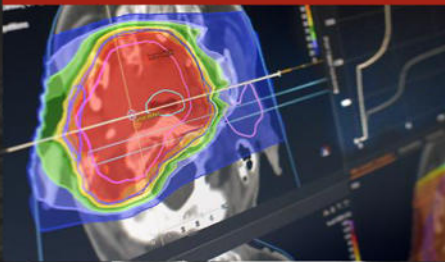


# The Family

Delta<sup>4</sup> TPV



Delta<sup>4</sup> Phantom+



Delta<sup>4</sup> Discover



**QA from prescription to final fraction**

**Delta<sup>4</sup> Family** one software platform  
for all your QA needs.

With Quality Assurance from prescription to final  
fraction you can now increase your workflow  
efficiency and be confident that the treatment  
dose delivered to your patient is safe.

**Delta<sup>4</sup>**  
by ScandiDos

*Innovative and Efficient QA*  
[www.delta4family.com](http://www.delta4family.com)



# GPU-accelerated bi-objective treatment planning for prostate high-dose-rate brachytherapy

Anton Bouter<sup>a)</sup>

*Centrum Wiskunde & Informatica, Science Park 123, 1098 XG, Amsterdam, The Netherlands*

Tanja Alderliesten, Bradley R. Pieters, and Arjan Bel

*Department of Radiation Oncology, Amsterdam UMC University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands*

Yury Niatsetski

*Elekta, Waardgelder 1, 3905 TH, Veenendaal, The Netherlands*

Peter A. N. Bosman

*Centrum Wiskunde & Informatica, Science Park 123, 1098 XG, Amsterdam, The Netherlands  
Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands*

(Received 25 February 2019; revised 22 May 2019; accepted for publication 7 June 2019;  
published xx xxxx xxxx)

**Purpose:** The purpose of this study is to improve upon a recently introduced bi-objective treatment planning method for prostate high-dose-rate (HDR) brachytherapy (BT), both in terms of resulting plan quality and runtime requirements, to the extent that its execution time is clinically acceptable.

**Methods:** Bi-objective treatment planning is done using a state-of-the-art multiobjective evolutionary algorithm, which produces a large number of potential treatment plans with different trade-offs between coverage of the target volumes and sparing organs at risk. A graphics processing unit (GPU) is used for large-scale parallelization of dose calculations and the calculation of the dose-volume (DV) indices of potential treatment plans. Moreover, the objectives of the previously used bi-objective optimization model are modified to produce better results.

**Results:** We applied the GPU-accelerated bi-objective treatment planning method to a set of 18 patients, resulting in a set containing a few hundred potential treatment plans with different trade-offs for each of these patients. Due to accelerations introduced in this article, results previously achieved after 1 hour are now achieved within 30 seconds of optimization. We found plans satisfying the clinical protocol for 15 of 18 patients, whereas this was the case for only 4 of 18 clinical plans. Higher quality treatment plans are obtained when the accuracy of DV index calculation is increased using more dose calculation points, requiring still no more than 3 minutes of optimization for 100 000 points.

**Conclusions:** Large sets of high-quality treatment plans that trade-off coverage and sparing are now achievable within 30 seconds, due to the GPU-acceleration of a previously introduced bi-objective treatment planning method for prostate HDR brachytherapy. Higher quality plans can be achieved when optimizing for 3 minutes, which we still consider to be clinically acceptable. This allows for more insightful treatment plan selection in a clinical setting. © 2019 American Association of Physicists in Medicine [<https://doi.org/10.1002/mp.13681>]

Key words: bi-objective optimization, evolutionary algorithms, GPU, HDR brachytherapy, treatment planning

## 1. INTRODUCTION

Treatment planning for prostate high-dose-rate (HDR) brachytherapy (BT)<sup>1</sup> consists of finding a specific dwell time for each dwell position, such that a desirable dose distribution is achieved. The main difficulty of treatment planning comes from the fact that the target volume(s) should receive sufficient dose, while the surrounding organs at risk (OARs) should receive as little dose as possible. For this reason, some form of automated optimization is often used to find a set of dwell times such that all clinical criteria are satisfied, or are as close as possible to being satisfied. Such clinical criteria

are generally defined in terms of dose-volume (DV) planning criteria (see Section 2.C).

Despite the fact that a treatment plan is often evaluated based on its DV indices, inverse planning methods frequently used in clinical practice do not directly optimize the DV indices.<sup>2,3</sup> Instead, for the purpose of efficiency, a penalty-based approach is used, which aims to minimize the sum of penalty values of all dose calculation (DC) points, with a penalty value assigned to a DC point when the dose at that point is outside the range of acceptable dose values. This reformulated problem is then optimized using, for example, simulated annealing,<sup>3</sup> linear programming,<sup>4</sup> or the Broyden–Fletcher–

Goldfarb–Shanno (BFGS) algorithm.<sup>2</sup> It was previously shown that such a penalty-based problem formulation weakly correlates with the quality of DV indices.<sup>5</sup> The optimal solution to a penalty-based problem formulation may not satisfy all clinical criteria, even if such a treatment plan does exist for the respective patient. Furthermore, finding optimal patient-specific weights for a penalty-based problem formulation is an arguably unintuitive and labor-intensive process.

Several studies have looked into directly optimizing DV indices using linear programming,<sup>6</sup> mixed-integer linear programming,<sup>5</sup> simulated annealing,<sup>7</sup> or evolutionary algorithms.<sup>8</sup> These approaches generally solve a relaxed version of the problem, or use a small number of dose calculation points, because solving plans to optimality directly using DV indices is known to take far too long for clinical practice,<sup>7</sup> necessitating state-of-the-art modern heuristics.

A core aspect of the difficulty of brachytherapy planning is caused by the inherent trade-off between tumor coverage and organ sparing, that is, any increase in dose to the target volumes also leads to an increase in dose to the OARs.

In order to gain more insight into the trade-off between coverage and sparing, a bi-objective treatment planning model and optimization method was previously introduced.<sup>8</sup> Multi-Objective Evolutionary Algorithms (MOEAs) are known to be among the state of the art when dealing with problems involving two or three conflicting objectives.<sup>9</sup> MOEAs are capable of finding large sets of potential solutions, with each solution having a different trade-off between the objectives of interest, that is, tumor coverage and organ sparing in this scenario. The optimization of a set of treatment plans allows for more insightful treatment plan selection, as it is directly clear how much sparing needs to be sacrificed in order to improve the tumor coverage by a certain amount. This can be especially useful in cases where it is not possible to satisfy all criteria defined in the clinical protocol, for example, due to suboptimal catheter placement, making it more difficult to find a desirable trade-off between coverage and sparing.

The aforementioned bi-objective treatment planning method uses a novel MOEA, named MO-RV-GOMEA,<sup>10</sup> as it performed better than other state-of-the-art MOEAs.<sup>8</sup> Moreover, in a recent observer study for 18 prostate cancer patients with three physicians as observers, plans resulting from this approach were found to be preferable compared to plans resulting from the current clinical workflow at our clinic.<sup>11</sup> However, this treatment planning method requires approximately 1 hour of computation time on a single CPU core, exceeding the acceptable amount of computation time in clinical practice.

In this article, we improve the quality of plans found by the previously introduced bi-objective treatment planning method<sup>8</sup> by improving the optimization model and method, and we accelerate the method to the extent that its execution time is clinically acceptable. Specifically, we apply large-scale parallelization on a graphics processing unit (GPU). By doing so, the method is sped up so much that it can even be used to intrinsically improve the approach by increasing the number of so-called dose calculation points used for

calculating the DV indices, decreasing the effect of overfitting and increasing the quality of treatment plans.

## 2. MATERIALS AND METHODS

### 2.A. Data

A dataset of 18 patients with intermediate- and high-risk prostate cancer, previously treated at the Amsterdam UMC in Amsterdam, the Netherlands, is used for all experiments in this article. These patients were treated between February 2015 and April 2017 with external beam radiotherapy on the prostate and base of the seminal vesicles to a dose of 44 Gy in daily fractions of 2.2 Gy followed by a single dose of 13 Gy HDR brachytherapy on the prostate. A median of 16 (range: 14–20) catheters were implanted, resulting in a median of 413 (range: 250–668) dwell positions. Catheter reconstruction and contouring of Regions Of Interest (ROIs) were done on three orthogonal pelvic T2-weighted turbo spin echo MRI (Ingenia 3 T Philips Healthcare, Best, the Netherlands) scans with a resolution of  $0.52 \times 0.52$  mm, and a slice thickness of 3.0 mm with a 0.3 mm gap. Three interpolated contours were added between each contoured slice of each ROI. The clinical plans that these patients were treated with were obtained through optimization with IPSA<sup>3</sup> or HIPO<sup>2</sup> using a standard set of objectives, followed by graphical optimization.

### 2.B. Clinical protocol

In clinical practice, the quality of a treatment plan is often assessed in terms of the DV histogram or a set of DV indices, and a visual inspection of the dose distribution. At the Amsterdam UMC in Amsterdam, the Netherlands, treatment plans were evaluated based on a clinical protocol consisting of a set of aspiration values for various DV indices. A dose index  $D_v^o$  defines the dose received by the most irradiated subvolume  $v$  of ROI  $o$ . A volume index  $V_d^o$  defines the subvolume of ROI  $o$  receiving a dose of at least  $d$ . Any amount of dose is expressed as a percentage of the prescription dose (PD). The clinical protocol for HDR prostate BT at the AMC is given in Table I.

### 2.C. DV index estimation

DV indices are generally estimated using a randomly generated (Monte Carlo-sampling) or a regularly distributed set of DC points.<sup>12</sup> When using Monte Carlo-sampling, ROIs are

TABLE I. Clinical protocol for HDR prostate BT at the AMC. Aspiration values are defined as a percentage of the prescription dose of 13 Gy

| Prostate         | Bladder            | Rectum             | Urethra               | Seminal vesicles |
|------------------|--------------------|--------------------|-----------------------|------------------|
| $V_{100} > 95\%$ | $D_{1cm^3} < 86\%$ | $D_{1cm^3} < 78\%$ | $D_{0.1cm^3} < 110\%$ | $V_{80} > 95\%$  |
| $V_{150} < 50\%$ | $D_{2cm^3} < 74\%$ | $D_{2cm^3} < 74\%$ |                       |                  |
| $V_{200} < 20\%$ |                    |                    |                       |                  |
| $D_{90} > 100\%$ |                    |                    |                       |                  |

represented by a large set of DC points  $\mathbf{d}$ , with  $\mathbf{d}_i$  describing the total dose received by the DC point with index  $i$ . The subset of DC points in an ROI  $o$  is defined as  $\mathbf{d}^o$ . In this work, an equal number of DC points is sampled in each ROI.

A volume index  $V_d^o$  is estimated by calculating the fraction of DC points in ROI  $o$  that receives at least a dose  $d$ , that is,

$$V_d^o = \frac{|\{\mathbf{d}_i \in \mathbf{d}^o | \mathbf{d}_i \geq d\}|}{|\mathbf{d}|}.$$

A dose index  $D_v^o$  is estimated by first selecting the subset of DC points in  $o$  that receives the highest dose, such that the equivalent volume of this set of points is equal to  $v$ . The value of  $D_v^o$  is then equal to the minimum dose received by any DC point in this set. Given the sequence  $\mathbf{d}^{s,o}$ , describing the sequence  $\mathbf{d}^o$  in descending order,  $D_v^o$  is defined as

$$D_v^o = \frac{\mathbf{d}_{\lfloor v/v^o \rfloor}^{s,o}}{PD}.$$

For an accurate estimation of DV indices, using a sufficiently large set of DC points is necessary.<sup>12</sup> This is especially important when DV indices are directly optimized using automated treatment planning, as this approach is prone to overfitting. This means that DV indices optimized using a certain set of DC points will always become worse when recalculated using a larger, more representative, set of DC points. The larger the set of DC points used for optimization, the more accurate the DV index calculation, and the smaller the effect of overfitting. Using a larger set of DC points does however increase the computational effort to estimate DV indices, making the optimization procedure more time consuming. Since the dose in all DC points can be calculated independently, the efficiency of DV index estimation can be greatly increased by the use of a GPU.

In this article, the reported DV indices of any optimized plan are recalculated after optimization on a set of 500 000 independently sampled DC points, identical to the standard setting in Oncentra Brachy (version 4.5, Elekta AB, Stockholm, Sweden).

In Section 3.A, we show how the number of DC points affects treatment plan quality of the automated treatment planning method used in this article.

## 2.D. Bi-objective treatment planning model

Previously, we proposed a bi-objective optimization approach<sup>8,13</sup> using two objectives that are directly defined in terms of the DV indices of a treatment plan. Dwell times of all activated dwell positions, those within a target volume plus 5 mm margin and outside of the urethra plus 1 mm margin, are optimized. This optimization approach was previously validated in a retrospective observer study,<sup>11</sup> showing that plans produced by solving this bi-objective model with the multiobjective evolutionary algorithm GOMEA<sup>10</sup> were preferred over the clinically used plan in all cases.

The two objectives of the bi-objective optimization model, the least coverage index (LCI) and least sparing index (LSI),

quantify how much the DV indices of a treatment plan deviate from the aspiration values specified in the clinical protocol, with regard to tumor coverage and organ sparing, respectively. Positive values for either LCI or LSI indicate that the clinical protocol has been satisfied with regard to coverage or sparing, respectively.

Given the protocol in Table I, the LCI and LSI of a treatment plan  $t$  are defined as

$$\begin{aligned} \text{LCI}(t) &= \min\{\delta_c(V_{100}^{\text{prostate}}), \delta_c(V_{80}^{\text{vesicles}})\}, \\ \text{LSI}(t) &= \min\{\delta_s(D_{1\text{cm}3}^{\text{bladder}}), \delta_s(D_{2\text{cm}3}^{\text{bladder}}), \delta_s(D_{1\text{cm}3}^{\text{rectum}}), \\ &\quad \delta_s(D_{2\text{cm}3}^{\text{rectum}}), \delta_s(D_{0.1\text{cm}3}^{\text{urethra}})\}, \end{aligned}$$

$$\delta_c(V_d^o) = V_d^o - V_d^{o,\min},$$

$$\delta_s(D_v^o) = D_v^{o,\max} - D_v^o,$$

where  $V_d^{o,\min}$ , and  $D_v^{o,\max}$  are aspiration values defined by the clinical protocol. We further refer to values of  $\delta_c$  and  $\delta_s$  as DV delta values.

The aspiration values for  $V_{150}^{\text{prostate}}$  and  $V_{200}^{\text{prostate}}$  are excluded from the LSI, as is done in previous work,<sup>8</sup> because these volume indices have a different unit than the remaining dose indices in the LSI, making them difficult to be compared to each other. Instead, the aspiration values for these volume indices are included as hard constraints, meaning that any plan produced by the optimization procedure must satisfy these aspiration values. Furthermore, the clinical criterion  $D_{90}^{\text{prostate}} > 100\%$  is not included in the LCI, because this criterion is always satisfied when  $V_{100}^{\text{prostate}} > 95\%$  is satisfied.

A benefit of optimizing the LCI and LSI is that their outcome strictly depends on the most violated DV index, which is the DV index with the lowest corresponding DV delta. All optimization effort is therefore spent on trying to optimize the most violated DV index. However, this means that any small improvement of the most violated DV index will be preferred over a much larger improvement of a different index, as long as this does not change which DV index is the most violated. Furthermore, optimization of remaining DV indices is no longer possible if the most violated DV index has reached its optimal value. Intuitively, it is desirable to spend some optimization effort on DV indices that are not the most violated, especially in the aforementioned cases. For this reason, we define adaptively weighted variations of LCI and LSI, denoted  $\text{LCI}_w$  and  $\text{LSI}_w$ .

To compute the  $\text{LCI}_w$ , the required DV deltas are first sorted in descending (DESC) order to find the order of DV indices from least to most violated. For the  $\text{LCI}_w$ , the sorted list of DV deltas is defined as

$$\Delta_c^s = \text{sort}^{\text{DESC}}([\delta_c(V_{100}^{\text{prostate}}), \delta_c(V_{80}^{\text{vesicles}})]).$$

Similarly, for the  $\text{LSI}_w$ , the sorted list of DV deltas is defined as

$$\begin{aligned} \Delta_s^s &= \text{sort}^{\text{DESC}}([\delta_s(D_{1\text{cm}3}^{\text{bladder}}), \delta_s(D_{2\text{cm}3}^{\text{bladder}}), \delta_s(D_{1\text{cm}3}^{\text{rectum}}), \\ &\quad \delta_s(D_{2\text{cm}3}^{\text{rectum}}), \delta_s(D_{0.1\text{cm}3}^{\text{urethra}})]). \end{aligned}$$

Given the sorted list of DV deltas, the  $LCI_w$  and  $LSI_w$  are then defined as a weighted sum of DV deltas, where the DV delta with the highest value, that is, corresponding to the least violated DV index, receives a weight of 1 and consecutive weights increase exponentially by a factor of  $\lambda = 10$ . All weights are then normalized to the range between 0 and 1. Note that the weights associated with individual DV deltas are dynamic throughout optimization, because the sorted order may change, and are not set by a user.

$LCI_w$  and  $LSI_w$  are defined as

$$\begin{aligned} LCI_w(\mathbf{t}) &= w_c(\delta_c(V_{100}^{\text{prostate}})) + w_c(\delta_c(V_{80}^{\text{vesicles}})), \\ LSI_w(\mathbf{t}) &= w_s(\delta_s(D_{1cm3}^{\text{bladder}})) + w_s(\delta_s(D_{2cm3}^{\text{bladder}})) \\ &\quad + w_s(\delta_s(D_{1cm3}^{\text{rectum}})) + w_s(\delta_s(D_{2cm3}^{\text{rectum}})) \\ &\quad + w_s(\delta_s(D_{0.1cm3}^{\text{urethra}})), \\ w_c(\delta) &= \frac{\lambda^{r-1}}{\sum_{i=1}^{|\Delta_c^s|} \lambda^{i-1}} \delta \quad \text{with } r \text{ s.t. } \Delta_c^s[r] = \delta, \\ w_s(\delta) &= \frac{\lambda^{r-1}}{\sum_{i=1}^{|\Delta_s^s|} \lambda^{i-1}} \delta \quad \text{with } r \text{ s.t. } \Delta_s^s[r] = \delta, \end{aligned}$$

where rank  $r$  is a one-based index that indicates the position of the respective DV index in the sorted list of indices.

For  $\lambda = 1$ ,  $LCI_w$  and  $LSI_w$  define equally weighted linear combinations of their respective DV deltas. This would result in a loss of focus on the most violated DV index, making it more difficult to find solutions satisfying all clinical criteria. As  $\lambda$  increases, the importance of the most violated DV index is increased. For  $\lambda$  approaching infinity,  $LCI_w$  and  $LSI_w$  are identical to the LCI and LSI. The value of  $\lambda = 10$  was empirically determined to produce equally good solutions in terms of LCI and LSI as the previously used bi-objective model,<sup>8</sup> while increasing the quality of DV indices that are not the most violated.

In this article, we use MO-RV-GOMEA with the  $LCI_w$  and  $LSI_w$  as optimization objectives, unless stated otherwise. Note that the  $LCI_w$  and  $LSI_w$  no longer have the property that a positive value indicates that all clinical criteria have been satisfied. For this reason, we generally report LCI and LSI in Section 3, even when  $LCI_w$  and  $LSI_w$  were used as optimization objectives.

In Section 3.B, we analyze how the results of the optimization are affected by using the  $LCI_w$  and  $LSI_w$  as optimization objectives, compared to the LCI and LSI.

## 2.E. Multiobjective evolutionary algorithm

Evolutionary Algorithms (EAs)<sup>14</sup> are computational heuristics that can be distinguished by the fact that a set, called the *population*, of potential solutions is maintained, rather than a single solution. In our case, a population of 96 solutions is used, where each solution represents a treatment plan. The *fitness* of a solution, in this case defined in terms of LCI and LSI, defines its relative quality. During each *generation* of an EA, *variation* is applied in order to generate a set of new potential solutions, called the *offspring*. By applying

*selection*, the solutions with the best fitness from the combined population and offspring form the population of the following generation, and non-selected solutions are discarded.

EAs are known to be among the state of the art in the field of multi-objective optimization.<sup>9</sup> The result of a multi-objective EA is a set of solutions that each have a different trade-off with regard to the objectives. In this article, we specifically use the Multi-Objective Real-Valued Gene-pool Optimal Mixing Evolutionary Algorithm (MO-RV-GOMEA),<sup>10</sup> as it was shown to achieve the best performance among a set of state-of-the-art multiobjective EAs on bi-objective treatment planning for prostate HDR BT.<sup>8</sup> MO-RV-GOMEA uses an archive with a maximum capacity of 1250 to store all plans for which no better plan was previously found. In this context, we consider a plan to be better when it Pareto-dominates the other plan in terms of LCI and LSI, that is, it has a better LCI value and an LSI value that is at least as good, or vice versa. This archive is constantly updated whenever new plans are found. At the end of an optimization run, all plans in the archive are presented as the results of the optimization.

We refer the interested reader to Section S-A of the Supporting Information for detailed parameter settings of MO-RV-GOMEA.

## 2.F. GPU-accelerated bi-objective treatment planning

GPUs are widely used for the purpose of high-performance computing, both within the field of medical physics,<sup>15</sup> for example, for dose calculation,<sup>16–19</sup> and in the field of evolutionary computation.<sup>20–23</sup> Because the design of GPUs is particularly suited for specific forms of large-scale parallel computation, they are very suitable for high-performance dose calculation, as the dose in all DC points can be calculated in parallel.

By far the most time-consuming procedures of the bi-objective treatment planning method are the dose calculation and the calculation of the DV indices of the treatment plans in the population. Due to the use of an EA, these procedures are always performed on the entire population, further increasing the potential for parallelization. We therefore use a GPU to improve the performance of the population-wide dose and DV index calculation. Specifically, the GPU that is used in this work is an NVIDIA Titan Xp, and all programming for the GPU is done in CUDA.<sup>24</sup> Using an NVIDIA Titan Xp GPU, our treatment planning method currently supports the optimization of sets up to 500 000 DC points.

We use a parallel dose calculation method that we previously introduced.<sup>25</sup> In this method, the dwell times of all treatment plans in the population are stored in a matrix  $T$ , where each row describes the set of dwell times of one treatment plan. The dose in all DC points is then calculated through the matrix–matrix multiplication  $D = TR^T$ , where  $R$  is a dose-rate matrix according to the TG-43 model.<sup>26</sup> Element  $R[i,j]$  describes the dose contribution in Gy/s of the source at dwell position with index  $i$  to the DC point with index  $j$ . Each row of  $D$  describes the dose in each DC point

of one potential treatment plan in the population. The number of rows of  $\mathbf{D}$  is therefore equal to the population size, and the number of columns is equal to the number of DC points.

The used optimization algorithm, MO-RV-GOMEA,<sup>10</sup> often finds new potential treatment plans by modifying a small number of dwell times of a potential treatment plan currently existing in the population. When this is the case, the dose distribution  $\mathbf{D}'$  of the modified treatment plan is calculated more efficiently through  $\mathbf{D}' = \mathbf{D} + \Delta\mathbf{TR}^T$ , with  $\Delta\mathbf{T}$  the matrix of differences in dwell times.<sup>25</sup> Matrix  $\Delta\mathbf{T}$  has few columns with non-zero elements, equal to the number of dwell times that have been modified. Only these columns with non-zero elements are considered for the calculation of  $\mathbf{D}'$ .

The estimation of dose indices requires sorting the dose in the DC points of each ROI for which a dose index must be calculated, for each potential treatment plan. For this purpose, each DC point  $d_i$  is assigned an identifier  $\text{ID}_i$ , such that DC points in the same ROI of the same treatment plan are assigned an identical ID. A list  $L$  of dose/ID pairs is then created, that is,  $L = [(d_0, \text{ID}_0), (d_1, \text{ID}_1), \dots]$ . The Thrust library,<sup>27</sup> which contains a large number of CUDA utility functions, is then used to sort  $L$  by dose value in ascending order. This is followed by the Thrust stable sort procedure, sorting  $L$  by ID in ascending order. The stable sort procedure sorts  $L$  by ID, but maintains the relative order of items with an identical ID. As the dose values of  $L$  were previously sorted, this means that  $L$  will be sorted by ID, but dose values with the same ID remain sorted. The above two sorting procedures therefore result in a list of DC points that are firstly sorted by key, and secondly sorted by dose. A dose index is then calculated according to Section 2.C.

The estimation of volume indices requires counting the number of DC points that receive at least a certain dose. For this purpose, all required DC points are first compared to their respective aspiration value in parallel, resulting in a large array of Boolean values. Thrust is then used to compute a cumulative sum of this array, after which the number of points achieving their respective aspiration value can be trivially computed for any subsequence of this array. A volume index is then calculated according to Section 2.C.

The sampling of DC points and the creation of the dose-rate matrix are not included in the optimization times presented in this article. Dose points are sampled using four CPU cores in parallel, and the dose-rate matrix is calculated on a single CPU core and then copied to the GPU memory. For a set of 100 000 DC points, this takes approximately 2 seconds on average.

In Section 3.C, we compare treatment plans optimized with the GPU-accelerated bi-objective treatment planning to the set of clinical plans.

### 3. RESULTS

In Section 3.A, we show how the outcome of optimization in terms of LCI and LSI depends on the number of DC points. Section 3.B then shows the effect of the introduction

of the  $\text{LCI}_w$  and  $\text{LSI}_w$  on the optimization. Finally, the quality of the results achieved by our bi-objective treatment planning method is shown in Section 3.C.

#### 3.A. DV index estimation

In Fig. 1, we illustrate the impact of the number of DC points on the optimization results. This figure shows the treatment plans found after completing 30 runs of the optimization with a time limit of 10 minutes on a set of 5000, 10 000, 20 000, 50 000, 100 000, and 200 000 DC points, respectively. A time limit of 10 minutes was used for all settings, as this time limit was found to be sufficient to obtain the maximum achievable performance for each of the tested number of DC points.

Figure 1(a) shows the perceived LCI and LSI, calculated using sets of 5000, 10 000, and 20 000 DC points used for optimization. The perceived LCI and LSI values, that is, as computed by the optimization algorithm, suggest that better treatment plans can be found when smaller numbers of DC points are used. However, Fig. 1(b) shows a substantially more accurate calculation of the LCI and LSI values of the identical set of plans, using a set of 500 000 independently sampled DC points. Because Fig. 1(b) shows a reduction in LCI and LSI compared to Fig. 1(a), this indicates that the quality of the resulting plans is substantially worse than perceived during optimization. This is caused by the fact that too small sets of DC points are not a good representation of their respective ROI. Due to this effect, no plan satisfying all clinical criteria was found for 8 of 30 runs when 5000 DC points were used for optimization. This was still the case for 2 of 30 runs when 10 000 DC points were used for optimization.

Figures 1(c) and 1(d) show that the difference between perceived and accurate LCI and LSI are much smaller when sets of at least 50 000 DC points are used for optimization. Furthermore, comparing Figs. 1(b) and 1(d) shows that the actual LCI and LSI obtained when solving for up to 20 000 DC points are clearly worse than the optimization algorithm perceives, whereas the difference between perceived and accurate LCI and LSI is much smaller for the set of 100 000 DC points.

As Fig. 1(d) shows a negligible difference between optimization outcomes for 100 000 and 200 000 DC points, a set of 100 000 DC points seems sufficiently large for a reliable calculation of the LCI and LSI during optimization.

#### 3.B. Bi-objective treatment planning model

Figure 2 shows how the use of  $\text{LCI}_w$  and  $\text{LSI}_w$  as optimization objectives, shown in the bottom row, affect optimization results, compared to the use of LCI and LSI, shown in the top row. In the middle plots, all DV indices of each treatment plan are displayed in a separate color. Each treatment plan is therefore represented by five different data points in different colors (online version only), by definition aligned vertically, as each plan is defined by a single value for its LCI. In the rightmost plots, the coverage of the prostate

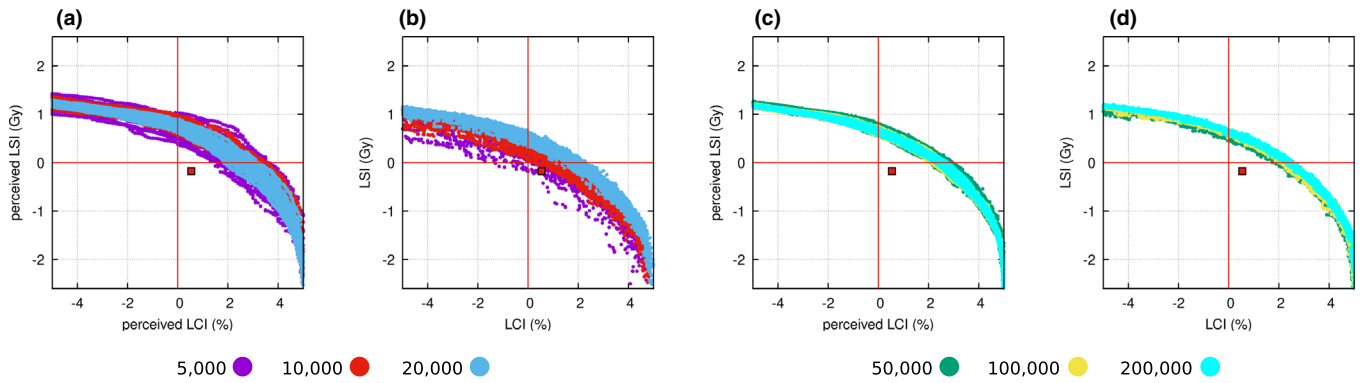


FIG. 1. Treatment plans found after 10 minutes for patient 1, shown in different colors (online version only) pertaining to the number of DC points used for optimization. All figures show the combined results of 30 independent runs (to show the degree of variance in outcomes) for each setting of the number of DC points. Perceived LCI and LSI ((a) and (c)) indicate the LCI and LSI values as calculated by the optimization algorithm on the given set of DC points. More accurate LCI and LSI values ((b) and (d)) are calculated using a set of 500 000 independently sampled DC points. The clinical plan is shown as a red square.

and seminal vesicles are shown for each treatment plan. Each treatment plan is represented by two data points in different colors (online version only), aligned horizontally.

The leftmost two figures show no perceivable difference of the LCI and LSI between weighted and nonweighted LCI and LSI. This means that both approaches find very similar values for the most violated DV indices in terms of coverage and sparing. However, the remaining subfigures show that optimizing the  $LCI_w$  and  $LSI_w$  results in better values for the DV indices that are not the most violated.

The middle subfigure of Fig. 2 shows slight improvements for  $D_{1cm3}^{bladder}$ ,  $D_{2cm3}^{bladder}$ ,  $D_{1cm3}^{rectum}$ , and  $D_{2cm3}^{rectum}$  in case of using  $LCI_w$  and  $LSI_w$  instead of LCI and LSI. In addition, much better values are found for the  $V_{80}^{vesicles}$ , showing that, for this patient, satisfying the clinical criterion for this index is possible regardless of the achieved LSI. Note that this is not the case for each patient, as shown in Figs. S1–S4, because this depends on the patient anatomy and catheter placement. As shown in Fig. 4, it is still difficult to achieve good vesicle coverage for patient 1, and for various other patients for which results are shown in Figs. S1–S4 of the Supporting Information. Overall, however, the results indicate that optimizing the  $LCI_w$  and  $LSI_w$  leads to better results in terms of individual DV indices than optimizing the LCI and LSI.

### 3.C. GPU-accelerated bi-objective treatment planning

We now show more detailed results obtained by optimizing for sets of 20 000 and 100 000 DC points, as the former setting was used when our bi-objective approach was introduced and validated,<sup>8,11</sup> and the latter setting is one we decided upon in Section 3.A. Results obtained by optimizing for sets of 5000, 10 000, 50 000, and 200 000 DC points are included in Table SI of the Supporting Information. Furthermore, Table SII of the Supporting Information contains results for 20 000 and 100 000 DC points using time limits of 30, 60, 180, and 300 seconds. In Fig. 3, we show the

results of 30 optimization runs, each with a newly sampled set of DC points, for patients 1, 2, and 3, optimized on 20 000 and 100 000 DC points.

The optimization with 20 000 DC points resulted in a set of 461 treatment plans on average. After reevaluation using 500 000 DC points, 236 plans remained on average, as any plan was removed from the set of results when it was Pareto-dominated by a different plan. The optimization with 100 000 DC points resulted in a set of 658 treatment plans on average. After reevaluation using 500 000 DC points, 428 plans remained on average.

When 20 000 DC points are used for the optimization, results obtained after 30 seconds are similar to those obtained after 10 minutes. Furthermore, variation in LCI and LSI can be fairly large for this number of DC points. When 100 000 DC points are used, a clear increase in LCI and LSI is observed for time limits up to 180 seconds. Most importantly, there is little variation in the 30 different runs using 100 000 DC points, meaning that this number of DC points more reliably results in a set of high-quality treatment plans. Although there is still a minor difference between optimizing for 180 or 600 seconds, we argue that this difference can be considered irrelevant. Optimizing for 300 or 600 seconds could be considered depending on the strictness of clinical time constraints.

Table II shows, for each patient used in this study, the clinical plan along with two plans selected from the result of the optimization on 20 000 DC points and 100 000 DC points. A time limit of 30 seconds was used for the optimization on 20 000 DC points, and 3 minutes for the optimization on 100 000 DC points. For both settings of the number of DC points, the optimization was performed 30 times per patient. The selected plan is the plan with the maximum achieved LCI that satisfied all sparing criteria, that is, achieved an LSI larger than 0. This plan is selected to show how much coverage can be achieved while all sparing criteria are satisfied. Although we show only one plan from each optimization run in Table II, each optimization run produced many plans with different trade-offs between coverage and sparing, as shown in Fig. 4.

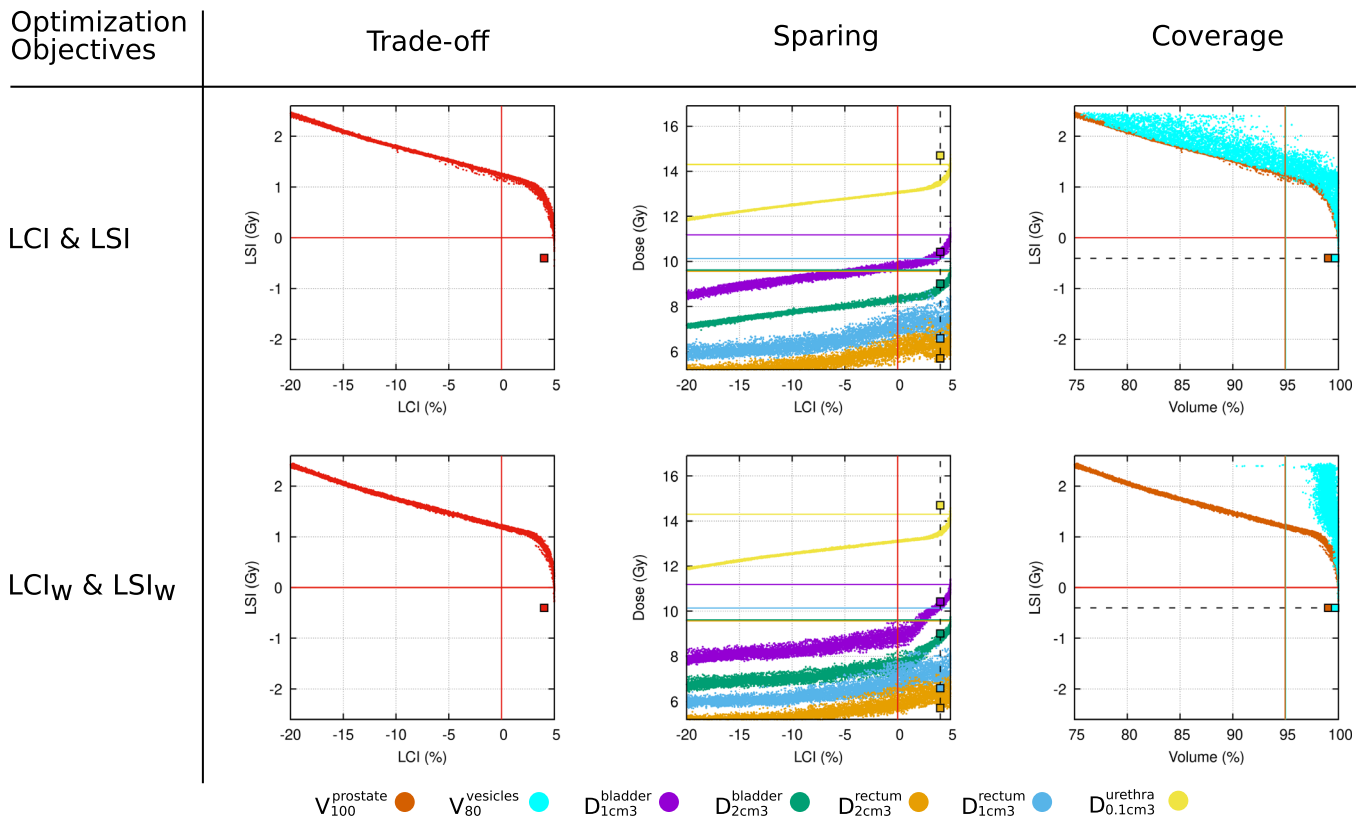


FIG. 2. Results for patient 2. The top row shows results found when the LCI and LSI are used as optimization objectives. The bottom row shows results found when the  $LCI_w$  and  $LSI_w$  are used as optimization objectives. Subfigures show all plans found by 30 independent optimization runs, to show variation, using a set of 100 000 DC points, and a time limit of 3 minutes. DV indices of the clinical plan are displayed as squares with identical colors (online version only) to the optimized treatment plans. In all subfigures, the (non-weighted) LCI and LSI are displayed, calculated using a set of 500 000 DC points.

Only 4 of 18 clinical plans satisfied all clinical criteria, whereas the optimization consistently (30 of 30 runs) found plans satisfying all clinical criteria for 15 of 18 patients. No plans satisfying the clinical protocol were found for the remaining three patients due to suboptimal catheter placement. The value of optimizing on a large number of DC points is mostly shown for patient 7, as all runs optimized on 100 000 DC points found a plan satisfying all clinical criteria, whereas this was not the case when optimizing 20 000 DC points.

A common characteristic of the clinical plans is that the urethra received a relatively large amount of dose, as the clinical criterion for the  $D_{0.1cm3}^{urethra}$  was satisfied for only 5 of 18 clinical plans. Compared to the clinical plans, some sparing of the rectum was sacrificed in many cases of the bi-objectively optimized plans. On average, the  $D_{1cm3}^{rectum}$  increased by 0.46 Gy, and the  $D_{2cm3}^{rectum}$  increased by 0.43 Gy. However, the  $D_{0.1cm3}^{urethra}$  was reduced by 0.68 Gy on average.

Figure 4 shows the combined results of 30 optimization runs for patient 1, compared to the clinical plan. All 30 runs are displayed in this figure to show the variance of the resulting treatment plans, caused by the stochastic nature of the optimization, and of random sampling of DC points. The variance of a DV index also depends on the difficulty of satisfying the corresponding criterion

defined in the clinical protocol, because DV indices that are more difficult to satisfy are assigned a larger weight in the  $LCI_w$  and  $LSI_w$ . Similar to Fig. 2, all DV indices of each treatment plan are displayed in a separate color according to the legend, and the clinical plan is displayed as a square. For each value of the LCI and LSI, Fig. 4 shows the corresponding values for all DV indices. Plots for all remaining patients are included in the supplementary material.

#### 4. DISCUSSION

In this article, we introduced a GPU-based acceleration of a previously introduced bi-objective evolutionary optimization approach for treatment planning for HDR prostate brachytherapy. This bi-objective treatment planning method directly optimizes DV indices, whereas conventional treatment planning methods, for example, IPSA<sup>3</sup> or HIPO,<sup>2</sup> use a penalty-based optimization model that weakly correlates with the quality of DV indices.<sup>5</sup>

In contrast to a more recently introduced multicriteria optimization approach<sup>28</sup> that requires pretrained regression models to guide the optimization toward clinically acceptable plans, our bi-objective evolutionary optimization approach requires no prior training,



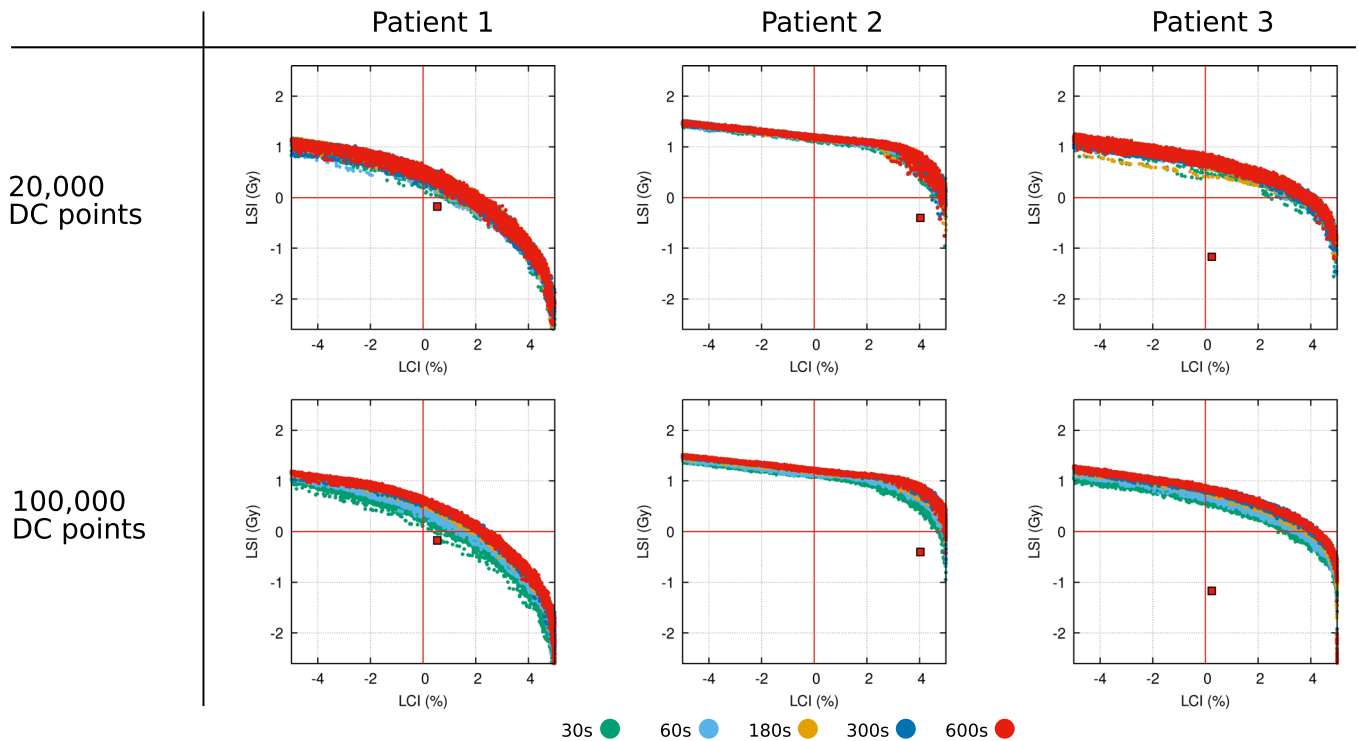


FIG. 3. Bi-objectively optimized plans for patients 1, 2, and 3, obtained after optimizing for the specified number of seconds, with 20 000 or 100 000 DC points. Results shown here are reevaluated on 500 000 DC points. To show the joint variation of the optimization and sampling of DC points, all treatment plans from 30 independent optimization runs are shown. The clinical plan is displayed as a red square.

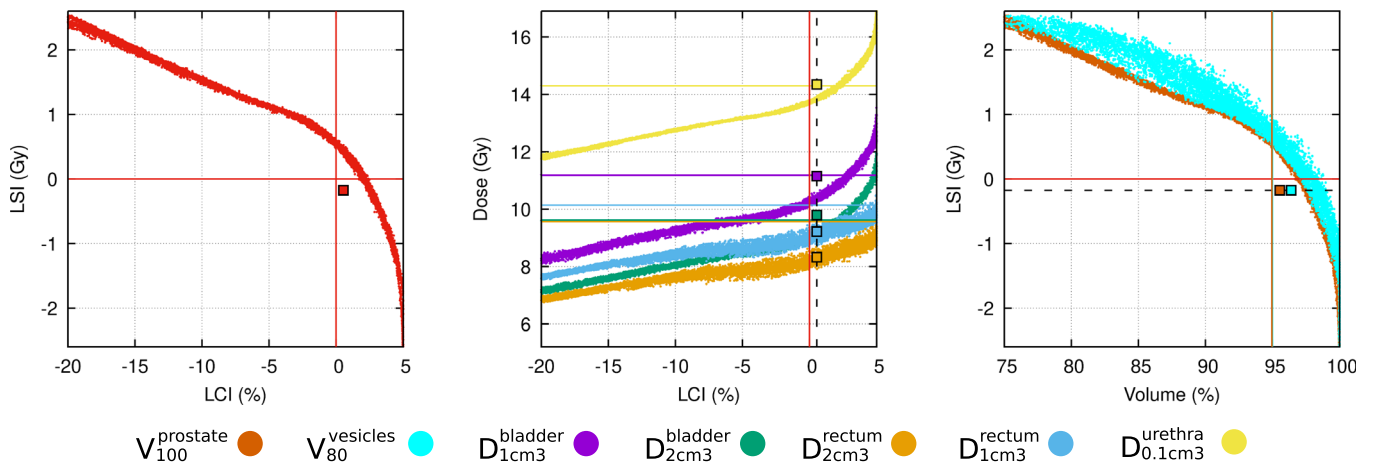


FIG. 4. Combined results of 30 optimization runs for patient 1, optimized for 3 minutes on 100 000 DC points with  $LCI_w$  and  $LSI_w$  as objectives, showing corresponding values of all DV indices for each value of the LCI and LSI.

because it directly calculates all relevant DV indices throughout optimization. As our bi-objective evolutionary optimization approach requires between 30 and 300 seconds, we have shown that it is indeed possible to approximate a surface of trade-off plans within a clinically acceptable time frame.

Increasing the accuracy of DV-index calculation by increasing the number of DC points further improves the DV indices achieved by the bi-objective optimization approach, with very good results obtainable in 3 minutes on 100 000 DC points. Even this more time-consuming approach is

arguably still within the time limit of clinical practice, because planning in our clinic currently takes 30–60 minutes, which is mainly spent on graphical optimization.<sup>11</sup> Increasing the number of DC points beyond 100 000 would further increase the required computation time for a negligible increase in plan quality.

From Section 3.A, we conclude that a sufficient number of DC points should be used to prevent the optimization from overfitting and thus to make sure the most benefit is achieved from using inverse planning. For sets of at least 100 000 DC points, the effect of overfitting was found to be negligible.

TABLE II. DV indices of the clinical plans, alongside means, and standard deviations (30 runs) of plans optimized on 20 000 (GOMEA20k) or 100 000 (GOMEA100k) DC points by the bi-objective optimization approach, for all 18 patients used in this study. Plans were selected from the trade-off curve to have maximum LCI while satisfying all sparing criteria. Volume indices and LCI values are shown as a percentage of ROI volume. Dose indices and LSI values are shown as a percentage of the PD of 13 Gy. Records shaded in gray do not satisfy the clinical protocol. Optimized results shaded in gray did not satisfy the clinical protocol in at least 1 of 30 runs. All underlined records are statistically significantly different from the corresponding index of the clinical plan (Wilcoxon signed-rank test,  $p < 0.05$ ; Bonferroni correction,  $n = 324$ )

| Patient | Plan      | $V_{100}^{\text{prostate}}$ | $V_{80}^{\text{vessicles}}$ | $D_{1cm3}^{\text{bladder}}$ | $D_{2cm3}^{\text{bladder}}$ | $D_{1cm3}^{\text{rectum}}$ | $D_{2cm3}^{\text{rectum}}$ | $D_{0.1cm3}^{\text{urethra}}$ | $V_{150}^{\text{prostate}}$ | $V_{200}^{\text{prostate}}$ | LCI        | LSI       |
|---------|-----------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|----------------------------|----------------------------|-------------------------------|-----------------------------|-----------------------------|------------|-----------|
| 1       | Clinical  | 95.5                        | 96.4                        | 85.8                        | 75.3                        | 70.9                       | 64.0                       | 110.4                         | 34.3                        | 13.4                        | 0.5        | -1.3      |
|         | GOMEA20k  | 96.4 ± 0.3                  | 97.7 ± 0.6                  | 83.3 ± 0.8                  | 73.6 ± 0.5                  | 70.3 ± 1.0                 | 64.1 ± 1.0                 | 108.9 ± 0.9                   | 27.8 ± 1.4                  | 11.2 ± 0.6                  | 1.4 ± 0.3  | 0.2 ± 0.2 |
|         | GOMEA100k | 97.0 ± 0.2                  | 97.7 ± 0.4                  | 83.5 ± 0.5                  | 73.7 ± 0.3                  | 71.8 ± 1.3                 | 65.5 ± 1.2                 | 109.7 ± 0.3                   | 28.5 ± 1.3                  | 11.5 ± 0.6                  | 2.0 ± 0.2  | 0.1 ± 0.1 |
| 2       | Clinical  | 99.0                        | 99.7                        | 80.2                        | 69.3                        | 50.6                       | 43.9                       | 113.1                         | 26.3                        | 9.9                         | 4.0        | -3.1      |
|         | GOMEA20k  | 99.8 ± 0.1                  | 100.0 ± 0.0                 | 85.0 ± 1.1                  | 72.6 ± 1.0                  | 55.7 ± 3.1                 | 48.3 ± 2.5                 | 108.8 ± 1.0                   | 19.0 ± 1.0                  | 7.2 ± 0.4                   | 4.8 ± 0.1  | 0.5 ± 0.5 |
|         | GOMEA100k | 99.9 ± 0.0                  | 100.0 ± 0.0                 | 85.3 ± 0.6                  | 72.8 ± 0.6                  | 56.1 ± 3.1                 | 48.7 ± 2.4                 | 109.4 ± 0.4                   | 19.3 ± 0.9                  | 7.4 ± 0.4                   | 4.9 ± 0.0  | 0.4 ± 0.4 |
| 3       | Clinical  | 98.3                        | 95.2                        | 92.5                        | 83.0                        | 78.6                       | 69.3                       | 109.0                         | 28.1                        | 9.8                         | 0.2        | -9.0      |
|         | GOMEA20k  | 98.5 ± 0.2                  | 98.9 ± 0.3                  | 81.4 ± 0.6                  | 73.7 ± 0.4                  | 76.5 ± 1.2                 | 67.4 ± 1.1                 | 108.7 ± 0.9                   | 22.8 ± 1.7                  | 8.5 ± 0.5                   | 3.5 ± 0.2  | 0.1 ± 0.1 |
|         | GOMEA100k | 99.0 ± 0.1                  | 99.1 ± 0.1                  | 81.5 ± 0.3                  | 73.9 ± 0.1                  | 77.1 ± 0.6                 | 67.7 ± 0.6                 | 109.5 ± 0.3                   | 24.0 ± 1.4                  | 9.1 ± 0.3                   | 4.0 ± 0.1  | 0.1 ± 0.1 |
| 4       | Clinical  | 99.3                        | 86.2                        | 89.2                        | 80.9                        | 67.8                       | 59.8                       | 114.4                         | 37.4                        | 17.6                        | -8.8       | -6.9      |
|         | GOMEA20k  | 93.1 ± 0.5                  | 93.3 ± 0.4                  | 82.0 ± 0.5                  | 73.8 ± 0.2                  | 72.6 ± 2.0                 | 63.2 ± 1.7                 | 109.0 ± 0.8                   | 25.6 ± 1.3                  | 10.9 ± 0.7                  | -2.0 ± 0.4 | 0.2 ± 0.2 |
|         | GOMEA100k | 93.7 ± 0.2                  | 93.7 ± 0.2                  | 81.9 ± 0.4                  | 73.9 ± 0.1                  | 72.9 ± 1.8                 | 63.4 ± 1.3                 | 109.4 ± 0.3                   | 26.3 ± 0.7                  | 11.4 ± 0.5                  | -1.4 ± 0.2 | 0.1 ± 0.1 |
| 5       | Clinical  | 96.1                        | 88.6                        | 86.3                        | 74.8                        | 70.8                       | 61.3                       | 110.4                         | 36.1                        | 16.0                        | -6.4       | -0.8      |
|         | GOMEA20k  | 97.0 ± 0.5                  | 98.6 ± 0.5                  | 83.0 ± 1.3                  | 73.3 ± 0.8                  | 75.4 ± 1.8                 | 66.8 ± 1.8                 | 108.8 ± 1.2                   | 28.1 ± 2.2                  | 10.4 ± 0.9                  | 2.0 ± 0.5  | 0.1 ± 0.1 |
|         | GOMEA100k | 97.8 ± 0.2                  | 98.7 ± 0.4                  | 83.5 ± 0.5                  | 73.7 ± 0.3                  | 76.3 ± 1.1                 | 67.7 ± 1.1                 | 109.6 ± 0.5                   | 27.8 ± 2.0                  | 10.2 ± 0.8                  | 2.8 ± 0.2  | 0.1 ± 0.1 |
| 6       | Clinical  | 97.8                        | 96.2                        | 80.0                        | 68.3                        | 80.3                       | 70.8                       | 121.1                         | 35.4                        | 14.2                        | 1.2        | -11.1     |
|         | GOMEA20k  | 96.0 ± 0.4                  | 100.0 ± 0.1                 | 82.4 ± 1.8                  | 71.7 ± 1.5                  | 77.4 ± 0.8                 | 68.4 ± 0.6                 | 109.4 ± 0.6                   | 27.5 ± 1.1                  | 10.3 ± 0.4                  | 1.0 ± 0.4  | 0.2 ± 0.2 |
|         | GOMEA100k | 96.7 ± 0.1                  | 99.8 ± 0.2                  | 83.7 ± 0.8                  | 72.7 ± 0.7                  | 77.7 ± 0.3                 | 68.7 ± 0.4                 | 109.6 ± 0.3                   | 27.2 ± 1.0                  | 10.3 ± 0.4                  | 1.7 ± 0.1  | 0.2 ± 0.2 |
| 7       | Clinical  | 97.5                        | 78.2                        | 84.2                        | 76.9                        | 64.7                       | 57.0                       | 113.9                         | 31.4                        | 14.7                        | -16.8      | -3.9      |
|         | GOMEA20k  | 95.2 ± 0.3                  | 95.5 ± 0.4                  | 81.1 ± 0.4                  | 73.8 ± 0.2                  | 70.5 ± 2.3                 | 62.5 ± 1.8                 | 108.5 ± 0.9                   | 26.3 ± 1.1                  | 10.9 ± 0.6                  | 0.1 ± 0.3  | 0.2 ± 0.1 |
|         | GOMEA100k | 95.9 ± 0.2                  | 96.1 ± 0.2                  | 81.0 ± 0.4                  | 73.9 ± 0.1                  | 70.5 ± 1.9                 | 62.5 ± 1.6                 | 109.3 ± 0.5                   | 25.8 ± 1.0                  | 10.7 ± 0.6                  | 0.9 ± 0.2  | 0.1 ± 0.1 |
| 8       | Clinical  | 98.3                        | 98.8                        | 75.7                        | 67.8                        | 81.1                       | 73.9                       | 127.1                         | 28.9                        | 11.3                        | 3.3        | -17.1     |
|         | GOMEA20k  | 96.3 ± 0.4                  | 99.0 ± 0.6                  | 81.1 ± 2.0                  | 72.4 ± 1.3                  | 77.5 ± 0.7                 | 70.5 ± 0.7                 | 109.2 ± 0.8                   | 17.9 ± 0.8                  | 6.7 ± 0.3                   | 1.3 ± 0.4  | 0.2 ± 0.3 |
|         | GOMEA100k | 97.0 ± 0.2                  | 98.9 ± 0.4                  | 82.2 ± 0.9                  | 73.3 ± 0.6                  | 77.7 ± 0.3                 | 70.5 ± 0.3                 | 109.8 ± 0.3                   | 17.9 ± 0.7                  | 6.8 ± 0.3                   | 2.0 ± 0.2  | 0.1 ± 0.1 |
| 9       | Clinical  | 97.8                        | 98.0                        | 80.5                        | 73.8                        | 53.1                       | 47.5                       | 113.9                         | 28.2                        | 12.3                        | 2.8        | -3.9      |
|         | GOMEA20k  | 97.7 ± 0.2                  | 99.3 ± 0.4                  | 80.2 ± 0.5                  | 73.8 ± 0.3                  | 58.4 ± 1.9                 | 52.4 ± 1.6                 | 108.2 ± 1.1                   | 22.7 ± 1.7                  | 9.0 ± 0.8                   | 2.7 ± 0.2  | 0.1 ± 0.2 |
|         | GOMEA100k | 98.3 ± 0.1                  | 99.5 ± 0.2                  | 80.0 ± 0.3                  | 73.9 ± 0.1                  | 58.8 ± 1.7                 | 52.5 ± 1.3                 | 109.3 ± 0.5                   | 22.8 ± 1.1                  | 9.0 ± 0.4                   | 3.3 ± 0.1  | 0.1 ± 0.1 |
| 10      | Clinical  | 96.7                        | 95.5                        | 76.5                        | 65.9                        | 67.1                       | 56.0                       | 109.6                         | 28.5                        | 10.5                        | 0.5        | 0.4       |
|         | GOMEA20k  | 98.8 ± 0.1                  | 99.9 ± 0.1                  | 84.2 ± 1.0                  | 73.0 ± 0.8                  | 73.4 ± 2.2                 | 61.3 ± 1.8                 | 109.7 ± 0.3                   | 21.5 ± 0.9                  | 7.7 ± 0.3                   | 3.8 ± 0.1  | 0.2 ± 0.1 |
|         | GOMEA100k | 99.2 ± 0.1                  | 99.9 ± 0.1                  | 84.8 ± 0.5                  | 73.5 ± 0.4                  | 74.8 ± 1.9                 | 62.3 ± 1.5                 | 109.9 ± 0.1                   | 21.4 ± 0.8                  | 7.7 ± 0.3                   | 4.2 ± 0.1  | 0.1 ± 0.0 |
| 11      | Clinical  | 98.3                        | 100.0                       | 82.8                        | 72.5                        | 64.5                       | 56.9                       | 109.7                         | 34.8                        | 15.1                        | 3.3        | 0.3       |
|         | GOMEA20k  | 99.2 ± 0.1                  | 100.0 ± 0.1                 | 82.7 ± 1.4                  | 71.9 ± 1.2                  | 64.7 ± 1.6                 | 56.7 ± 1.2                 | 109.6 ± 0.4                   | 30.6 ± 0.8                  | 13.4 ± 0.5                  | 4.2 ± 0.1  | 0.3 ± 0.4 |
|         | GOMEA100k | 99.5 ± 0.1                  | 100.0 ± 0.1                 | 84.4 ± 0.9                  | 73.2 ± 0.6                  | 65.6 ± 1.3                 | 57.6 ± 1.1                 | 109.9 ± 0.2                   | 30.5 ± 0.9                  | 13.4 ± 0.5                  | 4.5 ± 0.1  | 0.1 ± 0.1 |

TABLE II. Continued.

| Patient | Plan      | $V_{100}^{\text{prostate}}$ | $V_{80}^{\text{vesicles}}$ | $D_{1cm^3}^{\text{bladder}}$ | $D_{2cm^3}^{\text{bladder}}$ | $D_{1cm^3}^{\text{rectum}}$ | $D_{2cm^3}^{\text{rectum}}$ | $D_{0.1cm^3}^{\text{urethra}}$ | $V_{150}^{\text{prostate}}$ | $V_{200}^{\text{prostate}}$ | LCI        | LSI       |
|---------|-----------|-----------------------------|----------------------------|------------------------------|------------------------------|-----------------------------|-----------------------------|--------------------------------|-----------------------------|-----------------------------|------------|-----------|
| 12      | Clinical  | 98.8                        | 95.0                       | 76.0                         | 66.8                         | 63.0                        | 56.9                        | 116.8                          | 28.6                        | 11.2                        | 0.0        | -6.8      |
|         | GOMEA20k  | 98.7 ± 0.1                  | 99.8 ± 0.1                 | 82.5 ± 1.0                   | 72.8 ± 0.9                   | 69.3 ± 1.4                  | 62.3 ± 1.2                  | 109.6 ± 0.4                    | 23.9 ± 1.5                  | 9.0 ± 0.7                   | 3.7 ± 0.1  | 0.3 ± 0.3 |
|         | GOMEA100k | 98.9 ± 0.1                  | 99.8 ± 0.1                 | 83.1 ± 0.7                   | 73.2 ± 0.6                   | 68.9 ± 1.5                  | 62.0 ± 1.3                  | 109.9 ± 0.1                    | 23.2 ± 1.0                  | 8.9 ± 0.4                   | 3.9 ± 0.1  | 0.1 ± 0.1 |
| 13      | Clinical  | 97.5                        | 95.4                       | 81.5                         | 73.4                         | 62.7                        | 56.0                        | 111.5                          | 20.1                        | 8.5                         | 0.4        | -1.5      |
|         | GOMEA20k  | 97.2 ± 0.4                  | 97.8 ± 0.5                 | 81.3 ± 0.9                   | 73.5 ± 0.6                   | 69.0 ± 2.2                  | 61.9 ± 1.9                  | 108.8 ± 1.2                    | 17.8 ± 1.1                  | 6.5 ± 0.3                   | 2.2 ± 0.4  | 0.2 ± 0.2 |
|         | GOMEA100k | 97.9 ± 0.2                  | 98.0 ± 0.2                 | 81.6 ± 0.4                   | 73.9 ± 0.1                   | 70.9 ± 1.6                  | 63.6 ± 1.4                  | 109.4 ± 0.4                    | 17.8 ± 0.6                  | 6.5 ± 0.2                   | 2.9 ± 0.2  | 0.1 ± 0.1 |
| 14      | Clinical  | 93.4                        | 85.9                       | 65.9                         | 59.3                         | 70.0                        | 61.5                        | 122.7                          | 30.1                        | 11.8                        | -9.1       | -12.7     |
|         | GOMEA20k  | 94.4 ± 0.3                  | 99.9 ± 0.2                 | 80.6 ± 1.6                   | 72.8 ± 1.1                   | 73.9 ± 1.7                  | 66.3 ± 1.3                  | 109.5 ± 0.6                    | 27.1 ± 1.1                  | 11.1 ± 0.5                  | -0.6 ± 0.3 | 0.2 ± 0.2 |
|         | GOMEA100k | 94.9 ± 0.1                  | 99.9 ± 0.1                 | 81.2 ± 0.8                   | 73.3 ± 0.6                   | 74.4 ± 1.4                  | 66.5 ± 1.1                  | 109.9 ± 0.1                    | 26.6 ± 1.2                  | 10.9 ± 0.4                  | -0.1 ± 0.1 | 0.1 ± 0.1 |
| 15      | Clinical  | 95.7                        | 99.9                       | 84.7                         | 76.2                         | 76.2                        | 65.8                        | 111.9                          | 29.0                        | 12.8                        | 0.7        | -2.2      |
|         | GOMEA20k  | 98.0 ± 0.5                  | 99.8 ± 0.1                 | 82.2 ± 0.9                   | 73.5 ± 0.6                   | 75.6 ± 1.4                  | 66.5 ± 1.3                  | 108.7 ± 1.4                    | 25.0 ± 1.2                  | 10.0 ± 0.7                  | 3.0 ± 0.5  | 0.2 ± 0.2 |
|         | GOMEA100k | 98.7 ± 0.2                  | 99.8 ± 0.2                 | 82.5 ± 0.5                   | 73.6 ± 0.4                   | 76.3 ± 0.9                  | 67.2 ± 0.9                  | 109.7 ± 0.5                    | 25.2 ± 1.2                  | 10.3 ± 0.5                  | 3.7 ± 0.2  | 0.1 ± 0.1 |
| 16      | Clinical  | 96.2                        | 98.1                       | 75.2                         | 65.9                         | 53.6                        | 46.9                        | 107.1                          | 28.1                        | 9.4                         | 1.2        | 2.9       |
|         | GOMEA20k  | 99.9 ± 0.1                  | 100.0 ± 0.0                | 81.8 ± 1.5                   | 71.2 ± 1.2                   | 60.6 ± 1.6                  | 53.1 ± 1.4                  | 107.4 ± 1.0                    | 21.6 ± 1.8                  | 6.6 ± 0.5                   | 4.9 ± 0.1  | 2.2 ± 0.9 |
|         | GOMEA100k | 100.0 ± 0.0                 | 100.0 ± 0.0                | 82.0 ± 1.1                   | 71.4 ± 0.9                   | 63.0 ± 1.8                  | 55.1 ± 1.5                  | 107.5 ± 0.9                    | 21.3 ± 1.2                  | 6.8 ± 0.6                   | 5.0 ± 0.0  | 2.3 ± 0.9 |
| 17      | Clinical  | 95.3                        | 95.5                       | 72.2                         | 65.5                         | 67.9                        | 61.1                        | 109.3                          | 25.2                        | 8.6                         | 0.3        | 0.7       |
|         | GOMEA20k  | 99.6 ± 0.1                  | 99.8 ± 0.1                 | 81.8 ± 0.7                   | 73.7 ± 0.3                   | 75.1 ± 1.7                  | 68.1 ± 1.6                  | 108.8 ± 0.8                    | 21.2 ± 1.6                  | 6.5 ± 0.8                   | 4.5 ± 0.1  | 0.2 ± 0.2 |
|         | GOMEA100k | 99.8 ± 0.1                  | 99.9 ± 0.1                 | 82.1 ± 0.6                   | 73.8 ± 0.2                   | 75.9 ± 1.2                  | 68.9 ± 1.1                  | 109.7 ± 0.3                    | 21.1 ± 1.6                  | 6.5 ± 0.6                   | 4.8 ± 0.1  | 0.1 ± 0.1 |
| 18      | Clinical  | 96.8                        | 98.5                       | 100.7                        | 92.5                         | 69.4                        | 62.6                        | 127.8                          | 38.6                        | 15.5                        | 1.8        | -18.5     |
|         | GOMEA20k  | 93.7 ± 0.3                  | 95.7 ± 0.8                 | 78.8 ± 0.5                   | 73.7 ± 0.3                   | 73.8 ± 1.9                  | 65.8 ± 1.6                  | 109.2 ± 0.7                    | 26.4 ± 1.7                  | 8.9 ± 0.6                   | -1.3 ± 0.3 | 0.2 ± 0.1 |
|         | GOMEA100k | 94.4 ± 0.2                  | 95.5 ± 0.5                 | 78.8 ± 0.2                   | 73.9 ± 0.1                   | 73.7 ± 1.7                  | 65.8 ± 1.4                  | 109.6 ± 0.3                    | 26.0 ± 1.2                  | 8.8 ± 0.5                   | -0.6 ± 0.2 | 0.1 ± 0.1 |

These results were found using points that were sampled uniformly at random in each ROI, because random sampling was previously found to be a better estimator of DV indices than sampling on a regular grid.<sup>29</sup> Other sampling techniques have previously been used,<sup>6,3</sup> which could prove to require a smaller number of DC points for the accurate calculation of certain DV indices.

In this article, an equal number of DC points was sampled in each ROI, regardless of ROI volume. Preliminary experiments showed that distributing the DC points according to ROI volume resulted in plans with visibly worse LCI and LSI values. This is caused by the fact that  $LCI_w$  and  $LSI_w$  values mostly depend on the most violated DV index, and hence the number of DC points in the corresponding ROI. It is therefore valuable to have a relatively high number of DC points in the corresponding ROI of the most violated DV index. However, the ROI with the most violated DV index changes dynamically throughout optimization, and depends on many factors, including patient anatomy and clinical aspiration values. Further research is required to find a better distribution of DC points among ROIs, while avoiding a loss of obtained plan quality. The efficiency of the calculation of the values of the DV indices could also be improved by only using DC points in the rectum and bladder that are closest to the dwell positions, as these points are likely to receive the highest dose.<sup>6</sup>

As the use of multi-objective EAs results in a set of trade-off solutions, this inevitably leads to a decision-making process where a clinician must still decide which of the provided treatment plans is most suitable to actually use. Plots as in Fig. 4 can prove to be an aid for this decision-making process, as these plots help to visualize the trade-offs between all criteria of the clinical protocol, which should be complemented by the inspection of the dose distribution and dose-volume histograms for the evaluation of single plans. The plot showing only LCI and LSI can only give a rough indication of the trade-off between coverage and sparing, while the remaining plots tell the exact values of each DV index for a given LCI or LSI. Moreover, interactive software could further increase the intuitiveness of these plots by highlighting a selected treatment plan in all three plots. Note that the plots displayed in Fig. 4 display the combined results of 30 runs, for the sake of displaying the variation in the results of the optimization and the sampling of DC points, whereas a decision-maker would only see the results of one optimization run. Each run however covers the entire range of possible trade-offs shown in Fig. 4.

Based on previous work<sup>8</sup> and the results in Section 3, we conclude that MO-RV-GOMEA is an algorithm that is excellently suited for direct DV-index based bi-objective treatment plan optimization. The best results were found when the optimization objectives were the  $LCI_w$  and the  $LSI_w$ , and a set of 100 000 DC points, that is, 20 000 DC points per ROI, was used. For this configuration, up to 3 minutes were required for the optimization, as no substantial improvements were found for time limits up to 10 minutes.

Topics of future research include generalization to optimization on different clinical protocols, for example, the

American Brachytherapy Society consensus guidelines,<sup>30</sup> as the LCI and LSI could be constructed based on any clinical protocol consisting of a set of DV criteria. This also includes the addition of the  $V_{150}^{\text{prostate}}$  and  $V_{200}^{\text{prostate}}$  to the LSI, because these indices are currently included as hard constraints.

## 5. CONCLUSIONS

In general, bi-objective treatment planning provides a clinician with intuitive insight into the treatment planning process by being able to inspect a wide range of potential treatment plans having different trade-offs between coverage of the target organs and sparing of the OARs. We have developed a GPU-based parallel version of a previously introduced bi-objective treatment planning method<sup>8</sup> that optimizes for the worst target and sparing criterion. Furthermore, the optimization model has been adjusted to provide better results, particularly improving values of DV indices that are not the worst criterion. This bi-objective treatment planning method is now capable of producing a set of high-quality solutions, each with a different trade-off between the coverage of the target organs and sparing of the OARs, optimized for 100 000 DC points, in approximately 3 minutes, making it efficient enough to be used in a time-constrained clinical setting.

## ACKNOWLEDGMENTS

This work is part of the research programme IPPSI-TA with project number 628.006.003, which is financed by the Netherlands Organisation for Scientific Research (NWO) and Elekta. We gratefully acknowledge the support of NVIDIA Corporation with the donation of a Titan Xp GPU used for this research.

## CONFLICTS OF INTEREST

Anton Bouter, Tanja Alderliesten, Arjan Bel, Bradley R. Pieters, and Peter A.N. Bosman are involved in projects supported by Elekta.

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: anton.bouter@cw.nl.

## REFERENCES

1. Meigooni AS, Hoskin P, Baltas D, Venselaar JL. Introduction and innovations in brachytherapy. In: *Comprehensive Brachytherapy*. Boca Raton: CRC Press; 2012:30–35.
2. Karabis A, Giannouli S, Baltas D. HIPO: a hybrid inverse treatment planning optimization algorithm in HDR brachytherapy. *Radiother Oncol*. 2005;76:S29.
3. Lessard E, Pouliot J. Inverse planning anatomy-based dose optimization for HDR-brachytherapy of the prostate using fast simulated annealing algorithm and dedicated objective function. *Med Phys*. 2001;28:773–779.
4. Alterovitz R, Lessard E, Pouliot J, et al. Optimization of HDR brachytherapy dose distributions using linear programming with penalty costs. *Med Phys*. 2006;33:4012–4019.

5. Gorissen BL, Den Hertog D, Hoffmann AL. Mixed integer programming improves comprehensibility and plan quality in inverse optimization of prostate HDR brachytherapy. *Phys Med Biol.* 2013;58:1041.
6. Siau T, Cunha A, Atamtürk A, et al. IPIP: a new approach to inverse planning for HDR brachytherapy by directly optimizing dosimetric indices. *Med Phys.* 2011;38:4045–4051.
7. Deist TM, Gorissen BL. High-dose-rate prostate brachytherapy inverse planning on dose-volume criteria by simulated annealing. *Phys Med Biol.* 2016;61:1155.
8. Luong NH, Alderliesten T, Bel A, Niatsetski Y, Bosman PAN. Application and benchmarking of multi-objective evolutionary algorithms on high-dose-rate brachytherapy planning for prostate cancer treatment. *Swarm Evol Comput.* 2018;40:37–52.
9. Deb K. *Multi-Objective Optimization Using Evolutionary Algorithms*. Hoboken, NJ: John Wiley & Sons, Inc; 2001. <https://dl.acm.org/citation.cfm?xml:id=559152>
10. Bouter A, Luong NH, Witteveen C, Alderliesten T, Bosman PAN. The multi-objective real-valued gene-pool optimal mixing evolutionary algorithm. In Proceedings of the Genetic and Evolutionary Computation Conference. GECCO '17. New York, NY, USA: ACM; 2017:537–544.
11. Maree SC, Luong NH, Kooreman ES, et al. Evaluation of bi-objective treatment planning for high-dose-rate prostate brachytherapy—A Retrospective Observer Study. *Brachytherapy.* 2019;18:396–403.
12. Kirisits C, Siebert FA, Baltas D, et al. Accuracy of volume and DVH parameters determined with different brachytherapy treatment planning systems. *Radiother Oncol.* 2007;84:290–297.
13. Luong NH, Alderliesten T, Pieters BR, et al. Fast and insightful bi-objective HDR prostate brachytherapy planning. *Radiother Oncol.* 2018;127:S130.
14. Holland J. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press; 1975.
15. Pratz G, Xing L. GPU computing in medical physics: a review. *Med Phys.* 2011;38:2685–2697.
16. Hissoiny S, Ozell B, Bouchard H, Després P. GPUMCD: a new GPU-oriented Monte Carlo dose calculation platform. *Med Phys.* 2011;38:754–764.
17. Tian Z, Li Y, Hassan-Rezaeian N, Jiang SB, Jia X. Moving GPU-OpenCL-based Monte Carlo dose calculation toward clinical use: automatic beam commissioning and source sampling for treatment plan dose calculation. *J Appl Clin Med Phys.* 2017;18:69–84.
18. Ruciński A, Gajewski J, Olko P, et al. GPU-accelerated Monte Carlo code for fast dose recalculation in proton beam therapy. *Acta Phys Pol, B.* 2016;48:1625.
19. De Greef M, Crezee J, Van Eijk J, Pool R, Bel A. Accelerated ray tracing for radiotherapy dose calculations on a GPU. *Med Phys.* 2009;36:4095–4102.
20. Wong ML. Parallel multi-objective evolutionary algorithms on graphics processing units. In *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers*. GECCO '09. New York, NY, USA: ACM; 2009:2515–2522.
21. D'Agostino D, Pasquale G, Merelli I. A fine-grained CUDA implementation of the multi-objective evolutionary approach NSGA-II: potential impact for computational and systems biology applications. In: *International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics*. Berlin: Springer; 2014:273–284.
22. Ortega G, Filatovas E, Garzón EM, Casado LG. Non-dominated sorting procedure for Pareto dominance ranking on multicore CPU and/or GPU. *J Global Optim.* 2017;69:607–627.
23. Li SC, Yu TL. Speeding up DSMGA-II on CUDA platform. In *Proceedings of the Genetic and Evolutionary Computation Conference*. GECCO '17. New York, NY, USA: ACM; 2017:809–816.
24. NVIDIA Corporation. *CUDA C programming guide v10.0.130*; 2018.
25. Bouter A, Alderliesten T, Bel A, Witteveen C, Bosman PAN. Large-scale parallelization of partial evaluations in evolutionary algorithms for real-world problems. In *Proceedings of the Genetic and Evolutionary Computation Conference*. GECCO '18. New York, NY, USA: ACM; 2018:1199–1206.
26. Rivard MJ, Coursey BM, DeWerd LA, et al. Update of AAPM task group no. 43 report: a revised AAPM protocol for brachytherapy dose calculations. *Med Phys.* 2004;31:633–674.
27. Bell N, Hoberock J. Thrust: a productivity-oriented library for CUDA. In *GPU computing gems Jade edition*. Vol. 2. Amsterdam: Elsevier; 2011:359–371.
28. Cui S, Després P, Beaulieu L. A multi-criteria optimization approach for HDR prostate brachytherapy: I. Pareto surface approximation. *Phys Med Biol.* 2018;63:205004.
29. Niemierko A, Goitein M. Random sampling for evaluating treatment plans. *Med Phys.* 1990;17:753–762.
30. Yamada Y, Rogers L, Demanes DJ, et al. American brachytherapy society consensus guidelines for high-dose-rate prostate brachytherapy. *Brachytherapy.* 2012;11:20–32.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Data S1:** The Supporting Information consists of more extensive results for the complete set of 18 patients, as well as results for different time limits and different numbers of DC points.