Fast Gated Superconducting Nanowire Camera for Multi-Functional Optical Tomograph

Grant Agreement Nº 101099291

WP 7

D7.3: Data Management Plan

Lead Beneficiary: Politecnico di Milano (CUSBO)

Due date of deliverable: 30.09.2023

Type and dissemination level: DMP, Public

fastMOT is funded by the EU's HORIZON EUROPE programme under grant agreement number 101099291 and by the UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (grant number 10063660).

Funded by the European Union

UK Research and Innovation

Table of Contents

Disclaimer

About fastMOT

Traditionally, the monitoring of organs and deep body functional imaging is done by ultrasound, X-Rays (incl. CT), PET or MRI. These techniques only allow for very limited measurements of functionality, usually combined with exogenous and radioactive agents. In this project we are developing an innovative light sensing solution, a fast gated, ultra-high quantum efficiency single-photon sensor, to enable multi-functional deep body imaging with diffuse optics.

The new type of sensor is based on superconducting nanowire single-photon detectors that have shown to be ultra-fast and highly efficient. However, until now the active area and number of pixels has been limited to micrometres diameter and tens of pixels. We are using a combination of new techniques to overcome this limit and scale to 10,000 pixels and millimetre diameter.

In addition, we are developing new strategies for performing TD-NIRS and TD-SCOS to use this new light sensor optimally with Monte-Carlo simulations. We will implement the new light sensor in an optical tomograph and achieve a 100x improvement of SNR compared to using existing light sensors. With our Multifunctional Optical Tomograph we will be able to image deep organ and optical structures and monitor functions including oxygenation, haemodynamics, perfusion and metabolism.

Executive summary

The first release of the Data Management Plan of the project fastMOT is presented here. The document is drafted following the EU guidelines and public template for DMP, initially foreseen for Horizon 2020 and in force in Horizon Europe. The key issues on description of the dataset to be generated, compliance with the FAIR principles, general aspects related to conservation and protection of data, ethical issues are reported. This document can be updated to reflect new evolution in the DMP during the course of the Project.

Abbreviations

Abbreviation	Definition
DCS	Diffuse Correlation Spectroscopy
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable, and Reusable.
fNIRS	Functional Near InfraRed Spectroscopy
NIRS	Near InfraRed Spectroscopy
SCOS	Speckle Contrast Optical Spectroscopy
SNSPD	Superconducting Nanowire Single Photon Detector
SNR	Signal-to-Noise Ratio

1 Data Summary

The fastMOT project will hardly rely on previously acquired data. Rather, it will generate new data both in the development process and in the validation and first application phase. More specifically, the main categories of data that will be generated and specific datasets are:

- ***Characterisation of component and system and performances***
  - Lasers (pulse temporal width, spectral width, stability, power)
  - SNSPD (temporal response, spectral efficiency, gating specs)
  - Optics (collection efficiency, temporal broadening, speckle collection)
  - System (overall responsivity, temporal response, coherence measurements)
- ***System validation on tissue equivalent phantoms***
  - Tests for NIRS measurements (e.g. MEDPHOT [1], NEUROPT [2] protocols)
  - Tests for DCS/SCOS measurements (e.g. autocorrelation functions, measurement of viscosity, SNR assessment)
- ***In vivo applications on healthy volunteers***
  - Signal characteristics in vivo (signal level, basic optical properties, basal relative blood flow level)
  - Simple functional protocols (e.g. cuff occlusion, Valsalva manovre, cognitive or motor tasks)
- ***Simulated data***
  - Simulated measurements based on a theoretical modelling of the tissue response based either on Monte Carlo code or Diffusion Approximation of the Radiative Transport Equation including the effect of the measurement system.

The formats of the data generated during the Project will be:

- ***NIRS***: a custom binary file format adopted since decades at CUSBO labs, with tools for easy interpretation and load in different environments (e.g. Python, C++, …).
- ***DCS***: binary files with individual photon time-tagging encoding photon arrival time both at the microtime (ps – time-of-flight) and macrotime level (µs – correlation time).
- ***SCOS***: the format is still open and will be defined later on, depending on the specs of the new sensor, which will be available in the second half of the project.
- ***Component and system performance specifications***: text files.

- ***Additional sensors*** (e.g. monitoring of physiological parameters such as pulsatility, arterial saturation, respiration): the standard format provided by the vendor – usually text files.
- ***Simulations***: code (MatLab scripts, C++) and data (i)simulation images .jpeg or .tiff; (ii)distribution of time of flight of photons .txt ASCII; (iii)3D spatial distributions voxelised files .vox binary files;
- ***Metadata***: all measurements will be complemented with metadata, in the form of tables (tabulated text), verbose description (text), or structured description (json).

The size of the data is highly variable. By far, the largest datasets are those related to the DCS measurements where individual photon must be recorded. In this case, for a total count-rate of 1Mcounts/s, dataflows in the order of 10MB/s can be generated, leading to multi GB data for each measurement, easily reaching the TB level for whole measurement sessions. The issues of data conservation, together with the principle of limitation of long-term huge data storage will be considered. Either reduction of the raw datasets to synthetic forms (e.g. autocorrelation function) or cleansing of temporary files will be pursued.

These data will be of use to the very same partner of the consortium also after the conclusion of fastMOT for further development on the components and system, for alternative analysis of the acquired data, or to devise new approaches. Furthermore, also external users from the scientific community could benefit from the available dataset to simulate the advantages in using the SNSPD technology in various application areas, but also to exploit the combined use of NIRS and DCS/SCOS data to track in vivo functions, as well as to develop novel methods for the analysis of photon migration in biological tissues.

Within fastMOT project, TUD and SQ will collaborate to deliver arrays of superconducting nanowire single-photon detectors. The array size will be progressively enhanced in 3 steps:

1 - A SNSPD array of 6x6 SNSPDs: this array size is what we have sufficient experience with, and we can use as a safe choice to benchmark our readout schemes and optimize our electronics.

2 - A SNSPD array of 28x28 SNSPDs: this is an intermediate step which would help us to fully benchmark our readout scheme and address the scaling issues. Additionally, this array size is compatible with some of the standard image databases which would make the benchmarking easier and more objective.

3 - A SNSPD array of 100x100 SNSPDs: At this stage we expect the scaling issues for readout and sensor pixels are addresses so we could deliver the originally planned sensor.

In addition, the planned sensors have the possibility to be gate-able within 1ns which puts a limit to the size of SNSPD pixels. Our initial simulations and calculations indicate that a 5x5 μm^2 pixel should give the required speed. However, fixing the pixel size for the whole sensor array inevitably means different array fill-factor which is SNSPD active area to non-active/readout area ratio (this should not be confused with pixel fill factor which is the ratio of nanowire to the spacing between nanowires).

In addition, to design the arrays several simulation tools must be developed, and considerations and trade-offs must be taken into account. At TUD and SQ we are working on a stimulation code that couples the electrical and the thermal behaviour of the SNSPD and the readout mechanism. This code will allow the optimization of the different parameters for the readout circuit and the pixel size to achieve the expected performance. Naturally, there is a trade-off between the number of lines one must use to readout the complete chip and the complexity of the multiplexing mechanism used. We will use electro-thermal simulations to quantify this and make an informed decision for the final design.

2 FAIR Data

2.1 Making data findable, including provisions for metadata

The growing field of Diffuse Optics for clinical diagnostics does not have already a consolidated standardization of data and metadata. Some attempts have been tried in the Society for Functional Near Infrared Spectroscopy, but the proposed data format is not yet widely adopted and also will unnecessarily complicate the datastructure of fastMOT, still covering only the NIRS component. Yet, whenever possible, we will consider the recommendation of the fNIRS community also for what concerns definitions, tests and nomenclature. Also, we will make reference to existing International protocols for performance assessment of Diffuse Optics instruments, in particular, the BIP [3], MEDPHOT [1], NEUROPT [2] protocols.

Given the lack of standard formats for data and metadata, case specific metadata will be produced and linked to raw data. These will mostly specify:

- information on the specific configuration of the system in use (e.g. laser type and settings, optical networks, versioning on the detector);
- information on the sample under study (e.g. material, nominal optical properties);
- description of the protocol for in vivo measurements;
- link to the international protocol for phantom test and to a specific figure-of-merit.

The format of the Metadata will be adapted to make findability more easy, in particular by the following actions:

- use of summary tables and text files with a unique identifier linked to the raw data, but not tied to the huge raw data to ease harvesting;
- use of structured description of the setup, samples, and procedures with a json format to ease comparability of different measurement sessions and indexing of metadata.

A unique identifier is not foreseen ab initio on all datasets. Rather, it will be attributed to those datasets worth of long-term preservation as well as open availability, as specified in the following section.

2.2 Making data accessible

Only a fraction of the generated data will be made openly available, assigned a DOI, and preserved for long-term storage. In particular, the following datasets will be excluded from open access:

- Proprietary or developmental data generated during the design of the new components and systems to preserve know-how and patentability. In particular, this relates to the Company in the consortium which could get competitive disadvantages in openly sharing procedures, datasets, specifications both unveiling new concepts or internal methodologies.
- Data pertaining to exploratory or intermediate measurements sessions aiming at a first acquaintance with the new technology and identification of best testing conditions. Since data production can be huge (up to TB for measurement session), long preservation and curing of these temporary datasets not advantageous. Also, it will increase the amount of data rubbish contradicting the principles of DNSH.

In conclusion, the datasets suitable for long-term preservation and potentially open access are:

- Data related to systematic experiments, most often related to full journal publications. Also, a superset of presented data can be included to stage the whole experimental collection.

At the present stage, these datasets complying with the above requirements can be envisaged:

- Characterisation of the first laboratory release of the instrument.
- Complete phantom validation of the final system.
- In vivo measurements with the final system.

Consolidated open datasets will be deployed on a certified repository complying with EU requirements. As of now, Zenodo is the best option to host fastMOT open datasets. This choice could be revisited in the near future in case other preferable solutions are identified through the EOSC initiative.

2.3 Making data interoperable

As discussed in section 2.1, there is a lack of standards on common ontologies or vocabularies in the Diffuse Optics community. Whenever possible, data will be complemented with documents (e.g. scientific papers) explaining the meaning of specific measurements or protocols. The Metadata (in particular the structured ones in json format) will present the samples and protocols in simple terms, easily referrable to common understanding or basic scientific knowledge.

2.4 Increase data re-use

During the whole project, care will be taken in assuring best practices in laboratory measurements and definition of measurements. As already stated, for system validation on NIRS we will adopt the Internationally Agreed Protocols of BIP [3], MEDPHOT [1], NEUROPT [2]. For DCS/SCOS, initial attempts of interlaboratory comparison on agreed tests and figure-of-merits will be followed.

Data will be duly documented through metadata, in particular, in the Tabular section with indication of unit of measurements and proper naming of the variables. Reference to specific quality assessment and protocol definition will be addressed either in the structured json description of the experiment of in an accompanying collection of related documents.

3 Other research outputs

The fastMOT project will also produce software for:

- simulations of NIRS and DCS/SCOS signals;
- data analysis.

This software will be initially developed on a private GitHub repository. Then, disclosure and open access of more consolidated or useful tools will be considered. A balance between extra-effort to make the software user-friendly and effective interest for external users will be made. Overall, ICFO's GitHub server will be utilized for these purposes.

4 Allocation of resources

Costs for initial curing, maintenance and storing of data will be covered by standard procedures at local institutions (e.g. datawarehouses and access procedures in place at Politecnico di Milano hosting the CUSBO Facility). Long term preservation and open access will be granted by the service offered free of charge by external repositories (e.g. Zenodo).

5 Data security

The process for data generation of and storage is the following:

- data are saved locally on PC serving the workstation for best operational efficiency;
- consolidated datasets are stored at CUSBO on the datawarehouse of Politecnico di Milano (stored in a NAS disk space in a university server farm and protected by redundancy systems -RAID-, with daily safety replication on a second server farm; both server farms are located at the Polytechnic University of Milan (Leonardo and Bovisa) in special restricted access areas; access to the data is allowed only to certain people and protected by personal credentials; the resource can be reached via the VLAN of the Physics Department -within the institutional network of the Polytechnic of Milan-, protected by a firewall, and access from outside is only permitted via VPN);
- curated data and metadata worth for long-term preservation and open access (Section 2.2) will be deployed on a certified repository (Zenodo).

The process for software development is the following:

- The software code for simulations is stored to local computers (partners UCL, ICFO, CUSBO) and runs on a remote computer\cluster with several GPUs (UCL/ICFO).
- Users can connect to the remote computer inside MATLAB and/or Python from their local computer, securely via Internet (from anywhere) or LAN connection (when within the same institute as the remote computer). The code can also run using local computers (UCL, ICFO, CUSBO) as long as the local computers have GPUs.Simulation outputs will be stored locally and those outputs will can be later shared via repositories such as Zenodo.

6 Ethics

Metadata pertaining to in vivo measurements will contain generic demographic information (e.g. sex, body mass index, age). These data will be pseudoanonymized by assigning a unique identifier whose correspondence with the subject is not digitalised or accessible by digital means. This procedure was approved by the Ethical Committee which approved the study foreseen in fastMOT. The informed consent for data sharing and long term preservation is included in the form submitted to the volunteers.

7 Other issues

This DMP will be modified as needed also in agreement to new requirements or suggestions issued by:

- national and EU legislation;
- indications produced at EU level, in particular in the framework of Open Science and EOSC;
- best practices issued at the level of the LaserLab Europe network of laser infrastructures;
- local requirement set forth by Politecnico di Milano.

In the present stage, this field is quite dynamic with potentially new standards or practices in preparation. Therefore we will be flexible in adopting best practices as needed.

8 References

1. Pifferi, A., et al. (2005). *Performance assessment of photon migration instruments: the MEDPHOT protocol.* Applied Optics 44(11).
2. Wabnitz, H., et al. (2014). *Performance assessment of time-domain optical brain imagers, part 2: nEUROPt protocol.* Journal of Biomedical Optics 19(8).
3. Wabnitz, H., et al. (2014). *Performance assessment of time-domain optical brain imagers, part 1: basic instrumental performance protocol.* Journal of Biomedical Optics 19(8).