HIGH PERFORMANCE EVENT DISTRIBUTION USING HIPPI

T. Anguelov\textsuperscript{1}, W. Bozzioli\textsuperscript{2}, J-P. Matheys\textsuperscript{2}, P. McInnes\textsuperscript{3}, R.A. McLaren\textsuperscript{2}, N. McKay\textsuperscript{3}, A. Pastore\textsuperscript{4}, K. Peach\textsuperscript{3}, B. Renk\textsuperscript{5}, P. Vande Vyvre\textsuperscript{2}, A. Vascotto\textsuperscript{2} and A. Taurok\textsuperscript{1}

\textsuperscript{1} Österreichische Akademie der Wissenschaften, Vienna, Austria.
\textsuperscript{2} CERN, European Organization for Nuclear Research, Geneva, Switzerland.
\textsuperscript{3} Edinburgh University, Edinburgh, Scotland.
\textsuperscript{4} Digital Equipment Corporation, Joint Project Office, Geneva, Switzerland.
\textsuperscript{5} Institut für Physik, Universität Mainz, Mainz, Germany.

Abstract

Experiments now being planned, such as the NA48 Experiment (Cagliari-Cambridge-Dubna-CERN-Edinburgh-Ferrara-Mainz-Perugia-Pisa-Saclay-Siegen-Torino-Vienna Collaboration) on CP violation at the CERN SPS, require the collection of large volumes of data (up to 100 Mbytes/s) from several sources and at high instantaneous event rates (several thousands per second). Many of the problems encountered in such an experiment will have to be addressed by experiments at the very high energy hadron machines such as LHC (Large Hadron Collider) and SSC (Superconducting Super Collider). Events must be constructed from source modules and distributed to one of several third-level trigger processors. In this report we will address the problem of high-speed event distribution and show that the high-performance parallel interface (HIPPI) is well suited to the task of distributing the events, by a commercially available switch, to a farm of workstations used for third-level trigger processing.

Presented at CHEP'92, Annecy, France
21–25 September 1992

Mac/Word/6019F/PVV/sas
1. INTRODUCTION

The data-acquisition system of the NA48 Experiment [1] is designed to collect data from several sources, to assemble them into complete events and to inject these events into a third-level trigger workstation (L3WS) farm at high rates (up to 100 Mbytes/s). The design relies on standards such as high-performance parallel interface (HIPPI) [2] and uses commercially available hardware wherever possible. The overall system is shown in fig. 1.

![Diagram of the NA48 data acquisition system]

**Fig. 1** NA48 data acquisition system overview

The data from the front-end electronics of each detector is read out by local readout controllers, which accumulate (and perhaps manipulate) the data from a number of subdetector source modules. The readout controllers send the data independently over fibre optic links to an event builder, called "data merger" (DM), where the different parts of the same event are assembled into a single-data structure. Assembled events are then sent to the L3WS over a HIPPI link which includes a HIPPI switch that acts as a data-flow router.

The DM receives the "busy" or "free" status from the L3WS, via Ethernet. The HIPPI address of a free workstation is prepended to the event data and the HIPPI switch uses this information to route the event data to the corresponding workstation.

The L3WS farm will consist of Digital Equipment Corporation (DECstations) based on the TURBOchannel bus whose speed, i.e. 100 Mbytes/s, matches that of HIPPI. The TURBOchannel to HIPPI interface will feature high-speed transmission (as close as possible to the HIPPI and TURBOchannel maximum performances), scatter/gather direct
memory access (DMA) capabilities and detailed HIPPI error reporting. Software will also be provided both to control and test the interface.

Each workstation explicitly makes a request to the data distribution control software (running in the L3WS and in the DM) to participate to the data acquisition. This arrangement allows the insertion/removal of workstations without affecting the overall logic of the system. The software handling the HIPPI to TURBOchannel interface for the user application will include, amongst other things, buffer management, interface initialization, data transfer start/stop and HIPPI transmission error detection and (possibly) correction.

2. THE HIPPI STANDARD

HIPPI is an ANSI standard describing an efficient simplex high-performance point-to-point communication link. The physical-layer (HIPPI-PH) is designed for transmitting digital data at peak rates of 100 or 200 Mbytes/s using multiple copper twisted-pair cables at distances of up to 25 m. The two data rates are achieved with a data link carrying 32-bit or 64-bit words respectively.

A simple protocol is defined by HIPPI-PH: A “connection” between a “source” and a “destination” comprises one or more “packets”. There is no limit on the packet size which is fixed by the source. The packet itself is divided into “bursts” containing 1 to 256 words.

A destination shall send one “ready” indication for each burst that is prepared to accept from the source. A burst look-ahead function may be implemented in the destination.

The error detection scheme is based on a parity bit for each byte of data and one LLRC (Length/Longitudinal Redundancy Checkword) for each data burst. No error correction scheme is included in the standard.

HIPPI-PH does not support multiple destination configurations, but allows for a connection to be preceded by an “I field” which may be used by network equipment such as a switch to build 1-to-n configurations.

The speed of the link (100 Mbytes/s) and the simplicity of the protocol make HIPPI ideal as a point-to-point link in high-speed data-acquisition systems.
3. EVENT DISTRIBUTION IN NA48

The aim of our work is to develop the hardware and the software for a subset of the NA48 data-acquisition system using HIPPI links between the DM and the L3WS via a HIPPI switch. This includes the HIPPI to TURBOchannel interface, the associated software and the data distribution software handling the dialogue between the DM and the L3WS. This work is done in collaboration with DEC [3].

3.1 The HIPPI to TURBOchannel interface

This interface is a high-speed HIPPI destination TURBOchannel option [4]. It has been designed to optimize large-size DMA transfers. The interface contains two tables to avoid any CPU intervention during data transfer:

- an on-board scatter/gather table for DMA operations containing 64 K entries and allowing transfers of up to 256 Mbytes, and
- a history memory (32 K entries) recording transmission errors and protocol information.

The interface has been designed and a first prototype has been assembled. More details on the implementation of the interface can be found in ref. [5].

The software is divided in two layers: a device driver and a platform independent user-level library. This software has been specified [6] and its implementation has begun.

3.2 Data distribution software

The data distribution software controls the flow and distribution of data from the DM to the L3WSs. The data to be transferred will come in data blocks of ~ 100 Mbytes. There will be a data block every 12.5 s (as dictated by the accelerator cycle). The data distribution software will receive the status (busy or free) of each L3WS via Ethernet and will enable the DM to decide to which L3WS to send each data block to. The DM will insert the HIPPI I-field corresponding to the target L3WS and the data will automatically be routed to the correct workstation by the HIPPI switch.

The data distribution software has been specified [7] and the communication between the DM and the L3WS has been implemented.

3.3 First tests results

To verify the feasibility and performance of the proposed architecture, the test equipment presented in fig. 2 has been used. This arrangement includes two different VMEbus based modules, each a HIPPI source.
The first module, the second-level architecture test equipment (SLATE) [8, 9] has a memory which can be pre-loaded with simulated event data and delays. Once started, it can deliver data at the full HIPPI speed. The second module is a general-purpose VMEbus to HIPPI interface [10] which, whilst being more flexible than the SLATE, is limited by the VMEbus bandwidth to ~ 20 Mbytes/s.

Recent tests using the SLATE as HIPPI source, a switch from NSC [11] and two test boxes [12, 13] as destinations has allowed us to reach sustained speeds of up to 90 Mbytes/s when switching destination every 16 kbytes.

Acknowledgements

We wish to thank R. Mount of the L3 Collaboration, who has allowed us to use the HIPPI switch, E. Dénes for his help in the use of his SLATE program and A. Van Praag for his help in the setting-up of the HIPPI hardware test bench.
REFERENCES

[1] G.D. Barr et al., Proposal for a precision measurement of $e'/e$ in CP violating $K^0 \to 2\pi$ decays, CERN/SPSC 90–22, P253 (July 1990).


[8] T. Anguelov, A test system for second-level trigger and data acquisition architectures, these proceedings.


