First Steps towards data-based IPC for the EuXFEL optical synchronization System

Intelligent Process Control Seminar Speedtalk

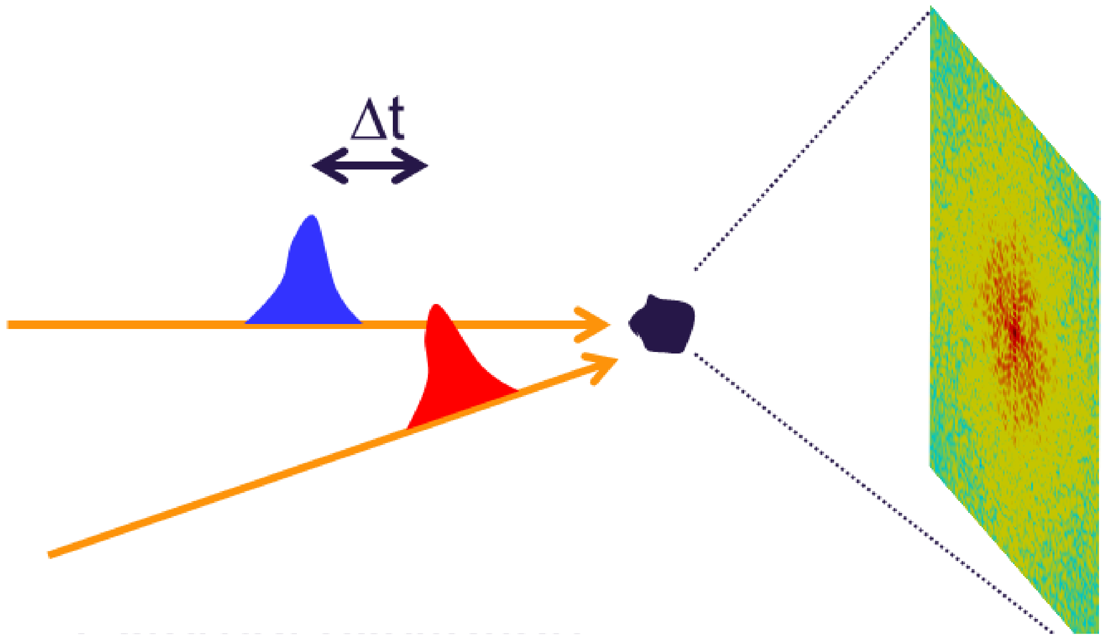
Maximilian Schütte on behalf of LbSync Team & MCS Team
LbSync – MSK - DESY
Hamburg, 17.11.2020

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

Pump-Probe Experimentation

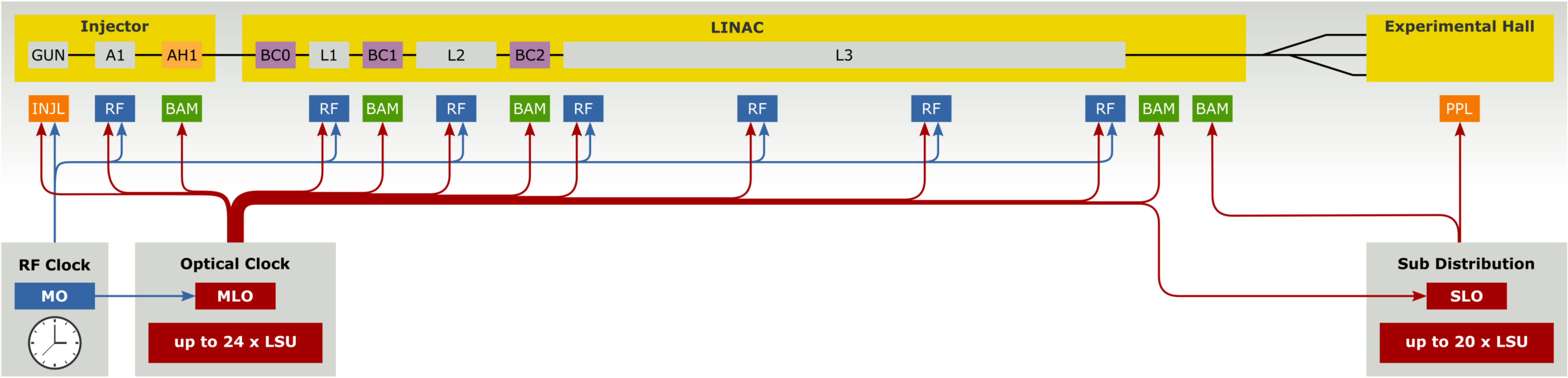
Principle of Operation

- An experimental sample is **activated** through a **pump** laser pulse.
- The process is **imaged** with a **probe** laser pulse at a specified time Δt relative to the activation.
- Repeating the process for a range of Δt, a movie of the process is created.
- The experiment may be repeated for identical Δt to increase brightness through image averaging.
- Accuracy of Δt ensures
  - steady movie “speed” (framerate)
  - sharpness after averaging.



Introduction to the Optical Synchronization System

Ultra Stable Timing for Ultra Sensitive Experiments

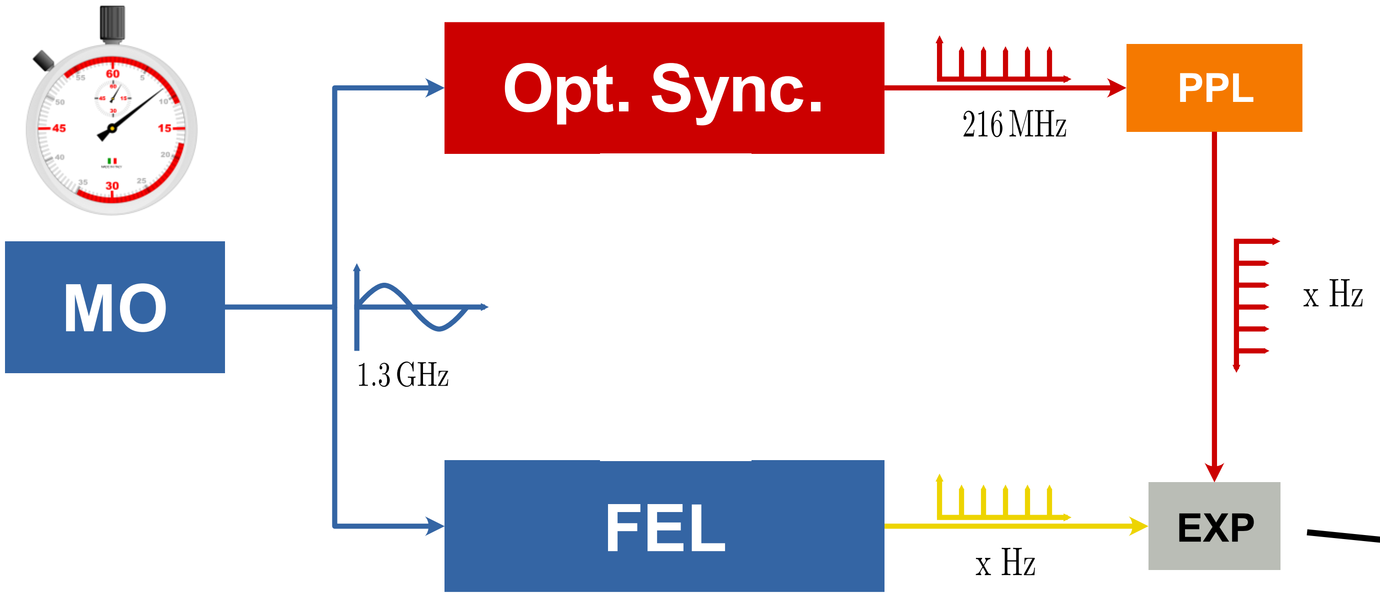


- The Master Laser Oscillator (MLO) acts as **pulsed optical clock** and is phase locked to the RF Master Oscillator (MO).
- Optical timing is distributed via **actively length-stabilized fiber links** and used for
  - Laser locking of injector and experiment pump-probe lasers through optical cross-correlation (**Laser-2-Laser**).
  - RF timing correction / stabilization at accelerator modules through Mach-Zehnder-Modulator (**Laser-2-RF / REFM-OPT**)
  - Bunch arrival time diagnostics (**BAM**)

Slide at courtesy of LbSync Team

Arrival Time Synchronization Problem

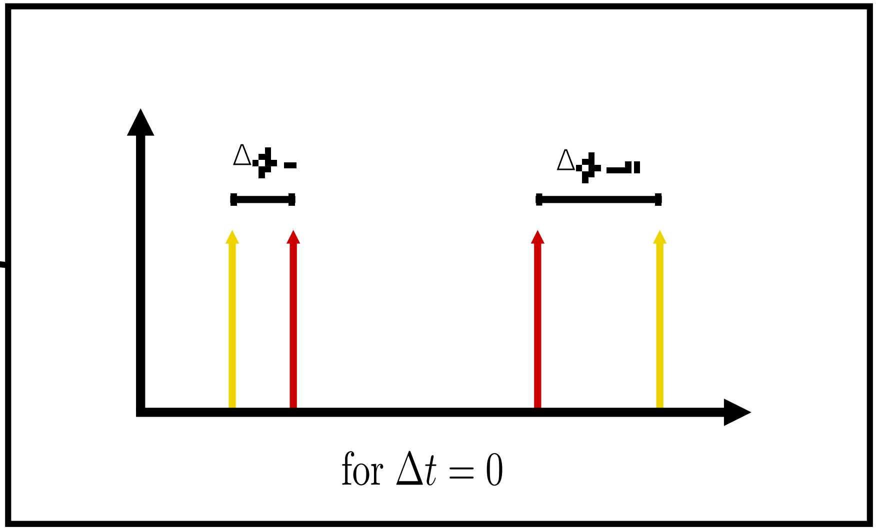
For Pump-Probe Experiments at FLASH and European XFEL



Objective is to minimize deviation $\Delta\phi$ from desired pulse arrival time difference Δt

$$\Delta t^* = \Delta t + \Delta\phi$$

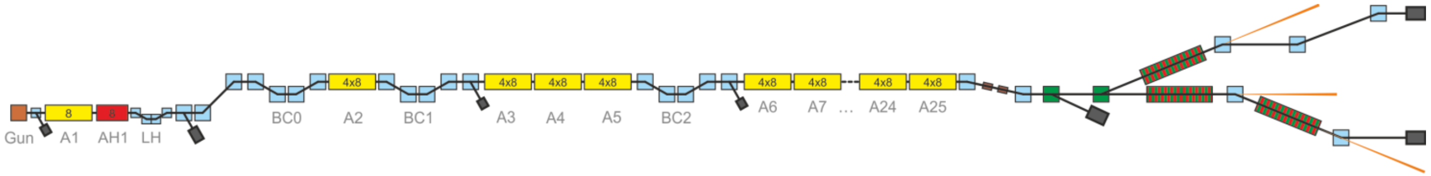
$$\min \lim_{k\to\infty} \frac{1}{k} \sum_{i=n}^{n+k} \Delta\phi_i \iff \min \lim_{T\to\infty} \frac{1}{T} \int_t^{t+T} \Delta\phi(t)\, \mathrm{d}t$$



Data Acquisition for Linear Accelerators

Current Situation at DESY / EuXFEL

- Large-scale accelerators provide huge amounts of data
  - ➢ And it's getting more
- > 10 Million data addresses in DOOCS control system for EuXFEL
  - ➢ Configuration, measurements, extracted features..
- > 20.000 high data-rate channels
  - ➢ Not feasible to completely transfer via network
- 30 TB/day of data recorded for EuXFEL in short term archive
  - ➢ In many cases not evaluated before deletion
- < 1 % of available data goes to central DAQ
  - ➢ Often channels required for certain conclusion are missing

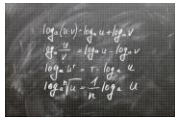


Motivation & Benefits

Why is data so popular?

- Data has become a new fundamental resource in science and the economy.

- Data based analysis goes beyond the complexity limitations of first-principles.



- Especially well suited for detecting anomalies and faults, which usually are not modelled.

- Can also skip analysis and go directly to data-driven algorithms („Machine Learning“).

- Powerful machines necessary, now available.

Why Optical Synchronization System?

- New technology, a lot to understand and learn
  - Long-term system health & aging
  - Fault sources
  - Curious effects & Crosstalk (Tidal influences, …)
- Good accessibility, even during operation
- Clear system boundaries
- Operates (more or less) independently from accelerator
- Major data source

fs LbSync

The Optical Synchronization DAQ

Definition & Purpose

„One Data Acquisition System to capture it all.“

Abstract

- In-house developed DAQ system
- Collecting data from system core, client and connected systems
- Permanent operation
- 100 TB / year quota

Goals

- Reconstruct the entire subsystem state
- All data in one place
- Multiple years into the past
- Show benefit of data analysis and algorithms on subsystem level

Evolution of Data Acquisition and DAQ Systems

Scope	Unsystematic data collection	Dedicated DAQ Software + Hardware for components and experiments	Subsystem (LbSync, LLRF) Acquisition	System (Accelerator) Acquisition
Time-Scale	Event Timescale	Days / Weeks Monitoring	Years	Project Lifetime
Resolution Volume	Couple of CDs	Tapes & HDDs	Data Servers	?

This Project

Core Questions for DAQ Design

New requirements need new solutions

How much and at which rate is data produced?	• Are the front-end CPUs up to the task? Protocols suited?
How much data can be transferred?	• Front-End upstream most likely bottleneck.
How much data can be stored?	• High reduction factor necessary!
How is it made accessible to users?	• Clearly documented API required!
How is data integrity ensured?	• Continuous validation desirable, much can be lost!
How is data identified?	• Labeling is inevitable, meaning of data needs to be clear!
Which data is NOT produced?	• Metadata problem – which data is required by algorithm?

The Optical Synchronization DAQ

Quick Facts

Quantitative

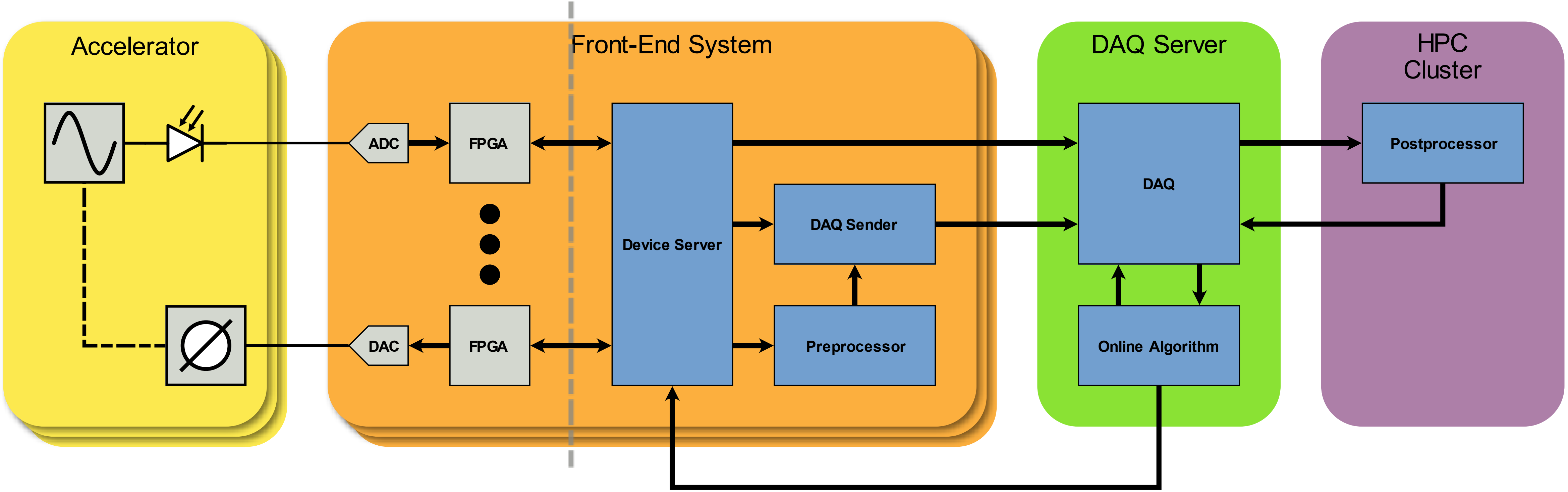
- 100 TB / year long term storage.
- Short ring-buffer with trip snapshot feature for excess data.
- > 280 MB / s input data rate (final).
- Front-end upstream up to or exceeding 1 Gb/s.
- Approx. 80’000 data channels (final count).
- ZeroMQ + Mutlicast (UDP) protocols.
- Multiple thousand distinct data channels manually labeled.

Qualitative

- Developed new DAQ configuration management and configuration data labelling workflow.
- Successfully commissioned for core systems and long term data channels.
- Final base commissioning around Winter 2020.
- Ongoing research on Metadata collection and storage.
- Ongoing research on data preprocessing and reduction.
- Data analysis already started.

Data Acquisition

Data Path Overview



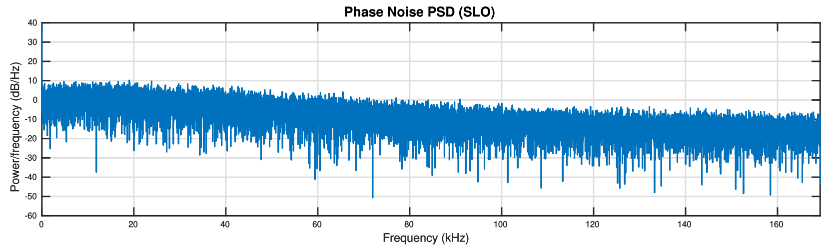
Ongoing Projects on LbSync Data

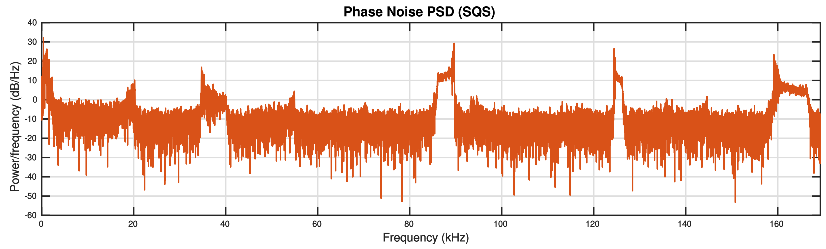
Bachelor Thesis D. Kraft "Link Monitoring"

- Initial experimentation with extracting & processing data from DAQ.
- Slow (< 10 Hz) data analysis for stabilized transmission fibers (days / weeks).
- Look for unexpected correlations & outliers.
- First steps towards slow-data long term health monitoring.

Helmholtz AI Voucher "Laser Health"

- Learn model for laser health based on fast (>300kHz) data.
- Cover wide range of operating conditions.
- Detect when laser leaves "healthy" regime.
- Automated counter-measures to maintain operation.

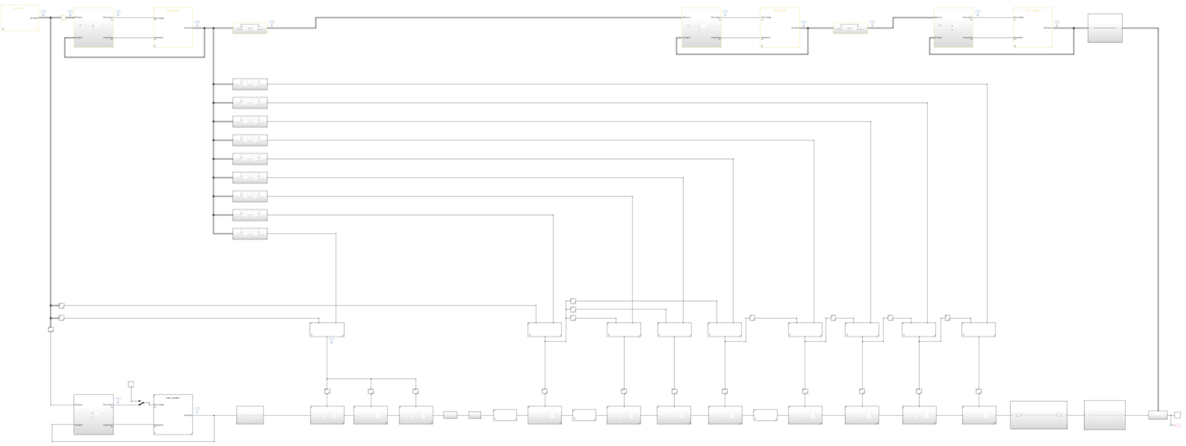




Ongoing Projects on LbSync Data

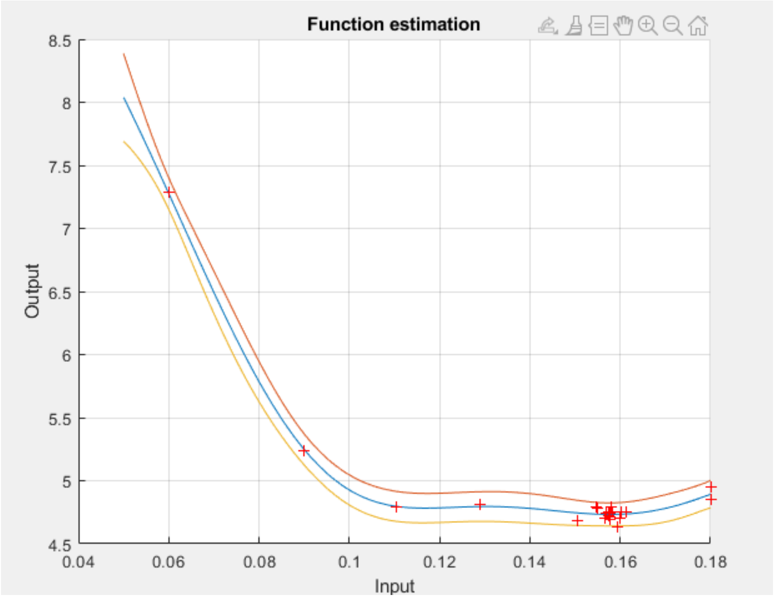
XFEL Arrival Time Jitter Model

- Model estimates arrival time jitter that is not observed continuously (dedicated timing tools not always available).
- Used to develop and test algorithms that are not safe enough for the real machine.
- Better understanding of arrival-time jitter propagation through machine.
- Data validates model.



Bayesian Controller Parameter Optimization

- Inexpensive tuning method for unknown cost / performance function.
- Can be designed with safety constraints.
- Goal: 10 dimensional input space optimization in reasonable time.



Future Projects on LbSync Data

Health-aware PLL Controllers

- High-performance controllers that continuously adapt to operating conditions and system health.
- Automatic synthesis from DAQ data.

Predictive Maintenance

- Join knowledge & models from current projects to predictive maintenance application, warning experts & operators of imminent synchronization system failure
- Optionally taking automated counter-measures.
- Extend to more subcomponents of system.
- Middle-layer service on DAQ data.

Thank you!

Contact

DESY. Deutsches
Elektronen-Synchrotron

www.desy.de

Maximilian Schütte
MSK / LbSync
maximilian.schuette@desy.de
+49 40 8998 1811