

# GPU-accelerated automatic identification of robust beam setups for proton and carbon-ion radiotherapy

F Ammazalorso<sup>1,2</sup>, T Bednarz<sup>3</sup> and U Jelen<sup>1</sup>

<sup>1</sup> Department of Radiotherapy and Radiation Oncology - Particle Therapy Center, University of Marburg, Marburg, Germany

<sup>2</sup> Department of Medical Physics in Radiation Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>3</sup> CSIRO Mathematics Informatics & Statistics, Sydney, Australia

E-mail: [filippo.ammazzalorso@uni-marburg.de](mailto:filippo.ammazzalorso@uni-marburg.de)

**Abstract.** We demonstrate acceleration on graphic processing units (GPU) of automatic identification of robust particle therapy beam setups, minimizing negative dosimetric effects of Bragg peak displacement caused by treatment-time patient positioning errors. Our particle therapy research toolkit, RobuR, was extended with OpenCL support and used to implement calculation on GPU of the Port Homogeneity Index, a metric scoring irradiation port robustness through analysis of tissue density patterns prior to dose optimization and computation. Results were benchmarked against an independent native CPU implementation. Numerical results were in agreement between the GPU implementation and native CPU implementation. For 10 skull base cases, the GPU-accelerated implementation was employed to select beam setups for proton and carbon ion treatment plans, which proved to be dosimetrically robust, when recomputed in presence of various simulated positioning errors. From the point of view of performance, average running time on the GPU decreased by at least one order of magnitude compared to the CPU, rendering the GPU-accelerated analysis a feasible step in a clinical treatment planning interactive session. In conclusion, selection of robust particle therapy beam setups can be effectively accelerated on a GPU and become an unintrusive part of the particle therapy treatment planning workflow. Additionally, the speed gain opens new usage scenarios, like interactive analysis manipulation (e.g. constraining of some setup) and re-execution. Finally, through OpenCL portable parallelism, the new implementation is suitable also for CPU-only use, taking advantage of multiple cores, and can potentially exploit types of accelerators other than GPUs.

## 1. Introduction

The finite range of protons and carbon ions (Bragg peak) contributes to their dose deposition precision, unmatched among external beam radiotherapy modalities [1], but also sets high requirements on patient setup reproducibility. Displacement of density interfaces (e.g. air-bone, common in the cranial region) can introduce changes in particle penetration depth, with consequent adverse dosimetric effects [2,3,4]. These may have extent larger than the positioning errors that caused them and may not be fully counteracted by the use of safety margins, a situation unknown to the more conventional treatment with high energy photons.

A possible solution, which well complements accurate patient immobilization and image-guided position verification, is the application of robust treatment planning approaches, for



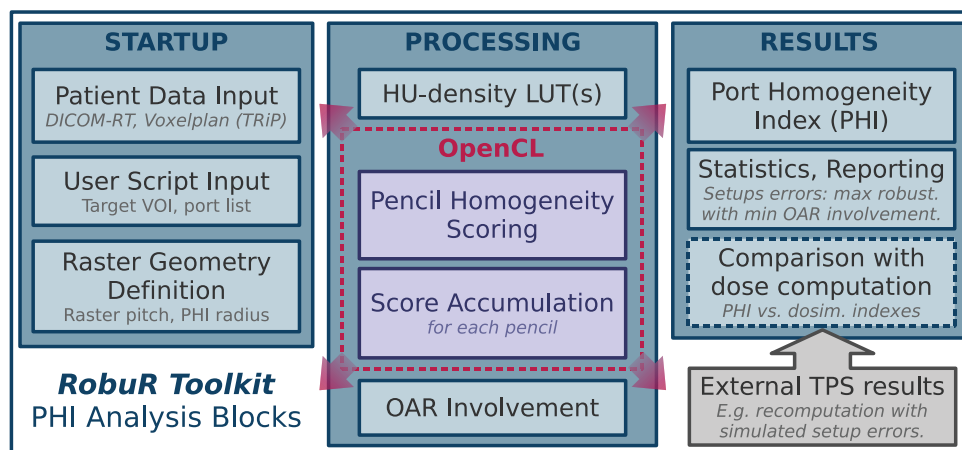
instance the choice of irradiation directions that minimize the potential effects of patient mispositioning, by avoiding severe inhomogeneities across entrance channels [2,3].

In real patient geometries the choice of robust beam setups is typically not straightforward and its verification through extensive dose computation and evaluation, with simulated uncertainties, computationally prohibitive, in the time-constrained clinical routine. An alternative is the use of computationally light surrogates, like the Port Homogeneity Index (PHI) [3], which can score the dosimetric stability of irradiation setups against mispositioning, by analyzing density patterns, prior to treatment plan optimization and computation.

## 2. Materials and methods

In the calculation of the PHI, running time is strongly dominated by the so-called *pencil homogeneity scoring*. During this phase, for each potential beam direction, non-divergent pencil beams from a high resolution raster are traced through the planning CT to the target volume and each is compared to the neighbouring ones using a distance metric that scores their relative density homogeneity [3]. To exploit the data-parallel nature of this and other radiotherapy computations, RobuR [5], a radiation therapy software toolkit developed at the University of Marburg, was recently extended with OpenCL [6] support and used to accelerate the PHI analysis on graphic processing units (GPU) (figure 1).

For 10 skull base cases, the new GPU-accelerated PHI analysis was employed to select particle irradiation setups that are robust against mispositioning. Numerical results and speed were compared against a previously available highly-optimized single-threaded native CPU implementation [5]. The test CPU was a 4-core Intel Xeon W3530 2.8 GHz on a workstation (host) with 12 GB RAM, while the test GPU was an Nvidia Tesla C2050 with 3 GB RAM. The software, written in C++, was compiled with the GCC version 4.8.0 and run under Linux using up-to-date stable OpenCL drivers from the respective vendors.



**Figure 1.** PHI analysis blocks within RobuR. The dashed box in the centre encompasses modules re-implemented with OpenCL and presented in this manuscript. The corner arrows indicate that OpenCL parallelization will be extended to other RobuR modules in the future.

## 3. Results

Numerical results were in agreement between OpenCL implementation and native CPU implementation. As previously validated for the CPU implementation, treatment plans adopting the selected beam setups were dosimetrically robust and maintained high mean target coverage

and dose homogeneity values, when fully recomputed with the TRiP98 treatment planning system [7] in presence of simulated positioning errors (figure 2).

Performance-wise, acceleration on the GPU was on average more than 90-fold. Table 1 reports mean computation times and speedups, with respect to the native CPU implementation, for a single beam port (42 were tested for each patient) in the individual and pooled cases. Each port analysis was performed 10 times to reduce statistical errors. Since OpenCL enables parallelization on different devices (including multicore CPUs) the accelerated PHI analysis was also run without GPU support, with execution times and speedups included in table 1.

Results in table 1 refer to raster resolution and in-depth pencil step both of 0.5 mm (resulting in 4k-40k pencils of 120-350 samples each) and a PHI neighbourhood radius (i.e. max expected setup error) of 2 mm and indicate pure computational time with single-precision floating point operations [6], neglecting overhead for data allocation and data transfer between host and GPU.

For comparison, previous dosimetric verification of each beam angle value by extensive recomputation with TRiP98 (22 selected combinations of 1-2 mm setup errors in 3D) required from several hours to one day, depending on tumour size, on a single CPU core.

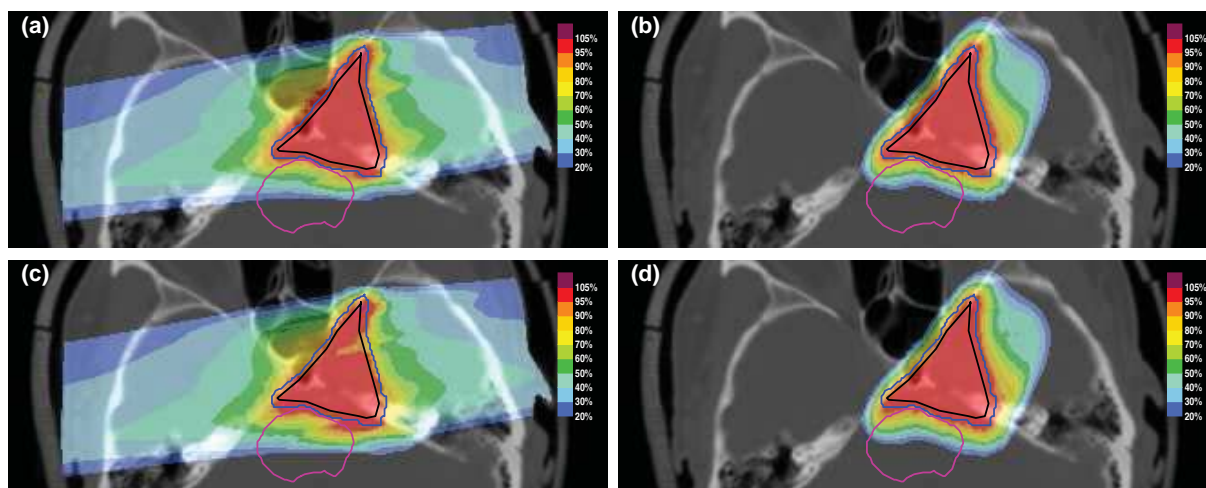
**Table 1.** Results for individual (rows) and pooled (last row) cases: mean(standard deviation) for a single port (in 420 computations). Second column: running times of the native CPU PHI analysis. Other columns: time and speedup (relative to the native CPU implementation) of the OpenCL (OCL) implementation, running respectively on GPU and on CPU (multi-threaded).

Patient number	Native CPU time [s]	OCL GPU time [s]	OCL GPU speedup	OCL CPU time [s]	OCL CPU speedup
P1	21.299 (6.628)	0.219 (0.109)	110.9 (32.6)	6.498 (1.996)	3.3 (0.0)
P2	9.891 (2.837)	0.122 (0.051)	91.4 (31.4)	2.552 (0.615)	3.8 (0.4)
P3	6.258 (1.485)	0.088 (0.038)	81.1 (27.2)	1.898 (0.438)	3.3 (0.0)
P4	15.768 (4.959)	0.148 (0.056)	114.2 (32.6)	4.810 (1.495)	3.3 (0.0)
P5	4.342 (1.136)	0.061 (0.024)	77.3 (22.7)	1.324 (0.331)	3.3 (0.0)
P6	10.888 (3.179)	0.160 (0.065)	75.2 (22.6)	3.280 (0.944)	3.3 (0.0)
P7	10.117 (2.849)	0.151 (0.064)	77.0 (27.8)	3.051 (0.841)	3.3 (0.0)
P8	5.464 (1.470)	0.052 (0.022)	116.0 (39.2)	1.396 (0.310)	3.9 (0.4)
P9	8.155 (2.195)	0.109 (0.046)	83.8 (28.3)	2.481 (0.649)	3.3 (0.0)
P10	11.388 (3.936)	0.126 (0.061)	99.5 (25.8)	2.935 (0.818)	3.8 (0.4)
Pool	10.357 (5.941)	0.124 (0.075)	92.6 (33.2)	3.022 (1.798)	3.5 (0.3)

#### 4. Conclusions

Selection of robust particle therapy beam setups through the Port Homogeneity Index can be effectively accelerated on a GPU and become an unintrusive part of the particle therapy treatment planning workflow, instead of a pre-processing offline step, as previously envisioned [3]. This opens new usage scenarios, like performing brute-force fine-grained robustness analysis of an entire treatment site (e.g.  $\geq 300$  beam angles for the whole head [3]) or re-running the entire analysis many times, interactively manipulating the results [5], e.g. including constraints.

From the implementation point of view, OpenCL enables portability to other kinds of accelerators [6] and brings benefit to (multicore) CPU-only use, too, making the new implementation the only one needed, independently of the hardware.



**Figure 2.** Exemplary dose distributions with a typical two lateral opposed beam setup (left) and a robust beam setup chosen through the PHI analysis (right). Effects on the planned dose (a, b) of a 2 mm patient positioning error in anterior direction are shown (c, d), demonstrating the greater dosimetric robustness of the PHI-selected plan.

The presented results do not include the overhead for data transfer between host and GPU, which can be non-negligible when compared to the pure computational time and affect the overall application performance, as observed for other radiotherapy GPU applications [8]. However, when processing a large number of irradiation setups, the impact of data transfer latency may be minimized by scheduling transfer operations concurrently with computations whose input data have already reached the GPU. This requires accurate transfer/execution synchronization implemented explicitly at application level, which is possible through OpenCL [6] and will be topic of further investigation. The project outlook includes the acceleration on GPU of additional high-level functionalities of RobuR (figure 1) that are complementary to the port robustness analysis, e.g. the assessment of OAR involvement [5].

## Acknowledgments

Work supported by a research grant of the University Medical Center of Giessen and Marburg (no. 35/2010MR). ICCR2013 attendance enabled by Verein zur Förderung der Strahlentherapie und Radioonkologie im Universitätsklinikum Marburg e.V.

## References

- [1] Amaldi U and Kraft G 2005 *Rep. Prog. Phys.* **68** 1861–1882
- [2] Lomax A J 2008 *Phys. Med. Biol.* **53** 1043–1056
- [3] Ammazalorso F, Jelen U, Krämer M, Straßmann G and Engenhart-Cabillic R 2009 *Radiother. Oncol.* **92** S109
- [4] Hopfgartner J, Stock M, Knäusl B and Georg D 2013 *Acta Oncol.* **52** 570–579
- [5] Ammazalorso F, Jelen U, Kraemer M and Zink K 2010 *49<sup>th</sup> Annual Conf. of the Particle Therapy Co-Operative Group* (Gunma, Japan) P6-16
- [6] Bednarz T 2010 Introduction to OpenCL *OzViz 2010 OpenCL Workshop* (Brisbane, Australia)  
<https://sites.google.com/site/ozvizworkshop/ozviz-2010/ozviz-2010-openc1-workshop-program>
- [7] Krämer M, Jäkel O, Haberer T, Kraft G, Schardt D and Weber U 2000 *Phys. Med. Biol.* **45** 3299–3317
- [8] Gu X, Jelen U, Li J, Jia X and Jiang S B 2011 *Phys. Med. Biol.* **56** 3337–3350