

93-191



СООБЩЕНИЯ  
ОБЪЕДИНЕННОГО  
ИНСТИТУТА  
ЯДЕРНЫХ  
ИССЛЕДОВАНИЙ  
ДУБНА

E10-93-191

V.Kotov, I.Aleksandrov, R.Pose, Yu.Yatsunenko

PIPE-PROCESSING ON SYSTOLIC STRUCTURES  
OF HOMOGENEOUS CALCULATING SYSTEMS  
FOR FAST PROCESSING  
OF TRACK INFORMATION

1993

# 1 Introduction

The paper describes a software-hardware option of the integral track criterion [1] for identification of muon track images in the 1st level trigger by constructing a systolic pipe-line on the basis of a one-bit massively processors. In particular, the version of the muon detector for the installation ATLAS [2] based on high precision drift tubes (HPDT) is considered as the cheapest one though with large information uncertainty in fixing track coordinates and thus the most difficult to be used in the muon trigger. It is shown that even with limited input information one can perform the required selection in the 1st level trigger in a startup time period of 1000 ns (depth of pipe-line) and get the result for each bunch. Thus, the integral method fits in with the calculation algorithms for one-bit massively processors, which allows the fullest possible paralleling of calculations, and this systolic pipe-line may be a basis for a fast trigger [3].

The paper considers processing of the data from one superlayer because presence of this processing stage in the parallel pipe-line is the fundamental feature of the given method. Systolic matrix calculators with bit processors are proposed to be built on the basis of the EPLDS (FLEX, ALTERA) technology. Basically, the idea of constructing a calculator consists in organization of serial-parallel calculations in a regular structure. The initial data are simplified preliminarily as much as possible so that their structure corresponded to the systolic pipe-line structure as fully as possible and was reduced to the smallest digit capacity by using increments (difference) in the structure of input data.

## 2 Description of the superlayer data model

A superlayer in the installation ATLAS has 6 layers (fig. 1). Each layer consists of about 30 HPDTs. The diameter of a tube is 30 mm. The tube position coordinates are known. From each layer one can have up to 6 values of the relative drift time  $t_i$ , measured with respect to the moment  $T_{0i}$  of occurrence of the event under investigation in the detector. These values also include the value

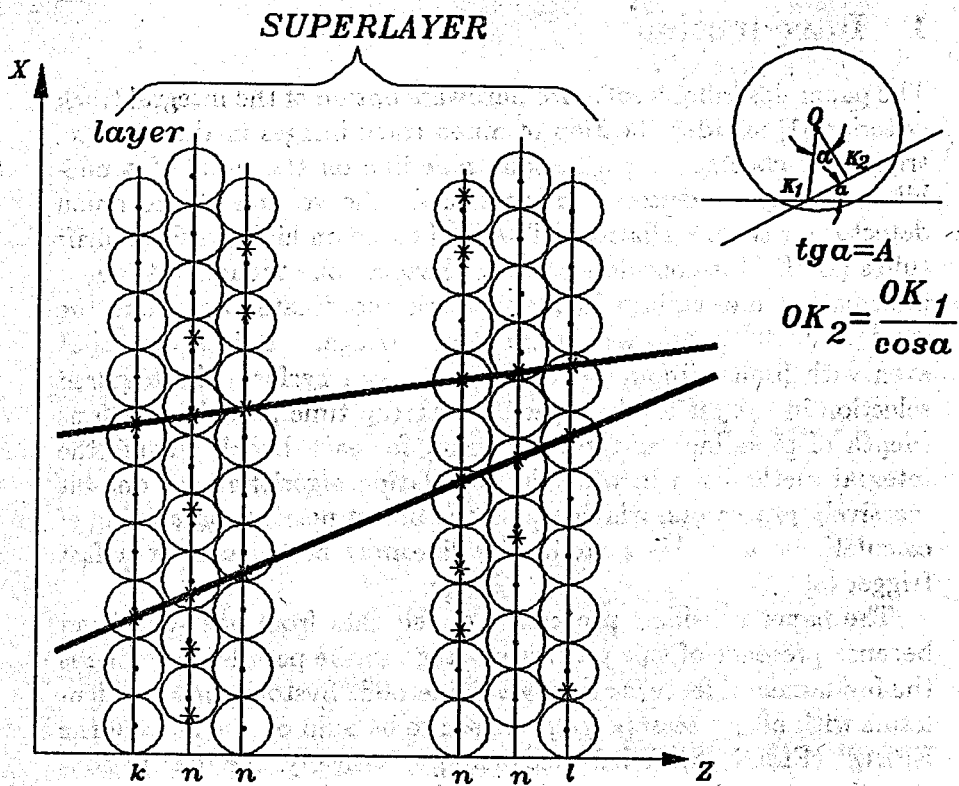


Figure 1. View of superlayer

$X_m(Z_n) = AZ_n + B$  - real point-ordinate on the layer  
 $a_{mn} = P_{mn} + S_{mn}V(t_{mn} - T_0)$  - measurement coordinate on the layer

Where:

- $T_0$  - general delay on the marking off time in superlayer
- $t_{mn}$  - measured drift time
- $V$  - drift velocity
- $S_{mn} = \pm 1$  - drift direction
- $P_{mn}$  - HPDT wire anode ordinate
- $Z_n$  - coordinate of layer  $n$
- $M_n$  - multiplicity per layer

of the delay in the transfer channels, which is known in advance. The values of  $t_i$  may arrive at the trigger mixed with respect to the source of their origin, i.e. in the trigger there can be data arrived from different bunch intersections. Thus the times  $T_0$  are not known in advance, and separation of the data in  $T_0$ , i.e. in bunches, is one of the results of trigger operation. It is also necessary to establish the presence of track from the given event in the muon detector and to find their slope angles.

The times  $t_i$  arrive at the pipe-line for each bunch, and it is not known either from which tubes the signals come or what the direction (rightward-leftward), of the charge drift in each tube is. Thus, we have a sign-position uncertainty. The drift rate in an HPDT [2] is  $50 \mu\text{m/ns}$ . The signals coming to the pipe-line can be a noise or result from passage of one or several tracks through the detector. Since the superlayer is small in radius and the muon energy is high, the muon trajectories in the superlayer can be considered straight. Another limitation of the model in question is a small range of particle entrance angles ( $\pm 10^\circ$ ).

### 3 Integral model of track images for the muon detector

The integral mathematical model of multitrack images is defined as a function of trajectory parameters and involves all measured coordinates [1]. In an individual superlayer (SL) muon trajectories can be approximated by straight line sections. The coordinates of the points on the layers are

$$X_m(Z_n) = AZ_n + B,$$

where  $n = 1, N$  is the number of layers  $Z_n$  is the coordinate of the layer  $n$ . For each real  $X(Z_n)$  we have a measured coordinate

$$a_{mn} = P_{mn} + S_{mn}V(t_{mn} - T_0),$$

where  $T_0$  is the total delay of time counting in a superlayer;  $t_{mn}$  is the measured drift time in a layer;  $V$  is the particle drift velocity;

$S_{mn} = \pm 1$  is the drift direction sign inside the tube;

$P_{mn}$  is the ordinate of the anode wire in HPDT mm.

It is assumed that on each  $n$ -th layer there can be  $M_n$  counts ( $m = 1, \dots, M_n$ ) belonging to real trajectories or having noise origin.

So, in this case the mathematical expression of the "Integral model" has the form:

$$R(A, B, T_0) = \sum_{n=1}^N \sum_{m=1}^{M_n} G(AZ_n + B - a_{mn}(T_0); \sigma_{mn}). \quad (1)$$

Here  $\sigma_{mn}$  is the measurement accuracy. The function  $G(Arg, \sigma)$  is defined as follows:

$$G(Arg; \sigma) = \begin{cases} 1 & , \text{ if } |Arg| \leq \sigma_{mn} \\ 0 & , \text{ if } |Arg| > \sigma_{mn}. \end{cases}$$

In the expression  $R(A, B, T_0)$  we have 3 unknown quantities. However, we can eliminate  $A$  and  $B$  in the following way. Let us take two reference layers (e.g.,  $k$  and  $l$ ). For a point on the track ( $A, B, T_0$ ) we write down the equalities:

$$\begin{cases} AZ_{\chi k} + B = P_{\chi k} + S_{\chi k} V(t_{\chi k} - T_0) \\ AZ_{\lambda l} + B = P_{\lambda l} + S_{\lambda l} V(t_{\lambda l} - T_0), \end{cases} \quad (2)$$

where  $t_{\chi k}, t_{\lambda l}$  are measured times,  $\lambda$  and  $\chi$  are indices corresponding to the chosen times on the layers  $l$  and  $k$ . Expressing  $A$  and  $B$  in terms of (2) with allowance for all available values of time  $t$ , we obtain a function of one variable  $F(T_0)$  instead of  $R(A, B, T_0)$ :

$$F(T_0) = \sum_{n=1}^N \sum_{\chi=1}^{M_k} \sum_{\lambda=1}^{M_l} \sum_{\mu=1}^{M_n} G[Arg(T_0); \sigma_n N], \quad (3)$$

$$Arg(T_0) = T_0[S_{\mu n} - S_{\lambda l} - h(S_{\chi k} - S_{\lambda l})] - \{[P_{\mu n} - P_{\lambda l} - h(P_{\chi k} - P_{\lambda l})]/V + S_{\mu n} t_{\mu n} - S_{\lambda l} t_{\lambda l} - h(S_{\chi k} t_{\chi k} - S_{\lambda l} t_{\lambda l})\}, \quad (4)$$

$$h \equiv \frac{Z_n - Z_k}{Z_k - Z_l}$$

For real trajectories one has  $F(T_0) \geq N - 2$ ; the value of  $F(T_0)$  may be larger than  $(N-2)$  because of the combinatorial effect. Thus,

calculating  $F(T_0)$  for a sufficient number of  $T_0$  (from 16 to 32, according to the real experimental conditions) and specifying the threshold value  $F_{min}$ , one can find maxima in  $T_0^*$  corresponding to the real trajectories and slope angles  $A$ . Thus, the data are sorted by  $T_0$  for bunches, and the track directions are found by angle  $A$ .

In formula (4) we do not know  $P$  and  $S$  for all indices. Knowing that  $P$  and  $S$  are related to each other and the drift tubes have a staggered arrangement, one can overcome their "position and sign" uncertainty by making a set of  $PS$  combinations for a single straight-line track in the given SL geometry. For the slope angle range  $(-10^\circ \leq A \leq 10^\circ)$  the number of these combinations is  $\sim 100$  (see the Table). Thus, by exhausting all  $PS$  combinations and all discrete values of  $T_0$  one can find all elements of tracks in the given superlayer, their slopes, and separator according to the bunches.

#### 4 Pipe-line of one-bit processor array

The function  $F(T_0)$  can be calculated for each  $PS$  combination and  $T_0$  independently, so the calculation algorithm will be described for one set  $(PS, T_0)$ . The values of  $P, S, T_0$  are known,  $P$  and  $S$  being constant on each layer with each particular  $PS$  combination. The trigger data coming from the detector for each bunch are transformed into quantities  $t_{l\lambda}, t_{k\chi}, t_{n\mu}$ , where the multiplicity for each layer  $l, n, k$  does not exceed 6. Thus, at the input we have a matrix

$$D = \begin{pmatrix} t_{11} & t_{k1} & t_{n11} \\ t_{12} & t_{k2} & t_{n12} \\ \vdots & \vdots & \vdots \\ t_{16} & t_{k6} & t_{n16} \\ & & t_{n21} \\ & & \vdots \\ & & t_{n26} \\ & & t_{n31} \\ & & \vdots \\ & & t_{n46} \end{pmatrix}$$



The rest of the algorithm is quite clear but, since the number of operations with the same variables is large, all the operations should be done in the one chip, leaving communication to the chip. The criteria for placing the algorithm in the package are:

- output of the result every 25 ns;
- realization of the algorithm inside chip with the greatest possible number of variants for  $PS$  combinations and a decrease in the data flow at the output by an order of magnitude.

It is proposed to calculate  $Arg$  and  $G$  by the systolic method by fulfilling the following operations in series:

Step 1.  $P_{\chi\lambda} = t'_{i\lambda} + t'_{k\chi}$ .

Step 2.  $H_{\lambda\mu\chi} = P_{\chi\lambda} + t'_{\mu n}$ .

Step 3. Calculation of  $|H_{\lambda\mu\chi}|$ .

Step 4.  $Q_{\chi\lambda\mu} = |H_{\lambda\mu\chi}| + (-\sigma)$ .

Step 5. Summation of all  $S_{\chi\lambda\mu}$ .

Step 6. Comparison of the result obtained at step 5 with the threshold; if it exceeded, a signal should be produced to indicate presence of a track for the given  $T_0$  and  $PS$ .

#### 4.1 Data processing with parallel code

A systolic matrix is built with 25 columns of 36 one-byte summator "triplets" each (in the first column of each line there is not a "triplet" but a single summator). We call the upper 6 lines of the matrix the first group, the next 6 lines the second group and so on. All 12 values of  $t_k$  and  $t_l$  come to the first column, each  $t_{k\chi}$  arriving at all 6 summators of group number  $\chi$  and each  $t_{i\lambda}$  arriving at summators number  $\lambda$  of each group. This is how step 1 is fulfilled. The results of operation 1 are fed through the lines to the inputs of the first summators in each "triplet" (i.e.  $P_{\lambda\chi}$  arrive at the first summator of all columns, starting from the 2nd one, in line  $\chi\lambda$  and so on). Simultaneously, all  $t_n$  come to the inputs of the same summators,  $t_{n\mu}$  being fed into inputs of all the first summators of the  $(\mu + 1)$  column and thus step 2 being fulfilled. The output signal from each first summator of the "triplet" is applied to the input of the second summator of the "triplet" and the absolute value of the number

at the input is calculated in these summators ("summator" is an arbitrary name of the logic unit which performs calculation of the absolute value). The output data of these summators are sent to the third set of summators in the "triplet" while their second inputs receive the values of  $(-\sigma_n)$ . Thus steps 3 and 4 are fulfilled.

Since we deal with straight-line tracks, for each line  $\chi\lambda$  within the columns "related" to one layer of a superlayer (let us call them group "n") no more than a unity can appear at step 4 in the columns that received  $t_{n\mu}$  from one layer n. This means that step 5 is fulfilled in two stages for each line. First the OR results of step 4 are combined for each line within each group "n", and these 4 bits of information are sent through lines to the counters behind the 25th column (the counters are also organized into columns of 36 elements). At the output of each line we have three-bit sums of step 4 results (a total of 36 numbers), and these numbers are added up in fast counters [6]. The value obtained is compared with the threshold one and if it is exceeded, an output signal is produced to indicate a candidate for the track required.

All steps are fulfilled as pipe-line processing, i.e. while the next step is being fulfilled, the input of the preceding one receives another portion of trigger data from the detector. For promising EPLDS technologies, the time to fulfill all operation for each of the above-mentioned steps may be below 25 ns with allowance for delay time to transmit information within the package. The total depth of the pipe-line with allowance for delay time to receive values of input data matrix  $D'$  is 40 steps (i.e. 1000 ns), but the result of processing of each data portion comes to the output of the flow calculator in a cycle equal to the bunch duration.

However, this approach has disadvantages. All calculations must be performed within packages of systolic matrices (at a clock frequency 100 MHz now and 250 MHz in 1998) [4]. Each summator requires 1 macrolocation (LAB) of the package. The total number of LABs in a package is 126 (in future the number will be increased by a factor of 8-10 [4]). So the complete realization of the algorithm for one value of  $T_0$  and one  $PS$  combination demands quite a lot of microcircuits and makes communication more difficult, which also increases the startup value (pipe-processing depth). That is why the

parallel-serial principle is proposed for the systolic pipe-line based on such systolic matrices.

#### 4.2 Data processing with serial-parallel code

In this case, the principle of the calculator is identical with the one described above. However, in each clock only 2 bits of each using of values of data are sent to the inputs of the "summatoms". Step 1 of the algorithm is performed by two-bit summators (see fig. 2) with end-around carry. For 6-bit words the whole sum will be obtained in 3 chip clocks (at this moment the high-order bit carry is set to zero). At steps 2, 3 and 4 two bits of data are recorded in each of the similar summators in accordance with the above algorithm, but step 3 and 4 calculations are paralleled: the left summator calculates the value of  $H_{\chi\lambda\mu} + (-\sigma_n)$  and the right summator calculates the value of  $H_{\chi\lambda\mu} + \sigma_n$ . The sign of the result appears in  $ZN$  and  $Z$  at step 4. By this moment the sign "K" is obtained for  $(t'_{k\chi} + t'_{l\lambda} + t'_{n\mu})$  at step 2, and the value required for step 5 is determined in the unit  $U = \bar{K} \wedge ZN \vee K \wedge \bar{Z}$ . In this algorithm steps 2-4 of each clock period take only 6 LE (1 LAB = 8 LE), which is much fewer than in the parallel code. However, the result will be obtained at steps 1-4 only in 3 clock periods. Thus, if the operations in LE are performed in less than 8 ns (with allowance for carry-over), or in 3 ns in the future, one can fulfill steps 1-4 for each set  $(t'_k, t'_l, t'_n)$  within the bunch time. So, the output of the result is not slower. Besides, in this case the requirements to outer commutation and input loads can be far less severe.

Internal exhaustion operations (steps 2-4) take 6 LE. In principle, it can be 4 LE; for this purpose each summator of steps 3-4 (left and right) must be replaced by a unit shown in fig. 3. In this case the remaining free FLEX ALTERA registers of each LE are used, because one needs only the sign of the calculation result after steps 3-4. It should be mentioned that in this algorithm the number of parallel input bits can vary from 1 to 6, according to the particular conditions of the problem, without violating regularity, matrix character and homogeneity of the systolic calculating structure.

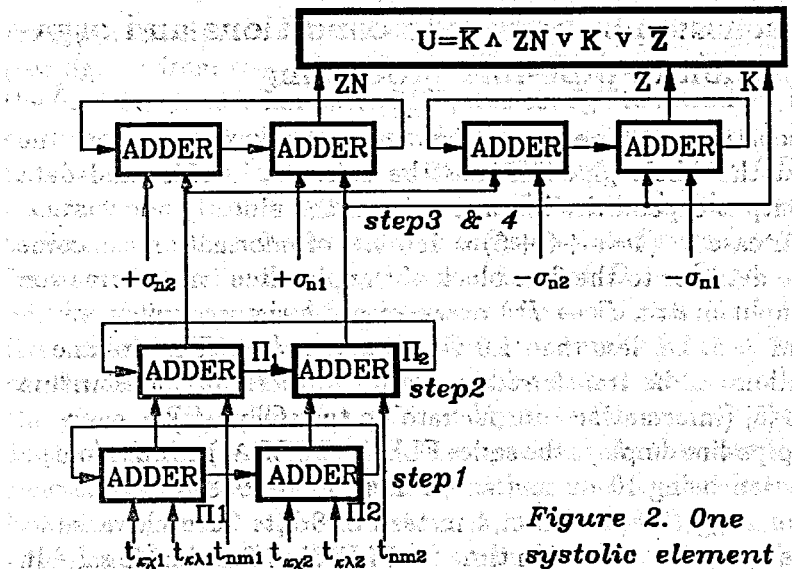
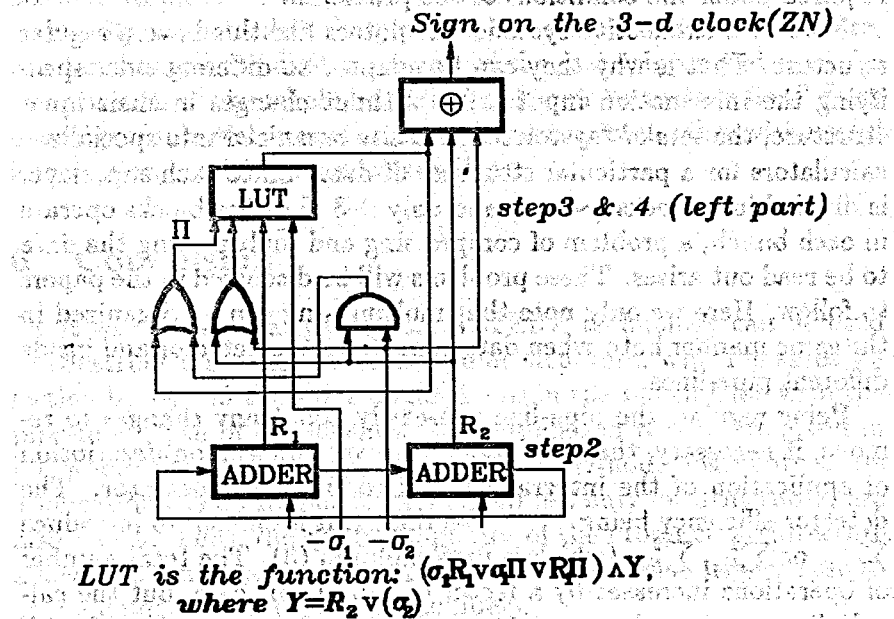


Figure 2. One systolic element



LUT is the function:  $(\sigma_1 R_1 \vee \alpha \Pi \vee R_2 \Pi) \Delta Y$ , where  $Y = R_2 \vee (\alpha_2)$

Figure 3. Part of one systolic element with use of combinatorial logic

## 5 Discussion of boundary conditions and organization of pipe-line processing

When constructing pipe-processing one must take into account the fact that the whole pipe-line must be balanced in rates and data input/output capabilities in some units of the pipe.

In our case  $5 * (4 * 6 + 6 + 6) = 180$  bits of information can come from the detector to the first block of the pipe-line (transformation of the input matrix  $D \rightarrow D'$ ) every 25 ns at the maximum multiplicity  $M = 6$ , i.e. less than 1.0 Gbyte per second. This volume of information can be transferred to the pipe-processing block with a SCI bus [5] (information transfer rate up to 1 Gbyte/s).

The pipe-line employs the series FLEX ALTERA, its input/output clock period being 10 ns now and 4 ns in 1996-1998 [4]. Thus one can enter trigger data even of 4 instead of 8 bits for each variable. Now the operation execution time for FLEX is 7.8 ns inside a LAB. So, inside the package the processing rate is not worse than it is required under the conditions of the problem.

Note that exhaustive systolic calculators like this have a regular structure. That is why they can be adapted to different rules specifying the information input order without changes in their inner structure, the sets of "systoles" are easily organized into specialized calculators for a particular structure of data. Since each superlayer is divided into blocks ( $\sim 800$ ) and only 2-3 % of the blocks operate in each bunch, a problem of compressing and multiplexing the data to be read out arises. These problems will be discussed in the papers to follow. Here we only note that multiplexing can be organized in the same manner both when data come from the detector and inside different pipe-lines.

Being regular, the pipe-line can easily adopt any changes to remove, if necessary, the restrictions that we imposed on description of application of the integral method to the muon detector. The detector efficiency being less than 1, it is enough to introduce  $\sum_{k=1}^N$  or  $\sum_{k=1}^2 \sum_{l=1}^2$  (which is preferable) in (3). The total number of operations increases by a factor of 4-6 in this case, but the calculation structure does not change despite a larger number of chip. Also, the particle drift in a tube was taken to be planar to the X-axis.

Actually, the signal is produced by particles arriving at the point O (see fig. 1) from the point K. So,  $OB = OK/\cos \alpha$ . Yet, owing to the fact that  $\alpha$  does not exceed  $10^\circ$ , it is enough to introduce 1 extra correction coefficient in  $PS$  combinations (1 value of the coefficient for  $\alpha = 0^\circ - 5^\circ$  another for  $\alpha = 6^\circ - 10^\circ$ ), and we obtain OB to an accuracy of  $\sigma$ . In this case the number of  $PS$  combinations becomes twice as large, but the exhaustion units do not change (coefficients for  $\alpha$  are taken into account at the stage of  $D \rightarrow D'$  transition). Finally, if  $t_i$  is measured together with the coordinate  $P_i$  of the tube from which the signal comes, the exhaustion units for particular  $PS$  combinations can be supplied only with those  $t_i$  that are related to the given  $PS$  combination. The delay time for this selection does not exceed 4 ns (one operation of comparison). In this case there are no changes in realization of the algorithm, and the integral method becomes even more powerful (the input multiplicity drops drastically, and the combinatorial output background decreases). Thus the calculation process in the pipe-line is sufficiently stable against possible variation of the initial data model.

Finally, we note that the pipe-line has a rather flexible structure, can be changed for different levels of the fast trigger and used for construction of the whole fast trigger. It can "smoothly flow" from one trigger sublevel to another complementing each other with useful information resulting from calculations.

## 6 Conclusion

So, it was shown that the FLEX ALTERA technology can be used to construct a pipe-line with a regular structure, which ensures selection by  $T_0$  and the slope angle  $A$  within the bunch time with the pipe-line depth of 40 bunches. The simulation results are given in [7], where the advantages of the integral method, possibility of its application to determination of  $P_1$ , and not only for drift tubes, are also shown. Thus, the systolic pipe-line of the proposed architecture allows the "integral method" to be used in the fast trigger system of the detectors for LHC, SSC colliders.

In the preceding section it was mentioned that the method becomes even more powerful if there is  $P_i$  for each  $t_i$ . Possibilities of



economical obtaining of this information will be discussed in future papers.

The authors are grateful to Z.V.Krumshtein and G.A.Chelkov (Dubna), E.S.Glushankov (St. Petersburg), V.S.Shmoilov (Lvov) for helpful discussions, to V.B.Rumyantsev and A.A.Kazakov (Dubna) for their assistance in preparation of the paper for publication.

## References

- [1] Yu.A.Yatsunenکو, NIM,A287(1990),422-430.
- [2] ATLAS, CERN/LHCC/92-4, LHCC/I 2, 1 October 1992.
- [3] V.Kotov and R.Poze: *Pipe processing for fast triggering*. CHEP92, Conference on computing in high energy, Physics—1992, 21-25 September 1992, Annecy, France, Abstracts.
- [4] ALTERA, Applications Seminar, Fall 1992.
- [5] SCI, Scalable Coherent Interface, P1596-23Jan91-doc210.
- [6] N.M.Nikityuk: *Use of parallel counters for triggering*, *Nuclear Instruments and Methods in Physics Research*. A321(1992)571-582, North Holland.
- [7] Yu.A.Yatsunenکو et al. To be published.

Received by Publishing Department  
on May 26, 1993.