NUCL SCI TECH (2021) 32:34
https://doi.org/10.1007/s41365-021-00872-4

Proton linac-based therapy facility for ultra-high dose rate (FLASH) treatment

Wen-Cheng Fang[1,2] · Xiao-Xia Huang[1] · Jian-Hao Tan[1] · Chao-Peng Wang[3] · Cheng-Cheng Xiao[1] · Yi-Xing Lu[1,2,4] · Yu Zhang[1,2,4] · Yu-Qing Yang[1,2,4] · Yi-Ming Xu[1,2,4] · Han-Yu Gong[1,2,4] · Jun-Qiang Zhang[1] · Shao-Peng Zhong[1] · Zhen-Tang Zhao[1,2]

Received: 11 January 2021 / Revised: 20 February 2021 / Accepted: 4 March 2021 / Published online: 8 April 2021

Abstract As an advanced treatment method in the past five years, ultra-high dose rate (FLASH) radiotherapy as a breakthrough and milestone in radiotherapy development has been verified to be much less harmful to healthy tissues in different experiments. FLASH treatments require an instantaneous dose rate as high as hundreds of grays per second to complete the treatment in less than 100 ms. Current proton therapy facilities with the spread-out of the Bragg peak formed by different energy layers, to our knowledge, cannot easily achieve an adequate dose rate for FLASH treatments because the energy layer switch or gantry rotation of current facilities requires a few seconds, which is relatively long. A new design for a therapy facility based on a proton linear accelerator (linac) for FLASH treatment is proposed herein. It is designed under two criteria: no mechanical motion and no magnetic field variation. The new therapy facility can achieve an ultra-high dose rate of up to 300 Gy/s; however, it delivers an instantaneous dose of 30 Gy within 100 ms to complete a typical FLASH treatment. The design includes a compact proton linac with permanent magnets, a fast beam kicker in both azimuth and elevation angles, a fixed gantry with a static superconducting coil to steer proton bunches with all energy, a fast beam scanner using radio-frequency (RF) deflectors, and a fast low-level RF system. All relevant principles and conceptual proposals are presented herein.

Keywords Proton therapy · FLASH · Proton linac · Deflector · Superconducting coil · Fast scanner

This work was supported by the Alliance of International Science Organizations (No. ANSO-CR-KP-2020-16) and the National Key R&D Program of China (No. 2018YFF0109203).

✉ Wen-Cheng Fang
fangwc@sari.ac.cn

Zhen-Tang Zhao
zhaozhentang@sinap.ac.cn

1 Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201204, China

2 University of Chinese Academy of Sciences, Beijing 100049, China

3 Shanghai APACTRON Particle Equipment Company Limited, Shanghai 201800, China

4 Shanghai Institute of Applied Physics, Chinese Academy of Sciences, Shanghai 201800, China

1 Introduction

Radiation-induced toxicity is a significant issue in irradiation therapy including proton therapy; hence, effective methods to protect normal tissues against toxicity have been investigated. In recent years, in vivo studies have shown that electrons, X-rays, or protons delivering all treatment doses of approximately 30 Gy in an ultra-short time (less than 100 ms), as defined by the requirement for ultra-high dose rate (FLASH) radiotherapy (hundreds of grays per second), can protect normal, healthy tissues compared with the conventional irradiation of the same total dose, which has be demonstrated by different radiotherapy of electron [1], proton [2] and X-ray [3]. In particular, the first clinical treatment has been accomplished by 15 Gy electron radiation within 90 ms [4]. Interest in FLASH radiation therapy as a new irradiation technique is currently increasing, but no commercially available therapy facility that includes protons, electrons, or X-rays can yield dose rates as high as that afforded by FLASH radiotherapy. Because of the precise depth dose distribution and ballistic properties of proton therapy, it may be an innovative approach for achieving the dose required for FLASH radiotherapy; however, many crucial techniques must be developed for this new generation of proton therapy.

Proton therapy, which constitutes approximately 85% of hadron therapy, is the primary method to treat tumors [5]. FLASH proton may be available with the use of modern equipment; however, to our knowledge, it has not yet been designed or implemented for FLASH radiotherapy. Two challenges for achieving FLASH proton in proton therapy facilities are as follows: energy must be switched repeatedly to obtain the spread-out of the Bragg peak (SOBP) and ultra-fast pencil beam scans. Cyclotrons and synchrotrons are currently the main proton facilities in operation worldwide. A cyclotron can only extract proton beams of fixed energy, and an additional degrader with mechanical motion must be used for energy change. In synchrotrons, all magnets must be cycled one or more times to eliminate the residual magnetic field before each energy change. Both cases require subseconds or a few seconds to switch each energy, which implies that it cannot complete one FLASH delivery within $\sim$ 100 ms. Meanwhile, pencil beam scanning is controlled by two *X*–*Y* scanning coils, with a scanning time of milliseconds for one point-to-point step, which is not sufficiently fast to deliver FLASH doses for the entire volume within 100 ms. To implement FLASH proton therapy, a new design for therapy facilities based on a proton linear accelerator (linac) for FLASH treatments is proposed herein. The linac accelerates protons in a straight line based on a simple magnet lattice, and it can change the proton energy pulse by pulse within milliseconds; it is a highly competitive and promising solution for achieving FLASH treatments. Under the two criteria of no mechanical motion and no magnetic field variation, the new therapy facility can achieve a FLASH of up to 300 Gy/s; therefore, a full FLASH treatment can be completed within 100 ms in proton therapy.

Based on the two criteria above, the conceptual design of a new therapy facility begins from a compact proton linac, through a fast beam kicker and a fixed gantry with static superconducting coils, to the end of a fast beam scanner. It includes many new techniques compared with conventional proton therapy facilities. To realize a more compact facility, the proton linac used was based on high-gradient technology developed via high operating frequencies; however, the proton energy with low phase velocity (β) might hinder the manufacture of radio-frequency (RF) structures with higher frequencies. Consequently, a compact linac was designed to include a proton injector comprising of a 714 MHz radio frequency quadrupole (RFQ) and a 714 MHz drift tube linac (DTL), an S-band DTL section, and an S-band linac, whereas all magnets of the S-band linac section, which encompass the energy range from 70 to 235 MeV, were permanent to force proton beams instead of electrical magnets. The fast beam switcher between the linac and gantry was implemented using an RF deflector with variable polarization, which can steer the proton beam in both azimuth and elevation angles. The gantry was implemented using many static superconducting coils that were distributed well in 12 azimuth directions, i.e., each can steer all-energy proton bunches from 70 to 235 MeV. At the end of the facility, a beam scanner was implemented using an RF deflector with variable polarization; therefore, an ultra-fast delivery of each scanning layer to the exact tumor location can be realized. Finally, the energy change or beam scanning was controlled by an ultra-fast low-level RF system (LLRF), which is faster than normal LLRF technology [6]. All details of the relevant principle and scheme are introduced and presented herein.

1.1 Overall requirements and conceptual design

FLASH treatment requires an ultra-high dose rate up to 300 Gy/s; therefore, it can deliver 30 Gy to the tumor in 100 ms. This indicates that the proton beam delivered to the tumor, based on a standard phantom measuring 10 cm $\times$ 10 cm $\times$ 10 cm, should reach an average current of 60 μA for the lowest energy layer of 70 MeV, and other energy layers above 70 MeV can spontaneously achieve a higher dose rate exceeding 300 Gy/s. Practically, in all treatment planning system (TPS) recipes, all 91 energy layers are appointed to specific energy points distributed from 70 to 235 MeV. A typical TPS recipe should enable many scanning layers (but less than 100 layers) for a tumor, in which the energy layers must be fewer than the scanning layers because the same energy typically exists in two or more neighboring scanning layers. To deliver a 30 Gy dose for different scanning layers within 100 ms, one scanning layer must be completed rapidly in approximately 1 ms, during which the proton beam is accelerated to the same energy. Because most parts within 1 ms are allocated for beam preparation and energy changing, while considering the current level of the pulse power, the maximum duration of beam radiation can be maintained at 10 μs to complete one scanning layer. Based on the analysis above, FLASH treatment can be implemented, as shown in Fig. 1, to determine the overall specifications for the conceptual design of proton therapy based on a linac.

A new proton RT facility based on a linac is proposed herein to perform the FLASH treatment illustrated in Fig. 1. The facility comprises a compact linac with energy ranging from 70 to 235 MeV, a fast kicker for proton beam separation, a static superconducting coil capable of steering

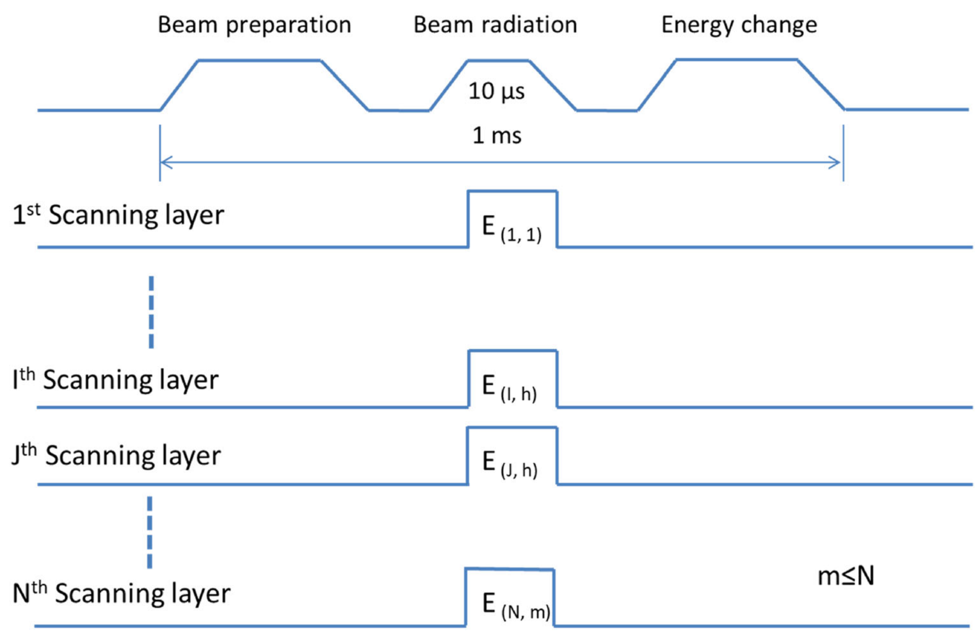

Fig. 1 Irradiation scheme of typical TPS recipe for proton FLASH therapy

all 91 energy points to the same final point, and a fast beam scanner for proton bunch scanning.

The compact linac will be developed based on high-gradient technology including RFQ, DTL, and an accelerating structure, and all the linacs should be operated at a repetition rate of 1 kHz with a pulse length of only 10 μs as the radiation length. The upstream of the linac is an injector comprising a 2.45 GHz ECR ion source, a 714 MHz RFQ-DTL, and four 2856 MHz DTLs, whose aim is to attain a fixed energy of 70 MeV. The ion source was designed to produce a 20 mA pulse current to capture a 10 mA peak current after the injector. The downstream of the linac is an S-band linac operated at 50 MV/m that can adjust the energy from 70 to 235 MeV while transferring a 10 mA pulse current to the gantry. The total length of the compact linac can be less than 18 m. In particular, all magnets in the linac are permanent without any magnetic field variation; consequently, the scanning and energy switch of one layer can be completed simultaneously within 1 ms.

The fast beam kicker and scanner were both implemented using an advanced deflector with variable polarization to steer the beam without any magnetic field variation, as in Criteria 2 above. The proton bunches from the linac can be kicked in both the azimuth and elevation angles by the fast kicker. Under the control of the low-level radio frequency (LLRF), the kicking angle can be switched pulse by pulse with the same repletion rate of 1 kHz, which is much faster than magnetic steering. Similarly, the fast scanner can perform faster in both the *X* and *Y* directions than the magnetic coil scanner to complete one-layer scanning within 10 μs.

The gantry was implemented using many superconducting coils located at different angles in one cycle, in which each identical coil was static without any magnetic field variation. It should be able to steer all proton bunches including 91 energy points, and finally to the isocenter point of the treatment room. Based on the gantry design, proton bunches with all 91 energy layers can be steered and delivered rapidly from different directions to the same point as the isocenter, thereby completing one full FLASH treatment within 100 ms. Because of the superconducting coil with an achievable 9 T magnetic field, the gantry was extremely compact compared with the conventional gantry.

The overall specifications of the conceptual design are listed in Table 1 based on the requirements of FLASH treatment. All the new techniques presented above can promote the development of new-generation proton therapy facilities. In the following sections, each part of the facility is introduced with more details based on reasonable principles and feasible technology.

1.2 Compact proton linac

In the proposed design, the compact linac comprises an injector and an S-band linac, as shown in Fig. 2. The maximum proton energy was targeted at 235 MeV, which is to be achieved at a length of 18 m; this can be realized using the high-gradient technology applied on RFQ, DTL, and S-band standing wave structures to achieve a compact proton linac. To rapidly deliver proton bunches and switch energy layers, the linac can be operated at a repetition rate of 1 kHz, with a pulse length of 10 μs. Meanwhile, the proton linac is fully configured using permanent

Table 1 Overall specifications of proposed conceptual design of FLASH radiotherapy facility

Parameter	Value or description
Peak dose rate (Gy/s)	300
Treatment duration (ms)	100
Energy range (MeV)	70–235
Repletion rate (kHz)	1
Irradiation pulse length (μs)	10
Peak current of ion source (mA)	20
Peak current after linac (mA)	10
Fast beam kicker	Deflector with variable polarization
Gantry type	Static superconducting coil, all energy
Beam scanner	Deflector with variable polarization

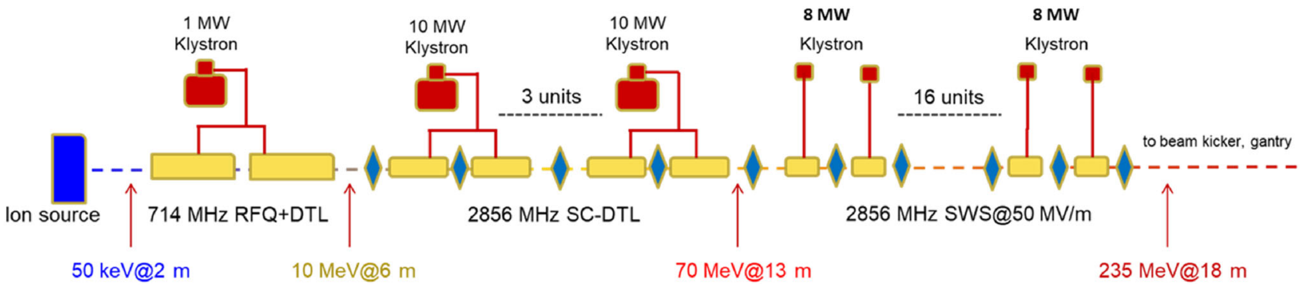

Fig. 2 (Color online) Schematic design of compact linac

quadrupole magnets without any variation in the magnetic field to focus and deliver all proton bunches in one FLASH treatment through the linac in less than 1 ms; hence, the consumption of electrical power is reduced compared with that of conventional accelerators.

Based on the configuration shown in Fig. 2, a repetition rate of 1 kHz is a difficult to achieve using the typical conducting technology for on an average RF power of 1.57 MW. However, because the duration of each FLASH treatment is only 100 ms, the total required RF energy of only 157 kJ can be reasonably powered by an electrical source and cooled by a water-cooling system. In particular, the klystron as a key equipment is extremely difficult to operate continuously at a 1 kHz repetition rate; however, a 100 ms treatment enables it to be an instantaneous pulse, which can only produce a low average power to simplify the configuration of the klystron collector.

In addition to reducing power consumption and space, all the quadrupoles were based on permanent magnets. Most importantly, the proton energy can be switched rapidly within 1 ms. Meanwhile, the rising time of all accelerating structures was 1–2 μs, i.e., three times the filling time, which is a short additional time not included in the irradiation time of 10 μs. Each 8-MW klystron depicted in Fig. 2 was driven and controlled by one independent low-level RF system; therefore, the klystron power can be changed ultra-fast in approximately a few microseconds. In principle, the linac can produce 100 energy layers at the most within 100 ms, adequate to create one typical TPS recipe including all scanning layers, which is less than 100 layers. In the upstream of the linac, as shown in Fig. 2, the injector was operating at a steady energy of 70 MeV as the lowest energy layer for the TPS recipe, rendering it relatively easy to design the linac lattice of permanent quadrupoles. However, downstream of the linac, which comprised 16 S-band accelerating structures achieving a 50 MV/m accelerating gradient higher than [7], 97 discrete steps from 70 to 235 MeV should be operated to configure different appointed energy layers; however, such a lattice design is complicated and poses a significant challenge as a crucial technology for the proton linac.

The linac proposed herein is similar to that in [8], but it requires a dedicated design and further optimization to achieve a high repetition rate and high gradient. The proton source based on a 2.45 GHz ECR ion source should produce a pulse current of 20 mA with a 10 μs pulse length, operated in 1 kHz repetitions. The RFQ and DTL operating at a higher frequency of 714 MHz, similar to the 750 MHz RFQ in [9], can achieve a higher gap voltage, thereby rendering the RFQ and DTL much more compact; additionally, a 10 MeV proton injector can be produced with only 4 m. Assuming the capture coefficient is 50% after the RFQ and DTL sections, the pulse current bunched can be 10 mA when the proton beam is delivered to the end of the DTL. Following the three units of the 2856 MHz SC-DTL, the bunched proton beams are then accelerated up to

70 MeV. Finally, the bunched proton beams are injected into the S-band high-gradient linac, including 16 accelerating structures independently driven by one 8 MW klystron. Each accelerating structure in the linac can achieve six different types of proton energy; consequently, the proton linac can generate a total of 97 discrete energy layers from 235 to 70 MeV, thereby satisfying the maximum requirement of 91 energy layers in one FLASH treatment. Figure 3 presents the preliminary beam dynamics of the proton linac.

1.3 Fast beam kicker

In the conventional facility for proton therapy, between the accelerator and gantry, a magnet is used to steer proton bunches with different energies in the elevation direction. However, it requires at least 10 ms to change the magnetic field to match with the proton energy, and this duration does not include the cycling time of the magnet. Meanwhile, in the azimuth direction, proton bunches are delivered via gantry rotation, which requires even more time (i.e., a few minutes); this will be presented in the next section. The beam delivery in both the azimuth and elevation angles in the conventional facility requires a considerable amount of time, and one FLASH treatment can be completed in 100 ms. To fast kick bunches in both the azimuth and elevation angles, a fast bunch kicker based on the new principle of an RF deflector with variable polarization will be equipped between the linac and gantry; this enables an ultra-fast kick rate of a few megahertz to be achieved primarily via LLRF technology.

The deflector as a beam diagnostic tool can deflect the beam and be used as an RF kicker [10]. The principle of the bunch length measurement is shown in Fig. 4. For a single-mode operating deflector, the HEM11 mode is the primary operating mode, which results in the rotation of the deflection plane in an axis-symmetric structure. To solve

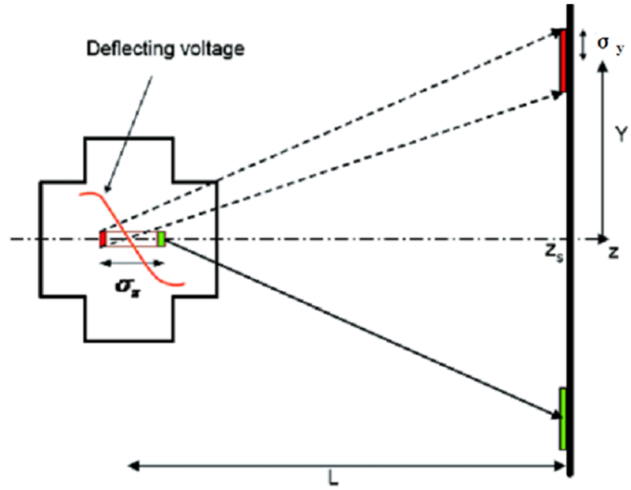

Fig. 4 Schematic principle of RF deflector

the degeneracy and prevent rotation on the deflection plane, some methods that break the symmetry were applied to cancel the degeneracy of the HEM11 mode. Once the deflection plane is fixed, the direction of deflection of the beam is fixed as well. However, the typical deflector shown in Fig. 4, which operates on a single mode, cannot fulfill the requirements of the fast kicker above. To satisfy the requirement of multidirectional deflection, a new design with variable polarization is proposed herein.

A new type of deflector operating in two modes was first proposed for the measurement of beam slice features in both the X and Y directions in 2015 [11]; similarly, such a deflector with variation polarization can be used as an ultra-fast bunch kicker for FLASH treatments. The deflector is driven by two power sources operating independently on two harmonics, i.e., the S-band (2856 MHz) and C-band (5712 MHz), which can establish two independent modes orthogonal to each other in the deflector cavity as HEM11 and HEM12, respectively, as presented

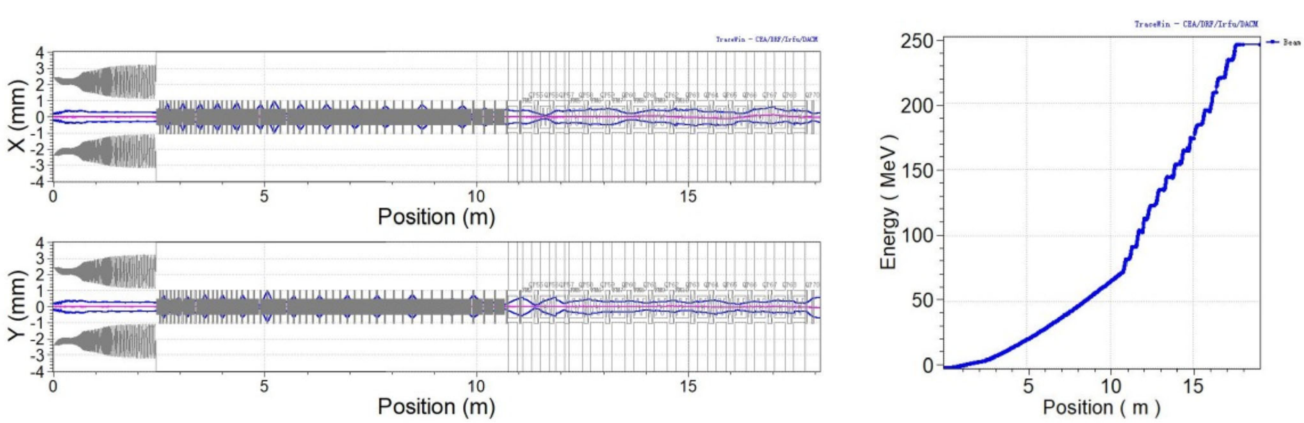

Fig. 3 (Color online) Preliminary beam dynamics simulation of compact linac. Left: *X*–*Y* beam size along linac; Right: Evolution of beam energy along linac

on the left side of Fig. 5. Based on two orthogonal modes exactly fixed as vertical and horizontal vectors, a specified solid angle comprising azimuth and elevation angles can be feasible and achievable, as shown on the right side of Fig. 5. Practically, any kicking angle can therefore be implemented under the control of an LLRF system as an RF power source, such as S-band and C-band klystrons. The LLRF is the most recently developed technique; it includes amplitude and phase modulations in many FEL synchrotron light source facilities and can achieve ultra-fast speeds for system control.

As shown in Fig. 5, under LLRF technology, the duration for status switching can be as short as a few microseconds. The fixed duration of one pulse length of the proton bunch energy, i.e., 10 µs, does not including the rising time of 0.3 µs, which is approximately three times the filling time, whereas the entire pulse duration is 1 ms, as presented in Fig. 1. Therefore, the kick angle should be fixed and maintained for at least 10 µs, as well as complete the kick angle switch for 1 ms, both of which are feasible and achievable by an RF power source under the LLRF technology. Based on the RF deflector with variable polarization, the key connection between the linac and gantry can be implemented to kick proton bunches at different solid angles at an ultra-fast rate. In FLASH therapy, it is essential to ensure that all ultra-fast proton bunches are delivered without any magnetic field variation or mechanical motion.

1.4 Fixed gantry comprising static superconducting coils

Typically, most recipes for tumors should be segregated into two or three radiation fields in different directions to protect healthy tissues; conventionally, this is performed by rotating the heavy gantry slowly for more than 1 min, which hinders a high dose rate to be achieved for FLASH treatments. For some simple tumors with only one radiation field, because of the multiple energy layers required to form the SOBP, the magnets on the gantry should be changed repeatedly within several seconds. The conventional gantry cannot satisfy the two criteria described above, and in principle, it cannot complete a FLASH treatment within 100 ms. Therefore, a new gantry without mechanical movement is proposed, as presented in Fig. 6. It comprises several identical superconducting coils for deliver proton bunches with different energies. The schematic illustration of one unit in such a gantry is presented in Fig. 6, showing the principle of beam delivery in only one unit between two superconducting coils, which is similar to the gantry design presented in [12]; however, the new design for steering bunches such as an α-orbit affords more space for patient treatment around a couch.

As shown in Fig. 6, the proton bunches were kicked by the fast kicker upstream at different elevation angles, such as θ_1 and θ_2, which correspond to different energies. After kicking at different angles, proton bunches with different energies were first injected on different orbits to the steering magnetic area between two superconducting coils. Eventually, they were focused into an identical point, such as the isocenter in the treatment room, after passing through the α-like orbit. Before the steering field, proton bunches with lower and higher energies passed along the upper and lower orbits, respectively. In the steering field, the orbit lengths for different energies differed; however, the orbit length for all energies should be controlled to ensure that all proton bunches can be prevented from internal collisions in the crossing. However, all proton bunches with only 10 ps length were distributed periodically, with a fixed period space of 350 ps; this implies that an extremely wide space existed between two bunches for the vertical bunches to pass through without any internal collision.

The fixed gantry developed based on the scheme shown in Fig. 6 is preferable for the FLASH facility instead of a rotating gantry to avoid mechanical motions, as shown in Fig. 7. The fixed gantry is a toroidal periodicity facility with several beam channels distributed uniformly in angle,

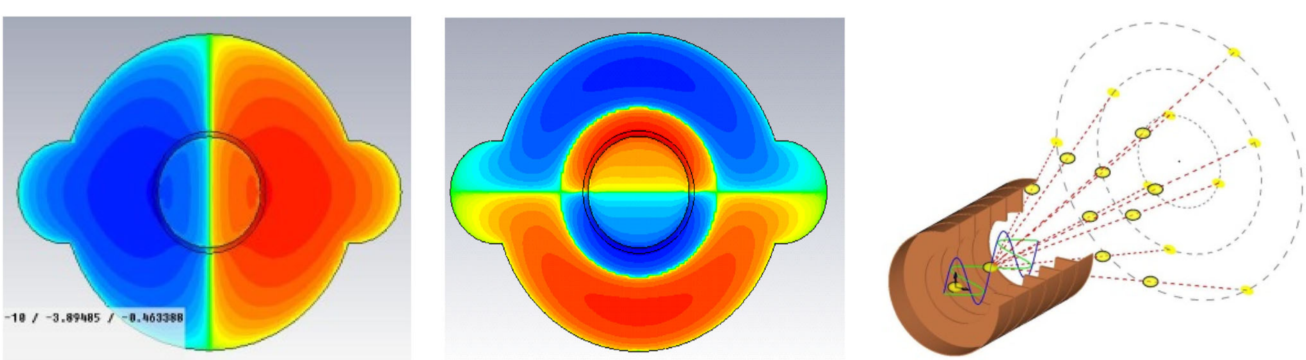

Fig. 5 (Color online) Schematic illustration of two-mode deflector as proton bunch kicker

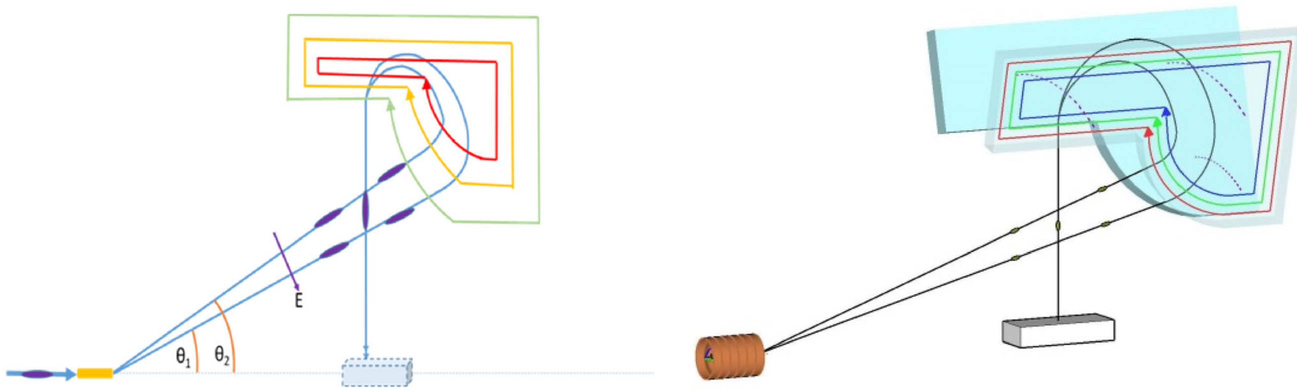

Fig. 6 (Color online) Schematic illustration of static gantry delivering proton beams from 70 to 235 MeV

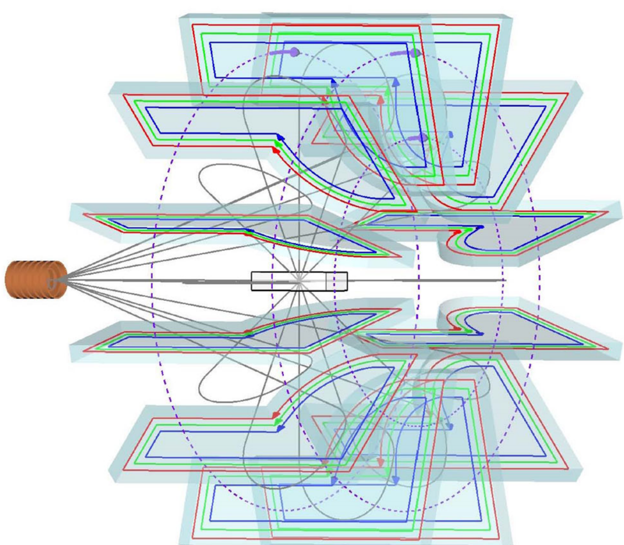

Fig. 7 (Color online) Schematic illustration of static superconducting coil covering 91 energy layers steered in different angles

by which all proton bunches of 91 energy layers can be delivered and then focused on an identical point in the treatment room. Considering the imaging and structural symmetry, the number of channels was initially expected to be 12. The gantry was developed based on a superconducting coil instead of a permanent magnet to reduce the weight of the facility and maintain a strong magnetic field. All coils yielded a circular magnetic field along the azimuth direction, steering all proton bunches from all azimuth angles kicked by the RF deflector along the α-like orbit in the vacuum channel. Based on the scheme shown in Fig. 6, all proton bunches from any azimuth angle were finally delivered to the isocenter of the treatment room.

The most significant property of the fixed gantry was the rapid delivery of bunches, as presented in Fig. 7; however, this must be assisted by the ultra-fast kicker achieved by the deflector with polarization variation, as shown in Fig. 5. In the transport channel, proton bunches with different energies were first kicked to different elevation angles and then entered the steering magnetic area at different initial points. The initial point must match exactly with the proton energy to pass through different α-like orbits, and finally, all bunches were guided to the same orbit after the superconducting coil. After delivery by the fixed gantry shown in Fig. 7, the proton bunches should be scanned specifically before it reaches the patient's couch. Because one full FLASH treatment cannot be scanned within 100 ms using the conventional scanning coil, a new proton bunch scanner developed from an RF deflector with variable polarization should be adopted, as will be presented in the following section.

1.5 Fast bunch scanner

An important equipment used in proton therapy for proton beam scanning is the beam scanner for *X*- and *Y*-direction scanning. Unfortunately it is impossible to complete one FLASH treatment within 100 ms by using two conventional scanning coils. According to the operation mode in Fig. 1, it's indicated that at most 10 μs can be spared to complete one layer scanning comprising approximately 1000 points, which can typically form a 10 cm × 10 cm scanning area by the proton beam with a transverse dimension of 3 mm (RMS) at the isocenter; therefore, one FLASH treatment can be completed in 100 ms. To improve the speed of beam scanning, a new beam scanner developed from an RF deflector with variable polarization, integrating simultaneous *X*- and *Y*-direction scanning, is proposed based on a scheme similar to that of a fast beam kicker. The fast beam scanner was designed with a filling time of approximately 10 ns; therefore, one scanning layer can be completed in less than 10 μs under a fast LLRF system. Currently, the pencil

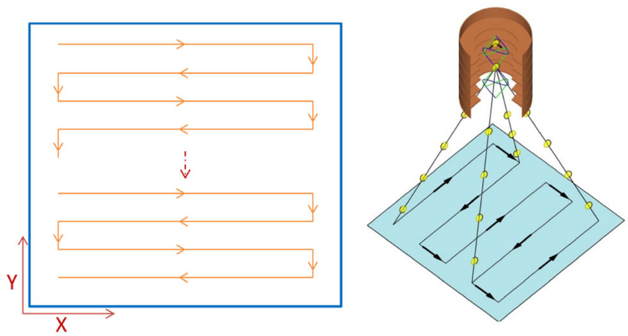

Fig. 8 (Color online) Continuous path scanning by fast beam scanner in one layer for FLASH treatment

beam is scanned point by point in most proton therapy. Therefore, additional time as long as 1 ms is required to stabilize each point location; in other words, a total duration of approximately 1 s is required to stabilize 1000 points in one scanning layer. Therefore, point scanning is unavailable to support the FLASH treatment by the fast beam scanner, except by continuously scanning the raster, as shown in Fig. 8.

Based on the beam scanning method shown in Fig. 8, at any time, the RF beam scanner scans in only one direction. This should be easily implementable using an RF deflector with variable polarization, under the control of a fast LLRF system. The scanning path begins with X-direction scanning from left to right, followed immediately by a Y-direction step jumping to the lower path. The X-direction scanning restarts from right to left and then jumps again to the lower path after arrival on the left side. All scanning paths are repainted as the scanning type above the rest path of the scanning layer and finally stop at the end of the scanning layer. The full beam scanning procedure is controlled by LLRF technology, and ultra-fast beam scanning can be realized in one direction within submicroseconds. Therefore, one full layer can be completed within 10 µs.

2 Conclusion

A compact proton therapy facility based on a high-gradient proton linac for FLASH treatments is proposed herein. This new proton therapy facility involves no mechanical motion and no magnetic field variation in its gantry system, and it can realize a FLASH of up to 300 Gy/s, thereby delivering a 30 Gy dose to a tumor in less than 100 ms for one full FLASH treatment.

The new design for proton therapy facilities comprises the following four crucial components for realizing FLASH treatments:

1. The proton linac is designed to operate at a repetition rate of 1 kHz; therefore, it can produce 100 scanning layers within 100 ms, in which the irradiation pulse length for each layer can be 10 µs or less. All proton bunches for one full FLASH treatment can be generated and delivered by the proton linac within 100 ms.
2. The fast kicker based on an RF deflector with variable polarization is a new type of equipment to switch proton bunches as a connection between the linac and gantry. The kicker is controlled by an LLRF system with a much faster switch rate than conventional pulsed magnets; in particular, the fast kicker is designed to kick proton bunches to different solid angles, including both azimuth and elevation angles. Based on this fast kicker, all bunches for one full FLASH treatment can be delivered from the linac to the gantry within 100 ms or less.
3. Comprising 12 superconducting coils, the gantry is fully static without any mechanical rotation or magnetic field variation. The gantry can accept and deliver all proton bunches with different energies and then deliver all bunches to the same orbit before the fast beam scanner.
4. The fast scanner is designed based on an RF deflector with variable polarization instead of a magnetic coil. Under the control of a low-level system such as a fast kicker, it can complete the irradiation of one layer scanning within 10 µs or less.

The conceptual design presented herein suggests that the FLASH treatment is achievable. However, many requirements of advanced technology beyond the current technological capability must be fulfilled to achieve new breakthroughs in proton therapy.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Wen-Cheng Fang, Xiao-Xia Huang, Jian-Hao Tan and Yi-Xing Lu. The first draft of the manuscript was written by Wen-Cheng Fang, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

1. V. Favaudon, L. Caplier, V. Monceau et al., Ultrahigh dose-rate FLASH irradiation increases the differential response between normal and tumor tissue in mice. Sci. Transl. Med. **6**, 24593 (2014). https://doi.org/10.1126/scitranslmed.3008973
2. A. Patriarca, C. Fouillade, M. Auger et al., Experimental set-up for FLASH proton irradiation of small animals using a clinical system. Int. J. Radiat. Oncol. Biol. Phys. **102**(3), 619–626 (2018). https://doi.org/10.1016/j.ijrobp.2018.06.403
3. P. Montay-Gruel, A. Bouchet, M. Jaccard et al., X-rays can trigger the FLASH effect: ultra-high dose-rate synchrotron light source prevents normal brain injury after whole brain irradiation in mice. Radiother. Oncol. **129**, 582–588 (2018). https://doi.org/10.1016/j.radonc.2018.08.016

4. J. Bourhis, M.C. Vozenin, J.F. Germond, C Bailat, et al. Clinical perspectives for FLASH Therapy. VHEE2020, CERN.
5. Particle Therapy Cooperative Group (PTCOG) Collaboration, http://www.ptcog.com.
6. C.C. Xiao, J.Q. Zhang, J.H. Tan et al., Design and preliminary test of the LLRF in C band high-gradient test facility for SXFEL. Nucl. Sci. Tech. **31**(10), 100 (2020). https://doi.org/10.1007/s41365-020-00806-6
7. H.Y. Li, X.M. Wan, W. Chen et al., Optimization of the S-band side-coupled cavities for proton acceleration. Nucl. Sci. Tech. **31**(3), 23 (2020). https://doi.org/10.1007/s41365-020-0735-7
8. S. Benedetti, A. Grudiev, A. Latina, High gradient linac for proton therapy. Phys. Rev. Accel. Beams **20**, 040101 (2017). https://doi.org/10.1103/PhysRevAccelBeams.20.040101
9. M. Vretenar, A. Dallocchio, A compact high-frequency RFQ for medical applications, in Proceeding of 27th Linear Accelerator Conference, LINAC2014: Geneva, Switzerland, August 31 September 5, 2014, pp. 935–938.
10. R. Akre, L. Bentson, P. Emma et al., A transverse rf deflecting structure for bunch length and phase space diagnostics, in PACS2001. Proceedings of the 2001 Particle Accelerator Conference (Cat. No. 01CH37268), Chicago, IL, USA, 2001, pp. 2353–2355. https://doi.org/10.1109/PAC.2001.987379.
11. J.H. Tan, W.C. Fang, Z.T. Zhao et al., Two-mode polarized traveling wave deflecting structure. Nucl. Sci. Tech. **26**(4), 040102 (2015). https://doi.org/10.13538/j.1001-8042/nst.26.040102
12. L. Bottura, E. Felcini, G. De Rijk et al., GaToroid: a novel toroidal gantry for hadron therapy. Nucl. Instum. Methods A **983**, 164588 (2020). https://doi.org/10.1016/j.nima.2020.164588