

## Variability resilient design of novel in memory compute architectures for machine learning

### 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.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		Indicate: N(ew data) or E(xisting data)	Indicate: D(igital) or P(hysical)	Indicate: Audiovisual Images Sound Numerical Textual Model SOfware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Simulation data	The functioning of integrated circuits is simulated with dedicated software, making use of physical and electrical models that are provided by the chip fabrication companies.	N	D	N	File formats of Cadence Spectre, Ansys HFSS, and Mentor Graphics QuestaSim	<100GB	NA
Measurement data	The functioning of integrated circuits is validated by performing measurements in the IC-lab, using equipment such as Arbitrary Waveform Generators, Vector Network Analysers and Digital Oscilloscopes to generate input signals and measure the corresponding output signals.	N	D	N	txt, csv, or similar	<100GB	NA
Design data	The integrated circuits are designed with specific software such as the Cadence design suite. Depending on the level of abstraction, certain formats are used, such as VHDL (high level circuit description) or GDSII (geometric shapes of the layout).	N	D	N	VHDL, Cadence database format	<100GB	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:

We will not reuse data.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- 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 work in this project will result in new concepts, architectures and designs of integrated electronic circuits and systems. These findings will certainly have potential for valorization. Valorization will most likely be implemented by means of follow-up research projects with industrial partners.

The design data will be protected by trade secrets. Patents are difficult to obtain in this field. The results of this project will form background knowledge in follow-up projects. The typical background licenses will be granted to the partners in such follow-up projects.

We will involve LRD if the need arises.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or 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

## 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).

### *Simulations data*

Raw simulation data will be collected per simulation test, including a text file with a clear description of what the data represent and how they were generated. The input files used for the simulation will be kept inside the same folder. The name of the folder will contain the simulation conditions. A text file explaining the naming will be maintained.

### *Design data*

Details on the conceptual, architectural and topological design of the circuits will be documented in word files. Links to the folders in which the design data are stored will be included, as well as all the necessary metadata to be able to extract and reuse

the design data: technology node, flavour, etc.

#### *Measurement data*

Raw measurement data will be collected per measurement test, including a text file with a clear description of what the data represent and how they were generated. The input-files used for the measurements will be kept inside the same folder. The name of the folder will contain the measurement conditions. A text file explaining the naming will be maintained.

**Will a metadata standard be used to make it easier to find and reuse the data?**

**If so, please specify which metadata standard will be used.**

**If not, please specify which metadata will be created to make the data easier to find and reuse.**

- No

There is no formally acknowledged metadata standard specific to our discipline. However, in our research group, we have a standardized method of structuring our data.

Our researchers are obliged to use this method. This method is available on our intranet and its importance is stressed during the yearly introduction session for new researchers.

#### **Data Storage & Back-up during the Research Project**

**Where will the data be stored?**

- Other (specify below)

We will use the dedicated RAID storage facilities of our research department ESAT.

**How will the data be backed up?**

- Standard back-up provided by KU Leuven ICTS for my storage solution

The data will be stored on our servers with automatic daily back-up procedures.

**Is there currently sufficient storage & backup capacity during the project?**

**If no or insufficient storage or backup capacities are available, explain how this will be taken care of.**

- Yes

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

Confidential data is stored on file servers which are only accessible by authorized people with specific account settings. The servers are located in a secured room with access limited to system administrators. For data related to specific, very advanced and exclusive technologies we have physically separate file servers.

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

The costs for data storage are internally accounted for at departmental level. MICAS carries a proportional part of the departmental IT costs.

## Data Preservation after the end of the Research Project

**Which data will be retained for 10 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 (explain below)

At least the following data will be retained for the expected 10 year period after the end of the project:

- the data needed to reproduce and verify published research results
- the data needed to prove and increase the value of research results that have valorization potential
- all design data

Retaining the data of every single simulation or measurement experiment would take too much physical storage space.

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

- Other (specify below)

We will use the dedicated RAID storage facilities of our research department ESAT.

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

The costs for data storage are internally accounted for at departmental level. MICAS carries a proportional part of the departmental IT costs.

## Data Sharing and Reuse

**Will the data (or part of the data) be made available for reuse after/during the project?**

**Please explain per dataset or data type which data will be made available.**

- Yes, as restricted data (upon approval, or institutional access only)
- Yes, as open data

The simulation, design, and measurement data will be used as a basis on which technology transfer activities can be initiated. Hence, careful IPR management will be needed, meaning that the data will not be shared outside the university without a prior agreement on confidentiality and IPR.

Data related to generic research results can be reused in follow-up projects.

We will publish in international journals, after careful consideration of valorization and patentability potential, during and/or after the project. We will follow the Green Open Access strategy for these scientific publications. In those cases where we do have to publish in journals that are behind a paywall (e.g. IEEE journals that are the top in the field), we will always make a digital copy of the accepted paper available through an online repository. We will ensure that every publication gets a Digital Object Identifier (DOI) and that we use our ORCID on every publication, so that the identification of the record and of the authors is unambiguous. Next to the Lirias document repository system of KU Leuven, we will also use arXiv, which is a free distribution service and open-access archive.

Data related to published results can be made available through KU Leuven's RDR.

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

The simulation, design and measurement data with restricted access will only be accessible within KU Leuven, and specifically only to those persons who have been granted access.

If a third party requests access to those data, this will be most likely in the frame of a collaboration agreement or a licensing agreement, and then it will be necessary to draw up data transfer or data sharing clauses as part of that agreement. In this agreement, the terms of use will be agreed upon.

**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 per dataset or data type where appropriate.**

- Yes, intellectual property rights

For the design, simulation and measurement data that are related to valorizable results, careful IPR management will be needed, meaning that the data will not be shared outside the university without a prior agreement on confidentiality and IPR.

**Where will the data be made available?**

**If already known, please provide a repository per dataset or data type.**

- KU Leuven RDR (Research Data Repository)
- Other data repository (specify below)

The simulation, design and measurement data will be made available for reuse through our internal archiving facilities. Publications will be accessible through the established channels.

The simulation, design and measurement data related to published results will be made available through KU Leuven's 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.**

- Data Transfer Agreement (restricted data)

If a third party requests access to those data, it will be necessary to draw up a data transfer or data sharing arrangement, most likely in the frame of a collaboration agreement. In this agreement, the terms of use will be agreed upon.

**Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.**

- Yes, a PID will be added upon deposit in a data repository

Data uploaded to the data repository will get a DOI (Digital Object Identifier).

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

The costs for data storage are internally accounted for at departmental level. Our research group carries a proportional part of the

departmental IT costs.

### **Responsibilities**

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

Wim Dehaene (supervisor) as end responsible

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

Ben Geeraerts (IT) as support + Wim Dehaene (supervisor) as end responsible

**Who will manage data preservation and sharing?**

Ben Geeraerts (IT) as support + Wim Dehaene (supervisor) as end responsible

**Who will update and implement this DMP?**

David Maes (valorization) as support + Wim Dehaene (supervisor) as end responsible