
Towards graphoepitaxy of semiconducting 2D single crystals on dielectric substrate patterns for next generation nanoelectronic devices

A Data Management Plan created using DMPonline.be

Creators: Iryna Kandybka, Annelies Delabie

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: Iryna Kandybka, Annelies Delabie

Grant number / URL: 1S60023N

ID: 197280

Start date: 01-11-2022

End date: 31-10-2026

Project abstract:

Semiconducting two-dimensional (2D) transition metal dichalcogenides like MoS₂ and WS₂ are attracting great interest for application in ultra-scaled nano-electronic devices because of their monolayer thickness, large band gap, low dielectric constants and lack of dangling bonds. A key prerequisite for their industrial exploitation is the availability of manufacturable deposition techniques like Chemical Vapor Deposition (CVD) to create semiconducting 2D monolayer crystals at wafer-scale. While epitaxial growth on monocrystalline templates ensures high quality 2D crystals, it requires an industrially incompatible transfer process from template to target substrate to enable device fabrication. On the other hand, direct growth on target dielectric substrates results in polycrystalline films with grain boundaries that degrade carrier mobility. The project's ambition is to explore a novel pathway to grow semiconducting 2D single crystals by graphoepitaxy directly on dielectric substrate patterns with topography. We will generate fundamental understanding of crystal growth for 2D WS₂ and MoS₂ during CVD and use that to design the topography of dielectric substrate patterns and choose optimal process conditions for graphoepitaxy. This method for direct growth is compatible with conventional nanoelectronic device fabrication flows and avoids a challenging transfer process. Moreover, this method broadens the application space of 2D semiconductors as it is compatible with 3D substrates.

Last modified: 05-04-2023

Towards graphoepitaxy of semiconducting 2D single crystals on dielectric substrate patterns for next generation nanoelectronic devices

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
		<i>Please choose from the following options:</i> <ul style="list-style-type: none"> Generate new data Reuse existing data 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> Digital Physical 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> Observational Experimental Compiled/aggregated data Simulation data Software Other NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> .por, .xml, .tab, .cvs, .pdf, .txt, .rtf, .dwg, .gml, ... NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <100MB <1GB <100GB <1TB <5TB <10TB <50TB >50TB NA 	
Samples SiO2	MoS2 or WS2 deposited on SiO2	Generate new data	Physical	NA	NA	NA	1000 samples
Samples sapphire	MoS2 or WS2 deposited on sapphire	Generate new data	Physical	NA	NA	NA	1000 samples
AFM data	AFM images of the surface	Generate new data	Digital	Experimental	.spm	<100GB	NA
RBS data	RBS characterization	Generate new data	Digital	Experimental	.docx	<1GB	NA
SEM data	SEM images of the surface	Generate new data	Digital	Experimental	.jpg	<100GB	NA
Raman/PL data	Raman/PL spectra from the samples	Generate new data	Digital	Experimental	.xlsx	<100MB	NA
GIXRD data	GIXRD spectra from the samples	Generate new data	Digital	Experimental	.txt	<100MB	NA
TEM data	TEM images of the surface	Generate new data	Digital	Experimental	.jpg	<1TB	NA
WCA data	WCA images	Generate new data	Digital	Experimental	.bmp	<100GB	NA
Reports	Internal and external summary of the results	Generate new data	Digital	Compiled/aggregated data	.docx	<100MB	NA
Manuscripts	Results to be published	Generate new data	Digital	Compiled/aggregated data	.docx	<100MB	NA
Results presentations	Regular summary on the results	Generate new data	Digital	Compiled/aggregated data	.pptx	<1TB	NA
Origin plots	Processing of the data	Generate new data	Digital	Other	.opju	<100MB	NA
Tool logbook	Log experiments	Generate new data	Digital	Observational	.xlsx	<100MB	NA

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.

- No

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.

- No

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).

Metadata will be provided in order to increase the potential for re-use. Since there is no formally acknowledged metadata standard specific to the PROJECT/CONSORTIUM discipline, an adapted Dublin Core Metadata standard will be used and consists of the following metadata elements:

- *Data ID, title, subject, name, creator, owner, related project ID, project summary, data abstract, keywords, type, size, format, source (e.g. ID of measurement device), language, related publications or patents, compliance, security and confidentiality aspects, rights*

The following metadata elements will be added in case the data relates to the results of experiments:

- *Experimental set-up, study design, sampling methodology, variable-level detail, reference to specific model, and all other information necessary for a secondary analyst to use the data accurately and effectively.*

Tabular data will contain SI units, these international Standard Scientific Units of Measurement will be used to denote values.

A folder and filing hierarchy will be set out and documented. Furthermore, an organisational standard workbook file will be used as temporary "data

catalog" during the project execution phase, where at the end of the project, easy overview of the created data as well as its potential value for reuse can be easily identified. At the end of the project, the data will be reviewed on reusability, and will be added to the institutional data catalog tool or data catalog excel in the main folder.

Folder set-up and naming will be used to deliver and communicate minimum necessary metadata.

- Folders will be used to collect metadata with the data, in case the data files cannot be annotated. In this case, a Readme file will be added, in .xls/xlsx format. Where possible, the metadata will be added directly to the data.
- Folder structure and naming will be set up according to the structure of the PROJECT/CONSORTIUM, as set out in the project description, by WORK PACKAGES->TASKS->EXPERIMENT->DELIVERABLE.

File naming conventions will be used to convey minimum necessary metadata as well.

- ProjectName_SourceInstitution_WorkPackage_DeliverableTitle_DataOfOriginYYYYMMDD_Version
- The file names together with the dates of creation allow quick reference to the labnotebook and project description, where the specific experiment(s) are described with more detailed parameters (time, consignee, protocol, samples names, conditions, ...).
- All report files will include the document version history, containing version, activity, date and responsible, as added to this document above.
- All communications (presentations) and meeting minutes and presentations will be identified by the following in the File title:

- o KOM = Kick Off Meeting
- o PM = Project Meeting
- o TM = Technical Meeting
- o RM = Review Meeting
- o Telco = Online meeting (Skype, GoToMeeting, WebEx, etc.)
- o AH = Ad-hoc meetings, with a specific topic

A codebook with abbreviations of variables and their definitions will be made and uploaded in the central repository for definitions at imec, in case these are not yet present.

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.

- Yes

At least the imec metadata standard will be used (see metadata information in the above answer). The following metadata will be kept at minimum, for which a data content catalog will be kept during the project.

- Data ID, name, creator, owner, related project ID, project summary, data abstract, keywords, type, size, format, source (e.g. ID of measurement device), experimental set-up, related publications or patents, compliance, security and confidentiality aspects.

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

Where will the data be stored?

Imec Data will be stored on the IMEC SharePoint (teams workspaces), or in specific cases on premise (Isilon). Data may be temporarily stored on researchers' computers or OneDrive's while processing the data. Storing sensitive/confidential (project) information on removable devices such as an external hard disk is not allowed by the imec acceptable use policy because it increases the risk of data loss (either through theft/loss of the disk or in case it breaks down).

How will the data be backed up?

During the project: data on SharePoint will benefit from the following security services:

- Automatic back-up
 - This is managed by Microsoft: The imec sharepoint is hosted on the Microsoft Cloud, which is a high available environment. Information on this environment is never automatically deleted, removal of specific information should be managed by the business.
 - If information is removed from this environment, it is moved to a "Recycle Bin". From this Recycle Bin it can still be restored by the user for a period of 93 days (or an admin in case the Recycle Bin was emptied).
 - After these 93 days, items are deleted, and Microsoft will keep a backup for 14 additional days. During this period restoration can still be requested through a Microsoft ticket.
 - After these 14 additional days, the data is permanently removed.

Other data on premises (Isilon) have a automatic back-up system. Data is backed up via snapshot technology, where all incremental changes in respect of the previous version are kept online.

**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

Office-type documents, smaller amounts of data (< 10 GB), targeted at humans?

- Yes -> imec Teams/SharePoint
- No -> Will the data be written or accessed by a machine without human login, or will the data be accessed outside of imec's internal network, or by external people?
 - Yes -> Azure data lake store
 - No -> Isilon (on-prem file servers)
- imec Teams (SharePoint):
 - Only to be used for smaller amounts of files (<10 Gb). Adding more data to a SharePoint is possible, but it is very costly for imec, and SharePoint will start performing badly.
- Azure data lake store: used to store larger amount of data, like "binary" data, "log" data, "measurement" data, ...
 - For larger amounts of files (the capacity is virtually unlimited)
 - Storage exists in many cost-ranges from normally priced cloud storage to very cheap archive storage for data that probably will not be used again.
- Isilon (on-prem file servers): Internal on-prem file shares to store larger amounts of data.
 - For larger amounts of files (but the total space is limited by the hardware that is available at imec, so there is a limit to what is possible)
 - Mainly for data that is actively used, because this storage is too expensive to use as archive storage.

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

During the project: as stated, data on SharePoint will benefit from the following security services:

- User-authentication, multifactor authentication can be activated
 - All logins to imec environments are protected by a strong password (minimal 14 characters) combined with multi factor authentication.
- Versioning system
 - Versioning is enabled by default on our SharePoint/Teams' environment. This can be modified by workspace owner and is under full responsibility by the business user
- System-encryption
 - This is managed by Microsoft as outlined in this article: [Cloud data security measures in SharePoint & OneDrive - SharePoint in Microsoft 365 | Microsoft Docs](#)
 - Making sure that data is shared with the right people is the responsibility by the business user.
 - Confidential data will be user encrypted in addition to this:
 - When creating documents (.docx, .xlsx, .pptx), the user is forced to assign an information classification label to the document (Public / Restricted / Confidential / Strictly Confidential) based on the sensitivity of the information.

While this label provides a (visual) marker on the sensitivity level of the information, it will not encrypt the document by default.

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

Storage is available at imec at no added cost to the project. In general can be calculated with 50€ per month for 1 terabyte of data. However, this will not be being charged internally and is included in overhead expenses.

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...).

All digital data. Samples will be disposed as 2D material is stable only for few months.

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

After the standard retention period of 5 years, data will be subjected to evaluation. This will include weighing the potential value versus the costs of

keeping it available. Decisions will be made by the data owners, in close collaboration with ICT service responsables for archiving.

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

In general can be calculated with 50€ per month for 1terabyte of data. However, this will not be being charged internally and is included in imec overhead expenses.

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.

- Yes, in an Open Access repository

For part of the data that is supporting publication

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

We do not exclude that the proposed work could result in research data with potential for tech transfer and valorization.

o It is expected to obtain data that can potentially be valorised in all workpackages. The data will be kept confidential until 1) a decision will have been made on patenting, and 2) a patent file or publication has been submitted.

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.

- No

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

Zenodo.org will be the preferred standard repository for open data, because this platform, part of the OpenAIRE collaboration, provides the minimum and recommended terms required by DataCite's Metadata Schema, including the assignment of an identifier. These services are furthermore free of charge and will be for the foreseeable future.

Software code will be placed in GitHub.com in an open accessible manner to the public.

Physical data and samples will only be stored until the intended analyses are finished and all information is gathered.

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.

CC-BY: By Attribution, which permits sharing and reuse of the material, for any purpose, as long as the original authors are credited. The material can be freely shared, redistributed, transformed, built upon and adapted for any purpose, including commercial use. Anyone using the material must provide credit to the original authors and indicate clearly any changes that were made, and there is a legal obligation to do this. These terms are widely used for open access articles in academic journals.

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

In case of the open publication of data, a PID or DOI will be provided. E.g. Zenodo (repository) creates a Digital Object Identifier (DOI) for each upload.
In case of restricted data that won't be published open, an accession number will be provided by the imec-specific metadata management framework (data catalog).

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

These costs will be covered in project overhead, except in case of circumstances that can be considered out of the ordinary.

6. Responsibilities

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

The project lead

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

The project lead with support from ICT, unless specified otherwise

Who will manage data preservation and sharing?

After project closure, ownership will be transferred to the applicable head of department at imec, and the owner will decide who will be responsible for operational data preservation and sharing.

Who will update and implement this DMP?

The project lead