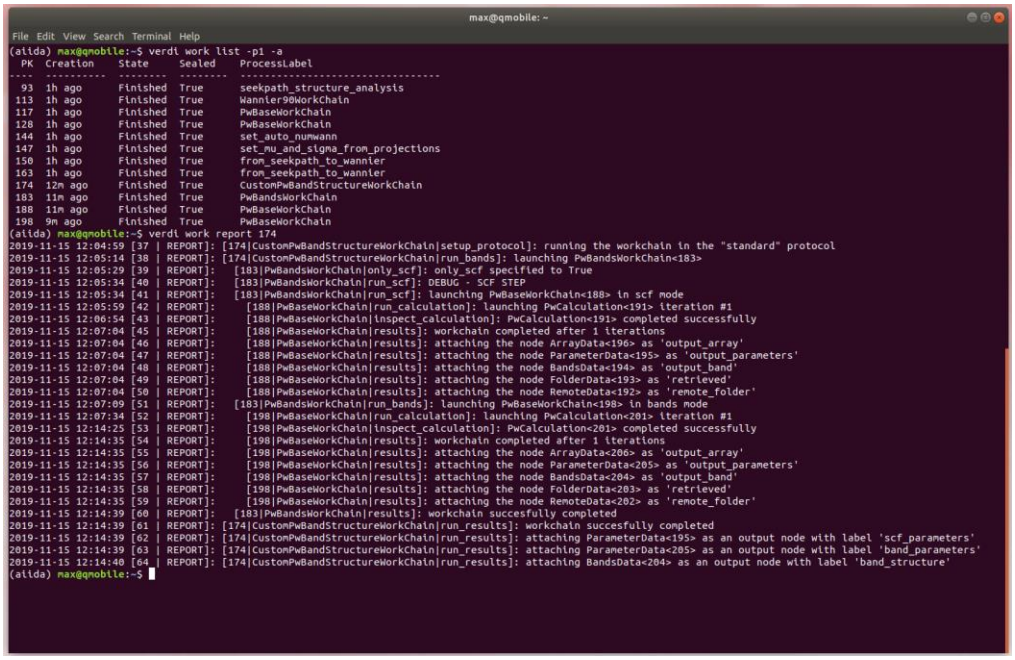
1. Run the script `~/run_DFT_bands.py 113` to start a workflow computing the bands using DFT on the same crystal structure. As usual, replace 113 with the Wannier WorkChain PK obtained at the beginning.
   Note that the Wannier90 WorkChain needs to have completed before you can run this script.



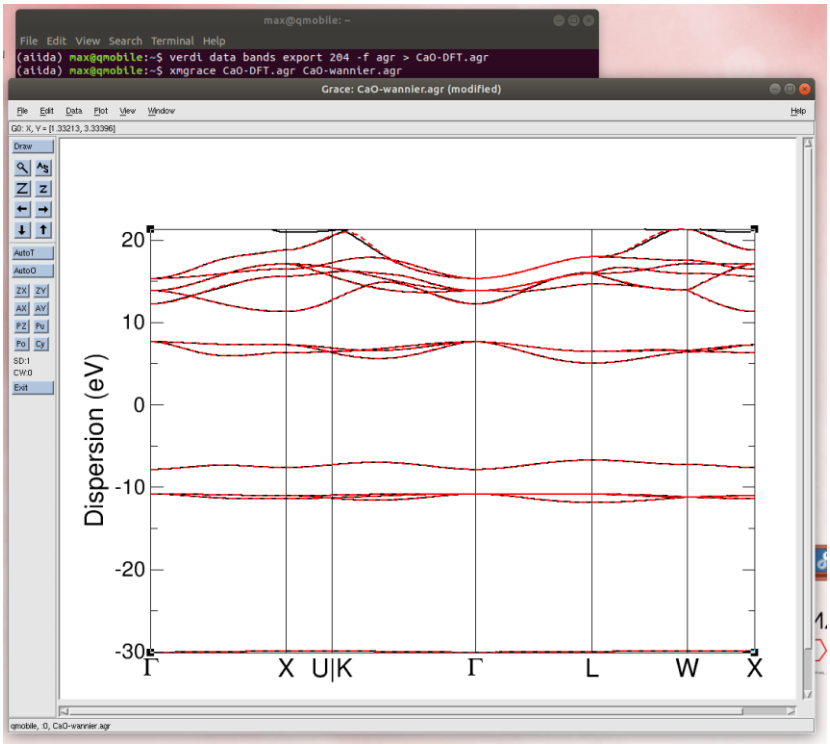
2. The output of the previous script instructs on how to plot the bands. In summary, you will have launched a new WorkChain (in this case for DFT bands, and in the example above with PK=174). You can use (as before) the commands `verdi work list -p1 -a` to check the workflow completion and `verdi work report 174` to see the report of the workflow.

   We show below the output of the two commands when the workflow has completed. In particular, in the output of the report, you can find (numbers refer to our specific example shown below):
   a. PwCalculation (Quantum ESPRESSO pw.x, SCF step): PK=191
   b. PwCalculation (Quantum ESPRESSO pw.x, NSCF step): PK=201

   In addition, we can get from the last line the PK of the output DFT bands (link label: `band_structure`). In our example: PK=204



3. We can now export the bands with
   `verdi data bands export 204 -f agr > CaO-DFT.agr`
   You can now plot both band structures overlapped using, for instance, xmgrace. (They both use the same k-points path computed by the seekpath[1] code.)
   You can run this with `xmgrace CaO-DFT.agr CaO-wannier.agr`
   Note: in the plot below we have then changed the color of the Wannier-interpolated bands to use red-dashed thicker lines, leaving the DFT bands as black (the default is to use black lines for the first file specified, i.e., the DFT bands; and random colors – one per band – for the second file, i.e., the Wannier bands).



[1] https://materialscloud.org/tools/seekpath/