P.N. Lebedev Physical Institute of the Russian Academy of Sciences

GPU-Accelerated processing for next generation nuclear emulsion scanning systems

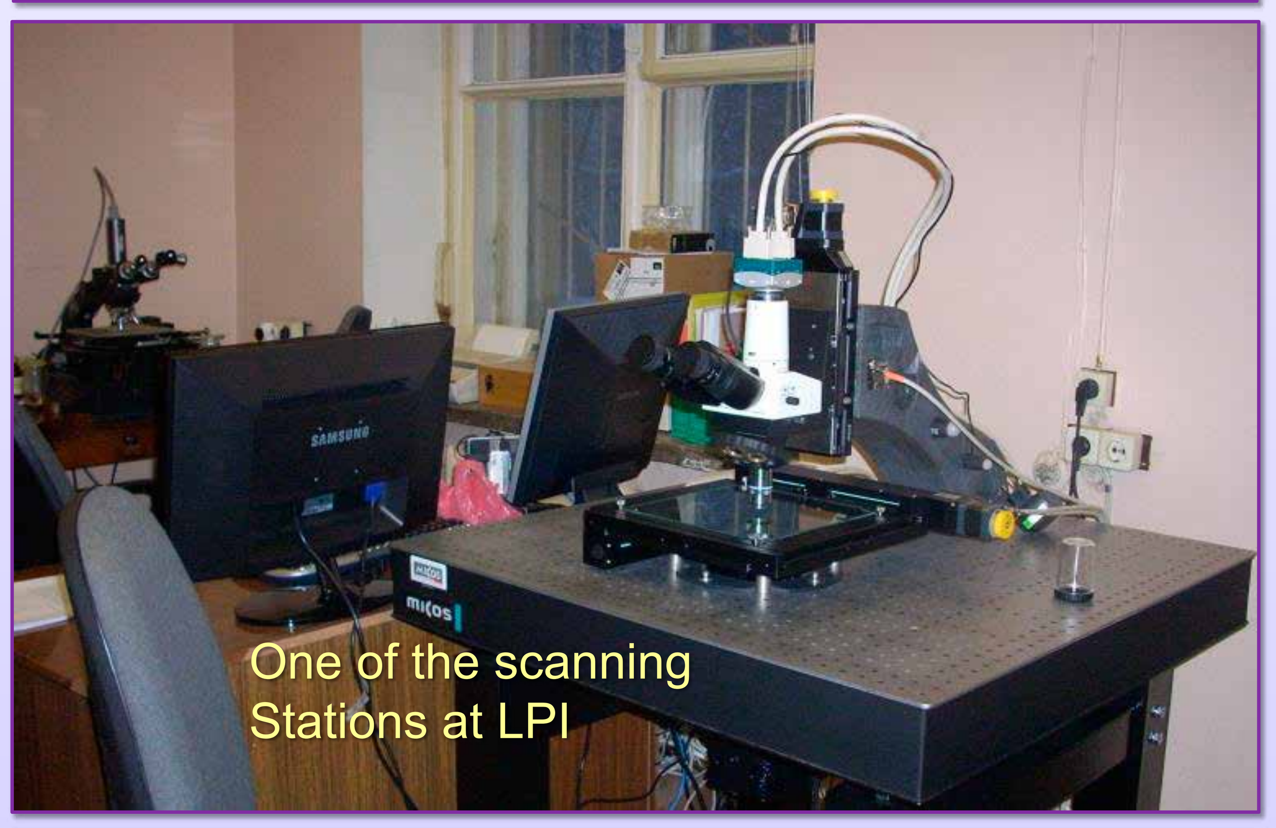
A. Aleksandrov[1,2], R. Mingazheva[1], N. Polukhina[1], T. Shchedrina[1,2], N. Starkov[1], V. Tioukov[2], M.Vladymyrov[1]
[1]-Lebedev Physical Institute, Moscow; [2]-INFN Naples department

Track detectors and automatic scanning systems

Track detectors, one type of which the nuclear photoemulsion is, are widely used in elementary particle physics over many decades. Long life of the technique is caused by a number of their advantages: the unique spatial (<1µm) and angular (~1mrad) resolution, visible spatial pattern of the particle interaction, relative simplicity and low price, the capability to accumulate information for a long time, etc.

More than 50 automated microscopes for track detector data processing operate over the world. Such an automated system is also working at the Lebedev Physical Institute - the completely automated measurement facility PAVICOM consists of three microscopes. The universality and great instrumental capabilities of the PAVICOM facility, implied during its development, allowed its application to a wide spectrum of problems: it is used to process almost all known types of solid tracking detectors.

Nowadays physical experiments that use nuclear emulsion need processing of huge area of it. For example OPERA experiment uses 110 000m^2 of nuclear emulsion films. Such amount of data must be processed in real time while scanning, preventing storage of data. Just to give an idea, the raw data of one scanned emulsion film (240cm^2) consumes about 700Gb of disk space, while the initial data-rate is 500Gbps. Currently scanning system like European Scanning System are scanning at speed of about 20cm^2/hour. Latest developments allow the increase of scanning speed up to 110cm^2/hour, making the processing speed to be a bottleneck. Here we will show the significant speed-up of the processing we've achieved by using GPU to keep the processing real-time.



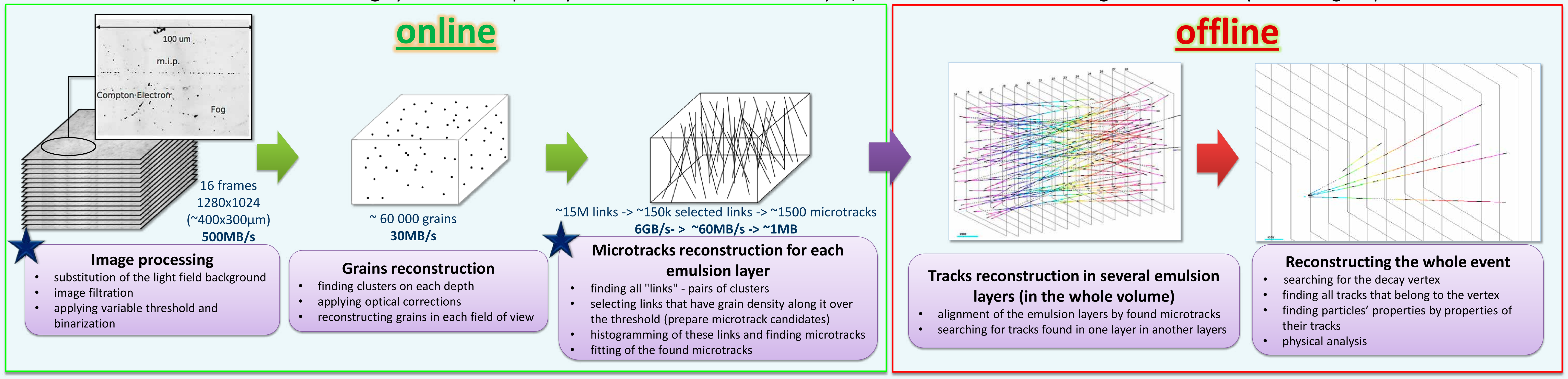
★ Image processing

Before the image processing was done on Matrox ODYSSEY XPro board. The whole processing time for one image of 1280x1024 px was 5.6 ms, which lead to 90ms per view (16 frames). The next generation scanning systems will have duty cycle less then 60 ms/view and resolution 2320×1726 px, making input data rate as much as 1.5Tbps. To keep real-time processing paradigm the whole image processing cycle was implemented in CUDA. The image processing is ideal for implementing on GPU: it satisfies requirements of data collocation and independence. So we have achieved great performance improvements: processing of one image on GTX 570 takes 0.46 ms and on Quadro 2000 1.2 ms. Quadro 2000 boards are present by default on Dell T7x00 workstations that are used for control of scanning procedure. Even using it we'll have processing time 63 ms/view with new camera, less then the scanning duty cycle, and suitable for next generations systems. 77% improvement!

Processing sequence in automatic scanning system

Charged particles while penetrating through the nuclear emulsion leave tracks which become visible in microscope after chemical development. Minimal ionization particles (m.i.p.) leave tracks that consist of separate grains.

The data obtained from the scanning system for 50 μm layer of nuclear emulsion is a sequence of 16 images grabbed at different Z-level for each field of view. Scanning is usually done for a region of thousands fields of view on several emulsion films. In online part of the processing microtracks (tracks in one emulsion layer) are reconstructed. Here is given the whole processing sequence:

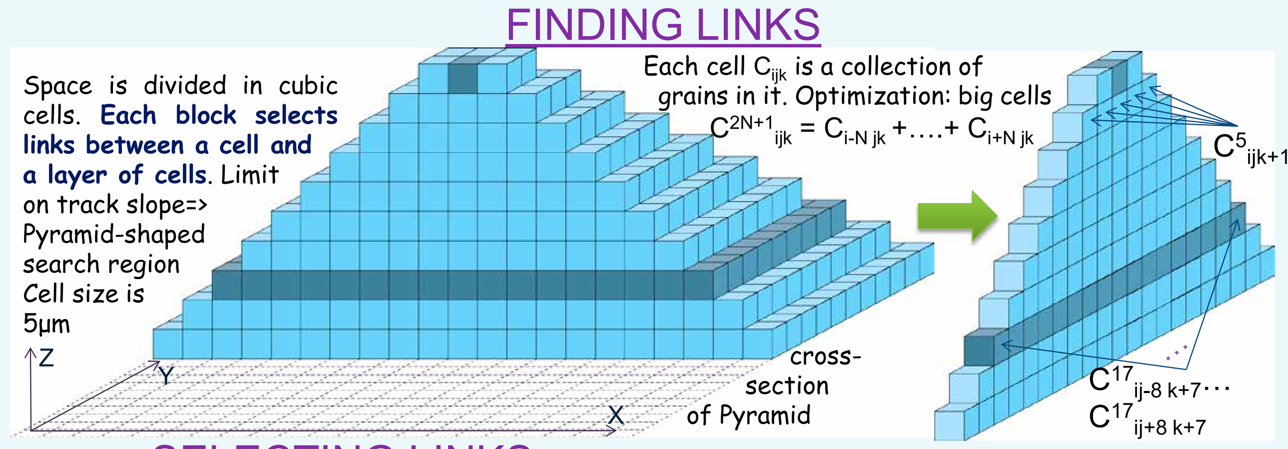


The most computationally intensive steps are image processing and tracking.

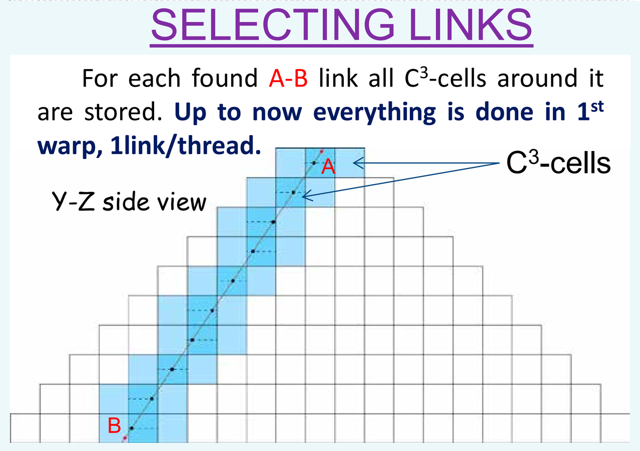
★ Hybrid CPU-GPU tracking

The most computationally intensive part of the processing cycle is microtrack reconstruction. A hybrid GPU-CPU tracking algorithm was successfully developed. Linking procedure is pretty suitable for GPU. It reduces the amount of data by factor 2, and is the most CPU-intensive.

FINDING LINKS



SELECTING LINKS



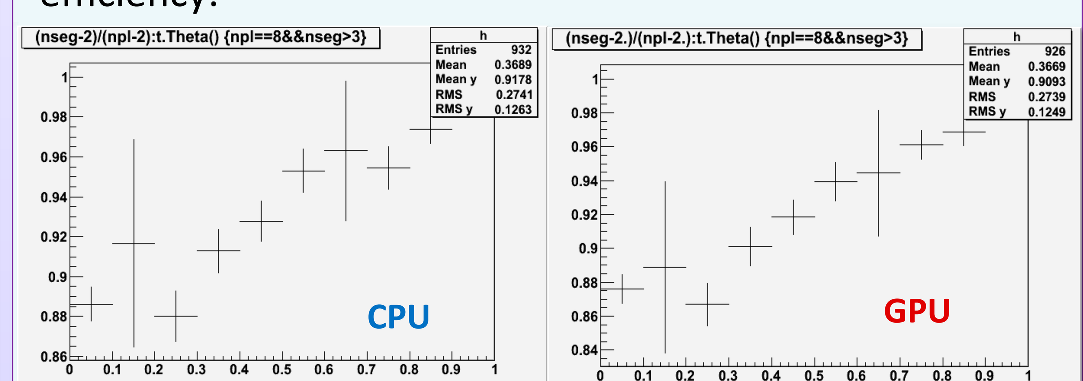
Sensitivity of nuclear emulsion is ~30grains/100µm, i.e. there will be 30 visible grains for 100µm of m.i.p. path.

While selecting links we can significantly reduce the background, by picking only those links which have number of grains within error cylinder higher then some threshold that corresponds to sensitivity of the emulsion. The problem is that the error cylinder's axes are not equal, more they depend on slope of the link. Thus finding if a grain is within the cylinder or not on the scale of 60k grains and 20M links will need pretty much of computation that involve trigonometrical calculations. The problem was faced in a tricky way: the whole space was rescaled according to grains' error ellipsoids so they become spheres.

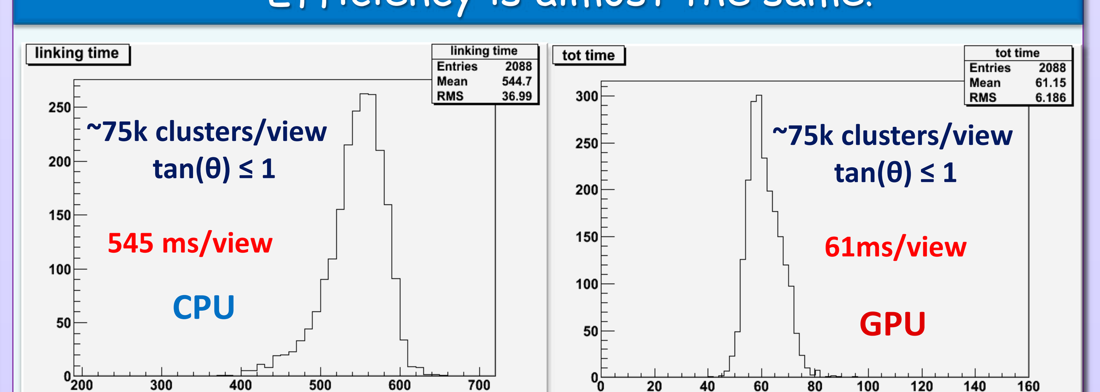
Using CC3.x shuffle functions we have implemented fast thread redistribution procedure to increase GPU utilization.

Testing and performance

To check the results we have processed same data set with both old CPU linking and GPU linking to compare efficiency:



Efficiency is almost the same!



Time is 89% less!

This processing was done on Dell T7500, Xeon X5650 2.66 GHz, 6 cores * 2 CPU *HT = 24 logical processors with GTX690 board. With CPU version of linking total time for tracks' slope limit tan(θ) ≤ 1 is ~550ms/view.

The algorithm is well scalable. A special task scheduling approach was developed to minimize overhead in many GPU systems. Running on 4 chips (2 GTX690) gives <5% overhead comparing to 1 chip processing.

This algorithm allows track reconstruction in fast scanning mode with just 2 GTX690 boards.

Large angle tracking

Many physical problems need reconstruction of tracks in emulsion that have large slope, up to tan(θ) ≤ 3 and even more in some cases. This is another important demand to next generation scanning systems.

We have developed n linking algorithm than can doesn't have slope limit (i.e. covers whole 2π sr.), with the link selection procedure described above.

Link search region geometry was revised to minimize shared memory needed for each particular subregion and optimize it's utilization.

Preliminary tests show that on 4 Titan boards it will be possible to perform tracking for tan(θ) ≤ 3 in real time.

Conclusion

The most computationally intensive parts of the processing were developed in CUDA. The performance and scalability achieved allow real-time data processing from images to microtracks at scanning speed (60ms/view, ~70cm^2/hour, data-rate 500Gbps) on only 1 workstation with 2 GTX690 boards, output data is at level of 20Mbps.

The presented work is used for processing data of the OPERA experiment in Lebedev Physical Institute and in Naples department of INFN, as well as for processing of data of muon radiography studies of Stromboli volcano and others. Also it will be used for further muon radiography studies.

Recent developments dedicated to next generation scanning systems will allow emulsion data acquisition with track reconstruction at large angles which is extremely important in some physical studies.

Acknowledgements

Speaker would like to thank his Closest Person for all the inspiration given.