

Hanne Kooy

8.1 A Brief History of Protons at the Harvard Cyclotron Laboratory

The first proton radiotherapy patient was treated in 1957 at the Berkeley Radiation Laboratory. At the Harvard Cyclotron Laboratory, treatments commenced shortly after in the early 1960s under the direction of the Massachusetts General Hospital neurosurgeon Dr. Raymond Kjellberg. Neurosurgeons were well equipped to use the precision of proton beams without the availability of 3D imaging technologies such as CT. Their appreciation of the 3D cranial anatomy projected on X-rays sufficed to treat neoplasms such as pituitary abnormalities and arterial venous malformations. Both Dr. Kjellberg in Boston and Dr. Leksell in Stockholm pioneered the use of protons in the cranial anatomy. Dr. Kjellberg's program, however, had ready access to the proton beam at the HCL (Fig. 8.1). Dr. Leksell's program did not have ready access which led to the invention of the Leksell Gamma Knife as an alternative therapeutic system for stereotactic radiosurgery. Protons were thus the first modality used in cranial stereotactic radiosurgery, while

the Gamma Knife made cranial stereotactic radiosurgery a standard modality.

A second program for eye treatments commenced under the direction of Dr. Evangelos Gragoudas from the Massachusetts Eye and Ear Infirmary in Boston. Again, the static and visually apparent anatomy of the eye and neoplasm afforded effective use of localized proton radiation.

Both programs were effective without the use of volumetric imaging or dose calculations; both were sufficiently served by manual calculation processes.

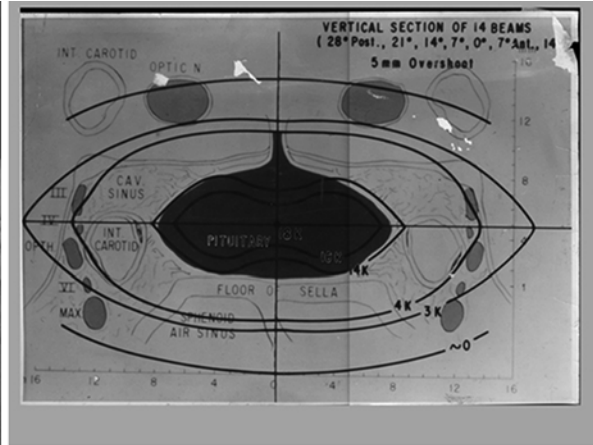
A third program in large-field "conventional" radiotherapy commenced in 1974 under the aegis of Dr. Herman Suit from the department of radiation oncology at Massachusetts General Hospital. The introduction of this program did require the use of volumetric data sets. Dr. Michael Goitein was one of the first to combine volumetric image data provided by the now available CT scanners and computational algorithms in a treatment planning system, dubbed Rx, for proton radiotherapy. Rx presented the physician and physicist with the necessary information and confidence to treat internal neoplasms reconstructed from the volumetric data.

All three programs continued at the HCL until 2001 and were transferred, uninterrupted, to the Northeast Proton Therapy Center (now the F.H. Burr Proton Therapy Center) on the campus of the Massachusetts General Hospital, the

H. Kooy, PhD
Department of Radiation Oncology, Massachusetts
General Hospital, Harvard Medical School,
Boston, MA, USA
e-mail: hanne.kooy@gmail.com



Fig. 8.1 Dr Raymond Kjellberg (and assistant) manipulating the stereotactic frame for a proton radiosurgery patient (*left*). An example dose calculation (not for this patient) shows the dose hand calculation for a pituitary



lesion using 14 proton beams that penetrate beyond the target volume, but their composite dose creates a focal region with sharp penumbra

second (the first was Loma Linda CA) proton facility within a hospital.

8.2 Implication for Modern Radiotherapy

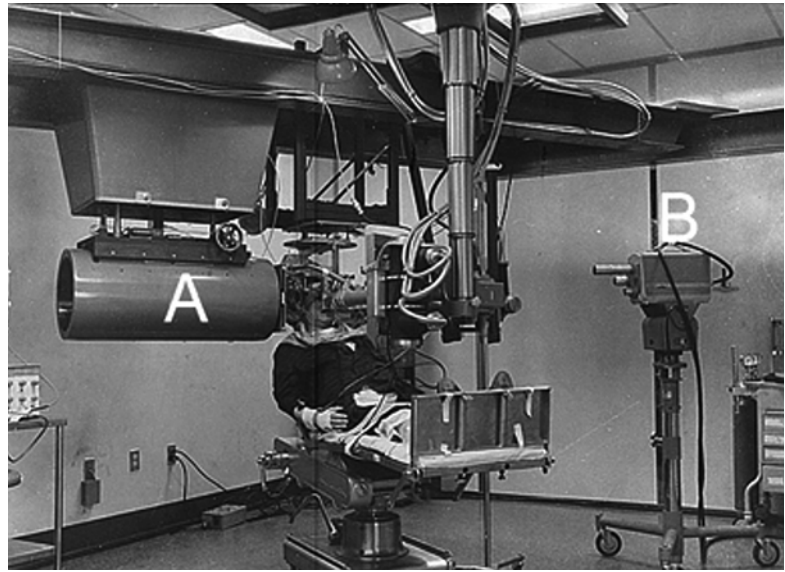
The precision of proton beams requires concomitant precision in treatment planning capabilities, in treatment delivery, and in patient positioning. Thus, proton radiotherapy in the 1980s already required and implemented these now assumed obvious requirements for precision radiotherapy. The limitation in proton energy up to 160 MeV at the HCL necessitated a focus on precisely those neoplasms that proved particularly suited for early application of proton radiotherapy: those in the head and neck, in the cranium, and soft tissue sarcomas. It is indisputable that proton radiotherapy, as demonstrated in those sites and especially for chordomas, proved the axiom of radiotherapy: increasing dose while sparing normal tissue increases cure. It should be noted that the HCL did not, per se, demonstrate the superiority of protons. It primarily demonstrated that when precision in dose delivery is achieved, outcome can

be improved. It secondarily demonstrated that protons well achieve such precision.

It was the well-understood physics of proton interactions in matter, i.e., scatter and energy loss, that allowed the precise manipulation of scattered and modulated (in energy and intensity) protons by mechanical means to create spread-out Bragg peak, SOBP, fields of variable range, and modulation. The use of apertures and range compensators, as modeled in the Rx treatment planning system, provided highly conformal 3D SOBP dose fields where the aperture served to provide lateral conformance, as in photon fields, and the range compensator served to provide distal target conformance to “stop” the proton field beyond the distal surface of the target.

Proton radiotherapy at HCL in the early 1980s was in sharp contrast to the parallel practice of radiotherapy. The first relied on now assumed obvious practices in precision therapy (Fig. 8.2), while the latter relied on X-ray simulation and simple 2D calculation methods. It would take two decades until photon radiotherapy would “catch up” and achieve the clinical performance of 1980 proton radiotherapy. Our question now is

Fig. 8.2 Patient setup and verification in the stereotactic treatment room at the Harvard Cyclotron Laboratory. A water-filled variable range shifter (A) permitted variable penetration into the cranium, while X-ray (B) permitted setup verification. The treatment technology thus implemented now assumed standards of precision



whether photon radiotherapy can maintain parity to proton radiotherapy to further the modern aims of radiotherapy.

8.3 Aims of Modern Radiotherapy and Proton Radiotherapy

Modern radiotherapy aims to optimize the response to radiation by minimizing dose to non-involved tissue to decrease normal tissue complications and to increase in target dose. These aims require improved differential imaging and identification of inter-target disease, improved localization (even in real time) with on-treatment imaging and target dose modulation, improved patient-specific response, and, now most importantly, improved quality of life for the patient.

These requirements require the application of in vivo and biological imaging capabilities, more advanced treatment planning and delivery systems, and, significantly, better understanding and utilization of differential biological response in the target and healthy tissues. Only some of these are furthered by external beam radiotherapy in the treatment management of the patient. All, however, need to synchronize over the course of treatment. It is this synchronization that is not promoted by the current treatment planning

architectures. Instead, treatment planning for modern radiotherapy must be deployed in an architecture that permits its functions to be distributed and accessed in their appropriate operational location. Current treatment planning systems are inconsistent with this requirement.

Radiotherapy operates in a safe zone of dose fractionation imposed by dose toxicity of the, at least originally, large volumes of irradiated healthy tissues. Modern conformal dose treatments reduce the volume of irradiated healthy tissues and offer the opportunity to increase fraction doses and modulate dose within a target. If such opportunity benefits the patient, proton and ion radiotherapy outperform photon radiotherapy.

It is recognized that physical, chemical, and genetic regional differences can exist within a single target. It is assumed that in such regions, differential dose delivery in terms of local ionization differentials (i.e., different lineal energy transfer, LET, distributions) can improve response through enhanced local biological response. Ion radiotherapy has a twofold advantage compared to photon radiotherapy. The application of different doses to different regions within a target will benefit from that modality that can achieve the sharpest regional dose gradients. The use of modulated LET

distributions can only be achieved with ion radiotherapy. Thus, proton and ion radiotherapy outperform photon radiotherapy where spatial and biological differentials are expected to improve outcome.

Imaging requirements prior and during treatment are core to the aims of modern radiotherapy. The original workflow model assumed that the treatment planning process preceded and remained static over a set of treatment delivery sessions. Treatment planning occurred as a stand-alone activity whose result was the treatment plan and the patient setup reference geometry to be applied and referenced daily. The daily effort to minimize inaccuracies and achieve presumed compliance was to reposition the patient to this static representation. Thus, the mitigation of all inaccuracies, and assumed worst case, had to be incorporated in the treatment plan to achieve acceptable dose coverage (and avoidance) over the time course of treatment. This process led to the definition of the planning target volume (PTV) as a geometric expansion of the target. This PTV expansion sufficed for photon treatments as the photon dose distribution in patient is invariant with respect to the patient's anatomy and geometric setup which equates dosimetric accuracy. An PTV expansion, or a pure geometric registration of the patient, does not suffice in proton treatments as the dose distributions is sensitive to geometry, and thus, geometry and dosimetry are strongly coupled, and one cannot serve as a surrogate for the other. Whereas geometric setup certainly serves as the best first corrective action, it does not in and of itself guarantee dosimetric accuracy even in XRT.

The optimal process re-images and re-plans the patient before treatment delivery and monitors during treatment for compliance to the adapted plan (and in the furthest extreme adapts the radiation field during treatment). Particle beams in these processes offer novel capabilities compared to photon beams. The particle beam, as a unit or as individual particles, can be readily detected in task-specific detection systems. These include geometric ionization chambers for position, solid state imaging planes for indirect spec-

troscopy of particle interaction products, and Faraday chambers for energy. These unique (compared to a photon beam) methodologies offer unique opportunities to obtain necessary inpatient information during treatment and can enable real-time feedback in the delivery process.

It is important to observe that the control parameters in ion radiotherapy – per spot charge position and energy – are more consistent than those – MLC leaf positions – in photon radiotherapy. For the latter, these leaf positions are an awkward (if necessary) transformation from the physical quantity – fluence – in patient. For ion radiotherapy, the spot parameters represent fluence in patient directly, are the parameters that are used to control the delivery, and are the parameters that are observed directly. There is no transformation from the intended to observed quantities, and treatment delivery observation can be directly translated to the dose in patient and used in a feedback control to maintain the correct dose in the patient under variable circumstances. Thus, particle radiotherapy can achieve better delivery performance when considering the modern aims of radiotherapy.

Effective use of particle radiotherapy is hampered by its label “expensive.” Any therapy has to be of course cost-effective. This is, unfortunately, a subjective debate within a society and between societies. Nevertheless, if considered as a debate in terms of cost-benefit in development, deployment, and clinical effectiveness, the outcome should favor particle radiotherapy as many of the aims are better (and thus likely cheaper) achieved with particle radiotherapy. Looking back, of course, one should question that because proton radiotherapy in the 1980s was superior to photon radiotherapy and because the cost to achieve parity between the two modalities has been significant over three decades, perhaps we should have adapted protons more broadly earlier. We should pose the same question now where significant expenditures will be made to achieve the latest aims of radiotherapy and where ion radiotherapy may significantly improve on the achievement and efficacy of those aims.

8.4 Requirements for Treatment Planning

It is necessary to challenge the current model for treatment planning given the above stated aims and the necessary integration of up-to-date technologies in the treatment management processes. There are three facets in the current model. The first continues to emphasize the pretreatment static (over time) treatment plan model without explicit consideration of the dynamics of the possible daily changes in treatment or other changes incurred by clinical realities. The second perpetuates the deployment of treatment planning systems with a 1980s model of a “workstation” with a functionally overburdened software application that aims to achieve too many workflow steps within its confines. The third emphasizes algorithmic components but not computational architectures that promote data management, communication, and workflow processes.

The deployment of a single, heavy, and shrink-wrapped application makes it difficult and contrived to move the various functions that accommodate the adaptive radiotherapy workflow to their optimal locations. In contrast, modern computing paradigms emphasize service-oriented architectures that promote logical decomposition of computational and data management domains and promote distributed access to these services. These architectures are based on distributed and disjointed processes connected by communication protocols.

Of particular interest are radiotherapy data management requirements characterized by large data sets and numerous temporal, logical, and computational associations between data. For example, a dose computation result should adapt when a treatment field parameter changes. The consequences of time have to be incorporated to model both motion effects and to model changes over the course of treatment.

Data management for radiotherapy has been often managed at the operating system file system level where individual patients are mapped to directories and patient data to sub-directories and files. Associations between data (i.e., a beam dependency of a dose calculation) were typically

not represented or implicit, and the state (i.e., the dose up to date *vis-à-vis* the beam state) was assumed. Relational databases have been commercially popular and readily available but are ill suited to radiotherapy data representations. Relational databases impose a rigid schema structure on the data and are coded to perform relational queries on large datasets of the type which are not useful for radiotherapy data (i.e., one is generally not interested in finding all patients with a gantry angle between 0° and 10°). Thus, neither the data casting nor the framework and its significant overhead are useful.

Instead, radiotherapy data is best managed by hierarchical structures linked by key-value association pairs to manage links between data instances. These key-value pairs and the desired query operations are better managed by much simpler systems such as NoSQL databases. Data is contained in XML documents, e.g., that allow for dynamic changes in the data definitions, dynamic management of associations, and ready replication and versioning. Most importantly, such databases readily scale to accommodate the problem and data size and promote distributed architectures.

Modern deployment models, such as cloud computing, are still absent in radiation oncology. It is clear, however, that the large computational requirements readily benefit from such models. A cloud deployment for computational and data management services would be to great benefit for both smaller and larger clinics. It provides affordable access to necessary computing resources on demand and distributed access in a hospital network. It frees clinical centers from the burden to invest in ever increasing computing hardware.

The limitations of the current, functionally monolithic, systems for treatment planning (TPS) and treatment management (TMS) have led to the introduction of ad hoc and institution-specific procedures to implement and manage workflow. After all, the clinical reality may require a patient to receive a new treatment plan which currently is essentially a repeat of the workflow for that patient’s treatment plan. Explicit tracking of such a repeat workflow, however, is not rendered in the TPS.

Of great significance is the work of the DICOM radiotherapy Working Group 7 which, in DICOM supplement 147, makes explicit the need for data objects that model the changing state of the patient as a consequence of treatment events and adaptable workflow tasks. In addition, IHE-RO, the Integrated Health Enterprise comprised of committees concerned with vendor and equipment interoperability and workflow within the radiation oncology domain, has defined specific workflow profiles that specify the details of various composite workflow processes necessary to accomplish tasks. An example workflow profile is “Integrated Positioning and Delivery Workflow,” which concerns itself with the positioning of a patient prior to treatment delivery, position monitoring (if any) during treatment delivery, and radiation delivery all managed by a single device.

In summary, modern requirements for treatment planning emphasize distributed computing as afforded by service-oriented architecture, explicit modeling of the workflow to dynamically connect the services needed as the treatment session unfolds.

8.5 Case Study

8.5.1 Volumetric Studies

Volumetric treatment planning requires the use of volumetric image studies for both photon (XRT) and proton radiotherapy (pRT). For pRT, the use of CT is axiomatic as CT is the only practical modality for which acceptable conversion from image voxel data, Hounsfield unit for CT, to voxel stopping power has been validated.

Volumetric image studies are necessary to define the physical computational space as represented by volumetric voxel elements and which allows the accurate computational transport of radiation through the patient’s anatomy.

Multimodality, cohesively registered, image studies allow the volumetric segmentation into discrete organs for dosimetric analysis of the internal anatomy. Multimodality image studies allow the geometric definition, in a known coordinate system, of target and organs at risk. Special

emphasis should be placed on accurate definition especially for the targets; failure negates the inherent precision of proton radiotherapy. Volumetric representations are derived from contour representations on image sections, while their computational representations are dictated by the computational requirements.

XRT and pRT differ notably in the use of a planning target volume (PTV). For XRT, the geometric expansion of a clinical target volume (CTV) into a PTV suffices to account for uncertainties equal to the geometric expansion. The lack of sensitivity of the photon field to these uncertainties, at least within the typical clinical magnitude, removes errors in dose and means that geometric accuracy equates to dosimetric accuracy. Thus, geometric positioning of the patient and geometric tracking of the patient suffice to maintain the dose distribution envelop within the desired specification. For a proton field, the field-patient interactions do depend on local geometry. Thus, geometric positioning does not suffice, or otherwise, and a change in geometry implies a change in dose. The use of a PTV is thus generally excluded. The PTV concept is nevertheless used, especially for spread-out Bragg peak (SOBP) fields. For the latter, the definition is typically obtained after a site-specific study reveals what margins yield in the desired tolerances after assessing dosimetric changes as a function of geometry [1].

Nevertheless, the use of PTV should be voided in favor of statistical approaches that model the treatment dynamics, in terms of geometric uncertainties in setup, patient motion, and patient changes. Bohoslavsky et al. [2], for example, define such a stochastic method and produce a margin prescription that improves on the use of a PTV margin. The improvement is a consequence of the fact that the dynamic consideration assesses the effect of statistical uncertainties, whereas the use of a PTV margin assumes that every treatment has to be within specification and thus represents is the worst case scenario.

The conversion of CT Hounsfield unit to relative (to water) stopping power is necessary as input to a proton dose algorithm, whether implemented as a heuristic (i.e., pencil beam) model or

as a Monte Carlo. This conversion has two basic problems. First, the conversion is based on a population-averaged conversion curve. This standard curve is applied to the CT Hounsfield unit value perhaps corrected or scaled to the specifics of a patient. Second, the conversion ignores details of the actual organ. Thus, different organs may have the same Hounsfield unit but different relative stopping powers. These limitations on the conversion to relative stopping power are an intrinsic limitation for pRT. The fundamental approach requires the use of proton transmission corrected data to allow the patient and site-specific derivation of stopping powers. Alternate methods, such as multispectral CT, are under active investigation to improve this conversion to voxel stopping power.

In practice, these uncertainties in stopping power assessment translate to a range uncertainty on the order of ~ 3 mm [3] and hence must be considered. For spread-out Bragg peak fields, the range and modulation width are increased by this amount which increases the longitudinal dimension by 6 mm! For pencil beam scanning (PBS) fields, the uncertainty needs to be considered in the optimization process, which is referred as robust optimization, i.e., the optimized result remains insensitive to the uncertainty [4].

8.5.2 Prescription and Course Considerations

The clinician's intent is expressed in statements that quantify the treatment course aims. There is, typically, a dichotomy between these statements and their use in a computational form. Most optimization algorithms are gradient based and use an objective function that is a quadratic summation of terms with heuristic weights to express relative term significance. Each term relates to an organ objective and thus a prescription statement. Such a form may be adequate for XRT where the number of optimization variables, i.e., leaf positions as a substitute for fluence profiles, is relatively (compared to pRT) small and the sensitivity of the optimal plan to these variables is relatively weak; i.e., the objective function has typically a

gentle varying minimum region. Even then, such forms often result in plans that, when evaluated by the physician, require heuristic tweaking of the weighting terms or the artificial introduction of "steering" volumes to meet the clinician's demand. Such algorithms therefore do not match well the supposed precision of a prescription.

The clinician's intent must be transformed into a computational form as input to an optimization algorithm. There are two issues. First, the prescription statement needs to be translated in to a computational form representative of the clinician's intent. Second, the nature of the computational form must yield a clinically optimal plan numerically consistent with the clinician's intent.

For prescriptions, we identify constraints and objectives (Fig. 8.3). Constraints are absolute statements such as minimum or maximum doses to a particular structure. Constraints are transparent – a constraint must be met and hence matches well to a computational form. Even so, gradient-based optimization methods often fail to meet the set of constraints. Objectives are clinical desires and often in competition with each other; that is, improving one objective must worsen another objective in a truly optimal plan. Objectives are constrained by the phase space of possibilities that remains after the constraints are satisfied. Objectives are not readily cast in computational form as their values are continuous and interdependent. It is the quantification and computation of objectives that often lack in optimization methods and impose iterations on the "optimal" plan as the clinician's intent is not represented. The use of multi-criteria optimization is a popular technique to manage objectives and inter-objective trade-offs [5].

In addition, we need to consider the time structure of a treatment; one reason already stated is to obviate the artificial use of a PTV. In XRT, the treatment session invariably requires a large set of beams that must be delivered as a single unit. This time structure is therefore invariably invisible and collapsed in a set of repeated identical fractions for a phase of the course. (Of note, we define a phase here as sub-course segment with its specific dose objectives such that all phase dose objectives meet the course dose objec-

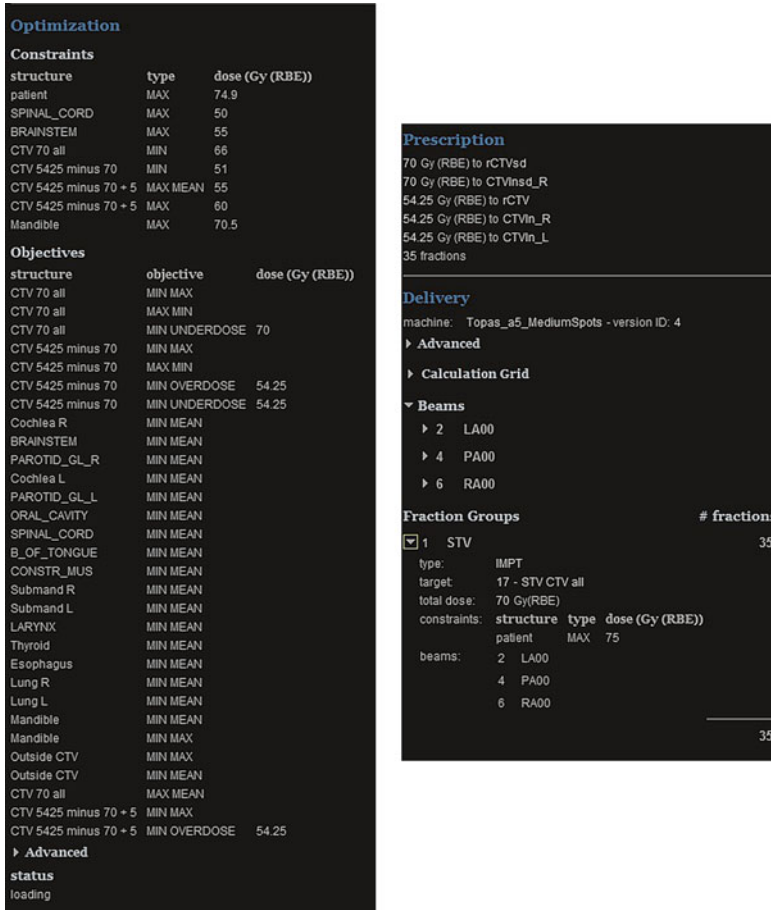


Fig. 8.3 The prescription, constraints, and objectives are shown for a nasopharynx case. The set provides a clinically complete specification for the multi-criteria optimization schema used in astroid. The prescriptions (total dose and fraction number) translate the total doses computed in the system to the appropriate fraction doses and spot parameters for treatment delivery. The constraints are absolute and must be achieved by the optimization (which otherwise fails to achieve a solution if the constraints can-

not be achieved). The objectives represent clinical desires within the bounds of constraints and are optimal trade-offs with respect to each other. That is, improving one objective worsens all others. The system uses fraction groups (see DICOM RT [Ion] Plan Fraction Scheme Module) that permit, for a particular course phase, to define and group subset of beam sets with individual constraints. The fraction group maps the total fraction group dose to individual fraction doses using the number of fractions

tives.) The practice of multi-beam sets within a XRT phase is rare to nonexistent.

The pRT practice does allow for intra-treatment session variability of the beam configuration. Such a configuration may be used for practical reasons, it is quicker to deliver two fields out of four per day, or biological reasons, where specific fractions can satisfy a sub-treatment phase biological objective compared to other fractions for the course phase.

The above observations concerning course and time management are consistent with DICOM Second Generation Radiotherapy (DICOM 2G; DICOM Supplement 147) [6]. DICOM 2G, first, recognizes the limitation of the first-generation objects and, second, makes explicit the above requirements in its data object definitions. The first-generation objects were time collapsed and ignorant of workflow order and simply served to move data between processes. The second-generation objects explicitly

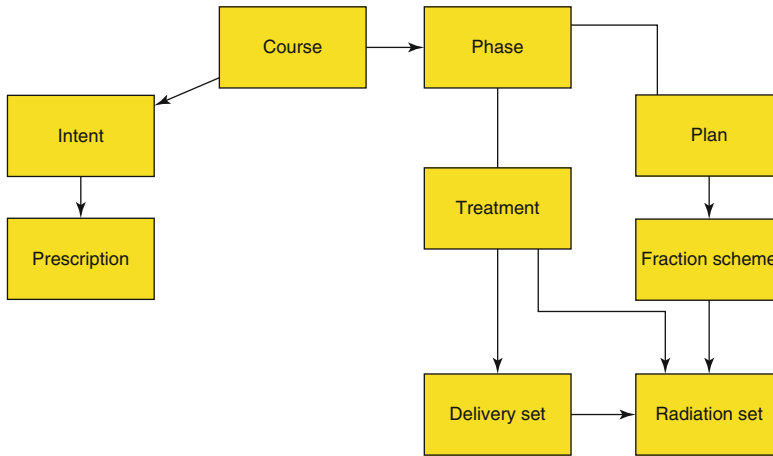


Fig. 8.4 DICOM second-generation object inspired model for treatment planning and delivery. A course defines its intent expressed as prescriptions. A course has one or more phases. Each phase has a plan representation in the treatment planning system and expressed by a fraction scheme (a set of one or more fraction groups) which

references radiation sets, a set of beams delivered in a fraction. A phase also has a parallel treatment representation expressed by the actual delivered radiation sets. The ability to represent and act on this model is a key to adaptive radiotherapy and complete documentation of the actual treatment versus planned

model the evolution of a treatment course and its delivery as a function of time. Adaptive radiotherapy (ART) requires DICOM 2G to consistently model and communicate patient changes and adaptations between computational and delivery services.

The second-generation model provides an RT course container structure that locates this course with respect to previous treatment courses and contains (1) prescriptions, (2) treatment phases that capture the time structure and differential objectives of the course, and (3) radiation sets (Fig. 8.4). A radiation set, in turn, captures fractionation schemas and the beams delivered in a fraction. The second-generation model, unlike the first-generation model which aggregated most data in the RT plan object, correctly decomposes data into orthogonal¹ object definitions, manages the time dimension, and captures the dynamics, i.e., variability, of the treatment course.

pRT is well positioned to use the second-generation structure and our planning methodology (astroid, a joint development between

Massachusetts General Hospital (MGH), Boston, MA and .decimal, Sanford, FL) explicitly models this structure. Astroid models a treatment course as above and has a radiation set as a fundamental planning unit. The user can model one or more radiation sets as a single set over which a (sub)set of the prescription statements and hence optimization are applied. Thus, optimization directly considers individual fractions and allows inter-fraction optimization to ensure that combinations of fractions (i.e., radiation sets) achieve a global course objective.

8.5.3 Field Considerations

Even a single proton field may provide the opportunity to achieve the dose objectives in contrast to the numerous fields required in XRT to achieve a measure of conformality. Thus, dose shaping does not rely per se on the number of fields. The choice of the number of fields and the orientation of the fields therefore remains an unsettled issue. Typically, the number of fields ranges from one to four per isocenter, and their orientation is best chosen to provide target coverage with the least lateral dimension, largest target to organ-at-risk

¹ Orthogonal design in data management implies that only a single data object represents a subset of data.

separation, and sharpest lateral penumbral falloff gradient between the target and organ at risk. Integral dose minimization is less of a practical concern as the integral dose is intrinsically low. It should be noted that these considerations are heuristic. The optimal number of beams and their optimal placement is a very difficult computational problem and beyond the current generation of treatment planning systems.

For our case, we assume a class solution and use a 3-field approach consisting of left and right oblique fields (at $\pm 45^\circ$) and a posterior field. Field size is typically not a limitation for scanned fields as for scattered fields where elongated targets require multiple isocenters. Multi-isocenter SOBPs fields are exceedingly cumbersome as they require feathered match lines.

The next consideration is the placement of “spots” defined as the (hypothetical) terminal point of the proton pencil beam. A spot point is defined in the, e.g., isocentric plane as a coordinate pair and in the longitudinal direction by energy/range. The choice of coordinates in the isocentric plane is typically chosen on a regular, rectangular or hexagonal, spaced grid. The choice of ranges may be constrained by the available ranges of the delivery device and is invariably constrained to sets of constant ranges, so-called energy layers, due to the long (compared to the lateral positioning of spots) time required to change energies. Thus, current systems require as many spots as possible to be delivered at the

same energy to minimize dead time between energy switches. The choice of layer spacing is “optimal” when consecutive layers are spaced proportional (or close to) the width of the layer pristine peaks. Pristine peaks, however, sharply decrease in width as a function of lower energy, and this optimal strategy causes a significant pileup of low-energy layers which results in a significant increase in treatment time because, again, energy switching time is a long process (on the order of seconds compared to milliseconds for spot lateral movement). Thus, our pragmatic approach, which contributes less than 2 % to the overall dose heterogeneity, is to space the energy layers by the width of the deepest and thus the broadest peak in the set.

It should be noted that the total number of spots (50,000 in our case for three fields, Fig. 8.5) for our case creates operational requirements for the delivery system. It is a rule of thumb that 10^8 protons deliver 1 Gy(RBE) to 1 cc of volume. The three fields, with a spot $\sigma \sim 5$ mm, deliver about 250×10^{11} protons in 50,000 spots with a mean of 5×10^6 protons/spot and a range of 10^4 – 10^8 protons/spot. Thus, a delivery system needs a dynamic range of about 10,000 in terms of spot charge control. A priori each spot should be delivered with high accuracy. Often, a specified charge may require the same spot to be delivered multiple times or at low (cyclotron) current to ensure the required charge precision. This, again, is an important treatment delivery system consid-

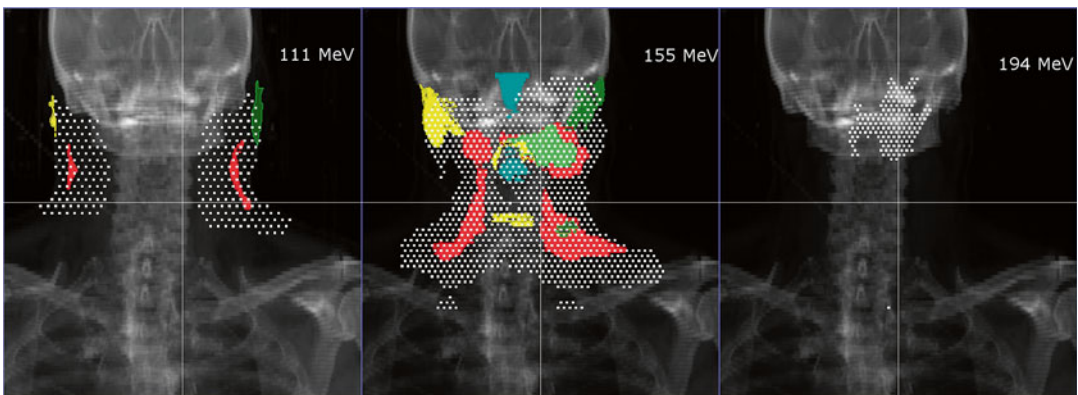


Fig. 8.5 Spot placement for three energy layers (111, 155, and 194 MeV) for the posterior-anterior field. Note that illustration of the respective target and organ-at-risk volumes is contained within the energy layer

eration. The number of protons per spot scales with the spot size when spots are spaced on a regular grid. For a spot with $\sigma \sim 10$ mm, the mean and ranges would be 2×10^7 and $4 \times 10^4 - 4 \times 10^8$. It is, of course, the lowest necessary charge in a problem that drives the performance of the delivery system. The ability of a proton production system, beamline, and scanning system to precisely deliver a spot of the minimum dose is a key performance requirement (Fig. 8.6).

The above spot placement strategy, as is the beam placement strategy, is heuristic. It ensures volumetric coverage of the set of spots but not that this set is optimal either to achieve the desired dose objectives or in terms of delivery. Again, a consideration of optimality for these parameters exceeds current practical computational abilities but is of considerable interest.

The resultant three fields result in 50,000 spots, each defined by a triplet of energy, isocenter plane position, and number of protons. The latter is converted to the reference ionization chamber moni-

tor unit to allow accurate control of the spot dose deposition. It is common to directly quantify the spot intensity by the equipment monitor unit in the treatment planning system. We cannot recommend this practice. Instead, we recommend the use of the number of protons as it is a device-independent and physically well-defined expression of the intensity. Thus, proton plans expressed by number of protons per spot can be readily inter-compared between different institutions.

Recent work in spot placement optimization [7] considers optimized placement of spots to minimize the number of spots while achieving prescription objectives. They allow for arbitrary spot placement, i.e., unconstrained in position and energy, and iterate over a set while continuously adding new spots and removing low-weight spots. Their analysis results in fewer spots and more charge per spot, more energy layers (as expected), reduction of dose to organs at risk (because spots can be removed), and (almost) an order of magnitude decrease in optimization

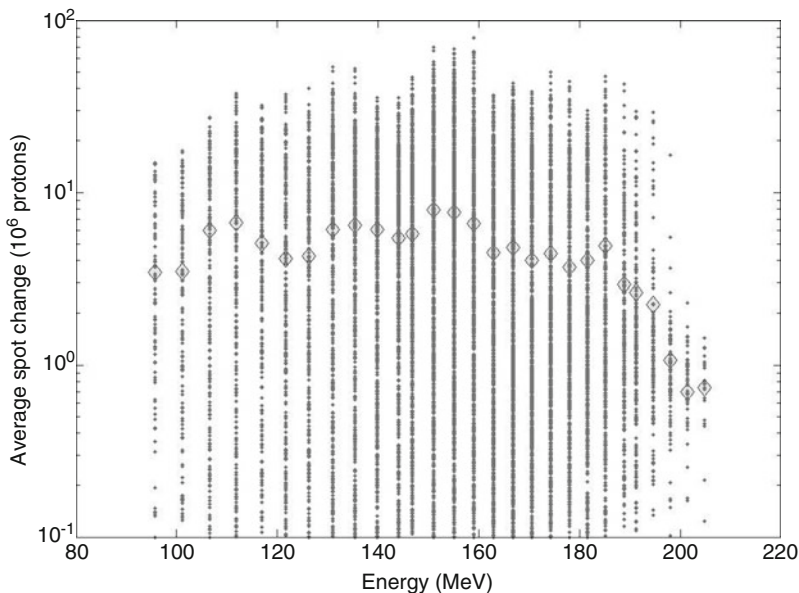


Fig. 8.6 The spot charge distribution within each energy layer (X-axis) for one field containing about 16,000 spots. The spot charge has an average of about 5×10^6 protons per spot. The range varies from (about) 100,000 to 100,000,000 protons per spot, a dynamic range of 1000. The ability for precise charge spot definition is an important requirement

of the proton production and delivery system. Considering the spot size of $\sigma \sim 5$ mm, the average spot charge is about $10^6 \cdot \text{mm}^{-2}$. That is, if the spot size increases (decreases), the average (and spreads) charge per spot increases (decreases). Thus, smaller spot size further increases the need for high-precision production and delivery

times. In addition, the technique improves as spot size decreases.

This strategy does require that spots can be of arbitrary energy, a capability not available with current generation of treatment planning and delivery systems. The next generation of delivery systems, such as the LIGHT linear accelerator (see Ugo Amaldi <http://cds.cern.ch/record/1312611>), promises such capability. It should be noted that the current practice of constant and separated energy layers intrinsically reduces the precision of proton target dose conformance by increasing the penumbral region around the target. Target dose inhomogeneity also increases as constant spot placement “misses” small target extensions and hence can cause an increase in local spot intensity as the spot is not well positioned relative to the target. Thus, the ability to place spots where needed and unconstrained by the current artifact of energy layers will improve dose conformance and homogeneity and the ability to use dose painting of targets.

It is, of course, the assessment of the optimal spot intensities that now remains as the core problem of the treatment plan optimization.

8.5.4 Plan Optimization

Our plan has the constraints and objectives listed in Figure 8.3. Our astroid system uses multi-criteria optimization ([8] and Fig. 8.7) to create a Pareto surface where each objective spans a dimensional axis and where the axis value range, i.e., minimum and maximum achievable objective value, is determined by the constraint values. That is, the optimization ensures (or fails otherwise) that every constraint is met and subsequently assesses the range of objective values. The objectives are correlated, i.e., their possible value sets span a Pareto surface in the multidimensional objective space, and improving one objective value (such as “minimized lung dose” and within the allowed range, see Fig. 8.6) necessarily worsens all the other objective values. That is, each set of objective values is best when considering the ensemble of values in the set. If the user wishes to improve one objective (say reduce surface dose to the brainstem), the set with that

new brainstem objective value out of necessity changes all other objective values. Algorithms that allow the user to change objective values are labeled as “Pareto surface navigation” algorithms and, themselves, are of a class of algorithms under investigation.

The astroid system allows the user to interactively change objective values within the constrained range, and the system will interactively update the dose displays to reflect the new set of objective values. Its navigation algorithm uses the new objective value as a constraint to find the set on the Pareto surface that contains that value.

Thus, the clinical practitioner, in effect, scrolls through all possible trade-offs and assures that each trade-off plan is the best given set of objective values. It is important to note that if the constraint values are changed, a completely different set of trade-offs can be considered. Thus, the Pareto surface optimization does not accommodate trading off a constraint. If so desired, the user can change a particular constraint to an objective and vice versa.

The result for our example case is shown in Fig. 8.8.

8.5.5 Plan Robustness

The dose computation on a static patient representation is considered incomplete for proton radiotherapy where the dose distribution is sensitive to uncertainties in geometry and range [9]. The latter is a consequence of the intrinsic systematic uncertainty in the conversion from CT Hounsfield unit to relative stopping power (among others). Figure 8.9 shows the dose-volume histograms (DVHs) for selected organs at risk and targets.

8.5.6 Dose Quality

The now practical availability of Monte Carlo computational methods will result in a shift, eventually complete, away from empirical pencil beam dose calculation models. Nevertheless, the pencil beam dose calculation models have been the basis for the clinical decision process and, in fact, have been quite accurate (except in

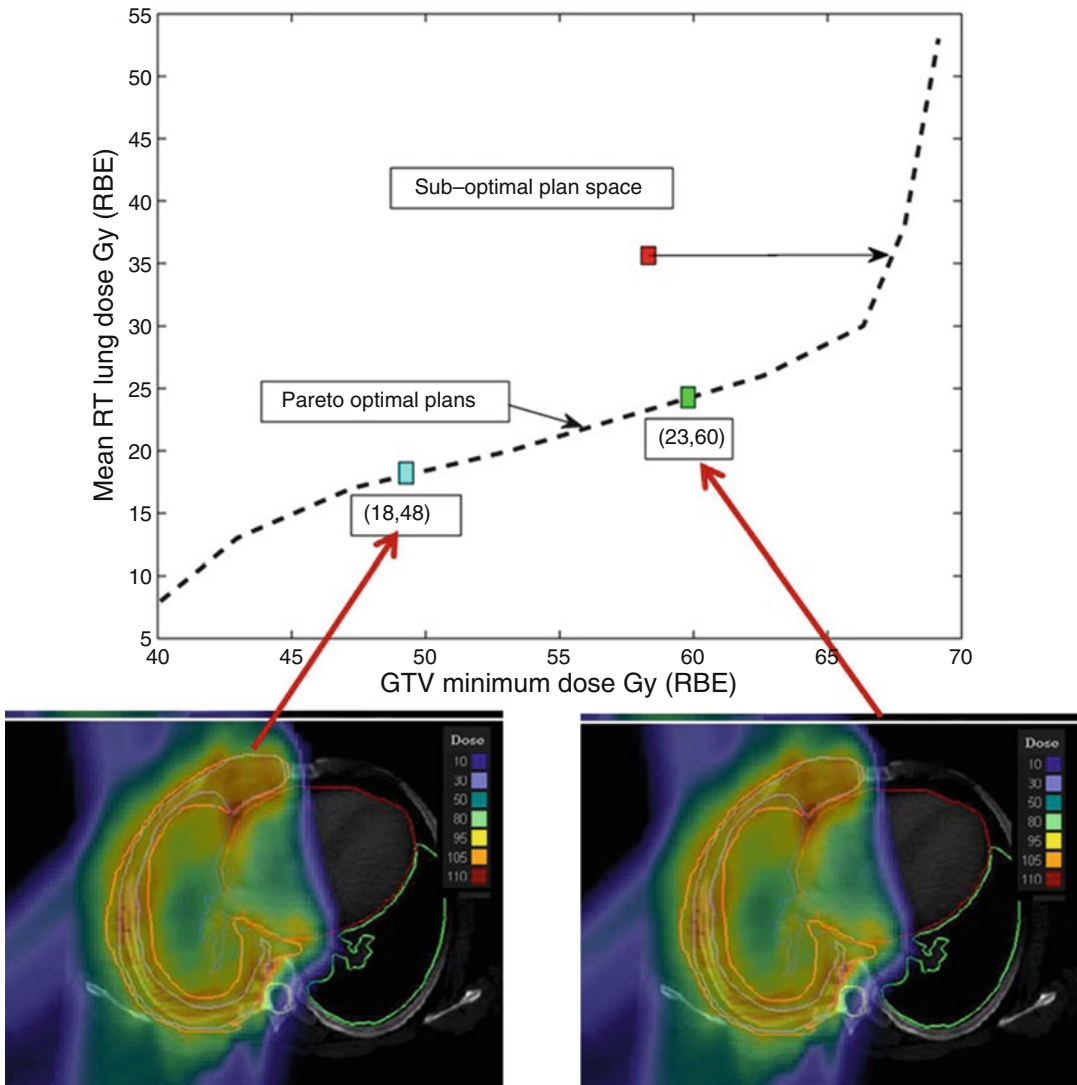


Fig. 8.7 Example trade-off scenario in a 3-field mesothelioma of the right lung pleura. The trade-off considers mean right lung dose (Y -axis) versus GTV minimum dose (X -axis). The curve that represents the possible trade-off pair values is illustrated in the graph above and forms a curve in the trade-off space. In this space, a hypothetical suboptimal plan (*red square*) lies above the curve. Its sub-

optimality is indicated by the observation that the GTV minimum dose can be significantly improved while maintaining mean right lung dose. The two isodose distributions represent the two points indicated on the curve. The clinical operator can move along the curve to assess the dosimetric consequence of a particular trade-off value pair

pathological cases such as where metallic implants are present in the patient). Figure 8.10 shows the results of dose computation with three different methods, one with a pencil beam model and two with a Monte Carlo, to illustrate the qualitative and quantitative differences and similarities between the methods.

8.6 Conclusion

Treatment planning requirements given the current aims of radiotherapy are not well implemented by the current commercial treatment planning system architectures. That is, proton radiotherapy permits a more dynamic evolution

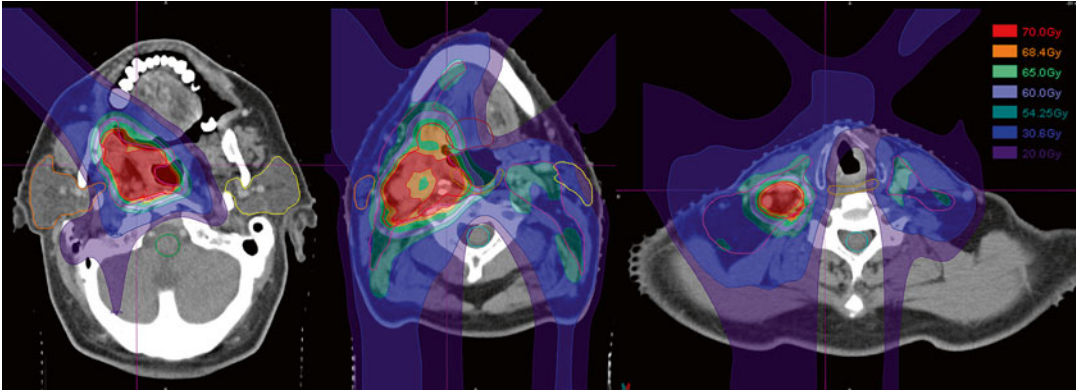


Fig. 8.8 The achieved dose distribution for the nasopharynx case example. The three fields contain about 50,000 spots (determined and placed heuristically without the

benefit of spot optimization). Note the ability for both dose avoidance and dose painting (the spot $\sigma \sim 5$ mm)

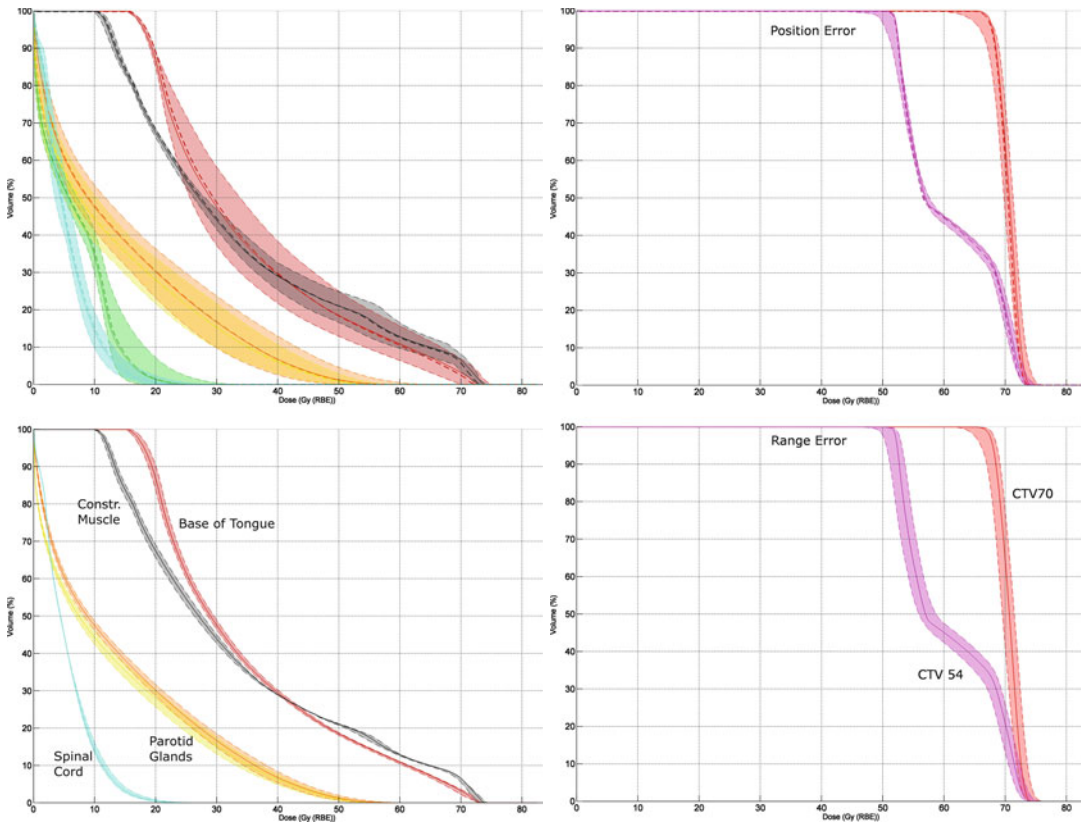


Fig. 8.9 The *left panels* show the DVHs for (selected) organs at risk and the *right* for the two CTV volumes. The *upper panels* show the effect of position error where the dose is recomputed on the error-modified patient with the nominal plan parameters. The *lower panels* show the effect of range over and undershoot of 3 %. The effect of position is more severe on the organs at risk as a consequence of the fact that these organs are invariably in high-dose gradients and thus are susceptible to geometric shifts of that gradient. The target volumes are much less affected. The range

error (*lower panels*) has almost no effect on the organs at risk but does affect the targets. The DVHs show the maximum and minimum bands with the *dashed line* representing the mean DVH of the error scenarios and the *solid line* representing the nominal DVH. One must observe that the positioning errors will average out to this mean over the large number of fractions for this site. Thus, the mean (or nominal) DVH is the representative as it is in XRT treatments. The range errors, however, are systematic and must be assessed within the band of the DVH

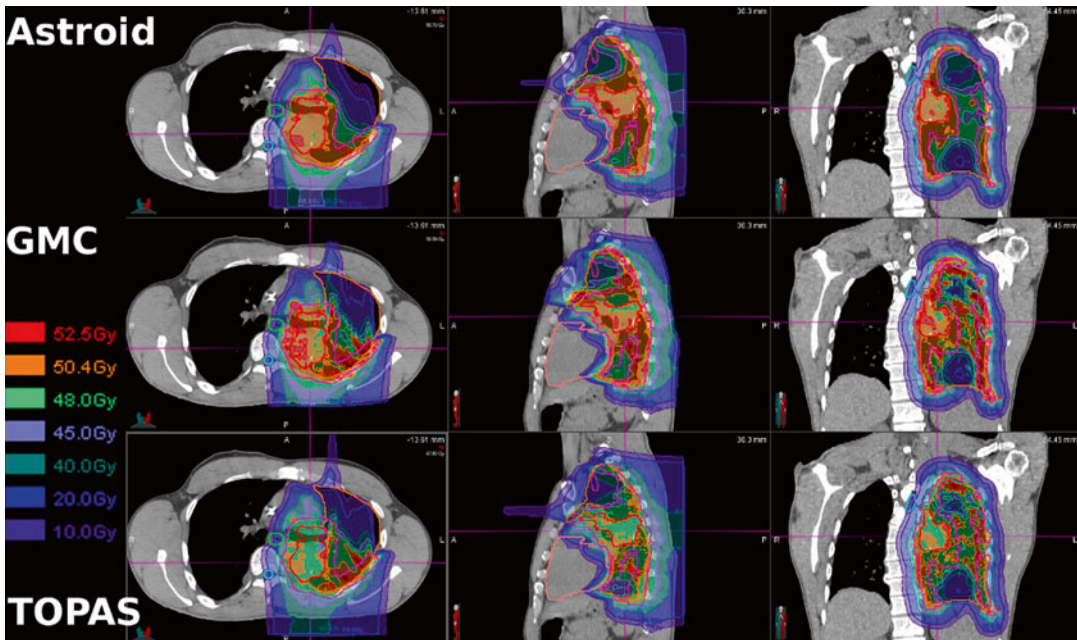


Fig. 8.10 Comparison of dose computation results between a pencil beam model implementation (astroid), a simple primary and secondary scatter Monte Carlo (GMC), and a complete Monte Carlo (TOPAS). Qualitatively, the distributions show the same features in this very heterogeneous geometry. Both Monte Carlos show, albeit with slight differences, more out-scattered protons in the target volume and hence more heterogeneous and lower dose. Monte Carlo calculations are criti-

cal as an independent verification. Their clinical accuracy, i.e., how will their differences impact clinical practice, needs to be clearly specified as they will imply changes in a practice established with “less” accurate computational means. Even so, the absolute accuracy of Monte Carlo in the tissue itself also requires more investigation because of patient tissue differences and calibration biases in the Monte Carlo (as in any dose algorithm)

and implementation of a treatment course as multiple radiation sets of only a few fields each can achieve competing dose objectives and thus permit a more tuned and adaptive approach to the planning and delivery process. The DICOM 2G definitions are a complete model for such approaches.

Proton treatment plans are computationally demanding because the number of optimization variables, namely, spot intensities, is very large (on the order of 10,000–100,000 per patient), because the set of “optimal” but competing solutions are (presumably) better, and because a treatment plan must explicitly model the uncertainty space (as compared to the PTV heuristic in XRT). These computational demands will greatly benefit from modern service-oriented and scalable architectures to provide the necessary computational horsepower.

Thus, just as proton radiotherapy in the 1980s led the way toward computational treatment

planning systems, proton radiotherapy in the coming decade again will push treatment planning toward more capabilities.

References

1. Engelsman M, Rietzel E, Kooy HM. Four-dimensional proton treatment planning for lung tumors. *Int J Radiat Oncol Biol Phys.* 2006;64:1589–95.
2. Bohoslavsky R, Witte MG, Janssen TM, van Herk M. Probabilistic objective functions for marginless IMRT planning. *Phys Med Biol.* 2013;58:3563–80.
3. Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Phys Med Biol.* 2013;2012(57):99–117.
4. Chen W, Unkelbach J, Trofimov A, Madden T, Kooy H, Bortfeld T, Craft D. Including robustness in multi-criteria optimization for intensity-modulated proton therapy. *Phys Med Biol.* 2012;57:591–608.
5. Halabi T, Craft D, Bortfeld TR. Dose–volume objectives in multi-criteria optimization. *Phys Med Biol.* 2006;51:3809–18.

6. DICOM Standards Committee, Working Group 7. Radiation therapy. Supplement 147: Second Generation Radiotherapy. Revision 42, March 28, 2014
7. van de Water S, Kraan AC, et al. Improved efficiency of multi-criteria IMPT treatment planning using iterative resampling of randomly placed pencil beams. *Phys Med Biol.* 2013;58:6969–83.
8. Monz M, Kufer KH, Bortfeld TR, Thieke C. Pareto navigation—algorithmic foundation of interactive multi-criteria IMRT planning. *Phys Med Biol.* 2008; 53:985–98.
9. Kraan AC, van de Water S, et al. Dose uncertainties in IMPT for oropharyngeal cancer in the presence of anatomical, range, and setup errors. *Int J Radiat Oncol Biol Phys.* 2013;87:888–96.