Title
A novel energy layer optimization framework for spot-scanning proton arc therapy

Journal
Medical Physics, 47(5)

ISSN
0094-2405

Authors
Gu, Wenbo
Ruan, Dan
Lyu, Qihui
et al.

Publication Date
2020-05-01

DOI
10.1002/mp.14083

Peer reviewed

HHS Public Access

Author manuscript

Med Phys. Author manuscript; available in PMC 2021 June 01.

Published in final edited form as:

Med Phys. 2020 June ; 47(5): 2072–2084. doi:10.1002/mp.14083.

A Novel Energy Layer Optimization Framework for Spot-Scanning Proton Arc Therapy

Wenbo Gu[1], **Dan Ruan**[1], **Qihui Lyu**[1], **Wei Zou**[2], **Lei Dong**[2], **Ke Sheng**[1]

[1]Department of Radiation Oncology, University of California—Los Angeles, Los Angeles, CA, 90095, USA

[2]Department of Radiation Oncology, University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract

Purpose: Spot-scanning proton arc therapy (SPAT) is an emerging modality to improve plan conformality and delivery efficiency. A greedy and heuristic method is proposed in the existing SPAT algorithm to select energy layers and sequence energy switching with gantry rotation, which does not promise optimality in either dosimetry or efficiency. We aim to develop a method to solve the energy layer switching and dosimetry optimization problems in an integrated framework for SPAT.

Methods: In an integrated approach, Energy Layer Optimization for Spot-scanning Proton Arc Therapy (ELO-SPAT) is formulated with a dose fidelity term, a group sparsity regularization, a log barrier regularization, and an energy-sequencing (ES) penalty. The combination of L2,1/2-norm group sparsity regularization and log barrier function allows one energy layer being selected per control point. The ES regularization term sorts the delivery sequence from high energy to low energy to reduce the total energy layer switching time (ELST) and subsequently the total delivery time. Within the ES penalty, the gradient of layer weights between adjacent beams is first calculated along beam direction and then along energy direction. The gradients indicate energy switch patterns between two adjacent beams. The time-wise costly energy switch-up is more heavily penalized in the ES term. This ELO-SPAT method was tested on one frontal base-of-skull (BOS) patient, one chordoma (CHDM) patient with a simultaneous integrated boost, one bilateral head-and-neck (H&N) patient and one lung (LNG) patient. We compared ELO-SPAT with Intensity Modulated Proton Therapy (IMPT) using discrete beams and SPArc by Ding et al. For the two arc algorithms, both the plans with and without energy sequencing were created and compared.

Results: ELO-SPAT reduced the runtime of optimization by 84% on average compared with the greedy SPArc method. In both the ELO-SPAT plans with and without ES, one energy layer per control point was selected. Without ES regularization, the energy sequence was arbitrary, with around 40 to 60 switch-up for the tested cases. After adding ES regularization, the number of energy switch-up was reduced to less than 20. Compared with the energy sequenced SPArc plans,

Corresponding author: Ke Sheng, Ph.D., UCLA Radiation Oncology, 200 Medical Plaza Driveway, Los Angeles, CA, 90095, USA, ksheng@mednet.ucla.edu.

Disclosure of Conflicts

The authors have no relevant conflicts of interest to disclose.

the ELO-SPAT plans with ES led to 24% less total energy layer switching time for synchrotron plans and 14% less for cyclotron plans. Both the ELO-SPAT and SPArc plans achieved better sparing compared with the IMPT plans for most OARs, with or without ES. Without ES, the ELO-SPAT plans achieved further improvement of the OARs compared with the SPArc plans, with an averaged reduction of OAR [Dmean, Dmax] by [1.57, 3.34] GyRBE. Adding the ES regularization degraded the plan quality, but the ELO-SPAT plans still had comparable or slightly better sparing than the SPArc plans with ES, with an averaged reduction of OAR [Dmean, Dmax] by [1.42, 2.34] GyRBE

Conclusion: We developed a computationally efficient spot-scanning proton arc optimization method, which solved energy layer selection and sequencing in an integrated framework, generating plans with good dosimetry and high delivery efficiency.

1. Introduction

To further reduce healthy tissue irradiation, the concept of proton arc therapy (PAT) has been proposed since 1997[1–3], to combine the unique dose deposition curve of protons and the benefit of rotating beams. Passive-scattering based proton arc therapy is not practical for clinical application mainly due to the difficulty of changing beam-specific compensator and range modulation wheel during gantry rotation. But with the recent development and increasing adoption of the spot-scanning technique[4–6], the modulation is integrated into the gantry, making PAT technically viable.

Using the modern scanning nozzle, the proton treatments are delivered spot-by-spot and layer-by-layer[7–9]. The time of spot scanning within the same energy layer is on the order of milliseconds, but it requires seconds to change energy to another layer[10,11], particularly from low to high energies. The slow energy layer switch is mainly due to magnetic hysteresis accompanying changing magnetic field strengths in the energy selection system. The energy layer switching time (ELST) cannot be easily reduced. Therefore, for practical proton arc delivery, reduction of the energy switching steps is an essential consideration besides the dosimetric quality.

Different delivery methods have been proposed for spot-scanning proton arc therapy (SPAT), like multiple static fields[12,13], distal edge tracking[14,15], and single energy modulation[16–18]. However, these delivering methods either cannot perform continuous rotation-delivery, or fail to fully utilize the freedoms in spot-scanning techniques, and the delivery efficiency is not optimized.

Ding et al[19] proposed a delivery-efficient and practical algorithm called SPArc. Similar to volumetric modulated arc therapy (VMAT)[20], this greedy algorithm starts with a coarse sampling of beams, also known as control points, and iteratively increases the sampling frequency while redistributing the energy layers, until reaching the desired sampling frequency. By this method, 1–3 energy layers remain active at each control point, ensuring acceptable delivery time. Later this algorithm was updated to optimize the energy delivery sequence from high to low instead of arbitrary switching to further shorten the delivery time[21]. Retrospective studies show the potential of SPArc plans that improve plan dosimetry compared with intensity-modulated proton therapy (IMPT) to lung cancer[19,22], prostate

cancer[23], and whole-brain radiotherapy[24]. Recently, the first prototype of SPArc delivery was performed on a clinical IBA Proteus One proton machine, with a Proton Dynamic Arc Delivery (PDAD) module. It demonstrated the feasibility of SPArc treatment within the clinical requirements[25].

However, in the current SPArc algorithm, the energy layer selection and optimization are greedy and heuristic. Due to the separate sequencing and plan optimization steps[21], the optimality of the plan delivery efficiency and the dosimetric quality cannot be promised. The alternating back-and-forth operation between the fluence map optimization and energy layer processing required for final plan creation is computationally inefficient.

To further improve SPAT, in this work, we present a novel optimization method to integrate energy layer selection and sequencing with scanning-spot optimization in a single framework, which affords a global search of all feasible energy layers and then simultaneously optimizes the energy sequence. The energy layer selection is achieved by the group sparsity regularization previously developed in our group, which were shown effective in beam orientation optimization[26,27]. An energy-sequencing regularization is developed to improve the SPAT delivery efficiency.

2. Materials and Methods

The proposed Energy Layer Optimization incorporated Spot-scanning Proton Arc Therapy (ELO-SPAT) optimization framework aims to select as few as possible energy layers from the available candidate layers and then encourage energy switch from high to low, for a predefined control point. Considering the continuity in gantry rotating, the optimization goal is exactly one active energy layer at each control point.

Motivated by this consideration, the ELO-SPAT framework is formulated with a dose fidelity, a group sparsity regularization, a log barrier regularization, and an energy-sequencing (ES) penalty term. The details are described in the following sections.

2.1. Notations

Before presenting the optimization framework, we establish the following notations.

- B is the number of static beams used as sampled control points. The fixed spacing of 2.5° or 2° is used in this study.
- E is the number of candidate energy layers in each beam. To simplify the notation, it is assumed that each beam has the same candidate energy layers. The infeasible layers can be eliminated during optimization. So the number of all candidate energy layers is $B \times E$.
- The vector $\boldsymbol{x}_{be}$ is the spot intensities of eth energy layer in bth beam. The length of $\boldsymbol{x}_{be}$, denoted as N_{be}, which is the number of scanning-spots in the specific layer, varies with beams and layers.

- The vector $\boldsymbol{x}_b$ is the concatenation of $\boldsymbol{x}_{be}$ (for $e = 1, \ldots, E$) with increasing energy, representing the spot intensities of bth beam. And the vector $\boldsymbol{x}$ is the concatenation of $\boldsymbol{x}_b$ (for $b = 1, \ldots, B$), following the sequence of gantry rotation.

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_b \\ \vdots \\ \boldsymbol{x}_B \end{bmatrix} \text{ and } \boldsymbol{x}_b = \begin{bmatrix} \boldsymbol{x}_{b1} \\ \boldsymbol{x}_{b2} \\ \vdots \\ \boldsymbol{x}_{be} \\ \vdots \\ \boldsymbol{x}_{bE} \end{bmatrix}. \quad (1)$$

- The vectors $\boldsymbol{y}$ and $\boldsymbol{y}_b$ are compact representations of $\boldsymbol{x}$ and $\boldsymbol{x}_{be}$ to eliminate the dimension of scanning-spot, with the element y_{be} being the sum of all elements in $\boldsymbol{x}_{be}$. Therefore, all the $\boldsymbol{y}_b$ (for $b = 1, \ldots, B$) are in the same length of E.

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_b \\ \vdots \\ \boldsymbol{y}_B \end{bmatrix}, \ \boldsymbol{y}_b = \begin{bmatrix} y_{b1} \\ y_{b2} \\ \vdots \\ y_{be} \\ \vdots \\ y_{bE} \end{bmatrix} \text{ and } y_{be} = \sum_{i=1}^{N_{be}} x_{bei}. \quad (2)$$

Equation (2) can also be written as matrix-vector multiplication:

$$\boldsymbol{y} = W\boldsymbol{x}, \quad (3)$$

where W is a summation matrix, to sum up $\boldsymbol{x}$ along the spot dimension.

- A new variable $\tilde{\boldsymbol{y}}_b$ is defined by replacing each element in $\boldsymbol{y}_b$ with 0 except the maximal element.

$$\tilde{\boldsymbol{y}} = \begin{bmatrix} \tilde{\boldsymbol{y}}_1 \\ \vdots \\ \tilde{\boldsymbol{y}}_b \\ \vdots \\ \tilde{\boldsymbol{y}}_B \end{bmatrix}, \ \tilde{\boldsymbol{y}}_b = \begin{bmatrix} \tilde{y}_{b1} \\ \vdots \\ \tilde{y}_{be} \\ \vdots \\ \tilde{y}_{bB} \end{bmatrix} \text{ and } \tilde{y}_{be} = \begin{cases} y_{be}, \ if \ y_{be} = \max(\boldsymbol{y}_b) \\ 0, \ \text{otherwise} \end{cases}. \quad (4)$$

- The matrix D_b^B is a discrete gradient operator for bth beam along the beam direction. For example, $D_b^B \boldsymbol{y}_b$ is a vector of the intensity difference between $\boldsymbol{y}_{b+1}$ and $\boldsymbol{y}_{b,}$

$$D_b^B \boldsymbol{y}_b = \boldsymbol{y}_{b+1} - \boldsymbol{y}_b. \quad (5)$$

- $\widetilde{D}_b^B$ is a specially designed gradient operator to make

$$\widetilde{D}_b^B \boldsymbol{y}_b = D_b^B \widetilde{\boldsymbol{y}}_b . \tag{6}$$

- The matrix D_b^E is a discrete gradient operator for bth beam along energy direction, while ignoring all zero elements. For example,

$$D_b^E \begin{bmatrix} 0 \\ \vdots \\ y_{be_1} \\ 0 \\ y_{be_2} \\ \vdots \\ 0 \\ y_{be_3} \end{bmatrix} = \begin{bmatrix} y_{be_2} - y_{be_1} \\ y_{be_3} - y_{be_2} \end{bmatrix}, \text{ and } \; y_{be_1}, y_{be_2}, y_{be_3} \neq 0 \; . \tag{7}$$

- The matrix A is the dose-calculation matrix. Each column of A is the vectorized dose to each voxel from a unit intensity spot, and A contains all the spots from the entire $B \times E$ candidate layers. Therefore, the product A and $\boldsymbol{x}$ gives the actual dose delivered to the patient.
- The vector $\boldsymbol{d}_0$ stores the ideal dose to each voxel, which is the prescription dose to the target volumes and zero to the OARs.
- A sigmoid operator $\mathcal{S}$ on a vector $\boldsymbol{u}$ of length K, is defined as

$$\mathcal{S}(\boldsymbol{u}) = \begin{bmatrix} s(u_1) \\ \vdots \\ s(u_k) \\ \vdots \\ s(u_K) \end{bmatrix} \text{ and } s(t) = \frac{2}{1 + e^{-\eta t}} - 1, \tag{8}$$

where $s(t)$ is a modified sigmoid function, as a smooth approximation of sign function, and η controls the level of smoothness. The function of $\mathcal{S}$ is to normalize each element in $\boldsymbol{u}$ to −1, 0 or +1.

2.2. Formulation of ELO-SPAT

The ELO-SPAT is formulated as follows:

$$\begin{aligned}\underset{\boldsymbol{x}}{\operatorname{argmin}} &\|A\boldsymbol{x}-\boldsymbol{d}_0\|_2^2+\sum_{b=1}^{B}\sum_{e=1}^{E}\alpha_{be}\|\boldsymbol{x}_{be}\|_2^{1/2}\\ &-\beta\sum_{b=1}^{B}\log\left(\sum_{e=1}^{E}y_{be}\right)+\gamma\sum_{b=1}^{B-1}h(D_b^E\mathcal{S}(\tilde{D}_b^B\boldsymbol{y}_b))\\ &\text{subject to}\quad \boldsymbol{x}\geq 0,\\ &\boldsymbol{y}=W\boldsymbol{x}\,.\end{aligned}\tag{9}$$

The first term is the dose fidelity term, penalizing the actual dose, calculated by $A\boldsymbol{x}$, from the prescription dose $\boldsymbol{d}_0$. A quadratic function is used in this study, but the choice of dose fidelity cost is flexible. The second term is an L2,1/2-norm group sparsity term. With proper tuning of the weighting hyperparameter α_{be}, the non-convex 1/2 norm effectively turns off most candidate layers. But the term alone can result in aggregated layers in some beam blocks and leaving some control points with no layers active, which does not fully utilize the rotating beams. In the third term, a log barrier regularization function is used to distribute the selected layers to the whole gantry rotating range. The term sums up the intensity of each beam and penalizes the zero intensities, therefore forcing each beam to keep at least one layer selected. β is the regularization parameter for the log barrier function. By picking a proper value of β and setting α_{be} large enough, one energy layer per beam can be ensured.

The fourth term regularizes energy-sequencing (ES) with a weighting parameter γ. ES regularization asymmetrically penalizes energy switching low-to-high harder than high-to-low. The details of ES regularization can be found in Section 2.3.

2.3. Energy-sequencing regularization

The method using the group sparsity to select a few layers out of the candidates is by gradually reducing the $\boldsymbol{x}_{be}$ of the layer with a lower weight to zero during the iterations. As a result, in each beam, the layer with the maximal intensity is most likely kept in each iteration. Therefore, in the energy-sequencing term, instead of $\boldsymbol{y}_b$, we control $\tilde{\boldsymbol{y}}_b$, which only keeps the maximal element in $\boldsymbol{y}_b$ and sets all others to zero, as defined in Section 2.1.

To better understand how energy sequencing works, we consider two adjacent beams during gantry rotation, beam b, and beam $b+1$. Assume the only nonzero elements of $\tilde{\boldsymbol{y}}_b$ and $\tilde{\boldsymbol{y}}_{b+1}$ are y_{be_1} and y_{be_2}, respectively, at the position of e_1 and e_2. As shown in Figure 1, if $e_2 > e_1$, meaning energy going up from beam b to beam $b+1$, the vector $\tilde{\boldsymbol{y}}_{b+1}-\tilde{\boldsymbol{y}}_b$, or $D_b^B\tilde{\boldsymbol{y}}_b$, shows a pattern of transitioning from a negative value to a positive value, with possible zeros before, between and after. A sigmoid operator $\mathcal{S}$ is suited to normalize each nonzero element in $D_b^B\tilde{\boldsymbol{y}}_b$ to −1 or +1.

With normalization, $\mathcal{S}(D_b^B\tilde{\boldsymbol{y}}_b)$ is a vector with only two non-zero elements −1, and +1, respectively. When the energy goes up, taking the difference of the non-zero elements along e direction (the D_b^E operator defined in Section 2.1) results in +2.

Similarly, if the energy goes from high to low, the result of D_b^E operation on $\mathcal{S}(D_b^B\tilde{\mathbf{y}}_b)$ is −2. Maintaining the same energy would result in zero values.

The above process can be written as $D_b^E\mathcal{S}(D_b^B\tilde{\mathbf{y}}_b)$ or equivalently $D_b^E\mathcal{S}\left(\widetilde{D}_b^B\mathbf{y}_b\right)$, as defined in Section 2.1. In summary, the value of $D_b^E\mathcal{S}\left(D_b^B\tilde{\mathbf{y}}_b\right)$ indicates the energy changing pattern between adjacent beams in the following relationship:

$$D_b^E\mathcal{S}\left(\widetilde{D}_b^B\mathbf{y}_b\right)=\begin{cases}+2, & \text{energy switch}-\text{up},\\ -2, & \text{energy switch}-\text{down},\\ 0, & \text{energy unchanged}.\end{cases} \tag{10}$$

For delivery efficiency, fewer energy switch-ups during gantry rotating are encouraged. Therefore, positive $D_b^E\mathcal{S}(\widetilde{D}_b^B\mathbf{y}_b)$ is more heavily penalized. In this work, the energy switch-down is less penalized than staying unchanged for two reasons. First, energy switching down has a small impact on the total delivery time. For example, the switching-down time is 0.6 s according to the IBA Proton Dynamic Arc Delivery module, which is considered a negligible increase compared with staying unchanged in this study. Second, doing so encourages more layers to be used for better dosimetry. Mathematically, a one-sided quadratic cost function is used to penalize energy switching. The cost function, $h(t)$, is defined as:

$$h(t)=\begin{cases}\frac{1}{4}(t+2)^2-1, & if\ t\geq -2,\\ -1, & \text{otherwise}.\end{cases} \tag{11}$$

The quadratic term makes the function $h(t)$ convex, smooth and differentiable, and the function definition is designed to give a value of 0 at 0 for simplicity.

Problem (9) is non-differentiable due to L2,1/2-norm. Alternatively, (9) can be solved by the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [28], an accelerated proximal gradient method. The details of solving the problem (9) using FISTA are shown in Appendix A.

2.4. Evaluation

ELO-SPAT was tested on one frontal base-of-skull (BOS) patient, one chordoma (CHDM) patient with the simultaneous integrated boost, one bilateral head-and-neck (H&N) patient, and one lung (LNG) patient. A full arc was used for the H&N case, and a partial arc was used for the rest of the cases. Gantry rotation was assumed clockwise. The control points for individual beams were spaced 2° in the LNG case and 2.5° otherwise. Dose calculation for the scanning spots covering the PTV and a 5 mm margin was performed using matRad[29,30], a MATLAB-based 3D treatment planning toolkit. The spot spacing was 3 mm in the beam direction, and 5 mm in the lateral direction. The dose calculation resolution was 2.5×2.5×2.5 mm^3. We assumed a constant RBE of 1.1. IMPT plans with 2~4 manually selected beams

were created for these tested patients for comparison. The prescription dose, target volume, arc range and IMPT beam angles for each patient are shown in Table I.

To investigate the effectiveness of ES regularization, we created two proton arc plans for each patient with or without ES regularization in (9). In the latter plan, the energy layers were selected by group sparsity and log barrier regularization, but not sequenced.

We compared the ELO-SPAT plans against the arc plans created using the SPArc method proposed by Ding et al[19]. Because robustness is not considered yet in our work, the robust optimization used in SPArc is replaced with a conventional PTV-based fluence map optimization for a fair comparison. We created the SPArc plans using arc setting and dose calculation identical to ELO-SPAT. To match the two plans, in SPArc optimization, we pushed the number of layers per beam to be 1 for as many beams as possible, leaving only a few beams to have two energy layers. Similarly, we created SPArc plans with[21] or without energy sequencing to compare with ELO-SPAT. The SPArc method with or without ES is denoted as SPArc-ES or SPArc-noES, and the ELO-SPAT method with or without ES is denoted as ELO-ES or ELO-noES

All plans were normalized to deliver prescription dose to 95% of volume. PTV homogeneity and D98% were evaluated for each plan. PTV homogeneity is defined as D95%/D5%. The mean and maximum doses for OARs were also evaluated.

The time spent on treatment planning and delivery of ELO-SPAT and SPArc were compared. For delivery, because energy layer switching time (ELST) is the major factor affecting the total delivery time, we use the total time spent on energy layer switching as the surrogate. The times required for switching energy up, down and keeping it unchanged were 5.5 s, 0.6 s, and 0 s according to the IBA Proton Dynamic Arc Delivery (PDAD) module. A constant ELST of 2.1 s was also used for calculation according to the M.D. Anderson proton therapy system[7] with a synchrotron accelerator.

Although robust optimization has not been incorporated into the current framework, the robustness of the arc plans were analyzed to investigate the influence of ELO to plan robustness compared against the conventional SPArc method. Nine scenarios were used, including one nominal case, two range uncertainty cases by scaling the CT number by ±3.5%, and six setup uncertainty cases by shifting the patient along with anteroposterior, superior-inferior, and mediolateral directions (±5 mm for the LNG patient and ±3 mm for the rest). The DVH bands of the PTVs were plotted to compare the PTV coverage of different arc plans under uncertainties.

3. Results

3.1. Optimization and delivery efficiency

The dose calculation and optimization were performed on a Xeon 28-core CPU server operating at 2.40 GHz clock, with Matlab and its Parallel Computing Toolbox. The energy layer delivery sequence for each patient using SPArc or ELO-SPAT with and without energy sequencing is shown in Figure 2. The number of energy switches for different arc plans is

plotted in Figure 3. The optimization time and expected delivery time are also shown in Figure 3.

Without ES regularization, although single energy layer at each control point is achieved, the energies are not ordered, resulting in 40 to 60 switch-up for the tested cases, adding a non-trivial amount of time to delivery. With ES regularization, the energy layer was sequenced to reduce the number of energy switch-up to fewer than 20, which are comparable to that of the SPArc plans with energy sequencing. Despite the similar number of energy up-switching, SPArc uses a regular sequencing pattern with the same number of down-switching between up-switchings. In comparison, ELO-SPAT sequencing patterns vary to meet dosimetric optimization needs. For the synchrotron plans, with a constant ELST of 2.1 s, the total ELST of ELO-ES plans was similar with the ELO-noES plans, with an averaged time reduction of 9%. For cyclotron plans, with an ELST-up of 5.5 s and ELST-down of 0.6s, the total ELST was reduced to around 2 min in ELO-ES from the 4–7 min in ELO-noES, with an averaged reduction of 61%. Meanwhile, considering both the ELO-SPAT and SPArc plans with ES, the ELO plans had 10–30 more unchanged energies between adjacent control points compared with SPArc, therefore leading to an averaged reduction of total ELST time by 24% for the synchrotron plans and by 14% for the cyclotron plans.

In addition to efficient delivery, the ELO-SPAT reduced the runtime of optimization by 84% on average, from the 0.5–2 hours in the SPArc plans to 5–30 min.

3.2. Dosimetry comparison

The DVH comparison of ELO-SPAT and SPArc without ES is shown in Figure 4, the DVH comparison of the two arc plans with ES is shown in Figure 5. The IMPT plan is plotted in both figures for comparison. The mean dose and max dose for several selected OARs were evaluated in the four arc plans and their differences from the IMPT plans are plotted in Figure 6.

All compared plans achieved similar PTV dose coverage. Qualitatively, both the ELO and SPArc plans achieved better sparing compared with the IMPT plans for most OARs, either with or without ES. But in the lung case, the low dose region of the right lung is larger in the arc plans compared with the IMPT plans.

Without ES, the ELO-SPAT plans achieved further improvement of the OARs compared with the SPArc plans. Lower DVH lines are observed in the ELO-SPAT plans. For example, in the CHDM case, the maximum dose to the left and right cochleas were reduced by 8.14 GyRBE and 6.42 GyRBE, respectively. In the lung case, the maximum dose to the spinal cord was reduced by 5.63 GyRBE. On average, the ELO-SPAT plans without ES reduced the [Dmean, Dmax] of the OARs by [1.57, 3.34] GyRBE from the SPArc plans without ES.

While adding ES regularization, the dosimetry of the quality of ELO-SPAT plans degraded but was still slightly better than the SPArc plans with ES. For example, in the CHDM case, the maximum dose to the left and right cochleas were reduced by 2.79 GyRBE and 4.39 GyRBE, respectively. In the lung case, the maximum dose to the spinal cord was reduced by

2.59 GyRBE. On average, the ELO-SPAT plans with ES reduced the [Dmean, Dmax] of the OARs by [1.42, 2.34] GyRBE from the SPArc with ES.

3.3. Convergence and effect of γ

A convergence plot of the ELO-SPAT method, for the BOS patient, is shown in Figure 7. The cost of each component in (9) during the iterations is also plotted. Dose fidelity, group sparsity, and log barrier all converged. The group sparsity is the component with the highest value because the tuning parameter α_{be} need to be large enough to make only one layer selected per control point. Since the differential matrices D_b^E and D_b^B were updated after every iteration, the cost on ES fluctuated during the iterations. But in general, the ES cost started from a high value, meaning lower delivery efficiency, converged to a low value, presenting higher delivery efficiency.

Figure 8 shows how the ES weighting parameter γ affects the number of switch-up and the value of dose fidelity cost. Generally, when γ increases from zero, the number of energy switch-up decreases and the dose fidelity increases. When gamma reaches a certain value, such as 16 in this case, the number of switch-up plateaus, while the fidelity cost is still in the trend of increasing. In this study, $\gamma = 16$ is picked for a minimum number of energy switch-up and the highest delivery efficiency.

3.4. Robustness analysis

Figure 9 shows the PTV DVH bands of the ELO and SPArc plans with and without ES under range and setup uncertainties. In these DVH band plots, the solid lines are the nominal DVHs without uncertainties, the bands bound the worst-case dose distributions, and the horizontal and vertical lines label the worst D95% of each method for each PTV. Qualitatively, the SPArc plans without ES have wider DVH bands and lower worst D95% for all PTVs, representing the worst PTV coverage under uncertainties. The PTV coverage of the rest three methods varies between different patients. The ELO without ES slightly outperforms the others. The averaged worst D95% is 83.5%, 84.6%, 83.5%, and 78.5% for the ELO-ES, ELO-noES, SPArc-ES, and SPArc-noES plans, respectively.

The DVH band width at D50% was also evaluated. It was observed that the ELO plans have narrower band width than the SPArc plans. The averaged band width is 6.7%, 6.4%, 8.9%, and 11.0% of the prescription dose for the ELO-ES, ELO-noES, SPArc-ES, and SPArc-noES plans, respectively.

4. Discussion

We present an integrated energy layer optimization method for scanning-spot proton arc therapy. The novel framework allows an integrated optimization of fluence map optimization, global search of candidate energy layers, and the delivery sequence. The energy-sequencing penalty is added as a soft regularization to dose fidelity term, therefore providing a flexible trade-off between dosimetry and delivery speed. In this work, the final ELO-SPAT plans were selected as the ones with the lowest achievable number of energy switch-up for best delivery efficiency, with slight scarification of dosimetry. In clinical

practice, the balance between dosimetry and delivery time can be tuned case-by-case. In the case of synchrotron where there is no difference between the time between energy layer switch up and down, the ES regularization can be removed for superior dosimetry.

The number of energy switch using the SPArc energy sequencing method is comparable to our optimization result, showing good performance with the heuristic method. On the other hand, energy switching patterns are distinctly different. Compared with the regular SPArc pattern, the energy sequencing pattern using ELO-SPAT is flexible to take advantage of the patient and arc geometry. For example, for the patient with a frontal BOS tumor, when the gantry rotates in clockwise from posterior to anterior in the first half arc, the overall energies of the candidate layers decrease because the tumor becomes shallower from the beam's eye view. The ELO-SPAT algorithm exploits the geometry and makes more switch-down before energy going up. The flexibility can facilitate future arc trajectory optimization to further enhance efficiency and dosimetry.

With a similar number of energy switch-up between the ELO-SPAT and SPArc plans, we observed 15–20% less total energy layer switching time in the ELO-SPAT method. This is because SPArc does not allow the energy to stay at the same level between control points due to the progressive sampling scheme and the way energy layers are distributed. Using ELO-SPAT, energy switching-down, and unchanged are both encouraged, thereby shortening the total energy layer switching time.

Another major benefit of ELO-SPAT is the significantly shortened optimization time by 5–10 fold from SPArc. In the progressive SPArc sampling scheme from coarse to fine control point resolution, repetitive fluence map optimization is required after either energy layer filtration or redistribution, resulting in long optimization runtime. In the ELO-SPAT optimization, although the algorithm starts with all candidate layers, the number of active layers is gradually reduced to the desired number, shrinking the size of the dose matrix needed for calculation during computation, shortening the time for each iteration. Furthermore, ELO needs only one run to obtain the final delivery sequence and fluence map. The optimization was further accelerated by FISTA, which converges at an optimal rate of $O\left(1/k^2\right)$ [28]. We expect additional acceleration using the graphics processing unit (GPU) platform and multi-resolution sampling of the dose matrix.

Due to the L2,1/2-norm for group sparsity and the sigmoid function for ES, problem (9) is highly non-convex. Originally, FISTA has been used to solve convex problems. However, recent FISTA work[26,27] and the convergence results (Figure 5) in this study suggest that FISTA can be used to solve certain non-convex problems with stable convergence. The differential matrices D_b^E and D_b^B need to be updated after every iteration, but their changes are gradual. Regardless of the ES cost fluctuation, the optimization converges in a few hundred iterations. On the other hand, because of the high non-convexity and the need to update D_b^E and D_b^B, a high weighting parameter on the ES term does not necessarily promote high delivery efficiency. As shown in Figure 7, the number of switch-up plateaus after reaching 16 for the BOS case. The main reason is that the problem is trapped in local minima due to the high non-convexity of the problem formulation. Still, the overall trend of

decreasing the number of switch-up and increasing fidelity cost is observed can still be used to guide parameter tuning. While a random initialization is used in current work to assign the initial spots intensities during optimization, which is common but not necessarily optimal, other initialization schemes can be explored to improve the convergence of the optimization problem and overcome the local minima problem.

A limitation of the current ELO-SPAT algorithm is that only one energy layer allowed per control point for simplicity, while a few more layers could lead to better dosimetry with small scarification of delivery time. Without the energy-sequencing term, multiple layers are readily achievable by tuning the group sparsity term. However, allowing multiple energy layers complicates energy sequencing in the current framework that only regulates the layer with maximal weight at each control point. This is a point for future improvement.

In the current problem formulation, the energy switching pattern is used as the surrogate of delivery time and the order of energy switching is penalized by a simple one-sided quadratic function (11). In future work, the function (11) can be designed to directly correlate the cost with machine-specific energy switching time, thereby allowing intuitive control of actual delivery time. Furthermore, the delivery time penalty can be incorporated as a hard constraint instead of the soft regularization in the current framework for the planner to specify the maximal permissible delivery time directly.

We compared the robustness of the ELO-SPAT method and the SPArc method in the presence of simple systematic range and setup uncertainties. The results show non-inferior PTV robustness using the proposed ELO-SPAT algorithm compared with the SPArc method (Figure 9). Because robust optimization is not explicitly incorporated in either framework, it is an area for future research. We recently developed a sensitivity regularization method to improve the robustness of IMPT plans against range and setup uncertainties[27,31]. This sensitivity-regularized robust optimization method is computationally efficient and resilient to worse-than-expected errors. The sensitivity regularization can be readily integrated into ELO-SPAT for increased plan robustness. Nevertheless, there is no consensus on how to evaluate the robustness of proton arc plans, which is more complicated than IMPT. For example, when the patient receives irradiation in a continuous arc, it becomes possible that the proton range is underestimated from one direction and overestimated from another direction, which can lead to more severe tumor underdosage or OARs overdosage than that from uniform range over- or under-estimation. There is also no established guideline or evidence-based clinical outcome to indicate a failure criterion of arc plan robustness. Therefore, important steps in proton arc therapy research are to investigate the difference of robustness between SPAT and IMPT and to develop tools to evaluate the robustness of arc plans in a clinically relevant way.

5. Conclusions

We developed a computationally efficient spot-scanning proton arc method, which solved energy layer selection and sequencing in an integrated optimization framework, generating plans with good dosimetry and high delivery efficiency.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

This research is supported by NIH Grants Nos. R44CA183390, R43CA183390, R01CA188300 and R01CA230278.

Reference

1. Deasy J, Mackie T, DeLuca P. Method and apparatus for proton therapy. 1997.
2. Sandison GA, Papiez E, Block C, Morphis J. Phantom assessment of lung dose from proton arc therapy. Int J Radiat Oncol Biol Phys. 1997. doi:10.1016/S0360-3016(97)00059-X
3. Schreuder AN, Shamblin J. Proton therapy delivery: what is needed in the next ten years? Br J Radiol. 2019. doi:10.1259/bjr.20190359
4. Kanai T, Kawachi K, Kumamoto Y, et al. Spot scanning system for proton radiotherapy. Med Phys. 1980;7(4):365–369. [PubMed: 6248752]
5. Pedroni E, Bacher R, Blattmann H, et al. The 200-Mev proton therapy project at the Paul Scherrer Institute: Conceptual design and practical realization. Med Phys. 1995. doi:10.1118/1.597522
6. Pedroni E, Scheib S, Böhringer T, et al. Experimental characterization and physical modelling of the dose distribution of scanned proton pencil beams. Phys Med Biol. 2005. doi:10.1088/0031-9155/50/3/011
7. Smith A, Gillin M, Bues M, et al. The M. D. Anderson proton therapy system. Med Phys. 2009. doi:10.1118/1.3187229
8. Gillin MT, Sahoo N, Bues M, et al. Commissioning of the discrete spot scanning proton beam delivery system at the University of Texas M.D. Anderson Cancer Center, Proton Therapy Center, Houston. Med Phys. 2010. doi:10.1118/1.3259742
9. Zhu XR, Sahoo N, Zhang X, et al. Intensity modulated proton therapy treatment planning using single-field optimization: The impact of monitor unit constraints on plan quality. Med Phys. 2010. doi:10.1118/1.3314073
10. Cao W, Lim G, Liao L, et al. Proton energy optimization and reduction for intensity-modulated proton therapy. Phys Med Biol. 2014. doi:10.1088/0031-9155/59/21/6341
11. Van De Water S, Kooy HM, Heijmen BJM, Hoogeman MS. Shortening delivery times of intensity modulated proton therapy by reducing proton energy layers during treatment plan optimization. Int J Radiat Oncol Biol Phys. 2015. doi:10.1016/j.ijrobp.2015.01.031
12. Seco J, Gu G, Marcelos T, Kooy H, Willers H. Proton arc reduces range uncertainty effects and improves conformality compared with photon volumetric modulated arc therapy in stereotactic body radiation therapy for non-small cell lung cancer. Int J Radiat Oncol Biol Phys. 2013. doi:10.1016/j.ijrobp.2013.04.048
13. Rechner LA, Howell RM, Zhang R, Etzel C, Lee AK, Newhauser WD. Risk of radiogenic second cancers following volumetric modulated arc therapy and proton arc therapy for prostate cancer. Phys Med Biol. 2012. doi:10.1088/0031-9155/57/21/7117
14. Flynn RT, Barbee DL, Mackie TR, Jeraj R. Comparison of intensity modulated x-ray therapy and intensity modulated proton therapy for selective subvolume boosting: A phantom study. Phys Med Biol. 2007. doi:10.1088/0031-9155/52/20/001
15. Oelfke U, Bortfeld T. Intensity modulated radiotherapy with charged particle beams: Studies of inverse treatment planning for rotation therapy. Med Phys. 2000. doi:10.1118/1.599002
16. Sanchez-Parcerisa D, Kirk M, Fager M, et al. Range optimization for mono- and bi-energetic proton modulated arc therapy with pencil beam scanning. Phys Med Biol. 2016. doi:10.1088/0031-9155/61/21/N565
17. Blanco Kiely JP, White BM. Dosimetric feasibility of single-energy proton modulated arc therapy for treatment of chordoma at the skull base. Acta Oncol (Madr). 2016. doi:10.3109/0284186X.2016.1170199

18. Langner UW, Eley JG, Guerrero M, et al. A method to deliver energy modulated planar proton Arc therapy (EMPPAT). J Prot Ther. 2017;3(312).
19. Ding X, Li X, Zhang JM, Kabolizadeh P, Stevens C, Yan D. Spot-Scanning Proton Arc (SPArc) Therapy: The First Robust and Delivery-Efficient Spot-Scanning Proton Arc Therapy. Int J Radiat Oncol Biol Phys. 2016. doi:10.1016/j.ijrobp.2016.08.049
20. Otto K. Volumetric modulated arc therapy: IMRT in a single gantry arc. Med Phys. 2008. doi:10.1118/1.2818738
21. Ding X, Li X, Liu G, Stevens C, Yan D, Kabolizadeh P. PO-0916 Energy layer switching sequence optimization algorithm for an efficiency proton arc therapy delivery. Radiother Oncol. 2019. doi:10.1016/s0167-8140(19)31336-2
22. Li X, Kabolizadeh P, Yan D, et al. Improve dosimetric outcome in stage III non-small-cell lung cancer treatment using spot-scanning proton arc (SPArc) therapy. Radiat Oncol. 2018. doi:10.1186/s13014-018-0981-6
23. Ding X, Li X, Qin A, et al. Have we reached proton beam therapy dosimetric limitations?–A novel robust, delivery-efficient and continuous spot-scanning proton arc (SPArc) therapy is to improve the dosimetric outcome in treating prostate cancer. Acta Oncol (Madr). 2018. doi:10.1080/0284186X.2017.1358463
24. Ding X, Zhou J, Li X, et al. Improving dosimetric outcome for hippocampus and cochlea sparing whole brain radiotherapy using spot-scanning proton arc therapy. Acta Oncol (Madr). 2019. doi:10.1080/0284186X.2018.1555374
25. Li X, Liu G, Janssens G, et al. The first prototype of spot-scanning proton arc treatment delivery. Radiother Oncol. 2019. doi:10.1016/j.radonc.2019.04.032
26. Gu W, O'Connor D, Nguyen D, et al. Integrated beam orientation and scanning-spot optimization in intensity-modulated proton therapy for brain and unilateral head and neck tumors. Medical Physics. 2018.
27. Gu W, Neph R, Ruan D, Zou W, Dong L, Sheng K. Robust Beam Orientation Optimization for Intensity-Modulated Proton Therapy. Med Phys. 0(ja). doi:10.1002/mp.13641
28. Beck A, Teboulle M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. SIAM J Imaging Sci. 2009;2(1):183–202. doi:10.1137/080716542
29. Cisternas E, Mairani A, Ziegenhein P, Jäkel O, Bangert M. matRad – a multi-modality open source 3D treatment planning toolkit. In: IFMBE Proceedings. Vol 51 ; 2015:1608–1611. doi:10.1007/978-3-319-19387-8_391
30. Wieser HP, Cisternas E, Wahl N, et al. Development of the open-source dose calculation and optimization toolkit matRad. Med Phys. 2017;44(6):2556–2568. doi:10.1002/mp.12251 [PubMed: 28370020]
31. Gu W, Ruan D, O'Connor D, et al. Robust optimization for intensity-modulated proton therapy with soft spot sensitivity regularization. Medical Physics. 2019.

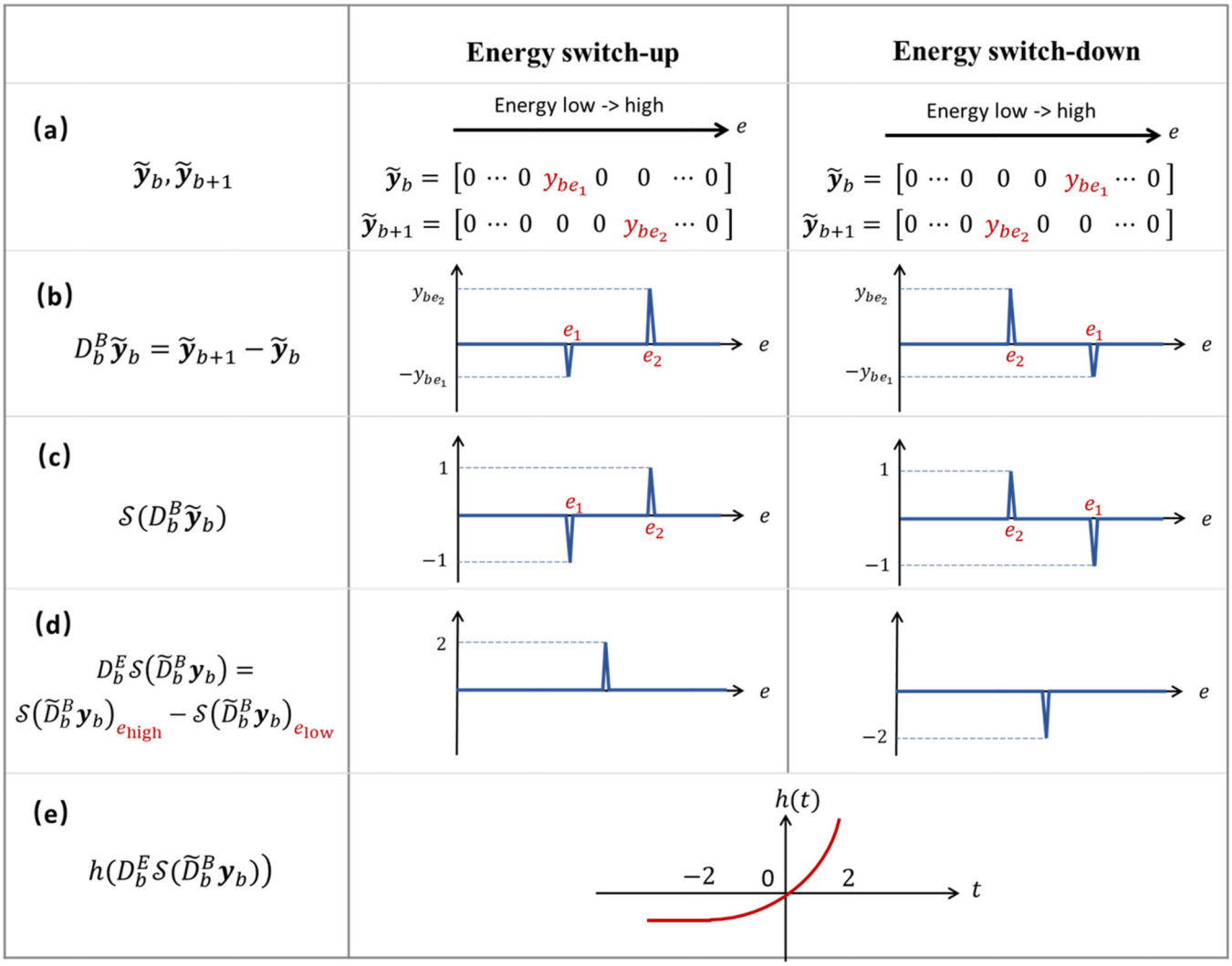

Figure 1.

Schematic workflow of energy-sequencing regularization. Two adjacent beams, beam b and beam $b+1$, are shown, with the situation of energy switch-up in the middle column and energy switch-down in the right column. (a) $\tilde{\mathbf{y}}_b$ and $\tilde{\mathbf{y}}_{b+1}$ are shown as row vectors, while energy increases from left to right. y_{be_1} and y_{be_2} are the sole nonzero element of $\tilde{\mathbf{y}}_b$ and $\tilde{\mathbf{y}}_{b+1}$, respectively. (b) $D_b^B \tilde{\mathbf{y}}_b = \tilde{\mathbf{y}}_{b+1} - \tilde{\mathbf{y}}_b$. (c) $D_b^B \tilde{\mathbf{y}}_b$ is normalized to −1 or +1. (d) Take the gradient of $\mathcal{S}(D_b^B \tilde{\mathbf{y}}_b)$ along e direction, which is the difference of the element at high e index and that at low e index. It yields a positive gradient when energy switches up and a negative gradient when energy switches down. (e) The positive gradient is penalized harder to encourage less energy switch-up.

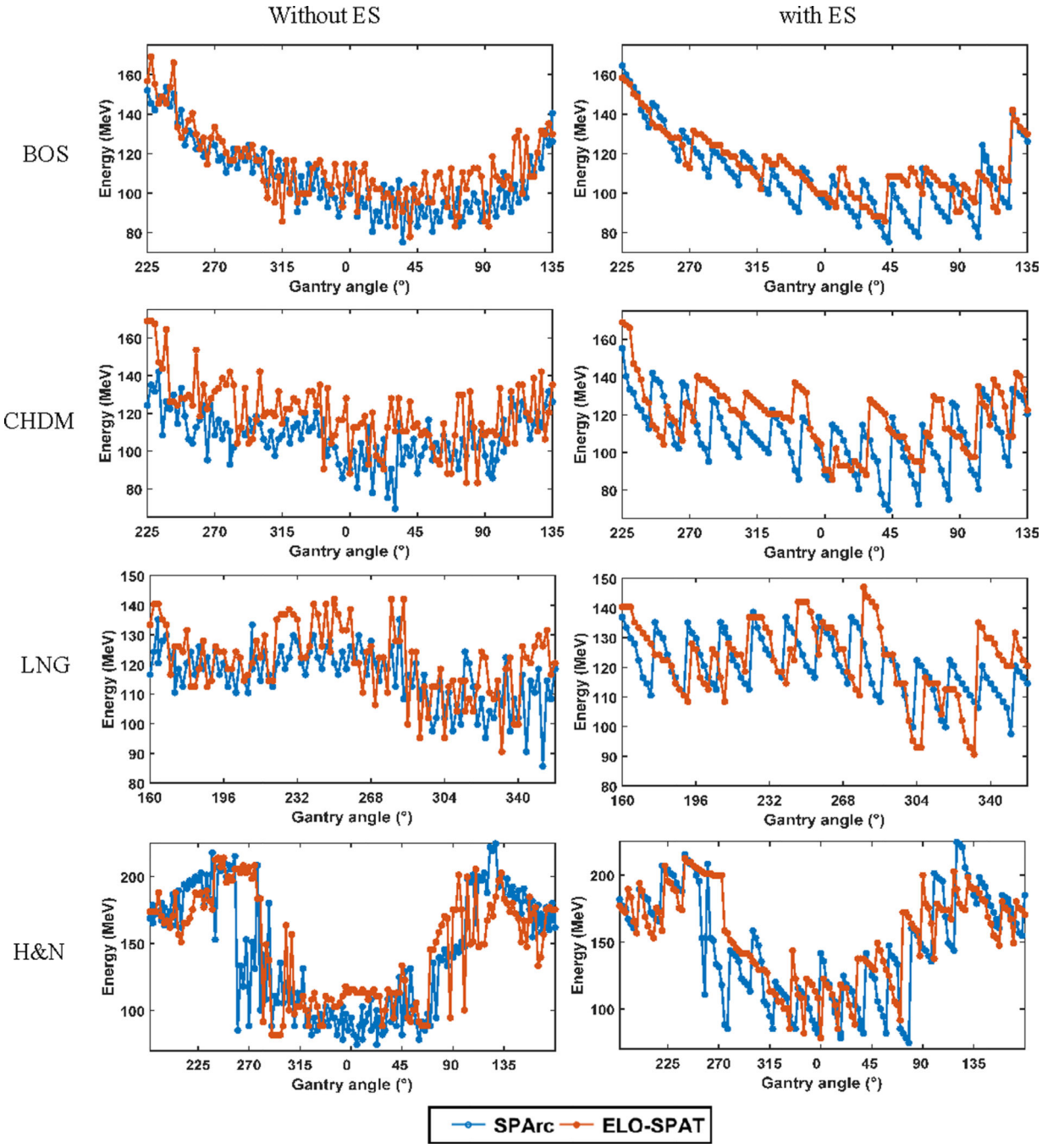

Figure 2.

Energy layer delivery trajectory comparison between SPArc (blue) and ELO-SPAT (red). The gantry rotates in clockwise following the angle of x-axis from left to right. The delivery sequences without ES are shown in the left column and with ES in the right column.

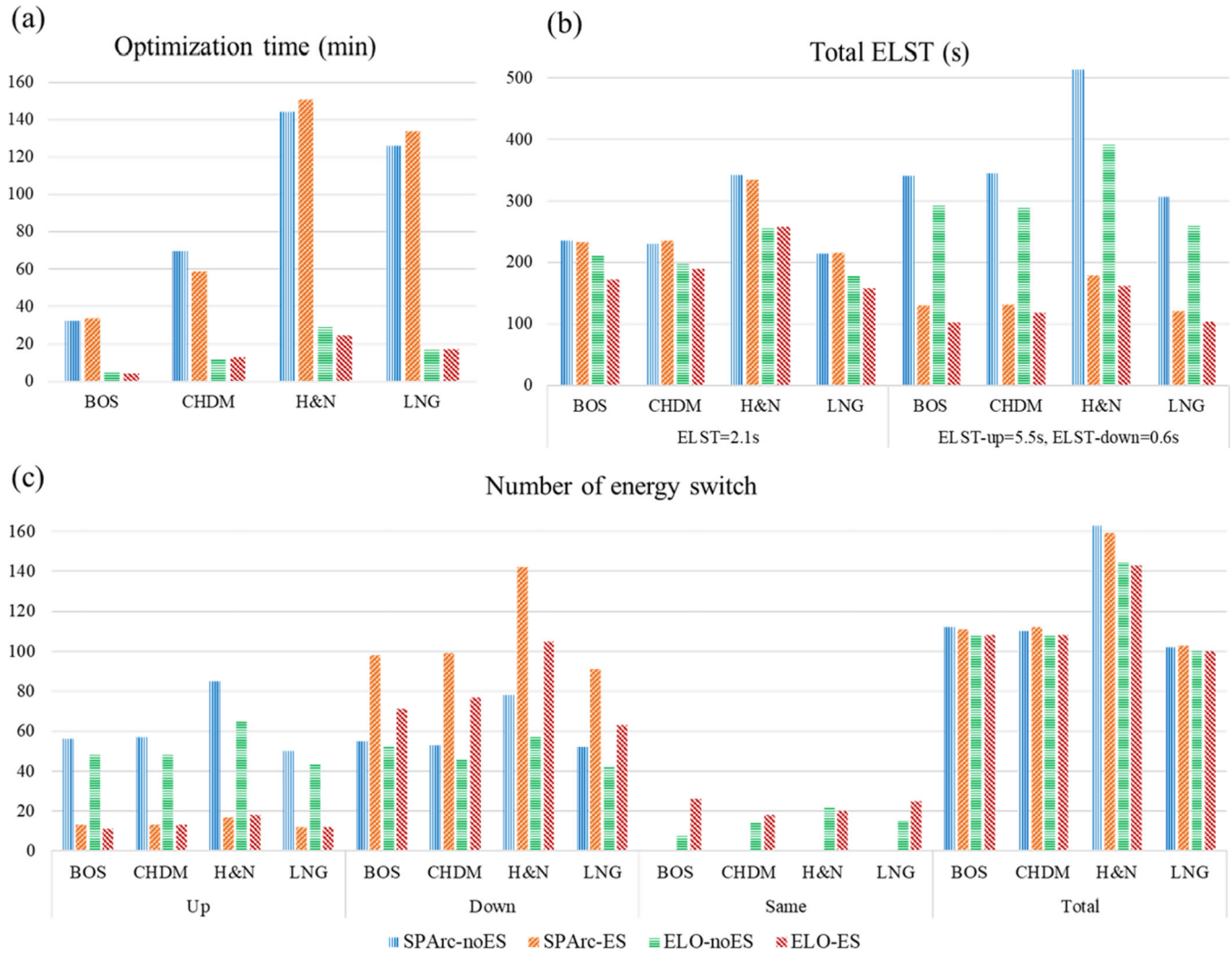

Figure 3.

(a) Optimization runtime of the four arc plans. (b) The total ELST time of the four arc plans when the ELST time is 2.1s (left) and the ELST-up is 5.5 s and ELST-down is 0.6s (right). (c) The number of energy switch up, down, and staying the same. The total number of energy switches is also plotted.

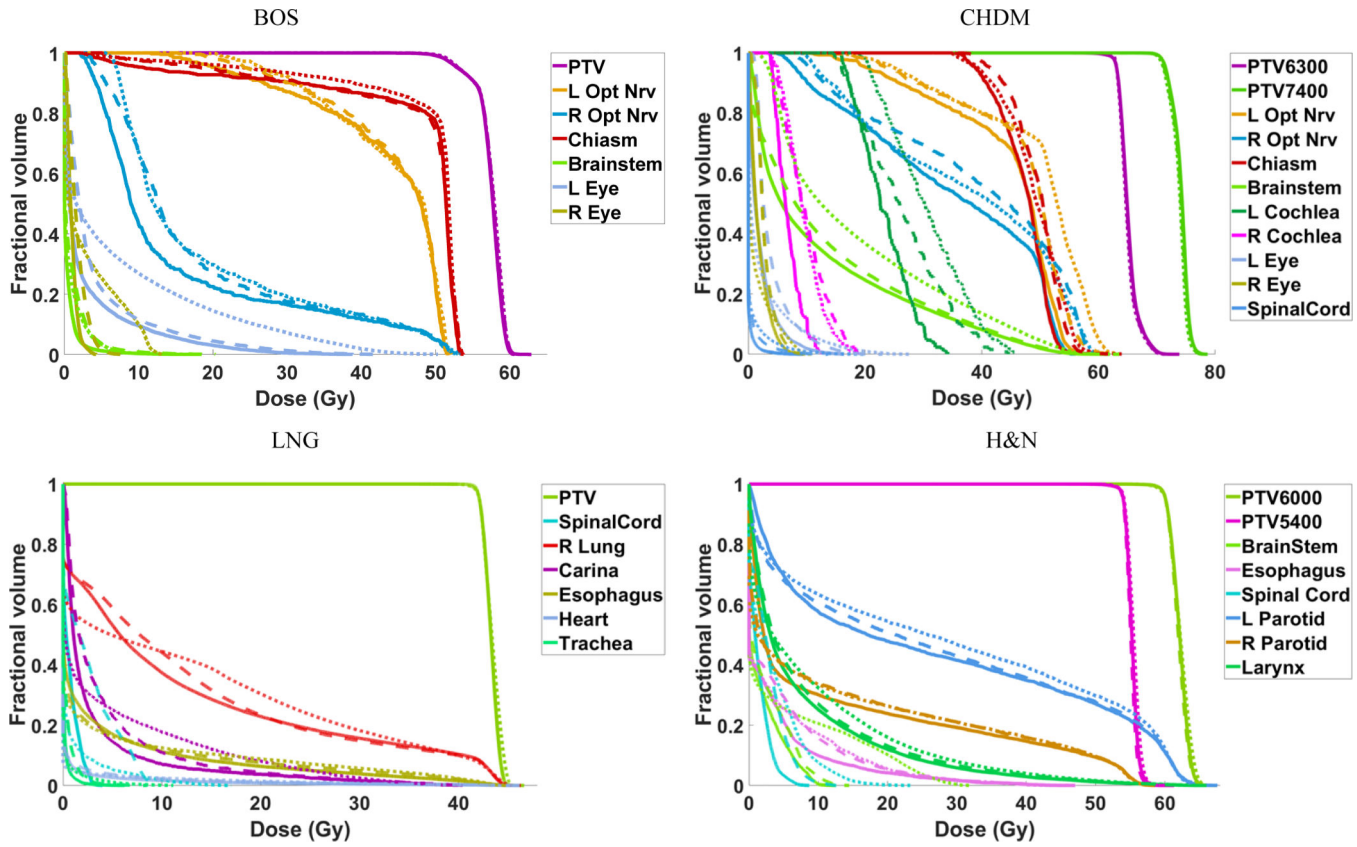

Figure 4.

DVH comparison of plans without ES. The ELO-SPAT plan is in solid line, the SPArc plan is in dashed line, and the IMPT plan is in the dotted line.

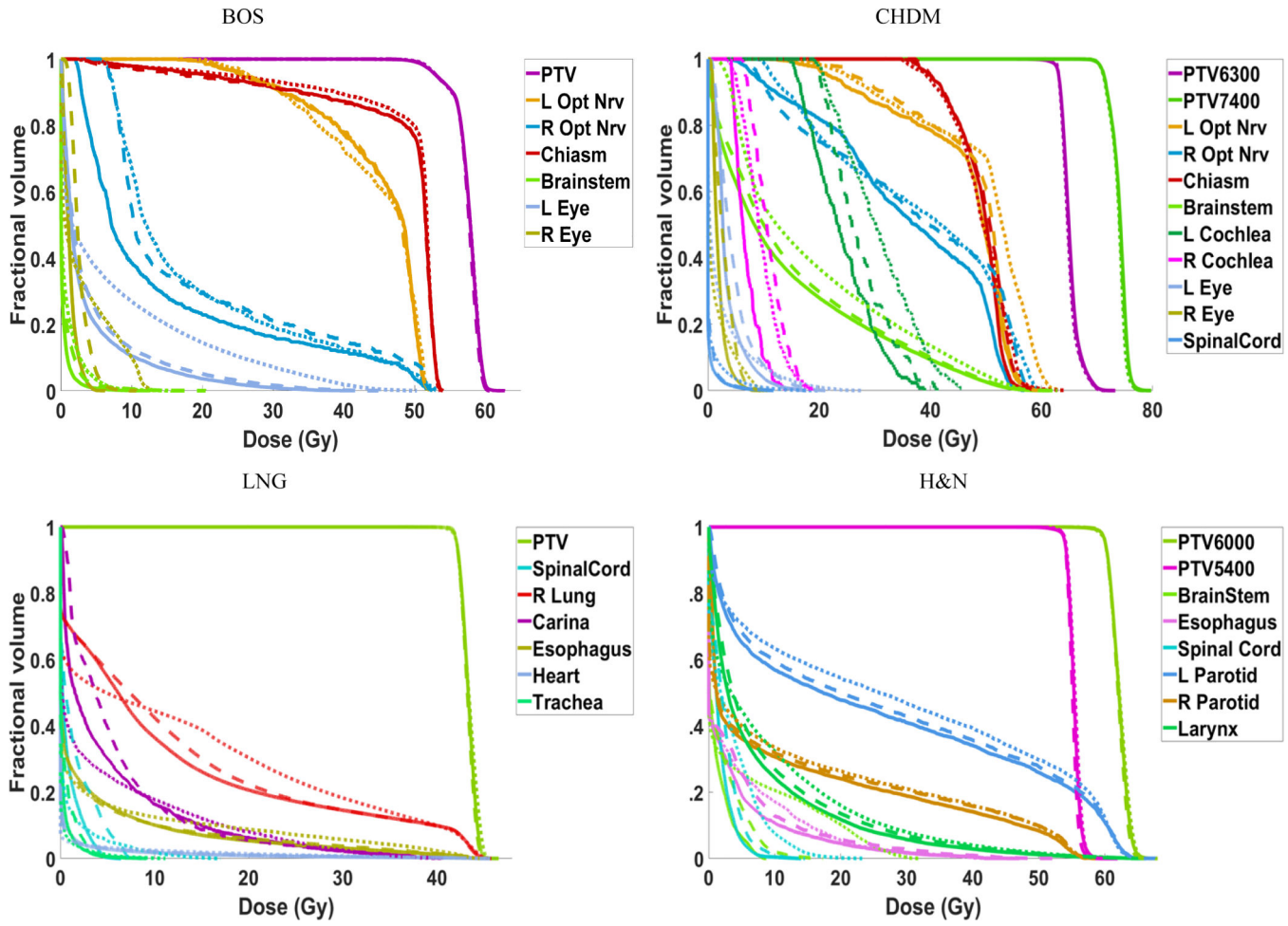

Figure 5.

DVH comparison of plans with ES. The ELO-SPAT plan is in solid line, the SPArc plan is in the dashed line, and the IMPT plan is in the dotted line.

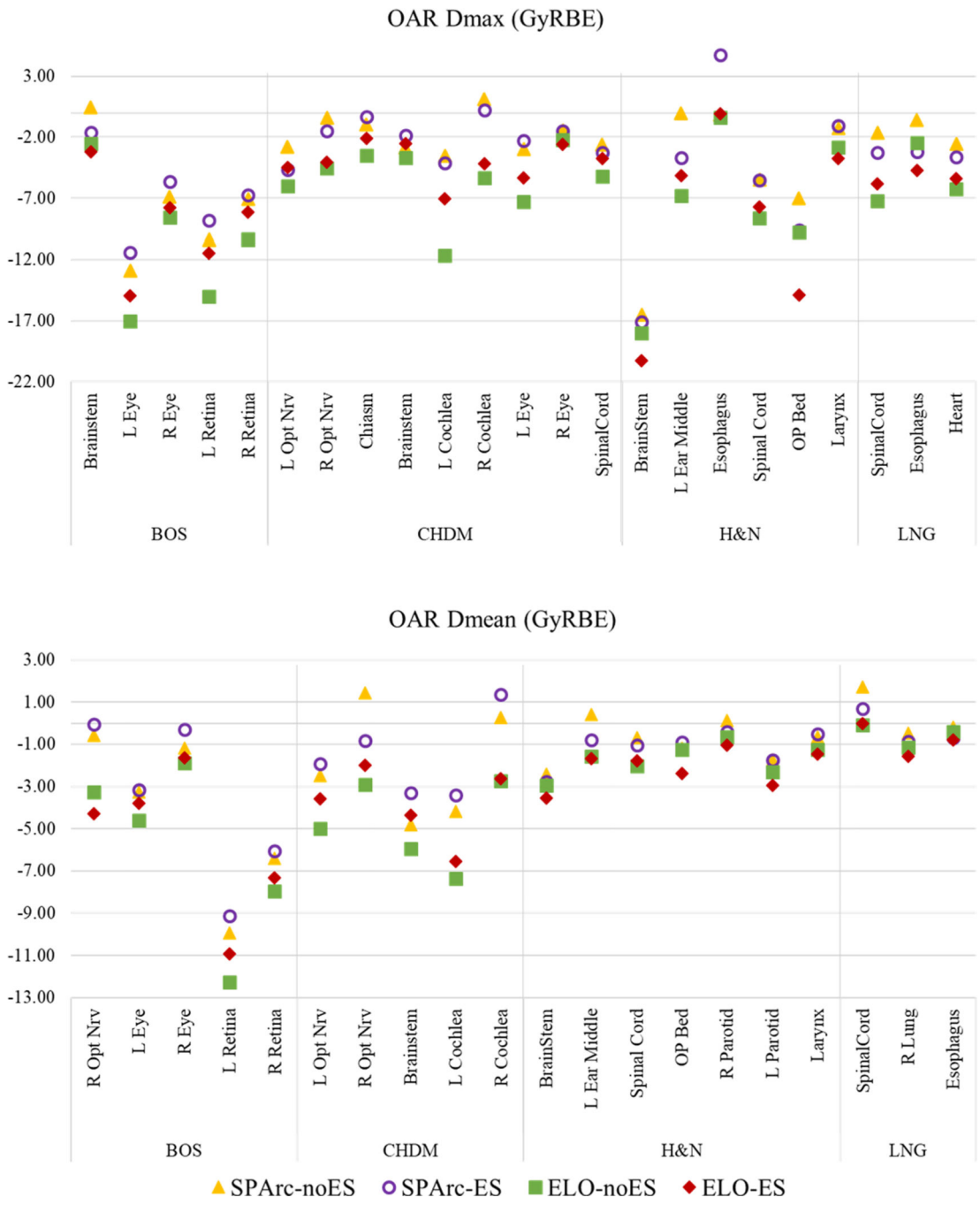

Figure 6.

The difference of OAR Dmax (top) and Dmean (bottom) in the four arc plans from IMPT. A negative value represents a reduction from the IMPT plan, and a positive value represents an increase.

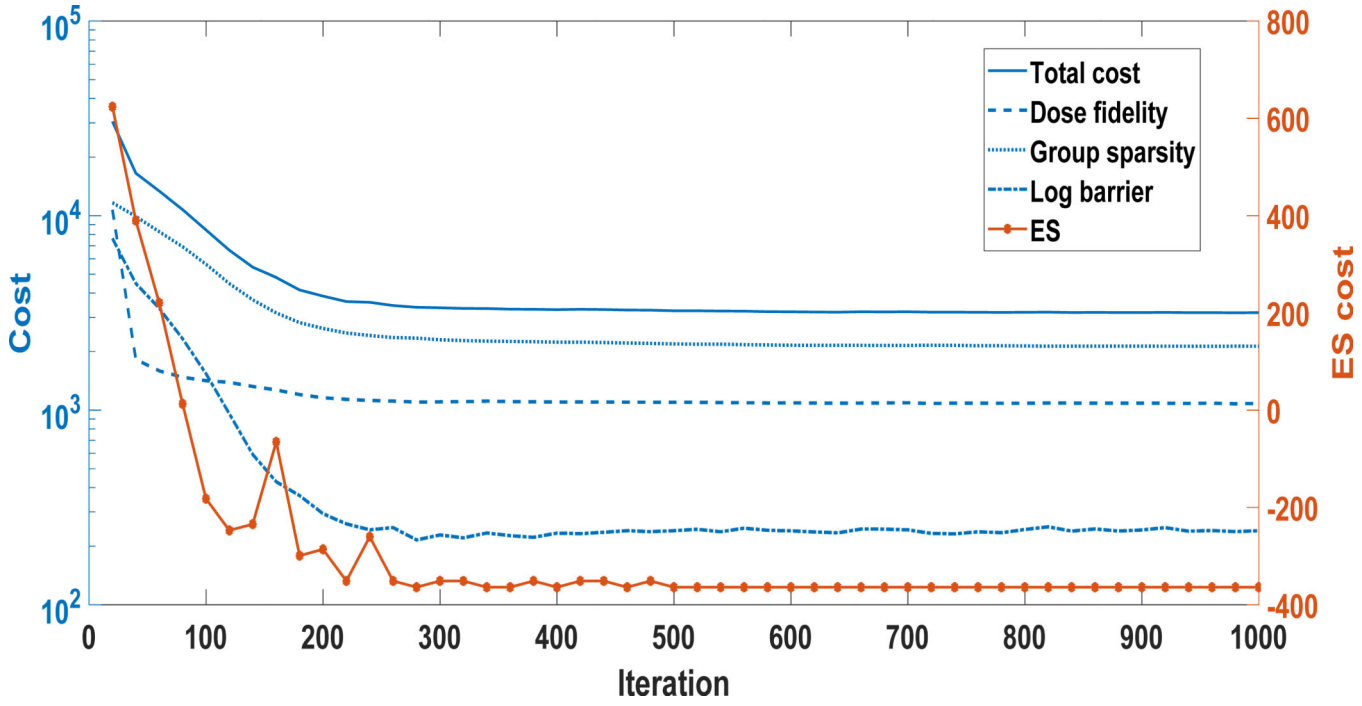

Figure 7.

Convergence plot of the ELO-SPAT method, for the BOS patient. The total cost and the cost components of dose fidelity, group sparsity, and log barrier are shown in the logarithm scale following the y-axis on the left. The ES cost is shown on a linear scale following the axis on the right.

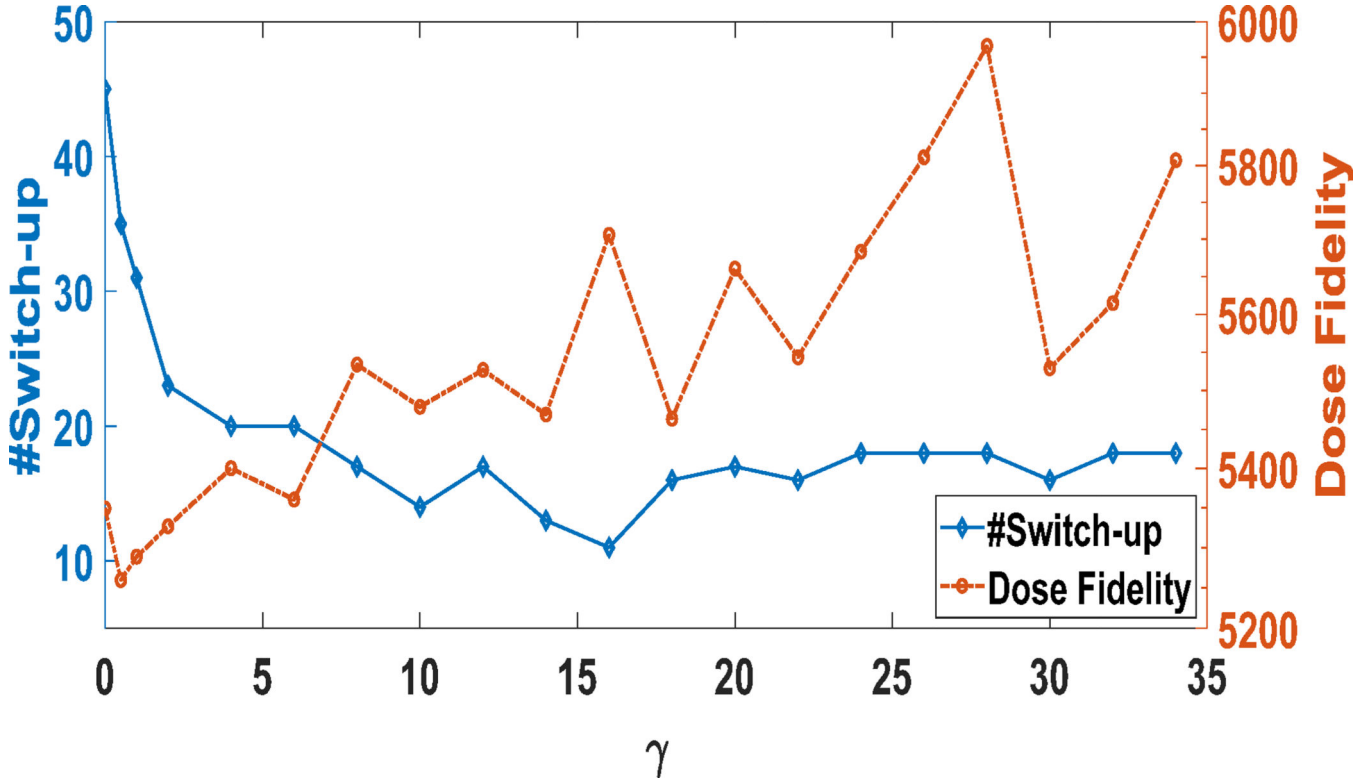

Figure 8.
The number of energy switch-up and the value of final dose fidelity versus γ, for the BOS patient.

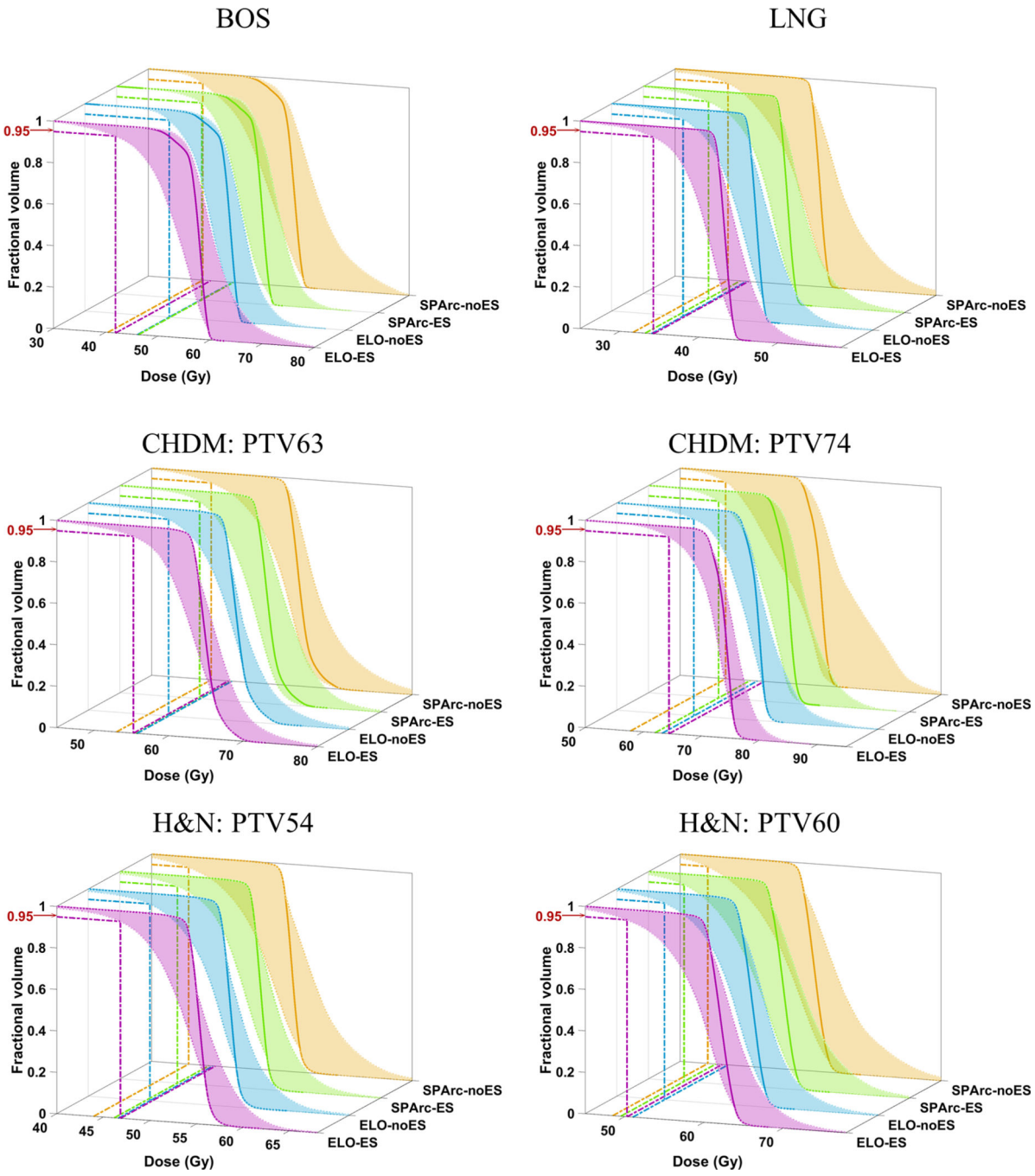

Figure 9.

PTV DVH band indicating the plan robustness. The four arc plans of the same PTV are shown in the same plot with different colors. ELOES is in magenta, ELO without ES in blue, SPArc with ES is in green and SPArc without ES is in khaki. The solid line is the DVH under the nominal situation and the band bounds the worst-case distribution. The worst D95% of each method is labeled by reference lines in the x-y plane. The two PTVs of the

CHDM patient are plotted in the third row, and the two PTVs of the H&N patient are plotted in the fourth row.

Table I.

Prescription doses, PTV volumes, arc range and IMPT beam angles for each patient.

Case		Prescription Dose (GyRBE)	PTV Volume (cc)	Arc angle (degree)			IMPT (gantry, couch) angle (degree)
				Start angle	Stop angle	Spacing	
BOS		56	66.8	225	135	2.5	(60, 273), (270, 0), (90, 0), (180, 0).
CHDM	PTV63	63	128.86	225	135	2.5	(60, 273), (270, 0), (90, 0), (180, 0).
	PTV74	74	26.6				
H&N	PTV54	54	179.1	180	180	2.5	(0, 0), (160, 0), (200, 0).
	PTV60	60	204.4				
LNG		42	297.8	160	0	2	(180, 0), (315, 0).