Available online at www.sciencedirect.com
SCIENCE DIRECT®
Nuclear Instruments and Methods in Physics Research A 558 (2006) 282–291
NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A
www.elsevier.com/locate/nima

Staging acceleration and cooling in a Neutrino Factory

C. Johnstone[a,*], M. Berz[b], K. Makino[b]

[a]Fermi National Accelerator Laboratory, P.O. Box 500, Batavia IL 60510, USA
[b]Department of Physics and Astronomy, Michigan State University, East Lansing MI 48824, USA

Available online 1 December 2005

Abstract

All schemes to produce intense sources of high-energy muons—Neutrino factories, beta beams, Colliders—require collection, RF capture, and transport of particle beams with unprecedented emittances, both longitudinally and transversely. These large initial emittances must be reduced or "cooled" both in size and in energy spread before the muons can be efficiently accelerated to multi-GeV energies. The acceleration stage becomes critical in formulating and optimizing muon beams; individual stages are strongly interlinked and not independent as is the case in most conventional acceleration systems. Most importantly, the degree of cooling, or cooling channel, depends on the choice of acceleration. In the current US baseline scenario, the cooling required for acceleration is about a factor of 10 in transverse emittance per plane. Longitudinal cooling is also required. In the proposed Japanese scenario, using an alternative acceleration scheme, no cooling is presumed. This work discusses two basic, but different approaches to a Neutrino Factory and how the optimal strategy depends on beam parameters and method of acceleration.
Published by Elsevier B.V.

PACS: 41.75–i; 41.85Ja; 07.77Ka

Keywords: Neutrino Factory; Muon acceleration; FFAGs

1. Introduction

The important stages in the US scheme for a Muon Collider [1] and Neutrino Factory [2–4] are outlined in Fig. 1 (left and right, respectively). Muons are created via the decay of pions, and pions are produced by directing an intense beam of protons onto a production target. The initial stage of a muon facility is considered to be a proton driver capable of delivering an ultra-short (3 ns long), high-intensity (10^{14} p/pulse) beam. Collection, capture and bunching of pions and muons following the production target are the next major systems. First, a 50 m long channel allows the majority of pions with momentum less than $\sim$1 GeV/c to decay into muons; the muons are then captured and bunched into 200 MHz RF buckets. Just after production, the captured particle distribution exhibits an enormous rms momentum spread of $\pm 55\%$. The tremendous energy spread of the muons is reduced through phase rotation in an induction linac or, more recently, in consecutive RF stations with varying frequencies. The combined bunching and phase rotation process produces a train of approximately a hundred 200 MHz bunches with a reduced $\delta p/p$ and a final rms value of about 10 cm in bunch length. (Both the induction linac and the RF-based bunching and rotation schemes produce similar final momentum spreads and bunch lengths so the criterion in choice of technique is merely cost.) The transverse emittance successfully captured is approximately 16 cm rad (full, normalized) at a momentum of 200 MeV/c (mainly due to the strong, large-aperture solenoid surrounding the production target). The effective range in momentum, however, of captured particles extends from approximately 25 MeV/c up to a cut off near 1 GeV/c. Since these large emittances cannot be efficiently accelerated, a "cooling stage" for emittance reduction precedes acceleration. To be effective, a cooling channel must be able to accept not only large transverse emittances (implying large-apertures common to both magnetic and RF components), but also a

*Corresponding author. Tel.: +1 630 840 3794; fax: +1 630 840 6311.
E-mail addresses: cjj@fnal.gov (C. Johnstone), berz@msu.edu (M. Berz).

0168-9002/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.nima.2005.11.078

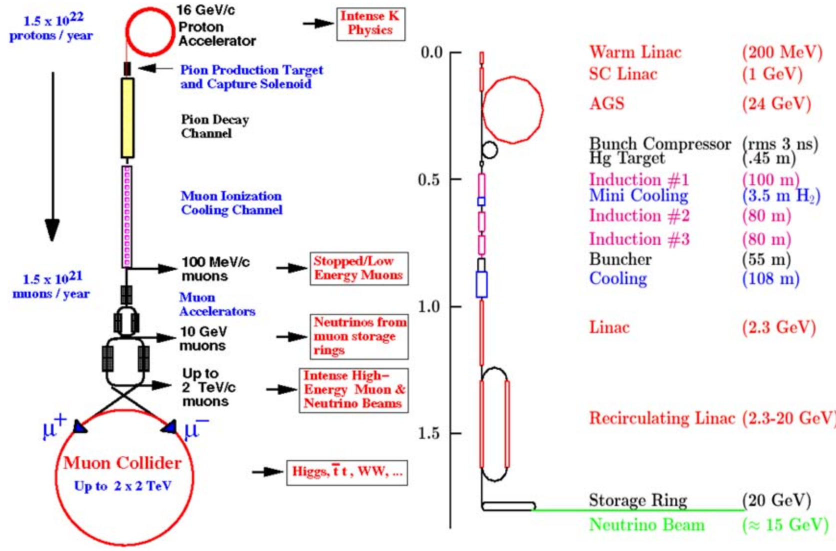

Fig. 1. Schematics of a Muon Collider [1] (left) and a Neutrino Factory [2] (right).

large (full) momentum spread of at least 40% $\delta p/p$ (generally quoted as $\pm 20\%$ $\delta p/p$ about a reference energy). The acceleration and storage rings—although nonconventional due to the large admittance and rapid cycle requirements that are imposed by large emittances and short muon lifetimes—represent the final stages of these facilities.

In all of the scenarios developed for Neutrino Factories and Muon Colliders, the captured transverse and longitudinal beam emittances are unprecedented. In comparison with high-energy hadron facilities, the transverse emittance is a factor of 1000 larger in each plane and the longitudinal emittance is 20–100 $\times$ larger even after bunching and phase rotation. Acceleration and collision of intense muon beams becomes impractical without a significant reduction, or cooling, of incipient emittances—transversely by a factor of 2.5–10 per plane for a Neutrino Factory [2–4] and at least a factor of 1000 for a Muon Collider [1]. (In the Neutrino Factory the required emittance reduction is tailored to the conditions for acceleration and in the Collider for the storage or collision ring.) Longitudinally, the degree of cooling differs drastically depending on the acceleration method, which is discussed here. The challenge in the design of these facilities, in particular the Neutrino Factory, lies primarily in accelerating the large beam emittances, a task which is further complicated by the short muon lifetime, or timescale on which these facilities must operate.

2. The acceleration stage

In a Neutrino Factory, the ability of, or limits to, accelerating large-emittance beams determines the specifications which upstream systems must meet, particularly the cooling. The downstream storage rings and experiments are presently not the limiting constraint. Acceleration proves, then, not only a difficult stage to develop, it becomes the pivotal one in the path to this facility. To further complicate issues, acceleration must occur rapidly because of potentially heavy losses from decay [1]. Linear accelerators are the optimal choice in this respect, but, above a few GeV, they become prohibitively expensive. Conventional synchrotrons cannot be used because normal conducting magnets cannot readily cycle in the ramping times [1] required by muon decay, nor do they support ultra-large beam emittances. In the past, the US baseline relied on recirculating linacs (RLAs) with separate, fixed-field arcs for each acceleration turn. Separate arcs allow control over the pathlength as a function of energy, allowing traversal times to be matched to the RF phase requirements for stable acceleration. Alternative approaches have focused on adapting the Fixed-Field Alternating Gradient (FFAG) accelerator first developed and tested at MURA [6], primarily because of its inherently large longitudinal acceptance. The Japanese approach (KEK) [7], for example, supports a radial-sector FFAG accelerator, but only in the context of a single-muon bunch and low frequency, broadband RF. Recent breakthroughs and improvements have demonstrated a new design for a FFAG accelerator [8,9] that can support stable, multi-turn acceleration for a high-frequency bunch train—the US scenario.

Transverse cooling (the upstream stage is termed precooling) is preferable in any approach after collection and capture to avoid enhanced component apertures, power levels, and, hence, cost of acceleration systems. However, it is the choice of accelerator that determines the longitudinal emittance that can be effectively accelerated and, therefore, also the degree of, or even the need for, longitudinal cooling. At a very fundamental level the overall design and staging of a muon facility depends on the method of acceleration chosen, and, for RLAs vs.

FFAGs, the longitudinal acceptance of the two machines embodies the most significant difference between the two scenarios. Since the RLA approach has been discussed extensively in past feasibility studies, and the intent of this paper is to focus on larger-acceptance accelerators, the RLA system requirements will be discussed only briefly. Instead this paper focuses on the FFAG approach, addressing the relaxed cooling requirements and re-formulation of upstream stages relative to the RLA. Additional references to FFAG acceleration are included in Ref. [10] for a more complete overview.

The acceleration stage is composed of two subsystems: a linear pre-accelerator and a recirculating-beam accelerator (in this case either the RLA or the FFAG machine). The preaccelerator will not be discussed in any detail here, but certain assumptions in its parameters are required to extrapolate to the next stage of acceleration.

2.1. Pre-accelerator

A linear accelerator is optimal for the first acceleration stage in order to bring the low-energy muon beam delivered by the cooling stage to an energy at which it becomes feasible to inject into a re-circulating-beam accelerator. It is advantageous to initiate acceleration after moderate cooling (the precooling stage) in order to mitigate aperture-related costs with components and power. Without precooling, the 16π cm rad emittance (full, normalized) which is captured gives rise to meter-scale apertures at 200 MeV/c, since peak beta functions can assumed to be $\sim$3 m (characteristic of 1–2 T solenoidal confinement [2] or quadrupole confinement in a short, 2 m, 90° FODO cell [11]). Even with minimal cooling (defined here as a factor of 2.5 per plane in transverse emittance), linac apertures decrease by 40%, which is a substantial reduction when applied to a meter.

3. Recirculating accelerators

Efficient injection into a recirculating accelerator precludes an extended transverse beam size; one-half to meter beam sizes are definitely problematic. The beam size at injection into the recirculating acceleration stage depends most strongly on the performance of the cooling stage in combination with the capability or acceleration installed in the pre-accelerator. Without the pre-accelerator, and even assuming the maximum cooling factor of 10, the beam size remains large: ± 0.15 m (full) for a beta function of 3 m. Furthermore, recirculating accelerator designs have peak beta functions near 10 m, which implies apertures must reach at least 0.5 m to accommodate injected beam. Clearly adiabatic cooling by accelerating to 2.5 GeV (implying a further reduction in geometric emittance relative to 200 MeV/c by a factor of 12) reduces the beam size substantially, by factor of 3.5. In order to achieve injection and more reasonable component apertures in the second acceleration system, a 2.5 GeV pre-accelerator is always assumed in this work. When combined with modest transverse cooling, the maximum beam size in the second accelerator becomes less than, or approximately, $\pm$8–16 cm (full) at 2.5 GeV, where the smaller number assumes a factor of 10 and the larger number a factor of 2.5 in cooling per plane. Beam sizes in the injection straight are lower than the maximum by about 25%, so injection remains challenging.

The transverse acceptance of the two proposed recirculators is described in the following sections. The emittance reduction factors are quoted relative to normalized units for convenience; they were calculated based on a 2.5-GeV pre-accelerator (which sets the injection energy) along with lattice design and practical assumptions for component apertures in the RLA and the FFAG cases, respectively. It is further assumed that a pre-accelerator can be designed that is capable of linearly accelerating both large transverse and large longitudinal emittances, with the latter particularly relevant to the FFAG scenario. (Linear acceleration implies that the effective normalized transverse emittance and the absolute momentum spread do not dilute significantly.)

3.1. Recirculating linear accelerator (RLA)

In an RLA, the beam is injected into a linac, accelerated, and returned by separate, fixed-field arcs on each acceleration turn, thereby achieving multiple passes of acceleration through the same linac. At the exit of each linac, the beam is sorted by energy and directed into a separate arc for transport on each acceleration pass. At the end of each arc, the trajectories from all arcs are recombined for acceleration in the opposing linac.

In spite of separate transport channels, however, the need to match to and maintain the RF bunching imposes the ultimate constraint on momentum acceptance within an arc and, indirectly, on its transverse acceptance. The implied condition on the longitudinal motion is that the value for M56 in each arc remain approximately constant over the accepted momentum range. For this purpose, strong families of sextupoles are used in the arcs, thereby achieving and fixing a maximum momentum acceptance in $\delta p/p$ of $\pm$5–10% for an RLA for a Neutrino Factory. Since sextupoles produce geometric aberrations, they also limit the useful dynamic aperture although with only a few turns, this is not a strong effect. Simply increasing the dispersion, thereby decreasing the strength of chromatic correction, does not appear to be effective in increasing either the longitudinal or the transverse acceptance of the RLA beyond this value due to aperture limitations and difficulty with high-order dispersion suppression in the linac sections.

Another, major difficulty in the design of the recirculating linacs lies in directing a beam with both a large transverse emittance and a large momentum spread into separate arcs on each acceleration pass. Clearly, to separate cleanly in a passive magnetic system (the only option for the microsecond circulation times), the energy width must be less than the energy difference between consecutive acceleration passes. A large transverse beam size implies a further increase in the distance required for achieving clean separation, promoting an unavoidable conflict between longitudinal and transverse acceptance. Transversely, the full acceptance achieved so far in the recirculating linac approach lies between 1.5π and 2π cm rad for the momentum spread quoted. The ability to separate this beam passively into independent component channels then sets the minimum acceleration requirement per turn and, therefore, the maximum number of recirculation turns achievable. Given these transverse emittances and quoted beam sizes (± 8 cm), the momentum spread that can be practically accelerated to extraction energy appears to be $\pm 1\%$ for each 400 MeV of acceleration per turn until the $\pm 10\%$ limit set by chromatic correction is reached. Correspondingly, only 4 acceleration turns have been reported for RLAs in order to accommodate large momentum spreads: $\sim\pm 5\%$ $\delta p/p$ for a 3–11 GeV RLA and $\sim\pm 10\%$ for a 3–20 GeV RLA [2]. The corresponding momentum spreads, 250 and 500 MeV/c, can be compared to the useful momentum width of the muon beam after bunching and phase rotation, which appears to be $\sim$600–700 MeV/c. (Much below 100 MeV/c beam is lost due to decay and above 700 MeV/c, the number of muons captured decreases rapidly.)

In conclusion, since the full initial normalized beam emittance, as defined by the present bunching scheme, is about 16π cm rad, this incipient emittance must be reduced or cooled by a factor of 8–10 transversely and at least 1.4–3 longitudinally before acceleration can be accomplished in an RLA designed for a Neutrino Factory.

3.2. FFAG acceleration option

Alternatively, a circular accelerator system can be devised with magnetic fields that remain constant during acceleration by adopting an alternating gradient focusing lattice. The arcs of such machines, composed of large aperture magnets, can be designed to accommodate the large-energy range in acceleration. The beam centroid orbit is not fixed as in a ramped machine, but rather moves across the magnet aperture during acceleration. Lattices have been developed which can contain an energy change of at least a factor of four, although current lattices have converged on a factor of 2 as most feasible technically. In a FFAG accelerator orbit length unavoidably changes with energy; this change can be substantial and can result in a significant phase-slip relative to the RF waveform (unless low-frequency RF is employed). The phase slip accumulates on a per turn basis and eventually prevents acceleration to the extraction energy. This effect limits the number of turns that can be supported under conditions of rapid acceleration when the RF phase cannot be adjusted on a corresponding timescale. Recent improvements in lattice design, however, have resulted in significant enhancement of the number of acceleration turns that can be supported over the RLA, 10–15 as compared with 4–5 turns for the same energy gain. A larger number of turns (>20) is not desirable in muon applications because of decay, especially at the lower energies. Consequently, a dramatic reduction in RF voltage is not gained using the FFAG, but there is significant cost reduction over its RLA counterpart.

There are several classifications of FFAGs which refer to the lattice and momentum dependence of the optics. Since the concern here is with machine admittance, only this general property will be advanced for the different machines. The so-called *scaling* FFAG accelerators, such as the radial or spiral sector, display an almost unlimited momentum acceptance, but transverse acceptance remains somewhat restricted. Another approach to FFAGs, referred to as *nonscaling*, employs only linear magnetic elements [12] (quadrupoles and dipoles). Although the transverse optics changes slowly with energy, this latest type of FFAG demonstrates both strong momentum acceptance and unlimited dynamic aperture (limited only by the physical restrictions of the components). However, conditions of rapid acceleration are required to avoid beam loss from betatron resonances, a condition which also applies to muon acceleration.

The *scaling* FFAG designs have successfully achieved dynamic apertures of a few centimeter-radians (full), thereby mitigating the transverse cooling requirements by about a factor of 2 relative to the RLA and yielding an overall transverse reduction factor of $\sim$5. *Nonscaling* FFAG designs exist with component apertures of 30–40 cm ($\sim$20 cm) horizontally (vertically) that accept a value of 6.4π cm rad for the full normalized emittance at the 2.5 GeV injection energy. The demonstrated emittance reduction factor is modest: only 2.5 in cooling prior to acceleration. Another advantage in both the *scaling* and *nonscaling* cases is total elimination of longitudinal cooling. In recent *nonscaling* FFAG designs, component apertures are now comparable to the RLAs, both horizontally and vertically. This is mainly due to the fact that the large dispersion in the RLA ($\sim$2–3 m required to separate beam with ± 5–10% $\delta p/p$ into different arcs) is comparable to the shifts of the beam centroid during acceleration in the FFAG, even in the presence of lower transverse cooling where beam sizes increase by 40–100%. (Beta functions are similar in the RLA and the FFAG.) To fully realize the reduced transverse cooling factor, however, the pre-accelerator must be capable of linearly accelerating both large transverse and large longitudinal emittances. (Again, linear acceleration implies that normalized emittances remain reasonably unchanged by the acceleration.)

At injection the momentum acceptance of the FFAG is, practically, about $\pm 14\%$, or ± 350 MeV/c. This limit is set by optimizing the FFAG design parameters which places the injection energy at a cell phase advance of 0.7π. The upper limit to phase advance for injection is about 0.8π in order to keep the injection optics stable and away from the π stability limit (the optics of the FFAG are FODO-like).

This upper limit on phase advance effectively sets the lowest momentum accepted for injection. Assuming an even spread about the central energy, gives the $\pm 14\%$ in $\delta p/p$. Even with a lattice optimized for acceleration and cost, the FFAG scheme does not appear to require any longitudinal cooling.

A simplistic view of cooling for a Neutrino Factory is illustrated in Fig. 2 including reference to different acceleration schemes. The technique of ionization cooling permits reduction of transverse emittances (4D phase space), or beam sizes, to levels acceptable for injection into accelerators with large momentum acceptance, or into ring coolers [4]. Ring coolers are multi-turn cooling channels designed to further reduce the transverse plus longitudinal emittance (6D phase space) to the smaller values required by a Neutrino Factory based on a RLA accelerator or by a Muon Collider.

3.3. Acceleration summary

Elimination of all longitudinal cooling, involving further muon decay, inevitable transmission losses, injection and extraction difficulties, and general R&D issues associated with both the optics and component design of advanced cooling channels, makes a persuasive argument to adopt the FFAG as the acceleration stage. Simple adiabatic damping during acceleration (if the dynamics are adequately conserving) is sufficient for the final transmitted emittances to meet the requirements of the storage ring and experiment. The present FFAG designs appear to satisfy this criterion, and, therefore, it appears sufficient to consider only a simple cooling stage for an FFAG: a straight or linear channel for transverse cooling (bends are required only for emittance exchange or longitudinal cooling). Such a linear channel must bring the initial, precooled emittance of 16π cm rad down to 3.2π cm rad for the *scaling* FFAG and 6.4π cm rad for the *nonscaling* FFAG accelerator. It should be noted, that a linear transverse precooler stage is relevant not only for the FFAG scenario, it is needed in RLAs and Muon Colliders as well with ring coolers representing a later stage of cooling. (Ring coolers also require "precooled" beam since their present design cannot accept large transverse emittances.)

4. The cooling stage

4.1. Ionization cooling

Emittance reduction, or ionization cooling, occurs because the muon beam loses momentum in all directions when traversing a target, or so-called absorber, and this energy loss can be replaced solely in the longitudinal direction by re-acceleration in an RF cavity; thus decreasing the beam's divergence for a given transverse dimension. The designs of stable optical configurations for cooling channels are particularly challenging because the straightforward cooling dynamics described above compete with the stochastic processes in the absorber; predominately multiple, or Coulomb, scattering which re-heat the beam. A net cooling effect can be achieved only if the cooling terms surpass the reheating ones, a state achieved through proper optics design in a cooling channel. The equation which follows represents the emittance increase in an absorber due to multiple scattering and, in the presence of cooling (reacceleration by an RF cavity), the minimum emittance [1] achievable for a specific channel design. From an optics standpoint, it is clear from equation 1 that the lower the average beta across the target, the proportionally lower the emittance increase from re-heating (multiple scattering), and therefore, the lower the achievable equilibrium emittance.

$$e_{\mathrm{N,min}} = \frac{\beta_{\perp}(14\,\mathrm{MeV})^2}{(2\beta m_{\mu} L_R\,\mathrm{d}E/\mathrm{d}S)} \tag{1}$$

where $\beta_{\perp}$ is the transverse beta function at the absorber, β the relativistic velocity, m_{μ} the mass of the muon, L_{R} the radiation length of the absorber material, and $\mathrm{d}E/\mathrm{d}s$ the energy lost per meter in the absorber.

Therefore, the performance of a cooling channel depends not only on the beta functions at the absorber and on their constancy across a large momentum range,

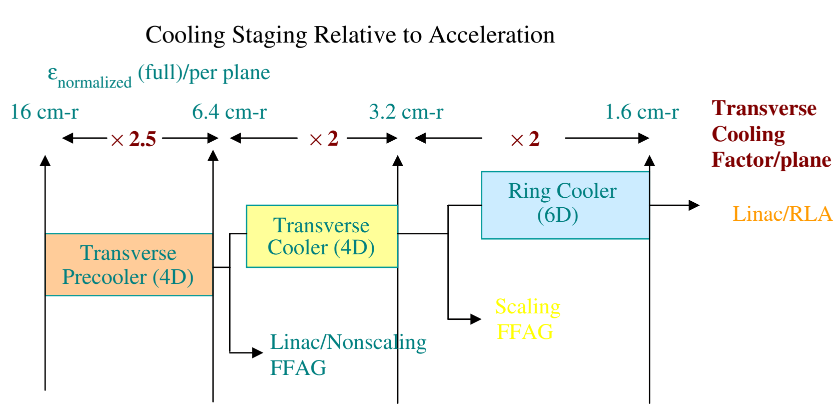

Fig. 2. Rough schematic of staged cooling relative to the acceleration model for a Neutrino Factory.

but, as important, also on the relative value of the starting emittance to the equilibrium emittance; the ratio needs only to be larger than $\sim$1.5 for effective cooling. This observation indicates that the extremely low beta functions required in the latter stages of cooling are not a prerequisite during the early stages and may actually be less optimal from both a technical and nonlinear standpoint. Relaxing the low-beta conditions at the absorber impacts tremendously the design, stability and strength of the elements used in the upstream cooling. This observation will later be used to optimize the design of the cooling channels and adapt to the mode of acceleration chosen.

4.2. Cooling channel specifications

Assuming that a full transverse acceptance of 1.5π to 2π cm rad, as required for an RLA, corresponds to 2.5σ of a Gaussian beam profile, the rms normalized emittance demanded from transverse cooling is 2.4π to 3.2π mm rad. This degree of cooling (according to Eq. (1)) corresponds to a cooling channel with an average beta function of <0.5 m at the absorber. In the FFAGs, the relaxed transverse conditions allow betas at the absorber to increase to $\leqslant 1$ and $\leqslant 2.5$ m for *scaling* and *nonscaling* machines, respectively.

In momentum the practical range for ionization cooling extends roughly from 100 MeV/c to a cutoff between 800 and 900 GeV/c. Application of all or a fraction of this range in the specification and design of a cooling channel depends on the longitudinal emittance capability of the acceleration system. The $\pm 5\%$ momentum acceptance of an RLA, for example, translates into a 250 MeV/c total width at an injection energy of 2.5 GeV, implying the optimal cooling range should potentially encompass 150–400 MeV/c. Early cooling channels [2,3] based on large-aperture solenoids with a 155–245 MeV/c cooling acceptance clearly were not optimally matched even to the requirements of the RLA.

In the case of FFAG acceleration one argues for a cooling channel which can accept and cool over as much of this momentum range as possible. The $\pm 14\%$ momentum spread accepted by current FFAG designs translates easily into an absolute momentum range of 700 MeV/c at the 2.5 GeV injection energy. Assuming no or slight increase in this spread, then all of the effective momentum spread of beam from the bunching/phase rotator stage can be utilized. In light of the relaxed transverse specifications and no prerequisites for longitudinal cooling, a simple, quadrupole-based channel [11] is found to be well-suited to FFAG acceleration. The concepts for this channel will be described in the following sections. (Such a channel could also serve as a precooling stage for RLA acceleration.) Since the baseline has been solenoidal, not quadrupole cooling, first a brief discussion of solenoidal cooling will be presented along with comparative arguments for quadrupole cooling.

4.3. Solenoidal cooling channel

The transverse cooling stages described in the feasibility studies have been based on extremely large aperture solenoids with strong, sometimes superconducting, fields (1.25–5 T), one baseline example being the sFOFO channel. This channel is well documented and will not be revisited in any detail [2,3]. The sFOFO channel is capable of achieving a value for beta at the absorber of about 0.4 m across a momentum range of 150–250 GeV/c. Substituting in the above equation produces an equilibrium emittance value of 1.7π mm mr (normalized, rms). The channel can be expected practically to deliver a muon beam with an rms normalized emittance of 2.5π mm mr (1.5 above equilibrium), clearly sufficient to drive an RLA acceleration stage.

For this value of low-beta, however, the momentum range is constricted to lying between the momentum limits of 155 and 250 GeV/c, with these two limits representing strong channel resonances, 2π and π, respectively [3]. Beam is essentially captured up to a momentum of close to a GeV, but in this channel the majority cannot be cooled or even transmitted. (In fact, the minimum in the re-heating terms which contribute to Eq. (1) actually occurs at a momentum of 400 MeV/c.) These and other concerns listed below exist with the solenoidal channels.

- discrete liquid hydrogen absorbers and RF cavities are housed physically within the solenoidal aperture increasing component aperture, cost, field nonlinearities, and power requirements;
- strong longitudinal to transverse correlations develop in solenoids;
- nonlinear dynamics increase dramatically with solenoid aperture–simple or low-order models are found to be inadequate[ ];
- strong, especially superconducting field strengths promote increased sensitivity of beam parameters at such low energies (0.1–1 GeV);
- the momentum reach (155–250 MeV/c) of the sFOFO is limited relative to beam from the bunching and phase rotation stage (later designs have improved acceptance [5]);
- operational problems have been discovered experimentally in the R&D of RF cavities operated in a strong solenoidal field.

These beam control issues and technical difficulties provide incentive to develop a channel based on more conventional magnetic elements: large-aperture quadrupoles rather than solenoids. Some of the technical motivations for a quadrupole-based channel are as follows:

- liquid hydrogen absorbers and RF cavities extracted from magnetic apertures;
- recent developments indicate the entire channel could be pressured with gaseous hydrogen (windowless absorber)

including the RF cavity which halves the length of the channel and increases dramatically the cavity gradient [13];

- magnetic fields are reduced to nonsuperconducting levels, with more stable beam dynamics, and less sensitivity to field imperfections and nonlinearities;
- quadrupoles are strong-focusing rather than "soft-focusing" elements with reduced nonlinear contributions from nonparaxial terms or kinematical effects.

4.4. Linear quadrupole precooler

When beta functions at the absorber exceed or approach 1 m, the focusing strength is dramatically reduced and the absorber no longer has to be located at the lowest or a very-low beta point, allowing more flexibility in the choice of optical structure and focusing elements. This observation represents the basis for designing a competitive cooling channel based on normal-conducting quadrupoles in a simple lens, or FODO-cell configuration. Its application is solely as an upstream stage of cooling and, being a linear channel with no bends, serves to reduce the large transverse beam size in preparation for acceleration in an FFAG or for injection into ring coolers.

4.4.1. Optical structure

With a sufficiently relaxed beta at the absorber, one can consider a short, alternating quadrupole lens structure. The advantages of a short FODO cell structure over a doublet or triplet quadrupole telescope are primarily in the acceptance and stability of optical parameters over a tremendous chromatic range. The dynamical range in telescope structures is about $\pm 5\%$ $\delta p/p$, beyond which there is no closed-orbit solution for off-axis beam. The limited momentum acceptance of the triplet/doublet quadrupole channels restrict their implementation to after longitudinal, or momentum, cooling has occurred and are not considered further here. However, in standard (implying repetitive) FODO-cell optics, the minimum beta in one plane is located at the maximum beta in the other. A minimum beta or beam size cannot be established simultaneously in both planes, and, therefore, the absorber cannot be located at the lowest beta point in this type of channel. The smallest beta for both planes combined is found halfway between the quadrupoles, at the "crossing point" in β_x and β_y. Due to this limitation, the valid application of a FODO-based cooling channel is just after capture and phase rotation.

4.4.2. Transverse cooling

For a short FODO cell, the average beta in both planes is equal and lies between 1 and 2 m for normal conducting quadrupoles and short ($\sim$0.5 m) spacing between them. The value of the beta functions at the crossing point is unusually stable over a large momentum range: from -20% to almost $+100\%$ if the phase advance is adjusted properly. The optics rationale for its design and stability will be discussed after a presenting the physical parameters chosen for this channel.

The physical parameters chosen for this channel shown in Fig. 3 were initially chosen to be comparable, or competitive with the sFOFO channel [2]. The aperture of the quadrupole was chosen somewhat conservatively—its length is constrained to be equal to its aperture—in order that the quadrupole field profile and therefore the optics are not fringe-field dominated.

The average beta achieved at the absorber in this channel is 1.6 m at 200 MeV/c (this is the defined central momentum of the sFOFO channel [2]). This absorber beta yields a design equilibrium emittance (rms, normalized) of 6.8π mm rad, or a practical rms final beam emittance of 10.2π mm mrad (factor of 1.5 above equilibrium). Assuming a 2.5σ Gaussian, the full final transverse beam emittance is then 6.4π cm rad, or a factor of 2.5 below the 16π cm rad emittance arriving from the upstream bunching stage.

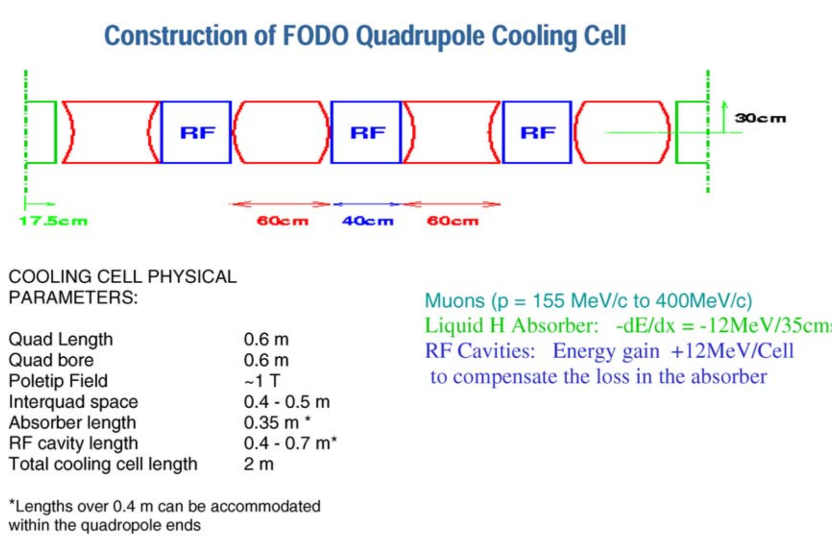

Fig. 3. Physical parameters and layout of the FODO-based quadrupole cooling channel.

4.4.3. Momentum acceptance

With the extreme demands placed on momentum performance, it is instructive to examine the FODO cell under the precepts of thin-lens conditions. First, it is useful to choose a reference momentum, p_0, and study the phase advance as a function of momentum relative to this reference in order to evaluate performance limits. For such a study, it becomes practical to assign a working point, or initial cell phase advance, to this reference momentum and one which is centrally located between stability limits: 0° and 180°. Clearly 90° is an obvious choice, hopefully optimizing the momentum reach of the channel. This choice of phase advance was applied to a p_0 of 200 MeV/c, a value chosen to be comparative with current cooling channel designs. The phase advance dependence, φ, on momentum can now be obtained in the thin-lens approximation (see reference)

$$\sin\frac{\varphi}{2} = \frac{(1/\sqrt{2})p_0}{p} \text{ with a clear lower limit of stability,}$$
$$p = \frac{1}{\sqrt{2}}p_0. \tag{2}$$

Differentiating gives

$$\frac{1}{2}\cos\frac{\varphi}{2}\,\mathrm{d}\varphi = -\frac{p_0}{\sqrt{2}p^2}\mathrm{d}p \quad \text{or} \quad \frac{\mathrm{d}\varphi}{\mathrm{d}p} = \frac{\sqrt{2}p_0}{p^2\sqrt{1-p_0^2/2p^2}}. \tag{3}$$

Notice that for a p_0 of 200 MeV/c, the above analysis (Eq. (2)) gives a lower momentum cutoff for the channel of ~140 MeV/c and, at large p, the phase advance varies more and more slowly, as $1/p^2$. The results of this analysis are graphed in Fig. 4 clearly demonstrating the large play in momentum of the simple-lens FODO cell. When compared with calculations, an almost constant factor of 0.8 was needed to translate the changes in phase advance from the thin-lens model to ones accurate for the channel as designed.

The slow variation in phase advance does not set restrictions on the length of the cell, but the variation of the peak beta function with momentum does. Using the definitions above, the peak beta function[ ] for a FODO cell is given by

$$\beta_{\max} = L\frac{\kappa(\kappa+1)}{(\kappa^2-1)^{1/2}},$$
$$\frac{\mathrm{d}\beta_{\max}}{\mathrm{d}p} = L\frac{(\kappa^2-\kappa-1)}{(\kappa-1)^{3/2}(\kappa+1)^{1/2}}\frac{\mathrm{d}\kappa}{\mathrm{d}p} = 0 \text{ for a minimum.} \tag{4}$$

The variable κ is defined by the following thin-lens equation.

$$\sin\frac{\varphi}{2} = 1/\varsigma \quad \text{where } \varsigma = f/L \text{ (thin lens).} \tag{5}$$

Here φ is the phase advance of the FODO cell, f is the focal length of $\frac{1}{2}$ of a full quadrupole, and L is the length of a half-cell from quadrupole center to center (see, for example references listed in Ref. [14]).

In the above Eq. (4), $(\kappa^2-\kappa-1)$ can only be set to 0 locally (at ~76°), but this does not guarantee stability in the beta function over a large range in momentum. The only approach that minimizes $\mathrm{d}\beta_{\max}/\mathrm{d}p$ over a broad spectrum is to let L approach 0. No drift between quadrupoles is optimal, but the choice of a short drift of ~0.5 m (which corresponds here to a half-cell length of 1 m) intentionally slows the variation of the maximum beam size with energy and at the same time insures a more feasible technical channel design. (Here absorbers and RF cavities are not installed inside magnet apertures.) The variation of the maximum beta with momentum for this design is shown below.

When the momentum dependence of the average beta at the absorber was studied, the change was found to increase slowly with energy and slowly relative to the increase in normalized admittance of the channel with energy (which is due to slowly varying peak beta values as a function of momentum). This increase in normalized emittance indicates cooling takes place over a large momentum range despite the increase in $\beta_\perp$ at the absorber.

Both the sFOFO and quadrupole cooling channels were fully modeled and tracked to high-order using the code COSY [15]. The simulation included:

- full nonlinear terms;
- with full solenoidal [16] and quadrupole fringe fields [17–19] (including different models);

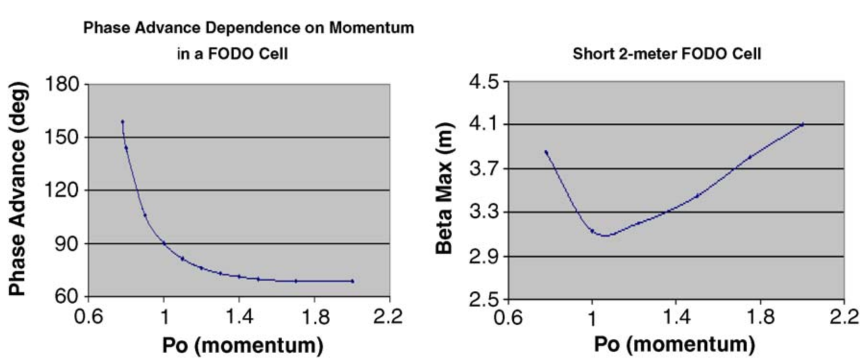

Fig. 4. On the left is the phase advance of a FODO cell plotted with respect to an arbitrary momentum, p_0, whose phase advance has been set to 90°. On the right is the variation of the peak beta function relative to p_0 for a half-cell length of 1 m.

- multiple scattering (absorbers + windows);
- energy loss including straggling and spin;
- dE/dx as a function of energy;
- 200 MHz sinusoidal RF;
- full quadrupole fringe fields (both Enge representation[ ] and actual measured fields[ ]).

For the tracking, particles were launched in 2 cm steps along both axes and along the diagonal starting at the center of the absorber with and without the cooling (the hydrogen absorber). With cooling turned on, the transmission losses in the quadrupole channel exclusive of muon decay were almost negligible—less than 1% over the momentum range accepted by the sFOFO channel (155–245 MeV/c). (This transmission corresponds to an rms bunch length of 7.5 cm, a σ_E of 12 MeV, $\varphi_s = 60^\circ$, $\Delta\varphi_s = \pm 54^\circ$, which corresponds to the 200 MHz RF bucket being about half filled.) Since the cooling rapidly reduces the emittance, the dynamic aperture is almost not relevant because a beam that fills the entire quadrupole aperture is cooled and is not lost even in the presence of fringe fields.

The predicted cooling behavior was observed; if a Gaussian distribution is launched which fills the quadrupole aperture, then the final rms of the distribution was found to be near 6.8 mm rad (normalized) for this specific channel (Fig. 5). The longitudinal losses of the quadrupole channel appear to be less than the solenoidal channel, the reason being the absence of longitudinal transverse correlations that plague solenoids.

As mentioned it is important to calibrate the expanded momentum reach of the quadrupole cooling channel. As noted above, the geometrical acceptance is almost constant, therefore, the normalized acceptance is increasing (since the relativistic velocity is not changing significantly). Hence, the absorber beta can be allowed to track the increase in normalized emittance acceptance. When benchmarked against the full cooling simulation performed at a p_0 of 200 MeV/c, the channel cools beyond 400 MeV/c. The momentum reach of the cooling in the quadrupole channel appears to be significantly larger than in the sFOFO, which extends from 155 to 245 MeV/c. Current effort is underway to launch a realistic beam from the bunching/phase rotation stage with varying momentum cuts to determine accurately the extent of the momentum cooling achieved in this channel.

4.5. Cooling summary

It is clear that a simple quadrupole cooling channel can be considered as the only cooling stage necessary for FFAG acceleration in a Neutrino Factory. At 200 MeV/c and a 60 cm diameter bore, it first appears that the aperture is insufficient to accommodate the full 16π cm rad emittance delivered by the upstream bunching and phase rotation system (assuming ~3 m peak beta function in the quadrupole). However, an elliptical beam pipe can be installed in the quadrupole which extends beyond the poletip diameter by 50% and still maintain good field quality if the poletips are properly designed and separated (F. Mills, private communication). This would accommodate full beam using FODO channel optics. It is also interesting to note that the 200 MeV/c central beam momentum was chosen due to the already strong solenoidal fields employed in the sFOFO channel and any increase in the central momentum implied an increase in field strength which is limited in this particular channel design. There is no reason not to accommodate a larger central momentum, and one which is more optimally matched to the production spectrum which extends from ~100 MeV/c up to ~1 GeV/c. Moreover, the minimum of the sum of the re-heating terms (multiple scattering plus

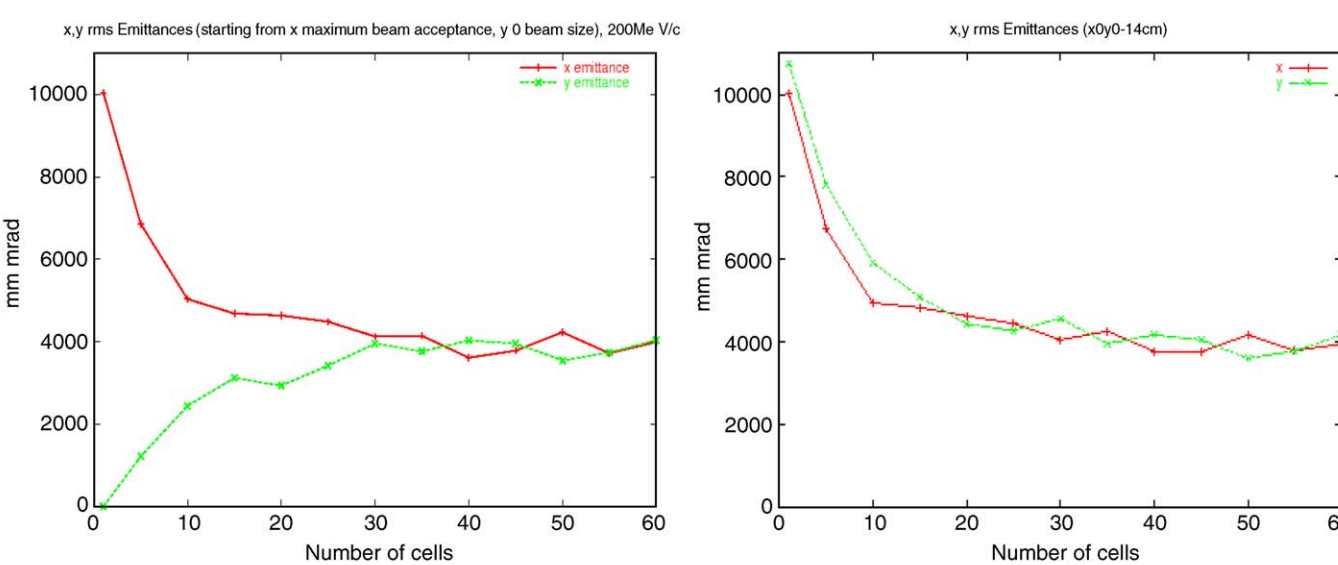

Fig. 5. Emittance reduction as a function of number of cooling cells. Left plot shows this starting emittance in x, no emittance in y with corresponding emittance growth up to the equilibrium emittance. Plot on right are particles launched in x,y Gaussian distributions, but along the diagonal.

straggling) actually occurs at 400 MeV/c. For the quadrupole channel design, the poletip field is not near a technical limit and can be increased. However, an increase in quadrupole strength is not required as this channel cools efficiently and well beyond the momentum reach of the sFOFO channel, from 150 to >400 MeV/c, so its central momentum is actually closer to 300 MeV/c. Further there are gains to be made by reducing losses from muon decay by propagating higher momenta, on average, through the channel.

5. Conclusions

In this paper, optimizing the stages in a Neutrino Factory has been presented. Clearly the staging and optimization are critically dependent on the choice and format of accelerator. It has been demonstrated that possibly the simplest, lowest-cost scenario is a *nonscaling* FFAG machine coupled to a linear (no bending) transverse cooling channel constructed from the simplest quadrupole lens system, a FODO cell. Transverse cooling demands are reduced by a factor of 4 and no longitudinal cooling is required relative to the RLA option. Detailed simulations further show that a quadrupole-based channel cools efficiently and much beyond the momentum range of a sFOFO [2] cooling channel with similar magnetic apertures. Current effort is underway to characterize the exact energy extent of the cooling. Applying different—both assumed and measured fringe fields—to represent the quadrupole elements fully has been an integral part of the simulations and ensures feasibility in quadrupole design and performance. One important observation is that such a channel cools effectively over a large variation in the fringe-field profile. Extensive design and simulation work are currently in progress on the proposed *nonscaling* FFAG and results are also encouraging. This cooling/acceleration scheme potentially represents the baseline scenario for the next US feasibility study.

References

[1] The μ^+ μ^- Collaboration, $\mu^+\mu^-$ collider: a feasibility study, BNL-52503, Fermi-Conf-96-/092, LBNL-38946, July, 1996.
[2] S. Ozaki, R. Palmer, M. Zisman, J. Gallardo (Eds.), Feasibility study II of a muon-based neutrino source, BNL-52623, June, 2001, available at http://www.cap.bnl.gov/mumu/studii/FS2-report.html.
[3] N. Holtkamp, D. Finley (Eds.), A feasibility study of a neutrino source based on a muon storage ring, Fermilab-Pub-00/108-E, 2000.
[4] Muon Collider/Neutrino Factory Collaboration, Phys. Rev. St. Accel. Beams 6 (2003) 081001.
[5] The Neutrino Factory and Muon Collider Collaboration, Neutrino factory and beta beams and development, Study 2-a, Lab Reports, BNL-72369-2004, FNAL-TM-2259, July 30, 2004, available at http://www.cap.bnl/mumu/study2-a and http://www.interactions.org/neutrinostudy.
[6] K. Symon, MURA-KRS-6, (MURA-43)(1954);
K. Symon, et al., Phys. Rev. 103 (1956) 1837.
[7] S. Machida, et al. Beam optics design of an FFAG synchrotron, (MOP1B20), Y. Sato, et al, Development of a FFAG proton synchrotron, (MOP1B21), EPAC 2000, submitted.
[8] C. Johnstone, S. Koscielniak, Nucl. Instr. and Meth. A 519 (2004) 472.
[9] S. Koscielniak, C. Johnstone, Nucl. Instr. and Meth. A 523 (2004) 25.
[10] M. Craddock, Cern Courier, V44, No. 6 (2004). FFAG Workshop, Vancouver, CA, April 15–24 2004, http://www.triumph.ca/ffag2004;
S. Koscielniak, M.K. Craddock, "simple analytic formulae for properties of nonscaling FFAG lattices", Proceedings of the Ninth European Particle Accelerator Conference, July 2004, Lucerne Switzerland (EPAC 2004);
S. Koscielniak, Comparison of COSY DA Maps with analytic formulae for orbit functions of a non-scaling FFAG accelerator, Proceedings of the Ninth European Particle Accelerator Conference, July 2004, Lucerne Switzerland (EPAC 2004);
C. Johnstone, S. Koscielniak, Optimizing non-scaling FFAG lattices for rapid acceleration, Proceedings of the Ninth European Particle Accelerator Conference, July 2004, Lucerne Switzerland (EPAC 2004).
[11] C. Johnstone, et al., Nucl. Instrum. and Meth. A 519 (2004) 472.
[12] C. Johnstone, A. Garren, in: Proceedings of the 1999 Particle Accelerator Conference, New York, NY, March 29–April 2, 1999 pp. 3068.
[13] R. Johnson, High pressure, high gradient RF cavities for muon beam cooling, in: Proceedings of the Linac2004, 17 August 2004, Lubeck Germany, available at http://www.muonsinc.com/TU203.pdf.
[14] E.D. Courant, H.S. Snyder, Ann. Phys. 3 (1958). Also, H. Bruck, "Accélérateurs de Perticules, Press Universitaires de France, Paris, 1966; English translation: Los Alamo Report LA-TR-72-10.
[15] M. Berz, K. Makino, COSY INFINITY Version 8.1—user's guide and reference manual. Technical Report MSUHEP-20704, Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, 2001, see also http://cosy.pa.msu.edu.
[16] K. Makino, M. Berz, D. Errede, C. Johnstone, Nucl. Intr. and Meth. A 519 (2004) 162.
[17] M. Berz, Modern Map Methods in Particle Beam Physics, Academic Press, San Diego, 1999;
Also, G. Hoffstatter, M. Berz, Phys. Rev. E 54(4) (1996) 5664;
G.H. Hoffstatter, Rigorous bounds on survival times in circular accelerators and efficient computation of fringe-field transfer maps, Ph.D. thesis, Michigan State University, East Lansing, Michigan, USA, 1994.
[18] B. Erdelyi, Simplectic Approximation of Hamiltonian Flows and Accurate simulation of fringe field effects, Ph.D. thesis, Michigan State University, East Lansing, Michigan, USA, 2001, http://bt.pa.msu.edu/cgi-bin/display.pl?name=erdelyiphd.
[19] S. Kowalski, H. Enge, RAYTRACE, Technical Report, MIT, Cambridge MA, 1985.