Muon Capture Schemes for the Neutrino Factory

Stephen Brooks
Trinity College, Oxford

Thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy at the University of Oxford

Hilary Term, 2010

Abstract

The proposed *neutrino factory*, a facility for precision measurements of neutrino oscillations, requires directional neutrino beams to reach the required sensitivities. Among the few sources of such beams is the decay of muons travelling at relativistic speeds, therefore an intense source of muons and subsequent rapid acceleration must be designed so they can reach the required energy before decaying.

This thesis considers several stages in this process: pions are produced from a proton beam hitting a target and pion yield optima are determined as a function of target design parameters and the proton energy. Issues related to producing the original proton beam are also discussed. The pions decay to a beam of muons, confined by a channel of solenoids and other components known as the *muon front end*. A design for this is found that meets the requirement of 10^{21} muons per operational year [1].

The computer code MARS15 [2] is used to simulate the target, with benchmarks against GEANT4 [3] and initial results from the HARP experiment [4]. The author's code Muon1 [5] is used for muon tracking, with its techniques also explained in the thesis.

To find the highest-yielding arrangement of magnets and accelerating components from the target onwards, Muon1 incorporates an optimisation feature where almost all parameters of the beamline can be varied. This produces a high-dimensionality search space where the best muon yield is sought using a genetic algorithm. As each individual evaluation of a design is itself a time-consuming simulation with tens of thousands of particles, the code has been deployed as a distributed computing project that is able to perform millions of simulations per optimisation.

Contents

Chapter 1

Introduction

The following sections describe the problem of detecting the very small mass of neutrinos and experimental methods for doing this, particularly the *neutrino factory* that is studied here. Attention then focusses on the *muon front end* section, as this is critical to the efficiency of such a facility. Lastly, the methodology used for the optimisation of front end designs in the rest of this thesis is outlined.

1.1 Neutrino Oscillations

The original theory of neutrinos [6] by Pauli in 1930 formulated them as massless particles with their momentum and energy alone sufficient to explain their interactions. Originally, these interactions were inferred from the continuous energy spectrum of electrons from beta decay but in 1956 a neutrino from a nuclear reactor was observed being absorbed by an atomic nucleus [7], confirming its status as a particle.

The first observations that did not agree with the massless neutrino model were early results from the Homestake solar neutrino detector [8], which from 1968 onwards showed a factor of $\sim$3 deficit in the number of electron neutrinos (ν_e) received on Earth compared with astrophysical models of solar neutrino production. At this stage it was unclear whether the deficit, referred to as the 'solar neutrino problem', was caused by errors in the astrophysical model, oscillation between the neutrino flavours (already proposed theoretically [9]), or something else. The Super-Kamiokande experiment [10], operating from 1996, was able to resolve this ambiguity by measuring oscillations in neutrinos produced from the decay of muons in the Earth's atmosphere. As the production sites of these neutrinos form a spherical shell around the Earth, the zenith angle of the incoming neutrinos at the detector determines how far away they were produced (up to $\sim$13000 km away) and this range of baseline distances is suitable for observing ν_μ oscillations. Further review of the evidence for neutrino oscillations is available in [11]; this section will now explain the physics behind the process.

For a stationary particle of rest mass m and wavefunction $|\psi\rangle$, the mass–(total) energy equivalence $E_t = mc^2$ becomes the quantum equation of motion $\hat{E}_t |\psi\rangle = mc^2 |\psi\rangle$, assuming that 'a particle' in this sense is always a pure mass eigenstate. The quantum energy

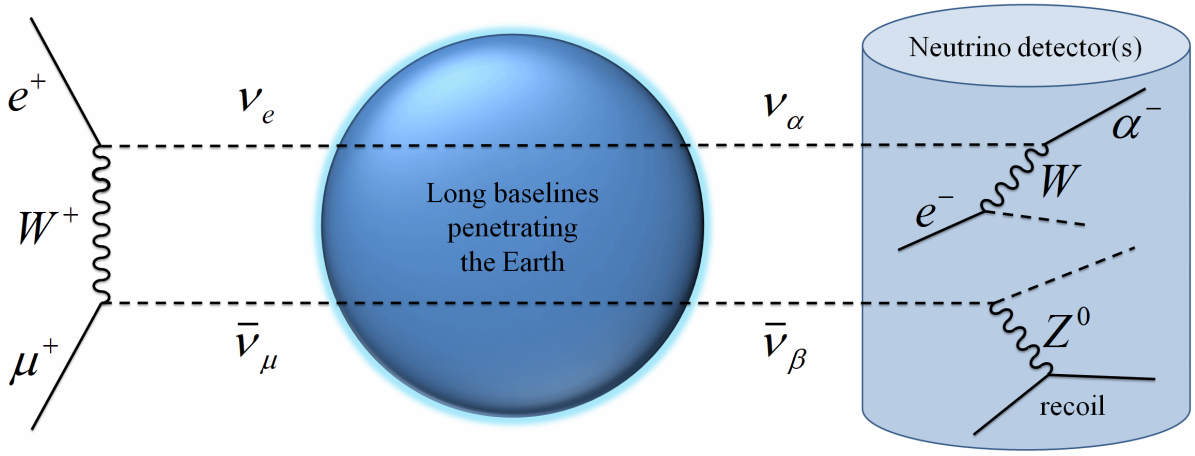


Figure 1.1: Decay of a positive muon via the weak nuclear force (left) as occurs in the neutrino factory decay ring. The neutrino and antineutrino produced may undergo further weak interactions in a distant neutrino detector (right), possibly after oscillating to a different flavour.

operator is given by $\hat{E}_t\,|\psi\rangle := \mathrm{i}\hbar|\dot{\psi}\rangle$, where dot denotes $\frac{\partial}{\partial t}$, so the time solution to the equation is $|\psi(t)\rangle = \exp(mc^2 t/\mathrm{i}\hbar)\,|\psi(0)\rangle$ with the exponential of a pure imaginary number producing a constant rate of phase precession with angular frequency $\omega_m := -mc^2/\hbar$.

Assuming that there are three neutrinos $|\nu_i\rangle$ with masses m_i for $i \in \{1,2,3\}$, consider the possible production mechanisms. A weak reaction such as the muon decay $\mu^+ \to \bar{\nu}_\mu e^+ \nu_e$ is mediated by the W^+ boson as shown on the left in figure 1.1. The important feature is that a weak vertex for a lepton of flavour $\alpha \in \{e,\mu,\tau\}$ will produce a flavoured neutrino that is a combination of the mass eigenstates $|\nu_\alpha\rangle := \sum_{i=1}^{3} U^*_{\alpha i}\,|\nu_i\rangle$, where U^* is the complex conjugate of the 3×3 *Maki–Nakagawa–Sakata* (MNS) *matrix*. The equivalent formula with antineutrinos on both sides uses U instead of U^*. U is unitary ($U^{-1} = U^{T*}$), so the inverse relation can be simplified to $|\nu_i\rangle = \sum_\alpha U^{-1*}_{i\alpha}\,|\nu_\alpha\rangle = \sum_\alpha U_{\alpha i}\,|\nu_\alpha\rangle$.

We can now model the production and detection of a neutrino in an oscillation experiment that produces ν_α and detects them using lepton flavour β. Transporting $|\nu_\alpha\rangle$ as defined above through proper times τ_i in the rest frames of the mass eigenstates $|\nu_i\rangle$ will apply independent phase advances to the components as follows:

$$|\nu_\alpha(\tau)\rangle = \sum_{i=1}^{3} U^*_{\alpha i} \exp(\mathrm{i}\omega_{m_i}\tau_i)\,|\nu_i\rangle\,.$$

The detection probability of a ν_β flavoured neutrino is $P_{\alpha\beta} = |\langle\nu_\beta|\nu_\alpha(\tau)\rangle|^2$. All observable operators in quantum theory are Hermitian and hence have orthonormal eigenvectors, meaning that $\langle\nu_i|\nu_j\rangle = \delta_{ij}$ and we can expand the probability formula taking advantage of some of these being zero:

$$P_{\alpha\beta} \;=\; \left|\sum_{i=1}^{3} U_{\beta i}\,\langle\nu_i|\sum_{j=1}^{3} U^*_{\alpha j}\exp(\mathrm{i}\omega_{m_j}\tau_j)\,|\nu_j\rangle\right|^2$$

$$
\begin{aligned}
&= \left| \sum_{i=1}^{3} U_{\beta i} U_{\alpha i}^{*} \exp(\mathrm{i}\omega_{m_i}\tau_i) \right|^2 \\
&= \sum_{i=1}^{3} U_{\alpha i}^{*} U_{\beta i} \exp(\mathrm{i}\omega_{m_i}\tau_i) \sum_{j=1}^{3} U_{\alpha j} U_{\beta j}^{*} \exp(-\mathrm{i}\omega_{m_j}\tau_j) \\
&= \sum_{i=1}^{3} |U_{\alpha i}^{*} U_{\beta i}|^2 + \sum_{i \neq j} U_{\alpha i}^{*} U_{\beta i} U_{\alpha j} U_{\beta j}^{*} \exp(\mathrm{i}(\omega_{m_i}\tau_i - \omega_{m_j}\tau_j)).
\end{aligned}
$$

At this point, note that

$$
\left| \sum_{i=1}^{3} U_{\alpha i}^{*} U_{\beta i} \right|^2 = \sum_{i=1}^{3} |U_{\alpha i}^{*} U_{\beta i}|^2 + \sum_{i \neq j} U_{\alpha i}^{*} U_{\beta i} U_{\alpha j} U_{\beta j}^{*}
$$

but because of the unitarity of U, the squared item on the left-hand side is the 'dot product' of two orthogonal rows: $\sum_{i=1}^{3} U_{\alpha i}^{*} U_{\beta i} = \delta_{\alpha\beta}$. Using this to replace the first term gives the result

$$
P_{\alpha\beta} = \delta_{\alpha\beta} + \sum_{i \neq j} U_{\alpha i}^{*} U_{\beta i} U_{\alpha j} U_{\beta j}^{*} \left(\exp(\mathrm{i}(\omega_{m_i}\tau_i - \omega_{m_j}\tau_j)) - 1 \right).
$$

The three mass eigenstates $|\nu_i\rangle$ are treated as remaining in superposition even though they propagate at slightly different speeds. By the uncertainty principle, the neutrino ν_i will not be perfectly confined in its rest-frame coordinates $(c\tau_i, \xi_i)$: there will be a small range $\Delta\xi_i$ around zero containing some amplitude for the neutrino, which all has the same phase $\omega_{m_i}\tau_i$. In most situations this is larger than the very small classical distance expected between the neutrinos. If the neutrinos are emitted at the common origin of the coordinate systems and received at (ct, L) in the laboratory frame (L being the *baseline length*), the Lorentz transform gives $\tau_i = \gamma_i t - \beta_i\gamma_i L/c$ for the neutrino ν_i that has relativistic factors β_i and γ_i in the laboratory frame.

If E_t is the total energy the neutrino was produced with, then

$$
\begin{aligned}
\omega_{m_i}\tau_i &= \left(-m_i c^2/\hbar\right)\left(\gamma_i t - \beta_i\gamma_i L/c\right) \\
&= -\left(m_i c^2 \gamma_i t - m_i c \beta_i \gamma_i L\right)/\hbar \\
&= -\left(E_t t - p_i L\right)/\hbar \\
\Rightarrow \quad \omega_{m_i}\tau_i - \omega_{m_j}\tau_j &= \left((p_i - p_j)L\right)/\hbar.
\end{aligned}
$$

The truncation of the power series $p_i = \sqrt{E_t^2/c^2 - m_i^2 c^2} \simeq E_t/c - m_i^2 c^3/2E_t$, valid in these ultra-relativistic cases may now be used to get

$$
\omega_{m_i}\tau_i - \omega_{m_j}\tau_j = \frac{1}{\hbar}\left(\frac{m_j^2 - m_i^2}{2E_t} c^3 L\right)
$$

and

$$
P_{\alpha\beta} = \delta_{\alpha\beta} + \sum_{i \neq j} U_{\alpha i}^{*} U_{\beta i} U_{\alpha j} U_{\beta j}^{*} \left(\exp(\mathrm{i}(c^3/\hbar)\Delta m_{ji}^2 (L/2E_t)) - 1 \right),
$$

where $\Delta m^2_{ij} := m^2_i - m^2_j$. This formula reveals the oscillatory nature of these detection probabilities as a function of L: each term is either a constant or a multiple of a linearly-advancing phase factor (the sum is real because the imaginary parts cancel on swapping i and j). If one is interested in a particular mass splitting, aiming for a phase advance of π for maximum signal in its corresponding term suggests the following baseline:

$$\begin{aligned} (c^3/\hbar)\Delta m^2_{ji}(L/2E_t) &= \pi \\ \Rightarrow \quad L &= \frac{2\pi\hbar}{c^3}\frac{E_t}{\Delta m^2}, \\ \text{or} \quad L\,[\text{km}] &\simeq 1.240\,\frac{E_t\,[\text{GeV}]}{\Delta m^2\,[(\text{eV}/\text{c}^2)^2]}. \end{aligned}$$

For example, neutrinos from nuclear reactions, either on Earth or in the Sun, have an energy spectrum of the order of a few MeV. For 3 MeV neutrinos in particular, L would be $\sim$1.4 km if Δm^2_{23} were used, or $\sim$47 km in the case of oscillations caused by the ν_1–ν_2 mass difference. The $\sim$1.4 km baseline forms the basis of nuclear reactor neutrino oscillation experiments [12, 13]. For experiments through the Earth with much longer baselines, the interaction of neutrinos with electrons in matter must be included [14], which changes the oscillation wavelengths from those given in vacuum by the formula above.

Oscillation experiments measure oscillation probabilities, which depend not just on Δm^2_{ij} but also the elements of the MNS matrix U. As U is unitary, the complex equivalent of a rotation matrix, it can be parameterised in terms of angles: the standard way of doing this is given on page 12 of [15]. There are three *mixing angles* θ_{12}, θ_{13} and θ_{23}, which determine the extent of neutrino mixing, plus one *CP-violating phase* δ [16]. It is the goal of neutrino oscillation physics to measure all these parameters as well as the Δm^2_{ij}, though individual experiments may be sensitive to different subsets of the parameters.

1.2 Neutrino Beam Facilities

Natural sources of neutrinos create an evenly-distributed but fairly weak flux over the entire Earth's surface, whereas nuclear reactors produce a concentrated flux from what is roughly a point source but emit neutrinos evenly in all directions. The best configuration for an oscillation experiment would be to have as many of the neutrinos as possible going through the detector, which requires a *neutrino beam* that stays concentrated over a long distance from the source, requiring anisotropic emission of neutrinos.

The only known way of practically achieving this is to take an interaction, such as a decay, which emits neutrinos isotropically in the rest frame of the parent particle, then pre-accelerate the parent particle to relativistic speeds so the Lorentz boost transforms the distribution of neutrino emission directions into a forward-directed beam in the experimental frame. For relativistic $\gamma \gg 1$ this produces a beam opening angle proportional to $1/\gamma$ and a flux enhancement factor of γ^2 relative to an isotropic source.

The source of beamed neutrinos taken together with the detector(s) at the other end of the baseline(s) is referred to as a *neutrino beam facility*.

Three main methods of producing boosted neutrino emission have been proposed.

- **Conventional neutrino beams**. Any experiment in which a beam of GeV-order energy hits a fixed target will generate large numbers of pions (π^+ and π^-), which if left alone will decay via $\pi^+ \to \mu^+\nu_\mu$, or the charge conjugate $\pi^- \to \mu^-\bar{\nu}_\mu$. The pions are mostly forward-directed because they retain some of the original beam's momentum and, as they are charged, may be focussed using elements such as solenoids or a magnetic horn. Examples include MINOS [17] and K2K [18].
  - **Superbeams**. An obvious way to increase the flux of the conventional beam is to increase the mean power of the primary beam. The term *superbeam*, for 'super-conventional beam', refers to experiments such as [19] that use megawatt-order (or higher) beams on the target.
- **Beta beams**. Beta decay and positron emission decay are naturally-occurring process that occur in certain radioisotopes. At the level of nucleons, these processes are $n \to p\, e^-\bar{\nu}_e$ and $p \to n\, e^+\nu_e$ respectively, producing electron-flavoured neutrinos complementary to the muon-flavoured ones from superbeams. A beta beam experiment [20] consists of a source of radioactive ions, which are then accelerated to a ring where they decay, emitting neutrino beams in their direction of travel. The decay ring incorporates straight sections pointing at the desired detector sites.
- **Neutrino factories**. Although muons are discarded in conventional neutrino beams, they also undergo a neutrino-producing decay $\mu^+ \to \bar{\nu}_\mu e^+\nu_e$ (and its charge conjugate), which is used in the proposed *neutrino factory* [21]. This produces neutrinos of two flavours with equal fluxes, which is a useful calibration. The muons survive for longer than pions, which provides (just) enough time for them to be accelerated, resulting in a more collimated beam from a decay ring similar to beta beams.
  - **Low-energy neutrino factories**. Removing some of the muon acceleration system from the conventional neutrino factory gives a smaller, cheaper facility that produces lower-energy neutrinos [22]. Detector technology can be optimised to perform better for these low energy interactions.

Figure 1.2 compares the sensitivities of these different facilities to neutrino oscillation parameters that have not currently been measured to high precision. The angle θ_{13} and CP-violating δ could still be zero consistent with current data and although there are some limits on $|\Delta m^2_{31}|$, it is not known whether the difference between the masses is positive or negative. A full discussion of the physics reach of the neutrino factory and other experiments is available in the ISS physics report [15]. The neutrino factory provides the best performance but requires the largest and most complex facility.

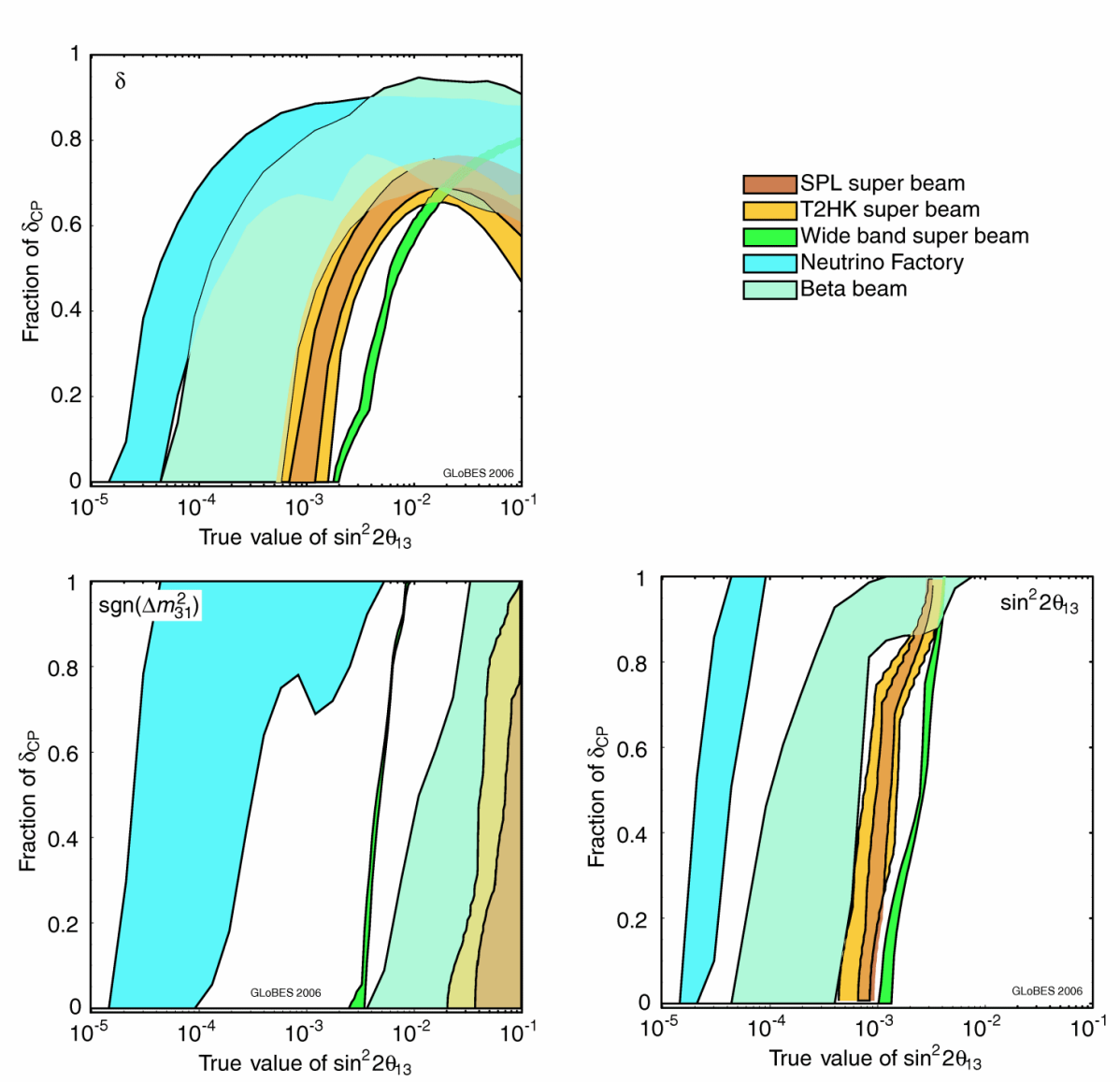


Figure 1.2: Sensitivity of neutrino beam facilities to $\delta \neq 0$ (top left), the sign of Δm^2_{31} (bottom left) and $\theta_{13} \neq 0$ (bottom right). The measurements become more difficult as θ_{13} approaches zero and for certain values of δ. Values to the right and below each coloured band can definitely be measured at that facility and within the band it depends on the specific variant of the design used. The low-energy neutrino factory (not shown) performs comparably to the beta beam, as shown in figure 21 of [23].

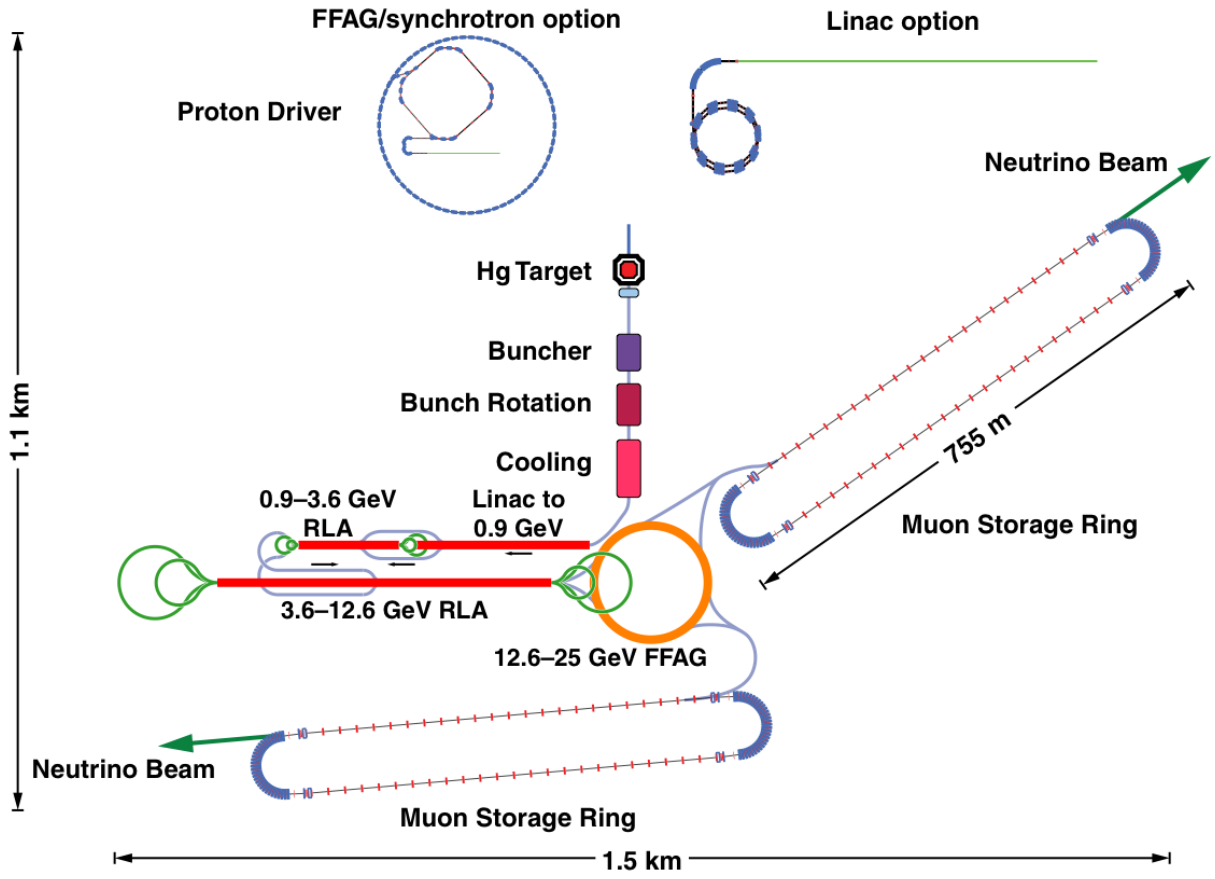


Figure 1.3: Possible neutrino factory layouts, drawn by C.R. Prior based on the conclusions of the ISS [24].

1.3 The Neutrino Factory

The *neutrino factory* is an accelerator-based neutrino source [21] intended to measure the mass splitting Δm^2_{31} and entries of the MNS matrix including $U_{e3} = e^{-i\delta} \sin\theta_{13}$, where θ_{13} is the ν_1–ν_3 mixing angle. As presented on pages 21–22 of [15], the best fit value of Δm^2_{31} using data available at the time of publication was $2.6 \times 10^{-3}\,(\mathrm{eV/c^2})^2$. Using this in the baseline formula derived in section 1.1 (and assuming matter effects are small) gives a length of 477 km per GeV of neutrino energy, suggesting that neutrinos of up to roughly 27 GeV can be used with baselines that form a chord across the Earth's interior. The high energy allows a directed neutrino beam to be created by the Lorentz-boosted decay of $\mu^+ \to \bar{\nu}_\mu e^+ \nu_e$ (and its charge conjugate), where the decay products acquire 105 MeV of kinetic energy in the muon rest frame. We therefore expect a neutrino beam of divergence 2–5 mrad when a beam of muons decays in-flight at these energies. This angle is proportional to the relativistic factor $1/\gamma$, so the flux at the detector is a factor of $\gamma^2 \sim 10^5$ above isotropic emission from sources such as nuclear reactors. A fast muon beam produces neutrinos with energies almost uniformly distributed between zero and the beam energy, so muons of energy 20–50 GeV are required.

The neutrino factory aims to produce 10^{21} muon decays per year [1], making it sufficiently sensitive to the mixing angle θ_{13}, which is currently constrained only by reactor

experiments. A schematic of a neutrino factory is shown in figure 1.3. So far, the most complete design study of a neutrino factory has been the Muon Collider Collaboration's Feasibility Study II [25] in 2001, which was updated with international consensus in 2007 by the International Scoping Study (ISS) [24].

It is difficult to produce an intense beam of muons at 20–50 GeV: muons have a mean life of only $\tau_\mu = 2.2\,\mu$s and must be produced as secondary particles from another accelerator. Multi-GeV protons are directed onto a high-Z target to maximise the number of strong interactions that produce pions. Each pion then decays to a muon ($\pi^+ \to \mu^+\nu_\mu$) in an average time of 26 ns. The mean beam power of the proton accelerator must be at least 4 MW for sufficient flux and additionally it must be pulsed to provide extracted bunches of a few ns duration for the time-dependent muon capture systems downstream. An accelerator with these properties is referred to as a proton driver [26].

The rate of neutrinos of each flavour produced is equal to the rate of muons decaying in the storage ring, given by

$$R_\mu = R_p \; e_{\pi/p} \; e_{\text{machine}} \; e^{-\tau/\tau_\mu}$$

(achieving 10^{21} per operational year requires $R_\mu = 10^{14}\,\text{s}^{-1}$). Here R_p is the rate of protons hitting the target, $e_{\pi/p}$ is the number of pions produced per incident proton and τ is the proper time experienced by a muon while going through the acceleration chain. The remaining term can be expressed as the product of the muon capture, acceleration and storage efficiencies:

$$e_{\text{machine}} = e_{\text{front}} e_{\text{acc}} e_{\text{store}}.$$

The work in this report mainly focusses on improving the efficiency of the target $e_{\pi/p}$ and the muon front end e_{front}. However, most terms in the equation cannot be considered in isolation: at constant beam power the proton energy will affect R_p and $e_{\pi/p}$ jointly. Also, the muon decay in the simulations described in this thesis is not handled in continuum as suggested by the equation. Instead, each muon decays individually, so the factor $e^{-\tau_{\text{front}}/\tau_\mu}$ will automatically be included in results from front end simulations.

Particular engineering challenges are involved in:

- **The muon front end**. The target creates a large, high divergence beam and pion decays produce additional momentum variation. Typically only a few percent can be captured both transversely and longitudinally. Superconducting solenoids of several Tesla are required for transverse capture, while RF phase rotation systems or magnetic chicanes are options for longitudinal matching.

- **Muon acceleration**. The lifetime of the muon is extended to $\gamma\tau_\mu$ in the laboratory frame, so time spent at low energies should be minimised and acceleration must be rapid. Conventional synchrotrons typically take at least 10 ms for the magnetic field to ramp from low to high, so are far too slow. Therefore fixed-field, multiple-pass machines such as RLA (Recirculating Linac) and FFAG (Fixed-Field Alternating Gradient) accelerators are under consideration. However, the most rapid acceleration possible is always achieved by a straight linac, which is used as the first stage.

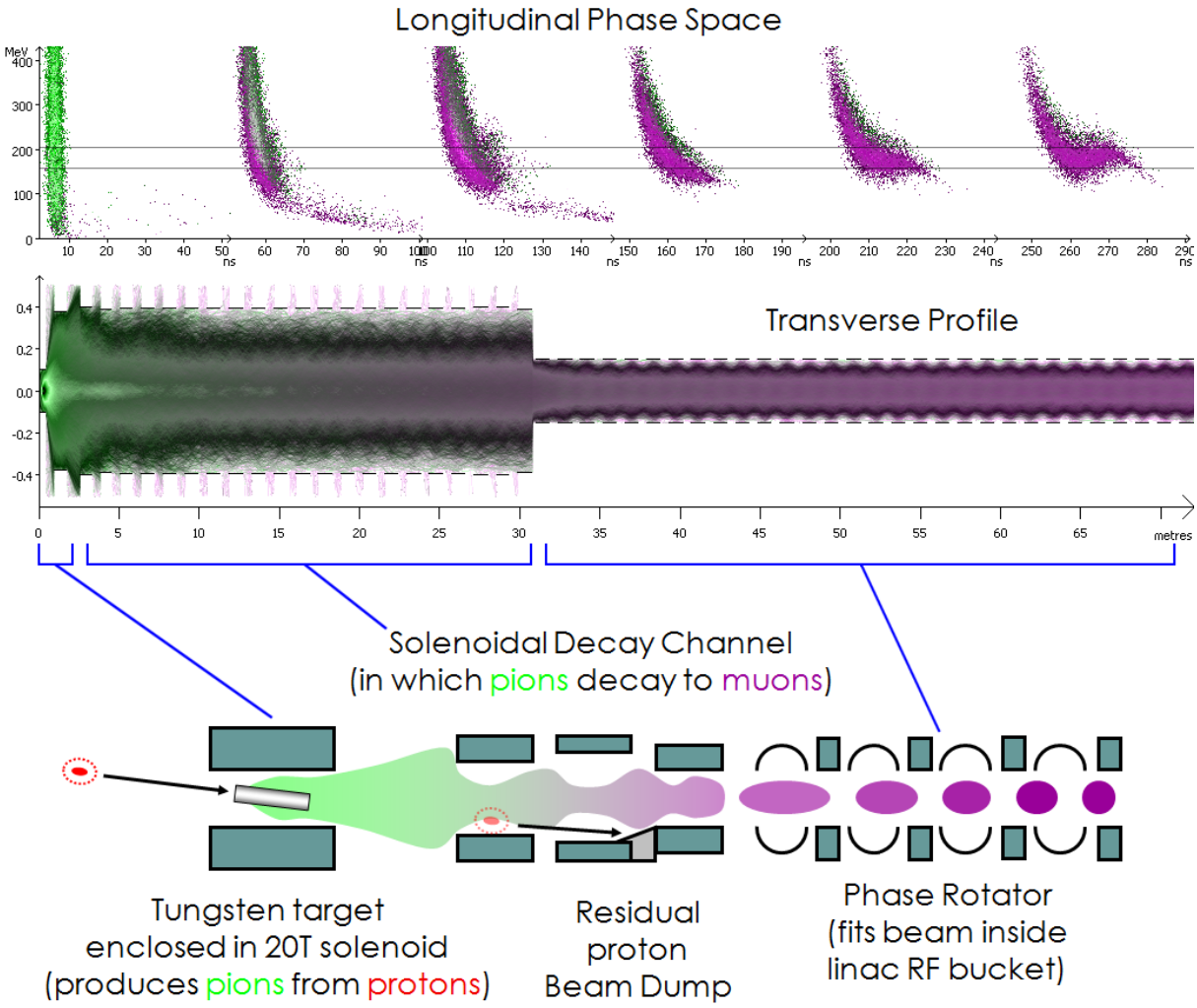


Figure 1.4: Parts of the muon front end as investigated in this thesis, in simulation (top) and schematic form (bottom). The phase rotator may work in either the bunching or antibunching direction depending on whether a bending chicane or RF cavities are used.

1.4 Muon Front End Structure

The *muon front end* is normally defined to be the piece of neutrino factory beamline extending from the target (which the protons hit) up to the start of muon acceleration. Most designs start with a solenoidal focussing channel (the *decay channel*) to capture pions from the target as they decay to muons, including a powerful solenoid of up to 20 Tesla (20 T) enclosing the target itself. Longitudinal (energy vs. time of flight) phase space is then manipulated, normally after pion decay, to have a beam suitable for the muon accelerator. This second process is known as *phase rotation*. A typical front end studied in this thesis is shown in figure 1.4.

Each stage of the front end improves one aspect of the beam distribution, which is initially a multidirectional spray of various particles including pions from the target (see sections 3.3 and 4.1.1). The steps listed below prepare a beam of muons that can be accelerated to the storage ring with low losses apart from the unavoidable muon decays.

- **The decay channel**. A strong magnetic field causes charged particles to spiral around the field lines instead of continuing in a straight line. This principle is used to deal with the particles emitted from the target, which can have hundreds of MeV/c of transverse momentum, making angles of tens of degrees with the forward direction. A field of 20 T along the accelerator axis can keep a large fraction of the particles within 10 cm of the axis, although it will not reverse the direction of particles emitted with a backward longitudinal momentum component. In a constant field the particles would continue on perfect helices with their large transverse momentum being conserved in magnitude. However, if the field decreases along the axis, the zero divergence property of magnetic fields implies there must be an outward-pointing component of the field vector. This produces a force on charged particles in the direction opposing their natural direction of spiralling in the solenoid field, decreasing the transverse part of their momentum. Since particle energy is conserved in magnetic fields, the decrease in transverse momentum must be accompanied by an increase in longitudinal (forward) momentum. All this means that if the magnetic field decreases significantly following the target, the transverse momentum and angle of travel relative to the accelerator axis will both be reduced. A magnetic field of some magnitude is still required to confine the particles in later sections, with decay channel fields in this thesis typically reducing to 3–4 T. Superconducting solenoids can produce the above sequence of fields.

  Out of the charged particles from the target, pions are the primary decay parents of muons, with kaons contributing a few percent when present. The pions have the longer mean lifetime of the two of roughly 26 ns [27] in their rest frame and the forward pions have relativistic factors of $\gamma \sim 3$, making the mean flight distance $\sim$20 metres before decay. This means that most decay channels are 30 metres long or more, though variation of this was allowed in the design process. The decay imparts a non-negligible momentum kick (tens of MeV/c) to the daughter muon, causing the beam to diverge while decays are still in progress. The end product of the decay channel is a beam of predominantly muons, with little decay activity, transversely confined within a channel of solenoids. Longitudinally the beam has 'debunched' since being produced at the target, i.e. the particles continue at different forward speeds, spreading out as time progresses.

- **Phase rotation**. Periodic RF systems such as those in the muon linac require the beam to occupy a certain stable region of the energy vs. arrival time phase space known as an *RF bucket*. Particles within this region will oscillate around the centre of the bucket, which follows the acceleration waveform, whereas those outside will lose synchronisation and not be fully accelerated. The bucket has a time span inversely proportional to the RF frequency and an energy range proportional to the applied voltage, but is always 'upright' in the sense that correlation is not desired between time and energy. The beam from the decay channel has a very strong correlation since the highest energy muons will arrive first: as such, it occupies greater time and energy ranges than its phase space area requires and attempting direct acceleration of it would result in RF voltages higher (and more expensive) than necessary.

  The phase rotation section removes this correlation by systematically offsetting the energy or time of arrival of the particles. Energy offset is achieved by another RF system before the linac that is phased to decelerate the leading muons and accelerate

the stragglers in the bunch (schemes exist with one or many bunches), leading to a smaller energy range at the output. Time-of-arrival offset, known as *reverse* phase rotation, uses magnetic bends to put the higher energy particles on longer paths, reducing the time spread at the output. Variants of both schemes are studied in this thesis: a reverse phase rotation scheme in section 8.2 and RF phase rotators in section 8.3 onwards.

The phase rotation channel typically continues the transverse beam confinement using solenoids as in the decay channel, with the RF cavities placed between them. The resulting beam has the stable transverse optics inherited from the decay channel in addition to a decorrelated longitudinal phase space.

- **Ionisation cooling**. For some proposed muon accelerators, such as FFAGs, the beam from the decay channel is still too large to fit in the *dynamic aperture* of the accelerator's magnets, within which the transverse dynamics are stable. Scattering of muons through materials placed in the beamline is used to drain transverse momentum, with RF acceleration replenishing forward momentum, leading to a smaller beam. The interaction is predominantly by knocking electrons from the atoms of the material, hence the name *muon ionisation cooling*. This technique is not used in the front ends studied herein, though chapter 6 explains the principle further and constructs a computer model of this beam scattering. The resulting transversely smaller beam would be compatible with smaller aperture muon linacs.

The UK neutrino factory design differs slightly from the ISS baseline [24] because the UK team have been investigating the possibility of a solid metal target instead of a mercury jet [28, 29]. The performance in terms of pion yield is quite similar between the two approaches, metals such as tungsten and tantalum being close to mercury in the periodic table, though some adaptations of the decay channel are necessary such as locating the residual proton beam dump further downstream (see section 3.5). Design issues related to this change are discussed throughout chapter 3.

The front ends in this thesis also differ from other baselines because they do not assume a muon ionisation cooling channel is used to reduce the beam size before entering the acceleration systems. A buncher and cooling system *may* be added to make the beam small enough to go through the optics of FFAGs but this is not attempted. Instead, it is suggested that a linear muon accelerator with a larger acceptance may be a practical option all the way to the decay ring energy (25 GeV in the ISS). The reasoning is that the neutrino factory shown in figure 1.3 requires building an estimated 2.04 km of beamline from the target to the storage rings, excluding transfer lines (0.87 km of this is in the eight RLA teardrop arcs). If the average gradient of the first ISS RLA's accelerating section (8.5 MV/m) was used to cover the range 0.9–25 GeV, that would require 2.84 km of linac to be added to a front end of 0.34 km given in section 8.5, for a total of 3.18 km and no internal transfer lines. The relativistic shrinkage of the beam in the high-energy half of the linac may allow this part to be shortened further, either with compact transverse optics or through high-gradient cavities (see end of section 8.1). Thus the linac option is not greatly longer than the multi-technique system, while the use of fewer different technologies could potentially lead to higher reliability or cost savings during manufacturing.

1.5 Guide to This Thesis

An automated approach is used to search for good front end designs. The basis for this is a particle tracking code called Muon1 [5], which calculates the number of muons that survive a given configuration of magnets and RF cavities.	Chapter 2
Tracking requires an initial distribution of pions from the target, which can be calculated with hadronic interaction codes such as MARS15 [2] used here. The target itself can be varied with a view to improving yields, so after a discussion of the design issues,	Chapter 3
the target parameters are optimised	Chapter 4
and the resulting pion production checked against some experimental data and another code.	Chapter 5
A module for simulating muon cooling in Muon1 is also developed.	Chapter 6
With the tracking code ready to use, an optimiser is constructed that loops Muon1 to search a specified design space for good muon yields using a genetic algorithm. This is deployed as a distributed computing project.	Chapter 7
Finally, different approaches for the front end and early muon linac are run as different design spaces in the distributed optimiser and the resulting best lattices discussed.	Chapter 8

Chapter 2

Particle Tracking Code (Muon1)

This chapter explains the principles of particle tracking as used in the Muon1 code [5], including some analysis of error sources. Models for magnetic fields are derived, together with algorithms for generating particle decays and locating collisions. The more specific aspects of specifying the accelerator as input to the code are discussed in section 7.1.

Muon1 was written beginning in 2001, at a time when two other codes for muon front end simulations, ICOOL [30] and G4MICE [31] were in development. ICOOL had many of the required features but tracked particles with beamline distance as the independent variable. It was observed that some pions emitted from the target went backwards (see section 4.1.1), so a code that tracked in time was preferred to investigate their behaviour including magnetic bottle or mirror effects around the target solenoids. G4MICE, a combination of GEANT4 [3] tracking and materials with magnetic field models for solenoids, did track in time but many of the GEANT4 material models proved not to be as accurate as those in other codes (see sections 5.1 and 6.3). It also had many external library dependencies and did not lend itself to a self-contained small distribution that could be downloaded and installed on Windows, as was used for the distributed computing project in chapter 7. Finally, neither of the other two codes had genetic algorithm optimisation capabilities, although these could conceivably have been wrapped around them in some way. Muon1 provided a more integrated package.

2.1 Equations of Motion

The force on a particle of charge q travelling through electric and magnetic fields $\mathbf{E}(\mathbf{x},t)$ and $\mathbf{B}(\mathbf{x},t)$ is given by the Lorentz law:

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}),$$

where $\mathbf{p}$ is the particle momentum. The time integration of this equation is complicated somewhat by the relativistic definition of momentum:

$$\begin{aligned} \mathbf{p} &= m\gamma\mathbf{v} \\ \gamma &= \gamma(v) = \frac{1}{\sqrt{1-(v/c)^2}}, \end{aligned}$$

where $v = |\mathbf{v}|$ and the particle has rest mass m. Particles must be represented by a *phase space* vector that encodes information about both their position and velocity (or momentum). There is some choice of how to parameterise the phase space, though it always has $3 + 3 = 6$ variables. The Muon1 tracking code represents position by a cartesian vector $\mathbf{x}$ relative to a fixed origin, the other options being some sort of polar coordinate system and/or changing coordinate system to align with the axis of the accelerator components. This is in the laboratory rest frame, another choice would have been to move the origin with some "central particle" in the beam. Finally, some codes when in the accelerator-aligned frame use "z" as the independent variable instead of time and store the arrival times of particles at a particular plane. These alternative techniques can provide faster tracking codes in some cases (for instance when tracking in z only requires a single component to be considered at a time) with variables that are more physically interpretable. Because the simulation of the muon front end starts from a pion production target, with particles going in many directions including backwards, a 3D approach that makes no assumptions is favoured for its generality. The large velocity spreads, even when going forwards, limit the usefulness of taking a frame co-moving with the beam.

For the three other components of the phase space vector, $\gamma\mathbf{v}$ is chosen. Early versions of the code used $\mathbf{v}$ but this started to lose accuracy for higher-energy particles. Since $E_t = m\gamma c^2$, inverting the expression for $\gamma(v)$ gives the velocity as a function of the particle's total energy:

$$v = c\sqrt{1 - \gamma^{-2}} = c\sqrt{1 - (E_t/mc^2)^{-2}}.$$

The -2 index on the ratio of the total energy to the rest energy means the amount subtracted from 1 becomes small at a quadratic rate, after which the square root moves it even closer to unity (and v to c). This closeness results in poor energy resolution when $\mathbf{v}$ is numerically quantised to triplets of floating point numbers. The vector $\gamma\mathbf{v}$ (actually equal to $\mathbf{p}/m$, so proportional to momentum) does not suffer from this problem and tends to infinity for larger energies, maintaining resolution in energy as seen in the ultra-relativistic limit $|\gamma\mathbf{v}| \simeq \gamma c = E_t/mc$ where the quantities become proportional. For low speeds $\gamma \simeq 1$ and $\gamma\mathbf{v} \simeq \mathbf{v}$ approaches the classical velocity.

It will be useful to have expressions for the derivative of the phase space vector $(\mathbf{x}, \gamma\mathbf{v})$ with respect to laboratory time t (the independent variable). By definition,

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{v} = \frac{\gamma\mathbf{v}}{\gamma}$$

but the γ in the denominator must be calculated from the speed information already stored in the phase space vector (i.e. $|\gamma\mathbf{v}| = \gamma v$) this time, not just from v as before. Although this can be done purely by algebraic manipulation of the expression for $\gamma(v)$, which produces quite messy formulae that later cancel to something simple, a physics argument can give the result directly. Start with the (classical, relativistic) total energy equation for a moving particle

$$E_t^2 = (pc)^2 + (mc^2)^2,$$

where $p = |\mathbf{p}|$ and note that $p = m\gamma v$ and $E_t = m\gamma c^2$. Substituting and then cancelling mc from within the brackets gives

$$(m\gamma c^2)^2 \;=\; (m\gamma vc)^2 + (mc^2)^2$$

$$\begin{aligned} \Rightarrow \quad (\gamma c)^2 &= (\gamma v)^2 + c^2 \\ \Rightarrow \quad \gamma &= \frac{\sqrt{(\gamma v)^2 + c^2}}{c}, \end{aligned}$$

which can be used as the expression for γ in

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \frac{\gamma\mathbf{v}}{\gamma} = \frac{\gamma\mathbf{v}c}{\sqrt{|\gamma\mathbf{v}|^2 + c^2}}.$$

The rate of change of the other component comes from the force law as opposed to kinematics:

$$\frac{\mathrm{d}(\gamma\mathbf{v})}{\mathrm{d}t} = \frac{\mathrm{d}(\mathbf{p}/m)}{\mathrm{d}t} = \frac{\mathbf{F}}{m} = \frac{q}{m}(\mathbf{E} + \mathbf{v} \times \mathbf{B}).$$

There is again the problem of having $\mathbf{v}$ instead of $\gamma\mathbf{v}$. However, the program can be arranged to calculate $\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}$ first and use the stored result in the expression

$$\frac{\mathrm{d}(\gamma\mathbf{v})}{\mathrm{d}t} = \frac{q}{m}\left(\mathbf{E} + \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} \times \mathbf{B}\right).$$

The complete time derivative of the phase space vector for the tracking problem can now be written out explicitly:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbf{x} \\ \gamma\mathbf{v} \end{bmatrix} = \begin{bmatrix} \frac{\gamma\mathbf{v}c}{\sqrt{|\gamma\mathbf{v}|^2+c^2}} \\ \frac{q}{m}\left(\mathbf{E}(\mathbf{x},t) + \frac{\gamma\mathbf{v}c}{\sqrt{|\gamma\mathbf{v}|^2+c^2}} \times \mathbf{B}(\mathbf{x},t)\right) \end{bmatrix},$$

where the functional dependencies have become unsuppressed to demonstrate how each element of the derivative depends on elements of the phase space vector and only one other variable, t. This last external dependency can be eliminated by making the current laboratory time part of the particle's vector, though in a real program this would not be stored repeatedly for each particle! As a result, the derivative becomes a proper function of the vector:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbf{x} \\ \gamma\mathbf{v} \\ t \end{bmatrix} = \mathbf{f}\left(\begin{bmatrix} \mathbf{x} \\ \gamma\mathbf{v} \\ t \end{bmatrix}\right) = \begin{bmatrix} \frac{\gamma\mathbf{v}c}{\sqrt{|\gamma\mathbf{v}|^2+c^2}} \\ \frac{q}{m}\left(\mathbf{E}(\mathbf{x},t) + \frac{\gamma\mathbf{v}c}{\sqrt{|\gamma\mathbf{v}|^2+c^2}} \times \mathbf{B}(\mathbf{x},t)\right) \\ 1 \end{bmatrix}.$$

2.2 Runge–Kutta Methods

The previous section produced a "time integration" problem of finding the solution $\mathbf{s}(t)$ of a differential equation with generic form

$$\dot{\mathbf{s}} = \mathbf{f}(\mathbf{s}),$$

where $\dot{\mathbf{s}}$ denotes the time derivative ($\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t}$) and the initial condition $\mathbf{s}(t_0)$ is specified to make the solution unique. In cases where there is not an exact analytic solution for the trajectory, as is usually the case for realistic magnetic fields, the calculation can proceed in a series of small timesteps, each of which calculates a numerical approximation to $\mathbf{s}(t+\delta t)$ from $\mathbf{s}(t)$.

Note that incremental time-stepping is generally the slowest class of method for trajectory calculation. In lower dimensionality cases (e.g. the two-dimensional phase space of longitudinal dynamics) the mapping $\mathbf{s}(t_0) \to \mathbf{s}(t_1)$ can be approximated by evaluating comparatively few tracks and then interpolating, so tracking can be avoided for the majority of the particles. Additionally, for beams that travel very close to the machine axis, power expansions in deviations from a central trajectory can give good results even for the full phase space (see for instance [32]). However, the pion beam to be captured in a neutrino factory front end has a very high emittance that is intrinsic to the way it is generated from nuclear interactions in the target, so particles may move into regions far from the central trajectory where a truncated power series is no longer a good approximation.

2.2.1 Principle

Considering a small but finite timestep of length δt, the definition of the derivative as a limit immediately suggests the following approximation:

$$\begin{aligned} \frac{\mathbf{s}(t+\delta t) - \mathbf{s}(t)}{\delta t} &\simeq \dot{\mathbf{s}}(t) \\ \Rightarrow \quad \mathbf{s}(t+\delta t) &\simeq \mathbf{s}(t) + \delta t\ \dot{\mathbf{s}}(t) \\ &= \mathbf{s}(t) + \delta t\ \mathbf{f}(\mathbf{s}(t)). \end{aligned}$$

This is known as Euler's forward difference method and is the simplest numerical integration method available for these problems. To obtain information about the error behaviour of this method, Taylor expand $\mathbf{s}(t+\delta t)$ about t:

$$\begin{aligned} \mathit{Error} &= \mathbf{s}(t) + \delta t\ \mathbf{f}(\mathbf{s}(t)) - \mathbf{s}(t+\delta t) \\ &= \mathbf{s}(t) + \delta t\ \dot{\mathbf{s}}(t) - (\mathbf{s}(t) + \delta t\ \dot{\mathbf{s}}(t) + \tfrac{1}{2}\delta t^2\ \ddot{\mathbf{s}}(t) + O(\delta t^3)) \\ &= -\tfrac{1}{2}\delta t^2\ \ddot{\mathbf{s}}(t) + O(\delta t^3), \end{aligned}$$

which shows that the error in a single timestep is proportional to δt^2 (plus terms that are $O(\delta t^3)$, i.e. order δt^3 and higher). Tracking through the entire time interval requires $(t_1 - t_0)/\delta t$ of these steps, giving a final error proportional to $\delta t\ (t_1 - t_0)$ provided the errors are additive. Because the rate of error accumulation is generally proportional to δt^1, Euler integration is known as a *first-order method*. This means that to increase its

tracking accuracy by an order of magnitude, an order of magnitude more computing time is required, which leads to very large compute times to produce accurate results.

Fortunately there are higher-order methods available that cancel more of the Taylor expansion of $\mathbf{s}(t+\delta t)$, leaving only higher powers of δt remaining. A simple example of how this can work is the method that uses $\dot{\mathbf{s}}$ from the central location $t+\frac{1}{2}\delta t$ instead of from t to determine the direction of the step, which gives it second-order (step error $\propto \delta t^3$) accuracy as the $\ddot{\mathbf{s}}$ term is now taken account of. The central location itself has to be worked out from an Euler step as shown below.

$$\begin{aligned}
\mathbf{s}(t+\delta t) &= \mathbf{s}(t) + \delta t\ \dot{\mathbf{s}}(t+\tfrac{1}{2}\delta t) + O(\delta t^3) \\
&= \mathbf{s}(t) + \delta t\ \mathbf{f}\Big(\mathbf{s}(t+\tfrac{1}{2}\delta t)\Big) + O(\delta t^3) \\
&= \mathbf{s}(t) + \delta t\ \mathbf{f}\Big(\mathbf{s}(t) + \tfrac{1}{2}\delta t\ \dot{\mathbf{s}}(t) + O(\delta t^2)\Big) + O(\delta t^3) \\
&= \mathbf{s}(t) + \delta t\ \mathbf{f}\Big(\mathbf{s}(t) + \tfrac{1}{2}\delta t\ \mathbf{f}(\mathbf{s}(t))\Big) + O(\delta t^3)
\end{aligned}$$

Notice how the approximation is expressed in a form that depends only on $\mathbf{s}$ evaluated at the current time (and not any other time, or its derivatives) and evaluations of the known function $\mathbf{f}$. This ensures that it is computable and leads to the general form of an N-step Runge–Kutta method:

$$\begin{aligned}
\mathbf{s}_0 &= \mathbf{s}(t); \\
\mathbf{s}_n &= \mathbf{s}(t) + \delta t\ \mathbf{f}\left(\sum_{i=0}^{n-1} \alpha_{ni}\mathbf{s}_i\right) \qquad \text{for } 1 \le n \le N; \\
\mathbf{s}(t+\delta t) &\simeq \sum_{i=0}^{N} \beta_i \mathbf{s}_i,
\end{aligned}$$

where the coefficients α_{ni} and β_i can be carefully chosen to cancel lower powers of δt in the error expression, leaving a method with a high order of accuracy. This technique was originally developed at the end of the 19$^{\text{th}}$ century for solving differential equations by hand [33] but for a more modern introduction with applications to computer programming, see [34].

The Muon1 code uses the fourth-order Runge–Kutta method defined as follows.

$$\begin{aligned}
\mathbf{s}_0 &= \mathbf{s}(t) \\
\mathbf{s}_1 &= \mathbf{s}(t) + \delta t\ \mathbf{f}(\mathbf{s}_0) \\
\mathbf{s}_2 &= \mathbf{s}(t) + \delta t\ \mathbf{f}(\tfrac{1}{2}\mathbf{s}_0 + \tfrac{1}{2}\mathbf{s}_1) \\
\mathbf{s}_3 &= \mathbf{s}(t) + \delta t\ \mathbf{f}(\tfrac{1}{2}\mathbf{s}_0 + \tfrac{1}{2}\mathbf{s}_2) \\
\mathbf{s}_4 &= \mathbf{s}(t) + \delta t\ \mathbf{f}(\mathbf{s}_3) \\
\mathbf{s}(t+\delta t) &\simeq \tfrac{1}{6}\mathbf{s}_1 + \tfrac{1}{3}\mathbf{s}_2 + \tfrac{1}{3}\mathbf{s}_3 + \tfrac{1}{6}\mathbf{s}_4
\end{aligned}$$

This order is chosen because of the relatively rapid convergence with decreasing step size and the fact that increasing to fifth order requires an additional two calculation steps, as opposed to orders up to 4 where only one more step is required for each successive order.

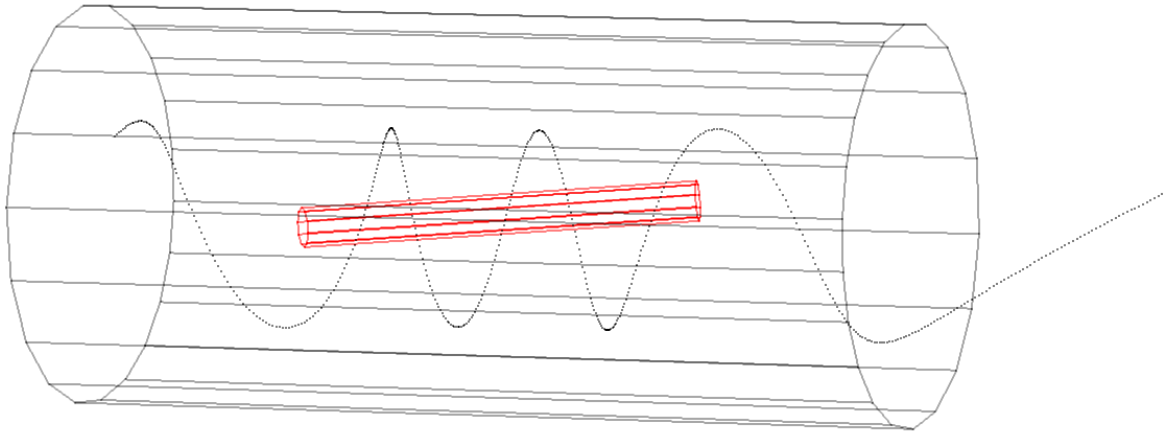


Figure 2.1: Path of the muon in the solenoid test problem.

2.2.2 Single Particle Error Study

The two main deviations from the ideal higher-order error behaviour in accelerator applications are that:

1. The errors are not always additive. When tracked paths diverge from each other, errors can be amplified on a timescale determined by the field gradients in the machine. A simple case of this is in a defocusing quadrupole magnet where off-axis distance and hence distances between particles increase exponentially.

2. In a tracking code, $\mathbf{f}(\mathbf{s})$ is not always analytic. Often the input for a code will specify $\mathbf{B}(\mathbf{x})$ as a 'field map' with values defined at the points of a grid. The interpolation used to generate field values between these points is generally not completely smooth, so steps that straddle the border will acquire an error of a lower power of δt than intended.

The first of these in particular makes creating a useful error analysis applicable to the general case impossible. Instead, some empirical tracking must be done when fixing a timestep, to ensure it is appropriate for the problem at hand. Below, this is done for two cases: one tracking through a solenoid field and the other tracking through a complex field map extracted from an OPERA [35] model of a muon cooling ring. In both of these, the final particle position and velocity obtained from timesteps ranging from 10^{-12} to 10^{-8} seconds is compared to an 'accurate' result using 10^{-13} seconds.

Test 1: Solenoid with Fringe Field

In this test an 180 MeV muon enters a B_∞=20 T solenoid that is 40 cm long and 10 cm in radius. The 20 T value is only an asymptotic central field for an infinitely long sheet coil solenoid with the same current density. The actual field calculation starts with the analytic expression for the on-axis field, $B_z = \frac{1}{2}B_\infty\left(\frac{a}{\sqrt{a^2+1}} + \frac{b}{\sqrt{b^2+1}}\right)$ where $a = z/R_{\text{sol}}$

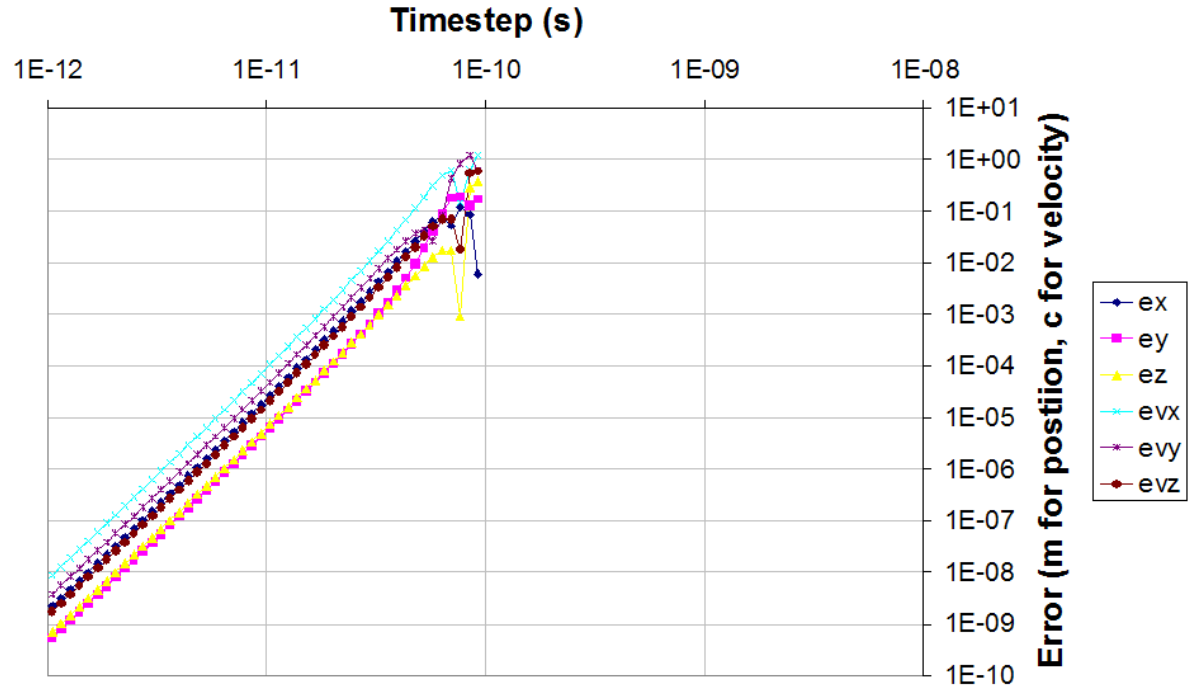


Figure 2.2: Dependence of errors on timestep size for the solenoid test problem.

and $b = \frac{L_{\text{sol}}-z}{R_{\text{sol}}}$. This is extrapolated off the central axis by a cubic power expansion in r derived from imposing $\boldsymbol{\nabla} \cdot \mathbf{B} = \boldsymbol{\nabla} \times \mathbf{B} = 0$ on a cylindrically symmetric problem (this is fully described in section 2.3.1). The muon is tracked for 5 ns, starting (3 cm,2 cm) off-axis at the entrance plane of the solenoid and moving at an angle of 48.5° to the central axis, so that it spirals around and then leaves via the other end (see figure 2.1).

Figure 2.2 shows the errors in the final phase space coordinates for the range of timesteps investigated. The right-hand side of the figure is blank because the muon hit the solenoid and was lost for those timesteps. The larger timesteps do not work due to the confined nature of this problem: the muon would travel from one side of the solenoid bore to the other in about 1 ns, so the timestep must be much smaller than that. Additionally, it must be much smaller than the Larmor period ($2\pi m_0\gamma/qB = 0.9985$ ns here) of its helical path within the solenoid, which is also short due to the large 20 T field.

Most striking is the very simple dependence of the error on the timestep below some initial irregularities. This comes purely from the δt^4 leading order error term and can be seen from the graph to decrease exactly four orders of magnitude for each order of magnitude in timestep. This dependency is so clean owing to the completely analytic nature of the $\mathbf{B}$ field that was used, for just one solenoid with no cut-offs or other irregularities.

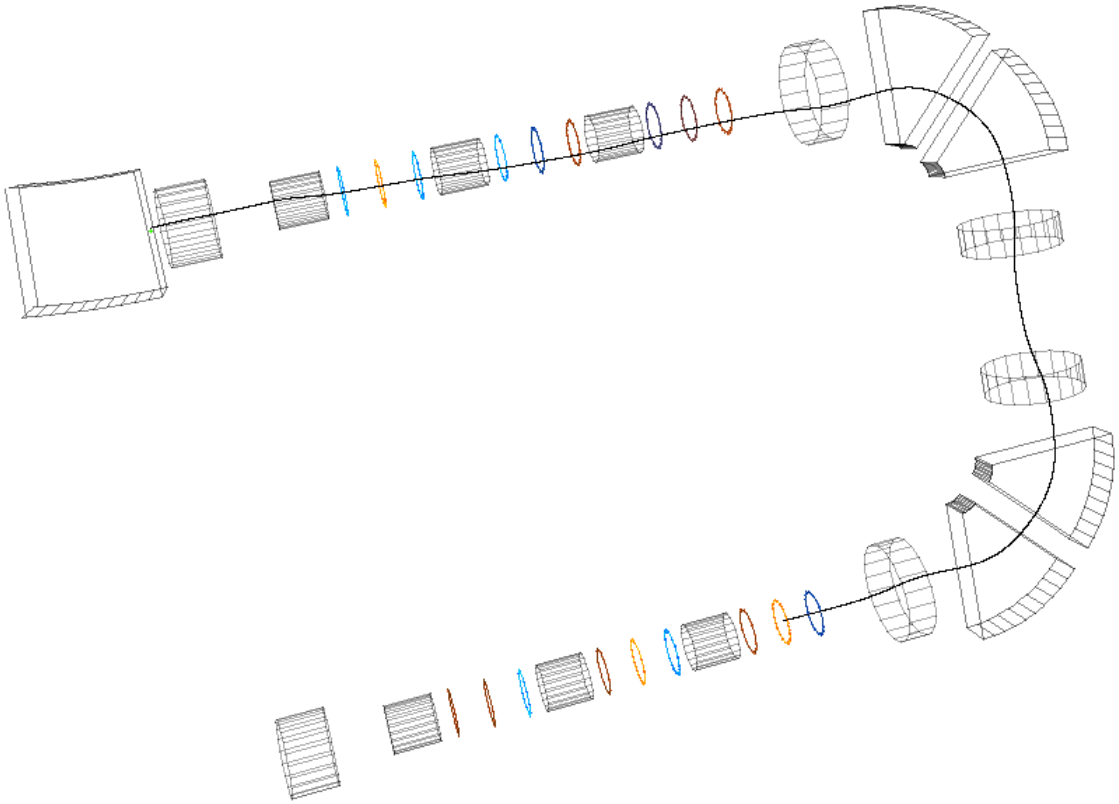


Figure 2.3: Muon path through the cooling ring field maps.

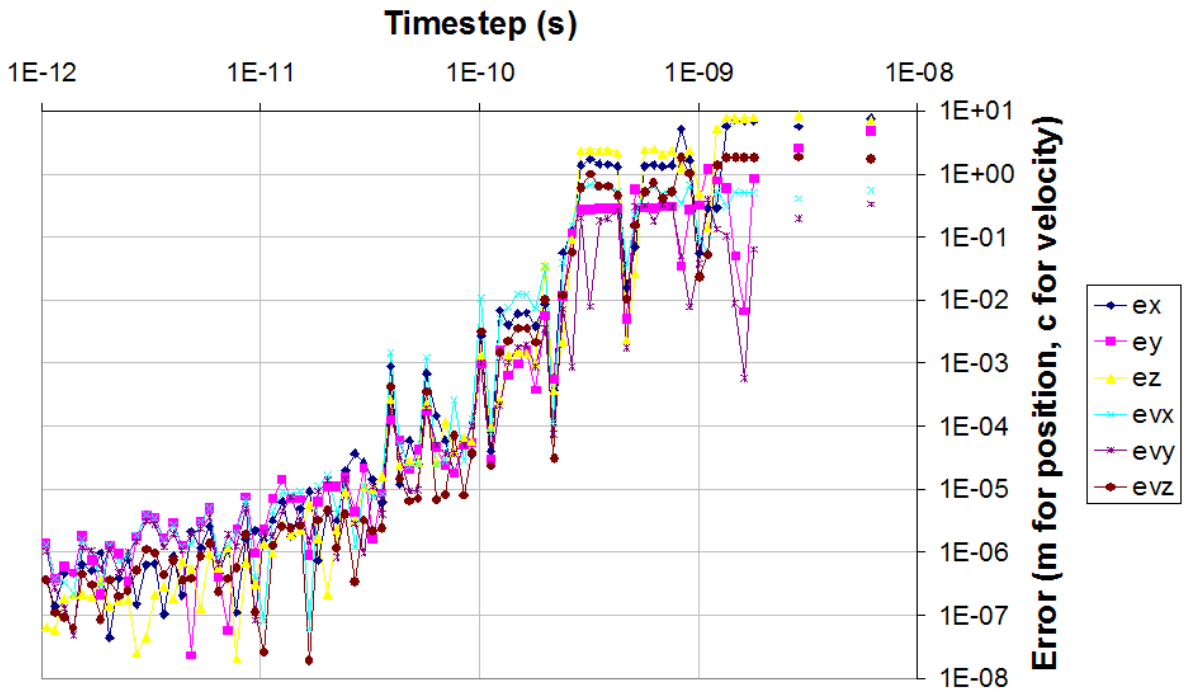


Figure 2.4: Dependence of errors on timestep size in the cooling ring field map test problem.

Test 2: Field Maps from OPERA

Here another 180 MeV muon is tracked for 50 ns through roughly one third of a turn in the Rees cooling ring† [36]. It starts (3 cm,2 cm) off-axis, but parallel, to test tracking of the transverse oscillations in the orbit visible in figure 2.3. The maps were generated in the magnetostatic code OPERA-3d and sampled at cubic grids of 5 cm spacing. Trilinear interpolation was used between the eight grid points at the corners of each cube to obtain the field values in the continuum. This gives a continuous **B** field but it is not differentiable at the cube boundaries.

The scale of the problem is larger here (metres) so many of the longer timesteps do work to some extent and now appear on the graph (figure 2.4). An initial steep decrease (possibly $\propto \delta t^4$) gives way to a slower decline as the error becomes dominated by lower terms. A δt^2 term is expected from the fixed number of crossings of field map cube boundaries, each of length δt and making a mis-estimate of the field proportional to that length, giving δt^2 overall. However, the final falloff looks even slower than that. An error of order δt will come from any field discontinuities, such as edges of the field maps where the field is nonzero. There, the mis-estimate is of fixed magnitude so the error decreases only as δt, which dominates as δt tends to zero.

The timestep of 10^{-11} seconds (0.01 ns) has been chosen as the default for Muon1 as this gives position errors of 25 μm in the first test and below 10 μm in the second, which are comparable to magnet alignment tolerances. The velocity errors decrease in step with the positional ones, demonstrating good convergence in all parameters. Energy conservation in magnetic fields is not enforced other than by the structure of the Lorentz force law, so small errors in particle energy occur with Runge-Kutta tracking but decrease in the same manner as the velocity errors.

2.2.3 Multi-Particle, Full Beamline Error Study

The previous section studied the error behaviour of a single particle trajectory, whereas there must be thousands of particles in a simulation if yield figures are to be statistically meaningful. The problem with running a numerical algorithm on a large number of inputs (initial particle positions here) is that it drastically increases the probability that one of those inputs encounters a pathological case, perhaps near a boundary, discontinuity or singularity in some function. Therefore we cannot believe the convergence behaviour is 'good' without also checking it for an entire particle beam.

A particular decay channel setup (roughly 30 m of beamline) was run for a pion input file consisting of 20936 particles; on one simulation 8963 particles got to the end of the channel. This number may include multiple decay products from a single particle, as discussed in section 2.4.1 and will vary slightly due to the random nature of the decays unless a fixed seed is used. Simulations were run for 120 ns, which is enough time for the majority of particles of interest to reach the end of the channel and then the distributions compared for timesteps of 5, 10, 15 and 20 ps (recalling that 10 ps was recommended from

†Muon1 tracking of this ring showed the optics to be unstable, which is why a full turn is not used and no cooling ring appears in this thesis.

the single particle work), giving the results shown in table 2.1.

Table 2.1: Distance errors between particles in pairs of beam distributions produced with different choices of Runge–Kutta timestep in the decay channel simulation.

Pair of timesteps compared	Maximum error (m)	RMS error (m)	Mean error (m)
10 ps, 5 ps	0.002339	0.000452	0.000266
15 ps, 5 ps	0.017983	0.003466	0.002050
15 ps, 10 ps	0.015655	0.003015	0.001784
20 ps, 5 ps	0.075524	0.014397	0.008522
20 ps, 10 ps	0.073302	0.013971	0.008268
20 ps, 15 ps	0.058233	0.011084	0.006550

The error measure between two entire distributions can be computed in a number of ways. Here, the positions of the particles in one distribution were compared with their counterparts in another, yielding a list of distances. The maximum, root mean square (RMS) and mean of these distances were calculated for lists coming from each possible pair of distributions. As the method being used is fourth order, in an ideal case each measure of error would be expected to follow the formula

$$Error = K|\delta t_1^4 - \delta t_2^4| = K\delta t_{\text{eff}}^4.$$

Here the 'effective timestep' has been defined as

$$\delta t_{\text{eff}} = |\delta t_1^4 - \delta t_2^4|^{\frac{1}{4}},$$

which is the timestep expected to have the same error from the true solution ($\delta t \to 0$) as the two compared finite timesteps have from each other, assuming ideal, linearly additive fourth order error scaling. δt_{eff} may be plotted against the three error measures and power law curves fitted as shown in figure 2.5.

These errors would be large for some accelerator applications but the comparison point for the muon front end is that the solenoids have an inner radius typically of the order of 30 cm. Thus the sub-millimetre RMS error at δt=10 ps and maximum error of 2.3 mm are good enough to evaluate the transmission. Note that the maximum error is always much larger than the averaged errors because it picks the 'worst' particle, with significant chances of initial trajectory errors being amplified in (for example) the rapidly-changing fringe fields of the solenoids.

The fact that the exponents in the power laws fitted to the graph are much closer to 5 than to 4 is a puzzle. It is possible that errors are not accumulating as $\delta t^5(T/\delta t) \propto \delta t^4$ but instead cancelling for most of the channel as the particle oscillates from one side to the other, leaving only a fixed number of significant erroneous timesteps ($\delta t^5 N$).

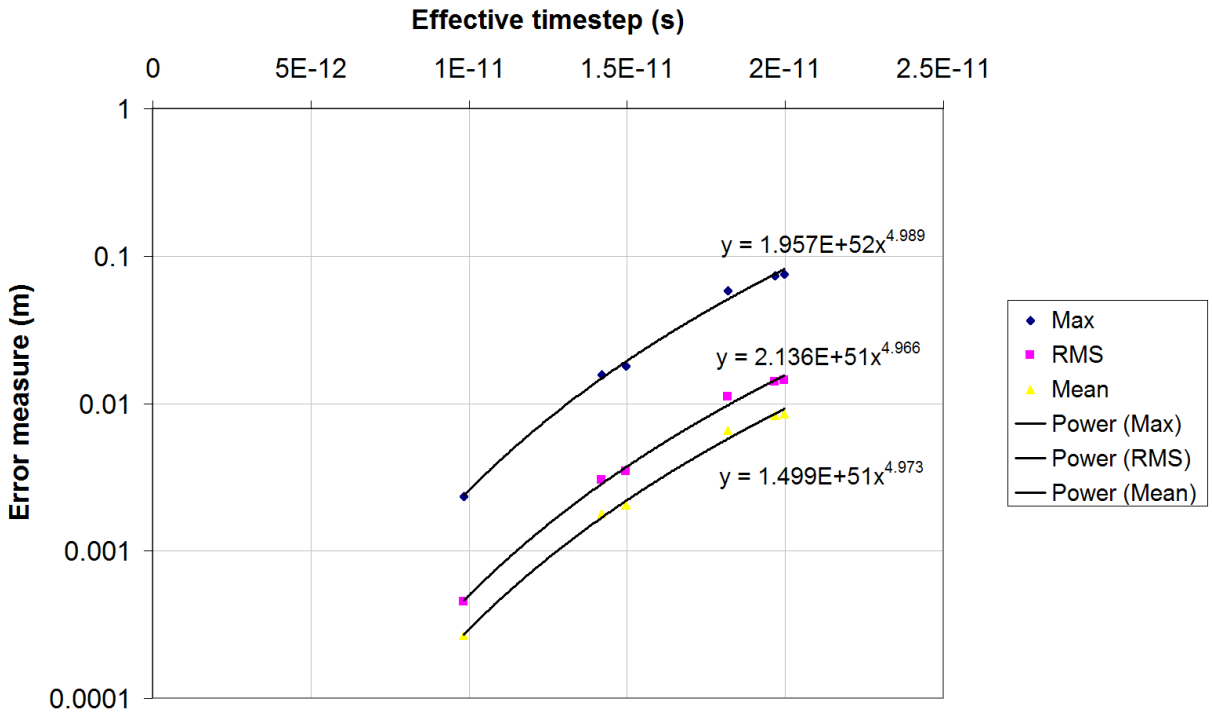


Figure 2.5: Power laws fitted to errors between tracking results for an entire beam under different timesteps.

2.3 Magnetic Fields

The tracking algorithm described in sections 2.1 and 2.2 requires the magnetic field $\mathbf{B}(\mathbf{x})$ to be evaluated at essentially arbitrary points: the Runge–Kutta evaluation nodes near each particle on any given timestep. There are several ways of providing this information on a computer, the following sections describe an algorithm for computing the field of solenoids and a field map representation for general magnets, requiring external data from a magnet design program.

2.3.1 Off-Axis Series Solution

A useful technique for calculating cylindrically-symmetric magnetic fields (expressed as functions $B^z(z,r)$ and $B^r(z,r)$) is that if the field on the axis $B^z(z,0)$ is known, Maxwell's equations in free space can be used to solve for a power expansion of the field in r up to the first conductor or object:

$$B^z(z,r) = \sum_{n=0}^{\infty} B_n^z(z) r^n \qquad B^r(z,r) = \sum_{n=0}^{\infty} B_n^r(z) r^n.$$

The functions $B_n^z(z)$ and $B_n^r(z)$ are to be determined for $n \geq 0$. Putting $r = 0$ in these equations makes $r^n = 0$ for $n \geq 1$, leaving only the r^0 $(n = 0)$ terms:

$$B^z(z,0) = B_0^z(z) \qquad B^r(z,0) = B_0^r(z).$$

The B^r component is the magnitude of the field pointing radially outwards but no such direction exists at $r = 0$, forcing $B^r(z,0) = 0$ to ensure continuity. Both $B_n^z(z)$ and $B_n^r(z)$ are now determined for $n = 0$.

For a static magnetic field in free space, Maxwell's equations simplify to $\boldsymbol{\nabla} \cdot \mathbf{B} = 0$, $\boldsymbol{\nabla} \times \mathbf{B} = \mathbf{0}$. Let $\mathbf{e}_z$ and $\mathbf{e}_r$ be the unit vectors in the z and r directions at any point in space. Taking first the scalar equation,

$$\begin{aligned} 0 &= \boldsymbol{\nabla} \cdot \mathbf{B} \\ &= \boldsymbol{\nabla} \cdot (B^z \mathbf{e}_z + B^r \mathbf{e}_r) \\ &= \boldsymbol{\nabla} B^z \cdot \mathbf{e}_z + B^z \boldsymbol{\nabla} \cdot \mathbf{e}_z + \boldsymbol{\nabla} B^r \cdot \mathbf{e}_r + B^r \boldsymbol{\nabla} \cdot \mathbf{e}_r \\ &= \partial_z B^z + B^z \boldsymbol{\nabla} \cdot \mathbf{e}_z + \partial_r B^r + B^r \boldsymbol{\nabla} \cdot \mathbf{e}_r. \end{aligned}$$

Note that $\boldsymbol{\nabla} \cdot \mathbf{e}_z = 0$ and $\boldsymbol{\nabla} \cdot \mathbf{e}_r = 1/r$. The series expansion may be substituted in to get

$$\begin{aligned} 0 &= \partial_z B^z + \partial_r B^r + \tfrac{1}{r} B^r \\ &= \sum_{n=0}^{\infty} B_n^{z\prime}(z) r^n + \sum_{n=0}^{\infty} B_n^r(z) n r^{n-1} + \sum_{n=0}^{\infty} B_n^r(z) \tfrac{1}{r} r^n \\ &= \sum_{n=0}^{\infty} B_n^{z\prime}(z) r^n + \sum_{n=0}^{\infty} B_n^r(z)(n+1) r^{n-1}, \end{aligned}$$

where $'$ is used to represent the derivative for a function of a single variable.

As the previous equation is true for all r, coefficients in r^n may be equated:

$$\begin{aligned} 0 &= B_n^{z\prime}(z) + B_{n+1}^r(z)(n+2) \\ \Rightarrow\ B_{n+1}^r(z) &= \frac{-1}{n+2}\, B_n^{z\prime}(z) \qquad \text{for } n \geq 0. \end{aligned}$$

This is one half of a recurrence relation to obtain the $B_n^{z,r}$ functions for $n > 0$. The other half comes from the magnetic vector equation:

$$\begin{aligned} \mathbf{0} &= \boldsymbol{\nabla} \times \mathbf{B} \\ &= \boldsymbol{\nabla} \times (B^z \mathbf{e}_z + B^r \mathbf{e}_r) \\ &= \boldsymbol{\nabla} B^z \times \mathbf{e}_z + B^z \boldsymbol{\nabla} \times \mathbf{e}_z + \boldsymbol{\nabla} B^r \times \mathbf{e}_r + B^r \boldsymbol{\nabla} \times \mathbf{e}_r. \end{aligned}$$

Here, $\boldsymbol{\nabla} \times \mathbf{e}_z = \boldsymbol{\nabla} \times \mathbf{e}_r = \mathbf{0}$. The cross product will select only the perpendicular parts of the gradient, so $\boldsymbol{\nabla} f(z,r) \times \mathbf{e}_r = \partial_z f(z,r) \mathbf{e}_\theta$ and $\boldsymbol{\nabla} f(z,r) \times \mathbf{e}_z = -\partial_r f(z,r) \mathbf{e}_\theta$. Thus

$$\begin{aligned} \mathbf{0} &= \boldsymbol{\nabla} B^z \times \mathbf{e}_z + \boldsymbol{\nabla} B^r \times \mathbf{e}_r \\ &= -\partial_r B^z \mathbf{e}_\theta + \partial_z B^r \mathbf{e}_\theta \\ &= \left(-\sum_{n=0}^{\infty} B_n^z(z) n r^{n-1} + \sum_{n=0}^{\infty} B_n^{r\prime}(z) r^n \right) \mathbf{e}_\theta. \end{aligned}$$

This is now in a form to equate coefficients in $r^n \mathbf{e}_\theta$, giving

$$\begin{aligned} 0 &= -B_{n+1}^z(z)(n+1) + B_n^{r\prime}(z) \\ \Rightarrow\ B_{n+1}^z(z) &= \frac{1}{n+1}\, B_n^{r\prime}(z) \qquad \text{for } n \geq 0, \end{aligned}$$

which is the other half of the recurrence relation. The relations zig-zag between the r and z functions as n increases.

Using these rules repeatedly,

$$\begin{aligned} 0 &= B_0^r(z) = B_1^z(z) = B_2^r(z) = B_3^z(z) = \cdots \\ B_1^r(z) &= -\tfrac{1}{2} B_0^{z\prime}(z) \\ B_2^z(z) &= -\tfrac{1}{2}\tfrac{1}{2} B_0^{z\prime\prime}(z) \\ B_3^r(z) &= \tfrac{1}{4}\tfrac{1}{2}\tfrac{1}{2} B_0^{z\prime\prime\prime}(z) \\ &\cdots \\ B_{2n}^z(z) &= \frac{(-1)^n}{2^{2n} n!^2} B_0^{z(2n)}(z) \\ B_{2n+1}^r(z) &= \frac{(-1)^{n+1}}{2^{2n+1} n!(n+1)!} B_0^{z(2n+1)}(z). \end{aligned}$$

Thus the full field can be calculated up to the radius of convergence of the series, provided that the n^{th} derivatives of the on-axis field function $B_0^z(z)$ are also computable.

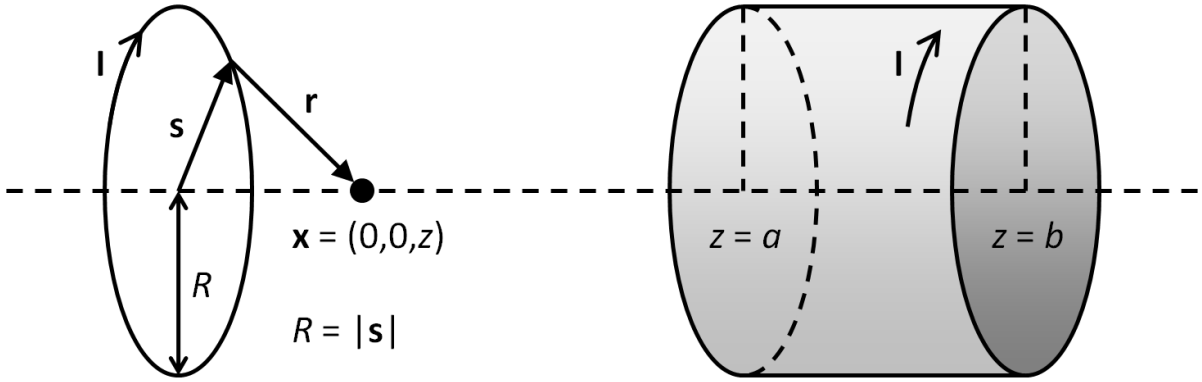


Figure 2.6: Geometry of the current loop and sheet solenoid field evalulation.

2.3.2 Axial Field of a Sheet Solenoid

The Biot-Savart Law gives the magnetic field generated by a loop carrying current $I = |\mathbf{I}|$ as

$$\mathbf{B}(\mathbf{x}) = \frac{\mu_0}{4\pi} \oint \frac{\mathbf{I} \times \mathbf{r}}{|\mathbf{r}|^3} \, \mathrm{d}s,$$

where $\mathbf{r}$ is the vector from the point on the loop to $\mathbf{x}$, $\mathbf{I}$ is the vector current (i.e. with direction parallel to the curve) and $\mathrm{d}s$ is the element of arc length.

Suppose there is a circular loop of radius R in the x–y plane, centred on the origin. The field value at the location $(0, 0, z)$ is of interest, because the on-axis field of the 'sheet solenoid' (of zero thickness but finite length) can be constructed by integrating the fields of many such loops. Parameterising $\mathbf{s}(\theta) = -R(\cos\theta, \sin\theta, 0) = -R\,\mathbf{e}_r$ for $\theta \in [0, 2\pi]$ gives $\mathbf{r}(\theta) = R\,\mathbf{e}_r + z\,\mathbf{e}_z$ and $\mathrm{d}s = R\,\mathrm{d}\theta$. The vector current is $\mathbf{I} = -I\,\mathbf{e}_\theta$. Therefore

$$\begin{aligned}
\mathbf{B}(0,0,z) &= \frac{\mu_0}{4\pi} \int_0^{2\pi} \frac{-I\,\mathbf{e}_\theta \times (R\,\mathbf{e}_r + z\,\mathbf{e}_z)}{(R^2+z^2)^{3/2}} \, R\,\mathrm{d}\theta \\
&= \frac{\mu_0 I R}{4\pi(R^2+z^2)^{3/2}} \int_0^{2\pi} -\mathbf{e}_\theta \times (R\,\mathbf{e}_r + z\,\mathbf{e}_z) \, \mathrm{d}\theta \\
&= \frac{\mu_0 I R}{4\pi(R^2+z^2)^{3/2}} \int_0^{2\pi} R\,\mathbf{e}_z - z\,\mathbf{e}_r \, \mathrm{d}\theta \\
&= \frac{\mu_0 I R}{4\pi(R^2+z^2)^{3/2}} \, 2\pi R\,\mathbf{e}_z \\
&= \frac{\mu_0 I R^2}{2(R^2+z^2)^{3/2}} \, \mathbf{e}_z.
\end{aligned}$$

For a sheet solenoid of the same radius occupying the interval $[a, b]$ in the z-axis, the on-axis field is the integral of the above expression:

$$\begin{aligned}
B_0^z(z) &= \int_{z-b}^{z-a} \frac{\mu_0 I_\ell R^2}{2(R^2+x^2)^{3/2}} \, \mathrm{d}x \\
&= \frac{\mu_0 I_\ell R^2}{2} \int_{z-b}^{z-a} \frac{1}{(R^2+x^2)^{3/2}} \, \mathrm{d}x,
\end{aligned}$$

where x is the offset of the field evaluation point from each infinitesimal current loop and I_ℓ is the current per unit length in the sheet. Integrating gives the final formula

$$\begin{aligned} B_0^z(z) &= \frac{\mu_0 I_\ell R^2}{2}\left[\frac{x}{R^2\sqrt{R^2+x^2}}\right]_{x=z-b}^{z-a} \\ &= \frac{\mu_0 I_\ell}{2}\left(\frac{z-a}{\sqrt{R^2+(z-a)^2}} - \frac{z-b}{\sqrt{R^2+(z-b)^2}}\right). \end{aligned}$$

Expression for the n^{th} Derivative

The axial field and its derivatives may be expressed more succinctly as follows

$$\begin{aligned} B_0^z(z) &= \frac{\mu_0 I_\ell}{2}\left(f(z-a) - f(z-b)\right) \\ B_0^{z(n)}(z) &= \frac{\mu_0 I_\ell}{2}\left(f^{(n)}(z-a) - f^{(n)}(z-b)\right), \end{aligned}$$

using the function $f(x) = x/\sqrt{R^2+x^2}$. To find $f^{(n)}(x)$ it simplifies matters to realise that

$$f^{(n)}(x) = \frac{p_n(x)}{(R^2+x^2)^n\sqrt{R^2+x^2}}$$

with $p_n(x)$ a polynomial and $p_0(x) = x$ for consistency in the $n = 0$ case. The recurrence relation for these polynomials is extracted in the following way.

$$\begin{aligned} f^{(n+1)}(x) &= \frac{\mathrm{d}}{\mathrm{d}x} f^{(n)}(x) \\ \frac{p_{n+1}(x)}{(R^2+x^2)^{n+1}\sqrt{R^2+x^2}} &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{p_n(x)}{(R^2+x^2)^n\sqrt{R^2+x^2}}\right) \\ &= \frac{{p_n}'(x)}{(R^2+x^2)^n\sqrt{R^2+x^2}} + p_n(x)\frac{-(n+\frac{1}{2})}{(R^2+x^2)^{n+1}\sqrt{R^2+x^2}}2x \\ &= \frac{{p_n}'(x)}{(R^2+x^2)^n\sqrt{R^2+x^2}} - \frac{(2n+1)xp_n(x)}{(R^2+x^2)^{n+1}\sqrt{R^2+x^2}}. \end{aligned}$$

Multiplying both sides by $(R^2+x^2)^{n+1}\sqrt{R^2+x^2}$ gives

$$p_{n+1}(x) = (R^2+x^2){p_n}'(x) - (2n+1)xp_n(x).$$

A computer would work out the coefficients of these polynomials once per solenoid by filling in a two-dimensional array, then calculate $f^{(n)}$ and hence $B_0^{z(n)}(z)$ to be used in the off-axis series found in section 2.3.1.

For additional speed, Muon1 has a version of this technique up to and including $n = 3$ (cubic terms) hard-coded into its solenoid field evaluation function. The off-axis series will not converge beyond the radius of the solenoid and high order terms beyond cubic extend the accurate field region by successively smaller amounts.

2.3.3 Field Maps

Sometimes a magnetic (or electric) field cannot be represented analytically, or perhaps the analytic series defining it is prohibitively time-consuming to evaluate. In these cases it is effective to precalculate a table of values of the field at certain points. More advanced techniques such as spline interpolation are also possible, improving accuracy at the expense of computation time.

Currently Muon1 includes the method of 'trilinear interpolation' on a regular cuboidal grid. Let the field map have grid spacings of δx, δy and δz in the three axes and contain values of $\mathbf{B} = (B^x, B^y, B^z)$ for locations from $(0, 0, 0)$ to $(P\delta x, Q\delta y, R\delta z)$ inclusive. This requires $3(P+1)(Q+1)(R+1)$ numeric values in total, which shall be labelled B^x_{000} to B^x_{PQR} for the x components and similarly for y and z.

Let the field value be required at (x, y, z) in coordinates relative to the field map, which may be a rotated or displaced component in the accelerator frame. This point is in the cuboid between $\mathbf{B}_{X_0Y_0Z_0}$ and $\mathbf{B}_{X_1Y_1Z_1}$ where $X_0 = \lfloor x/\delta x \rfloor$, $X_1 = \lceil x/\delta x \rceil$ and similarly for Y and Z. Trilinear interpolation derives the field at (x, y, z) from the field values at the eight corners of this cuboid, but is best explained by starting with a simpler case.

Simple (uni-)linear interpolation approximates the field value along one of the edges of our cuboid in the following way:

$$\mathbf{B}(x, Y\delta y, Z\delta z) \simeq (1-p)\mathbf{B}_{X_0YZ} + p\mathbf{B}_{X_1YZ}.$$

The fractional value $p \in [0, 1]$ is defined as $(x - X_0\delta x)/\delta x$ so that it equals zero at $\mathbf{B}_{X_0YZ}$ and tends to one when approaching $\mathbf{B}_{X_1YZ}$ (we only allow it to equal one when at the top edge of the last cell in the field map). So the field is approximated as a linear function in x that varies smoothly between the nodes. If we consider this equation for $Y = Y_0$ and Y_1 in turn, then linearly interpolate on the y-axis between those points, we get the *bilinear* approximation

$$\mathbf{B}(x, y, Z\delta z) \simeq (1-q)((1-p)\mathbf{B}_{X_0Y_0Z} + p\mathbf{B}_{X_1Y_0Z}) + q((1-p)\mathbf{B}_{X_0Y_1Z} + p\mathbf{B}_{X_1Y_1Z})$$

on the x–y planes of the grid, where $q = (y - Y_0\delta y)/\delta y$. The final step should now be obvious, we interpolate in z between any point on two adjacent x–y planes, setting Z to Z_0 and Z_1:

$$\begin{aligned}\mathbf{B}(x, y, z) \simeq (1-r)((1-q)((1-p)\mathbf{B}_{X_0Y_0Z_0} + p\mathbf{B}_{X_1Y_0Z_0}) + q((1-p)\mathbf{B}_{X_0Y_1Z_0} + p\mathbf{B}_{X_1Y_1Z_0}))\\ + r((1-q)((1-p)\mathbf{B}_{X_0Y_0Z_1} + p\mathbf{B}_{X_1Y_0Z_1}) + q((1-p)\mathbf{B}_{X_0Y_1Z_1} + p\mathbf{B}_{X_1Y_1Z_1})),\end{aligned}$$

with $r = (z - Z_0\delta z)/\delta z$. This is the formula for trilinear interpolation as used in Muon1. It may not be obvious from this expression that the order in which the interpolations were taken (here x then y, then z) does not affect the final result. The symmetry is revealed by multiplying out into individual terms

$$\begin{aligned}\mathbf{B}(x, y, z) \simeq p_0q_0r_0\mathbf{B}_{X_0Y_0Z_0} + p_0q_0r_1\mathbf{B}_{X_0Y_0Z_1} + p_0q_1r_0\mathbf{B}_{X_0Y_1Z_0} + p_0q_1r_1\mathbf{B}_{X_0Y_1Z_1}\\ + p_1q_0r_0\mathbf{B}_{X_1Y_0Z_0} + p_1q_0r_1\mathbf{B}_{X_1Y_0Z_1} + p_1q_1r_0\mathbf{B}_{X_1Y_1Z_0} + p_1q_1r_1\mathbf{B}_{X_1Y_1Z_1},\end{aligned}$$

where $p_0 = 1 - p$ and $p_1 = p$, with analogous definitions for q and r. All eight possible binary combinations of subscripts (corresponding to the cuboid vertices) are used to generate all the terms, without preference for a particular axis.

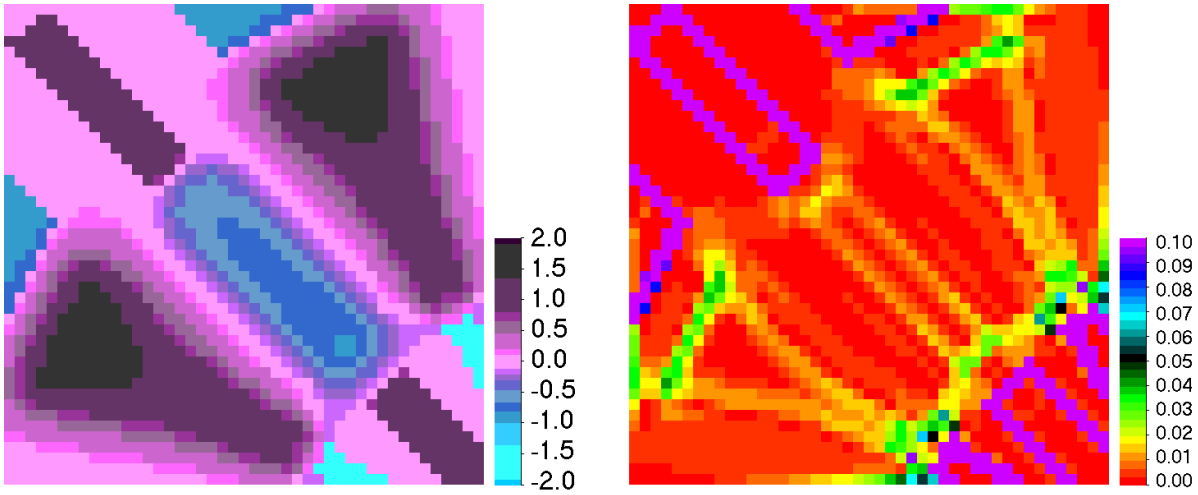


Figure 2.7: (Left) Plot of the on-plane magnetic fields in a coarse (10 cm grid spacing) field map, where the field vectors are perpendicular to the plane of the diagram and the signed magnitude is shown (scales are in Tesla). (Right) Plot of the error bound derived from second derivatives for each square in this plane. The desired path of the particles is from the bottom-left of the view to the top-right, through the regions of low error.

If $x < 0$, $x > P\delta x$, or a similar situation on the y- or z-axes, then the point is outside the range of the field map and $\mathbf{B} = \mathbf{0}$ is returned. This means one must be careful not to introduce discontinuities by having significant field magnitudes at the edges where particles may enter or leave the field map, because these can degrade the convergence of the tracking method (as discussed in section 2.2.2). However, Muon1 does permit multiple field maps (or other magnetic elements) to overlap with each other, in which case the field values from the different magnets are added together to give the total field. This field will still be physically valid because Maxwell's equations in free space are linear in $\mathbf{B}$ (this is known as the 'superposition principle' for magnetic fields).

The accuracy of linear interpolation depends on the second and higher derivatives of the $\mathbf{B}$ field and the size of the mesh cells. It is always true that if there are small features in the field that fall in between grid sampling points, they will not be approximated well, so we assume that the grid is fine enough for the error to be dominated by the leading term in the Taylor series that is different between the approximation and the actual function. Note that linear interpolation is exact for linear functions, so the terms of interest will be quadratic. Furthermore, for trilinear interpolation, even the quadratic mixed terms xy, xz and yz are approximated exactly because they are linear in x, y and z individually at each stage of the multidimensional interpolation procedure given above. So the quadratic terms of interest are in general:

$$\mathbf{Q}(x, y, z) = \mathbf{q}_x x^2 + \mathbf{q}_y y^2 + \mathbf{q}_z z^2,$$

where the $\mathbf{q}$ are arbitrary vector coefficients. The three terms above will be dealt with separately by the algorithm as it interpolates along each axis in turn. The error for x^2 will be a quadratic with the same leading coefficient but the linear interpolation subtracted so that it equals zero at the neighbouring points $X_0\delta x$ and $X_1\delta x$. There is only one such

function, $(x - X_0\delta x)(x - X_1\delta x)$, so the leading term error, $\mathbf{Error}(x, y, z)$, is

$$\mathbf{q}_x(x - X_0\delta x)(x - X_1\delta x) + \mathbf{q}_y(y - Y_0\delta y)(y - Y_1\delta y) + \mathbf{q}_z(z - Z_0\delta z)(z - Z_1\delta z).$$

This becomes more informative when we note that, considering the local Taylor series of $\mathbf{B}$ truncated to quadratic order, $\mathbf{q}_x \simeq \frac{1}{2}\frac{\partial^2\mathbf{B}}{\partial x^2}(x, y, z)$. Also we are only interested in values of $x \in [X_0\delta x, X_1\delta x]$, an interval of length δx with a root of $(x - X_0\delta x)(x - X_1\delta x)$ at either end. The extremum of this function is at the centre of this interval, $\frac{1}{2}\delta x$ distant from either root and the function has value $-\frac{1}{4}\delta x^2$ there. Thus if cubic and higher terms can be neglected, we may bound the error anywhere in the cuboid:

$$|\mathbf{Error}(x, y, z)| \leq \frac{1}{8}\left(\left|\frac{\partial^2\mathbf{B}}{\partial x^2}\right|\delta x^2 + \left|\frac{\partial^2\mathbf{B}}{\partial y^2}\right|\delta y^2 + \left|\frac{\partial^2\mathbf{B}}{\partial z^2}\right|\delta z^2\right).$$

The results of applying this formula to a field map as used in the Rees bending chicane [37] simulation is shown in figure 2.7. Large errors exist around where the field was sampled inside iron pole pieces of the magnet, but the errors are smaller (1% level) along the path that the particles are intended to traverse. The manifestation of field errors as particle position errors is dependent on the optics of the beamline in question, there is no good general-purpose error bound. More favourably, total bending is proportional to the integrated field $\int \mathbf{B}\ \mathrm{d}\ell$ along the particle trajectory, so there is some scope for error cancellation because negative and positive convexity regions, leading to under- and over-estimation in the field components, tend to alternate along the path.

2.4 Particle Decays

The muon front end is unusual in that the particles being transmitted (initially $\pi^{\pm}$) will decay (mainly into $\mu^{\pm}$) on a timescale comparable to how long it takes them to traverse parts of the accelerator, specifically the decay channel section. The mean lifetime of a stationary charged pion is $\tau_{\pi^{\pm}} = 26.033\,\mathrm{ns}$ [27], giving a typical 250 MeV pion a mean path length of $v\gamma\tau = 20.338$ metres.

In the rest frame of the particle, decay is quantum mechanically a perfect Poisson process. Therefore instances of actual decay times, measured as proper times since the particle's creation, can be generated using a probability distribution such that $P(t) \propto e^{-t/\tau}$. The cumulative distribution function for this is

$$F(t) = 1 - e^{-t/\tau}, \qquad (t \geq 0)$$

which can be inverted to give

$$F^{-1}(x) = -\tau\ln(1 - x). \qquad (0 \leq x < 1)$$

The reason F^{-1} is useful is that taking uniformly distributed numbers $x \in [0, 1)$ from the computer's (pseudo-)random number generator will ensure that $t = -\tau\ln(1 - x)$ has the desired distribution for the decay times [38]. These times are generated when the particle starts and compared to the particle's internal proper time clock, which is incremented by $\delta t/\gamma$ each timestep of the simulation, to determine when a decay is triggered.

2.4.1 Statistical Weighting and Multiple Decays

Chapter 4 describes how Muon1 takes output from a hadronic particle showering code as initial data for its tracking. The (macro-)particles in this sort of output are often 'weighted' according to the relative probability of each being produced, so these weights must be retained during tracking and used correctly in evaluating the muon transmission. For example, if each particle i has weight w_i,

$$\%\text{ Muon transmission } = 100\ \frac{\sum_{\substack{\text{Transmitted}\\ \text{muons } i}} w_i}{\sum_{\text{Muons } i} w_i} = 100\ \frac{T_\mu}{W_\mu}.$$

If the indices i are permuted in a random order, then the transmission of muon i is a random event with probability $p \simeq T_\mu/W_\mu$ (strictly this ratio is only an *estimate* of the true value of p). Thus the weight transmitted by this muon is w_i with probability p and 0 with probability $1-p$, giving a mean of pw_i and variance $pw_i^2 - (pw_i)^2 = p(1-p)w_i^2$. Since variances add linearly, the variance of the transmitted weight is

$$\text{Var}(T_\mu) = \sum_{\text{Muons } i} p(1-p)w_i^2 \simeq \frac{T_\mu(W_\mu - T_\mu)}{W_\mu^2} \sum_{\text{Muons } i} w_i^2,$$

which gives the standard deviation of the muon yield from tracking (ignoring pion production uncertainties on W_μ) as

$$\sigma_{\%\text{muons}} \simeq 100\ \frac{\sqrt{T_\mu(W_\mu - T_\mu)}}{W_\mu^2} \sqrt{\sum_{\text{Muons } i} w_i^2}.$$

For fixed transmission and total muon weight, the final term is made smaller by having more macro-particles with smaller weights. Many pions (roughly three quarters) are lost in the very early stages of the simulation before they have time to decay, leaving relatively few pions later on. If each pion with weight w is instead treated as N pions with weight w/N, this will produce a greater number of decay products (muons) with which to evaluate the transmission. However, as the original pion will remain on the same trajectory unless it decays, there is no reason to actually track N copies of it, so Muon1 treats it as a pion with weight w decreasing to $\frac{N-1}{N}w$, $\frac{N-2}{N}w$, etc. each time it decays. This does not increase the effective number of particles in the simulation by a full factor of N because the number of original pions is still fixed, but it does help, especially when working under an optimisation in which small differences in yield can give clues as to the direction of better designs.

2.4.2 Two-body Algorithm

The original implementation of particle decays in the Muon1 tracking code only admitted two-body processes. These can be dealt with by assuming the rest frame of the parent particle and producing two decay products with equal and opposite isotropically-selected

momenta and magnitude such as to conserve energy. This has now been replaced by the more general many-body decay algorithm described in section 2.4.3.

Even the simpler algorithm, however, is not free of implementation details. To transform the decay products' momenta back to the laboratory frame would be easy in one dimension, by Lorentz boosting the energy-momentum vector by the velocity $v = \beta c$:

$$\Lambda(\beta) \begin{bmatrix} E_t \\ pc \end{bmatrix} = \begin{bmatrix} \gamma & \beta\gamma \\ \beta\gamma & \gamma \end{bmatrix} \begin{bmatrix} E_t \\ pc \end{bmatrix} = \begin{bmatrix} \gamma E_t + \beta\gamma pc \\ \gamma pc + \beta\gamma E_t \end{bmatrix}.$$

Unfortunately, the decaying particle will not be travelling along a single axis and the boost formula for a general vector velocity $\mathbf{v} = \boldsymbol{\beta}c$ is required. To avoid dealing with multiple rotation matrices, the momentum can be decomposed into a part perpendicular to $\boldsymbol{\beta}$, which remains unchanged, and a parallel part that undergoes the transformation:

$$\mathbf{p} = \mathbf{p}_\parallel + \mathbf{p}_\perp \qquad \text{where} \qquad \mathbf{p}_\parallel = \frac{\mathbf{p}\cdot\boldsymbol{\beta}}{\beta^2}\boldsymbol{\beta} \qquad \text{and} \qquad \mathbf{p}_\perp = \mathbf{p} - \frac{\mathbf{p}\cdot\boldsymbol{\beta}}{\beta^2}\boldsymbol{\beta}.$$

Here, $\beta = |\boldsymbol{\beta}|$ and γ is defined from this scalar beta. Letting only $\mathbf{p}_\parallel$ be transformed, the one dimensional equivalent of p is $(\mathbf{p}\cdot\boldsymbol{\beta})/\beta$, which satisfies $\mathbf{p}_\parallel = p(\boldsymbol{\beta}/\beta)$. The result of the general Lorentz transformation is then

$$\Lambda(\boldsymbol{\beta}) \begin{bmatrix} E_t \\ \mathbf{p}c \end{bmatrix} = \begin{bmatrix} \gamma E_t + \gamma(\mathbf{p}\cdot\boldsymbol{\beta})c \\ \gamma\mathbf{p}_\parallel c + \boldsymbol{\beta}\gamma E_t + \mathbf{p}_\perp c \end{bmatrix},$$

where terms like $\beta\gamma E_t$ from the one-dimensional formula have been multiplied by $\boldsymbol{\beta}/\beta$ to give them the correct direction.

All that remains is to calculate the momentum of the decay products in the rest frame of the parent particle. Suppose they have masses m_1 and m_2 and will acquire equal and opposite momenta of magnitude p on decay. As momentum is conserved by construction, the one condition is energy conservation:

$$mc^2 = \sqrt{(m_1c^2)^2 + (pc)^2} + \sqrt{(m_2c^2)^2 + (pc)^2}.$$

This may be solved for p using the "$Q = \sqrt{R} + \sqrt{S}$" algorithm described in section 2.4.4.

A randomised vector direction $\hat{\mathbf{p}}$ is needed to generate the momenta $\mathbf{p}_1 = p\hat{\mathbf{p}}$ and $\mathbf{p}_2 = -p\hat{\mathbf{p}}$ in the parent particle's rest frame. In Muon1, $\hat{\mathbf{p}}$ is generated uniformly on the surface of a unit sphere by taking random points $\mathbf{x}$ within the cube $[-1, 1]^3$ until one falls within the unit sphere (i.e. $|\mathbf{x}| \leq 1$) and then returning $\hat{\mathbf{p}} = \mathbf{x}/|\mathbf{x}|$.

2.4.3 Many-body Decay Algorithm

The problem of an n-body decay imposes four constraints on the $3n$ degrees of freedom found in the momentum of the decay products: three for overall conservation of momentum and one for energy. The final states can therefore be found on a $(3n-4)$-dimensional manifold in the multi-particle phase space. This is two-dimensional for two-body decays (and can be made into a sphere by the method and symmetry argument above), but even

for three-body decays it is already five-dimensional and rotational symmetry can remove at most three of these.

Faced with this problem of a complicated surface to generate the distribution on, a Monte Carlo routine that covers the right surface with points (at some known density) is obtained first and then the density is corrected afterwards by rejecting a calculated proportion of them and re-running the original routine as necessary (this is known as *rejection sampling*). Formally, the first routine generates the primal probability density $F(\mathbf{p}_i)$ on the feasible manifold, then any other desired distribution $f(\mathbf{p}_i)$ can be derived from this by only accepting each primal point with probability $\propto f(\mathbf{p}_i)/F(\mathbf{p}_i)$. Note that this should not be larger than 1, so the ratio must be bounded above and scaled accordingly.

2.4.4 The Probability Density (spin averaged)

From the 2004 Particle Physics Booklet, section 38.4, the probability of a particle with 4-momentum P and randomly-selected spin decaying into n particles of 4-momenta $p_1, p_2 \dots p_n$ is governed by

$$\begin{aligned}\mathcal{P} &\propto \delta^4(P - \sum_{i=1}^{n} p_i) \prod_{i=1}^{n} \frac{\mathrm{d}^3\mathbf{p}_i}{E_i} \\ &= \delta(E - \sum_{i=1}^{n} E_i)\delta^3(\mathbf{P} - \sum_{i=1}^{n} \mathbf{p}_i) \prod_{i=1}^{n} \frac{\mathrm{d}^3\mathbf{p}_i}{E_i}.\end{aligned}$$

To simplify the derivation, in this and the next subsection only, the E and E_i will always be total energies (E_t elsewhere) and units will be used such that $c = 1$.

The delta functions can be multiplied into the main coefficient using the following rule, where f is a function with one root $f(a) = 0$:

$$\delta(f(x))g(x) = \frac{g(a)}{|f'(a)|}\delta(x - a).$$

Applying this three times, considering the probability expression as a function of p_n^x, p_n^y and p_n^z in turn, reduces it to:

$$\frac{\mathrm{d}^{3n}\mathcal{P}}{\mathrm{d}^3\mathbf{p}_1\mathrm{d}^3\mathbf{p}_2 \dots \mathrm{d}^3\mathbf{p}_n} \propto \left[\delta(E - \sum_{i=1}^{n} E_i) \prod_{i=1}^{n} \frac{1}{E_i}\right] \delta^3\left(\mathbf{p}_n - (\mathbf{P} - \sum_{i=1}^{n-1} \mathbf{p}_i)\right),$$

which when integrated over $\mathbf{p}_n$ gives

$$\frac{\mathrm{d}^{3n-3}\mathcal{P}}{\mathrm{d}^3\mathbf{p}_1\mathrm{d}^3\mathbf{p}_2 \dots \mathrm{d}^3\mathbf{p}_{n-1}} \propto \delta(E - \sum_{i=1}^{n} E_i) \prod_{i=1}^{n} \frac{1}{E_i},$$

after which the substitution $\mathbf{p}_n = \mathbf{P} - \sum_{i=1}^{n-1} \mathbf{p}_i$ should be used to determine the last momentum and E_n.

This needs to be integrated one more time to remove the last delta function, so a degree of freedom will be taken from $\mathbf{p}_{n-1}$. Let $\mathbf{p}_* = \mathbf{P} - \sum_{i=1}^{n-2} \mathbf{p}_i$ so that $\mathbf{p}_{n-1} + \mathbf{p}_n = \mathbf{p}_*$ to conserve 3-momentum. Now perform a Lorentz transformation into a primed frame where $\mathbf{p}'_* = \mathbf{0}$, so that $\mathbf{p}'_{n-1} = -\mathbf{p}'_n = p\hat{\mathbf{p}}$. Integrating over $p \geq 0$ will remove the delta function, leaving the two degrees of freedom in $\hat{\mathbf{p}}$ parameterising the remaining freedom of these last two momenta.

Starting with the expression written in the primed frame:

$$\frac{\mathrm{d}^{3n-3}\mathcal{P}}{\mathrm{d}^3\mathbf{p}'_1\mathrm{d}^3\mathbf{p}'_2\ldots\mathrm{d}^3\mathbf{p}'_{n-1}} \propto \delta(E' - \sum_{i=1}^{n} E'_i) \prod_{i=1}^{n} \frac{1}{E'_i}$$

...change $\mathbf{p}'_1$ through $\mathbf{p}'_{n-2}$ back to the original frame using $\mathrm{d}^3\mathbf{p}'_i = (E'_i/E_i)\mathrm{d}^3\mathbf{p}_i$:

$$\frac{\mathrm{d}^{3n-3}\mathcal{P}}{\mathrm{d}^3\mathbf{p}_1\mathrm{d}^3\mathbf{p}_2\ldots\mathrm{d}^3\mathbf{p}_{n-2}\mathrm{d}^3\mathbf{p}'_{n-1}} \propto \delta(E' - \sum_{i=1}^{n} E'_i) \frac{1}{E'_{n-1}E'_n} \prod_{i=1}^{n-2} \frac{1}{E_i}$$

...and then use the substitution, which implies $\mathrm{d}^3\mathbf{p}'_{n-1} = p^2\mathrm{d}p\mathrm{d}^2\hat{\mathbf{p}}$, to get:

$$\frac{\mathrm{d}^{3n-4}\mathcal{P}}{\mathrm{d}^3\mathbf{p}_1\mathrm{d}^3\mathbf{p}_2\ldots\mathrm{d}^3\mathbf{p}_{n-2}\mathrm{d}^2\hat{\mathbf{p}}} \propto \delta(E' - \sum_{i=1}^{n} E'_i) \frac{p^2\mathrm{d}p}{E'_{n-1}E'_n} \prod_{i=1}^{n-2} \frac{1}{E_i}.$$

Now that varying p is the concern, a "constant" part may be removed from the argument of the delta function

$$\delta(E' - \sum_{i=1}^{n} E'_i) = \delta(E'_* - (E'_{n-1} + E'_n)),$$

where $E'_* = E' - \sum_{i=1}^{n-2} E'_i$. The modulus of the derivative of this argument:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}p}(E'_{n-1} + E'_n) &= \frac{\mathrm{d}}{\mathrm{d}p}\left(\sqrt{p^2 + m_{n-1}^2} + \sqrt{p^2 + m_n^2}\right) \\
&= \frac{p}{\sqrt{p^2 + m_{n-1}^2}} + \frac{p}{\sqrt{p^2 + m_n^2}} \\
&= p\left(\frac{1}{E'_{n-1}} + \frac{1}{E'_n}\right) = p\frac{E'_{n-1} + E'_n}{E'_{n-1}E'_n},
\end{aligned}$$

...must be divided out when the argument of the delta function is linearised:

$$\begin{aligned}
\frac{\mathrm{d}^{3n-4}\mathcal{P}}{\mathrm{d}^3\mathbf{p}_1\mathrm{d}^3\mathbf{p}_2\ldots\mathrm{d}^3\mathbf{p}_{n-2}\mathrm{d}^2\hat{\mathbf{p}}} &\propto \delta(p - p_0)\frac{p\mathrm{d}p}{E'_{n-1} + E'_n} \prod_{i=1}^{n-2} \frac{1}{E_i} \\
&= \frac{p}{E'_{n-1} + E'_n} \prod_{i=1}^{n-2} \frac{1}{E_i}\bigg|_{p=p_0},
\end{aligned}$$

with p_0 being the value of $p \geq 0$ for which energy is conserved.

p_0 is the root of the delta function, so

$$E'_* = E'_{n-1} + E'_n = \sqrt{p_0^2 + m_{n-1}^2} + \sqrt{p_0^2 + m_n^2},$$

which is of the form

$$\begin{aligned} & Q = \sqrt{R} + \sqrt{S} \\ \Rightarrow \quad & Q - \sqrt{R} = \sqrt{S} \\ \Rightarrow \quad & Q^2 - 2Q\sqrt{R} + R = S \qquad (*) \\ \Rightarrow \quad & Q^2 + R - S = 2Q\sqrt{R} \\ \Rightarrow \quad & (Q^2 + R - S)^2 = 4Q^2 R. \qquad (*) \end{aligned}$$

This may be used on the expression now, but note that in the squaring operations marked (*) the implication goes only one way, so any value of p_0 derived using the following must be checked back to make sure the squaring operation did not equate things that originally had opposite signs! The squared-up expression is

$$({E'_*}^2 + m_{n-1}^2 - m_n^2)^2 = 4{E'_*}^2(p_0^2 + m_{n-1}^2),$$

giving the root explicitly as

$$p_0 = \sqrt{\frac{({E'_*}^2 + m_{n-1}^2 - m_n^2)^2}{4{E'_*}^2} - m_{n-1}^2}.$$

2.4.5 The Primal Distribution

The previous section gave the probability distribution with respect to an element of the momenta $\mathbf{p}_1 \ldots \mathbf{p}_{n-2}$ and the unit direction $\hat{\mathbf{p}}$ governing the last two particles. Thus if a uniform distribution is used for the $n-2$ momenta and an isotropic one for $\hat{\mathbf{p}}$ in the primal distribution, the $(3n-4)$-dimensional probability density will be exactly proportional to the selection probability.

Generating an isotropic $\hat{\mathbf{p}}$ is easy; what is not so obvious is how to choose a 'domain' in which the $\mathbf{p}_i$ are uniformly spread. As $|\mathbf{p}_i| \leq E_i \leq E = m_0$ (i.e. the mass of the parent particle), there is at least a crude limit on how large the momenta can be. Here note that many of these choices for $\mathbf{p}_1 \ldots \mathbf{p}_{n-2}$ miss the manifold entirely, a fact which becomes apparent when attempting to enforce conservation of energy leads to particles with negative energy, or taking square-roots of negative quantities. In these cases, the sample should be rejected, so in effect there are two levels of rejection sampling, one to get onto the manifold and a second to adjust the density. The first level does not affect the relative proportions between any of the probabilities within the manifold, so has no adverse consequences other than wasted CPU time.

The final ingredient for a workable implementation of rejection sampling is an upper bound on the probability density derived in the previous section, so that it may be divided out to give selection probabilities of no more than 1. By noting that $E_i \geq m_i \Rightarrow 1/E_i \leq 1/m_i$, one obtains

$$\frac{p}{E'_{n-1} + E'_n} \prod_{i=1}^{n-2} \frac{1}{E_i} \leq \frac{p}{E'_{n-1} + E'_n} \prod_{i=1}^{n-2} \frac{1}{m_i},$$

leaving only the front term to deal with. Remembering that $|\mathbf{p}'_{n-1}| = |\mathbf{p}'_n| = p$, it follows that $E'_{n-1} \geq p$ and $E'_n \geq p$ so that the denominator is at least $2p$, resulting in

$$\frac{p}{E'_{n-1} + E'_n} \prod_{i=1}^{n-2} \frac{1}{E_i} \leq \frac{1}{2} \prod_{i=1}^{n-2} \frac{1}{m_i}.$$

The probability to use is therefore

$$\mathcal{P}_1 = \frac{2p}{E'_{n-1} + E'_n} \prod_{i=1}^{n-2} \frac{m_i}{E_i}.$$

This is workable for several cases, but when there are particles of small mass (such as electrons or neutrinos) in the decay products, the ratio $m_i/E_i = 1/\gamma_i$ becomes very small when these acquire a significant fraction of the mass–energy of the heavier parent particle. The formula does not select any samples at all when photons are produced! The problem arises from the choice of a uniform momentum distribution for each of the first $n-2$ particles not being a very good 'fit' to the $1/E_i$ in the probability density, particularly when the energy is dominated by momentum rather than rest mass.

With this in mind, the bound can be modified to use the fact that $E_i = \sqrt{\mathbf{p}_i^2 + m_i^2} \geq \max\{|\mathbf{p}_i|, m_i\}$ so $1/E_i \leq \min\{1/|\mathbf{p}_i|, 1/m_i\}$. However, this is no longer a constant with respect to $\mathbf{p}_i$, so to have it divided out, a primal distribution must be generated with $F(\mathbf{p}_i) \propto \min\{1/|\mathbf{p}_i|, 1/m_i\}$. This is uniform for $|\mathbf{p}_i| \leq m_i$, so retaining the previous cutoff of $|\mathbf{p}_i| \leq m_0$, the ratio of the uniform part to the outer part is

$$\frac{4}{3}\pi m_i^3 \frac{1}{m_i} : \int_{m_i}^{m_0} 4\pi |\mathbf{p}_i|^2 \frac{\mathrm{d}|\mathbf{p}_i|}{|\mathbf{p}_i|} = \frac{m_i^2}{3} : \frac{m_0^2 - m_i^2}{2}.$$

Since $2|\mathbf{p}_i|\mathrm{d}|\mathbf{p}_i| = \mathrm{d}|\mathbf{p}_i|^2$, the outer part corresponds to choosing $|\mathbf{p}_i|^2$ uniformly on $[m_i^2, m_0^2]$. Generating events from these two parts in the ratio above, for each i up to $n-2$, gives the required primal distribution.

Now dividing out $\min\{1/|\mathbf{p}_i|, 1/m_i\}$ instead of just $1/m_i$ gives the new selection probability as

$$\mathcal{P}_2 = \frac{2p}{E'_{n-1} + E'_n} \prod_{i=1}^{n-2} \frac{\max\{|\mathbf{p}_i|, m_i\}}{E_i},$$

which is clearly better conditioned, as terms in the product part never fall below $1/\sqrt{2}$.

2.4.6 Results and Efficiency

The two versions of the algorithm described above were implemented and tested on a variety of decay modes for the kaon: the results are shown in table 2.2. Note that for 3-body and higher decays, there is sometimes a choice of which particles to associate the 'special' momenta $\mathbf{p}_{n-1}$ and $\mathbf{p}_n$ with. This affects the speed of the algorithms, particularly the first version, so these possibilities are broken down individually in the cases where the particles have different mass.

Table 2.2: Mean number of primal samples required to generate one valid decay. Figures are an average over 10000 successful decays except where another number of decays is indicated in parentheses. 'inf' denotes that it was infeasible to run the algorithm, as no valid decays were generated after 10^7 samples.

Decay (special pair)	Using $\mathcal{P}_1$	Using $\mathcal{P}_2$
$K^+ \to e^+\nu_e$	1	1
$K^+ \to \mu^+\nu_\mu$	1.0483	1.0485
$K^+ \to \pi^0 e^+\nu_e$		
(π^0, e^+)	inf	6.7499
(π^0, ν_e)	4123.71 (2425)	6.5441
(e^+, ν_e)	15.7419	6.4251
$K^+ \to \pi^+\pi^0\pi^0$		
(π^+, π^0)	151.285	60.5826
(π^0, π^0)	146.47	61.5364
$K^+ \to \pi^+\pi^+\pi^-$	184.181	75.1907
$K^+ \to \pi^+\pi^- e^+\nu_e$		
(π^+, π^-)	inf	666.969
(π, e^+)	inf	652.936
(π, ν_e)	8×10^5 (25)	654.098
(e^+, ν_e)	3746.72 (2669)	636.432

The second version of the algorithm using $\mathcal{P}_2$ displays a much more consistent speed across decays with the same number of daughter particles, with no decays in the test set it could not solve. Time increases rapidly with number of daughter particles but since no more than four are required for the particles in Muon1 and each calculation is fairly fast, the ∼650 iterations needed for 4-body decays are not a major problem.

2.4.7 Dalitz Plot Test

A standard plot that can be used to test the correctness of the routine is the *Dalitz plot* of m_{12}^2 against m_{23}^2, where $m_{ij}^2 = (m_i + m_j)^2 - |\mathbf{p}_i + \mathbf{p}_j|^2$. The correct phase-space distribution ought to be uniformly spread on the allowed region when projected into this plane. Such a plot is shown in figure 2.8 generated from many runs of the $\mathcal{P}_2$-based version of the algorithm.

The uniformity of the enclosed area is visually apparent, as is the sharpness of the edge in phase space of feasible decays. This provides additional confidence that no errors have been made in the coding of this mathematically complicated routine.

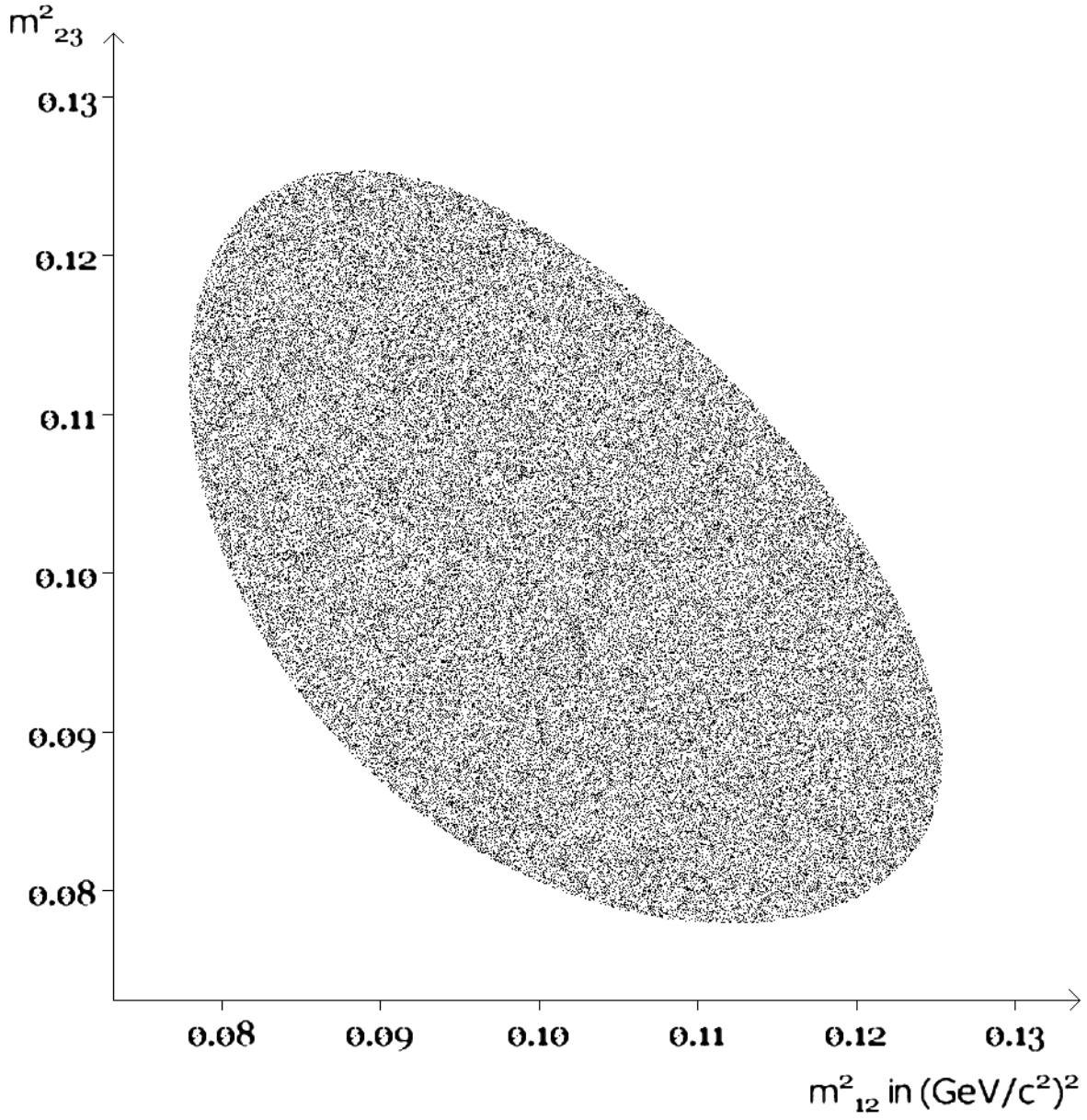


Figure 2.8: Dalitz plot for the $K^+ \to \pi^+\pi^+\pi^-$ decay, showing uniform coverage of the allowed area.

2.5 Collision Detection

When a particle goes outside the accelerator aperture, for instance into the coil of a solenoid, it is removed from the simulation and counted as a loss[†]. The code should be able to tell, similarly, when a particle has entered a region of scattering material or passed through a surface at which the beam is being logged or an aperture restriction imposed. Proximity detection is also needed to determine which components contribute significantly to the magnetic field at a given point in space. More complex collisional algorithms are needed for the scattering calculation itself, but this is presented in sections 6.4 and 6.5.

2.5.1 Proximity and Bounding Boxes

For tracking codes that step in z, it is often unambiguous which component a particle is inside, so the field and aperture restriction can be applied directly. For codes that step in t there will be particles at a range of z that may be in different components. The simplest method would be to check against every component to determine the field and aperture set (the set containing all allowed locations for particles):

$$\mathbf{B}(\mathbf{x}) = \sum_{n=1}^{\#\text{ Components}} \mathbf{B}_{\text{Component } n}(\mathbf{x}) \qquad A = \bigcap_{n=1}^{\#\text{ Components}} A_{\text{Component } n}.$$

A direct implementation of this on a computer unfortunately adds a calculation of length (# Particles)(# Components) to each timestep. This is bad scaling behaviour in the majority of situations where only a few components have any significant effect for each particle. Note that the superposition principle used for the magnetic fields here means that components containing saturating iron or other nonlinear materials must not be placed too far into the fringe fields of others; a solution is to treat 'interacting' objects as a single contiguous component using a field map. Note also that a collision is defined as a particle moving outside the allowed aperture A and hence into its complement $\bar{A} = \mathbb{R}^3 \backslash A$, which means that if the effects of each component are to be localised, the 'solid' collision zone $\bar{A}$ must be bounded, as opposed to some codes (particularly in z) in which the disallowed region may extend to infinity.

The efficiency problem mainly hinges on the detection of proximity to components. Once the set of nearby components has been determined, only the $\mathbf{B}$ fields from these need to be summed and only collisions with these components need to be checked for. If this detection is performed by a scan of all components on every timestep, it also risks bad scaling behaviour. However, the calculation that is being repeated is now just proximity detection rather than the full collision and field calculation; it is therefore desirable to make this as fast to compute as possible. The approach taken is of bounding cuboids C constructed such that

$$\{\mathbf{x} : |\mathbf{B}(\mathbf{x})| \leq \epsilon_B\} \cup \bar{A} \quad \subseteq \quad C = [x_1, x_2] \times [y_1, y_2] \times [z_1, z_2],$$

[†]In reality, some particles (particularly muons) are not absorbed and can scatter back into the aperture but Muon1 does not include this possibility. This makes a slightly pessimistic error in yield, but the probability of a muon that scatters out of the walls rejoining the beam for an extended period is minimal.

so only 6 comparisons (and possibly fewer) are required to determine whether a point (x, y, z) is within C, which contains all of the collision zone and significant field region.

Speed-of-Light Optimisation

An additional optimisation exists so that the scan for nearby components is only necessary on every N^{th} timestep, although there is some tradeoff. If a list of nearby components is to be valid for N timesteps (a time $N\delta t$) from the time it is constructed, it must refer to all components the particle could possibly enter into the bounding cuboid of during that time. The particle's trajectory is not known in advance but we would hope that it will remain within the light cone given by $\Delta x \leq c\Delta t \leq cN\delta t$. The set of all points that can reach C as defined above within time $N\delta t$ is

$$\begin{aligned}&\text{Expand}_{cN\delta t}(C) = \{\mathbf{x} : \exists \mathbf{y} \in C, |\mathbf{y} - \mathbf{x}| \leq cN\delta t\} \\ &\subseteq [x_1 - cN\delta t, x_2 + cN\delta t] \times [y_1 - cN\delta t, y_2 + cN\delta t] \times [z_1 - cN\delta t, z_2 + cN\delta t],\end{aligned}$$

so we may just use a larger cuboid, C extended by $cN\delta t$ outwards in every direction, as a valid bounding box for use on N successive steps. The tradeoff in this optimisation is that if N is made too large, the cuboids will also become large and time will be spent adding the fields of (and checking for collisions with) components that are far away from the particle but could have been reached by it in the time $N\delta t$ since the nearby component list was generated.

Proximity Grid Optimisation

The scan can be made even faster by using the principle of a *hash table*, one of the fastest known ways of speeding up search procedures, at the expense of some extra memory usage. A three dimensional cuboidal grid is constructed, in Muon1's case up to 50 cells on a side, covering the bounding cuboid of the entire accelerator. In each cell is put a list of the components that have bounding boxes somewhere in the volume of that cell. Thus when it is time to recalculate the nearby components for a particle, the cell in the grid containing that particle is looked up and the subsequent search for bounding boxes restricted to just those components listed in that cell.

The upshot of these optimisations, applied in Muon1 with $N = 10$ for the 'speed-of-light' optimisation, is that the time taken by proximity detection is very small compared to the time used for Runge-Kutta stepping.

2.5.2 Collision Detection Regions for Components

Muon1 defines collision regions automatically for certain components. The collision zone used for a sheet solenoid of length L and radius R is by default

$$\bar{A}_{\text{sol}} = \{\mathbf{x} : 0 \leq z \leq L, R \leq r \leq 5R\},$$

where (z, r) are $\mathbf{x}$ expressed in cylindrical coordinates coaxial with the solenoid, with origin at the centre of one end. The factor 5 can be changed by the user by specifying the

`OuterRadius` attribute of the `Solenoid` component type to be some length other than $5R$, see section 7.1 for how components are input to Muon1 using lattice files.

The other object that is specifically solid in the tracking code is the rod-shaped target, modelled as a solid cylinder that absorbs particles reentering it:

$$\bar{A}_{\text{rod}} = \{\mathbf{x} : 0.005L \leq z \leq 0.995L, r \leq 0.99R\}.$$

The coordinates are as before, but the peculiar 1% shrinkage of the zone is to make sure that particles starting exactly on the surface, which have some rounding in their coordinates as output from the particle production code, are never erroneously destroyed at their start location.

In addition to these two rules, Muon1 has a more general system that allows pure collision regions (or regions of material) to be specified separately from particular accelerator components. These are known as `Blockage` components. As the lattice format supports arbitrary displacements and rotations of objects from any point in the beamline and the beamline location may be left constant by not specifying a `Length` attribute for these components, complex shapes may be built up from the primitive blockages. The primitives supported as of Muon1 version 4.44 are cuboids, cylinders, spheroids (including spheres), annuli and toroids. Cones may have some use in the future.

As well as solid objects it is often desirable to impose a restriction on particles going through a plane in the accelerator, usually a plane perpendicular to the accelerator axis. Types of `Aperture` component were introduced for this purpose, which detect when a particle has crossed the plane, calculate the relative coordinates on the plane of the crossing point and check against the aperture restriction. Currently only circular and rectangular apertures are supported, although additionally the `MinEnergy` and `MaxEnergy` attributes may be set to stop particles outside a given energy range. Apertures are also used as the finishing component, which stops all particles on the plane and flags them as 'finished' (unless they were outside the aperture), which is used in figure-of-merit calculation.

2.5.3 High-accuracy Isosurface Method

Detecting exactly where a particle crosses a surface (an aperture plane or the boundary of a collision zone) is a common operation in all of the above mechanisms. Initial implementations used a simple method that checked whether the particle was inside a collision zone of any nearby component on each timestep and stopped it at its current location if it was. Apertures were dealt with by waiting for the current and old (previous timestep) positions of a particle to be on opposite sides of the plane and then taking a straight line between them to find the intersection with the aperture itself.

However these simple algorithms suffer from some problems. The point of using Runge–Kutta integration was to allow accurate tracking with a medium-sized timestep: for the chosen $\delta t = 10^{-11}$s, $c\delta t \simeq 3\,\text{mm}$ is large enough to jump over the corner of, for instance, a solenoid's annular region in a single step (see figure 2.9). This will change the transmission figure, which is constrained ultimately by the bore sizes, because the bore radius (which has been as small as 15 cm in simulations) is now up to 2% larger and 4% bigger in area. This becomes worse (up to 8% effect on transmission) when it

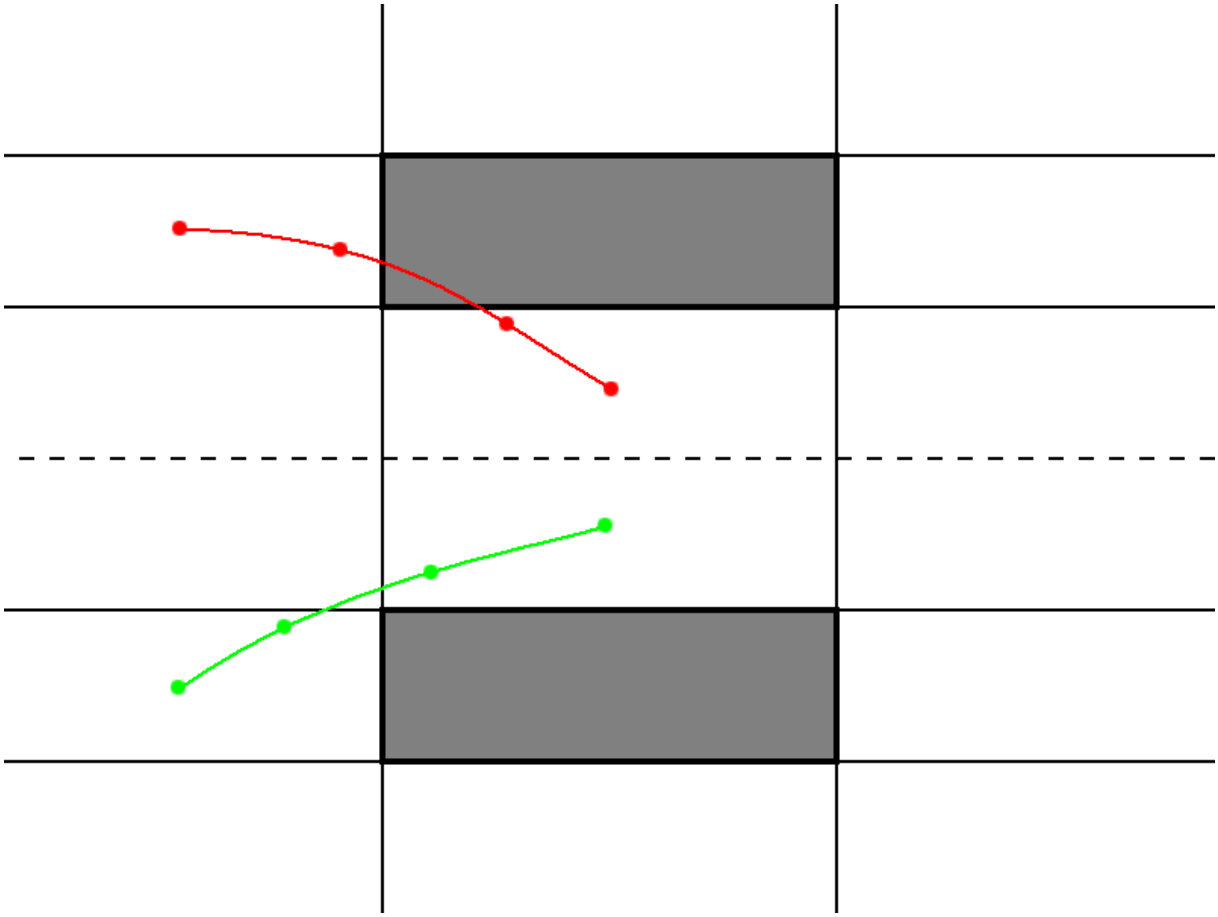


Figure 2.9: Diagram of how a (slightly exaggerated!) timestep may jump a particle over the corner of the collision region (red particle) or miss the collision region (green particle). The dotted line is the axis of symmetry of the solenoid being considered.

is considered that the transverse bores are really clipping the amplitude in 4-dimensional transverse phase space. The procedure of taking a straight line to detect the collision with an aperture plane is also inaccurate, though this is a second order effect in step length so smaller.

The solution is to detect when collisions with surfaces may have occurred and solve iteratively for the collision point. To detect collisions reliably for a solid object requires more than just checking if the particle is in $\bar{A}$ at each timestep. Instead, the solid object's surface is defined by smooth isosurfaces that extend outside the object from the corners if necessary: these are the thin lines shown in figure 2.9. An isosurface is the solution set of an equation $f(\mathbf{x}) = 0$ where f is some smooth function. If the particle has crossed *any* of these surfaces during a timestep, which can be easily seen by one of the fs changing sign, the intersection points are solved for and checked to see if they fall on the surface of the collision region or on some part of the isosurface extending into empty space. This method is still not perfect because it will not detect skimming collisions when a particle curves into and then out of an isosurface during a single timestep, but it improves things greatly, making missed collisions a second order (curvature) effect.

Solving for the collision point can be done by a number of standard methods. Repeated

bisection would work even for non-isosurfaces but as f is provided, a faster root finding technique like the secant method can be applied to $f(\mathbf{x}(t)) = 0$. In fact, Muon1 uses a combination of these two techniques, reverting to bisection if the secant method becomes divergent. Note that calculating $\mathbf{x}(t)$ for a general time in the timestep requires the Runge–Kutta step to be called as a subroutine of the collision point finder. Adding material scattering complicates the process further and the full stack of stepping routines is given in section 6.5.

2.6 RF Components and Phasing

Muon1 uses a simplified model for an RF cavity where the acceleration or deceleration takes place instantaneously as the particle crosses a plane placed where the midplane of the real cavity would be, usually perpendicular to the beamline. Cavities are defined by the peak voltage V_0 a particle can experience, their angular frequency $\omega = 2\pi f$ and an absolute phase ϕ. Note that the V_0 specified to Muon1 should already be reduced to account for the transit time factor of the cavity, since it is the maximum accelerating voltage given to *a particle* rather a definition of the electric field. Muon1 actually takes input in the form of the frequency f and relative phase ϕ_r and derives ω and ϕ (see below).

If the RF cavity has forward-pointing unit vector $\hat{\mathbf{n}}$ and a particle of charge q and velocity $\mathbf{v}$ passes through the cavity plane at simulation time t with offset $\Delta\mathbf{x}$ from the cavity centre, the following change in momentum is applied:

$$\Delta\mathbf{p} = \frac{qV_0}{\mathbf{v}\cdot\hat{\mathbf{n}}}\left(\cos(\omega t+\phi)\ \hat{\mathbf{n}} + \frac{\omega}{2c^2}\sin(\omega t+\phi)\ \mathbf{v}\times(\Delta\mathbf{x}\times\hat{\mathbf{n}})\right).$$

This is derived by assuming that in a cavity of notional length ℓ the particle experiences average electric and magnetic fields of

$$\begin{aligned} & & \langle\mathbf{E}\rangle &= \frac{V_0}{\ell}\cos(\omega t+\phi)\ \hat{\mathbf{n}} \\ \text{and} & & \langle\mathbf{B}\rangle &= \frac{V_0\omega}{2\ell c^2}\sin(\omega t+\phi)\ \Delta\mathbf{x}\times\hat{\mathbf{n}} \end{aligned}$$

during a total transit time of $\ell/\mathbf{v}\cdot\hat{\mathbf{n}}$ and using $\mathrm{d}\mathbf{p}/\mathrm{d}t = q(\mathbf{E}+\mathbf{v}\times\mathbf{B})$. The form of $\langle\mathbf{B}\rangle$ is the solution of Maxwell's wave equations that has $\mathbf{B} = \mathbf{0}$ on the $\hat{\mathbf{n}}$ axis and is consistent with the expression for $\langle\mathbf{E}\rangle$.

The absolute phase ϕ is calculated from the relative phase ϕ_r specified in the Muon1 input files by one of two possible methods. If a `Synchronous-particle` component has been defined and inserted at the beginning of the beamline, it is tracked through the simulation before all other particles, with the time that it crosses each RF plane defining the relative phase of zero. Absolute phases are then calculated from

$$\omega t + \phi = \omega(t - t_{\text{cross}}) + \phi_r \quad \Rightarrow \quad \phi = \phi_r - \omega t_{\text{cross}}.$$

If the synchronous particle is lost in the simulation, Muon1 calculates the crossing times from its initial velocity and the arc lengths of the beamline reference axis. If a synchronous particle is not specified, absolute phases are set equal to the phases in the input file.

Chapter 3

Target Design Considerations

The reference proton beam power used in the ISS [24] and most other neutrino factory studies is 4 MW. This is quoted as the mean beam power during nominal operation of the facility, including all inter-pulse gaps (the pulsed power is at the terawatt level). Although powers greater than 4 MW are possible, they would represent an increasingly large leap in proton accelerator technology over that currently demonstrated. Lower powers would produce so few muons that it would become hard to achieve the 10^{21} decays per 10^7 s of running time required for physics results in a reasonable timescale. The figure has perhaps a factor of two of elasticity in either direction but fixing it is essential to define the time and cost scale of the baseline design.

Several constraints mean that the concentration of power in the target becomes a considerable difficulty at the 4 MW level. Most of the heat deposition in the target will occur in the first two hadronic interaction lengths (around 20 cm for heavy metals); however, the radius should be much smaller than an interaction length (a few centimetres) to prevent pions being reabsorbed as they exit the target. This leaves a very small volume to dissipate the considerable heat load and this cannot be done with a static target as for radiation cooling the powers are a factor of ≥ 12 too high, as shown in section 3.3. Water cooling comes closer to feasibility but falls short due to lack of surface area in the

Table 3.1: Features of a solid pion production target compared with a liquid mercury jet.

Advantages of solid target	Disadvantages of solid target
Does not depend on hydrodynamics No liquid cavitation	Requires mechanical system for changing targets
Mercury vapour damages vacuum, require beam window	Mechanical joints may produce dust
No chance of contamination from spillage	Proton pulse intensity must not shatter target
Targets may be changed sideways at lower velocity than longitudinally	Requires the first solenoid coil to be split

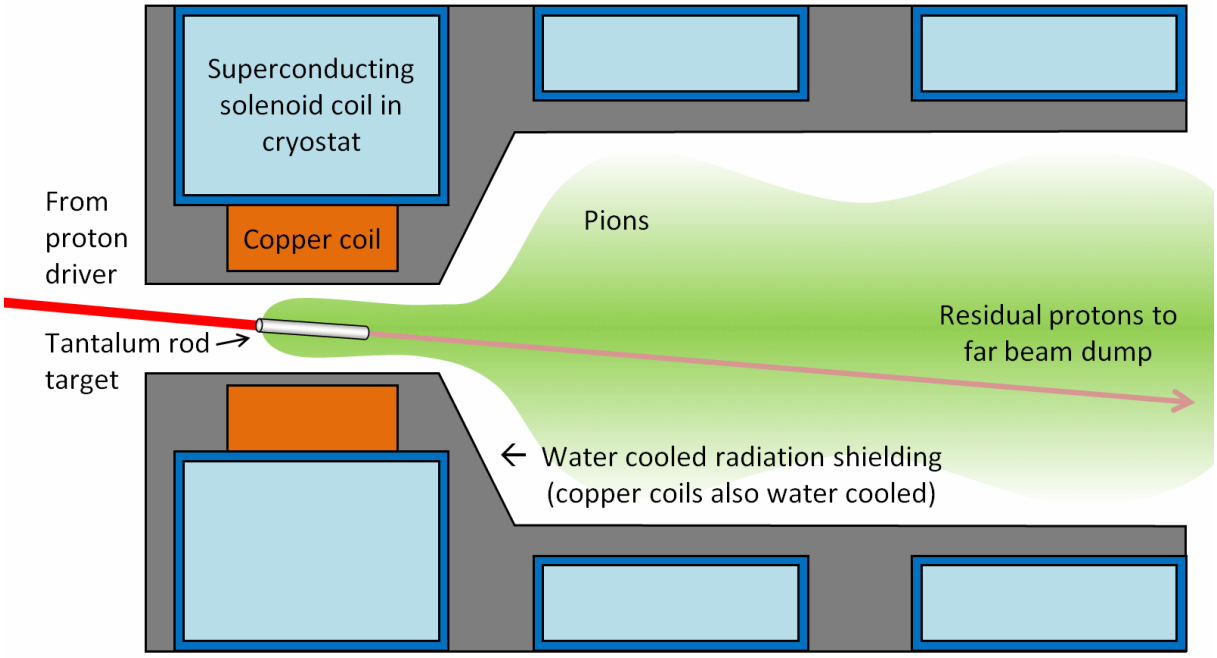


Figure 3.1: Main components of the UK solid target station.

target. Therefore the target material must be replaced for each pulse, either by moving the target material or steering the beam onto different targets, the former currently being considered more practical.

The UK neutrino factory design proposes a solid pion-production target as opposed to the liquid mercury jet suggested in the US study [25]. The solid 'target' would actually be a series of solid targets in a loop that take turns to be hit by the proton beam pulse, whereas the liquid jet vaporises after producing the pions and is flowing at such a rate that new liquid will be in place by the next pulse (which arrives at a 50 Hz rate in the liquid target baseline). This changes the set of limitations and challenges in the target as set out in table 3.1.

Proof-of-principle experiments in high power targets for the neutrino factory (and other high-power facilities) are being conducted at the time of writing, most notably MERIT [39] for the liquid mercury jet and wire shock test studies [29] for the solid target. Interested readers are encouraged to follow the references as the detail of these experiments is outside the scope of this thesis.

3.1 Proton Driver Parameters

With the beam power P fixed, the main remaining parameters are the proton beam kinetic energy (E_k) and the pulse repetition frequency f. Given these, the number of particles per pulse (N) can be derived from the formula for the mean power $P = fNE_k$. The energy per pulse will increase as f is decreased, so for a solid target it is unfavourable to use low repetition rates. Increasing the repetition rate eventually causes the power consumption of the pulsed RF systems in the muon accelerators to become unacceptable, with the ISS recommending a compromise value of $f = 50\,\text{Hz}$, which will also be used here. Given this, the number of protons per pulse is given by

$$N = \frac{P}{fE_k} = \frac{4 \times 10^6}{50E_k} = \frac{4 \times 10^6}{50(10^9 e)E_k\,[\text{GeV}]} = \frac{4.9932 \times 10^{14}}{E_k\,[\text{GeV}]}.$$

There is a threshold for pion production that puts a lower bound on E_k. Pions are produced at low energies by nucleon-on-nucleon collisions of the form $N+N \to N+N+\pi$ like $p + n \to p + p + \pi^-$, with other production channels becoming available at higher energies, see section 4.6.1 for an example. This means the centre of mass energy must exceed the masses of two nucleons and a pion (a charged pion for our uses), which is 2.016 GeV for $p + p + \pi^-$. The centre of mass energy satisfies

$$E_{cm}^2 = (E_1 + E_2)^2 - (p_1 + p_2)^2,$$

where E_i are the total energies of the two colliding particles and everything is measured in energy equivalents: GeV/c for momenta and GeV/c^2 for masses. The second particle is a nucleus at rest, so $p_2 = 0$ and $E_2 = m_2$. The incoming particle has total energy $E_1 = m_1 + E_k$ and momentum $p_1 = \sqrt{E_1^2 - m_1^2}$. Some rearrangement gives the required incoming proton energy as

$$E_k = \frac{E_{cm}^2 - (m_1 + m_2)^2}{2m_2}.$$

Thus E_k must be at least 286.5 MeV for the reaction $p + n \to p + p + \pi^-$ and at least 292.3 MeV for the reaction $p + p \to p + n + \pi^+$. Near to this threshold the cross section will still be very low and the outgoing pions may have such low energy that they become reabsorbed in the target. Previously [25] beam energies of tens of GeV have been found to achieve sufficient pion yield; a detailed investigation of where the optimum energy is can be found in section 4.2.

Dependent on the optimal energy is the choice of proton accelerator technology. Table 3.2 lists examples of machines that have particular energies, together with their current status and beam power.

At low energies, acceleration can be done with a single-pass linear accelerator, the most ambitious proposal in this class being the 8 GeV "Project X" concept from Fermilab [44]. The CERN SPL (Superconducting Proton Linac) design is still ambitious at 3.5 GeV in its second design report [42] (with the option to upgrade as far as 5 GeV if necessary) but has been designed to a much higher level of detail. The main technical disadvantage of using a linac as the neutrino factory proton driver is that the neutrino factory Scoping Study report [24] recommends a 1–3 ns RMS bunch length whereas a linac generally produces a

Table 3.2: Planned and existing high-power pulsed proton accelerators. Beam energy figures in brackets are not values on the primary sampling of the energy range in the simulations described in this chapter, which is given by the remaining figures. Repetition frequencies are 50 Hz unless otherwise stated.

Proton driver	E_k [GeV]	Power/Schedule
CERN SPL I [40]	2.2	Study, 4 MW, 75 Hz, superseded
ISIS upgrade RCS 1 [41]	3	Study, 1 MW
CERN SPL II [42]	(3.5)	Proposal (first 160 MeV approved), 4–5 MW ultimate
	4	
Green-field RAL synchrotron [43]	5	Study, 4 MW
ISIS upgrade RCS 2 [41]	6	Study, 5 MW
FNAL Project X linac [44]	8	Study, 2 MW, 10 Hz
Scoping Study FFAG option [24]	10	Study, 4 MW
ISR tunnel RAL synchrotron [43]	15	Study, 4 MW, 25 Hz
FNAL RCS proton driver [45]	(16)	Study, 1.2 MW, 15 Hz
	20	
BNL/AGS upgrade [25]	(24)	Study, 1 MW, 2.5 Hz
CERN slow-cycling synchrotron [46]	30	Study, 4 MW, 8.33 Hz
	40	
JPARC main ring [47]	50	Beam in 2009, 0.75 MW, 0.3 Hz
	75	
	100	
FNAL injector/NuMI [48]	120	Running at 0.4 MW, upgradable to 1.2 MW by 2013, 0.75 Hz

Table 3.3: Proton driver specification from the International Design Study for a Neutrino Factory (IDS-NF) baseline, version 1 [49] in 2008. Note: that study uses a liquid target.

Proton Driver	
Proton power	4 MW
Proton kinetic energy	5–15 GeV
Pulses per second	50
Bunches per pulse	3
Minimum time between bunches	17 μs
Maximum time for all bunches	40 μs
RMS proton bunch length	1–3 ns

'pulse' of order 1 ms in length on each repetition. This means that the pulse must first be *accumulated* in a ring at the end of the linac to reduce its length to the circumference of the ring, corresponding to a time closer to 1 μs. Once the beam is in a ring, irrespective of whether that ring is an accumulator for a linac or the acceleration ring of a synchrotron or FFAG, it must be further compressed to the final nanosecond timescale. This is achieved via careful use of RF systems in the ring to shorten the bunch to less than the natural bunch length during acceleration or accumulation. The main problem in this phase at low energies is the longitudinal 'space charge' repulsion of particles from each other, which is enhanced by shortening the bunch and thus increasing its charge density. The space charge is also reduced steeply by increasing the beam energy (relativistic time dilation and space contraction allow more charge in the same volume of the laboratory frame) and the decrease of beam current that comes with keeping the power constant at higher energies. Thus the SPL proposal has the most difficult bunch compression system in table 3.2, with a high voltage of RF in the compression ring. The advantage of linacs is that their overall intensity can be made higher than an accelerator chain containing rings and the associated injection and extraction, since there is no timing constraint on the acceleration: linacs may even be operated in continuous or "CW" mode if the components can cope with the heat loading. This is the reason that the SPL and Project X proposals are actually multi-use machines where proton intensity over and above that required for the neutrino factory is used for other applications such as nuclear physics experiments or spallation neutron sources.

At higher energies, circular accelerators dominate, as the length (and thus cost) of a linac become prohibitively large. There is also a trend for the repetition rate of these accelerators to decrease as energy increases, which is for two reasons: the amount of RF voltage around the ring limits the rate at which protons can attain high energies and the constant power requirement means less average current of protons is needed, so bunches can be accelerated less often. With the highest energy rings, cost is again driven up due to an increase in length, but this time because the available dipole magnetic fields are no longer able to bend the high-energy protons around on a smaller radius of curvature. At low energies, the main challenge for circular accelerators is the increased beam current and sharply-increased space charge. Thus there could be said to be an optimum towards the mid-range of options shown in the table. Finally, note that the circular accelerators come in two types: synchrotrons in which the magnetic fields are varied during the acceleration cycle to scale with the beam momentum and FFAGs in which the variation of magnetic field strength is provided spatially across the accelerating aperture that the beam orbit moves over with energy. FFAGs are still an active area of research and there is not yet a conclusive cost/benefit comparison of these two ring types. As of 2009, no FFAG has been demonstrated to work in the same power or space charge regime as existing high-power proton synchrotrons.

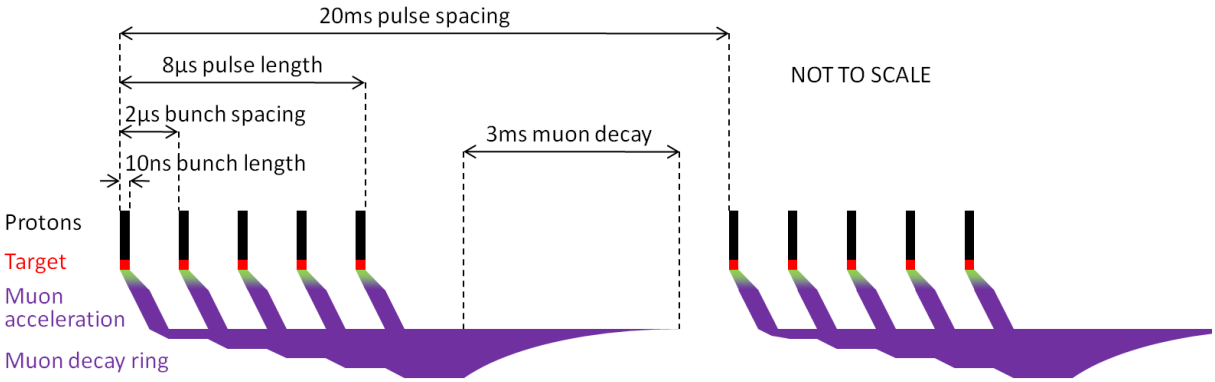


Figure 3.2: Bunch timing structure of a neutrino factory with a solid target, $f = 50\,\mathrm{Hz}$ and $h = 5$ bunches per pulse.

3.2 Bunch Spacing

The RF system of the final proton acceleration (or storage) ring provides an additional control over the bunch structure of the beam hitting the target. To make sure the longitudinal dynamics are consistent from turn to turn, the RF system must return to the same phase each time the centre of a bunch comes back to the cavity location, excluding slow adiabatic adjustments to the phase used in the acceleration programme. This requirement can be satisfied by making the RF frequency equal to the revolution frequency of the ring, but it can also be an integer multiple of this, with the multiplier known as h, the harmonic number. For example in the main ISIS ring [50] $h = 2$ and in the upgrade ring for a neutrino factory proton driver, $h = 6$ [41]. A consequence of having $h > 1$ is that another bunch at an angle of $2\pi/h$ behind the first in the ring experiences exactly the same RF programme and the same transverse dynamics too. It is therefore efficient to use all the stable phase space by filling all these RF 'buckets' in a ring if it is possible.

The consequence for the downstream systems is that if the final proton ring is designed with a harmonic number $h > 1$, then each machine pulse will produce up to h bunches that will hit the target with some delay from each other. If the entire turn is extracted at once then the spacing between bunches will be the same it was in the ring, usually less than $1\,\mu\mathrm{s}$. As the rise-time of the extraction kicker must be less than the time between bunches for successful extraction (unless there is an empty bucket used for this purpose), it is usually possible to have the kicker activate for only one of the bunches, which is sent to the target, then wait for a period of time before kicking out another and so on. Thus the delay between the bunches may be increased from the natural spacing, though in a synchrotron it cannot be increased so far as to make regular bunches at frequency fh because at some point the main magnet field must return to a low value in order pick up low-energy bunches to be accelerated in the next pulse. A storage ring would make regular spacing possible but at considerable cost; it must also be stressed that with regular extraction the frequency fh essentially becomes the repetition rate, so $fh \simeq 50\,\mathrm{Hz}$ would then be desirable. Usually the bunches are extracted close enough together that the resulting muon trains could all go through the muon acceleration system during a single pulse of its RF.

The target will react to this pulse substructure in a different way depending on whether it is solid or liquid. For a liquid jet, it is preferred that the bunches arrive close together (within 40 μs in table 3.3) before disruption or boiling of the mercury moves it out of the beam path. For a solid target, if the bunch spacing is sufficiently large to approach the oscillation time of acoustic (or shock) waves in the target rod, then several smaller shocks can be initiated at different phases, reducing or even cancelling the overall shock effect. This is shown in [51] where the oscillation time in a simulated 2 cm diameter tantalum rod is just under 10 μs and the overall stress amplitude is reduced by a factor of $\sim$5 by spreading the input shocks over a period of 10 μs rather than triggering them all at once. This would give the overall timing structure shown in figure 3.2.

One final aspect of the proton ring RF system that affects the bunches on the target is the bunch compression scheme that manipulates longitudinal phase space to reduce the duration of each bunch to $\sim$1 ns RMS. This is either done with programming of the main RF or an additional higher-frequency system and will act on all the bunches in the ring. It is required because introducing too much time spread into the distribution of muons from the target will increase their longitudinal emittance. It also means each proton bunch is virtually instantaneous from the point of view of the target material, meaning that thermal shock occurs, because the temperature has increased (from energy deposition) with no commensurate expansion of the atomic lattice. The shocks in fact propagate inwards from the target surface, which is able to expand into free space first.

3.3 Thermal Effects

The management of heating in the target caused by energy deposition from the proton beam and its secondary particles is one of the primary considerations in the target design. As mentioned earlier, the neutrino factory target cannot be made arbitrarily large in order to spread the heat load, so techniques for dealing with high energy densities are necessary. Additionally, surrounding the target in a coolant would lead to the reabsorption of pions and additional heat deposition, so radiative cooling schemes are considered.

The inner pie chart of figure 3.3 shows that the neutrino factory target is quite an efficient target for particle production, with most of the proton beam energy going into secondary particles rather than heat. The protons emerging at angles within 5 degrees of the rod's forward axis are singled out because they produce a 'hot spot' if they hit the target chamber wall at too close a range, with consequences discussed in section 3.5. These protons mostly consist of those that have been multiply scattered through the high-Z target material without experiencing a nuclear interaction. As such, they still have most of their original energy, $\sim$9.5 GeV for the 10 GeV proton beam example.

There is a terminology issue as to whether the neutrino factory rod is a 'stopping target' or not. The answer is that most of the primary protons undergo a nuclear interaction even in the 20 cm long case, with the proportion increasing as the rod becomes longer, though not attaining 100% because a few percent of the protons will always scatter out of the sides of the rod. However, 'stopping' in this sense does not imply that all of the power in the original beam is deposited as heat, as it is in neutron spallation targets, because such a large fraction of energetic secondaries escape permanently from the rod.

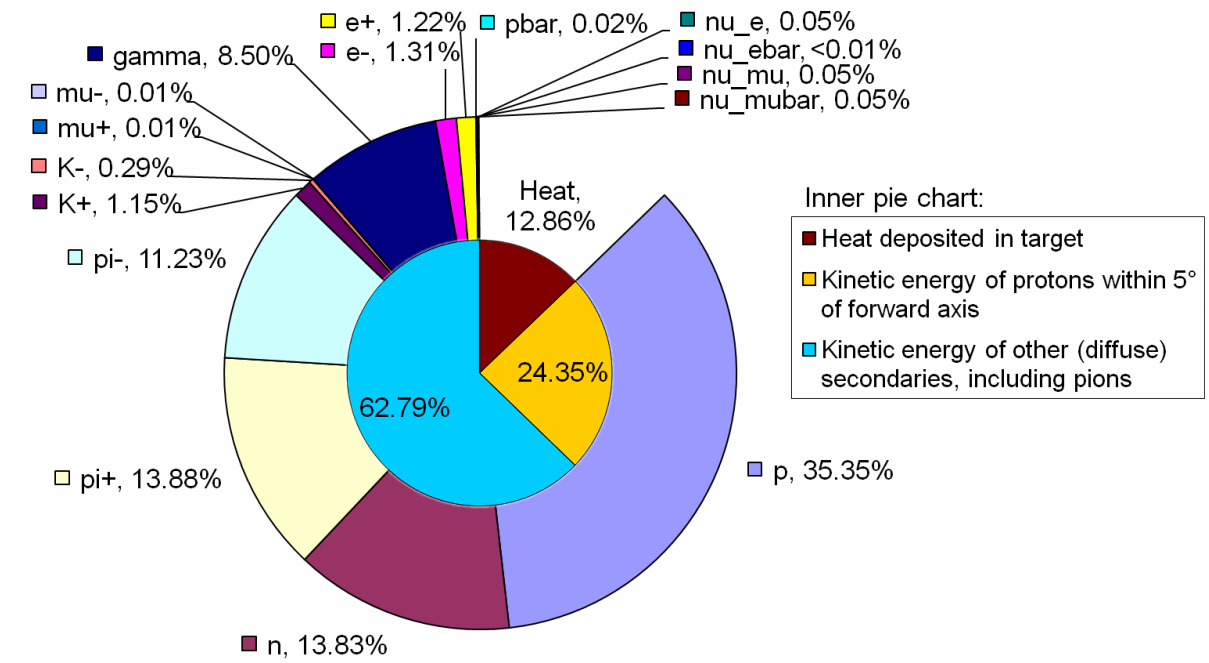


Figure 3.3: Where the energy from a 10 GeV proton beam goes when it hits a 20 cm long, 2 cm diameter tantalum rod target (MARS15 [2] calculation). For the variation of these figures with proton energy and length of rod, see sections 4.2.1 and 4.3.1 respectively.

The outer pie chart of figure 3.3 splits up the kinetic energy of the secondary particles by species. Positive and negative pions are the desired particles and constitute fairly large segments of the pie, carrying 24% of the original 4 MW power. Kaons also appear that can decay to usable pions later on. A final histogram by angle of emitted radiation (figure 3.4) shows that the target chamber will have to cope both with a diffuse component of the radiation and with the forward-directed 'hot spot' mostly coming from the protons. Apart from the target itself, the radiation will be most intense ($> 100\,\mathrm{W/cm^2}$) on the inner bore of the radiation shielding just inside the initial 20 T solenoid, meaning this shielding must contain extensive water cooling channels.

In the solid target design, the cooling of the target itself is achieved by radiation from the surface. The formula for thermally radiated power from an object of surface area A and temperature T is

$$P = \epsilon\sigma A T^4,$$

where ϵ is the emissivity, which may also depend on temperature, and σ is the Stefan-Boltzmann constant. In the approximation of continuous, volumetrically uniform irradiation and perfect thermal conductivity of the bar (i.e. an optimistic scenario), P can be set equal to the mean input power of heat deposited in the target. With the parameters used for figure 3.3 and a 4 MW proton beam, $P = 514\,\mathrm{kW}$ and $A = 0.0132\,\mathrm{m^2}$. The emissivity of tantalum or tungsten is ~0.32 at high temperatures [52] and decreases as T reduces. By rearrangement of the previous equation, the equilibrium temperature at which power radiated equals power input is

$$T = \sqrt[4]{\frac{P}{\epsilon\sigma A}} = 6808\,\mathrm{K}.$$

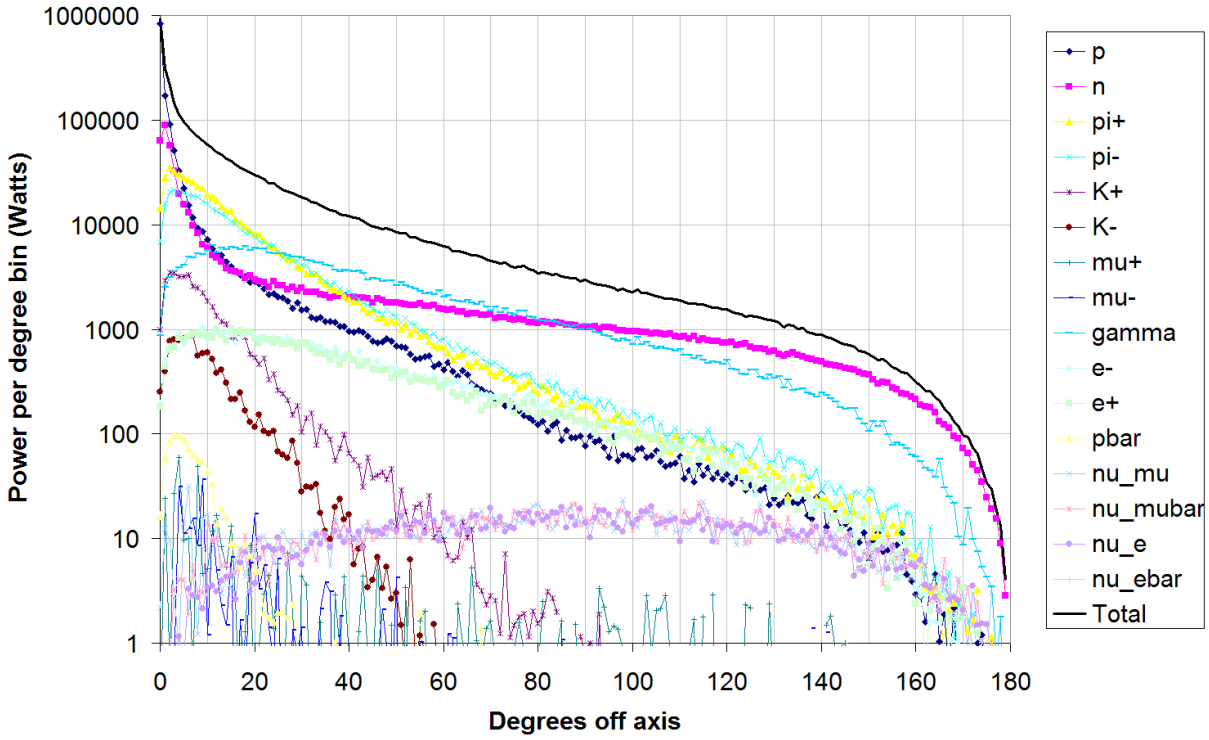


Figure 3.4: Graph showing the directionality of emitted power for each particle species, predicted by MARS15 [2] for the same setup as figure 3.3.

The melting point of tantalum is 3290 K and that of tungsten is 3695 K, so even under optimistic assumptions, a single bar of this size will not be able to radiatively dissipate the required amount of heat without melting.

It is now clear why the UK neutrino factory solid target design involves multiple bars that are replaced for each pulse. For N bars the above calculation can be redone with $P = 514/N$ kW, showing that $N \geq 12$ is required for tungsten not to melt and $N \geq 19$ for tantalum. These figures do not include margins for error and so on, with more realistic scenarios having $N \geq 200$ bars [53]. Tungsten demonstrates greater strength at high temperatures so is currently the preferred material option, though its pion production behaviour is very similar to the tantalum investigated in chapter 4 due to their proximity in the periodic table.

3.4 Geometry of the Solid Target Mechanism

Figure 3.5 shows a solid target chain concept in cross section, showing how the solenoid coils would have to be arranged such as to let the chain of rods pass through. Simulations in this thesis using Muon1 do not model the coils to this level of complexity (as the target design is still under development) but instead assume the magnetic field from a single 20 T initial coil.

To see how this affects the magnetic field profile, figure 3.6 shows how with coils set

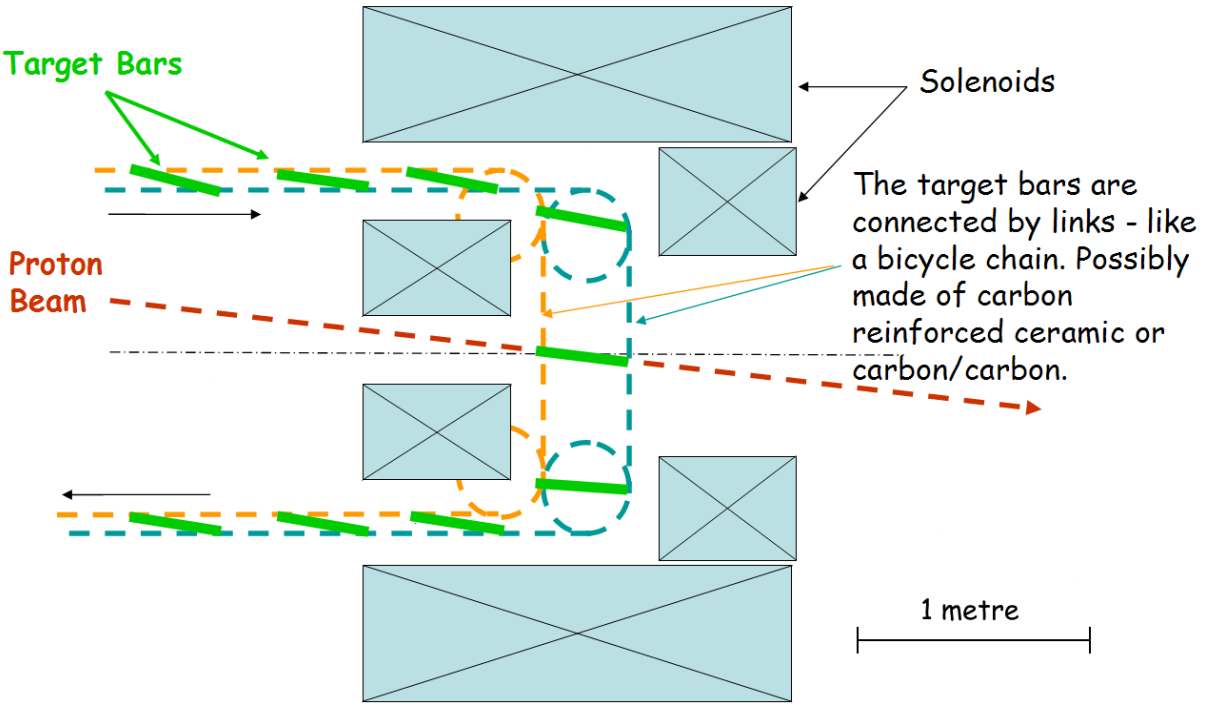


Figure 3.5: Schematic diagram of the solenoid arrangement surrounding the chain of solid targets, from [53].

at 50 A/mm^2, peak fields of 20 T can be created. One layout does this with a solid coil and the other splits the coil with enough room for the target and chain to pass through. The dip in on-axis field between the two coils in the second case goes no lower than 17 T. A few pions are tracked in the pictures, with one in the second case becoming trapped in a magnetic bottle in the field dip. However, that pion has a very small longitudinal momentum as evidenced by the steep pitch angle of its spiralling motion, so falls outside the region of 'interesting' pions defined in section 4.1.1.

Although the target station for the solid target mechanism is not yet engineered, we may usefully inherit features from the US Feasibility Study II [25] target cross section, reproduced in figure 3.7, which incorporates the correct amounts of shielding for the radiation fluxes produced. Any mechanism for removing the targets laterally will require the normal conducting initial coils as well as the superconducting coil and its cryostat to be split and the vacuum area extended to contain the chain or wheel of targets.

There are other options besides a chain for moving the solid target through the interaction area. Figure 3.8 shows the concept of a spokeless wheel or rim, which intersects through holes in the first solenoid and rotates continuously with targets attached so they are lined up with the proton beam for every 50 Hz pulse. Another option that is being considered is to have a powdered solid moving through the target station fast enough that each proton pulse hits a fresh volume of material. Figure 3.9 shows a tungsten powder jet formed by a proof-of-principle experiment at RAL [54, 55]. The powder may have advantages because it cannot shatter or fail mechanically in the same ways as a solid rod.

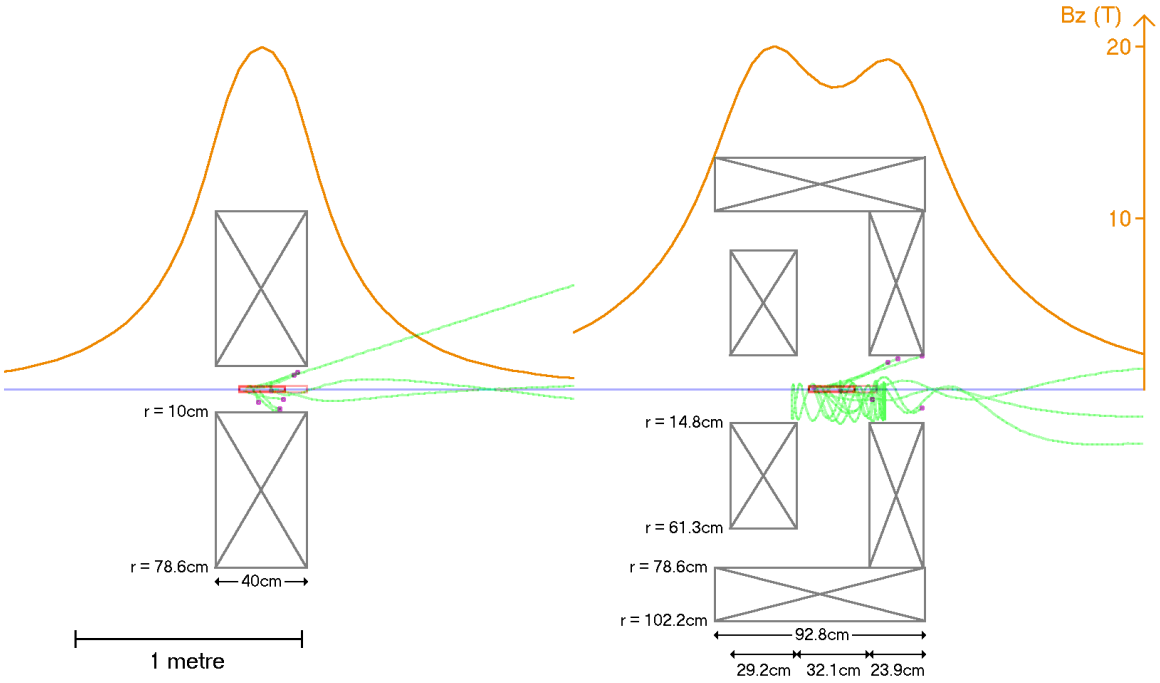


Figure 3.6: Cross sections of solid (left) and split (right) coil layouts around the target. The axial field profile $B_z(0, 0, z)$ is superimposed for a coil current of $50\,\mathrm{A/mm^2}$. The peak field in each case is 20 T. The red target rectangle is the 20 cm length case with the 30 cm, 3 cm diameter option shown by a fainter line.

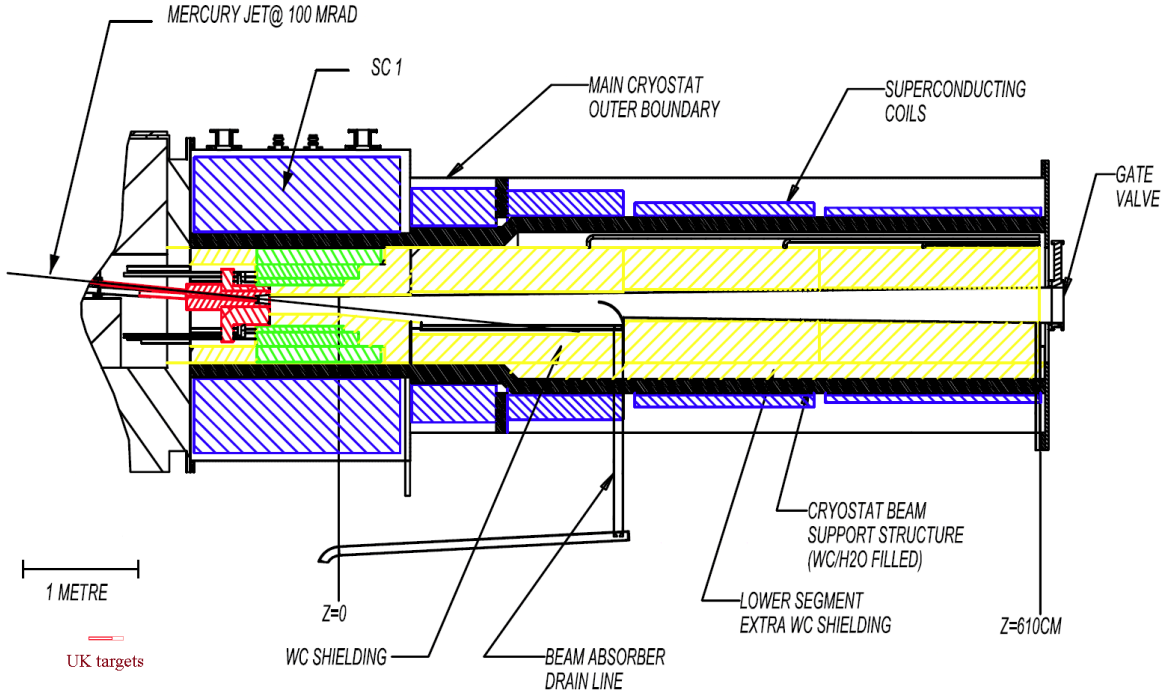


Figure 3.7: A cross section of the US Feasibility Study II [25] target station. Normal-conducting solenoids are shown in green and superconducting ones in blue. Water-cooled tungsten carbide shielding is yellow. The UK targets from figure 3.6 are shown to scale in the bottom-left corner.

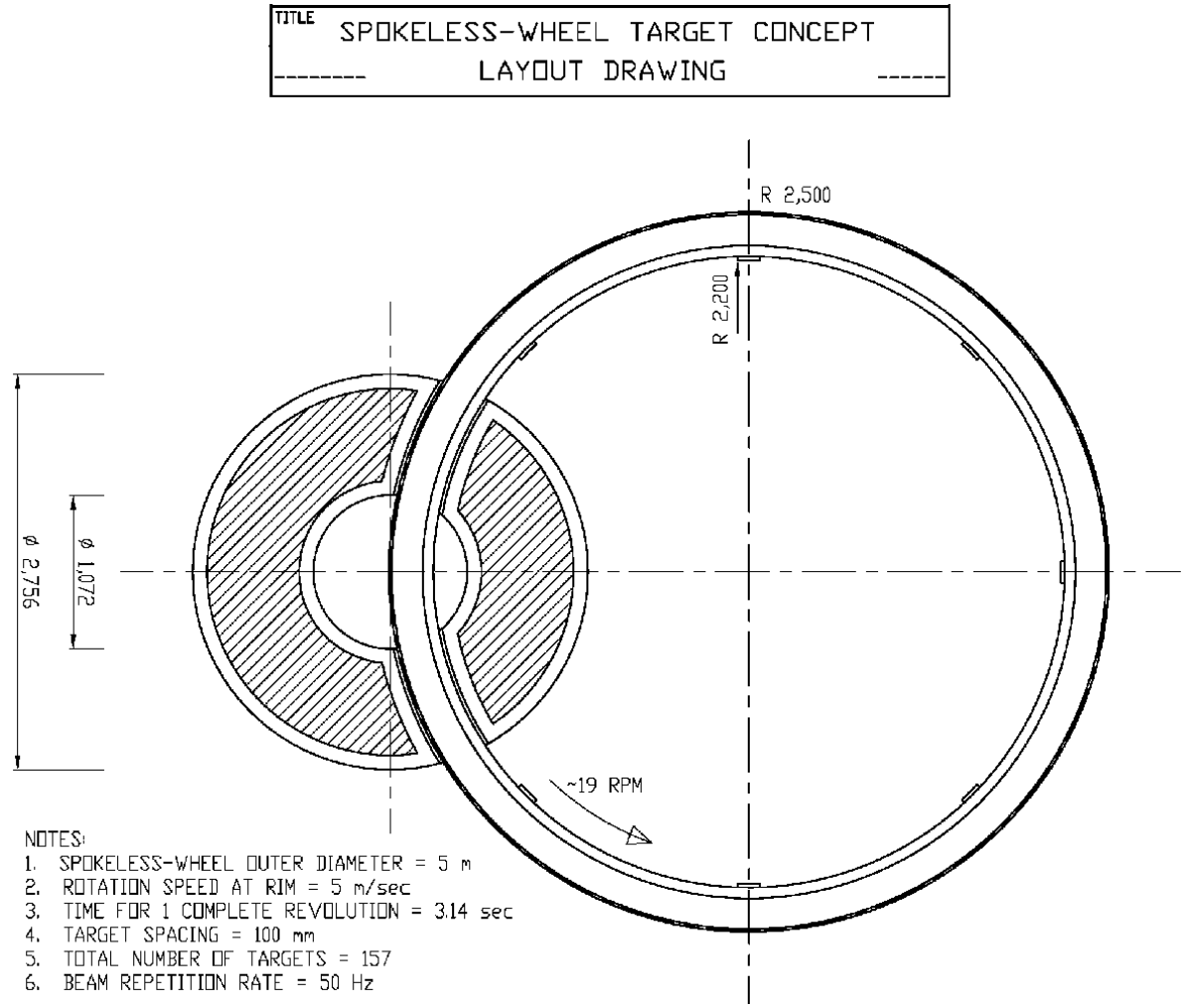


Figure 3.8: Cross-section of the rotating spokeless target wheel. The shaded annulus (left) is the coil of the first solenoid and target rods are attached to the wheel rim (right) with long axes pointing into the page. Adapted from [56].

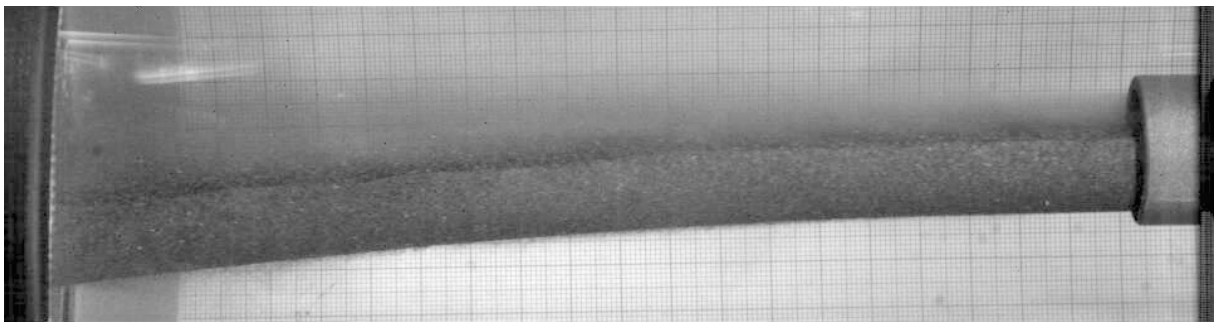


Figure 3.9: Photograph of an unconfined tungsten powder jet produced experimentally [54, 55]. The jet is flowing leftwards from the nozzle (right) at 3.7 m/s and the exposed portion of the jet is ∼20 cm long as measured by the 1/5/10 mm graph paper behind.

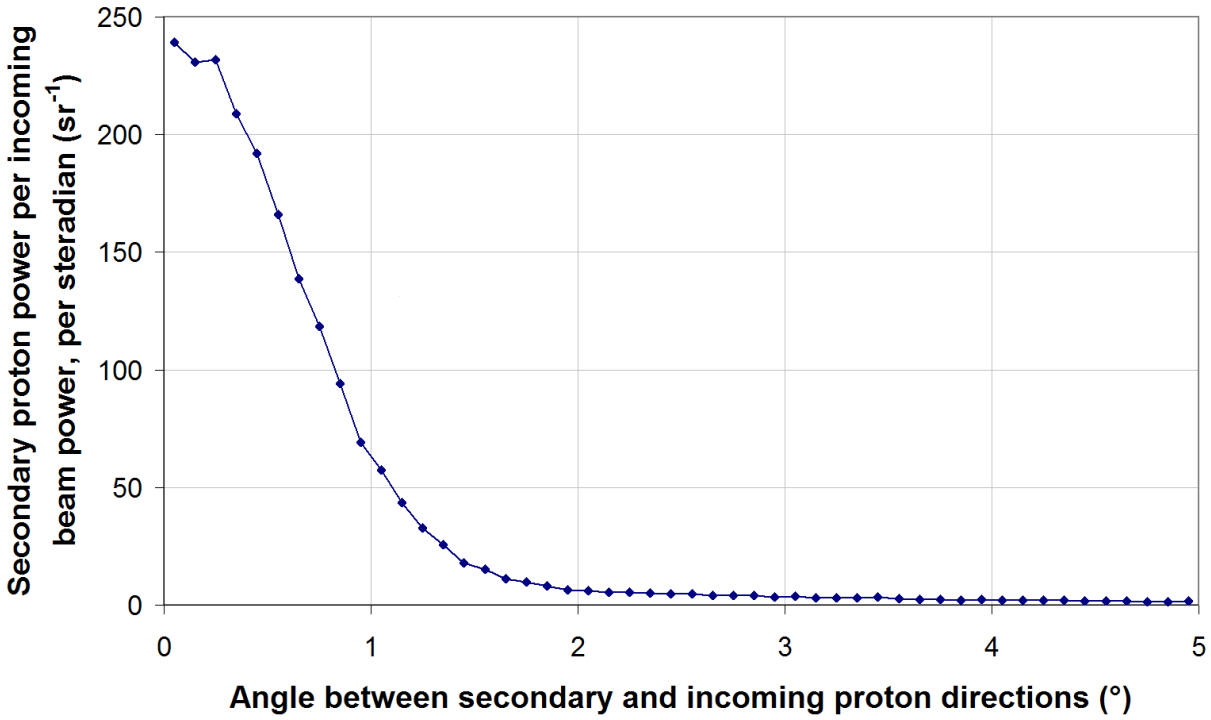


Figure 3.10: The angular distribution of remnant proton beam power within 5° of the incoming proton beam axis, from the same situation as in figure 3.3 (MARS15 simulation).

3.5 Beam Dumping

The UK target station beam dump differs from Study II, which uses the collection pool of the mercury jet to act as the beam dump for the 974 kW of forward directed protons (as defined in figure 3.3). The UK design uses a smaller angle between the incoming protons and the magnetic field axis, meaning that the protons will hit the side further down the muon decay channel. There a conventional water-cooled beam dump will absorb the energy over a larger area, as the protons impinge at a shallower angle. The large angle is needed in the liquid mercury design to keep the entire mercury flow loop within the initial cryostat, whereas solid target designs do not have this limitation.

Calculating the peak areal power density requires knowing something about the angular distribution of the forward-directed protons. In a given solid angle Ω (measured in steradians), the heating power H is the sum of the kinetic energies of protons arriving per unit time whose velocities lie within that angle. The rate of heat deposition will be directly proportional to P, the incoming proton beam power, so H/P is a constant independent of P. In the limit of small solid angles the dimensionless quantity $(1/P)\mathrm{d}H/\mathrm{d}\Omega$ evaluated at any given angle describes the concentration of secondary protons there. Figure 3.10 is a plot of $(1/P)\mathrm{d}H/\mathrm{d}\Omega$ against deflection angle and the peak of 239 $(\mathrm{W/W})\mathrm{sr}^{-1}$ is used in further calculations.

If the beam dump is situated a distance L from the target and the remnant proton

beam hits it with angle of incidence θ, the areal power density on the dump is

$$\frac{\mathrm{d}P_{\mathrm{dump}}}{\mathrm{d}A} = P\left(\frac{1}{P}\frac{\mathrm{d}H}{\mathrm{d}\Omega}\right)\frac{\sin\theta}{L^2}.$$

Values calculated using this formula are shown in table 3.4. This is not entirely accurate as the Study II figure is calculated using the tantalum rod target rather than the mercury jet, but shows the improvement gained by a larger distance to the dump and shallower incidence angle. Note that these power deposition figures are also strongly influenced by the target length (which reabsorbs protons), detailed in section 4.3.1.

Table 3.4: Comparison of the locations of the beam dump for the forward-directed proton 'hot spot' in Study II and UKNF designs. Peak areal powers are calculated using the formula given in the main text for a 4 MW incoming proton beam.

Parameter	Study II	UKNF	
Angle proton beam to magnetic axis	66.7	14	mrad
Distance beam dump downstream	3.75	25	m
Distance beam dump off axis	0.25	0.35	m
Peak areal power deposition	453	2.14	W/cm^2

Chapter 4

Pion Production Simulations

Muon front end simulations require an initial set of pion coordinates from the target. This chapter describes hadronic simulations that obtain pion distributions and allow these to be compared across different target parameters to find the best solution (sections 4.1–4.5). The target optimisation is decoupled from the muon front end in terms of yield, though some choices in the target have indirect effects on the later accelerator (section 4.6).

4.1 Modelling Particle Production in MARS15

The newest available version (15.04) of the MARS particle production code [2, 57] was used for the simulations. Later versions are likely to have updated models based on data from the HARP experiment [4]. As MARS15.04 does not, a benchmark of its results against some experimental data from HARP is given in section 5.2 to help gauge the level of accuracy of the current MARS models in the region of interest. The series of target simulations in the following sections uses the reference simulation in table 4.1 as the starting point from which to vary parameters of interest.

Table 4.1: Parameters of the target reference simulation. This configuration emerged early on as being somewhere near the optimum for pion production, so served as a starting point for investigations of variations in the parameters.

Code Used	MARS15
Target Material	Tantalum
Target Shape	Cylindrical
Target Dimensions	20 cm length $\times$ 1 cm radius
Protons Per Monte Carlo Run	10^5
Proton Beam Profile	Parallel beam, radius same as rod, parabolically distributed on circular target endface
Proton Direction	Parallel to target cylinder axis
Magnetic Field	None used (simulation only within target)

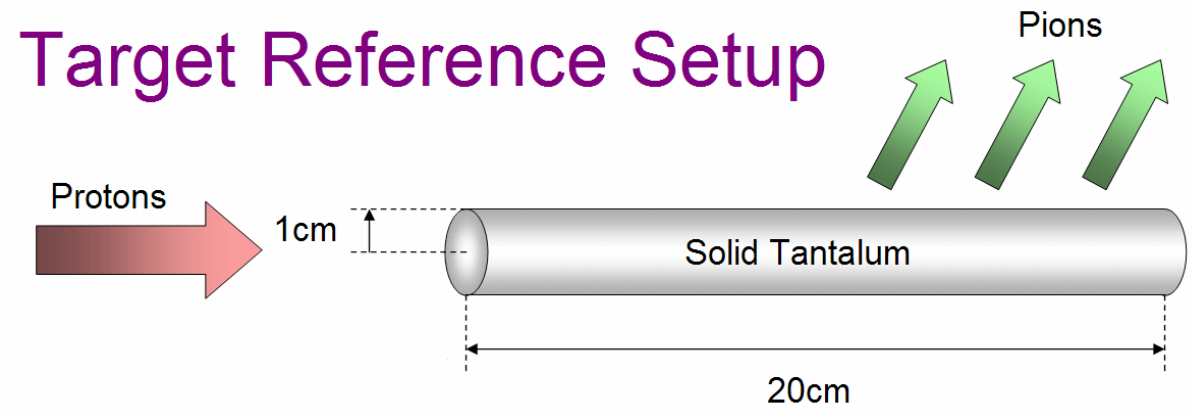


Figure 4.1: Schematic of the target reference simulation.

In the simulation, the proton beam hits the target end-on as shown in figure 4.1. MARS was configured to record all π^+ and π^- leaving through any of the target cylinder's surfaces. See appendix A for the input file used for these simulations. The target also produces copious quantities of neutrons; indeed, a stopping tantalum target is sometimes used in dedicated neutron production machines.

A radial parabolic distribution $\rho(r) \propto 1 - (r/R)^2$, $0 \leq r \leq R$, was used for the protons, with R the radius of the rod (see appendix B for generating routine). This is preferable to the rectangular and Gaussian beams that are defaults in MARS, as it is the same shape as the rod and does not lead to a proportion of the beam missing the target entirely. Another option frequently used in target studies is a uniform distribution on the circlular endface, but this is hard to achieve in practice with a proton extraction line. The parabolic distribution, however, is the projection into the plane of a 4D 'waterbag' phase-space distribution, which is often used as a model of proton bunches.

The incoming beam was assumed to be parallel in the simulation, as typical proton drivers have beams with too small an angular divergence to greatly affect the results of a hadron shower calculation in a block of material. For instance, the ISS baseline proton driver [24] has a beam with a 100% un-normalised emittance of 175 mm·mrad just after injection at a kinetic energy of 200 MeV, which becomes 10.3 mm·mrad at the 10 GeV extraction energy, scaling with $1/(\beta\gamma)$. At a beam focus or 'betatron waist', this emittance is the product of the beam radius and maximum particle angle (x'). Setting the radius to the 1 cm of the reference rod gives a beam divergence of 1.03 mrad. Even at the lowest considered energy of 2.2 GeV this is only 3.77 mrad, which would produce a 0.75 mm transverse offset in a 20 cm rod and little difference in interacting path length.

MARS can add a magnetic field as would be present in the rod in this situation. However, this may change in value if the magnetic design of the front end is to be optimised, so one field would not be correct for all simulations. The field strength also has little effect on the pion production, since over the width scale of the rod it produces only a slight curvature of the pion trajectories of interest. For instance, a typical pion spiralling on an $R = 5$ cm radius helix produced at the centre of an $r = 1$ cm radius rod will have its velocity vector rotated about the rod axis by $\theta \simeq r/R$ by the time it exits the rod. The displacement at the rod's surface will be to leading order $\frac{1}{2}\theta r$ but the difference in

path length Δl, important in particle production is given by $\Delta l/r = \frac{1}{6}\theta^2$ plus higher order terms. Here, $\frac{1}{6}\theta^2 \simeq \frac{1}{150}$ so the differences in interacting path length due to the magnetic field actually fall below the statistical errors in the Monte Carlo simulations.

The above simplifications mean that these pion production results are applicable quite generically for pion production from protons hitting a cylindrical target and the optima and trends discovered may be useful outside the context of the neutrino factory.

4.1.1 Pion Capture Probability Grid

To evaluate the number of 'useful' pions produced by the target, the pions leaving its surface have their longitudinal and transverse momenta (p_L, p_T) recorded and plots such as figure 4.2 show this momentum distribution and the later behaviour of the pions in the muon front end. The MARS code was not available to run on the distributed optimisation network, so it was not practical to run MARS in conjunction with every front end investigated. Instead, a pion distribution is simulated through two representative muon front ends, each producing a probability function for pion survival as a function of their initial (p_L, p_T) coordinates.

This makes optimisation of the target much quicker since a simple postprocessing step can now turn a target pion distribution into a figure for useful pions, rather than requiring tracking to be performed each time. The effect is similar to using momentum 'cuts' but is based more closely on the actual behaviour of the particles in the representative muon front ends.

Figures 4.2 and 4.3 display the fate of pions as a function of their initial (p_L, p_T) coordinates in two front ends chosen for their contrasting design principles. Figure 4.2 uses RF phase rotation to a low energy spread from section 8.3, whereas figure 4.3 uses a magnetic chicane buncher and linac from section 8.2. Since decay to a muon occurs in the front end, the end result is not deterministic: the range of possible decay muons originating from a single pion causes the blurring of the coloured regions in the figures. An additional small amount of blurring will come from any correlations of the behaviour with the pion's starting position on the target surface because reducing the initial five phase-space variables (assuming axial symmetry) into two (p_L, p_T) idealises the rod as a point source. However, the plots can still be used to produce a method for counting useful pions if they are treated as showing survival *probability* as a function of initial momentum.

To obtain probability figures, the plots were binned into $30 \times 30\,\mathrm{MeV/c}$ squares and the ratio of survived muons (white points on the plots) to the total number of initial pions was calculated in each one. In fact, since MARS generates particles with statistical weights, the ratios were calculated between sums of weights for the appropriate pions:

$$P_{ab} = \frac{\sum\limits_{\substack{\text{Transmitted muons } i: \\ (p_L,p_T)_i \in \text{Square}(a,b)}} w_i}{\sum\limits_{\substack{\text{Original pions } i: \\ (p_L,p_T)_i \in \text{Square}(a,b)}} w_i} = \frac{T_{ab}}{W_{ab}},$$

where the numerator and denominator of this expression define T_{ab} and W_{ab}.

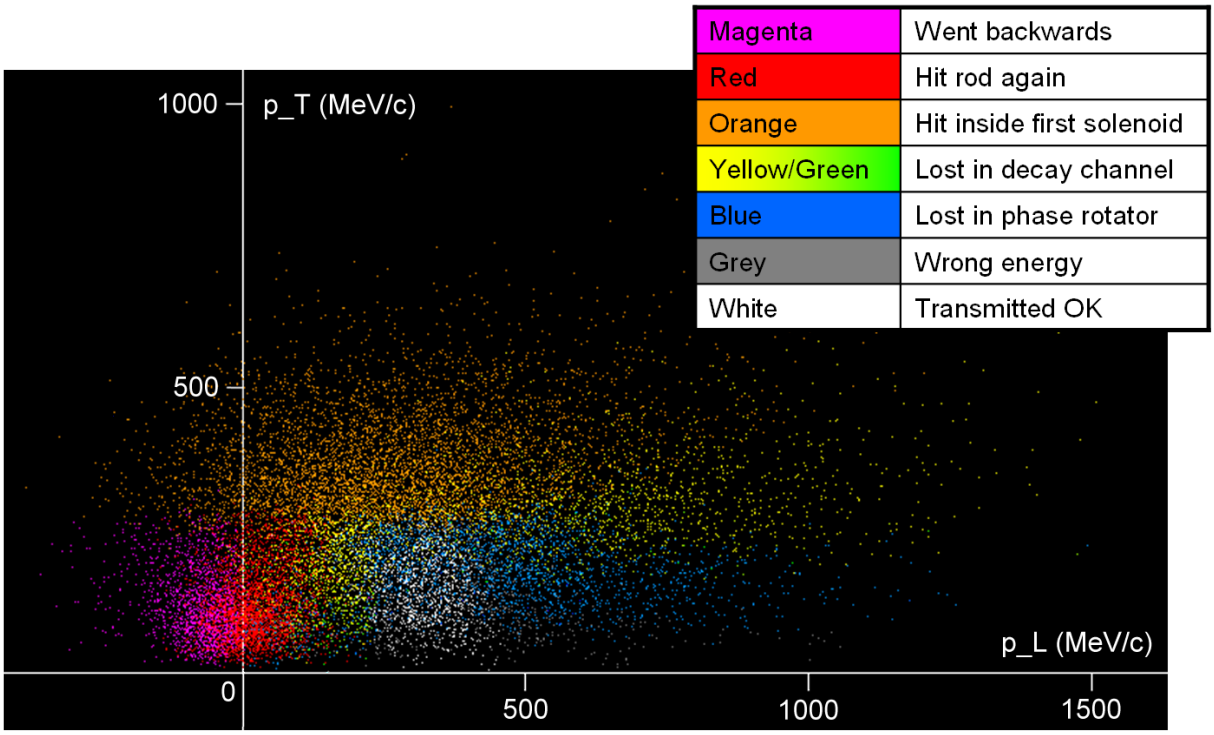


Figure 4.2: Scatter plot of the initial transverse and longitudinal momenta of pions at the target, coloured by their fate in a particular phase rotation front end. If the pion decays, the fate of its daughter muon determines the colour.

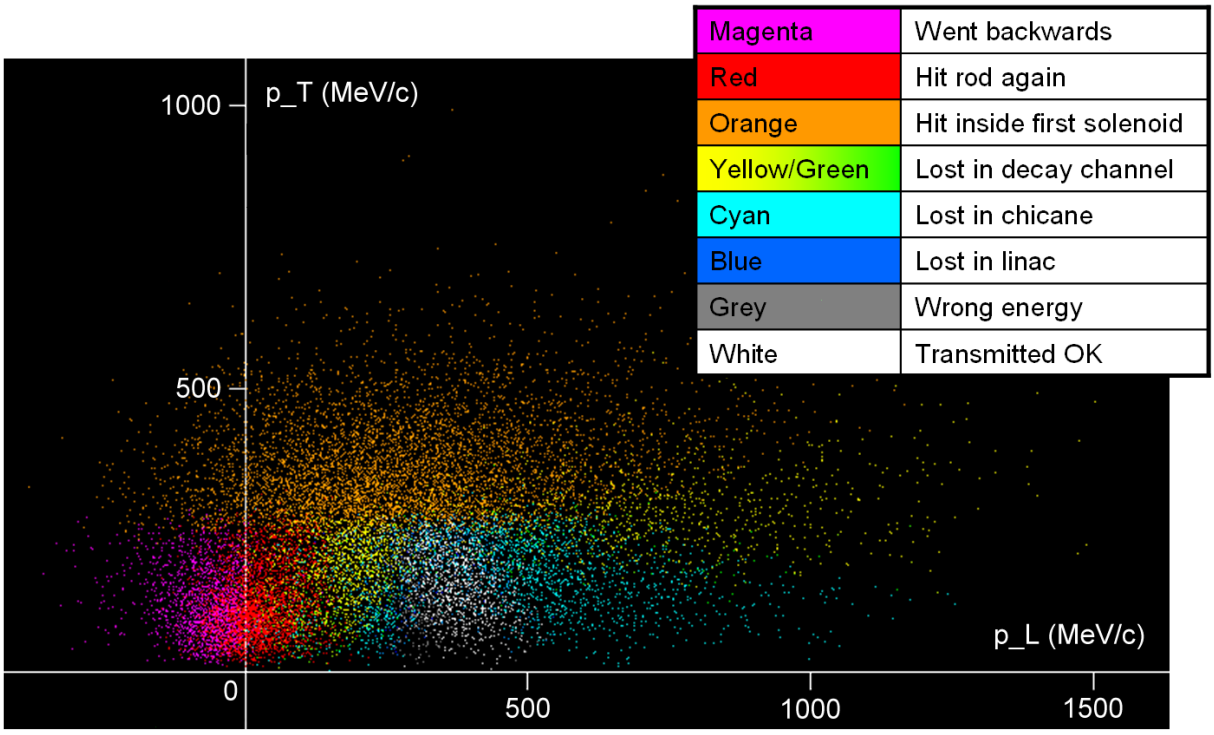


Figure 4.3: Using a front end with a chicane and pre-acceleration linac in figure 4.2 instead of a phase rotator gives different loss mechanisms but the white area of pion momenta that decay into transmitted muons is similar, with a slight difference in central energy.

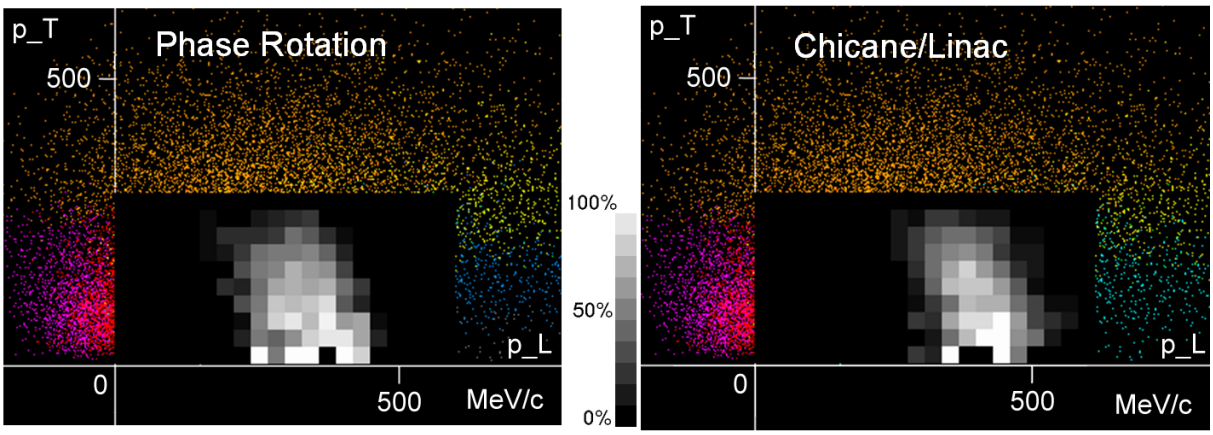


Figure 4.4: Probability grids derived from the particle acceptance data in figures 4.2 and 4.3. Lighter blocks represent higher probabilities of pions in that block decaying into transmitted muons; black is zero probability, white would be 100%.

The resulting matrices of survival probability $[P_{ab}]$ are shown in graphical form in figure 4.4. Both lattices had a few squares with this probability exceeding 90%, so there is a subpopulation of pions that will nearly always produce successful muons regardless of random decay effects or emission position on the rod.

Unfortunately the low statistics in some squares, particularly near the $p_T = 0$ axis where the true phase space volume represented by the square is smaller, generate unwanted noise and fluctuations in the probability map. Smoothing the values between adjacent squares appears to be one way of mitigating this. As it happens, often a square with bad statistics that is by chance nearly empty (W_{ab} is small) has neighbours with comparatively good statistics, which should be given more weight when calculating an average. Empirically, the following scheme produced a substantial improvement in the probability maps:

$$\tilde{P}_{ab} = \frac{T_{ab} + 0.1(T_{a-1,b} + T_{a+1,b} + T_{a,b-1} + T_{a,b+1})}{W_{ab} + 0.1(W_{a-1,b} + W_{a+1,b} + W_{a,b-1} + W_{a,b+1})}.$$

The resulting smoothed probability matrices $[\tilde{P}_{ab}]$ are plotted in figure 4.5.

Table 4.2 gives the numerical values of the probability grid squares for the phase rotation lattice. This grid is used to define *useful pions* in the remainder of this chapter; that is, if pion i is in square (a_i, b_i), then

$$\text{Useful pion weight} = \sum_i \tilde{P}^{\text{phaserot}}_{a_i b_i} w_i.$$

The range of pions transmitting muons with more than 50% probability corresponds roughly to the region with total momentum $250 \leq |\mathbf{p}| \leq 450\,\text{MeV/c}$ and transverse momentum $p_T \leq 200\,\text{MeV/c}$. The effect of using the chicane/linac grid instead of $[\tilde{P}^{\text{phaserot}}_{ab}]$ is demonstrated in section 4.1.2

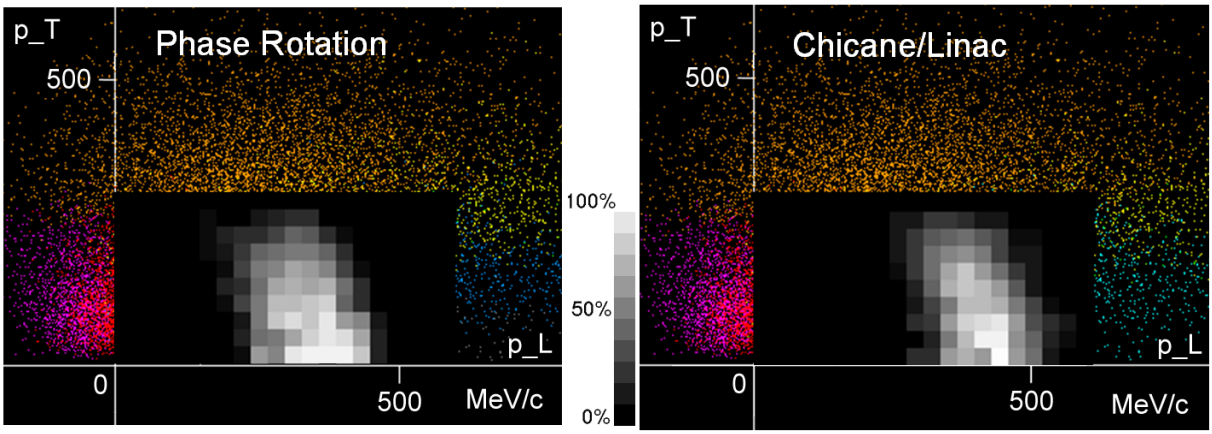


Figure 4.5: The tracking-derived grids in figure 4.4 smoothed to remove some aberrant values originating from statistical noise.

Table 4.2: Probability of successfully transmitting a muon through the phase rotation front end, as a function of bins in the (p_L, p_T) space of the original pion emitted from the target surface. The bins are 30 MeV/c in size and the lower end values are given in the top table row for p_L and the left-hand column for p_T, with zero transmission outside the region shown. Probability values are in percent.

MeV/c	0	30	60	90	120	150	180	210	240	270	300	330	360	390	420	450	480
0	-	-	-	-	-	-	-	6.0	58.8	54.2	91.9	92.6	92.3	92.3	76.8	15.6	-
30	-	-	-	-	-	-	1.1	14.2	45.2	47.6	64.6	79.9	90.5	86.3	65.7	13.0	0.8
60	-	-	-	-	-	-	0.8	12.2	49.3	78.7	85.3	77.5	88.5	62.7	51.4	9.2	0.5
90	-	-	-	-	-	0.2	6.9	28.6	49.4	74.4	74.5	77.4	80.6	39.7	17.0	1.8	-
120	-	-	-	-	-	1.5	14.4	19.4	42.6	64.8	70.0	66.3	67.0	41.9	20.5	1.5	-
150	-	-	-	-	0.4	2.1	5.8	27.5	47.0	53.5	65.3	59.4	53.8	26.5	5.0	0.2	-
180	-	0.1	-	0.4	2.9	5.6	12.0	24.2	41.5	45.3	49.7	49.1	38.2	17.2	1.8	-	-
210	0.1	1.4	0.3	1.7	2.6	6.6	15.0	18.5	20.9	26.3	41.4	34.9	18.6	6.7	0.5	-	-
240	-	0.2	-	0.2	1.6	4.5	2.6	3.5	9.7	13.9	20.1	20.7	4.0	0.6	-	-	-
270	-	-	-	-	0.1	0.5	0.1	0.1	1.1	3.4	1.6	1.5	1.5	0.2	-	-	-

In a constant **B** field and the absence of any other forces, the path of a charged particle will be a helix with a certain (constant) radius that depends on the momentum of the particle transverse to the magnetic field direction and the strength of that field. This is known as the *Larmor radius* and can be calculated directly from the equations of motion given in section 2.1 to be

$$r_{\text{Larmor}} = \frac{p_T}{|q||B_z|},$$

where q is the charge of the particle. Using MeV/c and units of the elementary charge gives a handier version of this formula for this sort of calculation:

$$\begin{aligned} r_{\text{Larmor}}\,[\text{cm}] &= \frac{p_T\,[\text{MeV/c}]\ (10^8/c)}{|q\,[\text{e}]|\ |B_z\,[\text{T}]|} \\ &\simeq \frac{p_T\,[\text{MeV/c}]}{3|B_z\,[\text{T}]|} \qquad \text{for a singly-charged particle.} \end{aligned}$$

The solenoid containing the target has $|B_z| = 20\,\text{T}$ and $q = \pm e$ for a pion, so the helix radius that the $p_T = 200\,\text{MeV/c}$ cut-off point corresponds to is $r_{\text{Larmor}} = 3.34\,\text{cm}$, using the exact version of the formula above. This is the largest helix radius at which more than half the pions are transmitted through the first solenoid. The centre of the helix will not be the solenoid axis: instead, the location the pion was emitted from the target surface will be at one side of the helix and the furthest excursion will be $2r_{\text{Larmor}} = 6.67\,\text{cm}$ from this point. This is approaching, though not equal to, the 10 cm radius of the solenoid bore but the particles are now sufficiently far off-axis to risk encountering the non-linear fringe field of the solenoid on the way out, in which case the helix follows the diverging field lines to larger radii rather than the solenoid axis, leading to loss. The largest possible transmittable helix would have a diameter of 11 cm for the standard 1 cm radius target (that is, from the far side of the target to the inner bore), so we would expect a complete removal of particles with $r_{\text{Larmor}} \geq 5.5\,\text{cm}$. This corresponds to $p_T \geq 330\,\text{MeV/c}$ and the transmitted particles disappear completely just below this value, with $300 \leq p_T < 330\,\text{MeV/c}$ being the first empty row of bins, just below those shown in table 4.2.

4.1.2 Difference in Yield Between Probability Grids

In the last section, two different probability grids were generated and it was stated that the phase rotation grid would be used for all subsequent yield evaluation rather than the one derived from the chicane/linac design. As evidence for this choice, some of the yield graphs presented in section 4.2 are evaluated here using both grids but with all other attributes kept the same; in fact they come from the same MARS particle output file. Specifically, the standard target with parameters in table 4.1 was used and the proton beam energy varied from 2.2 GeV to 120 GeV using the primary values in table 3.2.

Figure 4.6 shows the results for just the positive pions. The chicane/linac design has an overall lower efficiency than the phase rotator, for reasons discussed in the tracking results of chapter 8, which is one reason for it to be excluded. Another question would be whether the two different designs show different qualitative behaviour: as the purpose of the target optimisation is to find the location of the yield peak in such parameters as

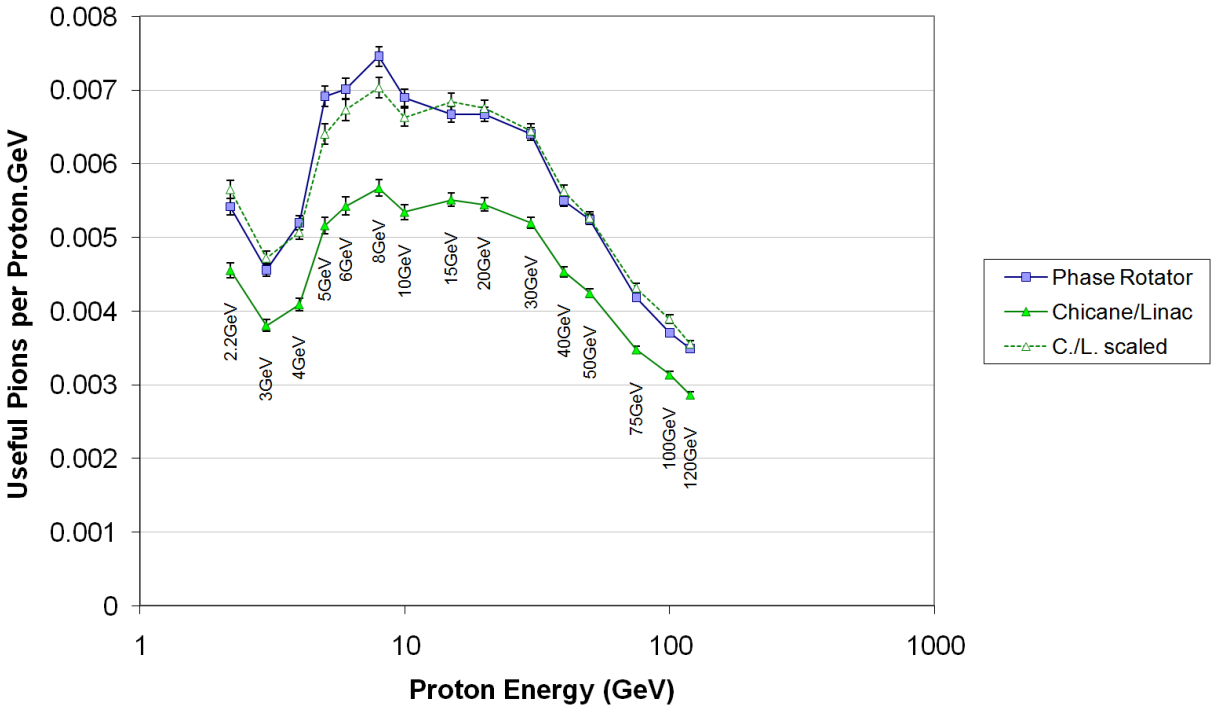


Figure 4.6: Comparison of useful π^+ yields calculated with two different probability grids over a range of energies. 1σ statistical error bars are shown.

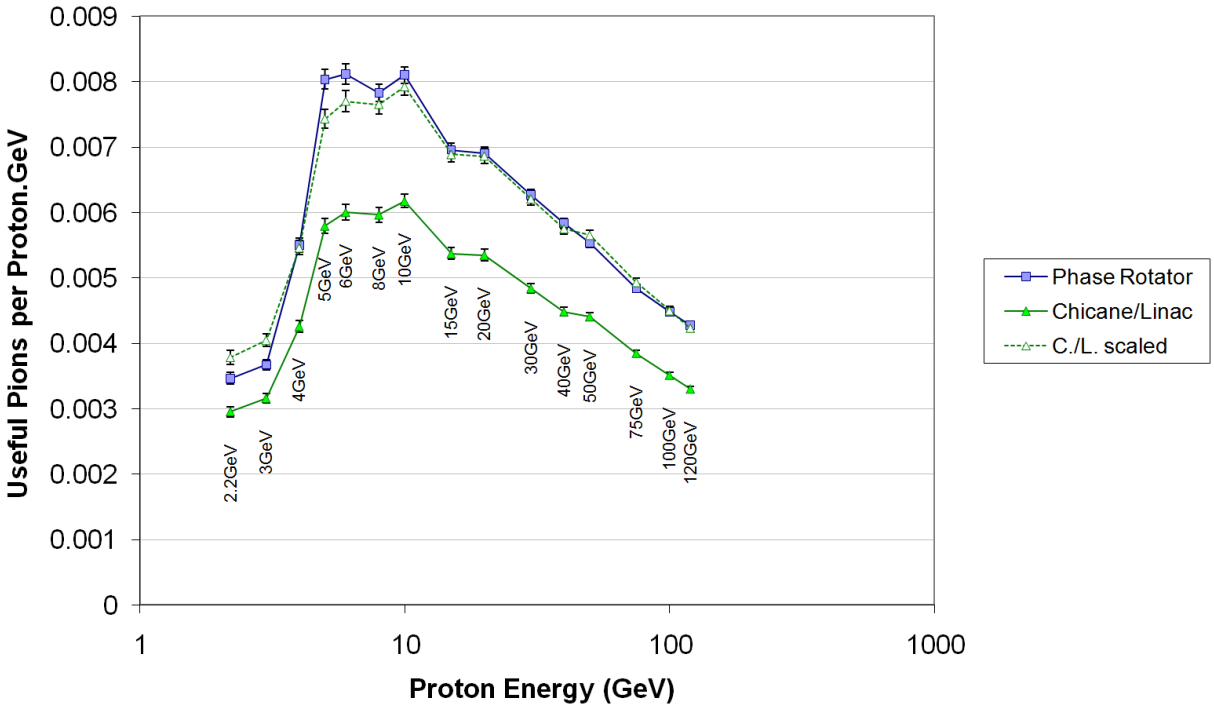


Figure 4.7: Comparison of useful π^- yields calculated with two different probability grids. 1σ statistical error bars are shown.

proton energy, rod length and so on, the absolute yield matters less than the position (and shape) of the optimum. To compare the shapes without the yield difference, the dotted line on the graph is a scaled version of the chicane/linac results. The scaling factor chosen was the average ratio between the two original yield series.

Figure 4.7 shows the same three lines for the negative pions, which have a slightly differently-shaped graph to the positives. In both cases, the scaled chicane/linac line stays very close to the phase rotation line, with a maximum difference of 10% for low proton energy (and low-yielding) simulations producing negative pions. These differences are only 2–3σ in terms of the statistical error on each data point but are probably systematic because they have the same sign in groups of nearby energies. However, the overall shapes of the graphs and locations of the optimum ranges are still similar, so the phase rotation grid is used as a representative figure of merit considering it both represents a higher-yield front end and evaluates design choices similarly to the chicane/linac grid.

4.2 Proton Energy and Rod Material

This section contains what are probably the most important findings about pion yield gained from the MARS15 studies. Section 3.1 described how the proton energy is a key parameter that has a large effect on the design of the proton driver, so the optimal energy must be determined before design work can continue. The proton energy is varied in the simulations here and the results plotted to show the location and shape of this optimum.

Although high-Z (Z = atomic number) targets from period 6 of the periodic table have conventionally been the choice for the neutrino factory, this study also tried alternative materials at intermediate and low-Z. Specifically, at high Z, the tantalum of the standard UK solid target configuration at the time (table 4.1) and the liquid mercury used in US Study II [25] were both run. For an intermediate-Z target, copper was chosen, more because it was the densest element in period 4 of the periodic table than for its suitability as a real target: it melts at perhaps too low a temperature to be considered for a solid target with radiative cooling. The objective was to change Z without changing the other parameters affecting the hadronic showering behaviour (primarily density) by too much. Carbon in the form of graphite is an establised low-Z target material that is finding applications in neutrino superbeam experiments [19]. Boron is a few percent denser (2.34 vs. 2.26 g/cm^3) but is not so commonly used, thus carbon is the element chosen from period 2.

There is still a significant difference in density between these elements so a legitimate comparison must give them each a different length of target so the protons will have the same probability of interacting. The relevant parameter that determines how frequently protons undergo nuclear interactions with path length in a material is its *hadronic interaction length*. This is used in table 4.3 to find a reference target length for each material that is used in the simulations of varying proton energy that follow. The original 20 cm tantalum rod corresponds to 1.79 hadronic interaction lengths.

The table also includes molybdenum (a candidate element from period 5 that was not simulated) and tungsten, which later became the preferred solid target due to superior

Table 4.3: Hadronic interaction length data from the Matprop tool at SLAC [58] allows reference lengths for other materials to be scaled from the 20 cm tantalum target.

Element	Z	Interaction length (cm)	Equivalent to 20 cm Ta (cm)	Rounded
Ta	73	11.18	20	20 cm
Hg	80	13.95	24.96	25 cm
C	6	36.92	66.05	66 cm
Cu	29	15.16	27.12	27 cm
W	74	9.54	17.07	17 cm
Mo	42	14.96	26.76	27 cm

mechanical strength at high temperatures. Its pion production behaviour will not be very different to the other period 6 elements plotted, which follow each other closely.

Figure 4.8 plots the useful pion yields (as defined in section 4.1.1) for all four materials over the primary range of energies given in table 3.2, showing each sign of pion separately. The units are chosen to be proportional to a constant beam power across all energies. The measure of "useful pions per incoming proton" may seem a sensible choice, but a 100 GeV proton carries much more kinetic energy and thus can create more hadronic particles than a 3 GeV proton. This is fixed by using the unit of "useful pions per proton·GeV", where the fast proton in the previous example would count for 100 p·GeV and the slow one for 3 p·GeV. As the denominator is now just a measure of beam kinetic energy, the numerical yield rate of useful pions can be calculated from the proton beam power P:

$$\text{Useful pion rate}\,[\text{s}^{-1}] = \frac{P\,[\text{W}]}{10^9 e}\ Y,$$

where Y is yield in these units.

There is an overall decreasing trend on all materials and signs above 10 GeV, with more marked differences in the low-energy behaviour. High-Z targets also fall off at low energy, exhibiting an optimum, whereas the carbon target continues to rise almost uninterrupted. This was naturally of interest, so later plots include an extension of the carbon simulation range to proton energies as low as 800 MeV. Unfortunately, as discussed in section 3.1, compressing the proton bunch to the order of nanoseconds becomes infeasibly difficult for energies below about 3.5 GeV, so this carbon peak must remain of theoretical interest unless breakthroughs are made in proton bunch compression schemes.

To clarify the optimisation, the pion yields of different signs may be summed to give the total yield as shown in figure 4.9 for each material. An optimal plateau in yield exists for the high-Z elements in the range 5–15 GeV. The 'interesting' increase of carbon yield at low energy finally reaches a maximum at 1 GeV before falling off. The situation for each sign individually, as examined in the NuFact conference paper [59] on these results, gives an optimal range (for high-Z targets) of 5–30 GeV for π^+ and 5–10 GeV for π^-. The lower end of the plateau at 5 GeV may not be as sharp in reality since MARS changes between different physics models over the range 3–5 GeV (see figure 5.3 and chapter 5 in general for comparisons between codes and physics models).

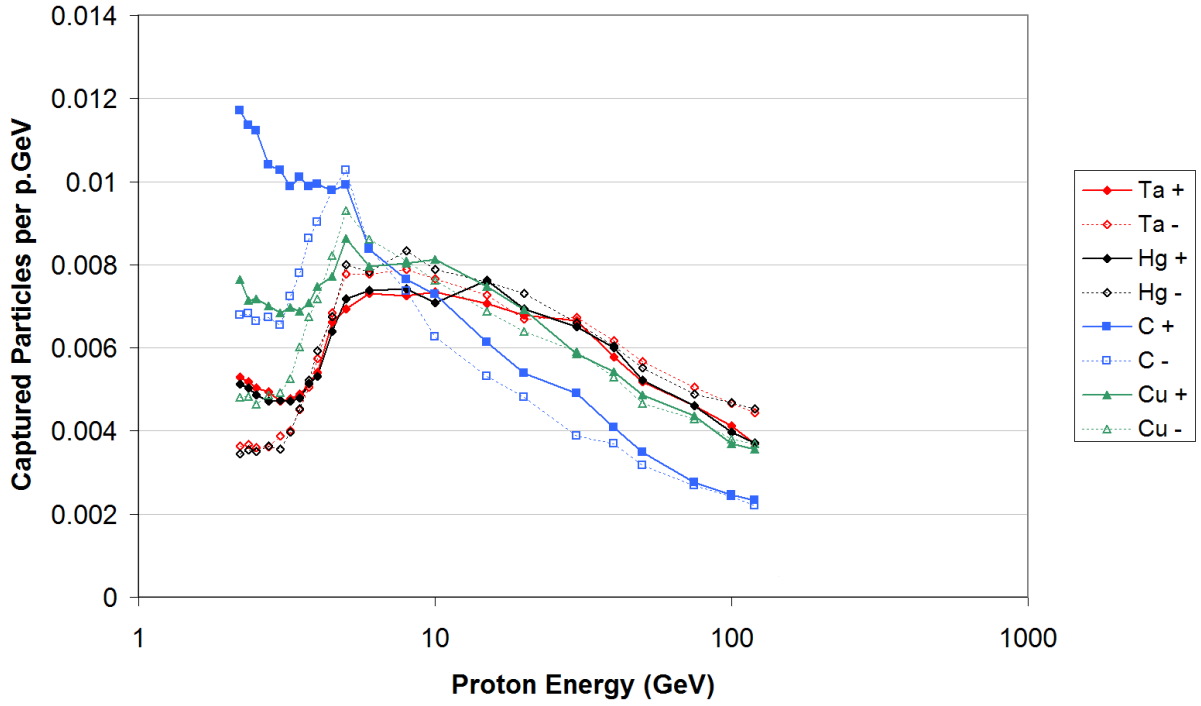


Figure 4.8: Variation of single-sign useful pion yields with incoming proton energy for four different materials. Note the log scale for proton energy. Error bars will be omitted from yield graphs from now on for clarity, but errors will be similar to those in figure 4.6.

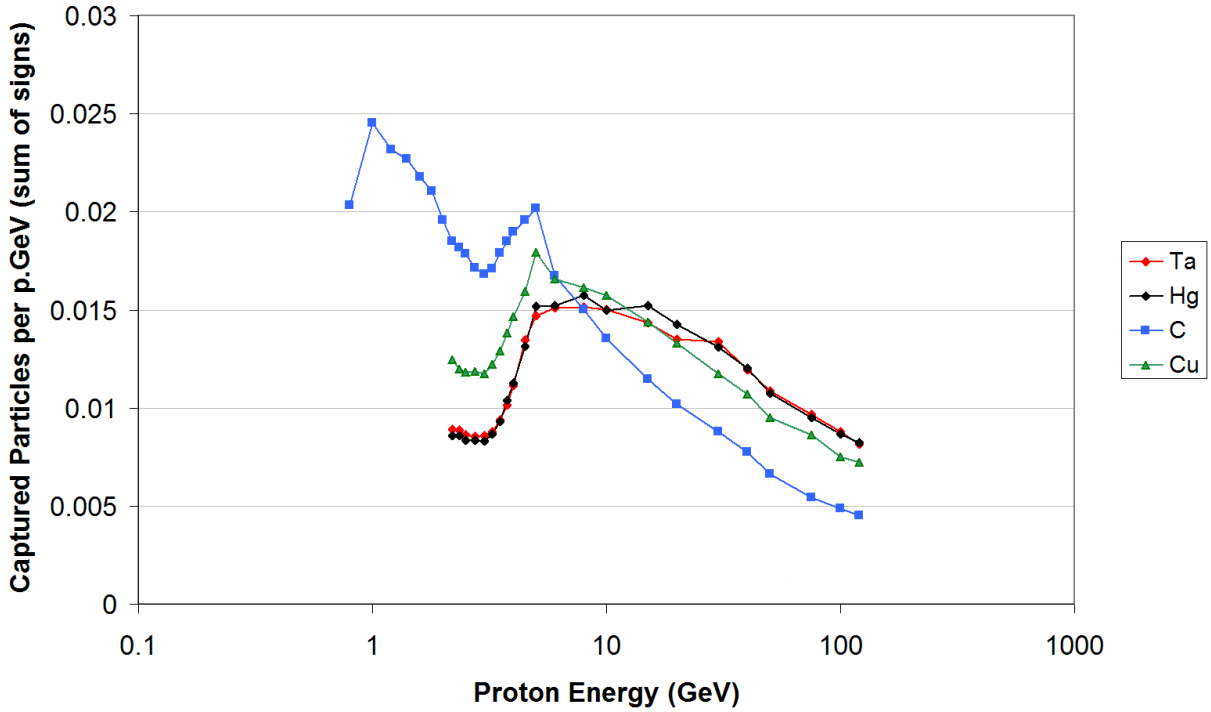


Figure 4.9: Total useful yield of pions with any sign as a function of energy, for four different target materials.

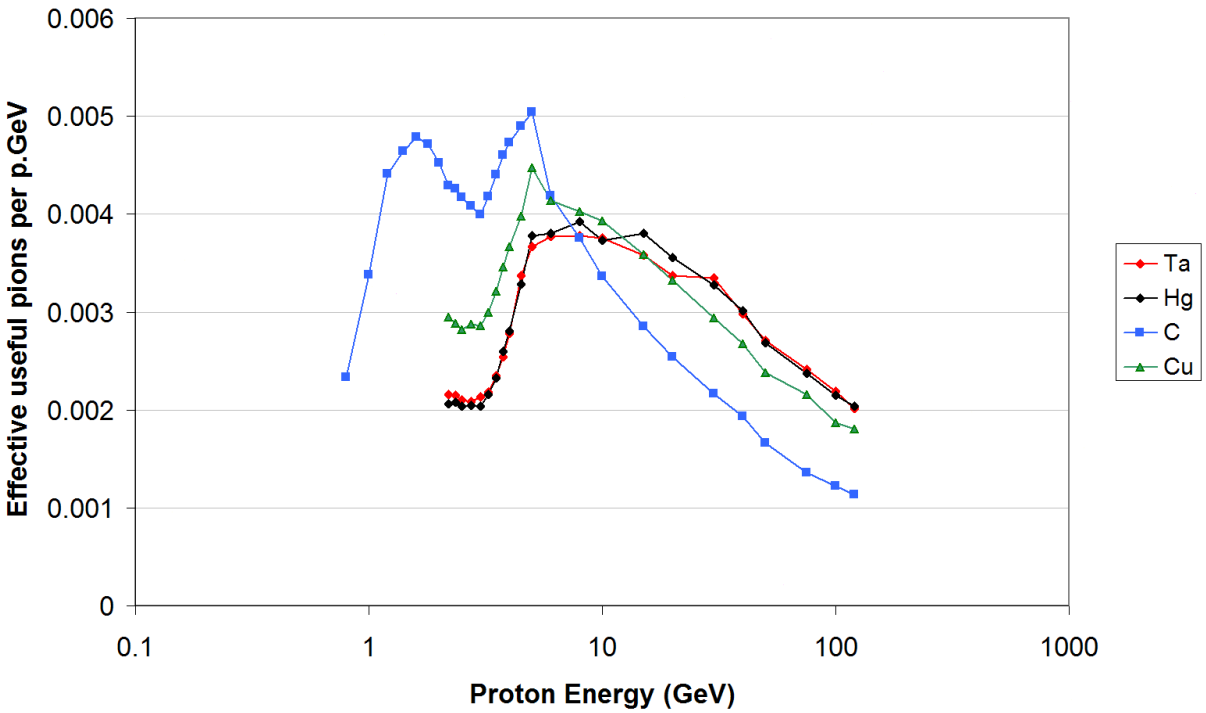


Figure 4.10: Effective useful yields of pions for CP violation calculations, which are calculated from results of both neutrinos and antineutrinos.

When the target produces different numbers of positive and negative useful pions and thus ultimately different numbers of neutrinos and antineutrinos of each flavour, it is sensible to ask what the impact of this lopsided mix of particles will be on the physics results. Evaluating the mass differences Δm^2_{ij} relies on measuring the oscillation length accurately and as particles and antiparticles have the same mass, the statistics from both neutrinos and antineutrinos undergoing the same oscillation channel may be added. So for this application the summed figure of merit in figure 4.9 is appropriate. However, the CP-violating angle δ is measured via the relation [60]

$$s_{12}c_{12}s_{23}c_{23}s_{13}c^2_{13}\sin\delta \propto P(\bar{\nu}_\alpha \to \bar{\nu}_\beta) - P(\nu_\alpha \to \nu_\beta),$$

where $s_{ij} = \sin\theta_{ij}$, $c_{ij} = \cos\theta_{ij}$ and α and β are some nonequal neutrino flavours. Thus the parameter $\sin\delta$ is calculated from the difference of two oscillation probabilities, one of which is measured using the neutrinos and the other with the antineutrinos. The individual statistical errors in these probabilities would be proportional to $1/\sqrt{Y_+}$ and $1/\sqrt{Y_-}$, which add in quadrature to make the statistical error in $\sin\delta$ proportional to $\sqrt{1/Y_+ + 1/Y_-}$. This is the same scaling behaviour as if the measurement had come from 'simple' statistics with an effective yield $Y_{\text{eff}} = (1/Y_+ + 1/Y_-)^{-1}$, half the harmonic mean of the two signed yields and less than either of them individually. Y_{eff} is plotted in figure 4.10, where although most of the graphs are a similar shape to the yield sum, the low-energy carbon peak is diminished because it was made up of a very uneven balance of positive and negative pions. Positive pions dominate in the low-energy regime, in part due to the charge excess coming from the incident protons when production of $\pi^+\pi^-$ and similar charged pairs is not completely dominant.

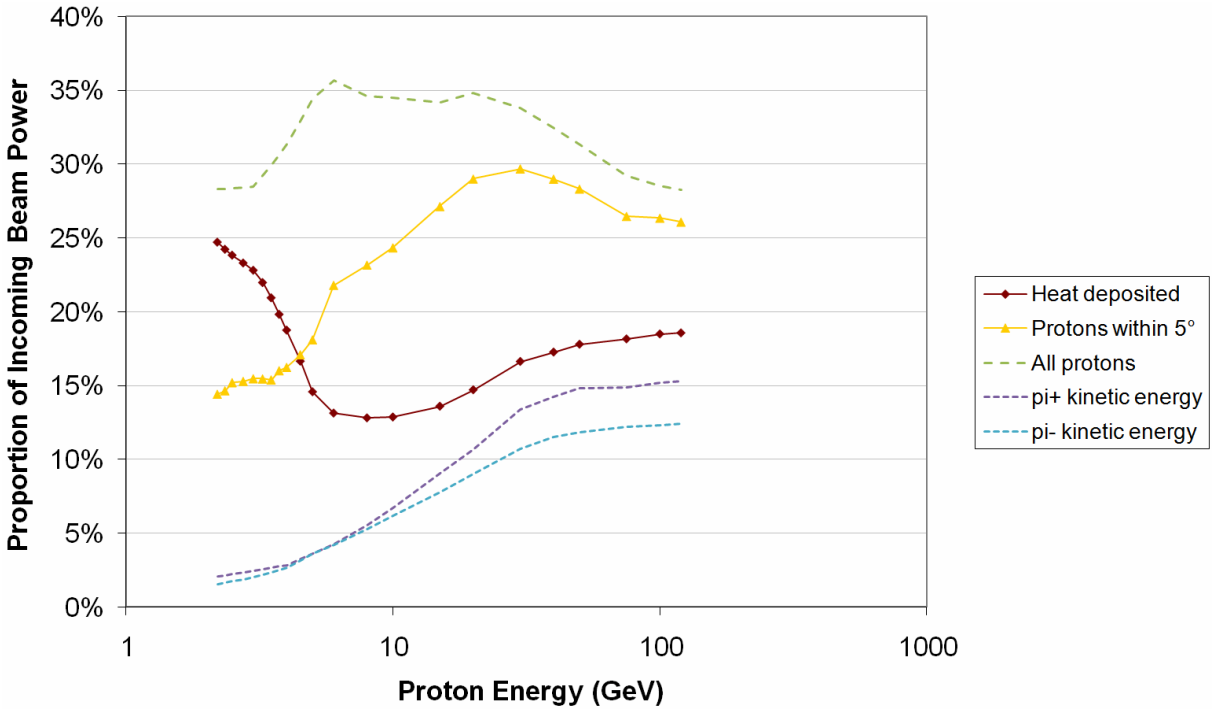


Figure 4.11: Variation of heating in the 20 cm-long tantalum target with incoming proton beam energy, which exhibits a minimum around 8 GeV. The kinetic energy of protons going to the forward beam dump and other selected secondaries are also plotted.

4.2.1 Variation of Heat Deposition with Energy

While figure 3.3 showed the division of incoming proton beam power for the standard target in table 4.1, this apportionment will change if the incoming protons have a different kinetic energy. MARS outputs energy deposition data by default in the file `MARS.OUT` and this subsection uses that file to examine the results from the energy scan of the tantalum target in the previous section.

Figure 4.11 plots the results of integrating the energy deposition within the target to get the total heat deposition at each proton energy. It also shows the variation of the proton power emitted in the forward cone of 5 degrees from the central axis, which was the other part of the energy flux that was separated out in figure 3.3's inner pie chart. Table 4.4 has the same results in numerical form. As section 3.3 described, the amount of heating in the target is a determining factor in the design, so minimising the heat deposited is an important goal when optimising the proton energy. Fortunately, comparing figure 4.11 with figure 4.9 reveals that the optima for particle production and low heating lie at almost the same energy. The range for which the summed useful pion yield for tantalum is within 10% of the optimum value is 5–15 GeV, whereas the range for which the rod heating is no more than 10% above the minimal value is 6–15 GeV.

Figure 4.12 shows which areas of the rod are heated the most by the incoming proton beam at various energies. The heating is not uniform, in fact uniform heating would be a fairly bad model of the situation so should only be used as an upper bound on the performance, such as in section 3.3. The heating density ϑ is plotted in units of kg^{-1}

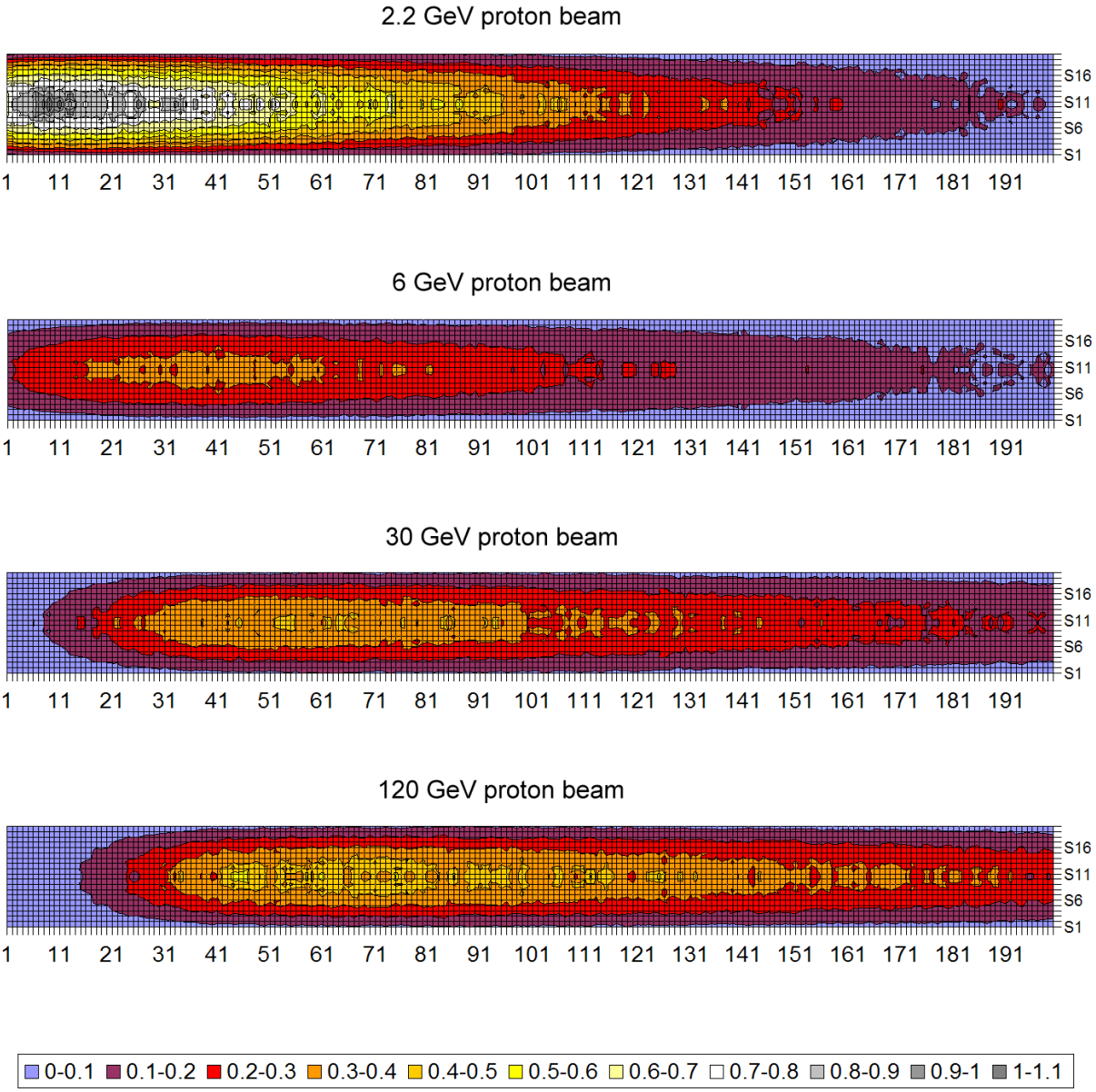


Figure 4.12: MARS15 simulations of heat deposition distributions for the standard tantalum rod with four different incoming proton beam energies. The grid cells are 1 mm squares and the proton beam enters from the left. Heating is measured in Watts per kilogram (of target material) per Watt of incoming beam.

Table 4.4: Table of heat depositions in the rod as a function of proton beam energy. The figures in kilowatts are scaled for a 4 MW proton source. The rightmost columns give the power going into the forward-directed protons of concern for the forward beam dump discussed in section 3.5. Compare with table 4.5.

E [GeV]	Heat deposited	kW	Protons within 5°	kW
2.2	24.7%	988.7	14.4%	577.0
3	22.8%	912.4	15.5%	619.2
4	18.8%	750.1	16.2%	649.8
5	14.6%	582.9	18.1%	724.5
6	13.1%	525.1	21.8%	872.3
8	12.8%	512.0	23.2%	927.0
10	12.9%	514.4	24.3%	974.0
15	13.6%	543.7	27.1%	1085.9
20	14.7%	588.3	29.0%	1160.4
30	16.6%	665.2	29.7%	1187.9
40	17.3%	690.5	29.0%	1160.2
50	17.8%	711.7	28.3%	1133.1
75	18.2%	726.3	26.5%	1059.0
100	18.5%	739.7	26.4%	1054.1
120	18.6%	742.7	26.1%	1043.4

because the plots are universal for heating that scales with incoming proton power or energy. Thus the unit may be thought of as either $(\mathrm{J/kg})/\mathrm{J_{pulse}}$ or $(\mathrm{W/kg})/\mathrm{W_{beam}}$. The real heating power H in a volume V may be recovered using the formula

$$H = P_{\mathrm{beam}}\, V \rho_{\mathrm{Ta}}\, \vartheta.$$

The peak of the heat deposition distribution moves further down the rod as proton energy increases. At high energies this is because a 120 GeV proton beam of a fixed power consists of fewer protons each carrying over 100 times its own rest mass in kinetic energy. This energy is gradually converted into the mass of more hadronic particles in successive interactions forming a shower through the rod, in which the kinetic energy of daughter particles successively decreases as more hadrons are dislodged from nuclei. This means that the bulk of the energy deposition is delayed in the case of higher-energy protons until a large number of slow (<GeV) hadronic secondaries are passing through the rod material. This happens further down the rod as more generations of the shower are required to break down the energy in this way. At low energies, the incoming protons are already of an energy where they produce slow secondaries, a significant number of which may even stop in the target, so the 2.2 GeV heating distribution is intense and has a peak almost at the front end of the rod. The overall intensity of heating in the rod is lowest for intermediate proton energies, reflecting the minimum in figure 4.11 as expected.

The distribution shapes are not so radically different from each other to change the previous conclusion (derived only from the total heating) about which proton energies are preferred from the point of view of providing a favourable thermal environment for the target, however.

4.3 Rod Length

The length of the rod can be varied to values other than the standard lengths in table 4.3. The optimisation consideration is to increase the rod length so that enough of the protons interact but not so much that the generated pions will spiral in the magnetic field and be reabsorbed by the rod. This is, however, dependent on the magnetic field value and requires evaluation of the pion production at least at the endplane of the target, which is not what the surface production simulations here with no field give. Instead, they will show only the production element of the situation, without the reabsorption.

Work on the reabsorption scenario has been done by Harold Kirk [61] and John Back [62], giving similar optimal lengths to those assumed in the reference scenario for materials in the previous section, which is in turn similar to the 'conventional wisdom' of using 2 interaction lengths for such target. Varying the length produces a broad optimum with little variation in yield around this point, this is in fact why the author assumed some scaled lengths instead of optimising them first. Thus the simulations in this section just provide some information about the pion source characteristics along the length of the rod and a 'sanity check' on the MARS results compared to theory.

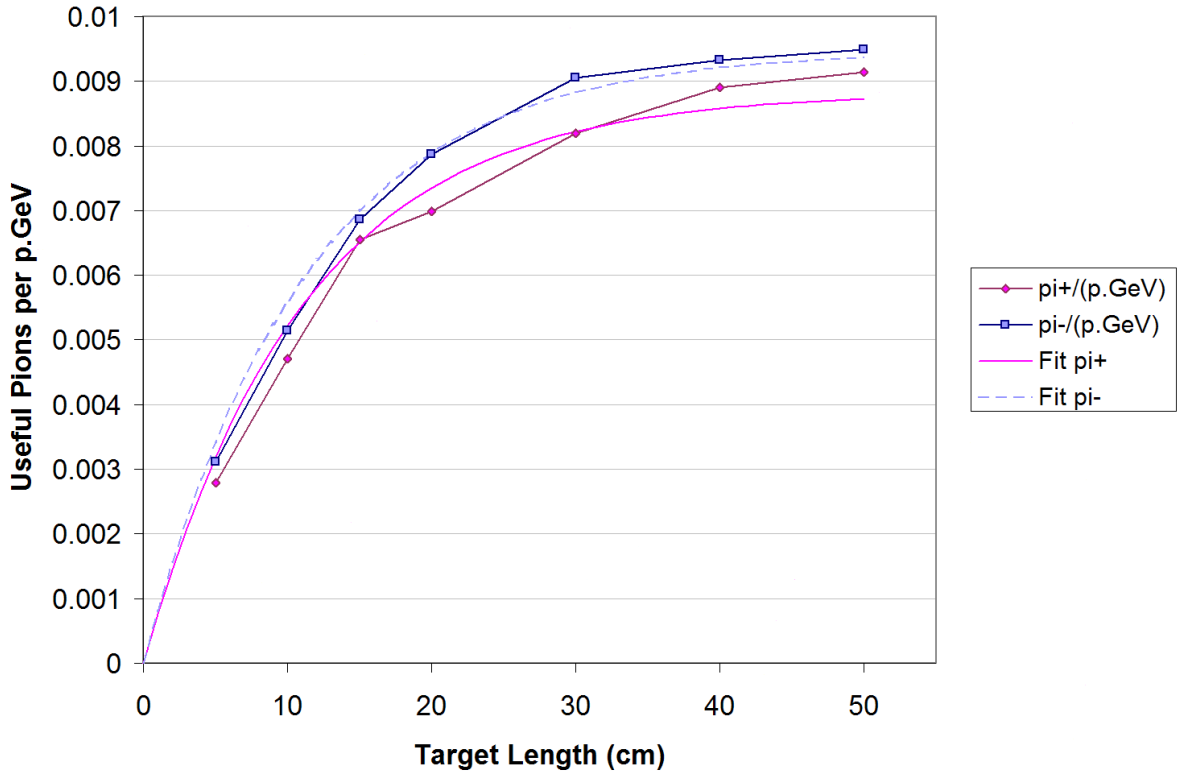


Figure 4.13: Useful pion production from tantalum rods of various lengths, excluding reabsorption, together with inverted exponential fits using the 11.18 cm hadronic interaction length of tantalum as the decay constant. All other simulation parameters are as in table 4.1.

Figure 4.13 shows the simulated production of pions weighted by the usefulness criterion developed in section 4.1.1. Without reabsorption, these should fit an exponential law that for a rod of length l gives

$$\text{Production} = A(1 - e^{-l/B}).$$

This model is derived from assuming the incoming protons have an equal probability of interacting in each length element of material and when they do interact, the same number of pions are produced at exactly that point along the rod. The number of protons remaining thus decays, as does the pion production at any position going along the rod, so the integrated production becomes an inverted exponential. A suitable choice for B would be the hadronic interaction length given in table 4.3 and this is used for the fit lines shown on the graph, though B can also be made a free parameter.

With B fixed at 11.18 cm, the best fit values of A for the two signs are $A_+ = 0.00883$ and $A_- = 0.00948$ useful pions per proton·GeV. The RMS deviations of data values from the fit are 3.87% for positives and 2.49% for negatives, expressed as percentages of the relevant A value, which sets the scale of the function. As the protons are at 10 GeV, there are roughly 0.076 useful pions per proton (for $l = 20$ cm) and thus the effective statistics are about 7600 pions from the original 10^5 protons. Therefore the statistical fit error would be expected to have size $1/\sqrt{7600} = 1.15\%$, which indicates this data has a small systematic deviation from the fit. This can be seen by carefully examining graph 4.13, which shows the fit lines overestimating production early on but underestimating it later.

Allowing the value of B to also become a parameter in the least squares fit improves the situation somewhat. The two signs now have parameters $A_+ = 0.00938\,\pi/(\text{p}\cdot\text{GeV})$, $B_+ = 13.83$ cm and $A_- = 0.00980$, $B_- = 12.70$ with RMS errors of 1.70% and 1.26% respectively. This is about as close a fit as the statistics will allow. It should be noted with caution that this does not mean the nuclear interaction length is 'wrong', just that there are other complexities in the pion-generation process taking place in a rod of finite thickness that are not modelled exactly by the simple exponential fit. For instance there is likely to be a forward emission angle so pions would be offset from their point of production when counted at the rod surface. For interpolation purposes, the fits with modified B are still useful.

4.3.1 Variation of Heat Deposition with Length

The simulations of different length targets in the previous section also generated output files of energy deposition and outgoing particle distributions. This means that the sectors of the inner pie chart in figure 3.3 can be reconstructed for each length and the dependency on target length examined. Again, these simulations do not include contributions from re-entrant pions that have spiralled back into the target due to the magnetic field, and in a similar way the criterion of '5 degrees angle of emission' is measured for protons at the rod surface assuming they continue in straight lines.

Figure 4.14 shows the variation of two energy measures with rod length. The amount of heat deposited in the target increases with length but levels off, which can be explained if only a few particles penetrate to large distances to deposit heat, with the majority of

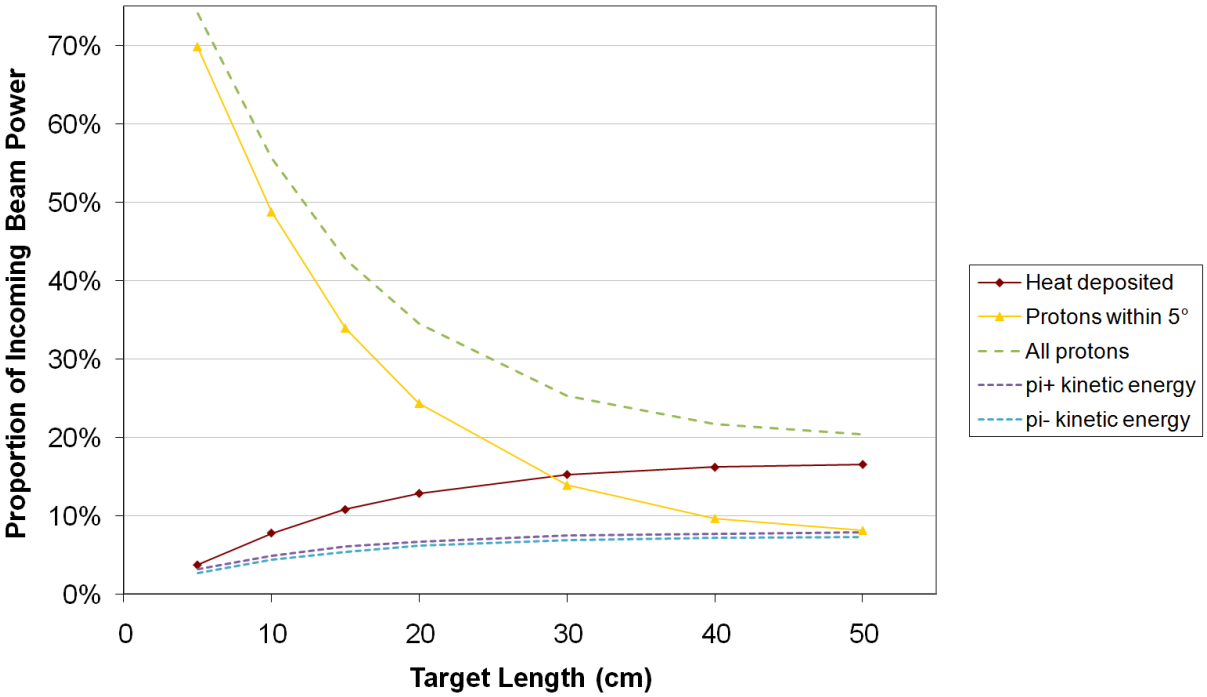


Figure 4.14: Variation of some of the heat loads shown in figure 3.3 with length of the tantalum rod. The 20 cm point corresponds to the reference simulation, with the points on either side retaining the baseline parameters from table 4.1 except for rod length.

energy having left the rod in the form of secondaries (corresponding to the inner blue segment in figure 3.3). It also looks like it follows an inverted exponential curve as the production figures did in the last section. Using the 11.18 cm nuclear interaction length to set the exponential scale again, the best fit is achieved for

$$H = 0.1575(1 - e^{-l/11.18}),$$

where l is the target length in centimetres and H is the heating in the target as a proportion of incident beam power. The RMS error in this fit is 7.28% so it is not particularly good. Allowing the length scale to vary gives $H = 0.1787(1 - e^{-l/16.75})$ with an error of 2.52%, which is better but the increase is still not fitted by a simple decaying exponential. The reason for this can be found in figure 4.12: if one of the central plots (there is no plot corresponding exactly to the 10 GeV used here) is truncated at a certain length, the heating in a rod of that shorter length will be the integral of the heat deposition over that smaller volume. The heating density takes a few centimetres to reach its peak, unlike an inverted exponential that has the highest rate of change at the beginning. So there may be no simple model for the heat load as a function of length here.

The other line in figure 4.14 counts heat from protons at small angles that have not undergone kinematically severe interactions so emerge either from the end or glancingly out of the side of the rod having only been slightly scattered. For short rods there will be a large contribution from protons that have gone from end to end and these are the 'uninteracted' protons used in the exponential model of yield in the last section. The decrease of this heat measure looks superficially like exponential decay but does not reduce fast enough at large rod lengths because of an accumulated population of protons that

Table 4.5: Heat deposition in the target and power in the forward proton beam dump for various lengths of tantalum rod. Kilowatt figures are scaled for a 4 MW proton source.

Target length (cm)	Heat deposited	kW	Protons within 5°	kW
5	3.7%	147.9	69.8%	2792.9
10	7.8%	311.4	48.8%	1950.1
15	10.8%	431.6	33.9%	1357.3
20	12.9%	514.4	24.3%	974.0
30	15.3%	611.1	13.9%	556.0
40	16.2%	648.1	9.7%	386.6
50	16.6%	662.8	8.1%	325.3

have scattered out of the sides of the rod and thus are no longer subject to progressive absorption by the rod material. A fit will not be attempted for this reason, but the values of this and the other heat measure are tabulated in table 4.5.

Both of these heat factors are of concern to the target station design: the heating for the integrity of the rod itself and the forward protons for their beam dump (see section 3.3 for more detailed discussion). The amount of heat in the forward protons varies quite steeply with increased rod length, meaning that if the forward beam dump becomes difficult, increasing the target length could reduce its heat load by a factor of 2–3×. This would not impact pion yield as steeply, though care must be taken with reabsorption effects not included here.

4.4 Rod Radius

The radius of the cylindrical target could be other than the 1 cm used so far. The advantage of using a wider target is that if the incoming proton beam size expands proportionately, the energy deposition density decreases by the square of the radius, potentially making thermal shock issues much more manageable. The disadvantage is that the thickened target can absorb more of the pions generated in its interior before they exit. To take advantage of the former point, the simulations in this section have kept the proton beam radius equal to rod radius as it varies, which also avoids excessive yield loss from absorption because protons will be producing pions all the way to the surface.

Output particle sets were obtained from MARS for both low (2.2 GeV) and high (30 GeV) proton beam energies, for both pion signs, with all other parameters as in table 4.1. The useful pion yield was calculated by the formula used throughout this chapter and the results plotted in figure 4.15. The different energies have different overall levels of yield, with a charge asymmetry in favour of positive pions for the 2.2 GeV case since at low energies the charge excess from positive protons is still visible above the pion charge-pair production. For radii of 1 cm and above, the graphs show a steady reduction as the radius increases (barring some statistical noise), which can be explained by absorption of pions by tantalum nuclei in the target. A more surprising feature is that the yield also starts

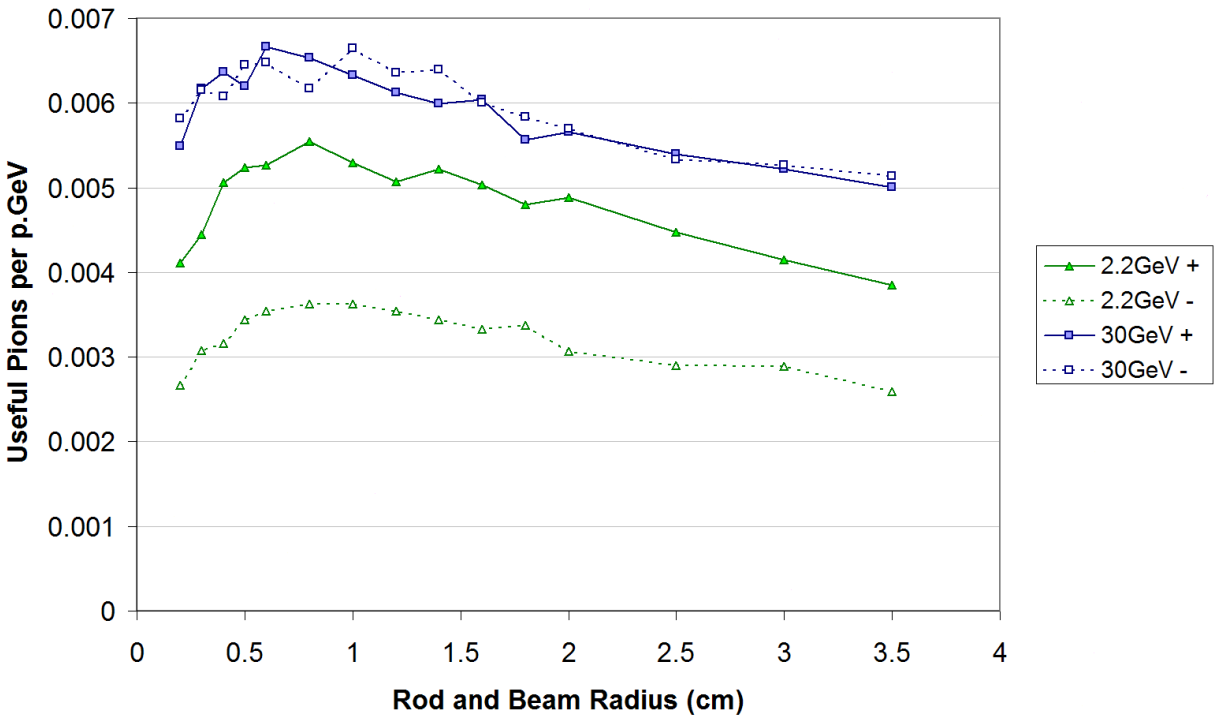


Figure 4.15: Effect of varying the radius of the cylindrical tantalum target on useful pion yield. Graphs are plotted for two different incoming proton energies and both pion signs.

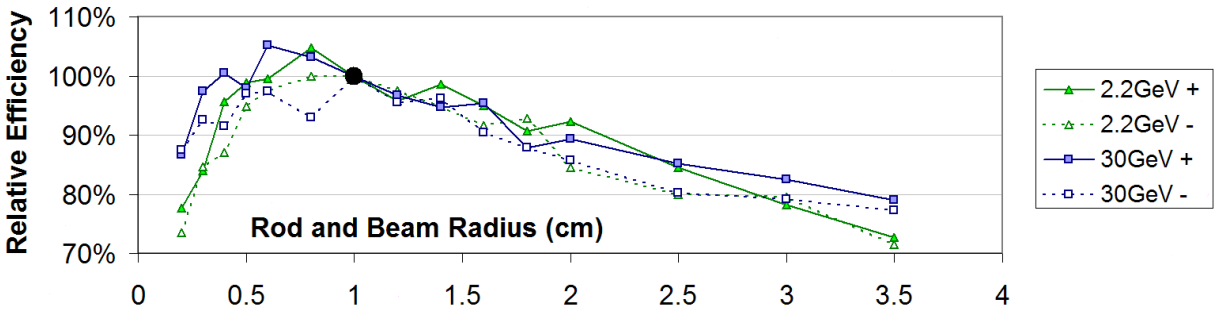


Figure 4.16: Useful pion yield variation with radius as a proportion of the $r = 1\,\text{cm}$ case.

to fall off for very thin rods. This is because tantalum and similar elements are high-Z targets and charged particles passing through a medium experience multiple Coulomb scattering with cross section proportional to Z^2 for the material. This cross section is also multiplied by the material density when calculating the number of interactions, so since tantalum is both high-Z ($Z = 73$) and dense ($\rho = 16.7\,\text{g/cm}^3$), it scatters the incoming protons quite strongly. For thin targets, this scattering is enough to knock a significant number of protons out of the target material by the time they reach the end 20 cm away, which results in a decrease in pion production.

Figure 4.16 isolates the radius dependence in the four scenarios by taking the ratio with the 1 cm situation in each case. It is difficult to optimise further than the 1 cm radius case; although all points between 0.5 and 1 cm radius have comparable yields, heat deposition density scaling as the inverse square of radius would be up to $4\times$ as large.

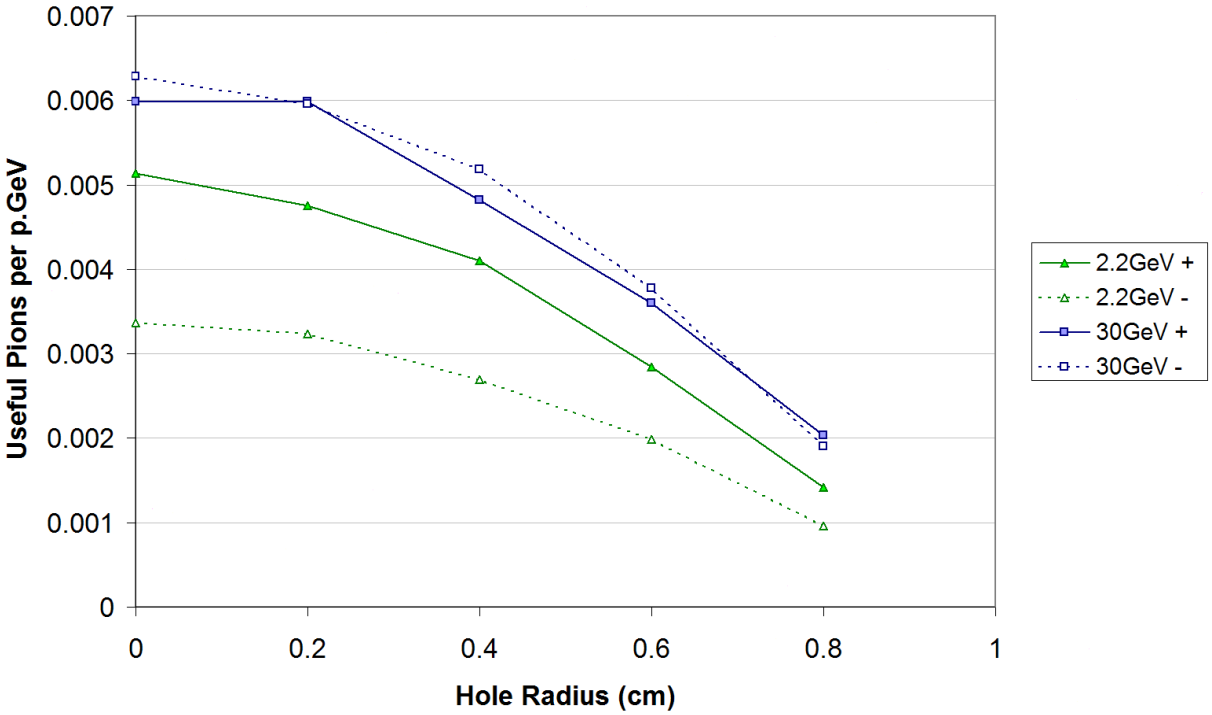


Figure 4.17: Useful pion yield as an increasingly-large hole is bored along the axis of the target, simulated with a uniform density proton beam (protons going through the hole are lost).

4.4.1 Rod with a Hole

During the investigation of the target yields, it was suggested that boring a hole along the axis of the rod might reduce on average the path length in material for emitted pions and thus decrease reaborption. This would waste protons that go through the hole, but it was suggested that a hollow phase space distribution could be created for the protons that minimised the number in this area. Therefore some MARS simulations were run for targets with various sized holes.

The standard target simulation (table 4.1) was modified in one important way aside from drilling a hole: the other simulations so far have used a parabolic proton distribution with a varying density, higher at the centre and decreasing to zero at the edges, used to represent a standard proton beam from a synchrotron. However, the beam for hollow targets will be optimised for the purpose, so this is no longer relevant. The simplest way to investigate any improvement from the hole without having to worry about proton accelerator issues is to use a uniform distribution, so the proton flux is the same everywhere on the front face of the target, including the hole where a known proportion $(r_{\text{hole}}/r)^2$ are wasted and miss the target completely. Thus in figure 4.17 the results for $0 \leq r_{\text{hole}} \leq 0.8\,\text{cm}$ are plotted for both signs and the low and high proton energies used in the last section. This figure on its own is not particularly useful as the curves look similar to scaled graphs of the non-wasted proton beam proportion $1 - (r_{\text{hole}}/r)^2$ and the question is whether the yield per *impacting* proton has been improved.

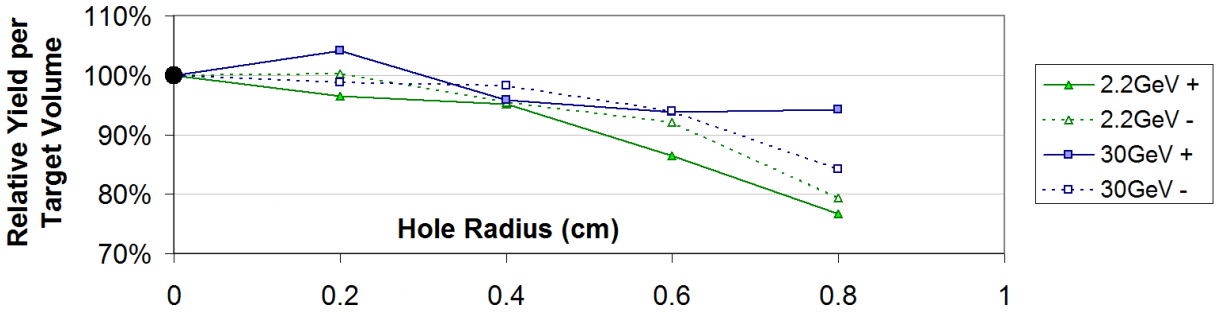


Figure 4.18: Useful pions per proton hitting the target rather than the hole, displayed as a ratios to the yield efficiency of the whole-target case.

Figure 4.18 shows the same pion yields divided by the proportion of protons hitting the target for each hole radius and normalised to the case with $r_{\text{hole}} = 0\,\text{cm}$ (a whole target). It shows that the efficiency of the target is, if anything, made worse by the incorporation of a hole, so this possibility for the target design may be excluded. The worsening of the efficiency for the targets with the largest holes could be explained by the out-scattering of protons from the remaining target, the same effect that made targets with very small radii inefficient.

4.5 Proton Beam Distribution

The previous subsection (4.4.1) alluded to the fact that the transverse distribution of the incoming proton beam was not in fact fixed but might be tailored to the target, to the extent possible, if this would increase the useful pion yield. This beam manipulation would probably be done with optical elements in the transfer line taking the extracted beam from the final proton ring to the target. Quadrupole magnets control the beam focussing in linear terms and can change the proton spot size in both axes (investigated in section 4.5.1) and higher order magnets such as octupoles can apply nonlinear distortions to the beam distribution, creating for example a beam with more particles towards the edges than in the circular parabolic case. A uniform beam is one such example and is compared in section 4.5.2.

4.5.1 Proton Beam Radius

Figure 4.19 shows the variation in useful pion yield when the target radius is kept at 1 cm but the proton beam radius is scaled to 3 mm or 1 mm as well as the full-size case. There is very little effect on either pion sign, so statistical error bars of one sigma from the simulation were plotted. The most significant difference is between the 3 mm and 1 cm beam radii for the negative pions and here the difference is 2.9σ in favour of using a full-sized beam. This would be consistent with the theory that pions coming from the smaller beam would always have to travel through at least 7 mm of material (6% of a hadronic interaction length for tantalum) before leaving the rod, so may be reabsorbed more, but the difference is too small to make a conclusion.

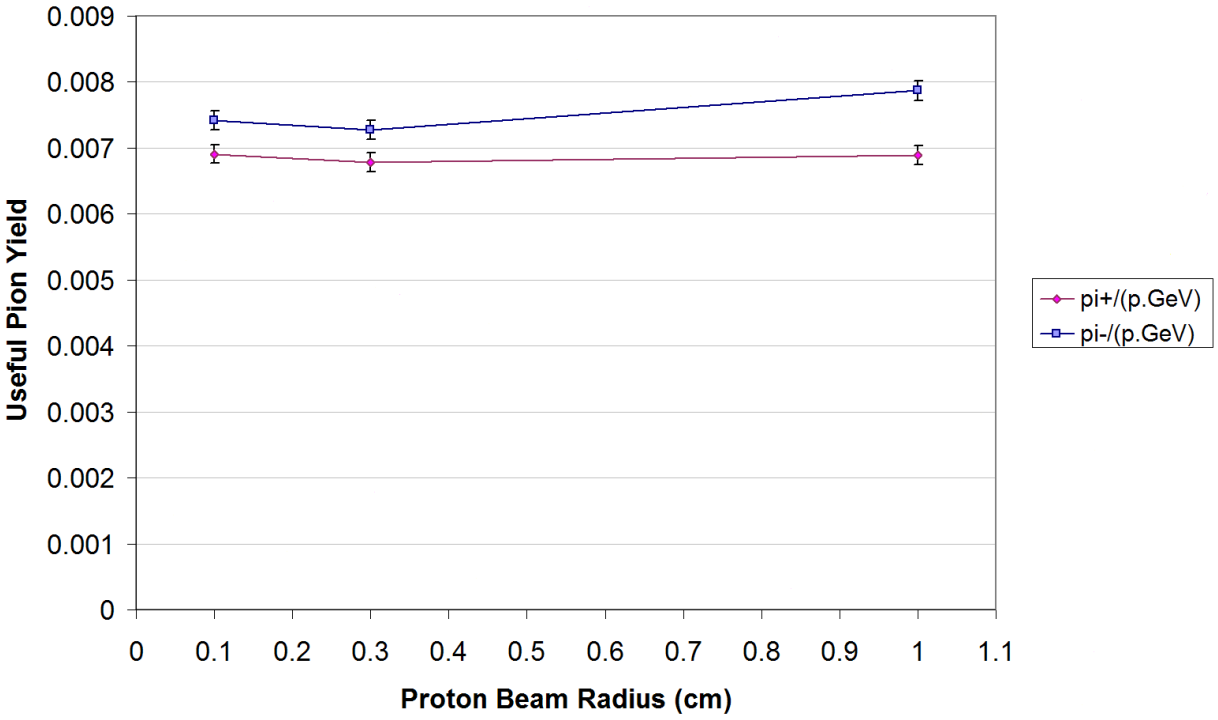


Figure 4.19: Effect (or rather, non-effect) on yield of having the incoming proton beam smaller than the 1 cm radius of a tungsten target cylinder. With 1σ statistical error bars.

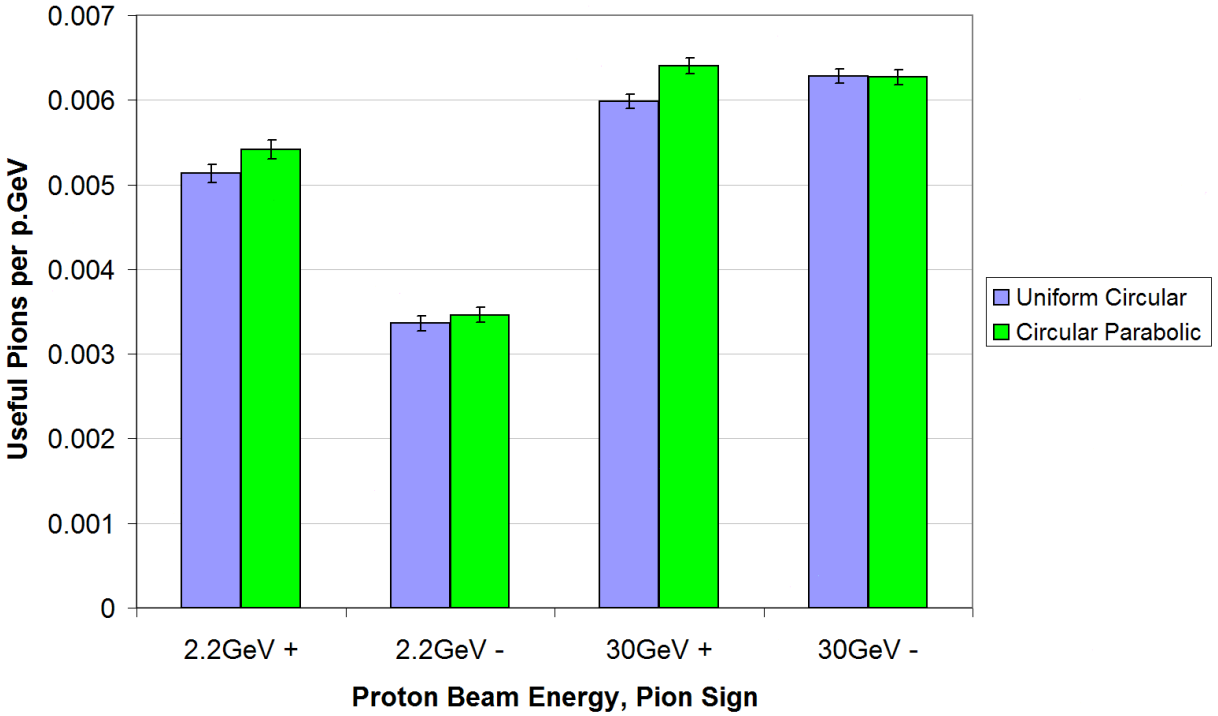


Figure 4.20: Comparison of pion yields generated by uniform versus parabolic transverse proton beam distributions hitting the tantalum target, at two energies. 1σ statistical error bars are shown.

4.5.2 Circular Uniform Beam vs. Circular Parabolic

Changing the proton beam distribution from the circular parabolic with density $\rho(r) \propto 1 - (r/R)^2$ to a uniform density ($\rho(r) \propto 1$) beam will mean more particles near the edge of the rod, though the difference will not be as dramatic as between the 3 mm and 1 cm radii parabolic beams in the last section. Figure 4.20 compares the yields for the two distributions, for both signs of pion and two proton energies, with one sigma statistical error bars shown on each value. The difference between the bars normally falls within the error bars, the most deviant case being the 30 GeV positive pions where the circular parabolic distribution yield is larger than the uniform by 3.4σ. As in the previous subsection, the conclusion is that the difference in yield from changing the proton distribution is too small to be significant in these simulations and is certainly less than 10%. Quite possibly this is too small to be significant at all in the design.

4.6 Downstream Effects

All the other sections in this chapter consider the optimisation of the target decoupled from the rest of the system, based on local variables such as the momenta of the pions at the rod surface. This section looks at two effects where there is nontrivial coupling between the proton beam energy on the target and the downstream front-end. Section 4.6.1 examines the contribution of the decay chain of kaons, only produced for higher proton energies and mostly decaying outside the target. Section 4.6.2 examines how the location of beam losses changes in the front end as the proton energy is varied.

4.6.1 Pion Production from Kaons

The charged kaons K^+ and K^- can be produced in a target bombarded by a high-energy proton beam and have a mean lifetime of 12.4 ns (compared to 26.0 ns for the charged pion). Since they are produced at relativistic speeds, they will travel several metres before decaying and so will not decay within the MARS target simulation.

Table 4.6: Principal decay channels of the kaon [27], together with the number of pion and muon products and thus the expected number of eventual muons from each channel.

K^+ decay to	Probability	$\#\mu$	$\#\pi$	Mean eventual muons
$\mu^+\nu_\mu$	63.39±0.18%	1	0	0.6339±0.0018
$\pi^0 e^+\nu_e$	4.93±0.07%	0	0	0
$\pi^0\mu^+\nu_\mu$	3.30±0.06%	1	0	0.0330±0.0006
$\pi^+\pi^0$	21.03±0.13%	0	1	0.2103±0.0013
$\pi^+\pi^0\pi^0$	1.757±0.023%	0	1	0.0176±0.0002
$\pi^+\pi^+\pi^-$	5.59±0.05%	0	3	0.1677±0.0015
Total	(100%)			1.0624±0.0028

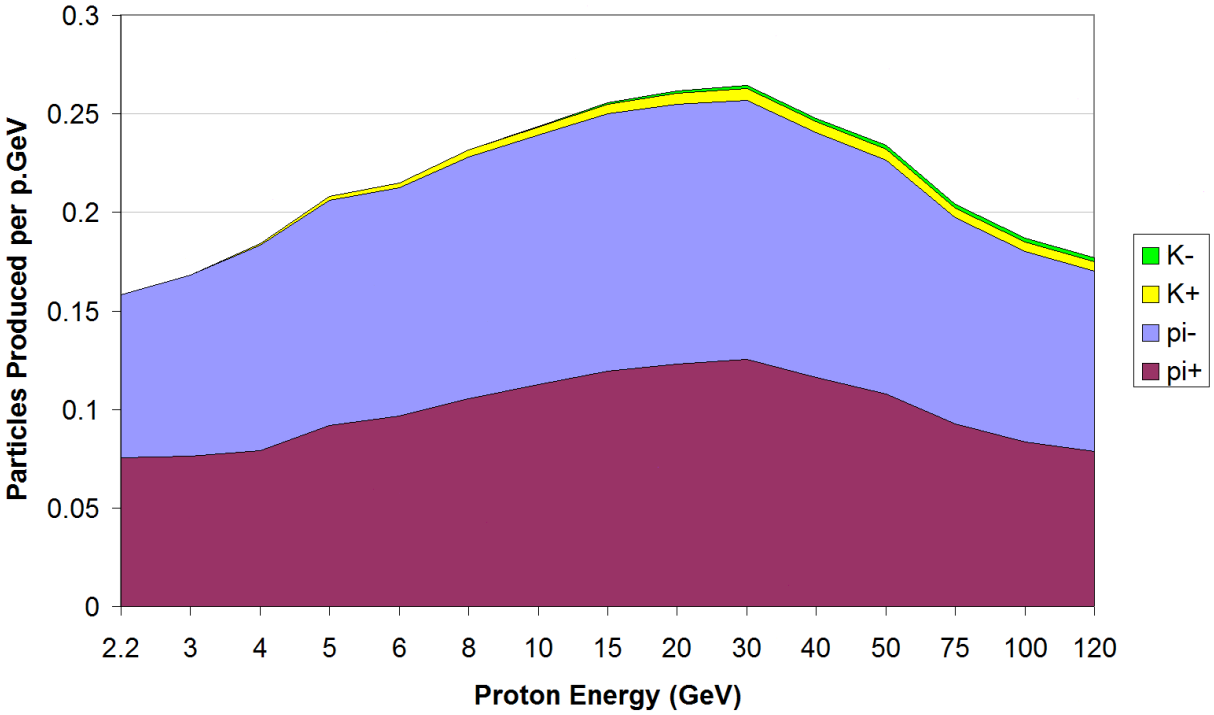


Figure 4.21: Comparison of the number of pions and kaons produced by the tantalum rod target at various proton beam energies.

The major decay channels of K^+ and their respective probabilities are shown in table 4.6. Many of the decay channels produce charged pions or muons, which will contribute to the neutrino factory yield. The charged pion itself has a 99.98770±0.00004% probability of decaying into a muon [27], so effectively the numbers of the two are added together, minus a tiny factor. This arithmetic reveals that on average, a charged kaon will ultimately produce about 1.06 muons and is therefore roughly equivalent to having a pion in the same location.

Figure 4.21 shows the total number of charged kaons as well as pions produced by the target. Below 4 GeV, no kaons are produced at all, whereas at the highest energies the positive kaons are adding another 6% to the total particle count and the more rarely produced negative kaons contribute 3%. There is a slight mixing between the signs introduced by the $K^+ \to \pi^+\pi^+\pi^-$ decay but this will only account for 5.6% of the decays of the few percent of kaons, so is negligible. Thus kaons produce a small but systematic increase in yields for higher proton beam energies, particularly for positive particles.

4.6.2 Redistribution of Losses

While previous sections dealt with beam transmission, the pions and muons that are lost will produce radioactive hot spots in the muon front end due to beam loss. The dumping of protons was discussed in section 3.5 and there will be a large level of general activation around the immediate target area from neutrons and other particles, but losses from pions and muons present a different problem because they are the intended particle for the focussing channel and can thus propagate a long way downstream before being lost.

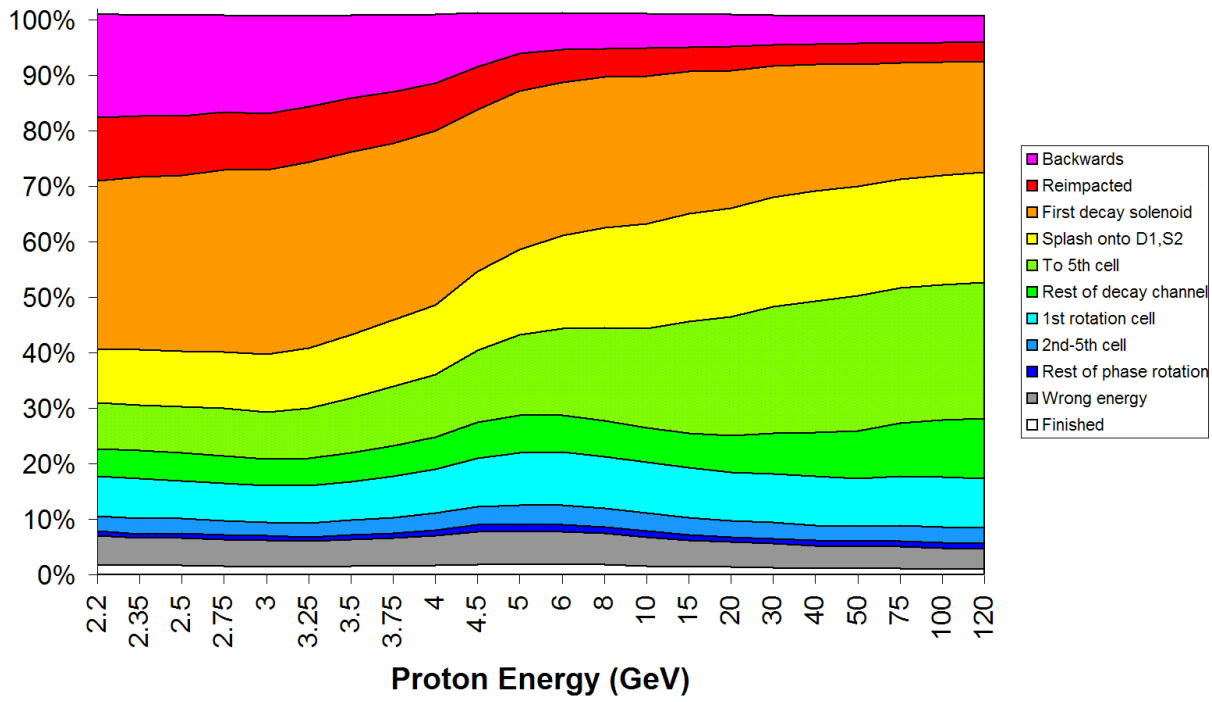


Figure 4.22: Losses of pions and muons redistribute themselves through the decay channel as the proton energy on the target is varied. Values are plotted as a fraction of the initial number of pions from the target; the total is slightly more than 100% due to decays such as $K^+ \to \pi^+\pi^+\pi^-$.

Therefore the Muon1 code was run on a particular pion decay channel and phase rotation system, specifically the optimal lattice from the `PhaseRotC` optimisation discussed in section 8.3, with a logging mechanism to count the weight of lost particles along the axis of the channel. Based on their location, the losses were put into categories and plotted as proportions of initial pions in figure 4.22, which also includes small strips at the bottom to represent on- and off-energy particles reaching the end of the beamline (the off-energy particles will be lost even further downstream). 'Backwards' losses are those from pions travelling in the wrong z direction, so they hit the beam plug before the `S1` 20 T solenoid (`D1` refers to the first drift space before `S2`, the lower field solenoid that follows). 'Reimpacted' losses are those that are reabsorbed in the tantalum cylinder target according to a simple absorption-length model, though [62] has a full study of this.

This exercise was repeated for all the proton energies for which MARS output particle sets were available (see [63] for the complete report). Kinematically it might be expected that the pion angular distribution will change as the forward proton momentum increases and this is consistent with the trends appearing in the graph. At low proton energies more (up to 61%) of the pions are lost within the first solenoid (`S1`) and target area, whereas at high energies this decreases to 28% but the losses in the decay channel excluding `S1` have more than doubled from 23% to 55%. So the increased boost from the proton is producing pions at more forward angles that are lost further down the decay channel. Beyond the decay channel, the amount of remaining beam at each stage exhibits a humped shape with an optimum in the mid-energies and this coupling to proton energy no longer changes, indicating that the beam has been cut down to an accepted region where only

the magnitude of yield within the region is important and not its precise distribution.

This suggests the beamline beyond the end of the decay channel is decoupled from the choice of proton beam energy other than by the scaling in pion yield, which means optimisation of those later sections can be done independently of the proton driver choice.

4.7 Summary of Optimal Target Parameters

The values in table 4.7 summarise many considerations about the performance of the target. The proton energy, target radius and lack of a hole are chosen to optimise pion yield, with the important proton energy checked again with a different code in the next chapter. The proton beam radius and rod angle have a weak effect on yield (plots for rod angle are shown on the final page of [62]), so the beam radius is chosen to reduce thermal shock and the rod angle to mitigate the proton hot spot on the beam dump. The proton distribution is likely fixed at roughly circular parabolic by the proton accelerator itself but this was found not to be a disadvantage for pion yield.

There remain many feasible choices for the target material, as the lines in figure 4.9 achieve comparable peak performances for achievable proton energies (this excludes the 1 GeV carbon peak). This report assumes a high-Z material similar to tantalum or tungsten, while [61] examines the optimisation in the carbon target case and figures for carbon are also given. Section 4.3 shows the exponential levelling off of yield with increasing number of hadronic interaction lengths in the target but does not account for reabsorption of pions hitting the target for a second time, since this requires tracking of pions outside the target including a magnetic field. As the magnetic field will be varied in a later optimisation using Muon1, it is unknown at the point of generating the initial pion distribution. However, other studies including [61, 62] find that the peak yield is for practical purposes very close to the ∼1.8 interaction lengths given in table 4.3.

Table 4.7: Recommended neutrino factory target parameters, assuming a solid target.

Component	Parameter	Recommended value	Source
Proton beam	Energy (E_k)	5–15 GeV	§4.2, §5.1, [59]
	Radius	Equal to target radius	§4.5.1
	Distribution	Circular parabolic	§4.5.2
Target	Material	Tantalum used here (others possible)	§4.2
	Length	20 cm for Ta, ∼60 cm for carbon	Table 4.3, [61]
	Radius	1 cm for high-Z, 1.25 cm for carbon	§4.4, [61]
	Shape	Solid cylinder with no hole	§4.4.1
	Angle	14 mrad to accelerator axis	§3.5, [62]

Chapter 5

Benchmarking Pion Production

When predicting pion production figures in the regime of the neutrino factory, it is important to realise that before the HARP experiment [4], there was no systematic attempt to measure pion yields in this energy domain. Thus the MARS15 cross section models had to be fitted to what little data there was available, with some interpolation based on how proton cross sections behave (see the MARS15 manual [64] section on hadron production). GEANT4 was in a similar situation. Section 5.1 sees how close these different models are to each other and 5.2 attempts a comparison of MARS15 with recent HARP data.

5.1 Comparison with GEANT4

Without a code available that incorporates the HARP data, comparing with the results of another code will roughly gauge the uncertainties present in the current models. The GEANT4 code [3] is also commonly used by particle physicists to model the interaction of particles with matter, so it was used as a comparison point for MARS.

Results in this section are from a collaboration with K. Walaron who wrote a more comprehensive report about benchmarking GEANT4 with MARS15 on this problem [65]. The findings were also presented in a paper [59] and the author's talk at the NuFact'05 conference.

The standard target problem specified in table 4.1 was reposed as a GEANT4 input data set and run for the same range of energies as the MARS15 simulations in section 4.2. The generated pions were weighted by the process defined in section 4.1.1 and the

Table 5.1: The proton energy ranges over which pion yield is near-optimal, as predicted by MARS15 and GEANT4 for each sign of pion.

Code	Optimal π^+	Optimal π^-
MARS15	5–30 GeV	5–10 GeV
GEANT4	4–10 GeV	8–10 GeV

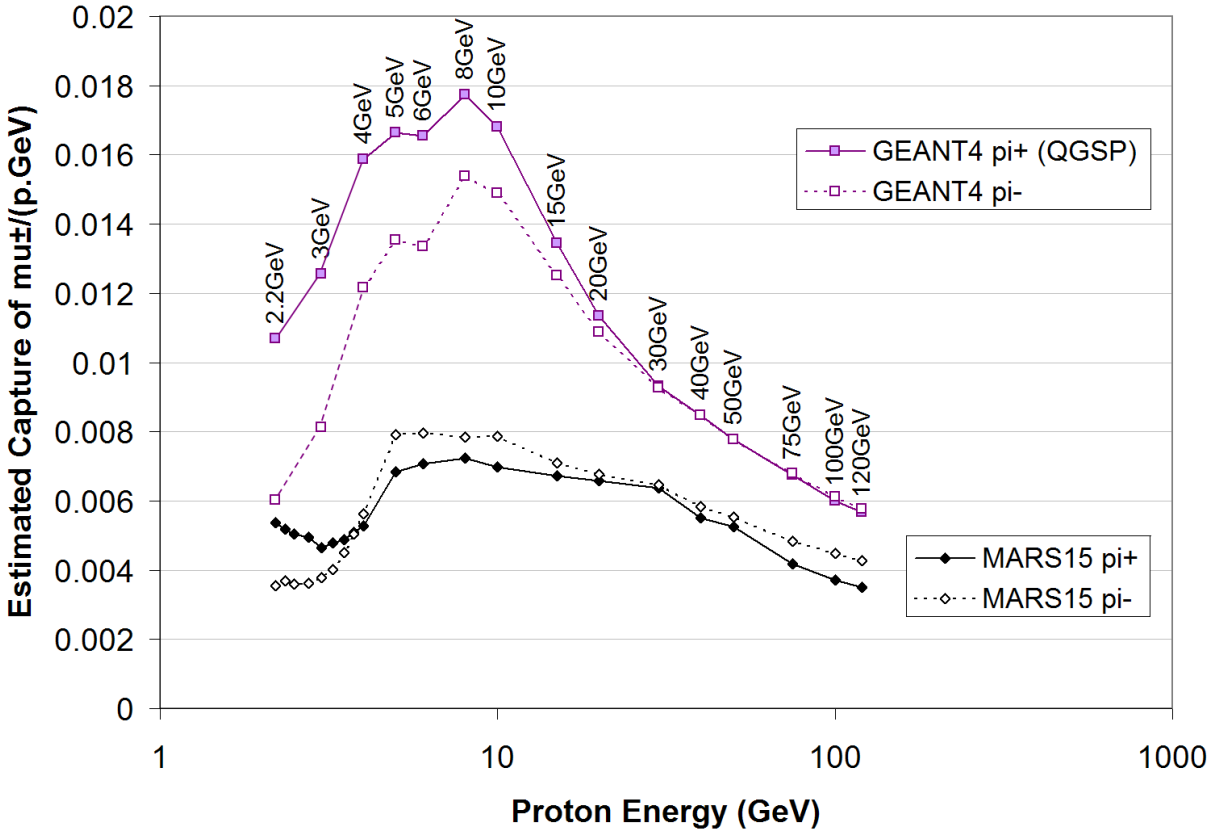


Figure 5.1: The results from two different Monte Carlo codes in predicting the yield of useful pions from the target as a function of incoming proton energy.

resulting figure of 'useful pions' plotted on the same graph in figure 5.1. Although the form of the graph with an optimum at intermediate proton energies is similar between the two codes, there is a factor of 1.35–3.05 (average 1.85) difference in absolute yield between the two, meaning that the predictive accuracy of these codes is still at a fairly rudimentary level in this situation. Both codes were run using default hadronic model settings, the effect of changing these in GEANT4 is explored in the next section.

Fortunately, the conclusions relevant to the neutrino factory about the optimal energy for the proton driver are not contradictory between the two codes and are summarised in table 5.1. The energy range of 8–10 GeV is safely within the optimal range for both signs, according to both of the codes used here. This is as good a recommendation as possible using tools at this stage of development and was the one given to the International Scoping Study [24] in 2006 when defining their baseline parameter set for the neutrino factory.

5.1.1 Energy Cutoffs in Hadronic Models

GEANT4 provides several modes called 'use cases' for selecting the models used at different energies for hadronic particle production models. Multiple models may be necessary because cross sections involving protons cannot be calculated from first principles and must use either a simplified physical model or a fit from measured data, both of which may only be applicable in a certain energy range.

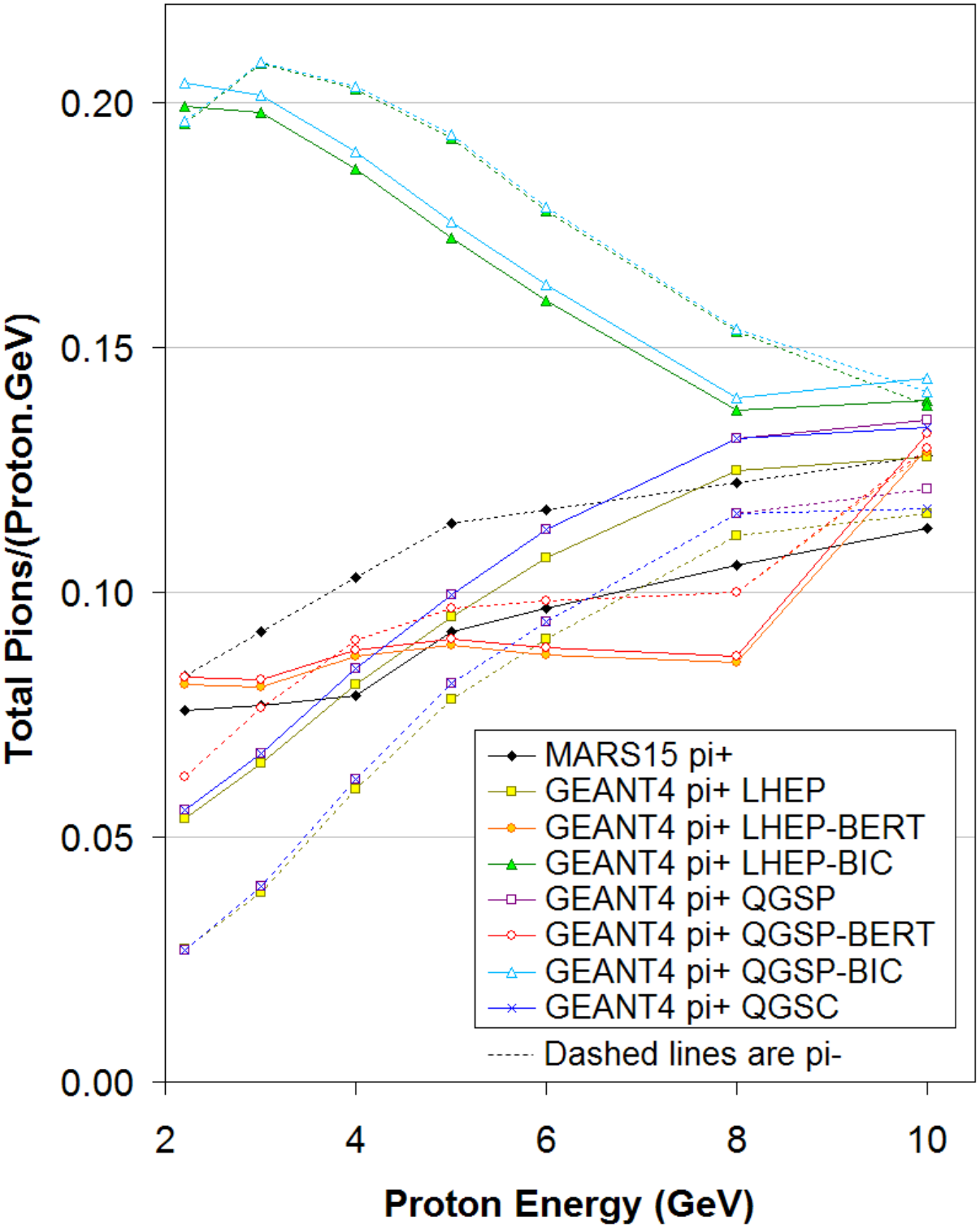


Figure 5.2: Comparison of all the available hadronic modes of GEANT4 together with MARS15 results for the total pion yield from the standard tantalum target.

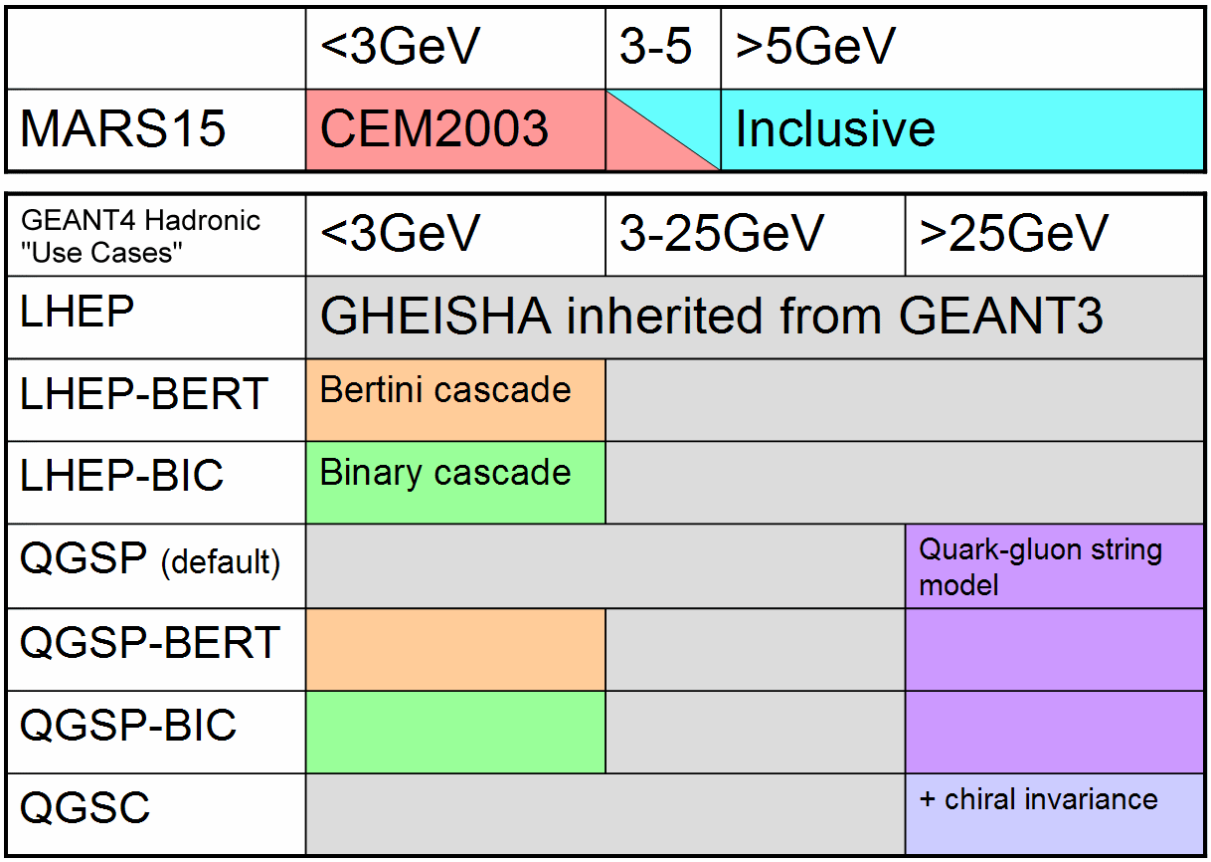


Figure 5.3: Diagram of energy cutoffs between hadronic models used in different GEANT4 'use cases'. The MARS15 model is linearly mixed from the cascade-exciton model (CEM2003) at 3 GeV to the MARS inclusive hadron production model at 5 GeV.

To see the effect of use cases on the results, all the possible choices were run for a range of energies and the total (uncut) pion yield plotted in each case, resulting in figure 5.2. It shows that the binary cascade (BIC) models diverge from the rest of the predictions, including MARS, as the proton energy decreases. The Bertini cascade (BERT) models also have a different shaped energy dependence but it is not as severe. These results are provided for information only here, as a detailed dissection of the GEANT4 code's hadronic sector is beyond the scope of this thesis.

Figure 5.3 shows how MARS15 and the various GEANT4 modes use different models at different energies; the 25 GeV cutoff in GEANT4 may not be as sharp as depicted because some difference at 10 GeV was found when changing the high-energy component in figure 5.2. The ramp between 3 and 5 GeV in MARS is also noteworthy when interpreting the yield graphs earlier in this chapter, as 'corners' in the plots at those values may be an artefact of the random mixing which the code makes between the two models, which uses proportions of each model that vary linearly in energy from zero to 100% over that energy range. Extrapolating the pure model lines from ≤ 3 GeV and ≥ 5 GeV in an example such as figure 4.7 would probably form two lines that do not meet, showing again that these models are not much better than a factor of 2 accurate in prediction of absolute pion yields.

5.2 Comparison with Early HARP Data

The HARP experiment [4] was designed to measure the production cross sections of π^+ and π^- from high-energy protons on various solid targets of different atomic number (Z). The applications for this data include studies of superbeams, neutrino factories and muon colliders, all of which generate a tertiary beam (of muons or neutrinos) from the pions that are produced in large numbers (compared to other possible parent particles) when hadrons hit a solid target. Importantly, the experiment was also designed to measure the double-differential cross section $\frac{\mathrm{d}^2\sigma}{\mathrm{d}p\mathrm{d}\theta}$, which may be integrated across an area in (p, θ) space to retrieve a partial cross section σ for the range of interest. This level of detail enables a full comparison with the output of Monte Carlo code such as MARS15.

Table 5.2: The four proton momenta used in the HARP experiment, converted to kinetic energies in GeV used for protons throughout the rest of this chapter.

p [GeV/c]	E_k [GeV]
3	2.205
5	4.149
8	7.117
12	11.098

HARP used protons of four different energies extracted from the CERN PS (Proton Synchrotron). The energies used are actually specified exactly in terms of the proton momentum in GeV/c but the GeV kinetic energy equivalents are also listed in table 5.2. This range is a subset at the low end of the energy range studied in the rest of this chapter. HARP had two detectors: a forward spectrometer that detects pions of low angle and relatively high energies and a Large Angle Spectrometer (LAS) that detects generally slower pions at higher angles. The locations of the detectors' apertures in (p_L, p_T) momentum space is plotted in figure 5.4, which also highlights the region of useful pion yield for the neutrino factory front end. Results from the forward spectrometer on an aluminium target are reported in [66] but this momentum range is of more interest for superbeams and misses the neutrino factory zone completely. The LAS overlaps with about half of the useful yield region, missing the areas below its lower θ limit of 0.35 radians. HARP LAS results have been reported for a tantalum target in [67], which is relevant for the solid neutrino factory target (and of a similar Z to a mercury jet) and forms the basis of the comparison with MARS15 in this section.

The HARP tantalum target is a cylinder and was replicated in a modified MARS simulation using the parameters in table 5.3. The short length of 5.6 mm is to produce approximately 5% of a hadronic interaction length in tantalum so that it behaves approximately as a 'thin target' (or foil) allowing extraction of the cross section without too much interference from cumulative absorption, energy loss or scattering of protons or pions. The dimensions were derived from measurements of the target recorded in the data sheet at http://www.hep.shef.ac.uk/research/harp/ta5.html.

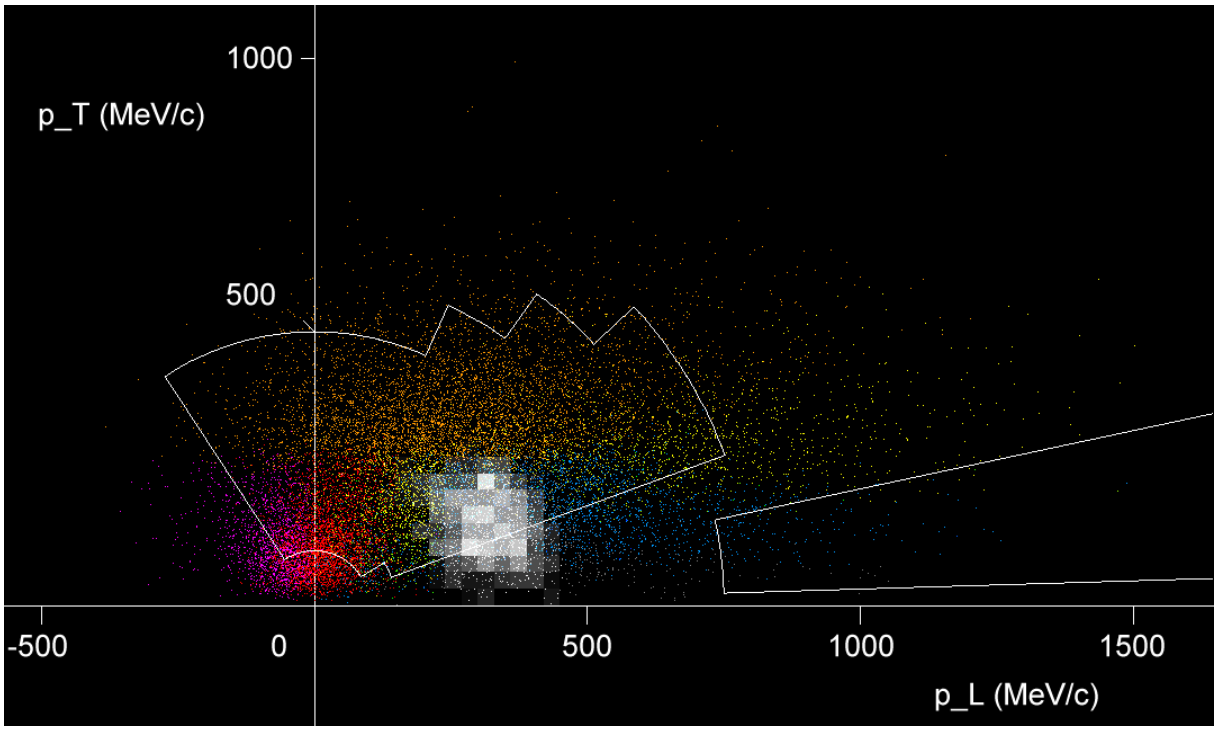


Figure 5.4: The pion momentum space plot of figure 4.2 with the acceptance ranges of the two HARP detectors superimposed. The Large Angle Spectrometer (LAS) is the central area and the forward spectrometer extends beyond the right edge of the plot to $|\mathbf{p}| = 6500\,\mathrm{MeV/c}$. The density of successful pions (white) leaving the channel is highlighted as grey squares, note this is different from the *probability* of transmission plotted in figure 4.5.

Table 5.3: Parameters of the MARS15 simulation of the (small) tantalum cylinder used in the HARP 5% interaction length study.

Code Used	MARS15
Target Material	Tantalum
Target Shape	Cylindrical
Target Dimensions	5.6 mm length × 15.1 mm radius
Target Mass	66.195 g ($\rho_{\mathrm{Ta}}^{harp} = 16.59\,\mathrm{g/cm^3}$ derived)
Protons Per Monte Carlo Run	10^6
Proton Beam Profile	Parallel pencil beam at centre of target end-face, 1% RMS Gaussian momentum spread.

The full range of results across all pion momentum and angular bins, for both pion signs and all four incoming proton energies are plotted in figures 5.5 and 5.6. Converting the MARS output, which is in terms of weighted numbers of pions in various bins for a known number of protons, into a cross section requires knowledge of the number density ρ_N of tantalum nuclei in the simulated target. The density of tantalum is $\rho_{\mathrm{Ta}}^{mars} = 16.68\,\mathrm{g/cm^3}$ and its atomic mass is $A = 180.9$, so

$$\rho_N = N_A \frac{\rho_{\mathrm{Ta}}^{mars}}{A} = 5.551 \times 10^{22}\,\mathrm{cm^{-3}},$$

where N_A is Avogadro's number. Next, given the path length l in centimetres and the observation that a total weight w of pions were produced from N incoming protons in the simulation, the cross section may be estimated as

$$\sigma\,[\mathrm{barns}] = \frac{w}{Nl\rho_N 10^{-24}}.$$

This formula is also correct if w and σ are only counted for pions within a particular zone of momentum space. If the zone is a small region such as one of the HARP momentum bins, the double differential cross section in that bin can be estimated by dividing σ by the area of that bin in (p, θ) space.

Some more consolidated (and easier to interpret!) graphs are presented in figure 5.7, which combines the forward and backward parts of the HARP detector, and in figure 5.8, which looks purely at the ratio between the signs in only the forward hemisphere. Figure 5.7 shows that MARS15 is underestimating the production of pions in backward directions by a factor of 2 or more, whereas the forward direction has a rather better fit, though MARS produces a peculiar kink around $p = 200\,\mathrm{MeV/c}$ in the $3\,\mathrm{GeV/c}$ proton case. Figure 5.8 shows that MARS is producing proportionately more negative pions than the HARP data suggests is the case. The HARP data shows an excess of positive pions, increasing at low energies, while the MARS15 data is more equal but also tends to include more positives as the proton energy decreases.

Finally, the overall pion cross section may be plotted as a function of proton kinetic energy, to provide a link with the results of section 4.2. This is not quite a 'pion yield' because the target here is much shorter than any neutrino factory would use, but the cross section in the relevant momentum zone may be compared across proton energies as a guide. The HARP report [67] also performed this analysis and chose the zone with $\theta < 0.95\,\mathrm{rad}$ and $250 \leq |\mathbf{p}| < 500\,\mathrm{MeV/c}$ to represent the neutrino factory, which encompasses the region on figure 5.4 where the acceptance overlaps with the detector. This is not totally correct because the yield from the most forward pions is missing, but being aware of this caveat, the comparison graph may still be plotted, giving figure 5.9 where MARS15 has used the same cut as HARP. The lines at the top of the graph are total yields in the forward hemisphere of LAS (excluding the $p \geq 700\,\mathrm{MeV/c}$ bin as in the HARP report).

The author's interpretation of figure 5.9 is as follows. In total, MARS produces too many π^- (by up to 24%) and too few π^+ (by up to 16%) at all energies, but the agreement at the $\sim$20% level is encouraging given the large discrepancy with GEANT4 discussed in section 5.1. When the cross sections are cut nearer to the domain of interest, the positive pion value never deviates by more than 13% between MARS and measurements, whereas the situation for π^- becomes worse, with MARS always overestimating the yield, sometimes by as much as 44% in the case of negative pions at 4 GeV.

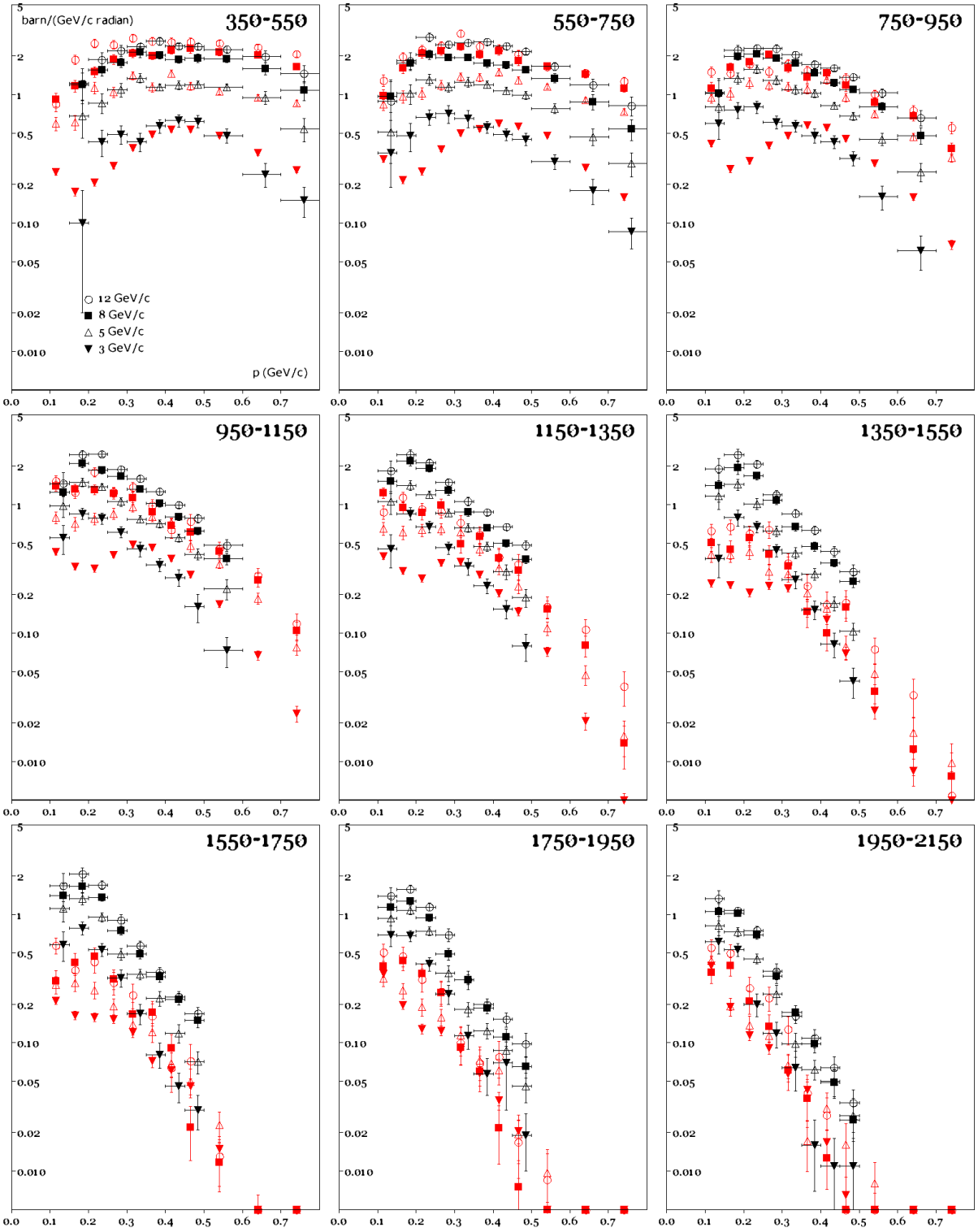


Figure 5.5: π^+ cross sections in all bins of the HARP Large Angle Spectrometer (black) together with MARS15 simulated values (red). Compare with figure 23 of the HARP report [67]. Numbers in the top-right of graphs are the angular bins in milliradians. MARS error bars are 1σ statistical; HARP error bars are 1σ but also include systematics.

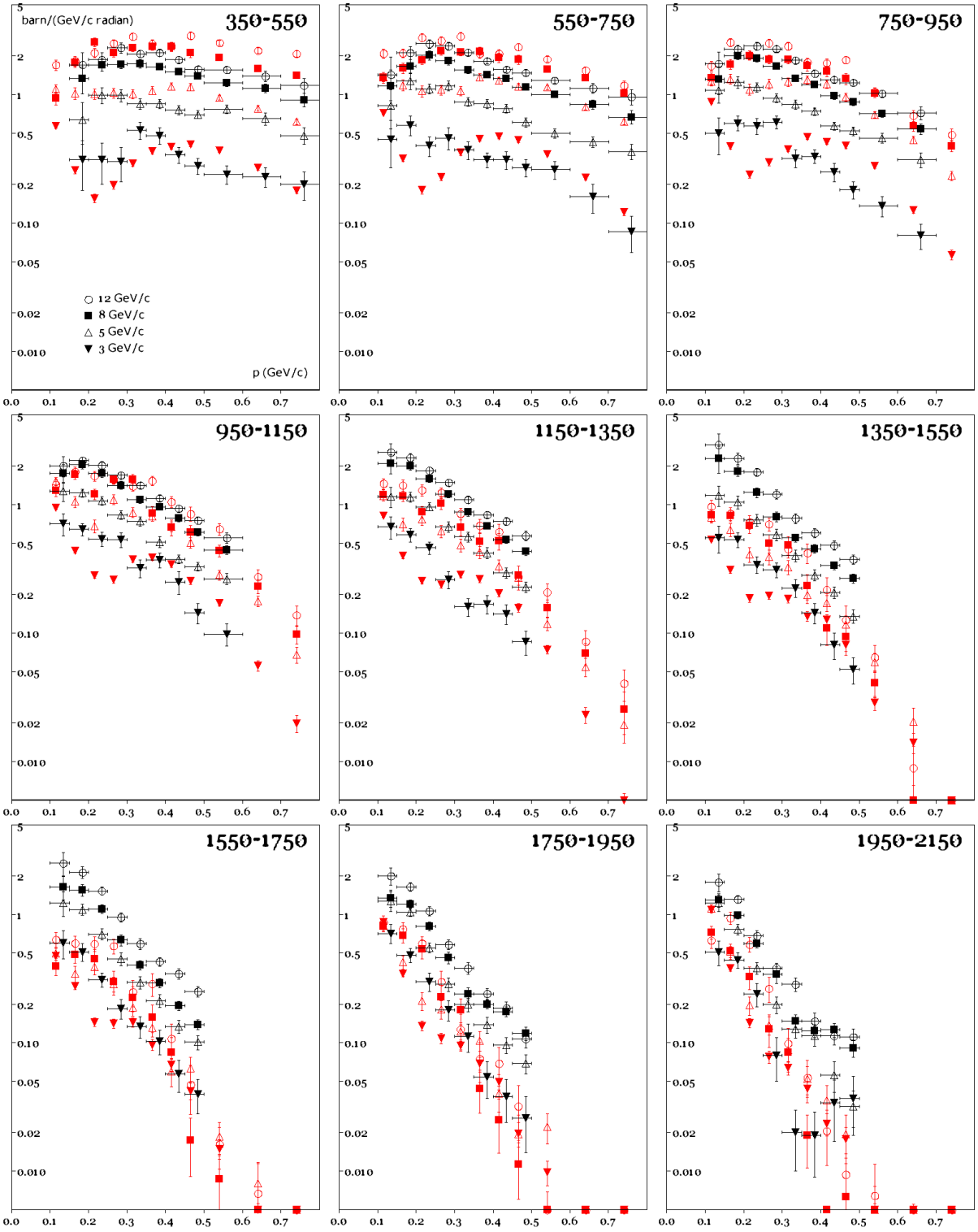


Figure 5.6: π^- cross sections in all bins of the HARP Large Angle Spectrometer (black) together with MARS15 simulated values (red). Compare with figure 24 of the HARP report [67]. MARS error bars are 1σ statistical; HARP error bars are 1σ but also include systematics.

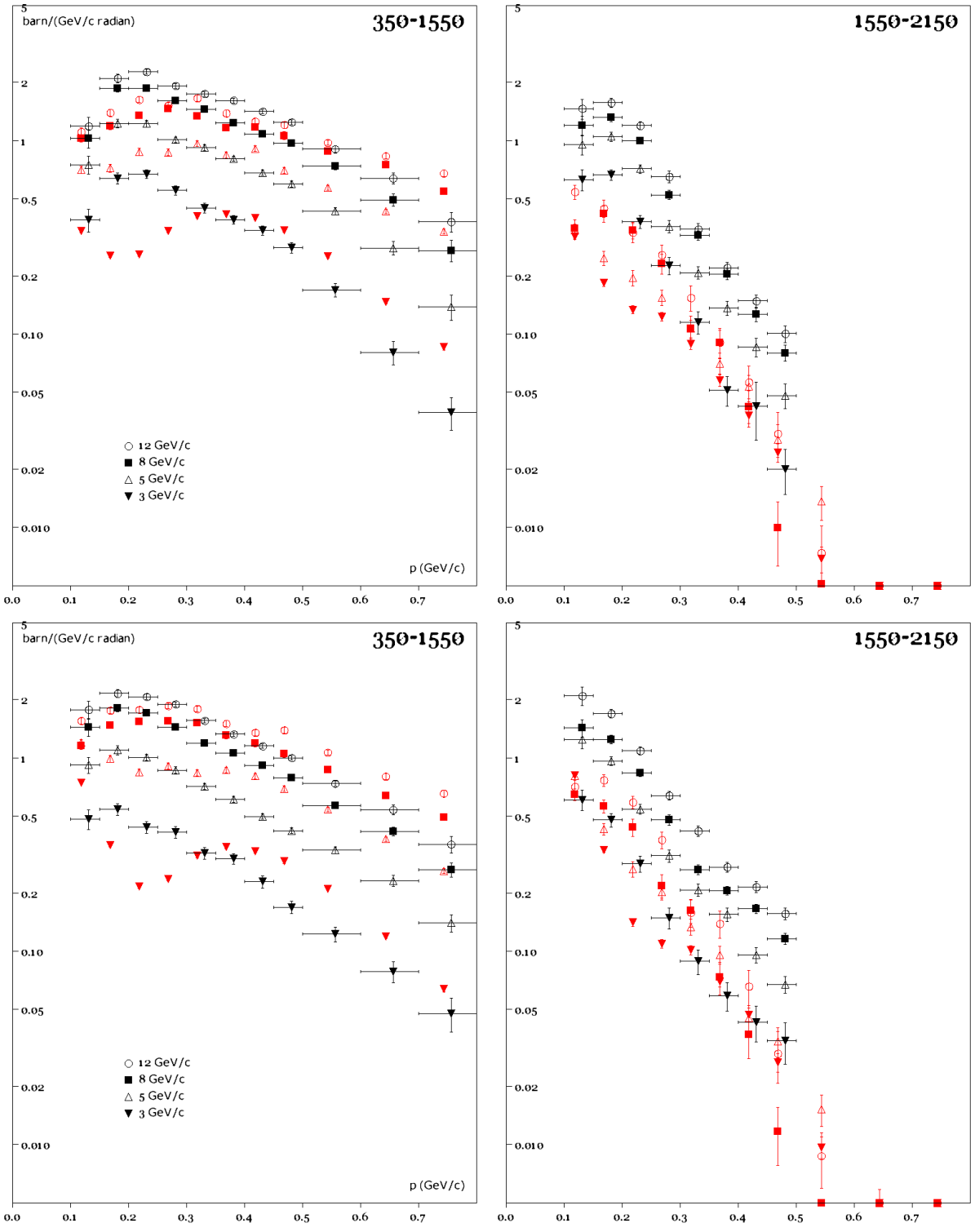


Figure 5.7: HARP measurements of forward (left) and backward (right) pion production cross sections for π^+ (top) and π^- (bottom), comparable with figure 26 in [67]. MARS15 simulation predictions for the same quantities are in red, with 1σ statistical error bars. HARP 1σ error bars also include systematics.

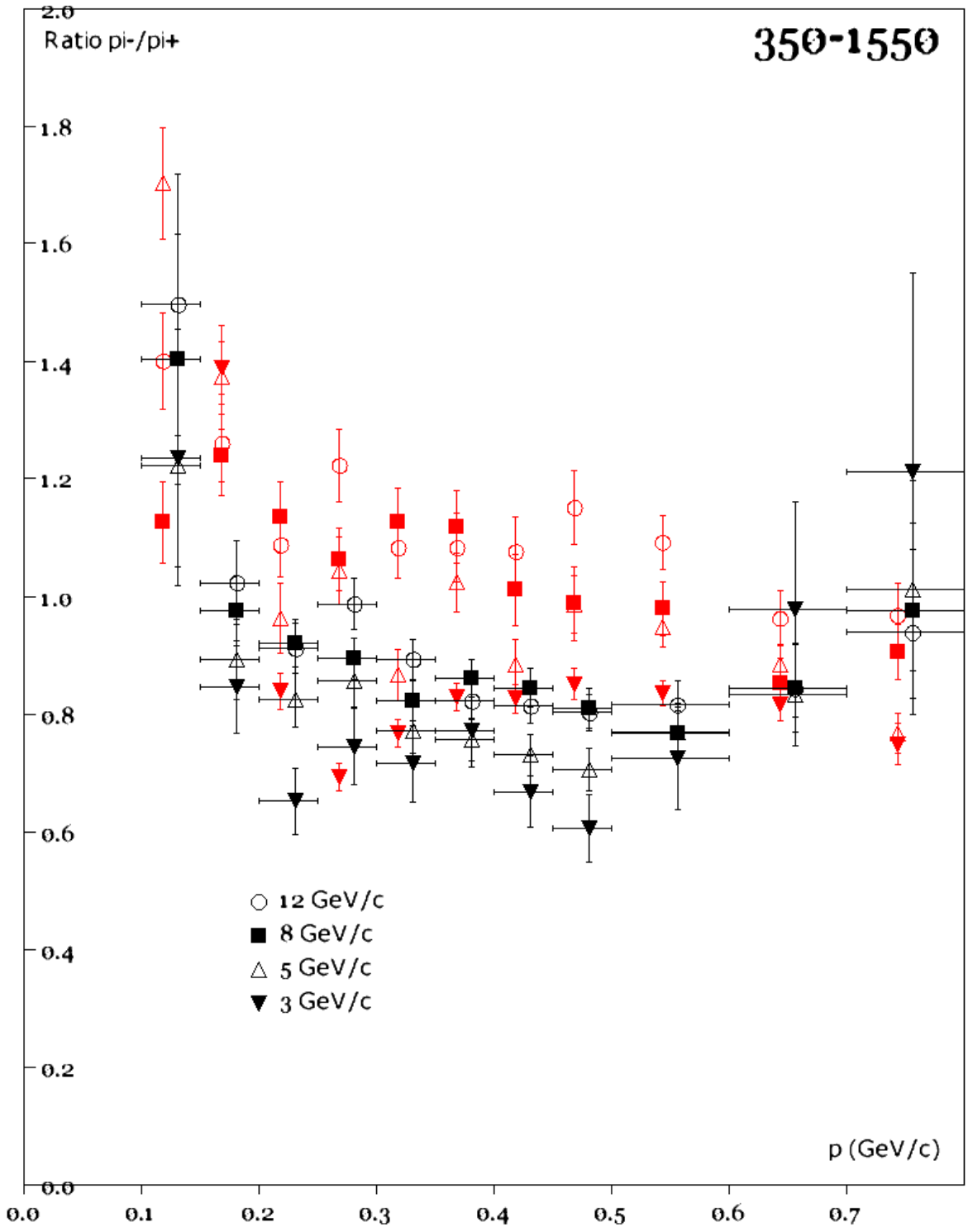


Figure 5.8: Ratio of negative over positive pions in the forward region of the HARP Large Angle Spectrometer. HARP measurements are in black and MARS15 simulation results in red. This is comparable with figure 27 of [67]. MARS error bars are 1σ statistical; HARP error bars are 1σ but also include systematics.

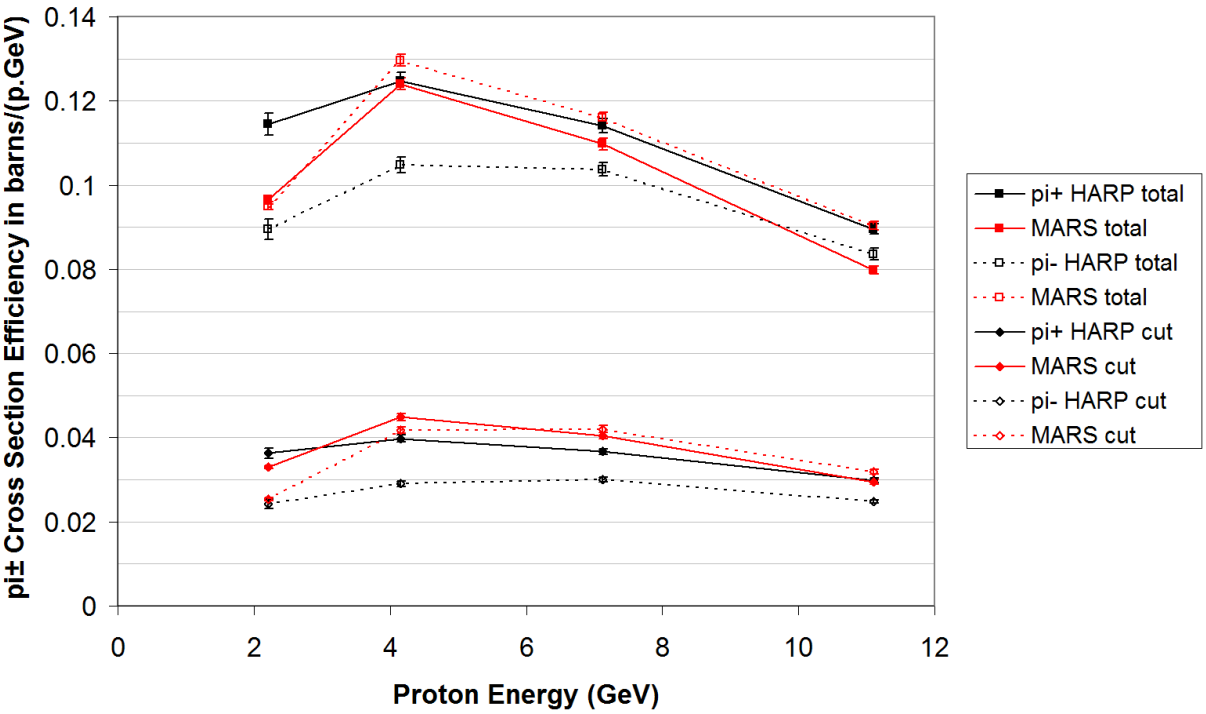


Figure 5.9: Pion cross section per GeV of proton energy as a function of the proton beam energy. Results are shown both for the entire forward LAS acceptance with $|\mathbf{p}| < 700\,\mathrm{MeV/c}$ and for a cut with $\theta < 0.95\,\mathrm{rad}$ and $250 \leq |\mathbf{p}| < 500\,\mathrm{MeV/c}$ in the area of the neutrino factory acceptance. This plot is comparable to figure 30 of [67] without the θ-only cut. MARS error bars are 1σ statistical; HARP error bars are 1σ but also include systematics.

The energy dependence characteristics for both MARS and HARP data suggest that the optimal proton energy is around 5 GeV, lower than previously concluded, with a slightly faster drop off as the energy increases. Note however that the exclusion of forward pions with $\theta < 0.35\,\mathrm{rad}$ from the calculation may systematically bias towards lower proton energies where the lack of forward momentum allows more pions to escape with larger angles. The fact that even the MARS results now favour this lower energy suggests this change may be due to the difference in cuts rather than a difference between the code and reality. Additionally, the slope between 5 GeV yields and those at 10 GeV was not very steep in previous MARS results, so it could be reversed by quite a small effect. One energy dependence feature that is definitely different between code and reality is how the MARS results fall much more rapidly than HARP when going from 5 to 3 GeV/c proton momentum. This is due to the interpolation between hadron models that kick in for MARS at about this energy, as illustrated in figure 5.3.

Overall, the agreement between MARS15 and HARP data in the pion momentum and proton energy ranges of interest, with the correct target material, is very encouraging for positive pions (deviations of less than 13% in these figures of merit), whereas MARS overestimates negative pion production here by about 30%, which should be taken into account in future. Later releases of MARS and GEANT will hopefully integrate more sophisticated fits from the HARP data and allow more accurate evaluation, perhaps at the 10% level or better, of the neutrino factory pion yield.

Chapter 6

Muon Energy Loss and Scattering Algorithm

The beam in the neutrino factory front end is larger and more divergent than those found in conventional accelerators. The *emittance* of a beam (defined in section C.4) is a measure of the beam size and divergence that is invariant under optical elements such as magnets, so can only be reduced via other means. D. Neuffer in 1983 therefore proposed *muon ionisation cooling* [68, 69], originally intended for muon colliders, that uses the energy loss of muons in a low-Z material to reduce the emittance. However, the material also causes scattering of the muons, which increases the emittance. After presenting analytic approximations of the resultant cooling rate along a section of beamline, the main result of this chapter is an accurate computer model of scattering using cross-sections of the muon–LH_2 interaction that can be integrated into tracking codes.

Although this cooling model has been integrated into Muon1, it was found that there exist promising front ends without cooling (see chapter 8), provided a muon linac is used for further acceleration. Thus the main results of this thesis do not include muon cooling but performance of the cooling algorithm on a 30 cm LH_2 block is presented in this chapter.

6.1 Principle of Muon Cooling

The muons from the decay channel have transverse momenta of up to 200 MeV/c, whereas beams in conventional accelerators have particles with much lower p_T, which can be handled with weaker magnets. For example the ISIS [50] proton accelerator covers the energy range 70–800 MeV (of the same order as the muons) but the p_T of the protons does not exceed 30 MeV/c. Optically, p_T can be decreased by defocussing the beam to a larger spatial size but the muon beam is already very large ($\sim$30 cm width) in the phase rotation channel, so that method of decreasing p_T is not an option.

Liouville's Theorem prohibits a system of pure electric and magnetic fields from reducing the amount of position-momentum phase space occupied by a beam with a very large number of particles. Without going into detail, this is because these systems can be modelled as a Hamiltonian system and Liouville's Theorem is a mathematical consequence

of Hamilton's equations [70].

For muon cooling and other applications where this phase space volume must be reduced, exceptions to the conditions of this theorem must be found. One way is to notice a beam is not a continuum of infinitely-many particles and statistical effects may be used. A famous example of this is stochastic cooling of protons [71], which relies on the small deviation of the centroid of the beam due to statistical fluctuations to create an opposing electric field, cancelling a small amount of the beam spatial spread and over very many revolutions reducing the beam size in a ring significantly. This method becomes slower in proportion to the square root of the longitudinal density of particles, typically taking seconds for proton beams, so would be unsuitable for the neutrino factory's high-intensity muon beam given the muon's short lifetime. †

Another class of exceptions to Liouville's Theorem is when the dynamics are not fully represented by a Hamiltonian system. Introducing an additional process, very often an interaction with matter, can give systems of this sort. One example is 'charge-stripping injection' in which an ion such as H^- encounters a thin foil and has its electrons stripped off, giving a proton. This is an irreversible process as protons reencountering the foil are just scattered slightly. The effect is that many copies of an original H^- beam may be superimposed on the same phase space for protons because they come from outside the protons' Hamiltonian system using the foil; superimposition of the two different-sign beams is easily done using a dipole magnet because they bend in opposite directions. Scattering is also a non-Hamiltonian process as the scattering kick given to each particle is random and not a function of its canonical coordinates: this increases the phase space volume and results in beam 'heating'.

The primary mechanism used in ionisation cooling is the energy loss of particles as they collide with those of matter, causing the matter to heat up slightly. This is a member of the class of non-Hamiltonian exceptions to Liouville's Theorem. It may not seem obvious that deceleration in matter is a non-Hamiltonian system, as it can be described by smooth differential equations under some circumstances. Assuming the material is homogenous and stationary, the formulae for $\dot{x}$ and $\dot{p}$, that is $\frac{\partial H}{\partial p}$ and $-\frac{\partial H}{\partial x}$, must not depend on x or t. If $\frac{\partial H}{\partial p}$ is only a function of p then $H = f(p) + g(x,t)$. Now $\dot{p} = -\frac{\partial H}{\partial x} = -\frac{\partial g}{\partial x}(x,t)$, but this must also depend only on p, thus is a constant, which is inconsistent with deceleration of particles in all directions because $\dot{p}$ must be negative for positive p and *vice versa*. So energy loss in materials is a possible candidate for non-Liouvillian reduction of the phase space volume ('cooling') of the beam.

The reduction of momentum regardless of its direction is the important feature of energy loss in the context of a cooling system. Removing energy from a nearly-parallel beam is equivalent to putting it through a decelerating electric field, which is known to be Liouvillian, so the important effects can only appear when the beam is put through the energy absorber with a range of angles. These are shown as effects on the particle's momentum vector in figure 6.1: the particle loses energy in its direction of travel, experi-

†The statistical exception to Liouville's Theorem can also allow emittance *growth*: repeated nonlinear forces (or dynamical distortions) acting on the beam can eventually cause the continuous volume in phase space containing the beam to become so twisted and intermixed with a larger external volume that it is no longer well-represented by the limited number of particles it contains. At this point, the beam is considered to have grown in phase space volume and fills a larger region with a lower density.

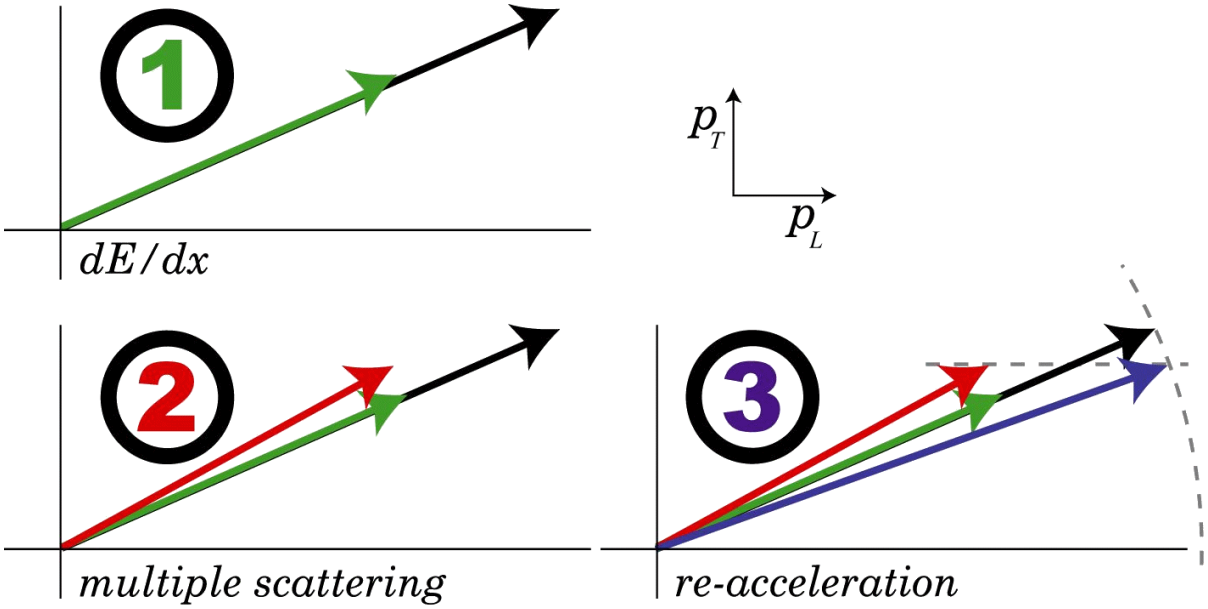


Figure 6.1: Muon momentum is reduced in magnitude (step 1) along the direction of travel by some material before being replenished by forward momentum by an RF cavity (step 3). This would reduce the p_T of all muons if it were not for scattering in the material (step 2) that counteracts the reduction when p_T is small enough. Diagram adapted from [74].

ences some scattering (an inevitable material effect that heats the beam) and then can be reaccelerated in a phase-space-conserving way by the electric field of an RF cavity. Notice that for sufficiently large incident particle angles (relative to the RF axis), or sufficiently small amounts of scattering, this results in a beam in which *all* particles have momentum vectors more parallel to the axis than before. Because the non-Liouvillian absorber is present, this no longer comes at the cost of an enlarged beam in positional coordinates, though it does come at the cost of magnets to provide large angles in the absorber.

The US Feasibility Study II [25] performed a detailed investigation of a channel of absorbers, magnets and RF cavities to achieve muon ionisation cooling, resulting in their design of a cooling unit cell shown in figure 6.2. As the Z of the absorber material increases, scattering increases faster than the energy loss, meaning that the optimal material to use is liquid hydrogen (gaseous hydrogen not being dense enough to have an effect within a practical length). This choice implies a complex cryogenic system and containment for the hydrogen, which must also be contained with low-Z (but not liquid) windows, beryllium being chosen there. To avoid a net deceleration of the beam, the amount of RF must be balanced to that of absorber and the channel must also be relatively short to reduce muon decays. Finally, the focussing magnets must be added, solenoids being used here for their axial symmetry and short focal length for relatively low-energy beams.

The LH_2 design is the one being tested in the MICE experiment [72]. However, the more recent ISS [24] and IDS [73] designs have used lithium hydride (LiH) absorbers, trading off some cooling performance in the higher Z material in exchange for a simpler cooling cell with less cryogenics. It is also possible to reclaim some of the lost performance by having thinner beryllium windows on the RF cavities because the windows are no longer structural: they are supported by the LiH absorber material.

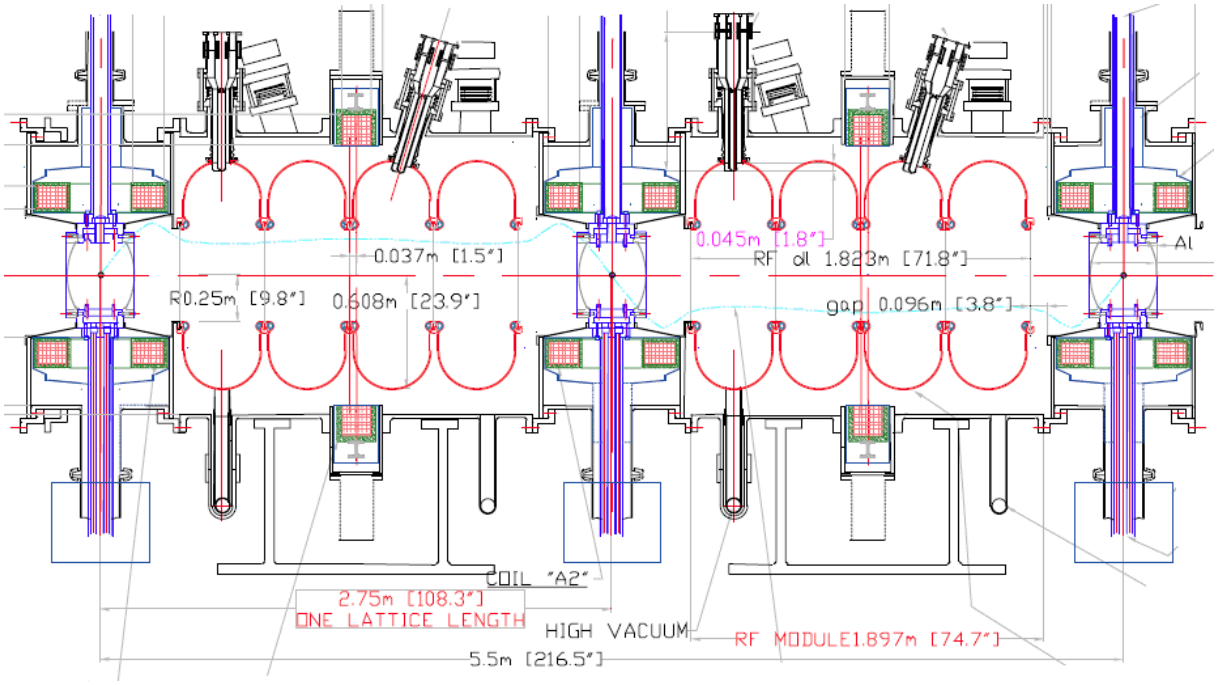


Figure 6.2: Cross section of the cooling lattice cell from US Study II [25], containing four-cell RF cavities with liquid hydrogen absorbers between. The surrounding solenoid coils can be seen as rectangular blocks.

6.2 Cooling Rates from Linearised Optics

The effect of the beam passing through matter may be approximated while staying within the linear optics model described in appendix C. A single transverse plane (x, x') will be considered and 'emittance' in this section will usually be the normalised RMS emittance of that plane. As explained in section 6.1, a cooling channel normally consists of reacceleration after the energy-absorbing material: thus the choice of normalised emittance, which is constant under acceleration. If the un-normalised emittance had been chosen instead, it would appear to decrease in the accelerating cavities but increase in the material from scattering, though the net effect would be the same as the material-only effect on the normalised value, provided the beam ends the channel with the same energy it started with.

The useful part of material interaction with regard to cooling is the energy loss per unit length of material $\frac{\mathrm{d}E}{\mathrm{d}z} < 0$, which can be derived from the Bethe–Bloch formula [75]. With pure energy loss, x' stays the same while x continues to evolve as in a drift space, so the only effect on $\varepsilon^{\mathrm{n}}_{\mathrm{rms}}$ is via the normalising factor $\beta\gamma$.

$$\begin{aligned}
\left(\frac{\mathrm{d}\varepsilon^{\mathrm{n}}_{\mathrm{rms}}}{\mathrm{d}z}\right)_{\mathrm{loss}} &= \frac{\mathrm{d}}{\mathrm{d}z}\left[\beta\gamma\varepsilon^{\mathrm{un}}_{\mathrm{rms}}\right] = \varepsilon^{\mathrm{un}}_{\mathrm{rms}}\frac{\mathrm{d}(\beta\gamma)}{\mathrm{d}z} \\
&= \varepsilon^{\mathrm{un}}_{\mathrm{rms}}\frac{\mathrm{d}(\beta\gamma)}{\mathrm{d}E}\frac{\mathrm{d}E}{\mathrm{d}z} = \varepsilon^{\mathrm{un}}_{\mathrm{rms}}\frac{1}{mc^2}\frac{\mathrm{d}(\beta\gamma)}{\mathrm{d}\gamma}\frac{\mathrm{d}E}{\mathrm{d}z} \\
&= \frac{\varepsilon^{\mathrm{un}}_{\mathrm{rms}}}{\beta mc^2}\frac{\mathrm{d}E}{\mathrm{d}z} = \frac{\varepsilon^{\mathrm{n}}_{\mathrm{rms}}}{\beta^2\gamma mc^2}\frac{\mathrm{d}E}{\mathrm{d}z}
\end{aligned}$$

$$= \frac{\varepsilon^{\mathrm{n}}_{\mathrm{rms}}}{\beta^2(mc^2+E_k)}\frac{\mathrm{d}E}{\mathrm{d}z}$$

Here, E_k is the kinetic energy of the particle and m is its rest mass ($\mathrm{d}E$ is identical for kinetic or total energy). The formula shows that the negative $\frac{\mathrm{d}E}{\mathrm{d}z}$ of the material causes a reduction in emittance, which cannot go below 0 due to the presence of $\varepsilon^{\mathrm{n}}_{\mathrm{rms}}$ on the right-hand side. If the particle were kept at a constant E_k (and β) by compensating acceleration, the predicted cooling *without scattering* in a uniform material can be evaluated by defining the cooling rate $k > 0$:

$$\begin{aligned} -k &= \frac{\mathrm{d}\ln\varepsilon^{\mathrm{n}}_{\mathrm{rms}}}{\mathrm{d}z} = \frac{1}{\varepsilon^{\mathrm{n}}_{\mathrm{rms}}}\frac{\mathrm{d}\varepsilon^{\mathrm{n}}_{\mathrm{rms}}}{\mathrm{d}z} = \frac{\mathrm{d}E/\mathrm{d}z}{\beta^2(mc^2+E_k)} \\ &\Rightarrow \varepsilon^{\mathrm{n}}_{\mathrm{rms}}(z) = \varepsilon^{\mathrm{n}}_{\mathrm{rms}}(0)e^{-kz}. \end{aligned}$$

The Bethe–Bloch curves give a point of minimum energy loss at approximately $\beta\gamma = 4$. This is of interest mainly because the energy loss does not change much with particle energy around this point, so beams with energy spread do not undergo nonuniform deceleration, to first order. For muons this corresponds to $E_k = 330\,\mathrm{MeV}$ and $\beta = 0.970$, a little higher than the energy in the muon front end. In liquid hydrogen at this point, the energy loss is $3.97\,\mathrm{MeV}$ per $\mathrm{gram/cm^2}$ of material [27], or $\frac{\mathrm{d}E}{\mathrm{d}z} = -28.2\,\mathrm{MeV/m}$ given its density of only $0.071\,\mathrm{g/cm^3}$. For graphite, the energy loss is $1.74\,\mathrm{MeV}$ per $\mathrm{g/cm^2}$ [76] and the density varies depending on grade but is approximately $2.2\,\mathrm{g/cm^3}$, giving $\frac{\mathrm{d}E}{\mathrm{d}z} = -383\,\mathrm{MeV/m}$. The cooling rates are therefore $k_{\mathrm{LH_2}} = 0.0687\,\mathrm{m^{-1}}$ and $k_{\mathrm{graphite}} = 0.935\,\mathrm{m^{-1}}$. At $\beta\gamma = 2.5$ in liquid hydrogen the energy loss per unit length is a few percent higher but $k_{\mathrm{LH_2}}$ increases to $0.120\,\mathrm{m^{-1}}$, mainly through the reduction of E_k to $179\,\mathrm{MeV}$, so the energy lost is a greater proportion of the muon's kinetic energy. $179\,\mathrm{MeV}$ is very near the front end energy. Continuing to decrease E_k would theoretically increase the cooling rate arbitrarily, but the benefits are lost from the nonuniform deceleration experienced far from the Bethe–Bloch minimum and scattering in the material, discussed next.

Collisions with the electrons and nuclei in a material cause frequent random changes to the incoming particle's momentum. In the case of muons, these collisions are almost entirely electromagnetic in nature (as opposed to pions that can also undergo hadronic processes). For the purposes of the linear model, these will be approximated by a continuous random walk in which $\langle x'^2\rangle$ increases linearly with distance. This generates an increase in beam emittance. By the Central Limit Theorem, the process will produce a Gaussian distribution if continued indefinitely from any starting distribution and a Gaussian distribution is the only one that will emerge unchanged apart from an increase in variance. It is also a small-angle approximation that neglects large-angle scatters that occur in rare head-on collisions with nuclei and will usually lead to the muon being lost.

The starting point is a formula [77] for the RMS angular deflection (projected into one transverse plane) of a particle emerging from a thickness z of material with radiation length X_0:

$$\theta_{\mathrm{rms}} = \frac{13.6\,\mathrm{MeV/c}}{\beta p}(q/e)\sqrt{\frac{z}{X_0}}\left(1+0.038\ln\frac{z}{X_0}\right),$$

where q/e is the charge of the particle in electron units and p is the forward momentum. This comes from a fit to the more comprehensive Molière theory of scattering; the logarithmic term in brackets to the right is not compatible with the simple random walk model

but fortunately has values within a few per cent of unity for thicknesses of similar order to X_0. For liquid hydrogen, $X_0 = 8.90\,\mathrm{m}$ and for graphite, $0.193\,\mathrm{m}$, which are similar to the characteristic cooling lengths $1/k$ defined previously, so this approximation is likely to be good for cooling systems. Neglecting this term and using the small-angle approximation $\theta \simeq x'$, the outgoing mean square muon angle is

$$\left\langle x'^2 \right\rangle = \theta_{\mathrm{rms}}^2 = \frac{(13.6\,\mathrm{MeV/c})^2}{\beta^2 p^2}(q/e)^2 \frac{z}{X_0},$$

which increases linearly with z by

$$\frac{\mathrm{d}\left\langle x'^2 \right\rangle}{\mathrm{d}z} = \frac{(13.6\,\mathrm{MeV/c})^2 (q/e)^2}{\beta^2 p^2 X_0}.$$

Recalling the definition of 2D RMS emittance from section C.5, the normalised emittance under consideration can be written in terms of beam moments

$$\varepsilon_{\mathrm{rms}}^{\mathrm{n}} = \beta\gamma\varepsilon_{\mathrm{rms}}^{\mathrm{un}} = \beta\gamma\sqrt{\left\langle x^2 \right\rangle \left\langle x'^2 \right\rangle - \left\langle xx' \right\rangle^2}.$$

The scattering process considered separately to energy loss does not affect the relativistic factors β and γ; nor does it immediately (without drift) affect the beam size $\langle x^2 \rangle$ or the correlation $\langle xx' \rangle$ because the scatters are uncorrelated with x position. Only $\langle x'^2 \rangle$ remains as a relevant variable and the effect of scattering on emittance may be evaluated directly using the chain rule:

$$\begin{aligned}
\left(\frac{\mathrm{d}\varepsilon_{\mathrm{rms}}^{\mathrm{n}}}{\mathrm{d}z}\right)_{\mathrm{scatter}} &= \frac{\mathrm{d}\varepsilon_{\mathrm{rms}}^{\mathrm{n}}}{\mathrm{d}\left\langle x'^2 \right\rangle} \frac{\mathrm{d}\left\langle x'^2 \right\rangle}{\mathrm{d}z} \\
&= \beta\gamma \frac{\mathrm{d}}{\mathrm{d}\left\langle x'^2 \right\rangle} \left[\sqrt{\left\langle x^2 \right\rangle \left\langle x'^2 \right\rangle - \left\langle xx' \right\rangle^2}\right] \frac{\mathrm{d}\left\langle x'^2 \right\rangle}{\mathrm{d}z} \\
&= \beta\gamma \frac{\left\langle x^2 \right\rangle}{2\sqrt{\left\langle x^2 \right\rangle \left\langle x'^2 \right\rangle - \left\langle xx' \right\rangle^2}} \frac{\mathrm{d}\left\langle x'^2 \right\rangle}{\mathrm{d}z} = \beta\gamma \frac{\left\langle x^2 \right\rangle}{2\varepsilon_{\mathrm{rms}}^{\mathrm{un}}} \frac{\mathrm{d}\left\langle x'^2 \right\rangle}{\mathrm{d}z}.
\end{aligned}$$

Recalling the definition of the Twiss function $\beta = \langle x^2 \rangle / \varepsilon_{\mathrm{rms}}^{\mathrm{un}}$ and using the earlier formula for $\mathrm{d}\langle x'^2 \rangle / \mathrm{d}z$,

$$\begin{aligned}
\left(\frac{\mathrm{d}\varepsilon_{\mathrm{rms}}^{\mathrm{n}}}{\mathrm{d}z}\right)_{\mathrm{scatter}} &= \frac{\beta\gamma\beta}{2} \frac{(13.6\,\mathrm{MeV/c})^2 (q/e)^2}{\beta^2 p^2 X_0} = \frac{(13.6\,\mathrm{MeV/c})^2 (q/e)^2 \gamma\beta}{2\beta p^2 X_0} \\
&= \frac{(13.6\,\mathrm{MeV/c})^2 (q/e)^2 \gamma\beta}{2\beta^3\gamma^2 m^2 c^2 X_0} = \frac{(13.6\,\mathrm{MeV/c})^2 (q/e)^2 \beta}{2\beta^3 (mc^2 + E_k) m X_0}.
\end{aligned}$$

Combining these results gives the overall change in emittance in a uniform absorbing material with both scattering and energy loss [78]:

$$\frac{\mathrm{d}\varepsilon_{\mathrm{rms}}^{\mathrm{n}}}{\mathrm{d}z} = \left(\frac{\mathrm{d}\varepsilon_{\mathrm{rms}}^{\mathrm{n}}}{\mathrm{d}z}\right)_{\mathrm{loss}} + \left(\frac{\mathrm{d}\varepsilon_{\mathrm{rms}}^{\mathrm{n}}}{\mathrm{d}z}\right)_{\mathrm{scatter}} = \frac{\varepsilon_{\mathrm{rms}}^{\mathrm{n}}}{\beta^2 (mc^2 + E_k)} \frac{\mathrm{d}E}{\mathrm{d}z} + \frac{(13.6\,\mathrm{MeV/c})^2 (q/e)^2 \beta}{2\beta^3 (mc^2 + E_k) m X_0}.$$

This consists of a negative part that is proportional to $\varepsilon_{\mathrm{rms}}^{\mathrm{n}}$ and a positive part that is constant. Ignoring kinematic factors, the negative (cooling) term is maximised by choosing a very negative $\frac{\mathrm{d}E}{\mathrm{d}z}$ and the positive (heating) term is minimised by choosing a large X_0

and small β in the cooling medium. Note that the β is actually a function of z depending on the optics of the beam in the absorber, but with many absorbers it can be replaced by an averaged constant value.

The cooling equation can be solved to give an offset exponential:

$$\begin{aligned} \frac{\mathrm{d}\varepsilon_{\mathrm{rms}}^{\mathrm{n}}}{\mathrm{d}z} &= -k\varepsilon_{\mathrm{rms}}^{\mathrm{n}} + h \\ \Rightarrow \quad \varepsilon_{\mathrm{rms}}^{\mathrm{n}}(z) &= (\varepsilon_{\mathrm{rms}}^{\mathrm{n}}(0) - h/k)\, e^{-kz} + h/k \\ &= (\varepsilon_{\mathrm{rms}}^{\mathrm{n}}(0) - \varepsilon_{\mathrm{eq}})\, e^{-kz} + \varepsilon_{\mathrm{eq}}, \end{aligned}$$

where the cooling rate k is defined as before, h is the heating term and the *equilibrium emittance*

$$\varepsilon_{\mathrm{eq}} = h/k = \frac{(13.6\,\mathrm{MeV/c})^2 (q/e)^2 \beta}{2\beta m X_0 (-\mathrm{d}E/\mathrm{d}z)}$$

is the value that $\varepsilon_{\mathrm{rms}}^{\mathrm{n}}$ converges to as $z \to \infty$. If $\varepsilon_{\mathrm{rms}}^{\mathrm{n}}$ starts below $\varepsilon_{\mathrm{eq}}$ then the scattering dominates the cooling from energy loss and the beam emittance grows. Note that $(q/e)^2 = 1$ for muons.

For solids and liquids, it is the reaccelerating RF systems and not the energy loss in the material that determine how large the average $\frac{\mathrm{d}E}{\mathrm{d}z}$ in a channel can be. It is also possible to cool using a large amount of gas but this time higher pressures ($\sim$100 atm) are needed to prevent the slow energy loss from making the cooling channel impractically long. When the beam emittance is much larger than $\varepsilon_{\mathrm{eq}}$, the average energy loss rate is the dominating factor in how rapidly the cooling progresses, so the channel is optimised for the limited apertures of the RF cavities to allow as much reacceleration as possible. Liquid hydrogen is often cited as a good cooling material [72] because it has the lowest X_0 of any solid or liquid, the low scattering allowing beams to have relatively low x' in the absorbers. To optimise the performance of cooling in the final, low-emittance stages, $\varepsilon_{\mathrm{eq}}$ can be reduced by choosing materials with the highest possible product $X_0(-\mathrm{d}E/\mathrm{d}z)$, which favours slightly higher-Z materials such as graphite. At this point the emittance is assumed to be small enough that optics with high x' (low β) in the absorbers can be produced without the beam later becoming too large for the RF apertures.

6.2.1 Example

The best yielding front end from chapter 8 produces a beam with RMS normalised emittance $\varepsilon_{\mathrm{in}} = 14812\,\mathrm{mm{\cdot}mrad}$ when projected into either transverse phase space plane (table 8.3). Suppose one wishes to add cooling to this front end until the beam can be transmitted by an FFAG-based accelerator that has a normalised acceptance of only $A_{\mathrm{out}} = 15000\,\mathrm{mm{\cdot}mrad}$ in each plane, instead of a muon linac. Transverse focussing optics suitable for cooling channels are described in detail in [79]; here the absorber and beta function from the start of the Study II [25] cooling channel are used.

The acceptance places a hard cut on particles outside the ellipse of π times that area, whereas the ellipse corresponding to the RMS emittance contains only $1-\sqrt{1/e} = 39.35\%$ of a beam with a Gaussian distribution. After more than one cooling length in a cooling channel, the beam will become very close to a Gaussian, so that distribution is assumed

here. If 95% of the beam is required to be within the 2D ellipses of both planes, it is sufficient to ensure that 97.5% is within each one. The amount of a 2D Gaussian outside n sigma is $e^{-n^2/2}$ and setting this lost amount to 2.5% gives $n = 2.716$. Since $A_{\text{out}} = n^2\varepsilon_{\text{out}}$, the required normalised RMS output emittance $\varepsilon_{\text{out}} = 2033\,\text{mm·mrad}$ is derived.

The central beam momentum in the channel is 200 MeV/c, corresponding to a kinetic energy $E_k = 120.5\,\text{MeV}$. Each 2.75 m cell contains a 35 cm thick liquid hydrogen absorber, plus enough RF cavities to restore the lost beam energy. The beta function minimum where the absorber is located is $\beta = 47\,\text{cm}$. Using the data for LH_2 given previously in this section and $X_{0,\text{LH}_2} = 8.81\,\text{m}$, calculated from Tsai's formula [80], the equilibrium RMS emittance is

$$\varepsilon_{\text{eq}} = \frac{(13.6\,\text{MeV/c})^2\ 0.47\,\text{m}}{2(0.884)(105.7\,\text{MeV/c}^2)(8.81\,\text{m})(28.2\,\text{MeV/m})} = 0.001873\,\text{m} = 1873\,\text{mm}\cdot\text{mrad},$$

which is definitely small enough for this example. The length of cooling material required can be found by setting $\varepsilon^{\text{n}}_{\text{rms}}(0) = \varepsilon_{\text{in}}$ and $\varepsilon^{\text{n}}_{\text{rms}}(z) = \varepsilon_{\text{out}}$:

$$\varepsilon^{\text{n}}_{\text{rms}}(z) = (\varepsilon^{\text{n}}_{\text{rms}}(0) - \varepsilon_{\text{eq}})\,e^{-kz} + \varepsilon_{\text{eq}} \quad \Rightarrow \quad z = \frac{1}{k}\ln\left(\frac{\varepsilon_{\text{in}} - \varepsilon_{\text{eq}}}{\varepsilon_{\text{out}} - \varepsilon_{\text{eq}}}\right).$$

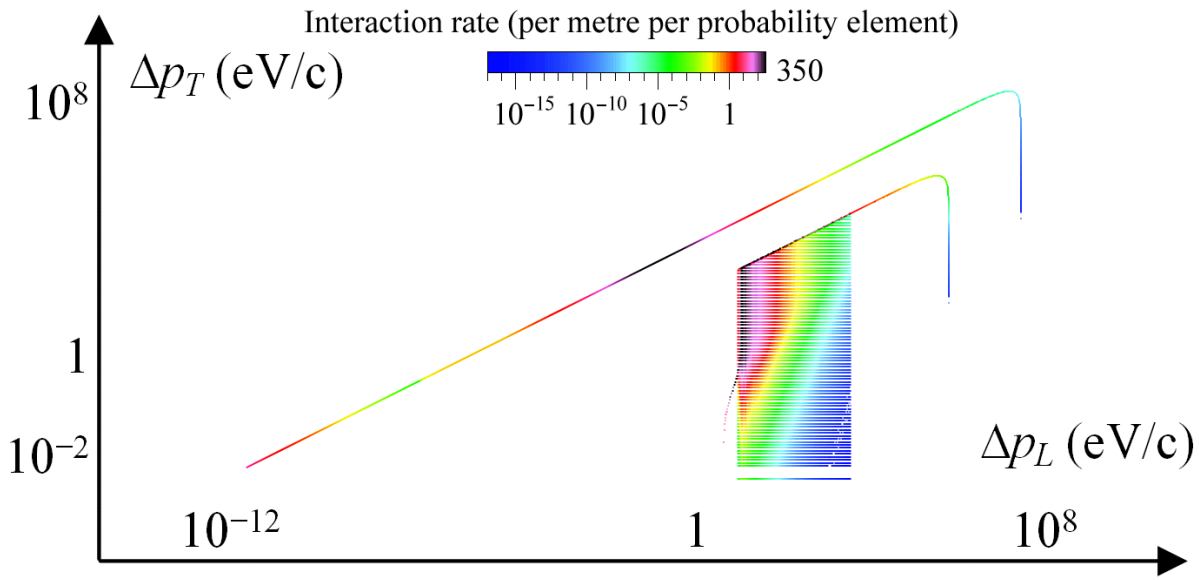
For the muon energy considered, $k = 0.159\,\text{m}^{-1}$ and the log term is $\ln(80.7) = 4.39$, giving a total material length of $z_{\text{LH}_2} = 27.53\,\text{m}$ required for cooling. To get this thickness requires 79 of the 35 cm Study II absorber units and therefore 79 cooling lattice cells of 2.75 m length each, giving a total cooling channel length of 217.25 m.

Finally, note that since $\beta\gamma = 1.89$, the input beam's 99% normalised emittance of 97455 mm·mrad corresponds to an un-normalised emittance $\varepsilon^{\text{un}}_{99} = 51485\,\text{mm·mrad}$ at the start of the cooling channel. The radius of the beam at the absorber may be calculated using the beta function: $r = \sqrt{\beta\varepsilon^{\text{un}}_{99}} = 15.6\,\text{cm}$. The absorbers specified in Study II have a window diameter of 360 mm ($r = 18\,\text{cm}$), which is large enough for the beam from the front end.

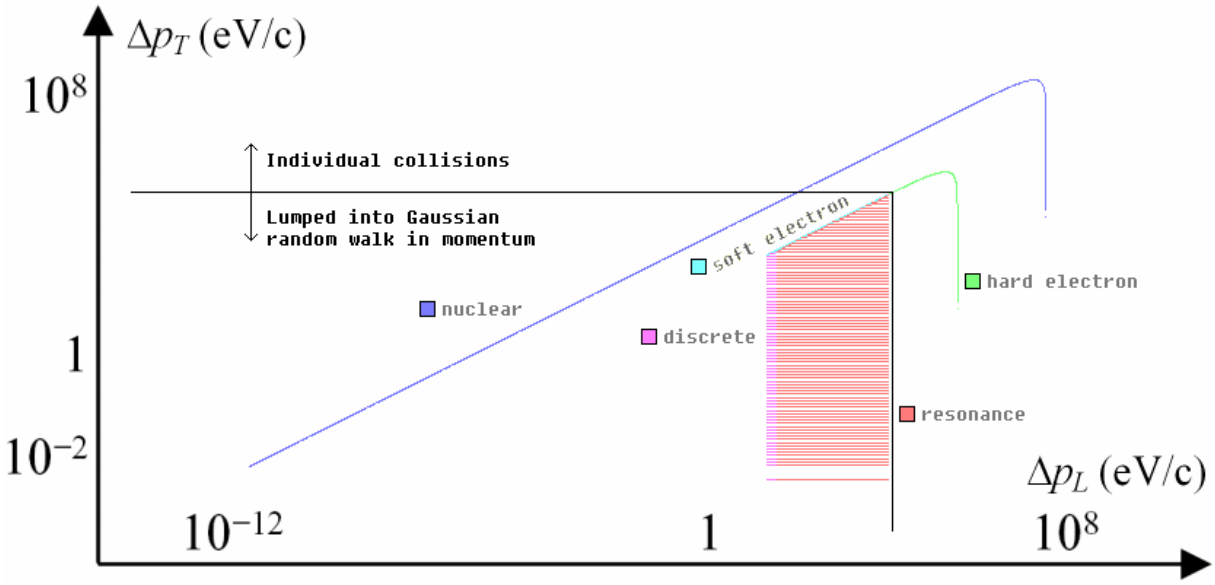
6.3 Using the ELMS Scattering Model for Liquid Hydrogen

A model of the interaction of muons with matter is required in order to simulate the cooling channel in Muon1. These interactions consist of different types of collisions by the muons with atomic electrons and nuclei, each of which transfers a certain amount of transverse and a certain amount of (backward) longitudinal momentum to the muon. These also depend on the incoming muon momentum, so abstractly a material M can be described by a function $F_M(\Delta p_L, \Delta p_T; p)$ where $\iint F_M(\Delta p_L, \Delta p_T; p)\text{d}\Delta p_L\text{d}\Delta p_T$ gives the frequency of collisions (in inverse metres) along the path of a muon with momentum p. Integrating over a specific range of $(\Delta p_L, \Delta p_T)$ momentum kicks gives the partial frequency of collisions in that range. This function is related simply to the double differential cross-section by the molecular number density ρ_M of the material:

$$F_M = \rho_M \frac{\text{d}^2\sigma_M}{\text{d}\Delta p_L\text{d}\Delta p_T}.$$



(a) Magnitudes of the delta function elements shown as colours.



(b) Location of the soft/hard collision cut.

Figure 6.3: Elements of the ELMS [81] muon–LH_2 cross-section for a 200 MeV/c muon on a log-log plot. The diagonal lines with a sharp downward bend at the end ('nuclear' and 'electron') would be half-ellipses on a linear graph.

The ELMS software [82] was developed at Oxford for the purpose of tracking muons in a realistic manner through a block of LH_2 material. Previous treatments condensed many physical processes into typically just two: longitudinal momentum loss of the muon as it is retarded by the material and multiple scattering, often modelled in the assumption of a large number of soft collisions, producing a Gaussian random walk in transverse momentum. ELMS uses photoabsorption cross-sections in conjunction with atomic physics to determine the full form of F_M (reported in [81]), in which scattering and energy loss are no longer independent processes but instead each collision will cause a little of each.

A log-log plot of the function F_{LH_2} (for liquid hydrogen) with $p = 200\,\mathrm{MeV/c}$ is shown in figure 6.3. The dotting in some regions is because F_M is a probability distribution and is approximated in ELMS by a sum of delta functions: colours in the plot represent the magnitude of each of these and are also on a log scale.

Given this input, tracking a muon through the material is just a matter of sampling the probability distribution and applying the momentum kick at each collision. These occur as a Poisson process along the path, with rate equal to the integral of F_M for the current muon momentum. Unfortunately, this rate is $9.7 \times 10^5\,\mathrm{m}^{-1}$ for a $200\,\mathrm{MeV/c}$ muon, which leads to slow tracking since this is typically 2500 collisions per Muon1 timestep.

6.4 Accelerated Scattering Tracking Algorithm

To increase the speed of the scattering model, Muon1 separates the relatively rare 'hard' collisions, in which the muon momentum and direction can change dramatically, from the frequently occurring softer collisions. For soft collisions, both the paraxial approximation (small angles) and the Central Limit Theorem (frequent collisions with no long tails in distribution) apply, meaning that the evolution of the muon momentum in the intervals *between* successive hard collisions can be treated as a Gaussian continuous random walk. This is exactly what simpler models such as the one used in section 6.2 do everywhere but here it has higher accuracy because the statistically ill-conditioned events have been removed. The expected momentum loss, transverse scattering variance and 'straggling' variance (statistical variation of $\mathrm{d}E/\mathrm{d}z$), which determine the Gaussian $\Delta\mathbf{p}$ distribution, can be reconstructed from 1st and 2nd moments of $F_M(\Delta p_L, \Delta p_T; p)$, where the integrals are performed over the soft region below the cut:

$$\begin{aligned}
\frac{1}{v_z}\frac{\mathrm{d}E}{\mathrm{d}z} = \frac{\mathrm{d}\langle p_L\rangle}{\mathrm{d}z} &= \iint \Delta p_L F_M \mathrm{d}\Delta p_L \mathrm{d}\Delta p_T \qquad \text{(momentum/energy loss);}\\
\frac{\mathrm{dVar}(p_T)}{\mathrm{d}z} &= \iint (\Delta p_T)^2 F_M \mathrm{d}\Delta p_L \mathrm{d}\Delta p_T \qquad \text{(multiple scattering);}\\
\frac{\mathrm{dVar}(p_L)}{\mathrm{d}z} &= \iint (\Delta p_L)^2 F_M \mathrm{d}\Delta p_L \mathrm{d}\Delta p_T - \left(\frac{\mathrm{d}\langle p_L\rangle}{\mathrm{d}z}\right)^2 \qquad \text{(straggling).}
\end{aligned}$$

Here, $\mathrm{Var}(x)$ is the variance operator defined as $\langle (x - \langle x\rangle)^2\rangle = \langle x^2\rangle - \langle x\rangle^2$, a quantity that adds linearly when independent random variables such as the $\Delta\mathbf{p}$ from individual scatters are added to each other. Note that $\mathrm{d}\langle p_T\rangle/\mathrm{d}z = 0$ by symmetry.

A suitable cut between soft and hard collisions (shown in figure 6.3(b)) was chosen

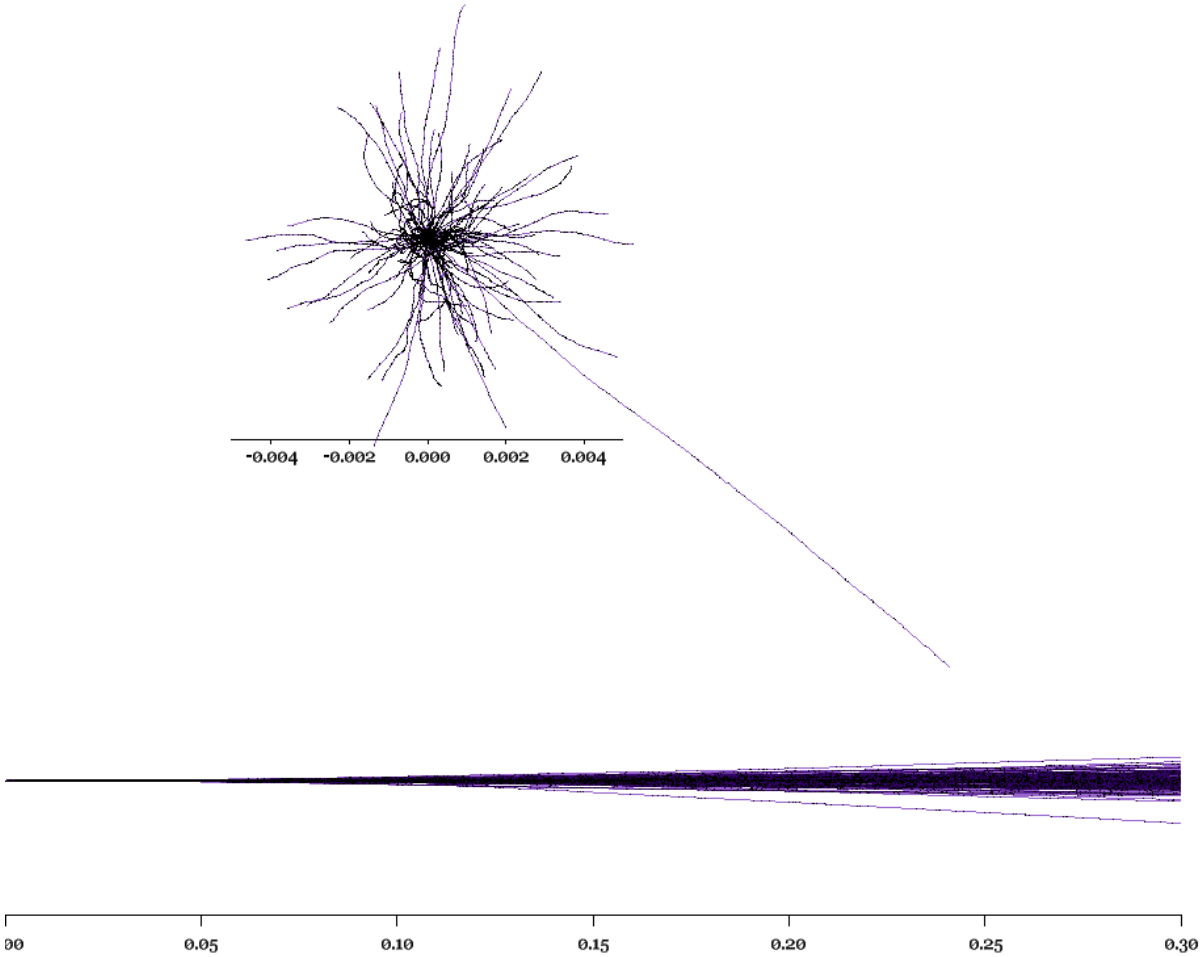


Figure 6.4: A pencil beam of muons is simulated entering an LH_2 absorber. (top) Transverse view of 100 simulated muon paths, looking into the absorber. (bottom) The same paths seen from the side, on a much larger scale (figures in metres).

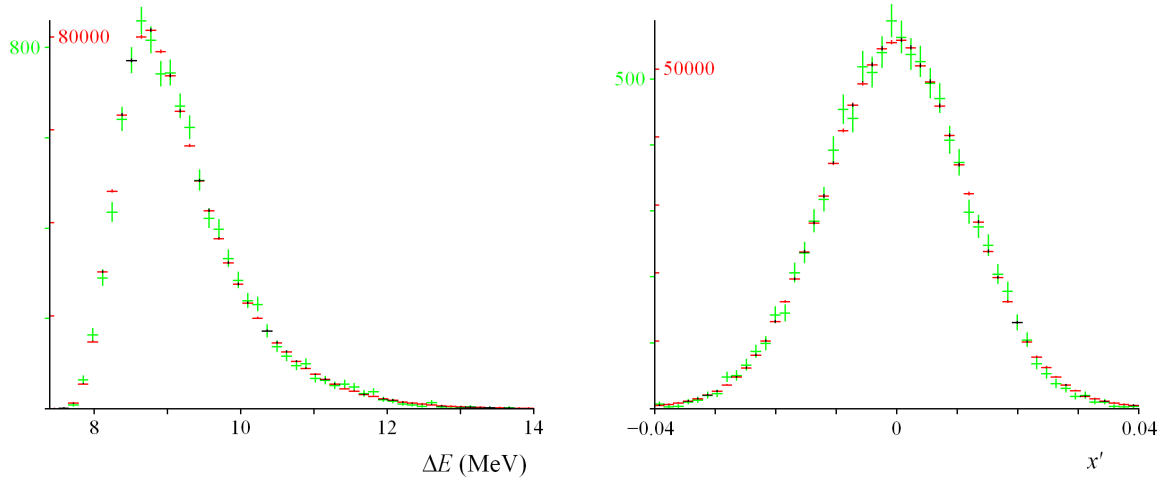


Figure 6.5: Distributions of energy loss and exit angle for 200 MeV/c muons simulated through 30 cm of LH_2. The green plot is 10^4 muons run with individual ELMS collisions, red points are 10^6 muons using the faster method. 1σ statistical error bars are shown.

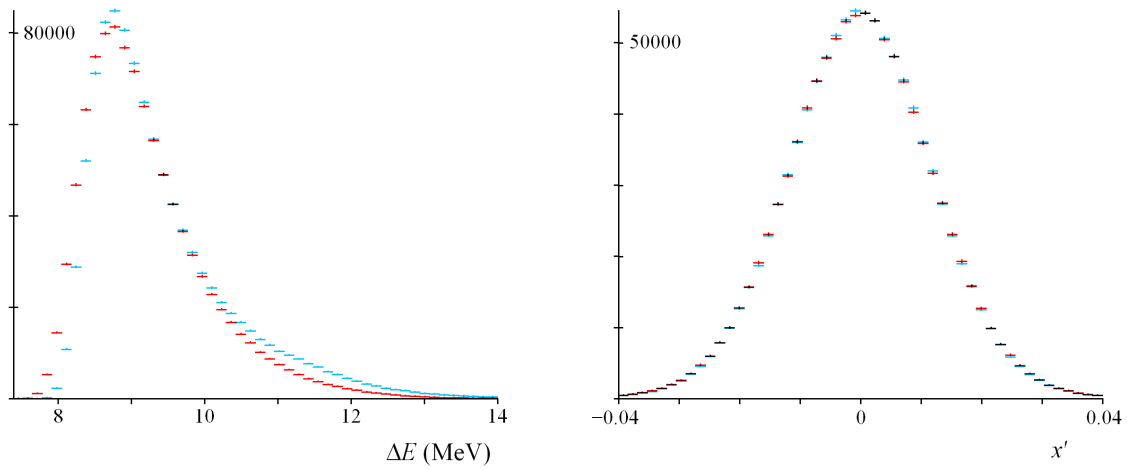


Figure 6.6: Distributions of energy loss and exit angle for one million 200 MeV/c muons simulated through 30 cm of LH_2. The red plot uses the faster ELMS-based method developed in this chapter; the blue plot is ICOOL [30] on default settings. 1σ statistical error bars are shown.

for the integrals. The tracking algorithm models the occurrence of hard collisions as a Poisson process in path length, calculating those collisions exactly and using the Gaussian momentum changes from soft collisions on the segments between. The new algorithm performed 1200 times faster than modelling every collision individually. Muon tracks such as those shown in figure 6.4 are produced, with distributions coming out of the absorber that are so far indistinguishable from complete tracking, as shown in figure 6.5.

6.4.1 Comparison with ICOOL

The same problem with 30 cm of liquid hydrogen was run using the code ICOOL [30], which also includes muon scattering and energy loss models. Figure 6.6 shows the result, with ICOOL agreeing with the ELMS-based method for scattering but having a differently-shaped tail in the energy loss distribution. It should be noted that ICOOL includes several options for these physics models but the defaults were selected here. The input file for this ICOOL run is listed in appendix D.

6.5 Integration with Muon1

The issue of combining sporadic events like collisions with the otherwise smooth Runge–Kutta tracking of a particle curving along its trajectory has been touched on briefly in section 2.5.3, which gives a procedure for finding the collision point with an isosurface. The exact intersection point is found by a root-finding algorithm in t but the position of the particle at each of the various t values tested by this algorithm has to be calculated by invoking the Runge–Kutta algorithm for a partial timestep.

Particle decays are another type of sporadic event and the time distribution of these

was described in section 2.4, without implementation details. Here Muon1 takes the easy option: since the timescale of pion decay is much larger than the 0.01 ns tracking timestep, rounding the decay time to the next timestep is not considered a bad approximation and decays are performed at the end of timesteps for any particles whose internal proper time clock exceeds their next scheduled decay time. Quantitatively, the error in this method is to increase the average pion lifetime by less than 1 part in 5000.

The reason for this recap about sporadic events is that the scattering tracking algorithm described in previous sections introduces quite frequent ones, which the Muon1 tracking logic must handle properly. In figure 6.4 the hard scatters are just visible as black dots marked along the purple trajectories and they occur with a similar frequency to the $\sim$3 mm Muon1 timesteps. The soft scatters also need to be added as a Gaussian perturbation to every substep inside a material.

To describe explicitly the hierarchy of steps and substeps required to take account of all four effects on the particle's motion (the three above plus bending from magnetic fields included in the Runge–Kutta integration), the actual stack of functions called by Muon1 is presented herein. The stack roughly reflects the size scales at which different phenomena take place, mnemonically: $fields \prec scattering \prec geometry \prec decays$.

- `Simulation_advance` is the top-level function called by the main loop of Muon1 repeatedly during a simulation. It advances the whole beam by one timestep, including all physics effects. In fact, because Muon1 detects the number of CPUs present on a computer and splits the beam into an equal number of particles for each, this function is mostly thread-synchronisation and wrapper code, spawning copies of `Simulation_advance_t` in parallel.

- `Simulation_advance_t` is a single-threaded function that advances a range of particles by one timestep including all physics effects. For each particle, it calls `collideparticle` repeatedly if necessary, advancing the particle through one contiguous geometry region at a time until its timestep is over. At this point, `decayparticle` is called on it.

  - `decayparticle` checks if a particle is due to decay by its internal proper time clock and if so, uses the algorithm described in section 2.4.3 to generate its decay products. It calls no further functions.

- `collideparticle` tracks a particle to the end of its current geometry region, including magnetic fields and scattering but not decays. The main purpose is to determine the first surface the particle hits, which marks where it leaves the region. The function loops through all components potentially in the vicinity of the particle (as per section 2.5.1) and calls the collision function for each, which varies from component to component; this in turn usually calls `scatterparticle_isocd` a number of times to see if the particle's path hits any of the isosurfaces described in section 2.5.3.

- `scatterparticle_isocd` determines where a particle hits an arbitrary isosurface, supplied as a function pointer argument. It uses a hybrid algorithm of Newton–Raphson root finding, root bracketing and bisection (in awkward cases) to find the value of t when the particle is on the surface. It is important to note that every

position of the particle used by this algorithm is calculated by a call to `scatterparticle` over a substep, so it takes full account of magnetic fields and scattering.

- `scatterparticle` advances a particle by a given δt through a known material. If the material is vacuum, this passes control directly to `advanceparticle`, otherwise it uses the algorithm described in section 6.4 to calculate a sequence of steps between successive 'hard' collisions. The intervening steps are each a call to `advanceparticle` with the 'soft' Gaussian scattering contribution added on to the momentum afterwards.

- `advanceparticle` is the core Runge–Kutta routine within Muon1 and tracks a particle through a magnetic field over a specified timestep using the algorithm in section 2.2. It does not take account of material scattering, geometry or particle decays. The required **B** field values at various points are obtained by summing the contributions from components suspected to be near the particle, each calculated by that component's magnetic field function (if any).

Chapter 7

Design and Optimisation System

The accelerator design must be specified to the Muon1 code so that the components and magnetic fields can be correctly placed in the simulation space and the algorithms described in chapter 2 applied. Section 7.1 describes the general format of Muon1 lattice files, which can be used to specify either a single accelerator or an optimisation range of designs. After simulation, designs are scored with the figures of merit in section 7.2 and logged in result files described in section 7.3, where they can be used as input for the genetic algorithm (section 7.4) that selects new designs to simulate. As this process of repeated simulation takes a large amount of computing time, it has been deployed as a distributed computing project (section 7.5) where the spare CPU time of volunteers from the Internet is used.

7.1 Muon1 Lattice Files

Accelerators are specified to Muon1 via 'lattice files', which are text files of a special form placed in the `lattices` directory, located in the same folder as the program `muon1.exe`. Lattices are simulated with a command such as `muon1 -c LatticeName`, which will run `lattices\LatticeName.txt` on the command-line. A full list of the command-line switches available is displayed by invoking the `muon1 -?` help switch.

Instead of providing an exhaustive reference to the lattice file format, which may evolve over time, this section will give a series of examples that show the general capabilities of the system.

7.1.1 A Simple Beamline of Components

In its basic form, a lattice file defines a set of components of different types having different physical attributes such as length, magnetic field, etc. These components are then placed successively along a theoretical reference axis for the accelerator built out of a sequence of straights and circular arcs. After the components have been defined, the order of the components along the beamline is given, as shown in the lattice file below.

```
{
A Simple Beamline of Components

Curly bracket sections may also be used as comments - in this case
be careful to respecify the component line format before use.
}

{Drift Length}
D1 0.25
D2 0.5
RD1 0.32

{Solenoid Field Length Radius}
S0 20 0.45 0.1
S1 -3.97 0.4 0.15

{RF Frequency/MHz Radius Voltage/MV Phase}
BRF1 31.4285 0.15 2.25 -90°
BRF2 201.14 0.325 0 -90°
RRFA 3.1428 0.2 0.8562 -180°
RRFB 201.14 0.2 0.5 -67.633°

{Dipole Length Angle Field HalfGap EdgeIn EdgeOut HalfWidth}
IB 1 -8° -0.123608 0.325 4° 0° 0.5

{RodSource Timespread Axial File X Y Z RodAngle}
TantalumRod 3 3 pi_plus_2G2eV.csv 0 0 288e-3 0

TantalumRod,S0,D1,S1,D2,S1,D1,BRF1,D1,S1,D1,BRF2,D2,IB,
   RD1,RRFA,RD1,RRFB;
```

Here, blocks of component definitions are introduced by a format line of the form `{Type Attrib1 Attrib2...}`, after which lines of the form `Name Value1 Value2...` will create a component of type 'Type' called 'Name' with attribute Attrib1 equal to Value1 and so on. Importantly, components are not inserted into the beamline when they are defined and the order of definition does not matter. The beamline order is defined by the comma separated list of component names at the end of the file, which is terminated by a semicolon. Components may be included multiple times, once, or not at all.

The attributes `Length` and `Angle` are special in that, regardless of the type of component they are an attribute of, they determine how the reference axis of the machine is built up. A length without an angle just extends the axis by that distance each time the component is included. A length L and angle θ will include a circular arc of curvature radius L/θ for that length, with negative angles producing reverse bends.

All values are in SI units by default, but Muon1 recognises values with units written on the end (such as nanoseconds or degrees in this file) as well as scientific 'e' notation (`288e-3`). The units may also be included in the format definitions to save writing them for each component, as MHz is here.

7.1.2 Cell Algebra

For machines with repeated beamline units, writing a straight list of all the components could be unnecessarily long or just undescriptive. To eliminate this redundancy, larger cells may be defined as blocks of the existing components with lines such as the ones shown below.

```
DFB1:QDT1,CLM,QFT1,DM3,BEND,DM3;
RCS2:8*(2*DFB1,3*DFB2,DFB0,DFB1,DF0,2*DF1);
A1:2*(ACHR,-ACHR);
FlipA1:~A1;
```

Each line starts with the new cell name, followed by a colon and the semicolon-terminated definition of that cell. This has the same format as the unprefixed list of machine components at the end of the file. In both cases, the operators of multiplication (`*`), negation (`-`) and mirror reflection (`~`) are available and can operate to any level with bracketing on the cells already defined. Multiplication repeats an identical sequence of cells the specified number of times; negation inserts the cells in reverse order and mirror reflection inserts the cells in the correct order but with the `Angle` attribute of each component changed in sign.

7.1.3 Optimisable Variables

The examples so far have specified a single accelerator. To optimise, the program requires a parameterised range of accelerators. Optimisation variables may be added by changing the definition of a component in the manner below.

```
{Solenoid Length Radius Field}
S1 [0.2,0.45] 0.1 [0,20]
```

This defines a solenoid `S1` whose length may be varied from 0.2 to 0.45 metres with a field from 0 to 20 Tesla. Its radius, however, is fixed at 0.1 metres since this attribute is specified as a single value rather than an interval range in square brackets. The replacement of a single value by a `[`min`,`max`]` bracket is all that is required to turn a point design into an optimisation ready for running.

The variables are named automatically and these names will reappear in the result files discussed in section 7.3. The variable name is the component name followed by the

first letter of the attribute name, all in lowercase, which would be `s1l` and `s1r` here. More letters from the attribute name are added as necessary if ambiguities arise.

It should be stressed that if multiple copies of `S1` are included in the beamline, they will all be identical. To have another solenoid with the same ranges for optimisation but potentially different values, which would generate additional optimisation variables, it would have to be defined separately with a different name like `S2`.

7.1.4 Example with Scripting Language

In more complex optimisations, the attributes of some components may have to be varied in a correlated way (for example to exclude impractical combinations), or the sequence of components in the beamline may change according to an optimisable parameter. Muon1 includes a C-like scripting language that can be used within the lattice files, with occurrences of the `#` symbol switching in and out of the script language on any given line. To show the various uses for this, the full lattice file for the `PhaseRotB` optimisation (a decay channel plus phase rotation system) is explained here.

```
{
PhaseRotB/~270
SolenoidsOnly (i.e. decay channel 30m long) plus 31.4MHz phase rotation
Solenoids now allowed up to 90cm long
Bugfix of PhaseRotA
May 21, 2004
}

!MINVER 4.41

{Parameter Value}
t_max 500e-9
n_decays 3
```

The initial comment block has the explanation and date in it. `!MINVER` is an optional control command that states the minimum version of Muon1 needed to run this lattice, important for the distributed project where users might not upgrade their software quickly enough. If running in distributed computing mode, the program searches for other compatible lattices to run instead. The last two lines are 'parameters', which are defined in the same way as components and determine general features of the simulation. Here they state that the simulation will terminate after 500 ns and each unstable particle (pions here) will undergo three multiple decays as described in section 2.4.1. The termination time is needed because some slow muons may be created in magnetic bottles near the solenoids and will orbit forever unless the program is stopped.

```
{RodSource Timespread Axial File X Y Z RodAngle}
TantalumRod 3 1 'pi_plus_2G2eV.csv' 0 0 [0,0.45] [0,0.5]
```

```
{Synchronous-particle t_start Energy Z}
SyncMuon #TantalumRod.Timespread?+5:0#ns 180MeV #TantalumRod.Z#
```

Two components are defined: a `RodSource` of pions, with optimisable angle and Z position, and a `Synchronous-particle` that is tracked to provide reference crossing times for the timing of the RF components. Note how the `File` attribute of the `TantalumRod` is a string, telling Muon1 to load the particles from the file `datafiles\pi_plus_2G2eV.csv`.

The attributes of `SyncMuon` contain the first example of the scripting language in use. Its start time is set to be +5 ns if the rod particles have been loaded with a timespread in emission and zero otherwise. Its Z position is set equal to the Z offset of the rod.

```
#z=0; ok=1;
#for (n=1;ok;n++)
#{
   {Solenoid Length Radius Field}
#  if (n==1) #S1 [0.2,0.45] 0.1 [0,20];#;
#  else {S[n]# [0.2,0.9] [0.1,0.4] #if (n<=4) #[-5,5];#; else #[-4,4];#;}
#  z+=S[n].Length;
#  if (z>=30) ok=0;
#  if (ok)
#  {
      {Drift Length Bore}
      #D[n]# [0.5,1] 0.5
#     z+=D[n].Length;
#  }
#}
#MaxSol=n-1;

DecayChannel: #for (n=1;n<=MaxSol;n++) {S[n] if (n<MaxSol) {#,#D[n]#,#}}#;
```

The construction of the decay channel is a more complex example, with most lines being scripted. There is a for loop and a `z` counter that is increased as solenoid and drift components are defined until it reaches 30 m in length. Each of the solenoids has optimisable length, field and (with the exception of `S1`) radius and the drifts have optimisable length too. The notation `D[n]` is a crude array implementation and would be equivalent to a variable called `D3` if `n` = 3, for example. The script variable `MaxSol` records the number of the last solenoid and then the list of components in the `DecayChannel` cell is printed by another for loop, putting drifts after all but the last solenoid.

```
{Parameter Value}
PhaseRotCells [20,40]; #// 30 in original
#PhaseRotCells=int(PhaseRotCells);

#for (n=1;n<=PhaseRotCells;n++)
```

```
#{
   {Parameter Value}
   PD#n# [0.5,1]
   {Drift Length Bore}
   PD#n#a #n==1?PD[n]*2/3-0.25:PD[n]*2/3# 0.15
   PD#n#b #PD[n]/3# 0.15
   {Solenoid Length Radius Field}
   #PS[n]# [0.2,0.9] 0.15 [-4,4]
   {RF Frequency/MHz Radius Voltage/MV Phase}
   #PRF[n]# 31.4285 0.15 [0,#(PD[n]+PS[n].Length)*2.25/1.4#] [-180°,+180°]
#}

PhaseRot:  #for (n=1;n<=PhaseRotCells;n++) {
PD#n#a,PRF#n#,PD#n#b,PS#n if (n<PhaseRotCells) #,#;
#}#;
```

The definition of the phase rotation channel starts with another use of 'parameters' in lattice files, that is to create an optimisation variable that is not tied to a specific attribute of a component but becomes a variable in the script that influences other things. Here, `PhaseRotCells` determines the number of RF-solenoid periods in the phase rotator and of course has to be rounded to an integer before use. The 'for' loop that follows is similar in principle to the decay channel one, with some additional algebra to position the RF kick component two thirds of the way through the drift, which becomes split into two drifts named like `PD3a`, `PD3b`. Another notable use of scripting is to impose an upper limit on the RF peak voltage of 2.25 MV per 1.4 m of drift length available to install the 31.4 MHz resonator into. This being based on a CERN design [83] where the large resonator also enclosed the solenoid, the solenoid's length is added to the drift available for this purpose.

```
{Finish-aperture radius MinEnergy MaxEnergy}
END 0.5 #180-23#MeV #180+23#MeV

TantalumRod,SyncMuon,DecayChannel,PhaseRot,END;
```

The final component to be defined is a `Finish-aperture` to collect the successful muons that have finished the channel in the kinetic energy range 180 ± 23 MeV. The file ends with the definition of the entire accelerator in terms of the cells defined already.

7.2 Figures of Merit

The transmission of muons, as a percentage of the particle weight of the original pions, is used to score an accelerator design once the simulation has finished. A simulation will finish either when there are no more active particles remaining, or the `t_max` time limit has been reached. In the first of these cases, all the particles will then be either marked as 'lost' or 'finished', whereas in the second some will still be 'active'. From the point of

view of the simulation, 'lost' and 'finished' particles are treated exactly the same except with a different label: they both have stopped moving.

So how does a particle acquire the special 'finished' status? Currently this is done using the `Finish-aperture` component as at the end of the example in the previous section. It produces a disc of the given radius, perpendicular to the beamline, that acts as a blockage. If energy bounds are not specified, then all particles hitting the disc are stopped and marked 'finished'. If, as in the example, there is an energy range, then particles outside that range are also stopped, but marked as losses.

At the beginning of the simulation, Muon1 normalises all the particle weights so that $\sum_i w_i = 1$. Thus after finishing, the percentage transmission score is simply

$$T = 100 \sum_{\text{finished muons } i} w_i.$$

However, for more difficult optimisations, a random element of the design space may be so poor that almost all such designs will give a zero score. It is useful to be able to distinguish between such designs to see which are nearer to being successful, so that the optimiser may climb out of the area of poor designs on its own without having to be seeded with an already good design. Thus two additional scores are defined to deal with the cases of beam not being transmitted through the whole design and beam being transmitted but mostly of the wrong energy.

The figure for a design that transmits less than a proportion $\epsilon = 0.1\%$ of the beam is defined by finding the distance along the beamline where at least ϵ weight of muons survived. That is the largest distance s, measured along the beamline axis, for which the weight of stopped muons with $s_i \geq s$ is at least ϵ. Call this distance s_ϵ and the total length of the beamline l. Then the distance score is defined as

$$D = s_\epsilon / l.$$

It ranges from near zero where almost all the particles stop in the early accelerator, to nearly one when the fraction of surviving muons is only reduced to below ϵ right near the end.

The figure for designs that transmit less than ϵ of the particles in the right energy range $[E_{min}, E_{max}]$ is determined from the energies E_i of the muons that hit the `Finish-aperture`. The ratio by which the energies are wrong, as a weighted average, is given by

$$W = \frac{\sum_{E_i<E_{min}} w_i(1 - E_i/E_{min}) + \sum_{E_i>E_{max}} w_i(1 - E_{max}/E_i)}{\sum_{\text{muons on aperture}} w_i}.$$

This can range from values near zero when most of the muons have almost the correct energy, to values of nearly one when the majority of the energies are very far outside the specified range.

The above figures T, D and W are blended together with linear ramps to make a smooth overall score S that increases with the performance of the design. If $T/100 \geq \epsilon$,

the transmission score is assumed to have enough statistics to be used on its own, whereas in the region $\epsilon/2 < T/100 < \epsilon$ it is linearly blended into the 'fallback' score S_1, as follows:

$$S = \begin{cases} T & \text{when} \quad T/100 \geq \epsilon \\ \frac{T/100-\epsilon/2}{\epsilon/2}\ T + \frac{\epsilon-T/100}{\epsilon/2}\ S_1 & \qquad\quad \epsilon/2 < T/100 < \epsilon \\ S_1 & \qquad\quad T/100 \leq \epsilon/2 \end{cases}$$

If more than ϵ weight of muons survive the channel, regardless of their energy, the fallback score rates the design between 0 and -1 depending on how wrong their energies are. If fewer than $\epsilon/2$ get to the end, the design is scored as $D-2$, which ranges between -2 and -1 depending on how far the muons survived down the channel. In between, linear interpolation is again used. Let f be the raw transmission fraction of muons, disregarding energies (so $f \geq T/100$). Then the fallback score is defined as:

$$S_1 = \begin{cases} -W & \text{when} \quad f \geq \epsilon \\ \frac{f-\epsilon/2}{\epsilon/2}(-W) + \frac{\epsilon-f}{\epsilon/2}(D-2) & \qquad\quad \epsilon/2 < f < \epsilon \\ D-2 & \qquad\quad f \leq \epsilon/2 \end{cases}$$

The combined score S as defined above is the one used in the Muon1 program when it outputs results. Note that for any value above 0.1 it is just equal to the muon transmission in percent. Lower yielding designs are given scores in a range down to -2.

7.3 Muon1 Result (and Queue) Files

In order for the optimiser to learn from previously simulated designs, the variables and score from each simulation are logged to one or more result files. Generally these will be `results.txt` and `results.dat` in the same directory as the Muon1 executable. All result files are sequences of records, one added for each simulation, which have the same two-line format as the example below.

```
#gen=8;d10l=075;d11l=237;d12l=013;d13l=000;d14l=011;d15l=000;d16l=024;
```

...

```
s8l=999;s8r=999;s9f=978;s9l=999;s9r=999;tantalumrodr=000;tantalumrodz=660;
0.365180 (368.8 Mpts) [v4.41f] {FD19E68427264FC7F20D451C} <PhaseRotB>
```

The first line can be very long so has been split for clarity. It contains a series of *variable*=*value*; pairs with the variable names coming from the lattice file as described in section 7.1.3. The value has been (linearly) normalised to three digits with 000 corresponding to the lowest and 999 the highest allowable value in the range. Accuracies of more than three significant figures were considered to be difficult to engineer, so by restricting the accuracy this representation prevents some degree of over-optimisation. Variables whose name begins with a `#` character are not from the lattice, they have been

added for other logging purposes, for instance `#gen` is automatically added to record which genetic algorithm mode produced the design (see section 7.4 for details). Hashed variables are standard integers that are not limited to three digits.

The second line always starts with the score that the design specified by the first line achieved. It must also contain the name of the lattice in angle brackets, `<PhaseRotB>` here, which specifies that the variables on the first line should be used with the lattice file `lattices\PhaseRotB.txt`. Other data is also usually supplied: in round brackets the amount of calculation required for the simulation in mega-particle-timesteps (Mpts) and the version of Muon1 in square brackets. This is useful for logging the overall amount of CPU activity on a distributed network of computers as well as benchmarking, while the version code can be used to filter out erroneous results in the case of a buggy Muon1 release. The hex code in curly brackets is a checksum produced by the program to make it more difficult to flood the distributed project with 'fake' results. The user may also add their own comments at the end of the second line, which are ignored by the program.

Normally on completion of a simulation, the two-line result is appended to both the files `results.txt` and `results.dat`, with the files being created if they do not yet exist. The reason for this duplication is that when running as part of the distributed computing network, `results.txt` contains only the buffer of results that have not yet been sent and is deleted after a successful upload. `results.dat` contains a complete history of results produced and may even be augmented with extra results generated on other machines, to serve as a knowledge base for the optimiser to generate new designs from.

A third file `queue.txt` may also contain results in a similar format and this has two purposes. Firstly, it acts as the method for a user to input specific designs (from a lattice's optimisation range) to test. These will specify the first line with the normalised values for the variables and the lattice tag on the second line, together with the string `TEST` in place of the score, since it has not been calculated yet. This form is shown in the example below, although the first line will often be much longer in reality.

```
s8l=916;s8r=896;s9f=895;s9l=909;s9r=976;tantalumrodr=000;tantalumrodz=446;
TEST <ChicaneLinacB>
```

If a lattice has no optimisable variables, it is not worth preparing a `queue.txt` file in this way before running Muon1 because specifying the lattice on the command-line will provide sufficient information[†]. The second purpose for the queue file is to recheck high-scoring results for both statistical accuracy and reliability against computer errors. Specifically, if a result has the highest score achieved so far for its lattice, Muon1 will run it an additional four times, storing the intermediate results in the `queue.txt` file and then finally add a result with an averaged score (excluding the highest and lowest) to `results.txt` and `results.dat`. In these cases, the intermediate result may have a comma-separated list of scores in place of the single score normally found at the start of the second line. Whatever the usage mode, the top result in the queue file is processed first and removed when completed, with the rest of the file being moved up.

[†]The reader may wonder what happens if an optimisable lattice is run from the command-line without a queue file. The answer is that Muon1 is always in continuous optimisation mode by default, so will choose either a random design (if `results.dat` does not exist) or continue optimising the designs based on the work accumulated in `results.dat` so far.

7.4 Genetic/Evolutionary Algorithms

When given an optimisable lattice, Muon1 has the task of choosing values for a potentially large number (often hundreds) of variables. Genetic and evolutionary algorithms are suitable for problems of large dimensionality where the optimum can be found without necessarily being given a nearby starting point. This is because they effectively perform a search on the design space that starts randomly and becomes focussed in higher-yielding areas as accumulated results reveal where these areas are. 'Brute force' searches where a grid in the search space is fully explored are not practical for high-dimensional problems because even if each variable is only sampled at 2 values, there are 2^D combinations for a D-dimensional grid, which exceeds 10^{30} with $D = 100$, for example. This means that 'intelligent' forms of brute force searching that subsequently use finer grids on areas with good results also cannot be used because even the initial coarse grid is too time-consuming to search. Another common optimisation method is to start from an initial design and evaluate the sensitivity of all parameters to small variations, then adjust these in the directions that increase the score. Unfortunately when the figure of merit is particle yield from a tracking simulation, the score has statistical noise that overwhelms the small differences in score produced by the small variations used to gauge the sensitivity.

The genetic algorithm terminology originated in early computer simulations of biological evolution [84]. These first examples of genetic algorithms followed a very close analogy with DNA replication, generating a 'population' of initially random binary strings (genomes), then performing artificial selection (as opposed to *natural selection*) to eliminate a large proportion of them before the remaining good strings were allowed to 'mate' in a process that mixes blocks of the parent binary strings together (genetic crossover) and occasionally inverts some of their bits (mutation). These offspring then form the population for the next generation and the process repeated many times, with the removal of the least-good designs biasing the distribution of strings towards areas in which the higher scoring designs are to be found.

In Muon1, the 'genome' is just the first line of the result as described in section 7.3, containing the normalised values for all the variables that specify the particular accelerator design within a lattice's optimisation range. 'Artificial selection' is performed on the grounds of higher scores (calculated as in section 7.2) being better and 'mating' by a variety of methods including analogues of the original mutation and genetic crossover.

However, the original formulation of the genetic algorithm is often inefficient in real engineering problems and requires some adaptation as reviewed in [85]. For Muon1, the presence of discrete generations means the designs and results from each generation are destroyed apart from their influence on the gene pool of the next generation. When each result is scored on the basis of a simulation that can take an hour or more to complete, this is considered wasteful. Instead, all the results are logged permanently but operations that generate new designs, such as mutation or crossover, are based on a subselection highly biased towards the higher scoring designs. This means low-scoring designs remain in the result log but hardly ever procreate.

The biased selection algorithm in Muon1 works as follows. A real number $\alpha \in [0, 1.25]$ is chosen uniformly at random and is used to define the subsample size $s = \lfloor N^\alpha \rfloor$ where

N is the total number of results in the `results.dat` history file. Random results are examined from this file s times over and the one having the highest score is selected for use. Methods such as crossover require two input genomes so will call this algorithm twice, or more if the second genome comes back the same as the first, until two differing ones are selected. The advantage of allowing values $\alpha > 1$ and hence $s \geq N$ is that nearly all of the results will be examined, so there is a finite probability of picking the very highest result in the file, which does not decrease appreciably with N. Allowing values of α near 0 means it is still possible for merely 'average' results to get picked, although with a reduced probability.

A second major modification to the binary genetic algorithm is needed because it has no concept of continuously-varying parameters, instead it would treat their floating point representation as just another binary string and produce 'corrupt' values that may be nowhere near either of the two parent values after a bit-wise crossover. Although this could be seen as introducing a little more mutation into the system and therefore harmless, Muon1 takes the approach of using continuous parameters and analogous operations to crossover (where values themselves remain intact, like individual bits) or mutation (where increments are chosen from a continuous distribution of values). This continuous parameter system also naturally allows other ways of generating new genomes, such as taking a weighted average of two sets of variables, an operation that has no analogue using binary strings.

With these modifications, Muon1's algorithm can more accurately be described as part of the broader class of 'evolutionary algorithms', which use the concept of learning from a pool of past results but do not impose a discrete genome analogous to DNA base pairs.

The full implementation contains several ways of generating new results from those selected from the `results.dat` file. As mentioned in section 7.3, each result is labelled via the `#gen` field to denote how it was created. These different methods (known as 'trial types') and their `#gen` values are described in the list below.

- `Random` (`#gen=0`) does not use any previous results and sets all variables to uniform random values. This is called when no results are in the `results.dat` history file, or with a decreasing probability $N^{-0.4}$ when it contains N results.
- `Mutate` (`#gen=1`) and `Crossover` (`#gen=2`) function roughly as described in the original genetic algorithm.
- `Interpolate` (`#gen=3`) takes two results and produces a weighted average of their parameters. A random number on the [0,1] interval determines how much weight is given to the first genome with the balance made up by the second.
- `Extrapolate` (`#gen=4`) is the same but uses numbers outside the [0,1] range to extend beyond one of the results' locations in design space.
- `MuSpherical` (`#gen=5`) produces a mutation according to a spherically symmetric distribution of offsets, if the variables are considered the axes of the design space with the Euclidean metric.
- `MuOne` (`#gen=6`) is like `Mutate` but only changes the value of a single variable.

- `Extreme` (`#gen=7`) is designed for the common situations when the optimal values of many parameters are at the extremal end of the allowed range (e.g. highest magnetic or accelerating field). It moves some of the variables towards the end of the range.

- `LocalGrad` (`#gen=8`) estimates the local gradient of the score function from a cluster of the nearest results to a selected one and produces a suggested improved design by moving 'uphill' a certain distance from there.

- `TopoSmooth` (`#gen=9`) uses the numbering that often appears in variable names to infer which apply to adjacent components (numbers differing by one) and smooths out each variable's value by averaging it with that of its neighbours.

- `CellDouble` (`#gen=10`), `CellDelete` (`#gen=11`) and `CellCommute` (`#gen=12`) also use the numbering of the variable names to perform the functions of duplicating a beamline cell, deleting a cell and swapping (commuting) two adjacent cells with each other.

The effect of adding these extra trial types to the algorithm was investigated with two benchmark problems.

7.4.1 Sine Wave Interpolation Test

The first benchmark problem was designed to have a figure of merit that depends on a large number of variables in a smooth but correlated way. A suitable problem is the interpolation of a sine wave with a linear spline $L(x)$ using the variables $(x_1, y_1) \ldots (x_{10}, y_{10})$ with $x_n \in [0, 2\pi]$ and $y_n \in [-1, 1]$ as the positions of the interpolating nodes. The score would then be

$$10 - \int_0^{2\pi} (\sin(x) - L(x))^2 \, \mathrm{d}x,$$

so that better spline approximations increase towards 10. Since the performance of the optimiser in a single run would depend on the random values chosen during various operations, a program was written to run the optimiser many times (with different random seeds) on the same problem and accumulate statistics on its average behaviour. Figure 7.1 shows this during operation for the sine interpolation problem.

After running the optimisation test problem 100 times, each of which was terminated either when the score reached 9.99 or when 10000 results had been generated, some average values (including error bars) were derived that show the true performance of the optimiser regardless of the random numbers used. The two most important statistics collected were the percentage of times the optimisation reached the target score before the 10000 result cutoff and the average number of iterations taken to get there in the cases where it did succeed. These are plotted in figure 7.2 for all $2^6 = 64$ combinations of the trial types with `#gen=1` through `6`. When multiple trial types were used, they were selected with equal probability for every non-random result. In the case where no trial types were used, only random results were generated, explaining the zero percent success rate in that case

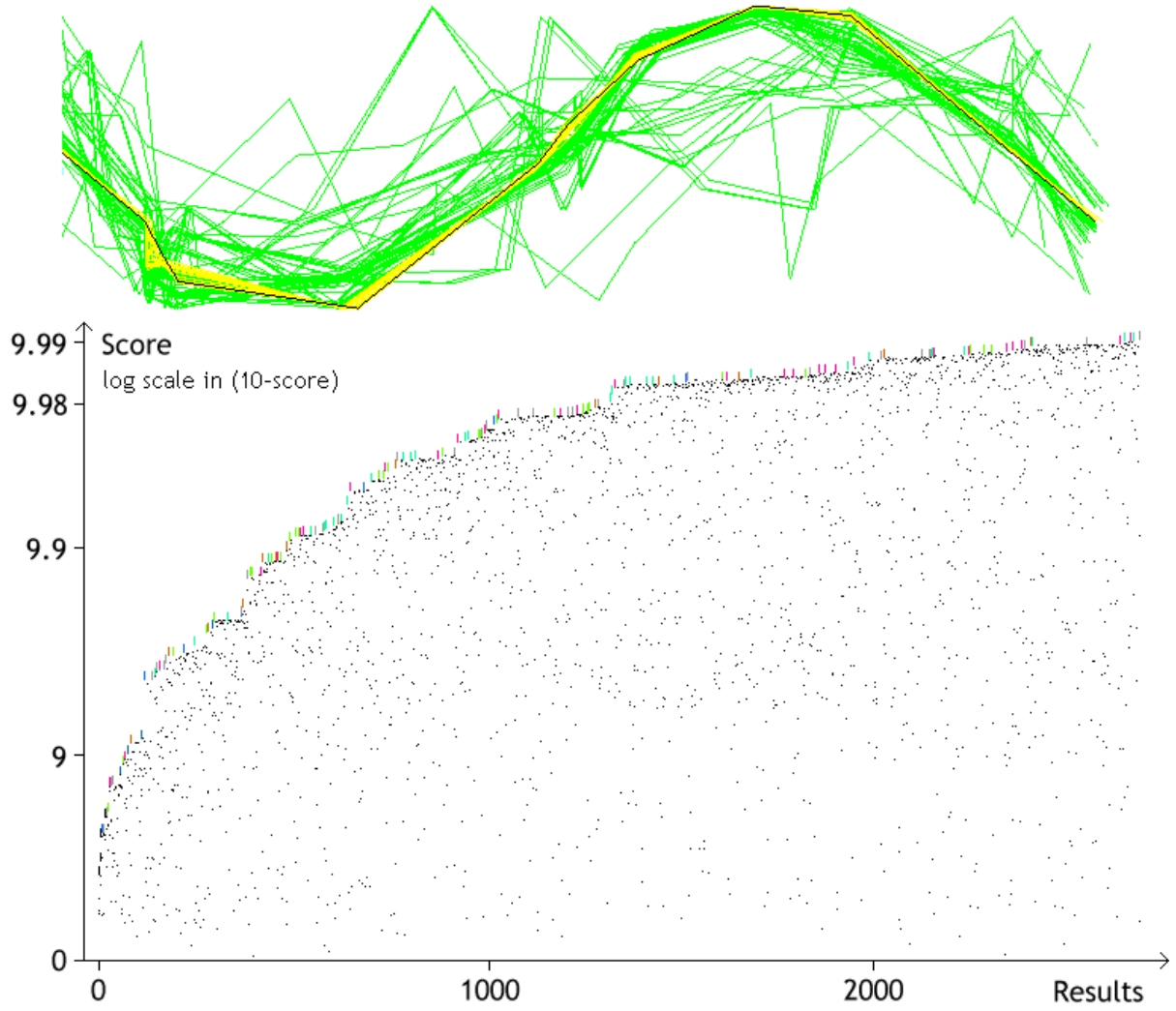


Figure 7.1: Genetic algorithm test run of the sine wave interpolation benchmark problem. The green graphic at the top shows successive best linear spline approximations to the sine wave with high scoring lines (≥ 9.98) in yellow and the final result (first above 9.99) in black. Below is a graph of the generally increasing scores of the results over time.

(the optimisation problem was *not* so easy that a million random guesses would produce a success!)

The success rate and average convergence time data can be amalgamated into a single figure of merit to represent the number of optimiser convergences expected in a million results ($10^6 *$ (success probability)/(average time)) and this is what is shown in figure 7.3. The general trend is that optimiser performance is patchy and often failing to converge when only a few trial types are combined and more robust when there are many. Crossover and interpolation get nowhere without adding other types because they only produce results 'between' those few random ones already produced. Adding a mutate type to anything guarantees at least some level of convergence, though spherical mutation is generally poor and single-parameter mutation is so good it almost beats all other combinations on its own (though with a big error bar). However, most mutation combinations are moderately enhanced by adding crossover or extrapolation.

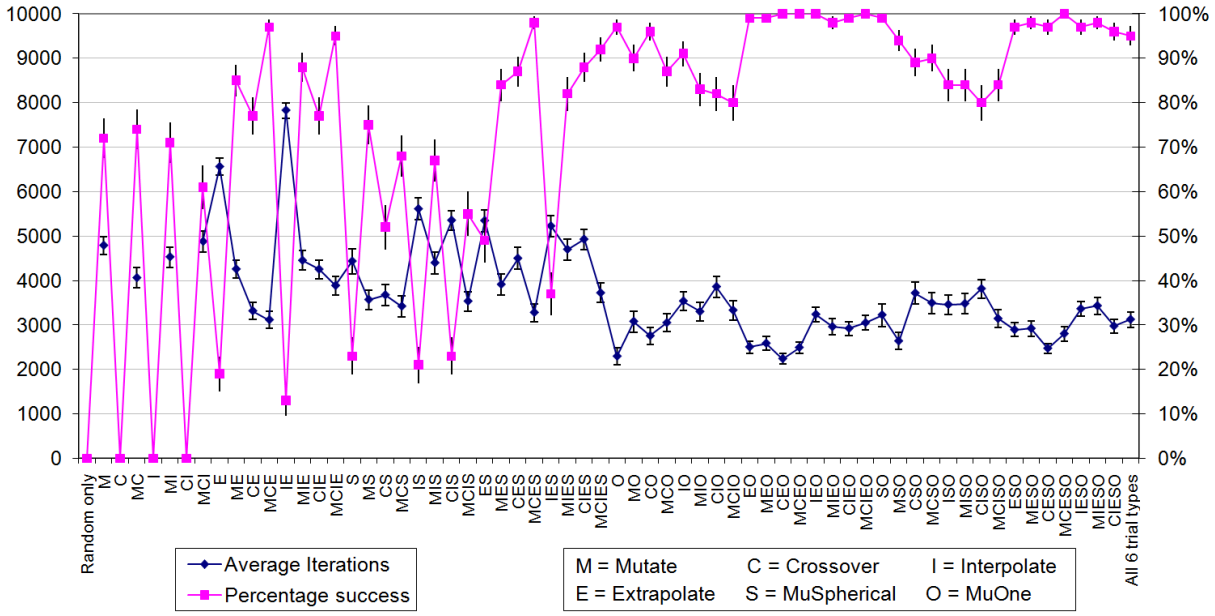


Figure 7.2: Success rate and average convergence time for the optimiser on the sine wave interpolation test problem with all possible combinations of the first six trial types being switched on and off. Error bars are one standard deviation.

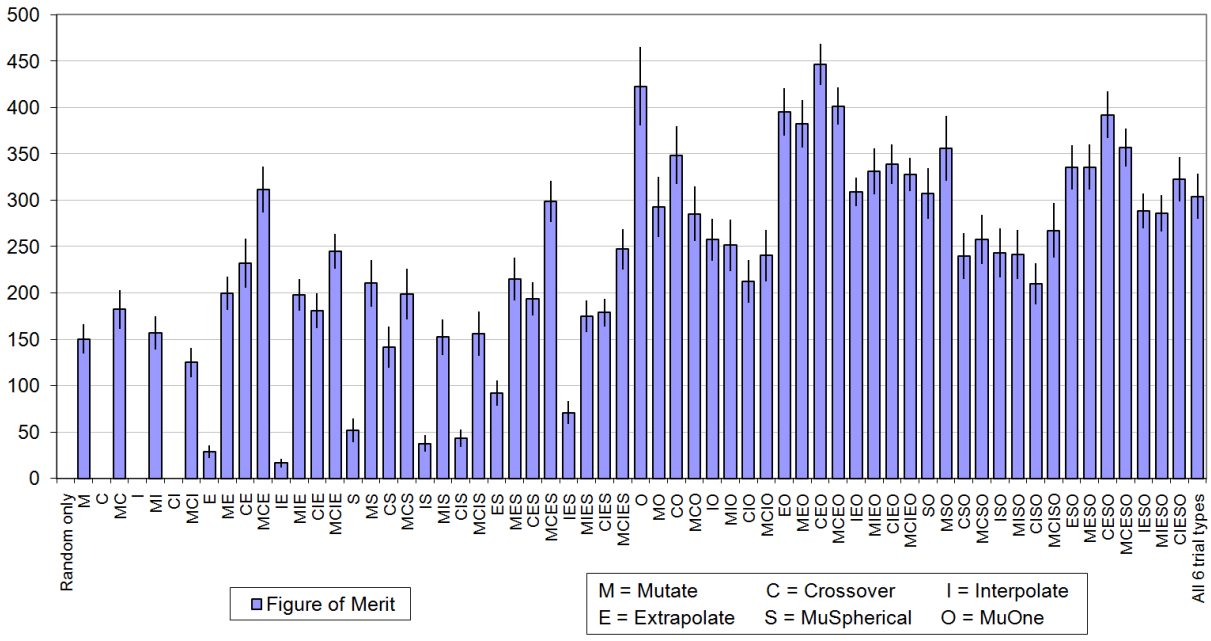


Figure 7.3: Figure of merit for optimiser performance running on the sine interpolation test problem, for each possible subset of trial types used (1σ statistical errors shown).

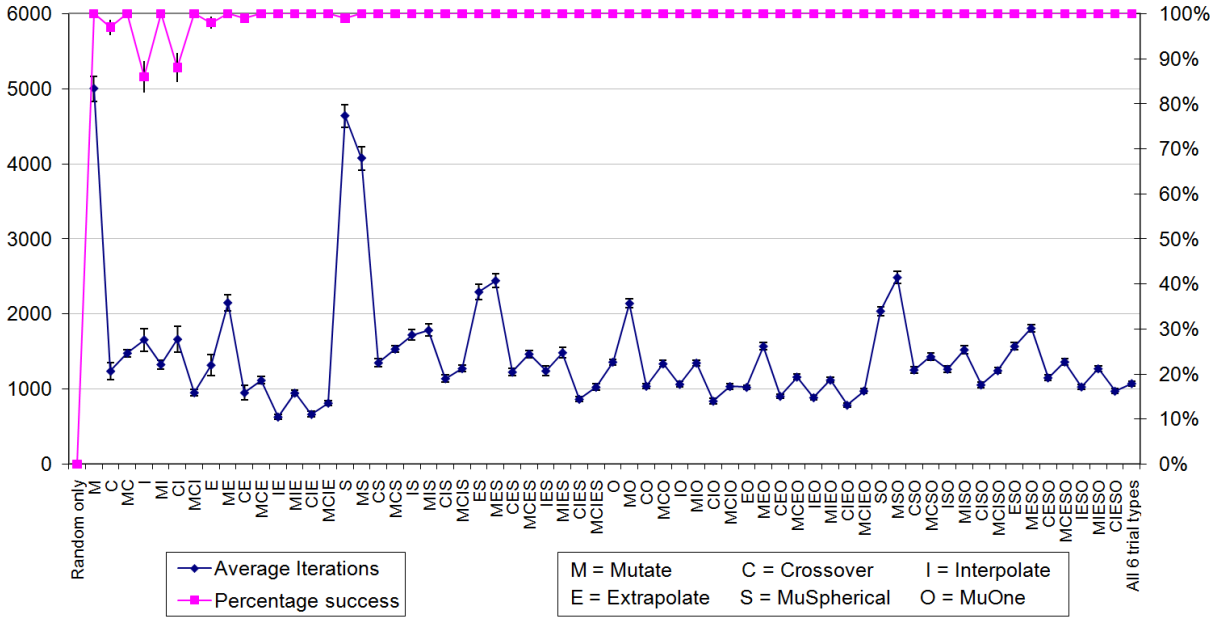


Figure 7.4: Success rate and average convergence time for the optimiser on the sequential parameters test problem, for each possible subset of trial types used (1σ statistical errors shown).

7.4.2 Sequential Parameters Test

A second benchmark problem was devised to test the optimiser in rather less smooth situations, including discontinuities and situations where some parameters become 'slack' unless activated by the proper choice of others. The latter situation of sequential parameters may occur when good choices in the early part of a beamline are required for the particles to propagate far enough to be affected by the variables controlling components further on. To represent this, some 'goal' values $g_1 \dots g_{10}$ are chosen randomly by the computer and kept secret from the optimiser, which will make guesses as vectors $(x_1, \dots x_{10})$ with $x_n, g_n \in [0, 1]$. If $|x_n - g_n| \leq 0.05$ for all n, the score is 10 and the successful optimisation is terminated. Otherwise let x_k be the first value such that $|x_k - g_k| > 0.05$ and put the score equal to $k + 1 - |x_k - g_k|$. This increases from slightly more than zero for guesses where x_1 is totally wrong, increasing fractionally as the guess becomes more accurate and jumping above each integer n once x_n becomes close enough to the goal. The success rates and average optimisation iterations for this second problem are plotted in figure 7.4 and show a significantly different pattern to that of sine wave test problem, though sharing the same feature that combining more trial types is generally better.

A single figure of merit can be produced for the sequential parameters test problem in the same way as for the sine interpolation test. To show the situations in which the optimiser performs well whichever sort of problem it is given, the geometric mean of the figures of merit for both problems is plotted in figure 7.5, the geometric mean being chosen so that a complete fail on either problem will ensure a zero score overall. The high performance of the `MuOne` trial type alone has become less significant than it was in the sine wave test, with the best overall performance coming from the combination of it with crossover and extrapolation. However, the mix of all six types still performs three-quarters

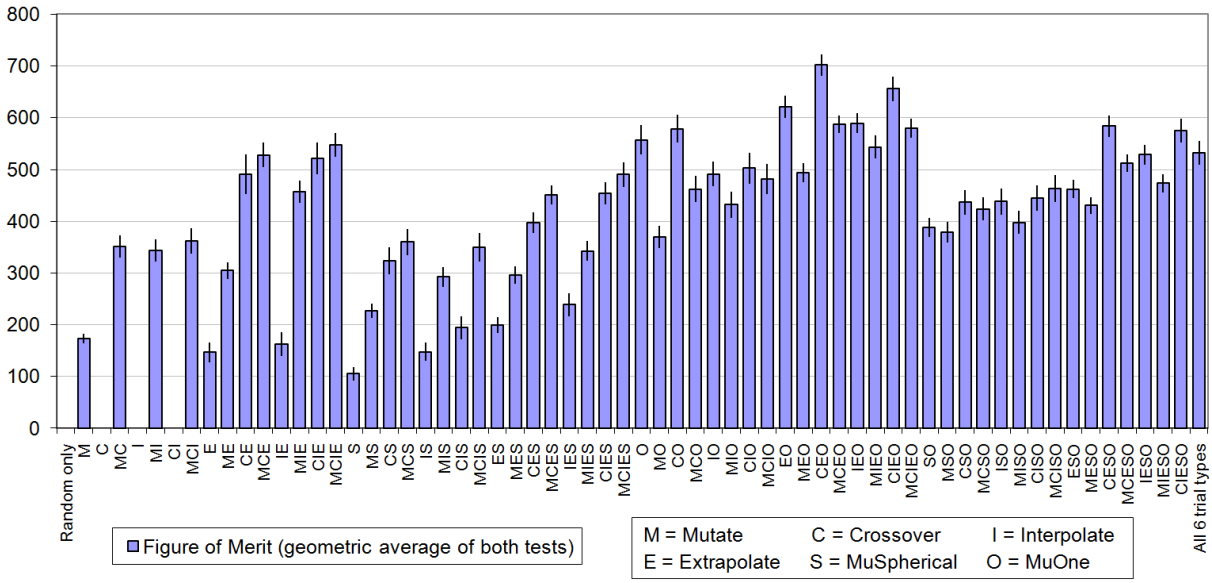


Figure 7.5: Combined figure of merit for optimiser performance across two different test problems, with 1σ error bars, as a function of the set of trial types used.

as well as anything else and comes with the insurance that if another, totally different, sort of problem was attempted that responds well to, for example, `MuSpherical`, then that algorithm would still be available. Muon1 by default runs with all these trial types active, together with `LocalGrad` and some of the 'topological' trial types but with a lower probability. The weighting of each type can also be customised in the `config.txt` file by the user.

An illustration of the distribution of results generated by having all these different methods of evolution active at the same time is figure 7.6. This is projection of one user's `results.dat` file onto the two principal axes of variation in the design space, when running the `SolenoidsTo15cm` lattice that optimises of the pion decay channel. Some diverging evolutionary pathways are clearly visible by virtue of the fact the user had several computers working independently for a period. The progression of another such evolution will be examined more closely in the next subsection; the resultant accelerator designs for all optimisations will be described in chapter 8.

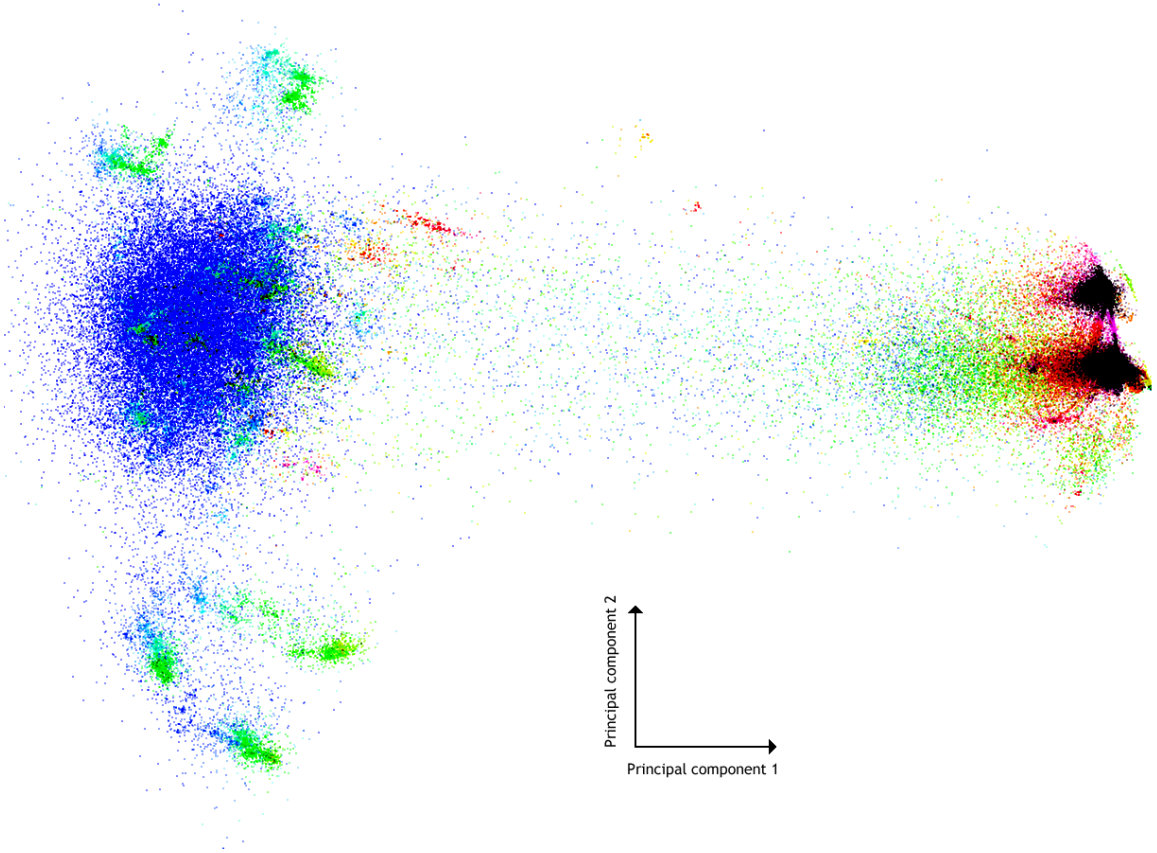


Figure 7.6: Projection of a user's result evolution through the design space. Principal component analysis (PCA) was performed to determine the two axes of largest variance in the results in the high (∼100) dimensional design space, which were used to plot each result as a point in two dimensions. Black/purple indicates the highest scoring results, through to blue for the lowest. Notable features include: clustering around the highest scores; bridging between two particular high scoring points in search of an optimum between; a diffuse blue ball of randomly-generated results at the left; and six separate pathways optimising out from the blue ball, caused by the program being run from scratch on six separate computers for a period, each one going its separate way.

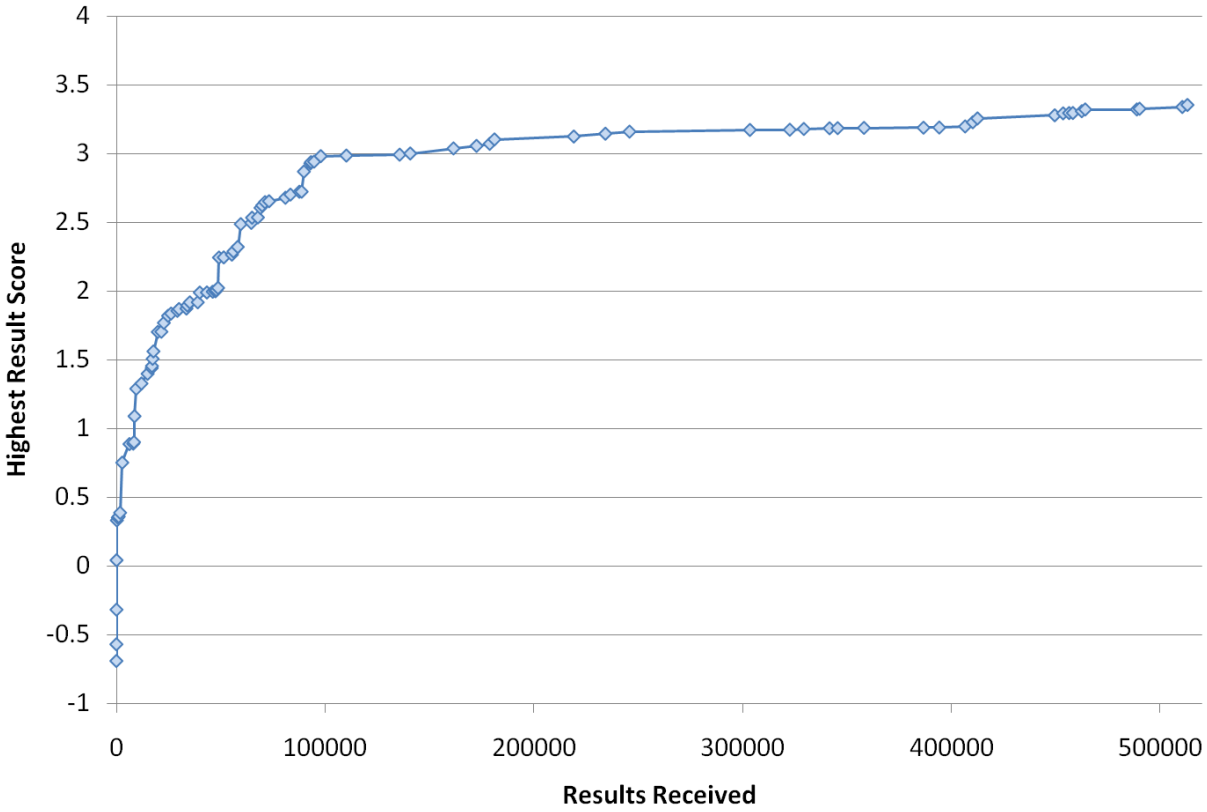


Figure 7.7: Progression of successive best scores attained by the `PhaseRotEb1` lattice.

7.4.3 Example Lattice Optimisation

The evolution of all the parameters in the `PhaseRotEb1` lattice is plotted in this subsection. How all these lattices differ from each other will be discussed in chapter 8 but for now it suffices to say that `PhaseRotEb1` optimises the decay channel and phase rotation sections of the muon front end, aiming for muons transmitted into an energy window of $180 \pm 23\,\mathrm{MeV}$. This optimisation is particularly easy to display because the later sections of the decay channel and the whole phase rotation are made from identical cells, which cuts down the number of variables significantly. Figure 7.7 plots the progress of the optimisation over the two month period it was running on the distributed network. Only the successive best scores received by the central server are plotted, so the very earliest results with scores between -2 and -1 are missed out, as by the time the participating computers uploaded their first batch of $\sim$50 results they had already achieved scores above -1.

The rate of increase in the muon yield slows down over time because the higher a score already is, the more difficult it is to generate a result that exceeds it. Despite saving on independent variables in later parts of the channel, `PhaseRotEb1` is still an optimisation of 48 variables. This means any sort of exhaustive search or guarantee that the eventual design is the best one in the whole design space is out of the question, since the design space consists of $1000^{48} = 10^{144}$ points. The optimisation only sampled 5.89×10^5 of these, while finding a design much better than the same number of random (or systematically spaced) samples would, though it is very much unknown how many points exceed the yield found.

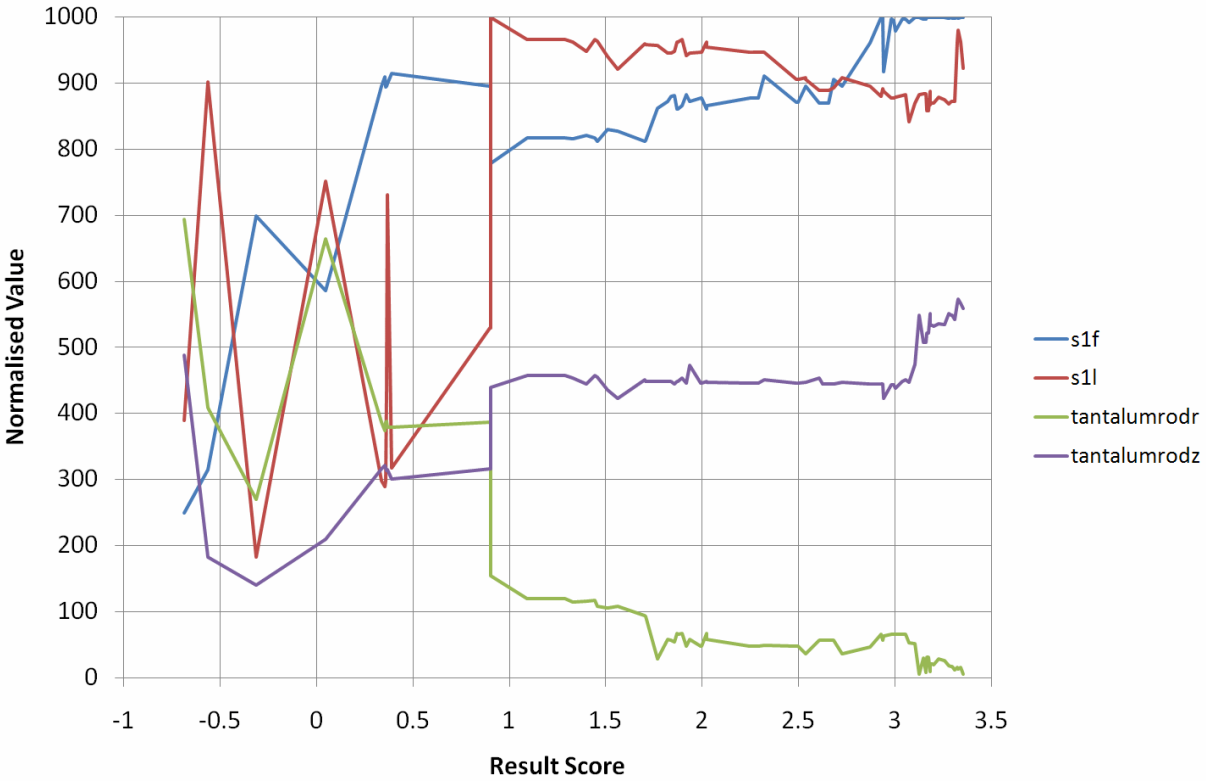


Figure 7.8: Evolution of variables concerning the target rod and its enclosing solenoid (`S1`) in the `PhaseRotEb1` lattice. The 0–999 range represents 0–20 T for the `S1` field, 20–45 cm for the `S1` length, 0–0.5 radians for the rod rotation angle and 0–45 cm Z displacements of the rod centre into the solenoid (measured from the front coil face).

To see how the variables had to be tuned to achieve increases in score, they are all plotted in groups on the seven graphs that follow (figures 7.8–7.14). Score is a more significant variable to plot against than time as it does not depend on the speed of computer calculation in any way.

Different types of variables are optimised while the score is in different ranges. To achieve a score above zero, at least 0.1% of the beam must finish within the correct energy range, so the first parameter to become close to its final optimised value is the RF phase used in the phase rotation on figure 7.10. For the transmission to reach 1%, the fields of the solenoids in the late decay channel and phase rotation (purple and black on figure 7.11) were given the same sign and the configuration of the target rod was changed so that the solenoid had a high field and length with the rod being more parallel to its axis. The discontinuity in many of the parameters at once around a score of 0.9 could be the optimiser jumping out of its current local minimum to another, as the evolution of results in another minimum overtook the first. Evolutionary algorithms are often put in the category of 'global optimisers' for this sort of reason, though as pointed out before, a truly global optimiser for completely general score functions does not exist.

Between 1 and 2% yield the main adjustments were to the voltage in the RF cavities and the length of the phase rotation section (figure 7.9), which were both increased. At a score of about 2.2 on figure 7.13 another jump occurred in which all the drifts in the

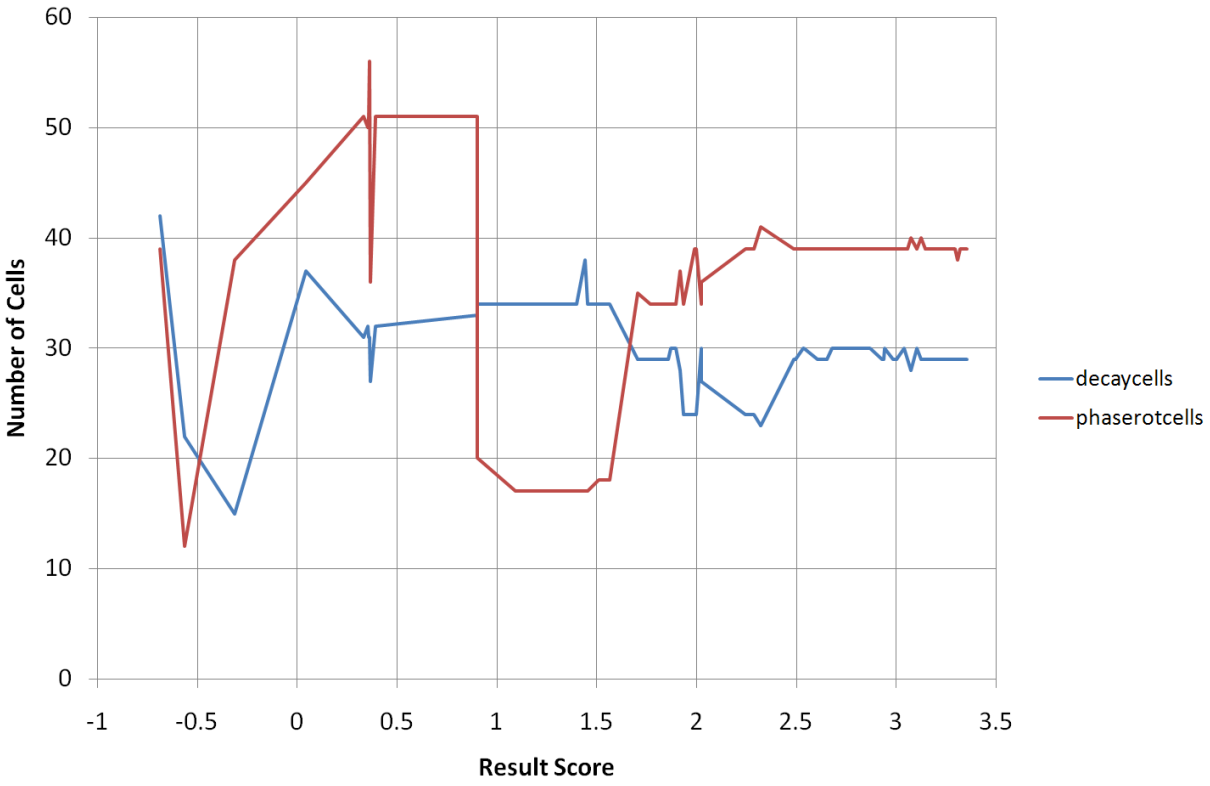


Figure 7.9: Evolution of the number of cells in the decay channel and phase rotation sections of the `PhaseRotEb1` beamline.

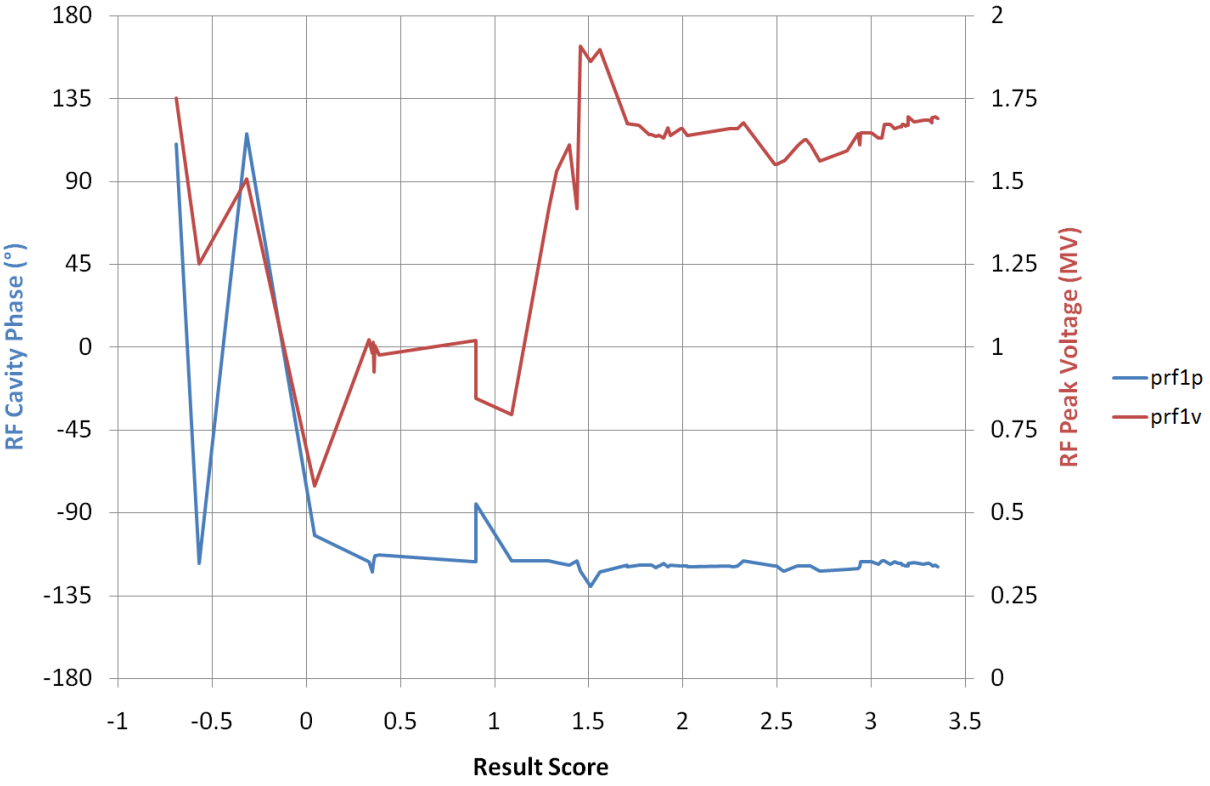


Figure 7.10: Evolution of the RF parameters in the phase rotation section of the `PhaseRotEb1` beamline.

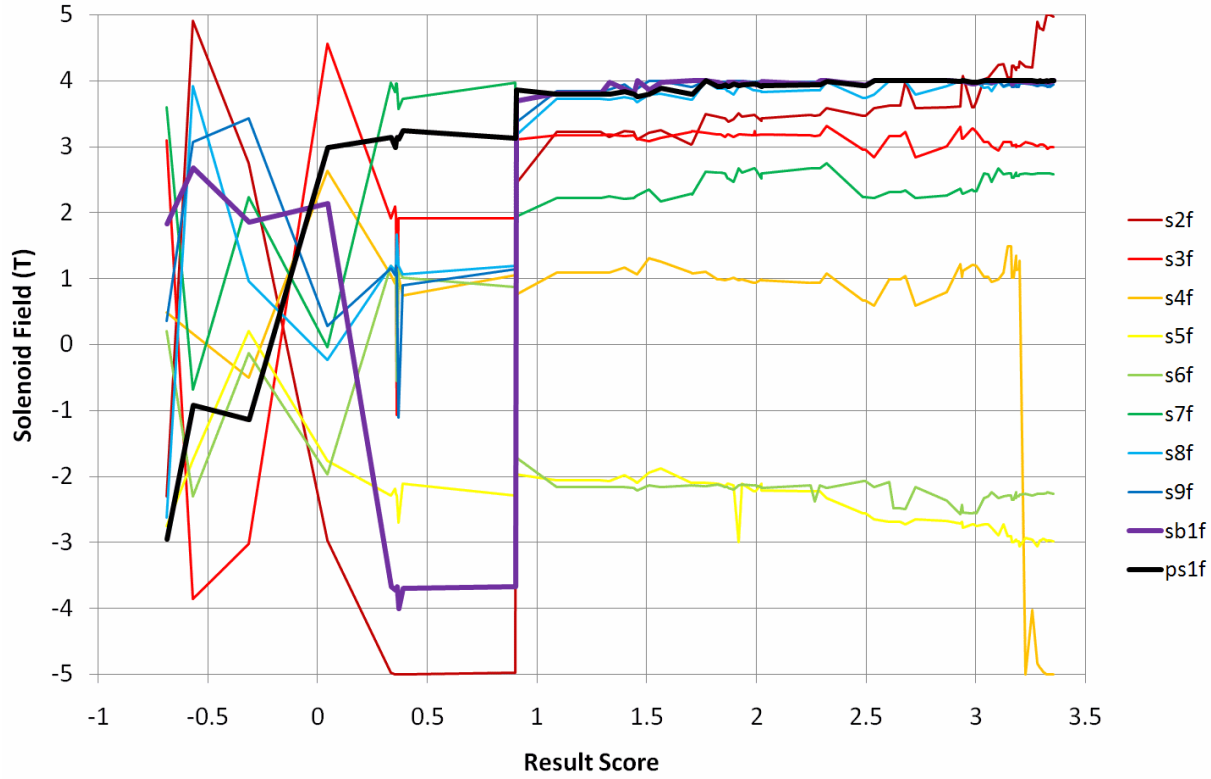


Figure 7.11: Evolution of solenoid fields in the `PhaseRotEb1` lattice.

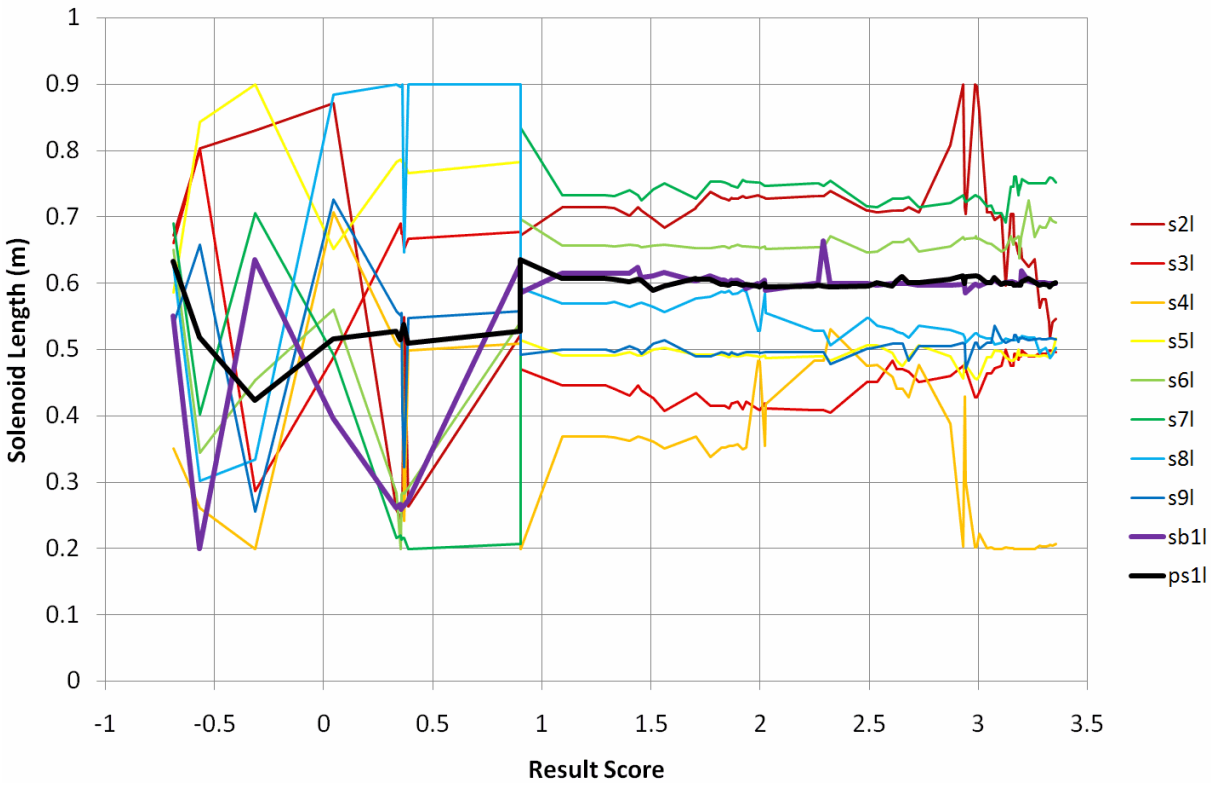


Figure 7.12: Evolution of solenoid lengths in the `PhaseRotEb1` lattice.

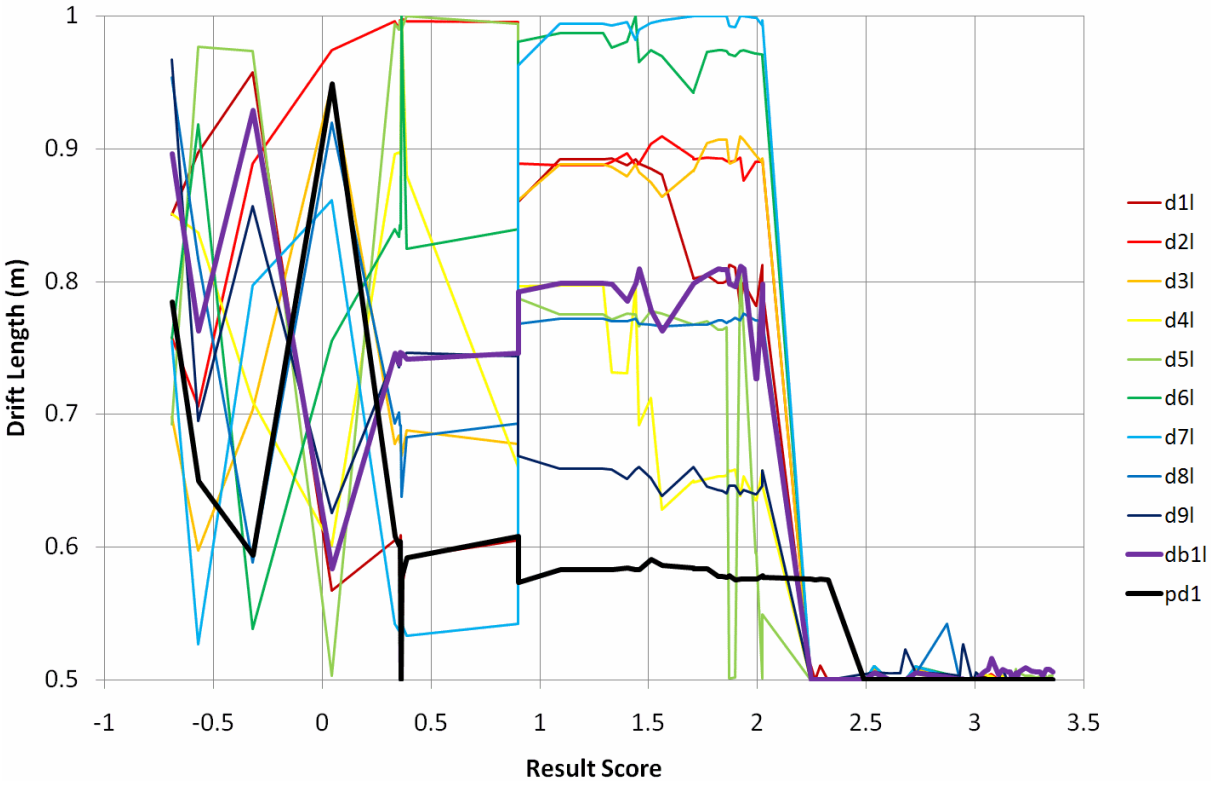


Figure 7.13: Evolution of drift lengths in the `PhaseRotEb1` lattice.

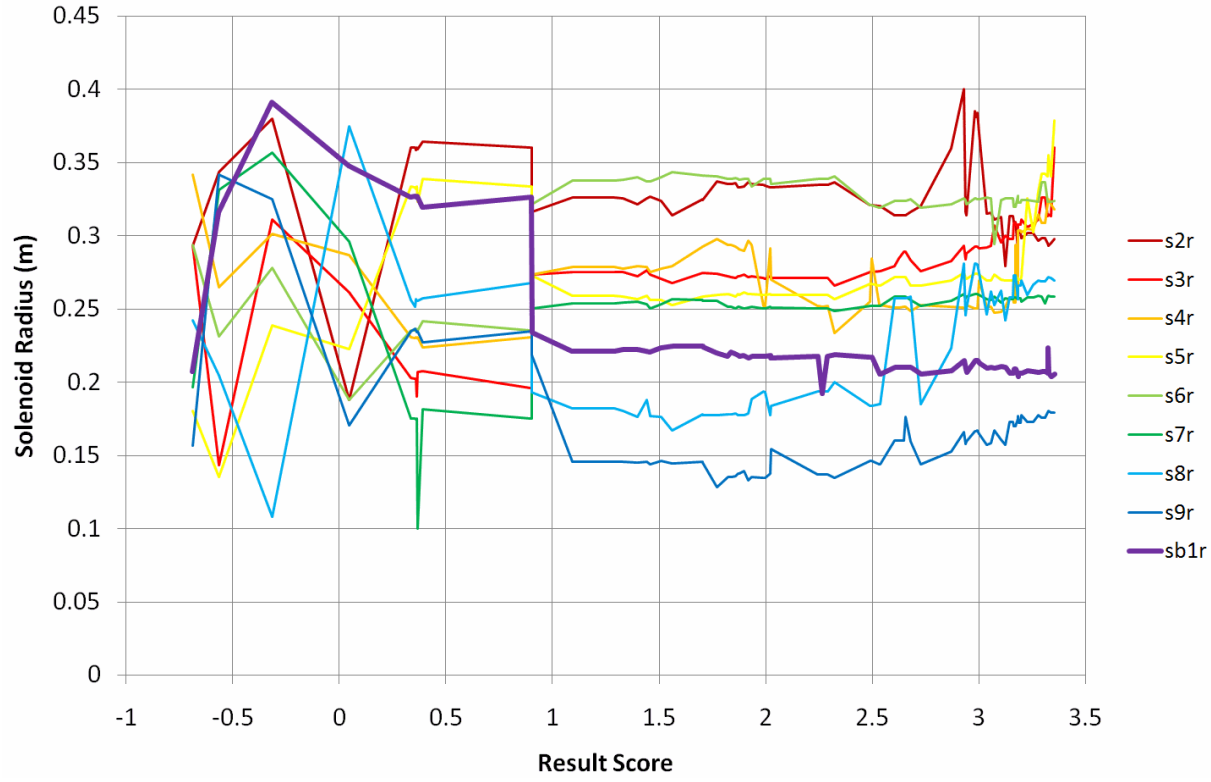


Figure 7.14: Evolution of solenoid radii in the `PhaseRotEb1` lattice.

decay channel were shortened to the minimum allowed length of 0.5 m. From here up to the final high of 3.355% most of the optimisation is concerned with gradually adjusting various parameters, in ways that appear to become quite correlated above 3%; one clear example is the anticorrelation of the rod Z offset and rotation in figure 7.8. The only further 'global-type' jumps are in solenoid number 4's length at 2.9% (figure 7.12) and field at 3.2% (figure 7.11), changes that presumably improve the optics accepting the beam from the target. These could likely be the result of a `MuOne` trial type in which only one parameter is changed, in this case to a value with higher yields.

7.5 Distributed Computing Project

To obtain the computing power required to run a 3D Monte Carlo simulation for each score evaluation within an optimisation, Muon1 can be run on many computers at once in a configuration known as 'distributed computing' that does not require particularly fast network connections between them. In fact, participating in such projects has become a popular pastime for many members of the public, as evidenced by the number listed on hub sites such as [86]. The additional features required to turn a scientific distributed program into one usable by the public are:

- A project website where the program can be downloaded (Muon1's is at [5]).
- Availability for a popular operating system, Windows in Muon1's case.
- Ease of installation: here, Muon1 simply requires extracting from its archive and once run will perform any additional setup by itself, downloading data from the Internet as appropriate.
- Smooth multi-tasking with other applications. Most such programs are designed to run at a very low task priority level, so only use 'spare' CPU cycles available when the user is not making demands of their computer. Muon1 also can work invisibly in the background.
- Some kind of feedback on the user's contribution or progress in the calculation. This can take the form of an overall scoring system measured in calculation units appropriate to the project and/or a graphical interface while a single calculation is running. Muon1 supports both of these and the calculation units used are mega-particle-timesteps, or Mpts, measured as 10^6 Runge-Kutta steps.

Once downloaded from [5] and installed on a computer, the behaviour of Muon1 (in its default distributed computing mode) is summarised by figure 7.15. This process combines all the elements described in previous sections of this chapter. The 'results storage' is the `results.dat` file and the 'results buffer' is `results.txt`. The reason Muon1 does not require computers to be connected by fast network connections is that individual simulations are run entirely on a single computer and will be so repeatedly until a large enough batch of results has been collected in the results buffer to be sent. This system also copes with intermittent connectivity by continuing to generate and store results suggested

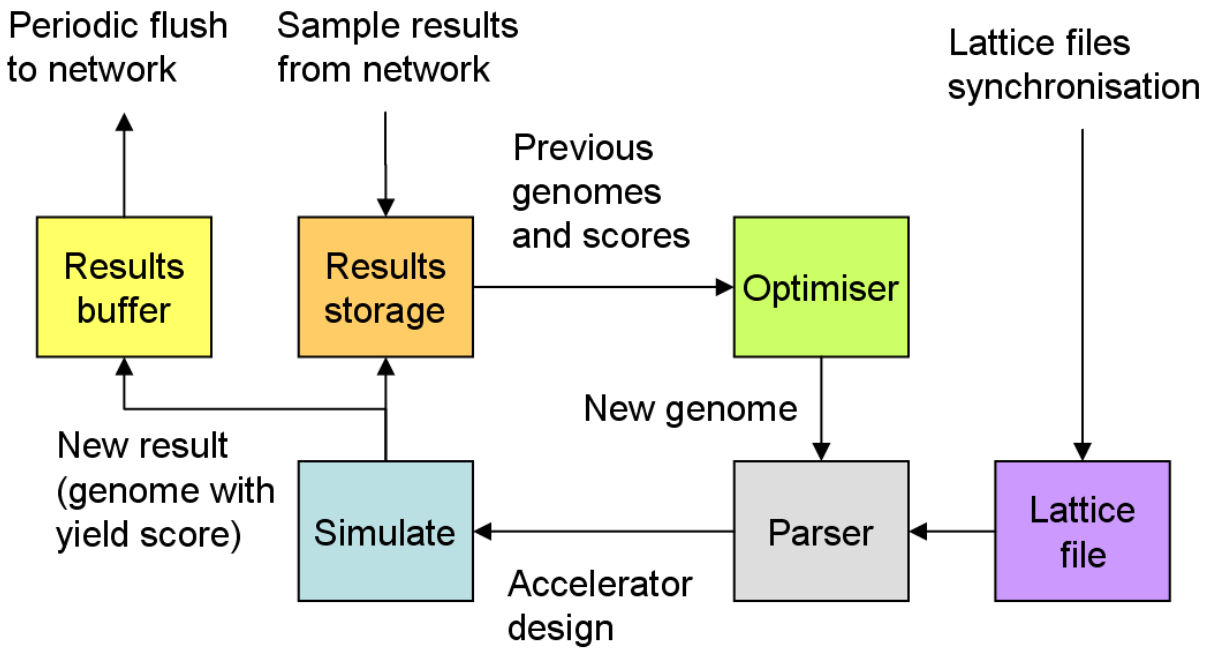


Figure 7.15: Cycle of operation of the Muon1 client program on a user's machine.

by the optimiser until a connection is available. Computers periodically check back with the central website to update the list of lattice files they should be working on.

As described so far, this system would be one of separate optimisations happening independently on many computers. However, another thing that Muon1 tries to download from the website is a 'sample file' for each optimisation, which contains a selection (typically 100) of the best results generated throughout the whole network. This provides a narrow channel of communication between the different installations, but it is effective: new high-scoring results and their variants are propagated across the network within hours because higher scoring results are far more likely to be selected out of `results.dat` for variation using the genetic algorithm.

Although most of Muon1's calculation is done with the program running in invisible mode (it is taking 95% of the CPU while I am writing this paragraph), a graphical mode is also provided which gives the fullscreen display shown in figure 7.16. This was originally written to help with the debugging and monitor the progress of the tracking code but is also now used by many project users as a screensaver.

The project as a whole also needs a 'back end' to receive and store results from the many users (typically more than 100 will be running the program at once). This additional infrastructure is represented in figure 7.17. Users have two methods to upload results: earlier versions of Muon1 used an FTP script but since some networks block FTP, an upload via HTTP (the protocol for viewing normal web pages) was added later. These go to an intermediate stage of a few FTP servers (also operated by volunteer users) that are provided as redundancy so the Muon1 clients always have somewhere to upload to. Every hour, the database computer at RAL retrieves all new results from these intermediates as part of a script that collects and counts the results submitted by each

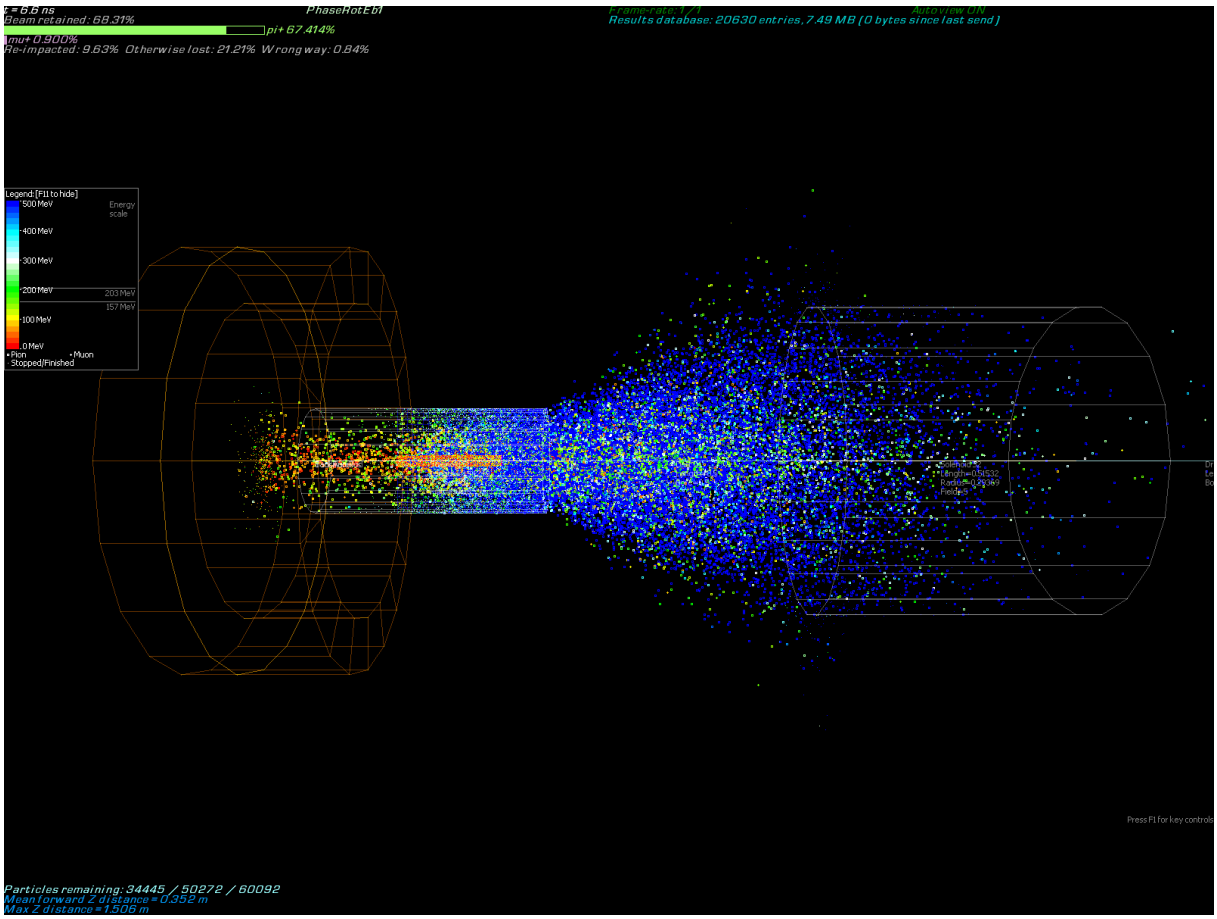


Figure 7.16: Muon1's 3D display of particles and components while running in graphical mode. This view shows the rod within the first solenoid and the second solenoid (shown as grey wireframe cylinders). Particles are mostly pions at this point and are coloured according to their energy, red being low and blue high. The brown annuluses on the left-hand side are additional geometry added to stop backwards pions from flying off into empty space (which wastes simulation time).

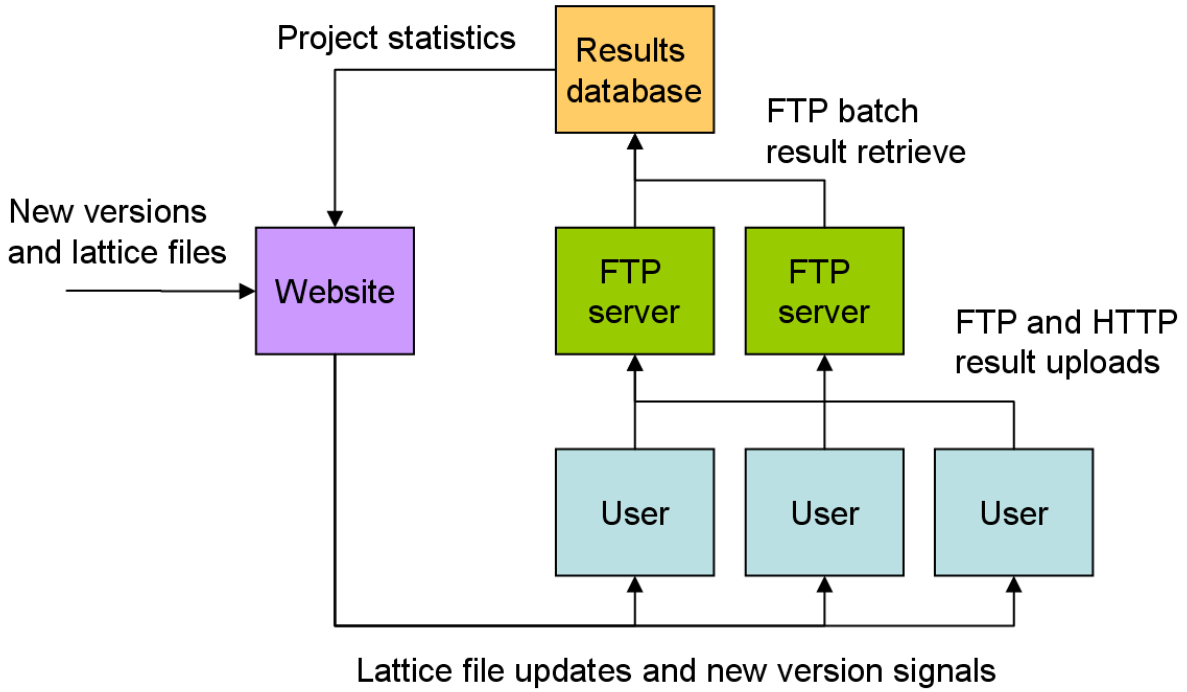


Figure 7.17: Flow of data through the Muon1 distributed project network.

user. This script also verifies the checksum of each result and removes any duplicated uploads. Lastly, the script publishes a table of the total number of results and Mpts that each user has computed on each optimisation, together with their highest result score on each optimisation. Users are identified by the filename given to their results during upload, which is taken from the `user.txt` they create in the Muon1 directory (if uploads are switched on and no `user.txt` file is found, there is a GUI prompt to enter the username). A second script on the database computer periodically counts through all the results for each active optimisation and generates a 'sample file' (as mentioned earlier) that is uploaded to the website.

With hundreds of active users, a considerable number of simulation results can be generated over time. Log files enabled the cumulative totals over most of the project lifetime to be plotted in figure 7.18. Observe that the number of results does not increase as smoothly as the amount of calculation. This is because simulations from different lattices can vary in length depending on how much beamline they simulate. Even results from the same lattice can vary significantly early on, as a low-scoring result in which most of the beam is lost will require less calculation than one where much of the beam travels further. A dip in the number of results is visible due to repeat removal and maintenance: the whole database currently stands at 183 GB and went through several software revisions to cope with the increasing size.

The particle-timesteps axis has much larger figures on it, by a factor of 10^8–10^9. This can be understood since each simulation has $\sim 10^4$ active particles undergoing 10^4–10^5 timesteps of 0.01 ns each, the total 100–1000 ns time corresponding to beamline lengths of the order of 100 m.

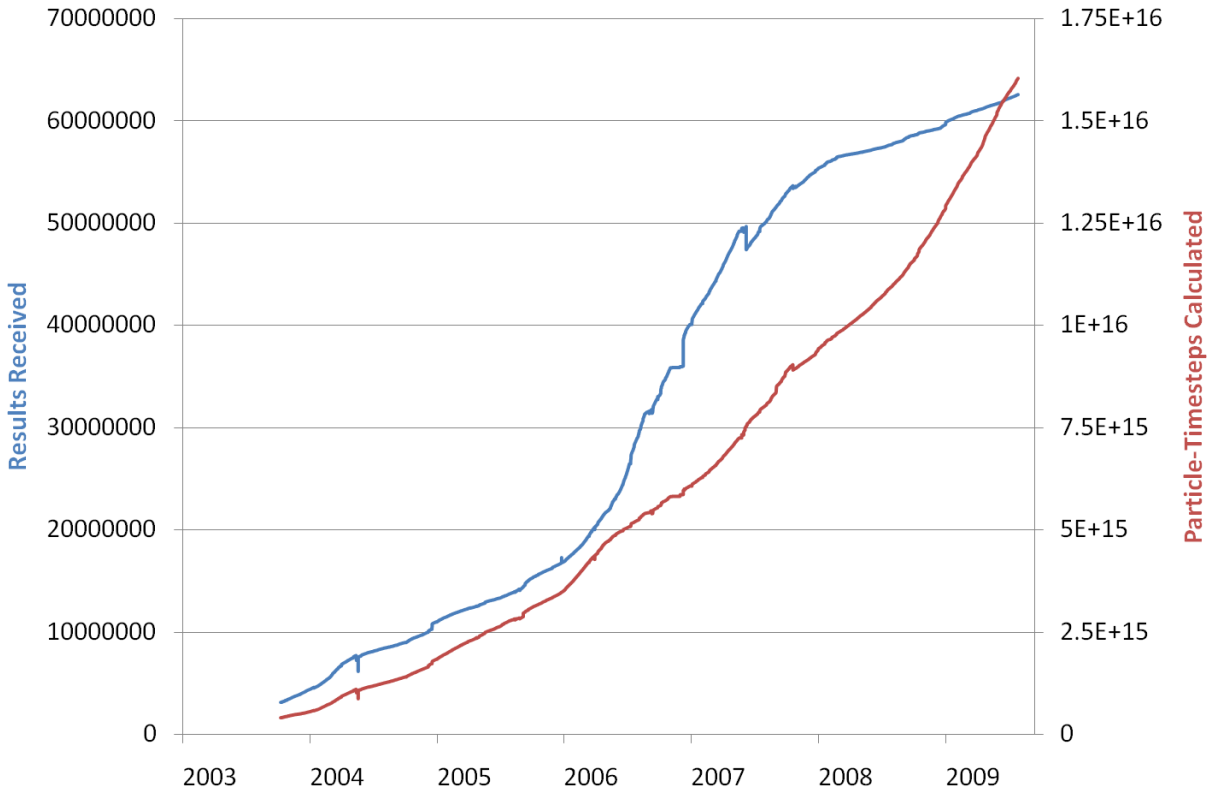
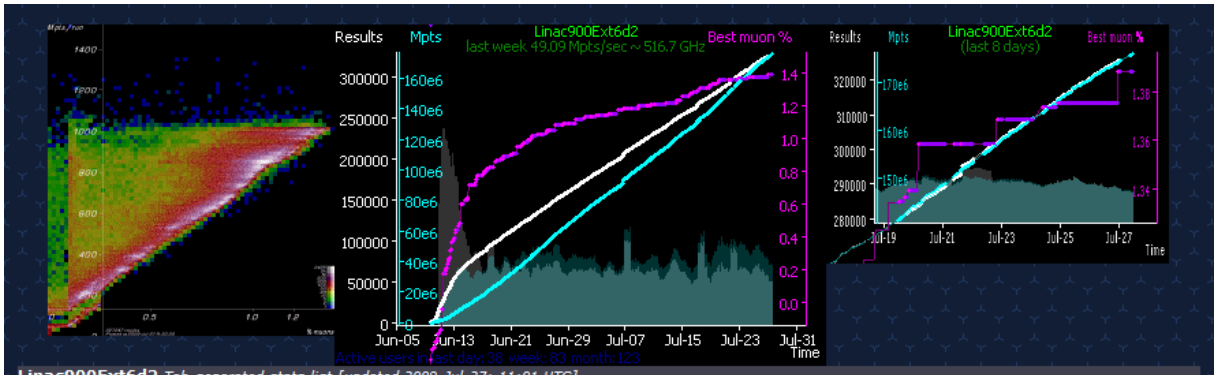


Figure 7.18: Total simulation results received and calculation performed across all optimisations run on the distributed computing network.

Users' scores and the progress of current optimisations are shown on the Muon1 website, a screenshot of part of this is shown in figure 7.19. The left-hand scatter plot shows result score against the number Mpts used to produce it, with a strong diagonal feature from the correlation of calculation time with result score mentioned earlier. This plot was originally produced as a curiosity but became part of the site when it was found that plotting Mpts as a second variable to score often separates different 'evolutionary tracks' so it becomes clear when a new design branch has been taken. It also shows up any erroneous results as these will very often have the wrong amount of calculation for a given score.

The distributed optimisation project allowed many different lattice files corresponding to ranges of possible muon front end designs to be uploaded to the project website and subsequently optimised. When the best muon yield achieved for a particular lattice stopped increasing for a long period of time (over a week), the lattice was retired from the website and the highest-scoring result extracted. Chapter 8 analyses the highest-scoring results from eight lattices covering a variety of front end configurations.



Linac900Ext6d2 *Tab-separated stats list [updated 2009-Jul-27; 11:01 UTC]*
Show users active in last **day**, **week**, **month**, **quarter**, **year**, or **ever**.

#	Username	v4.4 results	Mpts	Best muon percentage	Hours since last active
.	boinc (Boinc Wrapper)	105965	88`360577.0 (50.2%)	1.388295	1
1.	[TN]steinrar	5885	2`164745.8 (1.23%)	1.378031	0
2.	Pilgrim	205	163651.9 (0.0930%)	1.373975	34
3.	silent Float	76	84457.6 (0.0480%)	1.371918	15
4.	[TA]JonB	21261	11`034958.6 (6.27%)	1.371744	2
5.	[TA]Chocobo Navy	12312	5`468030.2 (3.11%)	1.368945	0
6.	[TA]viztech	2031	1`241840.6 (0.706%)	1.368500	1
7.	[TA]Silverthorne	7959	4`254772.2 (2.42%)	1.368011	2
8.	DWWehner	114	88868.7 (0.0505%)	1.366731	0
9.	Jay Pound	10	11811.1 (67.1µ)	1.366180	23
123 of 136 users active		***327660***	***175`908111.9 (1.10%)***	***1.388295***	***0***

Next › 50 100

Figure 7.19: Summary graphs for each optimisation and users' best results are updated hourly on the Muon1 website [5]. From left to right the graphs are: a scatter plot (calculation time against score) for all results of this lattice; history of successive best scores, total results and calculation; and a zoom into the last week of the middle plot. The scatter plot can be clicked for a more detailed version and the table below shows the current ten users with the highest submitted scores, along with their amounts of calculation and number of results submitted from this lattice.

Chapter 8

Results from Optimisation

This chapter describes the designs resulting from optimisation on the distributed computing network using the Muon1 code. Several different optimisation ranges of accelerators ('lattices') were submitted to the network, starting with just the decay channel and later including phase rotation and linac sections. These optimisation ranges are specified in tables and the best designs are shown with plots of beam behaviour and transmission figures. Full parameters for these designs are listed in appendix E.

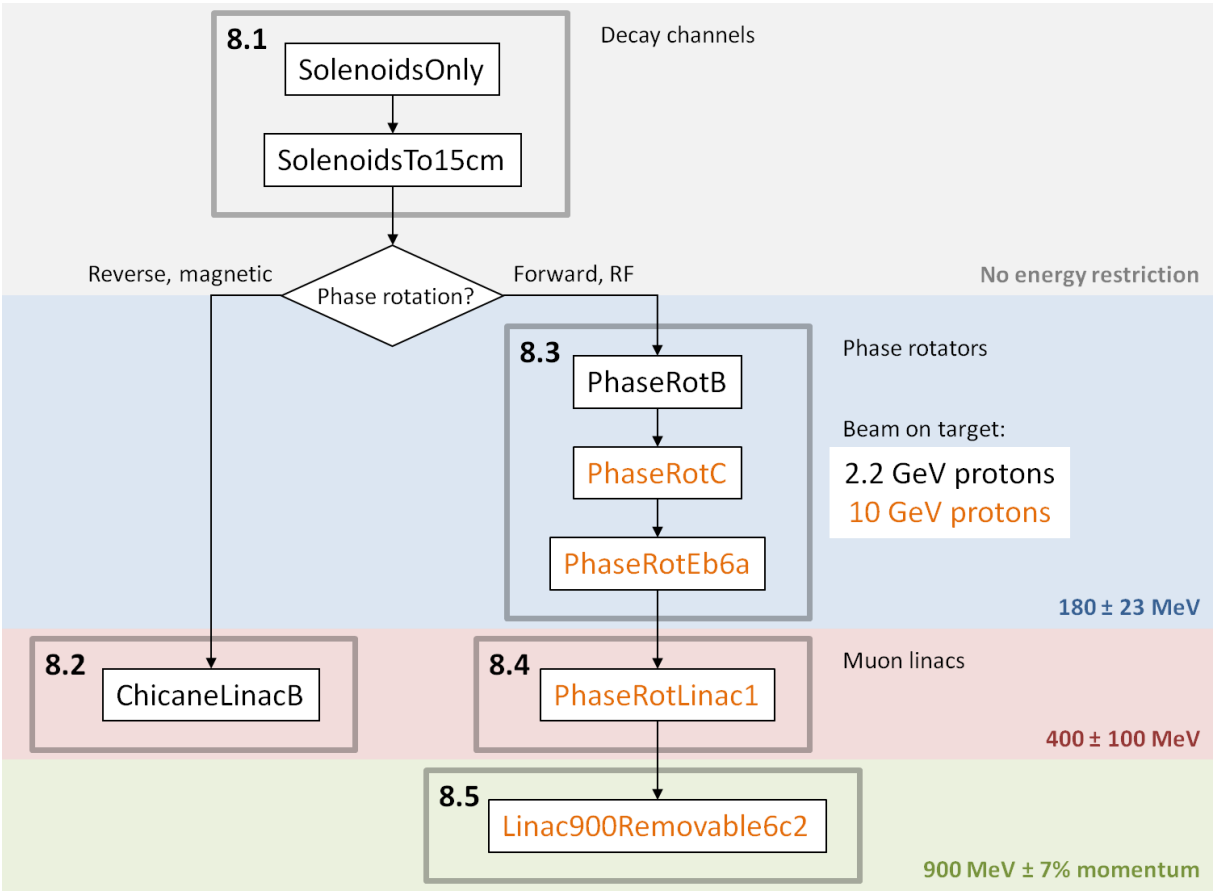


Figure 8.1: Lattices described in sections of this chapter, showing output energy criteria.

Table 8.1: Optimisation ranges and best muon yield for the `SolenoidsOnly` lattice.

Decay channel (to 30 m):		
Target	Angle	0–0.5 radians
	Z offset	0–0.45 m
`S1`	Length	0.2–0.45 m
	Radius	0.1 m (fixed)
	Field	0–20 T
Solenoid	Length	0.2–0.6 m
	Radius	0.1–0.4 m
	Field (`S2`–`S4`)	−5–5 T
	Field (`S5`–)	−4–4 T
Drift	Length	0.5–1 m
	Proton beam energy:	Muon energy criterion:
	2.2 GeV	Any
Best yield: μ^+/π	μ^+/p·GeV	μ^+/s (4 MW proton beam)
16.676%	0.012652	3.16×10^{14}
22.873% including pions		

8.1 Decay Channel

The first piece of the muon front end to be optimised consisted of the target and decay channel, with the target enclosed in the first (stronger) solenoid `S1` and the remaining solenoids centred on the same axis, in the direction the proton beam (and the majority of the pions) will be travelling. The lattice file for this was named `SolenoidsOnly` because it contains just the decay channel and no RF.

Table 8.1 defines the optimisation ranges of all the components in this lattice. The earlier optimisations, such as this one, used a MARS15 data set of pions generated from a 2.2 GeV proton beam because this was the energy of the CERN SPL design [40] at the time. Later on this switched to a 10 GeV proton beam, primarily because of the findings of the International Scoping Study in 2006 [24].

The target rod may be rotated by up to half a radian (probably an excessive angle but a suitable upper bound) and moved in Z so that its centre could be anywhere from the entrance to the exit of `S1`. Most optimisations chose a near-zero angle for the rod and placed it near the centre of `S1`, which was given the maximal allowed length of 45 cm. `SolenoidsOnly` is unusual in that it is the only optimisation examined that assigned a relatively large angle, 95.6 milliradians, to the rod. The figure of merit for this lattice is total weight of muons leaving the channel, with no energy cuts at all. Primarily it is concerned with the transverse focussing and magnetic capture from the target. However, since the decay channel is only 30 m long, a significant number (∼27%) of undecayed pions still remain in the beam, so the decay-channel-only lattices in this section have the transmission including the pions also printed for reference.

Ranges for the variables concerned with solenoids have been chosen to avoid impracti-

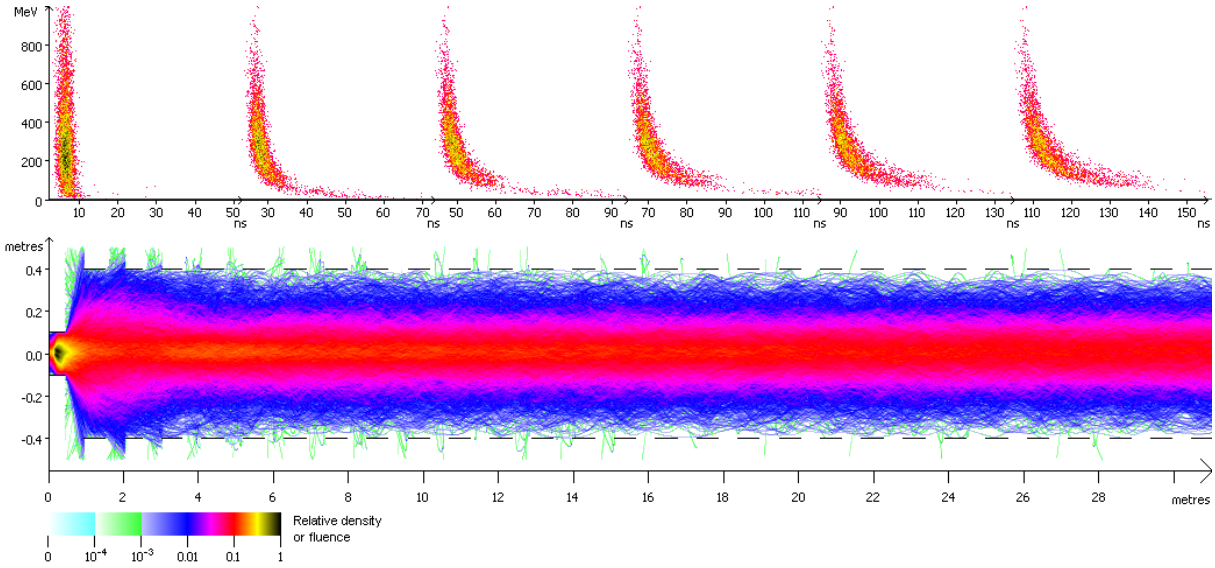


Figure 8.2: Longitudinal phase space evolution (top) and beam envelope plot (bottom) for the optimised design from the `SolenoidsOnly` lattice.

cally large field volumes. Only `S1` may have a field up to 20 T and the subsequent solenoids are limited to 4 or 5 T. Because of its high field a much tighter bore is anticipated for `S1` (10 cm radius), which cannot be increased. The remaining solenoids with only one quarter the field strength can have radii up to four times as big, which keeps the stored energy the same. Drifts of at least 50 cm are imposed between solenoids for independent cryostats, diagnostics and maintenance access, this is unlike Study II [25] and others in which the area near the target is a monolithic module. The solenoid-drift pattern in the channel repeats until a distance of 30 m is reached, where the channel is terminated and the muons within the final solenoid bore counted.

Figure 8.2 shows the behaviour of the beam from the target in this optimised decay channel. The beam maintains a constant size through most of the channel, coming from a regular focussing structure. All but two of the solenoids have been assigned the maximum possible radius, field and length and all drifts have been given the shortest allowed length. This means the technical constraints on the design are the primary limitation and suggests the yield would increase if solenoids were allowed to be larger, longer, higher in field or closer together. The exceptions are the third solenoid, which is only 48.3 cm long (60 cm max) and the fourth solenoid, which has a field of 4.24 T (5 T max) and a length of 55.3 cm. The effects of these non-regularities are visible in the beam envelope following that point. It is likely these are genuine optima related to the matching from the initial pion distribution because of these solenoids' proximity to the target and the fact that all the other solenoids have reliably been set at very close to maximal parameters.

Figures 8.4 and 8.5 show the RMS optical functions and emittances calculated from the beam in this design. The values are obtained by calculating moments of the particle distribution as it passes through planes of constant z spaced along the accelerator. Since only a finite number of particles are tracked, there is some statistical error visible for instance between the x and y planes for the α and β functions: these should be identical for a cylindrically symmetrical machine. Figure 8.4 also shows very low β at the location

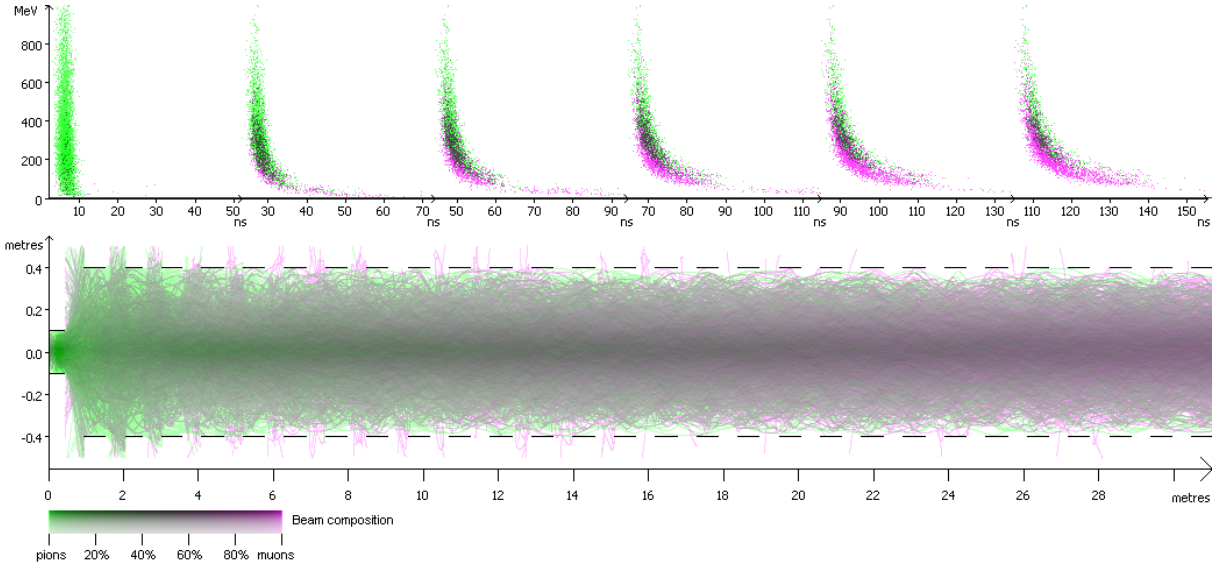


Figure 8.3: Pions (green) decaying to muons (purple) in the optimised design from the `SolenoidsOnly` lattice.

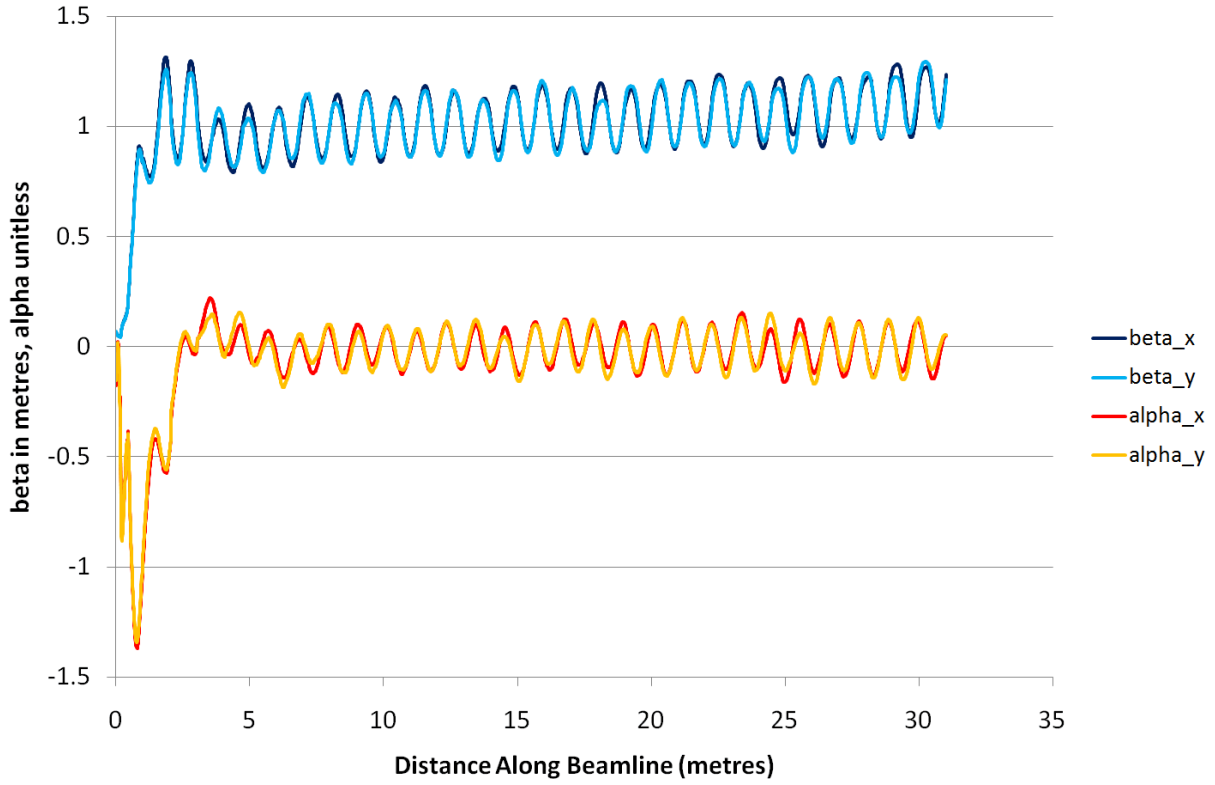


Figure 8.4: Optical functions in the optimised `SolenoidsOnly` design.

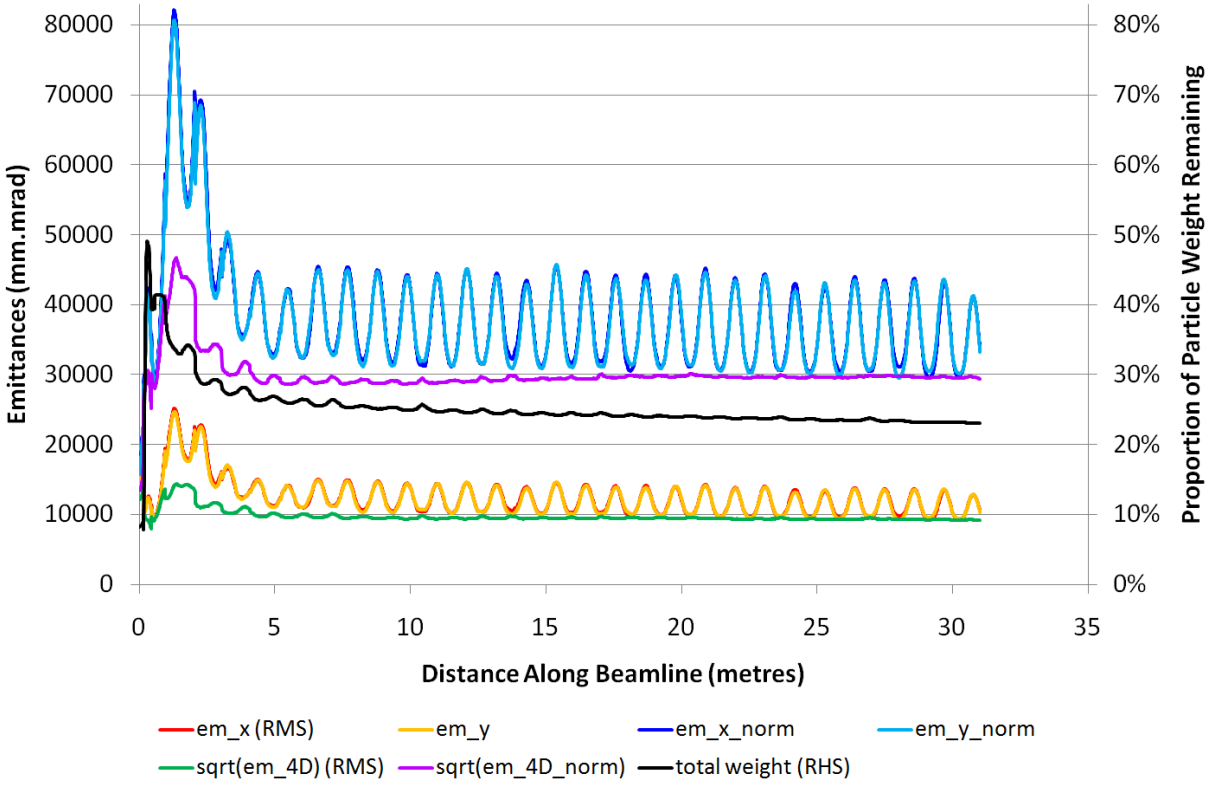


Figure 8.5: Emittance evolution in the optimised `SolenoidsOnly` design.

of the target, followed by negative α indicating a diverging beam, as would be expected. The functions then settle into a periodic cell, with an upward drift of β that may be explained because β is not a 'normalised' quantity and therefore varies with overall beam energy. Here, as the slightly higher-energy pions decay into muons, the average energy gradually reduces and unnormalised quantities therefore increase.

Figure 8.5 shows an artifact of the logging scheme whereby particles caught in magnetic bottles in the gaps between the solenoids are counted multiple times, making the remaining particle weight (and 4D emittance) appear to increase in these regions. Ignoring this, the normalised 4D emittance increases slightly in the middle section of the channel from pion decays, before becoming invariant as would be expected. The 2D emittances projected into the x and y planes oscillate above the 4D values because coupling between the planes, which is induced by solenoids, increases the projected emittances (intuitively: the beam appears bigger because information about its correlations has been lost).

The longitudinal phase space plots in figure 8.2 show the drift from the initial near instantaneous (1 ns RMS) spread of pion energies, with lower energy particles lagging behind more in time. This lag is nonlinear owing to the large energy spread involved. Figure 8.3 shows the same thing only with the beam coloured by composition, with the two particle species given different colours, instead of total intensity. The overall decay of pions to muons is visible in the transverse plot while the longitudinal plot shows that the pions tend to have slightly higher kinetic energies than their decay products at a given position in the beam. It also illustrates the time dilation phenomenon with higher-energy pions taking longer to decay than the slow ones near the bottom of the plot.

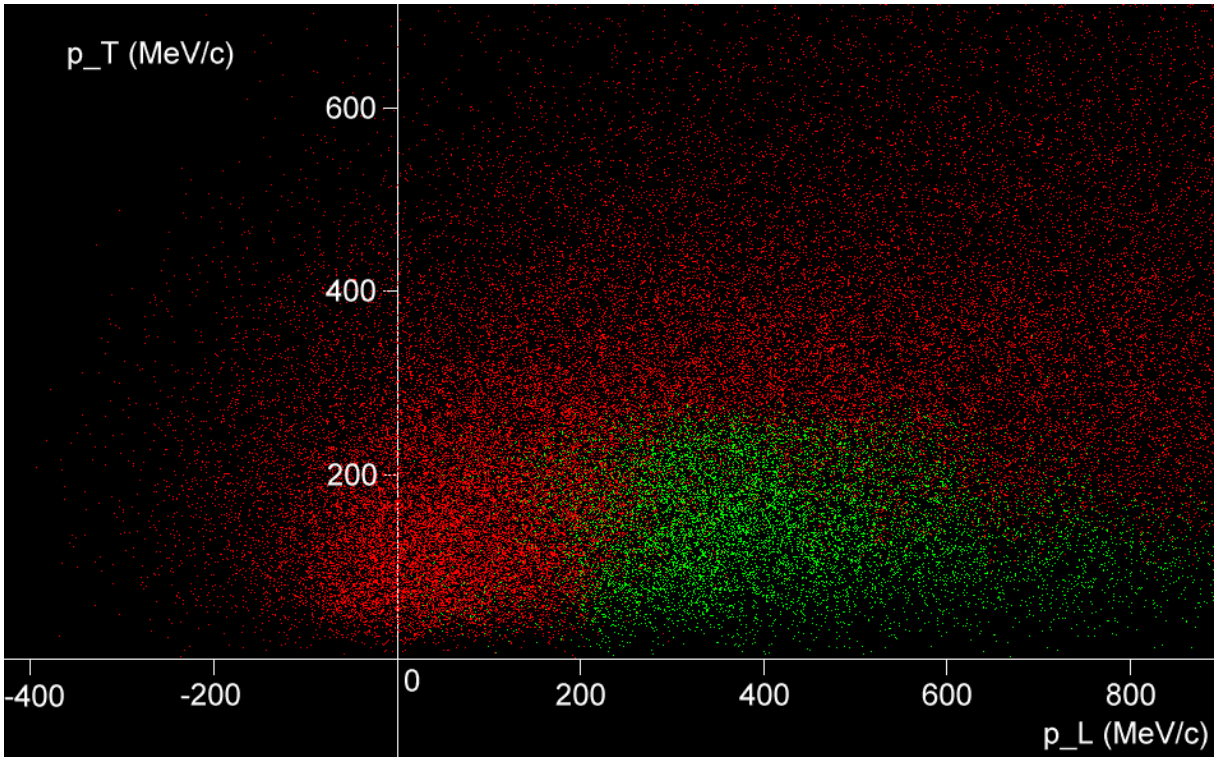


Figure 8.6: Pions in the initial (p_L, p_T) momentum space coloured green if they are transmitted through the decay channel and red otherwise.

The decay channel on its own does not attempt to select particles by energy. Figure 8.6 shows the momentum distribution of pions from the target, coloured according to whether they survive the decay channel. The solenoid fields and bores dictate an upper limit on p_T and there is no upper limit on p_L as a particle of high enough speed will always survive through a linear channel if it has a line of sight to the exit. While the reason for losses with $p_L < 0$ (backwards pions) is obvious, the poor survival for p_L less than about 200 MeV/c is because low forward momentum allows the pion to be reabsorbed in the target rod after spiralling in the 20 T field. Particles with very low p_L are also more easily diverted by the radial component of solenoid fringe fields, which can exert a force in the negative Z direction when the particle has a suitable transverse velocity. The force will reduce the p_L of the particle further, increasing the angle between its direction of motion and the accelerator axis; in extreme cases, particles can be reflected backwards (the magnetic bottle effect); losses become more probable in either situation.

Because the 40 cm radius of the decay channel is too large for most acceleration systems, an optimisation named `SolenoidsTo15cm` was started that forced the final solenoid (and aperture) to be 15 cm in radius. Table 8.2 shows the resulting yields, which on first sight appear encouraging given the cross-sectional area of the accelerator is reduced by 86% but the yield only by 33%. However, this lattice is an example of an optimiser finding an inadequacy in the figure of merit. The final solenoid's length has been reduced to the *minimum* possible here, which means its short length can accommodate a betatron focus of the beam. That is, although the beam is small in the final solenoid where the figure of merit is counted, the smallness has come at the expense of increased x' values and the

Table 8.2: Optimisation ranges and best muon yield for the `SolenoidsTo15cm` lattice.

Decay channel (to 30 m) *as for* `SolenoidsOnly` *(table 8.1) except*:		
Solenoid	Length	0.2–0.6 m
	Radius	0.1–0.4 m
	Radius (final solenoid)	**0.15 m (fixed)**
	Field (`S2`–`S4`)	−5–5 T
	Field (`S5`–)	−4–4 T
	Proton beam energy:	Muon energy criterion:
	2.2 GeV	Any
Best yield: μ^+/π	μ^+/p·GeV	μ^+/s (4 MW proton beam)
11.185%	0.008486	2.12×10^{14}
15.756% including pions		

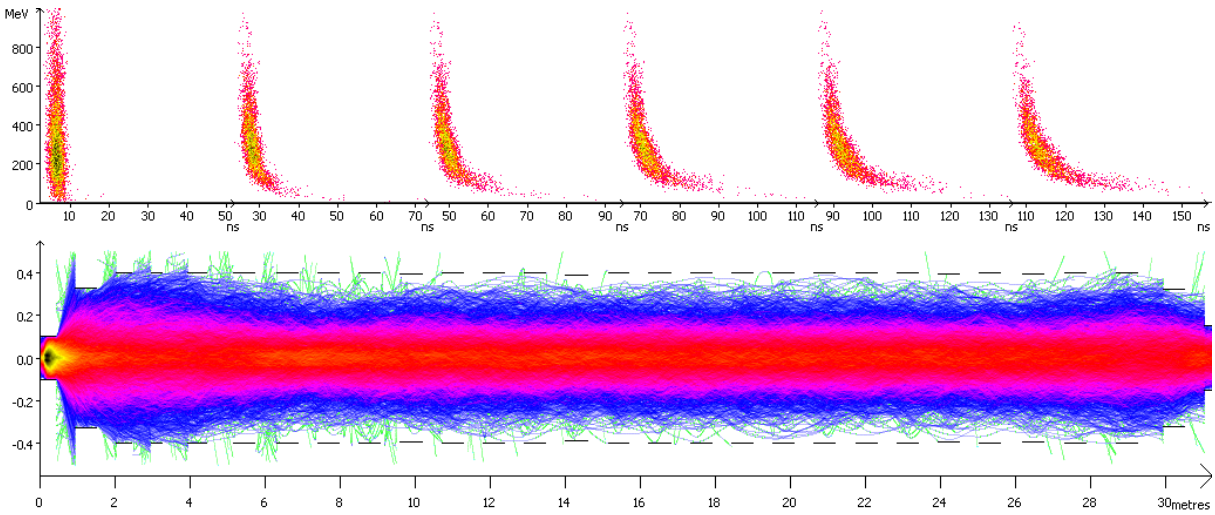


Figure 8.7: Longitudinal phase space evolution (top) and beam envelope plot (bottom) for the optimised design from the `SolenoidsTo15cm` lattice.

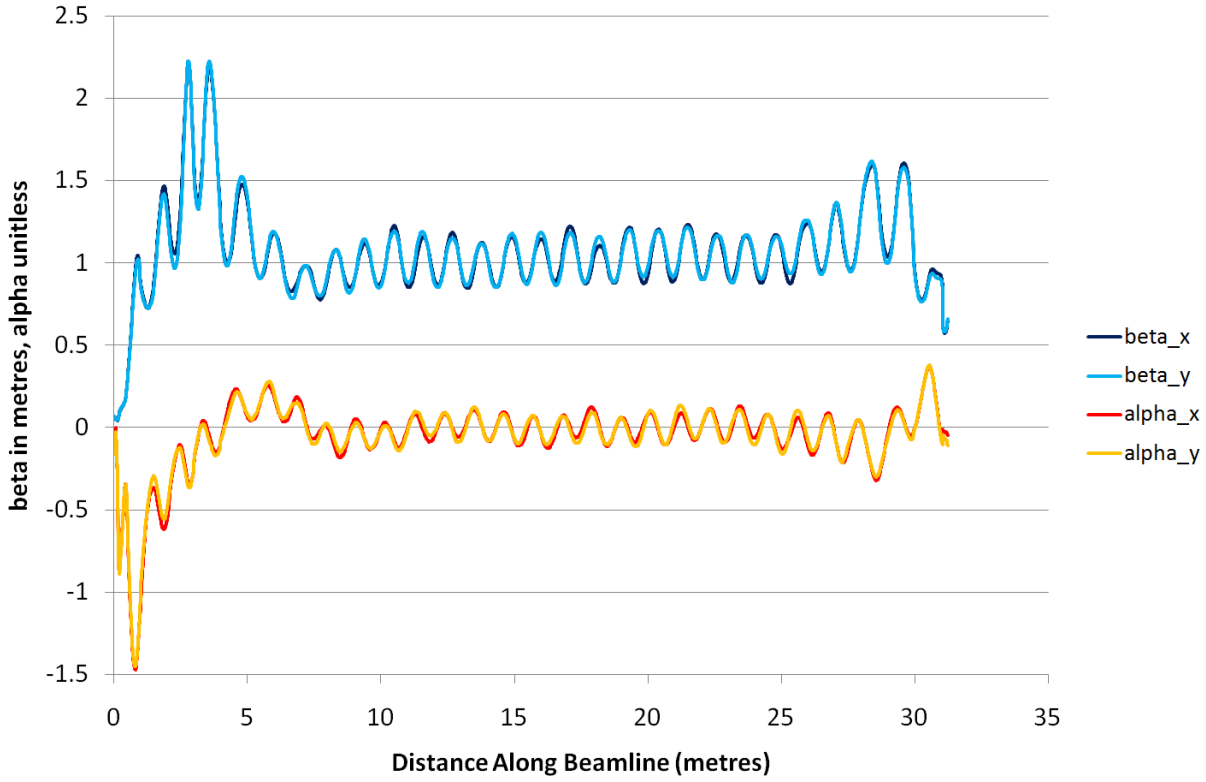


Figure 8.8: Optical functions in the optimised `SolenoidsTo15cm` design.

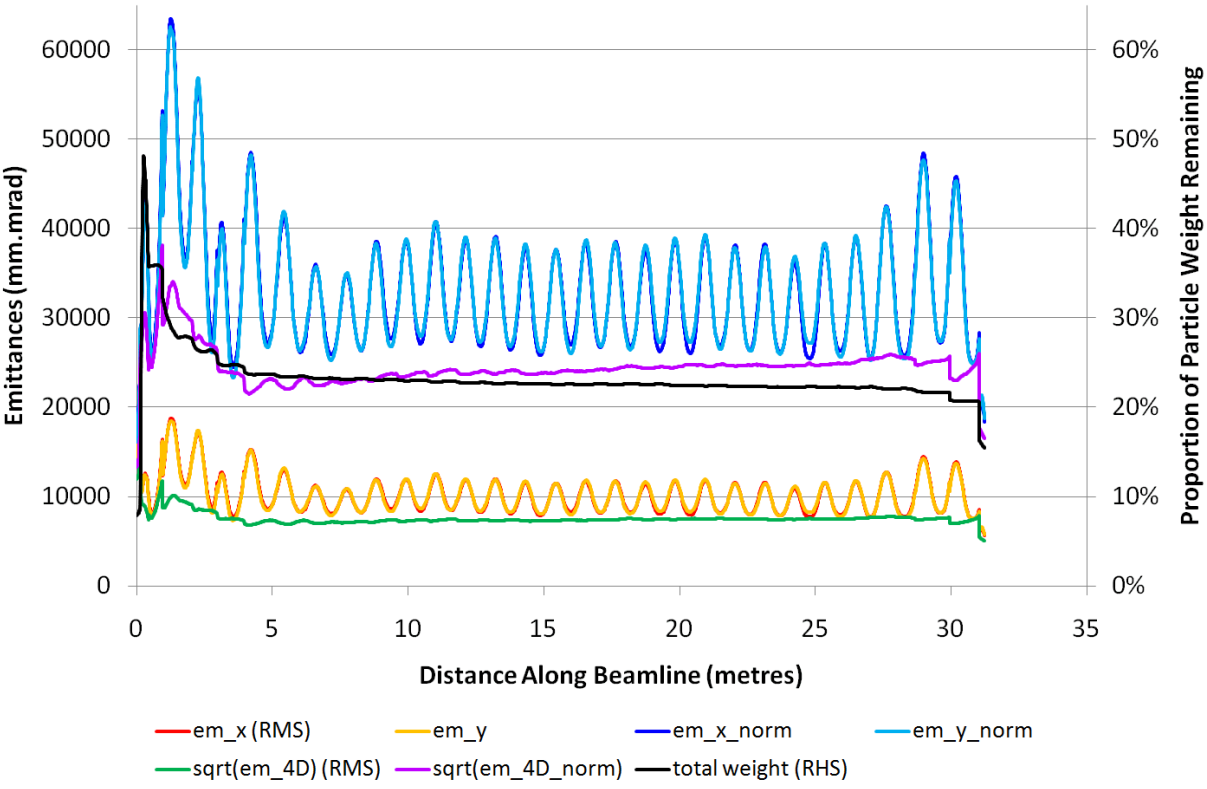


Figure 8.9: Emittance evolution in the optimised `SolenoidsTo15cm` design.

Table 8.3: Normalised emittances of output beams from all optimised lattices in this chapter when projected into the x–x' or y–y' plane (choice of plane does not matter as the beam is cylindrically symmetric).

(mm·mrad)	RMS	75%	90%	95%	99%	100%
`SolenoidsOnly`	34076	93666	177399	239913	343130	496991
`SolenoidsTo15cm`	18417	52940	84057	106017	156766	315720
`ChicaneLinacB`	8216	24031	37154	45431	59919	91079
`PhaseRotB`	9972	29677	42770	51120	64029	84877
`PhaseRotC`	10208	30366	44289	52548	65661	98812
`PhaseRotEb6a`	11592	34266	50288	60103	79165	114378
`PhaseRotLinac1`	13108	38696	58298	69791	88635	134122
`Linac900Removable6c2`	14812	44028	66225	78091	97455	140681

beam would blow up again in the next cell. The formation of the focus can be seen in the transverse plot on figure 8.7, towards the end of the channel.

Figures 8.8 and 8.9 also show the focus preceeded by a region of converging beam with positive α and the effect of beam scraping, which decreases the 4D emittance as the highest amplitude particles are removed. However, the 4D emittance also exhibits growth between the final two solenoids because the abrupt change in optics means the beam is mismatched there.

Although of limited use otherwise, this optimised design does confirm the nontriviality of the non-regular matching section from the target observed in `SolenoidsOnly`. `SolenoidsTo15cm` has found a different optimum with all of `S2`–`S5` deviating from maximal parameters in some way and `S4` having a flipped field. The channel then continues with the same regular structure as in `SolenoidsOnly`, with maximal solenoid fields, until a few cells before the final focus (with solenoid `S26` being given a flipped field and several non-minimal drift lengths). It is encouraging that the optimiser has produced 'matching sections' in a similar way to a human designer, even though every parameter was separately variable so it was equally possible to put such a section in the middle of the channel away from the boundary conditions at the ends.

In front ends with no cooling, the beam emittance will generally increase with the yield, since there is no way to pack more particles into a smaller emittance. The values for all the designs examined in this chapter are given in table 8.3 with various amounts of the beam contained in the ellipse. Since the next component after the front end will be a linac using quadrupoles, projected 2D emittances are given here instead of the 4D amplitude used for solenoid channels, as the dynamics of the two planes will become uncoupled after this point. As energy increases, the un-normalised emittance corresponding to these values will fall, permitting more conventional structures further down the muon linac. For example, above 8.7 GeV a beam of normalised emittance 80000 mm·mrad may fit through an ILC 1.3 GHz module of bore radius 35 mm and length 1256 mm. Lower frequencies with larger bores and shorter modules can fill the gap from 0.9–8.7 GeV following on from something like the `Linac900Removable6c2` option.

Table 8.4: Optimisation ranges and best muon yield for the `ChicaneLinacB` lattice.

Decay channel (to 30 m) *as for* `SolenoidsOnly` *(table 8.1) except*:		
Solenoid	Length	0.2–**0.9** m
	Radius	0.1–0.4 m
	Field (`S2`–`S4`)	−5–5 T
	Field (`S5`–)	−4–4 T
Chicane: *as specified in [37]* (fixed)		
Linac:		
Structure	Number of cells	60–120
Solenoid	Length	0.2–0.9 m
	Radius	0.15 m (fixed)
	Field	−4–4 T
Drift	Length	0.5–1 m
RF	Frequency	88 MHz (fixed)
	Voltage	0–$(L_{drift} + L_{sol})\times 4$ MV/m
	Phase	−180–180°
	Proton beam energy:	Muon energy criterion:
	2.2 GeV	400±100 MeV
Best yield: μ^+/π	μ^+/p·GeV	μ^+/s (4 MW proton beam)
1.400%	0.001062	2.65×10^{13}

8.2 Chicane and 400 MeV Muon Linac

This section examines an attempt to undo the distance–energy correlation that appears in the decay channel using a bending magnet chicane as shown in figure 8.10 and described in [37]. This idea from Grahame Rees was implemented in Muon1 with field maps and put on the distributed optimiser as the lattice `ChicaneLinacB`. Once the distance–energy correlation is undone, the rebunched beam is captured in an 88 MHz RF solenoidal linac to 400 MeV using cells similar to those in [83]. This frequency is favoured because experimental work at CERN (also described in the reference) has verified operation of such cavities up to the 4 MV/m gradient used in this optimisation. It is different to the 201 MHz used in many other studies [25, 24, 73] because it attempts to capture the muons in a single bucket rather than splitting into a train of bunches spaced at a higher RF frequency.

Table 8.4 shows which parameters were optimised in the decay channel and linac either side of the chicane. The maximum length allowed for solenoids in the decay channel has been increased from 0.6 to 0.9 m because these solenoids were almost always given the maximal length in the decay channel optimisations. The field and radius were also maximal but these would present much higher cost and engineering implications to increase than the length. The result is that `ChicaneLinacB` has a nontrivial optimum for solenoid length, about 83 cm, that is no longer at the end of the range, with the field converging to 3.25 T as well. Periodic solenoid channels exhibit resonances that cause particles to be lost around certain energies determined by the amount of focussing per cell. To increase transmission, it is likely that the amount of focussing (solenoid field and length) has been

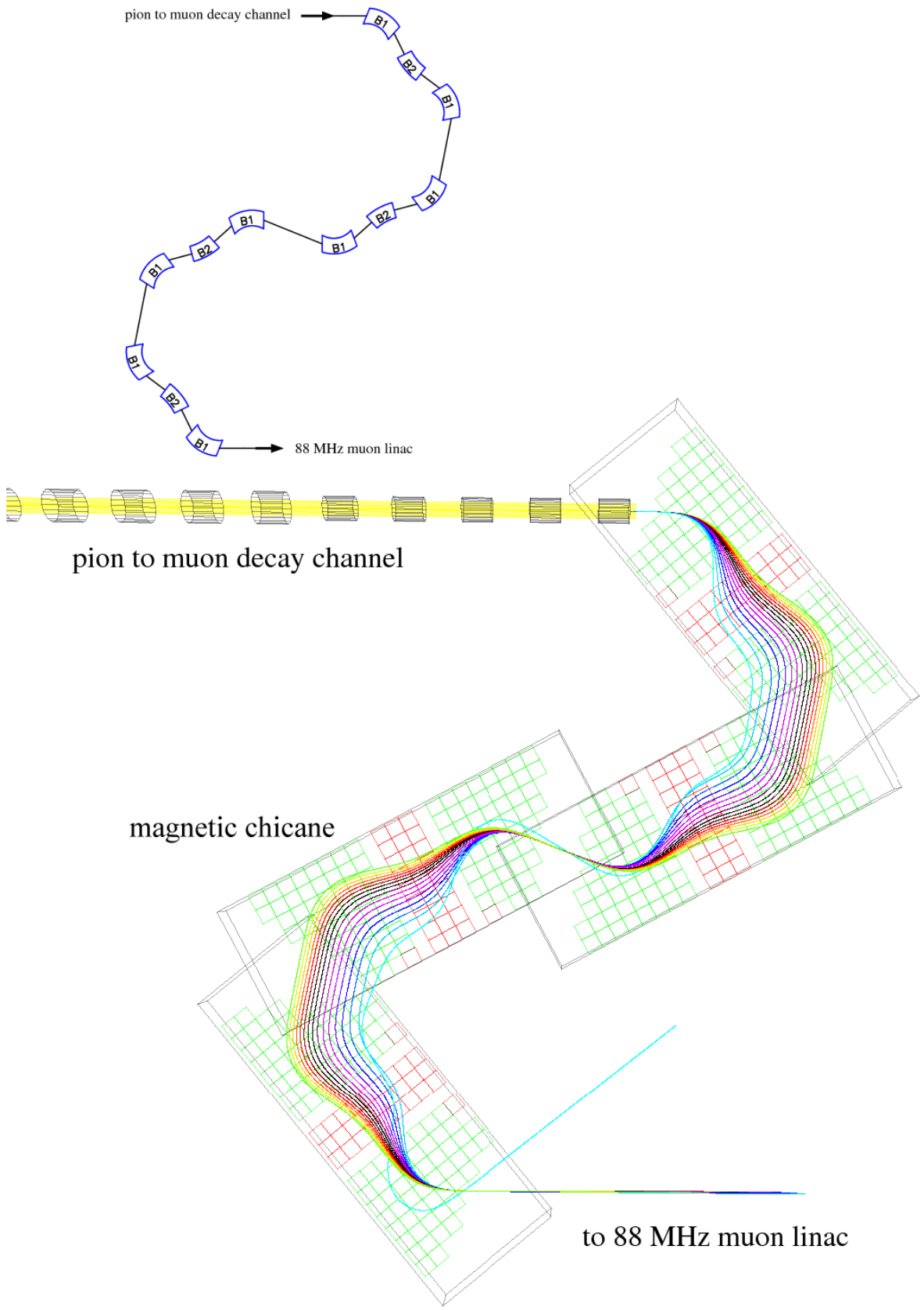


Figure 8.10: The structure of the magnetic bending chicane lattice (top) and muons of energy 120–270 MeV tracked through its field maps in Muon1 (bottom), adapted from [37]. Placement relative to the end of the decay channel is also shown.

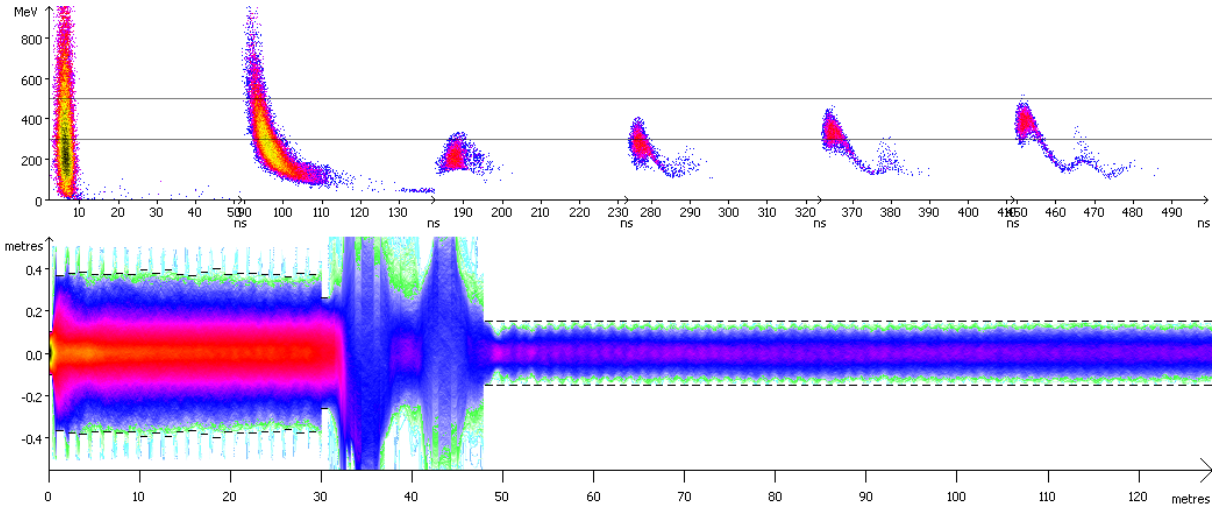


Figure 8.11: Longitudinal phase space evolution (top) and beam envelope plot (bottom) for the optimised design from the `ChicaneLinacB` lattice.

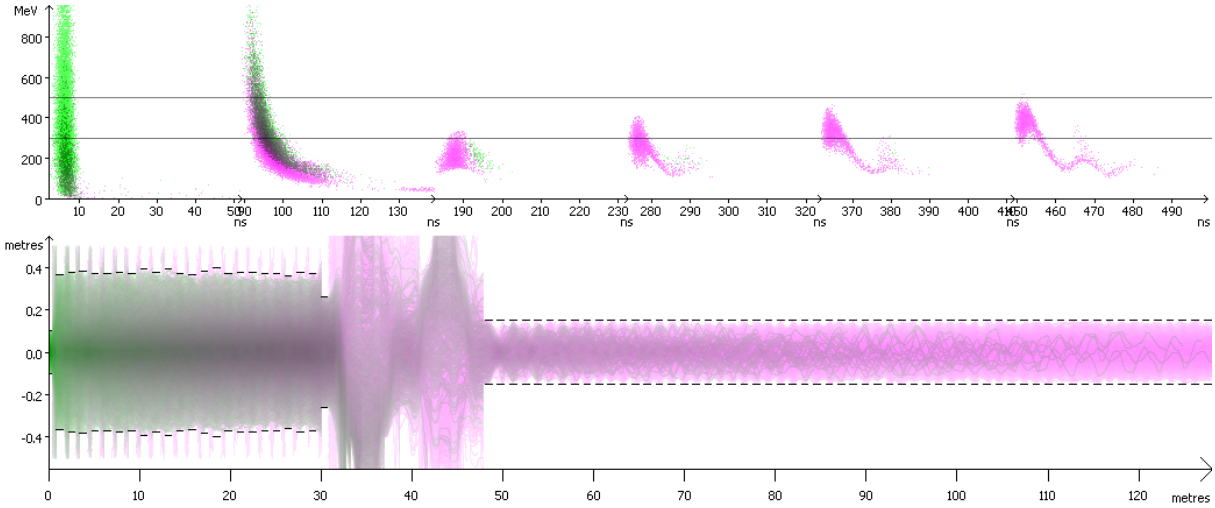


Figure 8.12: Pions (green) decaying to muons (purple) in the optimised design from the `ChicaneLinacB` lattice.

optimised to place these resonance energies far from the bulk of the particle distribution.

The linac section has a similar structure to the decay channel only now the solenoid radii are fixed at 15 cm to not go outside the RF bore. The drifts of $\geq$50 cm now become crucial for inserting the RF gaps, with the lattice assuming the RF is capable of 4 MV/m average gradient along the entire channel. For the figure of merit, muons are now only counted within a band of 400$\pm$100 MeV kinetic energy, a range that is chosen to contain the RF bucket of the muon accelerator [37].

Figure 8.11 shows how the RF bucket forms and moves up to within the target energy bands. It also shows how the optics of the magnetic chicane put a severe cut on the momentum of particles transmitted: compare the drifted (2^{nd}) longitudinal distribution to the much smaller (3^{rd}) one after the chicane, which is unwrapped in this plot and appears as a section of large transverse amplitudes. This leads to the chicane/linac design having

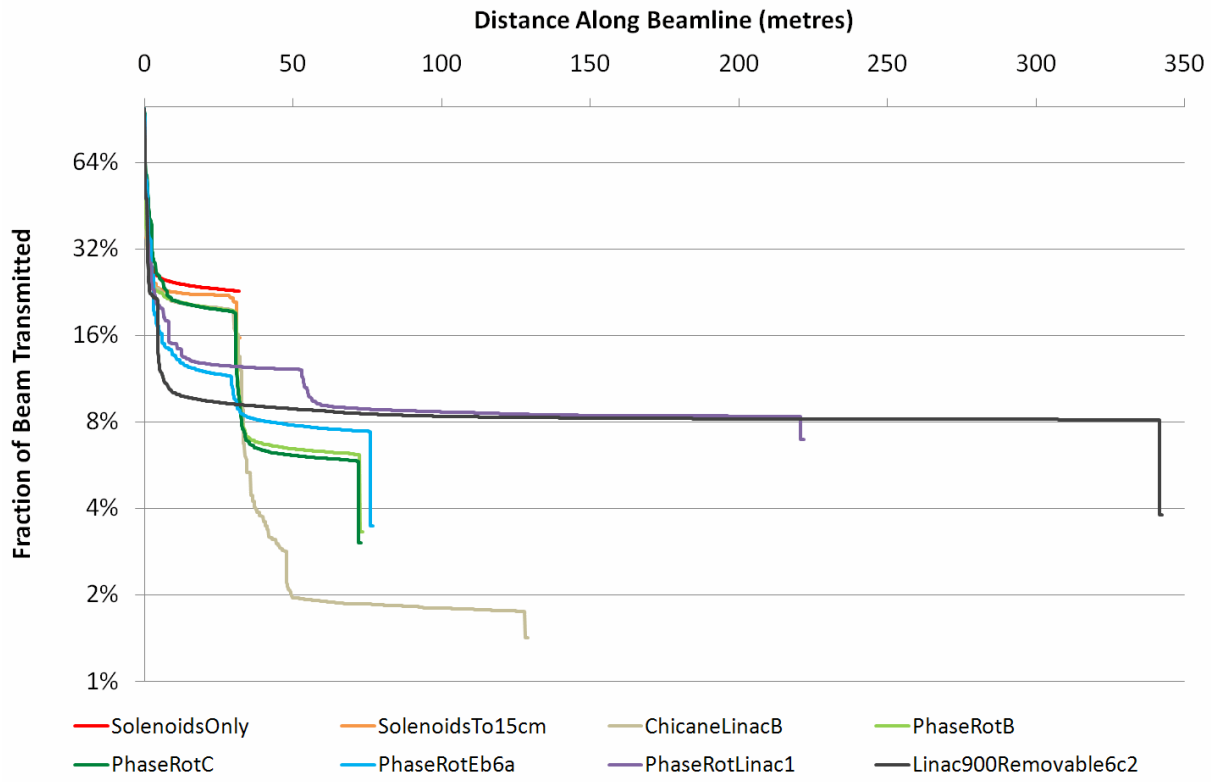


Figure 8.13: Particle transmission as a function of distance for all eight designs discussed in this chapter.

a low transmission of 2.65×10^{13} muons per second, approximately one quarter of the 10^{14} required for the neutrino factory to make 10^{21} neutrinos of each type in a 10^{7} second operating year. The large loss is caused by the limited acceptance of the chicane, partly longitudinal but more importantly in handling particles with large transverse amplitudes.

Examining the behaviour of undecayed pions in this system (figure 8.12) shows that they are mostly lost in the chicane because they have the wrong momentum. A greener tail of pions can be seen exiting the bottom of the transverse plot immediately after the chicane begins, though a few with similar energies to the muons survive through it. Almost no pions (one macro-particle in this simulation) remain at the 400 MeV energy. This plot also highlights a feature of the decay channel that the losses between the solenoids are almost entirely muons because these have the unpredictable directions resulting from pion decay. The original pion beam gets scraped into shape within the first ten solenoids meaning almost no pion losses after that.

Work moved on from the chicane idea after this optimisation because of the large losses. Figure 8.13 shows how the eight optimised designs discussed in this chapter retain their beam. Large losses are always present at the end of the wider, 30 m-long decay channel but the chicane takes the transmission three or four times lower than comparable points in other designs. Dog legs on the end of these graphs show the final reduction in yield from wrong-energy muons being rejected.

Table 8.5: Optimisation ranges and best muon yield for the `PhaseRotB` lattice.

Decay channel (to 30 m): *as for* `ChicaneLinacB` *(table 8.4)*		
Phase rotator:		
Structure	Number of cells	20–40
Solenoid	Length	0.2–0.9 m
	Radius	0.15 m (fixed)
	Field	−4–4 T
Drift	Length	0.5–1 m
RF	Frequency	31.4285 MHz (fixed)
	Voltage	0–$(L_{drift} + L_{sol}) \times \frac{45}{28}$ MV/m
	Phase	−180–180°
	Proton beam energy:	Muon energy criterion:
	2.2 GeV	180±23 MeV
Best yield: μ^+/π	μ^+/p·GeV	μ^+/s (4 MW proton beam)
3.332%	0.002528	6.31×10^{13}

8.3 Phase Rotation

The large energy spread of the beam in the decay channel causes problems for a lot of conventional accelerator designs, as focussing varies greatly between particles. To address this problem, a phase rotation section uses RF cavities to rotate the slanted longitudinal distribution (seen in figure 8.2) so that more of the particles finish in a smaller energy band, usable by systems downstream. It can achieve this mainly because after the drift in the decay channel, there is a correlation between arrival time and velocity, so if the RF voltage is made decelerating for the head of the bunch containing high energies and accelerating for the tail, this will cancel much of the variation. This is only effective over a limited range because of the sinusoidal nature of the RF voltage variation, thus ideally a phase rotation section is optimised to put this range in the densest region of muons.

It can also be said that phase rotation is doing the opposite of the bending chicane in section 8.2, which tried to reduce the variation in time of arrival, while the energy spread was kept necessarily the same as a consequence of using a purely magnetic system. Flight times were varied as a function of energy in the chicane, whereas RF systems can vary energies as a function of time of arrival.

The `PhaseRotB` optimisation specified in table 8.5 has the same decay channel as `ChicaneLinacB` and a phase rotation section similar to the linac from that lattice except using a lower frequency. The aim is not to produce an accelerating RF bucket but to have the RF period comparable to the spread in muon time of arrival from the decay channel. A frequency of 31.4285 MHz was chosen for compatibility reasons with the Rees cooling ring [36] but is a good choice irrespective of the downstream system (the frequency was not optimised because that requires parameterising the RF gradient limit, which is 2.25 MV per 1.4 m of beamline here). The goal kinetic energy spread of 180±23 MeV was also inherited from the ring's acceptance but the important point is that it is a lot

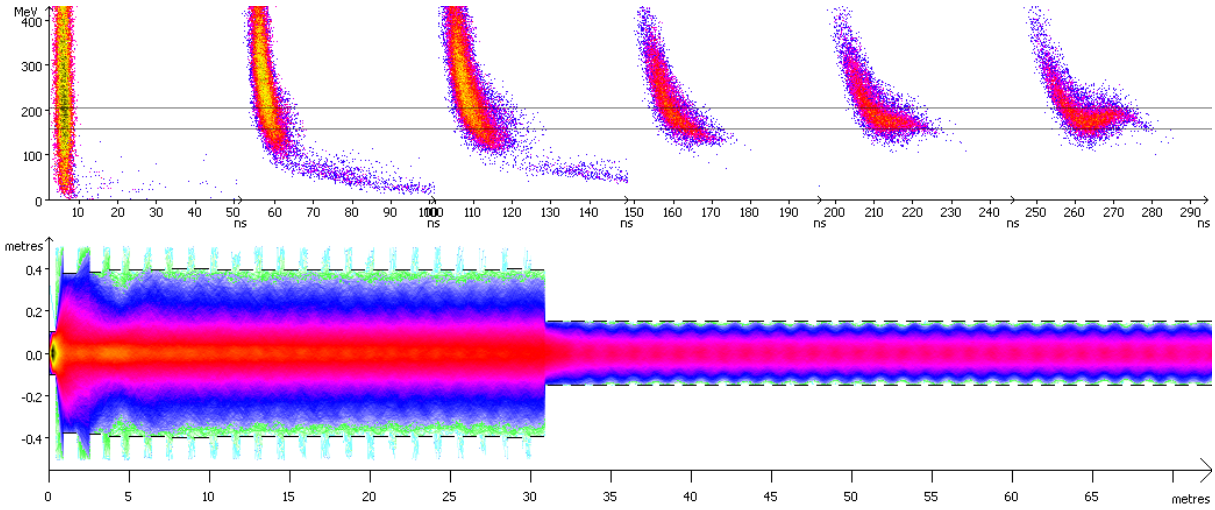


Figure 8.14: Longitudinal phase space evolution (top) and beam envelope plot (bottom) for the optimised design from the `PhaseRotB` lattice.

Table 8.6: Optimisation ranges and best muon yield for the `PhaseRotC` lattice.

Decay channel (to 30 m): *as for* `ChicaneLinacB` *(table 8.4)*		
Phase rotator: *as for* `PhaseRotB` *(table 8.5)*		
	Proton beam energy: 10 GeV	Muon energy criterion: 180±23 MeV
Best yield: μ^+/π 3.042%	μ^+/p·GeV 0.003435	μ^+/s (4 MW proton beam) 8.57×10^{13}

narrower than that for the 400 MeV linac in `ChicaneLinacB` and has a central energy near the mean of the particles in the decay channel. It provides a figure of merit that rewards narrowness in energy spread without requiring net acceleration or deceleration.

With the bending chicane removed, the muon transmission increased from 2.65 to 6.31×10^{13} muons per second into a proportionately narrower energy range. Figure 8.14 shows the resulting design and how the RF section flattens the energy distribution to get more muons into the goal range. In fact, even more can finish in this range if there is some slight over-flattening, resulting in the S-shaped energy distribution in the rightmost plot. The transverse plot shows that beam sizes in the RF section are similar to those in the `ChicaneLinacB` linac but many more particles have been retained, the regions of increased density represented by purple and red colours.

It was at this point that new results regarding the optimal proton beam energy on the target (see section 4.2) became available. As well as being integrated into the Scoping Study [24] recommendations, the new preferred value of 10 GeV was used in the Muon1 optimisations by replacing the old pion data file by a MARS calculation at that energy instead of 2.2 GeV. The data files were both generated with the same version MARS15.

The `PhaseRotC` lattice changes the proton beam energy to 10 GeV but leaves everything else from the `PhaseRotB` optimisation the same, as shown in table 8.6. It therefore

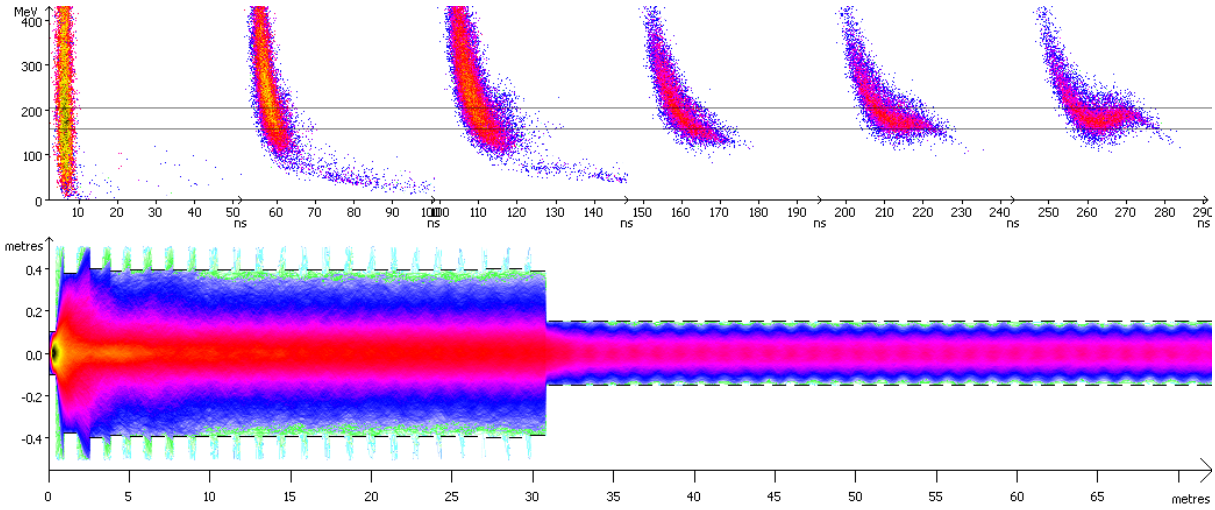


Figure 8.15: Longitudinal phase space evolution (top) and beam envelope plot (bottom) for the optimised design from the `PhaseRotC` lattice.

provides a direct comparison of the effect of changing the proton beam energy inclusive of the front end to the end of phase rotation being optimised. Perhaps surprisingly, the efficiency in terms of muons per original pion has decreased to only 0.913 times of that with the 2.2 GeV beam but the real yield in μ^+/p·GeV has increased by a factor of 1.359. This difference comes from the fact that more pions were produced per Watt of proton beam power in the target at 10 GeV, while the decrease in relative efficiency comes from proportionately more pions being produced in areas of the momentum phase space where they are not captured. The high energy tail is a likely candidate for the location of this wastage, since more kinetic energy is available from the 10 GeV protons in early collisions to produce high energy (multi-GeV) pions, which will not be captured by the phase rotation system.

The resulting yield of 8.57×10^{13} muons per second from `PhaseRotC` is becoming close to the 10^{14} neutrino factory requirement, though the later bunching and acceleration systems are at this point unspecified. The optimised design (listed in appendix E) has only subtle differences to that for `PhaseRotB`, leaving most of the yield improvement accounted for by the proton beam energy change.

Figure 8.15 shows that the transverse profile of the beam is very similar to the `PhaseRotB` channel in figure 8.14. The only visible difference being that `PhaseRotC` has a greater density of particles present in the first ∼5 solenoids of the decay channel near the target, as would be expected if the target produced more pions that were later lost. The pion/muon breakdown is shown in figure 8.16, with large pion losses visible between solenoids `S2` and `S3`; the continuation of pion to muon decay in the phase rotation section is also apparent.

Both of the `PhaseRotB` and `PhaseRotC` lattices had almost every attribute of each component in the beamline variable in the optimisation. This led to a large number of variables, 279 in the case of `PhaseRotB` and also a slow pace of optimisation: these lattices were still increasing in score very gradually after as much as 6 months running on the distributed network while individual components were tuned to gain minor increases in

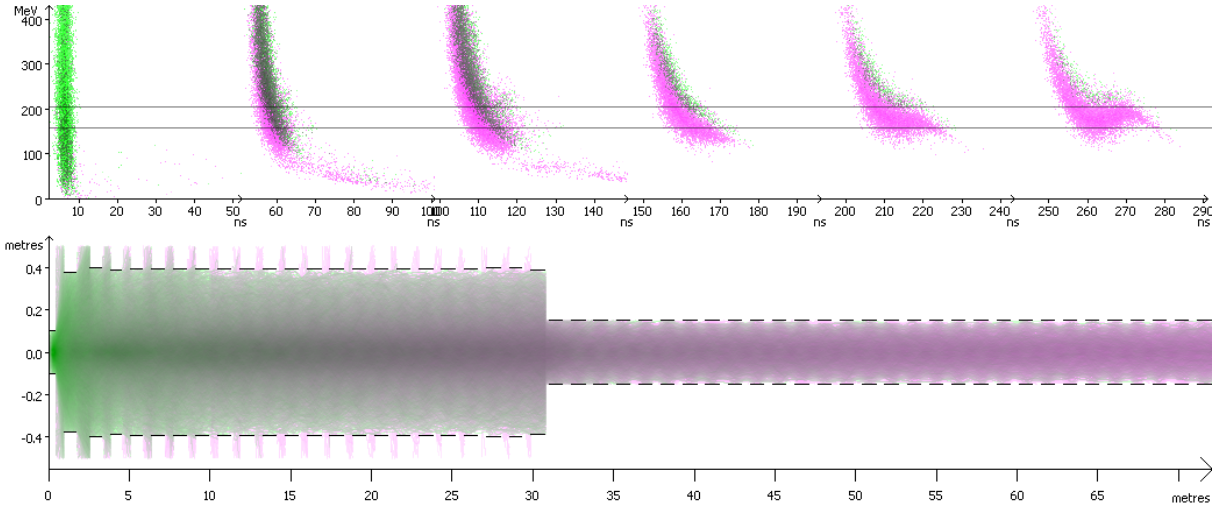


Figure 8.16: Pions (green) decaying to muons (purple) in the optimised design from the `PhaseRotC` lattice.

Table 8.7: Optimisation ranges and best muon yield for the `PhaseRotEb6a` lattice.

Decay channel *as for* `ChicaneLinacB` *(table 8.4) except:*		
Structure	Number of cells	15–50
	Blocks	6 after `S9`
Phase rotator *as for* `PhaseRotB` *(table 8.5) except:*		
Structure	Number of cells	6–60
	Blocks	6
	Proton beam energy:	Muon energy criterion:
	10 GeV	180±23 MeV
Best yield: μ^+/π	μ^+/p·GeV	μ^+/s (4 MW proton beam)
3.480%	0.003930	9.81×10^{13}

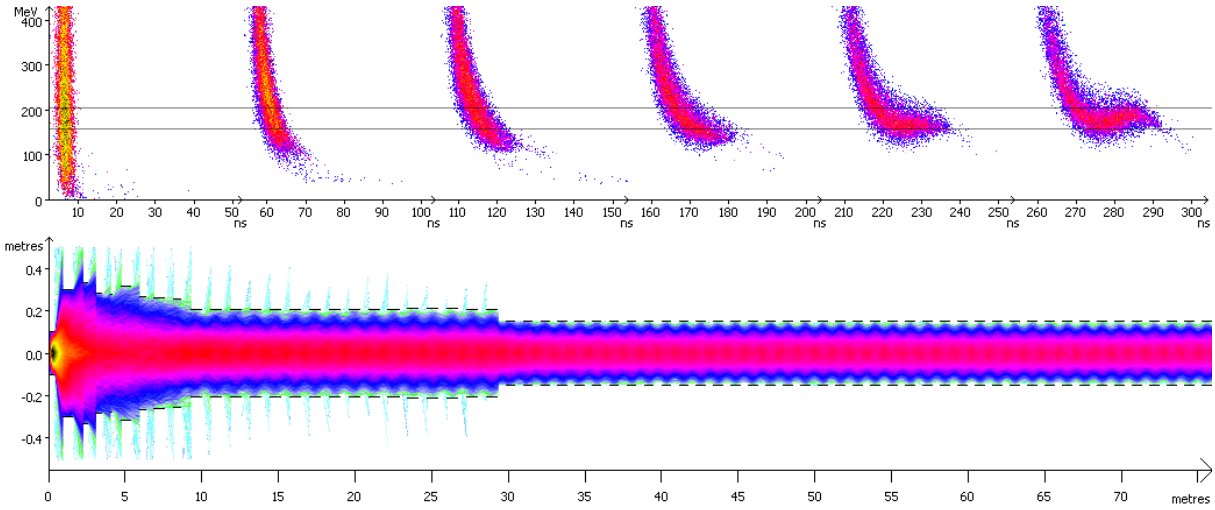


Figure 8.17: Longitudinal phase space evolution (top) and beam envelope plot (bottom) for the optimised design from the `PhaseRotEb6a` lattice.

yield. It was expected, then, that a reduction in the number of variables could quicken the pace of convergence. This was attempted in the lattice `PhaseRotEb6a`, presented in table 8.7, using the technique of 'blocking' some of the many cells in the late decay channel and phase rotation section together.

The blocking works by observing that long homogenous sections of beamline are usually optimised to regularly repeating cells with almost the same parameters as each other even though they are controlled be independent variables in the optimisation. Sections with RF may understandably incorporate some variation through the channel as the longitudinal phase space shape of the beam changes, however on local scales (i.e. without significant drift) the effects of RF are nearly additive in voltage, so a slow variation in cell parameters is expected to suffice. Only special sections such as matching near the target require further cell-to-cell adaptation. So with a blocked lattice, the majority of the beamline will be specified as blocks of identical cells that share the optimisation variables for only one cell. The `PhaseRotEb6a` lattice (other lattice variations were tried but it worked the best) splits the late decay channel starting from solenoid `S10` and the whole of the phase rotation section each into 6 blocks of equal length, the overall length of these sections being controlled by the `decaycells` and `phaserotcells` variables. Allowing the decay channel to vary in length is new but neither section can shorten to the point where each block consists of less than one cell. The number of optimisation variables was reduced threefold to 93.

The yield from this lattice was a further improvement to 9.81×10^{13} muons per second, essentially meeting the required rate. Figure 8.17 shows how the optimiser has chosen a decay channel of just under 30 m long without being forced to this length as in previous lattices. The only constraint was that the number of solenoids in the decay channel had to be between 15 and 50. This design also tapers the radii of the solenoids earlier on, something that may be favourable from a cost perspective. The overall technique of the phase rotation remains similar but with a slightly longer time interval of particles ending up in the goal energy range, leading to the increased yield.

8.3.1 Alternative Design for Two Signs

As shown in figure 3.3 and explained throughout chapter 4, both positive and negative pions are produced in the target in similar numbers. These will decay to both signs of muon in the decay channel, the motion of the negative particles being the mirror image of the positive ones provided the field remains axially symmetric. This can be proved by inspecting the cross product in the Lorentz magnetic force law $\mathbf{F} = q(\mathbf{v} \times \mathbf{B})$ under a change of sign in v_x, B_x and q and confirming the result is a change in the sign of F_x only. Thus the decay channel will work similarly on both signs of particle.

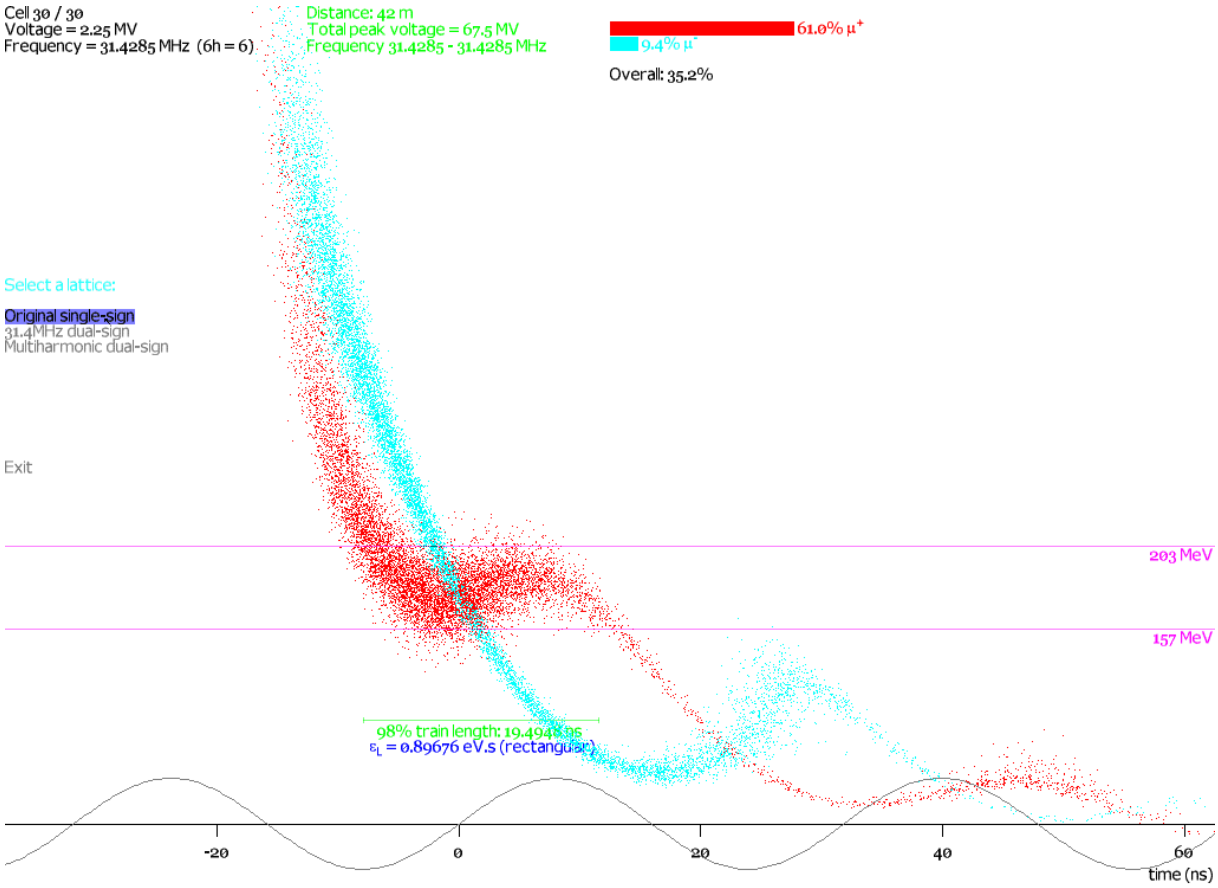


Figure 8.18: Effect of the phase rotator optimised for only one sign of muons (red) on the negative sign (blue).

The same is not true of the RF phase rotation sections because these will decelerate negative particles in the time periods when positive particles are accelerated. The result of a single-bunch phase rotation section (from lattice `PhaseRotC`) acting on both signs of muon is shown in figure 8.18. The positive muons have been rotated to a narrow energy range but the variation in energy of the negative muons has been doubled because the RF rotation has been acting in the wrong sense. This results in the very poor longitudinal transmission of negative muons (9.4%) into the goal energy band, as shown at the top of the figure. The results in this section were calculated with a fast 1D-only code so efficiencies do not include transverse losses.

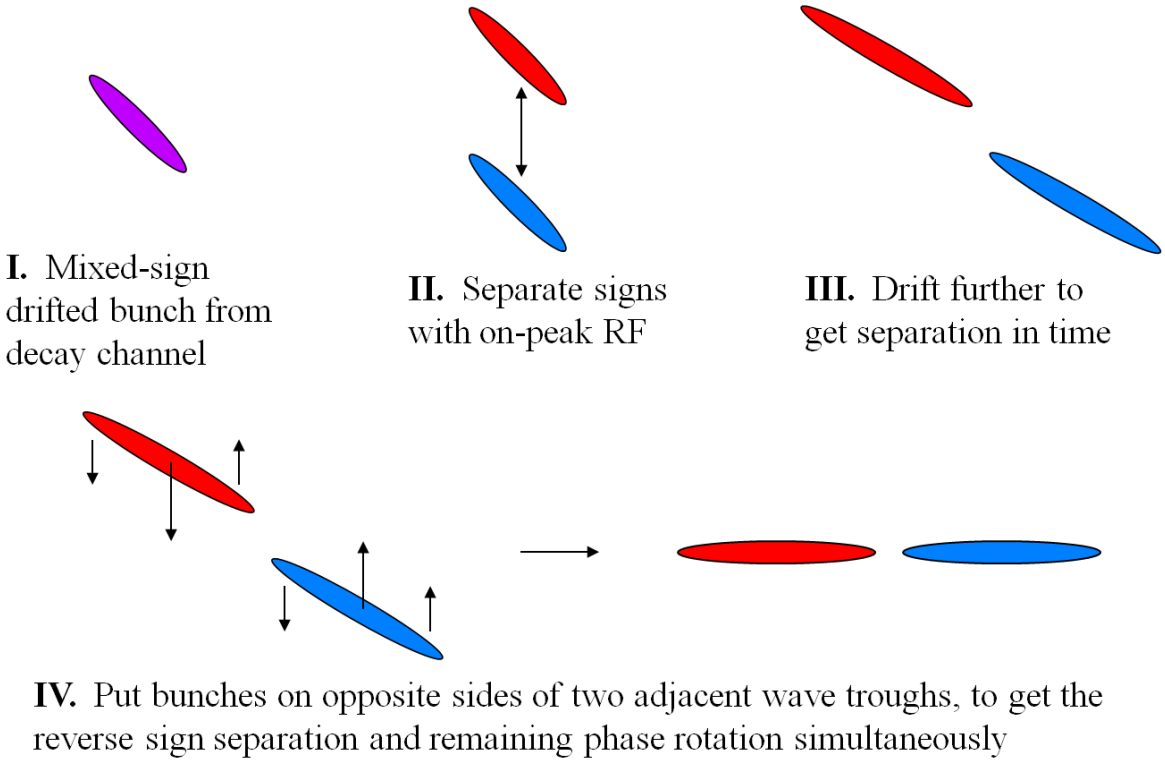


Figure 8.19: Proposed system for rotating both signs of muons with the same RF, even though they do not start with a 180° phase advance between them.

If the negative muons could somehow be delayed by 180° of the RF phase, then the **E** fields experienced by them would be the exact opposite of those for the positive muons and identical longitudinal dynamics would result, provided only one RF frequency was used in the channel. Time of flight differences could be produced by a dipole sending one sign of muons onto longer path-length trajectories but this is technically difficult with such a high-emittance beam. Another option that does not change the transverse dynamics, using purely RF, is illustrated in figure 8.19. This technique notices that a time of flight difference between the signs can be obtained after a drift section by accelerating one sign and decelerating the other, which is exactly what an on-peak RF cavity would do to the initial mixed sign beam. Once the two signs are suitably far apart the RF will no longer produce the opposite effect on each and there exists a phase that will reverse the acceleration and deceleration of the two signs while still rotating each bunch in the correct direction. This corresponds to a phase difference not equal to 180° in which the bunches are on the same sign of the RF voltage waveform but experience opposite slopes (so the effect will be opposite net accelerations with the same rotation direction).

One-dimensional simulations of such a scheme were attempted with the same input distribution of muons from the end of the decay channel and gave results such as shown in figure 8.20. In this scheme the longitudinal transmission of the two signs is more equal and the average transmission of both is higher. The disadvantage is that more complex phase space manipulations require a much longer RF channel, 183 m instead of the 42 m for the single-sign design, which would result in increased cost.

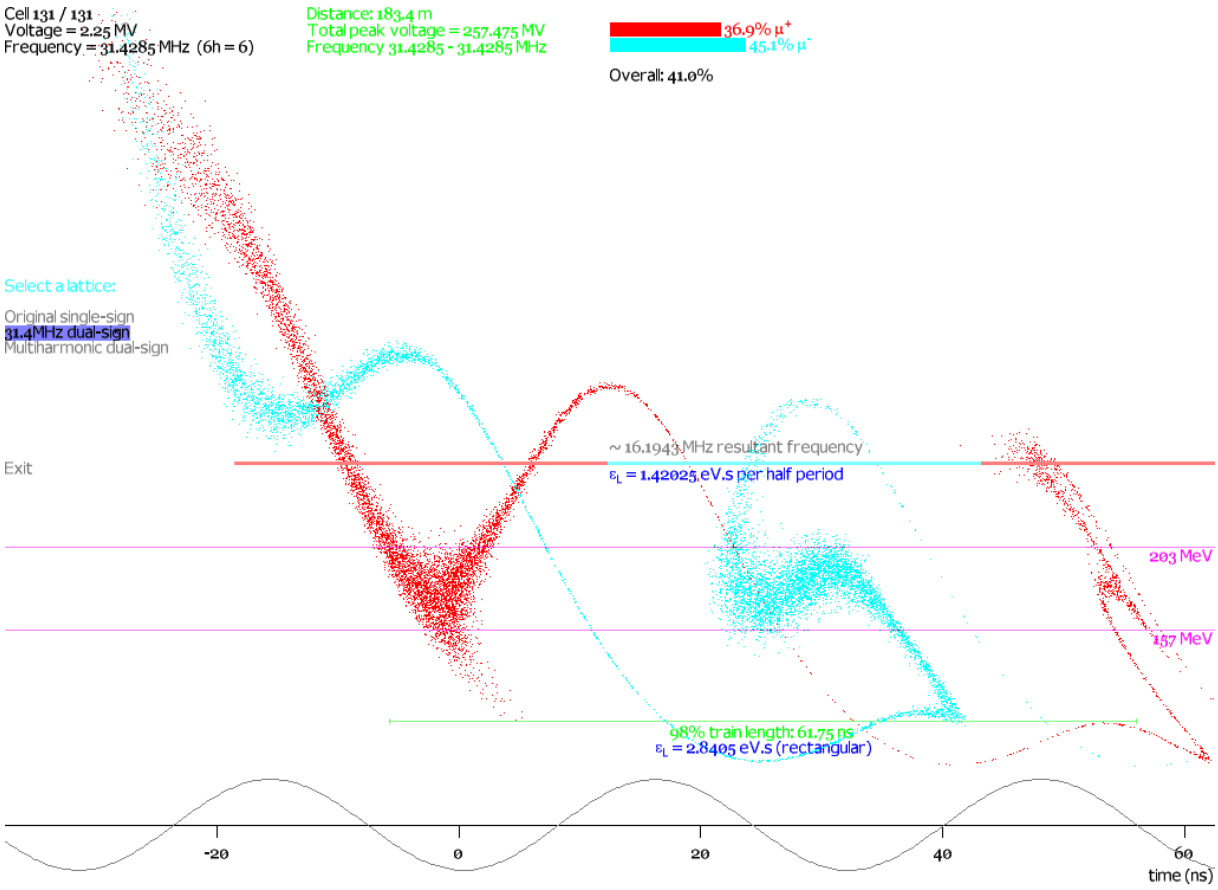


Figure 8.20: Improved phase rotation forming bunches of both signs of muons.

8.4 Phase Rotation and 400 MeV Linac

Using RF phase rotation instead of a chicane and changing the proton beam energy to 10 GeV improved greatly on the `ChicaneLinacB` design, but so far the goal energy has been 180±23 MeV, which does not include any overall acceleration of the beam. A more direct comparison could be made if a linac were added and the goal energy range changed back to the 400±100 MeV used in the chicane and linac design.

Table 8.8: Optimisation ranges and best muon yield for the `PhaseRotLinac1` lattice.

Decay channel *as for* `ChicaneLinacB` *(table 8.4) except:*		
Structure	Number of cells	10–50
	Blocks	1 after `S9`
Phase rotator *as for* `PhaseRotB` *(table 8.5) except:*		
Structure	Number of cells	1–60
	Blocks	1
Linac *as for* `ChicaneLinacB` *(table 8.4) except:*		
Structure	Number of cells	60–120
	Blocks	1
	Proton beam energy:	Muon energy criterion:
	10 GeV	400±100 MeV
Best yield: μ^+/π	μ^+/p·GeV	μ^+/s (4 MW proton beam)
6.944%	0.007840	1.96×10^{14}

The `PhaseRotLinac1` lattice, summarised in table 8.8, simply adds the 88 MHz linac from `ChicaneLinacB` onto the end of the existing phase rotation design. The linac now can inherit the blocking technique from the phase rotation (in `PhaseRotEb6a`) to reduce the number of variables. Setting the number of blocks to just one per section proved to be the most successful of various combinations tried, so the late decay channel, phase rotation and linac now each have a single cell that repeats throughout the structure.

The optimum (printed in appendix E) has an intermediate length decay channel with 39 solenoids covering a distance of just over 50 m, a maximal length 60 cell phase rotator but a minimal length 60 cell linac. The 88 MHz linac has become minimal in one additional way: all the RF voltages have been set to zero! This means that the resulting design, shown in figure 8.21, performs all its longitudinal manipulation using the 31.4 MHz cavities intended for phase rotation. This is possible because no constraint was ever imposed that theses cavities should be used at bunching phase rather than for net acceleration or deceleration. The 88 MHz linac section could therefore be removed to reduce the length of the design shown to under 140 m and give a marginal increase in yield by avoiding some muon decay. This rejection of the higher RF frequency is likely because once the hybrid phase rotation/acceleration is able to move the beam into the target energy range on its own, imposing 88 MHz RF just produces corrugations that overall move beam outside the range. For a longer linac, having fully formed bunches is a necessity and this happens in the 900 MeV linac optimisation of section 8.5.

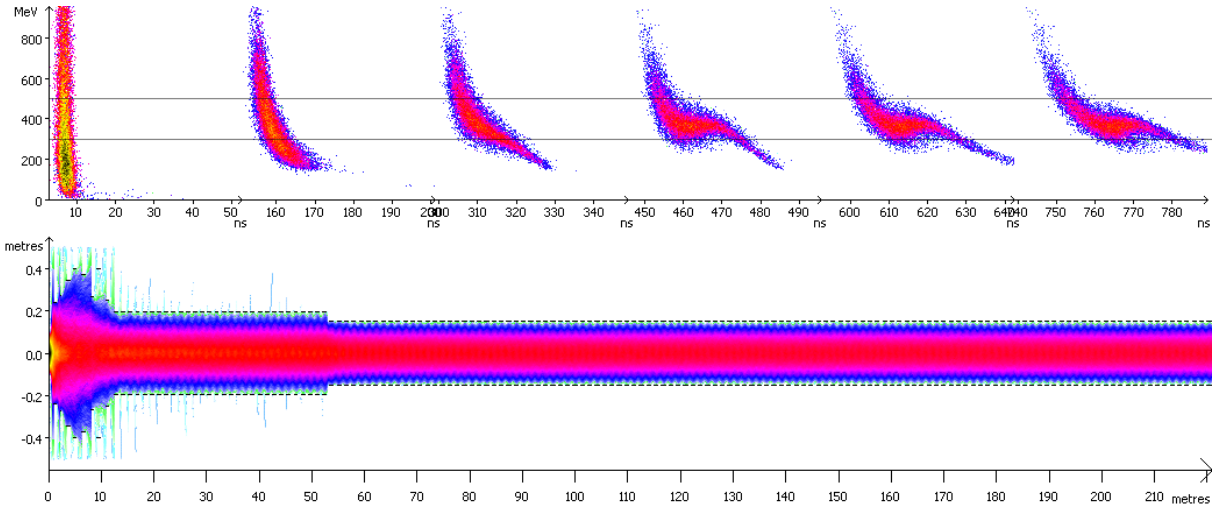


Figure 8.21: Longitudinal phase space evolution (top) and beam envelope plot (bottom) for the optimised design from the `PhaseRotLinac1` lattice.

The yield of 1.96×10^{14} muons per second is nearly twice the requirement for the neutrino factory. This has come about mainly by widening the energy range to be similar to `ChicaneLinacB`, which was intended to hold tall 88 MHz RF bucket(s) that are narrow in time. By contrast, this optimum has filled the band with a wide distribution of times using the lower frequency; any linac built beyond this design would have to use a similarly low frequency to catch this distribution in a single bunch, the alternative being a rebunching section at a higher frequency (like 201 MHz as used in [25]), which would add costs in terms of muon decays.

8.5 Direct 900 MeV Linacs

The goal energy range of 400±100 MeV, while useful for a comparison with `ChicaneLinacB`, is large compared to the energy ranges accepted by the downstream accelerators. Based on the ISS report [24] section 5.2, the first recirculating linac is assumed to accept an input kinetic energy of 0.9 GeV with a momentum spread $\Delta p/p$ of 7%. This translates into a kinetic energy range of 830.41–969.64 MeV, roughly 140 MeV wide instead of 200 MeV.

A longer linac with this goal energy is specified in optimisation `Linac900Removable6c2`, detailed in table 8.9. The phase rotation at 31.4 MHz and linac at 88 MHz are still present but the linac can be made much longer. All sections of the beamline can now be removed completely by the optimiser (apart from the first solenoid!) in the pursuit of higher yields.

The result this time is that the decay channel has been reduced to just four solenoids and the phase rotation segment has been removed entirely. Figure 8.22 shows that this design does all the work using the 88 MHz linac structure, which has been given the maximum allowed number of cells. Unlike `PhaseRotLinac1` in the previous section, this linac forms something much more resembling a conventional RF bucket in the longitudinal phase space. The yield of 1.07×10^{14} muons per second also meets the neutrino factory requirement, although the linacs beyond this design must have a wider bore to

Table 8.9: Optimisation ranges and best muon yield for the `Linac900Removable6c2` lattice.

Decay channel *as for* `ChicaneLinacB` *(table 8.4) except:*		
Structure	Number of cells	1–50
	Blocks	Up to 6 after `S9`
Phase rotator *as for* `PhaseRotB` *(table 8.5) except:*		
Structure	Number of cells	0–60
	Blocks	Up to 6
Linac *as for* `ChicaneLinacB` *(table 8.4) except:*		
Structure	Number of cells	0–240
	Blocks	Up to 6
RF	Phase	-360–$360°$
	Proton beam energy: 10 GeV	Muon energy criterion: 900MeV (k.e.) $\pm 7\%$ $\Delta p/p$
Best yield: μ^+/π 3.786%	μ^+/p·GeV 0.004274	μ^+/s (4 MW proton beam) 1.07×10^{14}

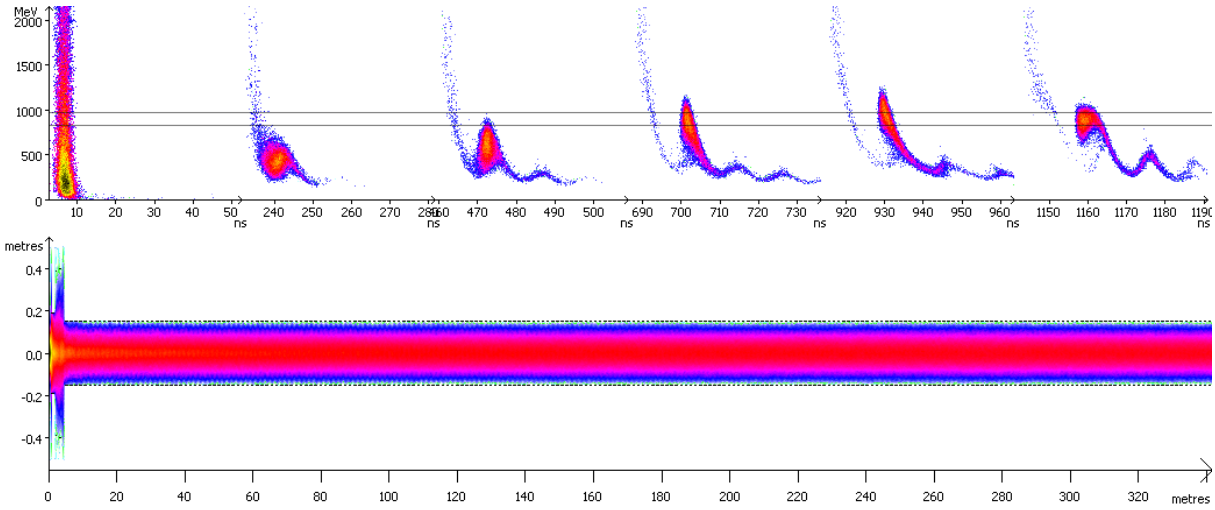


Figure 8.22: Longitudinal phase space evolution (top) and beam envelope plot (bottom) for the optimised design from the `Linac900Removable6c2` lattice.

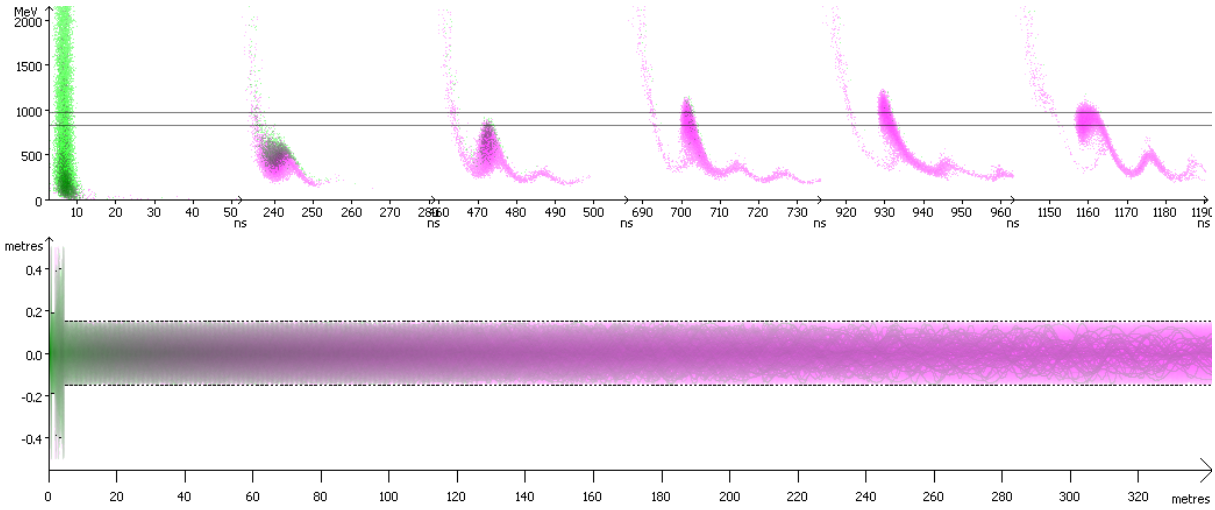


Figure 8.23: Pions (green) decaying to muons (purple) in the optimised design from the `Linac900Removable6c2` lattice.

accommodate an uncooled beam with normalised emittance as given in table 8.3.

This yield is lower than `PhaseRotLinac1` both because of the stricter final energy criterion and because the muons have to survive a nontrivial amount of acceleration. However, the 31.4 MHz combined phase rotation and acceleration approach used in `PhaseRotLinac1` could likely not be extended to higher energies because a stable RF bucket was never formed. This is why the entire linac was optimised again with the new energy goal and not extended from the previous system.

The lack of decay channel means that this accelerator copes with multiple species (pions and muons) at the same time. Figure 8.23 colours the beam by composition and gives an illustration of how this works in practice, with the pions occupying a higher kinetic energy level bucket than the muons. This is because the charged pions have a rest mass of 139.57 MeV/c^2 compared to the muons' 105.66 MeV/c^2, which means to have the same velocity (equivalently β or γ) they must be given 32% more kinetic energy. The velocity is what determines the RF phases when the particles arrive at the cavities so scaling up the kinetic energies in this fashion ensures correct phasing for both species, ignoring transverse path length effects and the fact the RF voltage will be less effective compared to the energy the pions already have in rotating the contents of the bucket. Fortunately the bucket only undergoes a fraction of a rotation before most of the pions have decayed so the effect of this discrepancy in bucket synchrotron frequency is small.

Removing the decay channel may not be practical to the extent the optimiser has done, since the area in the first few solenoids near the target has particularly high losses and radiation and this may not be conducive to stable operation of the RF cavities.

It is noticeable in the transverse plot of figure 8.22 that the beam does not undergo any adiabatic shrinkage in x as it is accelerated. Given the relation between un-normalised and normalised emittance, this may seem counterintuitive, but the shrinkage has happened elsewhere: figure 8.24 shows the distribution of x' in the beam along the length of the linac. Reduction of x'_{rms} is visible from the 10 m point until the end of the linac, with the change in distribution shape in the 4^{th} and 6^{th} sixths of the linac caused by flipped-sign solenoids

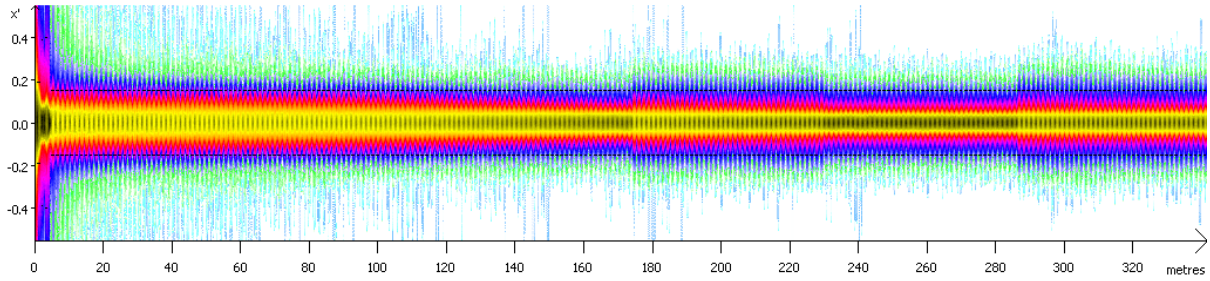


Figure 8.24: Plot of the x' distribution of the beam against z for the optimised design from the `Linac900Removable6c2` lattice.

being used there. Reduction before about 10 m is not considered genuinely adiabatic because it is mostly beam losses in the decay channel, where there is no acceleration.

This behaviour comes about because the magnetic field is constant throughout the linac and geometrically the RF acceleration will reduce x' by increasing p_z but leave x and p_x alone. In synchrotrons the beam shrinks in both x and x' because the focussing structure and Twiss parameters are kept constant by raising magnetic fields in proportion to the beam momentum. With linacs there is more flexibility and the emittance may be traded off between x_{rms} and x'_{rms} by varying the focussing field. Higher focussing, best attained with quadrupoles for higher energies, favours smaller x_{rms}.

Figure 8.25 visualises this same effect as an increase in β along the channel. Notice the reduction of un-normalised 4D emittance in figure 8.26 as the beam is accelerated, while normalised emittance stays constant once all the pions have decayed. The 2D emittances in this plot oscillate above the 4D values because of x–y correlation caused by the periodic solenoid fields. The sharp decrease of emittance at the beginning is due to losses as the beam enters the narrower linac section very soon after the target.

The component listing in appendix E shows that the first four of the six linac blocks have roughly the same RF parameters but the fifth turns off the RF voltage almost entirely. Then the final block switches it back on but at a completely different phase. The effect of this is the 'lopsided' rotation of the particles in the bucket shown in figure 8.22, with the different RF phase in the final sixth of the linac rotating particles distributed along a diagonal in phase space into the goal band. This nonconstant phase could allow a longer bunch of particles to finish in the goal band than would normally be expected, by shearing the distribution to reduce momentum spread and increase time spread.

Overall then, the optimised design from `Linac900Removable6c2` is a muon front end (without cooling) and initial acceleration system to 900 MeV, producing over 10^{14} muons per second of a single sign. Optimisation of a two-sign system similar to that in section 8.3.1 is possible but not yet attempted with the full Muon1 code, so it is not known how many opposite sign muons can be accelerated. For comparison, the ISS baseline [24] includes a cooling channel so ends with an RMS emittance of 7400 mm·mrad, which is about half of the 14812 mm·mrad produced here (see table 8.3 for more data). The ISS baseline transmits 0.0083 μ^+/p·GeV compared to 0.0043 in this design, though this value varies strongly with the linac energy range: for example `PhaseRotLinac1` produces 0.0078 μ^+/p·GeV in a range based on an 88 MHz linac rather than the ISS RLAs.

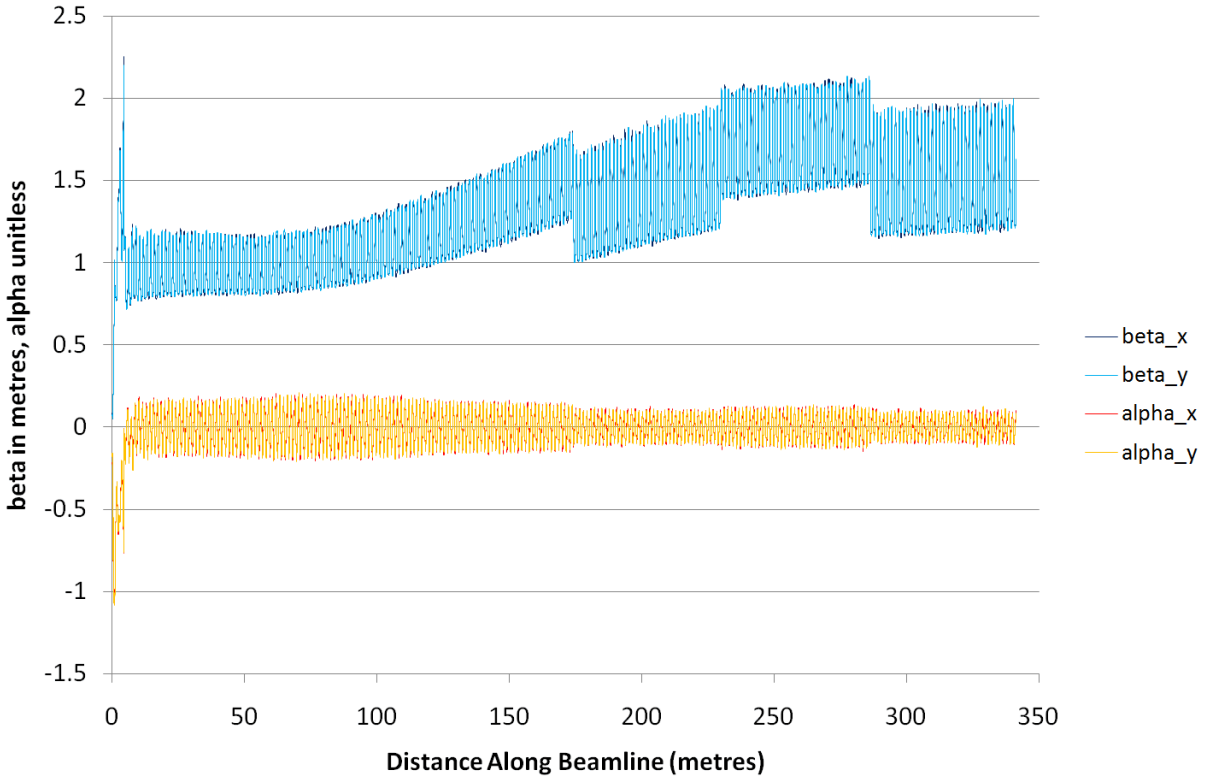


Figure 8.25: Optical functions in the optimised `Linac900Removable6c2` design.

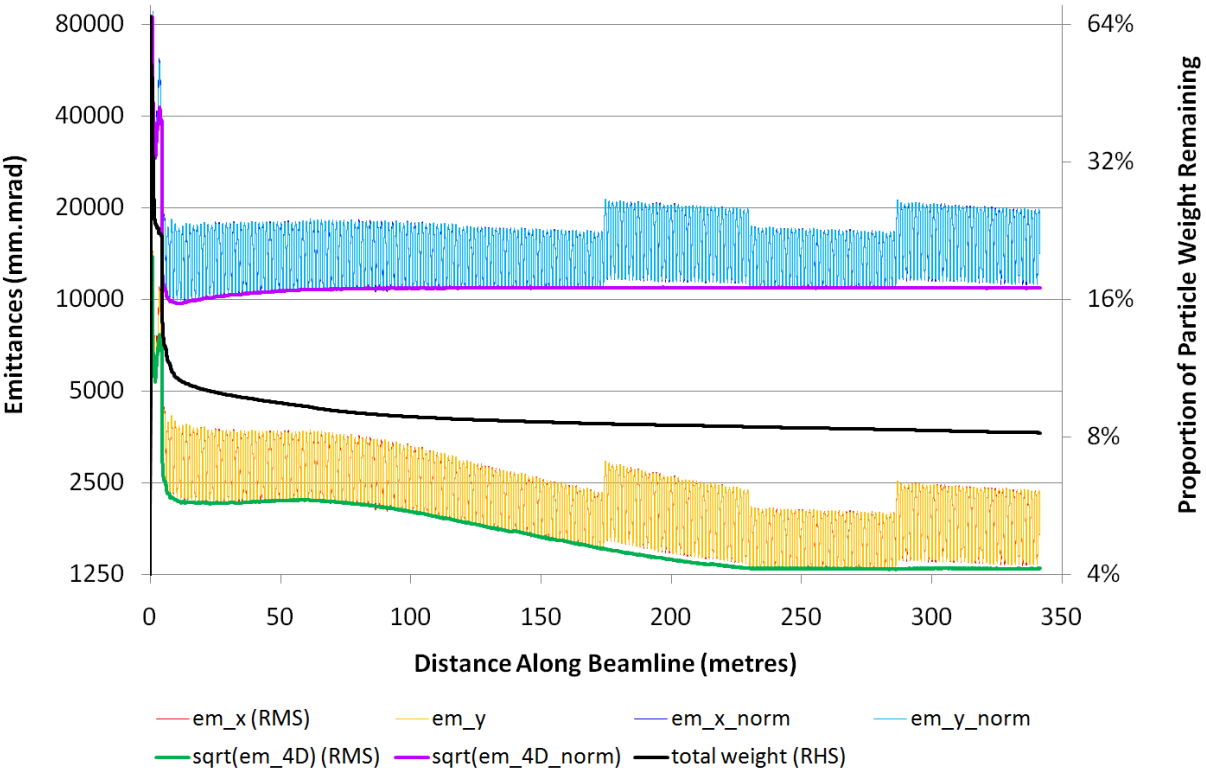


Figure 8.26: Emittance evolution in the optimised `Linac900Removable6c2` design.

Chapter 9

Conclusion

This thesis has described the design of a pion production target and muon front end for the neutrino factory, which meets the 10^{14} muons per second criterion derived from 10^{21} muons per operational year [1]. In the process, the effect of target design variations on useful pion yield has been characterised and a design system using a distributed genetic algorithm has been created to produce entire accelerator designs automatically. Therefore the work can be interpreted from the viewpoint of the three disciplines below.

Accelerator design. The baseline neutrino factory design [24, 25] has been well studied but includes components such as the muon cooling channel and a 6 m long magnet surrounding the target (figure 3.7) that may prove challenging to build. The designs in this thesis have no muon cooling and enforce 50 cm gaps between all solenoids, which sacrifices some muon yield, in exchange for a target area that can be taken apart and fewer beam windows to contain cooling material. The solid target option has also been used instead of the liquid mercury jet. The designs presented have taken an acceptable number of muons to 900 MeV, so it is suggested that the muon linac continues all the way to 25 GeV, providing a much simpler design than the baseline (figure 1.3) at a length and cost that is not necessarily much greater, as explained at the end of section 1.4.

Tracking of this design has been done with the author's Muon1 code that is fully 3D, includes particle decays and uses a particle set generated by MARS15 [2] from the target rather than an idealised beam distribution. Section 8.3.1 presents some promising work using a 1D code that shows adaptations of this design can accelerate both signs of muon.

Target design. The recommended design parameters for the pion production target are given in section 4.7. However, many of these 'optimal' values are surrounded by quite broad ranges in which the pion yield is almost as high. Thus, readers interested in designing a target station may find the graphs in sections 4.2–4.5 more useful since they allow other engineering constraints (of which there will be *many*) to be traded off against performance. The considerations that can be identified at the current level of design detail are discussed in chapter 3, including how the proton accelerator interfaces via the target with the rest of the neutrino factory. However, the chapter can only be an introduction to the problem of target design, which is a field of study in its own right [87].

The target designs were compared using calculations from the material interaction

code MARS15. However, hadronic interactions are not yet a 'solved problem' in particle physics and large discrepancies can remain between the empirical models available. Hence chapter 5 attempts to quantify the degree of uncertainty by comparing the MARS15 results with GEANT4 [3] and data available from the HARP experiment [4]. The codes agree quite well on the optimal proton energy for the neutrino factory, though the absolute pion yields from GEANT4 are too high by a factor of two. MARS15 is shown to have better (25% level) agreement with the HARP data in the region available for comparison.

Scientific computing. The search for high-yielding muon front end designs was made as automatic as possible by the use of computer optimisation. The user necessarily has to specify the figure of merit and boundary conditions such as the particle source, practical range of sizes for components and general structure of the beamline, with the rest of the problem being recast as maximising a function. The maximisation is done by a genetic algorithm, which requires many simulations, so the code was made to run on a distributed computing network to perform the large amount of work in parallel. The general description of a range of accelerators can be written in the Muon1 lattice file format and submitted to this network and optimisation will commence. The optimiser has been relatively successful on problems with hundreds of variables and the designs presented in chapter 8 are all examples of this.

Notably, the distributed network consisted mostly of volunteers who downloaded the client program from the project's website [5], with a few of the author's computers also running. The project has more than 100 users active at any one time, with thousands having run the program at some point and over 60 million muon front end designs have been simulated. This demonstrates the enthusiasm of the public for interacting with scientific projects, even complex ones like the neutrino factory.

Summary and Implications

The design of the front end in section 8.5 together with a muon linac to 25 GeV provides a useful complement to the current main thread of neutrino factory design work being pursued in the International Design Study (IDS) [73]. Neutrino factory design work globally has progressed through several stages, initially concentrating on increasing the muon yield as much as possible, leading to US Feasibility Study II [25], with a relatively aggressive design for a cooling channel using liquid hydrogen. The next stage focussed more on maintaining an acceptable muon yield while reducing the estimated hardware cost, thus came the lithium hydride cooling channel of the International Scoping Study (ISS) [24] and the replacement of the high-energy end of the recirculating linac (RLA) by an FFAG, which provided more turns through the same RF cavities and therefore reduced the number of cavities required to achieve a set level of acceleration.

Unfortunately, yield *vs.* cost was not the only trade-off being made during these design revisions. The rearrangement of the acceleration system to allow more passes per RF cavity has come at the cost of a much more complex machine. The current IDS baseline contains four muon accelerators of three different types (linac, RLA1, RLA2, FFAG), compared to the two types (linac, RLA) found in Study II. The RLAs require magnet switchyards at each end, all of which are different from each other, and the FFAG performs

an unconventional form of RF acceleration that relies on small time-of-flight differences in the orbit during acceleration.

It is possible that all of the components in the current IDS baseline design will work without major faults. However, the fact that complexity increases *risk* should not be ignored. That is partly the motivation of the proof-of-principle experiments MICE [72], EMMA [88], FETS [89] and MERIT [39], which test various parts or analogous models for the muon cooling channel, FFAG, proton driver linac and target, respectively. At the present point in time, the risks of the neutrino factory design cannot be addressed properly without admitting the possibility that one or more of those experiments may produce a negative result (or identify problems in scaling up to a full-size or full-power subsystem).

The design in this thesis, while not the best in muon yield or possibly cost, relies on fewer unconventional technologies than the current baseline and is simpler. No muon cooling or FFAG is involved at any point. Large, low-frequency RF cavities similar to those in cooling channels are used, but there is no scattering material or cryogenics involved, and the cavities are placed between solenoids rather than at the peak field, which has been shown to reduce cavity performance [90]. Linacs are proven technology and will not differ greatly for muons instead of protons, the only requirement being a larger bore at the low energy end, which other neutrino factory designs also have. The pion production target is unavoidably difficult because of the high powers involved but this thesis suggests how it could be reduced to the problem of a conventional solid target, provided the solid targets can be moved for each pulse.

In conclusion, achieving the 10^{14} muons per second goal with the simplified design presented here provides an important piece of risk management for the neutrino factory project. If the MICE or EMMA experiments point to problems with their novel technologies, fall-back designs that work without the corresponding subsystems are feasible. The solid target option provides an alternative to MERIT's liquid target (although it has its own distinct set of challenges) and the range of acceptable proton driver energies in section 4.2 provide additional flexibility to evaluate and make use of high-power proton facilities that may be built in the future.

Future Work

For the neutrino factory to be realised, a great deal (hundreds of person-years) of detailed engineering will have to be done. There are also some important conceptual questions that must be resolved first: what sort of target (solid or liquid) is desired? What are the implications for cost of a simplified design such as the one in this thesis compared to the baseline, which is shorter but more complex?

In terms of the Muon1 code and its associated optimisation work, it would be interesting to put the cooling components into the optimiser, or allow the solenoids near the target to overlap, to see if this produces a muon yield as high as the current IDS baseline.

Bibliography

[1] *GROUP REPORT: Physics at a Neutrino Factory*, C. Albright *et al.* (Eds. S. Geer and H. Schellman), Report to the Fermilab Directorate, FERMILAB-FN-692, `hep-ex/0008064` (2000).

[2] *The MARS Code System* version 15.07, by N.V. Mokhov, available from http://www-ap.fnal.gov/MARS/.

[3] *GEANT4—a simulation toolkit*, S. Agostinelli *et al.*, Nuclear Instruments and Methods A **506**, pp.250–303 (2003), available from http://geant4.web.cern.ch/geant4/.

[4] *Status and prospects of the HARP experiment*, M. Ellis, J. Phys. G **29**, pp.1613–1620 (2003).

[5] Muon1 Distributed Particle Accelerator Design project website, http://stephenbrooks.org/muon1.

[6] *Open Letter to Radioactive Persons*, W. Pauli (1930); reprinted in Physics Today **31**, p.27 (1978). Note that at the time he calls the new particle a 'neutron' rather than a neutrino.

[7] *Detection of the Free Neutrino: A Confirmation*, C. L. Cowan Jr., F. Reines, F. B. Harrison, H. W. Kruse and A. D. McGuire, Science **124**, p.103 (1956).

[8] *Measurement of the solar electron neutrino flux with the Homestake chlorine detector*, B.T. Cleveland *et al.*, Astrophys. J. **496**, pp.505–526 (1998).

[9] *Remarks on the Unified Model of Elementary Particles*, Z. Maki, M. Nakagawa and S. Sakata, Prog. Th. Phys. **28**, pp.870–880 (1962).

[10] *Evidence for Oscillation of Atmospheric Neutrinos*, Y. Fukuda *et al.* (Super-Kamiokande Collaboration), Phys. Rev. Lett. **81**, pp.1562–1567 (1998).

[11] *Review of Particle Physics*, Particle Data Group (2008), section 13: Neutrino Mass, Mixing and Flavour Change (pp.163–171).

[12] *Letter of Intent for Double-CHOOZ: a Search for the Mixing Angle* θ_{13} (2004).

[13] *A New Nuclear Reactor* ν *Experiment to Measure* θ_{13}, International Reactor θ_{13} Working Group (2004).

[14] *Neutrino oscillations in matter*, L. Wolfenstein, Phys. Rev. D **17**, p.2369 (1978).

[15] *Physics at a future Neutrino Factory and super-beam facility*, The ISS Physics Working Group, RAL-TR-2007-019, available from http://arxiv.org/abs/0710.4947v2.

[16] *Search for CP Violation with a Neutrino Factory*, Morimitsu Tanimoto, Phys. Lett. B **462**, pp.115–120 (1999).

[17] *Measurement of Neutrino Oscillations with the MINOS Detectors in the NuMI Beam*, P. Adamson *et al.*, Phys. Rev. Lett. **101**, p.131802 (2008).

[18] *Measurement of Neutrino Oscillation by the K2K Experiment*, The K2K collaboration, M. H. Ahn *et al.*, Phys. Rev. D **74**, 072003, `hep-ex/0606032` (2006).

[19] *First Operational Experience with the CNGS Facility*, E. Gschwendtner, A. Pardons *et al.*, contribution to NuFact'07, 6–11 August 2007, Okayama, Japan.

[20] *Topical review on 'beta-beams'*, Cristina Volpe, J. Phys. G **34**, R1–R44, `hep-ph/0605033` (2007).

[21] *Neutrino beams from muon storage rings: Characteristics and physics potential*, S. Geer, Phys. Rev. D **57**, pp.6989–6997 (1998).

[22] *A low energy neutrino factory for large θ_{13}*, Steve Geer, Olga Mena and Silvia Pascoli, Phys. Rev. D **75**, 093001 (2007).

[23] *EURONU WP6 2009 yearly report: Update of the physics potential of Nufact, superbeams and betabeams*, J. Bernabeu *et al.*, `arXiv:1005.3146` (2010).

[24] *The International Scoping Study of a Neutrino Factory and Superbeam Facility*, The ISS Collaboration, RAL-TR-2007-023.

[25] *Feasibility Study-II of a Muon-Based Neutrino Source*, ed., S. Ozaki, R. Palmer, M. Zisman, and J. Gallardo, BNL-52623 (2001).

[26] *Proton Drivers for a Neutrino Factory*, C.R. Prior, talk at the 1st NuFact Summer School (2002), available from http://hepunx.rl.ac.uk/uknf/wp1/prior-pdriver2002nfschool.doc.

[27] *Review of Particle Physics*, W.-M. Yao *et al.* (Particle Data Group), J. Phys. G **33**, p.1 (2006), available from http://pdg.lbl.gov.

[28] *Solid Target for a Neutrino Factory*, G.P. Škoro *et al.*, Proc. PAC 2009.

[29] *Solid target studies in the UK*, J.R.J. Bennett *et al.*, Proc. NuFact'05.

[30] *ICOOL: A Simulation Code for Ionization Cooling of Muon Beams*, R.C. Fernow, Proc. PAC 1999.

[31] *Simulation of MICE Using G4MICE*, C. Rogers and R. Sandstrom, Proc. EPAC 2006.

[32] *Transport, a Computer Program for Designing Charged Particle Beam Transport Systems*, K.L. Brown, D.C. Carey, Ch. Iselin and F. Rothacker. See yellow reports CERN 73-16 (1973) & CERN 80-04 (1980).

[33] Runge, C., *Math. Ann.* **46**, p.167 (1895); and Kutta, M.W.Z., *für Math. u. Phys.* **46**, p.435 (1901).

[34] *Numerical Recipes in C*, W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vettering, chapter 15.1: Integration of Ordinary Differential Equations - Runge–Kutta Method.

[35] *Magnet Design Using OPERA*, Vector Fields Inc., http://www.vectorfields.com/content/view/105/0/.

[36] *Features of a Muon Cooling Ring for a Neutrino Factory*, G.H. Rees *et al.*, Proc. NuFact'03.

[37] *Muon front-end chicane and acceleration*, G.H. Rees *et al.*, Proc. NuFact'02.

[38] This technique known as 'inverse transform sampling' is described in section II.2 of *Non-Uniform Random Variate Generation* by Luc Devroye (1986), available from http://cg.scs.carleton.ca/~luc/rnbookindex.html.

[39] *Studies of a Target System for a 4-MW, 24-GeV Proton Beam*, H.G. Kirk *et al.*, CERN-INTC-2004-016.

[40] *Conceptual design of the SPL, a high-power superconducting H^- linac at CERN*, B. Autin *et al.*, CERN-2000-012.

[41] *ISIS Megawatt Upgrade Plans—Neutrons and Neutrinos for Europe*, C.R. Prior *et al.*, Proc. PAC 2003.

[42] *Conceptual design of the SPL II, a high-power superconducting H^- linac at CERN*, F. Gerigk *et al.*, CERN-2006-006.

[43] *The Synchrotron Option for a Multi-Megawatt Proton Driver*, C.R. Prior, Nuclear Physics B (Proc. Suppl.) **155**, p.312 (2006).

[44] *Report on Accelerator Physics and Technology Workshop for Project X*, D. McGinniss *et al.*, FNAL Beams-doc-2970-v1 (2007).

[45] *The Proton Driver Design Study*, W. Chou *et al.*, FERMILAB-TM-2136 (2000).

[46] *A Slow-Cycling Proton Driver for a Neutrino Factory*, H. Schönauer *et al.*, Proc. EPAC 2000.

[47] *Present Status of J-PARC*, F. Naito, Proc. APAC 2007.

[48] *Beam Intensity Upgrade at Fermilab*, A. Marchionni, FERMILAB-CONF-06-225-AD (2006).

[49] https://www.ids-nf.org/wiki/FrontPage/Accelerator/Baselines/1 (2008).

[50] *Spallation Neutron Source: Description of Accelerator and Target*, ed. B. Boardman, Rutherford Appleton Laboratory technical report RL-82-006 (1982).

[51] *LS-DYNA calculations of shocks in solids*, G.P. Škoro *et al.*, Proc. NuFact'05.

[52] *Thermophysical Properties of Matter*, Volume 7: Thermal Radiative Properties—Metallic Elements and Alloys, Y.S.Touloukian and D.P. DeWitt (1970).

[53] *A Neutrino Factory Target Station Design based on Solid Targets*, J.R.J. Bennett, slides presented at the Third High-Power Target Workshop, Bad Zurzach, Switzerland (2007).

[54] *The Potential of Fluidised Powder Target Technology in High Power Accelerator Facilities*, C.J. Densham *et al.*, Proc. PAC 2009.

[55] *The production and anatomy of a tungsten powder jet*, T.W. Davies, O. Caretta, C.J. Densham and R. Woods, Powder Technology **201**, Issue 3, pp. 296–300 (2010).

[56] http://hepunx.rl.ac.uk/uknf/wp3/target_eng/-SpokelessWheelDiscussion.ppt.

[57] *MARS15 Overview*, N.V. Mokhov and S.I. Striganov, Hadronic Shower Simulation Workshop AIP Proceedings 896, Fermilab-Conf-07-008-AD (2007).

[58] *Matprop* web-based tool at http://www.slac.stanford.edu/comp/physics/matprop.html.

[59] *Computed Pion Yields from a Tantalum Rod Target: Comparing MARS15 and GEANT4 across Proton Energies*, S.J. Brooks and K.A. Walaron, Proc. NuFact'05.

[60] *Physics at a future Neutrino Factory and super-beam facility*, The ISS Physics Working Group (2007), section 4.4: Leptonic unitarity triangle and CP-violation, equations (205) and (206) involving J_{CP}.

[61] http://www.hep.ph.ic.ac.uk/uknfic/iss0406/talks/plenary/-2006-04-26-PlenaryParallel-5-Kirk.ppt, slide 13, optimal length of 60 cm.

[62] *Pion and muon yields from a 10GeV proton beam on various targets*, John Back, Proc. NuFact'08.

[63] *Secondary Particle Production and Capture for Muon Accelerator Applications*, S.J. Brooks, Proc. EPAC 2006.

[64] *The MARS Code System User's Guide Version 15(2006)*, Nikolai V. Mokhov and Catherine C. James, available from http://www-ap.fnal.gov/MARS/intro_manual.htm.

[65] UKNF Note 30: *Simulations of Pion Production in a Tantalum Rod Target using GEANT4 with comparison to MARS*, K.A. Walaron; http://hepunx.rl.ac.uk/uknf/wp3/uknfnote_30.pdf (2005).

[66] *Measurement of the production cross section of positive pions in p–Al collisions at 12.9 GeV/c*, The HARP Collaboration, Nucl.Phys. B **732**, pp.1–45 (2006).

[67] *Measurement of the production of charged pions by protons on a tantalum target*, The HARP Collaboration (M.G. Catanesi, *et al.*), Eur. Phys. J. C **51** no.4, pp.787–824 (2007).

[68] *Principles and Applications of Muon Cooling*, D. Neuffer, Proc. 12th International Conf. on High-Energy Accelerators, p.481 (1983).

[69] D. Neuffer, Part. Accel. **14**, p.75 (1983).

[70] *Classical Mechanics*, H. Goldstein, publ. Addison-Wesley (1950).

[71] *Stochastic Cooling*, D. Möhl, report CERN/PS/DL 78-25, online at http://www.slac.stanford.edu/econf/C781015/pdf/013.pdf (1978).

[72] *MICE Technical Reference Document*, The MICE Collaboration, available from http://www.mice.iit.edu/.

[73] IDS-NF Note 002: *Neutrino Factory: specification of baseline*, IDS-NF Steering Group, available from https://www.ids-nf.org/wiki/FrontPage/Documentation (2008).

[74] *The Physics of Muon Cooling for a Neutrino Factory*, S. Holmes - Ph.D. Thesis (Oxford), chapter 2: Muon ionisation cooling, figure 2.1 (2006).

[75] *Review of Particle Physics*, Particle Data Group (2008), equation (27.1) on p.268, available from http://pdg.lbl.gov.

[76] Value from http://durpdg.dur.ac.uk/HEPDATA/PART/-AtomicNuclearProperties/HTML_PAGES/087.html.

[77] *Review of Particle Physics*, Particle Data Group (2008), equation (27.12) on p.271, available from http://pdg.lbl.gov.

[78] *Multi-TeV muon colliders*, D. Neuffer, Advanced Accelerator Concepts '87, AIP Conf. Proc. **156**, pp.201–208 (1987).

[79] *Beam Envelope Equations in a Solenoidal Field*, G. Penn, NFMCC-doc-71-v1 (2000), available from http://nfmcc-docdb.fnal.gov/.

[80] *Review of Particle Physics*, Particle Data Group (2008), section 27.4.1: Radiation Length, equations (27.20,21) and table 27.2 on p.272, available from http://pdg.lbl.gov.

[81] *Ab initio liquid hydrogen muon cooling simulations with ELMS*, W.W.M. Allison *et al.*, J. Phys. G **34**, pp.679–685 (2007).

[82] Available from http://www-pnp.physics.ox.ac.uk/~holmess/ELMS/index.htm.

[83] *Cavity Design for the CERN Muon Cooling Channel*, Roland Garoby and Frank Gerigk, PS/RF Note 2001-014, CERN-NUFACT-NOTE 2001-087.

[84] *Simulation of genetic systems by automatic digital computers*, A.S. Fraser, Aust. J. Biol. Sci. **10**, pp.484–491 (1957).

[85] *Handbook of Genetic Algorithms*, L.D. Davis, publ. Van Nostrand Reinhold (1991).

[86] Active Distributed Computing Projects, http://distributedcomputing.info/projects.html

[87] 1st–3rd Oxford–Princeton High-Power Target Workshops. 2nd workshop website: http://www.physics.princeton.edu/~bridges/targetworkshop2/.

[88] *EMMA—the World's First Non-Scaling FFAG*, R. Edgecock *et al.*, Proc. EPAC 2008.

[89] *Status of the RAL Front End Test Stand*, A. Letchford *et al.*, Proc. IPAC 2010.

[90] *Effect of High Solenoidal Magnetic Fields in Breakdown Voltages of High Vacuum 805 MHz Cavities*, A. Moretti *et al.*, Proc. LINAC 2004.

[91] *Theory of the Alternating-Gradient Synchrotron*, E.D. Courant and H.S. Snyder, Ann. Phys. **3**, p.1 (1958).

[92] *A Selection of Formulae and Data Useful for the Design of A.G. Synchrotrons*, C. Bovet, R. Gourian, I. Gumowski and K.H. Reich (1970) CERN/MPS-SI/Int. DL/70/4.

[93] *Accelerator Physics*, S.Y. Lee (2004), chapter 2, section II.1: Transfer Matrix and Stability of Betatron Motion.

[94] *Particle Accelerator Physics I, 2nd Edition*, H. Wiedemann (1998), chapter 5.2: Matrix Formalism in Linear Beam Dynamics.

[95] C.T. Rogers - Ph.D. Thesis (Imperial College, London), chapter 3 (2007).

Appendix A

MARS.INP Template File

The token `$energy` in the file below is replaced by a number in GeV for each energy required. MARS uses kinetic energy and requires that this ends in a decimal point if it is a whole number (for instance `6.` for 6 GeV but `2.2` for 2.2 GeV). The token `$hole` is always replaced by `0.` in simulations where the rod is solid.

```
Tantalum Rod Pion Production [TEMPLATE]
/home/csf/brooks/mars15/dat
C GUI switch
CTRL 0

C Generate 100000 events
NEVT 100000

C 0.01mm boundary precision, 2cm initial trial step
SMIN 0.001 2.

C Beam is protons(1), parallel gaussian(2) [now modified in BEG1 subroutine]
IPIB 1 2
C 0.5cm RMS transversely (so rod is 2sigma), 1ns RMS longitudinally
BEAM 0.5 0.5 5=29.9792458
C Energy in GeV
ENRG $energy

C Two materials used: tantalum and vacuum
NMAT 2
MATR 'TA' 'VAC'

C Use 'R-sandwich' type geometry (default)
INDX 2=F
C Two R-sections: vacuum(#2) to $hole cm radius, tantalum(#1) to 1cm radius,
C ending at z=20cm
NLTR 2
RSEC $hole 1. 51=10 10 101=2 1
```

```
ZSEC 20. 1251=200

C We record particles at three surfaces, 1cm out, z=0-20cm;
C and r=0-1cm, z=0cm and z=20cm to fort.83 (81..90)
NSUR 3
RZTS 1. 1.  0. 20.  83.
     0. 1.  0.  0.  83.
     0. 1. 20. 20.  83.

STOP
```

Appendix B

Modification to User Subroutine BEG1 in m1504.f

This is the section of the long Fortran file 'm1504.f' that has been modified to generate the correct input proton distribution.

```
*      PARAMETER (CLIGHT=29979245800.D0)
      PARAMETER (PI=3.14159265358979238D+00)
C- - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
C+++ INSERT YOUR SOURCE TERM HERE +++
C      READ()...,W1,...
C      W=W*W1

      R=SQRT(1.D0-SQRT(RAND())) ! SJB: 1cm radius circular parabolic beam
      TH=RAND()*PI*2.D0
      X=R*COS(TH)
      Y=R*SIN(TH)

C++++++++++++++++++++++++++++++++++
      RETURN
      END
```

Appendix C

The Linearised Beam Optics Model

To understand and quantify the performance of cooling channels, it is informative to make the small-angle, or *paraxial* approximation to the transport of the muon beam, in which the optical elements have linear (and thus easy to analyse) effects. This approach also reveals the *Twiss parameters*[†] [91, 92] widely used in accelerator lattice design.

C.1 Equations of Motion in z

Section 2.2 shows how the transport of particles in the most general case can be described in terms of an augmented phase space vector $\mathbf{s}(t) = (x, y, z, \gamma v_x, \gamma v_y, \gamma v_z, t)$ for each particle and how it changes over time as given by a differential equation of the form $\dot{\mathbf{s}} = \mathbf{f}(\mathbf{s})$. If the particles are assumed to be travelling forwards so that z increases monotonically with t, it is possible to restate the problem with z as the independent variable instead of t, using a different set of phase-space variables $\mathbf{u}(z)$ including the transverse positions and velocities, forward velocity (or energy) and the time of arrival t as a new dynamical variable that depends on z. For a non-straight accelerator, this treatment uses curvilinear coordinates such that $x = y = 0$ represents a specified reference path through the machine, on which z is equal to the path length and the three axes remain orthogonal. Non-zero x and y are measured on planes of constant z perpendicular to the reference path. These conventions mean that the mapping $\mathbf{u}(z_0) \to \mathbf{u}(z_1)$ fully describes the behaviour of the physical piece of accelerator for which $z_0 \leq z \leq z_1$ (rather than the behaviour through a specific time interval). This mapping can be obtained from solving the new equation of motion in the form $\mathbf{u}' = \mathbf{g}(\mathbf{u})$, prime denoting differentiation with respect to z, noting that cases where $\dot{z} \leq 0$ are outside the scope of this method.

Just as in section 2.1, there is some choice of how to parameterise the phase space, a sensible option being $\mathbf{u} = (x, y, t, x', y', E_k, z)$ where E_k is the kinetic energy of the particle, in place of the peculiar $t' = \frac{\mathrm{d}t}{\mathrm{d}z} = \frac{1}{v_z}$ coordinate that would otherwise have been used. Here, the form of $\mathbf{g}$ will be derived in cartesian coordinates only, because the transformation to the curvilinear coordinates of choice is standard mathematics and would

[†]This name is historically inaccurate and they should be called Courant–Snyder parameters. However, Twiss is the more common usage and avoids confusion with the 'Courant–Snyder invariant'.

yield very long equations in the general case. This requires expressing the z-derivative of each component of $\mathbf{u}$ in terms of known constants and other components of $\mathbf{u}$. The constants in this case are the speed of light c and the mass m and charge q of the particle.

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}z} &= x' \\
\frac{\mathrm{d}y}{\mathrm{d}z} &= y' \\
\frac{\mathrm{d}t}{\mathrm{d}z} &= \frac{1}{v_z} \\
&= \frac{\sqrt{1+x'^2+y'^2}}{v} \\
&= \frac{(mc^2+E_k)\sqrt{1+x'^2+y'^2}}{c\sqrt{2E_kmc^2+E_k^2}}
\end{aligned}$$

This final expression will be written as $\frac{1}{v_z}$ as it appears frequently in the other terms. The derivatives of the primed variables need some manipulation to relate them to the time derivatives of $\gamma\mathbf{v}$ given in section 2.1.

$$\begin{aligned}
\frac{\mathrm{d}x'}{\mathrm{d}z} &= \frac{1}{v_z}\frac{\mathrm{d}x'}{\mathrm{d}t} \\
&= \frac{1}{v_z}\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\gamma v_x}{\gamma v_z}\right] \\
&= \frac{1}{\gamma v_z^2}\left(\frac{\mathrm{d}(\gamma v_x)}{\mathrm{d}t} - x'\frac{\mathrm{d}(\gamma v_z)}{\mathrm{d}t}\right)
\end{aligned}$$

It is now possible to substitute the Lorentz force law $\frac{\mathrm{d}}{\mathrm{d}t}(\gamma v_i) = \frac{q}{m}(E_i + (\mathbf{v}\times\mathbf{B})_i)$.

$$\begin{aligned}
\frac{\mathrm{d}x'}{\mathrm{d}z} &= \frac{q}{m\gamma v_z^2}(E_x + v_yB_z - v_zB_y - x'(E_z + v_xB_y - v_yB_x)) \\
&= \frac{q}{m\gamma v_z^2}(E_x - x'E_z) + \frac{q}{m\gamma v_z}(x'y'B_x - (1+x'^2)B_y + y'B_z) \\
&= \frac{qc^2}{mc^2+E_k}\left(\frac{1}{v_z^2}(E_x - x'E_z) + \frac{1}{v_z}(x'y'B_x - (1+x'^2)B_y + y'B_z)\right)
\end{aligned}$$

The derivative of y' can be calculated in a similar way, provided care is taken with the indices.

$$\frac{\mathrm{d}y'}{\mathrm{d}z} = \frac{qc^2}{mc^2+E_k}\left(\frac{1}{v_z^2}(E_y - y'E_z) + \frac{1}{v_z}((1+y'^2)B_x - x'y'B_y - x'B_z)\right)$$

Finally, the rate of change of energy comes only from the electric field.

$$\begin{aligned}
\frac{\mathrm{d}E_k}{\mathrm{d}z} &= \frac{1}{v_z}\frac{\mathrm{d}E_k}{\mathrm{d}t} \\
&= \frac{1}{v_z}q(\mathbf{v}\cdot\mathbf{E}) \\
&= q(x'E_x + y'E_y + E_z)
\end{aligned}$$

Taken together with $\mathrm{d}z/\mathrm{d}z = 1$, these expressions define $\mathbf{u}' = \mathbf{g}(\mathbf{u})$. The $\mathbf{E}$ and $\mathbf{B}$ fields also depend on z, which is why the independent variable z was appended to the definition of the vector $\mathbf{u}(z)$ so as to close the system, analogously to the 7-element $\mathbf{s}(t)$.

C.2 Transfer Matrices

Given an initial condition $\mathbf{u}(z_0)$ and a final position in the beamline $z_1 \geq z_0$, the first order system of differential equations $\mathbf{u}' = \mathbf{g}(\mathbf{u})$ derived in the last section may be solved to give $\mathbf{u}(z_1) = \mathbf{m}(\mathbf{u}(z_0))$ for some new function $\mathbf{m}$. This sort of function is known as a *mapping function* because it can be applied to the initial phase space location to map directly from one point in the accelerator to another. As t was included in the definition of the phase space vector $\mathbf{u}$, this even works for time-varying electromagnetic fields since $\mathbf{E}$ and $\mathbf{B}$ are still uniquely specified by the phase space point. However, mappings may generally not be used when there are interactions of the particles with each other, such as 'space charge' when the Coulomb interaction from other particles in the beam enters the definition of $\mathbf{E}$ to a significant extent. This is because $\mathbf{u}$ no longer contains all the relevant information about the state of the system.

Differential equations defined by a nonlinear function such as $\mathbf{g}$ will usually produce a nonlinear mapping function $\mathbf{m}$ that is even more complicated. A key technique for dealing with this complexity is to linearly approximate $\mathbf{m}$ by considering small perturbations from a 'central' point $\mathbf{u}_0$ at the beginning of the accelerator section: $\mathbf{u}(z_0) = \mathbf{u}_0 + \delta\mathbf{u}_0$, where terms in $(\delta\mathbf{u}_0)^2$ and above may be neglected. The result of the mapping is now

$$\begin{aligned}
\mathbf{u}(z_1) &= \mathbf{m}(\mathbf{u}(z_0)) = \mathbf{m}(\mathbf{u}_0 + \delta\mathbf{u}_0) \\
&= \mathbf{m}(\mathbf{u}_0) + (\delta\mathbf{u}_0 \cdot \nabla)\mathbf{m}(\mathbf{u}_0) + O((\delta\mathbf{u}_0)^2) \\
&\simeq \mathbf{u}_1 + \delta\mathbf{u}_1, \qquad \mathbf{u}_1 = \mathbf{m}(\mathbf{u}_0) \qquad \delta\mathbf{u}_1 = (\delta\mathbf{u}_0 \cdot \nabla)\mathbf{m}(\mathbf{u}_0)
\end{aligned}$$

assuming $\mathbf{m}$ is smooth enough to be Taylor expanded and defining the output central point and its perturbation term as shown. The components of this output perturbation are given by

$$\begin{aligned}
(\delta\mathbf{u}_1)_i &= \sum_j (\delta\mathbf{u}_0)_j \partial_j m_i(\mathbf{u}_0) \\
&= \sum_j A_{ij}(\delta\mathbf{u}_0)_j, \qquad A_{ij} = \partial_j m_i(\mathbf{u}_0)
\end{aligned}$$

where the matrix of constants $\mathbf{A}$ is defined so that in matrix product notation,

$$\delta\mathbf{u}_1 = \mathbf{A}\delta\mathbf{u}_0.$$

This matrix $\mathbf{A}$ is formally known as the Jacobian of the function $\mathbf{m}$ at the point $\mathbf{u}_0$. Since perturbations in the input or output z are zero by the definition of the problem, the 7^{th} row and column can be removed and the remaining 6×6 matrix is called a *(linear) transfer matrix* in accelerator terminology [93, 94]. It completely codifies the first order perturbative behaviour of particles near the trajectory that starts at $\mathbf{u}_0$ and ends at $\mathbf{u}_1$.

For beams, the central point is usually chosen to be the barycentre of the beam distribution, for example $\mathbf{u}_0 = \langle \mathbf{u}(z_0) \rangle$, which ensures that the $\delta\mathbf{u}_0$ values are 'small' for particles near the centre of the beam, or for the whole beam if it and the nonlinear terms in $\mathbf{m}$ are small enough.

C.3 Beam Shape and Covariance Matrices

Now that the actions of accelerator sections on particles close to the central point have been represented to first order as matrices, it makes sense to find a compatible model of the (sufficiently small) beam itself that these can act on. Consider first a vector of random variables $\boldsymbol{\zeta}$ such that

$$\left\langle \zeta_i^2 \right\rangle = 1 \qquad \mathbf{R}\boldsymbol{\zeta} \sim \boldsymbol{\zeta}$$

for all $1 \leq i \leq 6$, where angle brackets take the expected value of a random variable, $\sim$ means 'identical in distribution' and $\mathbf{R}$ is any 6×6 rotation matrix. These formulae imply that the ζ_i all have unit variance and the distribution of $\boldsymbol{\zeta}$ is invariant under rotations of the 6-dimensional space, thus the probability density function has the form $f(|\boldsymbol{\zeta}|)$. By symmetry arguments, these also imply that

$$\langle \boldsymbol{\zeta} \rangle = \mathbf{0} \qquad \langle \zeta_i \zeta_j \rangle = 0$$

for all $i \neq j$, with the first identity coming from considering $\mathbf{R}$ to be various 180° rotations that change the sign of some ζ_i and the second from considering a 90° rotation in the i–j plane, which changes $\zeta_i \zeta_j$ to $-\zeta_j \zeta_i$. This means that the ζ_i are not linearly correlated. Note that there are many possible choices for the distribution $\boldsymbol{\zeta}$ that satisfy the conditions: one is a unit Gaussian distribution, in which case all the ζ_i are independent and identically distributed; another is a uniform distribution on the interior of a 6-dimensional sphere, in which case they are still identically distributed but no longer independent.

The model of the beam shall be that

$$\delta\mathbf{u} \sim \mathbf{S}\boldsymbol{\zeta}$$

for some 6×6 matrix $\mathbf{S}$ that defines the beam shape, where $\delta\mathbf{u}$ is still the phase space offset from the central point (its 7^{th} entry is set to zero and removed) but interpreted now as a random variable statistically representing all particles in the beam. This distribution transforms through the section of the accelerator with transfer matrix $\mathbf{A}$ follows:

$$\delta\mathbf{u}_0 \sim \mathbf{S}_0\boldsymbol{\zeta} \qquad \Rightarrow \qquad \delta\mathbf{u}_1 \sim \mathbf{A}\mathbf{S}_0\boldsymbol{\zeta},$$

so $\mathbf{S}_1 = \mathbf{A}\mathbf{S}_0$ is the transformation rule for shape matrices $\mathbf{S}$.

Although shape matrices are a natural way of defining beam shape (and useful for generating specific distributions computationally starting from a normalised generator for $\boldsymbol{\zeta}$), they are not uniquely determined by the beam itself since

$$\mathbf{R}\boldsymbol{\zeta} \sim \boldsymbol{\zeta} \qquad \Rightarrow \qquad \mathbf{S}\mathbf{R}\boldsymbol{\zeta} \sim \mathbf{S}\boldsymbol{\zeta},$$

meaning that the shape matrices $\mathbf{SR}$ and $\mathbf{S}$ generate the same beam for any rotation matrix $\mathbf{R}$. This situation appears elsewhere in physics where there is a gauge group, for example the electromagnetic vector potential A^μ can be chosen in multiple ways to describe the same physical situation. It would be useful at this point to find the 'beam equivalent' of the physical field $F^{\mu\nu}$ that is uniquely determined by measurable parameters. The measurable parameters in this case are expectations of functions of the beam $\langle f(\delta\mathbf{u}) \rangle$.

If the central point is the beam barycentre then $\langle \delta\mathbf{u} \rangle = \langle \mathbf{S}\boldsymbol{\zeta} \rangle = \mathbf{S} \langle \boldsymbol{\zeta} \rangle = \mathbf{0}$, so the next simplest measurables are the quadratic functions, or covariances, equal to

$$C_{ij} = \langle \delta u_i \delta u_j \rangle .$$

These may be aggregated into a matrix and as the expectation operator treats matrix elements separately, the right-hand side can be expressed as a vector outer product, allowing some useful manipulation:

$$\begin{aligned} \mathbf{C} &= \left\langle \delta\mathbf{u} \delta\mathbf{u}^T \right\rangle \\ &= \left\langle \mathbf{S}\boldsymbol{\zeta} (\mathbf{S}\boldsymbol{\zeta})^T \right\rangle = \left\langle \mathbf{S}\boldsymbol{\zeta}\boldsymbol{\zeta}^T \mathbf{S}^T \right\rangle \\ &= \mathbf{S} \left\langle \boldsymbol{\zeta}\boldsymbol{\zeta}^T \right\rangle \mathbf{S}^T. \end{aligned}$$

The elements of $\boldsymbol{\zeta}\boldsymbol{\zeta}^T$ are just $\zeta_i \zeta_j$, so by the unit variance and zero linear correlation properties of $\boldsymbol{\zeta}$, $\left\langle \boldsymbol{\zeta}\boldsymbol{\zeta}^T \right\rangle = \mathbf{I}$, the 6×6 identity matrix. This gives the primary relation between $\mathbf{C}$ and $\mathbf{S}$:

$$\mathbf{C} = \mathbf{S}\mathbf{S}^T.$$

That $\mathbf{C}$ is gauge-invariant can be checked by replacing $\mathbf{S}$ by $\mathbf{SR}$ and noting that $\mathbf{R}\mathbf{R}^T = \mathbf{I}$ for rotation matrices. Using the transformation law $\mathbf{S}_1 = \mathbf{A}\mathbf{S}_0$ for shape matrices, the action of a section of accelerator on the covariance matrix can be derived.

$$\mathbf{C}_1 = \mathbf{S}_1 \mathbf{S}_1^T = \mathbf{A}\mathbf{S}_0 \mathbf{S}_0^T \mathbf{A}^T = \mathbf{A}\mathbf{C}_0 \mathbf{A}^T$$

C.4 Un-normalised and Normalised Emittances

Section 6.1 discussed Liouville's theorem: the fact that Hamiltonian systems conserve phase space volume. It will be useful to find a measure of phase space volume for the $\delta\mathbf{u} \sim \mathbf{S}\boldsymbol{\zeta}$ model of the (sufficiently small) beam developed so far. The distribution may not have a sharp edge in phase space because the random vector $\boldsymbol{\zeta}$ could be something such as a Gaussian that never exactly goes to zero. However, the variances $\langle \zeta_i^2 \rangle$ are defined to be unity, so in this case it is possible to take the distribution of $\boldsymbol{\zeta}$ as the definition of 'unit phase space volume'. Thus the distribution of particles $\delta\mathbf{u}$ will have volume $\det \mathbf{S}$ in these units, as the determinant of a matrix is the factor by which it scales volumes. This concept of phase space volume is known as *un-normalised RMS emittance*

$$\varepsilon_{\text{rms}}^{\text{un}} = \det \mathbf{S};$$

it can also be defined relative to the measurable covariance matrix using properties of the determinant:

$$\det \mathbf{C} = \det \mathbf{S}\mathbf{S}^T = \det \mathbf{S} \det \mathbf{S}^T = (\det \mathbf{S})^2 = \varepsilon_{\text{rms}}^{\text{un}\ 2} \qquad \Rightarrow \qquad \varepsilon_{\text{rms}}^{\text{un}} = \sqrt{\det \mathbf{C}}.$$

Note that the number of dimensions of $\delta\mathbf{u}$ in the phase space considered is often reduced to two or four, rather than the full six, in order to see the emittance effects in a subset of the variables of the system. This works best when coupling between pairs of phase space variables is weak: the longitudinal pair (t, E_k) are often only weakly coupled

to the transverse pairs (x, x') and (y, y'), so there exist varieties of (2D) *longitudinal emittance*. In a channel consisting only of quadrupole and dipole magnets, the transverse pairs are also decoupled from each other and 2D *transverse emittances* can be used for each pair of variables (or *phase space plane*). Solenoids, however, mix the x and y planes, so *4D transverse emittance* better represents the beam.

In cases where a beam without holes has a defined edge in phase space, that is the probability density function $f(\boldsymbol{\zeta}) = 0$ for $|\boldsymbol{\zeta}| > R$ but $f(\boldsymbol{\zeta}) > 0$ for $|\boldsymbol{\zeta}| < R$, the actual volume of phase space occupied by the beam (however diffuse it may be in some regions) can be calculated. In a D-dimensional phase space this is given by

$$V = S_D R^D \det \mathbf{S},$$

where S_D is the volume of a D-dimensional unit sphere, useful values being $S_2 = \pi$, $S_4 = \pi^2/2$ and $S_6 = \pi^3/6$. By convention, the factor of S_D is usually dropped, leading to the definition of *un-normalised 100% emittance* (also called *geometric emittance*)

$$\varepsilon_{100}^{\mathrm{un}} = R^D \det \mathbf{S} = R^D \sqrt{\det \mathbf{C}}.$$

For a known model distribution such as a uniform or parabolic beam, the value of R may be calculated from the requirement that the ζ_i have unit variance; note that in these cases the ζ_i are not independent, so a D-dimensional integral is required. Beams with holes ($f(\boldsymbol{\zeta}) = 0$ for some $|\boldsymbol{\zeta}| < R$) are also possible but there is no convention on whether the holes are counted in the total emittance.

Liouville's theorem states that phase space volume is conserved provided the system is described by a Hamiltonian and the correct measure of phase space volume is used. The emittances discussed so far use the variables x' and y', which only conserve phase space volume if the beam is not accelerated or decelerated. To make the theorem valid in general, x' and y' must be replaced by the 'canonical conjugate momenta' in the x and y directions. The end result of this calculation (described in [95]) is a just rescaling of the un-normalised emittances by the relativistic factor $\beta\gamma$ for each transverse plane:

$$\varepsilon_{\mathrm{2D}}^{\mathrm{n}} = \beta\gamma \; \varepsilon_{\mathrm{2D}}^{\mathrm{un}} \qquad \varepsilon_{\mathrm{4D}}^{\mathrm{n}} = (\beta\gamma)^2 \; \varepsilon_{\mathrm{4D}}^{\mathrm{un}},$$

giving a quantity known as the *normalised emittance*. This scaling applies to the RMS, 100% and other percentile variants of the emittance similarly. Note that if Liouville's theorem is obeyed, the incoming and outgoing emittances of a section of accelerator are the same, so $\det \mathbf{S}_1 = \det \mathbf{S}_0$, which places a restriction on the transfer matrix that $\det \mathbf{A} = 1$.

The units of emittance depend on the number of phase space dimensions considered. For a single plane such as (x, x') (2D emittance), x can be measured in millimetres and x' in milliradians, so mm·mrad is a commonly used unit. The basic SI unit would be metre-radians, or even just metres since radians are dimensionless, but this is a very large unit; however, mm·mrad is sometimes written as μm. For the two transverse planes taken together (4D emittance) the unit becomes $(\text{mm·mrad})^2$. In the longitudinal plane (t, E_k), the two variables are already conjugates in the Hamiltonian sense, so longitudinal emittance is always normalised and is commonly stated in units of eV·s (electronvolt-seconds). This makes the full unit for *6D emittance* $(\text{mm·mrad})^2$·eV·s. Sometimes when particles are tracked in the time domain, 6D emittances in $(\text{mm·mrad})^3$ appear because now the z direction is treated exactly like the other two.

C.5 Twiss Parameters

The covariance matrix $\mathbf{C}$ consists of measurable quantities that determine the shape matrix of the beam up to internal rotations $\mathbf{S} \to \mathbf{SR}$. It is symmetric because $C_{ij} = C_{ji} = \langle \delta u_i \delta u_j \rangle$, so in the case of 2D phase space for a single axis ($\delta\mathbf{u} = (x, x')$), there are three, not four, measurables: $\langle x^2 \rangle$, $\langle xx' \rangle$ and $\langle x'^2 \rangle$. The overall size of the beam has already been quantified by RMS emittance satisfying

$$\varepsilon_{\text{rms}}^{\text{un}\ 2} = \det \mathbf{C} = \left\langle x^2 \right\rangle \left\langle x'^2 \right\rangle - \left\langle xx' \right\rangle^2 ,$$

so it is possible to produce a scale-invariant matrix of the remaining measurable quantities that only relate to the beam shape:

$$\begin{aligned} \det(\mathbf{C}/\varepsilon_{\text{rms}}^{\text{un}}) &= 1 \\ \mathbf{C}/\varepsilon_{\text{rms}}^{\text{un}} &= \begin{bmatrix} \langle x^2 \rangle /\varepsilon_{\text{rms}}^{\text{un}} & \langle xx' \rangle /\varepsilon_{\text{rms}}^{\text{un}} \\ \langle xx' \rangle /\varepsilon_{\text{rms}}^{\text{un}} & \langle x'^2 \rangle /\varepsilon_{\text{rms}}^{\text{un}} \end{bmatrix} = \begin{bmatrix} \beta & -\alpha \\ -\alpha & \gamma \end{bmatrix}, \end{aligned}$$

where the last line defines the *beam Twiss parameters* α, β and γ. Note that these definitions mean $\beta, \gamma \geq 0$. Because $\mathbf{C}$ only had three independent variables and now one has been removed, the Twiss parameters only have two degrees of freedom between them, the unit determinant condition producing the constraint $\beta\gamma - \alpha^2 = 1$.

An *upright beam* is defined as being one in which $\alpha = 0$, or equivalently $\langle xx' \rangle = 0$ and is informative for understanding what the Twiss parameters mean (see figure C.1). A beam will pass through this configuration, in which there is (instantaneously) no overall expansion or contraction because the outermost particles have $x' = 0$, multiple times in a focussing system. Upright beams typically occur at all the maxima and minima of the beam size $x_{\text{rms}} = \sqrt{\langle x^2 \rangle}$ within the focussing channel. In the upright case the Twiss parameters and emittance are

$$\begin{aligned} \alpha &= 0, && \text{(definition of upright beam)} \\ \varepsilon_{\text{rms}}^{\text{un}} &= \sqrt{\langle x^2 \rangle \langle x'^2 \rangle} = x_{\text{rms}} x'_{\text{rms}}, \\ \beta &= \frac{\langle x^2 \rangle}{\varepsilon_{\text{rms}}^{\text{un}}} = \sqrt{\frac{\langle x^2 \rangle}{\langle x'^2 \rangle}} = \frac{x_{\text{rms}}}{x'_{\text{rms}}}, \\ \gamma &= 1/\beta. && \text{(by } \beta\gamma - \alpha^2 = 1\text{)} \end{aligned}$$

The RMS emittance emerges as the product of the RMS semi-axes of the phase space ellipse, which is consistent with being proportional to area. The β function, perhaps the most commonly used Twiss parameter, is the ratio by which the beam RMS size exceeds the RMS angle of a particle x'_{rms}. It has dimensions of length and is equal to the distance the beam would have to drift for the RMS change in particle x during the drift to be equal to the original x_{rms}, after which an elliptical phase space beam would be $\sqrt{2}$ times the original total size. This is the characteristic length of a focus from optics.

The definition of $\beta = \langle x^2 \rangle /\varepsilon_{\text{rms}}^{\text{un}}$ (regardless of α) can be rearranged as $x_{\text{rms}} = \sqrt{\beta \varepsilon_{\text{rms}}^{\text{un}}}$, meaning that for a constant $\varepsilon_{\text{rms}}^{\text{un}}$, β determines spatial beam size. Thus if a beam is required to be small in a certain region, the only pure optical technique available is to reduce β, as the emittance can only be changed by non-Liouvillian processes.

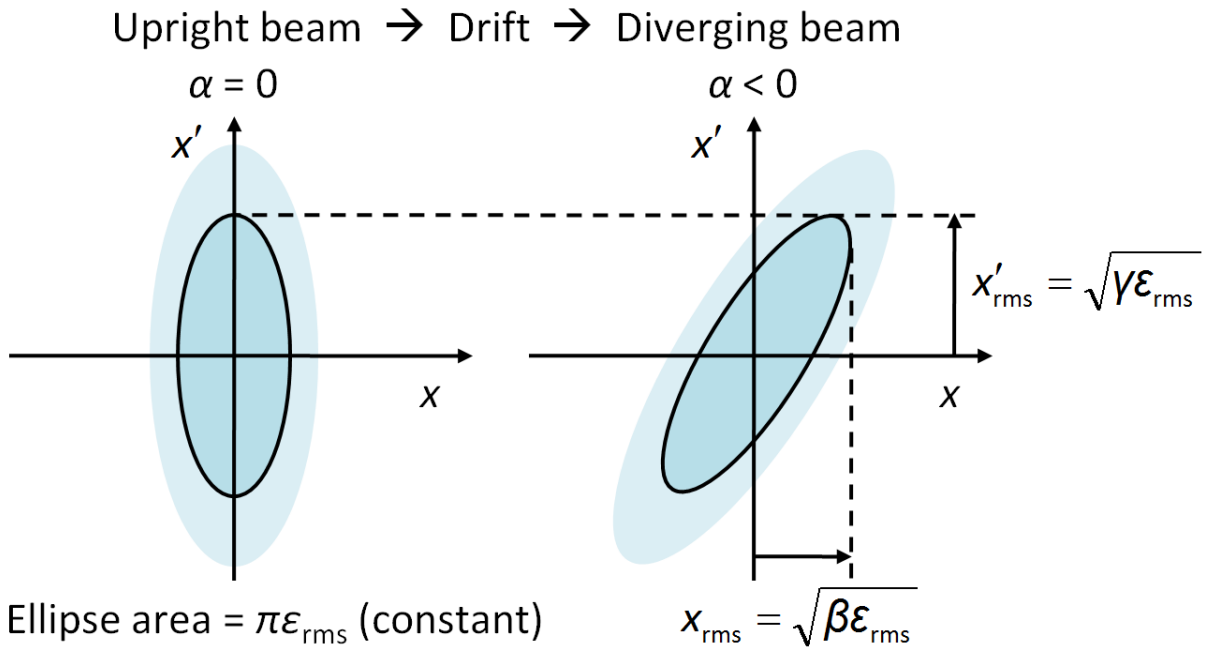


Figure C.1: RMS ellipse shapes in phase space, showing Twiss parameters.

When the beam is not upright, $\alpha \neq 0$, with negative α implying an expanding beam and positive α a contracting one. If a linear magnetic field is applied such that it reduces x' for positive x, this leads to an increase in α and the beam is focussed in that plane; a magnetic gradient of the opposite polarity would reduce α and defocus the beam. The Twiss parameters can also be used in multidimensional systems, typically by projecting the phase space into a single 2D plane and examining the RMS ellipse shape, though if there is correlation of the distribution between the different axes this may make the beam appear larger than it is.

Note that similar constructions apply to the 100% sizes of beams that are elliptical in phase space, with the Twiss parameters being the same but $x_{\max}$, $x'_{\max}$ being some multiple of the RMS values. Twiss parameters can also be found as a function of length in a periodic section of accelerator, from the condition that the input and output beam ellipse shapes in phase space must match. These are known as *matched Twiss parameters* and have the property that a beam with those values of α, β and γ should be *matched* (finish with the same distribution it started with) in that accelerator structure provided its emittance is small enough that the linear paraxial model holds.

Appendix D

ICOOL Input File

Below is the `for001.dat` input file for simulating a pencil beam of 200 MeV/c muons through 30 cm of LH_2. It uses the default ICOOL scattering model settings and was adapted from the `drif.f01` example included in the package. Particle output is written to `for009.dat`. The `npart` parameter may be reduced from `1000000` for fewer particles.

```
 Drift space example

 &cont npart=1000000 nprnt=3 prlevel=1 /
 &bmt /
1 2 1. 1
0.    0.0     0.        0.      0.      0.200  !mean: x y z px py pz
0.    0.      0.        0.      0.      0.000  !sigs
0
 &ints /
 &nhs  /
 &nsc  nscat =1 /
-0.10 5e-3 40 1 2   -0.10 10e-3 20 2 2
 &nzh  /
 &nrh  /
 &nem  /
 &ncv  /

SECTION
OUTPUT
SREGION
0.30   1  0.003
1  0.  0.10
NONE
  0. 0. 0. 0. 0.   0. 0. 0. 0. 0.     0. 0. 0. 0. 0.
LH
CBLOCK
   0. 0. 0. 0. 0.   0. 0. 0. 0. 0.
ENDSECTION
```

Appendix E

Listings of the Optimised Designs

Table 9.1: Muon1 component listing of cell `SolenoidsOnly`.

RodSource	File	RodAngle	Z
`TantalumRod`	'pi_plus_2G2eV'	0.0955956	0.24865
Solenoid	Length	Radius	Field
S1	0.45	0.1	20
S2	0.5996	0.3997	5
S3	0.48308	0.3994	4.95996
S4	0.55275	0.4	4.23924
S5	0.6	0.4	3.99199
S6	0.5992	0.4	3.96797
S7	0.5996	0.3994	3.97598
S8	0.6	0.3997	4
S9	0.6	0.4	4
S10	0.5992	0.3982	3.97598
S11	0.5996	0.4	3.97598
S12	0.5988	0.3994	3.99199
S13	0.5992	0.4	3.99199
S14	0.6	0.3997	4
S15	0.5988	0.4	4
S16	0.6	0.4	4
S17	0.5992	0.3997	3.99199
S18	0.5996	0.4	4
S19	0.6	0.4	4
S20	0.6	0.3997	4
S21	0.5996	0.3997	3.99199
S22	0.6	0.3985	4
S23	0.6	0.4	3.97598
S24	0.6	0.4	3.98398
S25	0.5988	0.397	3.96797
S26	0.6	0.4	4

continued from previous page

S27	0.5988	0.3997	3.99199
S28	0.5972	0.4	3.95996
S29	0.52873	0.3985	3.98398

Drift	Length	Bore
D1	0.5	0.5
D2	0.5	0.5
D3	0.5	0.5
D4	0.5	0.5
D5	0.5045	0.5
D6	0.5	0.5
D7	0.5	0.5
D8	0.5	0.5
D9	0.5	0.5
D10	0.5	0.5
D11	0.5	0.5
D12	0.5	0.5
D13	0.5	0.5
D14	0.5	0.5
D15	0.5	0.5
D16	0.5	0.5
D17	0.5	0.5
D18	0.5	0.5
D19	0.5	0.5
D20	0.5	0.5
D21	0.51101	0.5
D22	0.5	0.5
D23	0.5	0.5
D24	0.5	0.5
D25	0.5005	0.5
D26	0.5	0.5
D27	0.5	0.5
D28	0.504	0.5

Beamline order: TantalumRod, S1, D1, S2, D2, S3, D3, S4, D4, S5, D5, S6, D6, S7, D7, S8, D8, S9, D9, S10, D10, S11, D11, S12, D12, S13, D13, S14, D14, S15, D15, S16, D16, S17, D17, S18, D18, S19, D19, S20, D20, S21, D21, S22, D22, S23, D23, S24, D24, S25, D25, S26, D26, S27, D27, S28, D28, S29

Table 9.2: Muon1 component listing of cell `SolenoidsTo15cm`.

RodSource	File	RodAngle	Z
`TantalumRod`	'pi_plus_2G2eV'	0.0145145	0.22252
Solenoid	Length	Radius	Field
`S1`	0.44875	0.1	20
`S2`	0.6	0.32523	5
`S3`	0.42102	0.4	5
`S4`	0.33894	0.3976	-5
`S5`	0.52673	0.3991	3.94394
`S6`	0.5992	0.3991	3.7998
`S7`	0.598	0.3964	3.94394
`S8`	0.59399	0.3979	4
`S9`	0.5988	0.3979	3.96797
`S10`	0.5988	0.3958	3.86386
`S11`	0.5952	0.4	3.97598
`S12`	0.6	0.4	3.95195
`S13`	0.6	0.3997	4
`S14`	0.6	0.38709	3.88789
`S15`	0.58999	0.4	3.88789
`S16`	0.6	0.3976	3.8959
`S17`	0.5964	0.3997	3.93594
`S18`	0.6	0.4	3.91191
`S19`	0.5964	0.3988	3.9039
`S20`	0.5952	0.3967	3.8959
`S21`	0.59399	0.4	3.93594
`S22`	0.5972	0.3985	4
`S23`	0.5964	0.3955	3.9039
`S24`	0.596	0.3991	3.97598
`S25`	0.596	0.39459	3.79179
`S26`	0.6	0.3973	-3.61562
`S27`	0.6	0.3982	3.45546
`S28`	0.6	0.31952	4
`Sfinal`	0.2	0.15	4
Drift	Length	Bore	
`D1`	0.5	0.5	
`D2`	0.5	0.5	
`D3`	0.5	0.5	
`D4`	0.65415	0.5	
`D5`	0.66667	0.5	
`D6`	0.57307	0.5	
`D7`	0.5	0.5	
`D8`	0.5	0.5	
`D9`	0.5	0.5	
`D10`	0.50801	0.5	
`D11`	0.5	0.5	

continued from previous page

D12	0.5	0.5
D13	0.5	0.5
D14	0.53153	0.5
D15	0.5	0.5
D16	0.5	0.5
D17	0.5	0.5
D18	0.5	0.5
D19	0.51001	0.5
D20	0.51602	0.5
D21	0.5	0.5
D22	0.5	0.5
D23	0.5	0.5
D24	0.56356	0.5
D25	0.50651	0.5
D26	0.77628	0.5
D27	0.65766	0.5
D28	0.5	0.5

Beamline order: TantalumRod, S1, D1, S2, D2, S3, D3, S4, D4, S5, D5, S6, D6, S7, D7, S8, D8, S9, D9, S10, D10, S11, D11, S12, D12, S13, D13, S14, D14, S15, D15, S16, D16, S17, D17, S18, D18, S19, D19, S20, D20, S21, D21, S22, D22, S23, D23, S24, D24, S25, D25, S26, D26, S27, D27, S28, D28, Sfinal

Table 9.3: Muon1 component listing of cell `ChicaneLinacB`.

RodSource	RodAngle	Z	File
`TantalumRod`	0	0.2009	'pi_plus_2G2eV.csv'
Solenoid	Length	Radius	Field
`S1`	0.39494	0.1	19.9199
`S2`	0.84675	0.36847	4.81982
`S3`	0.71011	0.37508	4.0991
`S4`	0.79419	0.37988	4.07908
`S5`	0.87477	0.36877	3.2953
`S6`	0.83904	0.36907	3.27928
`S7`	0.83273	0.37808	3.1992
`S8`	0.84184	0.36907	3.11111
`S9`	0.83694	0.39309	3.16717
`S10`	0.82993	0.37658	3.17518
`S11`	0.83904	0.39069	3.17518
`S12`	0.83203	0.37147	3.15115
`S13`	0.82012	0.36366	3.15115
`S14`	0.82012	0.38018	3.15916
`S15`	0.83554	0.3961	3.18318
`S16`	0.83904	0.37327	3.25526
`S17`	0.81942	0.37538	3.23924
`S18`	0.82993	0.37628	3.39139
`S19`	0.83063	0.36907	3.43143
`S20`	0.76056	0.37207	3.33534
`S21`	0.82643	0.36246	3.3994
`S22`	0.82783	0.37658	3.43944
`S23`	0.83554	0.37147	2.67868
`S24`	0.82292	0.26306	2.18218
`LS1`	0.82152	0.15	2.74274
`LS2`	0.84254	0.15	3.14314
`LS3`	0.86426	0.15	3.23123
`LS4`	0.88318	0.15	3.27928
`LS5`	0.82853	0.15	3.38338
`LS6`	0.83343	0.15	3.27127
`LS7`	0.86076	0.15	3.2953
`LS8`	0.89019	0.15	3.24725
`LS9`	0.83203	0.15	3.28729
`LS10`	0.82222	0.15	3.39139
`LS11`	0.82643	0.15	3.36737
`LS12`	0.82222	0.15	3.35936
`LS13`	0.81942	0.15	3.27127
`LS14`	0.83904	0.15	3.13514
`LS15`	0.84184	0.15	3.17518
`LS16`	0.84254	0.15	3.27127
`LS17`	0.86076	0.15	3.22322

continued from previous page

LS18	0.83273	0.15	3.20721
LS19	0.8019	0.15	3.2953
LS20	0.80681	0.15	3.31932
LS21	0.81662	0.15	3.31131
LS22	0.83764	0.15	3.21522
LS23	0.85025	0.15	3.1992
LS24	0.84324	0.15	3.35135
LS25	0.83904	0.15	3.36737
LS26	0.82853	0.15	3.36737
LS27	0.83343	0.15	3.3033
LS28	0.83343	0.15	3.28729
LS29	0.82503	0.15	3.19119
LS30	0.83133	0.15	3.28729
LS31	0.83834	0.15	3.31131
LS32	0.83834	0.15	3.45546
LS33	0.83203	0.15	3.52753
LS34	0.84955	0.15	3.3033
LS35	0.83694	0.15	3.17518
LS36	0.83343	0.15	3.19119
LS37	0.83343	0.15	3.19119
LS38	0.83554	0.15	3.19119
LS39	0.84184	0.15	3.21522
LS40	0.82923	0.15	3.24725
LS41	0.82853	0.15	3.47948
LS42	0.84114	0.15	3.27127
LS43	0.86917	0.15	3.27127
LS44	0.85866	0.15	3.24725
LS45	0.89089	0.15	3.1992
LS46	0.84464	0.15	3.22322
LS47	0.81592	0.15	3.27928
LS48	0.84254	0.15	3.19119
LS49	0.85095	0.15	3.15916
LS50	0.82993	0.15	3.06306
LS51	0.84464	0.15	3.17518
LS52	0.82993	0.15	3.21522
LS53	0.82082	0.15	3.23924
LS54	0.83273	0.15	3.1992
LS55	0.84535	0.15	3.15115
LS56	0.85235	0.15	3.32733
LS57	0.85936	0.15	3.22322
LS58	0.86637	0.15	3.13514
LS59	0.87127	0.15	3.16717
LS60	0.87057	0.15	3.24725

Drift	Length	Bore
D1	0.5	0.5
D2	0.5	0.5

continued from previous page

D3	0.5	0.5
D4	0.5	0.5
D5	0.50501	0.5
D6	0.5	0.5
D7	0.5	0.5
D8	0.5	0.5
D9	0.5	0.5
D10	0.5	0.5
D11	0.5	0.5
D12	0.5	0.5
D13	0.5	0.5
D14	0.5	0.5
D15	0.503	0.5
D16	0.5025	0.5
D17	0.5	0.5
D18	0.5	0.5
D19	0.5	0.5
D20	0.5	0.5
D21	0.5	0.5
D22	0.5	0.5
D23	0.5	0.5
O2	0.2144	1
O1	0.4713	1
EDGE	0	1
HBF	0.176	1
HBD	0.176	1
HalfHBD	0.088	1
∼HBF	0.176	1
∼HBD	0.176	1
∼HalfHBD	0.088	1
LD1a	0.0833333	0.15
LD1b	0.16667	0.15
LD2a	0.33333	0.15
LD2b	0.16667	0.15
LD3a	0.33333	0.15
LD3b	0.16667	0.15
LD4a	0.33333	0.15
LD4b	0.16667	0.15
LD5a	0.33333	0.15
LD5b	0.16667	0.15
LD6a	0.33333	0.15
LD6b	0.16667	0.15
LD7a	0.33333	0.15
LD7b	0.16667	0.15
LD8a	0.33333	0.15
LD8b	0.16667	0.15

continued from previous page

LD9a	0.33333	0.15
LD9b	0.16667	0.15
LD10a	0.33333	0.15
LD10b	0.16667	0.15
LD11a	0.33333	0.15
LD11b	0.16667	0.15
LD12a	0.33333	0.15
LD12b	0.16667	0.15
LD13a	0.33333	0.15
LD13b	0.16667	0.15
LD14a	0.33333	0.15
LD14b	0.16667	0.15
LD15a	0.33333	0.15
LD15b	0.16667	0.15
LD16a	0.33333	0.15
LD16b	0.16667	0.15
LD17a	0.33333	0.15
LD17b	0.16667	0.15
LD18a	0.33333	0.15
LD18b	0.16667	0.15
LD19a	0.33333	0.15
LD19b	0.16667	0.15
LD20a	0.33333	0.15
LD20b	0.16667	0.15
LD21a	0.33333	0.15
LD21b	0.16667	0.15
LD22a	0.33333	0.15
LD22b	0.16667	0.15
LD23a	0.33333	0.15
LD23b	0.16667	0.15
LD24a	0.33333	0.15
LD24b	0.16667	0.15
LD25a	0.33333	0.15
LD25b	0.16667	0.15
LD26a	0.33333	0.15
LD26b	0.16667	0.15
LD27a	0.33333	0.15
LD27b	0.16667	0.15
LD28a	0.33333	0.15
LD28b	0.16667	0.15
LD29a	0.33333	0.15
LD29b	0.16667	0.15
LD30a	0.33333	0.15
LD30b	0.16667	0.15
LD31a	0.33333	0.15
LD31b	0.16667	0.15

continued from previous page

LD32a	0.33333	0.15
LD32b	0.16667	0.15
LD33a	0.33333	0.15
LD33b	0.16667	0.15
LD34a	0.33333	0.15
LD34b	0.16667	0.15
LD35a	0.33333	0.15
LD35b	0.16667	0.15
LD36a	0.33333	0.15
LD36b	0.16667	0.15
LD37a	0.33333	0.15
LD37b	0.16667	0.15
LD38a	0.33333	0.15
LD38b	0.16667	0.15
LD39a	0.33333	0.15
LD39b	0.16667	0.15
LD40a	0.33333	0.15
LD40b	0.16667	0.15
LD41a	0.33333	0.15
LD41b	0.16667	0.15
LD42a	0.33333	0.15
LD42b	0.16667	0.15
LD43a	0.33333	0.15
LD43b	0.16667	0.15
LD44a	0.33333	0.15
LD44b	0.16667	0.15
LD45a	0.33333	0.15
LD45b	0.16667	0.15
LD46a	0.33333	0.15
LD46b	0.16667	0.15
LD47a	0.33333	0.15
LD47b	0.16667	0.15
LD48a	0.33333	0.15
LD48b	0.16667	0.15
LD49a	0.33333	0.15
LD49b	0.16667	0.15
LD50a	0.33333	0.15
LD50b	0.16667	0.15
LD51a	0.33333	0.15
LD51b	0.16667	0.15
LD52a	0.33333	0.15
LD52b	0.16667	0.15
LD53a	0.33333	0.15
LD53b	0.16667	0.15
LD54a	0.33333	0.15
LD54b	0.16667	0.15

continued from previous page

LD55a	0.33333	0.15
LD55b	0.16667	0.15
LD56a	0.33333	0.15
LD56b	0.16667	0.15
LD57a	0.33333	0.15
LD57b	0.16667	0.15
LD58a	0.33333	0.15
LD58b	0.16667	0.15
LD59a	0.33333	0.15
LD59b	0.16667	0.15
LD60a	0.33333	0.15
LD60b	0.16667	0.15

Rectangular-aperture	x1	x2	y1	y2
B1A	-0.6	0.6	-0.15	0.15
B2A	-0.6	0.4	-0.15	0.15
∼B1A	-0.6	0.6	-0.15	0.15
∼B2A	-0.6	0.4	-0.15	0.15

BMagnet	File1	File2	File3	File4
L2_FIELD	'l2y0.txt'	'l2y5.txt'	'l2y10.txt'	'l2y13.txt'
∼L2_FIELD	'l2y0.txt'	'l2y5.txt'	'l2y10.txt'	'l2y13.txt'

RF	Frequency	Radius	Voltage	Phase
RF1	88 MHz	0.15	5.28079 MV	-64.6847°
RF2	88 MHz	0.15	5.31104 MV	-63.964°
RF3	88 MHz	0.15	5.45706 MV	-49.9099°
RF4	88 MHz	0.15	4.30324 MV	-42.7027°
RF5	88 MHz	0.15	3.88851 MV	-51.7117°
RF6	88 MHz	0.15	5.00805 MV	-44.1441°
RF7	88 MHz	0.15	4.6857 MV	-56.3964°
RF8	88 MHz	0.15	5.44943 MV	-69.7297°
RF9	88 MHz	0.15	5.20546 MV	-76.2162°
RF10	88 MHz	0.15	4.10299 MV	-70.4505°
RF11	88 MHz	0.15	3.84518 MV	-53.8739°
RF12	88 MHz	0.15	4.06064 MV	-54.2342°
RF13	88 MHz	0.15	5.03466 MV	-44.8649°
RF14	88 MHz	0.15	3.93535 MV	-39.0991°
RF15	88 MHz	0.15	4.15313 MV	-36.2162°
RF16	88 MHz	0.15	3.94565 MV	-33.3333°
RF17	88 MHz	0.15	3.66139 MV	-28.2883°
RF18	88 MHz	0.15	3.59131 MV	-22.5225°
RF19	88 MHz	0.15	3.70632 MV	-30.0901°
RF20	88 MHz	0.15	3.78307 MV	-35.8559°
RF21	88 MHz	0.15	3.86418 MV	-43.7838°
RF22	88 MHz	0.15	4.02229 MV	-52.7928°
RF23	88 MHz	0.15	3.86558 MV	-45.9459°
RF24	88 MHz	0.15	4.30806 MV	-50.6306°
RF25	88 MHz	0.15	3.75306 MV	-47.7477°

continued from previous page

RF26	88 MHz	0.15	4.05873 MV	-45.5856°
RF27	88 MHz	0.15	3.96693 MV	-45.5856°
RF28	88 MHz	0.15	4.39406 MV	-44.8649°
RF29	88 MHz	0.15	4.18596 MV	-40.9009°
RF30	88 MHz	0.15	3.90204 MV	-43.7838°
RF31	88 MHz	0.15	4.21195 MV	-45.2252°
RF32	88 MHz	0.15	3.97616 MV	-45.5856°
RF33	88 MHz	0.15	3.52008 MV	-50.991°
RF34	88 MHz	0.15	4.85243 MV	-62.5225°
RF35	88 MHz	0.15	4.20218 MV	-59.6396°
RF36	88 MHz	0.15	3.67328 MV	-61.4414°
RF37	88 MHz	0.15	4.09507 MV	-54.955°
RF38	88 MHz	0.15	4.786 MV	-50.2703°
RF39	88 MHz	0.15	1.81061 MV	-40.5405°
RF40	88 MHz	0.15	0.65464 MV	-39.0991°
RF41	88 MHz	0.15	3.67041 MV	-38.3784°
RF42	88 MHz	0.15	4.30132 MV	-36.9369°
RF43	88 MHz	0.15	4.72014 MV	-36.2162°
RF44	88 MHz	0.15	5.32583 MV	-32.973°
RF45	88 MHz	0.15	5.5023 MV	-37.2973°
RF46	88 MHz	0.15	5.29782 MV	-38.018°
RF47	88 MHz	0.15	4.56816 MV	-38.018°
RF48	88 MHz	0.15	4.35957 MV	-42.7027°
RF49	88 MHz	0.15	3.84054 MV	-26.8468°
RF50	88 MHz	0.15	3.78611 MV	-34.4144°
RF51	88 MHz	0.15	3.89799 MV	-31.5315°
RF52	88 MHz	0.15	3.82338 MV	-32.6126°
RF53	88 MHz	0.15	3.66498 MV	-29.009°
RF54	88 MHz	0.15	4.22632 MV	-23.2432°
RF55	88 MHz	0.15	4.09933 MV	-19.6396°
RF56	88 MHz	0.15	4.25064 MV	-37.2973°
RF57	88 MHz	0.15	4.10937 MV	-47.3874°
RF58	88 MHz	0.15	4.16338 MV	-71.8919°
RF59	88 MHz	0.15	3.94223 MV	-152.973°
RF60	88 MHz	0.15	3.79754 MV	-70.0901°

continued from previous page

Beamline order: TantalumRod, S1, D1, S2, D2, S3, D3, S4, D4, S5, D5, S6, D6, S7, D7, S8, D8, S9, D9, S10, D10, S11, D11, S12, D12, S13, D13, S14, D14, S15, D15, S16, D16, S17, D17, S18, D18, S19, D19, S20, D20, S21, D21, S22, D22, S23, D23, S24, 2*(O2, O1, EDGE, B1A, HBF, HBF, HBF, HBF, B1A, EDGE, O1, B2A, HBD, HalfHBD, L2_FIELD, HalfHBD, HBD, B2A, O1, EDGE, B1A, HBF, HBF, HBF, HBF, B1A, EDGE, O1, O2), 2*(O2, O1, EDGE, ∼B1A, ∼HBF, ∼HBF, ∼HBF, ∼HBF, ∼B1A, EDGE, O1, ∼B2A, ∼HBD, ∼HalfHBD, ∼L2_FIELD, ∼HalfHBD, ∼HBD, ∼B2A, O1, EDGE, ∼B1A, ∼HBF, ∼HBF, ∼HBF, ∼HBF, ∼B1A, EDGE, O1, O2), LD1a, RF1, LD1b, LS1, LD2a, RF2, LD2b, LS2, LD3a, RF3, LD3b, LS3, LD4a, RF4, LD4b, LS4, LD5a, RF5, LD5b, LS5, LD6a, RF6, LD6b, LS6, LD7a, RF7, LD7b, LS7, LD8a, RF8, LD8b, LS8, LD9a, RF9, LD9b, LS9, LD10a, RF10, LD10b, LS10, LD11a, RF11, LD11b, LS11, LD12a, RF12, LD12b, LS12, LD13a, RF13, LD13b, LS13, LD14a, RF14, LD14b, LS14, LD15a, RF15, LD15b, LS15, LD16a, RF16, LD16b, LS16, LD17a, RF17, LD17b, LS17, LD18a, RF18, LD18b, LS18, LD19a, RF19, LD19b, LS19, LD20a, RF20, LD20b, LS20, LD21a, RF21, LD21b, LS21, LD22a, RF22, LD22b, LS22, LD23a, RF23, LD23b, LS23, LD24a, RF24, LD24b, LS24, LD25a, RF25, LD25b, LS25, LD26a, RF26, LD26b, LS26, LD27a, RF27, LD27b, LS27, LD28a, RF28, LD28b, LS28, LD29a, RF29, LD29b, LS29, LD30a, RF30, LD30b, LS30, LD31a, RF31, LD31b, LS31, LD32a, RF32, LD32b, LS32, LD33a, RF33, LD33b, LS33, LD34a, RF34, LD34b, LS34, LD35a, RF35, LD35b, LS35, LD36a, RF36, LD36b, LS36, LD37a, RF37, LD37b, LS37, LD38a, RF38, LD38b, LS38, LD39a, RF39, LD39b, LS39, LD40a, RF40, LD40b, LS40, LD41a, RF41, LD41b, LS41, LD42a, RF42, LD42b, LS42, LD43a, RF43, LD43b, LS43, LD44a, RF44, LD44b, LS44, LD45a, RF45, LD45b, LS45, LD46a, RF46, LD46b, LS46, LD47a, RF47, LD47b, LS47, LD48a, RF48, LD48b, LS48, LD49a, RF49, LD49b, LS49, LD50a, RF50, LD50b, LS50, LD51a, RF51, LD51b, LS51, LD52a, RF52, LD52b, LS52, LD53a, RF53, LD53b, LS53, LD54a, RF54, LD54b, LS54, LD55a, RF55, LD55b, LS55, LD56a, RF56, LD56b, LS56, LD57a, RF57, LD57b, LS57, LD58a, RF58, LD58b, LS58, LD59a, RF59, LD59b, LS59, LD60a, RF60, LD60b, LS60

Table 9.4: Muon1 component listing of cell `PhaseRotB`.

RodSource	RodAngle	Z	File
`TantalumRod`	0	0.26802	'pi_plus_2G2eV.csv'
Solenoid	Length	Radius	Field
`S1`	0.43524	0.1	19.8599
`S2`	0.81592	0.37688	4.88989
`S3`	0.74164	0.38288	4.75976
`S4`	0.72062	0.3955	4.90991
`S5`	0.84464	0.39459	3.93594
`S6`	0.85375	0.3952	3.9039
`S7`	0.87898	0.397	3.83984
`S8`	0.88669	0.3967	3.80781
`S9`	0.89299	0.3952	3.91191
`S10`	0.89299	0.3955	3.93594
`S11`	0.89229	0.3952	3.91992
`S12`	0.89229	0.39429	3.91191
`S13`	0.89299	0.39429	3.9039
`S14`	0.89159	0.39399	3.91992
`S15`	0.88949	0.3952	3.91992
`S16`	0.8979	0.3952	3.94394
`S17`	0.88599	0.3952	3.91992
`S18`	0.9	0.39429	3.91191
`S19`	0.88458	0.39489	3.93594
`S20`	0.89439	0.3952	3.82382
`S21`	0.89159	0.3958	3.75175
`S22`	0.89089	0.3973	3.57558
`S23`	0.89159	0.39219	3.61562
`PS1`	0.89019	0.15	-2.96697
`PS2`	0.89089	0.15	-3.02302
`PS3`	0.89089	0.15	-3.04705
`PS4`	0.89369	0.15	-3.07107
`PS5`	0.89229	0.15	-3.07107
`PS6`	0.89229	0.15	-3.07908
`PS7`	0.8951	0.15	-3.06306
`PS8`	0.8951	0.15	-3.08709
`PS9`	0.89439	0.15	-3.07107
`PS10`	0.89439	0.15	-3.03904
`PS11`	0.8965	0.15	-3.04705
`PS12`	0.89439	0.15	-3.07107
`PS13`	0.8958	0.15	-3.07107
`PS14`	0.8986	0.15	-3.05506
`PS15`	0.89369	0.15	-3.07908
`PS16`	0.89439	0.15	-3.0951
`PS17`	0.89299	0.15	-3.07107
`PS18`	0.8965	0.15	-3.06306

continued from previous page

PS19	0.89369	0.15	-3.0951
PS20	0.89159	0.15	-3.08709
PS21	0.89229	0.15	-3.08709
PS22	0.8951	0.15	-3.07908
PS23	0.89369	0.15	-3.08709
PS24	0.89299	0.15	-3.11111
PS25	0.89229	0.15	-3.13514
PS26	0.89229	0.15	-3.15916
PS27	0.8958	0.15	-3.16717
PS28	0.89439	0.15	-3.11111
PS29	0.88949	0.15	-2.94294
PS30	0.89229	0.15	-3.12713

Drift	Length	Bore
D1	0.5005	0.5
D2	0.81281	0.5
D3	0.51301	0.5
D4	0.50801	0.5
D5	0.5	0.5
D6	0.5	0.5
D7	0.5	0.5
D8	0.501	0.5
D9	0.504	0.5
D10	0.5005	0.5
D11	0.5	0.5
D12	0.5	0.5
D13	0.5	0.5
D14	0.5	0.5
D15	0.5	0.5
D16	0.5	0.5
D17	0.5	0.5
D18	0.5	0.5
D19	0.5	0.5
D20	0.5	0.5
D21	0.5	0.5
D22	0.5	0.5
PD1a	0.0833333	0.15
PD1b	0.16667	0.15
PD2a	0.33333	0.15
PD2b	0.16667	0.15
PD3a	0.33333	0.15
PD3b	0.16667	0.15
PD4a	0.33333	0.15
PD4b	0.16667	0.15
PD5a	0.33333	0.15
PD5b	0.16667	0.15
PD6a	0.33333	0.15

continued from previous page

PD6b	0.16667	0.15
PD7a	0.33333	0.15
PD7b	0.16667	0.15
PD8a	0.33333	0.15
PD8b	0.16667	0.15
PD9a	0.33333	0.15
PD9b	0.16667	0.15
PD10a	0.33333	0.15
PD10b	0.16667	0.15
PD11a	0.33333	0.15
PD11b	0.16667	0.15
PD12a	0.33333	0.15
PD12b	0.16667	0.15
PD13a	0.33367	0.15
PD13b	0.16683	0.15
PD14a	0.33333	0.15
PD14b	0.16667	0.15
PD15a	0.33333	0.15
PD15b	0.16667	0.15
PD16a	0.33333	0.15
PD16b	0.16667	0.15
PD17a	0.33333	0.15
PD17b	0.16667	0.15
PD18a	0.33333	0.15
PD18b	0.16667	0.15
PD19a	0.33333	0.15
PD19b	0.16667	0.15
PD20a	0.33333	0.15
PD20b	0.16667	0.15
PD21a	0.33333	0.15
PD21b	0.16667	0.15
PD22a	0.33333	0.15
PD22b	0.16667	0.15
PD23a	0.33333	0.15
PD23b	0.16667	0.15
PD24a	0.33333	0.15
PD24b	0.16667	0.15
PD25a	0.33333	0.15
PD25b	0.16667	0.15
PD26a	0.33333	0.15
PD26b	0.16667	0.15
PD27a	0.33333	0.15
PD27b	0.16667	0.15
PD28a	0.33333	0.15
PD28b	0.16667	0.15
PD29a	0.33333	0.15

continued from previous page

PD29b	0.16667	0.15
PD30a	0.33333	0.15
PD30b	0.16667	0.15

RF	Frequency	Radius	Voltage	Phase
PRF1	31.4285 MHz	0.15	2.2074 MV	-114.414°
PRF2	31.4285 MHz	0.15	2.2197 MV	-113.333°
PRF3	31.4285 MHz	0.15	2.2197 MV	-107.928°
PRF4	31.4285 MHz	0.15	2.21744 MV	-113.694°
PRF5	31.4285 MHz	0.15	2.21297 MV	-112.973°
PRF6	31.4285 MHz	0.15	2.21297 MV	-112.252°
PRF7	31.4285 MHz	0.15	2.2309 MV	-113.694°
PRF8	31.4285 MHz	0.15	2.22865 MV	-114.414°
PRF9	31.4285 MHz	0.15	2.21183 MV	-114.054°
PRF10	31.4285 MHz	0.15	2.21407 MV	-113.694°
PRF11	31.4285 MHz	0.15	2.21966 MV	-113.333°
PRF12	31.4285 MHz	0.15	2.21407 MV	-113.694°
PRF13	31.4285 MHz	0.15	2.22832 MV	-112.973°
PRF14	31.4285 MHz	0.15	2.2275 MV	-113.694°
PRF15	31.4285 MHz	0.15	2.21296 MV	-113.694°
PRF16	31.4285 MHz	0.15	2.21407 MV	-112.252°
PRF17	31.4285 MHz	0.15	2.21409 MV	-114.054°
PRF18	31.4285 MHz	0.15	2.2219 MV	-111.892°
PRF19	31.4285 MHz	0.15	2.21969 MV	-111.892°
PRF20	31.4285 MHz	0.15	2.21186 MV	-112.973°
PRF21	31.4285 MHz	0.15	2.21969 MV	-114.414°
PRF22	31.4285 MHz	0.15	2.21743 MV	-113.333°
PRF23	31.4285 MHz	0.15	2.22193 MV	-83.0631°
PRF24	31.4285 MHz	0.15	2.23874 MV	-104.324°
PRF25	31.4285 MHz	0.15	2.21521 MV	-107.568°
PRF26	31.4285 MHz	0.15	2.21745 MV	-112.973°
PRF27	31.4285 MHz	0.15	2.22528 MV	-114.054°
PRF28	31.4285 MHz	0.15	2.2208 MV	-112.252°
PRF29	31.4285 MHz	0.15	2.20852 MV	-112.973°
PRF30	31.4285 MHz	0.15	2.22417 MV	-112.613°

continued from previous page

Beamline order: TantalumRod, S1, D1, S2, D2, S3, D3, S4, D4, S5, D5, S6, D6, S7, D7, S8, D8, S9, D9, S10, D10, S11, D11, S12, D12, S13, D13, S14, D14, S15, D15, S16, D16, S17, D17, S18, D18, S19, D19, S20, D20, S21, D21, S22, D22, S23, PD1a, PRF1, PD1b, PS1, PD2a, PRF2, PD2b, PS2, PD3a, PRF3, PD3b, PS3, PD4a, PRF4, PD4b, PS4, PD5a, PRF5, PD5b, PS5, PD6a, PRF6, PD6b, PS6, PD7a, PRF7, PD7b, PS7, PD8a, PRF8, PD8b, PS8, PD9a, PRF9, PD9b, PS9, PD10a, PRF10, PD10b, PS10, PD11a, PRF11, PD11b, PS11, PD12a, PRF12, PD12b, PS12, PD13a, PRF13, PD13b, PS13, PD14a, PRF14, PD14b, PS14, PD15a, PRF15, PD15b, PS15, PD16a, PRF16, PD16b, PS16, PD17a, PRF17, PD17b, PS17, PD18a, PRF18, PD18b, PS18, PD19a, PRF19, PD19b, PS19, PD20a, PRF20, PD20b, PS20, PD21a, PRF21, PD21b, PS21, PD22a, PRF22, PD22b, PS22, PD23a, PRF23, PD23b, PS23, PD24a, PRF24, PD24b, PS24, PD25a, PRF25, PD25b, PS25, PD26a, PRF26, PD26b, PS26, PD27a, PRF27, PD27b, PS27, PD28a, PRF28, PD28b, PS28, PD29a, PRF29, PD29b, PS29, PD30a, PRF30, PD30b, PS30

Table 9.5: Muon1 component listing of cell `PhaseRotC`.

RodSource	File	RodAngle	Z
`TantalumRod`	'pi_plus_10GeV.csv'	0	0.26396
Solenoid	Length	Radius	Field
`S1`	0.44274	0.1	19.9199
`S2`	0.81592	0.37898	4.83984
`S3`	0.75355	0.4	4.86987
`S4`	0.72062	0.38919	4.88989
`S5`	0.80821	0.39099	3.87988
`S6`	0.85025	0.3958	-3.81582
`S7`	0.87618	0.39459	3.85586
`S8`	0.88739	0.39489	3.84785
`S9`	0.88669	0.39309	3.91191
`S10`	0.89019	0.3952	3.88789
`S11`	0.88949	0.39279	3.87988
`S12`	0.88949	0.39249	3.87988
`S13`	0.89019	0.39189	3.87187
`S14`	0.89019	0.39129	3.87988
`S15`	0.88318	0.3952	3.85586
`S16`	0.87968	0.3952	3.86386
`S17`	0.88108	0.39159	3.84785
`S18`	0.88248	0.39159	3.8959
`S19`	0.9	0.39159	3.81582
`S20`	0.88949	0.39219	3.77578
`S21`	0.88669	0.3964	3.66366
`S22`	0.85866	0.3967	3.59159
`S23`	0.88739	0.38859	3.5996
`PS1`	0.87758	0.15	-3.05506
`PS2`	0.87828	0.15	-3.04705
`PS3`	0.87828	0.15	-3.06306
`PS4`	0.87828	0.15	-3.1031
`PS5`	0.89299	0.15	-3.08709
`PS6`	0.88178	0.15	-3.1031
`PS7`	0.88318	0.15	-3.0951
`PS8`	0.88248	0.15	-3.06306
`PS9`	0.88248	0.15	-3.07107
`PS10`	0.88038	0.15	-3.08709
`PS11`	0.89439	0.15	-3.07908
`PS12`	0.8951	0.15	-3.08709
`PS13`	0.89299	0.15	-3.06306
`PS14`	0.88739	0.15	-3.07107
`PS15`	0.88529	0.15	-3.0951
`PS16`	0.88739	0.15	-3.11111
`PS17`	0.88669	0.15	-3.08709
`PS18`	0.88529	0.15	-3.08709

continued from previous page

PS19	0.88318	0.15	-3.07908
PS20	0.88038	0.15	-3.08709
PS21	0.88108	0.15	-3.0951
PS22	0.88178	0.15	-3.07107
PS23	0.88388	0.15	-3.11111
PS24	0.88178	0.15	-3.15115
PS25	0.87968	0.15	-3.16717
PS26	0.87968	0.15	-3.21522
PS27	0.89299	0.15	-3.20721
PS28	0.89229	0.15	-3.14314
PS29	0.87828	0.15	-3.04705
PS30	0.87828	0.15	-2.96697

Drift	Length	Bore
D1	0.5	0.5
D2	0.80881	0.5
D3	0.51602	0.5
D4	0.50601	0.5
D5	0.503	0.5
D6	0.5	0.5
D7	0.501	0.5
D8	0.5	0.5
D9	0.5025	0.5
D10	0.5	0.5
D11	0.5	0.5
D12	0.5	0.5
D13	0.5	0.5
D14	0.5	0.5
D15	0.5	0.5
D16	0.5	0.5
D17	0.5	0.5
D18	0.501	0.5
D19	0.5	0.5
D20	0.504	0.5
D21	0.50651	0.5
D22	0.50501	0.5
PD1a	0.0833333	0.15
PD1b	0.16667	0.15
PD2a	0.33333	0.15
PD2b	0.16667	0.15
PD3a	0.33333	0.15
PD3b	0.16667	0.15
PD4a	0.33333	0.15
PD4b	0.16667	0.15
PD5a	0.33333	0.15
PD5b	0.16667	0.15
PD6a	0.33333	0.15

continued from previous page

PD6b	0.16667	0.15
PD7a	0.33333	0.15
PD7b	0.16667	0.15
PD8a	0.33333	0.15
PD8b	0.16667	0.15
PD9a	0.33333	0.15
PD9b	0.16667	0.15
PD10a	0.33333	0.15
PD10b	0.16667	0.15
PD11a	0.33333	0.15
PD11b	0.16667	0.15
PD12a	0.33333	0.15
PD12b	0.16667	0.15
PD13a	0.33333	0.15
PD13b	0.16667	0.15
PD14a	0.33333	0.15
PD14b	0.16667	0.15
PD15a	0.33333	0.15
PD15b	0.16667	0.15
PD16a	0.33333	0.15
PD16b	0.16667	0.15
PD17a	0.33333	0.15
PD17b	0.16667	0.15
PD18a	0.33333	0.15
PD18b	0.16667	0.15
PD19a	0.33333	0.15
PD19b	0.16667	0.15
PD20a	0.33333	0.15
PD20b	0.16667	0.15
PD21a	0.33333	0.15
PD21b	0.16667	0.15
PD22a	0.33333	0.15
PD22b	0.16667	0.15
PD23a	0.33333	0.15
PD23b	0.16667	0.15
PD24a	0.33333	0.15
PD24b	0.16667	0.15
PD25a	0.33333	0.15
PD25b	0.16667	0.15
PD26a	0.33333	0.15
PD26b	0.16667	0.15
PD27a	0.33333	0.15
PD27b	0.16667	0.15
PD28a	0.33333	0.15
PD28b	0.16667	0.15
PD29a	0.33333	0.15

continued from previous page

PD29b	0.16667	0.15
PD30a	0.33333	0.15
PD30b	0.16667	0.15

RF	Frequency	Radius	Voltage	Phase
PRF1	31.4285 MHz	0.15	2.16521 MV	-115.495°
PRF2	31.4285 MHz	0.15	2.16631 MV	-116.216°
PRF3	31.4285 MHz	0.15	2.17518 MV	-114.414°
PRF4	31.4285 MHz	0.15	2.19513 MV	-115.495°
PRF5	31.4285 MHz	0.15	2.17151 MV	-115.495°
PRF6	31.4285 MHz	0.15	2.19627 MV	-154.775°
PRF7	31.4285 MHz	0.15	2.16957 MV	-115.495°
PRF8	31.4285 MHz	0.15	2.17069 MV	-115.856°
PRF9	31.4285 MHz	0.15	2.19738 MV	-115.495°
PRF10	31.4285 MHz	0.15	2.15629 MV	-116.577°
PRF11	31.4285 MHz	0.15	2.1894 MV	-115.135°
PRF12	31.4285 MHz	0.15	2.22192 MV	-112.973°
PRF13	31.4285 MHz	0.15	2.18271 MV	-114.054°
PRF14	31.4285 MHz	0.15	2.17393 MV	-113.694°
PRF15	31.4285 MHz	0.15	2.16172 MV	-113.694°
PRF16	31.4285 MHz	0.15	2.15831 MV	-113.694°
PRF17	31.4285 MHz	0.15	2.1996 MV	-113.333°
PRF18	31.4285 MHz	0.15	2.20184 MV	-113.333°
PRF19	31.4285 MHz	0.15	2.16067 MV	-111.171°
PRF20	31.4285 MHz	0.15	2.15629 MV	-113.333°
PRF21	31.4285 MHz	0.15	2.20182 MV	-81.2613°
PRF22	31.4285 MHz	0.15	2.15848 MV	-113.333°
PRF23	31.4285 MHz	0.15	2.1907 MV	-113.694°
PRF24	31.4285 MHz	0.15	2.16737 MV	-112.613°
PRF25	31.4285 MHz	0.15	2.15298 MV	-113.333°
PRF26	31.4285 MHz	0.15	2.16407 MV	-113.333°
PRF27	31.4285 MHz	0.15	2.21857 MV	-111.892°
PRF28	31.4285 MHz	0.15	2.21521 MV	-112.252°
PRF29	31.4285 MHz	0.15	2.16409 MV	-111.532°
PRF30	31.4285 MHz	0.15	2.21509 MV	-111.892°

continued from previous page

Beamline order: TantalumRod, S1, D1, S2, D2, S3, D3, S4, D4, S5, D5, S6, D6, S7, D7, S8, D8, S9, D9, S10, D10, S11, D11, S12, D12, S13, D13, S14, D14, S15, D15, S16, D16, S17, D17, S18, D18, S19, D19, S20, D20, S21, D21, S22, D22, S23, PD1a, PRF1, PD1b, PS1, PD2a, PRF2, PD2b, PS2, PD3a, PRF3, PD3b, PS3, PD4a, PRF4, PD4b, PS4, PD5a, PRF5, PD5b, PS5, PD6a, PRF6, PD6b, PS6, PD7a, PRF7, PD7b, PS7, PD8a, PRF8, PD8b, PS8, PD9a, PRF9, PD9b, PS9, PD10a, PRF10, PD10b, PS10, PD11a, PRF11, PD11b, PS11, PD12a, PRF12, PD12b, PS12, PD13a, PRF13, PD13b, PS13, PD14a, PRF14, PD14b, PS14, PD15a, PRF15, PD15b, PS15, PD16a, PRF16, PD16b, PS16, PD17a, PRF17, PD17b, PS17, PD18a, PRF18, PD18b, PS18, PD19a, PRF19, PD19b, PS19, PD20a, PRF20, PD20b, PS20, PD21a, PRF21, PD21b, PS21, PD22a, PRF22, PD22b, PS22, PD23a, PRF23, PD23b, PS23, PD24a, PRF24, PD24b, PS24, PD25a, PRF25, PD25b, PS25, PD26a, PRF26, PD26b, PS26, PD27a, PRF27, PD27b, PS27, PD28a, PRF28, PD28b, PS28, PD29a, PRF29, PD29b, PS29, PD30a, PRF30, PD30b, PS30

Table 9.6: Muon1 component listing of cell `PhaseRotEb6a`.

RodSource	File	RodAngle	Z
`TantalumRod`	'pi_plus_10GeV.csv'	0.023023	0.22162
Solenoid	Length	Radius	Field
`S1`	0.42698	0.1	20
`S2`	0.67437	0.2994	4.57958
`S3`	0.31211	0.33213	3.37838
`S4`	0.46206	0.28288	3.46847
`S5`	0.2007	0.27688	1.26126
`S6`	0.66386	0.31742	-2.55856
`S7`	0.46416	0.26787	-2.7027
`S8`	0.75846	0.26156	2.58258
`S9`	0.52372	0.25586	4
`SB1`	0.78298	0.20811	-3.44745
`SB2`	0.78999	0.20751	-3.5035
`SB3`	0.78789	0.20721	-3.52753
`SB4`	0.77598	0.20541	-3.55155
`SB5`	0.77808	0.21141	-3.55956
`SB6`	0.79419	0.2045	-3.52753
`PS1`	0.62883	0.15	3.8959
`PS2`	0.62953	0.15	3.98398
`PS3`	0.64074	0.15	3.95996
`PS4`	0.63584	0.15	4
`PS5`	0.63724	0.15	4
`PS6`	0.63724	0.15	3.99199
Drift	Length	Bore	
`D1`	0.5	0.5	
`D2`	0.67017	0.5	
`D3`	0.5	0.5	
`D4`	0.5005	0.5	
`D5`	0.52252	0.5	
`D6`	0.5	0.5	
`D7`	0.5	0.5	
`D8`	0.52653	0.5	
`D9`	0.59309	0.5	
`DB1`	0.5	0.5	
`DB2`	0.5	0.5	
`DB3`	0.50501	0.5	
`DB4`	0.501	0.5	
`DB5`	0.5	0.5	
`DB6`	0.5	0.5	
`PD1a`	0.33333	0.15	
`PD1b`	0.16667	0.15	
`PD2a`	0.33367	0.15	
`PD2b`	0.16683	0.15	

continued from previous page

PD3a	0.33333	0.15
PD3b	0.16667	0.15
PD4a	0.33567	0.15
PD4b	0.16783	0.15
PD5a	0.33333	0.15
PD5b	0.16667	0.15
PD6a	0.33333	0.15
PD6b	0.16667	0.15

RF	Frequency	Radius	Voltage	Phase
PRF1	31.4285 MHz	0.15	1.68525 MV	-143.604°
PRF2	31.4285 MHz	0.15	1.79067 MV	-141.802°
PRF3	31.4285 MHz	0.15	1.72139 MV	-130.631°
PRF4	31.4285 MHz	0.15	1.72294 MV	-111.171°
PRF5	31.4285 MHz	0.15	1.75452 MV	-91.7117°
PRF6	31.4285 MHz	0.15	1.73074 MV	-81.6216°

Beamline order: TantalumRod, S1, D1, S2, D2, S3, D3, S4, D4, S5, D5, S6, D6, S7, D7, S8, D8, S9, D9, 3*(SB1, DB1), 3*(SB2, DB2), 2*(SB3, DB3), 3*(SB4, DB4), 3*(SB5, DB5), SB6, DB6, SB6, 7*(PD1a, PRF1, PD1b, PS1), 7*(PD2a, PRF2, PD2b, PS2), 7*(PD3a, PRF3, PD3b, PS3), 7*(PD4a, PRF4, PD4b, PS4), 7*(PD5a, PRF5, PD5b, PS5), 6*(PD6a, PRF6, PD6b, PS6)

Table 9.7: Muon1 component listing of cell `PhaseRotLinac1`.

RodSource	File	RodAngle	Z	
`TantalumRod`	'pi_plus_10GeV.csv'	0.02002	0.12928	
Solenoid	Length	Radius	Field	
`S1`	0.43749	0.1	20	
`S2`	0.72903	0.23904	4.92993	
`S3`	0.39059	0.27297	-5	
`S4`	0.7956	0.34414	-1.48649	
`S5`	0.72482	0.3997	3.55155	
`S6`	0.85375	0.37357	1.66967	
`S7`	0.48238	0.26426	-3.65566	
`S8`	0.88108	0.3997	-2.75876	
`S9`	0.7017	0.25255	-4	
`SB1`	0.86496	0.1979	-4	
`PS1`	0.8993	0.15	4	
`LS1`	0.9	0.15	-4	
Drift	Length	Bore		
`D1`	0.5	0.5		
`D2`	0.55455	0.5		
`D3`	0.7973	0.5		
`D4`	0.5	0.5		
`D5`	0.89139	0.5		
`D6`	0.9985	0.5		
`D7`	0.5	0.5		
`D8`	0.91041	0.5		
`D9`	0.83884	0.5		
`DB1`	0.5	0.5		
`PD1a`	0.33333	0.15		
`PD1b`	0.16667	0.15		
`LD1a`	0.33333	0.15		
`LD1b`	0.16667	0.15		
RF	Frequency	Radius	Voltage	Phase
`PRF1`	31.4285 MHz	0.15	2.24887 MV	41.6216°
`RF1`	88 MHz	0.15	0 MV	156.577°

Beamline order: `TantalumRod`, `S1`, `D1`, `S2`, `D2`, `S3`, `D3`, `S4`, `D4`, `S5`, `D5`, `S6`, `D6`, `S7`, `D7`, `S8`, `D8`, `S9`, `D9`, 29*(`SB1`, `DB1`), `SB1`, 60*(`PD1a`, `PRF1`, `PD1b`, `PS1`), 60*(`LD1a`, `RF1`, `LD1b`, `LS1`)

Table 9.8: Muon1 component listing of cell `Linac900Removable6c2`.

RodSource	File	RodAngle	Z	
`TantalumRod`	'pi_plus_10GeV.csv'	0.00500501	0.14144	
Solenoid	Length	Radius	Field	
`S1`	0.45	0.1	20	
`S2`	0.8986	0.19219	5	
`S3`	0.3044	0.38979	4.83984	
`S4`	0.82643	0.3988	5	
`LS1`	0.8993	0.15	-4	
`LS2`	0.9	0.15	-4	
`LS3`	0.8993	0.15	-4	
`LS4`	0.9	0.15	3.93594	
`LS5`	0.9	0.15	-3.95195	
`LS6`	0.89089	0.15	3.91992	
Drift	Length	Bore		
`D1`	0.5	0.5		
`D2`	0.5	0.5		
`D3`	0.5	0.5		
`D4`	0.5015	0.5		
`LD1a`	0.33333	0.15		
`LD1b`	0.16667	0.15		
`LD2a`	0.35569	0.15		
`LD2b`	0.17784	0.15		
`LD3a`	0.33333	0.15		
`LD3b`	0.16667	0.15		
`LD4a`	0.33333	0.15		
`LD4b`	0.16667	0.15		
`LD5a`	0.33567	0.15		
`LD5b`	0.16783	0.15		
`LD6a`	0.33333	0.15		
`LD6b`	0.16667	0.15		
RF	Frequency	Radius	Voltage	Phase
`RF1`	88 MHz	0.15	5.59159 MV	-11.5315°
`RF2`	88 MHz	0.15	5.72839 MV	0°
`RF3`	88 MHz	0.15	5.59159 MV	-5.76577°
`RF4`	88 MHz	0.15	5.57197 MV	10.0901°
`RF5`	88 MHz	0.15	0.0337178 MV	0°
`RF6`	88 MHz	0.15	5.25726 MV	-139.099°

Beamline order: `TantalumRod`, `S1`, `D1`, `S2`, `D2`, `S3`, `D3`, `S4`, `D4`, 40*(`LD1a`, `RF1`, `LD1b`, `LS1`), 40*(`LD2a`, `RF2`, `LD2b`, `LS2`), 40*(`LD3a`, `RF3`, `LD3b`, `LS3`), 40*(`LD4a`, `RF4`, `LD4b`, `LS4`), 40*(`LD5a`, `RF5`, `LD5b`, `LS5`), 40*(`LD6a`, `RF6`, `LD6b`, `LS6`)