A DATA DRIVEN PARALLEL PIPELINED HARDWARE RECONSTRUCTION PROCESSOR*

L. Borten, M. Church, E. Gottschalk, R. Hylton, B. Knapp, W. Sippach, B. Stern
Columbia University, Nevis Laboratories, Irvington, NY 10533

E. Hartouni, D. Jensen, M. Kreisler, M. Rabin, J. Strait
University of Massachusetts, Amherst, MA 01003

C. Avilez, W. Correa, J. Escalona, H. Morales, P. Salas, A. Zentella
University of Mexico Instituto de Fisica, 10000, Mexico DF, Mexico

C. Christian, G. Gutierrez, S. Holmes, R. Huson†, A. Wehmann
Fermilab, Batavia, Illinois 60510

ABSTRACT

An overall view is presented of Brookhaven E766 and Fermilab E690, which includes the construction and development of an on-line processor. The primary goal is to investigate the diffractive production of heavy quarks by studying exclusive multiparticle final states. The close relationship of the detector and read-out system with the processor and the physics goals that motivate this research program are discussed. In addition, we discuss the functions of the processor and explain an example of a module that forms part of the track fitter. We estimate we will reconstruct on-line 10^5 events/sec.

Introduction

By the end of the decade of the 60's it was expected that with the availability of high-energy beams, an explosion of data produced in high-energy collisions would lead to the systematic study (among other things) of many particle final states, hoping to learn more about hadron dynamics.[1] With the advent of the higher energy machines in the U.S. and Europe during the decade of the 70's, a wealth of events in high-energy collisions appeared; however, the exclusive multiparticle final states expected to provide more insight into the interaction of high-energy particles could not be studied. The problem was not only to produce those states, but to detect them and properly present them for further analysis.

---

*Presented by C. Avilez, on sabbatical leave at Fermilab, Fellow of the John Simon Guggenheim Memorial Foundation.

†Currently on leave at Texas A&M.

In order to detect and produce the appropriate information that leads to the full reconstruction of an exclusive reaction, the dead time of the apparatus has to be minimized and its read-out fast enough to achieve the proposed goals. Several levels of logical triggers have been found to be insufficient to identify the events under discussion, and the necessity of on-line processing is by now widely accepted.

With a highly segmented detector, fast read-out and on-line processing, it is possible to study the details of the dynamics of hadron physics in exclusive reactions.

E766 (BNL) and E690 (Fermilab)

The experimental high-energy physics program of the collaboration University of Massachusetts (Amherst)/Columbia University (Nevis)/ University of Mexico (IFUNAM)/Fermilab, is to explore the physics of the exclusive multiparticle final states.

Our primary physics goal is the comprehensive study of hadronic spectroscopy and production mechanisms, eventually covering strange, charm, and bottom particles at Tevatron energies. The construction of a multiparticle detector capable of measuring moderately complex reations (up to 20 particles) with high resolution and efficiency at moderately high rates is underway. A crucial feature of the experiment is detailed on-line reconstruction of up to 10^5 events/second chosen from more than 10^6 interactions/second. With the E690 spectrometer, this should allow the measurement of more than 10^4 fully reconstructed charm pairs and as many as 10^2 bottom pairs per hour of Tevatron beam.

The detector for this program is seen in Fig. 1, which also shows the part of the detector already in operation at Brookhaven. To perform accurate measurements over a wide range of particle types, angles, and momenta, the detector components must fit and function together.

The detector in operation at BNL consists of six drift chambers totaling 11,500 signal wires in a low-field, wide-aperture magnet (7kG, 4 ft×8 ft). Charged particle trajectories are efficiently reconstructed with momentum resolution $\delta p/p < 1\%$ (FWHM) from 200 MeV/c to 10 GeV/c. Mass resolutions for K_s and Λ^0 are presently 8 and 2 MeV (FWHM), with further improvement expected. Direct particle identification is provided by 102 time-of-flight counters and a segmented-threshold Cherenkov counter (96 channels). We have uniformly high acceptance for a wide range of topologies, particularly final states containing heavy particles. We intend to record a large number of events, fully reconstructed and identified as such, typically recording both raw and reconstruction results.

One of the main endeavors of E766 is to produce a sample of $\sim 10^6$ Ω^- for which the use of a real-time processor system is an absolute necessity. Our program at BNL also includes the search for highly-inelastic baryon resonances and further studies of hyperon polarization. In particular, we are interested in understanding Λ^0 polarization in exclusive reactions, since virtually nothing is known about the contribution of specific final states to this phenomenon.

The detector and processor form a hierarchy of independent subsystems linked to each other by high-speed buffers and linked to host computers through the Nevis transport system.[2] Transport system buffers between the detector and the processor allow data acquisition to bypass the on-line processor and allow data recorded on tape to be fed back to the processor.

The Hardware Processor

The processor is a way of dealing with very well-defined events, however rare, in a high rate environment, where very large-scale numerical computation is expected. It can adapt in the most elastic and flexible way to the computational needs of a wide variety of situations encountered in high-energy physics experiments. In a somewhat restricted way, it has been tested in Fermilab's E605,[3] and its first large-scale applications are E766 and E690.

The processor is a **data driven synchronous pipeline.** It has distributed memory and a modular structure with general interconnectability to match the computational needs of a given project. The pipeline is realized in terms of a parallel processor whose operands receive data when the data are present and the destination is available.* The synchronous aspect of the pipeline, in conjuction with two control bits, allows alignment of the data, avoiding juxtaposition of boundaries of blocks of information.

Data Processor For E766 and E690

A central feature of the processor is its extreme capability to match the computational requirements of a modern high-energy physics experiment. The problem we are interested in is the track reconstruction of all the particles in a moderate multiplicity (7 to 20) final state. The fitted track parameters provide us with charged particle momenta and trajectories which are used to reconstruct interaction and decay vertices. The track reconstruction consists of four stages: track finding, matcher, track fitter, and clean-up.

The track reconstruction by itself is most challenging and involves two extreme situations: pattern recognition to find tracks, which requires relatively little computation but a large number of combinatorial situations (actually dealt with by loops), and the track fitting, with a smaller number of decisions involved but requiring much more computation.

The track finding problem, which for straight lines has been discussed extensively,[3] is primarily solved with modules that involve bit parallel structures in a sequential word process. Experiment #605 at Fermilab has used the track finding in a single view of drift chamber with marginal constraints. In E766 (and E690) we look for tracks of charged particles in a magnetic field. In each view of the drift chamber, we obtain three wire numbers and the deviation from a straight line to represent the track. In Fig. 2 we show

---

*This is the data-driven principle that assures maximum hardware utilization, minimizing the number of idle components at any time.

the algorithm to find tracks in a moderate magnetic field by a single view. The results from different views are matched with strong constraints. For instance, the single-view curvature parameter is independent of view, providing a quick constraint. This pattern recognition, i.e., finding lines in single views and matching them to form track candidates, uses only wire addresses, while ignoring drift times.

We now concentrate on the other extreme, namely, the track fitter. In this part of the processor, as the computation proceeds, more bits are added to accurately represent the result of the calculation, changing thereby the size of the word involved.

The coordinates that specify the track are non-linear functions of the measured parameters. The process of fitting consists of finding the best set of parameters by minimizing a χ^2. Following Newton's method, a solution can be found to the non-linear problem, and in fact a few iterations, if an appropriate starting point is available, i.e., a zeroth order set of parameters $P^{(0)}$. The process of track finding provides us with this first approximate parameterization. In an operational way one first calculates,

$$P_i^0 = \sum_n A_{ni} X_i + A_{0i}, \; n = 1,.., 6; \; i = 1,..., 5$$

The wire numbers (X_i) are ten-bit words and the resulting five parameters (P_i^0) are 12-bit long. This kind of operation is done in hardware by a module called a sum multiplier which is explained below.

After having the zeroth order parameters we are interested in forming higher order nonlinear products of them, producing a set of 11 parameters (P_j, $j = 6, ..., 11$) to be used in a Taylor expansion beyond the linear terms. Again, using a sum multiplier we predict wire coordinates

$$X_i = \sum_{n=1}^{5} A_{in} P_n^{(0)} + \sum_{n=6}^{11} A_{in} P_n^{(1)}.$$

This produces 16-bit wire coordinates, 10-bit integer wire number and a 6-bit fraction. One may now get the difference of the predicted value from the nearest measurement within a Map, and by using a sum multiplier again produce a first correction to the zeroth order set of parameters. After a few iterations the drift time may be included to obtain the best predicted value for the coordinates within the measured precision of the experiment. See Fig. 3.

Comments on Modules; One Example of Bit Serial

The modules of the processor have three general requirements, namely, that they perform simple operations, have simple communication control, and if needed, have small storage independent of the size of the net. In the preceding talk, Dan Kaplan explained in more detail some features of the bit parallel modules and their protocol. Here we would like to explain the function of one of the most important modules of the track fitter, the sum multiplier. In an extremely simplified form showing the essence of the idea behind the design, let us do the following calculation

$$\Sigma\ a_n X_n,$$

where the X's are represented by three bits. The constants a_n can in principle be described by all the bits we like, independent of the "size" of the X's. We may write, assuming n=1,2

$$\Sigma\ a_n X_n = a_1 X_1 + a_2 X_2$$
$$= a_1\ (X_1^0\ 2^0 + X_1^1\ 2^1 + X_1^2\ 2^2 + \ldots)$$
$$+ a_2 (X_2^0\ 2^0 + X_1^2\ 2^1 + X_2^2\ 2^2 + \ldots).$$

We now rearrange this expression and factor out all powers of 2

$$\sum_n a_n\ X_n = (a_1 X_1^0 + a_2 X_2^0)\ 2^0$$
$$+ (a_1 X_1^1 + a_2 X_2^1)\ 2^1$$
$$+ (a_1 X_1^2 + a_2 X_2^2)\ 2^2.$$

The expressions in parentheses can be calculated in advance, in terms of the different possibilities of the weights of the binary expansion and loaded into a look-up table. The two vertical arrows show the two words that, serially fed into the module, provide the address to retrieve the precalculated combinations from the look-up table. In Fig. 4 we schematically show the structure of the sum multiplier.

The associative memory concept is extensively used both in the bit parallel and in the bit serial modules, in modules called Maps. Here the data are retreived by value and provide us with a considerable speeding-up factor. In the track finding problem, the use of a Map reduces an N^3 problem to an N^2 one. Using first and third chamber information to predict the position at the intermediate one allows us to retrieve by value the experimental result. In the least square process of the track fitting, we systematically have to compare results of evaluations with the experimental measurements in the first iteration, or in subsequent ones, with the result of the previous step. A Map module has two input ports, one to write on the memory of the module, the other a read port that receives whatever value was predicted and has to be tested for its validity. A third port of the module is just an output that provides a "road" of values around the tested one. Provision is made to mask the word at the read port to select fields of bits that are asked to be represented by the data written on memory. The Map itself does not "decide" if the tested value passed the test or not, but by providing a road of possible values around the predicted one, produces a word to be fed into a look-up table with precalculated results of the test as a function of the road.

Comments and Remarks

After recognizing the complexity of the problems of event selection and data analysis, we have set out to provide a general approach to large scale computation. Our first large scale operation will be at Brookhaven in 1985. This should provide an adequate demonstration of a more generally useful

inexpensive approach to the larger computation problems of high-energy physics experiments, as well as other fields.

References

[1]See for instance, C. N. Yang, "Hypothesis of Limiting Fragmentation" in the Third International Conference on High Energy Collisions (Gordon & Breach, 1969), p. 509.

[2]J. A. Crittenden, G. Beneson, H. Cunitz, Y. B. Hsiung, D. M. Kaplan, W. Sippach, and B. Stern, "A Data Acquisition System for Elementary Particle Physics," to be published in IEEE Trans. Nucl. Sci.

[3]See D. M. Kaplan, talk at this symposium. W. Sippach, G. Beneson, B. Knapp, IEEE Trans. Nucl. Sci., **NS-27**, 578 (1980); H. Cunitz, Y. Hsiung, B. Knapp, W. Sippach, "Data Driven Processing," in Proceedings of the International Conference on Instrumentation for Colliding Beam Physics, SLAC Report-250, June 1982.

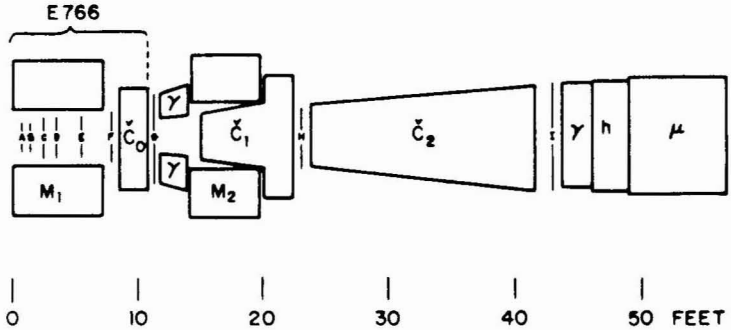

Fig. 1. Schematic layout of E690 (FNAL). A-I are drift chambers; M1 and M2 are large aperture modest field analyzing magnets; C_0, C_1, and C_2 are highly segmented threshold Cherenkov counters and downstream a subsystem of calorimeters. E766 is the part of the detector operating at Brookhaven National Laboratory.

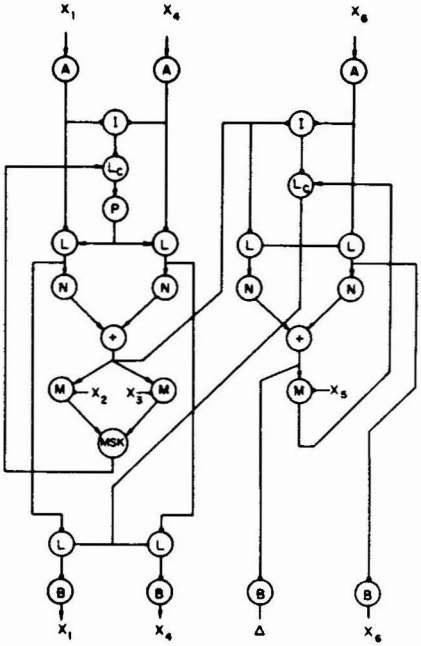

Fig. 2. Moderate magnetic field track finder. Uses wire numbers of single views of 6 drift chambers. X_1 and X_4 are used to predict values at either X_2 or X_3 (done at different cycles. This is why the straight line finder X_1-X_4 requires a page generator module [P]). The predicted value (X_2 or X_3) is used with X_6 to predict a value for the deviation from a straight line tested in a Map by X_5.

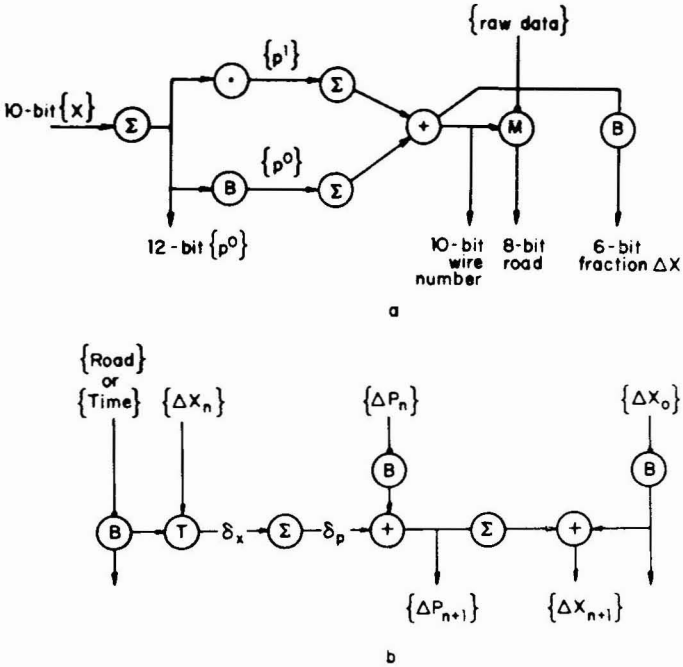

Fig. 3. Track fitter. Schematic diagram. The symbols stand for sum multiplier (Σ), buffer (B), adder (+), Table (T), Map (M), multiplier (•). For every high order product of $p^{(0)}$, one introduces a multiplier. In part a, (•) stands for all the multipliers needed to produce the required higher order terms of the parameters. The same applies to Σ.

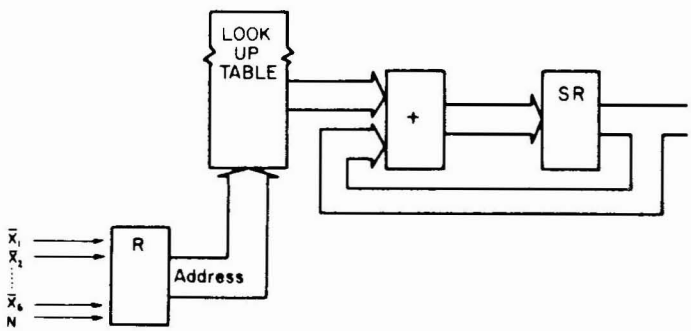

Fig. 4. Operation of the sum multiplier. Words are fed bit serially into a register. The nth bits of all the input words provide an address used to retrieve from a look-up table precalculated values of the linear combination $\Sigma\ a_n\ X_n^m$, where the X_n^m are the weights of the binary representation of X_n.

QUESTIONS AND ANSWERS

(Editor's Note: The question and answer sessions for the talks by Kaplan and Avilez were combined.)

Q: How much pay off do you get from the "refit" using the drift time information, compared to computation time it takes?

P. Lebrun

A: We need ~ 100μ to 200μ resolution. A normal, usual size PWC cannot give such a resolution. Therefore we do have to do this "refit."

C. Avilez

Q: Needs for off-line analysis?

A. Brenner

A: Most experiments will require only minimal off-line analysis.

M. Kreisler, C. Avilez

Q: Considering that typical off-line reconstruction codes consist of many pages of FORTRAN, how can you expect to build systems of that many modules?

M. Fischler

A: The mapping of FORTRAN in the most general sense into hardware modules is not the way to examine the problem. Rather the question is whether one can reconstruct a large number of tracks from complex topologies in a finite (i.e., small) number of modules? The answer to that is yes.

M. Kreisler

Q: How much does one of your modules cost?

M. Delfino

A: $100 to $300, depending on volume.

M. Kreisler

Q: How many experiments have the resources of talent to put together a data-driven system that will "fully" reconstruct their data?

T. Nash

A: At present, very few although we expect that to change as this technology is developed.

M. Kreisler

Comment: At E766, we currently write data limited by tape writing speed yielding approximately 4 tapes/hour. Such tapes take ~ 4 CPU hours on a Cyber. This would yield a mass of data not easily passed in toto through an off-line computer center. Thus the bulk of the data is filtered by the processor and then taken to an off-line system. Experiments such as ours will need computer time, of course, but **not** at the scale one would first guess by the magnitude of our data stream.

M. Kreisler

Q: A pipeline can only compute as fast as its slowest element. For either configuration, what is the efficiency of the pipeline for a typical event?

R. Fine

A: No answer