

A Comparison Between SIMDAQ and Testbed Measurements

R. Blair, J. Schlereth

Argonne National Laboratory, Argonne, Illinois 60148, USA

M. Dobson, A. Bogearns
CERN, Geneva, Switzerland

J.C. Vermeulen
NIKHEF, Amsterdam, Netherlands

R. Hauser
Michigan State University, East Lansing, Michigan 48824, USA

D. Calvet, I. Mandavidze, M. Huet
CEA Saclay, 91191 Gif-sur-Yvette CEDEX, France

Abstract

The discrete event simulation embodied in the SIMDAQ program plus a number of parameters form one view of the ATLAS Level 2 expected performance. Both as a test of SIMDAQ and as a way of better evaluating current parameters which can be extrapolated to values relevant for the future, we compare the testbed performance to that predicted by SIMDAQ configured similarly to the testbed setups. We do some rough adjustment of the parameters to better match the measured performance.

Introduction

The SIMDAQ program, written by Jos Vermeulen, is a discrete event simulation program written in C++ specifically to allow a detailed component level simulation of the ATLAS Level 2 system¹. It is designed to allow flexibility in component count and component performance. Because of this flexibility it is possible to use it to model even the small systems that have been assembled at CERN for the ATLAS testbed. SIMDAQ was developed for a complete ATLAS Level 2 simulation and is not designed to emulate precisely the system or software organization of current testbed/reference software system. Nonetheless it is close and using it to model the testbed systems is useful both as a test of SIMDAQ and as a check on our current understanding of the parameters used to estimate system performance.

Rather than rearrange the components in SIMDAQ to model the exact testbed implementation, we have taken the simple approach here of putting together a model of the testbed with process models as given in the standard SIMDAQ system and with component counts (processors, network layout and data sizes) as used in testbed measurements with an ATM switch and a system of 38 Intel based PC's made in October 1999. Because the PC's involved were acquired at different times and by different institutions they varied in a number of details. Clock speeds for the processors varied from 200 to 450 MHz. Most systems had clock speeds of 300 or 400 MHz. An inventory of systems used appears in table 1.

<i>Description</i>	<i>Number</i>	<i>Clock Speed</i>
PCATMAPP13-19	7	400 MHz
PCETB01-08	8	200 MHz
PCSCI01-07	7	350 and 450 MHz
PCET019-020	2	400 MHz
PCET021-022	2	dual 450 MHz
PCATMAPP01-08	8	300 MHz
PCATMAPP09-10	2	330 MHz
PCATMAPP11-12	2	dual 330 MHz

Table 1: System inventory used in testbed measurements

Because of the diversity of systems, motherboard chipsets as well as processor clock speeds, any attempt to reproduce the testbed results in detail would require implementing different parameters for many of the components in the system to be modeled. The spirit of what is attempted here is not to try to obtain a detailed picture of the system performance, but to show that to within *factors of two* the model and expected parameters yield the same performance as observed. This exercise has been useful in flushing bugs in SIMDAQ that effected the specialized measurements involved for the testbed and may have had an impact on the full system model. It was also instructive in suggesting areas where software and hardware improvements are needed to obtain the desired performance targets.

Measurements and Parameters

About 150 separate test runs were performed with the ATM based setup. In order to keep the comparisons manageable we have singled out three sets of runs that probe most system parameters. The measurements include a set of runs with no event queuing to test overall sums of system time, a set of runs with varying data sizes and a single ROB system and a set of runs with a large system with equal numbers of ROB and steering processors.

A number of parameters have been agreed upon as a reasonable guess at system parameters for the final ATLAS system. The ones that are relevant to these measurements are listed as nominal in table 2 along with a brief description. Component performance is expected to improve from present day systems and this improvement was reflected in the final system choices. The processor clock speed for final system processors was expected to be 1 GHz while processors used in these tests were typically 400 MHz. Since much of the performance parameterized scales, at least roughly, as clock speed it is reasonable to use as a

starting point parameters that reflect longer processing times by a factor of 2.5 (listed in table 1 as well). Not all parameters were simply scaled since the system being modeled did not have the same functionality and the link speed is not expected to rise significantly. Finally a parameter set was estimated by reviewing the measured performance in the runs with no queueing. Parameters were determined for scheduling time, link speed and processing times that tend to match these measurements given a simple system model. This falls short of a full fit of all the parameters using all the measurements, but given the spirit of obtaining a rough match and some guidance on where any large problems might be we felt this was adequate and can be evaluated by the agreement with the actual measurements. This last set of parameters is listed in table 2 as *measurement based*.

<i>Parameter</i>	<i>Description</i>	<i>Nominal</i>	<i>X 2.5</i>	<i>Measurement Based</i>
LinkSpeed	The time in microseconds required to transmit one byte through the various network interfaces	0.06667	0.06452	0.14493
ScheduleExecuteTime	Number of microseconds required to create (and tear down) a processing thread	5.0	12.5	42.64
Global Time	Time in microseconds required to process the global decision	55.55	138.9	226.
RoIProcessTime	Time in microseconds an RoI Processor (in the Supervisor) takes to process an RoI	8.0	20.0	6.6
DecisionBlockHandlingTime	Time required by ROB to handle a decision block	55.55	13.875	5.55

Table 2: Parameters used in SIMDAQ

All the measurements described here were performed with the reference softwareⁱⁱ even though alternative optimized software written by the Saclay group was used for a number of other measurements taken with the same system. There were significant differences between the optimized and reference software measurements. Much of the performance differences between SIMDAQ and the measurements can be made up by further optimization of the software as can be evidenced by both the significant improvement that the Saclay software made in performance and the fact that the link speed that optimizes some of the measurement agreement appears to indicate an effective link speed lower than the actual speed. One interpretation of this is that data copies that accompany the data transfers slow the transfers significantly (about a factor of two).

The SIMDAQ configuration used the ATM switch simulation for the switch since this best matched the actual switch. There were two configuration files that corresponded to two distinct setups. One setup involved data being transferred from the ROB's to the steering nodes. Another was used for runs where no data transfer was requested from the ROB's. The number of ROB's, RoI CPU's (in the supervisor) and steering nodes were adjusted to agree with the various runs. The parameters were also adjusted to correspond to the three choices outlined above.

Results

The first set of measurements involved a minimal system with a single component of each type. One supervisor processor, one ROB and one steering processor. Events were dispatched in free running mode by the supervisor and the data size was varied from no data (for this point there was no data request made to the ROB) to 4096 bytes per event. The queue depth in the steering node was set to a single event deep so no new event was dispatched until the last event had been processed by the steering node and an event

decision had reached the supervisor. The measured event rate and the three SIMDAQ parameter set results appear in figure 1 below.

No Queueing Rate

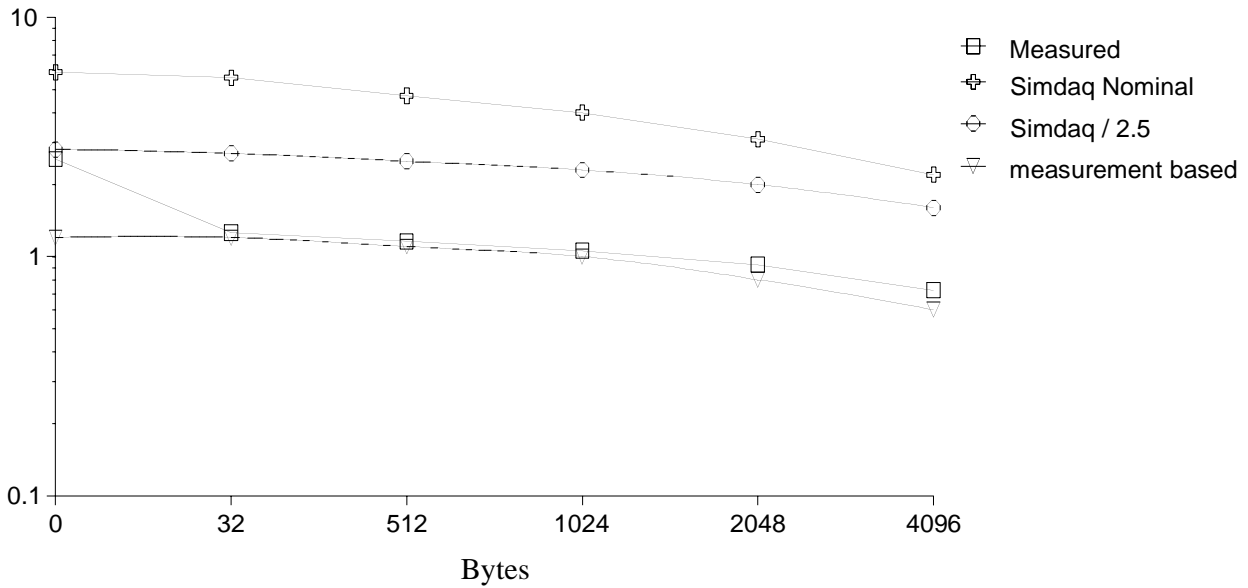


Figure 1: Event rates (kHz) versus data size for runs with a single component of each type and no events queued.

The second set of measurements involved testing the ROB (or ROB emulator) transfer ability. There were 3 RoI CPU's and more than 8 steering nodes and a single ROB. The event queue depth was set to 24 (which means that about 9 events were queued on average in each steering node). The total event rate was measured as a function of the size of the single buffer passed for each event. Figure 2 shows the rate versus data size measured and the corresponding SIMDAQ runs. The results clearly indicate that the degrading of the data link speed needed to match the no queueing runs is more pessimistic than is needed to match the measured rates for ROB performance.

ROB performance

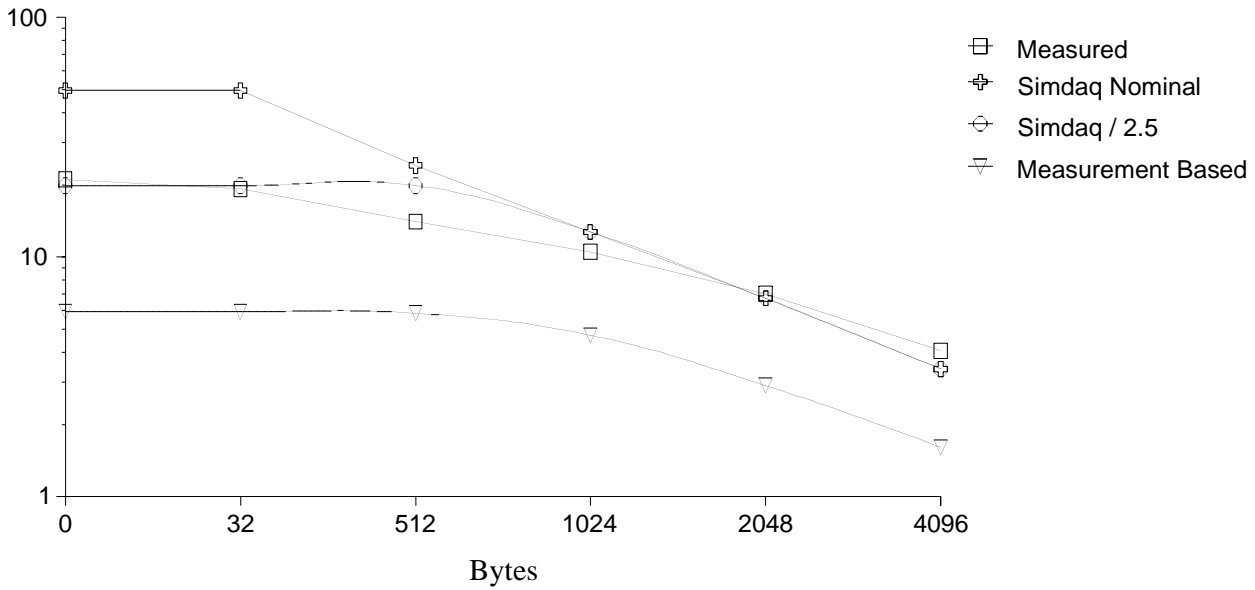


Figure 2: Rate in kHz versus data size for ROB performance runs.

In order to approximate as much as possible a full system a set of runs was taken with 14 ROB's, 14 steering nodes and 3 supervisor RoI CPU's. A single ROB was interrogated per event and 24 events were queued by each supervisor. The data size was varied and event rate was measured as a function of data size. The measured and predicted rates appear in figure 3.

System performance

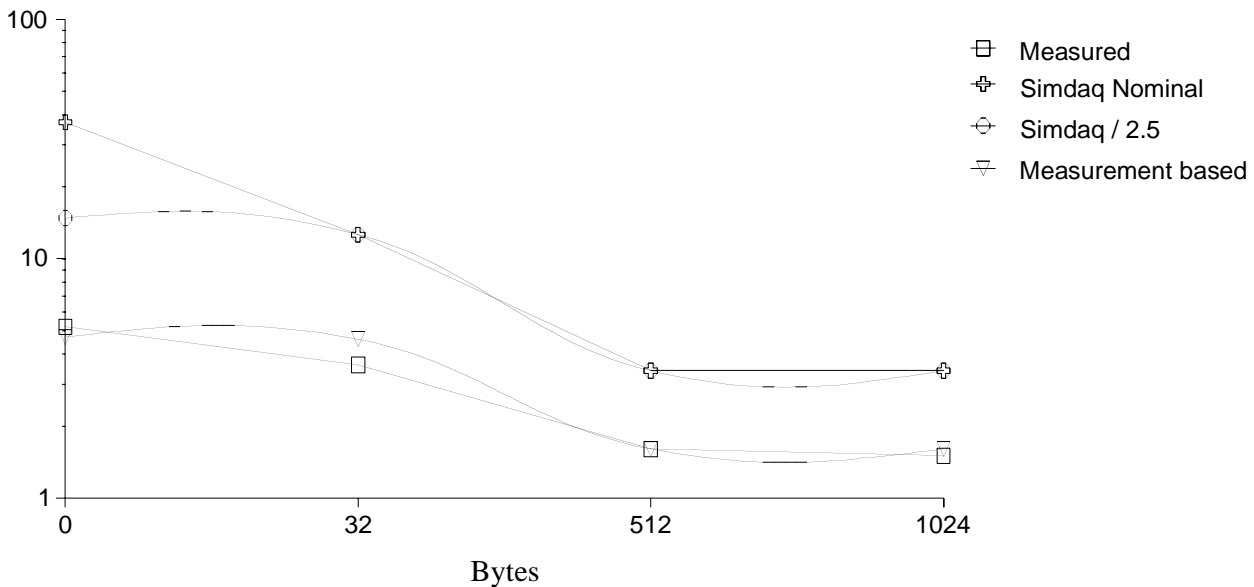


Figure 3: Performance measured for a 14X14X3 system.

One other set of runs of interest tested the rate capability of the supervisor in a system with a single ROB, 8 steering processors and a single RoI CPU. The queue depth was fixed at 24 events and the configuration determined the maximum rate that the supervisor could turn events around. The measured and SIMDAQ rates are summarized in table 3.

Conditions	Rate in kHz
measured	7.1
SIMDAQ nominal	26.3
SIMDAQ X 2.5	10.5

<i>Conditions</i>	<i>Rate in kHz</i>
SIMDAQ measurement based	3.8

Table 3: Summary of supervisor runs.

Conclusions

We have compared SIMDAQ to a number of testbed runs and find that the performance expected based solely on the current processor generation clock speed is in agreement with what was measured to within a factor of two. Given that the reference software used still has room for performance optimization this should probably be regarded as encouraging. At the very least a range of parameter is explored and some sense of how well the parameters are known is available from this comparison.

Appendix

Simdaq configuration file for cases where data is transferred from ROB's to steering node

```
#TestBed run with
#   histo file name: XXFileNameXX
#   linkspeed: XXLinkSpeedXX
#   scheduling times: XXScheduleExecuteTimeXX
#   number of steering: XXNumberOfGlobalXX
#   global time: XXGlobalXX
#   number of RoI Processors: XXNumberOfRoIProcessorsXX
#   RoI processor execute time: XXRoIProcessTimeXX
#   method of dispatch: XXPartitionManagementXX
#   decision block size: XXSizeOfDecisionBlockXX
#   number of ROB's: XXNumberOfROBInsXX
#   number of RoIs: XXNumberOfRoIsXX
#   data size: XXRoIDSizeXX
#   ROB decision block handling time: XXDecisionBlockHandlingTimeXX
#   queue number: XXMaxNumberOfRequestsXX
#   rate limit: XXLVL1TriggerRateXX
#####
# $Revision: 1.1 $
# $Date: 1999/11/03 02:45:16 $
# $Author: reb $
```

Header

```
ConfigFile FormatVersion 1 Name TestOfTheNewFormat DemoSimul
End Header
```

```
#=====
```

Objects

#Level-1

```
LVL1  LVL1TriggerSystem    1
LVL1  RoIBuilder           1
LVL1  FastBroadcaster      1
```

#EMcal - local

```
EM_CalLocal  ROD          1
EM_CalLocal  FastBroadcaster 1
EM_CalLocal  ROBIN        XXNumberOfROBInsXX
```

#Global

```
Global ModelCProcessor          XXNumberOfGlobalXX
Global BasicSwitch              2
```

#Supervisor

```
Super LVL2Supervisor           1
```

#Event Builder + Level-3

```
EBEF BasicSwitch              1
EBEF EFProcessor              1
```

End Objects

#=====

Associate
EM_CalLocal EMCalorimeter Global
End Associate

#=====

#EM_Calo
Assign EM_CalLocal ROBIN
End Assign

#=====

Parameters EventList
All EventWheelSize 32 WheelTimeSpan 64.0

Parameters TDAQSystem
All LinkSpeed XXLinkSpeedXX SizeOfDecisionBlock XXSizeOfDecisionBlockXX
All NumberOfEvents 100000 DumpHistoFileName XXFileNameXX NumberOfEventsPerInterval 100000
All LVL2AcceptProbability 0.0

Parameters SourceOfPhysEvents
All RoIDSize XXRoIDSizeXX NumberOfRoIs XXNumberOfRoIsXX NumberOfROBIns
XXNumberOfROBInsXX
All LocalCalEmRoI Delta ProcTimeAverage 1.0
All ROBINEmCalEmRoI Delta ProcTimeAverage 0.0
All Global Delta ProcTimeAverage XXGlobalXX

Parameters LVL1 RoIBuilder
All NumberOfRoIProcessors XXNumberOfRoIProcessorsXX RoIProcessTime XXRoIProcessTimeXX
IndividualAssign XXPartitionManagementXX
All DecisionHandling MaxNumberOfRequests XXMaxNumberOfRequestsXX FreeRunning
All SchedReqExecuteTime XXScheduleExecuteTimeXX SchedNextExecuteTime
XXScheduleExecuteTimeXX

Parameters LVL1 LVL1TriggerSystem
#All LVL1TriggerRate XXLVL1TriggerRateXX TypeOfDistribution EqualInterval RateLimited
All LVL1TriggerRate XXLVL1TriggerRateXX TypeOfDistribution EqualInterval

Parameters EM_CalLocal ROBIN
All GenerateProcTime Delta ProcTimeAverage 0.0 DecisionBlockHandlingTime
XXDecisionBlockHandlingTimeXX SchedReqExecuteTime XXScheduleExecuteTimeXX
SchedNextExecuteTime XXScheduleExecuteTimeXX
All ExtractByDMATransferSpeed 0.0 IndexingTime 0.

#0 = calorimeter data 1 = control traffic

#Parameters Global BasicSwitch

#Specific 0 StartUpTime 0.0 InternalByteTransferTime 0.0

#Specific 1 StartUpTime 0.0 InternalByteTransferTime 0.0

Parameters Global ModelCProcessor

All SchedReqExecuteTime XXScheduleExecuteTimeXX SchedNextExecuteTime
XXScheduleExecuteTimeXX HandlingAllRoisFlag RoIFormulateProcessTime 0.0 DecisionProcessTime
0.0
All MergeSpeed 0.0

#Parameters Super LVL2Supervisor

#All SuperTime 0.0 FreeRunning DecisionBroadcast

Parameters Analyser

All CurrentEventRate numberOfBins 500 minValue 0.0 maxValue 50.0
All LVL2DecisionTime-long numberOfBins 160 minValue 0.0 maxValue 40000.0

#Parameters EBEF FastBroadcaster

End Parameters

#=====

Connect

#Data Generation part

MultiConnect LVL1 LVL1TriggerSystem AnyOutPort To LVL1 FastBroadcaster InPort
SingleConnect LVL1 LVL1TriggerSystem 0 PortToROIBuilder To LVL1 RoIBuilder 0 LVL1InPort
MultiConnect LVL1 FastBroadcaster AnyOutPort To EM_CalLocal ROD InPort
MultiConnect EM_CalLocal ROD OutPort To EM_CalLocal FastBroadcaster InPort
MultiConnect EM_CalLocal FastBroadcaster AnyOutPort To EM_CalLocal ROBIN EventDataPortIn

#RoI distribution part

MultiConnectToRange LVL1 RoIBuilder AnyIndividualNetOutPort To Global BasicSwitch Sub 0 FOR 1
AnyInPort
MultiConnectToRange LVL1 RoIBuilder AnyDecisionOutPort To Global BasicSwitch Sub 1 FOR 1
AnyInPort
MultiConnectToRange Global ModelCProcessor OutPort To Global BasicSwitch Sub 1 FOR 1 AnyInPort

#Local parts

#EM_Calo

MultiConnectToRange EM_CalLocal ROBIN LVL2PortOut To Global BasicSwitch Sub 0 FOR 1
AnyInPort

#Local to global and global part

MultiConnectRangeTo Global BasicSwitch Sub 0 FOR 1 AnyOutPort To Global ModelCProcessor InPort
MultiConnectRangeTo NotLastOne Global BasicSwitch Sub 1 FOR 1 AnyOutPort To EM_CalLocal
ROBIN RoiPortIn
MultiConnectRangeTo Global BasicSwitch Sub 1 FOR 1 AnyOutPort To LVL1 RoIBuilder
AnyDecisionInPort

#ROBIN to EBEF, EF system

MultiConnect EM_CalLocal ROBIN EFPortOut To EBEF BasicSwitch AnyInPort
MultiConnect EBEF BasicSwitch AnyOutPort To EBEF EFProcessor InPort

End Connect

Simdaq configuration file for cases where no data is transferred from ROB's to steering node

```
#TestBed run with
#   histo file name: XXFileNameXX
#   linkspeed: XXLinkSpeedXX
#   scheduling times: XXScheduleExecuteTimeXX
#   number of steering: XXNumberOfGlobalXX
#   global time: XXGlobalXX
#   number of RoI Processors: XXNumberOfRoIProcessorsXX
#   RoI processor execute time: XXRoIProcessTimeXX
#   method of dispatch: XXPartitionManagementXX
#   decision block size: XXSizeOfDecisionBlockXX
#   number of ROB's: XXNumberOfROBInsXX
#   number of RoIs: XXNumberOfRoIsXX
#   data size: XXRoIDSizeXX
#   ROB decision block handling time: XXDecisionBlockHandlingTimeXX
#   queue number: XXMaxNumberOfRequestsXX
#   rate limit: XXLVL1TriggerRateXX
#####
# $Revision: 1.1 $
# $Date: 1999/11/03 02:45:16 $
# $Author: reb $
```

Header

```
ConfigFile FormatVersion 1 Name TestOfTheNewFormat DemoSimul
End Header
```

```
#=====
```

Objects

#Level-1

```
LVL1  LVL1TriggerSystem    1
LVL1  RoIBuilder           1
LVL1  FastBroadcaster      1
```

#EMcal - local

```
EM_CalLocal  ROD           1
EM_CalLocal  FastBroadcaster 1
EM_CalLocal  ROBIN        XXNumberOfROBInsXX
```

#Global

```
Global ModelCProcessor          XXNumberOfGlobalXX
Global BasicSwitch              2
```

#Supervisor

```
Super LVL2Supervisor           1
```

#Event Builder + Level-3

```
EBEF BasicSwitch              1
EBEF EFProcessor              1
```

End Objects

#=====

Associate
EM_CalLocal EMCalorimeter Global
End Associate

#=====

#EM_Calo
Assign EM_CalLocal ROBIN
End Assign

#=====

Parameters EventList
All EventWheelSize 32 WheelTimeSpan 64.0

Parameters TDAQSystem
All LinkSpeed XXLinkSpeedXX SizeOfDecisionBlock XXSizeOfDecisionBlockXX
All NumberOfEvents 100000 DumpHistoFileName XXFileNameXX NumberOfEventsPerInterval 100000
All LVL2AcceptProbability 0.0

Parameters SourceOfPhysEvents
All RoIDSize XXRoIDSizeXX NumberOfRoIs XXNumberOfRoIsXX NumberOfROBIns
XXNumberOfROBInsXX
All LocalCalEmRoI Delta ProcTimeAverage 1.0
All ROBINEmCalEmRoI Delta ProcTimeAverage 0.0
All Global Delta ProcTimeAverage XXGlobalXX

Parameters LVL1 RoIBuilder
All NumberOfRoIProcessors XXNumberOfRoIProcessorsXX RoIProcessTime XXRoIProcessTimeXX
IndividualAssign XXPartitionManagementXX
All DecisionHandling MaxNumberOfRequests XXMaxNumberOfRequestsXX FreeRunning
All SchedReqExecuteTime XXScheduleExecuteTimeXX SchedNextExecuteTime
XXScheduleExecuteTimeXX

Parameters LVL1 LVL1TriggerSystem
#All LVL1TriggerRate XXLVL1TriggerRateXX TypeOfDistribution EqualInterval RateLimited
All LVL1TriggerRate XXLVL1TriggerRateXX TypeOfDistribution EqualInterval

Parameters EM_CalLocal ROBIN
All GenerateProcTime Delta ProcTimeAverage 0.0 DecisionBlockHandlingTime 0.0
SchedReqExecuteTime 0.0 SchedNextExecuteTime 0.0
All ExtractByDMATransferSpeed 0.0 IndexingTime 0.

#0 = calorimeter data 1 = control traffic

#Parameters Global BasicSwitch

#Specific 0 StartUpTime 0.0 InternalByteTransferTime 0.0

#Specific 1 StartUpTime 0.0 InternalByteTransferTime 0.0

Parameters Global ModelCProcessor
All SchedReqExecuteTime XXScheduleExecuteTimeXX SchedNextExecuteTime
XXScheduleExecuteTimeXX HandlingAllRoisFlag RoIFormulateProcessTime 0.0 DecisionProcessTime
0.0
All MergeSpeed 0.0

#Parameters Super LVL2Supervisor
#All SuperTime 0.0 FreeRunning DecisionBroadcast

Parameters Analyser
All CurrentEventRate numberOfBins 500 minValue 0.0 maxValue 50.0
All LVL2DecisionTime-long numberOfBins 160 minValue 0.0 maxValue 40000.0

#Parameters EBEF FastBroadcaster

End Parameters

#=====

Connect

#Data Generation part
MultiConnect LVL1 LVL1TriggerSystem AnyOutPort To LVL1 FastBroadcaster InPort
SingleConnect LVL1 LVL1TriggerSystem 0 PortToROIBuilder To LVL1 RoIBuilder 0 LVL1InPort
MultiConnect LVL1 FastBroadcaster AnyOutPort To EM_CalLocal ROD InPort
MultiConnect EM_CalLocal ROD OutPort To EM_CalLocal FastBroadcaster InPort
MultiConnect EM_CalLocal FastBroadcaster AnyOutPort To EM_CalLocal ROBIN EventDataPortIn

#RoI distribution part
MultiConnectToRange LVL1 RoIBuilder AnyIndividualNetOutPort To Global BasicSwitch Sub 0 FOR 1
AnyInPort
MultiConnectToRange LVL1 RoIBuilder AnyDecisionOutPort To Global BasicSwitch Sub 1 FOR 1
AnyInPort
MultiConnectToRange Global ModelCProcessor OutPort To Global BasicSwitch Sub 1 FOR 1 AnyInPort

#Local parts

#EM_Calo
MultiConnectToRange EM_CalLocal ROBIN LVL2PortOut To Global BasicSwitch Sub 0 FOR 1
AnyInPort

#Local to global and global part
MultiConnectRangeTo Global BasicSwitch Sub 0 FOR 1 AnyOutPort To Global ModelCProcessor InPort
MultiConnectRangeTo NotLastOne Global BasicSwitch Sub 1 FOR 1 AnyOutPort To EM_CalLocal
ROBIN RoiPortIn
MultiConnectRangeTo Global BasicSwitch Sub 1 FOR 1 AnyOutPort To LVL1 RoIBuilder
AnyDecisionInPort

#ROBIN to EBEF, EF system
MultiConnect EM_CalLocal ROBIN EFPortOut To EBEF BasicSwitch AnyInPort
MultiConnect EBEF BasicSwitch AnyOutPort To EBEF EFProcessor InPort

End Connect

References

ⁱSIMDAQ 4.5 is described in a web based document available at <http://www.nikhef.nl/atlas/daq/Modelling-4-11-99/ComputerModel.pdf> and in DAQ-Note-98-086

ⁱⁱThe reference software is described in documents located at <http://www.cern.ch/Atlas/project/LVL2testbed/www/notes/>