



**MISSION
INNOVATION**

accelerating the clean energy revolution

POA MATERIALI AVANZATI PER L'ENERGIA

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

Workflow automatico di calcoli ab initio e
collezione dei dati di materiali catodici puri
e drogati

Francesco Buonocore, Claudio Ronchetti, Simone Giusepponi, Massimo Celino



D1.23, WORKFLOW AUTOMATICO DI CALCOLI AB INITIO E COLLEZIONE DEI DATI DI MATERIALI CATODICI PURI E DROGATI

Francesco Buonocore (ENEA), Claudio Ronchetti (ENEA), Simone Giusepponi (ENEA), Massimo Celino (ENEA)

Maggio 2023

Report MISSION INNOVATION

Ministero della Transizione Ecologica - 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: LA 1.20 Predisposizione di workflow per automatizzazione dei calcoli per accumulatori

Responsabile del Progetto: Massimo Celino, ENEA

Responsabile della LA: Francesco Buonocore, ENEA

Indice

1	SOMMARIO.....	4
2	INTRODUZIONE.....	5
3	DESCRIZIONE DEI METODI USATI	7
3.1	METODO DFT.....	7
3.2	WORKCHAIN	8
4	ESECUZIONE DEL WORKFLOW.....	11
4.1	INSTALLAZIONE	11
4.2	CONFIGURAZIONE ENVIRONMENT	12
4.2.1	<i>Conda environment</i>	12
4.2.2	<i>Singularity container</i>	12
4.3	ESECUZIONE.....	13
4.4	PARAMETRI DI ESECUZIONE.....	14
5	COLLEZIONE DEI DATI SUL DATABASE	15
5.1	STATISTICHE.....	17
6	CONCLUSIONI	22
7	RIFERIMENTI BIBLIOGRAFICI	22
8	ABBREVIAZIONI ED ACRONIMI	24

1 Sommario

È stato predisposto ed eseguito un workflow automatico di calcoli ab initio per materiali catodici basati su ossidi metallici lamellari del tipo MeTMO₂ puri e drogati i cui modelli periodici sono generati dal codice predisposto nella LA1.19 (Deliverable D1.22) e per i quali il protocollo di calcoli ab initio è stato definito nella stessa LA. Lo scopo di questo workflow è quello di calcolare le proprietà energetiche di strutture cristalline al variare del livello di doping e delle specie dopanti, in modo da generare un database per la ricerca dei materiali catodici più stabili. È stata usata la potenza computazionale del cluster CRESCO di ENEA per simulare accuratamente ed esplorare numerose composizioni e configurazioni cristalline tramite l'uso di questo workflow. I risultati dei calcoli sono raccolti in automatico e registrati sul database gestito da ENEA la cui realizzazione è in LA1.5.

2 Introduzione

Le simulazioni numeriche mirate allo sviluppo di materiali innovativi, grazie anche all'aumento delle prestazioni delle infrastrutture computazionali, possono essere utilizzate per realizzare ricerche e screening di materiali su larga scala. Ciò consente di identificare i candidati adatti e i parametri fisici desiderati, che meritano di essere ulteriormente esaminati mediante esperimenti. Uno svantaggio nell'uso intensivo di codici per il calcolo ad alte prestazioni riguarda la gestione di un elevato numero di simulazioni su infrastrutture HPC e come gestire enormi quantità di dati. Negli ultimi anni, sono stati sviluppati vari tool al fine di superare questi problemi con diverse applicazioni nel campo della fisica e chimica computazionale.

Sono di grande interesse gli studi sulle nuove batterie ricaricabili con elevata densità energetica e basso costo, considerato che la loro commercializzazione affronta ancora molteplici sfide riguardanti i materiali delle batterie e le interfacce. Alcune delle difficoltà incontrate dai materiali delle batterie sono che un singolo materiale spesso deve possedere diverse funzioni e deve essere anche economico, facile da preparare, sicuro ed ecologicamente sostenibile. Gli sviluppi recenti nei gestori di workflow, insieme all'aumento continuo della potenza di calcolo, hanno reso possibile i workflow computazionali automatizzati. Utilizzando questo metodo, il workflow di ricerca predefinito è seguito automaticamente per studiare decine di migliaia di materiali e selezionare quelli che soddisfano i molteplici requisiti.

Ad esempio, PyAdf è un framework di scripting per la chimica quantistica che può essere utilizzato per automatizzare flussi di lavoro quantistico-chimici [1]; ChemShell è un pacchetto modulare di chimica computazionale con un particolare focus sulle simulazioni ibride meccanica quantistica/meccanica molecolare (QM/MM) nell'ambito della biologia, chimica e scienza dei materiali [2]; FireWorks è un sistema di gestione dei flussi di lavoro per eseguire calcoli ad alta intensità computazionale in chimica computazionale e scienza dei materiali nei centri di supercalcolo [3]; ASE è un pacchetto software scritto nel linguaggio di programmazione Python con l'obiettivo di configurare, controllare e analizzare simulazioni atomistiche [4]; PyCDT è un toolkit Python che facilita la configurazione e l'analisi post-processing dei calcoli dei difetti puntuali di materiali semiconduttori e isolanti con suite DFT ampiamente disponibili [5]; QMflows consente agli utenti di scrivere flussi di lavoro complessi che coinvolgono più codici di chimica quantistica ed eseguirli in modo efficiente su computer paralleli di grandi dimensioni [6]; Automate (<https://atomate.org>) si basa sulle librerie Python open-source pymatgen, [7] e FireWorks. Il database Materials Project [8] è alimentato da Automate. Automate non solo può essere utilizzato come linguaggio per definire flussi di lavoro dinamici, ma fornisce anche molti flussi di lavoro comunemente utilizzati, come quelli per la struttura elettronica delle bande, il tensore elastico e il climbing-image nudged elastic band (CI-NEB). AiiDA è un'infrastruttura open-source in Python per aiutare i ricercatori ad automatizzare, gestire, conservare, condividere e riprodurre i complessi flussi di lavoro associati alla moderna scienza computazionale e tutti i relativi dati [9]. AiiDA può definire flussi di lavoro dinamici e il motore di AiiDA è in grado di sostenere carichi di lavoro ad alta capacità che coinvolgono decine di migliaia di lavori simultanei all'ora. AiiDA dispone anche di un registro dettagliato dell'esecuzione del flusso di lavoro chiamato grafo di provenienza. Vale la pena menzionare che il modulo plugin di AiiDA permette l'interfacciamento con software di simulazione comunemente utilizzati, come VASP, Quantum ESPRESSO, LAMMPS e GROMACS.

Il metodo automatizzato del workflow computazionale è stato applicato con successo nel campo dei materiali per batterie. Kahle et al.[10] hanno progettato un workflow ad alto rendimento a forma di imbuto per la ricerca di elettroliti a stato solido. AiiDA è stato utilizzato per la costruzione ed esecuzione del workflow. Dai

circa 1400 composti al litio distinti presenti nel database della Struttura Cristallina Inorganica (ICSD)[11] e nel database Crystallography Open Database (COD)[12], sono stati selezionati cinque conduttori di ioni veloci.

Nel campo dei materiali per batterie, i metodi automatizzati del workflow computazionale vengono utilizzati anche nello studio di elettroliti liquidi tradizionali e additivi. Poiché gli elettroliti liquidi sono principalmente composti da molecole, Halls et al. [13] hanno progettato un workflow ad alto rendimento per la simulazione delle proprietà molecolari fondamentali. Pipeline Pilot è stato utilizzato per la costruzione ed esecuzione del workflow.

Un altro lavoro incentrato sulle strutture molecolari e le proprietà degli elettroliti liquidi è stato condotto da Qu e collaboratori.[14] FireWorks è stato utilizzato per la costruzione ed esecuzione del workflow. Questi flussi di lavoro possono essere utilizzati nella selezione di elettroliti liquidi, determinando la finestra elettrochimica dal potenziale di ionizzazione e l'affinità elettronica.

Hanke et al. [15] hanno costruito un workflow computazionale automatizzato su più scale che consente simulazioni su più scale dalla composizione dell'elettrolita alle proprietà di trasporto dell'elettrolita e alle prestazioni della batteria. Pipeline Pilot (<https://www.3ds.com/products-services/biovia/products/data-science/pipeline-pilot>) è stato utilizzato per la costruzione ed esecuzione del workflow. Il workflow su più scale progettato da Hanke e colleghi consente l'esplorazione di elettroliti che soddisfano i requisiti desiderati senza compromettere le prestazioni della batteria.

Bölle et al. [16] hanno costruito un workflow di screening automatizzato per elettrodi di intercalazione in batterie basato sui calcoli della variazione di volume durante la carica/scarica, dell'energia di adsorbimento del portatore di carica e delle barriere di diffusione. MyQueue (<https://myqueue.readthedocs.io/en/latest>) è stato utilizzato per la costruzione ed esecuzione del workflow. Questo workflow è stato testato per identificare potenziali materiali catodici per batterie a ioni di magnesio dalla determinazione della stabilità degli elettrodi, la tensione a circuito aperto e la barriera di diffusione.

Oltre ai materiali per batterie, il metodo automatizzato di workflow computazionale è stato applicato anche in diversi altri campi, come la scienza delle superfici [17-18], i materiali bidimensionali [19-20] e i materiali fotoelettrochimici [21].

È anche importante migliorare l'interoperabilità e il riutilizzo dei dati pubblicati, e ciò è reso possibile dalla adozione dei principi FAIR [22], che richiedono che i dati scientifici siano reperibili, accessibili, interoperabili e riutilizzabili. Gli schemi del flusso di lavoro forniscono uno strumento essenziale per migliorare l'interoperabilità e il riutilizzo dei dati pubblicati. Esistono già esempi in cui i dati per i materiali sono raccolti in set di dati in repository che soddisfino i principi FAIR di condivisione dei dati come NOMAD (<https://nomad-lab.eu/>) o Materials cloud (<https://www.materialscloud.org/>).

Questa LA illustra il workflow automatico di calcoli ab initio per il calcolo e la collezione dei dati in un database centrale delle proprietà elettroniche di modelli materiali catodici basati su ossidi metallici lamellari del tipo MeTMO₂ (TM = transition metal) puri e drogati generati ed eseguiti secondo il protocollo predisposto nella LA1.19 (Deliverable D1.22). Il workflow calcola le proprietà energetiche di strutture cristalline al variare del livello di doping e delle specie dopanti, in modo da generare un database per la ricerca dei materiali catodici più stabili, ovvero che permettono un maggiore numero di cicli di carica/scarica, e valutare i potenziali di intercalazione (potenziali di cella aperta). Il workflow è stato implementato usando AiiDA, sul cluster CRESCO di ENEA.

Il caso d'uso scelto è quello del cristallo base dell'ossido metallico lamellare NaMnO_2 , per il quale sono stati generati in automatico i modelli atomici periodici dei materiali drogati in cui sono variati l'etero-atomo drogante e il livello di drogaggio. In ogni caso, il workflow qui presentato è applicabile a qualunque materiale catodico basato su qualunque tipo di ione e su diverse infrastrutture computazionali.

Il workflow usato adotta le strategie computazionali essenziali allo sviluppo e alla progettazione di nuovi materiali catodici per l'accumulo di energia, che sono state definite nella LA 1.19 (Deliverable 1.22). L'implementazione della workchain (catena di flussi di lavoro) permette lo studio automatico dei materiali catodici drogati partendo da alcuni cristalli base di partenza. Il workflow ha usato il codice Quantum Espresso per il calcolo di proprietà elettroniche a principi primi basato sui metodi della Density Functional Theory (DFT). Il protocollo definito adotta il metodo DFT+U ai fini dell'ottimizzazione geometrica della cella periodica e delle posizioni atomiche, e del calcolo delle energie totali. Infatti, l'approccio DFT+U è risultato avere un rapporto tra accuratezza e tempi di calcolo più efficiente (Deliverable 1.22). Il workflow calcola in automatico le energie di formazione dei cristalli per valutarne la stabilità ed i potenziali di intercalazione medi per determinarne il potenziale a cella aperta per le strutture diverse per drogaggio, e analizza, inoltre, la bontà di ciascun calcolo in termini di esito positivo e/o di presenza di errori. I risultati dei calcoli sono raccolti in automatico dal codice e memorizzati in formato json e csv, e registrati sul database gestito da ENEA prototipato in LA1.5.

3 Descrizione dei metodi usati

3.1 Metodo DFT

Il calcolo delle proprietà strutturali ed elettroniche dei sistemi target sfrutterà principalmente lo schema della teoria DFT [23], come implementato nell'applicazione PWscf del pacchetto Quantum ESPRESSO (QE) (www.quantum-espresso.org) [24].

La DFT è una teoria in principio esatta per lo studio delle proprietà di stato fondamentale dei sistemi di elettroni e ioni interagenti. Il metodo è basato sulle equazioni di Kohn-Sham [25] che sono risolte in modo autoconsistente, nel quale le funzioni d'onda sono sviluppate su un set di onde piane, vengono utilizzati pseudo-potenziali (PPs) e sono utilizzate le condizioni periodiche al contorno (PBC).

Il modello del materiale catodico base è il cristallo NaMnO_2 nella fase P2, determinata tramite diffrazione a raggi X da Paulsen e coautori [26]. La struttura è costituita da 8 strati esagonali sovrapposti, come è possibile osservare in Fig. 1. Questa struttura è stabile in un ampio intervallo di concentrazioni di atomi Na [27]. Le celle che abbiamo esaminato per i materiali catodici sono formate da 48 e 96 atomi, poiché dobbiamo introdurre l'effetto del doping con diverse concentrazioni.

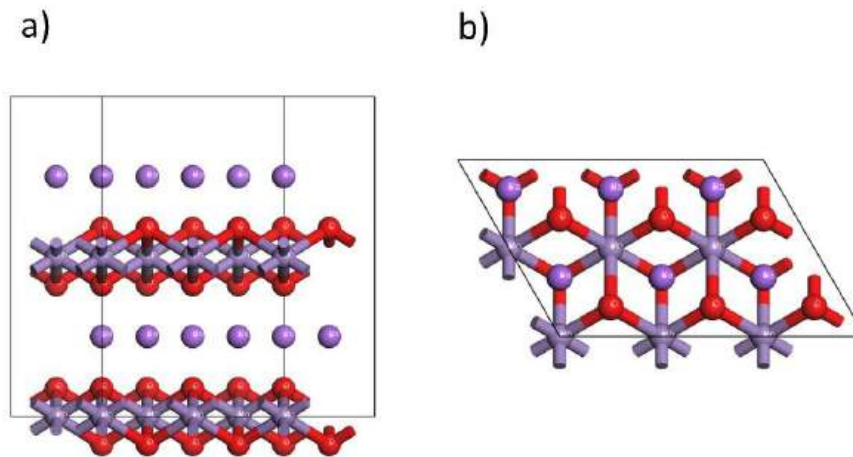


Figura 1. Modello atomistico della cella periodica del cristallo NaMnO_2 , nella fase P2 (qui è mostrata la cella di 48 atomi) nelle visuali a) laterale, b) da sopra (le sfere di diverso colore indicano le posizioni di specie atomiche diverse. Rosso: ossigeno; viola: sodio; grigio: manganese).

Il cut-off dell'energia usato per i calcoli è 80 Ry, la mesh di punti k usata per il sampling della zona di Brillouin è $8 \times 4 \times 4$ e $4 \times 4 \times 4$ per la cella di 48 e 96 atomi, rispettivamente. Lo scambio e correlazione è stato approssimato con il funzionale GGA-PBE [28]. È stato usato l'approccio DFT+U per avere una migliore descrizione dello scambio e della correlazione elettronica.

Lo scopo del drogaggio sostituzionale è quello di aumentare la ciclabilità rendendo il materiale più stabile. L'energia di formazione è stata calcolata dalla differenza dell'energia totale del sistema nella cella e quella dei suoi costituenti, presi nella forma stabile a temperatura ambiente. Le strutture su cui è stato operato il drogaggio sostituzionale aventi energia di formazione più bassa corrispondono alle conformazioni cristalline più stabili. La tensione media tra due limiti di intercalazione ($x_2=1$ e $x_1=0$) del materiale è stata anche calcolata a partire dall'energia totale per unità di formula del materiale e l'energia totale nella struttura cubica del sodio, come illustrato nel Deliverable 1.22.

3.2 Workchain

Una workchain, o catena di workflow, è l'automazione totale o parziale di un processo lavorativo, suddiviso in compiti più semplici (task) elaborati in sequenza o in parallelo secondo un modello logico al fine di ottenere gli stessi risultati in modo efficiente ed efficace. Nel nostro caso di studio, la workchain ha l'obiettivo di generare automaticamente delle strutture catodiche, variazioni della struttura di partenza (NaMnO_2) attraverso il drogaggio sostituzionale o la variazione del numero di ioni intercalati, eseguire il calcolo DFT+U e archiviare i dati ottenuti allo scopo di scoprire la struttura ottimale rispetto all'energia di formazione e al potenziale di intercalazione (redox).

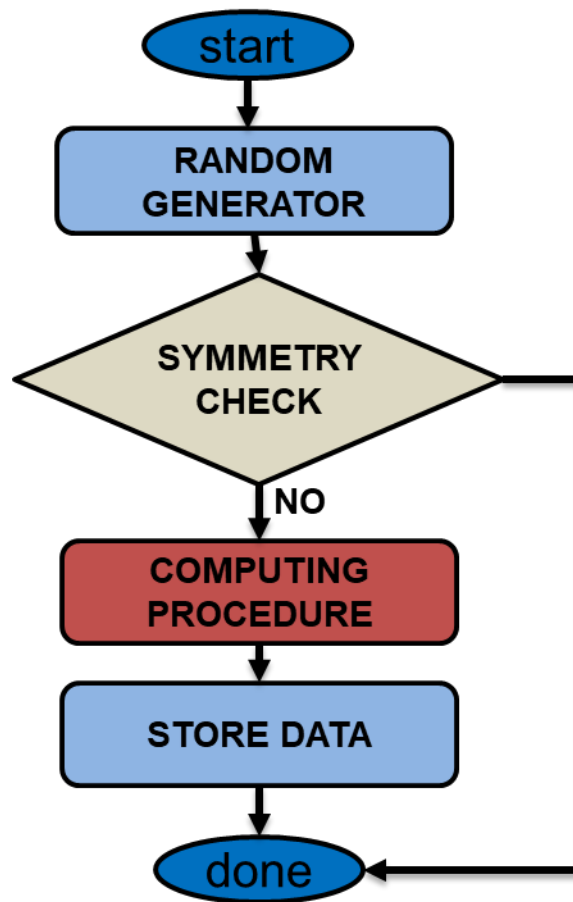


Figura 2. Diagramma di flusso della workchain.

La workchain, mostrata in Figura 2, è composta da quattro task principali:

1. **Random generator** – Creazione di una variazione dalla struttura di partenza in modo randomico;
2. **Symmetry equivalence check** – Verifica di equivalenza della simmetria tra il sistema random, generato nello step precedente, e i sistemi precedentemente calcolati;
3. **Computing procedure** – Esecuzione del calcolo atomico usando il codice Quantum ESPRESSO;
4. **Store data** – Memorizzazione dei dati nella base di dati.

Per l'implementazione della workchain si è deciso di utilizzare AiiDA (www.aiida.net), un'infrastruttura python open-source. AiiDA permette la costruzione, l'automazione e la gestione di workflow complessi associati alla scienza computazionale moderna e ai loro dati. Inoltre, permette di determinare la provenienza dei risultati prodotti attraverso la rappresentazione del processo di calcolo in una struttura di grafi aciclici.

AiiDA ha delle interfacce che le consentono di essere flessibile ed è supportata dalla community sia per lo sviluppo di nuovi plugin che per l'integrazione di codici di calcolo e delle infrastrutture di calcolo HPC. Nel nostro caso non è stato necessario implementare plugin in quanto già esistenti sia per il supporto al codice Quantum ESPRESSO che all'infrastruttura HPC ENEA CRESCO (maggiori informazioni nel D1.1).

I principali concetti di AiIDA usati nella nostra wokchain sono:

- Calculation (calcolo)
 - calculation function;
 - calculation job.
- Workflow
 - work function;
 - work chain.

Essi sono stato descritti nel Deliverable D1.22. Sulla base di questi concetti è stata definita una work chain che richiama, nell'ordine stabilito, tre calculation function e un calculation job. La logica dei task *Random system*, *Symmetry equivalence check* e *Store data* sono racchiusi nelle rispettive funzioni di calcolo mentre la logica di *Computing procedure* è stata racchiusa nel calculation job fornito tramite il plugin aiida-quantumespresso.

Nel seguito viene descritto come è stato eseguito effettivamente il workflow sulla infrastruttura di Cresco e la procedura di archiviazione dei dati sul database interno di ENEA.

4 Esecuzione del Workflow

Al fine di procedere con l'esecuzione del workflow è necessario procedere prima con l'installazione dell'environment.

Inoltre, sono state rese disponibili due modalità di esecuzione, la prima basata sull'utilizzo del tool Anaconda e la seconda basata sulla virtualizzazione tramite Singularity.

4.1 Installazione

Gli step richiesti per l'installazione sono:

1. Scaricare il repository dal GitLab ENEA in una propria directory eseguendo il comando:

```
git clone git@gitlab.brindisi.enea.it:<repository>/ai4mat.git
```

2. Configurare le credenziali di accesso al database tramite un file delle proprietà (l'utente deve avere i permessi di scrittura). All'interno vengono definiti il server, la porta di accesso, il nome del database, l'utente e la password, la tabella o collezione e il meccanismo di autenticazione.
3. Se non si è mai provveduto, configurare AiiDA definendo il file config.json e riportarlo nel percorso ~/.aiida/config.json. Di seguito un template di riferimento:

```
{
  "CONFIG_VERSION": {
    "CURRENT": <>,
    "OLDEST_COMPATIBLE": <>
  },
  "profiles": {
    "ai4mat": {
      "PROFILE_UUID": <UUID>,
      "AIIDADB_ENGINE": "postgresql_psycopg2",
      "AIIDADB_BACKEND": "django",
      "AIIDADB_NAME": "ai4mat",
      "AIIDADB_PORT": 5432,
      "AIIDADB_HOST": "aiida.service.com",
      "AIIDADB_USER": "aiidausr",
      "AIIDADB_PASS": <pwd>,
      "broker_protocol": "amqp",
      "broker_username": "aiidausr",
      "broker_password": <pwd>,
      "broker_host": "aiida.service.com",
      "broker_port": 5672,
      "broker_virtual_host": "",
      "AIIDADB_REPOSITORY_URI": "path/to/.aiida/repository/ai4mat",
      "options": {
        "logging.aiida_loglevel": "REPORT"
      }
    }
  }
}
```

```

    },
    "default_user_email": "aiida.user@email.com"
  },
  "default_profile": "ai4mat",
  "options": {
    "autofill.user.email": "aiida.user@email.com",
    "autofill.user.first_name": "Aiida",
    "autofill.user.last_name": "User",
    "autofill.user.institution": <institution>
  }
}

```

4.2 Configurazione environment

All'interno dell'infrastruttura di calcolo ENEA, sono state definiti due ambienti di esecuzione del workflow AiiDA. Il primo si basa sull'attivazione di un ambiente virtuale tramite Anaconda, mentre il secondo si basa sull'uso di un container tramite Singularity.

4.2.1 Conda environment

Conda consente di creare, esportare, elencare, rimuovere e aggiornare ambienti in cui sono installate diverse versioni di Python e/o pacchetti. Il passaggio da un ambiente all'altro si chiama attivazione dell'ambiente.

L'ambiente python è stato installato nell'ambiente virtuale conda 'aiida' ed è attivabile con il comando:

```
conda activate aiida
```

Se il comando conda o l'ambiente conda 'aiida' non vengono trovati è necessario rispettivamente importare il comando conda nella variabile dei path o impostare la cartella di progetto come repository di conda.

4.2.2 Singularity container

Un container è un'unità standard di software che impacchetta il codice e tutte le sue dipendenze in modo che l'applicazione venga eseguita in modo rapido e affidabile da un ambiente informatico all'altro. Un'introduzione ai container è presente al link <https://singularity-tutorial.github.io/00-introduction/>.

Il workflow invia il calcolo 'pw' alla coda LSF, quindi l'autenticazione Kerberos del contenitore deve essere impostata affinché il contenitore possa inviarlo.

È stato creato un contenitore Singularity utilizzando la seguente recipe:

```

Bootstrap: docker

From: aiidateam/aiida-core:1.6.5

%environment
# -----
  export TZ="Europe/Zurich"

%labels
# -----

```

```

Author aiidausr@enea.it
Version v.1.0

%post
# -----
# Update
apt-get update -y
# Install Kerberos
DEBIAN_FRONTEND=noninteractive apt-get install -y krb5-user
apt-get install -y libffi-dev libkrb5-dev
# Install pymongo and aiida-quantumespresso
PATH=/opt/conda/bin:$PATH
pip install aiida-core[atomic_tools,ssh_kerberos] -U
pip install pymongo aiida-quantumespresso
reentry scan

%runscript
# -----
verdi status

```

Una volta caricato il modulo Singularity è possibile eseguire il container come segue:

```

singularity exec --bind </usefull/path> $AIIDACONTAINER python main.py -cp
<PARAMETERS_FILE> -mp <CONNECTION_FILE> -xyz <XYZ_FILE>

```

4.3 Esecuzione

Una volta scaricato il processo e impostata l'autenticazione del database, è possibile usare il workflow AiiDA eseguendo lo script python in computation/main.py.

Lo script main.py necessita di tre file in input, ovvero:

- CALCULATION_PARAMS, file che definisce i parametri del calcolo
- MONGODB_PROPERTIES, file per l'autenticazione al database
- COORDINATES, file del sistema da elaborare.

In aggiunta, è possibile indicare se svolgere l'elaborazione direttamente sul sistema indicato oppure se applicare una serie di modifiche prima che venga elaborato tramite l'inserimento dell'opzione *default_generator*.

Tale script ha la seguente command line interface:

```

usage: main.py [-h] [-cp CALCULATION_PARAMS] [-mp MONGODB_PROPERTIES] [-xyz COORDINATES] [--default_generator]

```

AI4MAT - AiiDA Workflow

optional arguments:

```

-h, --help            show this help message and exit
-cp CALCULATION_PARAMS, --calculation-params CALCULATION_PARAMS
                        Path of the calculation parameters
-mp MONGODB_PROPERTIES, --mongodb-properties MONGODB_PROPERTIES
                        Path of the MongoDB properties
-xyz COORDINATES, --coordinates COORDINATES

```

Path of the material coordinates
 --default_generator If true the structure generator function returns the same input structure otherwise the random structure

4.4 Parametri di esecuzione

I parametri disponibili all'utente sono riportati in Tabella 1.

Tabella 1. Lista dei parametri

Name	Description	Note
parameters	It contains the Quantum ESPRESSO <i>pw</i> input (https://www.quantum-espresso.org/Doc/INPUT_PW.html)	
kpoints	Define the kpoints <i>mesh</i> and <i>value</i>	
pseudos	Define the <i>UPF file</i> for each element of the material	It isn't required
metadata.options.resources	Define the cores' number with <i>tot_num_mpirprocs</i>	
metadata.options.queue_name	Define the LSF queue to submit the Quantum ESPRESSO calculation	
metadata.options.prepend_text	Define the pre-run commands	
metadata.options.append_text	Define the post-run commands	
parallelization	Define the parallelization level with <i>pool</i>	Default value: 1
codename	It contains the <i>code</i> and the <i>computer</i> running the workflow	The format is code@computer
rules	It's the list of rules called in the material structure generator step	Choices: ['sodium_removing', 'manganese_replacement', 'all_sodium_removing']. Default value: ['manganese_replacement']
manganese_replacement_step	It's the step of replacements for each elements	Default value: 2
manganese_min_number	It's the minimum number of the manganese atoms in the system	Default value: $\text{int}(\text{len}(\text{Mn_idx})/2)$ (almost 50% of manganese atoms)
transaction_metals	It's the list of transaction metals used in the manganese replacement rule	Default value: ['Ni', 'Ti']

5 Collezione dei dati sul database

I risultati dell'elaborazione del workflow vengono infine memorizzati su database. Il formato utilizzato è JSON (JavaScript Object Notation) al cui interno le informazioni sono riportate nella struttura dati mostrata in Tabella 2.

Il dato prodotto da workflow tramite AiiDA è stato suddiviso in tre macro-gruppi, ovvero:

- **Inputs:** riporta la struttura del sistema, i parametri di esecuzione e i k-point e gli pseudo-potenziali da utilizzare.
- **Outputs:** riporta i risultati (e.g. energia totale, energia di fermi, ...), la traiettoria e le bande ottenute dall'elaborazione di Quantum ESPRESSO.
- **Metadata:** ingloba le informazioni aggiuntive riguardo l'elaborazione (e.g. la work directory, le opzioni di esecuzione, puntamenti agli oggetti di AiiDA, ...)

Rispetto ai campi riportati in Tabella 2, due proprietà legate al nostro caso d'uso vengono successivamente calcolate. Le due proprietà sono l'energia di formazione e il potenziale di intercalazione ricavabili dall'energia totale.

Tabella 2. Struttura dati

Name	Type	Index	Description
inputs.structure.CELL_PARAMETERS	Array of float	N	Informazioni sulla dimensione della cella in angstrom.
inputs.structure.POSITIONS	Array of float	N	Lista delle coordinate (espresse in angstrom) di ciascun atomo all'interno della cella.
inputs.structure.SYMBOLS	Array of string	N	Lista dei simboli atomici relativi a ciascuna posizione in inputs.structure.POSITIONS.
inputs.structure.formula	String	Y	Informazione sulle proporzioni chimiche degli atomi presenti all'interno della cella.
inputs.structure.xyz_file	String	N	Contenuto del relativo file .xyz.
inputs.parameters.CONTROL	Dictionary	N	Contiene tutti i parametri contenuti nella sezione <i>SYSTEM</i> del file di input in base alle disposizioni indicate da Quantum ESPRESSO.
inputs.parameters.SYSTEM	Dictionary	N	Contiene tutti i parametri contenuti nella sezione <i>SYSTEM</i> del file di input in base alle disposizioni indicate da Quantum ESPRESSO.
inputs.parameters.ELECTRONS	Dictionary	N	Contiene tutti i parametri contenuti nella sezione <i>ELECTRONS</i> del file di input in base alle disposizioni indicate da Quantum ESPRESSO.
inputs.parameters.IONS	Dictionary	N	Contiene tutti i parametri contenuti nella sezione <i>IONS</i> del file di input in base alle disposizioni indicate da Quantum ESPRESSO.

Name	Type	Index	Description
inputs.parameters.CELL	Dictionary	N	Contiene tutti i parametri contenuti nella sezione <i>CELL</i> del file di input in base alle disposizioni indicate da Quantum ESPRESSO.
inputs.kpoint.value	Array of float	N	Valore dei punti K.
inputs.kpoint.offset	Array of float	N	Offset dei punti K.
inputs.pseudo	Dictionary	N	Informazione del filepath UPF e del valore della massa atomica per ciascun atomo presente nella cella.
outputs.output_parameters	Dictionary	N	Risultati ricavati da AiiDA.
outputs.output_bands	List of array	N	Informazione sulla struttura a bande.
outputs.output_trajectory.cif_file	String	N	Contenuto del file CIF recuperato al completamento del calcolo. N.B.: Informazione di interesse in caso sia stato impostato un calcolo che prevede il rilassamento del sistema.
outputs.trajectories.positions	Array of array of float	N	Informazione sulle traiettorie ricavate da Quantum ESPRESSO.
outputs.remote_folder.remote_path	String	N	Percorso della macchina remota.
outputs.remote_folder.list_dir	List of string	N	Lista dei file e delle directory presenti in <i>outputs.remote_folder.remote_path</i> .
metadata.options	Dictionary	N	Informazioni aggiuntive inserite per la corretta esecuzione del calcolo.
metadata.aaida_references.calcjob	String	N	Identificativo dell'istanza di calcolo di AiiDA.
metadata.aaida_references.inputs	Dictionary	N	Identificativi delle istanze per ciascuna tipologia input di AiiDA.
metadata.aaida_references.outputs	Dictionary	N	Identificativi delle istanze per ciascuna tipologia output di AiiDA.
metadata.type	String	Y	Informazione sulla tipologia del calcolo. N.B.: Per i calcoli pw di Quantum ESPRESSO il valore da usare è 'quantumespRESSO.pw'
metadata.description	String	N	Informazione sulla tipologia dello studio. N.B.: Per lo studio sui catodi di sodio per le batterie il valore da usare è 'sodium cathodes'.
metadata.remote_workdir	String	N	Percorso di lavoro.
metadata.retrieved_list	List of string	N	Lista dei file ottenuti dal calcolo.
metadata.scheduler_info.job_id	Integer	N	Identificativo del job di calcolo.
metadata.scheduler_info.detailed_job_info	String	N	Informazioni sul job di calcolo.

Name	Type	Index	Description
<code>metadata.scheduler_info.scheduler_state</code>	String	N	Stato dello scheduler.
<code>metadata.scheduler_info.scheduler_last_checktime</code>	String	N	Data e ora in cui lo scheduler ha effettuato l'ultima verifica.
<code>metadata.scheduler_info.stdout</code>	String	N	Contenuto del file standard-output dello scheduler.
<code>metadata.scheduler_info.stderr</code>	String	N	Contenuto del file standard-error dello scheduler.

5.1 Statistiche

Al momento il workflow AiiDA ha generato 1107 elaborazioni che sono state memorizzate come documenti JSON nel database MongoDB. I calcoli sono stati eseguiti sull'infrastruttura di calcolo HPC di ENEA sfruttando il calcolo parallelo su 12 nodi per un totale di 576 core.

Le elaborazioni completate sono 766, mentre quelle concluse con errore sono 341. Gli errori sono dovuti per il 50% da limiti di tempo a disposizione per l'esecuzione (superando le 48 ore disponibili), il 30% da impossibilità del calcolo DFT di vertere a convergenza e il 20% da problemi manutenzione dell'infrastruttura di calcolo ENEA che hanno portato all'interruzione dell'elaborazione.

I calcoli completati con successo hanno richiesto in media dei tempi di elaborazione di circa 3h e 10 min utilizzando in media quasi 600GB di RAM totali. Nel dettaglio, Figura 5.1, Figura 5.2, Figura 5.3 e Figura 5.4 mostrano rispettivamente il numero di calcoli raggruppati per tempo di esecuzione richiesto, energia totale ed energia di fermi calcolate, e il numero di iterazioni SCF totali.

Infine, è stato riportato in Tabella 3 un estratto dei calcoli memorizzati su database. In Figura 5.5 sono mostrate una selezione di configurazioni visualizzate con ASE (usato da AiiDA per la visualizzazione delle strutture).

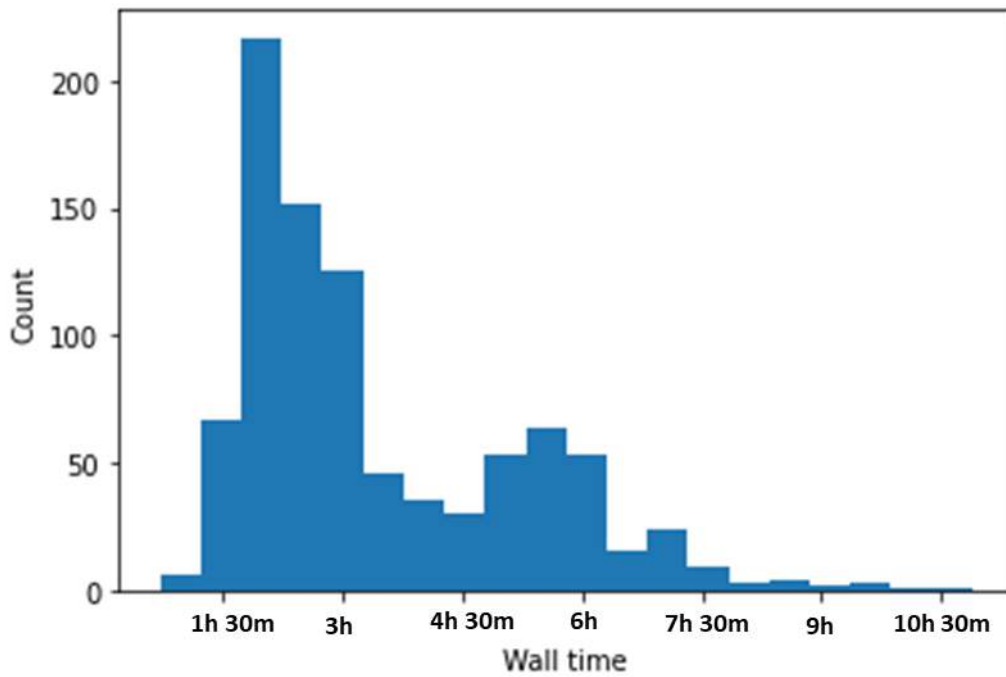


Figura 5.1 Numero elaborazioni per tempo di lavoro richiesto

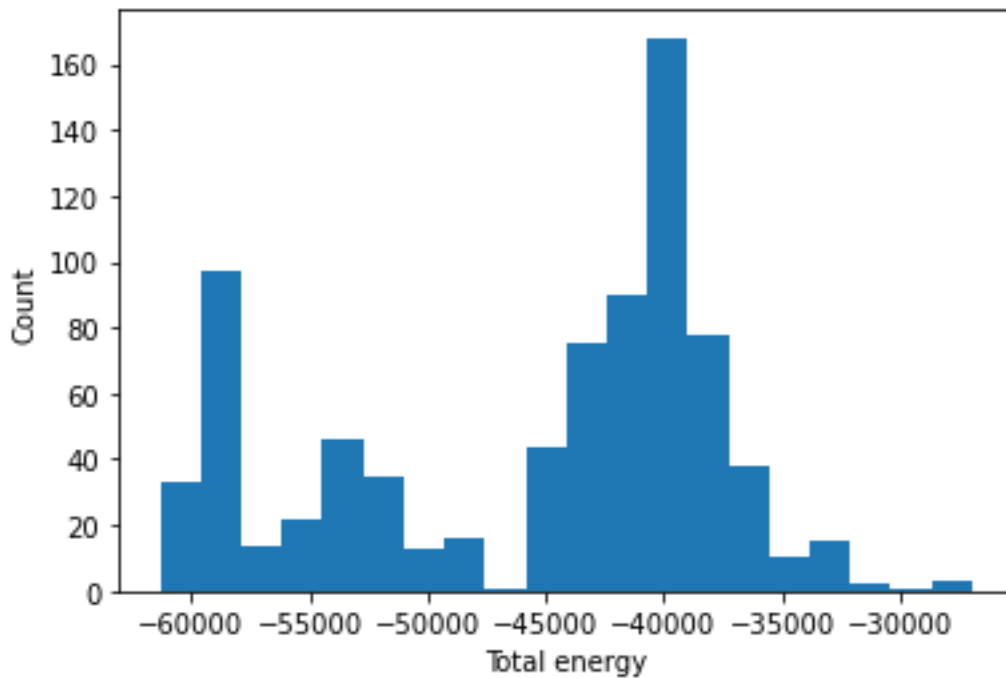


Figura 5.2 Numero elaborazioni per energia totale calcolata

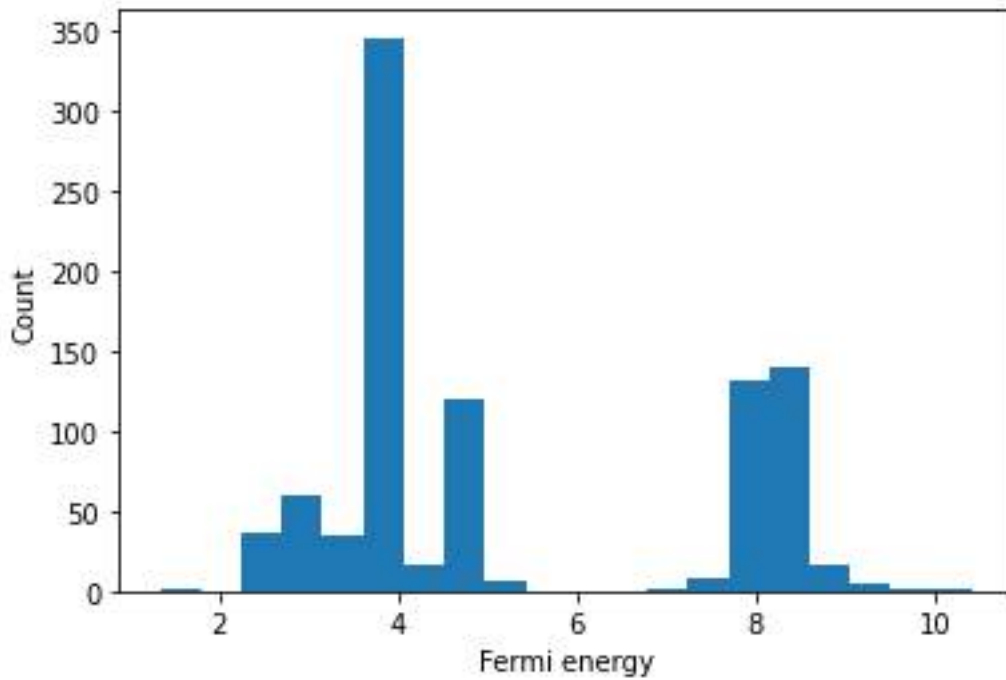


Figura 5.3 Numero elaborazioni per energia di Fermi calcolata

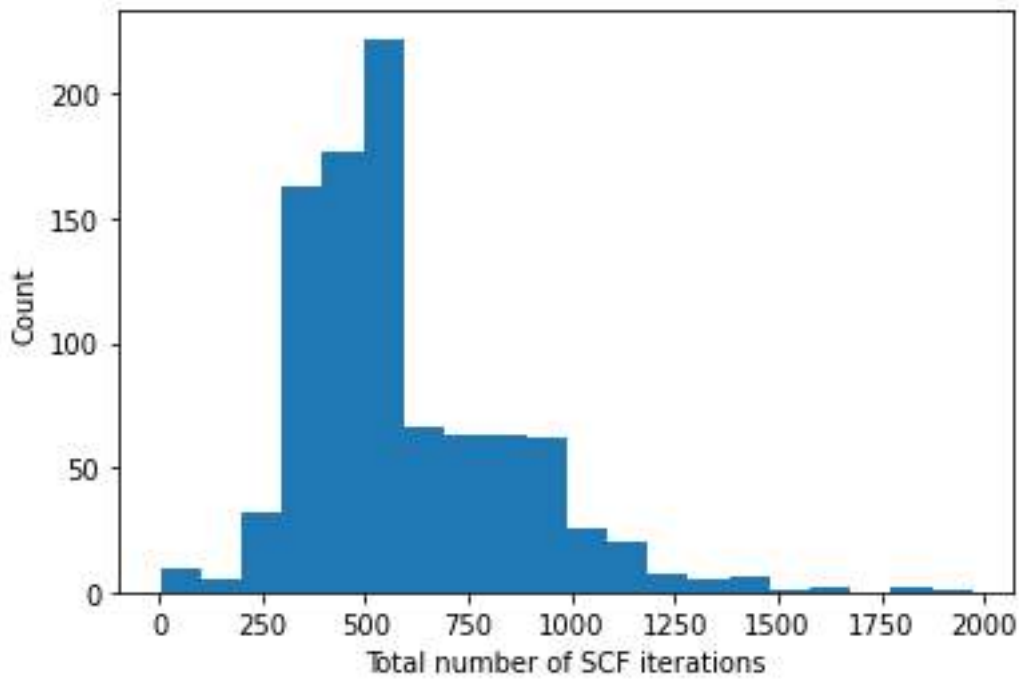


Figura 5.4 Numero elaborazioni per numero totale di iterazioni SCF svolte

Tabella 3 Esempio di subset elaborazioni memorizzati su database

ID	Fermi energy	Total energy	Wall time seconds	Estimated RAM total	Total number of SCF iterations	...
0	7.9021153210704	-61347.839300087	14820	52.43	500	...
1	7.9153679242804	-52098.217068389	16320	52.43	500	...
2	7.9988374202266	-58264.564356853	16920	52.43	500	...
3	7.6136504519543	-61347.839300087	16620	52.43	500	...

ID	Fermi energy	Total energy	Wall time seconds	Estimated RAM total	Total number of SCF iterations	...
4	7.743	-61347.839300087	12420	55.97	677	...
5	7.9347	-61347.839300087	17820	52.66	1382	...
6	7.7826367592424	-61347.839300087	10020	52.43	500	...
7	7.9422	-52097.904285781	16260	56.69	971	...
8	8.3898759912078	-55180.95244744	26700	52.43	500	...
9	7.9421063935514	-61347.839300087	15900	53.63	500	...
10	7.743	-51149.096020383	12360	55.97	677	...
11	7.743	-51096.64184782	12360	55.97	677	...
12	7.743	-52061.87832923	12420	55.97	677	...
13	7.743	-53148.744303817	12420	55.97	677	...
14	7.9603	-52617.551626267	16560	52.57	1252	...
15	7.853	-52617.425149797	15540	54.58	1081	...
16	7.9896194821973	-58790.843126452	12120	53.63	500	...
17	7.743	-53147.887075985	12420	55.97	677	...
18	8.0109310051818	-50062.428756243	19440	55.39	500	...
19	10.4157	-49527.592816496	11040	54.09	671	...
20	8.1207	-55172.076206405	28140	107.87	1140	...
21	7.8414	-55703.922449412	29640	98.98	1790	...
22	8.4593	-58790.843126452	13740	106.3	639	...
23	8.1575	-56235.501304249	17280	103.3	895	...
24	8.1205	-58790.700893639	18480	103.43	914	...
25	8.4605	-53678.411381684	18540	107.56	793	...
26	8.4799	-53149.303029031	18480	106.3	945	...
27	8.1802	-61345.9556137	11820	103.27	635	...
28	7.6535	-61345.9556137	13500	102.86	741	...
29	8.2856619705145	-55704.383395281	6360	104.14	500	...
30	8.0842	-52617.463831187	17160	100.25	888	...
31	7.7421674302888	-53678.513510769	6120	100.49	500	...
32	8.4895	-55702.984368125	15480	106.46	730	...
33	8.4605	-52085.451874256	18600	107.56	793	...
34	8.3502	-53677.034623235	16560	106.8	689	...
35	8.6121	-55172.425353898	20580	107.46	879	...
36	7.7107680085393	-58790.70261748	8220	100.49	500	...
37	9.0616	-55704.114828451	25920	106.1	1201	...
38	8.4807	-53149.009528273	13500	106.43	623	...
39	8.001	-53679.130407117	32220	103.55	1414	...
40	8.001	-42920.374819364	32700	103.55	1414	...
41	8.3275	-45475.479326874	32880	107.21	999	...
42	8.0777	-42920.51827737	18720	103.39	988	...
43	9.2875	-39831.726293762	24780	106.29	1160	...
44	8.6554	-40365.412232689	17640	105.42	1121	...
45	8.0652	-45475.479326874	8400	97.34	419	...
46	8.9783	-37276.58634601	31140	106.35	1425	...

ID	Fermi energy	Total energy	Wall time seconds	Estimated RAM total	Total number of SCF iterations	...
47	8.2379	-36743.030167831	21420	99.98	836	...
48	8.6075	-37811.08213775	23520	107.51	1036	...
49	8.015	-39831.541455678	33960	105.42	1780	...
50	8.3955	-39831.625051225	17940	106.43	867	...
...						

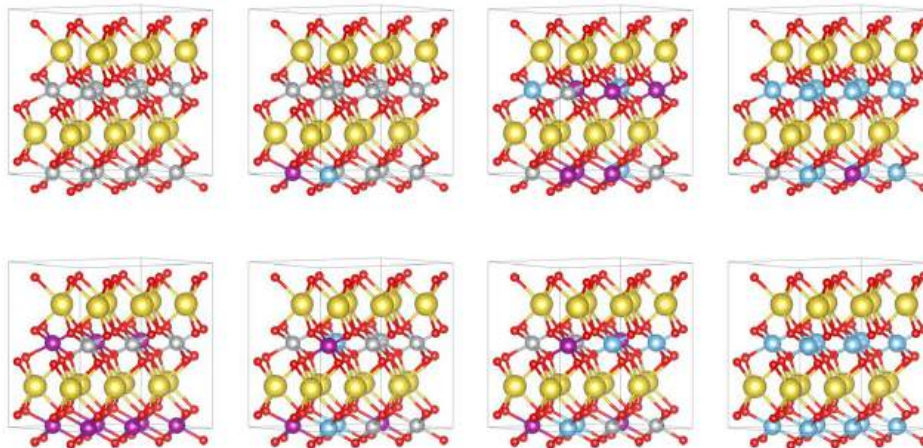


Figura 5.5. Selezione di strutture atomiche presenti nel database (le sfere di diverso colore indicano le posizioni di specie atomiche diverse. Rosso: ossigeno; giallo: sodio; celeste: titanio; grigio: nickel; viola:manganese).

6 Conclusioni

È stato sviluppato un workflow automatico di calcoli ab initio per materiali catodici basati su ossidi metallici lamellari del tipo MeTMO₂ (TM = transition metal). Abbiamo qui presentato le modalità di installazione dell'environment e di esecuzione, e mostrate alcune statistiche dei dati prodotti e archiviati.

Il workflow è stato implementato usando AiiDA, sul cluster CRESCO di ENEA, e ha usato il codice Quantum Espresso per il calcolo di proprietà elettroniche a principi primi basato sui metodi DFT. Il caso d'uso scelto è quello del cristallo base dell'ossido metallico lamellare NaMnO₂, per il quale sono stati generati in automatico i modelli atomici periodici dei materiali drogati. In questi ultimi sono variati l'etero-atomo drogante e il livello di drogaggio. Tuttavia, il workflow può essere esteso a qualunque materiale catodico basato su altri tipi di ione.

Due modalità di esecuzione sono state rese disponibili: la prima basata sull'utilizzo del tool Anaconda; la seconda basata sulla virtualizzazione tramite Singularity. Al momento il workflow AiiDA ha generato 1107 elaborazioni che sono state memorizzate come documenti JSON nel database MongoDB.

I workflow automatici che abbiamo illustrato nel presente rapporto giocheranno un ruolo essenziale nello svolgimento della ricerca di materiali per l'accumulo elettrochimico con le migliori prestazioni, in quanto forniranno ai ricercatori un insieme predefinito di parametri computazionali ed un insieme di dati che saranno elaborati tramite algoritmi predittivi per accelerare della ricerca.

7 Riferimenti bibliografici

- [1] C. R. Jacobs et al. PyADF - A Scripting Framework for Multiscale Quantum Chemistry. *J. Comput. Chem.* 32 (2011) 2328.
- [2] S. Metz, J. Kastner, A. A. Sokol, T. W. Keal, P. Sherwood. ChemShell - a modular software package for QM/MM simulations. *WIREs Comput. Mol. Sci.* 4 (2014) 101.
- [3] Jain et al. FireWorks: a dynamic workflow system designed for high throughput applications. *Concurrency Computat.: Pract. Exper.* 27 (2015) 5037.
- [4] H. Larsen et al. The atomic simulation environment - a Python library for working with atoms. *J. Phys.: Condens. Matter* 29 (2017) 273002.
- [5] D. Broberg, B. Medasani, N. E. R. Zimmermann, G. Yu, A. Canning, M. Haranczyk, M. Asta, G. Hautier. PyCDT: A Python toolkit for modeling point defects in semiconductors and insulators. *Computer Physics Communications* 226 (2018) 165.
- [6] F. Zapata, L. Ridder, J. Hidding, C. R. Jacob, I. Infante, L. Visscher. QMflows: A Tool Kit for Interoperable Parallel Workflows in Quantum Chemistry. *J. Chem. Inf. Model.* 59 (2019) 3191.
- [7] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* 2013, 68, 314.
- [8] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* 2013, 1, 011002.
- [9] K G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, B. Kozinsky. AiiDA: Automated Interactive Infrastructure and Database for Computational Science. *Computational Materials Science* 111 (2016) 218.

- [10] L. Kahle, A. Marcolongo, N. Marzari, High-throughput computational screening for solid-state Li-ion conductors. *Energy Environ. Sci.* 2020, 13, 928.
- [11] A. Belsky, M. Hellenbrandt, V. L. Karen, P. Luksch, New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Crystallogr. Sect. B Struct. Sci.* 2002, 58, 364.
- [12] S. Grazulis, A. Daskevicius, A. Merkys, D. Chateigner, L. Lutterotti, M. Quiros, N. R. Serebryanaya, P. Moeck, R. T. Downs, A. Le Bail, Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Res.* 2012, 40, D420.
- [13] M. D. Halls, K. Tasaki, High-throughput quantum chemistry and virtual screening for lithium ion battery electrolyte additives. *J. Power Sources* 2010, 195, 1472.
- [14] X. Qu, A. Jain, N. N. Rajput, L. Cheng, Y. Zhang, S. P. Ong, M. Brafman, E. Maginn, L. A. Curtiss, K. A. Persson, The Electrolyte Genome project: A big data approach in battery materials discovery. *Comput. Mater. Sci.* 2015, 103, 56.
- [15] F. Hanke, N. Modrow, R. L. C. Akkermans, I. Korotkin, F. C. Mocanu, V. A. Neufeld, M. Veit, Multi-Scale Electrolyte Transport Simulations for Lithium Ion Batteries. *J. Electrochem. Soc.* 2019, 167, 013522.
- [16] F. T. Böhle, N. R. Mathiesen, A. J. Nielsen, T. Vegge, J. M. Garcia-Lastra, I. E. Castelli, Autonomous Discovery of Materials for Intercalation Electrodes. *Batteries Supercaps* 2020, 3, 488.
- [17] J. H. Montoya, K. A. Persson, A high-throughput framework for determining adsorption energies on solid surfaces. *npj Comput. Mater.* 2017, 3, 14.
- [18] K. Tran, A. Palizhati, S. Back, Z. W. Ulissi, Dynamic Workflows for Routine Materials Discovery in Surface Science. *J. Chem. Inf. Model.* 2018, 58, 2392.
- [19] S. Hastrup, M. Strange, M. Pandey, T. Deilmann, P. S. Schmidt, N. F. Hinsche, M. N. Gjerding, D. Torelli, P. M. Larsen, A. C. Riis-Jensen, J. Gath, K. W. Jacobsen, J. J. Mortensen, T. Olsen, K. S. Thygesen, The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals. *2D Mater.* 2018, 5, 042002.
- [20] N. Mounet, M. Gibertini, P. Schwaller, D. Campi, A. Merkys, A. Marrazzo, T. Sohier, I. E. Castelli, A. Cepellotti, G. Pizzi, N. Marzari, Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds. *Nat. Nanotechnol.* 2018, 13, 246.
- [21] K. Kuhar, A. Crovetto, M. Pandey, K. S. Thygesen, B. Seger, P. C. K. Vesborg, O. Hansen, I. Chorkendorff, K. W. Jacobsen, Sulfide perovskites for solar energy conversion applications: computational screening and synthesis of the selected compound LaYS₃. *Energy Environ. Sci.* 2017, 10, 2579.
- [22] M. D. Wilkinson, et al., “The FAIR Guiding Principles for scientific data management and stewardship”, *Sci. Data* 3 (2016) 1.
- [23] P. Hohenberg, W. Kohn, “Inhomogeneous Electron Gas”, *Phys. Rev.* 136 (1964) B864-B871.
- [24] P. Giannozzi et al., “Quantum ESPRESSO toward the exascale”, *J. Chem. Phys.* 152 (2020) 154105.
- [25] W. Kohn, L. J. Sham, “Self-Consistent Equations Including Exchange and Correlation Effects”, *Phys. Rev.* 140 (1965) A1133–A1138.
- [26] J. M. Paulsen, R.A. Donaberger, J.R. Dahn, “Layered T₂-, O₆-, O₂-, and P₂-Type A₂/3[M¹₂+1/3M⁴+2/3]O₂ Bronzes, A = Li, Na; M¹ = Ni, Mg; M = Mn, Ti”, *Chem. Chem. Mater.*, 12 (2000) 2257.
- [27] R. J. Clement, P. G. Bruce, C. P. Grey, “Review—Manganese-Based P₂-Type Transition Metal Oxides as Sodium-Ion Battery Cathode Materials”, *Journal of The Electrochemical Society*, 162 (2015) A2589-A2604.

[28] J. P. Perdew, K. Burke and M. Ernzerhof, “Generalized Gradient Approximation Made Simple”, Phys. Rev. Lett., 77 (1996) 3865.

8 Abbreviazioni ed acronimi

CI-NEB = Climbing-image nudged elastic band

DFT = Teoria del funzionale della densità di carica

DFT+U = Teoria del funzionale della densità di carica con il parametro Hubbard U

FAIR (dati) = Dati che rispettano i principi di rintracciabilità (findability), accessibilità (accessibility), interoperabilità (interoperability) e riusabilità (reusability)

GGA = approssimazione del gradiente generalizzato

LA = Linea di attività

QE = Quantum ESPRESSO (software)

PBC = Condizioni periodiche al contorno

PBE = Funzionale di scambio correlazione di Perdew–Burke–Ernzerhof

PPs = Pseudo-potenziali usati nei calcoli DFT

XC = Potenziale di scambio-correlazione