DISSERTATION
Ionisation Cooling for a Neutrino Factory

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaften
unter der Leitung von

Univ.-Prof. Dr. Heinz Oberhummer
E141 – Institut für Kernphysik

under the supervision of
Dr. Alessandra Lombardi
CERN-PS

supported by the Austrian Doctoral Student Program

eingereicht an der Technischen Universität Wien
Fakultät für technische Naturwissenschaften und Informatik

von

Dipl.-Ing. Peter Gruber
8926475
1168 Av. de Jura, F-01630 Sergy

Genf, am 10. September 2001

Contents

To Nicole — for all her support,
her love and her patience.

Acknowledgements

I wish to thank Prof. Heinz Oberhummer, who has guided my studies for the last four years. He has encouraged me to apply for a CERN PhD position and he has proven to be literally a "Doktorvater".

I am indebted to my supervisor Alessandra Lombardi — the best "boss" one could imagine. She has left me a lot of freedom in doing my work and she always had an open door for my questions. Alessandra has taught me beam dynamics and Italian with competence and patience.

I want to thank my second supervisor and new boss Helmut Haseroth for all the interesting conversations and good ideas. I owe him my office, my computer and my new job as a fellow.

I also want to thank my friend Simone Gilardoni, for all the fun, all the coffees and all the insight into particle physics he has shared with me.

During my work, many people at the CERN PS division and at other places have supported me, have helped me with my work and found errors in my thesis: Bruno Autin, Michael Benedikt, Mauro Donega, Rob Edgecock, Adrian Fabich, Giuliano Franchetti, Barbara Holzer, Ed McKigney, Arnaud Perrin, Werner Pirkl, Philippe Royer, Richard Scrivens and Horst Schönauer.

Looking back at my studies I am grateful to three people who made me choose physics: my teacher Helmuth Mayr, who raised my interest for physics; Clemens, who gave me the right advice at the right time and my sister Marianne, who has set a good example.

Last but not least, I want to thank my parents who have made it possible for me to study physics.

Abstract

Ionisation cooling is one of the main ingredients of a Neutrino Factory – a high-intensity, high-energy neutrino machine. The cooling channel of the CERN 44/88 MHz baseline design is discussed. A sensitivity study varying various parameters and beam conditions is performed in order to establish critical values.
The question of how to compare different cooling scenarios is tackled by defining assessment criteria for the cooling performance. In the last chapter, "FASTCOOL", a small cooling experiment that would allow to observe ionisation cooling for the first time, is proposed.

Zusammenfassung

Diese Arbeit behandelt Ionisationskühlung für eine Neutrinofabrik ("neutrino factory"). Um die Eigenschaften der verschiedenen Neutrinoarten und besonders ihre Oszillationen studieren zu können, wird zur Zeit eine neue Art Teilchenbeschleuniger, die Neutrinofabrik, diskutiert [1, 2]. Die Neutrinofabrik (siehe auch Abbildung 7) ist ein mehrstufiger Beschleuniger: ein intensiver Protonenstrahl wird auf ein Target gesandt, erzeugt dort Pionen, die mit einer Halbwertszeit von 26 ns in Myonen zerfallen. Diese werden beschleunigt und in einen Speicherring injiziert, wo sie in zwei Neutrinos und ein Elektron zerfallen (siehe Gleichung 4). Das Maximum des Energiespektrums der Neutrinos liegt bei der kinetischen Energie des Myons – je mehr Energie dieses hat, desto höher ist die Energie der Neutrinos.

Bei den Myonen handelt es sich um sekundäre Teilchen, die in einem großen Phasenraumvolumen erzeugt werden. Um die Myonen effizient beschleunigen zu können und um eine geringer Strahldivergenz des Neutrinostrahls zu erreichen, ist es notwendig, dieses Phasenraumvolumen zu reduzieren – den Myonenstrahl zu kühlen. Die bisher praktizierten Verfahren zur Strahlkühlung sind für Myonen nicht geeignet, weil diese kaum Synchrotronstrahlung emittieren und nur eine sehr kurze Lebensdauer (2,2 μs) haben. Deswegen wurde bereits in den 1970er Jahren ein neues Verfahren von G. Budker[3] und A. Skrinsky[4] vorgeschlagen, die Ionisationskühlung. Diese erfolgt in zwei Schritten. Zuerst wird ein Strahl durch einen Absorber aus einem leichten Material — meistens Wasserstoff — gesandt. Dort wird der Teilchenimpuls p isotrop — also die transversale Komponente p_x um den gleichen Faktur wie die longitudinale Komponente p_z — reduziert. Im zweiten Schritt wird der Strahl in einer Radiofrequenzkavität wieder beschleunigt, allerdings nur in z-Richtung. Damit ist p_z am Ende unverändert, während p_x reduziert wurde, Die Strahldivergenz $x' = p_x/p_z$ sinkt.

In dieser Arbeit wird Ionisationskühlung im Hinblick auf eine Neutrinofabrik in fünf Kapiteln behandelt. Der Inhalt im Einzelnen:

Kapitel 2 – das Neutrino. Diese Kapitel ist eine Einführung in die Geschichte und Physik des Neutrinos, seine Entdeckung, seine Erzeugung und sein Nachweis. Nach den Mechanismen der Neutrinoproduktion werden die natürliche Quellen von Neutrinos behandelt: Fusionsreaktionen in Sternen, Pionenproduktion in oberen Schichten der Atmosphäre und β-Zerfälle in Kernkraftwerken. In Abschnitt 2.4 werden Neutrinooszillationen behandelt und die Oszillationswahrscheinlichkeit hergeleitet. Danach folgt ein Abschnitt über die Detektion von Neutrinos mit den wichtigsten Detektotypen, dem umgekehrten β-Zerfall und dem Wasser-Cerenkov Detektor. Danach werden die wichtigsten offenen Fragen in der Neutrinophysik aufgelistet: die Größe des Oszillationswinkels ϑ_{13}, das Vorzeichen der Massendifferenz Δm_{23}^2 und die Frage nach Beweisen für CP-Verletzung bei Neutrinos.

All diese Fragen könnten mit Hilfe einer Neutrinofabrik beantwortet werden, die sieben Kriterien erfüllen muß: hohe Intensität, hohe Myonenenergie, gute Kenntnis der Strahlparamters (und damit des Neutrinoflusses), beide Ladungen (μ^+ und μ^-), zwei Neutrinogenerationen (ν_e und ν_μ), ein sauberer Strahl und eine geringe Strahldivergenz.

Kapitel 3 – Die Neutrinofabrik. Alle in Kapitel 2 aufgestellten Forderungen erfüllt die Neutrinofabrik. Zu Beginn des Kapitels wird die detaillierte Funktionsweise dieses Beschleunigerkomplexes behandelt. Am Anfang steht ein hochintensiver Protonenbeschleuniger dessen Strahl auf ein Target gesandt wird. Dort entstehen über eine Δ-Resonanz Pionen. Diese haben im Mittel einen hohen transversalen Impuls (mehrere hundert MeV) und müssen mit einem starken Magnetfeld gesammelt werden. Nach einem 30 m langen Zerfallstunnel, in dem mehr als 95% der Pionen zerfallen, steht ein Myonenstrahl zur Verfügung. Dieser hat eine hohe Emittanz und eine große Bandbreite bei den Teilchenenergien. Zuerst wird der Energieunterschied mittels Phasenrotation reduziert, danach mit Ionisationskühlung das Phasenraumvolumen reduziert. Der nun etwas kompaktere Myonenstrahl kann nun auf Energien von bis zu 50 GeV beschleunigt und in den Speicherring injiziert werden.

In Abschnitt 3.3 wird das zur Zeit diskutierte Szenario einer Neutrinofabrik am CERN erörtert. Dieses basiert auf einem Linearbeschleuniger für Protonen mit 4 MW und einer Energie von 2,2 GeV. Für die Phasenrotation und Kühlung sind Radiofrequenzkavitäten mit 44 bzw. 88 MHz vorgesehen, die jeweils einen Solenoidmagneten für den Teilchentransport eingebaut haben. Eine solche Kavität ist in Abbildung 16 zu sehen. An Abschnitt 3.4 werden schließlich Möglichkeiten zur Steigerung der Intensität beim CERN Szenario vorgeschlagen: ein längerer Myonenstrahl oder ein kürzerer Zyklus der Neutrinofabrik.

Kapitel 4 – Ionisationskühlung. Nach einer Erörterung des Definition von "Emittanz" werden die vier bisher angewandten Strahlkühlungsverfahren präsentiert: Strahlungskühlung, stochastische Kühlung, Laserkühlung und Elektronenkühlung. All diese Verfahren können für Myonen nicht verwendet werden, nur Ionisationskühlung verspricht brauchbare Ergebnisse.

In Abschnitt 4.4 werden die bei Ionisationskühlung beteiligten physikalischen Prozesse beschrieben: Energieverlust (dE/dx), Moliere Streuung, Fokusierung und Beschleunigung mit rf. Danach wird eine neue Herleitung für die Rate der Kühlung gegeben. Ionisationskühlung funktioniert umso rascher, je größer die Strahlemittanz, je stärker die Fokusierung, je höher der Gradient der rf-Beschleunigung und je länger die Streulänge im Absorber ist. Am Ende jedes Kühlungsprozesses wird der kühlende Effekt von der Mehrfachstreuung aufgehoben und eine Gleichgewichtsemittanz erreicht. Am Ende des Kapitels wird der Kühlungskanal des CERN Szenarios beschrieben.

Kapitel 5 – Die Kühlkanal-Sensitivitätsstudie. In diesem Kapitel wird eine Studie zur Sensitivität des CERN Kühlkanals bezüglich der Variation verschiedener Strahlvariablen oder Kanalparameter präsentiert. Die Ergebnisse sind:

- Der CERN Kühlkanal operiert weit von der Gleichgewichtsemittanz entfernt und ist daher unempfindlich gegenüber Änderungen der Eingangsemittanz.
- Im Bedarfsfall könnte der CERN Kühlkanal mehr als doppelt so lang gemacht werden und würde noch immer funktionieren.
- Der Effekt des transversalen Kühlens ist unabhängig von longitudinalen Parametern wie Energie- oder Phasenunterschied. Werden diese

beiden Größen jedoch zu groß, so setzen Verluste ein, weil der Strahl nicht mehr transportiert werden kann.

- Die Wahl des Absorbermaterials ist nicht — wie bisher geglaubt — nur auf Wasserstoff beschränkt. Materialien wie Li, LiH oder Be liefern fast so gute Ergebnisse wie H, müssen jedoch nicht gekühlt werden.

Kapitel 6 – FASTCOOL. FASTCOOL ist der Vorschlag für ein Experiment zum Nachweis des Prinzips der ionisationskühlung. Es ist konstruiert, möglichst klein und einfach zu sein. FASTCOOL basiert auf folgenden Prinzipien: Mit einer hohen Meßgenauigkeit wird die Größe des benötigten Kühleffektes reduziert. Durch Einzelteilchenmessung bei der Emittanzbestimmung wird die Präzision weiter erhöht. Die Beschränkung auf zwei Dimensionen (x und x') sowie der Verwendung nur eines halben Strahles reduzieren weiter die Größe. Schließlich verwendet FASTCOOL starke Fokusierung statt einer großen Strahlemittanz und ist damit zu bestehenden Myonenquellen kompatibel.

Chapter 1

Introduction

This work relates to ionisation cooling in the context of a neutrino factory. Beam cooling is the reduction of the phase space volume of a charged particle beam. Ionisation cooling is a technique specially aimed a muon beams. In this thesis, it is treated in three ways: with analytic calculations, with single particle tracking simulations and with the proposal of FASTCOOL, a possible first experimental demonstration of ionisation cooling.

Chapter 2. About the neutrino, its discovery and its properties. Neutrino production, detection and neutrino oscillations. Open questions in neutrino physics and physics requirements for a new neutrino machine.

Chapter 3. The neutrino factory, the mechanisms behind it and CERN's baseline design.

Chapter 4. Ionisaton cooling. The theory of phase space reduction of a muon beam through energy loss and re-acceleration. Derivation of cooling rate and equilibrium emittance. The CERN 44/88 MHz cooling channel.

Chapter 5. The cooling channel sensitivity study. How the CERN cooling channel works and how it can be improved.

Chapter 6. The FASTCOOL experiment, a proposal for an easy to build, small first demonstration of ionisation cooling.

Chapter 7. Final conclusions.

Chapter 2

The Neutrino

2.0.1 History

After the discovery of radioactivity by accident by H. Becquerel in 1896 [6], radioactive decays and their "rays" (alpha, beta and gamma) were studied thoroughly. Radioactivity was understood as a two-body decay in which the initial nucleus I decayed into the final nucleus F plus the decay product, an electron e^- in the case of the β-decay.

$$I \to F + e^- \tag{1}$$

The problem with this interpretation was that it was not compatible with the observed energy spectrum of the electrons in the case of the beta decay. If I decays at rest into two particles, one would expect a discrete energy of the electron, as one had observed a discrete energy spectrum of the α particles. This can be calculated from the conservation of the momentum p and the total energy E_T, setting c=1:

$$\begin{aligned}
\text{before the decay: } E_\mathrm{T} &= m_0 \\
\vec{p} &= 0 \\
\text{after the decay: } E_\mathrm{T} &= E_1 + E_2 \\
\vec{p} &= 0 = \vec{p_1} + \vec{p_2} \\
\text{with } E_1^2 &= m_1^2 + p_1^2 \\
\text{we get } E_{1,2} &= \frac{E_\mathrm{T}^2 + m_1^2 - m_2^2}{2E_\mathrm{T}}
\end{aligned} \tag{2}$$

Here, m_0 is the mass of the pre-decay system, m_1 and m_2 are the masses of the post-decay nucleus and the electron and $E_{1,2}$ denote the kinetic energies

Figure 1: Wolfgang Pauli (1900-1958) postulated the neutrino in 1932 to explain the electron spectrum in the beta decay. It was discovered in 1956 by C. Cowan and F. Reines [5].

of the particles. Equation 2 gives one discrete energy for each particle. A two body decay cannot explain the continuous spectrum, and a third particle had not been observed. Nevertheless Wolfgang Pauli postulated in 1930 that there must be a third particle – as a last resort, and to "save" the laws of conservation. With three particles, the continuous energy spectrum can be explained. Pauli proposed the neutrino (which he initially called neutron) in the famous "letter to the radioactive Ladies and Gentlemen"[7] on December 4th, 1930:

> Dear Radioactive Ladies and Gentlemen:
>
> I beg you to receive graciously the bearer of this letter who will report to you in detail how I have hit on a desperate way to escape from the problems of the "wrong" statistics of the N and ^{6}Li and of the continuous β spectrum in order to save the "even-odd" rule of statistics and the law of conservation of energy. Namely the possibility that electrically neutral particles, which I would like to call neutrons might exist inside nuclei; these would have spin 1/2, would obey the exclusion principle, and would in addition differ from photons through the fact that they would not travel at the speed of light. The mass of the neutron ought to be about the same order of magnitude as the electron mass, and in any case could not be greater than 0.01 proton masses. The continuous β spectrum would then become understandable by assuming that in beta decay a neutron is always emitted along with the electron, in such a way that the sum of the energies of the neutron and electron is constant.

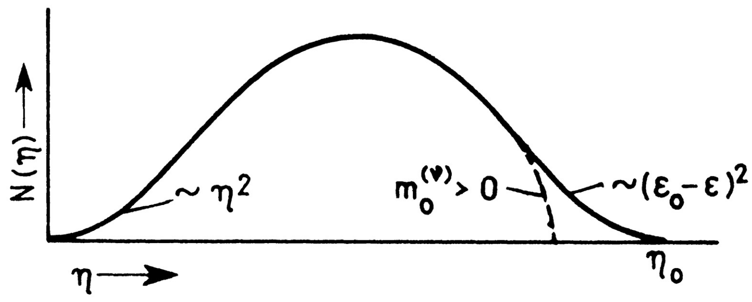

Figure 2: The continuous energy spectrum of the beta decay electrons, schematic illustration [8].

> ...
> I admit that my way out may look rather improbable at first since if the neutron existed it would have been seen long ago. But nothing ventured, noting gained. The gravity of the situation with the continuous β spectrum was illuminated by a remark by my distinguished predecessor in office, Mr. DeBye, who recently said to me in Brussels, "Oh, that's a problem like the new taxes; one had best not to think about it at all." So one ought to discuss seriously any way that may lead to salvation. Well, dear radioactive friends, weigh it and pass a sentence! Unfortunately, I cannot appear personally in Tübingen, for I cannot go away from Zürich on account of a ball which is held here on the night of December 6–7. With best regards to you and Mr. Baek,
>
> Your most obedient servant,
> W. Pauli

Pauli was remarkably right with some of his predictions: the neutrino has been found 26 years after Pauli wrote this letter and it is indeed chargeless and a fermion with spin 1/2. Yet Pauli was not right with his prediction of the neutrino's mass. Now that there is strong evidence that the neutrino has a mass, the upper limit of the mass of the electron neutrino Pauli refers to is five orders of magnitude smaller that the mass of the electron.

Further steps in the history of the neutrino have been:

- E. Fermi's quantitative theory of the β decay in 1934 [9], that could explain the continuous energy spectrum and predict the decay rate. It was Fermi who chose the Italian diminutive form "neutrino", as "neutron" had already been taken by the discovery of the neutron by J. Chadwick [10] two years earlier.

- The first evidence of neutrino existence by F. Reines and C. Cowan in 1956 [5], who detected neutrinos from the Savannah River nuclear reactor.
- The discovery of parity violation in the β decay by C. S. Wu et al. [11] in 1957.
- The presentation of a field theory of the weak interaction by R. Feynman and M. Gell-Mann in 1957 [12]. The big step forward of their theory was the description of the interaction via intermediate bosons as compared to the immediate point-like interaction in Fermi's theory.
- The discovery that the neutrino is left-handed by M. Goldhaber et al. [13] in 1958.
- The first measurement of solar neutrinos by R. Davis and D. Harmer in 1959 [14].
- The proposal of a muon neutrino by L. Pontecorvo in 1959 [15]. This lead to the proof of existence of at least two neutrino generations in 1962 by Steinberger et al. [16].
- The conclusion from LEP's precision measurement of the Z^0 decay in 1989 that there are only 3 light neutrino generations [17].
- The first report of strong evidence for neutrino oscillations in atmospheric neutrino data by the Super Kamiokande collaboration in June 1998 [18]. Although a neutrino deficit had been observed for some time, it was the Super Kamiokande experiment that showed with high statistical significance that this deficit depends on the path length and energy of the neutrino, as predicted by the theory.
- Upper limit of the ν_e mass found to be $mc^2 \leq 2.5\,\mathrm{eV}$ by Lobashev et al. in 1999 [19].
- Upper limits of the mass of ν_τ to be $mc^2 \leq 15.5\,\mathrm{MeV}$ established by LEP [20].
- And finally the discovery of the τ neutrino at the DONUT experiment at Fermilab in 2000 [21].

	mc^2	source
ν_e	$< 2.2\,\mathrm{eV}$	${}^3\mathrm{H} \rightarrow {}^3\mathrm{He} + e^- + \overline{\nu}_e$
ν_μ	$< 170\,\mathrm{keV}$	$\pi^+ \rightarrow \mu^+ + \nu_\mu$
ν_τ	$< 15.5\,\mathrm{MeV}$	$\tau \rightarrow 5\pi(\pi^0) + \nu_\tau$

Table 1: Upper limits of the neutrino masses. From [22].

2.1 The Neutrino

In the Standard Model, leptons are described by three left-handed flavour doublets

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix}_L \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}_L \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}_L$$

All these particles have already been observed (see previous section). The definition of flavour eigenstate is an experimental one. It is based on what we can "see" using a detector – this means it is based on the neutrino's interaction with matter (see Section 2.4). We have no access to what the neutrino really "is", we can only watch it interact. This superficial view leads to interesting phenomena such as neutrino oscillations. Neutrino oscillations are manifestations of the fact that what we "see" is not the whole picture of the particle, but only a projection of its wave function ψ on the flavour eigenstates. Neutrinos have a very low mass, as can be seen in Table 1.

The fact that there are only three light lepton generations has been verified by measurements of the width of the Z^0 boson peak at LEP [17]. Neutrinos only take part in the weak force.

2.2 Neutrino Production

Neutrinos are decay products. As the neutrino participates only in the weak force, the only way to produce a neutrino is through a weak decay. The most prominent processes in which neutrinos are produced are:

The weak decay of n and p in nuclei and of free n:

$$\begin{aligned} n &\rightarrow p + e^- + \overline{\nu_e} \\ p &\rightarrow n + e^+ + \nu_e \end{aligned} \tag{3}$$

The pion decay chain:

$$\begin{aligned}\pi^+ &\rightarrow \mu^+ + \nu_\mu \\ &\qquad \mu^+ \rightarrow e^+ + \overline{\nu}_\mu + \nu_e \\ \pi^- &\rightarrow \mu^- + \overline{\nu}_\mu \\ &\qquad \mu^- \rightarrow e^- + \nu_\mu + \overline{\nu}_e \end{aligned} \tag{4}$$

Neutrinos are furthermore produced in kaon, tauon and many other decays.

There are several processes in nature in which neutrinos are produced. Where not otherwise stated, the energy spectrum of the neutrinos is continuous, because they originate from a decay into three particles:

- Nuclear fusion reactions in stars. The fusion chain in stars starts with the weak reaction $p + p \rightarrow d + e^+ + \nu_e$. Other sources of neutrinos in stars are the less frequent reactions ^{7}Be+$e^- \rightarrow$ ^{7}Li+ν_e and ^{8}B$\rightarrow$ ^{8}Be+$e^+ + \nu_e$. These so-called "solar neutrinos" are always ν_e. They have kinetic energies of up to 10 MeV with discrete lines from the two-particle ^{7}Be decay at 0.5 and 0.9 MeV (see Figure 3).
  A special class of the solar neutrinos are neutrinos from super novae. They originate from $\nu\overline{\nu}$ pair production and appear in a burst from a super nova. They have energies in the range from a few MeV to several TeV.

- Nuclear power plants. Although not a natural source, nuclear power plants nowadays represent an important source of electron anti-neutrinos, which are produced in the beta decay of the fission products. They have energies in the few MeV range, up to 100 MeV. An example for the neutrino spectrum produced by a nuclear power plant is shown in Figure 4.

- Atmosphere. High-energetic cosmic rays hit nuclei in the upper atmosphere and produce pions (see Section 3.2.1 for a description of the pion production mechanism). They decay via a chain (see Equation 4) into muons and electrons. The total decay chain gives 2 ν_μ and 1 ν_e per pion. These neutrinos are in the MeV to GeV energy range.

A neutrino factory produces neutrinos similar to the way atmospheric neutrinos are produced, but with two differences. The muons are accelerated

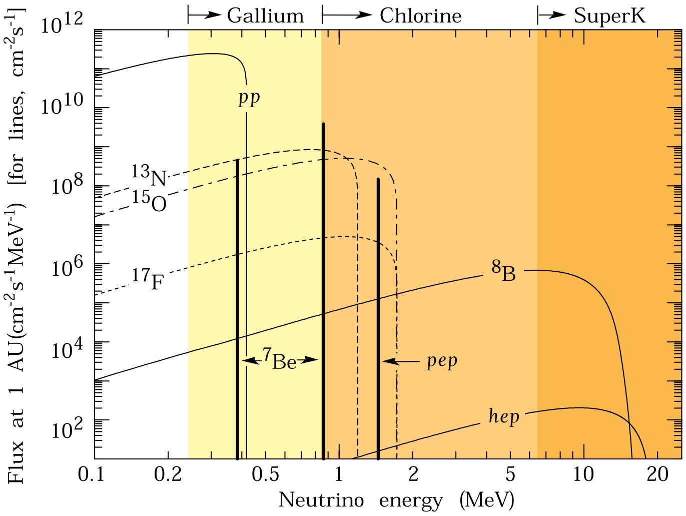

Figure 3: The solar neutrino spectrum. The color ranges show the sensitivity of the different detection techniques.

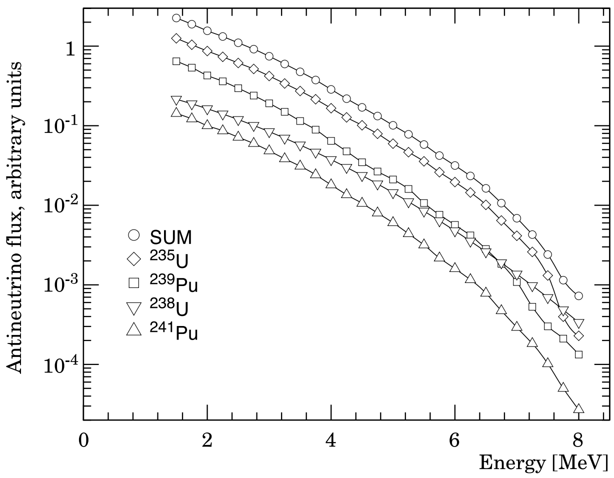

Figure 4: The neutrino spectrum seen by the CHOOZ experiment, which is located 1025m away from the 4.2 MW nuclear power plant of Chooz [23].

so that the neutrinos from the muon decay have on average a much higher energy than the neutrinos from the pion decay and the neutrino factory is sign selective, delivering only μ^+ or μ^- and their neutrinos at a given time.

As the pion has spin 0, its decay is isotropic ($d\sigma/d\Omega = 1/4\pi$). The muon decay, however, is a function of the muon's polarisation, with the directional neutrino spectrum being a function of the direction of the muon's momentum and spin.

2.3 Neutrino Oscillations

2.3.1 The solar Neutrino Problem

Our sun is a large thermonuclear reactor in which protons burn into helium and then heavier elements. Huge amounts of energy and of neutrinos are produced. The basic mechanism of the sun is that a gas contracts under the gravitational force, heating up to the point where nuclear reactions are ignited. As the cross-sections for all of these reactions rise sharply with the temperature, an equilibrium between thermal pressure and gravitation is established.

The standard solar model [25] describes the dependencies between nuclear physics, gravitation, thermo- and fluidodynamics of the sun. It is well tested and has recently been verified by helioseismology.

It is based on the assumptions that (a) the sun is in hydrostatic and thermal equilibrium, (b) its energy is produced by fusion and that (c) the energy transport inside the sun is dominated by radiation. This model predicts a neutrino flux of $4 \times 10^{10}\, \nu_e/(\mathrm{s\,cm}^2)$ on earth. Only half of this value is measured on the earth – a figure that is consistent for all solar neutrino experiments as shown in Figure 5. This difference between the measured and predicted flux of ν_e from the sun is called "solar neutrino problem".

The deficit of solar neutrinos has been a problem for neutrino physics for over 30 years. Now that the standard solar model is well established, and with evidence of neutrino oscillation from atmospheric neutrinos, the most probable solution for the solar neutrino problem are neutrino oscillations. This means that the missing ν_e oscillate into another neutrino flavour between the sun and the earth. Theory predicts that the oscillation probability depends on the distance sun-earth and the neutrino energy, and indeed, a

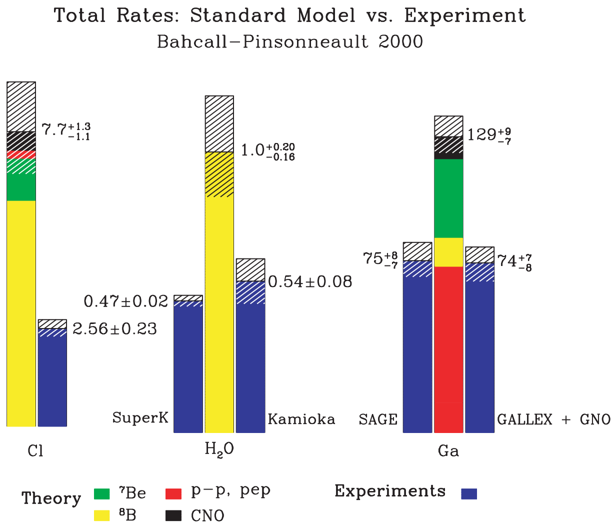

Figure 5: The solar neutrino deficit as seen with various detectors [24].

seasonal dependence in the number of observed ν_e has been found.

2.3.2 Oscillation Formalism

Neutrino oscillations can be explained by the difference between the mass eigenstates and the flavour eigenstates, like in the quark sector in the standard model. Whenever a neutrino is created or detected, this is due to the weak force, so its weak (or flavour) eigenstate determines the process. Its propagation, on the other hand, is described by the Schrödinger equation, which has different eigenstates that are called mass eigenstates. So, a neutrino is "born" in one flavour eigenstate. As soon as it starts propagating, the flavour eigenstate has to be projected on all three mass eigenstates. The superposition of the mass eigenstates is expressed using the mixing matrix U. The index l denotes the lepton flavour eigenstate and m denotes the mass eigenstate. In the standard model, the number of mass eigenstates N is 3. In the following derivation, $c = 1$:

$$|\nu_l\rangle = \sum_{m=1}^{N} U_{l,m}|\nu_m\rangle. \tag{5}$$

At the beginning of its journey, the neutrino ν_l can be described as a certain superposition of the three mass eigenstates:

$$\psi_l(x, t = 0) = \sum_m U_{l,m} \nu_m e^{ip_\nu x}. \tag{6}$$

These three eigenstates propagate with different wavelengths, because of their different masses. Thus, the various components of ν_l get out of phase and do not add up to ν_l anymore:

$$\psi_l(x, t) = \sum_m U_{l,m} \nu_m e^{ip_\nu x} e^{-iE_m t}$$

$$\text{with } E_m = E(\nu_m) = \sqrt{p_\nu^2 + M_m^2}. \tag{7}$$

As the neutrino mass is very small, we can approximate with $M_m \ll p_\nu$ that the neutrino velocity is c. If a neutrino is born at $t = 0$ and $x = 0$, then a time t, the distance is approximately $x = t$ (with $c = 1$). So we can write $\psi(x, x)$ instead of $\psi(x, t)$. With this we arrive at:

$$\psi_l(x, x) \simeq \sum_m U_{l,m} \nu_m e^{-i[M_m^2/2p_\nu]x}. \tag{8}$$

When the neutrino is detected, its state ν_m is projected on one flavour eigenstate ν_l', which can be different to the original one. The neutrino has oscillated. This projection can be expressed as (see Equation 6):

$$|\nu_m\rangle = \sum_{l'}^{N} U_{l',m}^* |\nu_{l'}\rangle \tag{9}$$

Hence the wave function of the neutrino ν_l is:

$$\psi_l(x, x) = \sum_{l'} \Big[\sum_m U_{lm} e^{-i[M_m^2/2p_\nu]x} U_{l'm}^* \Big] \nu_{l'} \tag{10}$$

In this way the wave function ψ_l of the neutrino ν_l can be expressed as a combination of the weak eigenstates $\nu_{l'}$. To calculate the probability $P(l \to l', x)$, that ν_l oscillates into $\nu_{l'}$, it is necessary to calculate $\psi\psi^\dagger$ and find the coefficient of $\nu_{l'}$. We get:

$$\begin{aligned} P(l \to l', x) &= \Big[\sum_{m'} U_{lm'}^* e^{iM_{m'}^2 x/2p_\nu} U_{l'm'}^* \Big] \cdot \Big[\sum_m U_{lm} e^{-iM_m^2 x/2p_\nu} U_{l'm}^* \Big] \\ &= \sum_m |U_{lm}|^2 |U_{l'm}|^2 \\ &\quad + \sum_{m' \neq m} Re(U_{lm} U_{lm'}^* U_{l'm'} U_{l'm}^*) \cos^2 \Big(\frac{M_m^2 - M_{m'}^2}{2p_\nu} x \Big) \\ &\quad + \sum_{m' \neq m} Im(U_{lm} U_{lm'}^* U_{l'm'} U_{l'm}^*) \sin^2 \Big(\frac{M_m^2 - M_{m'}^2}{2p_\nu} x \Big). \end{aligned} \tag{11}$$

The first two terms give the pure oscillation and the last term (with the imaginary part of the mixing matrix U) comes from the CP-violation. With the oscillation length $L_{mm'}$,

$$L_{mm'} = 2\pi \frac{2p_\nu}{M_m^2 - M_{m'}^2} = 2\pi \frac{2p_\nu}{\Delta M_{mm'}^2}. \tag{12}$$

we can write the oscillation probability in the case of CP-conservation (which means that U is real) as

$$P(l \to l', x) = \sum_m U_{lm}^2 U_{l'm}^2 + \sum_{m' \neq m} U_{lm} U_{lm'} U_{l'm'} U_{l'm} \cos^2\Big(\frac{2\pi x}{L_{mm'}}\Big). \tag{13}$$

The oscillation length $L_{mm'}$ is one of the most important input parameters for the design of a neutrino factory-detector ensemble. The oscillation is maximum for distances that are integer multiples of $L_{mm'}$. As the beam diverges, it is clearly best to choose $x = L_{mm'}$. The problem in choosing this distance is the fact that the mixing matrix U is not very well known. This leads to a circular problem: to better explore the mixing matrix, one needs to build a neutrino factory. To optimise the neutrino factory, one needs to know U.

2.3.3 Three Family Oscillations

In the case of three neutrino flavours, the mixing matrix U is a 3×3 orthogonal matrix.

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \tag{14}$$

U can be written as:

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13} & c_{12}c_{23} - s_{12}s_{23}s_{13} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13} & -c_{12}s_{23} - s_{12}c_{23}s_{13} & c_{23}c_{13} \end{pmatrix} \tag{15}$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$.

Equation 13 can be written explicitly as:

$$\begin{aligned} P(\nu_l \to \nu_l', x) = \quad & -4U_{l2}U_{l3}U_{l'2}U_{l'3}\sin^2\Delta_{32} \\ & -4U_{l1}U_{l3}U_{l'1}U_{l'3}\sin^2\Delta_{31} \\ & -4U_{l1}U_{l2}U_{l'1}U_{l'2}\sin^2\Delta_{21} \end{aligned} \tag{16}$$

where

$$\Delta_{ij} = \frac{1.27\delta m_{ij}^2\, x}{E}$$

and $\Delta m_{ij}^2 = (m_i^2 - m_j^2)$, having expressed the length in km, the mass difference in eV^2 and the energy in GeV. Here, the importance of the factor x/E (often referred as L/E) can be seen. The oscillation modulates with x/E; the higher the energy, the wider the oscillation pattern.

Using the Large Mixing Angle (LMA) solution with mass differences of $\Delta m_{12}^2 = 5\cdot 10^{-5}\,\mathrm{eV}^2, \Delta m_{23}^2 = 3.2\cdot 10^{-3}\ \mathrm{eV}^2$ and $\Delta m_{13}^2 = 3.25\cdot 10^{-3}\,\mathrm{eV}^2$, one can neglect the terms containing Δm_{12}^2 and write the oscillation probabilities for the various cases.

Survival of solar neutrinos:

$$P(\nu_e \to \nu_e, x) = 1 - \cos^4 2\vartheta_{13} \sin^2 \vartheta_{12} \sin^2 \frac{\Delta m_{13}^2 x}{4E} - \sin^4 \vartheta_{13} \tag{17}$$

Oscillation of atmospheric neutrinos:

$$P(\nu_\mu \to \nu_\tau, x) = \sin^2 2\vartheta_{23} \cos^2 \vartheta_{13} \sin^2 \frac{\Delta m_{23}^2 x}{4E} \tag{18}$$

Signal for a neutrino factory "wrong sign muons":

$$P(\nu_e \to \nu_\mu, x) = \sin^2 2\vartheta_{13} \sin^2 \vartheta_{23} \sin^2 \frac{\Delta m_{23}^2 x}{4E} \tag{19}$$

2.4 Neutrino Detection

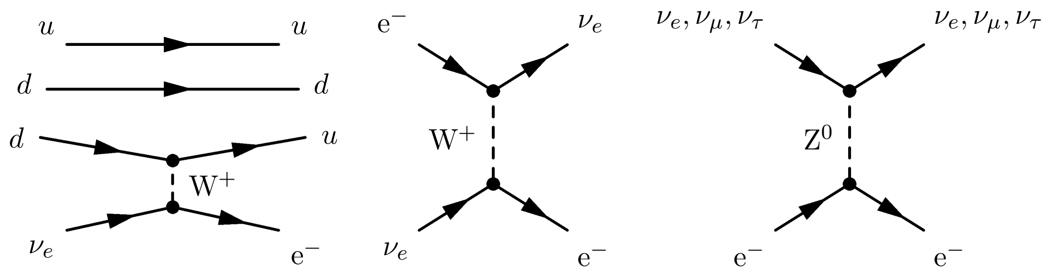

Figure 6: Feynman graphs for neutrino detection. On the left the inverse β-decay ($n+\nu_e \to p+e^-$), in the middle the charged current and on the right the neutral current reactions. In all cases a fraction of the momentum p_ν of the original neutrino is transferred to an electron. The electron's momentum $\vec{p}_e$ is a lower limit for $\vec{p}_\nu$.

Neutrinos can be detected through two different processes: the charged and the neutral current, with the inverse β-decay as the best-known example for the charged current.

The inverse beta decay is the inverse process of the neutrino production. A neutron n and a neutrino react to a proton and an electron. This reaction is endothermic, so the energy Q is needed to start it. This means that only neutrinos with a kinetic energy of at least Q can start this reaction – neutrinos with a lower kinetic energy cannot be detected.

$$n + \nu_e \rightarrow p + e^- - Q \tag{20}$$

The result is a beta active nucleus of a different element. An example is the C_2 Cl_4 detector in the Homestake mine in California, USA, which was the first detector for solar neutrinos. The chlorine in the detector can react as

$$^{37}\mathrm{Cl} + \nu_e \rightarrow \ ^{37}\mathrm{Ar} + e^- - 0.86\,\mathrm{MeV} \tag{21}$$

This reaction has been chosen because of its low Q value and because Ar is gaseous and thus easily to separate from the liquid C_2Cl_4. With this method it is possible to measure the number of neutrinos with $E_{\mathrm{kin}} > 0.86\,\mathrm{MeV}$, but no information about their direction or momentum is provided. This information is carried away by the electron, which is not detected with this kind of detector.

The other two neutrino detection methods, charged and neutral currents on electrons, provide more information about the neutrino's direction, because the electron, that carries a fraction of the neutrino's momentum, is actually detected. In the case of the Kamiokande detector this is done with a Cerenkov detector. From the direction and the momentum of the electron one can draw conclusions about the incident neutrino properties. The Feynman graphs for these two reactions are shown in Figure 6.

As the neutrino only interacts via the weak force, all detection techniques suffer from low cross-sections.

2.5 Open Questions in Neutrino Physics

Today, there are four main open questions in neutrino physics, three of which a neutrino factory could answer.

- The size of ϑ_{13}. This angle is believed to be very small. In the atmospheric oscillation $\nu_\mu \rightarrow \nu_\tau$ and in the solar oscillation $\nu_e \rightarrow \nu_x$, ϑ_{13} appears always in cos terms. Reactor neutrino experiments such as

CHOOZ [23] have given only an upper limit for ϑ_{13}. A neutrino factory would deliver enough statistics to measure even very small values of ϑ_{13}.

- The sign of Δm^2_{23}. This mass difference occurs in all oscillation probabilities only in $\sin^2$ terms. That is why from the observation of oscillation alone, it is not possible to tell the mass hierarchy of the neutrinos. Profiting from the matter effect, a neutrino factory makes it possible to measure the sign of Δm^2_{23}.

- CP violation. CP violation is manifested in a difference in the oscillation probabilities $P(\nu_1 \to \nu_2) \neq P(\overline{\nu_1} \to \overline{\nu_2})$. So far, there has been no evidence of CP violation in the leptonic sector. With a two charge approach of μ^+ and μ^- it is possible to prove the existence of CP violation in neutrinos and to measure the CP violation phase to about 10 degrees.

- Only the last question, whether the neutrino is a Majorana or a Dirac particle cannot be answered with the help of a neutrino factory. Here, physicists hope to find neutrino-less double β decays as a proof that the neutrino is a Majorana particle.

2.6 Physics Requirements for a Neutrino Machine

To answer the above questions, a dedicated neutrino beam that points to one or two detectors at a distance of about one oscillation length is needed. In order to find the answers within a reasonable amount of time ($\sim$ 5 years), such a "neutrino factory" has to have the following properties:

High intensity. As the cross-section for interaction with matter is extremely low, the only way to obtain a sufficient number of events is to produce a large number of neutrinos. The target figure is more than 10^{20} neutrinos/yr/detector which translates into 10^{21} muons/yr in the storage ring.

High energy. The cross section of the neutrino detection increases with E which makes higher-energy neutrinos easier to detect. In addition the plane opening angle of the muon decay – and with it the divergence of the neutrino beam – decreases inversely proportional to the relativistic factor $\beta\gamma$. For high energies, we can use the approximation $1/\beta\gamma \propto 1/E$ and thus the rate of neutrinos that hit the detector per solid angle scales proportionally to E^2. There is, however, a limit for the maximum desirable neutrino energy that comes from the L/E law for oscillations (see Equation 12). With the maximum baseline distance given as the earth's diameter, we get $E \leq 200\,\mathrm{GeV}$.

Knowledge of the beam. There are two ways to observe an oscillation $\nu_1 \to \nu_2$. One is to observe the appearance of ν_2, the other one is to observe the disappearance of ν_1. As the oscillation probabilities fall in the few percent range, it is necessary to know the flux of ν_1 better than that to observe disappearance. Furthermore, a better knowledge of the beam also increases the resolution for the more precise appearance measurements.

Two charges. To observe CP violation ($P(\nu_1 \to \nu_2) \neq P(\overline{\nu_1} \to \overline{\nu_2})$), it is necessary to have beams of ν_1 and $\overline{\nu_1}$ which corresponds to both charges of the mother particle.

Two flavours. A second flavour allows a cross-checking of the oscillation and CP-violation parameters and to decrease systematic errors. Using the CPT–theorem that $P(\nu_1 \to \nu_2) = P(\overline{\nu_2} \to \overline{\nu_1})$, a new expression for the CP violation can be derived: $[P(\nu_1 \to \nu_2) = P(\overline{\nu_2} \to \overline{\nu_1})] \neq [P(\overline{\nu_1} \to \overline{\nu_2}) = P(\nu_2 \to \nu_1)]$. Two neutrino flavours can be produced with neutrinos from the μ decay. The matter effect of the transport of the electron neutrino can be used to measure the sign of Δm_{23}.

Pure beam. A beam from a neutrino factory has to be very pure, which means that only muons of one charge should produce neutrinos at a time. If there was a background of muons of the other sign, it would not be possible to distinguish oscillated neutrinos from the beam background. In order not to degrade the precision on the oscillation parameters, the purity must be much higher than the oscillation probability.

Divergence. In order to keep the number of neutrinos that actually hit the detector as high as possible, the neutrino beam's divergence should be as small as possible. There are two components in the divergence: the *rms* opening angle from the muon decay, which scales with $1/\gamma$, and the divergence of the muon beam itself. It is the second one that should be kept small in the decay ring and can be influenced by previous beam cooling.

Chapter 3

The Neutrino Factory

3.1 Neutrino Factory Layout

To satisfy the desires listed in Section 2.6, a totally new type of particle accelerator complex has to be conceived: the neutrino factory [2]. The challenge is to produce a large number of high energy neutrinos of two flavours in a well-known, well collimated beam. As discussed in Section 2.2, neutrinos are decay products. The concept of a neutrino factory is to mass-produce particles that decay into neutrinos. The typical Q-value of such a reaction is in the range of 1–100 MeV, which is below the required neutrino energies of tens of GeVs. As neutrinos are neutrally charged, they cannot be accelerated. The solution is to add energy to the pre-decay system.

The choice of the pre-decay particle is determined by its production mechanism, its life time and the requirement of two neutrino flavours. As a neutrino factory aims at producing 10^{21} neutrinos per year, the mother particle must be easy to produce. Its lifetime must be long enough to survive

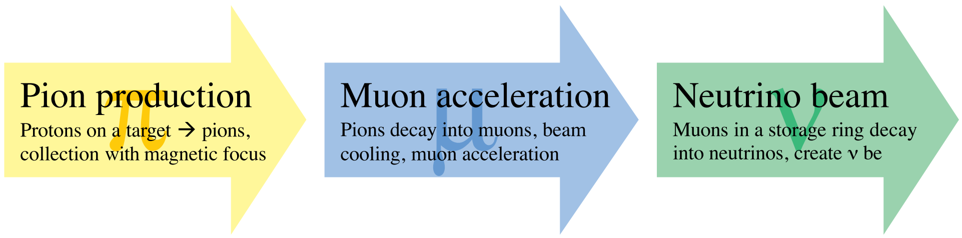

Figure 7: The general principle of a neutrino factory: Mass-production of pions that decay into muons. The muons are accelerated and injected in a decay ring, where they decay into neutrinos, producing a well-defined neutrino beam.

acceleration, but short enough so as not to block the machine. There is only one neutrino production mechanism that meets all these requirements: muons from the pion decay chain:

$$\begin{aligned} \pi^+ &\rightarrow \mu^+ + \nu_\mu \\ & \qquad \mu^+ \longrightarrow e^+ + \overline{\nu}_\mu + \nu_e \end{aligned}$$

This decay chain already gives the rough structure of a neutrino factory (see Figure 7): pion production, muon acceleration, neutrino beam. By switching from π^+ to π^-, one can produce ν_μ and $\overline{\nu}_e$ instead of $\overline{\nu}_\mu$ and ν_e.

3.2 Neutrino Factory Physics

3.2.1 Pions

The pion $\pi^\pm$ is the lightest charged meson. Its mass is 139.57 MeV, its mean life time is 26 ns [26]. A pion is produced via a Δ resonance when two nucleons collide. The best way to produce pions is to send a proton beam onto the target. The minimum proton energy to produce a pion on a target is 600 MeV. A plot of the production efficiency is given in Figure 8. The lower two curves in this plot show the pion yield divided by the proton energy. This line is nearly flat, which means that the number of pions produced depends only on the beam power and not on the beam energy. This leads to the fact that the energy distribution of the produced pions does not depend on the proton beam energy, the maximum of the pions produced have a kinetic energy of about 300 MeV. This can be explained by re-interaction. A proton with a low energy ($E \approx$ 1–2 GeV) produces one pion and looses most of its kinetic energy in the process. A high energy proton ($E > 5$ GeV) produces a first pion loosing some energy, then re-interacts with the target and produces another pion and so forth.

In order to achieve the desired 10^{21} neutrinos per year, a mean beam power of 4 MW of protons during 10^7s/yr is needed. This is equal to 10^{16} protons/s at 2 GeV or 4×10^{14} protons/s at 50 GeV.

Target. The target material and geometry influence the pion production yield. Currently, the pion production cross-section is only known to 25% [27]. To increase the precision on the cross section and to help choose the right target material, the HARP experiment has been started [28].

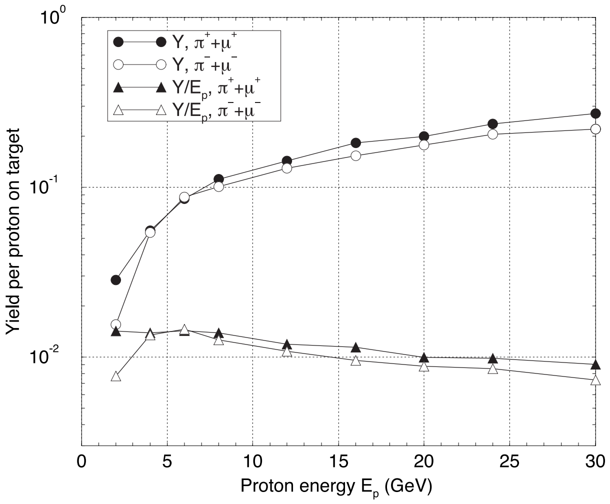

Figure 8: Pion yield of an 80 cm long, 1.5 cm diameter ^{12}C target as a function of the energy of the incident proton. The upper two lines show the π^+ and π^- yields per proton as a function of the proton energy, the lower two lines are normalised to the beam power. The pion yield per beam power does not depend strongly on the proton beam energy [27].

The pion yield is not the only argument for choosing a target material. The high beam power of 4 MW leads to a number of difficulties: up to 25% of the beam power heat the target and have to be cooled away. The protons pass through the target at nearly the speed of light and heat the target locally. The thermal expansion, which propagates at the speed of sound, cannot keep up with this process and so stress builds up in the target. Furthermore the protons activate the target, which makes disposal critical.

Transverse collection. The pion production is isotropic in the centre of mass system ($d\sigma/d\Omega = 1/4\pi$). In the lab frame, the pion distribution is boosted in the direction of the proton beam. The resulting pion distribution is far from being a beam. The average transverse momentum $p_{\rm t}$ of the pions is 220 MeV/c and far too big for a classical beam transport. The $p_{\rm t}$ has to be reduced or converted into longitudinal momentum p_z. There are two focusing devices that can perform this conversion: the magnetic horn, which was invented in 1960 by Simon van der Meer at CERN [29] and a tapered solenoid. This process is called "transverse collection".

A **magnetic horn** (Figure 9) provides a toroidal magnetic field ($B_\phi = I/2\pi r$) which is perpendicular to $p_{\rm t}$ and effectively rotates $p_{\rm t}$ into p_z. This ef-

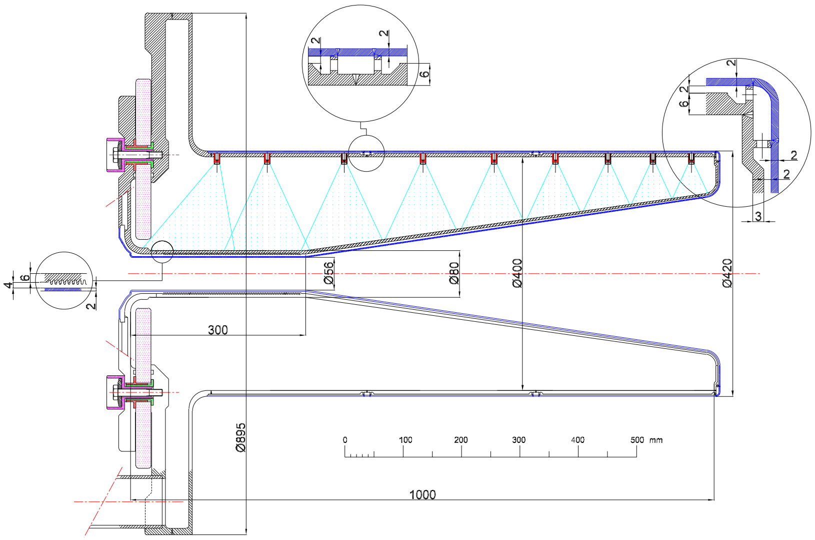

Figure 9: Example of a magnetic horn: the inner part of the proposed horn for the CERN neutrino factory. [30].

fect is sign-selective. One charge is focused while the other one is defocused. The horn itself is made of a single layer of conductor, mostly Aluminium. The magnetic filed is only inside the horn volume, while the area around the target is field-free. This means that particles fly away from the target, pass through the inner wall of the horn, get bent inside the horn and leave it through the inner or the front wall. The passage through the horn walls makes it necessary to make them as thin as possible. Unlike magnetic coils that normally have N windings and profit from the fact that the B field increases with the number of windings as $B = NI/2\pi r$, magnetic horns have only one "winding" and thus need to have extremely high currents in the range of hundreds of kA.

The **tapered solenoid** works completely different than the horn. It is based on the fact that the kinetic energy E and the total momentum p are conserved in a magnetic field. The transverse momentum p_{t} is reduced in an adiabatically decreasing B_z and p_z is increased to keep $p = \sqrt{p_{\mathrm{t}}^2 + p_z^2}$ constant.

$$\frac{p_{\mathrm{t}}^2}{B} = const \tag{22}$$

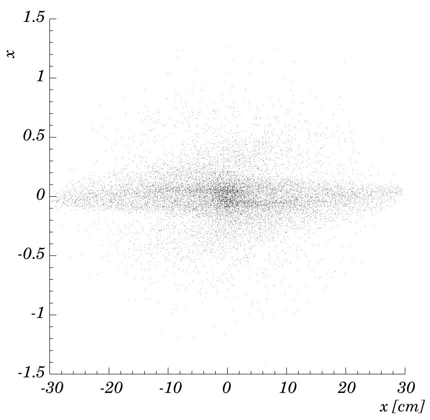

Figure 10: The distribution of pions produced by 2.2 GeV protons on a 30 cm liquid mercury target after the collection with the CERN horn [31].

During this process, the size of the beam increases with

$$\frac{r^2}{B} = const \tag{23}$$

The tapered solenoid is only limited by size. By reducing B_z from 20 T to 2 T, the transverse momentum p_{t} is reduced by a factor $\sqrt{10}$, while the beam radius increases by the same factor. In principle, it would be possible to reduce p_{t} further, but in the extreme one would have a parallel beam with $r = \infty$.

3.2.2 Muons

The result of the transverse collection is a pion beam with an extremely large normalised emittance, up to $2000\,\pi$ cm mrad *rms*. The pion half-life is 26 ns at rest which corresponds to a decay length of 7.8 m . After 30 m, more than 95% of the pions are decayed into muons while more than 95% of these muons ($\tau = 2.2\,\mu$s) are still alive. The ν_μ produced in the pion decay are of a comparatively low energy (<200 MeV) and are not used for physics. As the pion decay is isotropous, some of the pions decay towards the beam centre, decreasing their p_{t} while the same number increases its p_{t}. The result is an increase in *rms* p_{t} which translates into an emittance increase of up to 50%. As some pions decay in the forward direction and some backwards, the energy spread is also increased, but by a lesser fraction, as it is already very big.

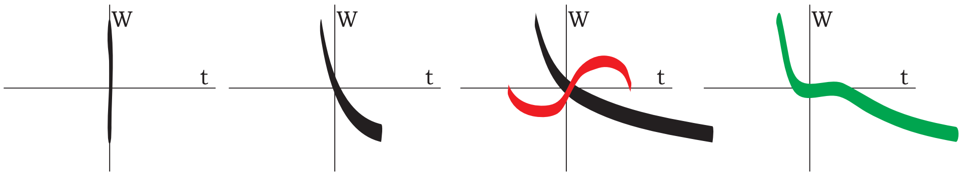

Figure 11: The principle of phase rotation: a non-relativistic beam drifts in order to build up a time-energy correlation. Rf cavities set at –90° decelerate early arriving fast particles and accelerate late arriving slow particles. The result is a flattened out energy distribution (green).

Phae rotation. The resulting muon beam has a large transverse emittance and a large energy spread. As mentioned above, the pions are produced around a peak energy of 300 MeV. The muons have similar energies and are not fully relativistic ($\beta \approx 0.9$). If the energy spread remained, bending the beam in the upstream accelerator and decay ring would be difficult. Furthermore, the particles of the beam travels with different speeds and the beam would get longer and longer. To reduce this energy spread, phase rotation has to be applied.

As particles with different energies travel at different speeds, an energy-time (or energy-space) correlation builds up. In a phase rotation section, rf cavities are set in the way that early arriving particles (which are faster) are decelerated, while particles that arrive later are accelerated. The result is a decreasing energy spread (see Figure 11). In the case of the neutrino factory, the energy spread of the muons is so large that it is impossible to completely remove it. The spectrum is only flattened out at the energy where the most particles are. Only these particles are of interest for the downstream machine.

Transverse cooling. The muon beam that comes out of the phase rotation is nearly mono-energetic, so transporting it will not increase its length. Yet, this beam has a transverse emittance of about 2500 π cm mrad, which is a factor of 10^6 to 10^8 larger than for example the emittances of the LEP beam [32] (horizontal: $3 \times 10^{-3} \pi$ cm mrad; vertical: $3 \times 10^{-5} \pi$ cm mrad). Transporting and accelerating such a beam is inconvenient and costly. Furthermore, it would not meet the requirement for a low divergence of the neutrino beam. The solution is to apply beam cooling to the muon beam at the earliest possible stage, which is after pion decay and phase rotation.

Due to the short half-life of the muon, this process must be very fast. The only known cooling technique that allows fast phase space reduction of a high-emittance muon beam is ionisation cooling. It is described in detail in Chapter 4.

Cooling adds a difficulty to the overall design of the neutrino factory, but it is worthwhile as it makes the downstream acceleration section and storage ring easier to build.

Acceleration. After the cooling, the muons can be accelerated in order to have enough kinetic energy in the pre-decay system. Again, this process is dominated by the short lifetime of the muons. A fast acceleration of muons seems only possible with recirculating linacs.

3.2.3 Neutrinos

The accelerated muons are stored in a decay ring. The decay of a high-energy muon produces 2 neutrinos that go on average in the same direction as the muon, with a maximum opening angle of $1/\gamma$. A muon storage ring is made of two long straight sections and the necessary bends to close the ring. All the muons that decay in one straight section form a neutrino beam. The idea is to point each straight section towards a detector. A detailed calculation of the efficiency of a decay ring and the relation between decay ring geometry and detector location is performed in Appendix A.

As the distances under consideration are fairly wide (700 km to 7000 km), the neutrino beam will have to travel through the earth. This is a unique feature of a neutrino beam, as it interacts only weakly with the matter of the earth and survives even 7000 km of travel virtually without losses.

3.3 The CERN Neutrino Factory Design

3.3.1 Time Structure

The base of the CERN design [33] is the time structure of the whole complex. The aim is to produce as many neutrinos per year as possible. There are three hierarchical levels of time structures in the neutrino factory:

Muon lifetime. The storage ring is filled with muons, they decay and ideally then the storage ring is filled immediately again. The muons in the

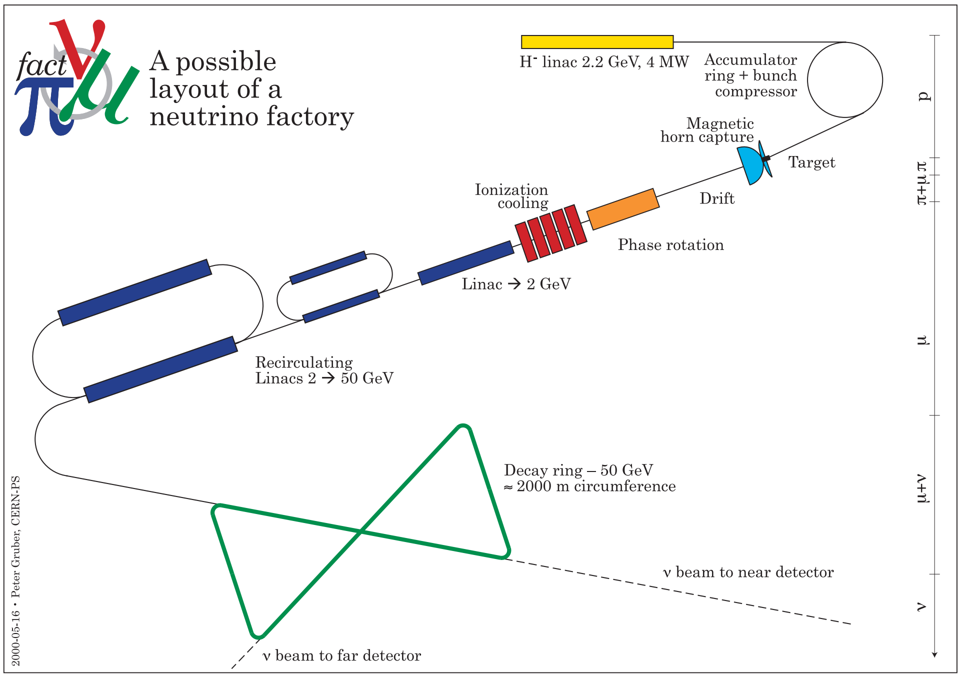

Figure 12: Schematic view of a neutrino factory. The particles in the various stages of the accelerator are shown on the line close to the right edge.

storage ring have an energy of 50 GeV ($\gamma = 473$) and a life time of 1.04 ms. After 5 (10) lifetimes, 99.3% (99.996%) of the muons have decayed and new beam could be injected. This implies a possible maximum repetition rate of 100–200 Hz for the whole system.

This repetition rate cannot be quite achieved due to other limits: the numerous rf cavities in phase rotation, cooling and acceleration have to be pulsed due to heat production and their power consumption. In fact they are only needed during the $\approx 3.3\,\mu$s passage of the muon beam. Leaving the cavities on all the time would lead to enormous power consumption. As most of the power in a nomal conducting cavity is transformed into heat, cooling the cavity would also be difficult. It has been established that the maximum feasable repetition rate for these components is 50 Hz (=20 ms).

Beam length. The second time structure comes from the length of the beam. The protons are almost instantaneously converted into pions and then muons, so in first approximation the muon beam is as long as the proton beam. All circular elements limit the beam length. These are: the accumu-

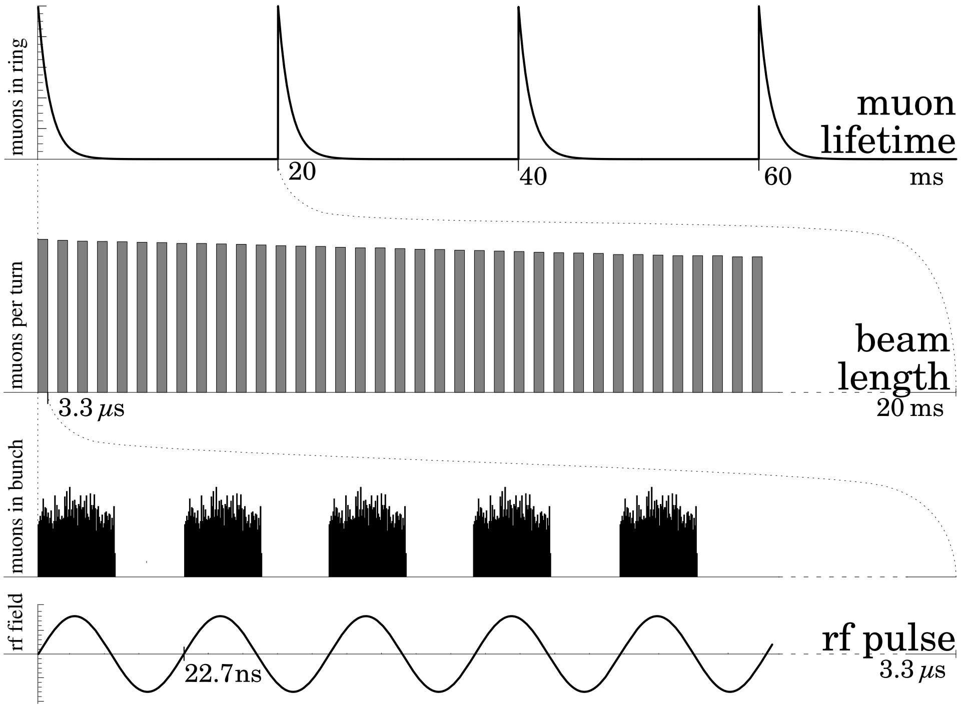

Figure 13: The time structure of the neutrino factory on the three levels as seen in the decay ring. ***Muon lifetime:*** *every 20 ms, a new muon beam is created that decays in the decay ring.* ***Beam length:*** *A 1 km (3.3 μs) long beam cycles in the 2 km long storage ring. At every turn the number of muons decreases with* $\exp(t/\gamma\tau)$. ***Rf pulse:*** *140 mini bunches of 22.7 ns fit in the 1 km long beam.*

lator ring (1 km circumference) and the decay ring (2 km circumference). So the maximum length of the beam is 1 km or 3.3 μs.

Rf pulse length. The smallest time structure of the neutrino factory is the period of the 44 MHz cavities of the phase rotation and the first part of the cooling channel, which corresponds to 22.7 ns. The CERN scheme is based on keeping mini bunches within the rf bucket during the whole chain from the proton driver to the decay ring to avoid re-bunching. The spacing between the mini-bunches from the target has to be a multiple of this period and the proton bunches should be as short as possible. The CERN scheme is based on filling every rf bucket. With a beam length of 3.3 μs there is space for 140 mini-bunches.

Mean beam power	4 MW
Kinetic energy	2.2 GeV
Repetition rate	50 Hz
Pulse duration	2.2 ms
Number of protons per pulse (per second)	$2.2\times10^{14}(1.1\times10^{16})$
Mean current during a pulse	13 mA
Overall length	799 m
Bunch frequency	352.2 MHz

Table 2: Parameters of the superconducting proton linac (SPL) [34].

3.3.2 The Superconducting Proton Linac (SPL)

The proton driver of the CERN design is a superconducting linear accelerator that produces 2.2 GeV H^- ions [34]. The advantages of a linac are the high current and the possibility for high repetition rates. The SPL can deliver H^- ions at the required 50 Hz and could be run in principle at even higher repetition rates.

The output of the SPL cannot, however, be directly used to produce pions. The reason is the low peak current. That is why the accelerated H^- ions have to be collected over a 2.2 ms period in an accumulator ring. This is done by charge exchange injection: the negatively charged ions are injected into the accumulator where they are stripped to protons. The protons that are already in the accumulator ring are not affected by this process. Through phase-space painting, the whole circumference of 1 km of the accumulator ring is filled with protons.

After the accumulation, the bunch structure has to be adapted to the downstream parts. To make phase rotation possible, the single bunches must be as short as possible ($\approx$ 3 ns). To match to the 44 MHz structure, the bunch spacing must be 22.7 ns. The short bunches are produced in a compressor ring of 1 km circumference. This is done by inverse phase rotation. It is intended to place both rings in the old ISR tunnel (Figure 14), while the rest of the neutrino factory could fit on the Prevessin site (Figure 15), with a transfer line for the proton beam.

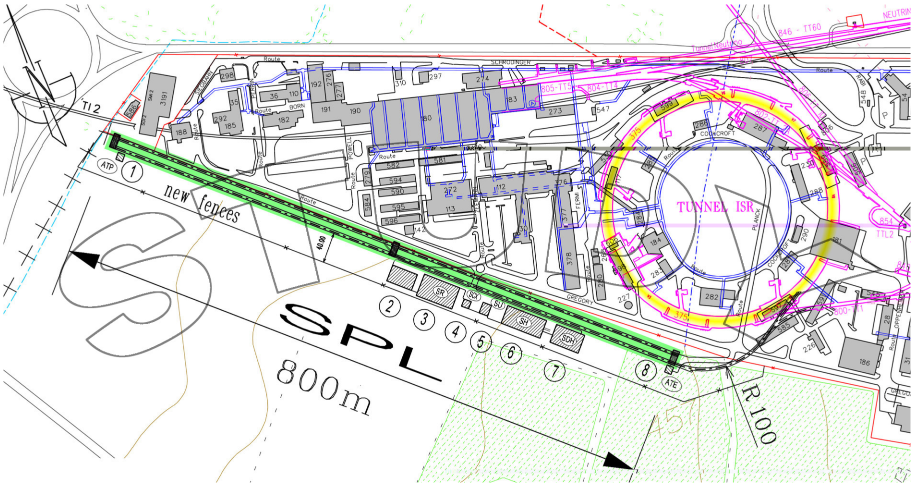

Figure 14: A possible location for the superconducting proton linac (green) and the accumulator/compressor rings (yellow) at the CERN Meyrin site [35].

3.3.3 Target

The proposed target for the CERN scheme is a liquid metal jet. Liquid mercury is pressed with 25 m/s through a nozzle in order to form a 1.5 cm diameter jet that serves as target. The proton beam deposits on average 14 J/g in the mercury [36]. This leads to a temperature rise of 80 degrees, which causes rapid thermal expansion and rupture of the jet, which is destroyed at every passage of a proton beam (every 20 ms).

A liquid metal jet has been chosen exactly for this feature: unlike a solid target that remains in the target station for a long period of time and that has to withstand 5×10^8 blasts/yr, the liquid target is formed and destroyed at every passage of the beam. The heat load is carried away with the material and can be cooled at a more convenient place away from the interaction point. The high speed of the mercury jet is directly related to the destruction rate. For an optimum pion yield, interaction of the primary beam and re-absorption of the secondary pions have to be balanced. For Hg, the optimum target length has been found to be 30 cm, which is two interaction lengths. The target has to be replaced at 50 Hz, which gives a minimum speed of 15 m/s. To allow some time to stabilise the jet, a speed of 25 m/s has been chosen.

Another advantage of liquid targets is their easy refine-ability, which makes it possible to separate activation products from the target material.

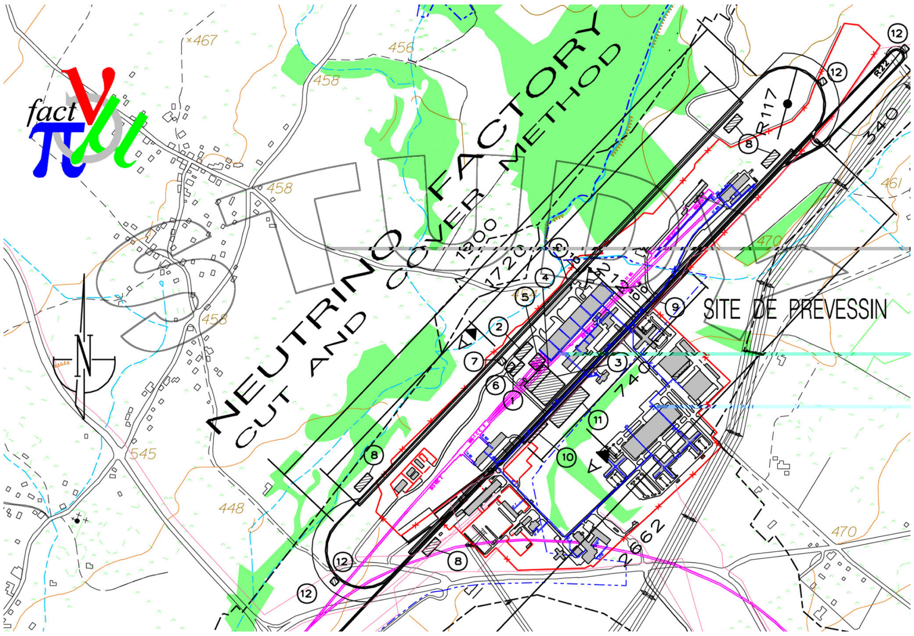

Figure 15: Possible locations of the neutrino factory front end in the CERN Prevessin site [35].

Mercury has been chosen because of its good pion yield, its low melting point and its high heat capacity. An initial test of a mercury jet in a proton beam has been performed at BNL, where the rupture of a prototype Hg–jet during the passage of the proton beam has been observed with a high speed camera [37].

3.3.4 Transverse Collection

The pions that are produced in the target have a wide angular distribution. To transport them, they have to be focused. In the CERN scheme, it has been chosen to use a magnetic horn for this task. The CERN horn is a variation of the classic idea of the magnetic horn, as it is made of two independent magnetic volumes. This increases the total p_t acceptance. The inner volume focuses the particles with lower p_t, while the outer volume those with medium to higher p_t. Particles with $p_\mathrm{t} > 80\,\mathrm{MeV}$ are lost.

The horn conductor is made of Al, with a "sandwich" construction of one layer of Al (or Ti), free space for water cooling and a second layer of Al. Figure 9 shows a sketch of the horn. The specifications are listed in Table

	inner horn	outer horn
inner radius	4 cm	20 cm
outer radius	20 cm	40 cm
length	1 m	
material	Al	
average thickness	1 cm	
current	300 kA	600 kA

Table 3: Specifications of the magnetic horn for the CERN neutrino factory scheme [38]

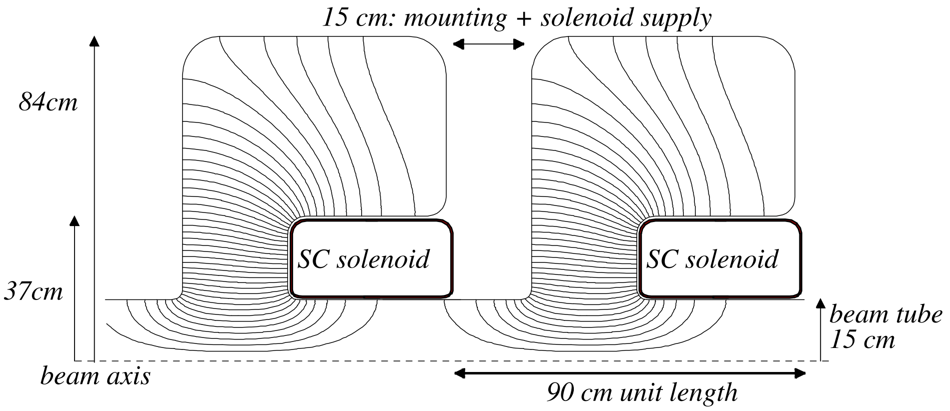

Figure 16: The CERN 88 MHz cavity with built-in solenoids for the second part of the cooling channel [39].

3.

3.3.5 Pion Decay and Phase Rotation

The CERN scheme is based on solenoidal transport of the large emittance beam. Immediately after the magnetic horn is the starting point of a solenoidal decay channel for the pions. This is a 30 m long transport channel made of one continuous, 60 cm diameter, 1.8 T solenoid. At the beginning of the decay channel, the 2D transverse $(x - y)$ emittances of the pions are 1740–1690 π cm mrad rms. The pion decay increases this emittance, because some of the released energy increases the transverse momentum. At the end of the decay channel, the 2D transverse emittances have increased by 40% to 2380–2440 π cm mrad rms [41].

cavity		
frequency	44 MHz	88 MHz
length	140 cm	90 cm
diameter	240 cm	168 cm
bore radius	30 cm	15 cm
rise time	318 μs	156 μs
gradient	2 MV/m	4 MV/m
kilpatrick	2.5	2.3
P_{max}	1.57 MW	2.19 MW
P_{mean}	53 kW	56 kW
material	normal conducting Cu	
solenoid		
length	88 cm	40 cm
inner radius	30 cm	15 cm
outer radius	52 cm	37 cm
B_z	1.8 T	5 T
material	superconducting	

Table 4: Properties of the 44 and 88 MHz cavities with built-in solenoids. Data shown per cavity [39].

During this 30 m of drift, an energy-time correlation builds up in the beam (see Figure 17). This correlation is used for phase rotation (see Section 3.2.2). The phase rotation channel is 30 m long and is made of 44 MHz normal conducting cavities with integrated solenoids. An example for a cavity with integrated solenoid is shown in Figure 16. The technical data of the cavities are summed up in Table 4. The cavities have an accelerating gradient of 2 MV/m and are set to –90°. As they are normal conducting, they consume a peak power of 1.57 MW. It is difficult to both provide and cool away such an amount of power. That is why the cavities have to be pulsed at a maximum rate of 50 Hz. In fact, these cavities are one of the limiting factors for the repetition rate of the neutrino factory. The maximum de- and acceleration voltage is 60 MV. Thus, the energy difference can be reduced by up to 120 MeV. The energy spread of the muons produced is much higher: there are all energies between a few MeV and the maximum proton energy. The phase rotation part cannot completely remove this energy spread, but "flatten out" a certain part of the energy-phase diagram.

In order to achieve the best intensity, it has been chosen to optimize the phase rotation for the energy range with the highest particle density, which is $E_{\mathrm{kin}} \approx 200\,\mathrm{MeV}$. The muon energy spectra before and after the phase rotation are compared in Figure 17.

3.3.6 Cooling

After the phase rotation, a 188 m long cooling channel is foreseen to reduce the 2D transverse emittance of the muons by a factor of 2.5 to $1000\,\pi$ cm mrad *rms*. At the end of the cooling channel, the average muon energy is 360 MeV. The CERN cooling channel is described in detail in Section 4.7.

3.3.7 Acceleration

The cooled muons are accelerated to their final energy of 50 GeV. This is done in three steps with a linac up to 2 GeV and two recirculating linacs up to 20 and 50 GeV.

3.3.8 Decay Ring and Baselines from CERN

The decay ring starts immediately after the second recirculator. Its geometry depends on the locations of the detectors. These have not yet been

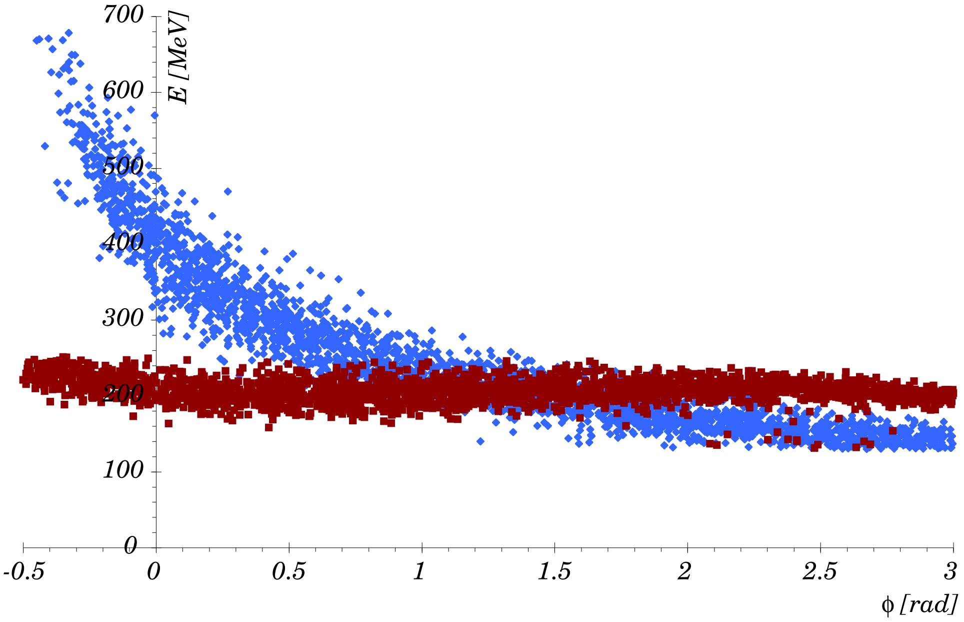

Figure 17: Muon energy spectrum before (blue) and after (red) phase rotation. Particles with energies below 130 MeV have been omitted.

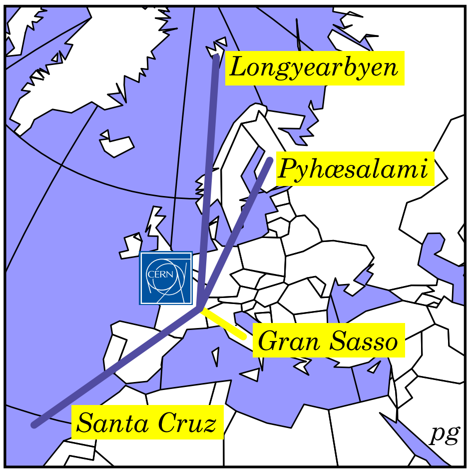

Figure 18: Baselines under consideration for a CERN based neutrino factory. Depending on the baseline length and the angle the efficiency of a neutrino factory varies between 30% and 70%. An optimal configuration is a straight line between the two detectors and the neutrino factory.

chosen. Possible neutrino beams going through the earth from CERN to various existing or planned neutrino detectors are shown in Figure 18. A more detailed discussion of decay ring scenarios for CERN can be found in Appendix A.

3.4 Ideas for Improving the CERN Scenario

3.4.1 Two Accumulator rRings

Doubling the number of accumulator and compressor rings could be a way to increase the intensity of the neutrino factory by a factor of 2. These two rings have been designed to fit into the old ISR tunnel that has a circumference of 1 km. Other factors, mainly the heat load on the stripping foil for the charge exchange, limit the size [40].

For the moment, a 1 km long proton beam produces a 1 km long muon beam that is injected into a 2 km long decay ring. By firing two pairs of rings one immediately after the other, the muon beam can be made longer. Assuming the same number of protons per rf pulse, the total number of pions (muons) produced per cycle would be doubled. The decay ring could be filled nearly completely. To visualize this option one can imagine filling the holes in the "beam length" part of Figure 13.

The following points analyze the impact of this scenario on the other components of the CERN scheme:

- SPL. Currently, the SPL runs for 2.2 ms every 20 ms to fill the accumulator. To fill a twice as big accumulator with twice the number of protons, it would have to run for 4.4 ms, a duty cycle of 22%.

- Target. The beam would hit the target during 6.6 μs instead the previous 3.3 μs. This is still much shorter than the observed break-up time of more than 750 μs [37]. The break up would be more violent and the Hg would be heated more, but the target would remain stable until the end of the passage of the proton beam.

- Cavities in phase rotation and cooling. These cavities could run for 100 μs every 50 Hz [42], so the new 6.6 μs duty cycle would not harm them.

- Recirculating linacs. Beam loading in the recirculating linacs would be a problem. The length of the first recirculator may be too short.
- Decay ring. The decay ring has to have a circumference of 2 km anyway because of efficiency considerations. It would be filled to nearly 100% instead of the currently planned 50% and thus be used more efficiently.

By building two accumulator and compressor rings the intensity of the neutrino factory could be doubled with reasonable cost. Components that have been used at low duty cycles, such as the cooling channel, could be used more efficiently.

3.4.2 Increasing the cycle speed

From the muon life time point of view, the cycle of the neutrino factory could be as fast as 200 Hz. All other parameters constant, the intensity could be increased by a factor of up to 4. To visualize this option one can imagine reducing the space between the peaks in the "muon lifetime" part of Figure 13.

The following points discuss the consequences of increasing the cycle speed towards 200 Hz, both in the case of the 3.3μs proton beam and the "double accumulator ring" variant.

- SPL. The duty factor would increase to 44% for the short proton beam and to 88% for the long proton beam. As the SPL could run in principle c.w., this seems feasable. There could be conflict for beam with other customers of the SPL.
- Target. A beam power on the target of 16 MW (in the case of the short proton beam) to 32 MW (long proton beam) is beyond any proposed parameter for high power proton targets. The speed of the Hg jet would have to be increased to up to 100 m/s — faster than the maximum speed of the TGV. The horn would have to be pulsed faster. A solution with multiple targets [43] would help overcome this problem. Still, a machine with 200 or more explosions per second is not completely unthinkable: The motor of the Jordan formula 1 team is a 3 litre V10 at 18 000 rpm, which is 300 Hz [44].
- Cavities. The cavities have been identified as the limiting factor for the cycle speed. The problem is not the extensive use of the cavities —

	current duty factor		max. duty factor
SPL	2.2 ms in 20 ms	11%	c.w.
target	30 cm of 50 cm	60%	?
44/88 MHz cavities	3.3 μs in 20 ms	1.6×10^-4	10^-3
recirculators	$8 \times 3.3\,\mu$s in 20 ms	1.3×10^-3	?
decay ring (length)	1 km of 2 km	50%	$\approx 100\%$
decay ring (time)	5 ms in 20 ms	25%	$\approx 100\%$

Table 5: Duty factors of the components of the CERN scheme.

they are only used during 3.3μs (6.6μs) per cycle, the problem is the long rise time (318 μs for the 44 MHz cavity) that adds to the effective duty cycle. Making the rise time shorter would keep the duty factor within tolerable limits, even though the cycle is faster. This is possible with more rf power, but at a cost.

- Recirculators. The recirculators would definitely suffer from severe beam loading. Additional beam loading compensation would have to be foreseen.

- Decay ring. Running at 200 Hz would just make efficient use of the decay ring.

Increasing the cycle speed towards 200 Hz would imply a lot of difficulties, but it is one of the few ways to increase the intensity by a factor of 4. It is a question of optimization if one wants to encounter the difficulties associated with this step.

Chapter 4

Ionisation Cooling

4.1 Emittance

The emittance ε is the phase space volume of a beam. Emittances can be defined for the whole six dimensional phase space or subspaces. Common expressions are the 2D transverse emittance in one transverse plane, the 4D transverse emittance in both transverse planes and the full 6D emittance.

Liouville's theorem states that the phase space volume occupied by a group of particles stays constant with time, if only conservative forces are applied:

$$f\left(\vec{q}(t), \vec{p}(t), t\right) = f\left(\vec{q}(t'), \vec{p}(t'), t'\right) \tag{24}$$

The direct consequence of Liouville's theorem is the fact that the emittance is a conserved quantity, if it is defined using conjugate variables, e.g., $\vec{p}$ and $\vec{q}$. It has been, however, a tradition in beam physics to use a different set of variables to define the transverse emittance. These are x (the same as q) and x', which is the angle between the particle's momentum vector and the beam line axis. With the usual convention that the longitudinal momentum component is denoted as p_z, x' is defined as:

$$x' = \mathrm{atan}\left(\frac{p_x}{p_z}\right) \tag{25}$$

The divergence x' is usually given in mrad. As x' is normally a very small quantity, one can make the approximations $\mathrm{atan}(\delta) \approx \delta$ and $p = p_z / \cos x' \approx p_z$. This leads to the easier-to-use expressions

$$x' = \mathrm{atan}\left(\frac{p_x}{p_z}\right) \approx \left(\frac{p_x}{p_z}\right) \approx \left(\frac{p_x}{p}\right) \tag{26}$$

The choice of non-canonical variables leads to the problem that ε is no more conserved. The best example for the non-conservation of ε is adiabatic damping that occurs when accelerating a beam. In a perfect rf cavity, the longitudinal momentum p_z of a particle is increased by an electric field E_z, while p_x is unchanged. This results in a decreasing $x' = p_x/p_z$. The volume xx' seems to shrink.

To compensate the choice of non-canonical variables for ε, the concept of the normalised emittance ε_{N} has been introduced:

$$\varepsilon_{\mathrm{N}} = \beta\gamma\varepsilon \tag{27}$$

The relativistic factors $\beta\gamma$ compensate the division by p in x'. With $p = m_0 c\beta\gamma$ the normalised emittance can be written as:

$$\varepsilon_{\mathrm{N}} = \beta\gamma\varepsilon = \beta\gamma x x' = x\beta\gamma\frac{p_x}{p} = x\frac{p}{m_0 c}\frac{p_x}{p} = x p_x \frac{1}{m_0 c}, \tag{28}$$

where ε_{N} is the product of two conjugate variables divided by the constant factor $m_0 c$. Therefore ε_{N} is a conserved value under Liouville's theorem.

The phase space volume of a beam can be expressed in various ways. One is the total or 100% emittance, which is the smallest ellipse (or hyper-ellipsoid) that contains all the particles. This measure is often distorted by errant particles that increase ε to virtually any value. Better definitions are the 95% and 90% emittances, which are the smallest volumes that contain the respective percentage of beam particles.

It is convenient to define an *rms* emittance. This is the so-called statistical definition of ε, for two dimensions it is given as:

$$\varepsilon^2 = \left\langle x^2 \right\rangle \left\langle x'^2 \right\rangle - \langle x x' \rangle^2 \tag{29}$$

The above definition gives the 1σ emittance. In order to minimise losses, accelerators are normally built to accommodate several times the *rms* emittance.

The dimension of the emittance is meter, although it is more common to use cm mrad, which is the product of practical units for x and x'. Sometimes ε is given in units of π mm mrad or π cm mrad. This comes from $A = ab\pi$, the expression for the area of an ellipse with the two axes a and b.

In all simulations in this study, the emittance has been calculated with the program PARMILA [45], that uses the statistical definition of the emittance and uses π cm mrad as units.

4.2 Cooling

Cooling a charged particle beam means reducing its normalised emittance. This seems to be violating Liouville's theorem in the first place, but it is not. Only the phase space of *the whole system* is conserved, according to Liouville. If one wants to cool a beam, one has to find a system that contains more than just the beam and to "shift" phase space volume from the beam to the rest of the system. It is much like a refrigerator works: the system is more than the refrigerating compartment. While water freezes inside, the kitchen air is heated. The second law of thermodynamics is observed.

Cooling beams has been used for over 30 years in accelerator physics to make beams brighter and to achieve better physics results [46]. Cooling has been a successful ingredient especially in colliders, in which the luminosity increases with $1/\varepsilon^2$.

Depending on whether the transverse or the longitudinal phase space densities are increased, we talk about transverse or longitudinal cooling. There are four standard methods of cooling today:

Radiation cooling has been the first cooling technique. It profits from the synchrotron radiation emitted by charged particles in a storage ring. The power P_S of the synchrotron radiation depends on the rest mass m_0 and the energy E of the particle as well as the radius R of the ring:

$$P_S = \frac{e^2 c}{4\pi\varepsilon_0} \frac{E^4}{R^2 (m_0 c^2)^4}. \tag{30}$$

Radiation cooling can cool a beam transversally and longitudinally, effectively increasing its 6D phase space density. The synchrotron radiation is emitted in the particle's direction, decreasing its momentum $\vec{p}$. The energy loss is restored with rf cavities, that produce a field parallel to the beam. Only the longitudinal momentum component p_z is restored, while the transverse component p_t is left unchanged. The result is a decreasing $x' = p_\mathrm{t}/p_z$ and thus transverse cooling. On top of that, the energy loss is proportional to E^4, which means that particles with higher energy loose more energy than lower energetic ones. As all particles receive the same amount of energy $E = eU$ in the cavity, the result is a reduction of the momentum spread, which is longitudinal cooling. As the synchrotron radiation is inversely proportional to m_0^4, radiation cooling works only for electrons with high energy. It has successfully been used at LEP.

Stochastic cooling has been proposed Simon van der Meer in 1972 [47] as a solution to increase the luminosity in the $p\overline{p}$-collider ISR, because radiation cooling would not work for the low energetic and heavy (anti-)protons. The idea of stochastic cooling is to divide the beam in small chunks and to measure the sidewards deviation of each chunk at one point, to amplify, invert this signal and to apply it back to this chunk. If y part of the beam is a little bit too left, it will be kicked right and vice versa. Unlike radiation cooling, stochastic cooling acts on the beam as a whole in a statistic manner. It is best suited for low intensity, high emittance beams.

Electron cooling can be imagined as temperature exchange between a "hot" ion beam and a "cold" electron beam. As it is much easier to produce a cold electron beam, electrons can be used to take up some transverse momentum of the ions via elastic scattering. It is best used for heavy particles with low energies.

Laser cooling is a special cooling technique for partially ionised atoms. It uses that fact that transitions between two energy levels of the bound electrons are very sharp. Ions at different speeds see the laser light at different wave lengths due to the Doppler shift ($\omega = (1 \pm v/c)\gamma\omega_{\text{laser}}$). This means that only ions with the "right" momentum absorb the laser light and are accelerated by $E = \hbar\nu$. During laser cooling, all particles are slowly decelerated. The particles with a too high kinetic energy are decelerated while the particles that fall in the resonant region, are re-accelerated. The result is a longitudinally cooled beam with a very small energy spread. Laser cooling can only be used for partially ionised atoms with a suitable meta–stable energy state.

All of the above techniques are proven and have their applications. None of them, however, is suited for cooling muons. The reasons are the short life time (τ=2.197 μs) and the medium mass (mc^2 = 105.7 MeV) of the muon. That is why a fifth cooling technique is under consideration for a neutrino factory and a muon collider:

Figure 19: The principle of transverse ionisation cooling: a particle passes through an absorber and looses momentum uniformly (1), but only its longitudinal momentum p_z is restored (3). The divergence $x' = \mathrm{atan}(p_\mathrm{t}/p_z)$ is reduced. Unfortunately scattering (2) reduces the cooling effect or may even lead to heating.

4.3 Transverse Ionisation Cooling

Ionisation cooling has been conceived by Budker and Skrinsky in the early 1970s [3], [4], already with cooling muon beams for a muon collider in their minds. The principle of transverse ionisation cooling (see figure 19) is the reduction of x' by repetitive moderation and acceleration. A particle is sent through an absorber, where it looses uniformly momentum. After this, only the longitudinal component of the momentum is restored. If p_t is reduced and p_z restored, the divergence $x' = \mathrm{atan}\frac{p_\mathrm{t}}{p_z}$ decreases – much like radiation cooling.

Without any perturbation, cooling would be an exponential process leading to infinitely small emittances. Unfortunately there are effects that increase the emittance. As they counteract cooling, they are called heating. Heating effects are multiple scattering in the absorber, nonlinearities in the focusing and rf defocusing. The multiple scattering is independent of the emittance while cooling is proportional to the emittance. For low emittances, the cooling effect decreases until an equilibrium emittance is reached that cannot be further reduced, as calculated in Section 4.5.

4.4 Effects in Cooling

4.4.1 Energy Loss – dE/dx

When a charged particle passes through matter, it looses energy due to ionisation, which is the result of an inelastic scatter between the particle and the electrons in the atomic shell. The cross-section of this process depends highly on the particle momentum and is shown in figure 20. Muons

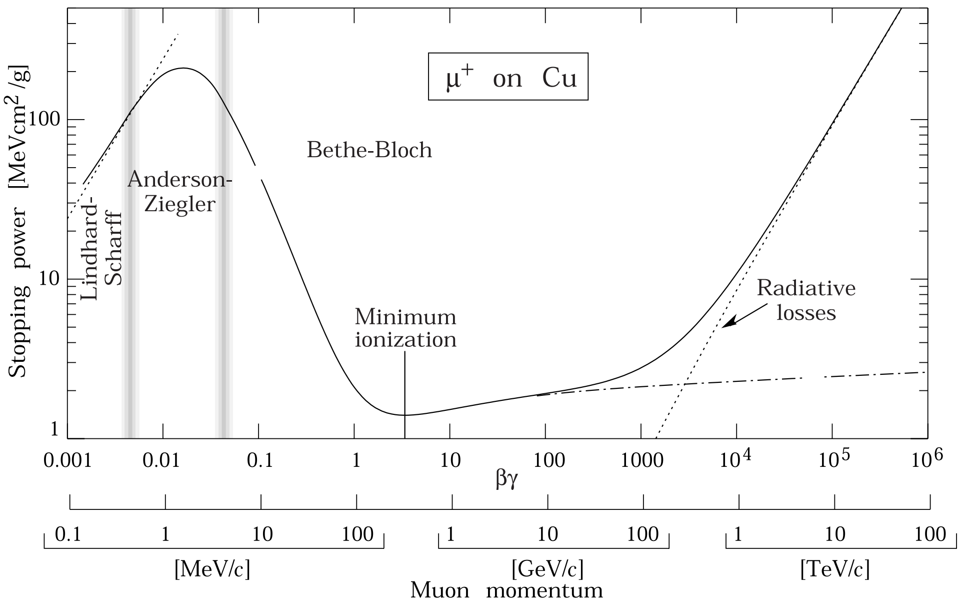

Figure 20: Energy loss of muons in matter. For a wide range of energy, muons loose the same amount of energy per cm of matter transversed. Longitudinal cooling is possible in the energy ranges with a positive slope. From [48]

in the energy range of 150 MeV to 400 MeV are so-called minimum ionising particles. In the case of a liquid hydrogen absorber, they loose 0.28 MeV per cm.

4.4.2 Moliere Scattering

When a charged particle passes through matter, it is scattered elastically off the nuclei of the absorber, usually by a small angle. This process occurs very often during the passage through the absorber. The standard way to describe the effect of multiple scattering is Moliere's statistic theory [49].

Moliere describes the result of the multiple scattering with one parameter, the *rms* plane scattering angle ϑ_0. With the thickness z of the absorber and the radiation length X_0 of the material, ϑ_0 is defined for a particle with the unit charge q and a momentum p as[1]:

$$\vartheta_0 = \frac{13.6\mathrm{MeV}}{\beta c p} q \sqrt{z/X_0}[1 + 0.038 \ln(z/X_0)] \tag{31}$$

[1]In the review of Particle Physics [48], the charge is expressed with z and the length of the absorber with x. In order to maintain a consistent notation, here the charge is denoted with q and the width of the absorber with z

This equation can be simplified in good approximation for muons with $q = 1$ to

$$\vartheta_0 = \frac{13.6\mathrm{MeV}}{\beta c p}\sqrt{z/X_0}. \tag{32}$$

The distribution follows with good approximation the Gaussian distribution.

Moliere theory is already over 50 years old and there is an ongoing discussion about its accuracy for hydrogen and other low Z materials. A test of Moliere theory and competing new models is currently performed at the MUSCAT experiment [50].

4.4.3 Focusing

Focusing influences cooling via the relation between ε and x'. This relation is defined using the β-function, which is denoted as $\beta_\perp$ in this thesis to avoid confusion with the relativistic β.

For an upright beam ($\alpha = 0$), the divergence is related to the emittance as:

$$x' = \sqrt{\frac{\varepsilon}{\beta_\perp}} \tag{33}$$

In a solenoid, the β-function can be given as approximately

$$\beta_\perp = \frac{2pc}{ecB_z} \tag{34}$$

For a muon with $p = 200\,\mathrm{MeV/c}$ in a magnetic field of $2\,\mathrm{T}$, $\beta_\perp = 0.667\,\mathrm{m}$. For an *rms* emittance $\varepsilon = 2000\,\pi\,\mathrm{cm\,mrad}$ one can calculate an *rms* x' of $x' \approx 100\,\mathrm{mrad}$.

4.4.4 rf–Acceleration

A perfect rf cavity has an accelerating field E_z, that is parallel to the beam axis. Depending on the particle's phase φ, it is de- or accelerated:

$$\Delta E = eU\cos\varphi \tag{35}$$

In a real cavity, the field is not perfectly parallel to the beam axis and there are other field components apart from E_z. This leads to effects such as rf defocusing. Furthermore, the design U is only achieved on the beam axis.

before cooling	p	p_x	$x' = p_x/p$
after the absorber	$p - dp$	$p_x - x'dp$	$x' = (p_x - x'dp)/(p - dp) = p_x/p$
after the rf cavity	p	$p_x - x'dp$	$x' = (p_x - x'dp)/p$

Table 6: Steps in the transverse cooling process. The divergence x' is reduced in the acceleration. The reduction of the normalised emittance ε_{N}, however, takes place in the absorber: here, the $\beta\gamma$ is reduced with constant x'.

4.5 Transverse Cooling Rate

The following derivation of the transverse cooling rate per unit length λ starts from the viewpoint that transverse cooling is x' reduction (under constant conditions). This statement is similar to the standard notion that cooling is emittance reduction (as $x' = \sqrt{\varepsilon/\beta_\perp}$ for $\alpha = 0$), but leads to a more intuitive derivation.

The derivation is based on a cooling cell that consists of an absorber that reduces the total momentum p by dp, a set of cavities that restore p_z and constant focusing throughout the whole cell. The cell length is dz. The following equations are for one 2D plane, the $x - x'$ plane. p_x is one component of p_{t}. The paraxial approximation $p = p_z$ is used and

$$x' = \operatorname{atan}\left(\frac{p_x}{p_z}\right) \approx \frac{p_x}{p_z} \tag{36}$$

Even though the divergences that occur in a cooling channel can be quite big, the above approximation is still valid, as the maximum x' is 300 mrad and atan(0.3) is 0.291.

In the absorber, all components of p are reduced uniformly. If dp/dz is the momentum loss in the absorber and one rewrites Equation 36 as $p_x = x'p_z$, then

$$\frac{dp_x}{dz} = x'\frac{dp_z}{dz} \approx x'\frac{dp}{dz} \tag{37}$$

After the absorber, $p_z \approx p$ is restored in an rf cavity, while p_x remains unchanged. The change of x' per unit length can be written as total differential:

$$\frac{dx'}{dz} = \frac{d}{dz}\left(\frac{p_x}{p_z}\right) = \underbrace{\frac{1}{p_z}\frac{d}{dz}p_x + p_x\frac{d}{dz}\frac{1}{p_z}}_{absorber} + \underbrace{\frac{1}{p_z}\frac{d}{dz}p_x + p_x\frac{d}{dz}\frac{1}{p_z}}_{cavity} \tag{38}$$

The second and the fourth term cancel out, because p_z is restored after the rf cavity. In a perfect cavity, the third term is zero, as the electrical field has only a z component. The remaining term can be written as:

$$\frac{1}{p_z}\frac{dp_x}{dz} \approx \frac{x'}{p}\frac{dp}{dz}. \tag{39}$$

The factor dp/dz is again the momentum loss in the absorber. Without multiple scattering, cooling can be described as the following x' reduction:

$$\frac{dx'}{dz} = \frac{x'}{p}\frac{dp}{dz} \tag{40}$$

This reduction is indirectly proportional to x' and inversely proportional to p. This means that cooling works best for high x' and low momenta.

To include heating, one has to pass to rms quantities, as multiple scattering is only defined using the rms scattering angle $\langle\vartheta_0\rangle$. The rms geometric emittance is defined as:

$$\varepsilon^2 = \left\langle x^2 \right\rangle \left\langle x'^2 \right\rangle - \langle xx' \rangle^2 \tag{41}$$

Supposing there are no correlations between x and x', the second term can be set to zero. Furthermore, a short absorber is supposed, thus $dx/dz = 0$. The derivation to z is split into the contributions of cooling and heating:

$$\begin{aligned} 2\varepsilon\frac{d\varepsilon}{dz} &= \left\langle x^2 \right\rangle \left(\frac{d}{dz}\left\langle x'^2 \right\rangle_{\text{cool}} + \frac{d}{dz}\left\langle x'^2 \right\rangle_{\text{heat}} \right) \\ &= \left\langle x^2 \right\rangle 2x' \frac{dx'}{dz}_{\text{cool}} + \left\langle x^2 \right\rangle \frac{dx'^2}{dz}_{\text{heat}} \end{aligned} \tag{42}$$

The expression for the x' reduction (Equation 40) and the simplified multiple scattering (Equation 32) are inserted Equation 42:

$$2\varepsilon\frac{d\varepsilon}{dz} = \left\langle x^2 \right\rangle 2x' \frac{x'}{p}\frac{dp}{dz} + \left\langle x^2 \right\rangle \frac{13.6\text{MeV}^2}{\beta^2 c^2 p^2 X_0} \tag{43}$$

with $\langle x^2 \rangle = \varepsilon\beta_\perp$ the second term can be rewritten as:

$$2\varepsilon\frac{d\varepsilon}{dz} = 2\left\langle x^2 \right\rangle \left\langle x'^2 \right\rangle \frac{1}{p}\frac{dp}{dz} + \varepsilon\beta_\perp \frac{13.6MeV^2}{\beta^2 c^2 p^2 X_0} \tag{44}$$

Using $\varepsilon^2 = \langle x^2 \rangle \langle x'^2 \rangle$ and dividing by 2ε we get:

$$\frac{d\varepsilon}{dz} = \frac{\varepsilon}{p}\frac{dp}{dz} + \frac{\beta_\perp 13.6\text{MeV}^2}{2\beta^2 c^2 p^2 X_0} \tag{45}$$

Using $\varepsilon_{\mathrm{N}} = \beta\gamma\varepsilon$, the relativistic relations $E = cp/\beta$ and $E = \gamma m_0 c^2$ and inverting the sign of the energy loss, Equation 45 can be transformed into the form of the Fermilab status report [1]:

$$\frac{d\varepsilon_N}{dz} = -\frac{\varepsilon_N}{\beta^2 E}\frac{dE}{dz} + \frac{\beta_\perp 13.6\mathrm{MeV}^2}{2\beta^3 m E X_0} \tag{46}$$

This equation can be solved in the absence of heating ($X_0 = \infty$):

$$\varepsilon_N(z) = \varepsilon_N(0)e^{-\frac{1}{\beta^2 E}\frac{dE}{dz}} \tag{47}$$

The solution with heating is:

$$\varepsilon_N(z) = \varepsilon_N(0)e^{-\frac{1}{\beta^2 E}\frac{dE}{dz}} + \frac{\beta_\perp 13.6\mathrm{MeV}^2}{2\beta m X_0}\frac{1}{\frac{dE}{dz}} \tag{48}$$

In the first case, cooling would asymptotically reach zero, which is unphysical. In the second case, cooling would asymptotically reach the so-called equilibrium emittance. This is the emittance at which cooling and heating cancel out.

$$\begin{aligned}\frac{\varepsilon_{EQ}}{\beta^2 E}\frac{dE}{dz} &= \frac{\beta_\perp 13.6\mathrm{MeV}^2}{2\beta^3 m E X_0} \\ \varepsilon_{EQ} &= \frac{\beta_\perp 13.6\mathrm{MeV}^2}{2\beta m X_0}\frac{1}{\frac{dE}{dz}}\end{aligned} \tag{49}$$

4.6 Longitudinal Ionisation Cooling

On top of transverse cooling, longitudinal ionisation cooling is possible in some energy ranges where the energy loss ($\frac{dE}{dx}/E$) has a positive slope. In this case particles with a higher kinetic energy are decelerated more than particles with a lower kinetic energy. The energy spread is reduced. The mechanism is similar to radiation cooling, but the effect is much smaller. For muons, the energy loss curve has a significant positive slope only below 1 MeV and above 1 TeV (see figure 20). These energy ranges are of no practical interest for the neutrino factory or a muon collider.

Another way of performing longitudinal cooling is emittance exchange with wedge absorbers. The idea here is to create a $p - x$ correlation in a bent solenoid. In a bending section, particles with higher momentum have a larger bending radius. If a wedge absorber is placed at the end of the bending section in a way that muons with higher energy pass through the thicker end of the absorber, while the lower energetic ones pass through the thinner end, the energy spread is reduced. However, the resulting beam is wider and has a bigger transverse emittance. That is why this procedure is called emittance exchange – the 6D emittance is not reduced, only its components are exchanged. This procedure is still valuable, because it can be combined with transverse cooling to achieve real cooling in all three planes.

Emittance exchange occurs in any solenoidal channel, when the focusing strength is changed, as the transverse and the longitudinal planes are coupled.

4.7 The CERN 44/88MHz Cooling Channel

The CERN 44/88MHz cooling channel [51] consists of 2 parts and 5 main sections (see figure 21):

1. The 30 m long decay channel. This is a 1.8 T solenoidal channel right after the target. There is no rf in this part.

2. The 30 m long phase rotation channel. Here the focusing is performed by 1.8 T solenoids integrated in 30 44 MHz cavities, each 1m long, operating at -90°and delivering 2 MV/m to perform the phase rotation.

3. The first part of the cooling channel (Cooling I). Here are 11 cooling cells, each consisting of 4 44 MHz cavities (operating at 0°) and a 0.24 m long hydrogen absorber. Focusing as in (2.).

4. The acceleration part, which takes the average beam particle energy from 200 MeV to 280 MeV. It consists just like the previous section of 44 cavities, again at 44 MHz operating on crest. Focusing as in (2.).

5. The second part of the cooling channel (Cooling II), which operates at 88 MHz. This part is focused by 5 T solenoids. With the effect of the first cooling, the acceleration and the stronger focusing, the beam is now much smaller, so that it can fit into the bore of the 88 MHz cavities. These operate at 4 MV/m. There are 28 cooling cells in this section, each consisting of 8 cavities that are 0.5 m long and a hydrogen absorber of 0.4 m length.

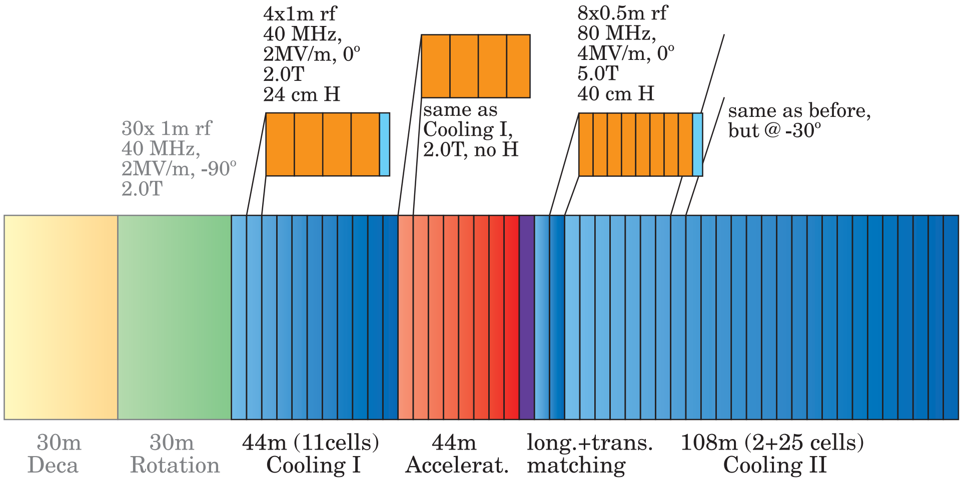

Figure 21: The CERN 44/88MHz cooling channel. The length of each section reflects only the rf. Immediately after the target there is a 30 m long decay channel for the pion decay and a phase rotation section. These two parts are not included in the sensitivity study in chapter 5. The cooling channel itself consists of a first cooling section at 44 MHz, an acceleration section and a second cooling section at 88 MHz.

Chapter 5

The Cooling Channel Sensitivity Study

5.1 Sensitivity and Stability

The sensitivity of a system expresses how much the output function $S = S(x_1, x_2, \ldots)$ of the system changes, if one of the input parameter x_i is varied. Mathematically this is expressed with the partial derivative:

$$\text{sensitivity} = \frac{\partial S}{\partial x_i} \tag{50}$$

The sensitivity of a known function can be easily calculated. For a real system, however, this is a too theoretical concept. First, we suppose that all parameters are independent ($\partial x_i / \partial x_j = 0$) and that we can form the partial derivatives. Second, in most cases we do not know not even know the function S and we have to rely on simulations. With data from simulations, it is not possible to give $\partial S / \partial a$, but only a linear approximation around a value x_i^*:

$$\left.\frac{\partial S}{\partial x_i}\right|_{x_i = x_i^*} \approx \frac{S(x_i^*) - S(x_i)}{x_i^* - x_i} = \frac{\Delta S}{\Delta x_i}. \tag{51}$$

Therefore any result found with simulations is only valid for x_i near the reference parameter x_i^*.

The main goals of studying the sensitivity of a system are: to determine the stability of a system, to find tolerances for input parameters and to find room for improvement.

Stability means that the output parameter changes slower than the input parameter. A system is called stable, if

$$\left|\frac{S(x_i^*) - S(x_i^* \pm \Delta x_i)}{S(x_i^*)}\right| = \left|\frac{\Delta S}{S}\right| \leq \left|\frac{\Delta x_i}{x_i^*}\right|. \tag{52}$$

If a system is stable, there are large tolerances for the input parameters. If it is not stable, it is easy to influence and hence can be improved easily.

5.2 Aim of this Study and Strategy

The overall goal is to understand the 44/88 MHz cooling channel better – especially how the output changes as a function of the input variables and channel parameters – in order to use this information in an optimisation process.

There are two sets of quantities that can be influenced in the design of the cooling channel: input variables and channel parameters. The first set describes the beam at the beginning of the cooling channel. These variables can be influenced by changing the design of the upstream parts of the neutrino factory. The channel parameters describe the design and settings of the hardware of the cooling channel. An overview of both types of quantities is given in Table 7. It is the aim of this study to understand how the many variables and parameters that can be changed influence the overall performance.

5.2.1 Simulation Strategy

The strategy of this study is to change one single input variable or channel parameter in a sensible range and to study the changes of the output variables that are listed in Table 10. It is then established if the system is stable against changes of this variable/parameter and how much the various output variables depend on it. The output is compared to a reference run with the baseline parameters.

5.2.2 Simulated Beam Method

It is not the goal of this study to establish how the variables/parameters can be achieved technically. For example, no effort is taken to find an actual upstream channel that produces the used input variables. No technical design work has been performed to check the praticabilty of the used channel parameters.

Input variables		
ε_{r}	radial emittance	$3130\,\pi$ cm mrad
$\overline{E}$	mean kinetic beam energy	180 MeV
ΔE	energy spread	±15%
$\Delta\varphi$	phase spread	±45°
Channel parameters — 44 MHz part		
U	accelerating gradient	2 MV/m
φ_{rf}	rf phase	0°
B_{z}	focusing field strength	2.0 T
l	Length of this part of the channel	44 m
X_0	Radiation length of the absorber material	866 cm (H_2)
Channel parameters — 88 MHz part		
U	accelerating gradient	4 MV/m
φ_{rf}	rf phase	–30°
B_{z}	focusing field strength	5.0 T
l	Length of this part of the channel	116 m
X_0	Radiation length of the absorber material	866 cm (H_2)

Table 7: Input variables and beam parameters that have been studied in the sensitivity study and the values for the CERN baseline design.

In the standard simulation approach, a pion distribution is created using the MARS code system [52]. The particles are tracked through the decay channel and the phase rotation. The particle distribution at the end of the phase rotation is the input for the cooling channel. This approach is called *simulated particles method*, because it tracks particles that have been created by a hadronic interaction simulation code. As this distribution is sent through the cooling channel, only muons and only the interesting part of the energy spectrum are used. In reality, other particles such as kaons, neutrons and protons are produced at the target and propagate through the cooling channel.

The above approach does not work for hypothetical variations of the input beam variables, because they can only be changed in the upstream parts of the neutrino factory such as target, transverse capture, decay and phase rotation. As these parts are not the subject of this study, the *simulated beam method* has been developed. In the simulated beam method a beam is

	sim. particles	sim. beam
At the beginning of the channel		
ε_{r}	3089	3089
ε_z	314	336
At the end of the 44 MHz part		
ε_{r}	2231	2404
ε_z	368	405
T	0.968	0.997
At the end of the 88 MHz part		
ε_{r}	916	1122
ε_z	496	515
T	0.526	0.505

Table 8: Results of the calibration run for the simulated beam method. The simulation has been performed on the CERN baseline design as shown in Figure 21. the transverse emittance ε_{r} is defined in Equation 54.

generated by the beam generator PARMILA [45] using the twiss parameters α and β and the 2D-emittances $\varepsilon_{x,y}$ for the transverse phase space and the pahse φ, the phase spread $\Delta\varphi$, the mean kinetic beam energy $\overline{E}$ and the energy spread ΔE for the longitudinal phase space. PARMILA creates a Gaussian distribution with these *rms* values.

To match these two methods, a calibration run has been performed. The beam variables for the simulated beam method have been chosen in order to match the results of the simulated particle method as close as possible. There are, however, differences that can be attributed to the form of the distribution: the simulated particles come from the big phase space of the pions that has been transversally cut by the acceptance of the phase rotation and cooling channel and longitudinally in the simulation. On the transverse plane the simulated particle method will always give better results because it can profit from cutting away few halo particles, which results in a significant reduction of the *rms* emittance. There is no way to exactly reproduce the transverse phase space, but to better represent the longitudinal cuts, a larger distribution is created in PARMILA and the cut in phase. In the calibration run with 10,000 particles the simulated beam method proved to be precise to the 10% level. The results of the calibration run are compared to the

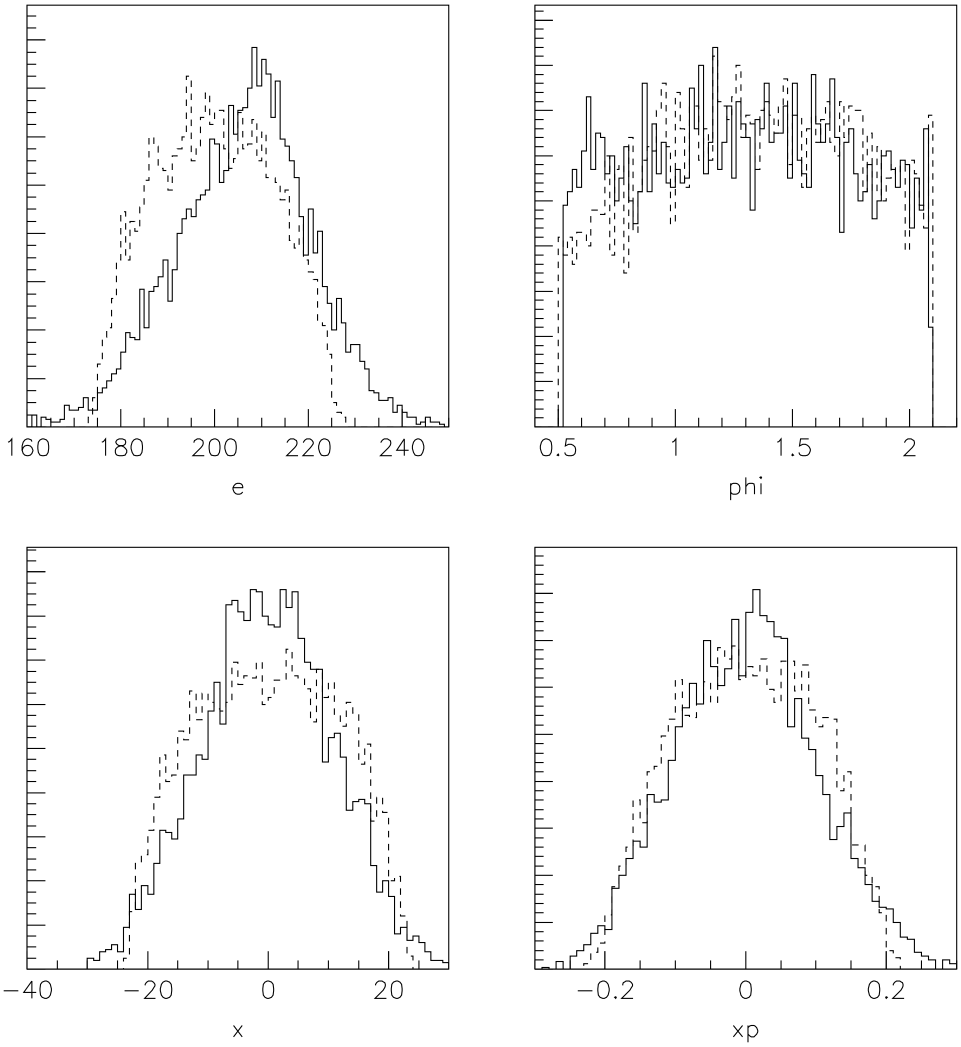

Figure 22: Calibration for the simulated beam method: histograms comparing the simulated beam (dashed) with the MARS particle production (line). The units are E[MeV], φ[rad], x[cm], x'[rad]. The number of particles per bin is given in arbitrary units.

simulated particles method in Table 8 and a phase space plot comparing the input variables is shown in Figure 22. The results are close to the original beam, but there are some subtile differences. Therefore the results of this study should only be interpreted as relative to the reference run with the simulated beam method.

The simulated beam method has also been used for the runs with the changed channel parameters to allow for better comparison.

5.3 Assessments for Cooling Devices

If one wants to find the "best" cooling channel, criteria for assessing a cooling channel – so called *figures of merit* – have to be defined. To judge the cooling performance, one has to keep in mind the reasons for cooling. There are two main reasons why cooling is applied. First, to fit a large beam into a given acceptance and second to increase the luminosity of an accelerator. Depending on the scope and on the question, whether losses or longitudinal heating are acceptable, two different types of figure of merit can be defined: emittance and number of particles in an acceptance.

5.3.1 Assessments Based on the Emittance

The most basic quantity for assessing a cooling channel is the 2D transverse cooling factor Λ_2:

$$\Lambda_2 = \frac{\varepsilon_{\mathrm{r,in}}}{\varepsilon_{\mathrm{r,out}}} \tag{53}$$

The transverse emittance ε_{r} is defined as

$$\varepsilon_{\mathrm{r}} = \sqrt{\varepsilon_{\mathrm{x}}\varepsilon_{\mathrm{y}}}. \tag{54}$$

For the study of the CERN cooling channel it is better to use ε_{r} than $\varepsilon_{\mathrm{x,y}}$, because the $x-y$ coupling of the solenoidal field produces oscillations between ε_{x} and ε_{y}. The 4D transverse cooling factor Λ_4 is then defined as:

$$\Lambda_4 = \frac{\varepsilon_{\mathrm{x,in}}\varepsilon_{\mathrm{y,in}}}{\varepsilon_{\mathrm{x,out}}\varepsilon_{\mathrm{y,out}}} = (\Lambda_2)^2 \tag{55}$$

$\Lambda_{2,4}$ are > 1 for cooling and < 1 for heating. Λ_2 is a simple concept to assess a cooling channel, but there are two cases in which Λ_2 overestimates the transverse cooling:

Collimation. It is very easy to reduce the transverse emittance of a beam by simply collimating it. If only the transverse cooling factor Λ_2 is studied, the channel would seem to cool. Even longitudinal collimation is possible in a bent section, which would pretend longitudinal cooling.

In a lossless or almost lossless channel, Λ_2 is a good measure for cooling. In all other cases the transmission $T = n_{\mathrm{out}}/n_{\mathrm{in}}$ has to be taken into account to prevent an interpretation error. The solution is to define a new quantity Ξ_4, the transmission corrected 4D cooling factor:

$$\Xi_4 = \frac{\varepsilon_{\mathrm{x,in}}\varepsilon_{\mathrm{y,in}}}{\varepsilon_{\mathrm{x,out}}\varepsilon_{\mathrm{y,out}}}\frac{n_{\mathrm{out}}}{n_{\mathrm{in}}} = \Lambda_4 T \tag{56}$$

The 2D corrected cooling factor Ξ_2 is analogously defined as

$$\Xi_2 = \Lambda_2 \sqrt{T}. \tag{57}$$

As $T \leq 1$, we conclude $\Xi_{2,4} \leq \Lambda_{2,4}$. This is the reason why $\Lambda_{2,4}$ can only overestimate, but never underestimate the cooling performance.

This definition is based on the assumption that all losses occur in the transverse plane. If this is not the case, the longitudinal losses have to be accounted for separately or another definition has to be used.

Longitudinal heating. Liouville's theorem (Equation 24) states the conservation of the 6D phase space. It is, however, perfectly compatible with Liouville that the shape of this 6D volume changes. This may lead to a reduction of the transverse phase space, while the longitudinal one grows by the same amount. If only the transverse cooling factor Λ_2 is looked at, a system may seem to cool, even though the 6D phase space is conserved. This phenomenon is called emittance exchange (see Section 4.6).

To cover the full phase space, a 6D cooling factor Λ_6 can be defined as:

$$\Lambda_6 = \frac{\varepsilon_{\mathrm{x,in}}\varepsilon_{\mathrm{y,in}}\varepsilon_{\mathrm{z,in}}}{\varepsilon_{\mathrm{x,out}}\varepsilon_{\mathrm{y,out}}\varepsilon_{\mathrm{z,out}}} \tag{58}$$

If losses have to be taken into account, Ξ_6 can be defined as

$$\Xi_6 = \Lambda_6 T. \tag{59}$$

Again, all losses are attributed to the transverse plane in this definition, as $\Xi_6 = \Xi_4 \frac{\varepsilon_{\mathrm{z,in}}}{\varepsilon_{\mathrm{z,out}}}$.

There may be, however, cases in which an ε_r reduction is desired and a ε_z reduction does no harm. In this case, Λ_2 or Ξ_2 are sufficient measures for the cooling performance. This is the case in the CERN baseline design: the longitudinal acceptance of the downstream elements is large enough that a certain amount of longitudinal heating can be tolerated, while the transverse cooling must be maximised.

Looking only at Λ_6 is not the *ultima ratio.* From the point of view of Liouville's theorem, this might be the right measure, but from a practical point of view, the possibility to actually perform emittance exchange is limited (see Section 4.6). It is not enough if a cooling channel reduces the 6D phase space volume, it has to reduce it in the *right plane*, as well.

In this section, six different figures of merit based on emittance have been defined. There is no general rule which is the "best" to use. In any case, the figure of merit has to be matched to the goals of the cooling system.

5.3.2 Assessment Based on the Intensity within an Acceptance

There are two reasons why one wants to cool a muon beam: first, because the subsequent accelerators would be of an immense transverse dimension and second because the resulting neutrino beam would have a too large divergence. The first criterion can be translated into the problem of stuffing the maximum possible number of muons into a given acceptance. Cooling is then be defined as the increase of the number of particles within a given acceptance. The intensity-based cooling factor Π is defined as:

$$\Pi = \frac{n_{\mathrm{out}}\big|_{\text{particle in } \varepsilon_{\mathrm{acc}}}}{n_{\mathrm{in}}\big|_{\text{particle in } \varepsilon_{\mathrm{acc}}}} \tag{60}$$

If the reference acceptance $\varepsilon_{\mathrm{acc}}$ is 2-dimensional, Π_2 is the 2D intensity-based cooling factor and in the same way Π_4 and Π_6 are defined. For all cases, $\Pi >$ '1 for cooling and $\Pi < 1$ for heating.

This definition is based on the assumption that there is a point in the cooling channel, where the acceptance decreases and losses occur. The question is, how many times more particle fit into the lower acceptance with cooling than without cooling. It is important to have the application in mind when establishing $\varepsilon_{\mathrm{acc}}$, because choosing the wrong $\varepsilon_{\mathrm{acc}}$ can lead to artifacts:

Too large $\varepsilon_{\mathrm{acc}}$. If $\varepsilon_{\mathrm{acc}}$ is larger than the beam's original emittance, then $\Pi_2 = 1$. Cooling cannot be observed, even though it might occur.

Too small $\varepsilon_{\mathrm{acc}}$. If $\varepsilon_{\mathrm{acc}}$ is smaller than the equilibrium emittance of the cooling channel, then $\Pi_2 < 1$ even if $\Lambda_2 \gg 1$. The channel seems to heat. The reason for this is the fact that cooling happens only on the envelope of the beam, while the core might even be heated. As the envelope has a larger weight, the heating of the core does not change the cooling performance. If this edge is not taken into account because of a small $\varepsilon_{\mathrm{acc}}$, then cooling cannot be observed.

code	description
3	drift section
6	slit
10	absorber
19	solenoid
22	rf cavity
9	repeat tag

Table 9: PATH elements used in this simulation. A detailed description can be found in [53].

The factor Π is a very good figure of merit for channels with particle losses, but for the above mentioned reasons it can only be used for one specific problem at at time. Comparing different scenarios is not possible, because $\varepsilon_{\mathrm{acc}}$ would not be matched. Therefore, Π is not used in this study.

5.4 PATH

All particle tracking in this study has been carried out with PATH [53]. PATH is a multi particle tracking code based on TURTLE which has been developed at Los Alamos. PATH has been originally created for Linac design and has gradually been expanded at CERN for the simulation of cooling channels. Each particle is represented by 6 coordinates $(x, x', y, y', E_{\mathrm{tot}}, \varphi)$ plus the rest mass m_0 and the charge q as well as a "dead particle" flag. All particles are tracked one step (the step size is variable) and then collective effects (such as space charge) are calculated. Then all particles are tracked through the one more step and so on.

The geometry is given by aligning elements one after each other in an input file. The elements that have been used in this simulation are listed in Table 9. Each element is represented by a numeric code followed by the respective parameters. There is one element per line in the input file.

PATH works in two modes: fast mode and field map mode. In the fast mode, the magnetic and electric fields are calculated in ideal approximations. In this mode, only the E_z component is calculated for rf cavities and the hard edge approximation is used for the magnetic field of the solenoid. To get a more accurate simulation, the field map mode uses magnetic and electric

Part of the cooling channel [MHz]		44+88	44	88
ε_{r}	radial emittance [π cm mrad]	1122	2404	1122
ε_z	long. emittance [π deg MeV @ 44 MHz]	515	406	515
T	Transmission	0.525	0.97	0.84
Λ_4	4-D cooling factor	7.78	1.7	5.92
Ξ_4	4-D phase space density increase	3.93	1.6	4.4
Ξ_6	6-D phase space density increase	2.55	1.4	2.8

Table 10: Results of the baseline simulation using the simulated beam method.

field maps created in other programs such as POISSON. This mode is much slower.

In this study, only the fast mode has been used because of the high number of individual runs that have been necessary. Real field maps for solenoids always contain nonlinearities that – according to Liouville's theorem – are responsible for emittance growth. The results of this study can be seen as best case limits, with respect to the focusing effects.

5.5 Baseline Simulation

As a base for any comparison, a baseline simulation has been performed with both the simulated beam method and the simulated particle method. The latter is for reference only because it is not comparable to the other results. Because of this, the results from the simulated beam method are displayed in grey in the graphs.

The three parts of the cooling channel can be identified in the graphs. Cooling I ranges from 0 to 44 m, the acceleration to 76 m and Cooling II to 184 m. The most striking thing in Figures 24-26 are the "bumps" at 76 m. Here, the frequency is doubled from 44 to 88 MHz and nearly half the beam is lost. This is a purely longitudinal phenomenon and has nothing to do with cooling. To better separate the effects of frequency doubling from cooling, the results of this study have been analysed for three regions: the whole channel, Cooling I and Cooling II. Between Cooling I and Cooling II, there is the acceleration and the frequency doubling. This is the reason why the results for Cooling I and Cooling II do not add (or multiply) up to the whole channel in Table 10.

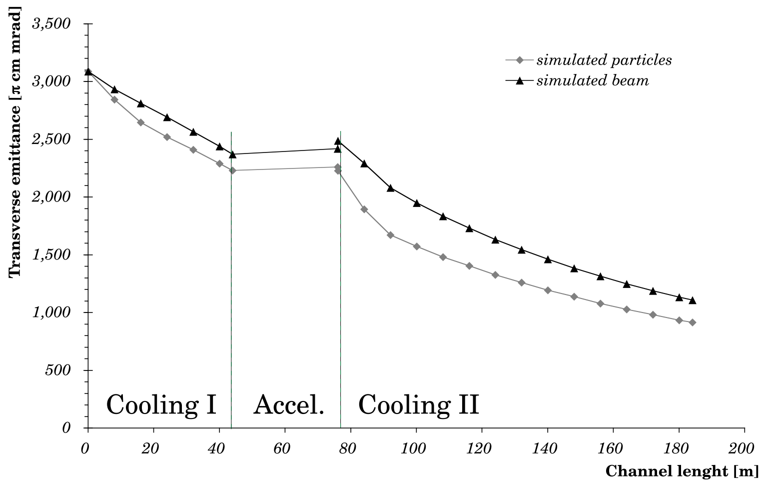

Figure 23: The transverse 2D emittance ε_r versus the length of the cooling channel in the baseline simulation.

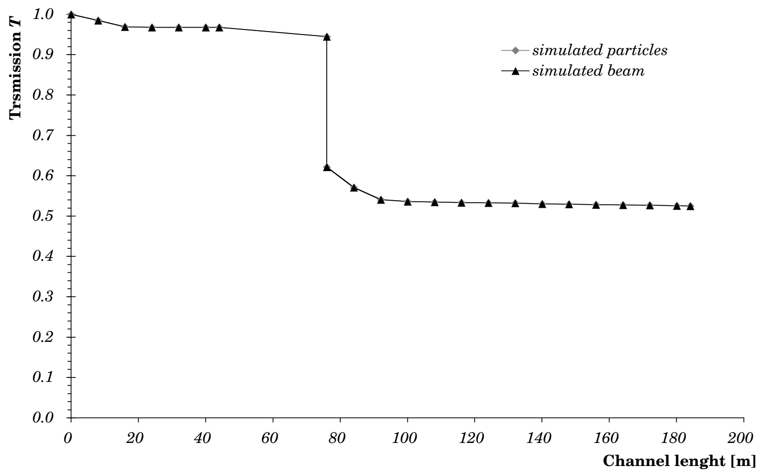

Figure 24: The transmission T versus the length of the cooling channel in the baseline simulation. The losses at the frequency doubling can be seen at 76 m.

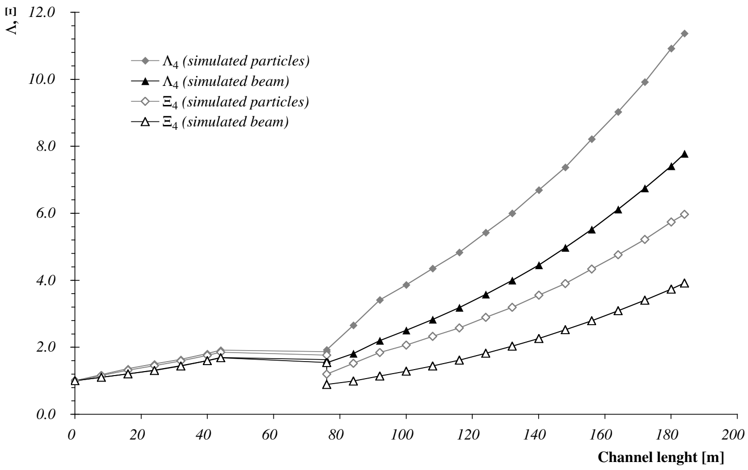

Figure 25: The 4D transverse cooling factor Λ_4 and the phase space density increase Ξ_4 versus the length of the cooling channel in the baseline simulation.

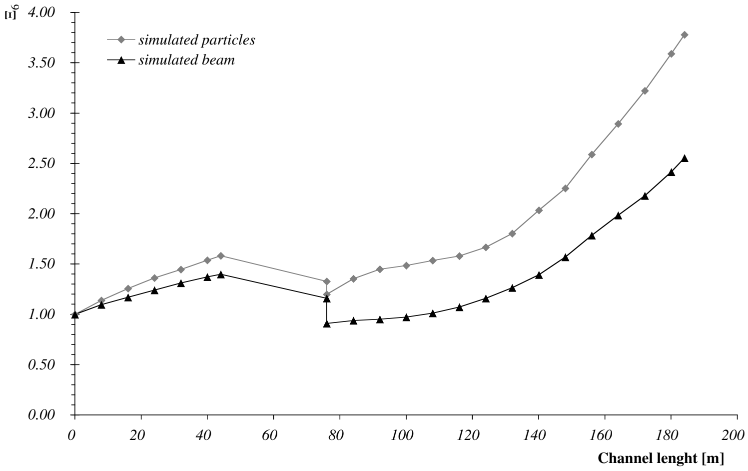

Figure 26: The 6D phase space density increase Ξ_6 versus the length of the cooling channel in the baseline simulation.

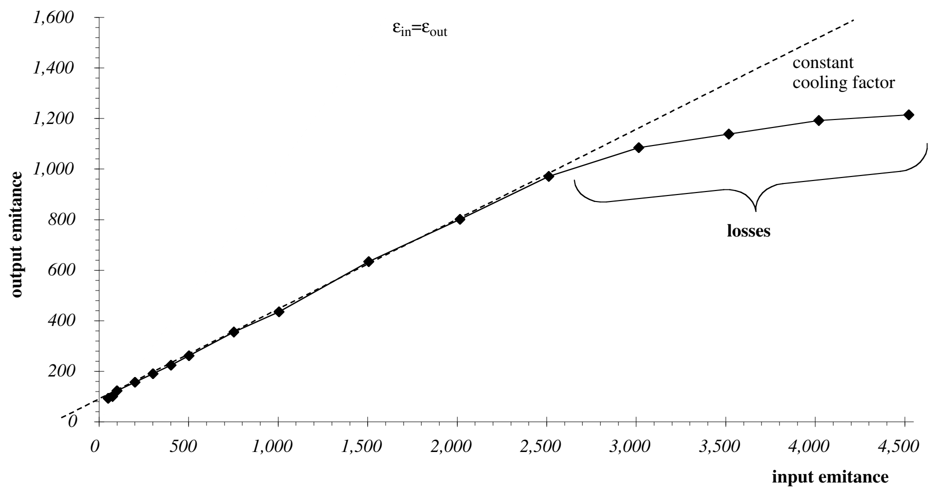

Figure 27: Output emittance versus input emittance; We can distinguish three regions: the heating region for very low $\varepsilon_{\rm in}$, then the linear region and finally the limit of the transport capacity of the cooling channel. The flat end of the curve for high $\varepsilon_{\rm in}$ is not a cooling phenomenon, but is due to beam losses, because the emittance is larger than the acceptance of the channel. The units are π cm mrad normalised emittance.

5.6 Results I: Change of Input Variables

5.6.1 Change of ε_{in}

A series of simulations with various rms $\varepsilon_{\rm in}$ has been performed. The reference value is $\varepsilon_{\rm r,in} = 3089\,\pi$ cm mrad. Simulations were run with values for $\varepsilon_{\rm r,in}$ of 50, 75, 100, 200, 300, 400, 500, 750, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500 and 5000 π cm mrad. The input variables were the same as in the reference run, with the exception of the input transverse emittance. The results are shown in Figure 27.

Interpretation. The low equilibrium emittance shows that multiple scattering is of no concern in the parameter range of the CERN 44/88 MHz channel. $\varepsilon_{\rm out}/\varepsilon_{\rm in}$ is basically a straight line which is only limited by the acceptance of the solenoidal transport channel. Yet, going to larger emittances gives a better cooling factor as the flat influence of the multiple scattering is divided by a higher emittance.

dE/dx is flat.

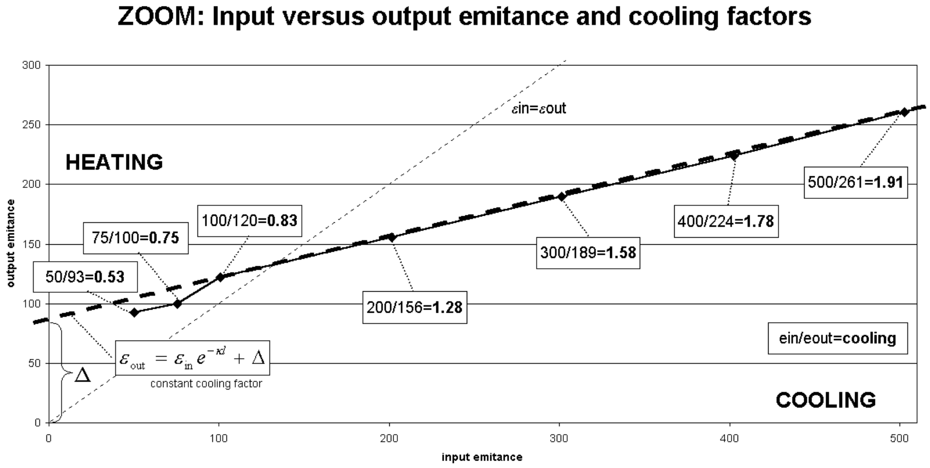

Figure 28: Zoomed in view of Figure 27. The values in the boxes are $\Lambda_2 = \varepsilon_{\mathrm{in}}/\varepsilon_{\mathrm{out}}$.

5.6.2 Change of input Δ E

Simulations have been carried out with ΔE_{in} equal to 0, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 MeV. The reference value is $\Delta E_{\mathrm{in}} = \pm 33\,\mathrm{MeV}$. It was shown that the cooling factor does not change as a function of the energy spread, but the overall transmission, is a strong function of the energy spread (see Figure 30).

Interpretation. As one would expect, a longitudinal change (the ΔE) does not influence transverse cooling. That is why from a cooling point of view, the energy spread has little influence. Beam transport, however, gets more difficult with an increasing energy spread as particles tend to fall out of the already full rf bucket, if the ΔE is too large. The reason for this is the moderate energy ($E = 400\,\mathrm{MeV}$) of the particles, which translates in a relativistic β of about 0.95. This means that particles with different energies still have different speeds.

Explain losses between 2 and 2a.

5.6.3 Change of input $\Delta\varphi$

Simulations have been carried out with phase spreads of ± 5 to $\pm 100°$. The cooling factor is virtually unaffected by the phase spread, while the 4D

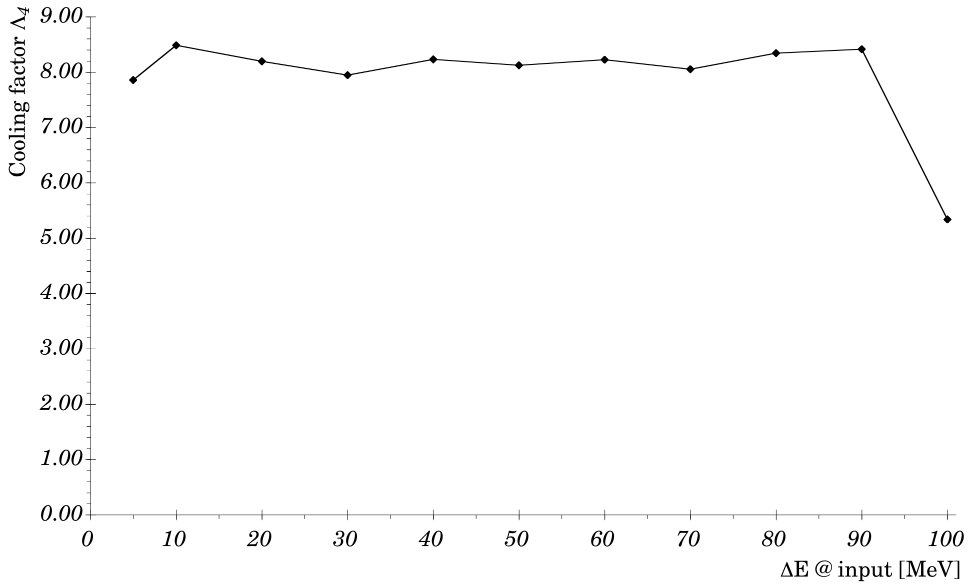

Figure 29: Cooling factor versus ΔE_{in}

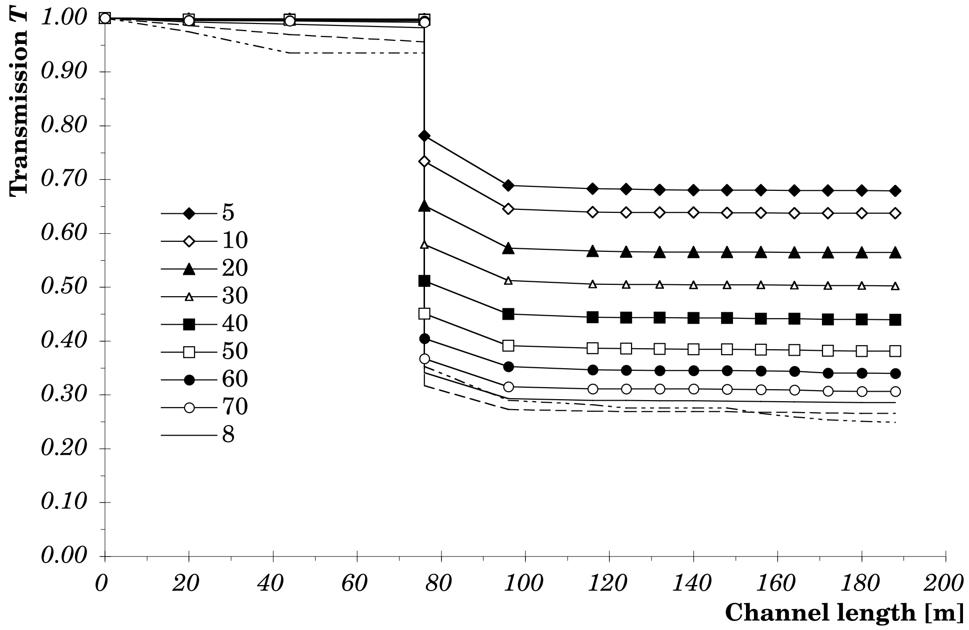

Figure 30: Transmission versus ΔE_{in}

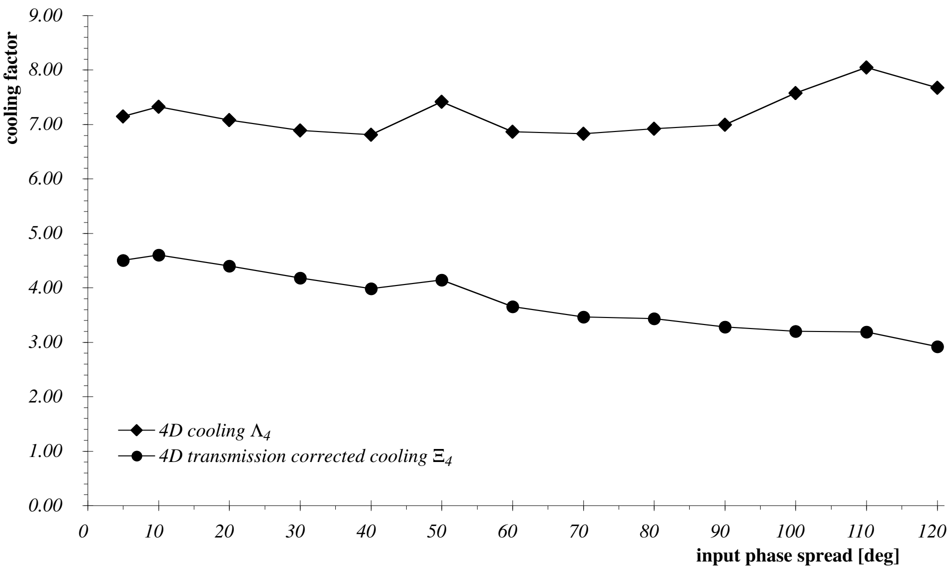

Figure 31: Cooling factor and 4D phase space density increase versus $\Delta\varphi$

density increase suffers from longitudinal losses.

Interpretation. Again, a change of longitudinal parameters does not affect the transverse cooling. The big effect of the $\Delta\varphi$ on T comes from the frequency step from 44 to 88 MHz, where, depending on the bunch length, up to 70% of the particles are lost longitudinally.

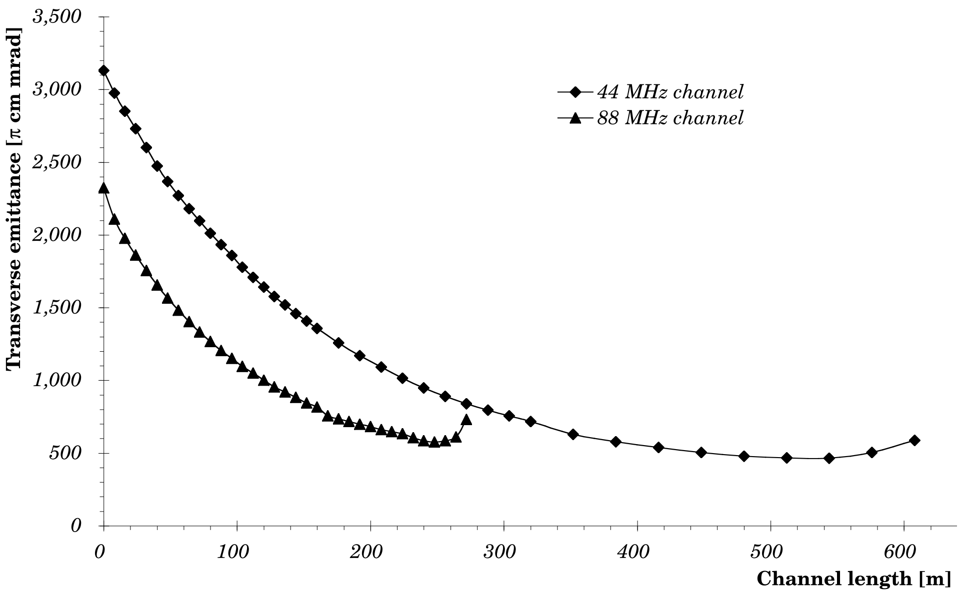

Figure 32: Change of length – ε_r. Both parts of the cooling channel cool well below the values that are needed at the neutrino factory.

5.7 Results II: Change of Channel Parameters

5.7.1 Change of the Length l

The length of each part of the cooling channel has been changed. The 44 MHz part has been tracked through 608 m; the 88 MHz part has been tracked trough 272 m. In the 44 MHz part, the equilibrium emittance has been found at 512 m and in the 88 MHz part at 248 m. Subsequent emittance growth as predicted in [54] has been observed.

Results. Both parts of the cooling channel cool well below the values forseen for the neutrino factory. If more cooling is needed, it can be provided by making the 88 MHz channel longer.

5.7.2 Change of the Absorber Material

The following absorber materials have been tested: H, LiH, Li, Be, C, Al for the second part of the cooling channel (88 MHz), starting from a transverse emittance $\varepsilon_r = 2600\pi$ cm mrad. The beam has been tracked through 112 m of cooling channel.

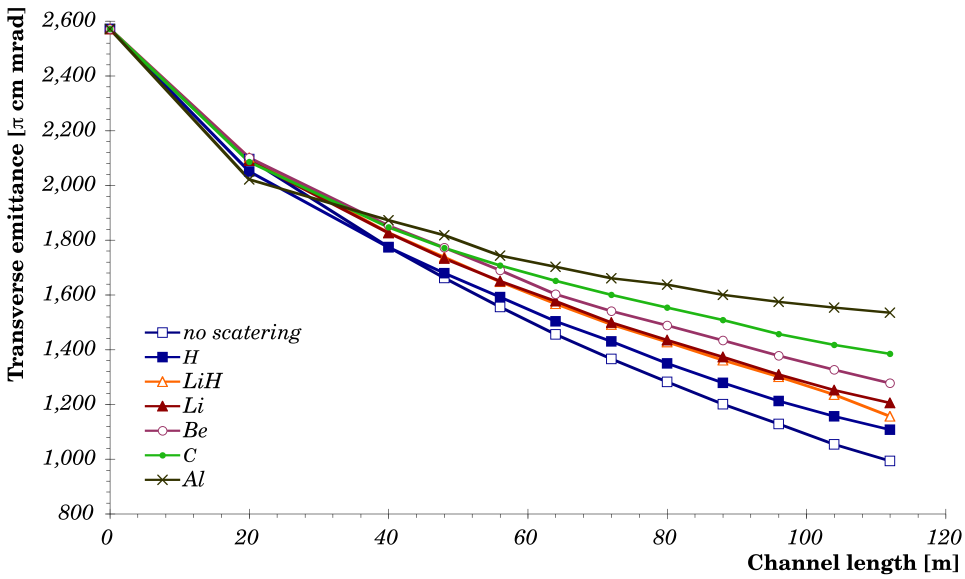

Figure 33: Cooling for various absorber materials

Interpretation. The surprising result of this test is the fact that the absorber material and the multiple scattering is not so important as often thought. While there are still good reasons for hydrogen as an absorber, materials such as LiH, Li and Be, maybe even C can be considered (at least for the first part of the cooling channel). This leaves some alternatives to a liquid absorber.

There is an alternative way to read Figure 33: 95 m of coolig channel with H cool just as well as 105 m of cooling channel with LiH. Now it is possible to weigh the difficulties associated with liquid hydrogen absorbers with the cost of 10 m extra meters of cooling channel.

Chapter 6

FASTCOOL - a Proposed Cooling Experiment

6.1 A Cooling Experiment

Ionisation cooling is one of the most compelling parts of a possible future neutrino factory [33]. Although ionisation cooling has been intensely discussed, it has never been shown experimentally. There is no known experiment of ionisation cooling with muons and there has been just one experiment with friction cooling at very low energies[2] [55].

A generic cooling experiment consists of the following four components:

- Muon source. There are several laboratories in the world that provide muon beams. It is much simpler to move a possible experiment to an existing beam than to construct a muon source at the experiment's location. Most notable muon sources are PSI in Villigen (CH), RAL in Didcot (UK) and TRIUMF in Vancouver (CAN).
- Beam definition. As discussed in Chapter 4, cooling works only for high ε_{r}. No muon source exists that provides beams with sufficiently large emittances. The beam has to be degraded ("heated") in a layer of high Z material, profiting from multiple scattering.

[2]Friction cooling is based on the same principle as ionisation cooling, but the absorber is solid. The lightest practical solid absorber material is carbon, which has been used for this experiment. However, it was carried out at muon energies of few keV, this is five orders of magnitude away from the energies under consideration for a neutrino factory.

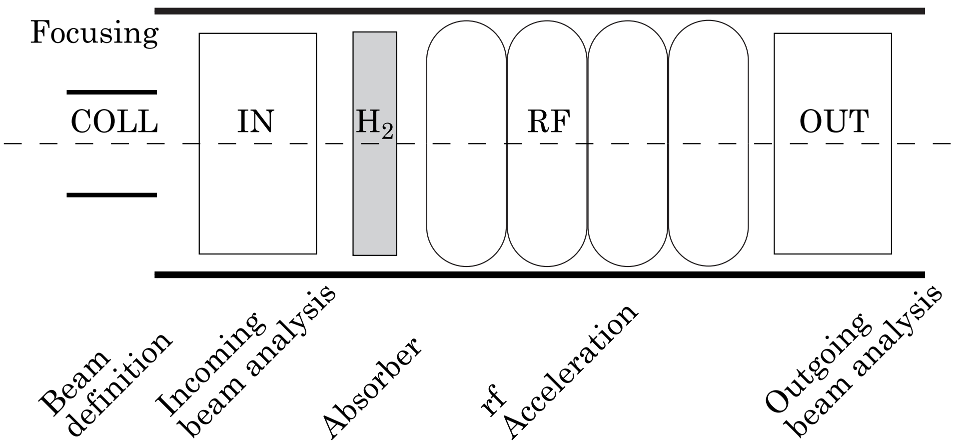

Figure 34: A generic cooling experiment consists of beam definition to increase the emittance, incoming and outgoing analysis as well as the cooling device made of absorber and rf acceleration.

- Incoming and outgoing beam analysis. To prove cooling, the emittance (or particle density) has to be measured before and after the experiment. The two measurement units can be identical as the beam parameters do not change significantly between input and output.
  An alternative design foresees a single analysis section for the outgoing beam and two measurements with the cooling apparatus turned on and off [56].

- Cooling apparatus. The cooling apparatus is the central piece of each cooling experiment and consists of one or more absorbers and rf cavities, as discussed in Chapter 4.

Based on the 88 MHz part of the CERN cooling channel a rough estimate of the dimensions of a cooling experiment can be given: to achieve a $\approx 10\%$ emittance reduction, 2 cooling cells are needed. The cooling apparatus alone would be 8 m long, plus at least 2 m of length for the incoming and outgoing beam analysis. The power requirements would be severe, building such an experiment a major undertaking.

At this point, the question can be asked: "Why do we need a cooling experiment [57]?" The four phenomena involved in ionisation cooling described in Chapter 4.3 are all well known or known. There are ongoing efforts to simulate ionisation cooling. Still, there are several arguments for building a cooling experiment:

- to prove the principle ionisation cooling
- to prove cooling and to check simulations against the experiment
- to prove that a certain apparatus performs ionisation cooling on a range of muon beams and to check the simulations
- to measure the cooling factor Λ_2 and the transmission T as a function of the incoming beam and the apparatus setup
- to prove that a baseline design works, to measure cooling factor and transmission under various conditions and to compare the results to simulations

The arguments have been ordered by difficulty. It is not the scope of this thesis to decide which of these questions should be answered. But ***if*** there is a desire to build a small and simple cooling experiment, that just proves the principle of ionisation cooling, ***then*** the FASTCOOL proposal could meet it at moderate cost.

If the requirement is to plan a small experiment only to prove the principle of cooling, then the question must be: "How can one downsize a cooling experiment?"

6.2 Downsizing a Cooling Experiment

It has been a tradition in high energy physics to build a small and simple test device in order to explore a new concept. Unfortunately, downsizing a cooling experiment is impossible without major constraints to its validity. There are a number of physics limits to downsizing a cooling experiment:

Length. As shown in Equation 46, cooling is a small effect. Without heating we calculate for relativistic particles ($\beta \approx 1$): $\Delta\varepsilon/\varepsilon = \Delta E/E$. For 300 MeV muons, this is $\Delta\varepsilon/\varepsilon = 0.33\%/\mathrm{MV_{acc}}$. With a measurement precision of 2% [58], a 5σ effect could be shown with 30 MV of accelerating gradient, which means 7.5 m of cavities at 4 MV/m plus 1 m of hydrogen absorber.

Transverse emittance. As the beam size is $x_{\max} = \sqrt{\varepsilon\beta_\perp}$, at least the diameter of the device could be reduced by running with a smaller emittance.

Unfortunately, cooling works best at higher emittances and does not work at all below the equilibrium emittance.

Focusing. Another way to reduce the transverse size of the apparatus would be a smaller β-function, which means stronger focusing, as $x = \sqrt{\varepsilon\beta_\perp}$. This way cannot be exploited, however, as the focusing fields in the 88 MHz part are already 5T and stronger fields are extremely difficult to produce [59]. If it were possible, there would be an additional advantage due to the lower influence of multiple scattering at low $\beta_\perp$.

Muon energy. One way to reduce the size of a cooling experiment is lowering the energy of the muons. Thus, one can win twice: as shown in Equation 46, the cooling factor is higher for low-energy muons and on top of that Equation 34 shows that focusing works better for lower energies[3].

Energy reduction is a valid scenario for a smaller cooling experiment, but the validity of such an experiment for the final neutrino factory is limited.

Detector resolution. The best way to downsize a cooling experiment is to increase the detector resolution. In the absence of systematic errors, the size of the needed effect is proportional to the measurement resolution.

6.3 FASTCOOL Design Principles

FASTCOOL has been designed to be the smallest possible muon cooling experiment in the energy range of $p = 150\,\mathrm{MeV/c}$. Its design is based on the fact that cooling is always x' reduction. If the whole experiment is put in a drift region ($d\beta_\perp/dz = 0$), it is sufficient to measure a reduction of x' to prove emittance reduction:

$$\begin{aligned} \varepsilon &= \beta_\perp x'^2 \\ \frac{d\varepsilon}{dz} &= x'^2 \frac{d}{dz}\beta_\perp + \beta_\perp \frac{d}{dz} x'^2 \end{aligned} \tag{61}$$

The first term in the above equation is zero in a drift region. The idea is to measure x' before and after the cooling apparatus. This is done with

[3] This is the reason why the only known muon cooling experiment [55] was carried in the energy range of a few keV.

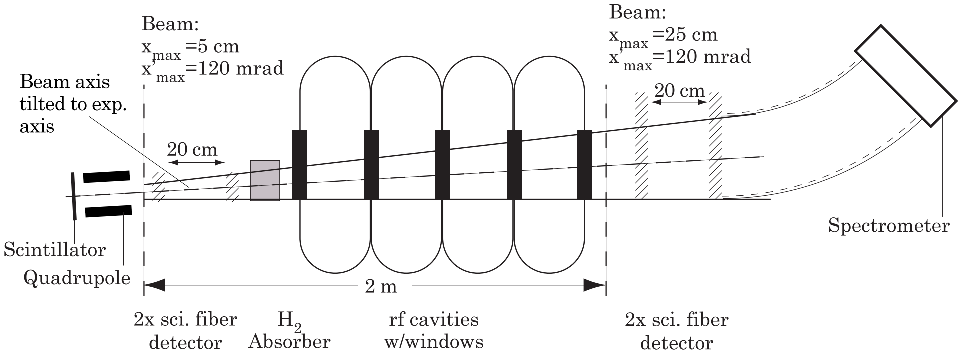

Figure 35: Experimental setup for the FASTCOOL proposal. The core experiment is only 2.2 meters long, thus making focusing unnecessary. The beam enters the apparatus off-axis, thus simulating the upper half of a beam with twice the emittance.

two pairs of detectors that measure x, profiting from the fact that in a drift region $x = x_0 + x'd$ and $x' = \frac{\Delta x}{d}$.

To make FASTCOOL as small as possible, the following design principles have been employed:

- Precision. Detector precision is traded for the size of the effect. The detectors are outlined in Section 6.5.2

- Single particle method. To make best use of the fine detectors and to increase the precision, each particle is detected separately and the emittance is reconstructed in the computer using the particle coordinates. The single particle method is explained in Section 6.5.1

- Only two-dimensional. The experiment measures cooling only in one plane, the x-plane. This makes quadrupole defocusing possible and makes FASTCOOL compatible with existing muon beams without degrader.

- Low $\beta_\perp$. Instead of using a diffuser to increase the emittance, FASTCOOL uses a quadrupole to produce a low-$\beta_\perp$ region.

- Tilted half beam. To pretend the large x' needed to prove cooling, the beam is sent through the experiment tilted at half its divergence. Like this the beam behaves like half a beam with twice the x'. Only the positive x-coordinated are measured. Using only half a beam is dangerous if the systematic errors are not under control, but gains

total length	<3 m
total width	<3 m
p_μ	150 MeV/c
rf voltage	6 MV
rf gradient	3 MV/m
rf frequency	46 MHz
cavity bore diameter	50 cm
absorber	19 cm H_2
$rms\, x'_{\text{in}}$	122 mrad
$rms\, x'_{\text{out}}$	125 mrad
resolution	0.3 mrad
x'_{max}	250 mrad

Table 11: Parameters of FASTCOOL

another factor of 2 in size. The tilted half beam method is explained in Section 6.5.3.

6.4 Experiment Setup

The setup is shown in figure 35. The beam is defocused in a long quadrupole magnet, passes through a pair of scintillating fibre detectors, then a 20 cm long liquid hydrogen absorber and four 50 cm long rf cavities with 3 MV/m and the second pair of scintillating fibre detectors. At the end, the particle's momentum is measured with a spectrometer.

Beam line. The experiment is based on a beam line providing 150 MeV/c muons with a momentum resolution of 1% or better. The rate must be compatible with the speed of the detectors, about 1 kHz. This design is based on an *rms* emittance of the muon beam of $\varepsilon = 300\,\pi$ cm mrad.

Beam definition. The beam is defocused in a 24 cm long quadruople magnet with a field of 0.9 T and an aperture of 7 cm. This increases the $rms\ x'$ to 125 mrad. With an acceptance x'_{max} of 250 mrad, 2σ of the beam distribution are accomodated.

Incoming beam anaylsis. The incoming beam analysis is performed with a pair of scintillating fiber detectors separated by a 40 cm drift. They have a resolution of 120μm. Using $x' = \frac{\Delta x}{d}$ one can calculate an angular resolution of 0.3 mrad. A pair of detectors adds 0.24% of a radiation length to the scattering. This is low compared to the 2.3% of a radiation lenght of the absorber.

Hydrogen absorber. The liquid hydrogen absorber is 19 cm wide. It is contained in a cryostat with 300 μm thin Al windows.

Accelerating cavities. The cavity is based on a design study by W. Pirkl [60]. It has a frequency of 46 MHz and a bore diameter of 50 cm to accomodate the beam. The outer diameter is 220 cm. Each of the four cavities is 50 cm long with a 16 cm long gap. A 50 μm window of Al makes it possible to achieve a field uniformity of better than 0.5% even for this geometry. Each cavity has a peak power requirement of 3.5 MW.

Outgoing beam analysis. The outgoing beam analysis is based on the same detectors as the incoming beam analysis, they are just wider as the beam defocuses.

Spectrometer. At the end of the experiment there is a spectrometer magnet to measure the particle's momentum which is needed for the $\beta\gamma$ factor in the normalized emittance.

The experiment itself is in a drift region. The defocusing occurs before the first x' measurement is carried out. Unlike the CERN channel, there are no solenoidal fields in this experiment. This means, that there is no coupling between x and y and no orbit dispersion. This is the basis for making it possible to use only one coordinate and only half a beam.

6.4.1 Experiment Simulation

The experiment has been simulated with PATH with a 10.000 particles sample. The simulation is based on perfect cavities. Multiple scattering in the absorber, the Al windows and the detectors have been taken into account. The results of the simulation are summed up in Table 12.

	input	output
emittance	$303\,\pi$ cm mrad	$298\,\pi$ cm mrad
$rms\, x'$	125 mrad	122 mrad
cooling factor Λ_2	1.6%	

Table 12: Results of the FASTCOOL simulation with PATH.

6.4.2 Potential of FAST COOL

FASTCOOL has the following advantages:

- Fast and comparatively easy to achieve proof of ionisaton cooling for 150 MeV/c muons.
- Comparison of simulation and experiment for *one* specific case of ionisation cooling.
- Possibility to gain experience with some of the machinery involved in the CERN cooling channel, namely low frequency rf cavities and liquid hydrogen absorbers.

6.5 Methods and Technologies for FASTCOOL

6.5.1 The Single Particle Method

To make use of a precise detector, it is mandatory to use the single particle method. This means that only one particle at a time passes through the apparatus. The particle's track is reconstructed using standard methods of particle physics. In the case of FASTCOOL the reconstruction is very simple, as there is no electric of magnetic field between the two incoming and outgoing beam detectors.

The full beam is reconstructed from a number of particle tracks. With this method it is possible to calculate the emittance following its statistical definition, eliminating the detour of beam profiles. With the single particle method, one can achieve far greater resolutions than with any collective emittance measurement with a full beam.

One limitation to the single particle method is the requirement that the beam and the experimental apparatus remain stable during the data taking period. With typical data rates of about 1kHz and statistic samples of

10.000 to 1.000.000 particles, this data taking period is between 10 s and 16 min long.

A part from the above limitation, there are some extra possibilities that are opened by the single particle method:

Software collimation. As the beam is reconstructed from single particle tracks in the computer, it is possible to use only particles that lie within a certain emittance and to ignore (=collimate) the remaining ones. With this technique it is possible to analyse the behaviour of the cooling apparatus for various input emittances in a single measurement.

Various beam distributions. By changing the weight of individual particles it is furthermore possible to simulate various beam distributions like rectangular or Gaussian distributions.

Couplings. Couplings can be identified by analyzing projections with software that are normally not accessible with beam monitors, like the $x - y'$ plane.

6.5.2 Scintillating Fiber Detectors for FASTCOOL

A detector for a cooling experiment must fullfill two demands: high tracking precision and low scattering. The detector proposed for FASTCOOL is a scintillating fiber detector based on the detector for MUSCAT [50] and the detector proposed for the CERN cooling test facility [61].

As only the x coordinate is of interest for this experiment, the detector is made of just one layer of square scintillating fibers that are 0.5 mm thick. This reduces the number of channels and the scattering in the detector. The detector is based on the scintillation light emitted when a muon passes through a fiber. This light is collected and transported to the edges. Clear fibers are attached to transport the light to the read out electronics. As optical fibers are nearly 100% transparent, it is possible to transport even only a few photons over meters. This makes it possible to put all the read-out electronics in a "safe place", away from magnetic fields and radiation sources.

The desired active area of the detector is a square. The length of a side depends on the detector position: 12.5 cm for the first, 22.5 cm for the

Detector Geometry	square
Scintillator	0.5×0.5 mm fiber
Spatial Resolution	$\approx 120\ \mu$m
Time Resolution	500 ps
$X_{\rm det}/X_0$	$\approx 0.12\%$
Vacuum Compatible	Yes

Table 13: Characteristics of the proposed scintillating fibre detector

second, 50 cm for the third and 60 cm for the last detector. These sizes are comparable to the existing 30 cm diameters of the MUSCAT detector.

The readout is perfomed through a multianode photomultiplyer tube (PMT). The signal from the PMT is amplified and digitized in a VME card. The electronics has not been designed yet, but there are designs for similar purposes in [50] and [61].

6.5.3 The Tilted Half Beam Method

FASTCOOL has been designed to use just one half of a highly divergent beam. The high divergence is needed to produce a significant effect compared to the detector resolution, as in the case of FASTCOOL $\Delta E/E = \Delta\varepsilon/\varepsilon \propto \Delta x'/x'$. To get a big $\Delta x'$, a big x' is needed. This would make the transverse dimension of the experiment large, as $x = x_0 + x'd$. The solution is to use only one half of the beam with positive x-coordinated. As the setup is symmetric, no information except about systematic errors is lost. Here, a careful setup and surveying are needed to compensate the loss. Following an idea of H. Haseroth [56], there is a way to make the setup simpler: instead of using half a beam of x', FASTCOOL uses a whole beam of $x'/2$, tilted by $x'/4$. The result is exactly the same, but without losses. A sketch of the tilted half beam is shown in Figure 35.

Chapter 7

Conclusions

Physics case for a neutrino factory. The physics case for a neutrino factory has been discussed in Chapter 2. There is a strong case for a neutrino factory, as three out of the four main questions in neutrino physics can be answered with such a machine.

Increased intensity. The CERN baseline scheme for a neutrino factory has been described in Chapter 3. An analysis of the time structure lead to two ways to increase the intensity by up to a factor of 8. A factor of two can be gained by foreseeing two instead of one accumulator and compressor rings — this would double the length of the muon beam and thus double the intensity. More difficult would be an increase in the repetition rate from the current 50 Hz towards the physics limit of 200 Hz — this would increase the intensity by another factor of up to four.

Cooling theory. A new derivation of the cooling rate has been give in Chapter 4. The new approach starts from the microscopic view, that cooling is always x' reduction and gives the same results as the previous calculations.

Sensitivity study. The CERN scheme has been studied for sensitivity with respect to various beam variables and channel parameters in Chapter 5. The results in short:

- Couplings between the longitudinal and transverse plane have been found to be very low. This is manifested in a low dependence of the transverse cooling performance of longitudinal parameters.

- Changes of the beam variables that are caused by cooling phenomena have been separated from changes caused by other phenomena. The shift from 44 to 88 MHz has been identified to be the largest source of beam losses. This frequency doubling is also sensitive to changes in longitudinal input parameters.

Alternatives to hydrogen. The CERN scheme for the neutrino factory would work with a few solid absorber materials, as well. Various absorber materials have been tested and it has been established that LiH, Li and Be are viable alternatives to H as an absorber material. The solid materials would imply an $\approx$10% increase in length of the cooling channel, which has to be weighed against the complexity of a cryogenic liquid absorber.

Cooling experiment. A simple first cooling experiment, called FAST-COOL, has been proposed and simulated in Chapter 6. This experiment could prove the principle of ionisation cooling at a comparatively low cost.

Appendix A

Decay Ring Geometry and Baseline Efficiencies

A.1 Decay Ring Efficiency

The figure of merit of a neutrino factory is the number of neutrinos per year that actually reach one of the detectors. This number depends on the number of muons in the decay ring, the opening angle of the neutrino beam and the geometry of the decay ring.

The influence of the decay ring on the performance of the neutrino factory can be expressed using the efficiency η_{DR}. The muons in the decay ring have very high energies ($\gamma \approx 500$), this means their decay length is much longer than the circumference of the decay ring. The probability for a muon to decay is the same everywhere in the decay ring in this case. Muons that decay in a long straight section pointing towards a detector give "useful" neutrinos, muons decaying in a short straight section or in a bend give "lost" neutrinos. The efficiency η_{DR} can be defined with the length of a straight section l and the circumference c as:

$$\eta_{\mathrm{DR}} = \frac{2l}{c} \tag{62}$$

If a long straight section for a baseline length L points into the earth, it must be tilted by an angle τ.

$$\tau = \arcsin\frac{L}{r_e} \tag{63}$$

Here, r_e denotes the radius of the earth. If both straight sections in a decay ring point to a detector, they both point into the earth and the angle α^*

Figure 36: A large number of baselines for the CERN neutrino factory have been studied. The current favourites are shown in full lines.

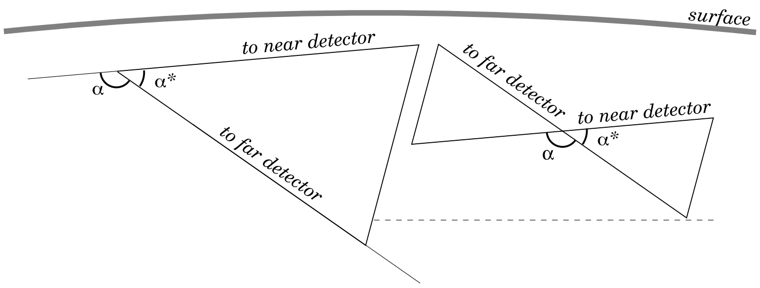

Figure 37: Geometry of a bow-tie and triangular decay ring. One of the advantages of the bow tie is the lower depth it requires.

between them is at least the sum of $\tau_1 + \tau_2$. The exact value of α^* depends on two factors: apart from the two baseline lengths on the geography of the detector locations. If they are exactly opposite, $\alpha^* = \tau_1 + \tau_2$. If the angle between the directions (in the plane) is less than 180°, then α^* becomes larger.

In a first step, only a triangular storage ring is considered. In the asymptotic case, the bendings can be neglected. With α^* the asymptotic efficiency η_{as} for two baselines can be given as:

$$\eta_{\mathrm{as}} = \frac{2}{2 + \sqrt{2(1 - \cos\alpha^*)}} \tag{64}$$

The efficiency is best for low values of α^*. For a given set of baseline lengths, α^* is lowest for exactly opposite directions from the lab to the detector locations.

In fact there are two scenarios to close the ring: a triangle or a bow-tie (see Figure 37). In the asymptotic case, they give the same efficiency, as a bow-tie consists of two similar triangles. If the bending radius r is taken into consideration, the two configurations give different efficiencies.

For a bow-tie, $\eta_{\bowtie}$ is

$$\eta_{\bowtie} = \frac{2l}{2l + r(2\pi + \alpha^*) + (l + 2r\tan\frac{\pi-\alpha^*}{4})\sqrt{2(1 - \cos\alpha^*)} - 4r\tan\frac{\pi-\alpha^*}{4}} \tag{65}$$

Whereas for a triangle, $\eta_{\triangle}$ is

$$\eta_{\triangle} = \frac{2l}{2l + r2\pi + (l + r\tan\frac{\pi-\alpha^*}{2} + r\tan\frac{\pi-\alpha^*}{4})\sqrt{2(1 - \cos\alpha^*)} - 2r\tan\frac{\pi-\alpha^*}{4}} \tag{66}$$

These functions are plotted in Figure 38. These two functions cross and therefore it is possible to calculate the angle α^* at which the triangle and the bow-tie have the same efficiency $\eta_{\triangle} = \eta_{\bowtie}$:

$$\tan\frac{\pi - \alpha^*}{2}\sqrt{2(1 - \cos\alpha^*)} = \alpha^* + \tan\frac{\pi - \alpha^*}{4}\sqrt{2(1 - \cos\alpha^*)} - 2\tan\frac{\pi - \alpha^*}{4} \tag{67}$$

Numerically we find $\eta_{\bowtie} = \eta_{\triangle}$ for $\alpha^* = 92.2756°$. Generally, a low α^* is desirable, because of Equation 64. Thus for all *sensable* neutrino factories, the bow-tie is better.

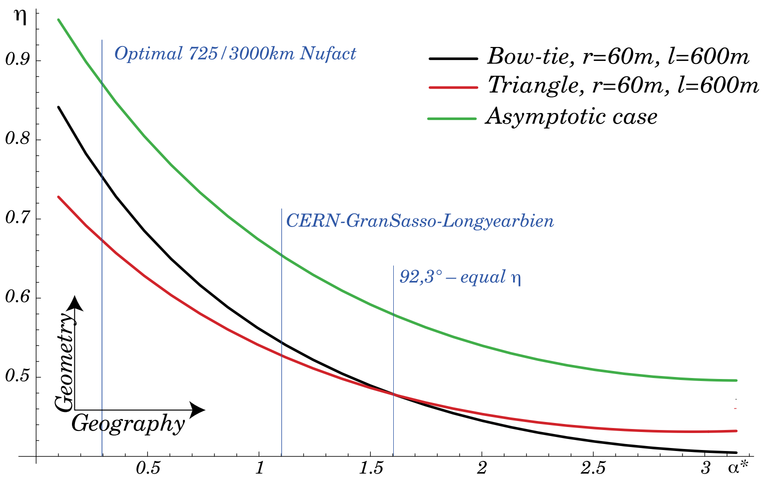

Figure 38: Efficiencies of ideal and realistic decay rings as a function of the angle α^* between the long straight sections.

Location	longitude	latitude	baseline L	inclination τ
CERN	46°45'N	6°48'E		
Gran Sasso	42°25'N	13°31'E	725 km	3.3°
Pyhaesalami	63°00'N	26°00'E	2225 km	10.0°
Santa Cruz	21°6'N	15°30'W	2731 km	12.4°
Longyearbyen	78°00'N	16°00'E	3504 km	16.0°

Table 14: Coordinates of CERN and the detector locations under investigation [62].

A.2 Geographical Circumstances

This ring has two long straight sections that aim to two detectors. There are three big high energy physics laboratories and three neutrino detectors in discussion for a neutrino factory with a short and a long baseline experiment. In the following, we shall discuss geographical implications on the decay ring design and on the efficiency of the decay ring. The coordinates of CERN and the detectors in this survey are shown in Table 14.

All calculations are based on a model of the earth with a radius r_e depending on the latitude ϕ (in degrees) as

$$r_e = r_0\Big(1 + \frac{|\phi| - 45}{90}\Delta r\Big) \tag{68}$$

with $r_0 = 6367.46$ km and $\Delta r = 21.36$ km. The sea level has not been taken

Near detector	Far detector	Baseline angle α	efficiency η	theoret. max. efficiency $\eta_{\max}$
Gran Sasso	Pyhaesalami	95.6°	59.8%	89.6%
Gran Sasso	Santa Cruz	106.5°	62.6%	88.0%
Gran Sasso	Longyearbyen	116.8°	65.6%	85.6%

Table 15: Scenarios of neutrino beams originating from CERN.

into account, as it is small compared to Δr.

The coordinates of the points have been translated into vectors in polar coordinates. The inclination of each of the straight sections relative to the earth surface and the angle between the two straight sections have then been calculated. If all three points (laboratory, near and far detector) are on a line, these three angles add up to 180°. This is the optimum case as the efficiency of the decay ring maximises.

A.3 Possible Scenarios

In line with the physics requirements, three possible detector pairs have been investigated. Gran Sasso is in each of these pairs as it provides the short baseline. For each detector pair, the angles between the baselines and the efficiency have been calculated. To allow for a better assessment, the theoretical maximum efficiency $\eta_{\max}$ is given in Table 15. The value $\eta_{\max}$ is the efficiency of a hypothetical detector pair with the two baseline lengths L_1 and L_2 and exactly opposite locations. The value of $\eta_{\max}$ can be calculated by combining Equation 63 and Equation 64:

$$\eta_{\mathrm{as}} = \frac{2}{2 + \sqrt{2\left(1 - \cos\left(\arcsin\frac{L_1}{r_e} + \arcsin\frac{L_2}{r_e}\right)\right)}} \tag{69}$$

Conclusions. A comparison between η and $\eta_{\max}$ in Table 15 shows that none of the studied options is a perfect solution. In all cases, more than 20% of the intensity is lost due to the bad geographical circumstances.

List of Figures

List of Tables

Bibliography

[1] N. Holtkamp, D. Finley (Ed.): A Feasibility study for a Neutrino Source Based on a Muon Storage Ring, Fermilab 2000

[2] B. Autin, A. Blondel, J. Ellis: Prospective study of muon storage rings at CERN, CERN Yellow Report 99-02

[3] G. I Budker in: Proc. of the 15^{th} Internat. Conf. on High energy Physics, Kiew (1970)

[4] A. N. Skrinsky: Intersecting Storage Rings at Novosibirsk, in Proc. of Morges Seminar (1971)

[5] C. Cowan, F. Reines et al: Detection of a Free Neutrino: a Confirmation, Science **124**, 3212 (1956)

[6] Figures in Radiation History, http://www.orcbs.msu.edu/radiation/radhistory/antoinebecquerel.html, accessed 2001-07-31

[7] W. Pauli, Open Letter to the Radioactive Ladies and Gentleman, (1930); (translated into English) Physics Today **31**, 27 (1978)

[8] T. Mayer-Kuckuk: Kernphysik, B. G. Teubner, Stuttgart 1992

[9] E. Fermi Z. Phys. **88**, 161 (1934)

[10] J. Chadwick, Nature **129**, 312 (1932)

[11] C. S. Wu et al, Phys Rev **105**, 1413 (1957)

[12] R. Feynman, M. Gell-Mann: Theory of the Fermi Interaction, Phys. Rev. **109**, 1 (1958)

[13] M. Goldhaber et al: Phys. Rev. **109**, 1015 (1958)

[14] R. Davis, D. S. Harmer, Buss. Am. Phys. Soc. **4**, 217 (1959)

[15] B. Pontecorvo, JETP **10**, 1236 (1960)

[16] Schwartz, Steinberger et al: Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos, Phys. Rev. Lett. **9**, 1 (1962)

[17] The LEP collaboration: A combination of preliminary electroweak measurements and constraints on the Standard Model, CERN/EP/ 2001-021, 61

[18] Y. Fukuda et al., Phys. Rev. Lett. **81**, 1562 (1998)

[19] V. M. Lobashev et al: Direct search for neutrino mass and anomaly in the tritium beta spectrum, Physics Letters B **460**, 227 (1999)

[20] Aleph Collaboration: An upper limit for the tau neutrino mass from three- and five-prong tau decays, Eurp. Phys. Journal C **2**, 395 (1998)

[21] DONUT Collaboration: Nucl. Phys. Proc. Suppl. **9**, (1999)

[22] A. de Santo: An experimentalist's view of neutrino oscillations, to be published in Intern. J. Mod. Phys. A (2001)

[23] The CHOOZ collaboration: Limits on Neutrino Oscillations from the CHOOZ experiment, PLB **466**, 415 (1999)

[24] J. Bahcall: Solar Neutrinos: Where We Are, Where We Are Going, Astrophysical Journal **467**, 475 (1996)

[25] J. N. Bahcall Rev. Mod Phys. **50**, 881 (1978)
J.N. Bahcall and M. Pinsonneault, Phys. Rev. **D58**, 6016 (1998)

[26] Review of particle physics, Eur. Phys J C **15**, 27 (2000)

[27] N. V. Mokhov: Particle Production for a Muon Storage Ring: I. Targetry and π/μ Yield, FERMILAB-Conf-00/208 (2000)

[28] M. G. Catanesi et al: Proposal to study hadron production for the neutrino factory and the atmospheric neutrino flux, CERN-SPSC P315 (1999)

[29] S. Van der Meer: A directive device for charged particles and its use in an enhanced neutrino beam, CERN-61-07 (1961)

[30] S. Rangod, private communication 2001

[31] S. Gilardoni: Review and comparison of solenoid and horn capture scheme, CERN-NUFACT 86 (2001)

[32] Review of particle physics, Eur. Phys J C **15**, 161 (2000)

[33] B. Autin et al: The CERN Neutrino Factory Working Group Status Report and Work Plan, CERN-OPEN 334 (2000)

[34] The SPL study group, M. Vretnar (Ed.): Conceptual design of the SPL – a high-power superconducting proton H^- linac at CERN, CERN 2000-012

[35] M. Poehler: Possible layouts of a Neutrino Factory at CERN, talk given at the Meeting of the Neutrino Factory Working Group on 2001-08-22

[36] A. Blondel, J. P. Fabre, S. Gilardoni, N. Vassilopoulos: A Thin Target Scheme for the Muon Source, CERN-NUFACT 6 (2000)

[37] A. Hassenein et al: An R&D Program for Targetry and Capture at a Neutrino Factory and Muon Collider Source, Preprint submitted to Elsevier Science

[38] J. M. Maugain, S. Rangod, F. Voelker: Technical Results of the Horn Study for the Nufact, CERN-NUFACT 80 (2001)

[39] F. Gerigk: Cavities and Solenoids for a Muon Cooling Channel, presentation given at the Workshop on Instrumentation for Muon Cooling Studies, Imperial College, London, 23-24. 02. 2001

[40] H. Schönauer, private communication 2001

[41] E. B. Holzer: Simulation of the pion decay channel of a neutrino factory, proceedings of NUFACT '01 in Tsukuba, Nucl. Inst. A, to be published

[42] P. Gruber, K. Hanke et al: The muon ionisation cooling experiment, to be published

[43] S. Gilardoni, K. Hanke, A. Perrin: Proposal for a four horn capture scheme, CERN-NUFACT 77 (2001)

[44] http://www.technicalf1.com/article.php?sid=186;
accessed 2001-09-05

[45] Computer codes for particle accelerator design and analysis, A compendium. Los Alamos, LS-UR-90-1766

[46] D. Möhl: Beam Cooling: Past, Present and Future, CERN-PS-2001-021-DR

[47] S. van der Meer: Stochastic Damping of Betatron Oscillations, CERN ISR-PO **72/31** (1972)

[48] Review of particle physics, Eur. Phys J C **15**, 163 (2000)

[49] H. Bethe: Moliere's Theory of Multiple Scattering, Phys. Rev. **89**, 1256 (1953)

[50] W. Allison et al: The Muscat proposal, RAL-TR-2001-015 (2001).

[51] A. Lombardi: A 40-80 MHz system for phase rotation and cooling, CERN-NUFACT 34 (2000)

[52] N.V. Mokhov, "The `MARS` Code System User's Guide", Fermilab-FN-628 (1995)
N.V. Mokhov, "`MARS` Code Development, Benchmarking and Applications" Fermilab-Conf-00-066(2000)
O.E. Krivosheev and N.V. Mokhov, "A New `MARS` and its Applications", Fermilab-Conf-98/43 (1998)
N.V. Mokhov, S.I. Striganov, A. Van Ginneken, S.G. Mashnik, A.J. Sierk and J. Ranft, "`MARS` Code Developments"

[53] A. Lombardi, A. Perrin, A. Berg: PATH handbook. Not published.

[54] G. Franchetti, Single Particle Description of Ionisation Cooling, CERN-NUFACT 051 (2000)

[55] M. Mühlbauer: Kühelung eines Strahls niederenergetischer Myonen durch Moderation und Beschleunigung: – Reibungskühlung –, Herbert Utz Verlag, München 1997

[56] H. Haseroth, private communication

[57] A. Tollestrup: Why do we need a cooling experiment?, Muon Collider Note MUC/NOTE/COOL_EXP/PUBLIC/0138 (2000)

[58] P. Janot: Emittance reduction measurement : update *with* cooling, presentation given at the CERN Neutrino Factory Working group meeting, 2001-06-14

[59] K. Hanke, private communication 2001.

[60] W. Pirkl, private communication 2000.

[61] P. Gruber, E. McKigney: A first study of a scintillating fibre detector for a muon ionisation cooling experiment, CERN-NUFACT 79 (2001)

[62] The Times atlas of the world; 9th ed., London 1992

Curriculum Vitae

Dipl.–Ing. Peter Gruber
geb. 16.03.1971 in Wien / Österreich

Schulbildung:

1977–1981	Volksschule Herbststraße, Wien XVI
1981–1989	Gymnasium Auf der Schmelz, neusprachl. Zweig, Wien XV
1989	Matura mit Auszeichnung

Studium:

1989–1998	Studium der Technischen Physik an der TU–Wien
1991–1995	Tutor am Institut für Experimentalphysik und am Institut für theoretische Physik, TU Wien
seit 1992	Studium der Philosophie an der Universität Wien
1997–1998	Diplomarbeit am Institut für Kernphysik, TU–Wien
1998–2000	Mitarbeit am EU-Projekt "ESPACE" (European System for Particle and Astrophysics Cyber Education)
1999-8/2001	Doctoral Student am Europ. Kernforschungszentrum in Genf / Schweiz (CERN)
seit 9/2001	Fellow am Europ. Kernforschungszentrum in Genf (CERN)

Militärdienst:

10/1997-6/1998	Abteilung für Leistungsmedizin und Wehrergonomie, Heeresspital, Wien