The DAQ of the COMPASS Experiment

L. Schmitt, H. Angerer, N. Franz, B. Grube, B. Ketzer, I. Konorov, R. Kuhn, W. Liebl, S. Paul, H. Fischer, A. Grünemaier, F.-H. Heinsius, K. Königsmann, T. Schmidt, U. Fuchs, and M. Lamanna

***Abstract*—COMPASS is a fixed target experiment at CERN's SPS. In 2002, a first physics run was completed with 260 TB of data recorded, corresponding to 5 billion events. The data acquisition architecture is based on custom frontends, buffers based on PCI cards, and server PCs networked via Gigabit Ethernet. A custom timing and trigger distribution system provides unique event identification and time synchronization. Results on the performance of the system and an outlook to improvements using online filtering will be given.**

***Index Terms*—Buffer memories, COMPASS, data acquisistion, optical fiber devices, S-Link, trigger control system.**

I. Introduction

COMPASS is a new experiment [1] at the CERN SPS aimed at the study of structure and spectroscopy of hadrons with multiple types of high intensity beams.

A polarized muon beam is used for deep inelastic scattering on a polarized ^{6}LiD target to study the spin structure of nucleons. The primary goal here is to measure the gluon contribution to the nucleon spin by measuring hadrons produced by the photon–gluon-fusion process. In addition, a fraction of the beam time is used to measure the transverse spin structure of the nucleon by polarizing the target in transverse direction.

In a second phase of COMPASS, hadron beams will be used to study Primakoff scattering, exotic mesons, glueballs, and charmed hadrons.

The COMPASS experiment uses a double forward spectrometer for best momentum resolution. In 2002, COMPASS had a RICH detector after the first spectrometer magnet, two hadronic calorimeters and muon filters, and an electromagnetic calorimeter after the second magnet for particle identification. In the final setup, both spectrometer parts will be equipped with RICH detectors, electromagnetic and hadronic calorimeters, and muon filters.

The tracking system is composed of a number of tracking stations, each made up of three nested detector types with increasing granularity and rate capability toward the center of the beam: drift chambers, straw tubes, and MWPC are used for the outer regions, Micromegas, and GEM detectors for the intermediate region and scintillating fibers for the inner region with the beam passing through. Upstream of the target additional silicon detectors and scintillating fibers are used for beam detection. Fig. 1 shows the setup of the COMPASS spectrometer.

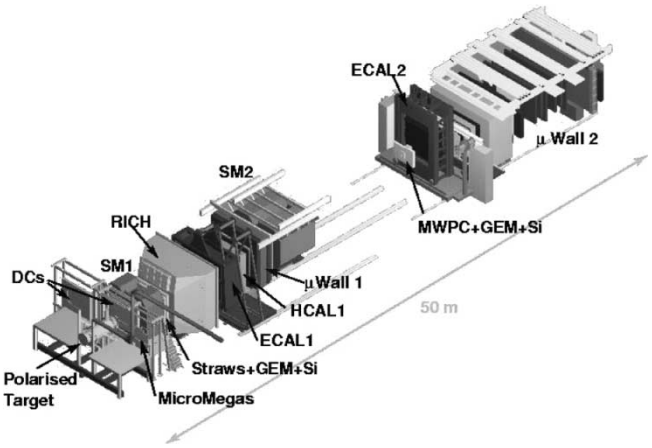

Fig. 1. Setup of the COMPASS spectrometer.

In 2002, COMPASS had its first 100 day long physics run. During this time its apparatus consisting of various detector types with a total of 190 000 electronics channels was read out at a rate of 5 kHz.

II. Requirements to the DAQ

The COMPASS experiment has around 190 000 detector channels distributed over various subsystems and a 50 m long setup. It runs at the SPS accelerator at CERN which has a spill structure with a 4.8 s long particle extraction during which the experiment sees a more or less uniformly distributed, random flux of 2×10^8 particles. This extraction is repeated every 16.8 s thus giving a duty cycle of about 30%. Depending on the physics program, during the extraction the readout electronics has to cope with a trigger rate from 5 kHz to a design maximum of 100 kHz corresponding to an average readout time of 10 μs.

The detectors use four different types of frontends depending on their specific needs. The RICH detector uses the GASSIPLEX ASIC developed at CERN [2]. It has a 1200 ns risetime and a multiplexer stage presently requiring a fixed readout deadtime of 3–5 μs. Most tracking detectors are read out via the F1-TDC chip developed for COMPASS by ACAM and the University of Freiburg [3]. Its excellent time resolution is in particular needed for the scintillating fiber trackers which determine the most precise track timing in COMPASS. Its design is inherently pipelined and thus deadtime-free.

The GEM [4] and silicon trackers [5], on the other hand, use a readout based on the APV25 chip [6] developed for the silicon tracker of CMS. This chip has an analog pipeline and a

Manuscript received June 23, 2003; revised October 8, 2003. This work is supported by the German Ministry for Education and Research, BMBF.

L. Schmitt, H. Angerer, N. Franz, B. Grube, B. Ketzer, I. Konorov, R. Kuhn, W. Liebl, and S. Paul are with the Physik-Department, TU München, Garching, D-85747 Garching, Germany (e-mail: Lars.Schmitt@ph.tum.de).

H. Fischer, A. Grünemaier, F.-H. Heinsius, K. Königsmann, and T. Schmidt are with the Fakultät für Physik, Universität Freiburg, D-79104 Freiburg, Germany.

U. Fuchs and M. Lamanna are with CERN, IT Division, CH-1211 Geneva, Switzerland.

Digital Object Identifier 10.1109/TNS.2004.829386

TABLE I
NUMBER OF CHANNELS, OCCUPANCY, EVENT SIZE OF DETECTORS

Detector	*Channels*	*Occ. (%)*	*kB/evt.*
BMS	*256*	*12.2*	*0.35*
SciFi	*3,936*	*6.5*	*2.71*
Silicon	*9,168*	*5.3*	*1.94*
Micromegas	*12,288*	*2.9*	*4.39*
GEM	*30,720*	*5.5*	*6.78*
Small DC	*4,224*	*3.3*	*1.14*
Straws	*6,912*	*1.4*	*1.26*
RICH	*82,944*	*3.5*	*13.13*
MWPC	*25,592*	*0.8*	*3.00*
Large DC	*976*	*15.1*	*0.72*
Calorimeters	*1,024*	*1.9*	*0.21*
Myon-Filters	*10,128*	*0.8*	*1.26*
Hodoscopes	*1,536*	*5.2*	*0.80*
Counters	*576*	*94.5*	*2.42*
Total	*190,280*	*3.9*	*40.1*

multievent buffer followed by a multiplexer. Since the buffer is 10 events deep and at COMPASS three amplitudes per event are read out via the multiplexer at 20 MHz, a limitation of maximum 10 triggers in 200 μs arises.[1] Finally, the calorimeters are read out by gated, fast integrating ADC modules (called FIADC) which can digitize up to three consecutive events within 30 μs. Appropriate safety margins to these deadtime requirements should be applied.

Table I lists the number of channels, occupancies, and data sizes of each detector. The total event size is around 40–50 kB depending on noise and trigger settings. This size, SPS duty cycle and trigger rates lead to data rates ranging from 900 MB to 18 GB per SPS spill. While the first figure can be still safely stored to tape, rates above about 1 GB have to be filtered online before mass storage.

III. ARCHITECTURE OF THE DAQ

The mostly pipelined detector readout electronics feeds the data into concentrator modules, 9U VME boards called *CATCH*, which are being developed at the University of Freiburg. In a similar function, but specialized for the highly integrated readout of GEM [4] and silicon [5] detectors, the *GeSiCA* board, developed at TU Munich, is used.

From these boards, the data are transferred via optical links to the DAQ computers. These links follow the *S-Link* standard [7] developed at CERN. The data received via S-Link are buffered on PCI cards called *spillbuffers* to profit from the accelerator's duty cycle to reduce the sustained data rate to a third of the onspill rate.

The computers housing the spillbuffers (*readout buffers*, ROBs) transmit the data via Gigabit Ethernet to eventbuilder computers, which combine the subdetector data to complete event blocks and transfer it to CERN's central data recording system.

The general architecture is sketched in Fig. 2.

[1]At a later time, the multiplexer will run at 40 MHz to allow for trigger rates up to 100 kHz.

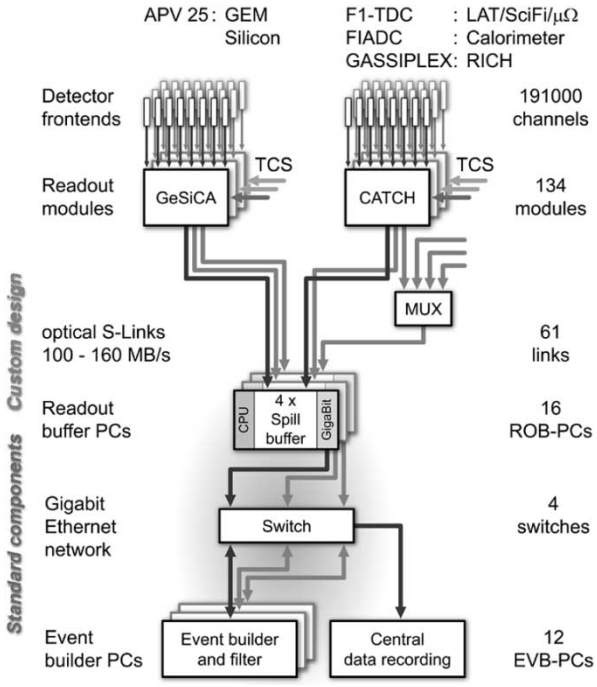

Fig. 2. General architecture of the DAQ system.

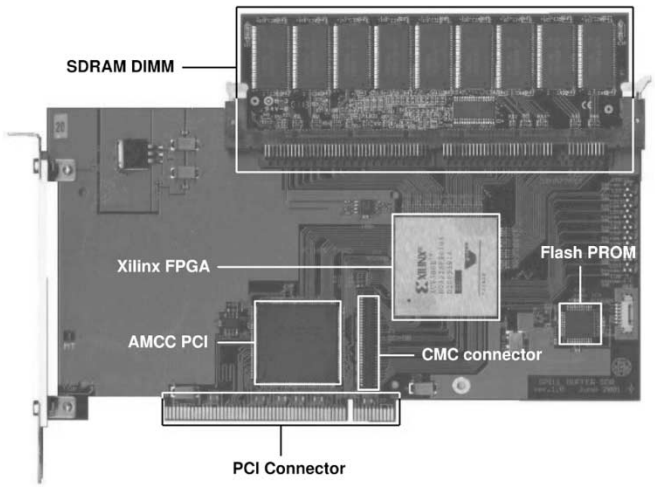

Fig. 3. Photograph of the COMPASS spillbuffer PCI card.

IV. COMPONENTS OF THE SYSTEM

A. Spillbuffer

The spillbuffer pictured in Fig. 3 is a PCI card which was developed at TU Munich to provide the buffering of data from one full spill for all detectors. The card employs the S-Link standard to receive data via a mezzanine card which is plugged to an IEEE 1386 CMC connector. Several versions of these mezzanines with optical links, a fiber channel implementation with 100 MB/s, a single G-Link version with 128 MB/s, and a dual G-Link version with 160 MB/s, are in use in COMPASS to adapt to the varying bandwidth requirements of different detectors.

The spillbuffer card has 512 MB of SDRAM memory which is addressed via a Xilinx FPGA as FIFO. This FPGA also performs the data reception via the S-Link protocol and controls the PCI interface which is based on an AMCC 5935 chip. The PCI

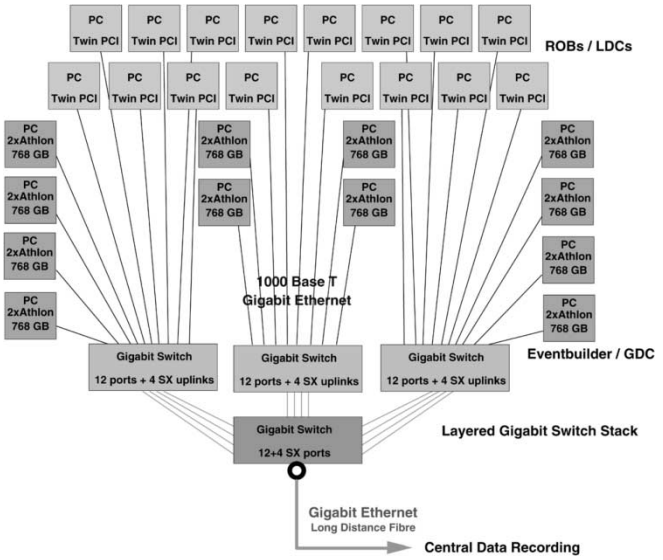

Fig. 4. The COMPASS eventbuilding network in 2002.

interface allows DMA transfers from the spillbuffer memory directly to the host PC's main memory.

B. DAQ Computers

In 2002, we used 16 computers as readout buffers, each with 1 GB ECC SDRAM as main memory, two 866 MHz or 1266 MHz PIII CPUs, and a 3COM Gigabit Ethernet interface. They have two PCI buses to read in the data on the first 32 bit bus and then transfer it through the second 64 bit bus via Gigabit Ethernet to the eventbuilder computers. This effectively allows a simultaneous reading from the detectors and writing to the eventbuilding network without overhead or bandwidth losses. Four spillbuffer cards are mounted per readout buffer.

As eventbuilders 12 computers with two Athlon MP 1900+ processors, 1 GB ECC DDR-SDRAM, and an IDE-RAID with net 640 GB per machine were used. In total 7.68 TB disk space is therefore available as buffer in case of problems with tape recording giving a safety margin of up to two days. These computers are also meant to provide limited CPU power for a simple online filter.

C. Eventbuilding Network

All DAQ computers are connected to a set of three switches 3COM 4900 with twelve 1000 BaseT ports and four 1000 BaseSX uplinks (Fig. 4). These uplinks connect the frontend switches to a backbone switch 3COM 4900SX which provides the necessary crosswise connectivity for all DAQ machines. To achieve a balanced configuration, four eventbuilder computers are connected to each frontend switch matching the number of uplinks to the backbone. ROB computers are connected according to their average output. The stack of switches provides a total number of 36 1000 BaseT ports to connect DAQ computers. The total eventbuilding bandwidth amounts to 12 Gb/s. The buffered data files are asynchronously transmitted to the central data recording system through the same network via 1 Gb uplink.

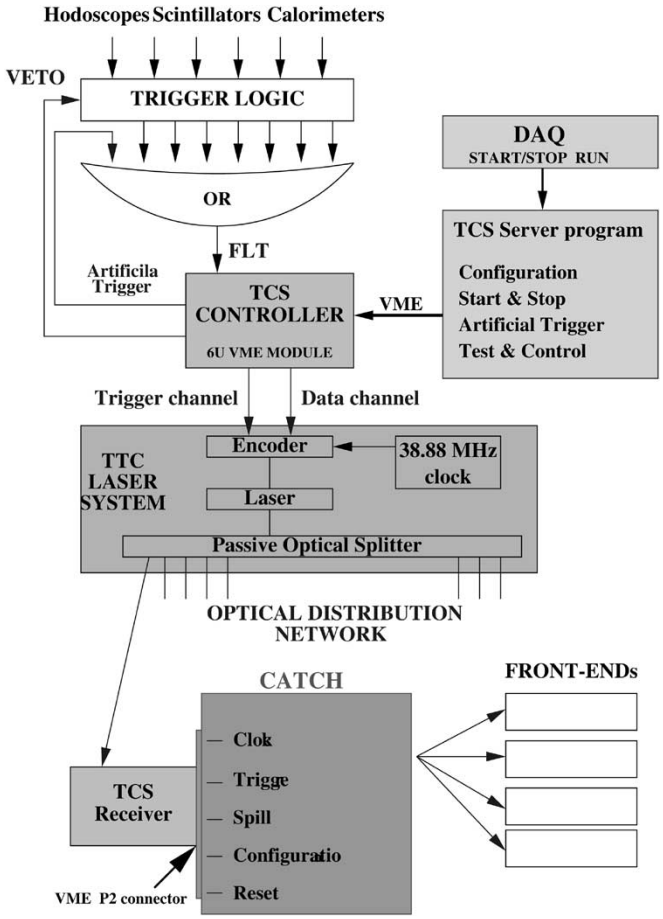

Fig. 5. Schematic view of the TCS system.

Presently, the data is transmitted using the safer TCP protocol. At the current rates, no congestion and very rare retransmissions are observed. To cope with the maximum possible trigger rate it will be however necessary to use lower level protocols and zero-copy network software. A moderate extension of the switch stack may be necessary as well, if online filtering cannot provide the data reduction to stay within present recording rates.

D. Trigger Control System

To ensure data integrity in the pipelined readout architecture each trigger has to be labeled by a unique event number. The synchronous transmission of the trigger and its identification to all detectors is done by means of optical signals distributed by the Trigger Control System (*TCS*) developed at TU Munich [8], schematically shown in Fig. 5.

The TCS system provides a stable clock to the entire experiment. This is necessary for precise time measurements of particles passing through drift detectors and scintillators to correctly correlate detector hits at very high beam rate.

The TCS controller, a 6U VME board, encodes the trigger input and splits it into a synchronous trigger signal and an asynchronous event data block (event number, type, etc.). The encoded pulses are transmitted by high-power laser diodes and distributed via glass fibers and passive optical splitters to the detectors, a system developed at CERN for the LHC experiments

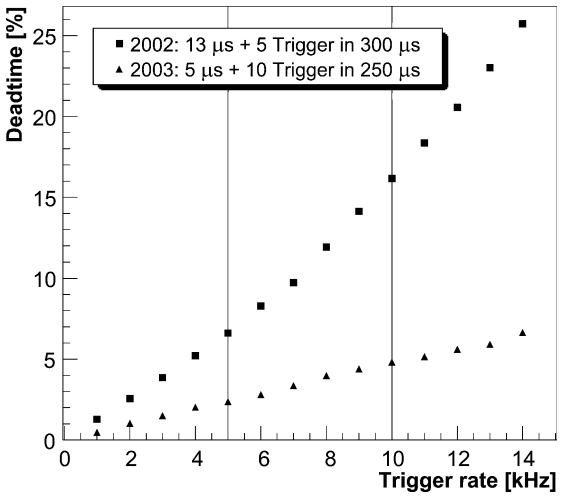

Fig. 6. Deadtime of the COMPASS DAQ, 2002 versus 2003.

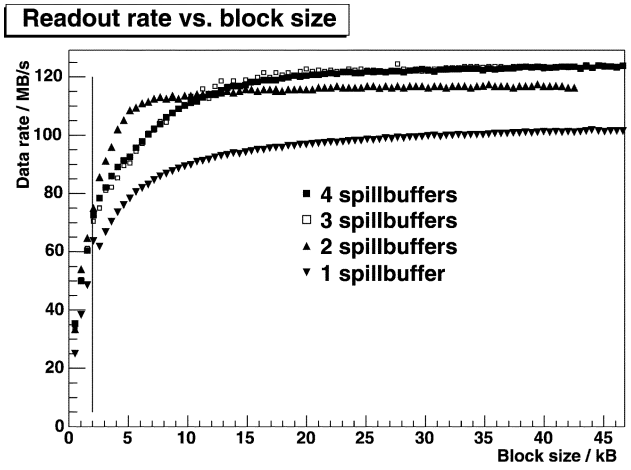

Fig. 7. Readout rate versus block size for one to four spillbuffer cards. The event rate decreases from 240 kHz at 0.5 kB to 1.6 kHz at 75 kB per spillbuffer card. The line at 2 kB indicates a typical upper event size in COMPASS.

[9]. At the heart of the module is a powerful Xilinx Virtex2 3000 FPGA chip. This module also controls two different types of deadtimes, a fixed deadtime to accommodate the RICH readout and two variable deadtimes based on the number of triggers in given time intervals for calorimeter and APV readout. All values are programmable. Fig. 6 shows the relation between total deadtime and trigger rate with the settings employed in 2002 and 2003. The main improvements came from modifications of the Micromegas readout, which in 2002 still required a fixed deadtime of 13 μs and, and the APV readout, imposing in 2002 a limit of 5 triggers per 300 μs. In addition, the introduction of two separate variable deadtimes removes rare glitches in the FIADC readout due to more than three close triggers within 30 μs.

The TCS receiver is a VME P2 transition board, which receives the optical signals and decodes the TCS information, in particular the trigger, the clock, SPS spill information, and control signals. It consists of an optical receiver, a clock recovery chip, an FPGA, and some other discrete components. It can be programmed via broadcasts to mask or block specific triggers from the attached readout concentrator module.

The TCS system was operated successfully and a very low time jitter of about 43 ps rms between TCS receivers was measured.

In addition to physical, also artificial triggers can be generated and directed to individual detectors by the TCS system. They serve for calibration or noise measurements.

An additional function of the system is to provide separate deadtimes and triggers for up to eight independent trigger systems connected to the same readout electronics. This allows the sectioning of the DAQ system to test detectors independently before integration into the full system. Since all signals have to be broadcast via the same optical network, the division into sub-DAQ systems is done via programmable timeslices.

E. Software

The central flow-control software of the COMPASS DAQ is based on DATE written by the ALICE DAQ group [10]. DATE provides software for eventbuilding, run control, information logging, and event sampling. The latter was used to implement an experiment-wide online monitoring program based on ROOT [11].

Modifications have been made to the eventbuilding to optimize input/output (I/O) and accommodate an interface to the online filter. The run control was adapted to the needs of COMPASS to integrate further functionality like a trigger interface and a run logbook based on MySQL. The latter logs run specific information from various components of the experiment (beamline, polarized target, magnets, trigger, etc.) along with automatically generated standard monitoring histograms and information provided by the shift crew. All information is made available via a Web interface.

DATE also provides a framework to perform the actual detector readout. Within this framework, a PCI driver for the spillbuffer card was developed. It is based on a custom Linux kernel module managing DMA transfers from the spillbuffer card to the host computer's main memory. The module structures the data stream by receiving interrupts from the spillbuffer and accordingly creates a dynamic event directory accessible by the DATE readout program from user space.

On the ROB computers, the local data collector (LDC) software of DATE reads the subevent data from the spillbuffers, checks and formats it, and sends it via the network to the eventbuilder computers in a round-robin manner based on the unique event number present in the data. On the eventbuilders it is received by the global data collector software (GDC) and then buffered on the local RAID arrays until it is transfered to the central data recording. This transfer is controlled via daemons written in PERL and interfaced to DATE via a number of bookmark files. After transfer via the network to CERN's computer center the data in 2002 was formatted in Objectivity/DB before being stored in the CASTOR hierarchical storage manager [12]. In 2003, the raw data was stored directly in CASTOR, whereas a catalog of metadata is stored in an Oracle 9i database.

All readout concentrator modules of COMPASS are housed in VME crates for power and control. A custom client server software controls and configures the modules. It receives its information from a MySQL database which is maintained through a Web interface.

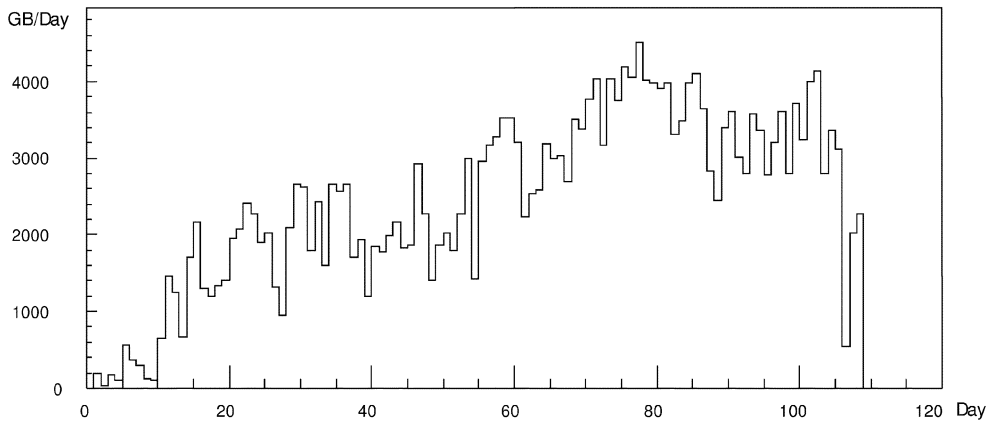

Fig. 8. Daily data rate in 2002.

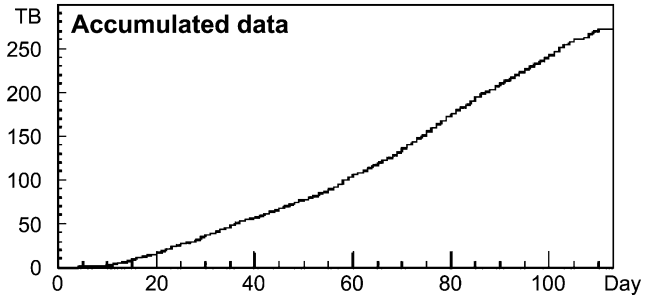

Fig. 9. Total accumulated data in 2002.

To control the TCS system and interface it with the DAQ system software a server process running on a Linux VME CPU was written. It sends all necessary signals to the TCS controller module and keeps track of all sub-DAQ systems.

V. Performance of the System

The spillbuffer PCI card was tested in a full scale test bench consisting of four GeSiCA boards connected via four S-Link cards of the single G-Link version to an ROB computer and the standard TCS system. In this test, the block size was increased and at the same time the event rate decreased to stay at a constant input flow of close to 120 MB/s during extraction. For the readout, the interrupt driven Linux DMA driver and the DATE software were used. The curves in Fig. 7 show that at larger block sizes the system is able to reach with a continuous rate of about 125 MB/s nearly the full PCI bus speed. With a single card about 100 MB/s can be read out whereas with three and four cards full performance is reached. The case of two PCI cards is particular since it is able to exceed the performance of three and four cards at block sizes below 10 kB when it reaches close to 120 MB/s. This may be explained by the efficient handshake between interrupts and readout in case of handling two cards in a dual processor machine.

After tailoring and debugging of network and S-Link PCI drivers and tuning disk I/O of the system, full production stability could be reached in summer 2002. In this phase, the DAQ exceeded the maximum anticipated data rate of 35 MB/s (3 TB/day) by up to 50% (see Fig. 8). On average 25 000 triggers were recorded per SPS spill with data sizes between 40 and 50 kB.

TABLE II
Comparison of Figures of the COMPASS DAQ in 2002 and 2003

Item	2002	2003
Channels	190k	200k
CATCH/GeSiCA	134	144
S-Links	61	66
Event size	42 kB	34 kb
Max. sustained rate	120 MB/s	180 MB/s
Deadtime	7% @ 5 kHz	2% @ 5 kHz, 4.5% @ 10 kHz

This corresponds to a sustained data rate of 60 MB/s, an onspill rate of 200 MB/s, and a disk writing rate of 5 MB/s. During the beamtime in 2002, COMPASS recorded 260 TB of data corresponding to about 5.5×10^9 events (Fig. 9). In test runs, a sustained rate of 120 MB/s to disk could be reached in 2002, without disk writing even 220 MB/s were possible, corresponding to about 20 kHz trigger rate.

VI. Outlook for 2003

Further improvements in readout and stability make a production trigger rate of 10 kHz onspill possible at a deadtime of only 4.5% compared to 7% at 5 kHz in 2002 (see Fig. 6). By implementing a scheme supressing unneeded headers from empty detector readout cards the average event size could be reduced to about 34 kB. First tests in 2003 showed sustained disk writing rates for the full system of up to 180 MB/s by means of a more efficient file system and optimized disk access. The benchmark figures of the years 2002 and 2003 are compared in Table II.

However, since by now the main bottleneck of the system is the limited bandwidth for tape writing, this amount of data has to be filtered online and reduced to the nominal tape writing speed. Therefore, at TU Munich an online filter algorithm is being developed which takes advantage of the fact that the eventbuilding uses only 20% of the available CPU power. The remaining power will be used to reduce the trigger rate by up to a factor of two from 50 000 events per spill, given a time quota of approximately 4 ms to decide about each event when running two filter threads per machine (see Fig. 10).

At a later stage, it is foreseen to have filter clients running on nodes of a separate filter farm thus making even higher first level trigger rates up to the design maximum of 100 kHz possible with

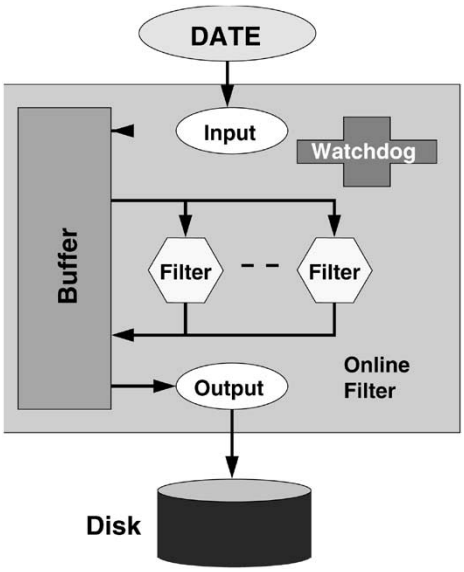

Fig. 10. Schematics of the COMPASS online filter.

a reduction factor of 10–20. In this scenario, a processing time of the order of 100 ms is available per event.

The filter program is multithreaded and works with a large I/O buffer. The input data is received via an IP socket connection, the output is written in large blocks optimized for fast disk I/O. A watchdog thread restarts stuck filter threads and guards necessary timeouts.

VII. Conclusion

The data acquisition system of the COMPASS experiment has been successfully deployed and recorded 5 billion events with 260 TB of data in 2002. The hardware concept has been proven to fulfill the needs of COMPASS and the road to higher trigger rates by employing online filtering is laid out.

Several customizations of the system have been made, increasing its flexibility and reliability. The hardware, based on mass marktet components for the eventbuilding network and computers and on custom frontend modules, has proven its stability and power to cope with the data.

The scalability of the system will allow a smooth upgrade to accomodate also the needs of the second phase of COMPASS which is foreseen to start in 2006 and reach well into the LHC era.

References

[1] G. Baum and the COMPASS Collaboration, "COMPASS: A Proposal for a Common Muon and Proton Apparatus for Structure and Spectroscopy," Tech. Rep., CERN-SPSLC-96-14.

[2] G. Baum *et al.*, "The COMPASS RICH-1 read-out system," *Nucl. Instrum. Methods A*, vol. 502, pp. 246–246, 2003.

[3] H. Fischer *et al.*, "Implementation of the dead-time free F1 TDC in the COMPASS detector readout," in *Proc. 8th Pisa Meeting Advanced Detectors 2000*, vol. 461, Nucl. Instrum. Methods A, La Biodola, Italy, May 21–27, 2001, pp. 507–507.

[4] B. Ketzer *et al.*, "Triple GEM tracking detectors for COMPASS," *IEEE Trans. Nucl. Sci.*, vol. 49, pp. 2403–2410, Oct. 2002.

[5] R. de Masi *et al.*, "Present status of silicon detectors in COMPASS," in *Proc. 9th Eur. Symp. Semiconductor Detectors—Schloss Elmau 2002*, vol. 512/1–2, Nucl. Instrum. Methods A, June 2003, p. 229.

[6] M. Raymond *et al.*, "The APV25 0.25 μm CMOS read out chip for the CMS tracker," in *Proc. 2000 5th Workshop Electronics for the LHC Experiments (LEB'99)*, Snowmass 1999, Electronics for LHC Experiments, p. 162, Snowmass, CO, Sept. 20–24, 1999, preprint IC/HEP-003.

[7] H. C. van der Bij, R. A. McLaren, O. Boyle, and G. Rubin, "S-LINK, a data link interface specification for the LHC era," *IEEE Trans. Nucl. Sci.*, vol. 44, pp. 398–402, June 1997.

[8] I. Konorov *et al.*, "The trigger control system for the COMPASS experiment," in *Proc. Conf. Rec. IEEE Nuclear Science Symp.*, San Diego, CA, 2001.

[9] B. G. Taylor, "TTC distribution for LHC detectors," *IEEE Trans. Nucl. Sci.*, vol. 45, pp. 821–828, June 1998.

[10] "DATE 3.7, ALICE DATE User s Guide, v.2," CERN ALICE DAQ group, ALICE Internal Note/DAQ ALICE-INT-2000-31, 2001.

[11] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework," *Nucl. Instrum. Methods A*, vol. 389, p. 81, 1997.

[12] J. P. Baud, O. Barring, and J. D. Durand, "CASTOR Project status," in *Proc. Int. Conf. Computing in High-Energy Physics and Nuclear Physics (CHEP'00)*, CHEP 2000, Computing in High Energy and Nuclear Physics, p.365, Padova, Italy, Feb. 7–11, 2000.