Available online at www.sciencedirect.com

SciVerse ScienceDirect

Physics Procedia 37 (2012) 1884 – 1891

Physics Procedia

TIPP 2011 - Technology and Instrumentation in Particle Physics 2011

Use of GPUs in trigger systems

Gianluca Lamanna[a,*]

[a] *CERN, Geneva, Switzerland*

Abstract

In recent years the interest for using graphics processor (GPU) in general purpose high performance computing is constantly rising. In this paper we discuss the possible use of GPUs to construct a fast and effective real time trigger system, both in software and hardware levels. In particular, we study the integration of such a system in the NA62 trigger. The first application of GPUs for rings pattern recognition in the RICH will be presented. The results obtained show that there are not showstoppers in trigger systems with relatively low latency. Thanks to the use of off-the-shelf technology, in continous development for purposes related to video game and image processing market, the architecture described would be easily exported to other experiments, to build a versatile and fully customizable online selection.

Keywords: GPU, Trigger, RICH, NA62, rare decay

1. Introduction

In High Energy Physics (HEP) experiments the trigger system is used to on-line select interesting events, in order to reduce the bandwidth requirements and the disk space needed to collect data. In the last years, thanks to the growing of computing power in the PCs and the larger bandwidth of the networks, the possibility to use systems based on commercial equipment is becoming feasible. In particular the trigger software level in the nowadays systems, is a relevant part of the total computing farm in an experiment, while the hardware custom part is smaller. The "triggerless approach", in which all data are sent to the computing farm, is becoming a possible solution for many experiments in the next future. There are many advantages of using a trigger system based on PCs: versatility, flexibility, scalability, updatability, offline reproducibility and possibility to profit of the continuous development for the IT industry. In any case the high event rate in HEP experiment requires a huge computing power in the on-line farm. The GPUs offer the possibility to cover the gap existing between the present PCs and the computing speed needed for the on-line processing. This is due to the different architecture of the GPU with respect to the standard processors (CPU): a larger silicon area is devoted to computing units instead of control and caching (fig.1).

In this paper we discuss the benefits and the issues of a trigger system based on GPU. In particular we focus on the application in real time processing for the NA62 experiment. This experiment, briefly described

[*] Corresponding authors
Email address: gianluca.lamanna@cern.ch (Gianluca Lamanna)

doi:10.1016/j.phpro.2012.04.107

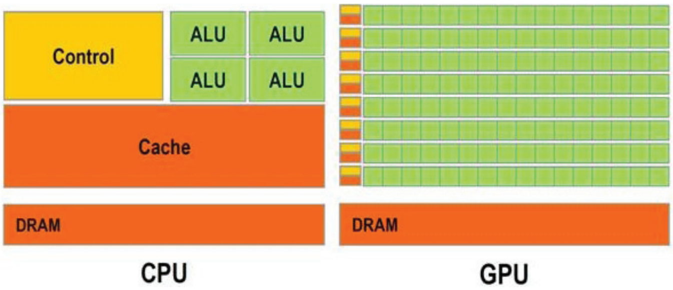

Fig. 1. CPU and GPU: different distributions of resources.

in par.4, aims at collecting O(100) events of the ultra rare kaon decay $K^+ \rightarrow \pi^+ \nu\bar{\nu}$, in two years of data taking. As first example of possible GPU's application, we discuss the 10 MHz pattern matching and ring fitting in the NA62 RICH (see [1] for more details).

2. The GPU (Graphics Processing Unit)

The problems related to the 3D and 2D rendering, video editing, and, in general, to image processing, drove the development of a new architecture of powerful dedicated processors. In the modern video cards, indeed, everything concerning graphics is done on board, leaving the CPU on the hosting computer free from this resources consuming task. The peculiarity of the application fits well in a high parallel structure of the processor, where several computing cores work together on different sets of data. This architecture is called SIMD (Single Instruction Multiple Data). Different levels of parallelization can be exploited in this scheme by grouping the multicore structure in different layers. Thanks to this structure the computing power can easily exceed the Tera Flop level in commodity systems (like desktop PC). In recent years the interest for this kind of architecture arose for the use of GPU in general purpose applications (GPGPU) [2], outside the field of image processing. Several examples can be found in literature going from lattice QCD calculation [3], to seismology and medical physics. The relative simplicity to program these devices contributes to encourage the spread of this new way of computing. The two main vendors (AMD and NVIDIA) are putting a lot of effort to provide, together with the hardware, a comprehensive framework to exploit the GPU features [4].

While the application for generic computing is nowadays common place, the possibility to use the GPU for real time purposes, like trigger of HEP experiment, has to be proven. The main difficulty is related to the latency of the processing: the high bandwidth of data transmission and the possibility to process several events at the same time, can "hide" the latency for single event, as will be discussed in the following. In table 1 we show the characteristics of the video cards used in our tests. The AMD Radeon HD5970 and the NVIDIA Tesla C2050 are in the top level of the market, while the NVIDIA C1060 comes from the previous

	NVIDIA Quadro 600	NVIDIA Tesla C1060	NVIDIA Tesla C2050	AMD Radeon HD 5970
Number of multiprocessors	2	30	14	20
Total number of cores	96	240	448	3200
Core frequency (GHz)	0.64	1.3	1.15	0.725
Main mem. (GB)	1	4	3	2
Main mem. bandw. (GB/s)	25.6	102	144	256
Fast on-chip mem. (KB/multiprocessor)	48	16	48	32
Computing power (TFLOPS)	0.246	0.93	1.03	4.6

Table 1. Characteristics of the device used in the test. The AMD board is a dual GPU device. In any case for GPGPU purposes only one processor can be presently used.

generation of devices (two years old) and the NVIDIA QUADRO 600 is an entry level video card for a standard desktop PC.

3. Use of GPUs in a trigger system

In standard trigger systems the complexity and the quality of the trigger primitives used to take a trigger decision is limited by latency requirements and possibility to sustain the given rate. Usually the trigger is subdivided in hardware and software levels. The hardware levels are, in general, faster and the decision is based on hit patterns of the event, while in software levels some kind of preliminary event reconstruction can help to obtain the required reduction factor. The idea to use GPUs (possibly hosted in standard PCs) in computing units could help to build high quality primitives for more selective decision. In this regard some critical point must be taken into account in designing such a system:

- data transfer from detector to PC: fast, reliable and time-deterministic dedicated links must be employed in order to exploit a large bandwidth;
- small latency and high rate: high computing power in a limited number of machines in order to cope with a high rate avoiding the use of complex network infrastructure;
- stable latency: mainly in synchronous application, the latency spread should be small with a low level of tails.

A GPUs based system can easily address the two last points: the video card processors have a huge computing power and, since the functionality inside the chip are limited to computing, a quasi deterministic behavior. On the other hand the use of PC to host the GPU cannot be avoided. Standard PC, running the operating system, have somewhat unpredictable timing characteristics: several precautions should be used in order to decrease the possible fluctuations introduced by the hosting PC, like the use of real time operating system and protocol offload in the Network Interface Card (NIC). Anyway in order to decrease the contribution of the latency and jitter, the buffering of the events should be foreseen in readout board or in interface card. The events have to be grouped in packets in order to optimize the data transmission through the PCI express bus and to allow the concurrent processing of several events at the same time.

4. The NA62 experiment

In this section we will briefly introduce the NA62 experiment, focusing in particular on the trigger system. Further details on the experiment can be found in [5]. The NA62 experiment aims at measuring O(100) $K^+ \to \pi^+ \nu \bar{\nu}$ events in two years of data taking. This decay mode has a Branching Ratio (BR) of $(8.7 \pm 0.7) \times 10^{-11}$ precisely predicted in the Standard Model (SM), with a theoretical irreducible error of few percent. For this reason $K^+ \to \pi^+ \nu \bar{\nu}$ is a "golden mode" to test the CKM structure of the SM and an excellent probe of new physics beyond the SM, in a complementary way with respect to the direct search.

From an experimental point of view, the measurement is very challenging due to the smallness of the BR and the presence of a large background, mainly from $K^+ \to \pi^+ \pi^0$ and $K^+ \to \mu^+ \nu$ (with BR of 20.7% and 63.5% respectively). The present measurement of this decay is based on 7 candidates collected by E787+E949 Brookhaven experiments [6] leading to a value of $BR = (1.74^{+1.3}_{-0.89}) \times 10^{-10}$. Differently with respect to these experiments, NA62 will exploit the kaons decay in flight technique using an unseparated hadron beam (kaon ~ 6%) of 75 GeV/c, produced from 400 GeV/c protons from the CERN SPS impinging on a fixed beryllium target. The decay region will be housed in a ~ $70m$ long ~ 2.5 m in diameter vacuum tube, in order to reduce the secondary interactions of both the decay products and the primary beam. In fig.2 the layout of the experiment is shown. The background rejection will be based both on high resolution kinematics reconstruction and on veto systems. The former will be achieved by measuring with high precision both the out coming pion momentum (STRAWS tracker) and the incoming kaon momentum (GIGATRACKER), while the photon veto rejection will be done using rings made of lead glass blocks (LAV) for the photons with a large angle, an electromagnetic calorimeter (LKr) for the photons in the forward

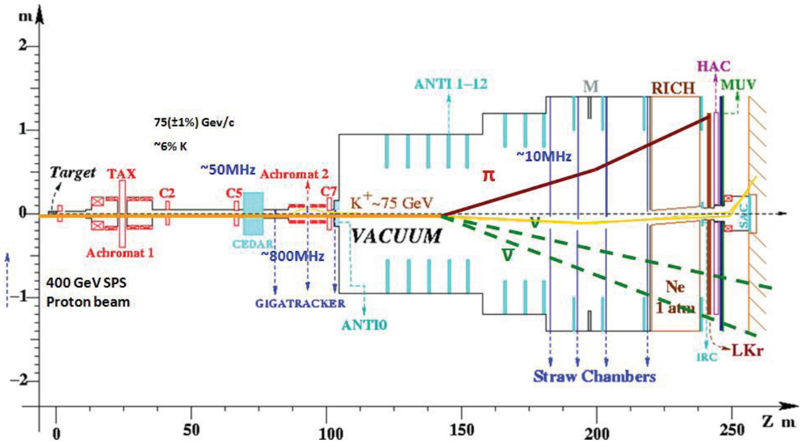

Fig. 2. Layout of the NA62 experiment.

direction and other small calorimeters (IRC and SAC) for the photons close to the beam line. The particles identification system (CEDAR, RICH and MUV) will help in rejecting the non kinematically constrained component of the total background.

The RICH, in particular, must identify pions and muons in the momentum range 15 GeV/c to 35 GeV/c, giving a μ suppression factor better than 10^{-2} with a good time resolution. Čerenkov light is produced in a 18 m long, 3.7 m wide tube filled with neon at atmospheric pressure. The light is reflected by a composite mirror of 17 m focal length, focused on two separated spots. The two spots are equipped with ~ 1000 PMs of 1.8 cm in diameter each. After amplification and discrimination [7], the PM signal time is digitized by high resolution TDCs. A typical pion ring, for averaged accepted momentum, is identified with ~ 20 firing PMs, as predicted by Monte Carlo and confirmed with a full-length prototype [8]. The time resolution was measured to be better than 100 ps for all momenta in the considered range.

4.1. The NA62 Trigger system

In order to collect O(100) SM events an intense beam and a reliable data acquisition and trigger system (TDAQ) are needed. An efficient on-line selection of candidates represents an important issue for this experiment because of the large reduction to be applied on data before tape recording. On the other hand, a loss-less data acquisition system is mandatory to avoid adding artificial detector inefficiencies when vetoing background particles; this last requirement is less common in standard readout and trigger systems. For the above reasons the NA62 and DAQ system are integrated in a completely unified digital system. In order to reduce the event rate from 10 MHz to tens of kHz, the TDAQ is structured in a three level system. The first level (L0) will be completely hardware based while the other levels (L1 and L2) will be based on software: the L1 decision is taken on single-subdetector reconstructed quantities, while the L2 decision is taken on the fully reconstructed event with high resolution. The L0 trigger primitives are constructed in the same board (TEL62) in which the data are stored to wait for the trigger decision. The TEL62 board is a general purpose board (an upgraded version of the TELL1 board developed by EPFL for the LHCb experiment [9]) with 5 FPGA and large buffers to store the data, waiting for the trigger decision delivered to the board through the CERN standard TTC interface. On the TEL62, up to 4 daughter boards can be mounted. For the majority of the detector in NA62, time is the most important information to provide. Therefore a daughter board with 4 HPTDC [10] has been designed, in order to have 512, 100 ps time resolution, channels in a single TEL62 board. The trigger primitives produced in the TEL62 are sent to L0 trigger processor (L0TP) using an Ethernet connection with low level protocol. The reduction factor of 10, at L0, will be obtained using positive information from RICH and CHOD the veto information from LKr, MUV and LAV. The L0 trigger decision is broadcasted through TTC to all the TEL62 with a fixed latency of 1 ms. The interesting events will go into the L1 via Ethernet connection. The L1 will apply, in software, a further reduction factor of 10.

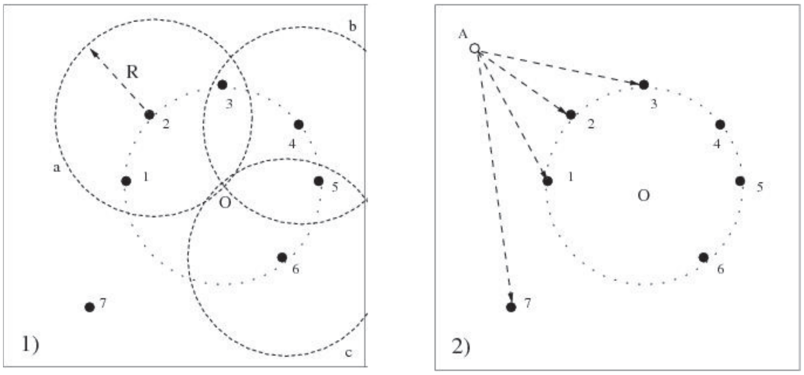

Fig. 3. 1) GHT algorithm. 2) DOMH algorithm.

The decision at this level is based on the data coming from single detector, in particular STRAW and RICH. In the L2 the full event is reconstructed and a more complete event selection is applied in order to reduce the rate to tens of kHz. The latency in the software levels is not defined, but all the events have to be processed before the next accelerator burst (order of 20-30 s).

5. Fast ring reconstruction for the NA62 RICH detector

As a first example of GPU application in the NA62 trigger system we studied the possibility to reconstruct rings in the RICH. The center and the radius of the Čerenkov rings in the detector are related to the angle and the velocity of the particle. This information can be employed at trigger level to increase the purity and the rejection power for many triggers of interest. The ring reconstruction could be useful both at L0 and L1. In both cases, because of the high rate of 10 and 1 MHz respectively, the computing power required is significant. The GPUs can offers a simple solution of the problem. The use of video cards in the L1 is straightforward: the GPU can act as "coprocessor" to speed up the processing. On the other hand the L0 is a small latency synchronous level, and the possibility to use the GPU must be verified. In order to test feasibility and performances, as a starting point we have implemented five algorithms for single ring finding in a sparse matrix of 1000 points (centered on the PMs in the RICH spot) with 20 firing PMs ("hits") on average. Since most of the events in kaon decays have a single charged track, we focused in the beginning on single ring recognition, the case of multi-rings will be treated later in the section.

In order to achieve the best performance the GPU architecture must be carefully considered. The computing cores are grouped in "multiprocessors", sharing memory and instruction pool. The access to on-chip memory is very fast (up to 1 TB/s) if read and write conflicts are carefully avoided. The processes (the "threads") running in each core must be synchronized at the multiprocessor level to maintain concurrent execution without divergences and partial serialization. The parallelization of the algorithm is easily obtained exploiting the parallel structure inside a multiprocessor, while several multiprocessors can be used to process several events at the same time.

The first algorithm we tested is based on a Generalized Hough Transform (GHT) (fig.3). In this approach each hit is considered as the center of a probe circle with a fixed radius. The point with the largest number of intersecting circles, varying the radius over the range 5 to 11 cm in steps of 2 cm, is considered as the center of the Čerenkov ring. The limitation for this algorithm comes from the amount of the on-chip fast memory available. This fact put a limit on the size of the three dimensional parameters space (center position and radius) used for the maximization procedure. The advantage is that, for each event, only a limited number of threads (equal to the number of hits) has to run concurrently on the GPU.

In the POMH (Problem Optimized Multi Histograms) and in the DOMH (Device Optimized Multi Histograms) approaches each point in the grid is considered as candidate for a center (fig.3). An histogram of the distances between the center candidate and the hits is constructed in order to identify the true center and

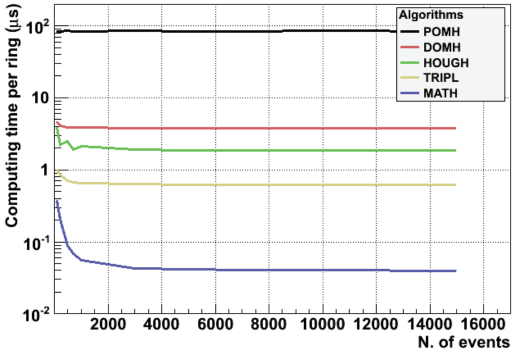

Fig. 4. Computing time for different algorithms measured with Tesla C1060 board.

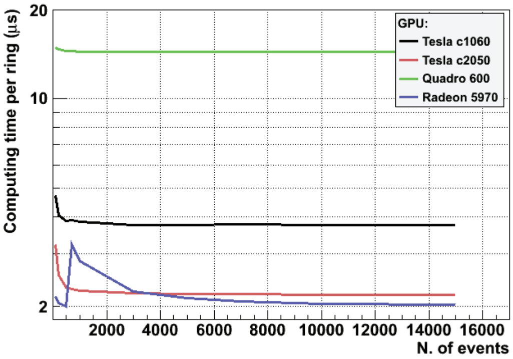

Fig. 5. DOMH algorithm: comparison on different boards.

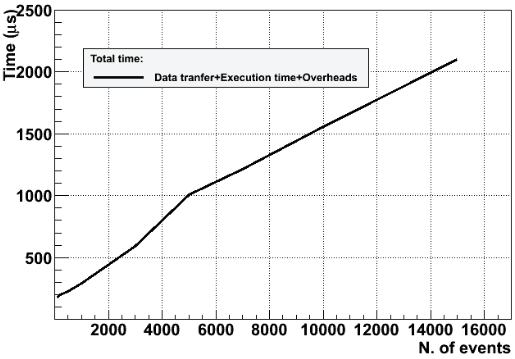

Fig. 6. Total time to process a batch of N events, including all the contributions.

the corresponding radius. The difference between the two algorithms is in the management of the parallelization structure in the GPU. In the POMH case each core has to make very simple operations (calculate the distance) but the whole processor in used only for one event; in the DOMH the single core operations are a little bit more complicated in order to optimize the number of concurrent processes with the read and write procedure in the fast memory, allowing multiple processing of events at the same time.

The last two algorithms we tested exploit more the computing power than the fast memory access with respect to the previous algorithms. In the TRIPL algorithm the center of the ring is obtained averaging the intersecting points of the axis of the segments connecting random sets of three hits ("triplets"). To reduce the noise contribution and to increase the resolution of the method, we consider many triplets in each event. In the last algorithm we tested, called MATH, the least-squares method is applied in a coordinate system in which the problem can be analytically solved [11] with a linear inversion.

In fig.4 the computing times are shown (Tesla C1060). For a packet of more than 1000 events the MATH algorithm takes ∼ 55 ns/event. In fig.5 the comparison between different video cards is shown in case of DOMH algorithm (most sensitive to the different GPU characteristics).

In a synchronous system the processing time jitter is an important issue. In case of the MATH algorithm this has been measured as ∼ 0.5% with negligible tails. The computing time determines the maximum event rate achievable by the system. The maximum performances are obtained with bunches of many events, as shown in the plot above. Since the time to copy data between PC and video cards has been measured to scale almost linearly with the number of events, the total latency is highly influenced by the size of the packet. In fig.6 the total time to process a certain number of events is shown: in addition to computing and transfer

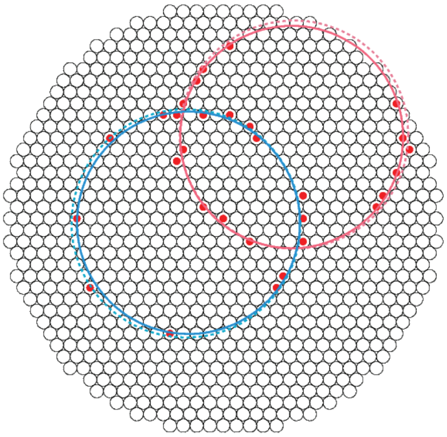

Fig. 7. Results of the ALMAGEST ring search algorithm. The dashed lines represents the generated rings while the solid lines the fitted rings.

time, the contribution of the time spent by the host PC to set the GPU operation by the driver (overheads) is also considered. For the application in the NA62 L0 trigger level, other times "external" to the GPU operations have to be taken into account. The time to transfer data through the Ethernet as been measured to be ~ $50\mu s$ for packets of 1.5 kB, while the transfer time, for a similar batch of data, from the ethernet card to the RAM is measured to be order of ~ $10\mu s$. Including all the contributions, as an example, with a batch of 100 events (that at 10 MHz are collected in $10\mu s$) the total time to have the trigger decision is ~ $250\mu s$, well below the 1 ms requirements for the L0 latency.

5.1. Fitting two rings

After the L0, only 50% of the remaining events has a single track while the others are multi-track ones. The majority of this components comes from $K^+ \to \pi^+\pi^+\pi^-$. Due to the RICH detector acceptance practically all the 3 tracks events are separated in the two spot rings generating not more than 2 rings per spot. The interest to search for 2 or more rings, at the trigger level, is related to both the suppression of the 3 charged pions background and the positive detection of other interesting decay modes, like, for instance, $K^+ \to \pi^+\mu^+e^-$.

At L1 an algorithm to search for two rings has been implemented. Given the structure of the trigger architecture, there is no possibility to have information from the spectrometer at this early time. For this reason the multi-ring searching algorithm has to be "trackless" without assumptions on the ring position. To cope with the high rate at L1 (1 MHz) the algorithm has to be "fast" and with "high resolution" in order to have good quality of the selection. We developed a new algorithm with these characteristics suitable to be parallelized on the GPU. This algorithm, called ALMAGEST, is based on the Ptolemy's theorem, that gives a criteria to decide if 4 points lie on a circle. In a given thread on the GPU, three points are selected randomly and other points belonging to the same ring are then identified. In the GPU several processes can run concurrently: this can be exploited to decrease the inefficiency due to wrong random choice of the first three points. After this pattern recognition phase, a fit procedure (like MATH described above) can be applied in order to obtain the circle parameters with high resolution. In fig.8 the computing time, on TESLA C1060 board, as a function of the number of the events is shown. In case of L1 the latency issues are less relevant since the requirements are less stringent with respect to the L0. Using the same idea it is possible to find an arbitrary number of rings, with an iterative procedure.

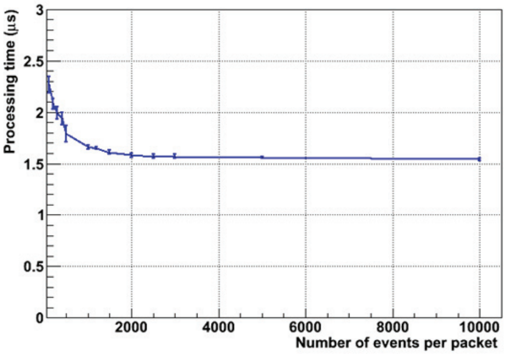

Fig. 8. Computing time to fit 2 rings using the Almagest algorithm.

6. Conclusions

The use of GPUs in the trigger of High Energy Physics experiments gives the possibility to design a very flexible and versatile system based on commodity equipment. The computing power of the GPU, exceeding of orders of magnitude the one of a CPU, offers a valid solution to address the problems of an high quality on-line selection. The latency introduced by the unavoidable procedure to transfer the data from the host PC on the computing device is not an issue in case of synchronous hardware levels with 1 ms of maximum latency and 10 MHz of maximum input rate. The software level farms can benefit from the use of GPUs as mathematical coprocessors, to reduce dimensions and costs. In this work we studied the GPU application in the trigger system of the NA62 experiment. In particular we focused on the ring finding in a rich detector both at L0 and L1 triggers levels. For the first trigger level we implemented a procedure to process events with a single ring in $\sim 55ns$ per event, keeping a total latency $\sim 250\mu s$. For the L1, where the latency is not an issue, we studied a completely new ring search algorithm based on the Ptolemy's theorem, giving a processing time of $1.5\mu s$ per event. No showstoppers are evident, at the moment, to implement a real trigger system based on video cards.

References

[1] G. Collazuol, G. Lamanna, J. Pinzino, M. Sozzi, Nucl. Instrum. Meth. A **662** (2012) 49
[2] `http://www.gpgpu.org/`
[3] G. I. Egri, Z. Fodor, C. Hoelbling, S. D. Katz, D. Nogradi and K. K. Szabo, Comput. Phys. Commun. **177** (2007) 631 [arXiv:hep-lat/0611022].
[4] `http://www.nvidia.com/cuda/`
[5] The NA62 collaboration - NA62 Technical Design - NA62-10-07 CERN, Geneva (2010).
[6] S. Adler *et al.* [The E949 Collaboration and E787 Collaboration], Phys. Rev. D **77** (2008) 052003 [arXiv:0709.1000 [hep-ex]].
[7] F. Anghinolfi, P. Jarron, F. Krummenacher, E. Usenko and M. C. S. Williams, IEEE Trans. Nucl. Sci. **51** (2004) 1974.
[8] B. Angelucci *et al.*, Nucl. Instrum. Meth. **A621** (2010) 205-211.
[9] G. Haefeli, A. Bay, A. Gong, H. Gong, M. Muecke, N. Neufeld and O. Schneider, Nucl. Instrum. Meth. A **560** (2006) 494.
[10] M. Mota, J. Christiansen, JSSC, vol 34, no. 10, Oct 1999
[11] J. F. Crawford, Nucl. Instrum. Meth. **211** (1983) 223-225.