2020 HGF – OCPC – Programme

for the involvement of postdocs in bilateral collaboration projects

Title of the project:

Finding Anomalies in the Heart Rhythm of the World-Largest Linear Accelerator:
Data Analysis for the Optical Synchronization System at the European XFEL

Helmholtz Centre, division/group:

DESY, MSK

Project leader:

Annika Eichler

Contact Information of Project Supervisor: (Email, telephone)

annika.eichler@desy.de, +49 (0)40 8998 4041

Web-address:

https://msk.desy.de/

Department/Group: (at the Helmholtz centre or Institute)

Machine beam controls

Programme Coordinator (Email, telephone and telefax)

Dr. Frank Lehner
DESY Head of Directorates Office
Phone: +49 40 8998 3612
Email: frank.lehner@desy.de

Description of the project (max. 1 page)**:**

Since the invention of particle accelerators, nearly 30% of all Nobel Prizes in Physics, on average one Prize every 3 years, had a direct contribution from accelerators. The European XFEL, at 3.4-km-long the largest currently operated linear particle accelerator in the world, is capable of breaking several world records concerning the repetition rate of up to 27,000 light flashes per second, as well as the quality and intensity of the produced X-ray photon pulses. With this, cutting-edge research opportunities in molecular and material science and system biology are opened on atomic scale. Those precise measurements require timing with an error margin in the femto-second range for most subsystems within the facility. In order to provide this high precision timing, an optical synchronisation system is installed at the European XFEL to stabilize critical RF stations and the experimental lasers in time, resulting in a relative jitter of approximately 30 fs rms.

To further improve the performance but especially the availability of the optical synchronization system, currently a data acquisition (DAQ) system for the optical synchronization system is built. Up to now, the data corresponding to the optical synchronization system was only accessible online without any possibility to access histories. With this no root cause analysis of occurring failures was possible nor failure statistics. To enable this, the DAQ system for the optical synchronization system is currently set up with a capacity of 100TB yearly for long-term storage. This capacity limitation allows for 3.17 MB/s of stored data in average, which is only a percentage of the produced data. The main laser for example, the master laser oscillator, has 1000 thousand attributes. While the majority of attributes are slowly varying up to 10Hz, there are some percentages of fast attributes (328 kHz); only two of these would fill the storage capacity alone. Therefore, data reduction is a key question for long-term archiving.

At the start of this project, the newly built DAQ system will have been running for almost a year. This is a great basis for data analysis on the optical synchronization system facing towards anomaly detection and a predictive maintenance system. The first step will be a detailed analysis of the existing data. This includes the review of the data pre-processing. Pre-processing consists of data aggregation, cleaning, resampling, dimensionality reduction, normalization and transformation. Resampling is unavoidable here, since different attributes are sampled with different frequency. As already mentioned, due to the high dimensionality, dimensionality reduction is of major importance. Here classical approaches like principal component analysis will be reviewed, and more advanced methods like feature extraction by autoencoders will need to be analysed.

As one exemplarily field of data analysis, anomaly detection methods to detect failures in the system are to be applied and further developed. Here, the challenges lie in the heterogeneity of the attributes, e.g. different sampling rates, data types and physical meaning. The majority of attributes will be analysed as time series but there are also some in frequency domain. Due to the large number of attributes (approximately 80.000) multivariate tools are to be applied. Anomalies of interests will be contextual or collective outliers.

The algorithms will be trained and developed on the historical data that will have been recorded at this stage. For this, DESY provides computing power on the HPC clusters. Based on those results an online health monitoring and predictive maintenance system is to be developed. The project is finalised by the online validation of the developed algorithm.

The optical synchronization system is used here as a demonstrative example, while data analysis for accelerator physics in general will be of main importance for the whole accelerator operation in the future. The project will be a great step towards accelerator operation with new tools based on data analysis. The advantage of this project is that the optical synchronization system is an enclosed subcomponent, where consistent data access will be provided thanks to the current activities.

Description of existing or sought Chinese collaboration partner institute (max. half page)**:**

There is already a strong collaboration with the Institute of High Energy Physics, Chinese Academy of Science. The existing collaboration is currently focused on machine learning (ML) for HEPS and on MicroTCA implementation for the timing system. For ML, a Chinese student is coming to DESY this year, who has worked on ML supported lattice design, and also for the MicroTCA project there is already a successful exchange of students.

Based on this, the goal is to extend this collaboration to the topic of data analysis for accelerator physics with the optical synchronization system as an example. Here, the goal is to build on the existing successful collaboration.

Required qualification of the post-doc:

- PhD in data science, mathematics, machine learning or similar disciplines.
- Experience with python is required.
- Additional skills in accelerator physics are welcome but not required. Interest in this application domain is presumed.
- Language requirement: English