A GAUGE MODEL OF DATA ACQUISITION

M.W. Krasny, LPNHE Pierre et Marie Curie University, Paris, France

Abstract

A novel model of architecture of the data acquisition and data analysis is presented. This model attempts to optimize the data processing for a facility-like type of experiment in which diverse physics groups impose their, often mutually exclusive, physics-goal-optimal data handling architectures with minimal mutual interference. Details of a possible incarnation of the model projected to the ATLAS hardware capacities have been documented in a series of notes [1].

INTRODUCTION

Traditionally, in the pre-LHC multi-purpose high-energy experiments, the diversification of physics programs has been largely decoupled from the process of data-taking - the physics-optimal choice of the content of the sub-detector byte-stream data, and the implementation of refined event selection method(s), have been made in the offline analysis. For the author of this contribution, such a scheme cannot be continuously extended to the LHC environment, without significant sacrifices in the scope and in the quality of the experimental program. The rigid architecture of the present data acquisition and the data analysis models will, very likely not withstand the internal pressure of diverse, and often mutually exclusive, physics programs competing for their optimal data handling schemes. Such a pressure may result in a phase transition to an alternative architecture based upon factorisable physics-goal-oriented sub-systems bound loosely together through the common hardware maintenance, data calibration and basic data reconstruction resources. A model of such an architecture is sketched in this contribution. It is constructed in close analogy to the construction of the gauge models in particle physics. The gauge-dependent degrees of freedom of this model include: the format and content of the detector raw data, the event building method, the event filtering strategies, and the event analysis scheme. The model specifies gauge-invariant methods of handling of the data on the virtual, physics-goal-oriented slices of the data taking and data analysis system.

At the heart of the proposed model is the freedom and the responsibility, given to the physics groups, to effectively impose the data-taking configuration of the sub-detectors and of the TDAQ-system on an event-by event basis, allowing for a physics-group-dependent, optimal use of the LHC detectors capacities. As an example, a physics group analyzing rare, large E_T events is given the freedom to choose recording the most complete front-end electronic information for each of the sub-detectors. This group may implement a strategy based on inclusive, low-purity but high-efficiency, region-of-interest-guided on-line selection methods. Within the same run, a physics group interested in large cross-section processes is given a freedom to choose recording highly-compressed front-end electronic information restricted to a subset of sub-detector partitions. This group may employ on-line event selection methods based upon global topological criteria, rather than on region-of-interest-guided inclusive selection criteria, and may optimize the purity of the selected sample, rather than the efficiency of the event selection. Flexibility is given to each of the groups to run, if necessary, a group-dependent, optimal software selection-framework implemented on a subset of the level 2 and the event filter processors.

The model presented in this note changes the invariant character of the content of recorded events from an experiment-invariant one to a physics-goal-invariant one. The notion of a universal recorded event sample is thus lost and replaced by the notion of physics-goal-optimal event-streams.

The model changes significantly the borders and the interfaces between the domains of the central data acquisition, sub-detectors, general data reconstruction and physics groups. Each of the physics group can choose independently their preferred raw data form, the event selection and event building strategy, the hardware capacity of their TDAQ slice, and their optimal data reconstruction software in all the aspects except for the detector calibration and basic reconstruction. The disaggregation of the data taking and data analysis process into the domains of responsibilities is shown in Fig.1.

THE MODEL

The model absorbs event-by-event variation of the data selection and event filtering process within the three variability domains shown in Fig. 2.

Any offline task, attempting to analyze collected events, is, within the present model, exposed only to precisely-defined quantum-reactions of the data-taking process to variable data-taking conditions. These reactions are fully encapsulated within the three variability domains and are represented by quantum transitions between an allowed discrete set of states. The model is constructed in three steps. The first step consists of defining the complete set of *eigenstates* of each of the three variability domains. The second step consists of projecting these eigenstates onto the detector-partition and the TDAQ-partition granularity.

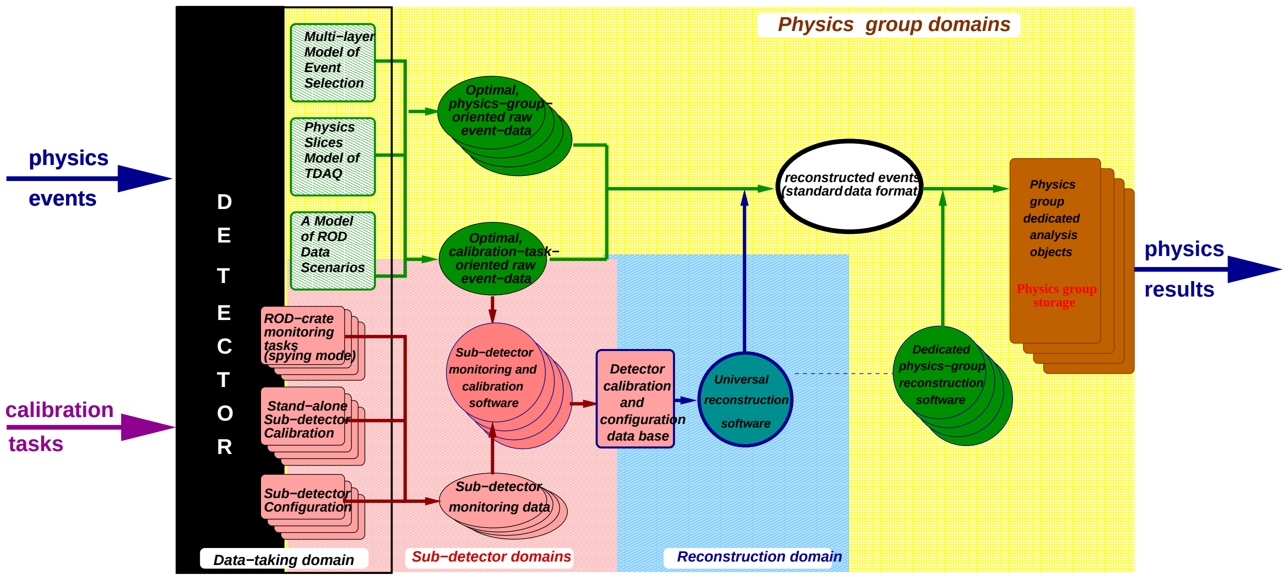

Figure 1: The domains of the data taking and data analysis.

The third step consists of specifying the model dynamics in terms of the causality and the time-granularity pattern of quantum transitions between the allowed eigenstates.

The above three steps are described in the following three sections.

THE EIGENSTATES

The eigenstates of the raw data

The eigenstates of the raw data content are specified in terms of: the data content; the data compression method; the zero suppression scheme; the channel addressing mode; the format and the content of the raw-data summary blocks.

In the presented model a broad spectrum of sub-detector-allowed operation modes is confined to a **small** set of eigenstates spanning the full space of compromise between compactness and completeness of information. These eigenstates are referred to as *data scenarios*. Reactions of the sub-detector operation modes to a variable data-taking environment are **confined** in the proposed model, to transitions (quantum-jumps) between the allowed data scenarios.

The eigenstates of the event-selection tools

The eigenstates of the event selection-tools are specified in terms of: the allowed level 1 , level 2 and event filter trigger signatures, and the allowed layers of the event-selection algorithms.

The trigger signatures are modeled in terms of trigger elements, trigger thresholds, and pre-scaling factors.

All algorithms are grouped into layers. The main goal of such a grouping is to create a supplementary fine-structure of latencies for the event-selection process.

The eigenstates of each of the algorithm layers are specified in terms of *the type* and *the granularity* of data they are using, and in terms of their *function* in the data selection. The data type-reflects the stage of the raw data unpacking and preparation. Three data-types are defined in the presented model: (1) the byte-stream data containing a fixed-format and fixed-position level 2 trigger summary blocks; (2) the raw-data objects; (3) the reconstructed-data objects. The allowed granularities of the data which are requested by an algorithm are: (1) one read-out-link; (2) one region-of-interest group of read-out-links; (3) a predefined (constant) group of read-out-links (e.g. those of a given sub-detector, or those corresponding to a predefined η-regions). The type and granularity of the data used by an algorithm determines unambiguously how they are aggregated into algorithm subsets called in this note as **layers**. The algorithm layers are the basic entities (algorithm quanta), which could be implemented on any of the level 2 trigger or event filter processors.

Each data-selection algorithm fulfills one of two functions:

1. Verifying, whether the full (partial) condition corresponding to a trigger element is fulfilled (this includes verification of the trigger-menu-predefined *veto-elements*);

2. Flagging of *infected data structures* in a way which is independent of the preloaded trigger menus. The algorithms fulfilling this functions are called the *T-algorithms*.

In the proposed model, the allowed range of the eigenstates of the event selection tools is confined to a discrete set of allowed trigger signatures and a discrete set of allowed algorithm layers. These eigenstates will be used to define the eigenstates of the TDAQ configurations discussed in the following section.

The concrete implementation of the model is based upon seven algorithm layers described in details in [1].

The eigenstates of the TDAQ-configurations

In the gauge model the clone-like slices of the TDAQ system are dynamically mapped onto **physics-goal optimized slices**. A physics slice consists of a subset of the level 2 trigger and event filter processing units. Their association is virtual, task-driven, rather than hardware-driven. Each of the physics slices has its own identity determined by the type of the level 1 trigger accepted events which will be directed to this slice. The slices may be considered as a physics working group encapsulated playground, where their members will be allowed, within well-defined rules, to make real-data exercises and eventually converge to their physics-optimal detector-data handling schemes. Each of the physics slices is given the freedom to implement, if necessary, its own event-selection framework using the full set of, or a fraction of algorithm layers. Such a framework could be optimally adapted to both the concrete physics-goal and to the actual data-taking environment. It may include a particular run-time configuration of standard algorithms confined to one slice. Each of the physics slices is also given the freedom of choosing the implementation place of the chosen subset of the algorithm layers. Last, but not least, each of the physics slices is given the freedom to choose the event building mode (i.e., the event data which will be permanently stored).

The concept of physics slices is central to the proposed model of the TDAQ architecture. This concept implements a vision according to which the quest for the physics-goal-oriented flexibility of the data-taking process will eventually result in a physics-specialized partitioning of the overall TDAQ capacity.

The physics-slices eigenstates of the allowed TDAQ-configurations are expressed in terms of:

- the list of physics slices - each slice being unambiguously defined by the allowed level 1 trigger type of events which could be processed in the slice;
- the assignment method of the level 1 trigger accepted events to the physics slices;
- the processing capacity and event building capacity of each of the slices;
- the configuration of the event-selection framework in each of the slices;
- the run-time configuration of algorithms in each of the slices;
- the allowed event-building-modes implemented in each slice.

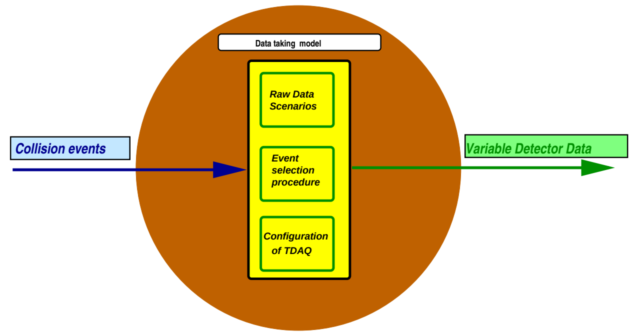

Figure 2: The variability domains

THE GRANULARITIES

The detector granularity

The projection of the data scenario eigenstates onto the detector granularity consists of defining the smallest partition of the detector, for which a given scenario is active within a given time interval. In the proposed model, the data content and the data-compression methods are allowed to take any of the allowed eigenstates on a channel-by-channel basis. The zero-suppression eigenstates are allowed to vary also on a channel-by-channel basis. The addressing mode eigenstates and the summary block eigenstates are confined to the read-out-link granularity.

The TDAQ granularity

The minimal unit of the TDAQ granularity in the proposed model is the physics-slice. The level 2 trigger and the event filter trigger menu eigenstates, the eigenstates of implemented algorithms layers, the data-selection framework eigenstates, the event-building eigenstates, and the run-configuration eigenstates of the selected algorithms are allowed to vary on a slice-by-slice basis.

THE DYNAMICS

Quasi-static implementation

In the quasi-static implementation of the model, the physics run is the basic time unit. Transitions between the eigenstates of the variability domains are allowed on a run-by-run basis. In the quasi-static implementation, the shift crew is given the responsibility to choose a particular set of data-taking eigenstates on the basis of the previous run experience and on the basis of the anticipated detector and TDAQ performance in the actual data-taking environment. The chosen set of run-defined-eigenstates is written to the configuration database and drives subsequently the configuration-dependent features of the data-reconstruction process. A change in the data-taking environment results in stopping and starting the run with the new eigenstates.

Dynamic implementation

The dynamic implementation of the model is based upon three time-scales: the event-by-event time-scale, the run-by-run time-scale, and the period-by-period time-scale. The modeling consists of defining the minimal time-scale at which the transitions between the eigenstates of the variability domains may occur and of specifying the allowed mechanisms which could activate these transitions.

In the dynamic implementation, the data scenarios are allowed to change on event-by-event basis. The read-out-link -granularity transitions are driven by the bit pattern of the level 1 trigger word. The channel-by-channel granularity transitions are restricted at present to those driven by the channel content. The eigenstates of the level 1 , level 2 , and event filter trigger signatures, the eigenstates of the implemented algorithm layers and the run-time configuration eigenstates implemented on each of the TDAQ slices are allowed to vary on a run-by-run basis. The number of the TDAQ slices and the corresponding data-processing capacity of each of the slices, the method of assignment of the level 1 accepted events to the slices, the configuration of the event-selection framework and event building implemented on the slice are allowed to change on a period-by-period basis. The run- and period-dependent settings of the TDAQ system, the trigger menus, and the implemented layer structure define the spectrum of allowed environments in which the events are dynamically selected.

In the dynamic implementation of the model each of the level 1 accepted events is directed to a predefined TDAQ slice and exposed to a slice-dependent event-selection framework. Each event will be confronted there with a dedicated selection path composed of activated algorithm layers and slice-activated high-level trigger trigger menus. An optimal path reflects, simultaneously, the sensitivity of the rate of the level 1 trigger accepted events to the data-taking environment conditions, and specific wishes of the physics group interested in analyzing events of a given type.

The active algorithm layers implemented in the slice determine the latencies of the decision steps in rejecting and accepting the events. The activation of the T-algorithms on a particular slice is driven by the local, slice-specific dead time.

THE GAUGE DEPENDENT EVENT PATH FOR GAUGE INDEPENDENT PHYSICS ANALYSIS

Within the proposed event selection framework, each level 1 triggered event is confronted with a dedicated gauge-path shown in Fig. 3. A level 1 trigger word determines the data contend of the sub-detector byte-streams for the event and determines the TDAQ slice to which this event is directed. The event is exposed to the slice-specific event selection framework and event building scheme. Selected events coming form a slice form the slice specific data-stream which is reconstructed using general reconstruction software. This software is used only for calibration of the detector response and for reconstructing the space points for the trackers and the cluster energies for the calorimeters. The pre-reconstructed events are processed further using the physics group dedicated programs.

A physicist exposed to the purely offline analysis environment and analyzing the data taken by one of the TDAQ slices could ignore the data-taking environment and the presented model altogether. In other words, the gauge-dependent-path of (her)his favorite events, from the LVL1-accept decision up to inclusion in the final physics plot will be fully encapsulated for (her)his gauge independent analysis. The encapsulation mechanism is exactly the same as the one for the QED-events generated in the technically most convenient gauge. Any physicist asking a valid physics question need not care which gauge is used by the event generator. At the worst, he may be confronted with a very ineffective event generation leading to very large analysis errors.

On the other hand, a curiosity-driven physicist will be provided with the full gauge-path information for each of the accepted events. This will enable him to search for the most efficient gauge event-path, to study the gauge symmetry breaking phenomena, and to modify the gauge invariance rules implemented on his (her) slice.

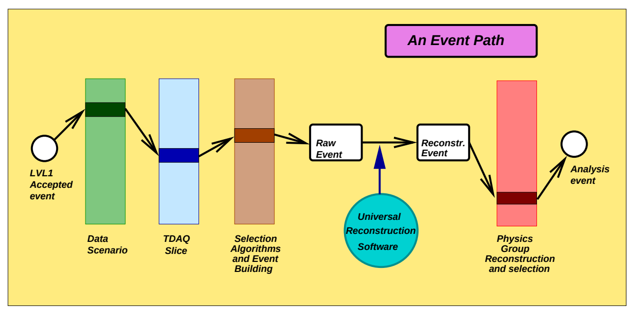

Figure 3: An event path.

For the detiled description of the model see [1].

REFERENCES

[1] M.W. Krasny, A Model of Dynamic Integration of ATLAS detector, Trigger and Software in LHC Data-Taking Environment: NOTE I - Prolegomenon - ATL-COM-GEN-2003-002; NOTE II -A Gauge Model of Data Taking - ATL-COM-GEN-2003-004; Note III - A Model of Data Scenarios - unpublished; Note IV - An Event Data Model of Variable Raw Data - unpublished; Note V - A Multi-Layer Model of the LVL2 Event Selection Architecture - unpublished; Note VI A Physics-Slices Model of the TDAQ architecture - unpublished; Note VII - a Dynamic Model of Integration Phases - unpublished.