ELSEVIER

Nuclear Instruments and Methods in Physics Research A 453 (2000) 397–404

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A

www.elsevier.nl/locate/nima

The STAR level-3 trigger system

J.S. Lange[a,*], C. Adler[a], J. Berger[a], M. Demello[b], D. Flierl[a], J. Landgraf[c], M.J. LeVine[c], A. Ljubicic Jr.[c], J. Nelson[d], D. Roehrich[e], J.J. Schambach[f], D. Schmischke[a], M.W. Schulz[g], R. Stock[a], C. Struck[a], P. Yepes[b]

[a]*University of Frankfurt, August-Euler-Straße 6, D-60486 Frankfurt, Germany*
[b]*Rice University, Houston, Texas 77251, USA*
[c]*Brookhaven National Laboratory, Upton, New York 11973, USA*
[d]*University of Birmingham, Birmingham B15 2TT, United Kingdom*
[e]*University of Bergen, Allegaten 55, 5007 Bergen, Norway*
[f]*University of Texas, Austin, Texas 78712, USA*
[g]*University of Heidelberg, Philosophenweg 12, 69120 Heidelberg, Germany*

Accepted 20 June 2000

Abstract

The STAR level-3 trigger is a MYRINET interconnected ALPHA processor farm, performing online tracking of $N_{\text{track}} \geq 8000$ particles ($N_{\text{point}} \leq 45$ per track) with a design input rate of $R = 100$ Hz. A large-scale prototype system was tested in 12/99 with laser and cosmic particle events.

Keywords: Trigger; Data acquisition; Nucleus–nucleus collisions; Proton–proton collisions

1. Introduction

The RHIC accelerator at Brookhaven National Laboratory, USA, will investigate Au + Au collisions with $\sqrt{s} \leq 200$ A GeV and p + p collisions with $\sqrt{s} \leq 500$ GeV. The STAR experiment [1] is a large scale, cylindrical, symmetric 4π-detector. Physics data taking will start in 2000 with a full size Time Projection Chamber (TPC), $R_{\text{in}} = 0.6$ m, $R_{\text{out}} = 2$ m) with 24 sectors, 6912 pads each. TPCs are specifically suitable for detecting high-density charged particle fluxes in high-multiplicity nucleus–nucleus events.

* Corresponding author.
E-mail address: soeren@bnl.gov (J.S. Lange).

2. Architecture

2.1. The STAR trigger architecture

The STAR trigger system is subdivided into 4 hierarchic levels. The level-0 output rate is 10^5 Hz, levels-1 and -2 as well as coincidence with the TPC gating reduce the rate by one order of magnitude each. Level-3 trigger is supposed to reduce an input rate of 10^2 Hz to the final DAQ rate of $R_{\text{tape}} = 1$ Hz at an expected TPC event size of $\simeq 15$ MB. Task examples for the level-0/-1/-2 trigger stages are (a) selection of central and peripheral Au + Au events based upon multiplicity (function of impact parameter) and (b) rejection of beam-gas events with a vertex far from the interaction point. The tasks of level-3 trigger are event selections

0168-9002/00/$ - see front matter
PII: S0168-9002(00)00664-1

based upon the online reconstructed track parameters of each particle. Several applications have been proposed for Au + Au collisions, being on the one hand high p_T trigger applications as enrichment of heavy (anti-)fragments (e.g. $\overline{He^4}$), STAR EMC (Electromagnetic Calorimeter) calibration using tagged high p_T π^-, QCD hard parton scattering (leading high p_T hadrons in jets), and c and b quark decays (e.g. high p_T leptons). On the other hand, online invariant mass reconstruction of J/ψ, $\Upsilon \to e^+e^-$ is proposed, as suppression of $c\bar{c}$ and $b\bar{b}$ production is commonly regarded as one the most promising signatures of the quark-gluon plasma. Additionally, for p + p collisions, a level-3 trigger algorithm for filtering of ≤ 700 pile-up events in the TPC (at highest Luminosity $\mathscr{L} = 2 \times 10^{31}$ $cm^{-2}s^{-1}$) per one level-0 trigger is being developed.

2.2. The STAR DAQ architecture

Level-3 trigger architecture is closely related to the STAR DAQ architecture, in which one VME crate is mapped onto each physical TPC sector (two TPC sectors in the first stage), containing a Sector Broker, i.e. Motorola MVME-2306 VME board, carrying a PowerPC 604 (300 MHz, VxWorks), as the TPC sector master controller. The Sector Broker carries a MYRINET interface (cf. Section 5) for (a) raw data transfer to the main STAR event builder and (b) connection to the level-3 track finder CPUs. Moreover, each DAQ crate also contains six VME receiver boards, each carrying three mezzanine cards with

- one Intel i960 CPUs (33 MHz, VxWorks) for (a) data formatting and (b) running the level-3 cluster finder,
- 4 MB of dual-ported VRAM for buffering and pipelining of raw data of 12 events.

Further details about the hardware are described elsewhere [2,3].

2.3. The STAR level-3 trigger architecture

Level-3 trigger scheme consists of two main components:

- The *sector level*-3 part ("Sector-L3") is mapped onto one physical TPC sector. It contains (a) level-3 cluster finder (cf. Section 3) and (b) the level-3 track finder (cf. Section 4). Data transfer of cluster data and track data is performed by MYRINET (cf. Section 5). Typical data sizes per event are $\simeq 85$ MB for raw data, $\simeq 15$ MB after zero suppression (DAQ taping event size), $\simeq 3$ MB after cluster finding and $\simeq 0.5$ MB after track finding.
- The *global level*-3 part ("Global-L3") consists of $1, \ldots, n$ ($n = 1$ in the first stage) master CPUs for the whole STAR TPC, collecting all track data via MYRINET and issueing the level-3 decision.

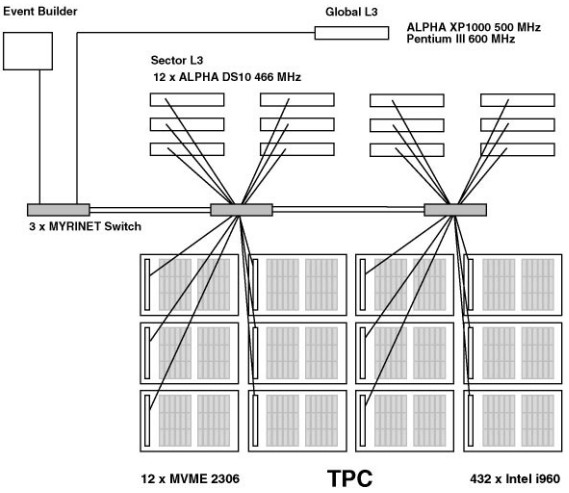

Fig. 1. STAR level-3 trigger system arhitecture as used in the system test 12/99(cf. Section 8). Event building of level-3 specific events has been performed locally on the Global-L3 CPU, integration into the STAR event building is foreseen.

Fig. 1 shows the schematic level-3 trigger architecture, as installed for the system test in 12/99 (cf. Section 8). The project development of level-3 trigger can be subdivided into two main stages.

- In the first stage (installed 12/99), for both DAQ and level-3 trigger system two physical TPC sectors are mapped onto one logical level-3 sector, which only contains one track finder CPU. Thus, all trigger rate design values are to be multiplied by a factor of $\frac{1}{2}$. Level-3 trigger will employ TPC data only, and the input trigger rate is $R \leq 25$ Hz.

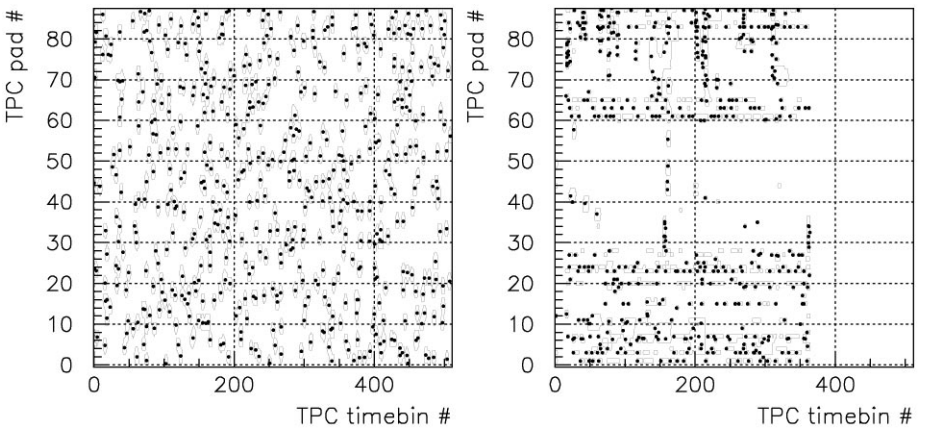

Fig. 2. Level-3 cluster finder example for one TPC sector. The black spots indicate the reconstructed centers-of-gravity. *Left*: 600 simulated clusters on the most inner TPC padrow, reconstructed in $\tau_{\text{cluster}} = 7.5$ ms on the Intel i960. *Right*: Beam-gas event, recorded with STAR in 07/99 (all padrows, view from top, TPC clock set to timebin $t_{\max} = 348$).

- In the second stage, envisaged for 2001, one TPC sector maps one level-3 sector, which then contains up to 4 parallelized track finder CPUs. Level-3 trigger will employ additional track information of SVT (Silicon Vertex Tracker) [4], and the final design value for the input trigger rate is $R = 100$ Hz (limited by TPC frontend readout rate).

3. Cluster finder

For a TPC readout, one ADC channel is indexed by a pad number ($r\varphi$-direction, e.g. 88 pads for the most inner padrow at $R_{\text{in}} = 0.6$ m) and a drift timebin number (z-direction, 512 timebins per pad). Clusters are continuous $(r\varphi, z)$ regions with an ADC value above threshold. In a first step, for each TPC cluster the center-of-gravity (weighted mean according to ADC values) is calculated to obtain particle hit xyz-coordinates.

The cluster finder algorithm runs on the Intel i960 CPUs, implemented on the DAQ receiver boards (cf. Section 2.2). The number of i960s is 18 per TPC sector, 432 for the whole TPC. Input to the cluster finder are zero-suppressed TPC raw data, stored in the VRAM buffer. The output cluster data, i.e. (a) cluster center-of-gravity and (b) cluster total charge (ADC sum), are sent via VME to the Sector Broker, which itself ships the data via MYRINET to the level-3 track finder CPU (expected data transfer rate of $\simeq 3$ MB/s per TPC sector).

The cluster finder time constraint is $\tau_{\text{cluster}} \leq 10$ ms (input rate $R = 100$ Hz). Benchmarks on the i960 were performed for 600 clusters (realistic Au + Au scenario) on the TPC's most inner padrow. The position resolution of $\Delta(r\phi) \simeq 37$ μm and $\Delta z \simeq 13$ μm could be obtained with an algorithm within $\tau_{\text{cluster}} = 7.5$ ms (absolute cluster finding efficiency $\varepsilon = 93\%$). The clusters and reconstructed centers-of-gravity are shown in Fig. 2 (*left*). If two clusters are merged, an additional deconvolution subroutine must be started, consuming 6.0% more CPU time than in case of two separated clusters. Fig. 2 (*right*) shows the reconstructed clusters for a STAR beam-gas event, recorded in 07/99. Different TPC clock timing lead to a maximum drift timebin $t_{\max} = 348$ in that case (x-axis).

4. Track finder

Monte-Carlo simulations[1] predict for a central Au + Au collision (impact parameter $b \leq 2.0$ fm)

[1] Monte-Carlo simulations were performed using the event generator HIJING 1.31, which is based on a QCD-inspired model for jet production [5].

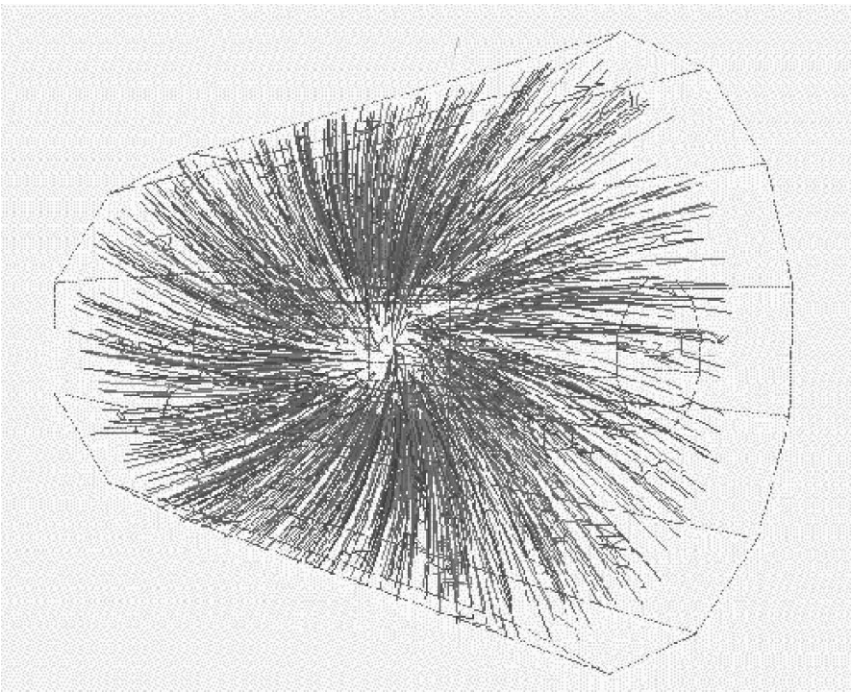

Fig. 3. Level-3 track finder example. 1000 simulated π^+/π^- particle tracks in the STAR TPC are shown. Tracking was performed on an ALPHA XP1000 within $\tau_{\text{track}} = 9.6\,\text{ms}$ per TPC sector.

that the track finder algorithm must be able to fit at least $N \simeq 400$ tracks per event per TPC sector, each consisting of $N_{\text{point}} \leq 45$ points (given by the number of padrows). This shall be referred as "Au + Au bench-mark event" hereafter.

The fast track finder algorithm has specifically been developed for level-3 trigger project [6]. It employs *conformal mapping*, i.e. a transformation of a circle[2] into a straight line, followed by a fit with a *follow-your-nose* method. A given space point (x, y) is transformed into a conformal space point (x', y') according to the equations $x' = (x - x_0)/r^2$ and $y' = (y_0 - y)/r^2$, using $r^2 = (x - x_0)^2 + (y - y_0)^2$. The transformation requires the knowledge of one point (x_0, y_0) on the track trajectory, either (a) the interaction point (vertex constraint for primary tracks) or (b) the first point associated with the track (no vertex constraint for secondary tracks). Fig. 3 shows an example of 1000 simulated level-3 trigger π^+/π^- particle tracks ($p_{\text{T}} \leq 1.0$ GeV/c). Fig. 4 shows the track finder efficiency as a function of p_{T} (*top*) and pseudorapidity η (*bottom*) for 50,000 Monte-Carlo generated tracks (independent of particle type). The efficiency of $\varepsilon \geq 90\%$ for $|\eta| \leq 1.2$ and $p_{\text{T}} \geq 400$ MeV/c is well suited for high p_{T} trigger applications (cf. Section 2.1), the p_{T} resolution being e.g. $\Delta_{p_{\text{T}}} = 14.9$ MeV/c (RMS) for $p_{\text{T}} = 500$ MeV/c.

The track finder time constraint is given by $\tau_{\text{track}} \leq \tau_{\text{buffer}} - \tau_{\text{cluster}}$, while $\tau_{\text{cluster}} \leq 10$ ms is given by the time being necessary for cluster finding (cf. Section 3). In the first project state τ_{buffer} is given by the buffer time of 12 pipelined events ($\tau_{\text{buffer}} = 12 \times 10$ ms), in the final stage by one event only ($\tau_{\text{buffer}} = 10$ ms). Timing benchmarks for different CPU platforms have been described in detail in Ref. [2]. The fastest available CPU platform is

[2] In the STAR solenoid magnetic field of $B = 0.5$ T charged particle tracks can be parametrised as helices, being visible as circles in an xy-projection.

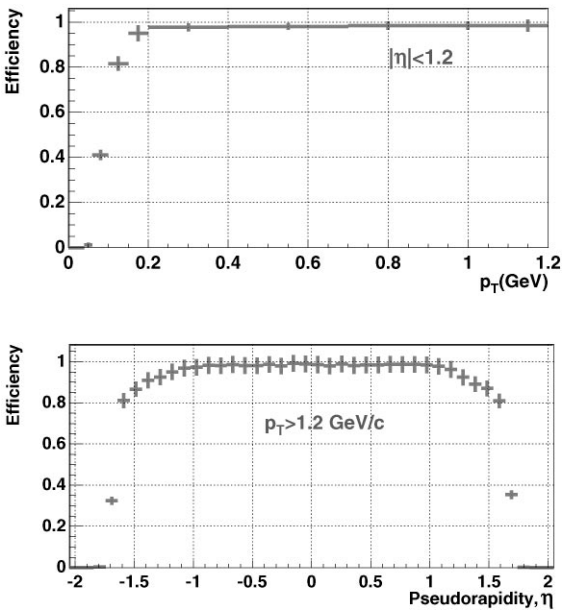

Fig. 4. Level-3 track finder efficiency as a function of p_T(*top*) and pseudorapidity η (bottom) for 50,000 Monte-Carlo generated tracks (assumed hit resolutions of $\delta(r\phi) = 500\,\mu m$ and $\delta z = 2$ mm).

the ALPHA 21264 (64 bit). For the Au + Au benchmark event, an ALPHA XP1000 (500 MHz) prototype machine gave a result of $\tau_{track} = 88$ ms, to be compared with $\tau_{track} = 135$ ms for e.g. a Pentium II (450 MHz). Fine tuning of initialization parameters (e.g. number of η/ϕ-slices for follow-your-nose search range) lead to a significant improvement of the processing speed ($\tau_{track} = 39$ ms). If additionally for each track, a dE/dx *truncated mean*[3] value is calculated, an additional time of $\Delta\tau = 8$ ms is needed.

The current system consists of 12 ALPHA DS-10 (466 MHz) machines. The ALPHA 21264 chip provides two additional significant advantages: Firstly, it is the first ALPHA chip with a hardware `sqrt()` function implementation (~ 30 `sqrt()` calls per track). Secondly, it has 128 bit wide memory access to 2 MB of level-2 cache on chip, important for fast "data digging" (only 64 bit wide access to external cache for Intel Pentium). Further ALPHA 21264 hardware details are described elsewhere [2,7]. Linux was chosen as operating system, with the kernel Linux 2.2.12 running stable on ALPHA. Each DS-10 machine is booting Linux diskless, console messages being routed via serial port and ethernet to any arbitrary terminal, and thus enabling a single user to remotely control the whole processor farm.

5. Network

Level-3 trigger system requires for each TPC sector a high bandwidth network connection between a PMC adapter (on the Sector Broker VME board) and a PCI adapter (in the Sector-L3 ALPHA). The estimated whole system throughput is $R = 52$ Mbytes/s (level-3 track data, Au + Au benchmark event), to be added to $R \simeq 15$ Mbytes/s DAQ data throughput. The two different point-to-point, full-duplex networks SCI and MYRINET were tested for application in the level-3 trigger. Due to missing DMA capabilities, [4] Gigabit Ethernet was not considered. Scalable Coherent Interface (SCI) [8] utilizes transfer of small data payload of 16 or 64 bytes at a high clock frequency of 200 MHz. Signals are transmitted via 18 pin differential ECL. Prototype D310 PCI-SCI interface are available from DOLPHIN [9], the C2D PMC-SCI interface was available from VMETRO [10] (product line stopped in 1999). MYRINET [11] supports variable data payloads ≤ 4 kB (LaNai interface chip operating at 33 MHz PCI clock). Signals are transmitted via two different cable types, i.e. SAN cable (PMC adapter to switch, 20 pin single-ended signal) and LAN cable (PCI adapter to switch, 37 pin Low Voltage Differential Signal 1.2 ± 0.4 V). Both PMC and PCI host adapters (32 bit, 2 Mbyte memory in the level-3 trigger configuration) are available from MYRICOM [12].

[3] A *truncated mean* is calculated by (a) ADC value into dE/dx transformation (e.g. gain calibration, pedestal subtraction) (b) list sorting and (c) calculation of the mean of the lowest 70% (tale truncation).

[4] Background DMA (Direct Memory Access) data transfers do not consume any CPU time (except for bus arbitration and interrupt handling). Thus, the processor is free for tasks as track finding or data formatting.

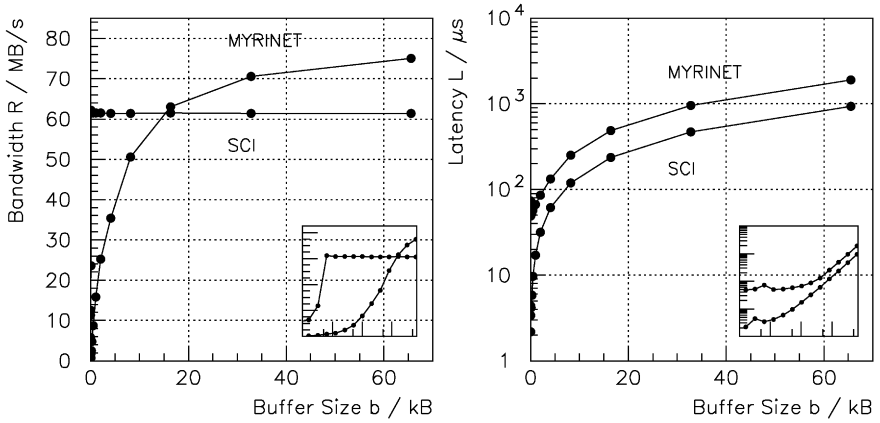

Fig. 5. Bandwidth (left) and Latency (right) as a function of buffer size for both MYRINET and SCI (Level-3 PCI-PCI point-to-point DMA, logarithmic buffer size axis in small pictures).

The basic architectures differ significantly, i.e. SCI utilizes a ring topology, MYRINET a switch topology. Thus, e.g. SCI requires the Sector Broker to carry two PMC adapters (different rings), as while MYRINET requires the existence of (16-port) switches between the PMC and the PCI side. Fig. 5 shows the bandwidth and latency[5] as a function of buffer size for both MYRINET and SCI (PCI-PCI point-to-point DMA benchmark). Typical buffer sizes for the level-3 trigger are $b = 128$ byte for messages and $b \geq 20$ kbyte for data transfers. In both cases, the bandwidth is limited by PCI bus to $R \simeq 60$–70 Mbytes/s. In case of SCI, the maximum bandwidth is already achieved for small buffer sizes $b \geq 64$ bytes (corresponding to SCI payload), but saturates at a lower level ($R = 62$ Mbyte/s). However, former D310 versions also achieved $R = 72$ Mbyte/s (cf. Fig. 6 in Ref. [2]), only a recent hardware revision lead to a bandwidth reduction of $\Delta R/R \simeq 14\%$. For both cases, the CPU usage is as low as $\leq 12\%$ due to DMA. In case of SCI the (one-way) latency of $L = 2$–3 μs is smaller than in case of MYRINET ($L \simeq 30$–40 μs) due to an extra MYRINET software layer. The latency limit for the STAR DAQ design is $L \leq 100$ μs. The final decision for the usage of MYRINET was driven by (a) long-term test stability issues[6], (b) hardware revision status and availability and (c) free availability of MYRINET driver software as "open source" for numerous platforms (e.g. Linux/Intel, Linux/ALPHA, VxWorks).

6. Global Level-3 trigger

The Global-L3 CPU performs (a) track data collection from all Sector-L3 machines, (b) a level-3 decision algorithm based on event characteristics (e.g. invariant mass reconstruction, further examples in Section 2.1) and (c) issues the level-3 yes/no decision to the event builder (MYRINET message). Merging of low p_T tracks split by TPC sector boundaries is foreseen for the future. Multiple Global-L3 processes and/or CPUs will cover different physics decision tasks simultaneously, the yes/no decision being issued as logical OR. Both Pentium III (600 MHz) and ALPHA XP1000 (500 MHz)

[5] Latency $L = t_1 - t_2$ is defined by the time t_1 for issuing an interrupt (e.g. end-of-package) on the receiver and the sender acknowledge time t_2.

[6] In a ring topology, any failure (e.g. a single faulty cable) affects the complete system, in a switch topology only a single point-to-point connection.

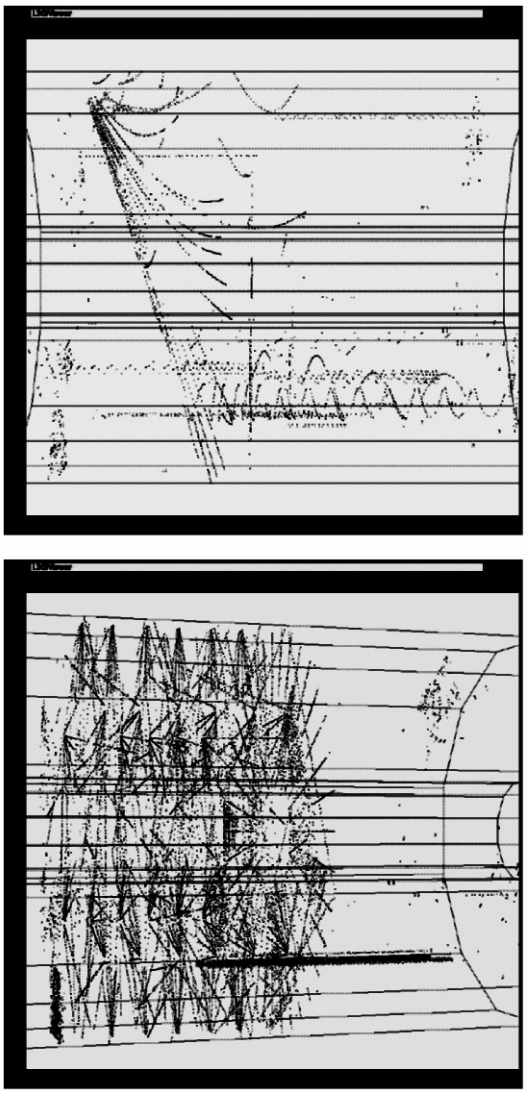

Fig. 6. 3D event display (cf. Section 7) for a level-3 processed cosmic particle event (*top*) and a laser event (*bottom*), recorded in system test (cf. Section 8) in 12/99 (level-3 cluster centers-of-gravity, *zy* sideview).

have been tested as Global-L3, showing no difference in MYRINET performance. As an example for a global decision algorithm, invariant mass reconstruction for 100 particle pairs ($J/\psi \to e^+e^-$, high p_T, cut $p_T \geq 1.5$ GeV/c pre-applied) requires a CPU time of $t = 0.4$ ms on ALPHA.

7. 3D event display

In order to visualize and browse a large number of events (level-3 accept and reject cases) quickly, a fast 3D event display has been developed. The C++ program is based upon the high-performance graphics language *OpenGL* (Mesa 3.0 library). The graphical user interface has been designed using the Qt 2.0 Library. The hardware, a Pentium III (600 MHz) with (*OpenGL* supporting) nVidia RIVA TNT2 graphics adapter, allows the display of 250,000 clusters ($N_{track} = 8000$) as well as mouse controlled operations like rotating and zooming in quasi-realtime without significant delay.

8. System test

In 12/99, a detailed system test of the level-3 trigger system (cf. Fig. 5) has been performed in two stages.

(1) While the STAR TPC was switched off, empty events (even size identical to Au + Au collision event) have been used for tests of high-rate and long-term stability of one single sector subpath (one VME crate, one Sector-L3, one Global-L3). For messages only (72 messages per event), a trigger input rate of $R \simeq 600$ Hz could be processed stably for 1.2 million events. For messages and data transfer, a bandwidth of $R = 48$ Mbytes/s was achieved (buffer size $b = 196$ kbyte), being lower than the MYRINET bandwidth (Fig. 5, *left*) due to the message protocol.

(2) While the STAR TPC and the STAR magnet ($B = 0.5$ T) were switched on, a test of the complete[7] level-3 trigger prototype system (432 Intel i960, 12 Sector-L3, one Global-L3) was performed with laser[8] and cosmic particle

[7] The step from one up to 12 Sector-L3 CPUs also marks at the same time the step from synchronous to asynchronous mode (arriving data sequence by random).

[8] For laser events the track finder was not operational, because straight tracks result in unreasonable $p_T \to \infty$ values.

events. Event displays for one event of each type (using the 3D event display, cf. Section 7) are shown in Fig. 5.

For single cosmic particle tracks, a very high track finding efficiency $\varepsilon \geq 95\%$ was achieved, although loose χ^2 cuts (in order to be able to find tracks with z_{vertex} even outside the TPC) lead to a number of reconstructed tracks $N_{\text{track}} \geq 1$ in $\sim 30\%$ of all cases. In total, 15,000 level-3 specific events (cluster and track data) were recorded successfully. Further cosmic data taking is planned for 01-03/2000. The start of the RHIC physics program is envisaged for 04/2000.

9. Summary

The STAR level-3 trigger system will perform online track finding for high multiplicity Au + Au collider events ($N_{\text{track}} \geq 8000$, $N_{\text{point}} \leq 45$ per track), utilizing high-performance ALPHA 21264 CPUs and high-bandwidth MYRINET network interfaces (expected data transfer rate $\simeq 52$ Mbytes/s). A global level-3 CPU will perform tasks like e.g. online invariant mass reconstruction and issue an accept/reject decision with a design input rate $R = 100$ Hz ($R \leq 25$ Hz for a prototype system in 12/99).

Acknowledgements

This work has been supported by the *Bundesministerium für Bildung und Forschung*, Germany (contract ♯06OF 840 I). We would like to thank MYRICOM, Inc. for excellent technical and documentational support. J.S. Lange would like to thank the Alexander-von-Humboldt organization for support of INSTR99 conference participation.

References

[1] STAR collaboration, STAR Conceptual Design Report, Lawrence Berkeley Laboratory, University of California, PUB-5347, June 1992.
[2] J.S. Lange et al., The proposed Level-3 Trigger system for STAR, 11th IEEE Real Time Conference, 6/14-18/99, Santa Fe, USA, IEEE Trans. Nucl. Sci., accepted for publication.
[3] A. Ljubicic Jr. et al., Design and implementation of the STAR experiment's DAQ, 10th IEEE Real Time Conference, 9/22-26/97, Beaune, France.
[4] J. Takahashi et al. (STAR SVT collaboration), Silicon drift detectors as tracking devices, Nucl. Instr. and Meth., these proceedings.
[5] X.N. Wang, M. Gyulassy, Phys. Rev. D 44 (1991) 3501.
[6] P. Yepes, A fast track pattern recognition, Nucl. Instr. and Meth. 380 (1996) 582.
[7] http://ftp.digital.com/pub/Digital/info/semiconductor/
[8] SCI (Scalable Coherent Interface) Standard Compliant, ANSI/IEEE 1596-1992.
[9] http://www.dolphinics.no.
[10] http://www.vmetro.com.
[11] MYRINET American National Standard, ANSI/VITA 26-1998.
[12] http://www.myri.com.