Plan Overview

A Data Management Plan created using DeiC DMP

Title: Application Specific Integrated Circuit for New Generation of MEMS Acoustic Sensors

Creator: Ivan Jørgensen

Principal Investigator: Ivan Jørgensen, Jakob Kenn Toft

Data Manager: Ivan Jørgensen

Affiliation: Danmarks Tekniske Universitet / Technical University of Denmark

Template: DCC Template

ORCID iD: 0000-0003-2667-6925

Project abstract:

Dette projekt vil søge, i samarbejde mellem DTU og Knowles Electronics, at udvikle et nyt fuldt integreret kredsløb. Knowles har udviklet en ny MEMS akustisk sensor, med en aldrig før set tilgang til MEMS baseret akustisk måling. Sensoren vil tillade et hidtil uset spring i ydelse af MEMS baserede mikrofoner. Til at drive dette nye MEMS modul, skal der udvikles en state-of-the-art fuldt integreret høj-spændings bi-directionel step-up DC-DC converter, som skal opnå en ydelse der er hidtil uset.

ID: 1668

Last modified: 30-06-2023

Grant number / URL: 8053-00044B

Application Specific Integrated Circuit for New Generation of MEMS Acoustic Sensors

Data Collection

What data will you collect or create?

In the integrated circuit research field, primary materials and data is collected by creating new and improved integrated circuits (ICs) and documenting how these perform. The data is therefore the result of observing the behavior of a prototype that was created in the project. In addition to the circuit diagrams describing the integrated circuits, simulation and measurement data is generated in the design and qualification process, along with scripts that facilitate the design of ICs.
Referring to the document "Guidelines and Procedures (GaP) of the retention of primary materials and data at DTU Elektro" (GaP), the primary material and data is defined for both experimental research and computational research. This project is a combination of both experimental research and computational research. In this DMP, the following classification is used. For the later description of the data management for each item, no distinction is made between primary materials and data, only between physical objects and data.
Primary materials

- Integrated circuit prototypes
  - The physical samples
  - Schematics
  - Layouts
- Printed circuit boards for test setups
  - The physical boards
  - Schematics
  - Layouts
- Raw measurement data
- Raw simulation data
  - Temporary simulation data used in the design process
  - Final signoff simulation data for use in scientific publications
- Scripts for capturing and processing measurement data
- Scripts for optimizing and synthesizing integrated circuits

How will the data be collected or created?

- Processed data
- Analyzed data (aggregated statistics, diagrams, and plots)
- Outputs from the optimization and synthesis scripts

The following is a list of the types of primary materials and data that are generated in the project. Under each type of data is listed the possible data formats used, the estimated amount of data, the structure of the data, and the data versioning.

- Schematics and layout drawings of integrated circuits and the corresponding testbenches
  - **Description**: the components and interconnection of these that make up the integrated circuit. This data defines the implementation of the prototypes from which the measurement data is obtained and the input to the circuit simulator that generated the simulation data. The testbenches define the interconnection of the circuit under test and various electrical sources and sinks used for testing the circuit performance.
  - **Formats**: defined by the industry standard tools that are used in the design of integrated circuits. Currently, the we use the Open Access database format defined by the IC manufacturer Open Access Coalition. The final chip layouts are also stored in the GDSII database format that is used by the IC manufacturers.
  - **Data amount**: typically less than 1 GB and 50,000 files for a prototype chip. This project will produce 1-2 chips.
  - **Structure**: Defined by the Open Access database format.
  - **Versioning**: The design data is under revision using git on a GitLab server hosted on the groups Linux system.
- Schematics and layout drawings of printed circuit boards
  - **Description**: the components and interconnection of these that make up the printed circuit boards used for testing the integrated circuits in the lab. The layout is the input for the printed circuit board manufacturer for manufacturing the physical boards.
  - **Formats**: defined by the tools used. Currently, the software Altium Designer is used for designing the printed circuit boards, and the data is stored in a proprietary file format. The layout is also stored in the Gerber format, which is the industry standard used by the manufacturers.
  - **Data amount**: typically, less than 100 MB and less than 100 files per PCB.
  - **Versioning**: The Altium software has integrated support for git, which is used for revision management.
- Simulation data from circuit simulators
  - **Description**: electrical signals (voltages and currents) of the circuit nodes and branches for the different types of circuit simulations carried out. This data is used for evaluating the performance of the designed circuits. Two types of simulation data is generated: temporary design-time data, and final signoff-data.
  - **Formats**: The simulation data format is defined by the circuit simulator used. In this project Cadence Spectre is used, which saves the data in a proprietary database format. The data is both used in the development of the prototypes ICs and as final qualification data that is compared with the measurements. The final qualification simulations are also stored in .csv files and saved along with the papers where they are used.
  - **Data amount**: This depends on simulator settings, the types of simulations performed, and the simulations times required to ensure enough coverage. Typically, between 1-5,000 GB and less than 1 million files. The data is generated in the design process and most is therefore only useful in the instant it is generated as the circuit is typically reconfigured based on the results of the simulations. Only the simulation data from the final circuit is retained as this forms the basis of the paper manuscripts.
  - **Structure**: The data is stored for each testbench that tests parts of the circuit.
  - **Versioning**: The data is not versioned and only the final data is to be saved for later. The data can be recreated by re-simulating the testbenches.
- Measurement data from tests of prototype integrated circuits and printed circuit boards
  - **Description**: measurements of electrical signals (voltages and currents). The data is obtained by using lab instruments to measure signals on the test PCB where the prototype chip is mounted. The data is used in papers to show the performance if the designed integrated circuit
  - **Formats**: Several intermediate file formats are used depending on the equipment used. The final results are saved in numpy arrays (.npy) or compressed numpy arrays (.npz) or as raw (.csv) files.
  - **Data amount**: Typically, 1-500 GB.
  - **Structure**: The data is structured in folders in sub folders under each prototype IC and each experiment carried out for each prototype IC.
  - **Versioning**: The data is not versioned.
- Scripts for synthesizing, optimizing, drawing, and evaluating electrical and electronic circuits at various abstraction levels.
  - **Description**: Part of the research in integrated circuit design deals with developing optimizing methods and synthesis tools to allow for designing better circuits. These methods are implemented as scripts.
  - **Formats**: Python, Matlab, Cadence Skill, Cadence Ocean. The open source interpreted language Python is preferred.
  - **Data amount**: The source code for the scripts is typically less than 10 MB and less than 1000 files. The output of the scripts is saved as numpy arrays (.npy) and the amount varies greatly with the type of script. From 1 kB (for a component sizing script) to 100 GB (for an exhaustive topology generation script).
  - **Structure**: The source code is stored in folders for each project.
  - **Versioning**: The source code is under revision in git hosted on the ELE gitlab server or the DTU gitlab server, or possibly both, but at least one.
- Scripts for performing lab measurements and for analyzing the simulation and measurement data.
  - **Description**: The lab measurements are often automated using scripts.
  - **Formats**: Python.
  - **Data amount**: Typically less than 100 files and less than 10 MB for the source code.
  - **Structure**: The scripts used for performing measurements on a script is located under the folder for each integrated circuit project.
  - **Versioning**: The scripts are under revision with git.

Documentation and Metadata

What documentation and metadata will accompany the data?

There are not metadata standards in the integrated circuit design field of research. Below are outlined the types of metadata that are manually generated per research tradition.

- Schematics and layout drawings of integrated circuits and the corresponding testbenches
  - Metadata in the form of author, date, format, and size is generated by the software tools used (Cadence Virtuoso suite) and stored in the Open Access database format.
- Schematics and layout drawings of printed circuit boards
  - Metadata in the form of author, date, format, and size is generated by the software tools used (Altium Designer) and stored in the Open Access database format.
- Simulation data from circuit simulators
  - Simulation logfile describing simulator configuration is stored along with the simulation results.
- Measurement data from tests of prototype integrated circuits and printed circuit boards
  - Readme files including a description of the lab setup including instrument types and serial numbers.
- Scripts for synthesizing, optimizing, drawing, and evaluating electrical and electronic circuits at various abstraction levels.
  - No metadata
- Scripts for performing lab measurements and for analyzing the simulation and measurement data.
  - No metadata
- Script outputs
  - Appropriate metadata written in readme file.

Ethics and Legal Compliance

How will you manage any ethical issues?

The project does not deal with human beings, human materials, laboratory animals, genetically modified organisms, or any other biological organisms. The use of the designed electronics is in commercial products for powering e.g. laptops, LEDs lamps, or audio amplifiers, and no personal data or ethical issues are involved.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

The design of an integrated circuits is described in the form of schematics and layout drawings. The layout drawings are used to fabricate the IC in a specialized semiconductor fabrication plant. To gain access to the process design kits required for designing physical ICs requires signing non-disclosure agreements (NDAs) as the IC manufacturers are secretive about their processes. These NDAs pose restrictions on what data can be shared publicly as will be covered in detail later. The integrated circuit design databases can e.g. not be shared publicly.

The results of this research project might lead to patentable ideas. This limits what can be shared publicly in the time until the intellectual property has been properly protected. This will be handled case-by-case by involving the IPR department at DTU.

If contracts are made between DTU and an external collaborating party in this project, such a contract might further limit what can be shared externally.

This project is an Industrial PhD project, and the intellectual property rights are therefore governed by the terms of the Industrial PhD agreement, and on the collaboration agreement between the company and DTU.

Storage and Backup

How will the data be stored and backed up during the research?

First, the different roles in the project in relation to data management are listed. Next, the different storage solutions used are listed with a description of backup and access control for each of them. Finally, the storage solutions used for each type of research data is listed.

Roles

- Main researcher: Jakob Kenn Toft
- ELE system administrator: Allan Jørgensen
- ELE group IT security responsible/ELE group leader: Michael Andersen
- Project administrator: Dorte Bettina Svejstrup
- Project participants: Jakob Kenn Toft, Mohammad Shajaan (Company supervisor, Knowles)and Ivan Jørgensen (associate professor, at DTU).

Storage solutions

The local guidelines at the department for provided infrastructure ("Guidelines and Procedures (GaP) of the retention of primary materials and data at DTU Elektro" appendix 6: "ELE specific procedures on storage, IT, documentation, long-term preservation and publication of research data") list the following storage solutions:

- Paper notebooks
- Gitlab
- Sharepoint
- OneNote
- M- and O-drives
- Cadence server

In this project, the "cadence server" (which refers to the home drives described below), gitlab (referring to either ELE's own gitlab server or DTU's common gitlab server) are used for storing data. In the following list, the backup and access control for each type of storage is described.

- ELE linux server home directory
  - **Backup**: Hosted on an AIT NAS server. Snapshots of the home drive is taken four times per day. Backup of the home directory is taken once per day to an AIT NAS server.
  - **Access control**: The main researcher and ELE system administrator has access.
- ELE gitlab server
  - **Backup**: Hosted on a virtual server at AIT. Daily backup to an AIT NAS server.
  - **Access control**: By default, only the main researcher and ELE system administrator have access. The main researcher can grant access to other ELE group members to specific projects or integrated circuit designs.
- DTU gitlab server:
  - **Backup**: handled by AIT.
  - **Access control**: by the default, only the primary researcher has access.

- O-drive:
  - **Backup**: handled by AIT.
  - **Access control**: By default, the main researcher, ELE group leader, and the project administrator have access. The access rights are handled by the ELE group leader.

How will you manage access and security?

- Schematics and layout drawings of integrated circuits and the corresponding testbenches.
  - **Storage**: Stored on the ELE linux home directory belonging to the main researcher and on the ELE gitlab server for sharing with project collaborators.
  - **Sharing within project**: The data can be accessed by other ELE group members via gitlab if the main researcher grants access.
- Schematics and layout drawings of printed circuit boards
  - **Storage**: On DTU network drive (O-drive)
  - **Sharing within project:** Data on the network drive can be accessed by the group leader, the project participants, and the project administrator. Access rights is handled by the ELE group IT responsible.

- Simulation data from circuit simulators
  - **Storage**: Temporary simulations are stored on the ELE linux home drive belonging to the main researcher.
  - **Sharing within project:** The raw simulation data is used in the development project and can only be accessed by the main researcher and system administrator. Final simulation data is transferred to the DTU network drive (O-drive) for broader access.

- Measurement data from tests of prototype integrated circuits and printed circuit boards
  - **Storage**: stored on the O-drive.
  - **Sharing within project:** via general O-drive access rights.

- Scripts for synthesizing, optimizing, drawing, and evaluating electrical and electronic circuits at various abstraction levels.
  - **Storage**: on the O-drive and on DTU gitlab server. For projects shared publicly, the code repository is also shared on an open source hosting service such as github. For code used in publications, the IEEE Supplemental material service is used.
  - **Sharing within project**: via general O-drive access rights.

- Scripts for performing lab measurements and for analyzing the simulation and measurement data
  - Storage:
  - Sharing within project:

Security of the data is ensured by only using services hosted by DTU.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

- Schematics and layout drawings of integrated circuits and the corresponding testbenches
  - Stored at least 5 years on the ELE home drives hosted at AIT.
- Schematics and layout drawings of printed circuit boards
  - Stored at least 5 years on the O-drive.

- Simulation data from circuit simulators
  - Only specific simulation data used in papers are preserved for at least 5 years.

- Measurement data from tests of prototype integrated circuits and printed circuit boards
  - Stored at least 5 years on the O-drive.

- Scripts for synthesizing, optimizing, drawing, and evaluating electrical and electronic circuits at various abstraction levels.
  - Stored at least 5 years on the gitlab server.

- Scripts for performing lab measurements and for analyzing the simulation and measurement data.
  - Stored at least 5 years on the gitlab server.
- Physical integrated circuit prototypes
  - Stored at least 10 years in labeled boxes in ELE storage facilities.
- Physical printed circuit boards.
  - Stored at least 10 years in labeled boxes in ELE storage facilities.

What is the long-term preservation plan for the dataset?

- The current price for storage on the O-drive is 1000 kr./TB/quarter and is currently covered by the department budget.
- Storage on the BIGNAS system costs 250 kr./TB/quarter, and therefore, this solution should be considered for long term storage after the end of the project.

Data Sharing

How will you share the data?

- Schematics and layout drawings of integrated circuits and the corresponding testbenches
  - Shared internally using the ELE gitlab server.
- Schematics and layout drawings of printed circuit boards
  - Shared on the O-drive at the moment. Later, also using a shared server solutions such as Altium Vault.

- Simulation data from circuit simulators
  - Shared internally on the O-drive.

- Measurement data from tests of prototype integrated circuits and printed circuit boards
  - Shared internally on the O-drive.

- Scripts for synthesizing, optimizing, drawing, and evaluating electrical and electronic circuits at various abstraction levels.

- Shared internally on the ELE gitlab server.

- Scripts for performing lab measurements and for analyzing the simulation and measurement data.
  - Shared internally on the ELE gitlab server.

Are any restrictions on data sharing required?

Data own by the company will not be shared by DTU.

Responsibilities and Resources

Who will be responsible for data management?

- Data management planning: Ivan Jørgensen
- IT security: Michael Andersen (IT Security responsible at DTU Elektro)
- Lab management: Hans Christians Andersen
- Documentation: Jakob Kenn Toft
- Michael Andersen (as group leader at ELE).

What resources will you require to deliver your plan?

Existing Sw tools are used. Maintained by ELE System administrator and AIT.