A First-level Event Selector for the CBM Experiment at FAIR

J de Cuveland and V Lindenstruth for the CBM Collaboration
Frankfurt University, FIAS Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, DE
E-mail: cuveland@compeng.uni-frankfurt.de

Abstract. The CBM experiment at the upcoming FAIR accelerator aims to create highest baryon densities in nucleus-nucleus collisions and to explore the properties of super-dense nuclear matter. Event rates of 10 MHz are needed for high-statistics measurements of rare probes, while event selection requires complex global triggers like secondary vertex search. To meet these demands, the CBM experiment uses self-triggered detector front-ends and a data push readout architecture.

The First-level Event Selector (FLES) is the central physics selection system in CBM. It receives all hits and performs online event selection on the 1 TByte/s input data stream. The event selection process requires high-throughput event building and full event reconstruction using fast, vectorized track reconstruction algorithms. The current FLES architecture foresees a scalable high-performance computer. To achieve the high throughput and computation efficiency, all available computing devices will have to be used, in particular FPGAs at the first stages of the system and heterogeneous many-core architectures such as GPUs for efficient track reconstruction. A high-throughput network infrastructure and flow control in the system are other key aspects. In this paper, we present the foreseen architecture of the First-level Event Selector.

1. Introduction

1.1. The CBM Experiment
The goal of the Compressed Baryonic Matter (CBM) experiment at FAIR is to explore the QCD phase diagram in the region of high baryon densities. The most promising observables are rare probes that carry information on the matter properties, such as particles containing charm quarks (D-mesons and charmonia), low-mass vector mesons decaying into dilepton pairs ($\rho$, $\omega$ and $\phi$ mesons), or $\Omega$ hyperons (consisting of 3 strange quarks). The CBM research program requires the systematic and comprehensive investigation of these observables [1].

The experimental task is to identify hadrons and leptons in collisions with up to 1000 charged particles at event rates of up to 10 MHz. These measurements require fast self-triggered readout electronics, a high-speed data acquisition architecture, and an appropriate high-level event selection concept.
1.2. CBM Overview

The CBM experimental setup consists of a set of detectors optimized for measurement of different properties of the generated particles, which are combined to detect a specific observable. One of the possible setups is shown in Figure 1. The Silicon Tracking System (STS) is the main tracking detector of the experiment. With its multiple layers and high channel density, it generates a large fraction of the overall data output. The Ring-Imaging Cherenkov (RICH) Detector and the Transition Radiation Detector (TRD) also contribute major amounts of data.

One of the experimental challenges in heavy-ion experiments is the large number of charged particle tracks generated in each collision event. In CBM, we expect about 700 charged tracks in the aperture of the experiment. Another challenge in CBM are the extremely high reaction rates of up to \(10^7\) MHz, which are required for sufficient statistics of rare probes. In combination with the large number of tracks, this demands not only for fast and radiation hard detectors, but also for a high-speed data acquisition system and efficient online event selection. As many relevant online event selection conditions require complex computational tasks such as the identification of leptons and hadrons or high-precision vertex reconstruction, CBM takes a novel approach to data acquisition and online event selection.

2. The CBM Readout Architecture

As many event selection criteria in CBM require complex analyses, there is no fast online trigger signal [2]. This is in contrast to existing heavy-ion experiments, such as ALICE, where a Level-1 trigger based on limited information from specific trigger detectors decides on readout of the front-end buffers. Instead, all front-end electronics are self-triggered. All hits are detected autonomously and forwarded, together with a time stamp, via the data acquisition chain for analysis in a high-level processing farm. As no trigger is available, the association of hits with events is done later using time correlation. The generic readout chain of the CBM experiment is presented in Figure 2.

Self-triggering Front-end Electronics. The triggerless system design has important consequences. CBM’s self-triggering front-end produces a high data rate of approximately 1 TByte/s. Data is transmitted off the detectors in a data-push architecture, which requires fast readout links and large buffer sizes to accommodate rate fluctuations and the online event selection (Level-1)
decision time. As even with today’s technology many GBytes of buffer memory are possible, the
allowable L1 decision latency can be up to many milliseconds or even seconds.

Data Synchronization  As the front-end generated time stamps in lieu of a global trigger signal
are required to associate the acquired data to an event, all front-end electronics have to be
synchronized. In CBM, the T-Net [3] supplies precise timing information without immediate
synchronization to the accelerator and also implements a controlled throttling mechanism to
cope with possible situations of congestion.

Data is aggregated and forwarded in a hierarchical readout structure. The components of the
readout tree buffer and merge data, ensuring correct time ordering of the hits at all stages. For
some detectors, they may also perform initial feature extraction.

High-level Processing  The name of the online processor farm, First-level Event Selector (FLES),
implies that there is no low-level event filtering performed before this stage. Data transmitted
to FLES will be concise; we expect the CBM raw data and event summary data (ESD) to be of
similar size. Therefore, local feature extraction before event building is of limited use. Instead,
the FLES must be able to perform interval/event building on the full input rate of 1 TByte/s.
After this, the online event analysis includes clusterization and feature extraction, tracking in
four dimensions (including time) and finally event selection.

Assuming an archiving rate of 1 GByte/s and an average event size of at least 40 kByte for
minimum bias Au+Au collisions, a maximum event rate of 25 kHz can be accepted for permanent
storage. At a primary event rate of 10 MHz, this requires the FLES event selection algorithms
to reject background events at a factor of at least 400.

3. Architecture of the First-level Event Selector

3.1. FLES Overview

The First-level Event Selector will be implemented as a scalable high-performance computer. All
available types of computing devices will have to be used to achieve the required throughput and
computation efficiency. This includes in particular FPGAs at the first stages of the system and
heterogeneous many-core architectures such as GPUs to enable efficient track reconstruction.
The system will contain approximately 1000 input nodes with fast PCIe interfaces receiving
data from the detector at a total rate of 1 TByte/s. Online event analysis will be performed by
processing nodes estimated to require in total approximately 60 000 cores (year 2010 equivalent).
A high-throughput network infrastructure will connect all nodes and enable interval building.
at the full input data rate. The architecture of the First-level Event Selector is summarized in Figure 3. In total, there are 1000 links at 10 GBit/s transmitting data into FLES.

3.2. Input Interface to FLES

The interface between the custom data acquisition network and the FLES PC farm is implemented as an add-on card to the host PC (see Figure 4). This add-on card has to feature high-bandwidth links both from the readout electronics and to the host PC. In addition, a large and high-bandwidth buffer memory is desirable. The current baseline assumption is to use a custom FPGA-based add-on card with a PCIe interface at the maximum achievable PCIe data rate.

The FPGA design implements the data acquisition protocol for receiving data, merges input links, performs memory management and controls the PCIe transfer. To allow for efficient time-indexed data access, it additionally analyzes the incoming data’s time stamp information and builds corresponding look-up tables.

3.3. Interval Building

The CBM experiment has several detector configurations adapted to the measurement of different observables. Depending on the trigger scenario, different detectors are required for event selection. While an effective dielectron trigger uses data from most detectors, other observables require only a subset of detectors for online selection. Thus, while there are scenarios that require full event building prior to selection, other selection scenarios are feasible with a reduced load on the event building network. This is especially important during the early building phases of the experiment.

To comply with the specific demands of different running scenarios, the FLES implements a two-staged event building and event selection system. The main data flow in the First-level Event Selector is illustrated in Figure 5. Data from all detectors that contribute to the initial stage of event selection is forwarded to the interval-building network between input nodes and processing nodes. The goal here is to concentrate data that has a similar time stamp from all detector links in a single processing node. For that reason, the input nodes package all input data corresponding to fixed-time intervals (containing, e.g., 1000 events). Each processing node receives and analyzes data from a different time interval. At the borders of these intervals, information is duplicated according to the corresponding detector’s time precision. Thus, the intervals can be processed independently. This interval building corresponds to the event building process commonly used in similar experiments. To optimize for maximum throughput, several routing concepts are currently under investigation. Using a network technology such as InfiniBand QDR, FLES has to implement high-throughput event building at the full data rate of 1 TByte/s.
3.4. Online Event Analysis

For each given time interval, the associated processing node performs the required feature extraction and online analysis. At CBM, interactions are distributed randomly over time with a given average rate. At an average event rate of 10 MHz, the average time between events is as low as 100 ns. Due to the random distribution of events over time, the actual time between events is in many cases much shorter. Discarding all events that cannot be cleanly separated in time would result in a considerably reduced efficiency of the detector. To avoid this, the track reconstruction algorithm applied by the processing node directly matches tracks not only in the three spatial dimensions, but also in time as an additional dimension. After interesting events have been identified, the two-staged event building process allows the processing node to request a corresponding sub-interval from additional detectors not participating in the initial interval building.

To efficiently utilize resources, the FLES building network is also used as an interface to the mass storage system.

The challenge in online event reconstruction is the high event rate of 10 MHz in combination with non-trivial trigger criteria requiring partial or full event reconstruction. For open charm production, event selection will be based on an online search for secondary vertices, which requires high-speed tracking and event reconstruction of STS and MVD data. The highest suppression factor has to be achieved for \( J/\psi \) mesons, where a pair of high-energy electrons or muons is required in the TRD or MUCH detectors [1].

To reach the necessary computing power, all modern computing architectures have to be utilized, including many-core CPUs and graphics cards (GPUs). Track reconstruction, which is the most time consuming combinatorial stage of the event reconstruction, will be based on parallel track finding and fitting algorithms, implementing the Cellular Automaton and Kalman filter methods. Novel languages such as OpenCL are used for parallel programming on the heterogeneous CPU/GPU online event selection system. To fully exploit the hardware potential of today’s heterogeneous many-core systems, reconstruction code is being rewritten with respect to efficient parallel processing [4].
3.5. Prototype System: LOEWE-CSC

LOEWE-CSC is a 830-node heterogeneous supercomputer designed and realized at the Goethe University of Frankfurt [5]. In many regards, this computer approaches the cluster required for FLES, and thus can be used as a prototype.

The LOEWE-CSC comprises 772 compute nodes each with two 12-core AMD Opteron 6172 processors. In addition, each of these nodes is equipped with an ATI Radeon HD 5870 GPU. The nodes are interconnected by a 40-GBit/s InfiniBand QDR network in a fat tree architecture. At the Linpack benchmark, the total processing power amounts to $285.2 \text{ TFlop/s}$ (Position 22 in Top500, Nov. 2010). The use of GPUs and an innovative cooling design result in an excellent energy efficiency of $737 \text{ MFlop/J}$ (Position 8 in Green500, Nov. 2010).

Regarding number of nodes and networking architecture, LOEWE-CSC almost completely resembles the planned FLES supercomputer. For fundamental system tests, the lack of external data inputs can be overcome by using, as a replacement for the interface cards, the GPUs as test data sources that are largely independent from the host system. The availability of Loewe-CSC as a prototype for FLES will allow for validation of the developed concepts.

4. Summary

The Compressed Baryonic Matter (CBM) experiment at FAIR requires high event rates ($10^7 \text{ Hz}$) and complex global triggers. To realize these requirements, data from self-triggered detector front-ends is forwarded in a data push readout architecture to a central physics selection system, the First-level Event Selector (FLES). To be able to process the input data stream of $1 \text{ TByte/s}$, this HPC processor farm will include FPGAs at the input stages and heterogeneous many-core architectures (e.g., GPUs). The FLES interval building network has to be capable of handling the full input data rate as relevant trigger scenarios require at least partial event building before selection. To maximize efficiency, parallelized reconstruction algorithms will perform fast 4-dimensional online event reconstruction.

References

[1] Friman B, Höhne C, Knoll J, Leupold S, Randrup J, Rapp R and Senger P (eds) 2010 The CBM physics book 1st ed (Lecture notes in physics no 814) (Berlin: Springer) ISBN 978-3-642-13292-6
[2] Müller W F J 2006 The CBM experiment @ FAIR — new challenges for front-end electronics, data acquisition and trigger systems J. Phys. Conf. Ser. 50 371–6
[3] Lemke F, Slogsnat D, Burkhardt N and Brüning U 2010 A unified DAQ interconnection network with precise time synchronization IEEE Trans. Nucl. Sci. 57 412–8 ISSN 0018-9499
[4] Kisel I 2006 Event reconstruction in the CBM experiment Nucl. Instrum. Methods Phys. Res. A 566 85–8 ISSN 0168-9002
[5] Goethe University of Frankfurt, Center for Scientific Computing The LOEWE-CSC Cluster URL http://csc.uni-frankfurt.de/?51