



**MISSION
INNOVATION**

accelerating the clean energy revolution

POA MATERIALI AVANZATI PER L'ENERGIA

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

Implementazione della piattaforma QMflows per l'automatizzazione di calcoli di chimica computazionale

Ivan Infante, Liberato Manna

D1.10, Implementazione della piattaforma QMflows per l'automatizzazione di calcoli di chimica computazionale

Ivan Infante, Liberato Manna

Febbraio 2023

Report MISSION INNOVATION

Ministero dell'Ambiente e della Sicurezza Energetica - ENEA

Mission Innovation 2021-2024 - II 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: Liberato Manna, IIT

Indice

1	Introduzione	4
2	Implementazioni di QMFlows	4
2.1	Aggiunta di un template per calcoli di analisi vibrazionali: <code>qmflows.freq</code>	4
2.2	Aggiunta di un template per calcoli di ottimizzazione di cella: <code>qmflows.cell_opt</code>	5
2.3	Aggiunta di supporto per calcoli di tipo “unrestricted”	6
2.4	Aggiunta di supporto per più tipi di set di base e funzionali di scambio-correlazione	6
2.5	Aggiunta di diverse proprietà agli output di CP2k	6
2.6	Altre implementazioni	7
3	Implementazioni di Nano-QMFlows	7
3.1	Aggiunta di un supporto per il calcolo delle proprietà elettroniche di una serie di sistemi differenti 8	
3.2	Aggiunta di supporto per calcoli di tipo “unrestricted”	8
3.3	Aggiunta di supporto per più tipi di set di base e funzionali di scambio-correlazione	8
3.4	Altre implementazioni	8
4	Conclusione	8
5	Abbreviazioni	9
6	Riferimenti bibliografici	9

1 Introduzione

La moderna chimica quantistica computazionale richiede un insieme di strumenti computazionali per eseguire calcoli sempre più complessi, che si basa sull'intercomunicazione tra i suddetti strumenti. Tale comunicazione avviene di solito tramite i cosiddetti “shell scripts” che hanno l'obiettivo di automatizzare le azioni di input/output come: il lancio dei calcoli in un cluster di calcolo, la lettura dell'output risultante e l'invio del relativo risultato numerico a un altro programma. Tali script sono difficili da mantenere ed estendere e richiedono una notevole esperienza di programmazione per poter essere utilizzati. È quindi auspicabile un insieme di strumenti automatici ed estensibili che permettano di eseguire simulazioni complesse su piattaforme hardware eterogenee, mantenendo la massima flessibilità e limitando la manutenzione.

In questo contesto è emerso il pacchetto Python QMflows [1] per l'automazione di workflows della chimica computazionale. QMflows è costruito sulla libreria PLAMS [2], che è stata progettata per fornire un'interfaccia interfaccia generale ai pacchetti di simulazione quantomeccanica e consiste in una serie di moduli per automatizzare i seguenti compiti: (i) generazione di input, (ii) gestione delle dipendenze dei compiti con Noodles [3], (iii) capacità avanzate di manipolazione molecolare con con rdkit [4], (iv) rilevamento e recupero dei calcoli falliti, (v) memorizzazione dei dati numerici con h5py [5]. Ciò consente ai chimici computazionali di utilizzare gli emergenti ambienti di calcolo parallelo di massa in modo semplice e di concentrarsi sull'interpretazione dei dati scientifici piuttosto che sulle tedious procedure di invio dei calcoli e sull'elaborazione manuale dei dati.

Per quanto riguarda, più nello specifico, la ricerca sui nanomateriali, è invece emerso il pacchetto Python Nano-QMFlows che offre una serie di workflow per il calcolo di diverse proprietà elettroniche critiche per il loro effettivo impiego nei settori emergenti della tecnologia. I workflow disponibili permettono calcoli di tipo: (i) “single points”, ovvero di struttura elettronica per una data configurazione strutturale, dal quale è possibile ricavare, ad esempio, il valore del bandgap, (ii) “Crystal Orbital Overlap Population”, (iii) “Inverse Participation Ratio”, (iv) spettri di assorbimento (energie degli stati eccitati, momenti di dipolo di transizione e forza dell'oscillatore) e (v) vettori di accoppiamento non adiabatico, con vari software di chimica quantistica.

Nell'ambito di questa attività si è lavorato per espandere le funzionalità e migliorare la generalizzabilità e l'estensibilità dei pacchetti software di QMFlows e Nano-QMFlows in base alle esigenze dei progetti correlati che coinvolgono simulazioni atomistiche (tipicamente di chimica quantistica ma anche force fields) di sistemi quali materiali allo stato solido e nanocristalli semiconduttori. Di seguito, proponiamo una panoramica completa delle implementazioni effettuate, riassunta nei documenti CHANGELOG.md inclusi nelle repositories sopra citate.

2 Implementazioni di QMFlows

2.1 *Aggiunta di un template per calcoli di analisi vibrazionali: qmflows.freq*

Un importante implementazione di QMFlows consiste nell'introduzione di un template per l'analisi termochimica dei modi normali di vibrazione con vari livelli di teoria e software di chimica

quantistica: calcoli DFT e DFTB [6] in ADF [7] (**Figura 1**), calcoli di chimica quantistica e force-field in CP2k [8], calcoli in ORCA [9].



```

235 #: Templates for frequency analyses calculations.
236 freq = Settings(yaml.load("""
237 specific:
238   adf:
239     analyticalfreq:
240     basis:
241       core: None
242       type: DZP
243
244     numericalquality: good
245
246   ams:
247     ams:
248       Task: SinglePoint
249       Properties:
250         NormalModes: Yes
251
252   dftb:
253     dftb:
254       resourcesdir: "DFTB.org/3ob-3-1"

```

Figura 1: Estratto del template per l’analisi termochimica dei modi normali di vibrazione a livello DFT (“adf”) e DFTB (“dftb”) [6] in ADF [7] contenuto nel repository di QMFlows.

2.2 Aggiunta di un template per calcoli di ottimizzazione di cella: *qmflows.cell_opt*

Un'altra fondamentale implementazione di QMFlows riguarda l’aggiunta di un template per effettuare calcoli force-field di ottimizzazione di cella con CP2k [8] (**Figura 2**), in cui vengono rilassati sia le posizioni atomiche che i parametri di cella del sistema.

```

master  qmflows / src / qmflows / templates / _templates.py
Code    Blame  431 lines (391 loc) · 10.8 KB
378     cell_opt = Settings(yaml.load("""
412         subsys:
413             cell:
414                 periodic: xyz
415                 topology:
416                     coord_file_format: 'OFF'
417                     center_coordinates:
418                         center_point: 0.0 0.0 0.0
419         motion:
420             cell_opt:
421                 type: direct_cell_opt
422                 optimizer: lbfgs
423                 max_iter: 500
424         print:
425             cell_low:
426                 filename: ''
427         global:
428             print_level: low
429             project: cp2k
430             run_type: cell_opt
431     """, Loader=UniqueSafeLoader))

```

Figura 2: Estratto del template per calcoli force-field di ottimizzazione di cella con CP2k [8] contenuto nel reposirio di QMFlows.

2.3 Aggiunta di supporto per calcoli di tipo “unrestricted”

Sono inoltre state implementate la gestione e validazione delle proprietà e dei workflow che coinvolgono calcoli di tipo “unrestricted”, estendendo così gli strumenti disponibili in QMFlows anche a sistemi contenenti ioni “open-shell”.

2.4 Aggiunta di supporto per più tipi di set di base e funzionali di scambio-correlazione

Sono anche state effettuate una serie di modifiche che garantiscono maggiore flessibilità nella scelta dei set di base e funzionali di scambio-correlazione più appropriati al sistema in esame, tra le quali:

- Aggiunta delle keywords `basis` e `potential` nella sezione `cp2k_general_settings`.
- Aggiunta di un supporto per l'analisi di set di base di CP2k composti da molteplici esponenti.
- Aggiunti nuovi alias per i potenziali di CP2k.

2.5 Aggiunta di diverse proprietà agli output di CP2k

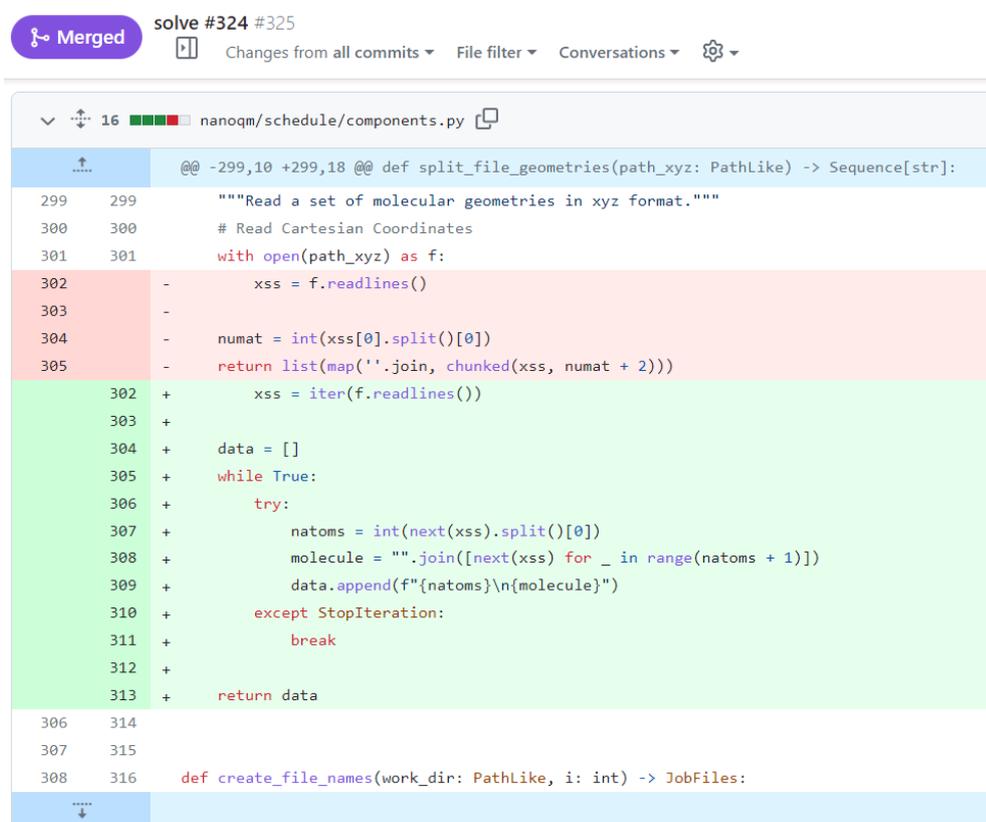
L’ultima importante implementazione di QMFlows ha contribuito all’estensione dei risultati memorizzati dopo calcoli di chimica quantistica e force-field in CP2k [8] con 6 nuove proprietà generiche:

- volume
- forze
- coordinate
- temperatura
- reticolo
- pressione

2.6 Altre implementazioni

Diverse altre modifiche sono state effettuate per aggiornare QMFlows all'evoluzione dei software di chimica quantistica (ad esempio l'aggiunta di supporto per la lettura degli orbitali molecolari ottenuti con versioni di CP2K ≥ 8.2), dei cluster di calcolo (ad esempio la modifica dell'estensione predefinita dell'eseguibile CP2K da `.popt` a `.ssmp`) e di Python (tra le altre, fine del supporto per Python 3.6 e 3.7 e inserimento di quello per Python 3.11). Sono state apportate anche altre modifiche minori e vari fix, elencati nel documento `CHANGELOG.md`.

3 Implementazioni di Nano-QMFlows



```

@@ -299,10 +299,18 @@ def split_file_geometries(path_xyz: PathLike) -> Sequence[str]:
299 299     """Read a set of molecular geometries in xyz format."""
300 300     # Read Cartesian Coordinates
301 301     with open(path_xyz) as f:
302 -         xss = f.readlines()
303 -
304 -         numat = int(xss[0].split()[0])
305 -         return list(map(''.join, chunked(xss, numat + 2)))
+
+     xss = iter(f.readlines())
+
+     data = []
+     while True:
+         try:
+             natoms = int(next(xss).split()[0])
+             molecule = ''.join([next(xss) for _ in range(natoms + 1)])
+             data.append(f"{natoms}\n{molecule}")
+         except StopIteration:
+             break
+     return data
306 314
307 315
308 316     def create_file_names(work_dir: PathLike, i: int) -> JobFiles:

```

Figura 3: Modifiche apportate al repository di Nano-QMFlows per la gestione di una serie di sistemi differenti, le cui configurazioni strutturali sono contenute, una dopo l'altra, in un file di coordinate nel formato xyz.

3.1 Aggiunta di un supporto per il calcolo delle proprietà elettroniche di una serie di sistemi differenti

La più importante implementazione di Nano-QMFlows consiste nell'aggiunta di un supporto per il riconoscimento, la gestione e il calcolo delle proprietà elettroniche di interesse (tramite il dedicato workflow) per una serie di sistemi differenti, le cui configurazioni strutturali sono contenute, una dopo l'altra in un file di coordinate nel formato xyz ([Figura 3](#)). Questo costituisce un notevole passo avanti per l'utilizzo di QMFlows e Nano-QMFlows nell'ambito del "material discovery", dove una serie di sistemi "candidato" semplicemente elencati in un file xyz possono ora essere filtrati in base ad alcune loro specifiche proprietà elettroniche, calcolate in modo efficiente ed automatizzato tramite tali pacchetti software.

3.2 Aggiunta di supporto per calcoli di tipo "unrestricted"

È inoltre stato ridefinito il concetto di "active space" per il supporto di calcoli di tipo "unrestricted" nei workflows di Nano-QMFlows, permettendo di sfruttare le implementazioni apportate in QMFlows per tali calcoli.

3.3 Aggiunta di supporto per più tipi di set di base e funzionali di scambio-correlazione

In modo simile a quanto attuato in QMFlows, sono inoltre state effettuate una serie di modifiche che garantiscono maggiore flessibilità nella scelta dei set di base e funzionali di scambio-correlazione più appropriati al sistema in esame, ad esempio:

- Aggiunta delle opzioni `basis_set_file_name` e `potential_file_name`.
- Eliminazione dell'"hard code" per i nomi dei set di base per il calcolo dell'indice HOMO.
- Aggiunta sistematica del suffisso `-q` quando si leggono i set di base da `.hdf5`.
- Aggiunta di un supporto per l'analisi di set di base di CP2k composti da molteplici esponenti.
- Codifica dei coefficienti `BASIS_MOLOPT` nel file `.hdf5`.
- Aggiunta di impostazioni per la scelta manuale dei funzionali di scambio correlazione GGA, MGGA e ibridi.

3.4 Altre implementazioni

Infine, come per QMFlows, sono state apportate altre modifiche per aggiornare Nano-QMFlows all'evoluzione dei software di chimica quantistica, dei cluster di calcolo e di Python. Sono state implementate anche altre modifiche minori (come la variazione del criterio di convergenza predefinito `eps_scf` da $5e-4$ a $1e-6$ nei templates) e vari fix, elencati nel documento `CHANGELOG.md`.

4 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 LA,

sono state estese le funzionalità dei pacchetti software QMFlows e Nano-QMFlows per assistere la ricerca di materiali semiconduttori (sia allo stato solido che sotto forma di nanocristalli semiconduttori) sempre più performanti, permettendo così di includere materiali con composizioni ioniche sempre più complesse (ad esempio sistemi “open-shell”) e con livelli di teorica sempre più elevati (ad esempio funzionali di scambio-correlazione ibridi) e anche di estrarre un numero crescente proprietà dalle simulazioni atomistiche effettuate. È importante notare che tutte le implementazioni elencate in questo documento hanno preparato QMFlows al prossimo release della sua versione 1.0.0.

5 Abbreviazioni

ADF = Amsterdam Density Functional

DFT = density functional theory

DFTB = Density Functional based Tight Binding

PLAMS = Python Library for Automating Molecular Simulation

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] <https://github.com/SCM-NV/PLAMS>
- [3] <https://nlesc.github.io/noodles/>
- [4] <http://www.rdkit.org/>
- [5] <http://www.h5py.org/>
- [6] Grimme, S., Bannwarth, C. & Shushkov, P. A (2017) Robust and Accurate Tight-Binding Quantum Chemical Method, *J. Chem. Theory Comput.* 13, 1989–2009.
- [7] te Velde, G. *et al.* (2001) Chemistry with ADF. *J. Comput. Chem.* 22, 931–967.
- [8] 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.
- [9] Neese, F. *et al.* (2012) The ORCA program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2, 73– 78.