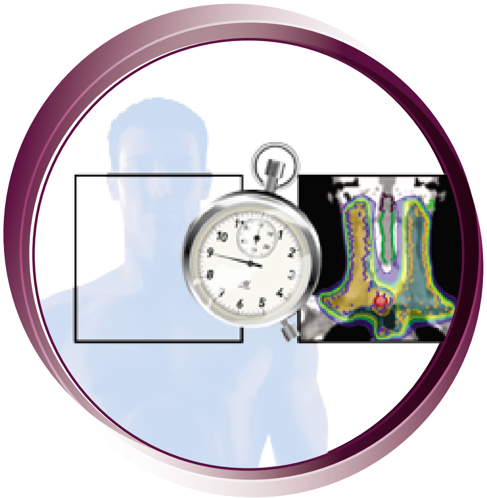
ACCURAY®

VoLO™(VOXEL-LESS OPTIMIZATION) TECHNOLOGY

(Latin: Volo = I fly)



TomoTherapy®

VoLO™ (VOXEL-LESS OPTIMIZATION) TECHNOLOGY

The new Accuray VoLO™ technology represents an exciting step forward for the TomoTherapy® treatment planning process, and promises to make treatment planning faster, more flexible and more interactive, making optimal TomoTherapy treatments available to more patients. It combines the power of graphics processing units (GPUs) traditionally used in the computer gaming world with a new Non-Voxel Broad Beam (NVBB) algorithm, an innovative plan optimization framework.

The NVBB framework is the result of re-engineering the TomoTherapy System's plan optimization algorithm, altering the way in which beam and patient geometries are visualized and also eliminating the need for creation and storage of large amounts of data. This has made it possible to take advantage of the raw processing power of the GPU platform whose strength is non-memory intensive yet highly-parallelized computations.

VoLO Technical Attributes

The key properties of VoLO can be summarized as follows:

- GPU implementation using hundreds of processor cores for highly-parallelized, ray-by-ray dose computations and beam element updates
- Continuous, non-voxel broad beam (NVBB) representation of beam and patient geometry
- Collapsed cone convolution superposition (CCCS) combined with fluence convolution broad beam (FCBB) dose calculations for accuracy and speed
- Use of FCBB during iterations eliminates the need for a "beamlet calculation" preprocessing step
- Direct treatment parameter optimization (DTPO) for determination of machine control parameters
- A high degree of flexibility in altering machine parameters during optimization

GPU Implementation and the NVBB Framework

Each GPU card used by the VoLO system contains 448 processors (cores) that work simultaneously to calculate dose and optimize the parameters that control the treatment delivery system. There are two cards per GPU node mounted in the data server/optimizer assembly. The standard configuration (single node) enables planning on up to four TomoTherapy Planning Stations at one time. The high performance configuration (two nodes) enables planning on up to five Planning Stations at one time. This simultaneous usage includes both dose calculation and optimization.

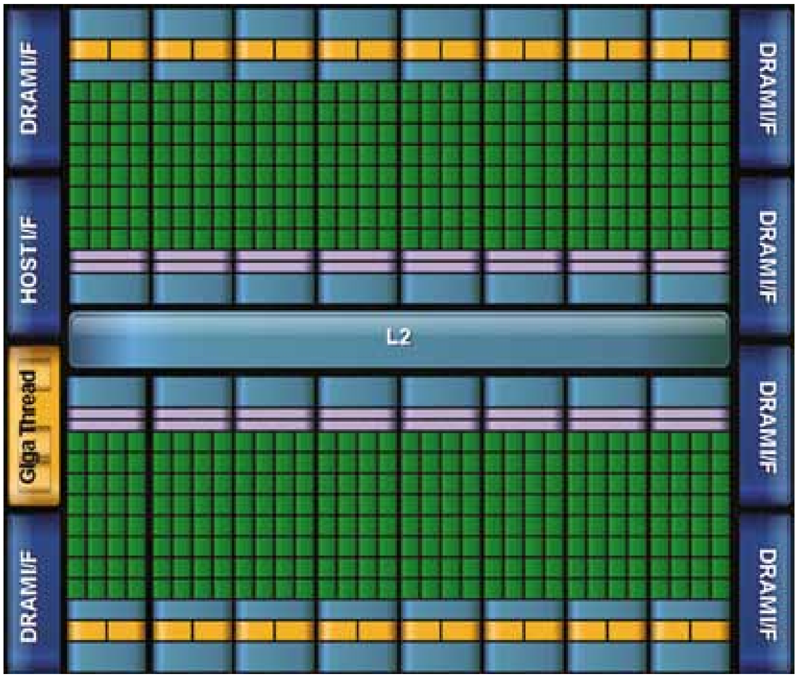


Figure 1. Schematic of a Fermi GPU card. Each small square represents a single processing unit.

Parallelization is suited to computational problems that involve a very large number of discrete elements, whose properties need updating rapidly to contribute to the overall result. While GPU implementation of a radiotherapy optimization problem is ideal in principle, data storage requirements must be minimized due to the limited memory available in the GPU architecture. The NVBB computational geometry and optimization approach used in the VoLO™ technology were specially adapted for a GPU-based implementation.

The NVBB framework is very economical in memory utilization because it does not involve storage of pre-computed "beamlet" dose distributions to be used during iterative optimization. Instead, innovative on-the-fly dose calculations are used for the large number of 3D dose evaluations that are necessary during optimization. These calculations are accurate, fast and have minimal memory overhead.

In the NVBB framework, a "beam's eye view" perspective of the patient geometry is used. Dose calculations and optimization of system parameters are done by considering rays diverging from the trajectory of beam source locations during either a TomoHelical™ or TomoDirect™ delivery. One such diverging beam geometry is shown in Figure 2. Rays intersect a plane perpendicular to rays joining the source and machine isocenter, known as the beam's eye view plane, on which beam fluence entering the patient is defined. The interaction of these rays with the patient geometry and elements of the delivery system such as the jaws and the leaves of the binary multi-leaf collimator are inherent in the algorithm. A direct treatment parameter optimization (DTPO) approach is used to update machine parameters. For further reading see reference [1].

The specific dose calculation algorithms used within the NVBB framework are known as fluence convolution broad beam (FCBB) and collapsed cone convolution superposition (CCCS). These are used for different aspects of the optimization process.

As shown in Figure 3, the NVBB framework involves full dose, iteration dose and final dose calculations. In addition, an adaptive full dose correction step is used as part of each iteration dose step to increase accuracy.

The fluence convolution broad beam algorithm is used during the iteration dose steps. This algorithm is designed to allow rapid 3D dose calculations, while making certain time-saving approximations about radiation transport in heterogeneous material. For further reading see reference [2].

The collapsed cone convolution superposition algorithm is used during the full dose and final dose steps. This 3D dose calculation algorithm is more rigorous than FCBB in modeling radiation transport in heterogeneous material.

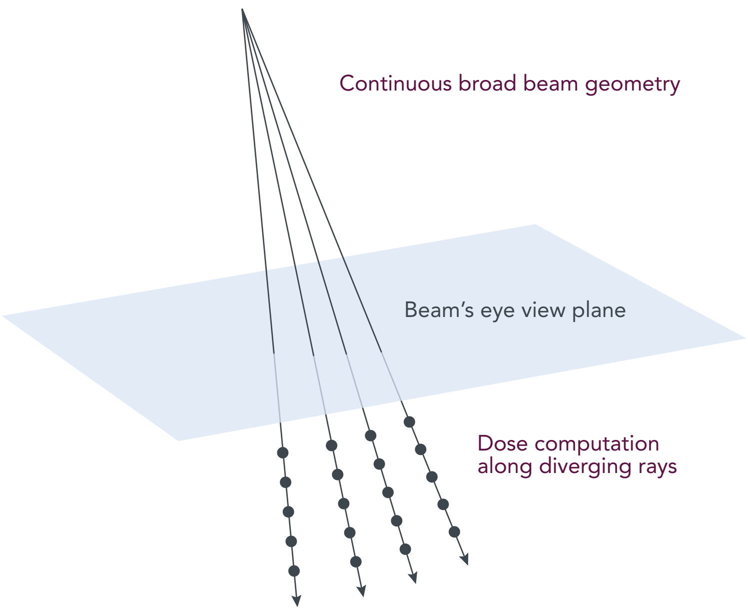


Figure 2. Diverging rays from a source position intersecting a beam's eye view plane. Dose is calculated at points along each ray.

The adaptive full dose correction employs known differences between FCBB and CCCS to make an additive correction to each iteration dose. The difference is determined every tenth iteration and improves the accuracy of other iterations without the need for a full dose calculation. This is a key element of the overall optimization strategy.

When the user is satisfied with the plan result, a final dose is calculated that fully takes into account constraints of the delivery system and sets delivery parameters according to the number of fractions chosen for the treatment course. Machine parameters used for system control are then stored ready for patient quality assurance and subsequent delivery.

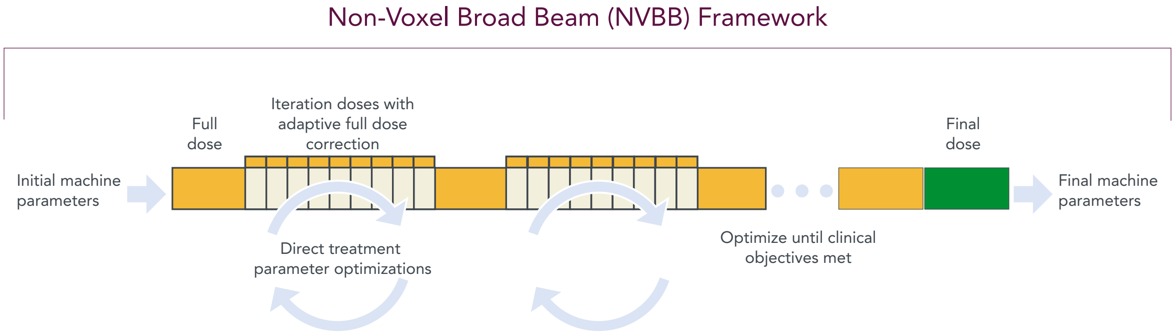


Figure 3. Process used for optimization in the NVBB framework. After initial machine parameters are used as inputs, direct treatment parameter optimization is performed iteratively until a satisfactory dose distribution is attained. Full dose calculated every tenth iteration is used in an adaptive full dose correction during other iterations.

Clinical and Operational Advantages

Interactivity and flexibility are the two key attributes of the VoLO™ planning process. In particular, the VoLO technology provides:

- Fast overall planning time, especially for cases involving large or complex treatment volumes. TomoTherapy® planning for large and complex geometries was always conceptually simple, but now high speed is added to simplicity for all cases.
- Optimization that can begin as soon as contouring is completed and essential machine parameters are defined by the user.
- Extra flexibility to explore alternative parameters such as field width, helical pitch and dose grid resolution interactively. Because machine parameter-specific beamlets are not required with NVBB, options for improving plan quality can be investigated without incurring time overheads.

These advantages bring the potential to plan more patients faster, thereby making optimal TomoTherapy treatments available to additional patients within the clinic.

Efficiency

In the peer-reviewed paper describing the NVBB algorithm [1] a simple planning efficiency test was performed for a range of Helical TomoTherapy cases with different levels of complexity, number of dose calculation points and number of beamlets. Here, number of beamlets refers to the number of “open” MLC (multileaf collimator) leaf instances occurring during the treatment delivery. This number ranges from approximately 4,700 for the prostate case to 117,000 for the total marrow irradiation (TMI) case. In previous versions of the TomoTherapy optimizer, a 3D dose distribution would have been created and stored for each beamlet during the preprocessing step, typically requiring from several minutes to several tens of minutes. Preprocessing is virtually eliminated with the NVBB implementation. For these examples, the number of dose calculation points ranges from approximately 2 million to 8 million, depending on the irradiated volume and dose grid resolution.

Cumulative time components are shown in Figure 4 for each case. Durations are shown for (i) preprocessing, (ii) 100 optimization iterations, (iii) full dose, and (iv) final dose. Note that iteration times include full dose calculations as part of the adaptive full dose correction process. These plots are simple indications of planning time and would be affected by human interaction in a clinical setting. As an indication of how these times compare with the previous CPU-based optimization strategy including beamlet calculations, the breast case and TMI cases had planning times of approximately 30 minutes and 150 minutes respectively. This represents a time reduction of 10-20 times, mostly due to the elimination of the beamlet calculation step. It is clear that preprocessing, full dose and final dose are not significant time components with this new strategy. Iteration times are fast also due to the extreme parallelization of the computation process via GPU implementation. Note that the relative variability in overall planning time is surprisingly small considering the large variation in number of beamlets and dose calculation points. The variation in absolute time in minutes is also small. Cases traditionally considered “complex” can be planned quickly using this new framework.

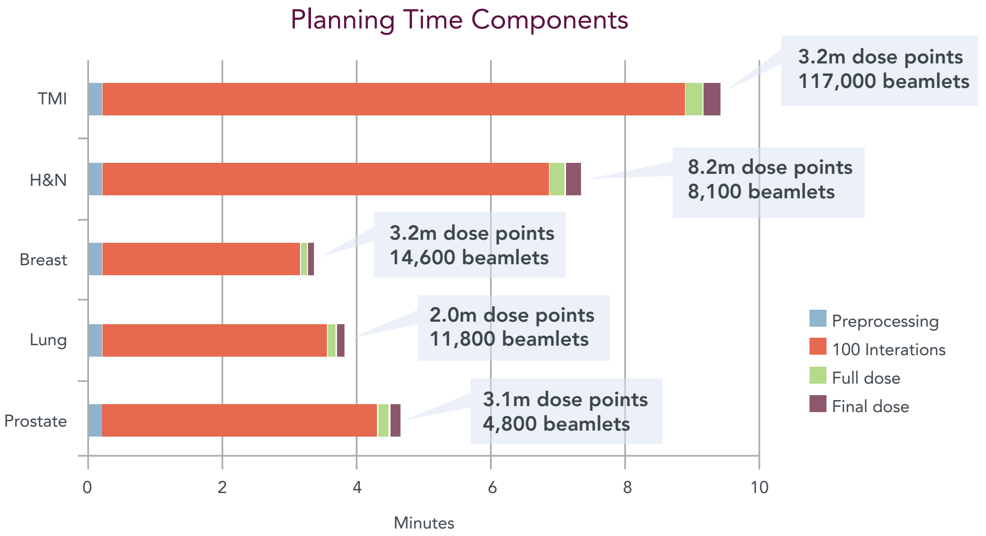


Figure 4. Planning time components for a range of cases including preprocessing, 100 optimization iterations, full dose calculation and final dose calculation Note that the 100 iterations step includes a full dose calculation every ten iterations as part of the adaptive full dose correction process. Number of beamlets (MLC leaf opening events) and dose calculation points are indicated for each case. Data from reference [1].

Extra insight into the performance improvements brought about by the VoLO™ technology can be gained via a comparison of optimization times with those for previous generations of the TomoTherapy® optimizer. Shown in Figure 5 are the times for a 200-iteration optimization session for a head and neck cancer case for VoLO along with four older generation systems. Gen 5 is the "Generation 5" CPU-based hardware currently shipped as the standard configuration on all Hi·Art® Systems. The VoLO (labeled GPU) technology can be seen to be approximately three times as fast as Gen 5 for this case and more than ten times faster than optimization using original TomoTherapy hardware. Except for the GPU case, times include the pre-optimization beamlet calculation step that is unnecessary with the VoLO technology.

Some of the time reduction seen with the VoLO technology is due to the absence of the beamlet calculation step, some is due to the parallelization inherent in GPU operation, and some is due to other increases in algorithm efficiency.

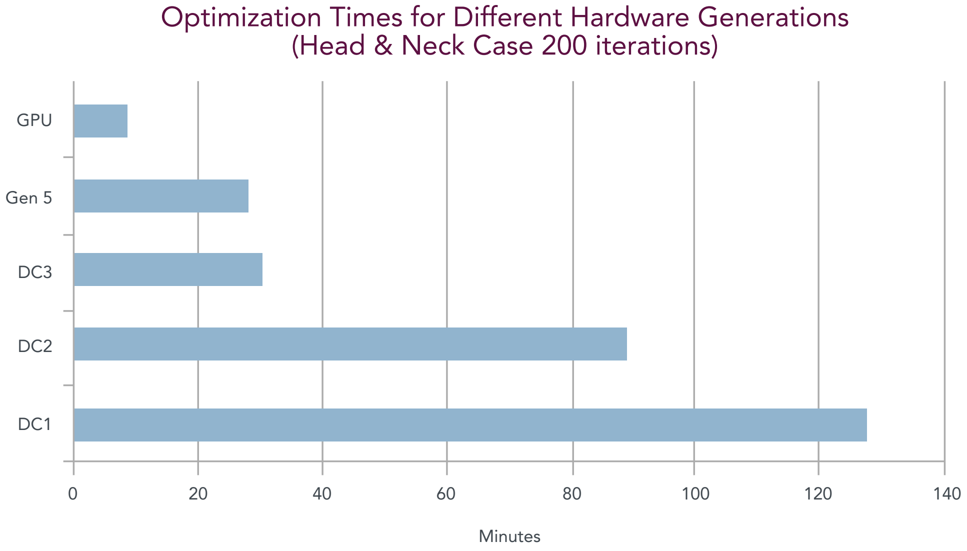


Figure 5. Optimization times for VoLO (GPU) versus previous generations of the TomoTherapy optimizer for an example head and neck cancer case. Gen 5 is the "Generation 5" CPU-based hardware currently shipped as the standard configuration on all Hi·Art systems. Each time plotted includes 200 optimization iterations. Except for GPU, times include a pre-optimization beamlet calculation step.

Further comparing VoLO™ planning speed with that of the generation 5 hardware, Figure 6 shows optimization times for breast, head and neck and craniospinal cases. Again, 200 iterations were performed for each case. The head and neck case is the same one used to generate the data in Figure 5. The breast example used the TomoDirect™ delivery technique so the total number of "beamlets" employed is much less than if this had been a TomoHelical™ delivery. As with the head and neck case, optimization times for the breast and craniospinal plans are reduced by approximately a factor of three between Gen 5 and GPU.

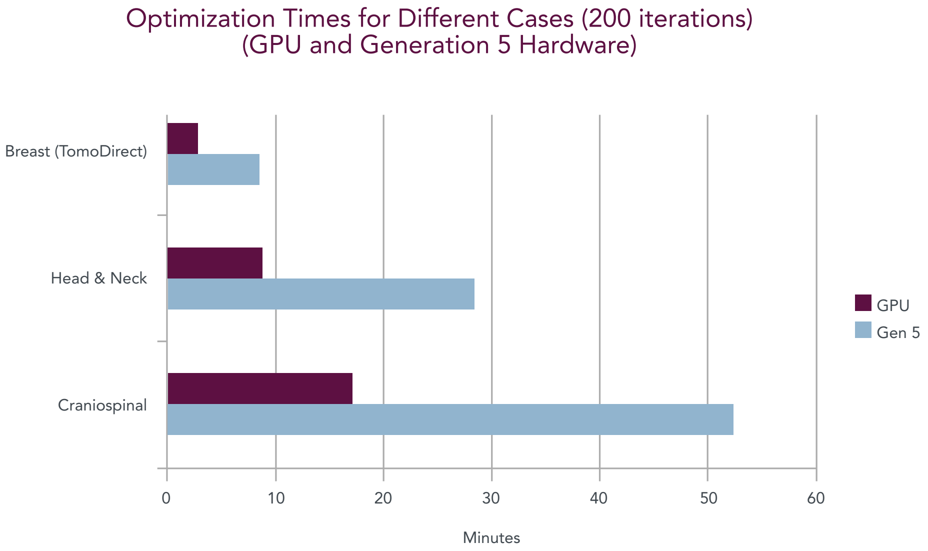


Figure 6. Optimization times for breast, head and neck and craniospinal cases for the VoLO (GPU) and Generation 5 hardware. 200 iterations were performed for each case. The head and neck case is the same one used to generate the data in Figure 5. For Gen 5, these times include the pre-optimization beamlet calculation step that is unnecessary with the VoLO technology.

Although a large contribution to the increase in speed between Gen 5 and VoLO is the absence of an initial beamlet calculation, optimization iterations and other dose calculations are also faster. Figure 7 shows a comparison of average individual iteration time for the same three cases. In the case of VoLO, iteration time is an average over ten iterations where one is a "full dose" calculation. As described above, one in every ten iterations is a full Collapsed Cone Convolution Superposition (CCCS) calculation and the others use the more approximate Fluence Convolution Broad Beam (FCBB) algorithm.

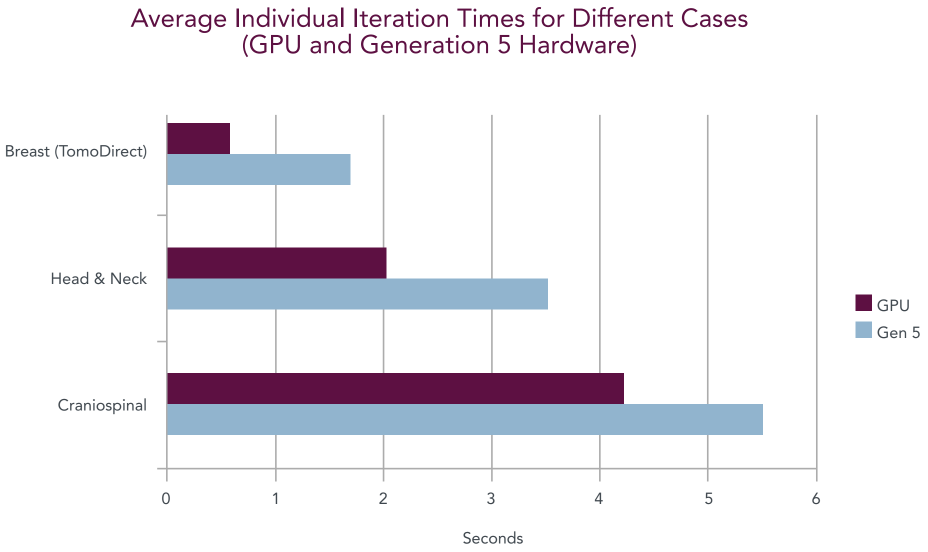


Figure 7. Average individual iteration times for the VoLO (GPU) and Gen 5 configurations for the same cases as in Figure 6. This illustrates that reduction in iteration time contributes to the overall increase in optimization speed.

Finally, Figure 8 shows the reduction in "final dose" computation times between Gen 5 and GPU configurations. For the head and neck plan as well as the craniospinal plan this graph shows a more dramatic difference than for the iteration times. Final dose is computed once at the end of the plan optimization process. This calculation step takes account of practical machine characteristics and creates final instructions for machine control per fraction of the treatment course. Time includes saving of the dose and delivery instructions to the central data server. The dramatic reduction in time is mostly due to the extreme parallelization of the algorithm enabled by the use of a GPU.

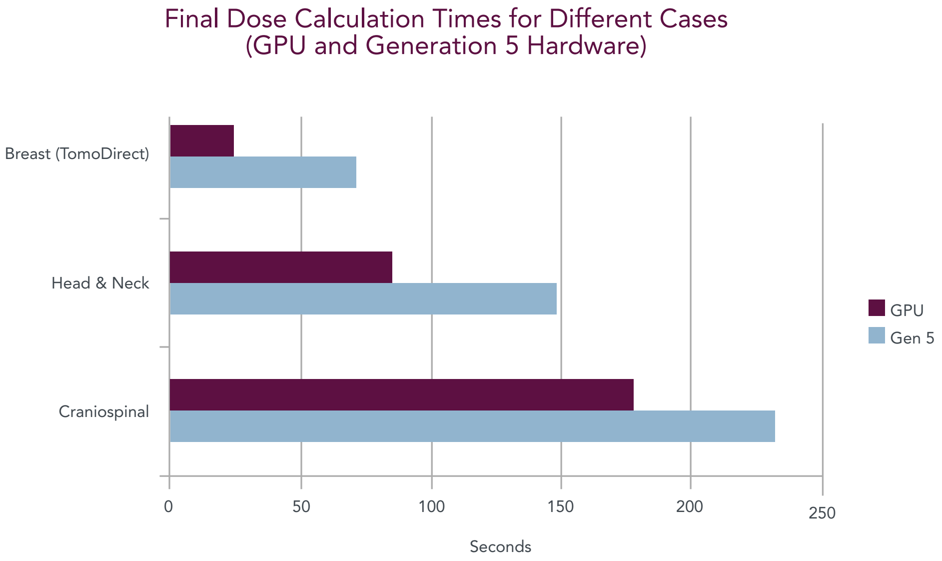


Figure 8. Average "final dose" calculation times for the VoLO™ (GPU) and Gen 5 configurations for the cases in Figure 6 and Figure 7. This calculation step takes account of practical machine characteristics and creates final instructions for machine control per fraction of the treatment course. Time includes saving of the dose and delivery instructions to the central data server.

To close the loop with this new treatment planning platform, rigorous tests of its accuracy have been carried out, including consistency with the traditional CPU-based platform used by current clinical systems. Agreement with measured doses, highly-accurate Monte Carlo dose calculations, as well as the CPU platform has been shown to be excellent [3].

References

1. Lu W. A non-voxel-based broad-beam (NVBB) framework for IMRT treatment planning. Phys Med Biol. 2010 Dec 7;55(23):7175-210.
2. Lu W, Chen M. Fluence-convolution broad-beam (FCBB) dose calculation. Phys Med Biol. 2010 Dec 7;55(23):7211-29.
3. Chen Q, Lu W, Chen Y, Chen M, Henderson D, Sterpin E. Validation of GPU based TomoTherapy dose calculation engine. Med Phys. 2012 Apr;39(4):1877-86.

TomoTherapy®

ACCURAY®

UNITED STATES

Accuray Corporate Headquarters
1310 Chesapeake Terrace
Sunnyvale, CA 94089
USA
Tel: +1.408.716.4600
Toll Free: 1.888.522.3740
Fax: +1.408.716.4601
Email: sales@accuray.com

TomoTherapy
1240 Deming Way
Madison, WI 53717
USA
Tel: +1 608 824 2800
Fax: +1 608 824 2996

ASIA

Accuray Japan K.K. - Tokyo
Shin Otemachi Building 7F
2-2-1 Otemachi, Chiyoda-ku
Tokyo 100-0004
Japan
Tel: +81.3.6265.1526
Fax: +81.3.3272.6166

Accuray Asia Ltd. - Hong Kong
Suites 1702 - 1704, Tower 6
The Gateway, Harbour City
9 Canton Road, T.S.T.
Hong Kong
Tel: +852.2247.8688
Fax: +852.2175.5799

Accuray Chengdu
#39 Huatai Road, Longtan
Industrial Zone,
Section 2 East, 3rd Ring Road,
Chengdu, Sichuan 610051
China

EUROPE

Accuray Europe
Tour Atlantique 25e
1 Place de la Pyramide
92911 Paris La Défense Cedex
France
Tel: +33.1.5523.2021
Fax: +33.1.5523.2039

Tomotherapy Belgium BVBA
Pegasuslaan 5
1831 Diegem
Belgium
Tel: +32 (0)2 400 4400
Fax: +32 (0)2 400 4401/02