

Planning and Optimizing Treatment Plans for Actively Scanned Proton Therapy

**Francesca Albertini, Sylvain Gaignat,
Matthias Bosshardt, Antony J. Lomax***

Centre for Proton Radiation Therapy
Paul Scherrer Institute
Villigen PSI, Switzerland
tony.lomax@psi.ch

1 Proton Therapy: An Introduction

There are many sophisticated and advanced methods currently being proposed and introduced into radiotherapy. One of the most promising is proton therapy. When quasi-monoenergetic proton beams are applied to a medium, they display the rather attractive property of having a well-defined range, beyond which no energy is deposited, and a concentrated and focused high dose volume (dose = deposited energy) in a narrow region known as the “Bragg peak.” This property was recognized by Robert Wilson in 1946^[1] as being almost ideal for radiotherapy, in that by exploiting the three-dimensional localization of dose in the Bragg peak, increased dose can potentially be applied to the tumor whilst simultaneously sparing surrounding normal tissues. It took approximately 8 years however before the potential of proton therapy could be demonstrated on patients, with the first being treated at the Lawrence Berkeley National Laboratory in 1954.^[2]

One of the challenges facing the early pioneers of proton therapy was how to modulate the typical proton beams that are produced by particle accelerators so as to produce a clinically useful field size large enough that arbitrarily sized tumors could be irradiated. For instance, although the Bragg peak has many desirable characteristics, it is typically only a few millimeters wide in depth. In addition, proton beams produced by modern accelerators have a rather narrow cross section, again typically of only a few millimeters. Thus, in order to irradiate anything but the smallest tumors, such proton beams must be broadened (or “spread”) both along, and orthogonal to, the beam direction. For the first proton treatments this was done using what has subsequently been called the passive scattering technique. In depth, the field size (i.e., the volume over which the dose can be considered homogenous) can be extended by the simple technique of delivering a

*Corresponding author.

number of range adjusted and appropriately weighted individual Bragg peaks. When the weights (or more correctly stated, the relative fluences) of the individual Bragg peaks are correctly calculated, then a uniform dose profile along the beam direction can be produced, known as the “Spread-Out-Bragg-Peak” (SOBP).^[3]

Orthogonal to the beam direction, the narrow pencil beam emitted from the accelerator nozzle is broadened through the insertion of “scattering-foils” into the beam.^[4] As protons are scattered predominantly by interactions with orbiting electrons as they pass through any medium, and the amount of scatter is dependent on the atomic number of the material, a narrow proton beam can be effectively spread by the insertion of high- Z foils directly in the beam. By working in the highest portion of this Gaussian profile, a lateral uniformity of $\pm 2\text{--}5\%$, depending on the required field size, can be achieved. Finally, the resulting delivered fields are shaped to better conform the delivered dose to the tumor through the use of patient- and field-specific collimators (to conform the cross-sectional shape of the delivered field to the projected cross-sectional shape of the target volume) and compensators (to match the distal end of the SOBP to the distal end of the tumor).

Although a relatively simple approach, passive scattering has been a remarkably successful delivery technique for proton therapy, with more than 55,000 patients having been treated worldwide with this technology. Indeed, even today, most new proton facilities being planned or built will still be based on the passive scattering approach. Nevertheless, passive scattering must slowly be considered to be somewhat limited in its capabilities, and if proton therapy wants to preserve its edge in comparison to the rapidly developing delivery techniques being introduced for radiotherapy using photons, then new delivery methods for proton therapy must also be developed.

2 Proton Therapy Using Active Scanning

One such approach is active scanning. This was first suggested by Kanai et al.,^[5] and has subsequently been implemented in our clinical proton therapy facility at the Paul Scherrer Institute (PSI) in Switzerland.^[6] The principle of active scanning is deceptively simple, and is based on the fact that protons, being charged particles, are subject to Lorentz forces. That is, when subjected to an electric field, they are accelerated and when subjected to a magnetic field, they are deflected. It is this second characteristic that is exploited in active scanning to spread the beam laterally. Thus, in this approach, the narrow proton “pencil beam” is no longer broadened through scattering, but is “scanned” across the tumor using a scanning magnet. In depth, the Bragg peaks are stacked by altering the initial proton energy at the accelerator source. Through this combination of scanning and energy variation, the Bragg peak can be effectively placed anywhere in three dimensions within the tumor. Dose uniformity is then achieved through a mathematical optimization of the individual fluences of each pencil beam (see below).

Active scanning has a number of advantages over the more traditional passive scattering approach. First, it can be fully automated and computer controlled, meaning that, as only Bragg peaks that stop in the tumor volume are delivered, it is not necessary to use collimators and com-

pensators to achieve dose conformation. Second, it is more efficient than passive scattering in that less protons need to be delivered in order to achieve a given total dose to the tumor and therefore less dose is delivered to the normal tissues. Last, active scanning is an inherently more flexible technique than passive scattering.

Although active scanning has many advantages, it is a rather new delivery technique, and at the time of writing there are only two facilities worldwide that are using active scanned protons in clinical practice, with an additional center using a similar approach with carbon-ions. Of the proton centers, PSI has the widest experience, with more than 500 patients having been treated, all by using active scanning. Initial clinical results are promising,^[7,8,9] and based on this success, most new proton therapy centers being currently built or being planned will have active scanning as an option, if not the only delivery approach.

3 SFUD and IMPT

As implemented at PSI, active scanning can be applied in two “flavors.” The first essentially “mimics” passive scattering delivery, such that each individually applied field applies a more or less homogenous dose across the selected target volume, much as the scatterer and SOBPs of passive scattering apply a uniform dose across the target.^[10] However, the active scanning approach adds some flexibility, and can be considered more efficient in that only Bragg peaks delivered within (or close to) the target are delivered. One can think of this as the delivery of a number of SOBPs-shaped beamlets, but where the extent of each SOBP is matched to the thickness of the target along the beam direction. This is, however, somewhat of an oversimplification. In practice, it is necessary to individually optimize the weight (fluence) of each delivered beamlet to truly deliver a homogenous dose to arbitrarily complex shaped target volumes. A full treatment plan then consists of the simple linear addition of such individually optimized fields. For this reason, we call this method Single Field, Uniform Dose (SFUD) delivery,^[10] which is the direct active scanning equivalent of open-field photon plans or passive scattered proton treatments. Indeed, the SFUD approach was the method of choice for the first patients treated at PSI, and is still used in about 60% of the currently applied plans at PSI.

Ultimately however, scanning also provides the possibility to deliver so called “intensity modulated proton therapy” (IMPT).^[11,12] In this approach, the optimization process is extended such that all Bragg peaks from all selected field directions are optimized simultaneously. The major difference in this approach is that the “uniform, single field” constraint is removed, and the optimization is given full reign to weight Bragg peaks regardless of the final form of the individual field dose distributions, as long as the total dose (the addition of all the individual field dose distributions) combine to give the desired result. Although the difference to SFUD planning is rather subtle, the IMPT approach best exploits the full potential of scanned proton therapy, in that it provides even more flexibility in how the target is irradiated in comparison to SFUD. Consider for instance Figure 1. A three-field SFUD plan applied to the defined target volume would (schematically) look like the dose distribution in Figure 1a. It would achieve a good dose homogeneity over the target

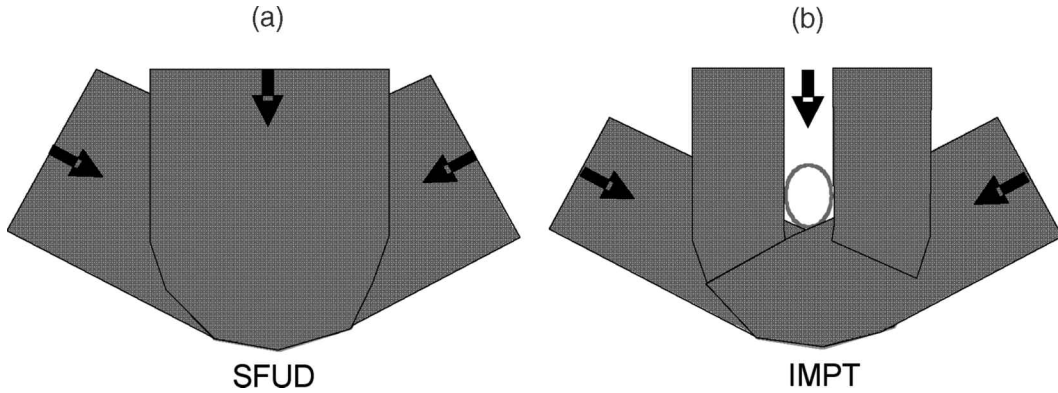


Figure 1. (a) Schematic representation of a SFUD (Single Field, Uniform Dose) plan. Each of the three fields (indicated by the arrows) delivers a homogenous dose to the whole target volume. (b) Schematic representation of an IMPT plan. The posterior field (from the top) has been modulated such as to avoid irradiating the red critical structure. Consequently, the two lateral oblique fields must also be modulated in order to still obtain a homogenous dose across the target volume. The advantage of IMPT lies in its ability to individually modulate fields in order to make up for under- or overdosing from the other fields due to their sparing of neighboring critical structures. SEE COLOR PLATE 1.

volume, but inevitably, the organ at risk (OAR) close to the target would receive some dose due to the plateau dose being delivered from the posterior beam and the anterior edges of the anterior-lateral oblique beams. By relaxing the constraint that each field must uniformly irradiate the whole target volume however, we could irradiate the target in the way indicated in Figure 1b. Here, the beamlets passing through the OAR from the anterior beam would be (ideally) completely switched off, thus sparing this organ completely. If also the anterior edges of the anterior-lateral beams are also switched-off (again to avoid irradiating the OAR), then a homogenous dose could still be delivered to the target volume by modifying the individual fields as indicated in the figure. For scanning, such a field and dose arrangement would be achieved through IMPT, or the simultaneous optimization of all beamlets from the three field directions, in that the missing dose from one field (i.e., the posterior field where the beamlets have been reduced in weight where they pass through the OAR) are compensated for by contributions from the other fields. For the optimization to achieve this, then it is obvious that it must know the positions and spatial whereabouts of all beamlets of all fields.

As an aside, such a plan as indicated in Figure 1b can also be applied using passive scattering, by “patching” fields together. Indeed, this technique has been done for many years in proton therapy centers worldwide, and can be considered to be a “manual” IMPT approach. The advantage of scanning (and IMPT), however, lies in the ease with which such plans can be calculated (the patching is performed by the optimization algorithm), particularly when the problem is more complex than that shown in Figure 1 (i.e., with multiple neighboring critical structures).

In summary, regardless of whether one uses SFUD or IMPT planning, an optimization algorithm for the calculation of a set of fluences that provides the desired dose distribution is a prerequisite for active scanned proton therapy. The optimization approach used for such calculations at our institute will be outlined in the following section.

4 The Optimization Problem in Proton Therapy

As described above, active scanning consists of the delivery of a three dimensional pattern of individually weighted proton pencil beams, where in this context a pencil beam's weight can be considered to be its fluence (i.e., the number of protons delivered). However, given that pencil beams are delivered with a spacing of 5 mm in all directions (both laterally and in between successive Bragg peaks in depth), then many thousands of individual pencil beams are required to fully cover the tumor volume. Thus, the only tractable solution to finding the "optimal" set of relative fluences for all these pencil beams is to apply an optimization algorithm. For the system at PSI, we have adopted the following optimization approach.

Our planning system uses a quasi-Newton technique, based on that developed by Pedroni^[13] for pion therapy and subsequently modified by Lomax^{[11][14]} for spot scanning and IMPT. The basic approach is also similar to that described by Bortfeld et al.^[15] and Wu and Mohan^[16] for the optimization of intensity-modulated radiotherapy with photons. The cost function $F(\dot{u})$, to be minimized, is defined as

$$F(\dot{u}) = \chi^2 = \sum_i^N g_i^2 (P_i - D_i)^2 \quad (1)$$

where $\dot{u} = (\omega_1, \omega_2, \dots, \omega_M)$ are pencil beam fluences, D_i is the calculated dose to the grid point i , P_i^1 is the prescribed dose at that point, and g_i^2 is the associated importance factor, based on the planner's prescriptions.

When the pencil beam weights $\dot{u} = (\omega_1, \omega_2, \dots, \omega_M)$ are updated during the iterative process, then the dose $\mathbf{D} = (D_1, D_1, \dots, D_N)$ can be calculated through

$$D_i = \sum_{j=1}^M d_{ij} \omega_j, \quad (2)$$

¹The prescription dose is assumed to be homogeneous within the volume apart from close to the surface, where it has a Gaussian fall-off calculated in such a way that the dose to the target volume's surface is 90% of the maximum prescribed dose to the target (this is done to provide a more realistic template that approximately matches with the lateral fall-off of the individual proton pencil beam).

²For the weighting matrix (\mathbf{g}) all dose grid points within the target volume are assigned to a fixed weight of 1.

where d_{ij} is the unweighted dose contribution of pencil beam j to the grid point i , M is the number of pencil beams, and N is the total number of grid points.

The inverse problem to be solved here is to find the set of pencil beam fluences (\hat{u}), which minimizes the cost function $F(\hat{u})$. As stated by Bortfeld,^[15] there exists a huge variety of iterative algorithms for the solution of such problem which are similar to the Newton iteration:

$$\omega(k+1) = \omega(k) - \alpha_k (\nabla^2 F(\omega(k)))^{-1} \nabla F(\omega(k)). \quad (3)$$

The main problem to be solved is how to invert the Hessian matrix $\nabla^2 F(\omega(k))$, which is difficult due to the huge size of \mathbf{d} ; therefore the matrix has to be approximated to make the problem tractable. A common approximation of this matrix is by a diagonal matrix \mathbf{S} whose elements are those of the Hessian matrix: $S_{j,j} = \sum_l^N D_{jl}^2$ (see e.g., Bortfeld et al.^[15]).³

In one dimension, for a single pencil beam j , the iterative update of the beam weight ω_j from iteration (k) to iteration ($k+1$) then becomes

$$\omega_j(k+1) = \omega_j(k) - \alpha_k \begin{pmatrix} \frac{dF}{d\omega_j} \\ \frac{d^2F}{d\omega_j^2} \end{pmatrix} = \omega_j(k) + \alpha_k \begin{pmatrix} \sum_l^N g_i^2 (P_i - D_i) d_{ij} \\ \sum_l^N g_i^2 d_{ij}^2 \end{pmatrix}. \quad (4)$$

where α_k is a damping factor to ensure convergence⁴ for the iterative process.

All $\omega_j(k)$ should fulfill physical constraints which require that only positive fluencies can be delivered. Often this request is satisfied by using a positive operator, which turns to zero all negative constraints, $\hat{u} = \begin{cases} \hat{u} & \text{if } \omega \geq 0 \\ 0 & \text{if } \omega < 0 \end{cases}$ (see e.g., Bortfeld et al.^[15]).

³The use of the diagonal matrix \mathbf{S} to approximate the Hessian matrix implies that there are only a few pencil beams delivering dose to the same voxel.^[17,18] Off-diagonal elements of the Hessian matrix are non-zero if two beamlets deliver dose to the same voxel. By setting these elements to zero every beamlet is treated independently during the optimization process. As this assumption is patently incorrect for particle therapy (there is large overlap between beamlets laterally, and particularly along the beam directions where many plateau doses overlap), convergence can only be achieved by introducing a damping function.

⁴Often α_k is chosen to satisfy the **Wolfe conditions**. The Wolfe conditions are a set of inequalities that guarantee convergence in (unconstrained) optimization especially in quasi-Newton methods for performing inexact line-searches. Inexact line-searches provide an efficient way of computing an acceptable step length that reduces the cost ‘‘sufficiently,’’ rather than minimizing the cost function exactly.

However, as mentioned by Pflugfelder^[17,18], the use of the positive operator can slow down the algorithm since it can repeatedly provide the same solution and can result in solutions which are worse than the previous iteration.

Our planning system uses a damping factor introduced by Lomax et al.^[11,14] with the following form:

$$\alpha_{ij}(k) = \frac{\omega_j(k) d_{ij}}{D_i(k)}, \quad (5)$$

which represents the fraction of the total calculated dose at the k^{th} -iteration contributed by spot j to grid point i . Equation (4) then becomes

$$\begin{aligned} \omega_j(k+1) &= \omega_j(k) - \alpha_k \left(\frac{\sum_l^N g_l^2 (P_i - D_i) d_{lj}}{\sum_l^N g_l^2 d_{lj}^2} \right) = \omega_j(k) + \omega_j(k) \left(\frac{\sum_l^N g_l^2 \frac{d_{lj}}{D_i} (P_i - D_i) d_{lj}}{\sum_l^N g_l^2 d_{lj}^2} \right) \\ &= \omega_j(k) \left[1 + \frac{\sum_l^N g_l^2 \frac{d_{lj}^2}{D_i} P_i - \sum_l^N g_l^2 \frac{d_{lj}^2}{D_i} D_i}{\sum_l^N g_l^2 d_{lj}^2} \right] = \omega_j(k) \left[1 + \frac{\sum_l^N g_l^2 \frac{d_{lj}^2}{D_i} P_i}{\sum_l^N g_l^2 d_{lj}^2} - 1 \right] \\ &= \omega_j(k) \frac{\sum_l^N g_l^2 d_{lj}^2 \frac{P_i}{D_i}}{\sum_l^N g_l^2 d_{lj}^2}. \end{aligned} \quad (6)$$

With the use of the damping factor described in equation (5), our optimization algorithm fulfills the physical request of positive fluences without adding additional constraints to the problem.⁵

⁵ The damping factor not only guarantees convergence to the iterative process but also elegantly transforms the constrained problem to an unconstrained one. The absence of additional constraints is probably the reason for the rapid convergence observed with our optimization algorithm. Matsinos et al.^[19] have in fact shown for a single-case study that PSI optimization algorithm produces results comparable to conjugate gradient (CG) and simulated annealing (SA) optimization methods, but converging to a lower value of the objective function and, most importantly, in a faster way ($t_{\text{CG}} \sim 1.5 t_{\text{PSI}}$ and $t_{\text{SA}} \sim 2.2 t_{\text{PSI}}$).

Both the target-dependent prescription dose P_i and the weighting function g_i can be modified by the presence of critical structures. All critical organ volumes of interest (VOIs) can be assigned individual prescription doses and weights, depending on their relative importance in the planning process. The weighting function for critical structures has been defined in the same manner as described by Bortfeld et al.^[15] So to express it explicitly, the objective function $F(\dot{u})$ maybe be rewritten as

$$F(\dot{u}) = \sum_I^{N_T} g_i^2 (P_i^T - D_i)^2 + \sum_I^{N_{OAR}} g_i^2 (P_i^{OAR} - D_i)^2 H(P_i^{OAR} - D_i) \quad (7)$$

where N_T and N_{OAR} are respectively the number of voxels in a target volume and in an organ at risk. $H(P_i^{OAR} - D_i)$ is the step function defined as

$$H(P_i^{OAR} - D_i) = \begin{cases} 1 & D_i > P_i^{OAR} \\ 0 & D_i \leq P_i^{OAR} \end{cases} \quad (8)$$

In other words, only the voxels in the OAR with dose greater than the tolerance dose will contribute to its objective function components. This concept has proven to result in a fast and stable optimization.^[15,20]

5 Degeneracy in SFUD and IMPT Optimization

In a typical SFUD or IMPT plan, there are many thousands of applied beamlets, each of which can be individually weighted by the optimization algorithm. As the number of beamlets that can be varied determines the degrees of freedom (DOF) available to the optimization process, in comparison to other optimization problems in radiotherapy, the DOFs available when optimizing SFUD, and in particular IMPT, plans are enormous. The main reason for this is the fact that in scanned proton therapy, beamlets are not just distributed in a plane orthogonal to the field direction, but also *in depth* along the beam direction. It is this shifting (and weighting) of Bragg peaks which allows for construction of the SOBP in passive scattering, and this same technique can be exploited in SFUD and IMPT optimization to modify the depth dose curve individually along the beam direction, such as to deliver *different* effective depth dose curves through different parts of the target. However, as stated above, even in the optimization of IMRT with photons (where the modulation matrix per field is

“only” two dimensional), the optimization problem is rather degenerate. That is, there are a large number of different beamlet fluence combinations for each field which result in very similar calculated dose distributions. Due to the addition of a third dimension to the problem in proton therapy, it is perhaps not surprising that degeneracy of the solution for SFUD and IMPT is even larger.

To illustrate the degeneracy of the optimization problem in the optimization of proton plans, two examples will be presented.

6 An Example of Degeneracy 1: Spot Reduction

The first example is an investigation into the possibilities of reducing the number of delivered Bragg peaks as part of the optimization process (so called “spot reduction”). A previous analysis of a series of SFUD plans clinically delivered at the PSI facility showed that the optimization procedure resulted in a few relatively highly weighted Bragg peaks per field enhanced by a large number of beamlets with very low weights.^[21] For instance, this analysis found that more than 90% of all delivered beamlets had weights of 10% or less of the maximum beamlet weight in each field. Thus, we wanted to investigate to what extent these low-weighted Bragg peaks could be reduced (“switched-off”) during the optimization procedure, whilst still preserving the same dose homogeneity and dose coverage to the target volume. The optimization scheme was thus modified, such that after each m iterations, the $n\%$ lowest weighted beamlets were set to zero, and then the optimization loop continued. As can be seen from equation (6) above, an important characteristic of the update function of our optimization algorithm is that, as it is a product of the weight from the last iteration rather than an addition, if the weight of a beamlet is set to zero, then it must stay zero. Thus, by setting the $m\%$ lowest weighted beamlets to zero each n iterations, the number of non-zero-weighted Bragg peaks is gradually reduced.

In Figure 2, a plot of the cost function $[F(\hat{u})]$, see equation (1) above] is shown for this algorithm. The iterations where the lowest weighted spots are discarded are clearly seen as large spikes in the cost function. This is easy to understand—as every four iterations suddenly 20% of spots are removed, and this has a substantial effect on the resultant dose distribution (\mathbf{D}). However, this is almost immediately recovered in the next iteration of the optimization, as $F(\hat{u})$ quickly returns to a value very similar to that immediately before the discard step. Indeed, if the “spikes” of the spot discards are ignored, the cost function monotonically reduces, and seems to be little affected by the fact that 20% of spots are being switched-off every four iterations. Figure 3 shows two resulting single-field dose distributions calculated with the conventional algorithm (left hand panels) and with the algorithm modified to perform the spot reduction (right hand panels). Visually, the two dose distributions are almost identical. However, for the conventional algorithm, the plan consists of 1690 non-zero beamlets per field, whilst the plan resulting from the spot reduction

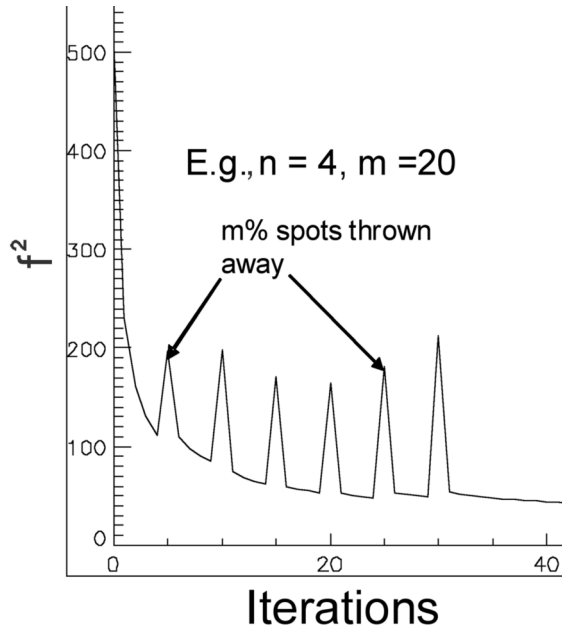


Figure 2. A plot of $F(\hat{u})$ (the cost function) against iteration number for the spot reduction algorithm. For this example, the 20% lowest weighted spots are “discarded” (set to zero weight) every four iterations. Each spot discard is seen as a spike in the cost function, but this quickly returns to a value similar to that before the spot discard. By 40 iterations, $F(\hat{u})$ is not very different to that one would expect when optimizing normally (i.e., without discarding spots during the process).

algorithm reduced this to only 770 non-zero beamlets. This is a very clear indication of the degeneracy of the optimization. The two dose distributions in Figure 3 are, to all intents and purposes, identical, but the distribution and weightings of the individual beamlets used to calculate them are very different. In this study, 15 different cases were studied from the point of view of spot reduction, and, whilst maintaining the same target homogeneity and coverage, spot reductions of between 75% and 85% could be achieved.^[22]

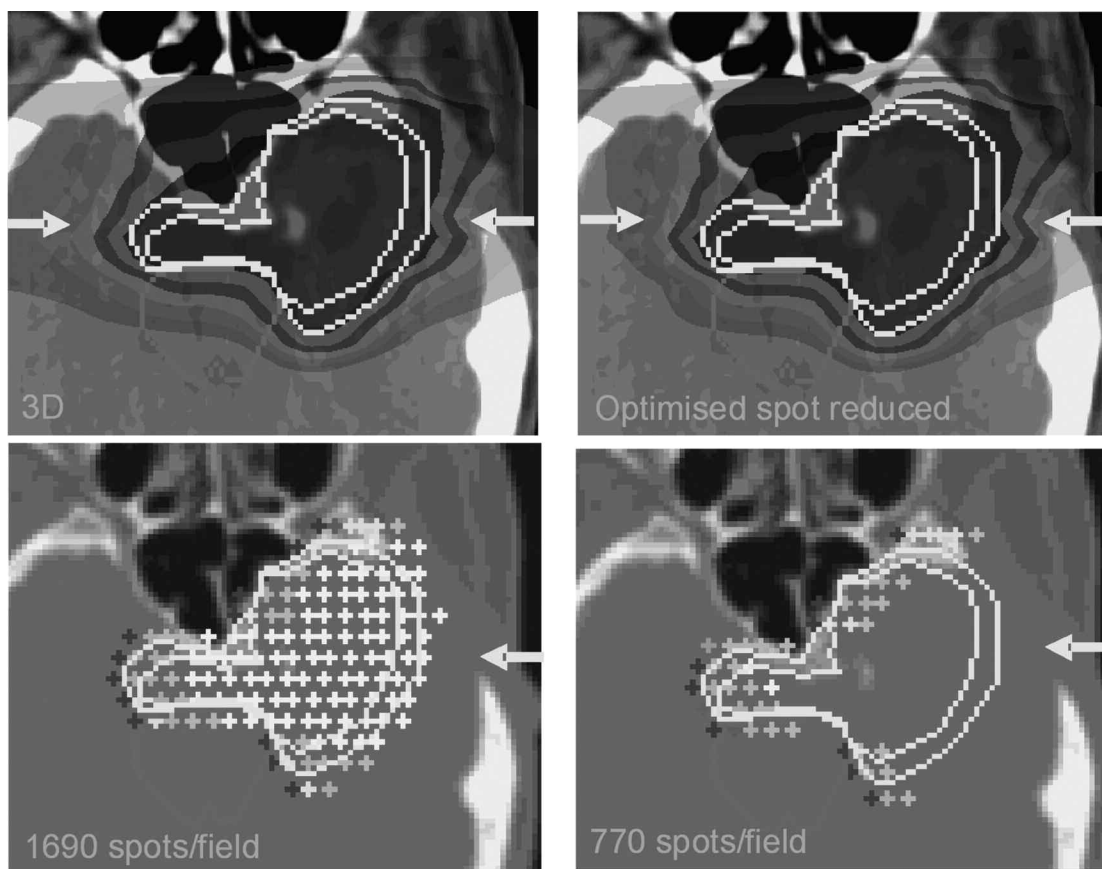


Figure 3. The results of the spot reduction algorithm. The top left panel shows the dose distribution for a two-field plan calculated using the normal (3D) algorithm. Bottom left shows the corresponding spot distribution for the field incident from the right. The highest weighted Bragg peaks are shown in the non-white colors (red, blue, turquoise, etc.). The white Bragg peaks are the lowest weighted. Note, although the Bragg peak weights vary considerably, there is an even covering of the whole target volume. In the right-hand panels are the corresponding figures for the spot reduction algorithm. Although the dose distribution is almost identical to that on the left, this is achieved using almost three times less Bragg peaks than the normal approach. Mainly, the lowest weighted (white) Bragg peaks have been removed. SEE COLOR PLATE 2.

7 An Example of Degeneracy 2: Field Numbers and Orientations in IMPT

As a second example of degeneracy, we present a study looking into the effects of field directions and number of fields in IMPT plans calculated for a set of skull-base chordoma cases. These are radiation resistant tumors originating in the bony structures of the skull base. Although relatively slow growing, these tumors are particularly problematic due to their radioresistance and very close proximity to dose limiting critical structures such as the brainstem and optic structures (optic nerves and chiasm). Although relatively rare tumors, these have become a standard indication for proton therapy, which allows for high levels of conformation allowing for dose escalation to the tumor bed, whilst simultaneously sparing the neighboring critical structures. Figure 4a shows a typical case, with a central and somewhat anteriorly positioned target volume, slightly wrapping around the brain stem. The optic structures are not seen in this figure, as they are positioned just cranially of the target volume. The aim of this study was to look into a set of different field arrangements for treating such cases, to investigate if a “standard” set of beam arrangements could

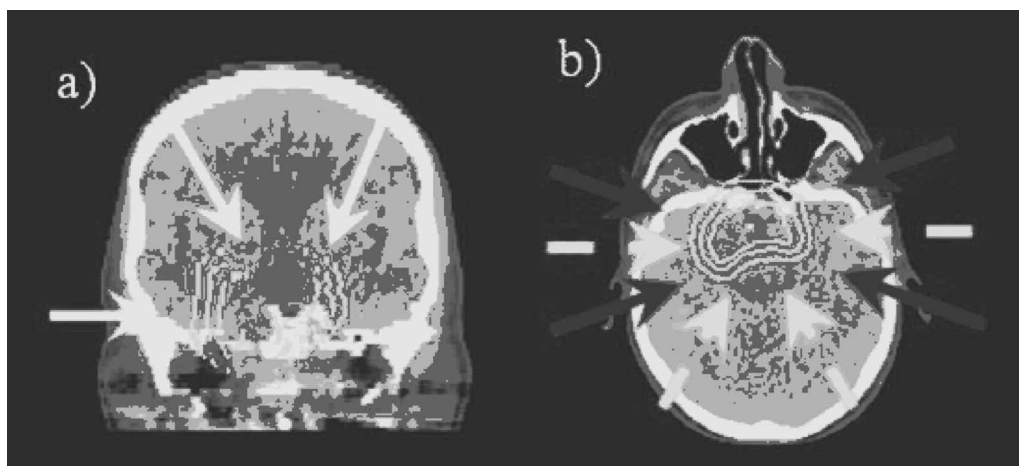


Figure 4. An example of a skull base chordoma used for analyzing the effect of field directions and number of fields on IMPT treatment plans. The inner contour is the gross tumor volume (GTV), the middle the clinical target volume (CTV), and the outer contour the planning target volume (PTV). The aim is to cover the PTV contour with as homogenous a dose as possible. The different field directions are as shown in the two figures. (a) The field configuration of plan A. (b) The other field directions used to construct plans B–E. The red fields show the four fields of the 4F-star arrangement and the green fields, the four fields of the 4F-Pi arrangement. The other two arrangements are combinations of these red and green fields. SEE COLOR PLATE 3.

be determined. The field arrangements investigated are shown in Figure 4b. Five arrangements were investigated:

1. Three fields consisting of one lateral field and two non-coplanar, superior-lateral oblique fields (3F).
2. Four-field star arrangement consisting of two posterior lateral oblique and two anterior lateral oblique fields (4F-star).
3. Four-field “Pi” arrangement consisting of two lateral fields and two posterior-lateral oblique fields, forming a Pi arrangement (4F-Pi).
4. Six-field star, combining the four fields of the 4F-star arrangement plus an additional two lateral fields.
5. Six-field Pi arrangement, consisting of the four fields of the 4F-star plan, plus the two posterior fields of the Pi arrangement.

For each field arrangement, IMPT plans were calculated assuming a target dose of 74 gray (Gy) and with exactly the same dose constraints on the neighboring critical structures—maximum doses of 64 Gy to the brainstem, 53 Gy to the center of the brainstem, and no more than 60 Gy to the optical structures. Priority was given to the critical structure constraints, and the plans were evaluated looking at the various dose-based parameters representing the dose coverage and dose homogeneity to the various target volumes.

Figure 5 shows the results of this analysis. For each target (PTV, CTV, and GTV)⁶, the V95 and D98 for each plan are plotted. The V95 is simply the percentage of the target volume that receives 95% of the prescription dose or more, whereas D98 is the dose (as percentage of the prescription dose) that 98% of the target volume receives. Both parameters are a measure of the minimum dose coverage of the target volume, which is, for such tumors, considered to be a strong indicator for tumor control.^[23] In addition, also in Figure 5, the absolute maximum dose in each plan is plotted.

The results in Figure 5 again clearly show the degeneracy of the optimization problem in IMPT. Despite a wide range of different field arrangements (and different numbers of fields), all the measured parameters are more or less identical, indicating that, regardless of if we take a simple three-field approach or a more complicated six-field arrangement, the target coverage (for the same maximum doses in neighboring critical structures) will be the same. Clearly then, in the absence of any other factors, one would tend towards the simplest three-field approach, as this would be the easiest and fastest to deliver, and increasing the number of fields (or changing their arrangement) brings little or no advantage.

⁶ PTV = planning target volume; CTV = clinical target volume; GTV = gross target volume.

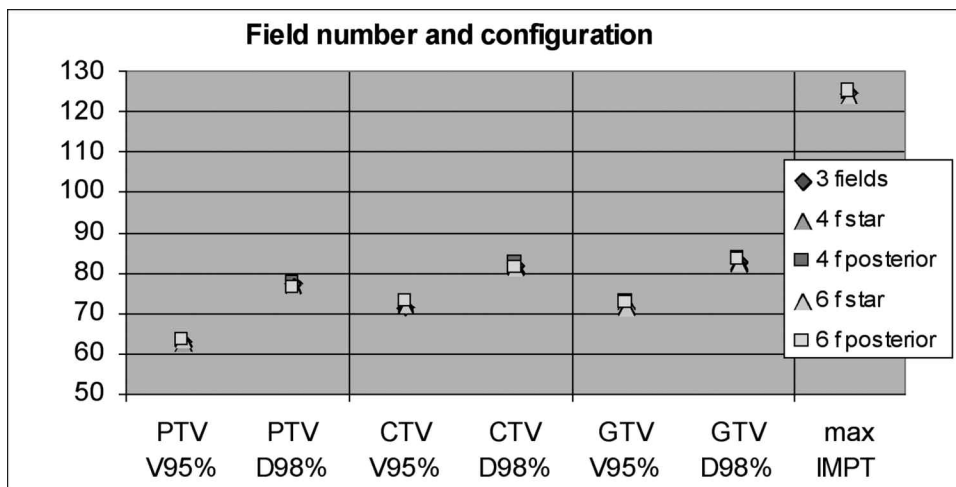


Figure 5. Various plan evaluation parameters calculated for the five field arrangements (A–E). V95 is the volume that receives 95% of the percentage dose. D98 is the dose that 98% of the volume receives, and max is the maximum dose anywhere in each plan. Despite the various field arrangements and different numbers of fields (between 3–6), all parameters are almost identical across all plans (including the maximum dose). SEE COLOR PLATE 4.

However, for this study, we were also interested in investigating not just the nominal dose distributions resulting from the plans, but also how “robust” each field arrangement is against possible delivery uncertainties. One of the most important uncertainties in proton therapy is the range. It is perhaps a truism to state that the advantage of protons is that they stop whereas the disadvantage of protons is that we do not always know where. Nevertheless, potential uncertainty in the range must be taken seriously, particularly as this error could very well be systematic—that is, it is the same for every fraction.

Figure 6 shows an analysis of the robustness for field arrangements B and C (4F-star and Pi arrangements). Figure 6a shows the maximum doses of the nominal plans and the maximum doses when both plans are re-calculated assuming a $\pm 3\%$ systematic change to the CT values of the planning CT. In this case, reducing CT values by 3% will result in a systematic *overshoot*, whereas increasing them by 3% results in a systematic *undershoot* of delivered proton pencil beams. As would be expected from the discussion above, the maximum doses for the nominal plans are essentially identical, as are the maximum doses for the condition of systematic undershoot (+3% HU). However, when simulating an overshoot, the star arrangement is substantially more robust, with the

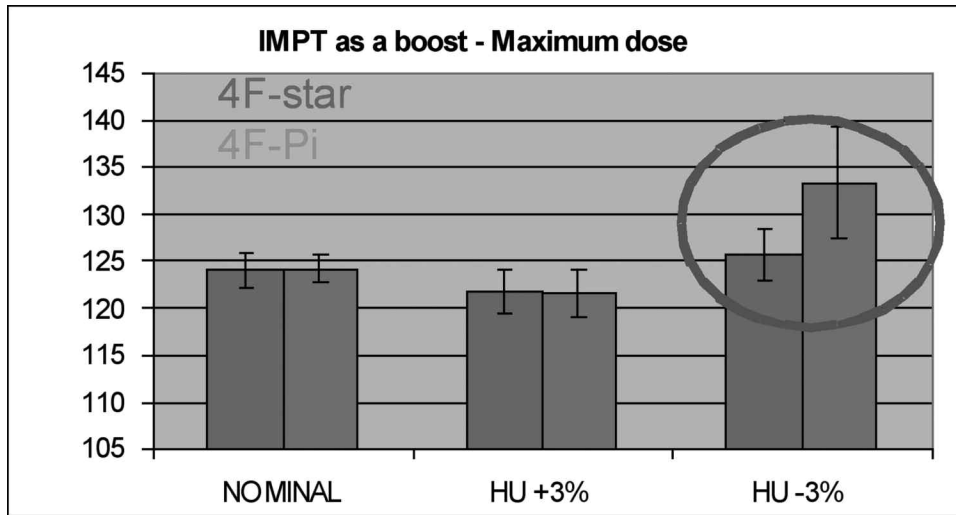


Figure 6. Changes in the maximum dose in the 4F-star and 4F-Pi field arrangements when recalculated on CT datasets with all Hounsfield units (HU) altered by either +3% or -3%. These simulate approximately 3% undershoot (+3% HU) or 3% overshoot (-3% HU), respectively. As seen in Figure 5, there is no difference in maximum dose for the nominal plans, and only a very minimal difference for the undershoot plan (+3% HU). However, for the -3% plans (overshoot), there is a large and significant difference in the maximum dose between to two field arrangements. Clearly, when using the absolute maximum dose in the plan as a criteria, the 4F-Pi arrangement is much more sensitive to 3% overshoot than the 4F-star arrangement. SEE COLOR PLATE 5.

maximum dose increasing only 1% to 2% in comparison with the nominal plan. For the Pi arrangement on the other hand, the maximum dose increases substantially by 9% to 10%, an alarming amount given the relatively modest range uncertainty modeled. This supports the results of a recent, detailed study of the sensitivity of IMPT plans to range and spatial uncertainties published by our group^[24,25], and both confirm the sensitivity of certain types of IMPT plans to different types of delivery errors.

In summary, from both studies reported above, it can be seen that optimization for proton therapy has a very large degeneracy associated with it. However, the second study also shows that, once other plan characteristics are taken into account, differences in plans do begin to become evident. The question is therefore: Can the degeneracy of the solution from the point of view of target coverage and critical structure dose sparing be utilized to design plans that can be differentiated through other, less obvious characteristics?

8 Can Degeneracy Be Utilized?

We believe that degeneracy can be utilized in a number of ways to drive the plans to more desirable results based on non-dosimetric-based criteria. Such an idea has already been touched on by Wilkens et al.^[26,27], who incorporated additional information about relative biological effectiveness (RBE) into the optimization procedure such as to make biologically uniform distributions, and to show that certain classes of plan (in their example, Distal Edge Tracking) were much more sensitive to changes in the RBE than other plans (e.g., 3D).

One other example has already been discussed above; that is, the design of plans that, whilst being nominally identical, are more robust to potential delivery errors. In the example above, a number of plans were calculated with different starting conditions (in this case, field arrangements), and the robustness of each plan assessed *post priori*. When applied to a large number of sample cases, this approach can lead to “case-solutions” for the selection of beam arrangements for given classes of cases. Indeed, based on the work discussed above, we have adopted the four-field star arrangement as our arrangement of choice for skull base chordomas in our clinic, mainly because of the relative robustness of these plans to range errors. Alternatively, one can think of incorporating robustness constraints directly into the optimization process, an approach taken by both the Boston and Heidelberg groups.^[28,29]

Alternatively, one can also think of exploiting degeneracy to construct plans that are more easily deliverable. This was the motivation for the first study described above. The intention was to see to what extent the numbers of pencil beams per field could be reduced, in the hope that this could lead to significantly reduced field delivery times. It is unfortunate that, on analyzing this in detail, the gain in time from spot reduction is rather small, as the main consequence for the delivery is a reduction in the number of energy levels that have to be applied, whilst the other delivery dead-times, and the total beam-on time, remain more or less the same. As for our existing gantry, the time to change energies is only a few ms (50 to 90 ms), then the reduction in overall time by reducing the number of delivered Bragg peaks is small (of the order of 3% to 4%). However, one can certainly think of other ways of forcing the result to more deliverable results. One simple idea is to restrict the dynamic range (i.e., the ratio of the maximum weight to the minimum) of the delivered fields. This, indeed, is a consequence of the spot reduction algorithm described above, with reductions in the dynamic range of a factor 4 or more being observed. In particular, this reduction in dynamic range is mainly due to an increase in the magnitude of the minimum spot weight in each field, which, given the difficulties of accurately measuring low numbers of particles, could have some advantages in the design of delivery hardware and monitoring. Alternatively, as it was found that during the spot reduction process mainly Bragg peaks in the proximal portion of the target were removed, then the resulting plans also had a reduced spread of energies that had to be delivered. Again, in practice, this could lead to a reduction in the number of energy changes that need to be applied per field. Given that for some proton delivery systems, the time to change energies is of the order of 2 seconds or more, then this could potentially indirectly reduce the overall delivery time.

9 Summary: Giving Power to the Planner

In this chapter, we have outlined the concepts behind scanned proton therapy and IMPT. We have also shown that the optimization problem in scanned proton therapy is highly degenerate from the point of view of the resultant dose distributions. However, as one looks in more detail at the individual plans, then degeneracy begins to disappear and could therefore also be utilized to design plans that fulfill other, non-dose based criteria. These include plan robustness and delivery based constraints. However, these are only two possibilities, and many other plan “characteristics” could perhaps be studied and incorporated into the planning process. This could be utilized by incorporating successively more “constraints” into the optimization process. On the other hand, and this is our philosophy, one can give control of the outcome of the planning system to the planner. For example, it is a characteristic of the optimization process presented here, that the resulting dose distribution is heavily dependent on the starting conditions. If these can be stipulated in a straightforward and intuitive way, then this leads to the concept of “planner driven” optimization, where the user drives the optimization towards plans with desirable characteristics by steering it through the manipulation of these starting conditions. We are just beginning to investigate the power of this approach.

References

1. Wilson, R.R. (1946). “Radiological use of fast protons.” *Radiology* 47:487–491.
2. Sisterson, J. (2003). “World wide charged particle patient totals 2003.” *Particles* Newsletter.
3. Koehler, A.M., R.J. Schneider, and J.M. Sisterson. (1975). “Range modulators for protons and heavy ions.” *Med Phys* 31:437–440.
4. Koehler, A.M., R.J. Schneider, and J.M. Sisterson. (1977). “Flattening of proton dose distributions for large fields.” *Nucl Instrum Meth* 4:297–301.
5. Kanai, T., K. Kawachi, Y. Kumamoto, H. Ogawa, T. Yamada, H. Matsuzawa, T. Inada et al. (1980). “Spot scanning system for radiotherapy.” *Med Phys* 7:365–369.
6. Pedroni, E., R. Bacher, H. Blattmann, T. Böhringer, A. Coray, A. Lomax, S. Lin et al. (1995). “The 200 MeV proton therapy project at PSI: Conceptual design and practical realization.” *Med Phys* 22:37–53.
7. Weber, D.C., H. P. Rutz, E. Pedroni, A. Bolsi, B. Timmermann, J. Verwey, A.J. Lomax, and G. Goitein. (2005). “Results of spot-scanning proton radiation therapy for chordoma and chondrosarcoma of the skull-base: The Paul Scherrer Institute experience.” *Int J Radiat Oncol Biol Phys* 63:401–409.
8. Rutz, H.P., D. Weber, S. Sugahara, B. Timmermann, A.J. Lomax, A. Bolsi, E. Pedroni, A. Coray, M. Jermann, and G. Goitein. (2007). “Extracranial chordoma: Outcome in patients treated with function-preserving surgery followed by spot-scanning proton beam irradiation.” *Int J Radiat Oncol Biol Phys* 67:512–520.
9. Timmermann, B., A. Schuck, F. Niggli, M. Weiss, A. J. Lomax, E. Pedroni, A. Coray, M. Jermann, H.P. Rutz, and G. Goitein. (2007). “Spot-scanning proton therapy for malignant soft tissue tumours in childhood: First experiences at the Paul Scherrer Institute.” *Int J Radiat Oncol Biol Phys* 67:497–504.

10. Lomax, A.J. “Intensity Modulated Proton Therapy” in *Proton and Charged Particle Radiotherapy*. T. Delaney and H. Kooy (eds). Boston: Lippincott Williams & Wilkins. 2008.
11. Lomax, A.J. (1999). “Intensity modulated methods for proton therapy.” *Phys Med Biol* 44:185–205.
12. Lomax, A.J., T. Böhringer, A. Coray, G. Goitein, M. Grossman, P. Juelke, S. Lin et al. (2001). “Intensity modulated proton therapy: A clinical example.” *Med Phys* 28:317–324.
13. Pedroni E. “Therapy Planning System for the SIN-Pion Therapy Facility” in *Treatment Planning for External Beam Therapy with Neutrons*. G. Burger, A. Breit, J.J. Broerse (eds.). Munich: Urban and Schwarzenberg, pp. 60–69, 1981.
14. Lomax, A.J., E. Pedroni, B. Schaffner, S. Scheib, U. Schneider, and A. Tourovsky. “3D Treatment Planning for Conformal Proton Therapy by Spot Scanning” in *Proceedings of 19th L H Gray Conference*. London: British Institute of Radiology, pp. 67–71, 1996.
15. Bortfeld, T., J. Buerkelbach, R. Boesecke, and W. Schlegel. (1990). “Methods of image reconstruction from projections applied to conformation radiotherapy.” *Phys Med Biol* 35:14231434.
16. Wu, Q., and R. Mohan. (2000). “Algorithms and functionality of an intensity modulated radiotherapy optimization system.” *Med Phys* 27:701–711.
17. Pflugfelder, D., J.J. Wilkens, S. Nill, and U. Oelfke. (2008). “A comparison of three optimization algorithm for intensity modulated radiation therapy.” *Z Med Phys* 18:111–119.
18. Pflugfelder, D. Risk Adapted Optimization in Intensity Modulated Proton Therapy (IMPT). Ph.D. Dissertation, Heidelberg, 2008.
19. Matsinos, E., B. Schaffner, and W. Kaissl. Optimization Algorithms for Intensity Modulated Proton Therapy (IMPT). Presentation at 44th AAPM Annual Meeting, Montreal, Quebec, July 14–18, 2002.
20. Thieke, C., T. Bortfeld, A. Niemierko, and S. Nill. (2003). “From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning.” *Med Phys* 30(9):2332–2339.
21. Lomax, A.J., T. Böhringer, A. Bolsi, A. Coray, F. Emert, M. Jerman, S. Lin, E. Pedroni, H.P. Rutz, O. Stadelmann, B. Timmermann, J. Verwey, D.C. Weber, and G. Goitein. (2004). “Treatment planning and verification of proton therapy using spot scanning: Initial experiences.” *Med Phys* 31:3150–3157.
22. Bosshardt, M., and A.J. Lomax. Optimising Spot Numbers for IMPT. European Society of Therapeutic Radiology and Oncology (ESTRO) Physics Congress, Lisbon, Portugal, September 2005.
23. Qu, W., R. Mohan, A. Niemierko, and R. Schmidt-Ullrich, (2002). “Optimisation of intensity modulated radiotherapy plans based on the equivalent uniform dose.” *Int J Radiat Oncol Biol Phys* 52:497–504.
24. Lomax, A.J. (2008). “Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: The potential effects of calculational uncertainties.” *Phys Med Biol* 53:1027–1042.
25. Lomax, A.J. (2008). “Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: The potential effects of inter-fraction and inter-field motions.” *Phys Med Biol* 53: 1043–1056.
26. Wilkens, J.J., and U. Oelfke. (2004). (A phenomenological model for the relative biological effectiveness in therapeutic proton beams.” *Phys Med Biol* 49:2811–2825.
27. Wilkens, J.J., and U. Oelfke. (2005). “Optimization of radiobiological effects in intensity modulated proton therapy.” *Med Phys* 32:455–465.
28. Unkelbach J., T.C.Y. Chan, and T. Bortfeld. (2007). “Accounting for range uncertainties in the optimization of intensity modulated proton therapy.” *Phys Med Biol* 52:2755–2773.
29. Pflugfelder, D., J.J. Wilkens, and U. Oelfke. (2008). “Worst case optimization: A method to account for uncertainties in the optimization of intensity modulated proton therapy.” *Phys Med Biol* 53:1689–1700.