Atomic layer deposition of two-dimensional transition metal dichalcogenides and ovonic threshold switching materials: A new computational framework for precursor design

A Data Management Plan created using DMPonline.be

Creator: Bram van der Linden

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 1SH9E24N

ID: 205485

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Atomic Layer Deposition (ALD) is an advanced thin film deposition technique that has become essential for industrial production of nano-electronic devices, omnipresent in our society. ALD uses self-limiting surface reactions of gas-phase precursors and is ideally suited to provide nanoscale thickness control. Still, ALD of Ovonic Threshold Switching (OTS) materials, promising selector materials in memory arrays, is a huge challenge. A second challenge is ALD of two-dimensional (2D) transition metal dichalcogenides (TMDCs), which suffer from nano-crystallinity. The development of new ALD processes and precursors is expensive and complex. This project therefore proposes an efficient computational framework based on first principles and thermodynamic simulations. The computational framework identifies the optimal precursors and process conditions and as such creates fundamental understanding of the properties and reactivity of ALD precursors. The method starts with precursor screening assuming gas-phase reactions. Next, the thermal stability of precursors is studied, as well as physisorption and chemisorption. The computational framework will be validated and applied to 2D TMDCs where we will not only investigate the ideal precursors but also the best surface inhibitor molecules to limit nano-crystallinity. We will also apply our method to ternary and quaternary OTS materials, including a study of dissociative precursor adsorption that can lead to surface site blocking.

Last modified: 28-03-2024

Atomic layer deposition of two-dimensional transition metal dichalcogenides and ovonic threshold switching materials: A new computational framework for precursor design

FWO DMP (Flemish Standard DMP)

---

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .csv,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Python library to automate Density Functional Theory (DFT) calculations	Developed within Imec. Automates the creation of input files for Density Functional Theory	Reuse existing data	Digital	Software	.py	< 1 GB	
Python library for thermodynamic simulations	Developed within Imec. Contains algorithms to determine equilibrium composition of molecules and solids	Reuse existing data	Digital	Software	.py	< 1GB	

Raw data DFT	All the DFT input and output files generated throughout the project. The input files contain the parameters used to do the DFT calculations.	Generate new data	Digital	Simulation data	.txt files but with different extensions which reflect the content of the file (e.g .cpki denotes input files, .cpko output file, .restart is the file that can be used to restart the calculation)	< 1 TB	
Raw data thermodynamic simulations	All input and output data of thermodynamic simulations generated throughout the project.	Generate new data	Digital	Simulation data	.json	< 100 GB	
Data processing	Jupyter notebooks and processed data (e.g. images). Jupyter notebooks are a combination of Python code and written explanation. In the Python code the DFT and thermodynamic simulations are set up and the results are analyzed afterwards.	Generate new data	Digital	Simulation data	.ipynb, .png, .jpeg	< 100 MB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

NA

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

The project might result in a new ALD process with other process conditions and/or precursors than before, thus affecting all aforementioned data sets. If this process is significantly different from already documented or patented processes, an opportunity arises to valorize the project. The valorization process would be done in close collaboration with the Intellectual Property experts from Imec and the Intellectual Property Unit of KU Leuven Research & Development (LRD).

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- Yes

The Python libraries for automation of DFT calculations and for automation of thermodynamic simulations are owned by Imec. These libraries can only be accessed by members of the modelling team at Imec and other groups within Imec if access is provided by Geoffrey Pourtois, head of the modelling team. The Jupyter notebooks use the Python libraries and can therefore only be accessed by the modelling team and close collaborators of the project, i.e. the supervisor, promotor and colleagues which are strongly involved in the project.

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

All the data generation and data processing are done in Jupyter notebooks. The notebooks are clearly structured in chapters, thus making the complete workflow from data generation to data processing clear. Python code is accompanied by written text which explains the approach taken and the assumptions made. To enhance readability, the code is divided into blocks with written comments that explain the function of such a block, the meaning of a variable name or why and how a specific Python function is used.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

The metadata standard of Crystallographic Information Framework (CIF) is used for crystals: these CIF files serve as an input for DFT, and the Python library to automate DFT enables transformation of DFT output into CIF structures.
The DFT calculations are ordered in folders with relevant names. Each folder contains an input file (a text file) with all the DFT settings. Such a file is standardized and can be directly fed into the DFT software CP2K to redo the calculation. The thermodynamic input and output data are stored as json files in folders with clear names. If not clear from the folder name, the Jupyter notebooks can be consulted to retrieve the content of a folder. The Python code in the Jupyter notebook contains written comments and use clear variable names. The written text in the Jupyter notebooks makes the work understandable and reproducible.

3. Data storage & back-up during the research project

Where will the data be stored?

The data will be stored on a personal Imec OneDrive and on the Imec supercomputing infrastructure. On both systems Imec security measures and back-ups are in place.

How will the data be backed up?

The OneDrive is automatically backed up. The supercomputing infrastructure is backed up by Imec.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- Yes

The storage solutions have the following capacity:

- OneDrive: 1TB
- Supercomputing infrastructure: 300 GB, which can be increased if needed.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

- The OneDrive is strictly personal and can only be accessed on a personal Imec laptop which is secured with a password. Two-factor authentication is required when accessing on a different device.
- The files on the supercomputing infrastructure can only be accessed by Linux users with the right permissions. Such permission is granted by the supervisor of the project, Geoffrey Pourtois.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

No costs are expected.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- Certain data cannot be kept for 10 years:

The Raw DFT data are too large in size to be kept for 10 years and include many irrelevant output files. Only the essential input and output data will therefore be kept, including molecular/solid structures and (Gibbs free) energies, for example.

Where will these data be archived (stored and curated for the long-term)?

The data will be stored on the supercomputing infrastructure of Imec with enough storage in place. Powerpoint presentations will be put on a

Microsoft SharePoint from Imec.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

No costs are expected.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Other, please specify:
- Yes, in a restricted access repository (after approval, institutional access only, …)
- No (closed access)

Papers and their Supplementary Information will be put in Lirias to make them available through green open access. The Supplementary Information could, for example, be the convergence of DFT parameters. Raw DFT data will be made available "upon reasonable request" by the corresponding author of the publication. The same applies to the thermodynamic data as long as it contains no reference to the Python library for thermodynamic simulations. The Python libraries and files that contain information regarding the Python libraries such as the Jupyter notebooks are closed access.

If access is restricted, please specify who will be able to access the data and under what conditions.

Data is made available "upon reasonable request" if the aim is to employ it for research purposes only, no commercial use is allowed. The data which contain information regarding the Python libraries is only shared with the promotor, supervisor, members of the modelling team at Imec and other Imec colleagues involved in the project.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- Yes, Intellectual Property Rights

The Python libraries for automation of DFT calculations and for thermodynamic simulations are owned by Imec and have closed access. Access to the libraries has to be approved by Geoffrey Pourtois and can only be granted to employees within Imec.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Data will be included in the Supplementary Information of publications. The essential raw DFT and thermodynamic data can be made available "upon reasonable request". These data are put on the KU Leuven RDR.

When will the data be made available?

Upon publication of research results.

Which data usage licenses are you going to provide? If none, please explain why.

The license CC-BY-NC-SA-4.0 would apply to the raw DFT and thermodynamic data.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

An identifier is generated automatically by the KU Leuven RDR.

What are the expected costs for data sharing? How will these costs be covered?

No costs are expected.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Bram van der Linden

Who will manage data storage and backup during the research project?

Bram van der Linden, Geoffrey Pourtois and Annelies Delabie

Who will manage data preservation and sharing?

Geoffrey Pourtois and Annelies Delabie

Who will update and implement this DMP?

Bram van der Linden