

# Report of the December 17, 2004 Run2b Event Builder Upgrade Review

William Badgett, Guillermo Gomez-Ceballos, Steve Nahn, Jim Patrick, Mel Shochet

*Run 2b Event Builder Upgrade Review Committee*

## Abstract

This is a report on the findings of the Review committee for the Run 2B Event Builder based on the presentations of the [review of December 17, 2004](#).

## Contents

<b>1 Introduction</b>	<b>1</b>
<b>2 General</b>	<b>2</b>
<b>3 Comments and Recommendations from the committee</b>	<b>3</b>
3.1 Throughput calculations . . . . .	3
3.2 Error Handling and Efficiency of Operation . . . . .	3
3.3 Commissioning and Operations Plan . . . . .	4
3.4 Additional Crates . . . . .	5
<b>4 Summary</b>	<b>6</b>

## 1 Introduction

Part of the CDF Run 2b upgrade to cope with higher luminosity is an upgrade to the Event Builder (EVB) system, which is responsible for taking event fragments in “DAQ” buffers after a L2 Accept (L2A) and combining them into a full event for further trigger processing by the L3 system. For Run 2A, the current EVB runs comfortably at 300 Hz with an event size of  $\approx 250$  kB, for a throughput of  $\approx 75$  MB/sec. Due to increased luminosity and subsequent increased occupancy, the Run 2b system will need to operate at 1 kHz with event sizes of  $\approx 500$  kB, for a throughput  $\approx 500$  MB/s. Indeed in 2004 during high luminosity running the EVB was a source of high deadtime, achieving  $\approx 390$  Hz with 30% deadtime. Stop gap measures had to be applied to the trigger scheme to alleviate the high deadtime and return to reasonable data taking efficiency.

A schematic of the EVB system is shown in figure 1. Roughly speaking, after a L2A, raw data from the front end crates are concatenated into  $\approx 72$  VRBs housed in 15 VRB crates. The L2A is passed from the Trigger Manager process (TM) to the Scanner Manager process (SM) and Scanner CPUs (SCPU), all of which are implemented as embedded Motorola CPUs running VxWorks, via a proprietary network called SCRAMnet. Upon receiving the L2A, each SCPU reads out and concatenates the event fragments from the VRBs in its crate, and sends the resulting fragment through the ATM switch to one converter node, which receives the data and passes it on to the L3 subfarm for processing.

In the upgraded EVB system, the plan is not only to increase maximum rate capability but also to remove outdated technology in favor more widely available and better supported architectures. In this vein, the SCRAMnet control network will be replaced by Gigabit ethernet, the ATM switch by a Cisco 6509 Gigabit ethernet switch, the SCPUs by faster VMIC 7805 processors running real time Linux, and the SM functionality will be moved to a commodity PC also running the EVB proxy process which is responsible for relaying Run Control Commands to the EVB system. In addition, the software for the entire EVB system is undergoing a complete overhaul to accommodate the new architecture.

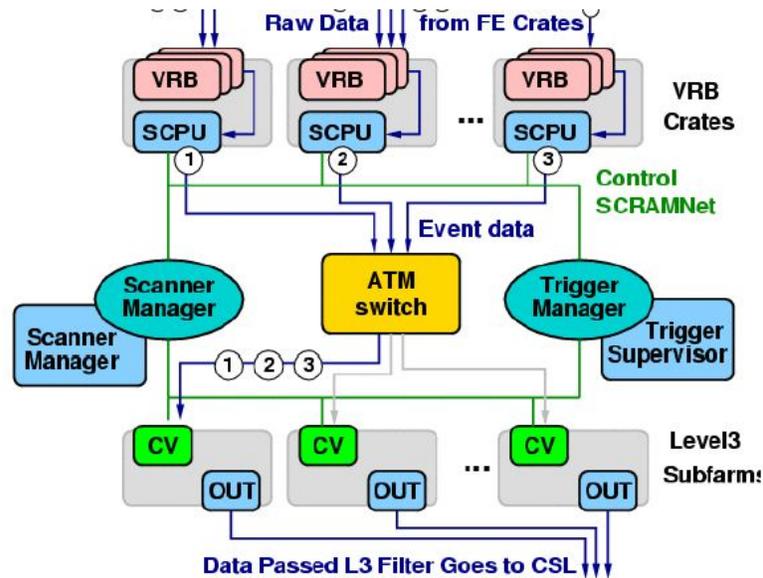


Figure 1: Schematic of the EVB system

## 2 General

The committee heard overview reports on the software status, rate measurements and benchmarks, commissioning plans, and also limitations to the L2A rate from other sources. The talks can be found at

[http://www-cdf.fnal.gov/internal/WebTalks/Archive/0412/041217\\_run\\_2b\\_event\\_builder\\_review/](http://www-cdf.fnal.gov/internal/WebTalks/Archive/0412/041217_run_2b_event_builder_review/).

In general the project is very well advanced- it is clear that the project will be completed in time, more likely well ahead of schedule. It appears to be well staffed currently and the plan for implementation is sound. The committee commends the proponents so far for their admirable progress.

Clearly, there are two outside influences which dictate the apparent performance of the EVB; the L2A rate and the event size in each of the VRBs crates. Outside of the EVB upgrade several projects to control the L2A rate and reduce and balance the event size in the VRB crates are either recently completed or in the works. The latter was also discussed at the review.

### **3 Comments and Recommendations from the committee**

The review committee did have some specific recommendations to be followed up to help get the system in place as soon as possible.

#### **3.1 Throughput calculations**

The committee believes that the design is sound, but concurs with the presenters that work needs to be done to optimize the code to reach the throughput required. This is of course what the next few months are for, but the committee will expect to see real measurements which back up some of the estimates of gain from optimization at the next review.

#### **3.2 Error Handling and Efficiency of Operation**

Several members thought that since the SCPU code and EVB proxy code are being rewritten the EVB group should take this opportunity to reexamine the error handling philosophy. This includes making sure the relevant data gets saved for expert diagnosis, including potentially information from the previous event, and also that the irrelevant data gets suppressed - because of the data driven structure, once one event is corrupted, subsequent events have a high probability of generating a substantial spew of errors, which only serves to obscure the initial failure. The proposed philosophy of not stopping for errors probably won't work in most cases, at least for bad non-SVX VRB data.

In addition, the current EVB is known to take considerable time to recover after a failure, and cost CDF downtime. The committee encourages the proponents to consider not only reducing the rate of this necessity, but also making sure that the recovery procedure is not excessive. For example:

1. Is it really necessary to reboot all SCPUs when only one fails?

2. Is there a way to avoid taking the entire CDF system back to the PARTITION state?

When CDF is running well, any recovery times requiring the entire DAQ to be reset are significant, and this is the time to revisit and possibly address those associated with the EVB.

### 3.3 Commissioning and Operations Plan

The committee thinks that it will be very hard to commission the system if the current EVB system has to be compromised to do it, since there are only typically a few hours between stores, which would allow little time for actual tests if the changeover takes some time, and no contingency should a switchover encounter trouble. Moreover, as part of the Commissioning, the committee fully expects the new system to take real BEAM data from the real detector during End of Store studies for example. With significant setup time this will be much more difficult. The minimal litmus test for readiness will be getting real beam data from the VRBs through at least the reformatter at normal data taking rates (say 300 Hz) with less BUSY deadtime than the current system. Only after several stores of with all CDF components participating in data taking with the new EVB system will decommissioning of the old EVB be conceivable, even if this test period spans the 2005 shutdown. Finally, scheduling commissioning time with the new EVB system will be further complicated because many other detector groups require a working EVB to complete their work.

To alleviate if not solve these problems, the committee considered a scenario where the DAQ VRB Test CPUs currently residing in the DAQ VRB crates could be replaced with the new EVB SCPUs, by default, with the proviso that the DAQ group can reinstall one of their CPUs to run Tracer-VRB tests, which should occur infrequently (The VRB crates dedicated to Silicon readout already have new EVB SCPUs installed). This was thought to be the only impediment to being able to run with the new EVB without disabling the old EVB. If this is not the case, measures should be taken to alleviate the other restrictions, and *clear procedures* on how to switch from the old EVB to the new EVB **AND BACK!** should be written down and available on web pages. The goal here is to enable earlier and more frequent testing with the real front-end systems. These tests should also include rate/data volume torture tests with the real front-end as soon as possible to find any problems early.

If the hardware impediments are relieved, the committee whole-heartedly supports the dedication of significant testing time to the EVB upgrade, and recommends a formal strategy with the Operations group for commissioning the system as soon as feasible with minimal cost to Physics data taking. This could include a set of necessary criteria for each test, such as

- $L_{inst} < x$
- Stated goal of test

- Results of last test understood and presented (if applicable)
- ...

While such a list may appear to be rather bureaucratic, with rotating shift crews and Operations managers, and extended periods of time between tests it simplifies negotiating for testing time if both Operations and EVB Commissioning know what is expected.

While the current manpower seems sufficient for putting together the system, of the four people involved only one is a potential pager carrier during operations, which is clearly not enough. Therefore the committee thinks it is *crucial* to get more people who will be responsible for the operation involved in the commissioning so that expertise is spread among more people. For instance, assigning a new person to document the system would not only produce a new expert but would also provide a start at the missing documentation. Assigning incoming graduate students responsibility for the commissioning the system would also provide an expansion of the expertise base for future operation as well.

### 3.4 Additional Crates

In order to reduce the throughput load on any one SCPU, there is a proposal to add more VRB crates to the system. Previously the maximum number of crates was 16, limited by the number of input spigots to the ATM switch. With the CISCO switch this restriction is alleviated, such that additional crates could be added to further parallelize the data flow. The committee didn't feel comfortable recommending how many additional crates, although it agreed the number should be  $\geq 3$ . To understand how many more crates are needed, we recommend:

1. Detailing what are the impediments to getting n Crates, where  $n=3,6$
2. Contacting [Yale people](#) to collaborate on simulations of downtime with all Upgrades (SVT, L2, TDC format) simulated, with event sizes of 40, 30, and 20 kB/vrb crate.

A decision on how many VRB crates are in the system should be made on a timescale of month, meaning by the end of January, 2005. Waiting longer could have serious impact on the schedule. In the meantime, because of a long lead time and the eventuality of getting some crates we suggest starting to prepare racks (via Operations) in Row 30 of the 1st floor for future installation. It is our understanding that power, water cooling, and rack protection installation are likely to be necessary for this row. Along these lines an additional 6 VMIC processors should be procured in time to service the additional crates. Should less crates be necessary, the excess processors can serve as spares.

## 4 Summary

Overall the committee is confident that the Run 2b upgrade for the Event Builder will satisfy the requirements for CDF and the schedule appears to be sufficient for completion by the 2005 shutdown, and in fact may be ready well before that date. Specific suggestions were made to help make the EVB a better tool and to ease the commissioning process. The committee looks forward to the lack of BUSY deadtime.