1 A Review of Nonscaling CW FFAs for Proton and Ion Therapy Applications

Carol Johnstone, Fermilab, Batavia, IL, USA
Mail to: cjj@fnal.gov

Reinhard Schulte, Loma Linda University, Loma Linda, CA
Mail to: rschulte@llu.edu

Abstract

Significant progress on compact, variable-energy versions of non-scaling fixed field accelerators with alternating strong-focusing gradients (nsFFA) has been made and adapted to proton and ion therapy. Not only has isochronous (CW) capability been demonstrated in a realizable design, an isochronous racetrack format has evolved which supports lengthy, synchrotron-like straight sections. Long straight insertions promote low-loss injection and extraction systems and further facilitate extracting lower-energy orbits using a bipolar bump-magnet system. To efficiently extract variable energy using this method, the accelerator complex requires separated accelerator stages, each stage with limited but therapeutically optimized energy ranges targeting specific cancer types and penetration depths. Different energy stages further support multiple treatment rooms increasing patient throughput and reducing treatment cost. The staged system proposed here also realizes cost savings in gantries by tailoring delivery to different therapeutic energies and beam requirements, promoting the requirement for only one high-energy gantry in the facility. The proposed facility consists of 3 stages, a 30 MeV injector, a 30 - 90 MeV/nucleon FFA for shallow cancers, and 90 - 250/330 MeV/nucleon stage for deep tumors, pelvis for example. The higher energy of the final stage, 330 MeV/nucleon is preferred to support Proton Computed Tomography. Isochronous FFA stages can be designed to accelerate either protons or ions with charge to mass of $1/2$.

1.1 Introduction

Cancer is the second-largest cause of death in the U.S. with about half of all patients receiving definitive radiation therapy; overall approximately two-thirds of all cancer patients will receive some form of radiation therapy. The majority of radiation treatments are still performed with linacs that generate energetic electron beams directed onto targets producing secondary X-rays for photon therapy. Particle beam therapy using proton and ion beams, however, have better dose specificity and conformal deposition compared with conventional radiotherapy and has rapidly evolved into a frontier in cancer therapy over the last decade. Clinical experience with protons and ions has produced remarkable local tumor control rates in single-institution studies.

However, despite the promise of gross reduction in acute and late effects and overall quality of life after treatment, the expansion of proton and especially ion therapy centers has been limited by initial capital cost, lengthy treatment course, insurance reimbursement schedule issues, and overall economic sustainability of centers, especially in the competitive U.S. environment. This paper addresses limitations of current accelerator systems and single-purpose centers for therapy, and proposes a more economically sustainable, multipurpose center model, in addition to reviewing technical advances in accelerators and beam delivery represented by recent developments in nonscaling FFA technology.

---

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.
1.2 Dual Accelerator Particle Therapy Model

To address non-competitiveness with photon therapy, multiple applications and sources of revenue must be incorporated into the hadron therapy center, optimally at the civil planning stage. As discussed in the DOE report “Accelerators for America’s Future” [1], the key to further progress in cancer therapy is through research and development. A sustaining facility “would enable biological studies; development of imaging technologies; exploration of scanning and controls for safe, flexible beam delivery; investigation of small-mass dosimetry instrumentation; and, perhaps most important, hosting of clinical research, clinical trials, and advanced patient protocols for newly developed technologies.” In 2013 NCI jointly with DOE organized a workshop on ion beam therapy where more than 60 experts from diverse fields related to radiation therapy were asked to define research and technical needs for advancing charged particle therapy, producing a detailed final report [2]. Table 1 summarizes technical accelerator and beam delivery specifications derived from this report. In summary, in addition to progress in technical systems, a sustainable center needs to support research in advanced protocols to reduce overall cost of treatment course (without compromising clinical outcomes), a plan for particle-based imaging and dose verification dosimetry, and have the capability to deliver high intensities ($\geq 20$ Gy/min) for radiobiology and hypofractionation research – in addition to including opportunistic commercial revenue streams.

To effectively stage a facility for an economically optimum therapy program, it is critical to first assess therapeutic needs in terms of beam parameters and delivery technology. Table 2 [3] is a model-based treatment using 70 MeV as the lower limit since many nozzles (or energy degraders) only work in the 70 to 250 MeV range. (Lower energies are generally obtained using plastic range shifters placed close to the skin and aperture, for example with breast, pediatric patients, and parotid tumors in the jaw.) The data clearly show a breakdown of cancer by beam energy and argues for a dual energy accelerator system as optimal - where $E \leq 150$ MeV can be used for roughly 50% of the patients. Higher energy is only required for deep tumors such as prostates. Two accelerator stages are therefore optimal for proton therapy. A dual-stage accelerator model has the following applications and advantages.

- Conclusions from patient data
  - 1st Accelerator Stage: $\sim 30$ to ($\sim 90/150$) MeV/nucleon for ions/protons
    Lung, breast, CNS, rectum, pediatric, head & neck : $\sim 50\%$ of patients
  - 2nd Accelerator Stage: ($\sim 90/150$) to ($250/330$) MeV/nucleon for ions/ protons
    Deep tumors, pelvis, and prostate

- Gantry costs
  - Only one high-energy gantry needed
  - Lower-cost, much smaller gantry delivers beam from 1st stage
  - Solves problem of oversubscribed high energy gantry

- Patient Throughput and efficiency
  - Allows simultaneous operation of $\geq 2$ treatment rooms
  - Multi-room operation will reduce treatment costs

1.3 Radioisotope Production in a Therapy Facility

In addition to radiotherapy, an economic opportunity for medical radioisotope production can be implemented exploiting the infrastructure and injector accelerator at incremental additional cost to the overall facility. Radioisotopes can provide substantial income and an ongoing revenue stream to support operational and maintenance overheads for the entire facility, well in excess
Table 1: Ion therapy facility clinical requirements from the NCI/DOE report [2] are summarized.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-ion capability</td>
<td>p, He, Li, B, C (O and Ne also desirable)</td>
</tr>
<tr>
<td></td>
<td>Fast switching between ion species (≤ 1 sec)</td>
</tr>
<tr>
<td>Energy range</td>
<td>60 MeV/nucleon to 430 MeV/nucleon for ions; 250 MeV for protons. Depths up to 30 cm for carbon ions</td>
</tr>
<tr>
<td>Field size</td>
<td>At least 20 × 20 cm² (optimally up to 40 × 40 cm²)</td>
</tr>
<tr>
<td>Real-time imaging (radiography and CT):</td>
<td>For tumor position verification and motion control</td>
</tr>
<tr>
<td></td>
<td>For patient sizes up to 60 cm in depth. (For protons this is 330 MeV)</td>
</tr>
<tr>
<td>Dose delivery rates:</td>
<td>Minimum requirement</td>
</tr>
<tr>
<td></td>
<td>Hypofractionation treatments in under 1 minute, (ideally in one breath-hold for motion control)</td>
</tr>
<tr>
<td></td>
<td>20 Gy/minute up to 5-8 Gy/8 sec (breath-hold) for a cubic liter</td>
</tr>
<tr>
<td></td>
<td>(corresponding to 4 × 10¹² p/sec)</td>
</tr>
<tr>
<td>Pencil beam scanning:</td>
<td>Transverse scanning rate of 1-10 cm/msec</td>
</tr>
<tr>
<td></td>
<td>Energy step time of 10-100 msec (These are present state-of-the-art for NC and SC magnetic components)</td>
</tr>
<tr>
<td>Transverse beam size:</td>
<td>3 mm to 10 mm FWHM</td>
</tr>
<tr>
<td>Energy step size</td>
<td>Protons: 2 MeV (~ 0.25 cm in range)</td>
</tr>
<tr>
<td></td>
<td>Carbon: 2 MeV/nucleon (~ 0.1 cm in range)</td>
</tr>
<tr>
<td>Lateral targeting accuracy at the Bragg peak</td>
<td>Protons: ±0.5 mm</td>
</tr>
<tr>
<td></td>
<td>Carbon: ±0.2 mm</td>
</tr>
<tr>
<td>Dose accuracy/fraction</td>
<td>2.5% monitored at ≥ 40 kHz during dose deposition</td>
</tr>
</tbody>
</table>

Table 2: Proton therapy data from Paul Scherrer Institute categorized per cancer (based on discussions with G. Coutrakon, 2009.)

<table>
<thead>
<tr>
<th>Site</th>
<th>Percentage</th>
<th>Energy Range (MeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung</td>
<td>9%</td>
<td>70 - 170</td>
</tr>
<tr>
<td>Breast</td>
<td>3%</td>
<td>70 - 140</td>
</tr>
<tr>
<td>CNS</td>
<td>15%</td>
<td>70 - 150</td>
</tr>
<tr>
<td>Pediatric</td>
<td>8%</td>
<td>70 - 150</td>
</tr>
<tr>
<td>Head &amp; Neck</td>
<td>15%</td>
<td>70 - 150</td>
</tr>
<tr>
<td>Prostate</td>
<td>45%</td>
<td>200 - 250</td>
</tr>
<tr>
<td>Other</td>
<td>2%</td>
<td></td>
</tr>
</tbody>
</table>

of any added impact. Production of radioactive isotopes has many applications with the most direct benefits realized in medical diagnosis and therapy - impacting the quality of life for therapy patients [1, 4]. The benefits of expanded availability of key isotopes domestically are considered substantial and even critical. Such a capability provides added value in that it is aligned closely
with the interests of the oncology user community.

One example is Radioimmunotherapy (RIT) with radiolabeled antibodies, a promising, newly emerging cancer treatment modality that has achieved response rates of up to 80 percent in a limited number of cancers. RIT selectively delivers radionuclides that emit \(\alpha\)-particles, \(\beta\)-particles, or Auger electrons to tumors. The isotope group of the Nuclear Science Advisory Committee (NSAC), recognizing the gap between production and demand particularly of \(\alpha\)-particle-emitters, advises in their long-range plan that the United States should “invest in new production approaches of \(^{211}\text{At}\) emitters with high priority for \(^{211}\text{At}\) and \(^{225}\text{Ac}\)” [5]; these are produced with an ion and proton beam, respectively.

Specifically, for RIT, production of \(^{225}\text{Ac}\) and \(^{211}\text{At}\) is considered a top priority for the DOE isotope program. The focus of R&D and one of the main programmatic goals of the DOE Isotope Program therefore is improving and making commercially feasible production of \(\alpha\)-emitters \(^{225}\text{Ac}\) and \(^{211}\text{At}\) in addition to increasing their availability. Both rare isotopes can be produced in an ion injector stage with 30 MeV/nucleon and high currents - hundreds of microamps. A separated sector with high-gradient cavities would support the highest, lowest-loss currents and extraction to a target system. A separated sector would also support internal targets. Features of integrated isotope production are listed below.

**Integrated Radioisotope Production**

- **“Dual-function” Injector:**
  - Injector to 1\(^{st}\) treatment accelerator stage
  - \(~0.5 - 30\) MeV/nucleon (100 microamps)
  - Radioisotope production (outside of treatment hours)

- **Internal targets select optimal energy for production**
  - Examples:
    - \(\text{H}_2^+\) ion: PET isotopes, \(^{99}\text{Mo}\) (\(^{99}\text{Tc}\), FDG
    - \(\text{He}^2^+\): PET isotopes
    - \(\text{He}^2^+\): alpha emitters (\(^{211}\text{At},^{225}\text{Ac}\) for radioimmunotherapy)

**1.4 Accelerator Options for Therapy**

**Background**

Cyclotrons dominate proton therapy in part due to fixed rather than ramped magnetic fields along with fixed-frequency RF systems, which simplifies machine cost and operation, but mainly due to their ability to deliver variable-intensity, continuous (CW) beam in a compact format. Continuous beam delivery holds a distinct advantage from an operational and treatment standpoint because it not only simplifies accelerator systems, operational overhead, and beam control, it further reduces technical complexity and response times required for accurate dose delivery and real-time verification – impacting critical dosimetry systems and integrated dose control during treatment. CW beam is desirable for pencil beam scanning (PBS) [6], and is a requirement for compatibility with Proton Computed Tomography technology and image reconstruction [7]. Proton Computed Tomography (pCT) has noted advantages in imaging that include lower dose and enhanced tissue identification [8]; current estimates project a factor of 10 less dose to patient for equivalent resolution compared to conventional CT. Reduced dose indicates adaptive therapy could be implemented if pCT were available in proton therapy facilities, with corresponding improved local control over the tumor and more conformal dose distribution. However, pCT
requires a maximum energy of 330 MeV protons to fully transit all regions of the human anatomy and current iso-cyclotrons do not presently support this energy: present compact isochronous cyclotrons for proton therapy reach only 250 MeV as required for treatment, but too low for full pCT capability. Even if a 330 MeV iso-cyclotron were available, degrading the higher fixed-energy output beam to therapy energies is not an attractive solution requiring much higher output primary currents, increased beam emittances, and residual activation. Imaging in general is inherently challenging as variable energy is required combined with precision intensity control (single-proton buckets) – and the degrader complicates both for a cyclotron-based facility.

The ultra-compact synchro-cyclotrons, such as the new S2C2 offered by IBA, are also common therapy machines specifically optimized for PBS (with kHz RF frequency sweep times). However, variable energy still requires a degrader and the single-bucket proton pulse delivered at kHz rates is incompatible with practical imaging systems. The synchrotron can attain the higher energies, but again the low duty cycle does not support useful imaging reconstruction times, nor is the synchrotron slow-spill easily adapted to ultra-low current. Conventional CT scans take 90 seconds and only isochronous accelerators are competitive. Given all the advantages of CW beam in particle therapy and imaging, the next section reviews the challenges of maintaining isochronism in higher-energy proton and ion accelerators while retaining a compact footprint.

**Compact isochronous Accelerators**

Compact medium to high energy proton and ion accelerators are a challenging, critical technology. Historically, cyclotrons have been the most compact accelerator technology, but only at energies up to a few hundred MeV. As the energy increases, stronger acceleration is required to minimize beam losses and radiation during acceleration and especially during beam extraction. Orbit separation decreases with energy so higher-energy machines require increased acceleration gradients to provide sufficient radial separation between beams that comprise different acceleration turns in order to efficiently extract. A common solution is to separate the magnetic sectors to insert higher-gradient RF structures. However, once space is inserted between the magnetic sectors of the cyclotron, the footprint of the cyclotron grows rapidly. The compromise to retaining compactness in cyclotrons is beam loss and activation during extraction - which can vary from 20% for the 250 MeV Varian cyclotron (used at PSI) and up to 60% for the IBA 230 MeV Cyclone. Losses present operational, maintenance, and personnel exposure issues even at low therapy currents. The vault shielding required is significant and increases rapidly with energy – 15’ is typical of proton cyclotron vaults and projected to be ~ 23’ for ion cyclotron vaults. For a compact AVF iso-cyclotron that provides a 330 MeV beam, degrading to the lowest treatment energies (~150 MeV and lower) would likely be prohibitively lossy. Even with the present 230/250 MeV output energy, beam transmission and delivery at the lower energies is a fraction of a percent combined with significantly higher associated emittances. The inherent limitations of the cyclotron at higher energies and the possibility of variable energy without a degrader has stimulated interest in the potential of FFAs.

Towards this end, the concept of isochronous orbits coupled to constant machine tune has been explored and developed for the most general type of FFA (termed non-scaling) using powerful new methodologies in fixed-field accelerator design [9–11]. The property of isochronous orbits again enables the simplicity of fixed RF and by inference, CW operation, as in the cyclotron, but with strong focusing. By tailoring a nonlinear radial field profile, the FFA can remain isochronous with stable and strong machine tunes, well into the relativistic regime achieving 330 MeV for protons and 430 MeV/nucleon for ions - maintaining the constant dose characteristic of cyclotrons. Specifically, isochronous compact racetrack designs have been developed for therapy and security applications with the advantage that lower energy beams can be ef-
ciently extracted for patient treatment without changes to the acceleration cycle and ring magnet strengths. The long synchrotron-like opposing straight sections can be exploited for efficient injection and extraction in addition to providing for rapid energy variation, which has also been explored in other nonscaling FFA designs [12]. The methodology used to design FFA isochronous dynamics is discussed next.

1.5 Isochronous Nonscaling FFAs

Isochronous Dynamics in an FFA

The weak-focusing nature of traditional cyclotron fields does not permit long (several meters) straight sections without a significant scaling up of machine radius and size. However, the addition of strong focusing gradients (and corresponding strong beam envelope control) to conventional cyclotron fields – including reversed gradients to capture both transverse planes – does allow insertion of long synchrotron-like straight sections and thus efficient implementation of high-gradient, multiple-cavity RF modules, even SCRF cryomodules. Further, the nonscaling nonlinear FFA designs have evolved into a very compact racetrack shape – essentially a recirculating linear accelerator with FFA arcs. This basics of this new generation of ultra-compact nonscaling FFAs with constant machine tunes are described here. First an overview of the dynamics of FFAs with alternating strong-focusing gradients is reviewed and compared with cyclotrons and synchrotrons, highlighting the potential for isochronous orbits with stable synchrotron-like performance at high energies.

In addition to isochronous central orbits, a key dynamics issue is resonance avoidance nominally through stable, approximately constant machine tunes over the entire acceleration energy range – which can be accomplished in a FFA by applying alternating strong-focusing gradients. Conventional isochronous cyclotron design (both sector and spiral sector) cannot maintain isochronous orbits and stable tunes simultaneously at relativistic energies. Nonscaling FFAs with optimized nonlinear field profiles, however, are capable of both; further they exhibit the strong-focusing machine tunes, tune footprints, and space-charge tune shifts characteristic of synchrotrons. The potential for high-intensity operation and strong-tune tolerance of space charge effects has been reported in published simulations [13].

There are several important dynamical consequences in relativistic cyclotrons that complicate their design, impact the physical parameters of the machine, and limit their energy reach. These are briefly discussed below without derivation. In cyclotrons the azimuthal B field profile does not change with radius that is, the extent of the “hills and valleys” scale in direct proportion to radius. Due to the fixed azimuthal field profile, \(B(\theta)\), the magnitude of the radial B field must increase as \(B = \gamma B_0\), to maintain isochronous orbits, with \(\gamma\) the relativistic factor and \(B_0\) the magnetic field at injection (\(\gamma \rightarrow 1\) at nonrelativistic energies giving the conventional cyclotron).

The increasing peak B field results in a gradient term which changes the machine tune; the tune cannot be held constant in relativistic cyclotrons - it rises radially and decreases vertically. A spiral shape is introduced to increase the vertical edge crossing effect (flutter) with energy to maintain a stable vertical machine tune. Eventually the vertical tune becomes unsustainable. Tune approximations from equation (1) give \(\nu_r \approx \gamma\) and \(\nu'_z \approx 1 - \gamma^2 + F(1 + 2\tan^2 \epsilon)\) where \(F\) is the flutter and \(\epsilon\) the spiral angle.

\[
F = \left( \frac{B(\theta) - B_{av}}{B_{av}} \right)^2
\]

(1)

where \(\epsilon\) defines the axis \(R = R_0 \epsilon^\theta \cot \epsilon\).

As \(\gamma\) increases, \(\nu_r\) increases and \(\nu'_z\) decreases. The machine tune changes and crosses betatron resonances with potential for beam blowup and losses. Stable beam properties, dynamics
and low-loss operation can in general not be achieved in high-energy cyclotrons unless the acceleration is strong enough to jump across resonances and separate orbits sufficiently for extraction (alternatively, the machine size can increase significantly decreasing the gradient, hence the large size of the 590-600 MeV cyclotrons at PSI and TRIUMF).

The dynamics of FFAs, both scaling and nonscaling are dominated by synchrotron-like dynamics [14]. In a FFA with alternating gradients all conventional focusing terms are utilized as given in the following thin-lens approximation:

$$\frac{1}{f_F} = k_F \ell + \frac{\theta}{\rho_F} + \frac{\eta}{\rho_F},$$

where $f_F$, $\rho_F$ in equation (2) are the focal length in the horizontal plane and bend radius, respectively, $\ell$, the horizontally-focussing magnet half length, $k_F$ the “local” horizontally focusing gradient for an arbitrary field order, $\theta$ the sector bend angle, $\eta$ the edge crossing angle (the tangent is approximated). In the vertical (non-bending) plane, the sector bend term does not contribute, and the vertical tune depends on only the strong-focusing and the edge-focusing terms. This simple thin-lens expression highlights how the three machines fundamentally operate and determines their ultimate flexibility in format and machine characteristics.

The synchrotron relies on the first strong-focusing gradient term, the cyclotron on the last two terms which are the centripetal (weak focusing, bend plane only) and edge focusing terms, but the FFA utilizes all three terms for beam envelope control and dynamical stability. Further, unlike the scaling version, the non-scaling FFA optimizes the normal and reverse-field gradients and edge angles independently; this is critical for achieving compactness and simultaneous isochronous (CW) dynamics. This last point is very important because individual control over these parameters allows the field, orbit location, and important machine parameters such as tune, footprint, and aperture to be approximately independent and more strongly controlled than in a cyclotron. The reverse gradient is especially critical to a stronger vertical tune than can be achieved in a cyclotron. This increase is not solely due to the addition of strong focusing, but also from the radial increase in edge focusing strength (due to increasing magnetic field). In the non-scaling FFA (only), these different focusing principles can be combined in different and varying relative strengths through the acceleration cycle – this varying composition can be exploited to control the machine tune without applying the field scaling law. The terms can be further inter-played and optimized to achieve stable dynamics in highly flexible and ultra-compact lattice structures.

The design of the non-scaling FFA with its inherent flexibility is even more powerful when a high-order nonlinear field expansion is applied. The nonlinear gradient has the advantage of providing increased focusing in both transverse planes as a function of radius – more than a linear field gradient – thereby providing a strong constant tune in both planes while preserving considerable freedom in physical parameters. When components are designed with independent normal and reverse gradients, the orbital path length can be constrained such that the revolution time at each momentum scales with velocity, and simultaneously the machine tune can be controlled through edge and weak focusing effects independent of this path length – thus impacting tune but not revolution frequency. Expressed in terms of linear dynamics, a focusing “quadrupole-like” gradient is generated locally about each isochronous orbit as a feed-down from the nonlinear radial field profile; by following a nonlinear radial field profile a FODO-like cell tune is controlled and maintained as a function of energy. The terms and field-order in the nonlinear radial magnetic profile must be matched to the isochronous orbit which is dictated by the central field of and trajectory length through the magnet.

Unlike the cyclotron, which relies on a predominately dipole field or fixed $B$-field scaling with $\gamma$, and is therefore limited in adapting path length to velocity as the energy becomes relativistic, the non-scaling FFA can maintain isochronous orbits well into relativistic energy regimes as
shown in Figure 1.

Only the nonscaling FFA can be CW as there is no flexibility in the scaling law for scaling FFAs. Further, a linear-field nonscaling FFA cannot maintain constant tune, both horizontally and vertically, as a function of energy since relativistic effects are inherently nonlinear. Only a nonlinear gradient can therefore provide increased focusing in both transverse planes stably as a function of energy through the radially-increasing gradient, edge, and centripetal focusing terms. The order of the expansion provides the degrees of freedom necessary to maintain a strong constant tune in both planes, decoupling it from and preserving considerable freedom in physical parameters. Strong tunes, especially needed in the vertical, imply mitigation of space-charge effects and stable acceleration to potentially higher currents than the cyclotron for the same footprint.

Based on optimized nonlinear field profiles, several FFA designs with alternating gradient magnets are presented in this report as follows with emphasis on the racetrack format.

1. 0.2 - 1 GeV proton circular and racetrack variant;
2. 30 - 90 MeV racetrack FFA for therapy;
3. 90 - 250/330 MeV racetrack FFA for deep tumor therapy and pCT.

The 1 GeV proton racetrack is relevant because field strengths and layout for both superconducting and normal conducting versions correspond in physical size and field strengths to a 400 MeV/nucleon ion therapy accelerator. (The applications of 1 GeV protons are irradiation and security applications.)

**CW nonscaling FFA Designs**

The nonlinear design principles described above have been applied utilizing advanced design algorithms to produce a 200 MeV to 1 GeV quasi-isochronous circular FFA with a small footprint ($dt/t = \pm 3.5\%$) [15] and utilizing superferric magnetic fields of 5 T or less. The machine footprint is shown in Figure 2 (left), with four 2 m straights and each cell composed of an FDF triplet. Initially all FFA designs were completely periodic. As the demands for compactness increased and higher energies were requested for protons (1 GeV) and ion therapy (430 MeV/nucleon), eventually all inter-component straight sections became too short for effective extraction or even injection. The number of straights also increased with energy as periodicity was required to increase causing an undesirable increase in footprint. Several of these initial circular FFAs for medical therapy and imaging have been designed and documented [16] but will not be discussed further in this report in preference to the racetrack format. The racetrack configuration optimizes compactness yet provides for RF modules to be inserted in one straight and injection extraction systems in the opposite straight— an optimal hardware and overall layout configuration. To this end we have designed a racetrack variant that initially eliminated two of the straights in a ring with fourfold symmetry and lengthened the remaining two opposing straights to preserve the total length of straight section as shown in Figure 2 (middle). This preserved both the circumference and the machine tune permitting a dynamical comparison between the circular and racetrack designs. The beam orbits are identical through the common bend or arc structures.
and are shown in Figure 2 (right) for the range of machine energies. These design orbits were computed with COSY [17].

Figure 2: Lattices for 200 MeV to 1000 MeV protons circular and full racetrack FFA layouts. (left, right) shows the normal (blue) and reverse gradient (red) components for a periodic compact FFA and one with two long straight insertions on opposing sides (middle). Particle tracks through the FDF bend structure (right) are identical for both versions since straights are inserted at symmetry points in both lattices.

The ring tunes for the ring and the racetrack lattice versions, as computed from the initial hard-edge design lattice using a fringe field model based on Enge functions and simulated in PyZgoubi [18] are shown in Figure 3. Note the tunes are essentially constant with energy in the horizontal but have some vertical variation that will be addressed in future work. Interestingly, the lattices are almost identical in machine tunes and dynamics if the circumference, focusing strength, and total straight section lengths are preserved. Tracking simulations for both versions are discussed next.

Figure 3: Machine tunes for the 2m ring (left) and a racetrack nsFFA (right) as a function of beam energy.

Tracking

Given that a nonlinear field profile is required for all compact, fixed magnetic field accelerators to achieve high energies (including synchrocyclotrons), efficient beam transmission is an overriding concern, so high-order simulation was initiated again using new methodologies in fixed-field accelerator modeling [9–11]. Figure 4 depicts the tracking pictures at high-energy using COSY[17] for singe-particle tracking. In the FFA designs, both circular and racetrack, the geometric emittance was found to be preserved as a function of energy, apparently due to the approximately constant tune and avoidance of resonances. When geometric emittance is preserved, the dynamical acceptance or DA (which is a strong indicator of robust approximately
linear dynamics about reference orbits) actually increases and beam transmission appears predicted to be near 100%. The elliptical phase space portraits at the highest nonlinear gradient reference orbit of 1 GeV in Figure 4 support this observation. This is in contrast to the DA predicted for the 8-sector version of an 800 MeV cyclotron [19] which decreases with energy in simulations. Comparison of tracking results indicates about a factor of 4 increase in DA for the racetrack FFA described in this work combined with a factor of 4 decrease in footprint when compared to the published tracking results for this specific cyclotron design [19].

![Figure 4: Single-particle tracking in COSY INFINITY showing results for the 1 GeV reference orbit in the racetrack nsFFA. The orbits represent 1 cm steps in the horizontal and 1 mm in the vertical. The aperture limit is reached in the vertical.](image)

**Dynamic Aperture**

We now consider the dynamic aperture of these FFAs and present an algorithm, implemented in PyZgoubi, for the calculation. The presence of nonlinearities in the accelerator lattice can give rise to unstable motion, and over a very large number of particle turns through the accelerator a particle’s motion may become unbounded. The approximate boundary between stable and unstable motion, and the particle amplitude at which this boundary occurs, is termed the dynamic aperture (DA). The period of time over which particles must be followed in simulation to predict the dynamic aperture depends upon the application of that accelerator. For the ultra-compact FFAs discussed here if implemented with SRF, then the protons remain in the accelerator for only around 40 turns, but with NCRF, as in the PSI cyclotron, the number of turns increase by an order of magnitude.

Dynamic aperture is expressed here in terms of the limit in transverse particle amplitudes or, equivalently, the limit in transverse single particle actions $J_{x,y}$ over the number of acceleration turns and is calculated by increasing $J_x$ and $J_y$ until the particle is lost (particle loss defined by the actual physical apertures). We therefore consider increments of $J_x$ and increments of $J_y$ or some combination of both defined the angle in the $(J_x, J_y)$ plane. For the large horizontal and small vertical apertures of fixed-field machines and the midplane symmetry, the DA is best described in terms of these two planes. Single-particle trajectories are generally elliptical at small amplitudes as the dynamics are approximately linear locally about a closed reference orbit, but become distorted and non-elliptical at larger amplitudes as nonlinear magnetic fields cause increasingly nonlinear motion. Hence we compute the dynamic aperture by combinations of particle coordinates based on the particle action e.g. we increase $J_x$ by increasing the single particle coordinates along the four axes $(+x, 0), (0, +x'), (-x, 0)$ and $(0, -x')$ and similarly for $J_y$. The DA has been computed in the horizontal and vertical using this method over 40 and 200 turns to represent a strong gradient, SC RF version and the impact of a slower acceleration.
cycle, respectively. To model a more realistic magnet configuration, the DA is simulated using the PyZgoubi framework, with a series of random horizontal misalignments applied to all magnetic elements to study the DA reduction from a perfect lattice. The misalignments are modelled as a Gaussian distribution cut off at $3\sigma$. The DA is shown as the mean over the seeds, with an error bar corresponding to the RMS variation of the seeds. Table 3 shows the DA achieved for the 2 m circular and racetrack versions of the nsFFA in the horizontal and vertical planes computed with PyZgoubi and an Enge function fringe field. Figure 5 shows the fractional decrease of dynamic aperture with respect to the ideal case for the circular 2 m FFA in terms of normalized stored emittance as a function of misalignment width (left). The DA shows a slow decline for misalignment widths up to 600 $\mu$m misalignments. The same calculation for the racetrack is shown in Figure 5 (right) and shows a slightly faster decline of DA with misalignment width. However, both machines are surprisingly tolerant to misalignments in the horizontal plane given the nonlinear field profile of these machines.

Table 3: The dynamic aperture in unnormalized emittance units for the circular FFA and the racetrack FFA computed with PyZgoubi.

<table>
<thead>
<tr>
<th>Dynamic aperture</th>
<th>Circular FFA</th>
<th>Circular FFA</th>
<th>Racetrack FFA</th>
<th>Racetrack FFA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Horizontal</td>
<td>Vertical</td>
<td>Horizontal</td>
<td>Vertical</td>
</tr>
<tr>
<td></td>
<td>$\pi$ mm.mrad</td>
<td>$\pi$ mm.mrad</td>
<td>$\pi$ mm.mrad</td>
<td>$\pi$ mm.mrad</td>
</tr>
<tr>
<td>200 MeV 40 turns</td>
<td>74700</td>
<td>107</td>
<td>25000</td>
<td>46</td>
</tr>
<tr>
<td>200 MeV 200 turns</td>
<td>72600</td>
<td>91</td>
<td>23700</td>
<td>43</td>
</tr>
<tr>
<td>800 MeV 40 turns</td>
<td>156300</td>
<td>364</td>
<td>63000</td>
<td>277</td>
</tr>
<tr>
<td>800 MeV 200 turns</td>
<td>155100</td>
<td>356</td>
<td>63000</td>
<td>223</td>
</tr>
</tbody>
</table>

Figure 5: The relative drop in dynamic aperture in percent as a function of misalignment for the circular FFA (left) and racetrack FFA (right) computed with PyZgoubi. Both show stability and similar performance in the presence of large relative 600 $\mu$m misalignments of main magnets.

To summarize, a 0.2-1 GeV nonscaling FFA design indicates a form factor that is 9 m diameter
and a racetrack that is 7 m wide by 11 m long for normal conducting magnets; 6 to 8 combined function magnets and 2 to 4 m long straight sections. Simulations demonstrate that the elongated configuration possesses the same stable beam tunes as the circular FFA configuration. The racetrack accelerators have the potential to approach the low losses of the linac and synchrotron and achieve cleaner separation of beam for extraction given the high gradient acceleration cavities supported. Both designs are fairly insensitive to horizontal misalignments, an important result for achieving a working machine with realistic engineering tolerances. Although this work was for a 1 GeV CW proton accelerator, the field strengths and physical dimensions are similar and can be applied to a 400 MeV/nucleon ion therapy accelerator for a charge to mass ratio of $1/2$ - the nonlinear field radial profile would need to be adapted to different isochronous conditions, as dictated by the relativistic effects and different velocities of the ions traveling similar reference orbits.

1.6 Staging ion Accelerators

The future of particle therapy rests on developing cost-efficient treatment and improved effectiveness; a goal which can for many cancers be accomplished by utilizing higher-RBE ions, and high Linear Energy Transfer (LET) as provided by ion and carbon therapy facilities; in turn the future of versatile light-ion therapy facilities relies critically on the development of compact, operationally manageable accelerator and beam delivery systems that can be implemented economically.

A staged accelerator complex applies to ions as well as protons with a major consideration being variable energy extraction for ions. The use of degraders for changing beam energy becomes even more problematic with ion beams over proton beams given the increased potential for radioisotope production and potential transport of other species during beam delivery [20]. For CW beam a variable energy extraction system requires a system of extraction “bump” magnets and a septum or Lambertson similar to extraction scenarios in synchrotrons. With helium and carbon ions of particular interest, a variable-energy, CW ion accelerator system was designed that reaches an energy of 250/330 MeV/nucleon for acceleration of ions with a charge to mass ratio of $1/2$ – a therapeutic energy for H or the higher energy for imaging. He$^{2+}$ is of specific interest and has a 30 cm range for a treatment energy of 250 MeV/nucleon; however treatment with either protons (in the form of H$^+$) and carbon and other light ions is also supported by this accelerator system. Helium has become of significant interest for pediatric cancers given the reduced multiple Coulomb scattering effects on the lateral penumbra and reduced dose to surrounding healthy tissues relative to proton beams. The long straights of the racetrack format again allow insertion of both high-gradient RF modules and an efficient injection and extraction systems in a highly compact footprint. Variable-energy extraction is achieved by varying the strength of fast-sweeping bipolar “bump” extraction magnets. This approach eliminates the high activation and unavoidable radioisotope products from the degrader-based energy control used in a conventional iso-cyclotron.

A dual-stage approach with a lower energy injector (70–90 MeV/nucleon) is necessary to facilitate a technically feasible variable-energy extraction system which occupies one of the long straight sections in the higher-energy stage. High-gradient fixed-frequency RF cavities (~0.3-0.5 MeV/turn per cavity) would be installed in the opposing straight section. Peak fields in the arc sections of the racetrack will determine the width or shortest dimension of the higher energy stage. The injector can be a simpler accelerator, possibly a cyclotron, although rapid switching between ion sources is important. Below are two tables that represent the goals of optimized ion therapy accelerator and beam delivery design. Table 4 lists the clinically-relevant properties for the accelerator system and Table 5 presents realistic design parameters based on normal-conducting guide fields. A comparison of an equivalent circular layout with the compact
The racetrack form remains important to understand if there are any fundamental performance differences.

Table 4: Clinical specifications for the ion accelerator

<table>
<thead>
<tr>
<th>Description</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-ion Capability</td>
<td>H\textsuperscript{2+}, D\textsuperscript{+}, He\textsuperscript{2+}, Li, C\textsuperscript{6+}; Fast switching time between ions (\sim 1 sec)</td>
</tr>
<tr>
<td>Maximum Energy</td>
<td>330 MeV/\mu</td>
</tr>
<tr>
<td></td>
<td>60 cm for imaging, protons;</td>
</tr>
<tr>
<td></td>
<td>[\sim 23 m for treatment, carbon;</td>
</tr>
<tr>
<td></td>
<td>&gt; 30 cm (250 MeV) for treatment, helium]</td>
</tr>
<tr>
<td>Injection Energy (preliminary)</td>
<td>79 - 90 MeV/\mu</td>
</tr>
<tr>
<td>Dose Delivery Rates</td>
<td>5 Gy/min/cubic liter (minimum)</td>
</tr>
<tr>
<td></td>
<td>Up to 7 Gy/8 sec/cubic liter (hypofractionation)</td>
</tr>
<tr>
<td>Energy Step Size</td>
<td>2 MeV</td>
</tr>
<tr>
<td>Dose Accuracy per fraction</td>
<td>2.5%</td>
</tr>
</tbody>
</table>

Table 5: Machine specifications for the higher energy stage ion accelerator.

<table>
<thead>
<tr>
<th>Description</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Format</td>
<td>Racetrack: 2-fold periodicity</td>
</tr>
<tr>
<td></td>
<td>Comparable circular: 4-fold periodicity</td>
</tr>
<tr>
<td>Beam structure</td>
<td>CW</td>
</tr>
<tr>
<td>Isochronous Level</td>
<td>\sim 10^{-4} - 10^{-5} (initial design 10^{-2} - 10^{-3} only)</td>
</tr>
<tr>
<td>Extraction Energy Range</td>
<td>\sim 70-90 MeV - 330 MeV/u</td>
</tr>
<tr>
<td>Extraction arc radius</td>
<td>\sim 5 m radius</td>
</tr>
<tr>
<td>Maximum B field</td>
<td>&lt; 2.7 T for NC water-cooled magnets</td>
</tr>
<tr>
<td>Acceleration Gradient</td>
<td>1 MeV</td>
</tr>
<tr>
<td>Dose Accuracy per fraction</td>
<td>2.5%</td>
</tr>
<tr>
<td>Nucleon flux</td>
<td>10^{12} u/sec</td>
</tr>
<tr>
<td>Machine reliability</td>
<td>&gt;95%</td>
</tr>
<tr>
<td>Transverse Emittance (\sigma)</td>
<td>\sim 1 \pi mm.mrad (horz and vert, normalized)</td>
</tr>
<tr>
<td>Straight section racetrack</td>
<td>\geq 4 m</td>
</tr>
<tr>
<td>Equivalent circular</td>
<td>\geq 2 m</td>
</tr>
</tbody>
</table>

Designing for He\textsuperscript{2+} presents enormous savings in overall facility cost due to the reduced 250 MeV/nucleon (the same optimal energy per nucleon as in proton therapy but at a lower charge to mass ratio) versus the 430 MeV/nucleon required for 30 cm depth carbon therapy as can be seen from Figure 6 [21]. An increase in energy to 330 MeV/nucleon, however, would be required for comprehensive imaging which again requires one particle per RF bucket at a CW rate for practical reconstruction times. Such a machine will accelerate C\textsuperscript{6+} to a penetration
depth of 23 cm, making this a viable carbon therapy option for many cancers. Also, protons in
the form of $H_2^+$; i.e. charge to mass of $1/2$, can be accelerated in this system as the small
differences in nuclear mass are not significant enough to change the isochronous performance.

A dual stage accelerator system is optimal from a beam dynamics perspective as it limits
the impact of relativistic effects on orbital revolution time by narrowing the acceleration range.
This in turn lowers the order of the field expansion in radial profiles and decreases difficulty in
solving for constant machine tunes and adiabatic lattice functions. Dual stage and the racetrack
format are required to implement variable energy extraction promoting access to lower energy
orbits using technically feasible bump magnets. The low-energy stage further supports an eye
line, and treatment of melanoma, mesothelioma, and other skin or shallow tumors. The two
stages proposed for proton or ions could comprise the following energy ranges and support listed
applications.

- 18 to $\sim$250-330 MeV $H^-$ or protons
  - Fixed-frequency RF, CW beam
  - Low intensity for pCT
  - Stripping controls extraction energy and intensity for $H^-$
  - Racetrack is required for variable energy proton beams

Or, for a dual-stage ion accelerator:

1\textsuperscript{st} stage:
- 9 to $\sim$70/90 MeV charge to mass ratio of $1/2$
  - Fixed-frequency RF, DC beam for all ions
  - Variable energy extraction
  - Upstream injector for high-energy ring

2\textsuperscript{nd} stage:
- 70/90 to 250/330 MeV/nucleon
  - Variable energy extraction
  - Adjustable, fast orbit bump magnets/extraction septum in long straight
    - DC extracted beam
    - Variable energy on scale of tens of microseconds
    - Investigating extracted energy range

Note both stages require an injector which could be an off-the-shelf cyclotron.

It is important to understand the relationship between isochronous orbits and aperture which
can be approximated in terms of average radius at injection and extraction and approximate
main arc magnet apertures using the following equations, where $\beta$ is the relativistic velocity.

$$\bar{R}_{\text{extraction}} - \bar{R}_{\text{injection}} = \text{Aperture.}$$

For the isochronous condition, the average extraction radius can be expressed in terms of the
average injection radius.

$$(\frac{\beta_{\text{extraction}}}{\beta_{\text{injection}}} - 1) \bar{R}_{\text{injection}} = \text{Aperture.}$$

Using a normal conducting, extraction-energy radius of $\sim$ 5 m for the arc (2.5 m for supercon-
ducting arc version), the long straight section of the racetrack then defines the required magnet
Figure 6: The beam rigidity for fully-stripped, charge to mass of $1/2$, light ions as a function of the kinetic energy required to achieve a therapeutic penetration depth of 30 cm in water (human tissue equivalent) computed using SRIM/TRIM [22].

aperture as shown in Table 6 for isochronous orbits. A longer straight section increases aperture, but it allows longer extraction magnets to successfully extract the inner, lower energy orbits. Straight sections from 4 to 5.5 m has been successfully incorporated, changing the circular FFA in Figure 7 (top, left) into a racetrack (bottom, left). However stability of the machine tune and overall dynamics were greatly improved by adding a short 1 m straight section as also shown in Figure 7, (bottom, left). This short straight section is extremely useful for diagnostics and potentially additional RF (such as higher harmonic flattop cavities). One additional advantage is only two types of magnet designs are required for the racetrack, eliminating the very large normal-bend magnet at the center of the arc. A comparison of a circular ring design with 2 m straights was also studied for comparison with the reference CW nsFFA. Layouts and radial field profiles are shown in Figure 7.

Optimized radial field profiles were imported into COSY and fields expanded at magnet edges using an Enge function fallow for performance evaluations – the integrated magnetic field is kept constant to maintain the isochronous reference orbits. A field heat map is shown in Figure 8 (left) and the corresponding reference orbits as a function of energy also in Figure 8 (right). Using the COSY field map, machine tunes were then calculated as shown in Figure 9. At injection, near 90 MeV, the fields from the individual magnets merge due to proximity and tunes cannot be calculated. This lack of azimuthal field variation also occurs in an AVF and compact separated-sector cyclotron in the central or injection region, nominally incurring low beam capture efficiency and high loss at injection.

Table 6: Machine horizontal aperture dependence on length of racetrack long straight section for 330 MeV/ nucleon (pCT) and 430 MeV/nucleon (full-depth carbon therapy) ion accelerators.

<table>
<thead>
<tr>
<th>Energy Range MeV/nucleon</th>
<th>Straight 0 meters</th>
<th>Straight 2 meters</th>
<th>Straight 4 meters</th>
</tr>
</thead>
<tbody>
<tr>
<td>70 - 330</td>
<td>2.28</td>
<td>2.67</td>
<td>2.86</td>
</tr>
<tr>
<td>90 - 330</td>
<td>1.96</td>
<td>2.21</td>
<td>2.63</td>
</tr>
<tr>
<td>70 - 430</td>
<td>2.2</td>
<td>2.80</td>
<td>3.12</td>
</tr>
<tr>
<td>90 - 430</td>
<td>2.5</td>
<td>2.47</td>
<td>2.75</td>
</tr>
</tbody>
</table>

For a cross check, the racetrack was also modeled in the code Cyclops [23], extracting tunes
Figure 7: Layout of the circular 90-330 MeV/nucleon with 2 m straight sections (top, left) and a 5 m straight-section racetrack (bottom, left). The associated field profiles are also shown (right) as a function of $B(T)$ versus radius. Note that the extent of the magnet aperture is marked between the blue lines.

Figure 8: Field gradient strengths for half of the 90-330 MeV/nucleon racetrack (left) showing the extended field falloff at the edges and the corresponding closed isochronous reference orbits at different helium kinetic energies (right) for a field map generated by COSY [17].
Figure 9: Machine tunes for a 90-330 MeV/nucleon ion accelerator circular with 2 m straight sections (top) and the racetrack (bottom) with 5 m and 1 m straight sections as shown in the previous Figure 7. The horizontal axis is full kinetic energy for a He$^{2+}$ ion.

and time of flight (ToF) deviation as a function of energy. These are shown in Figure 10 and are consistent with the COSY calculation. The ToF deviation is ±2%, a reasonable level of isochronism for a preliminary lattice design for a CW machine. The isochronism level will need be reduced to $10^{-4} - 10^{-6}$ depending on the acceleration installed. The final adjustment is nominally done by shimming in the manufactured machine.

An example of a working extraction system layout for a 4 - 5.5 m long straight section is shown in Figure 11. High energy beam (blue trajectory) is extracted for an inward-bending polarity of the bump magnets. As the polarity reverses, inner, low-energy orbits (red trajectory) are swept outward and into the septum extraction magnet.
Figure 10: Machine tunes for a 90-330 MeV/nucleon ion accelerator circular with 5 m and 1 m straight sections and corresponding TOF deviation in 10% units relative to the He\(^{2+}\) orbit at 800 MeV. Modeling performed in Cyclops [23].

Figure 11: Layout of a 4-bump bipolar extraction system capable of variable energy extraction.

1.7 Summary

The FFA with alternating gradients amalgamates the best features of the synchrotron (strong focusing) and the cyclotron (fixed-field magnetic fields) and, for isochronous designs, high-gradient, fixed frequency RF. The high reliability and turnkey operation of fixed-field – and now fixed-frequency RF – accelerators make them ideal for medical and commercial purposes.

With isochronous performance, and strong-focusing optics, these new, advanced non-scaling FFAs have the potential for a low-loss CW machine for combined application in proton and ion therapy and imaging. Significant work has progressed on a stable, CW 30-330 MeV/nucleon isochronous FFA that can support both pCT and proton or ion therapy with minimal or no interruption for imaging diagnostics. This work presents the first CW designs of a relatively compact
accelerator for pCT and proton therapy and one which could potentially replace cyclotrons in many existing facilities. The evolution of the circular CW nsFFA into a racetrack format has significant potential for not only accommodating higher gradient RF modules, including SCRF, but also the potential for variable energy extraction using bipolar orbit-bump magnets in a long, ~5 m straight section using an extraction Lambertson or septum, as in synchrotron extraction. The settings of the orbit bump magnets select the internal orbit to extract and therefore the extraction energy. With such long straight sections, variable energy can also now be implemented for ions with a charge to mass of $1/2$. For variable-energy without a degrader and, ideally, for therapy optimization, two accelerator stages covering different penetration depths and targeting specific cancers are proposed. Further, an injection energy of 30 MeV was chosen to be compatible with a 30 MeV injector cyclotron, preferably separated-sector, and envisioned for radioisotope production to provide additional revenue capacity to a therapy facility.

In summary nonscaling, nonlinear versions of FFA with alternating gradients provide a new accelerator technology that can deliver continuous beam with high efficiency, but also reliably with “turnkey” operation from the standpoint of fixed magnetic fields and fixed RF frequency. These compact machines support dual energy extraction for therapy and imaging and are envisioned for successful integration into conventional cyclotron facilities for proton therapy; i.e. compatible with existing degrader, beamlines, and treatment rooms (including gantries). The racetrack and dual stage approach further support an ultra-compact variable-energy light-ion facility.

ACKNOWLEDGMENTS

This review paper includes important contributions to the development of nonscaling FFAs from Martin Berz and Kyoko Makino (Michigan State University), Pavel Snopok (Illinois Institute of Technology), Jordan Taylor, David Bruton, and Rob Edgecock (University of Huddersfield), Rob Appleby, Hywel Owen, and Sam Tygier (University of Manchester), and Shane Koscielniak and Thomas Planche (TRIUMF).

References


