The MINOS Data Acquisition System

A. Belias, G. J. Crone, E. Falk Harris, C. Howcroft, S. Madani, T. C. Nicholls, G. F. Pearce, D. E. Reyna, N. Tagg, and M. A. Thomson

Abstract—The MINOS long-baseline neutrino experiment consists of two detectors separated by 730 km. Both are equipped with identical data acquisition (DAQ) systems, based on continuous, dead time free readout. Data are read from the untriggered front-end electronics by VME single board computers and transferred across high-speed PCI data links for consolidation by data routing processors. An array of Linux computers selects events of interest using software-based trigger algorithms. We present the design of the DAQ system and report on experience gathered during early operation of the experiment.

Index Terms—Data acquisition, data buses, computer networks, programming.

I. INTRODUCTION

RECENT analyses of solar, atmospheric, and accelerator neutrino data from a number of experiments [1]–[5] have presented compelling evidence for the existence of neutrino oscillations. The MINOS experiment [6], [7] is designed to perform a number of precision measurements of oscillation phenomena over a long baseline with an accelerator-based neutrino source.

The MINOS experiment, currently under construction, consists of two underground detectors situated 730 km apart. A new beamline will produce neutrinos from 120 GeV protons extracted from the Fermilab Main Injector. Passing through the near detector, also situated at Fermilab, the neutrino beam propagates through the earth before reaching the far detector located in the Soudan mine, MN (see Fig. 1). Neutrino oscillations in flight will yield a difference in the absolute rate and energy distributions observed by the two detectors.

The far detector was completed in July 2003 and the near detector installation is scheduled to finish in October 2004, prior to commissioning of the neutrino beam in early 2005. A third, smaller detector module has been employed since 2001 in an extensive programme of test and calibration measurements at the
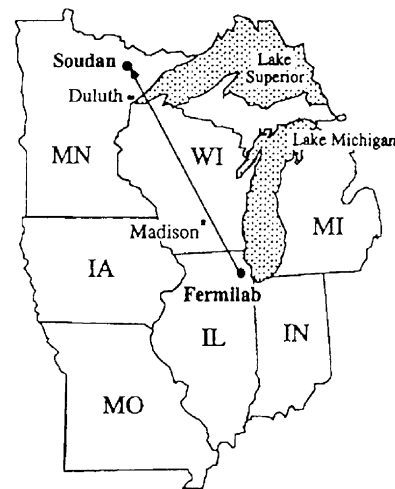
Fig. 1. The MINOS neutrino beam baseline, which stretches 730 km from Fermilab, IL, to the Soudan mine, MN. The beam passes through the earth between the two locations and reaches a maximum depth of approximately 10 km midway.

European Organization for Nuclear Research (CERN) Proton Synchrotron.

II. THE MINOS DETECTORS

The near detector has a total mass of 0.96 kt and is situated approximately 1 km downstream of the neutrino source at a depth of 100 m below ground level. The far detector has a total mass of 5.4 kt and is situated at a depth of 713 m, where the rock overburden provides effective shielding from background cosmic ray events.

The detectors are of a magnetized tracking calorimeter design, with vertical planes of alternating octagonal steel absorber plates and plastic scintillator strips. While the segmentation of both detectors is the same, the geometry of each is optimized to the incident particle flux. The near detector has 282 planes, not all populated with scintillator, of size 3.8 m high by 4.8 m wide, while the far detector has 484 fully populated planes of size 8 m × 8 m. Fig. 2 shows a schematic view of one half of the far detector.

Optical fibers transport scintillation light from the planes to multianode photomultiplier tubes (PMT). At the far detector, eight fibers are multiplexed into one PMT pixel, yielding a total of 23 232 PMT pixels. In the near detector some fibers are multiplexed four-to-one, while others are not multiplexed, leading to 11 616 pixels in total. The output signals from each pixel are digitized and time-stamped by custom-built, VME-based front-end electronics.
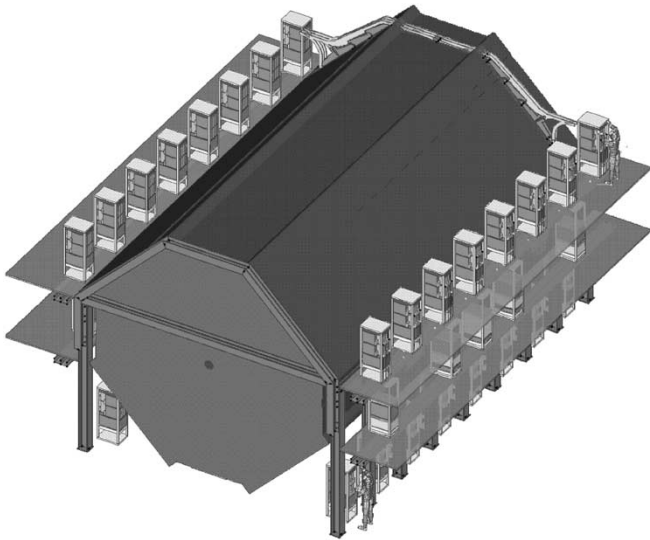
Fig. 2. Schematic view of one half of the MINOS far detector, showing the octagonal steel/scintillator plane assemblies and support structure, the optical fiber transport paths and the racks of readout and data acquisition electronics.

The event rates expected at the two detectors are considerably different. At the far detector the rate is uniformly of the order of 1 MHz and is dominated by single, uncorrelated hits arising from PMT noise and radioactive decay. At the near detector, the rate when there is no neutrino beam production (out-of-spill) is approximately 0.5 MHz, due to PMT noise, radioactive decay and cosmic muons; during the 10 μs beam spill period, the instantaneous hit rate is approximately 60 MHz. The design of front-end electronics has therefore been individually optimized for each detector.

The far detector front-end electronics operate in a free-running mode. All pixels on a PMT are digitised by a 14-bit analog-to-digital converter (ADC), when the dynode signal from the PMT exceeds a programmable threshold. The sampled channels are time-stamped by an incrementing 30-bit counter with 1.5 ns precision. For each channel, a pedestal value is subtracted from the ADC sample, and the data sparsified by comparison with a programmable threshold table. The sparsified data are written as 8-byte values into output buffers for readout by the data acquisition (DAQ) system, yielding a total raw data rate of 8 MB/s. All front-end modules are synchronised to a common, high-precision 40 MHz clock reference, which is distributed optically with control signals by a central timing unit. The clock is also used to synchronise the readout of data from the front-end electronics by the DAQ system through the generation of VME interrupts at a programmable rate. The clock is derived from the Global Positioning System (GPS), which also provides an absolute time standard for other detector systems.

In the near detector front-end electronics, all PMT pixels are digitised every 19 ns, without dead time, by an autoranging ADC to give an 11-bit floating-point value. A beam spill signal triggers an acquisition period of 10 μs, while out of spill a discriminated dynode signal triggers a 150-ns acquisition period. Once the trigger is complete, all hits stored during the acquisition are processed; pedestal subtraction and linearization

are performed for each channel in a single step via a lookup table, yielding a 16-bit integer value. The hits are sparsified, time-stamped and stored in a readout buffer. The hits are 8 bytes in size, leading to a total raw data rate of 4 MB/s out of spill. The near detector timing system is synchronised to the accelerator clock system, distributing a 53 MHz reference, a spill gate and other control signals to all front-end modules. Again, the clock is used to synchronise the readout of data by the DAQ. The phase of timing signals from a GPS system relative to the accelerator clock is also measured, allowing accurate reconstruction of the absolute event time.

III. THE DAQ SYSTEM

The MINOS DAQ system reads out the front-end electronics in an untriggered, dead time free manner. The data from all front-end modules are consolidated and transferred to an array of trigger processors, where flexible software algorithms time sort the data and select events of interest based on spatial and temporal clustering of hits. In spite of the difference in detector topology, channel count and front-end electronics at the two detectors, both DAQ systems are functionally identical. The DAQ system is constructed entirely from commercially available components. Fig. 3 shows the layout of the DAQ system, using the far detector as an example. The components shown therein are explained in Section III-A–F.

A. System Architecture

The primary data transfer mechanism in the MINOS DAQ system is the Creative Electronics Systems [8] PCI Vertical InterConnect (PVIC) system. This is a multinode, high-speed PCI-to-PCI interconnect, which provides transparent data transfer and messaging capabilities. A maximum of fifteen nodes can be connected in one PVIC chain (branch), with a number of physical layer options available: a GTL bus for short-range, intracrate connections, a PECL differential bus (DIFF) for medium length chains (15 m at 66 MHz PVIC bus speed, 30 m at 33 MHz) and a 1.5 Gbit/s serial optical link (OPT) for longer distances (200 meters on 850 nm multimode fiber). The PVIC is available in a number of form factors, including PCI short-format card, PCI Mezzanine Card (PMC) and 6U VME.

A Direct Memory Access (DMA) engine on each PVIC node allows autonomous block transfers at full PCI speed (132 MB/s theoretical at 66 MHz), while interrupt-linked mirrored memory provides low-latency, deterministic messaging between nodes. The PVIC system was selected due to its high throughput with minimal CPU overhead and, with a common high level Application Programming Interface (API) available for a number of operating systems, its relative ease of programming.

In the DAQ system, up to four Read Out Processors (ROPs), responsible for acquiring data from the front-end electronics, are daisy-chained into a 30-m DIFF PVIC branch. At the end of the chain, the PVIC is converted to the OPT layer, which transfers data over a distance of up to approximately 75 m to the Branch Readout Processor (BRP). Four BRPs consolidate data from each of the four input PVIC branches. A DIFF PVIC chain forms the output branch, across which the BRPs transfer
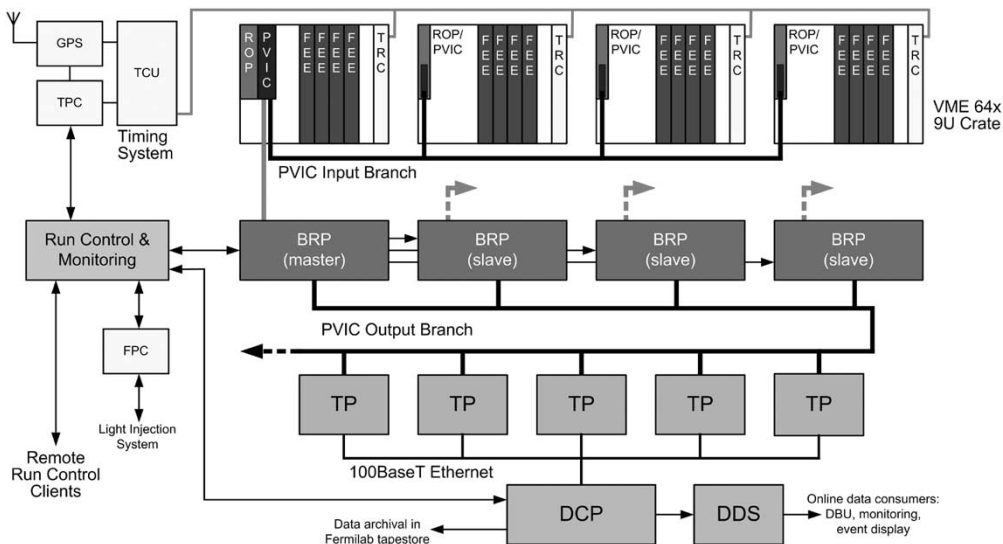
Fig. 3. The architecture of the MINOS DAQ system. The far detector front-end electronics and clock system are shown as an example. For clarity, only one of the four PVIC input branches is illustrated. Abbreviations are explained in the text.

data to an array of Trigger Processors (TPs), where data of interest are selected for storage. A maximum of eleven TPs can be supported on the output branch; five are typically deployed.

The DAQ system was designed to handle considerably higher data rates than those presented in Section II. Sixteen ROPs are present in the far detector system, each designed to support a mean data rate of 2.5 MB/s from the front-end electronics. At the near detector, there are eight ROPs, each handling 5 MB/s of data. Each input branch is therefore required to support 10 MB/s data transfer to the BRP, while the output branch must sustain 40 MB/s to the TPs. The expected output rate from the TP array is sufficiently low (100 kB/s maximum) that 100 Mbit Ethernet is sufficient for transferring data downstream.

B. The Read Out Processor

The Read Out Processor consists of a CES RIO3 single board computer, with a 400-MHz PowerPC CPU and 64 MB of memory, running the Wind River Systems [9] VxWorks 5.4 real-time operating system. There is one ROP per 9U VME 64x front-end crate. The ROPs run as diskless nodes, booting the operating system via File Transfer Protocol (FTP) across the local network.

The ROP reads data from alternating buffers in the front-end electronics, allowing digitization to continue during readout and thus incurring no dead time. The readout is synchronised across all front-end crates by the Timing system Central Unit (TCU).[1]

This fans out clock and control signals to the Timing Receiver Card (TRC), which generates VME interrupts at a programmable rate (typically 40 Hz). Upon receipt of the interrupt, the ROP transfers data using VME Multiplexed Block Transfers (MBLT) from the currently active buffer of each front-end module into memory.

The data from all front-end modules in a crate, known as time blocks, are buffered in the ROP memory and assembled into

convenient units, termed time frames, which are typically one second in length. Each time frame is labeled with header information, including the full date and time, crate number and time frame identifier. Data blocks from monitoring tasks or online calculation of pedestal values are appended to the time frame as necessary.

Time frames are buffered by the ROP until requested by the BRP using PVIC mirrored memory messaging. The time frame is subsequently transferred into the memory of the BRP by DMA across the PVIC.

The TRC interrupt rate and time frame length are both variable quantities, which can be separately tuned to optimize the efficiency of both the VME MBLT and the PVIC DMA transfers. Furthermore, both types of transfer occur autonomously under the control of dedicated logic and therefore incur minimal CPU overhead on the ROP, independent of the data throughput.

C. The Branch Readout Processor

The Branch Readout Processor is responsible for receiving time frames of data from all ROPs on a given input branch, assembling the data in memory and transferring it to the TPs for processing. The BRP takes the form of an Intel Pentium-based PC running the GNU/Linux operating system.

One BRP functions as a master, coordinating its operations with those of the other three slave BRPs via TCP/IP. The master BRP instructs the BRPs to request a given time frame from the ROPs. The requests are issued sequentially to the ROPs on an input branch via PVIC mirrored-memory messages, causing the time frame to be transferred via DMA into the memory of the BRP. By transmitting the appropriate target address to each ROP in turn the time frame is assembled contiguously in the BRP memory.

All BRPs assemble and buffer data from the ROPs in parallel. Once complete, the master BRP instructs each BRP in turn to transfer the time frame to a TP via the output PVIC branch. Again, by appropriate choice of the DMA target address, the entire time frame is assembled contiguously in the memory of

---

[1]The detailed implementation and nomenclature of the timing systems differs at the near and far detectors, although the basic principles of operation are the same. The far detector system is described here.

the TP. This "construction by transfer" methodology avoids the need for copying data within the memory of any component in the data path, allowing system performance to be maximized.

The master BRP also controls the operations of the TPs, communicating with them via TCP/IP and keeping track of the number of time frames currently being queued and processes in each TP. The distribution of time frames to the TPs is controlled by the master BRP using selectable least-loaded or round robin algorithms.

D. The Trigger Processor

The Trigger Processor performs a number of flexible processing tasks on complete time frames to select candidate events. The TP also takes the form of a Pentium-class machine running the GNU/Linux operating system.

Upon receipt of a time frame, the hits are time sorted. A Quicksort algorithm [10] takes advantage of local structure in the data to sort the hits on a crate-by-crate basis. Having sorted the data, the TP processing then searches for a range of event types of interest. A light injection system [11] periodically injects controlled pulses of light into the detector in order to provide a stable calibration reference. A direct signal from this system is fed into the front-end electronics and used to define a time reference for each pulse. All hits within a well-defined time window around this reference are designated by the TP as a so-called flasher event. These are identified and passed to a dedicated processing task, which analyzes the detector response to the light pulses and generates summary data, which are used offline to determine calibration constants. This module also retrieves information about such events from the Flasher PC (FPC), which controls the light injection system. The calibration summary data are periodically queued to an output buffer for transmission downstream.

In parallel with the identification of flasher events, the TP scans the time frame to locate events of physics interest. This is done on the basis of spatial and temporal clustering of hits in the detector. Time clusters of hits, typically within a window of 200 ns, are identified as candidate events and passed through a trigger test. This test takes the form of a logical OR of a number of algorithms; the primary algorithm for MINOS requires at least m planes out of any group of n contiguous planes in the detector contain one or more hits. All hits observed within a configurable window of the trigger cluster are written to the output buffer; this includes hits up to, typically, 5 μs prior to the trigger to assist with offline interpretation of events. Detailed studies of simulated MINOS data have demonstrated [6] that the values $m = 4$ and $n = 5$ provide an optimal trigger. At the near detector, all data flagged as occurring within a beam spill are written out without further selection.

Additional algorithms, such as null and random triggers, are implemented in order to facilitate calibration of the detector and continuous sampling of background conditions. The TP also identifies data, such as ROP monitor blocks, which require no processing to be performed and places them directly into the output buffer.

This approach to triggering imposes minimal bias on the data by reducing only uncorrelated single hits arising from noise and background sources, while greatly reducing the data rate and af-fording maximum flexibility through straightforward reconfiguration of the trigger software.

The TP also performs a number of data integrity checks, ensuring that the time frames from all ROPs are correctly synchronised and structured. Monitoring statistics, such as the rate of singles hits per crate, are also accumulated and placed in the output buffer. All data output by the TP are transmitted via TCP/IP to the data collection processor (DCP).

E. The DCP

The DCP receives and merges the output data streams from all TPs. The event stream is ordered and formatted as a ROOT [12] tree before being written to a file on disk.

In order to avoid trigger inefficiencies at the boundaries of time frames, the ROPs create an overlap by copying one time block from the end of a time frame into the start of the next. This leads to a finite but small probability that a candidate event can be found by two TPs in adjacent time frames. The DCP is therefore configured to identify and filter out duplicate events.

A background process copies completed data files to the Fermilab mass storage facility for archival and offline analysis. The active output file is also made available to other online data consumers, such as the database update (DBU) task, online monitoring and event displays, via the data distribution system (DDS).

F. Run Control and Monitoring

The Run Control and Monitoring system provides overall control of the DAQ system. All components implement a well-defined state model and, under the command of Run Control, make controlled transitions between states. These state transitions define a datataking run, which represents data acquired with a particular detector configuration.

Run Control passes configuration parameters to components during the run-prepare phase, allowing the operation of the DAQ and front-end electronics to be steered appropriately. It is possible to define complex sequences of runs in which normal datataking is interspersed with special calibration and monitoring runs. Run Control can execute these sequences autonomously, allowing unattended operation of the experiment.

The Run Control and Monitoring system is implemented using a client/server model. A single server controls the DAQ, while multiple clients can be connected to the server via TCP/IP sockets, allowing remote operations. A mechanism for handover of control from one client to another is provided in order to avoid contention.

The client is based on the ROOT framework and presents a GUI to the operator. This displays the current state of the DAQ, allows the operator to start and stop runs and sequences and displays variable-priority status messages from all components. The server also receives and accumulates monitoring statistics from various DAQ components, which are presented to the operator in the form of instantaneous values and histograms.

IV. PVIC PERFORMANCE CHARACTERIZATION

Fig. 4 shows the DMA transfer performance of the PVIC branches, measured in the DAQ prototype system. Both input
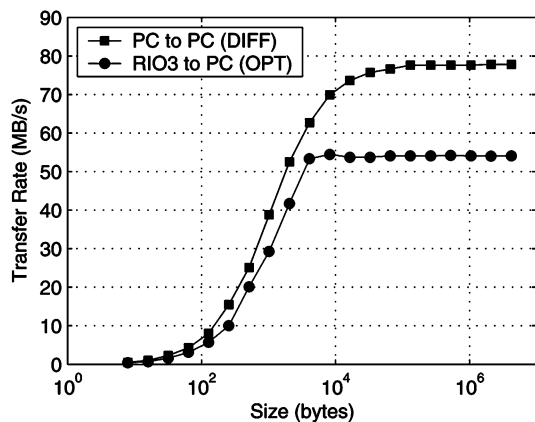
Fig. 4.   Transfer rate measured on RIO to PC (OPT) and PC to PC (DIFF) links as a function of transfer size. Both demonstrate adequate performance for transfer sizes greater than 1 kB.
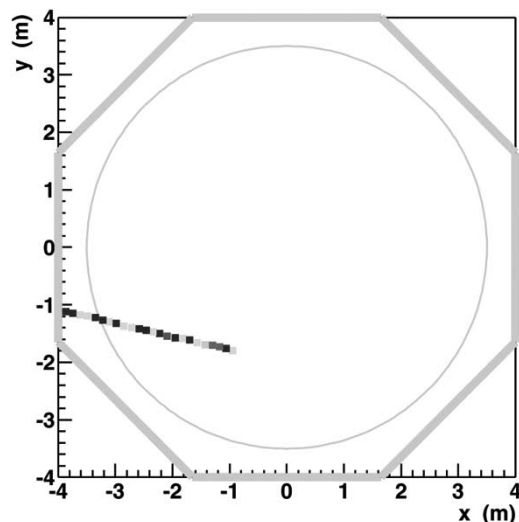

Fig. 5.   Axial view of an event observed in the MINOS far detector. The topology and timing of the event is consistent with an upward-going muon originating from a neutrino passing through the earth.

and output branches are observed to demonstrate adequate performance for transfer sizes greater than approximately 1 kB. The RIO3 to PC connection using the OPT layer (input branch) saturates at a lower maximum rate due to the 33-MHz clock speed, while the PC to PC link over the DIFF layer (output branch), running at 66 MHz, is limited by the performance of the PCI bus in the PC architecture.

V. INSTALLATION AND OPERATIONAL EXPERIENCE

The far detector DAQ system was successfully installed and commissioned during 2001/2002. The system comfortably handles the typical data rates observed in the detector, where 10 MB/s of raw data reduces to approximately 5 kB/s output rate from the DCP, yielding a typical event rate of 8 Hz. A typical flasher event rate of 50 Hz yields an additional 2 MB/s of data. Under these conditions, each TP (with a 700-MHz Pentium III processor) requires approximately 3 s to process each time frame, of which 0.5 s is consumed by flasher data analysis.

The far detector DAQ system is now in routine operation, supporting detector commissioning and nonbeam datataking Fig. 5 shows an event in the far detector observed during 2002.

VI. CONCLUSION

A novel DAQ system for the MINOS experiment has been presented. A modular design based entirely on commercially available technologies has been adopted, allowing functionally identical DAQ systems to be interfaced to the front-end electronics specific to each of the two detectors. The system enables continuously sensitive acquisition of data from the detectors while incurring no dead time. A fully software-based triggering scheme allows flexible selection of candidate events and provides integrated calibration and monitoring capability. The DAQ system has been successfully deployed at the far detector and proven to meet the requirements of the experiment.

REFERENCES

[1] Y. Fukuda *et al.*, "Evidence for oscillation of atmospheric neutrinos," *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, Aug. 1998.
[2] W. W. M. Allison *et al.*, "The atmospheric neutrino flavor ratio from a 3.9 fiducial kiloton-year exposure of Soudan 2," *Phys. Lett. B*, vol. 449, pp. 137–144, Mar. 1999.
[3] S. Fukuda *et al.*, "Determination of solar neutrino oscillation parameters using 1496 days of super-kamiokande-I data," *Phys. Lett. B*, vol. 539, pp. 179–187, July 2002.
[4] Q. R. Ahmad *et al.*, "Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory," *Phys. Rev. Lett.*, vol. 89, p. 011 301, July 2002.
[5] M. H. Ahn *et al.*, "Indications of neutrino oscillation in a 250 km long-baseline experiment," *Phys. Rev. Lett.*, vol. 90, p. 041 801, Jan. 2003.
[6] "The MINOS Detectors Technical Design Report," The MINOS Collaboration, Fermilab, NuMI-L-337, 1998.
[7] K. Lang, "MINOS detectors for neutrino interactions," *Nucl. Instrum. Meth.*, vol. A461, pp. 290–292, Apr. 2001.
[8] Creative Electronics Systems, Petit-Lancy, CH, Switzerland. [Online]. Available: http://www.ces.ch
[9] 500 Wind River Way. Wind River Systems, Inc., Alameda, CA. [Online]. Available: http://www.wrs.com
[10] B. W. Kernighan and D. M. Ritchie, *The C Programming Language*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1988, p. 87.
[11] P. Adamson *et al.*, "The MINOS light injection system," *Nucl. Instrum. Meth.*, vol. A492, pp. 325–343, Oct. 2002.
[12] R. Brun and F. Rademakers, "ROOT—An object oriented data analysis framework," *Nucl. Instrum. Meth.*, vol. A389, pp. 81–86, Apr. 1997.