



**MISSION
INNOVATION**

accelerating the clean energy revolution

POA MATERIALI AVANZATI PER L'ENERGIA

**PROGETTO IEMAP - Piattaforma Italiana Accelerata per i Materiali per
l'Energia**

Il software QMflows sarà mantenuto all'interno di GitHub in formato open source e verrà documentato e distribuito professionalmente agli utenti del codice

Ivan Infante, Liberato Manna

D1.11, Il software QMflows sarà mantenuto all'interno di GitHub in formato open source e verrà documentato e distribuito professionalmente agli utenti del codice

Ivan Infante, Liberato Manna

Febbraio 2023

Report MISSION INNOVATION

Ministero dell'Ambiente e della Sicurezza Energetica - ENEA
Mission Innovation 2021-2024 - Il annualità
Progetto: Piattaforma accelerata per i Materiali per l'Energia
Work package: IEMAP: Italian Energy Materials Acceleration Platform
Linea di attività 1.10: Espansione della piattaforma di workflow a diversi codici
Responsabile del Progetto: Massimo Celino, ENEA
Responsabile della LA: Manna, Liberato, IIT

Indice

1	Introduzione	4
2	Preparazione dell'input	4
3	Set up del calcolo.....	6
4	Interpretazione dei risultati.....	8
5	Conclusione	8
6	Riferimenti Bibliografici	8

1 Introduzione

Uno dei principali obiettivi dei pacchetti software QMflows [1] e Nano-QMFlows è quello di consentire a utenti non esperti di concentrarsi sugli aspetti scientifici dei workflow (la definizione dei sistemi da studiare, il livello di teoria da utilizzare per ogni attività e le proprietà chimico-fisiche da recuperare e analizzare) invece di risolvere i problemi tecnici legati a specifici pacchetti software di chimica quantistica o all'esecuzione parallela dei workflow. Per questo tipo di utilizzo è richiesta solo una comprensione di base del linguaggio di scripting Python, una competenza che la maggior parte dei chimici computazionali oggi possiede.

In questo contesto, è stata inserita nella documentazione di Nano-QMFlows un tutorial dedicato ad ogni workflow (con vari livelli di difficoltà) che ripercorre passo passo la preparazione dell'input, l'esecuzione del workflow e l'interpretazione dei risultati. Questo ha facilitato l'utilizzo di QMFlows e Nano-QMFlows per la preparazione e l'esecuzione di calcoli anche per utenti non esperti, permettendo loro di focalizzarsi soltanto sull'aspetto di interpretazione dei risultati ottenuti confrontandoli, ad esempio, con le evidenze sperimentali disponibili. Di seguito, proponiamo di ripercorrere uno di questi tutorials in relazione anche alle implementazioni apportate a Nano-QMFlows e descritte nel precedente deliverable D1.10. In particolare, si è scelto l'"advanced example" del tutorial "single points calculation" che permette il calcolo della struttura elettronica, cioè autovettori e autovalori (energie), per ogni punto di una traiettoria di dinamica molecolare pre-calcolata del modello di nanocristallo Cd₃₃Se₃₃.

2 Preparazione dell'input

Nella cartella di lavoro, creare un file "input_test_single_points_advanced.yml" e copiare al suo interno l'input ([Figura 1](#)) ricordandosi di rispettare l'indentazione. Copiare localmente anche la traiettoria di dinamica molecolare pre-calcolata "[Cd33Se33 fivePoints.xyz](#)".

Prestare attenzione ai seguenti parametri, comuni ai file di input di tutti i workflow:

- `workflow`: Il workflow desiderato per i calcoli, in questo caso "single_points".
- `project_name`: nome del progetto per i calcoli. Si può scegliere quello che si preferisce.
- `active_space`: Intervallo di orbitali molecolari (occupati e virtuali) da calcolare. Ad esempio, se nei calcoli si devono considerare 50 occupati e 100 virtuali, lo spazio attivo deve essere impostato su [50, 100].
- `compute_orbitals`: Specifica se gli autovettori e autovalori (energie) degli orbitali selezionati devono essere calcolati. L'impostazione predefinita è True.
- `path_hdf5`: Percorso in cui deve essere creato o può essere trovato il file .hdf5: hdf5 è il formato utilizzato per memorizzare gli orbitali molecolari e altre informazioni.
- `path_traj_xyz`: Percorso al file contenente le coordinate del sistema, in questo caso la traiettoria di dinamica molecolare pre-calcolata, fornito in formato xyz.

- `path_hdf5`: Per eseguire i calcoli è necessario un percorso scratch. Per i sistemi di grandi dimensioni, i file `.hdf5` possono diventare molto grandi (centinaia di GB) e i calcoli vengono invece eseguiti nell'area di lavoro scratch. Anche i risultati finali saranno memorizzati qui.

```

workflow:
  single_points

project_name: Cd33Se33
active_space: [50, 50]
dt: 1
path_hdf5: "Cd33Se33.hdf5"
path_traj_xyz: "Cd33Se33_fivePoints.xyz"
scratch_path: "/tmp/singlepoints_advanced"
calculate_guesses: "first"

cp2k_general_settings:
  basis: "DZVP-MOLOPT-SR-GTH"
  potential: "GTH-PBE"
  cell_parameters: 28.0
  periodic: none
  executable: cp2k.popt

cp2k_settings_main:
  specific:
    template: pbe_main
    cp2k:
      force_eval:
        dft:
          scf:
            eps_scf: 1e-6

cp2k_settings_guess:
  specific:
    template:
      pbe_guess

```

Figura 1: Estratto della documentazione di Nano-QMFlows, in particolare del "advanced example" del tutorial "single points calculation", che illustra la struttura di un file di input per il calcolo della struttura elettronica, cioè autovettori e autovalori (energie), per ogni punto di una traiettoria di dinamica molecolare pre-calcolata del modello di nanocrystallo Cd33Se33.

In `cp2k_general_settings` è possibile personalizzare le impostazioni utilizzate per generare l'input di CP2k [2], il software di chimica quantistica impiegato. Qui è possibile specificare il livello di teoria che si desidera utilizzare nei calcoli (set di base e funzionale di scambio-correlazione) e le caratteristiche principali del sistema (parametri e angoli di cella, periodicità, carica, molteplicità, ...). Nelle sottosezioni `cp2k_settings_guess` e `cp2k_settings_main` è possibile fornire informazioni più dettagliate sulle impostazioni di input di CP2k da utilizzare rispettivamente per calcolare la funzione d'onda guess e per eseguire il calcolo principale. In questo esempio, abbiamo utilizzato uno dei template disponibili, specificamente destinato a calcoli con un funzionale di scambio-correlazione PBE. In questo esempio, mostriamo anche come personalizzare ulteriormente le impostazioni di

cp2k_general_settings. In particolare, viene aggiunta una sottosezione per sovrascrivere alcuni parametri del template pbe e per restringere il criterio di convergenza eps_scf a 1e-6 (il valore predefinito nel template pbe_main è 5e-4). Si noti che viene utilizzata un'indentazione specifica per riprodurre la struttura di un tipico file di input cp2k.

NOTE: Questo tutorial mostra come sia possibile sfruttare le implementazioni di Nano-QMFlows descritte nel deliverable D1.10. Ad esempio, l'aggiunta di supporto per più tipi di set di base e funzionali di scambio-correlazione è accessibile rispettivamente tramite le keyword basis e potential nella sezione cp2k_general_settings. È inoltre possibile effettuare calcoli di tipo "unrestricted" tramite la keyword multiplicity. Questo tutorial è stato scritto prima dell'ultima implementazione di Nano-QMFlows, nella quale il valore predefinito eps_scf nel template pbe_main è stato modificato a 1e-6.

Nel file di input ([Figura 1](#)), prestare particolare attenzione ai seguenti parametri che sono stati aggiunti rispetto alle impostazioni comuni di tutti i workflow:

- dt: Il valore del passo temporale utilizzato nella simulazione di dinamica molecolare (in fs).
- calculate_guesses: Specifica se calcolare la funzione d'onda guess solo nel primo punto della traiettoria ("first") o in tutti ("all"). In questo caso si mantiene il valore predefinito, "first".

NOTE: È adesso possibile sfruttare il workflow single point per il calcolo della struttura elettronica e proprietà correlate (come il valore di bandgap) di una serie di sistemi differenti grazie alla relativa implementazione di Nano-QMFlows descritta nel deliverable D1.10. In questo caso, sarà sufficiente: (i) inserire in path_traj_xyz il percorso ad un file nel formato xyz contenente, una dopo l'altra, le coordinate dei sistemi di interesse e (ii) impostare la keyword calculate_guesses su "all". In questo modo verrà calcolata una funzione d'onda guess (differente) per ognuno dei sistemi.

3 Set up del calcolo

Quando si utilizza questo tutorial, assicurarsi di avere installato in un environment di conda l'ultima versione di QMFlows e Nano-QMFlows. I calcoli di chimica quantistica vengono eseguiti con il pacchetto CP2k, che deve quindi anche essere installato.

Il calcolo può essere eseguito semplicemente tramite l'attivazione del environment di conda e l'uso del comando run_workflow.py con:

```
conda activate qmflows
```

```
run_workflow.py -i input_test_single_points_advanced.yml
```

```
In [1]: import h5py

In [2]: f = h5py.File('Cd33Se33.hdf5', 'r')

In [3]: f.keys()
Out[3]: <KeysViewHDF5 ['coefficients', 'cp2k', 'eigenvalues', 'energy']>

In [9]: c_ao = f['coefficients/point_0'][(0)]

In [10]: c_ao.shape
Out[10]: (1254, 100)

In [11]: e_ao = f['eigenvalues/point_0'][(0)]

In [12]: e_ao.shape
Out[12]: (100,)
```

Figura 2: Estratto di un notebook Jupyter che illustra come ritrovare, tramite semplici comandi nel linguaggio di scripting Python, le informazioni conservate nel file .hdf5 relative alla struttura elettronica, cioè auto-vettori e autovalori (energie), per il primo punto della traiettoria di dinamica molecolare del nostro modello di nanocristallo Cd33Se33.

```
In [18]: e_gap = (e_ao[50]-e_ao[49])*27.211

In [19]: e_gap
Out[19]: 1.1416835714280604

In [23]: import matplotlib.pyplot as plt

In [32]: plt.barh(e_ao*27.211, width=1, height=0.02)
Out[32]: <BarContainer object of 100 artists>
```

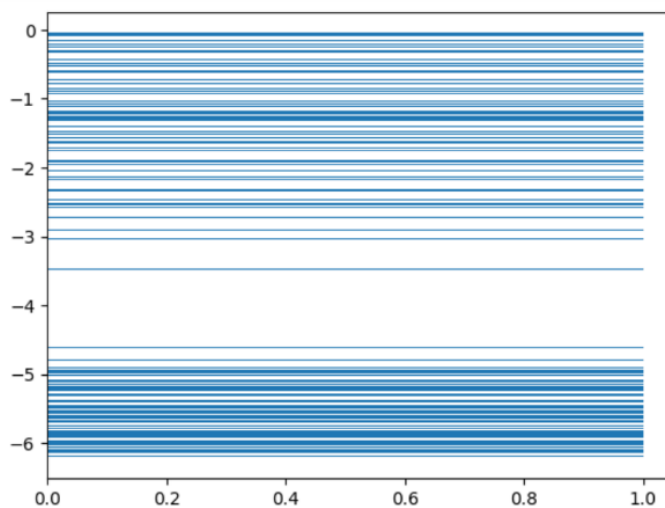


Figura 3: Estratto di un notebook Jupyter che illustra come ottenere, tramite semplici comandi nel linguaggio di scripting Python, il valore di band-gap e la densità degli stati per il primo punto della traiettoria di dinamica molecolare del nostro modello di nanocristallo Cd33Se33 grazie al file .hdf5.

4 Interpretazione dei risultati

Una volta completato il calcolo con successo, si troverà un file Cd33Se33.hdf5 nella directory di scratch indicata in input. Questo contiene tutte le informazioni conservate da (Nano-)QMFlows relative alla struttura elettronica, cioè autovettori e autovalori (energie), per ogni punto della traiettoria di dinamica molecolare del nostro modello di nanocristallo Cd33Se33. Tali informazioni sono facilmente accessibili tramite comandi base nel linguaggio di scripting Python, eseguiti ad esempio in un notebook Jupyter (**Figura 2**).

Dal file Cd33Se33.hdf5 si possono inoltre ottenere altre proprietà rilevanti come il valore di band-gap e la densità degli stati di ogni punto della nostra traiettoria (**Figura 3**).

5 Conclusione

I software descritti rappresentano in generale un'innovazione rispetto allo stato dell'arte, in quanto programmi di questo tipo, per l'automatizzazione e la gestione dei workflow di simulazioni atomistiche su nanocristalli semiconduttori, non esistono nella comunità scientifica. In questa attività, sono stati preparati i pacchetti software QMFlows e Nano-QMFlows al loro utilizzo anche da parte di utenti anche non esperti tramite l'inserimento di molteplici files di test e esempi e l'ampiamiento della documentazione, che comprende adesso, per ogni tipo di workflow, una dedicata sezione di "tutorial". Quest'ultimo punto si è rivelato essenziale per un efficace diffusione di questi strumenti alla comunità scientifica, in eventi quali conferenze e workshop computazionali.

6 Riferimenti Bibliografici

- [1] Zapata, F., Ridder, L., Hidding, J., Jacob, C. R., Infante, I., & Visscher, L. (2019). QMflows: a tool kit for interoperable parallel workflows in quantum chemistry. *Journal of chemical information and modeling*, 59(7), 3191-3197.
- [2] Hutter, J. *et al.* (2020) CP2K: An Electronic Structure and Molecular Dynamics Software Package - Quickstep: Efficient and Accurate Electronic Structure Calculations. *J. Chem. Phys.*, 19.