



Fermi National Accelerator Laboratory

FERMILAB-Conf-92/310

Run Control and Resource Management in the DØ Run Time System

**B. Gibbard
for the DØ Collaboration**

*Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510*

October 1992

Talk presented at the *Conference on Computing in High Energy Physics - 1992*,
Annecy, France, September 21-25, 1992

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

██████ Talk presented at Conf. on Computing in
High Energy Physics - 1992 (CHEP92), Annecy, France
Sept. 21-25

Run Control and Resource Management in the D0 Run Time System*

B. Gibbard

Brookhaven National Laboratory

Upton, New York, USA, 11973

For the D0 Collaboration[†]

The D-Zero run time system consists of a number of distinct subsystems which perform the basic functions of triggering, digitizing, data gathering, and data recording. The allocation and high level coordination of the elements in this system are accomplished by a server task which maintains and consults a detailed model of the system. The strengths and weaknesses of this method of operation and the underlying data taking architecture are discussed.

Introduction

D-Zero is a large general purpose detector which began operation at the Fermilab Tevatron in the spring of 1992. It includes 7 major detector subsystems with approximately 115,000 channels of electronics. The detector's run time system consists of a multi-tier trigger integrated with a high performance data acquisition system. The original specification was that it be capable of digitizing and collecting 300 KBytes events at rates of 200 to 400 Hz. These events are then subjected to high level software filters capable of reducing the rate to approximately 2 Hz which is recorded. The system must be capable of operating in multi-user mode during calibration and debugging and in a monolithic mode during physics data taking. These requirements make the allocation of detector resources and the control of runs relatively complex. An introduction to the elements which must be allocated and controlled is useful to an understanding of this process.

*This work was supported in part by the U.S. Department of Energy and the National Science Foundation.

[†]The D0 collaboration includes: Universidad de los Andes (Colombia), University of Arizona, Brookhaven National Laboratory, Brown University, University of California at Riverside, CBPF (Brazil), CINVESTAV (Mexico), Columbia University, Delhi University (India), Fermilab, Florida State University, University of Hawaii, University of Illinois at Chicago, Indiana University, Iowa State University, Lawrence Berkeley Laboratory, University of Maryland, University of Michigan, Michigan State University, Moscow State University (Russia), New York University, Northeastern University, Northern Illinois University, Northwestern University, University of Notre Dame, Panjab University (India), IHEP-Protvino (Russia), Purdue University, Rice University, University of Rochester, CEN-Saclay (France), SUNY at Stony Brook, Superconducting Supercollider Laboratory, Tata Institute of Fundamental Research (India), University of Texas at Arlington, Texas A & M University.

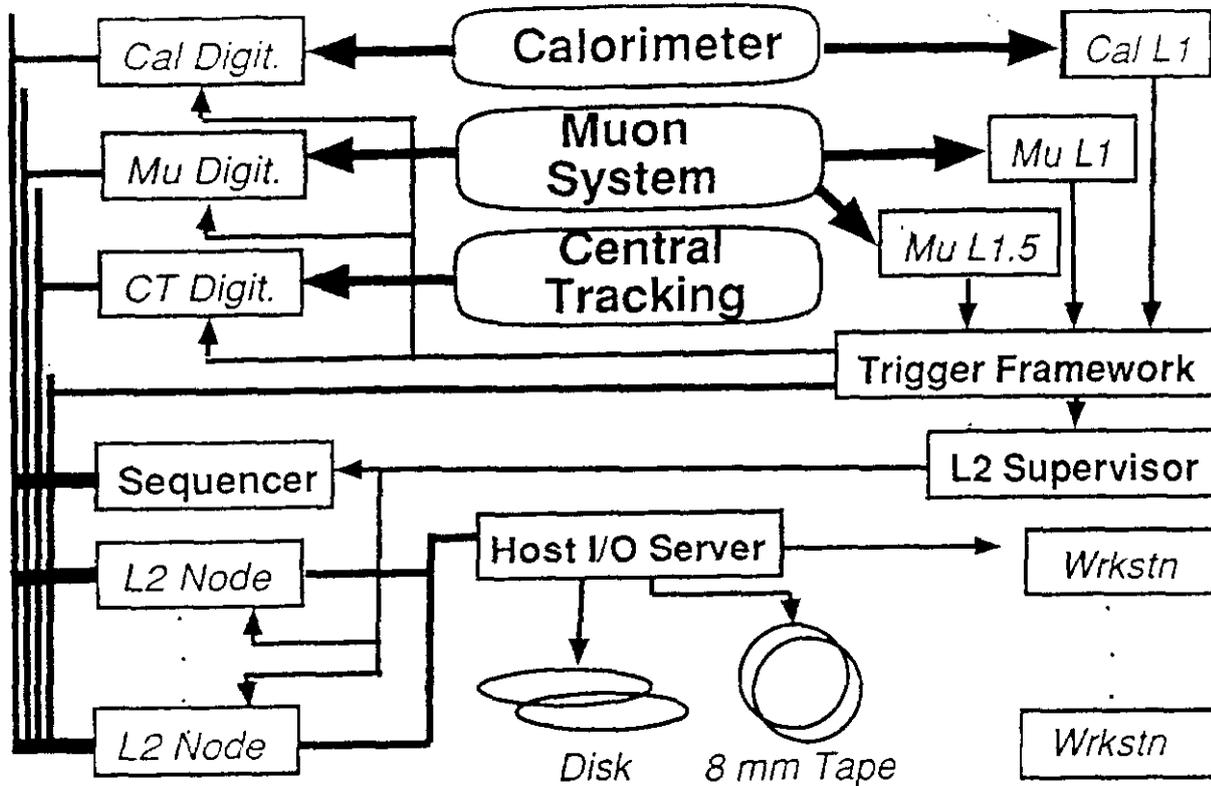


Figure 1. Schematic of the D0 data taking system.

Hardware Trigger

Level 1/1.5 triggers are logical combinations of a number of 256 available "trigger terms".¹ These trigger terms consist of information regarding physics signatures observed in the detectors. They include counts of EM and Jet Towers above E_t thresholds, missing and total scalar E_t above thresholds, and the presences of track-like hit distributions in muon drift chambers. Most of the trigger terms have settable thresholds of various sorts associated with them.

The trigger framework which coordinates the Level 1/1.5 trigger is capable of supporting 32 independent triggers as described above and labels each event with a 32 bit trigger mask indicating which triggers fired. The trigger framework deals with the digitizing system as 32 independent geographic sectors. It is capable of using any of the 32 Level 1 triggers to initiate readout of any combination of these 32 geographic sectors.

Digitizing System

The VME based digitizing system is housed in 80 crates, each assigned to one of the 32 geographic sectors described above. There are a variety of settable parameters associated with each digitizing crate including the digitizing mode, zero suppression parameters, pedestal values, etc. There are in addition a number of auxiliary devices located in VME crates including pulsers and timing/gating logic.

Data Collection

The data collection system, as indicated in Figure 1, consists of a set of high speed data cables, VME crate resident cable drivers, multi-port memories in filter processors which accept data, and sequencers which control their readout.² The cable driver in each crate contains a list processing engine which performs data collection local to that crate. There are 8 data cables each of which has a sequencer which controls readout for that cable. The sequencer instructs different crates to read out for different events depending on the value of the trigger mask for that event.

Data from the cables are received by multi-port memories located in a farm consisting of 50 VAX 4000/60 processors. The decision as to which cables will be read out for which events and to which of the processors an event will be routed is made by the Level 2 Supervisor and is based again on the value of the trigger mask.

Filtering

Once the event is fully assembled in the address space of one of the processors, software criteria can be applied to further determine if the event is of interest. This decision-making is organized around a set of "Tools". When a tool is invoked it calculates some quantities, makes a comparison and then returns either a pass or fail result for that event. Tools associated with the identification and measurement of jets, muons, electrons, photons, and missing E_t have been developed. For a particular filter certain tools with associated parameters are applied in a sequence. This list of tools and parameters is called a script. The system is capable of working with up to 128 such filter scripts at a time. The decision of which scripts should be applied to an event is determined by the value of its triggers mask. Each applied script is run either until one of the tools fails the event or the entire filter has been satisfied. At the completion of the Level 2 filtering a 128 bit filter mask is added to the event indicating which of the filters were satisfied. If the event has any set bits in the filter mask it will be retained and passed on to the host computer for recording.

Data Recording

The data logging task on the host directs events to various recording streams on the basis of their filter masks. The logging task can also direct events for express line analysis on the basis of this mask. The express line is a small farm of processors capable of doing the full reconstruction of 10 per cent of the events in near real time.

Operation

Control of the D0 data taking spans the various run time subsystems, configuring them for coordinated activity. This function is performed by a task running in one of the host cluster VAXes as a detached server to which clients establish connections in order to request configurational and operational changes in the detector. The coordinating task itself establishes connections to the various run time subsystems to down load configuring parameters and to request operational changes. This task maintains a model of the run

time system which contains the values of downloaded parameters as well as the intrinsic and established connectivity across subsystems. Comparisons of change requests with this model are used to determine what allocations should be made and what download sequences are required to reconfigure and/or alter operation of the detector. The basic unit of configurational request is a trigger. A trigger request is accompanied by a complete description of what trigger and filter criteria apply, what portions of the detector should be read out, the values of associated pulsers and gating logic, and the disposition of the events among the various recording streams and the express line. The complete description must be consistent with the current model of the run time system for the request to be accepted.

Support for multi-user operation was an important original specification of this system. This mode of operation is primarily used for calibration, testing and debugging of the detector while physics data taking is most usually done monolithically. Allowing for either the shared or exclusive use of detector elements facilitates flexible and efficient utilization of the detector during such operations. Multi-user operation consists of independent simultaneous runs including separate triggers, filters and recording streams.

The simultaneous runs which occur in multi-user operations are only directly visible at the highest level of the run time system. At the lower levels, such as triggering, digitizing, data collection and filtering, events are each treated independently on the basis of their trigger masks. The starts and stops of individual runs are seen as pauses in the data flow during which parameters are changed. It is only at the highest level, where the data logging task in the host I/O server distributes events on the basis of their filter mask to the runs and streams of the various individual client users, that the multi-user nature of operations is clearly visible.

Conclusions

The resource allocation and run control of the D0 system is characterized by centralized configuration of a data driven-like architecture involving distinct subsystems and dedicated busses. It features logical rather than physical partitioning for multi-user operation. The sharing of data path resources during multi-user operation does result in pauses in data flow as the system adapts to new users and occasions of less than optimal load balancing between simultaneous users. In general the run time system is performing well in both monolithic and multi-user (typical 4 to 6 users) operation and all design specification goals are currently being approached and in some cases have already been exceeded.

References

- [1] M. Abolins, D. Edmunds, P. Laurens, J. Linnemann, B. Pi, "A High Luminosity Trigger Design For The Tevatron Collider Experiment in D0," *IEEE Transactions on Nuclear Science*, Vol. 37, No. 1, 1989, pp. 384-389.
- [2] D. Cullen-Vidal, D. Cutts, J. Hoftun, D. Nesci, C. Johnson, R. Zeller, "D0 Level-2/Data Acquisition; The New Generation," presented at the Conference on Computing in High Energy Physics, Tsukuba, Japan (March, 1991).