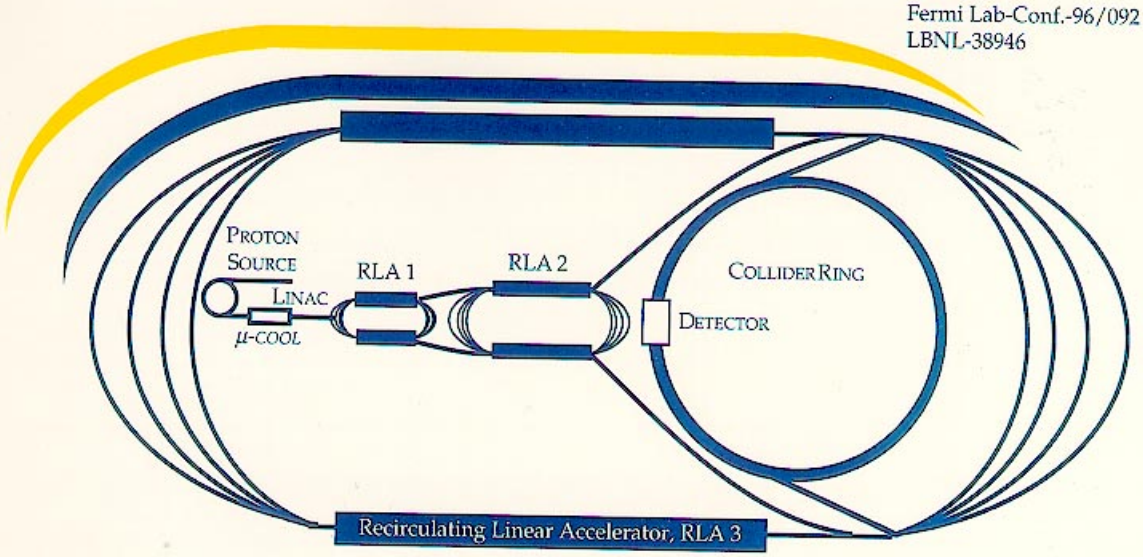
BNL-52503
Fermi Lab-Conf.-96/092
LBNL-38946



$\mu^+\mu^-$ COLLIDER
A FEASIBILITY STUDY

Collaborators
from the following Institutions

Argonne National Laboratory
Budker Institute of Nuclear Physics
Brookhaven National Laboratory
Columbia University
Deutsches Electronen-Synchrotron
Fairfield University
Fermilab
Indiana University
National Lab. for High Energy Physics, KEK
Lawrence Berkeley Laboratory
Stanford Linear Accelerator Center
State University of New York, Stony Brook
University of California, Berkeley
University of California, Davis
University of California, Los Angeles
University of Mississippi
University of Virginia
University of Wisconsin

July 1996

This research was supported by the U.S. Department of Energy under contracts DE-AC02-76-CH00016, DE-AC02-76-CH03000 and DE-AC03-76-SF00098.

MUON MUON COLLIDER: A FEASIBILITY STUDY

The $\mu^+\mu^-$ Collider Collaboration[1]

June 20, 1997

[1]This research was supported by the U.S. Department of Energy under contracts DE-AC02-76-CH00016, DE-AC02-76-CH03000 and DE-AC03-76-SF000098.

Contents

List of Figures

List of Tables

INTERLABORATORY AND UNIVERSITY COLLABORATION

Argonne National Laboratory, Argonne, IL 60439-4815, USA
Brookhaven National Laboratory, Upton, NY 11973-5000, USA
Fermi National Accelerator Laboratory, Batavia, IL 60510, USA
F. Bitter National Magnet Laboratory, MIT, Cambridge, MA 02139, USA
Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
Stanford Linear Accelerator Center, Stanford, CA 94309, USA
Center for Advanced Accelerators, UCLA, Los Angeles, CA 90024,USA
Columbia University, New York, NY 10027, USA
Fairfield University, Fairfield, CT 06430-5195, USA
Indiana University, Bloomington, IN 47405, USA
SUNY, Stony Brook, NY 11974, USA
University of California, Berkeley, Berkeley, CA 94720-7300, USA
University of California, Davis, CA 95616, USA
University of Mississippi, Oxford, MS 38677, USA
University of Virginia, Charlottesville, VA 22901, USA
University of Wisconsin, Madison, WI 53706, USA
BINP, RU-630090 Novosibirsk, Russia
KEK, Tsukuba-shi, Ibaraki-Ken 305, Japan
DESY, Hamburg, Germany

INSTITUTIONAL AFFILIATION OF SCIENTISTS IN THE COLLABORATION

R. Palmer	Brookhaven National Laboratory, Upton, NY 11973-5000
A. Tollestrup	Fermi National Accelerator Laboratory, Batavia, IL 60510
A. Sessler	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
A. Skrinsky	BINP, RU-630090 Novosibirsk, Russia
C. Ankenbrandt	Fermi National Accelerator Laboratory, Batavia, IL 60510
A. Baltz	Brookhaven National Laboratory, Upton, NY 11973-5000
V. Barger	University of Wisconsin, Madison, WI 53706
O. Benary	Tel-Aviv University, Ramat-Aviv, Israel
M. Berger	Indiana University, Bloomington, IN 47405
A. Bogacz	Center for Adv. Accelerators, UCLA, Los Angeles, CA 90024-1547
S. Caspi	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
P. Chen	Stanford Linear Accelerator Center, Stanford, CA 94309
W-H. Cheng	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
Y. Cho	Argonne National Laboratory, Argonne, IL 60439-4815
D. Cline	Center for Adv. Accelerators, UCLA, Los Angeles, CA 90024-1547
E. Courant	Brookhaven National Laboratory, Upton, NY 11973-5000
D. Ehst	¿Argonne National Laboratory, Argonne, IL 60439-4815
R. Fernow	Brookhaven National Laboratory, Upton, NY 11973-5000
M. Furman	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
J. Gallardo	Brookhaven National Laboratory, Upton, NY 11973-5000
A. Garren	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
S. Geer	Fermi National Accelerator Laboratory, Batavia, IL 60510
I. Ginzburg	Institute of Mathematics, Novosibirsk, Russia
H. Gordon	Brookhaven National Laboratory, Upton, NY 11973-5000
M. Green	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
J. Griffin	Fermi National Accelerator Laboratory, Batavia, IL 60510
J. Gunion	University of California, Davis, CA 95616

R. Gupta	Brookhaven National Laboratory, Upton, NY 11973-5000
A. Hershcovitch	Brookhaven National Laboratory, Upton, NY 11973-5000
T. Han	University of California, Davis, CA 95616
C. Johnstone	Fermi National Accelerator Laboratory, Batavia, IL 60510
S. Kahn	Brookhaven National Laboratory, Upton, NY 11973-5000
D. Kahana	Brookhaven National Laboratory, Upton, NY 11973-5000
H. Kirk	Brookhaven National Laboratory, Upton, NY 11973-5000
T. Kycia	Brookhaven National Laboratory, Upton, NY 11973-5000
P. Lebrun	Fermi National Accelerator Laboratory, Batavia, IL 60510
Y. Lee	Brookhaven National Laboratory, Upton, NY 11973-5000
D. Lissauer	Brookhaven National Laboratory, Upton, NY 11973-5000
L. Littenberg	Brookhaven National Laboratory, Upton, NY 11973-5000
A. Luccio	Brookhaven National Laboratory, Upton, NY 11973-5000
H. Ma	Brookhaven National Laboratory, Upton, NY 11973-5000
A. McInturff	Fermi National Accelerator Laboratory, Batavia, IL 60510
F. Mills	Fermi National Accelerator Laboratory, Batavia, IL 60510
N. Mokhov	Fermi National Accelerator Laboratory, Batavia, IL 60510
A. Moretti	Fermi National Accelerator Laboratory, Batavia, IL 60510
G. Morgan	Brookhaven National Laboratory, Upton, NY 11973-5000
M. Muhlbauer	Technische Universitat Munchen, Germany
M. Murtagh	Brookhaven National Laboratory, Upton, NY 11973-5000
D. Neuffer	Fermi National Accelerator Laboratory, Batavia, IL 60510
K-Y. Ng	Fermi National Accelerator Laboratory, Batavia, IL 60510
R. Noble	Fermi National Accelerator Laboratory, Batavia, IL 60510
J. Norem	Argonne National Laboratory, Argonne, IL 60439-4815
B. Norum	University of Virginia, Charlottesville, VA 22901
I. Novitski	Fermi National Accelerator Laboratory, Batavia, IL 60510
K. Oide	KEK, Tsukuba-shi, Ibaraki-Ken 305, Japan
F. Paige	Brookhaven National Laboratory, Upton, NY 11973-5000
Z. Parsa	Brookhaven National Laboratory, Upton, NY 11973-5000
J. Peterson	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
V. Polychronakos	Brookhaven National Laboratory, Upton, NY 11973-5000
M. Popovic	Fermi National Accelerator Laboratory, Batavia, IL 60510
S. Protopopescu	Brookhaven National Laboratory, Upton, NY 11973-5000
Z. Qian	Fermi National Accelerator Laboratory, Batavia, IL 60510
P. Rehak	Brookhaven National Laboratory, Upton, NY 11973-5000

T. Roser	Brookhaven National Laboratory, Upton, NY 11973-5000
R. Rossmanith	DESY, Hamburg, Germany
R. Scanlan	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
L. Schachinger	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
Q-S Shu	DESY, Germany
G. Silvestrov	BINP, RU-630090 Novosibirsk, Russia
S. Simrock	CEBAF, Newport News, VA 23606
I. Stumer	Brookhaven National Laboratory, Upton, NY 11973-5000
D. Summers	University of Mississippi, Oxford, MS 38677
M. Syphers	Brookhaven National Laboratory, Upton, NY 11973-5000
H. Takahashi	Brookhaven National Laboratory, Upton, NY 11973-5000
H. Takai	Brookhaven National Laboratory, Upton, NY 11973-5000
V. Tchernatine	Brookhaven National Laboratory, Upton, NY 11973-5000
Y. Torun	SUNY, Stony Brook, NY 11974
D. Trbojevic	Brookhaven National Laboratory, Upton, NY 11973-5000
W. Turner	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
A. Van Ginneken	Fermi National Accelerator Laboratory, Batavia, IL 60510
T. Vsevolozhskaya	BINP, RU-630090 Novosibirsk, Russia
R. Weggel	F. Bitter National Magnet Laboratory, MIT, Cambridge, MA 02139
E. Willen	Brookhaven National Laboratory, Upton, NY 11973-5000
W. Willis	Columbia University, New York, NY 10027
D. Winn	Fairfield University, Fairfield, CT 06430-5195
J. Wurtele	University of California, Berkeley, Berkeley, CA 94720-7300
Y. Zhao	Brookhaven National Laboratory, Upton, NY 11973-5000

MUON MUON COLLIDER: A FEASIBILITY STUDY

The $\mu^+\mu^-$ Collider Collaboration[1]

September 27, 1996

[1]This research was supported by the U.S. Department of Energy under contracts DE-AC02-76-CH00016, DE-AC02-76-CH03000 and DE-AC03-76-SF000098.

EXECUTIVE SUMMARY

Introduction

A feasibility study is presented of a $\mathbf{2} + \mathbf{2}\,\mathrm{TeV}$ muon collider with a luminosity of $\mathcal{L} = \mathbf{10^{35}}\,\mathrm{cm^{-2}s^{-1}}$. The resulting design is not optimized for performance, and certainly not for cost; however, it does suffice–we believe–to allow us to make a credible case, that a muon collider is a serious possibility for particle physics and, therefore, worthy of **R&D** support so that the reality of, and interest in, a muon collider can be better assayed. The goal of this support would be to completely assess the physics potential and to evaluate the cost and development of the necessary technology.

The muon collider complex consists of components which first produce copious pions, then capture the pions and the resulting muons from their decay; this is followed by an ionization cooling channel to reduce the longitudinal and transverse emittance of the muon beam. The next stage is to accelerate the muons and, finally, inject them into a collider ring wich has a small beta function at the colliding point. This is the first attempt at a point design and it will require further study and optimization. Experimental work will be needed to verify the validity of diverse crucial elements in the design.

Muons because of their large mass compared to an electron, do not produce significant synchrotron radiation. As a result there is negligible beamstrahlung and high energy collisions are not limited by this phenomena. In addition, muons can be accelerated in circular devices which will be considerably smaller than two full-energy linacs as required in an $e^+ - e^-$ collider. A hadron collider would require a CM energy 5 to 10 times higher than 4 TeV to have an equivalent energy reach. Since the accelerator size is limited by the strength of bending magnets, the hadron collider for the same physics reach would have to be much larger than the muon collider. In addition, muon collisions should be cleaner than hadron collisions.

There are many detailed particle reactions which are open to a muon collider and the physics of such reactions–what one learns and the necessary luminosity to see interesting events–are described in detail. Most of the physics accesible to an $e^+ - e^-$ collider could be studied in a muon collider. In addition the production of Higgs bosons in the s-channel will

allow the measurement of Higgs masses and total widths to high precision; likewise, $t\bar{t}$ and W^+W^- threshold studies would yield m_t and m_W to great accuracy. These reactions are at low center of mass energy (if the MSSM is correct) and the luminosity and $\Delta p/p$ of the beams required for these measurements is detailed in the Physics Chapter. On the other hand, at $\mathbf{2} + \mathbf{2}\,\mathrm{TeV}$, a luminosity of $\mathcal{L} \approx \mathbf{10^{35}}\,\mathrm{cm^{-2}\,s^{-1}}$ is desirable for studies such as, the scattering of longitudinal W bosons or the production of heavy scalar particles. Not explored in this work, but worth noting, are the opportunities for muon-proton and muon-heavy ion collisions as well as the enormous richness of such a facility for fixed target physics provided by the intense beams of neutrinos, muons, pions, kaons, antiprotons and spallation neutrons.

To *see* all the interesting physics described herein requires a careful study of the operation of a detector in the very large background. Three sources of background have been identified. The first is from any halo accompanying the muon beams in the collider ring. Very carefully prepared beams will have to be injected and maintained. The second is due to the fact that on average 35 % of the muon energy appears in its decay electron. The energy of the electron subsequently is converted into EM showers either from the synchrotron radiation they emit in the collider magnetic field or from direct collision with the surrounding material. The decays that occur as the beams traverse the low beta insert are of particular concern for detector backgrounds. A third source of background is $e^+ - e^-$ pair creation from $\mu^+ - \mu^-$ interaction. Studies of how to shield the detector and reduce the background are addressed in the Detector Chapter.

Polarization of the muons allows many very interesting measurements which are discussed in the Physics Chapter. Unlike the electron collider in which the electron beam is highly polarized and the positron beam unpolarized, both muon beams may be partially polarized. It is necessary to select forward moving muons from the pion's decay and thus reduce the available number of muons and hence the luminosity. The necessary machine technology needed to achieve such a collider is discussed in the Option Chapter; at the moment it is not part of our point design, although such capability would almost certainly be incorporated into an actual device.

The Machine

A major portion of this report is devoted to the details of a muon collider complex. The *driver* of a muon collider is a 30 GeV proton synchrotron capable of providing 2.5×10^{13} protons per bunch with four bunches per pulse and 15 Hz pulse rate. The repetition rate, but not the number of protons, is beyond that of any existing machine, but not so far beyond as to seem unrealistic. In fact, the criteria are almost met by the design of KAON. The protons are driven into a target, most likely a liquid target, where copious pions are produced (about one pion per proton). Questions of target survivability are discussed in the

Target Chapter. The target is surrounded by a 20 T solenoidal field, which is adiabatically matched to a 5 T solenoid in the decay channel. The captured pions have a wide range of energy, with a useful range from 100 MeV up to 1 GeV. A strong *phase-rotating* rf field is used to reduce this energy spread as well as the longitudinal extent of the beam. This results in approximately 0.3 muons per proton with mean energy of 150 MeV and a $\pm 20\,\%$ rms energy spread. The muons (about 8×10^{12}) are subsequently cooled by means of ionization cooling which is achieved in a periodic channel consisting of focusing elements, solenoids and/or lithium lenses and absorber at places of small beam size (but corresponding large transverse beam angles) and rf cavities to make up for the energy loss. In some locations along the channel, dispersion is introduced and wedge shaped absorbers are used to produce longitudinal cooling. This is described in the Cooling Chapter. We allow for further loss, beyond natural decay, between the number of captured muons and the final number of muons at the collider ring; at the entrance of the acceleration system is 3×10^{12} per bunch.

After cooling, the muons are accelerated in a cascaded series of recirculating linear accelerators, as described in the Acceleration Systems Chapter. A conventional synchrotron cannot be used as the acceleration is too slow and the muons will decay before reaching the design energy. On the other hand, it is possible to consider synchrotron-like pulsed magnets in the arcs of a recirculator. It should be noted that the primary cost of a muon collider complex is in the acceleration, so care and attention must be devoted to this matter. However, the process is reasonably straight-forward.

The collider ring is injected with two bunches of each sign of 2×10^{12} high energy muons. Approximately 1000 turns occur within a luminosity lifetime, thus making a ring (in contrast with a single collision) advantageous. In order to reach the desired high luminosity, it is necessary to have a very low beta, of the order of 3 mm, (and associated very large betas in the focusing quadrupoles) at the insertion point. Since the muons only live about 1000 turns, numerical simulations can easily provide us with quantitatively correct information. It is necessary to run the ring nearly isochronously so as to prevent bunch spreading and yet keep the rf impedance low enough as to prevent collective instabilities. Space charge effects, and beam-beam effects, in the collider ring are being studied and some conclusions are presented in the Collider Ring Chapter. Such a ring has never been built, but should be possible to construct and operate.

The muon complex requires numerous superconducting magnets. These are needed in the capture section, in the decay channel, in the arcs of the recirculating accelerators, and in the collider ring. Attention has been given to these magnets, as well as to the very special magnets required for the interaction region, and these various considerations may be found in the appropriate chapters.

A study of the scaling laws governing muon colliders is presented in the Options Chapter. Naturally, one would, if the concept is shown to be of interest, initially construct a lower energy machine (perhaps in the hundreds of GeV region) and thus the scaling laws are of special interest. In particular, a *lower energy demonstration machine* of $\mathcal{L} = \mathbf{10^{33}}\,\mathrm{cm^{-2}s^{-1}}$ at $250 + 250\,\mathrm{GeV}$ could serve as a breadboard for exploring the properties and technologies needed for this class of colliders while providing useful physics.

Conclusions

We suggest that to make sensible decisions about the future, the potential of a muon collider must be explored as rapidly and aggressively as possible. The accompanying document furnishes a solid base for identifying areas where more study and/or innovations are needed. In particular, **R&D** needs to be done related to the muon cooling channel, recirculating superconducting magnets or pulsed magnets for the accelerator in order to arrive at a design that minimizes cost. The magnets for the collider ring have a high heat load from muon decay electrons. Configurations other than a $\cos(\theta)$ magnet, such as a C-magnet, require study and modeling. The performance of rf cavities in the presence of intense radiation needs to be measured.

A sustained, extensive and integrated program of component development and optimization will have to be carried out in order to be assured that the design parameters can be attained and the cost minimized. The technology for the most part already exists within the High Energy Physics community and the work should involve the US, Europe, Russia, Japan and the international HEP community as a whole.

FOREWORD

This is a first attempt to gather in a single document the technical options and status of an ever evolving prospective high-energy (2+2 TeV), high-luminosity ($\mathcal{L} = 2 \times 10^{35} cm^{-2} s^{-1}$) $\mu^+\mu^-$ Collider.

This report is the compendium of the collaborative effort of scientists from Brookhaven National Laboratory (BNL), Fermi National Accelerator Laboratory (Fermilab) and Lawrence Berkeley National Laboratory (LBNL) with significant contributions from individual researchers from SLAC, KEK, CERN and US universities.

The first organizational meeting took place in October 1995, during the *9th Advanced ICFA Beam Dynamics Workshop* held in Montauk, NY. After some discussions, a steering committee was named to write and edit the various chapters of this *book*.

Three steering committee meetings were held (BNL, February; Fermilab, April and LBNL, May) to assess the progress of the chapters and to consider new promising technical alternatives.

A system connected to the INTERNET was implemented in a server (`http://www.bnl.gov/`), which can be easily reached with a WEB browser, with entry at the BNL Muon Collider Study Group WEB site:

`http://www.bnl.gov/~cap/mumu/mu_home_page.html`

Studies of the physics goals and requeriments of a $\mu^+\mu^-$ Collider began formally, with several workshops and symposiums, after the Port Jefferson Third Advanced Accelerator Concepts Workshop, June 14-20, 1992, where a working group on *Physics Opportunities* considered the merits of such a collider.[1] Inmediately after Port Jefferson, a special workshop was held in Napa, California, in the fall of 1992.[2] This was followed by the second workshop

[1] P. Chen and K. MacDonald, *Summary of the Physics Opportunities Working Group*, AIP Conference Proceedings **279**, Advanced Accelerator Concepts, Ed. J. Wurtele, 853 (1993)

[2] *Proceedings of the Mini-Workshop on $\mu^+\mu^-$ Colliders: Particle Physics and Design*, Napa CA, Nucl. Inst. and Meth., **A350** (1994); Ed. D. Cline

on *Physics Potential and Development of $\mu^+\mu^-$ Colliders*, Sausalito California, 1994.[3], the 9th Advanced ICFA Beam Dynamics Workshop, Montauk, New York in October 1995[4] and the Symposium on Physics Potential and Development of $\mu^+\mu^-$ Colliders at San Francisco, CA December 1995.[5]

I would like to acknowledge the contribution of Kathleen Tuohy, Patricia Tuttle and "Sam" Vanecek for their attention to details, that contributed greatly to making this report readable.

Juan C. Gallardo
For the $\mu^+\mu^-$ Collaboration

[3] *Physics Potential and Development of $\mu^+\mu^-$ Colliders, 2^{nd} Workshop*, Sausalito, CA, Ed. D. Cline, AIP Press, Woodbury, New York (1995)

[4] *Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop*, Ed. J. Gallardo, AIP Press (1996)

[5] *Proceedings Symposium on Physics Potential and Development of $\mu^+\mu^-$ Colliders*, San Francisco, CA December 1995, Supplement to Nucl.Phys. B, Ed. D. Cline and D. Sanders, to be published

INTERLABORATORY AND UNIVERSITY COLLABORATION

Argonne National Laboratory, Argonne, IL 60439-4815, USA
Brookhaven National Laboratory, Upton, NY 11973-5000, USA
Fermi National Accelerator Laboratory, Batavia, IL 60510, USA
F. Bitter National Magnet Laboratory, MIT, Cambridge, MA 02139, USA
Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
Stanford Linear Accelerator Center, Stanford, CA 94309, USA
Center for Advanced Accelerators, UCLA, Los Angeles, CA 90024,USA
Columbia University, New York, NY 10027, USA
Fairfield University, Fairfield, CT 06430-5195, USA
Indiana University, Bloomington, IN 47405, USA
SUNY, Stony Brook, NY 11974, USA
University of California, Berkeley, Berkeley, CA 94720-7300, USA
University of California, Davis, CA 95616, USA
University of Mississippi, Oxford, MS 38677, USA
University of Virginia, Charlottesville, VA 22901, USA
University of Wisconsin, Madison, WI 53706, USA
BINP, RU-630090 Novosibirsk, Russia
KEK, Tsukuba-shi, Ibaraki-Ken 305, Japan
DESY, Hamburg, Germany

INSTITUTIONAL AFFILIATION OF SCIENTISTS IN THE COLLABORATION

R. Palmer	Brookhaven National Laboratory, Upton, NY 11973-5000
A. Tollestrup	Fermi National Accelerator Laboratory, Batavia, IL 60510
A. Sessler	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
A. Skrinsky	BINP, RU-630090 Novosibirsk, Russia
C. Ankenbrandt	Fermi National Accelerator Laboratory, Batavia, IL 60510
A. Baltz	Brookhaven National Laboratory, Upton, NY 11973-5000
V. Barger	University of Wisconsin, Madison, WI 53706
O. Benary	Tel-Aviv University, Ramat-Aviv, Israel
M. Berger	Indiana University, Bloomington, IN 47405
A. Bogacz	Center for Adv. Accelerators, UCLA, Los Angeles, CA 90024-1547
S. Caspi	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
P. Chen	Stanford Linear Accelerator Center, Stanford, CA 94309
W-H. Cheng	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
Y. Cho	Argonne National Laboratory, Argonne, IL 60439-4815
D. Cline	Center for Adv. Accelerators, UCLA, Los Angeles, CA 90024-1547
E. Courant	Brookhaven National Laboratory, Upton, NY 11973-5000
D. Ehst	¿Argonne National Laboratory, Argonne, IL 60439-4815
R. Fernow	Brookhaven National Laboratory, Upton, NY 11973-5000
M. Furman	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
J. Gallardo	Brookhaven National Laboratory, Upton, NY 11973-5000
A. Garren	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
S. Geer	Fermi National Accelerator Laboratory, Batavia, IL 60510
I. Ginzburg	Institute of Mathematics, Novosibirsk, Russia
H. Gordon	Brookhaven National Laboratory, Upton, NY 11973-5000
M. Green	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
J. Griffin	Fermi National Accelerator Laboratory, Batavia, IL 60510
J. Gunion	University of California, Davis, CA 95616

R. Gupta	Brookhaven National Laboratory, Upton, NY 11973-5000
A. Hershcovitch	Brookhaven National Laboratory, Upton, NY 11973-5000
T. Han	University of California, Davis, CA 95616
C. Johnstone	Fermi National Accelerator Laboratory, Batavia, IL 60510
S. Kahn	Brookhaven National Laboratory, Upton, NY 11973-5000
D. Kahana	Brookhaven National Laboratory, Upton, NY 11973-5000
H. Kirk	Brookhaven National Laboratory, Upton, NY 11973-5000
T. Kycia	Brookhaven National Laboratory, Upton, NY 11973-5000
P. Lebrun	Fermi National Accelerator Laboratory, Batavia, IL 60510
Y. Lee	Brookhaven National Laboratory, Upton, NY 11973-5000
D. Lissauer	Brookhaven National Laboratory, Upton, NY 11973-5000
L. Littenberg	Brookhaven National Laboratory, Upton, NY 11973-5000
A. Luccio	Brookhaven National Laboratory, Upton, NY 11973-5000
H. Ma	Brookhaven National Laboratory, Upton, NY 11973-5000
A. McInturff	Fermi National Accelerator Laboratory, Batavia, IL 60510
F. Mills	Fermi National Accelerator Laboratory, Batavia, IL 60510
N. Mokhov	Fermi National Accelerator Laboratory, Batavia, IL 60510
A. Moretti	Fermi National Accelerator Laboratory, Batavia, IL 60510
G. Morgan	Brookhaven National Laboratory, Upton, NY 11973-5000
M. Muhlbauer	Technische Universitat Munchen, Germany
M. Murtagh	Brookhaven National Laboratory, Upton, NY 11973-5000
D. Neuffer	Fermi National Accelerator Laboratory, Batavia, IL 60510
K-Y. Ng	Fermi National Accelerator Laboratory, Batavia, IL 60510
R. Noble	Fermi National Accelerator Laboratory, Batavia, IL 60510
J. Norem	Argonne National Laboratory, Argonne, IL 60439-4815
B. Norum	University of Virginia, Charlottesville, VA 22901
I. Novitski	Fermi National Accelerator Laboratory, Batavia, IL 60510
K. Oide	KEK, Tsukuba-shi, Ibaraki-Ken 305, Japan
F. Paige	Brookhaven National Laboratory, Upton, NY 11973-5000
Z. Parsa	Brookhaven National Laboratory, Upton, NY 11973-5000
J. Peterson	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
V. Polychronakos	Brookhaven National Laboratory, Upton, NY 11973-5000
M. Popovic	Fermi National Accelerator Laboratory, Batavia, IL 60510
S. Protopopescu	Brookhaven National Laboratory, Upton, NY 11973-5000
Z. Qian	Fermi National Accelerator Laboratory, Batavia, IL 60510
P. Rehak	Brookhaven National Laboratory, Upton, NY 11973-5000

T. Roser	Brookhaven National Laboratory, Upton, NY 11973-5000
R. Rossmanith	DESY, Hamburg, Germany
R. Scanlan	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
L. Schachinger	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
Q-S Shu	DESY, Germany
G. Silvestrov	BINP, RU-630090 Novosibirsk, Russia
S. Simrock	CEBAF, Newport News, VA 23606
I. Stumer	Brookhaven National Laboratory, Upton, NY 11973-5000
D. Summers	University of Mississippi, Oxford, MS 38677
M. Syphers	Brookhaven National Laboratory, Upton, NY 11973-5000
H. Takahashi	Brookhaven National Laboratory, Upton, NY 11973-5000
H. Takai	Brookhaven National Laboratory, Upton, NY 11973-5000
V. Tchernatine	Brookhaven National Laboratory, Upton, NY 11973-5000
Y. Torun	SUNY, Stony Brook, NY 11974
D. Trbojevic	Brookhaven National Laboratory, Upton, NY 11973-5000
W. Turner	Lawrence Berkeley National Laboratory, Berkeley, CA 94720
A. Van Ginneken	Fermi National Accelerator Laboratory, Batavia, IL 60510
T. Vsevolozhskaya	BINP, RU-630090 Novosibirsk, Russia
R. Weggel	F. Bitter National Magnet Laboratory, MIT, Cambridge, MA 02139
E. Willen	Brookhaven National Laboratory, Upton, NY 11973-5000
W. Willis	Columbia University, New York, NY 10027
D. Winn	Fairfield University, Fairfield, CT 06430-5195
J. Wurtele	University of California, Berkeley, Berkeley, CA 94720-7300
Y. Zhao	Brookhaven National Laboratory, Upton, NY 11973-5000

Contents

Chapter 1

INTRODUCTION AND OVERVIEW

Contents

1.1 General Considerations

This report describes the theory and technology needed for muon colliders and gives a consistent set of parameters for a 2+2 TeV machine with a luminosity of $10^{35}\,\mathrm{cm}^{-2}s^{-1}$ as well as for a 250+250 GeV collider with luminosity of $10^{33}\,\mathrm{cm}^{-2}s^{-1}$. The higher energy machine would be the upgrade of the lower energy machine since the muon source has common properties. In addition, a *demonstration* machine is discussed, which could serve as a breadboard for exploring the properties of this class of colliders before committing large sums to the construction of the final complex.

The possibility of muon colliders was introduced by Skrinsky et al.[1] and Neuffer[2]. More recently, several workshops and collaboration meetings have greatly increased the level of understanding[3],[4]. After the workshop at Sausalito, in December 1994, a collaboration was formed by BNL and FNAL to study the concept and prepare this document for Snowmass. This effort has expanded to include LBNL, ANL and several other individuals from KEK, DESY and various universities. Subsequently, three mini-workshops were organized and attended by over sixty scientists, to discuss the several technical options and assess the

progress and status of the study for a prospective muon collider. Their contributions are gathered in this document.

1.2 Technical Considerations

Hadron collider energies are limited by their size, and technical constraints on bending magnetic fields. At very high energies it will also become impractical to obtain the required luminosities, which must rise as the energy squared. In fact, lepton colliders in general, offer the advantage that the interaction energy is given by twice the machine energy, because they undergo simple, single-particle interactions, compared to the hadron collider where the effective energy is much lower than that of the proton. Even worse, the gluon-gluon background radiation makes it increasingly difficult to sort out the complicated decay schemes envisaged for the SUSY particles. The lepton collider on the other hand offers clean production of charged pairs with a cross section comparable to $\sigma_{\rm QCD} = 100/s\,$fb where s is the energy squared in TeV^2.

Extension of e^+e^- colliders to multi-TeV energies is severely performance-constrained by beamstrahlung, and cost-constrained because two full energy linacs are required[5] to avoid the excessive synchrotron radiation that would occur in rings. Muons ($\frac{m_\mu}{m_e} = 207$) have the same advantage in energy reach as electrons, but have negligible beamstrahlung, and can be accelerated and stored in rings with a much smaller radius than a hadron collider of comparable energy reach, making the possibility of high energy $\mu^+\mu^-$colliders attractive.

The answer to the question of *Why study muon colliders?* is therefore driven by the following two facts:

- Muon colliders can reach much higher energy than e^+e^- colliders due to the much reduced synchrotron radiation. The beamstrahlung and initial state radiation is also smaller leading to better energy definition of the initial state.

- For cases where the coupling is proportional to the mass, as in the case of s-channel Higgs production, muons have an advantages of $\approx (207)^2$ over electrons.

There are however, several major technical problems with muon colliders:

- Muon decay with a lifetime of 2.2×10^{-6} s. This problem is partially overcome by rapidly increasing the energy of the muons, and thus benefiting from their relativistic γ factor. At 2 TeV, for example, their lifetime is 0.044 s which is sufficient for approximately 1000 storage-ring collisions.

- another consequence of the muon decay is that the decay products heat the magnets of the collider ring and create backgrounds in the detector.

- Since the muons are created through pion decay into a diffuse phase space, some form of cooling is essential. Conventional stochastic or synchrotron cooling is too slow to be effective before they decay. Ionization cooling can be used, but the final emittance of the muon beams will remain larger than that possible for electrons in an e^+e^- collider.

- The machine represents an *untried* technology. It will require an aggressive **R&D** program before a conclusion can be reached. This document should help to define the course of the necessary work.

Despite these problems, it appears possible that high energy muon colliders might have luminosities comparable to or, at energies of several TeV, even higher than those in e^+e^- colliders[5]. Because the $\mu^+\mu^-$machines would be much smaller[6], and require much lower precision (the final spots are about three orders of magnitude larger), they may be significantly less expensive. However, e^+e^- colliders are at a technologically more advanced stage of development and likely will be built before a demonstration muon collider. Hence, it is relevant to ask *what is it that a muon collider may contribute to our understanding of the energy frontier that cannot be achieved with an electron collider?*. That is briefly summarized next and discussed in details in the Physics Chapter.

1.3 Physics Considerations

There are at least two physics advantages of a $\mu^+\mu^-$collider, when compared with an e^+e^- collider:

- Because of the lack of beamstrahlung, a $\mu^+\mu^-$collider can be operated with an energy spread of as little as 0.01 %. It is thus possible to use the $\mu^+\mu^-$collider for precision measurements of masses and widths, that would be very difficult, if not impossible, with an e^+e^- collider.

- The direct coupling of a lepton-lepton system to a Higgs boson has a cross section that is proportional to the square of the mass of the lepton. As a result, the cross section for direct Higgs production from the $\mu^+\mu^-$system is 40,000 times that from an e^+e^- system.

However, there are liabilities:

- It will be relatively difficult to obtain both high polarization and good luminosity in a $\mu^+\mu^-$collider, whereas good polarization of one beam can be obtained in an e^+e^- collider without any loss in luminosity. However, in the muon case moderate polarization could be obtained for both beams which compensate for the lower luminosity.

- Because of the decays of the muons, there will be a considerable background of photons, muons and neutrons in the detector. This background may be acceptable for some experiments, but it cannot be as clean as in an e^+e^- collider.

1.4 Overview of Components

The basic components of the $\mu^+\mu^-$collider are shown schematically in Fig.1.1. Tb.1.1 shows parameters for the candidate designs. Notice that more precisely a factor of π must appear in the dimensions of emittance (i.e. π mm mrad). The emittance ϵ is defined as the *rms* transverse phase space area divided by π and the normalized emittance is $\epsilon^N = \beta\gamma\epsilon$.

A high intensity proton source is bunch compressed and focused on a pion production target. The pions generated are captured by a high field solenoid and transferred to a solenoidal decay channel within a low frequency linac. The linac serves to reduce, by phase rotation, the momentum spread of the pions and of the muons into which they decay. Subsequently, the muons are cooled by a sequence of ionization cooling stages. Each stage consists of energy loss, acceleration, and emittance exchange by energy absorbing wedges in the presence of dispersion. Once they are cooled, the muons must be rapidly accelerated to avoid decay. This can be done in recirculating accelerators (à la CEBAF) or in fast-pulsed synchrotrons. Collisions occur in a separate high field collider storage ring with a single very low beta insertion.

Each one of these components is described in details in the following chapters.

1.5 Discussion

The physics reach of a $\mu^+\mu^-$collider is well outlined by the studies that have been done for a e^+e^- collider. It is reasonably clear that an actual realization of a muon collider has both technical advantages and disadvantages when compared with an e^+e^- machine. Similarly, it has specific physics advantages and disadvantages. Thus, it seems reasonable to consider $\mu^+\mu^-$colliders as complementary to e^+e^- colliders just as e^+e^- colliders are complementary to hadron machines.

It is worthwhile at this point to face some *what if* questions:

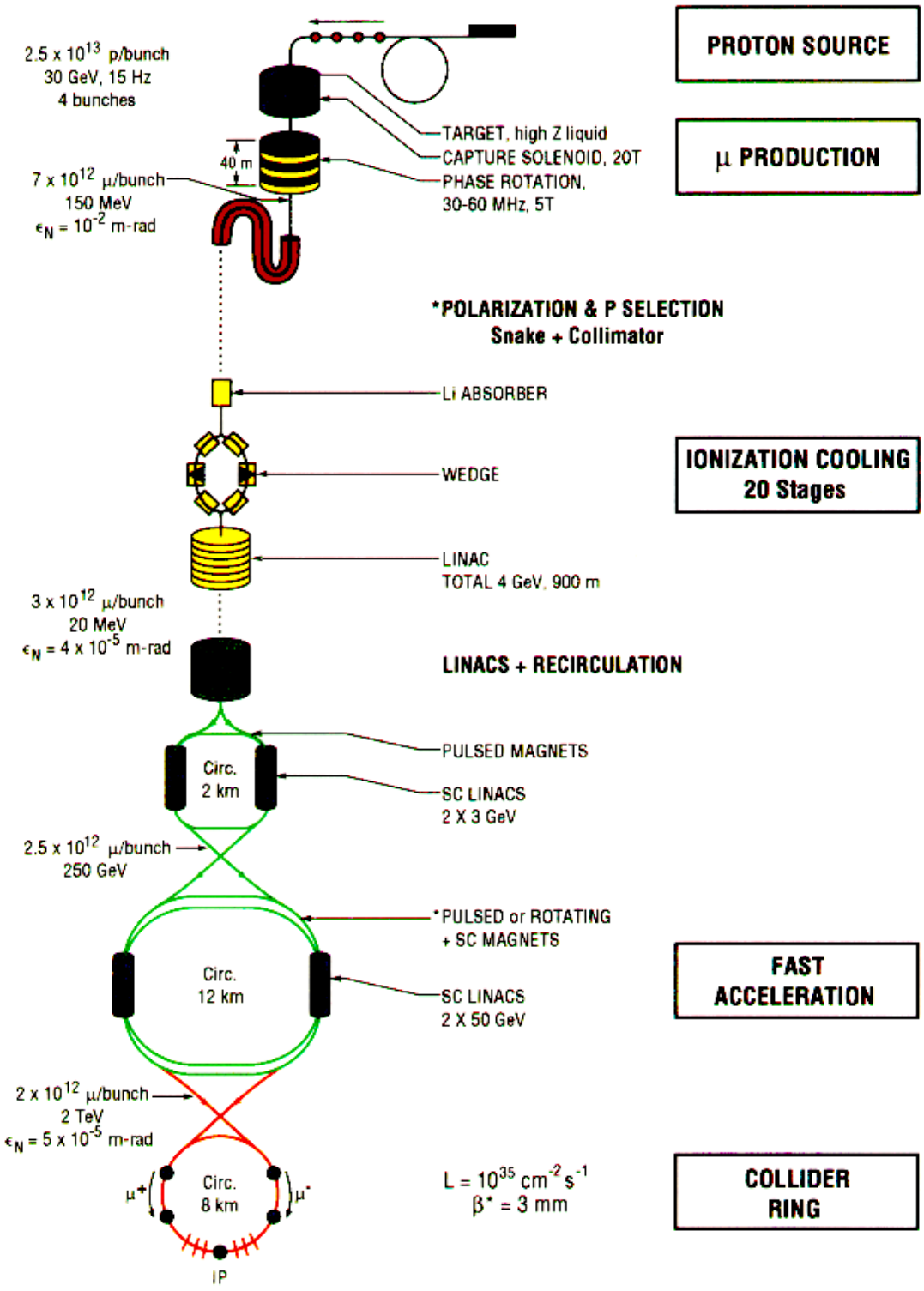


Figure 1.1: Schematic of a muon collider.

Table 1.1: Parameters of collider rings

		4 TeV	.5 TeV	Demo.
Beam energy	TeV	2	.25	.25
Beam γ		19,000	2,400	2,400
Repetition rate	Hz	15	15	2.5
Muons per bunch	10^{12}	2	4	4
Bunches of each sign		2	1	1
Normalized *rms* emittance ϵ^N	$10^{-6}\pi$ m $-$ rad	50	90	90
Bending Field	T	8.5	8.5	7.5
Circumference	km	7	1.2	1.5
Average ring mag. field B	T	6	5	4
Effective turns before decay		900	800	750
β^* at intersection	mm	3	8	8
rms beam size at I.P.	μm	2.8	17	17
Luminosity	$\mathrm{cm}^{-2}\mathrm{s}^{-1}$	10^{35}	5×10^{33}	6×10^{32}

- What if the next machine is TESLA or JLC? Clearly, we would support either one of them as good citizens of the international High Energy community. Nevertheless, is there then a complementary machine that could be built in this country?

- What if Nature is different from the scenario presented by SUSY of new physics opening up below 500 GeV and higher energy is required?

- What if the *next machine* is not built for more than ten years? An aggressive muon collider **R&D** during that period may show that it is a natural add-on to existing facilities with rich physics possibilities for an accelerator complex that is affordable in a staged manner.

The studies of the past year are contained in this document which outlines in detail what is known about this class of machine. It appears that many of the problems have been solved or at least have solutions. On the other hand, it is also clear that much more work needs to be done -including experimental work. The present technologies are being pushed to the limit in some cases; on the other hand, new inventions to solve various problems have regularly occurred, showing a healthy tension between challenges and the capacities of the scientists to produce innovations; room exists for more discoveries that can lead to reduced cost, increased luminosity, polarization and simpler configurations.

The present report furnishes a solid base for identifying the main areas of study. The

machines described have internally consistent sets of parameters but no optimization of the various components have been attempted. An important part of the optimization procedure involves extensive testing of real components. This is the ultimate objective of the *demonstration machine* and it is hoped that the same collider can also be a useful physics tool, although it is too early to visualize exactly how this would come about. However, even before this, there will have to be an extensive and integrated program of component development.

Finally, the question of the cost of a $\mu^+\mu^-$collider *is not addressed* in this report. Obviously, the next phase of the work will be to optimize the many pieces of the machine in order to minimize the cost. On that regard, the investment in the muon source is the first and most important step. The potential for systematically raising the energy depends on the muon source. A low energy collider ($250 + 250\,\mathrm{GeV}$) with a luminosity of $10^{33}\,\mathrm{cm}^{-2}s^{-1}$ not only has the possibility of interesting physics but also provides the technical base for the higher energy versions in a scenario where the upgrade is achieved by integrating a modest budget over time.

Bibliography

[1] E. A. Perevedentsev and A. N. Skrinsky, Proc. 12th Int. Conf. on High Energy Accelerators, F. T. Cole and R. Donaldson, Eds., (1983) 485; A. N. Skrinsky and V.V. Parkhomchuk, Sov. J. of Nucl. Physics **12**, (1981) 3; *Early Concepts for $\mu^+\mu^-$ Colliders and High Energy μ Storage Rings*, *Physics Potential & Development of $\mu^+\mu^-$ Colliders. 2^{nd} Workshop*, Sausalito, CA, Ed. D. Cline, AIP Press, Woodbury, New York, (1995).

[2] D. Neuffer, *Colliding Muon Beams at 90 GeV* , FermiLab Physics Note FN-319 (1979), unpublished; D. Neuffer, Particle Accelerators, **14**, (1983) 75; D. Neuffer, Proc. 12th Int. Conf. on High Energy Accelerators, F. T. Cole and R. Donaldson, Eds., 481 (1983).

[3] *Proceedings of the Mini-Workshop on $\mu^+\mu^-$ Colliders: Particle Physics and Design*, Napa CA, Nucl Inst. and Meth., **A350** (1994) ; Proceedings of the Muon Collider Workshop, February 22, 1993, Los Alamos National Laboratory Report LA- UR-93-866 (1993) and *Physics Potential & Development of $\mu^+\mu^-$ Colliders 2^{nd} Workshop*, Sausalito, CA, Ed. D. Cline, AIP Press, Woodbury, New York, (1995).

[4] Transparencies at the *2 + 2 TeV $\mu^+\mu^-$ Collider Collaboration Meeting*, Feb 6-8, 1995, BNL, compiled by Juan C. Gallardo; transparencies at the *2 + 2 TeV $\mu^+\mu^-$ Collider Collaboration Meeting*, July 11-13, 1995, FERMILAB, compiled by Robert Noble; Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop, Ed. J. C. Gallardo, AIP Conference Proceedings 372, 1996.

[5] *Overall Parameters and Construction Techniques Working Group report*, *Proceedings of the Fifth International Workshop on Next-Generation Linear Colliders*, Oct 13-21, 1993, Slac-436, pp.428.

[6] D. V. Neuffer and R. B. Palmer, Proc. European Particle Acc. Conf., London (1994); M. Tigner, in Advanced Accelerator Concepts, Port Jefferson, NY 1992, AIP Conf. Proc. **279**, 1 (1993).

[7] R. B. Palmer et al., *Monte Carlo Simulations of Muon Production*, *Physics Potential & Development of $\mu^+\mu^-$ Colliders 2^{nd} Workshop*, Sausalito, CA, Ed. D. Cline, AIP Press, Woodbury, New York, pp. 108 (1995).

Contributors

- R. B. Palmer, (BNL/SLAC)
- A. Tollestrup, (FermiLab)
- A. Sessler, (LBNL)

List of Figures

List of Tables

Chapter 2

PHYSICS
(4 TeV AND 500 GeV)

Contents

2.1 Abstract

We discuss the capabilities of future muon colliders to resolve important particle physics questions. A collider with c.m. energy $\sqrt{s} = 100$ to 500 GeV offers the unique opportunity

to produce Higgs bosons in the s-channel and thereby measure the Higgs masses, total widths and several partial widths to high precision. At this same machine, $t\bar{t}$ and W^+W^- threshold studies would yield superior precision in the determination of m_t and m_W. A multi-TeV $\mu^+\mu^-$ collider would open up the realm of physics above the 1 TeV scale, allowing, for example, copious production of supersymmetric particles up to the highest anticipated masses or a detailed study of the strongly-interacting scenario of electroweak symmetry breaking.

Techniques and strategies for discovering and measuring the properties of Higgs bosons via s-channel production at a $\mu^+\mu^-$ collider, and the associated requirements for the machine and detector, are discussed in detail. The unique feature of s-channel production is that, with good energy resolution, the mass, total width and partial widths of a Higgs boson can be directly measured with remarkable accuracy in most cases. For the expected machine parameters and luminosity the Standard Model (SM) Higgs boson h_{SM}, with mass $\lesssim 2m_W$, the light h^0 of the minimal supersymmetric Standard Model (MSSM), and the heavier MSSM Higgs bosons (the CP-odd A^0 and the CP-even H^0) can all be studied in the s-channel, with the heavier states accessible up to the maximal $\sqrt{s}$ over a large fraction of the MSSM parameter space. In addition, it may be possible to discover the A^0 and H^0 by running the collider at full energy and observing excess events in the bremsstrahlung tail at lower energy. The integrated luminosity, beam resolution and machine/detector features required to distinguish between the h_{SM} and h^0 are delineated.

2.2 Particle Physics Opportunities at $\mu^+\mu^-$ Colliders

2.2.1 Introduction

There is increasing interest recently in the possible construction of a $\mu^+\mu^-$ collider[1, 2, 3, 4]. The expectation is that a muon collider with energy and integrated luminosity comparable to or superior to those attainable at e^+e^- colliders can be achieved[5, 6, 7]. An initial survey of the physics potential of muon colliders has been carried out[8]. In this report we summarize some of the progress on the physics issues that has been made in the last year; a more comprehensive report is in preparation[9].

One of the primary arguments for an e^+e^- collider is the complementarity with physics studies at the LHC. The physics potential of a muon collider is comparable to that of an electron collider with the same energy and luminosity. However, electron colliders are at a technologically more advanced stage and will likely be built before muon colliders. Hence a very relevant issue is what can be done at a muon collider that cannot be done at an electron collider.

The advantages of a muon collider can be summarized briefly as follows:

- The muon is significantly heavier than the electron, and therefore couplings to Higgs bosons are enhanced making possible their study in the s-channel production process.
- The limitation on luminosity from beam-beam interactions that arises at an e^+e^- collider is not relevant for muon beam energies below about 100 TeV; very small/flat beams are unnecessary. Instead, large luminosity is achieved for $\sim 3\,\mu$m size beams by storing multiple bunches in the final storage ring and having a large number of turns of storage per cycle. Radiative losses in the storage ring are small due to the large muon mass. Thus, extending the energy reach of these colliders well beyond the 1 TeV range is possible.
- The muon collider can be designed to have finer energy resolution than an e^+e^- machine.
- At a muon collider, $\mu^+\mu^+$ and $\mu^-\mu^-$ collisions are likely to be as easily achieved as $\mu^+\mu^-$ collisions.

There are two slight drawbacks of a muon collider. The first is that substantial polarization of the beams can probably not be achieved without sacrificing luminosity. The second drawback is that the $\gamma\gamma$ and $\mu\gamma$ options are probably not feasible. At future linear e^+e^- colliders, the possibility exists to backscatter laser photons off the electron and/or positron beams. The resulting back-scattered photons are highly collimated and could serve as a photon beam, thus converting the e^+e^- collider to a $e\gamma$ or $\gamma\gamma$ collider. The collisions from the back-scattered photons have center-of-mass energies that range up almost to that of the parent e^+e^- collider. Including this option at a $\mu^+\mu^-$ collider is problematic from kinematic considerations. The highest photon energy ω attainable from a lepton with energy E is

$$\frac{\omega_{\max}}{E} = \frac{x}{x+1}\,, \tag{2.1}$$

where

$$x = \frac{4E\omega_0}{m_\mu^2 c^4}\,. \tag{2.2}$$

For a muon collider $x \ll 1$ unless a laser photon energy ω_0 of the order of keV is possible, which seems unlikely.

A proposed schematic design for a muon collider is shown in Fig. 1.1 in chapter 1. Protons produce π's in a fixed target which subsequently decay giving μ's. The muons must be collected, cooled and subsequently accelerated to high energies. Since the muon is so

much heavier than the electron, synchrotron radiation is much less so that circular storage rings are feasible even at TeV energies.

The monochromaticity of the beams will prove critically important for some of the physics that can be done at a $\mu^+\mu^-$ collider. The energy profile of the beam is expected to be roughly Gaussian in shape, and the rms deviation R is expected to naturally lie in the range $R = 0.04\%$ to 0.08%[10]. Additional cooling could further sharpen the beam energy resolution to $R = 0.01\%$.

Two possible $\mu^+\mu^-$ machines have been discussed as design targets and are being actively studied [2, 3, 4]:

(i) A first muon collider (FMC) with low c. m. energy ($\sqrt{s}$) between 100 and 500 GeV and $\mathcal{L} \sim 2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ delivering an annual integrated integrated luminosity $L \sim 20\ \mathrm{fb}^{-1}$.

(ii) A next muon collider (NMC) with high $\sqrt{s} \gtrsim 4$ TeV and $\mathcal{L} \sim 10^{35}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ giving $L \sim 1000\ \mathrm{fb}^{-1}$ yearly.

2.2.2 *s*-Channel Higgs Physics

The simplest Higgs sector is that of the Standard Model (SM) with one Higgs boson. However, the naturalness and hierarchy problems that arise in the SM and the failure of grand unification of couplings in the SM suggest that a single Higgs boson is probably not the whole story of electroweak symmetry breaking. Therefore, it is crucially important to understand and delineate experimentally various alternative possibilities.

Supersymmetry is an especially attractive candidate theory in that it solves the naturalness and hierarchy problems (for a sufficiently low scale of supersymmetry breaking) and in that scalar bosons, including Higgs bosons, are on the same footing as fermions as part of the particle spectrum. The minimal supersymmetric model (MSSM) is the simplest SUSY extension of the SM. In the MSSM, every SM particle has a superpartner. In addition, the minimal model contains exactly two Higgs doublets. At least two Higgs doublet fields are required in order that both up and down type quarks be given masses without breaking supersymmetry (and also to avoid anomalies in the theory). Exactly two doublets allows unification of the SU(3), SU(2) and U(1) coupling constants. (Extra Higgs singlet fields are allowed by unification, but are presumed absent in the MSSM.) For two Higgs doublets and

no Higgs singlets, the Higgs spectrum comprises 5 physical Higgs bosons

$$h^0, H^0, A^0, H^+, H^- \ . \tag{2.3}$$

The quartic couplings in the MSSM Higgs potential are related to the electroweak gauge couplings g and g' and the tree-level Higgs mass formulas imply an upper bound on the mass of the lightest Higgs boson, $m_h \leq M_Z$. At one loop, the radiative correction to the mass of the lightest Higgs state depends on the top and stop masses

$$\delta m^2_{h^0} \simeq \frac{3g^2}{8\pi^2 m_W^2} m_t^4 \ln\left(\frac{m_{\tilde{t}_1} m_{\tilde{t}_2}}{m_t^2}\right) \ . \tag{2.4}$$

Two-loop corrections are also significant. The resulting iron-clad upper bounds on the possible mass of the lightest Higgs boson are

$$m_{h^0} \lesssim 130 \text{ GeV} \qquad \text{MSSM}, \tag{2.5}$$

$$m_{h^0} \lesssim 150 \text{ GeV} \qquad \text{any SUSY GUT}, \tag{2.6}$$

$$m_{h^0} \lesssim 200 \text{ GeV} \qquad \text{any model with GUT and desert.} \tag{2.7}$$

In the largest part of parameter space, e.g. $m_{A^0} > 150$ GeV in the MSSM, the lightest Higgs boson has fairly SM-like couplings.

The first discovery of a light Higgs boson is likely to occur at the LHC which might be operating for several years before a next-generation lepton collider is built. Following its discovery, interest will focus on measurements of its mass, total width, and partial widths. A first question then is what could be accomplished at the Large Hadron Collider (LHC) or the Next Linear Collider (NLC) in this regard.

At the LHC, a SM-like Higgs can be discovered either through gluon fusion, followed by $\gamma\gamma$ or 4ℓ decay,

$$gg \to h \to \gamma\gamma \ , \tag{2.8}$$

$$gg \to h \to ZZ^\star \to 4l \ , \tag{2.9}$$

or through associated production

$$gg \to t\bar{t}h \quad \hookrightarrow \gamma\gamma \ , \tag{2.10}$$

$$q\overline{q} \to Wh \quad \hookrightarrow \gamma\gamma \ . \tag{2.11}$$

The LHC collaborations report that the Higgs boson is detectable in the mass range $50 \lesssim m_h \lesssim 150$ GeV via its $\gamma\gamma$ decay mode. The mass resolution is expected to be $\lesssim 1\%$. At the NLC the Higgs boson is produced in the Bjorken process

$$e^+e^- \to Z^\star \to Zh \tag{2.12}$$

and the h can be studied through its dominant $b\bar{b}$ decay. At the NLC (which may be available prior to a $\mu^+\mu^-$ collider) the mass resolution is strongly dependent on the detector performance and signal statistics:

$$\Delta m_h \simeq R_{\rm event}({\rm GeV})/\sqrt{N}\ , \tag{2.13}$$

where $R_{\rm event}$ is the single event resolution and N is the number of signal events. The single event resolution is about 4 GeV for an SLD-type detector[11], but improved performance as typified by the "super"-LC detector would make this resolution about 0.3 GeV[12, 13]. The uncertainty in the Higgs boson mass for various integrated luminosities is shown in Fig. 2.1. For a Higgs boson with Standard Model couplings this gives a Higgs mass determination of

$$\Delta m_{h_{SM}} \simeq 400\ {\rm MeV} \left(\frac{10\ {\rm fb}^{-1}}{L}\right)^{1/2}\ , \tag{2.14}$$

for the SLD-type detector. Precision measurements of the Higgs total width and partial widths will be necessary to distinguish between the predictions of the SM Higgs boson h_{SM} and the MSSM Higgs boson h^0. Can the total and partial widths be measured at other machines? This is a complicated question since each machine contributes different pieces to the puzzle. The bottom line[14] is that the LHC, NLC, and $\gamma\gamma$ colliders each measure interesting couplings and/or branching ratios, but their ability to detect deviations due to the differences between the h^0 and h_{SM} is limited to $m_{A^0} \lesssim 300$ GeV. Further, a model-independent study of all couplings and widths requires all three machines with consequent error propagation problems.

The s-channel process $\mu^+\mu^- \to b\overline{b}$ shown in Fig. 2.2 is uniquely suited to several critical precision Higgs boson measurements [15, 16]. Detecting and studying the Higgs boson in the s-channel would require that the machine energy be adjusted to correspond to the Higgs mass. Since the storage ring is only a modest fraction of the overall muon collider cost[17], a special-purpose ring could be built to optimize the luminosity near the Higgs peak.

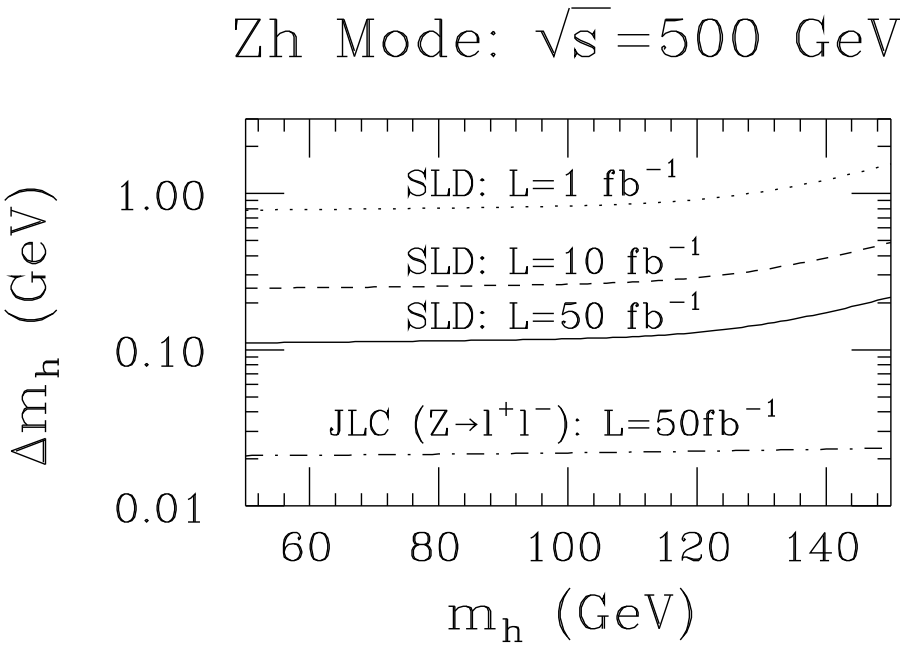


Figure 2.1: The uncertainty $\pm\Delta m_h$ in the determination of m_h for a SM-like Higgs boson using Zh production and a ± 4 GeV ("SLD") or ± 0.3 GeV ("JLC") single event mass resolution for m_h.

The s-channel Higgs phenomenology is set by the $\sqrt{s}$ rms Gaussian spread denoted by $\sigma_{\sqrt{s}}$. A convenient formula for $\sigma_{\sqrt{s}}$ is

$$\sigma_{\sqrt{s}} = (7 \text{ MeV}) \left(\frac{R}{0.01\%}\right) \left(\frac{\sqrt{s}}{100 \text{ GeV}}\right) . \tag{2.15}$$

A crucial consideration is how this natural spread in the muon collider beam energy compares to the width of the Higgs bosons, given in Fig. 2.3. In particular, a direct scan measurement of the Higgs width requires a beam spread comparable to the width. The narrowest Higgs boson widths are those of a light SM Higgs boson with mass $\lesssim 100$ GeV. In the limit where the heavier MSSM Higgs bosons become very massive, the lightest supersymmetric Higgs typically has a mass of order 100 GeV and has couplings that are sufficiently SM-like that its width approaches that of a light h_{SM} of the same mass. In either case, the discriminating power of a muon collider with a very sharp energy resolution would be essential for a direct width measurement.

A quantitative examination of Fig. 2.3 shows that for typical muon beam resolution

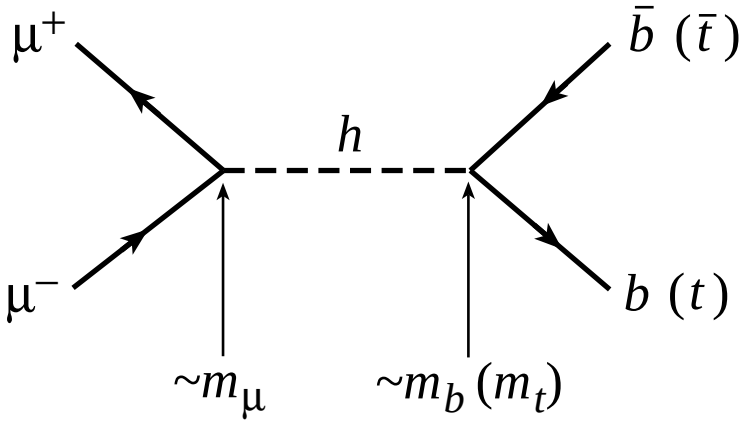


Figure 2.2: Feynman diagram for s-channel production of a Higgs boson.

$(R = 0.06\%)$

$$\sigma_{\sqrt{s}} \gg \Gamma_{h_{SM}} \text{ , for } m_{h_{SM}} \sim 100 \text{ GeV} , \tag{2.16}$$

$$\sigma_{\sqrt{s}} \sim \Gamma_{h^0} \text{ , for } m_{h^0} \text{ not near } m_{h^0}^{\max} , \tag{2.17}$$

$$\sigma_{\sqrt{s}} \lesssim \Gamma_{H^0}, \Gamma_{A^0} \text{ , at moderate } \tan\beta , \tag{2.18}$$

$$\text{for } m_{H^0,A^0} \sim 400 \text{ GeV} ,$$

$$\ll \Gamma_{H^0}, \Gamma_{A^0} \text{ , at large } \tan\beta , \tag{2.19}$$

$$\text{for } m_{H^0,A^0} \sim 400 \text{ GeV} .$$

To be sensitive to the $\Gamma_{h_{SM}}$ case, a resolution $R \sim 0.01\%$ is mandatory. This is an important conclusion given that such a small resolution requires early consideration in the machine design.

The s-channel Higgs resonance cross section is

$$\sigma_h = \frac{4\pi\Gamma(h \to \mu\mu)\,\Gamma(h \to X)}{(\hat{s} - m_h^2)^2 + m_h^2[\Gamma_h^{\text{tot}}]^2} , \tag{2.20}$$

where $\hat{s} = (p_{\mu^+} + p_{\mu^-})^2$ is the c. m. energy squared of the event, X denotes a final state and Γ_h^{tot} is the total width. The effective cross section is obtained by convoluting this resonance form with the Gaussian distribution of width $\sigma_{\sqrt{s}}$ centered at $\sqrt{s}$. When the Higgs width is much smaller than $\sigma_{\sqrt{s}}$, the effective signal cross section result for $\sqrt{s} = m_h$, denoted by $\overline{\sigma}_h$,

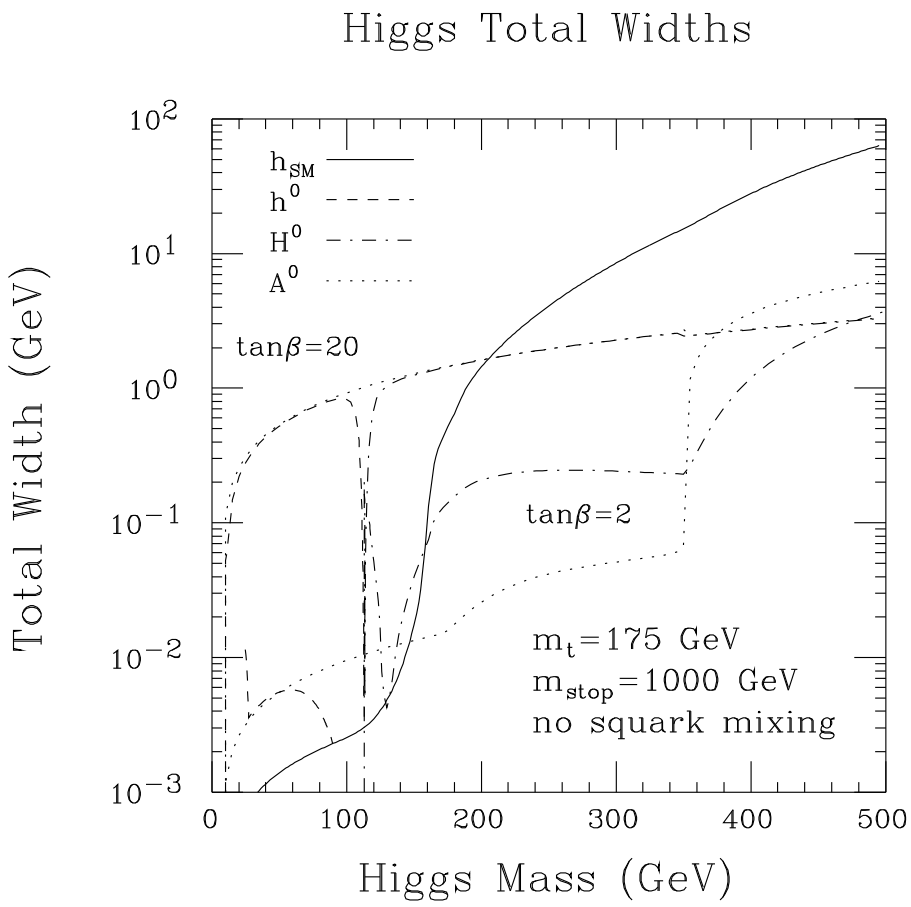


Figure 2.3: Total width vs mass of the SM and MSSM Higgs bosons for $m_t = 175$ GeV. In the case of the MSSM, we have plotted results for $\tan\beta = 2$ and 20, taking $m_{\tilde{t}} = 1$ TeV and including two-loop corrections following Refs. [18, 19] neglecting squark mixing; SUSY decay channels are assumed to be absent.

is

$$\overline{\sigma}_h = \frac{2\pi^2 \Gamma(h \to \mu\mu)\, BF(h \to X)}{m_h^2} \times \frac{1}{\sigma_{\sqrt{s}}\sqrt{2\pi}} \,. \tag{2.21}$$

In the other extreme, where the Higgs width is much broader than $\sigma_{\sqrt{s}}$, at $\sqrt{s} = m_h$ we obtain

$$\overline{\sigma}_h = \frac{4\pi BF(h \to \mu\mu) BF(h \to X)}{m_h^2} \,. \tag{2.22}$$

Figure 2.4 illustrates the result of this convolution as a function of $\sqrt{s}$ for $\sqrt{s}$ near m_h in the three situations: $\Gamma_h^{\rm tot} \ll \sigma_{\sqrt{s}}$, $\Gamma_h^{\rm tot} \sim \sigma_{\sqrt{s}}$ and $\Gamma_h^{\rm tot} \gg \sigma_{\sqrt{s}}$. We observe that small R greatly

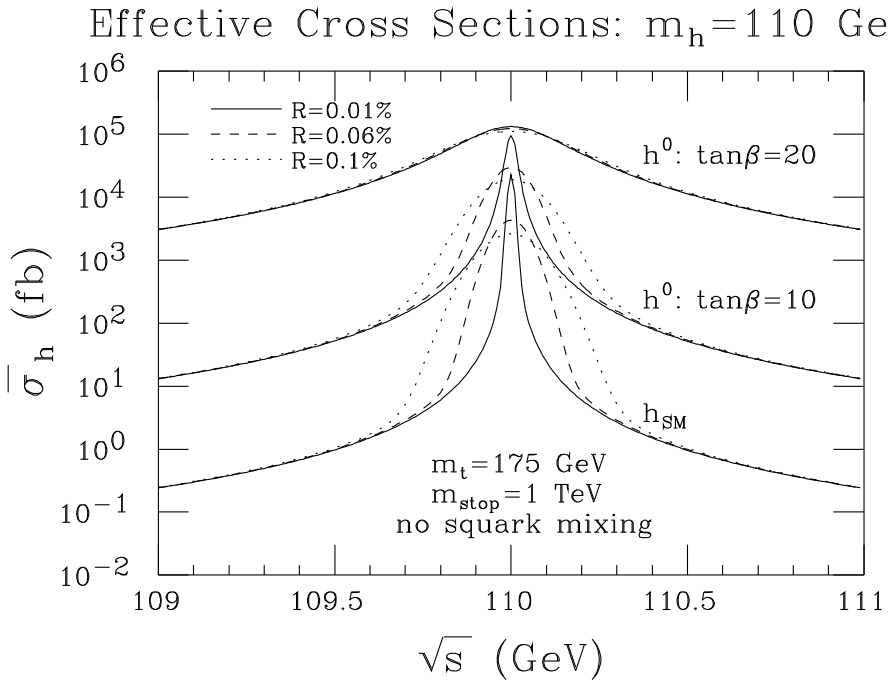


Figure 2.4: The effective cross section, $\overline{\sigma}_h$, obtained after convoluting σ_h with the Gaussian distributions for $R = 0.01\%$, $R = 0.06\%$, and $R = 0.1\%$, is plotted as a function of $\sqrt{s}$ taking $m_h = 110$ GeV.

enhances the peak cross section for $\sqrt{s} = m_h$ when $\Gamma_h^{\rm tot} \ll \sigma_{\sqrt{s}}$, as well as providing an opportunity to directly measure $\Gamma_h^{\rm tot}$.

As an illustration, suppose $m_h \sim 110$ GeV and h is detected in $e^+e^- \to Zh$ or $\mu^+\mu^- \to Zh$ with mass uncertainty $\delta m_h \sim \pm 0.8$ GeV (obtained with luminosity $L \sim 1$ fb^{-1}). For a standard model Higgs of this mass, the width is about 3.1 MeV. How many scan points and how much luminosity are required to zero in on $m_{h_{SM}}$ to within one rms spread $\sigma_{\sqrt{s}}$? For $R = 0.01\%$ ($R = 0.06\%$), $\sigma_{\sqrt{s}} \sim 7.7$ MeV (~ 45 MeV) and the number of scan points required to cover the 1.6 GeV mass zone at intervals of $\sigma_{\sqrt{s}}$ will be 230 (34), respectively. The luminosity required to observe (or exclude) the Higgs at each point is $L \gtrsim 0.01$ fb^{-1} ($L \gtrsim 0.3$ fb^{-1}) for $R = 0.01\%$ ($R = 0.06\%$). Thus, the total luminosity required to zero in on the Higgs will be ~ 2.3 fb^{-1} (~ 10.2 fb^{-1}) in the two cases.

More generally, the L required at each scan point decreases as (roughly) $R^{1.7}$, whereas the number of scan points only grows like $1/R$, implying that the total L required for the scan decreases as $\sim R^{0.7}$. Thus, the $\mu^+\mu^-$ collider should be constructed with the smallest

possible R value with the proviso that the number of $\sqrt{s}$ settings can be correspondingly increased for the required scan. It must be possible to quickly and precisely adjust the energy of the $\mu^+\mu^-$ collider to do the scan.

To measure the width of a SM-like Higgs boson, one would first determine m_h to within $d\sigma_{\sqrt{s}}$ with $d \lesssim 0.3$ and then measure the cross section accurately at the wings of the excitation peak, see Fig. 2.4. The two independent measurements of $\sigma_{\rm wings}/\sigma_{\rm peak}$ give improved precision for the Higgs mass and determine the Higgs width. It is advantageous to put more luminosity on the wings than the peak. Thus, to extract the total width we propose the following procedure[16]. First, conduct a rough scan to determine m_h to a precision $\sigma_{\sqrt{s}}d$, with $d \lesssim 0.3$. Then perform three measurements. At $\sqrt{s}_1 = m_h + \sigma_{\sqrt{s}}d$ expend a luminosity L_1 and measure the total rate $N_1 = S_1 + B_1$. Then perform measurements at

$$\sqrt{s}_2 = \sqrt{s}_1 - n_{\sigma_{\sqrt{s}}}\sigma_{\sqrt{s}} \tag{2.23}$$

and one at

$$\sqrt{s}_3 = \sqrt{s}_1 + n_{\sigma_{\sqrt{s}}}\sigma_{\sqrt{s}} \tag{2.24}$$

yielding $N_2 = S_2 + B_2$ and $N_3 = S_3 + B_3$ events, respectively, with luminosities of $L_2 = \rho_2 L_1$ and $L_3 = \rho_3 L_1$. The backgrounds can be determined from measurements farther from the resonance or from theoretical predictions. Next evaluate the ratios $r_2 = (S_2/\rho_2)/S_1$ and $r_3 = (S_3/\rho_3)/S_1$, for which the partial decay rates in the numerator in Eq. (2.55) cancel out. Since the excitation curve has a specific shape given by convoluting the denominator in Eq. (2.55) with the Gaussian distribution, these measured ratios determine the mass and total width of the Higgs boson. We find that the choices $n_{\sigma_{\sqrt{s}}} \simeq 2$ and $\rho_2 = \rho_3 \simeq 2.5$ are roughly optimal when $\sigma_{\sqrt{s}} \gtrsim \Gamma_h^{\rm tot}$. For these choices and $R = 0.01\%$, a total luminosity $L = L_1 + L_2 + L_3$ of 2 fb^{-1} (200 fb^{-1}) would be required to measure $\Gamma_h^{\rm tot}$ with an accuracy of $\pm 30\%$ for $m_h = 110$ GeV ($m_h = m_Z$). An accuracy of $\pm 10\%$ for $\Gamma_h^{\rm tot}$ could be achieved for reasonable luminosities provided m_h is not near m_Z.

It must be stressed that the ability to precisely determine the energy of the machine when the three measurements are taken is crucial for the success of the three-point technique. A mis-determination of the *spacing* of the measurements in Eqs. (2.66) and (2.67) by just 3% would result in an error in $\Gamma_{h_{SM}}^{\rm tot}$ of 30%. This does not present a problem provided some polarization of the beam can be achieved so that the precession of the spin of the muon as it circulates in the final storage ring can be measured. Given this and the rotation rate, the

energy can be determined to the nearly 1 part in a million accuracy required. This energy calibration capability must be incorporated in the machine design from the beginning.

The other quantity that can be measured with great precision at a $\mu^+\mu^-$ collider for a SM-like Higgs with $m_h \lesssim 130$ GeV is $G(b\bar{b}) \equiv \Gamma(h \to \mu^+\mu^-)BF(h \to b\bar{b})$. For $L = 50$ fb^{-1} and $R = 0.01\%, 0.06\%$, $G(b\bar{b})$ can be measured with an accuracy of $\pm 0.4\%, \pm 2\%$ ($\pm 3\%, \pm 15\%$) at $m_h = 110$ GeV ($m_h = m_Z$). By combining this measurement with the $\pm \sim 7\%$ determination of $BF(h \to b\bar{b})$ that could be made in the Zh production mode, a roughly $\pm 8 - 10\%$ determination of $\Gamma(h \to \mu^+\mu^-)$ becomes possible. ($R = 0.01\%$ is required if $m_h \sim m_Z$.)

Suppose we find a light Higgs h and measure its mass, total width and partial widths. The critical questions that then arise are:

- Can we determine if the particle is a SM Higgs or a supersymmetric Higgs?
- If the particle is a supersymmetric Higgs boson, say in the MSSM, can we then predict masses of the heavier Higgs bosons H^0, A^0, and $H^\pm$ in order to discover them in subsequent measurements?

In the context of the MSSM, the answers to these questions can be delineated.

Enhancements of $\Gamma_h^{\rm tot}$ of order 30% relative to the prediction for the SM h_{SM} are the norm (even neglecting possible SUSY decays) for $m_{A^0} \lesssim 400$ GeV. A 10% measurement of $\Gamma_h^{\rm tot}$ would thus be relatively likely to reveal a 3σ statistical enhancement. However, using the deviation to determine the value of m_{A^0} is model-dependent. For example, if $m_h = 110$ GeV and there is no stop mixing, then the percentage deviation would fairly uniquely fix m_{A^0}, whereas if $m_h = 110$ GeV and there is maximal stop mixing, as defined in Ref. [14], then the measured deviation would only imply a relation between $\tan\beta$ and m_{A^0}.

$\Gamma_h^{\rm tot}$ could be combined with branching ratios to yield a more definitive determination of m_{A^0}. For instance, we can compute $\Gamma(h \to b\bar{b}) = \Gamma_h^{\rm tot} BF(h \to b\bar{b})$ using $BF(h \to b\bar{b})$ as measured in Zh production. It turns out that the percentage deviation of this partial width for the h^0 from the h_{SM} prediction is rather independent of $\tan\beta$ and gives a mixing-independent determination of m_{A^0}, which, after including systematic uncertainties in our knowledge of m_b, would discriminate between a value of $m_{A^0} \leq 300$ GeV vs. $m_{A^0} = \infty$ at the $\geq 3\sigma$ statistical level.

Returning to $\Gamma(h \to \mu^+\mu^-)$, deviations at the $\gtrsim 3\sigma$ statistical level in the prediction for this partial width for the h^0 as compared to the h_{SM} are predicted out to $m_{A^0} \gtrsim 400$ GeV. Further, the percentage of deviation from the SM prediction would provide a relatively

accurate determination of m_{A^0} for $m_{A^0} \lesssim 400$ GeV. For example, if $m_h = 110$ GeV, $\Gamma(h^0 \to \mu^+\mu^-)$ changes by 20% (a $\gtrsim 2\sigma$ effect) as m_{A^0} is changed from 300 GeV to 365 GeV.

Deviations for other quantities, e.g. $BF(h \to b\bar{b})$, depend upon the details of the stop squark masses and mixings, the presence of SUSY decay modes, and so forth, much as described in the case of $\Gamma_h^{\rm tot}$. Only partial widths provide a mixing-independent determination of m_{A^0}. The $\mu^+\mu^-$ collider provides, as described, as least two particularly unique opportunities for determining two very important partial widths, $\Gamma(h \to b\bar{b})$ and $\Gamma(h \to \mu^+\mu^-)$, thereby allowing a test of the predicted proportionality of these partial widths to fermion mass independent of the lepton/quark nature of the fermion.

Thus, if $m_{A^0} \lesssim 400$ GeV, we may gain some knowledge of m_{A^0} through precision measurements of the h^0's partial widths. This would greatly facilitate direct observation of the A^0 and H^0 via s-channel production at a $\mu^+\mu^-$ collider with $\sqrt{s} \lesssim 500$ GeV. As discussed in more detail shortly, even without such pre-knowledge of m_{A^0}, discovery of the A^0, H^0 Higgs bosons would be possible in the s-channel at a $\mu^+\mu^-$ collider provided that $\tan\beta \gtrsim 3-4$. With pre-knowledge of m_{A^0}, detection becomes possible for $\tan\beta$ values not far above 1, provided $R \sim 0.01\%$ (crucial since the A^0 and H^0 become relatively narrow for low $\tan\beta$ values).

Other colliders offer various mechanisms to directly search for the A^0, H^0, but also have limitations:

- The LHC has a discovery hole and "h^0-only" regions at moderate $\tan\beta$, $m_{A^0} \gtrsim$ 200 GeV.

- At the NLC one can use the mode $e^+e^- \to Z^\star \to H^0A^0$ (the mode h^0A^0 is suppressed for large m_{A^0}), but it is limited to $m_{H^0} \sim m_{A^0} \lesssim \sqrt{s}/2$.

- A $\gamma\gamma$ collider could probe heavy Higgs up to masses of $m_{H^0} \sim m_{A^0} \sim 0.8\sqrt{s}$, but this would quite likely require $L \sim 100$ fb^{-1}, especially if the Higgs bosons are at the upper end of the $\gamma\gamma$ collider energy spectrum[20].

Most GUT models predict $m_{A^0} \gtrsim 200$ GeV, and perhaps as large as a TeV[21]. For large $m_{A^0} \sim m_{H^0}$, s-channel searches can be made at a $\mu^+\mu^-$ collider up to $\sim \sqrt{s}$, whereas the $Z^\star \to H^0A^0$ mode at an e^+e^- collider fails for $m_{A^0} \sim m_{H^0} \gtrsim \sqrt{s}/2$. In particular, at a muon collider with $\sqrt{s} \sim 500$ GeV, scan detection of the A^0, H^0 is possible in the mass range from 200 to 500 GeV in s-channel production, provided $\tan\beta \gtrsim 3-4$, whereas an e^+e^- collider

of the same energy can only probe $m_{H^0} \sim m_{A^0} \lesssim 220$ GeV. That the signals become viable when $\tan\beta > 1$ (as favored by GUT models) is due to the fact that the couplings of A^0 and (once $m_{A^0} \gtrsim 150$ GeV) H^0 to $b\overline{b}$ and, especially to $\mu^+\mu^-$, are proportional to $\tan\beta$, and thus increasingly enhanced as $\tan\beta$ rises.

Although the $\mu^+\mu^-$ collider cannot discover the H^0, A^0 in the $\tan\beta \lesssim 3$ region, this is a range in which the LHC *could* find the heavy Higgs bosons in a number of modes. That the LHC and the NMC are complementary in this respect is a very crucial point. Together, discovery of the A^0, H^0 is essentially guaranteed.

If the H^0, A^0 are observed at the $\mu^+\mu^-$ collider, measurement of their widths will typically be straightforward. For moderate $\tan\beta$ the A^0 and H^0 resonance peaks do not overlap and $R \lesssim 0.06\%$ will be adequate, since for such R values $\Gamma_{H^0,A^0} \gtrsim \sigma_{\sqrt{s}}$. However, if $\tan\beta$ is large, then for most of the $m_{A^0} \gtrsim 200$ GeV parameter range the A^0 and H^0 are sufficiently degenerate that there is significant overlap of the A^0 and H^0 resonance peaks. In this case, $R \sim 0.01\%$ resolution would be necessary for observing the double-peaked structure and separating the A^0 and H^0 resonances.

A $\sqrt{s} \sim 500$ GeV muon collider still might not have sufficient energy to discover heavy supersymmetric Higgs bosons. Further, distinguishing the MSSM from the SM by detecting small deviations of the h^0 properties from those predicted for the h_{SM} becomes quite difficult for $m_{A^0} \gtrsim 400$ GeV. However, construction of a higher energy machine, say $\sqrt{s} = 4$ TeV, would allow discovery of A^0, H^0 in the $b\overline{b}$ or $t\overline{t}$ channels (see the discussion in Section 2.2.5).

We close this section with brief comments on the effects of bremsstrahlung and beam polarization. Soft photon radiation must be included when determining the resolution in energy and the peak luminosity achievable at an e^+e^- or $\mu^+\mu^-$ collider. This radiation is substantially reduced at a $\mu^+\mu^-$ collider due to the increased mass of the muon compared to the electron. In Fig. 2.5 we show the luminosity distribution before and after including the soft photon radiation. These bremsstrahlung effects are calculated in Ref. [16]. A long tail extends down to low values of the energy.

For a SM-like Higgs boson with width smaller than $\sigma_{\sqrt{s}}$, the primary effect of bremsstrahlung is a reduction in the peak luminosity. The ratio of the luminosity peak height after and before including the bremsstrahlung is shown in Fig. 2.6. The conclusions above regarding s-channel Higgs detection are those obtained with inclusion of bremsstrahlung effects. The low-energy bremsstrahlung tail provides a self-scan over the range of energies below the design energy, and thus can be used to detect s-channel resonances. The full

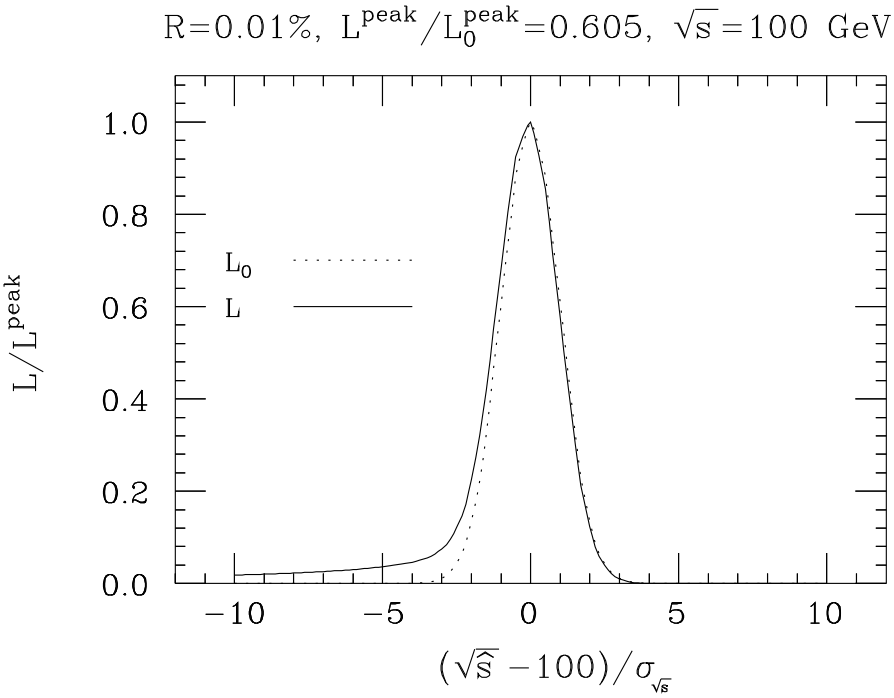


Figure 2.5: $d\mathcal{L}/d\sqrt{\hat{s}}$ relative to its peak value at $\sqrt{\hat{s}} = \sqrt{s}$ is plotted before and after soft-photon radiation. We have taken $\sqrt{s} = 100$ GeV and $R = 0.01\%$. The ratio of peak height after including soft-photon radiation to that before is 0.605.

luminosity distribution for the tail is shown in Fig. 2.7. Observation of A^0, H^0 peaks in the $b\bar{b}$ mass distribution $m_{b\bar{b}}$ created by this bremsstrahlung tail may be possible. The region of the $(m_{A^0}, \tan\beta)$ parameter space plane for which a peak is observable depends strongly on the $b\bar{b}$ invariant mass resolution. For an excellent $m_{b\bar{b}}$ mass resolution of order ± 5 GeV and integrated luminosity of $L = 50$ fb^{-1} at $\sqrt{s} = 500$ GeV, the A^0, H^0 peak(s) are observable for $\tan\beta \gtrsim 5$ at $m_{A^0} \gtrsim 400$ GeV (but only for very large $\tan\beta$ values in the $m_{A^0} \sim m_Z$ region due to the large s-channel Z contribution to the $b\bar{b}$ background).

In the s-channel Higgs studies, polarization of the muon beams could present a significant advantage over the unpolarized case, since signal and background come predominantly from different polarization states. Polarization P of both beams would enhance the significance of a Higgs signal provided the factor by which the luminosity is reduced is not larger than $(1 + P^2)^2/(1 - P^2)$. For example, a reduction in luminosity by a factor of 10 could be compensated by a polarization $P = 0.84$, leaving the significance of the signal unchanged[22].

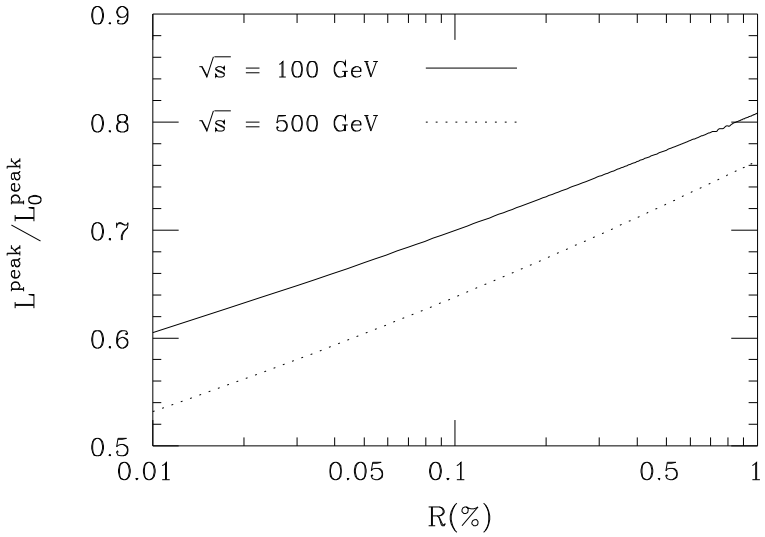


Figure 2.6: $\frac{d\mathcal{L}}{d\sqrt{\hat{s}}}/\frac{d\mathcal{L}_0}{d\sqrt{\hat{s}}}\Big|_{\sqrt{\hat{s}}=\sqrt{s}}$ as a function of R for $\sqrt{s} = 100$ and 500 GeV.

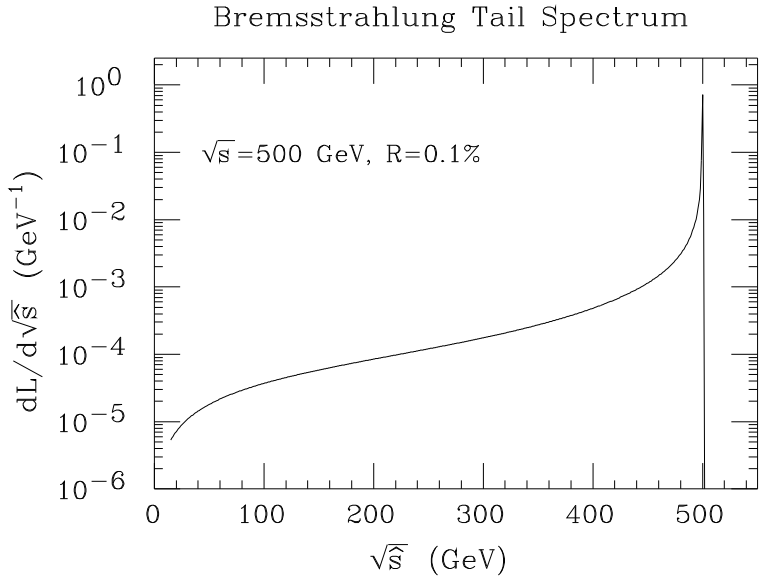


Figure 2.7: $\frac{d\mathcal{L}}{d\sqrt{\hat{s}}}$ as a function of $\sqrt{\hat{s}}$ for $R = 0.1\%$ and $\sqrt{s} =$ 500 GeV. The integral under the curve is normalized to 1.

Furthermore, *transverse* polarization of the muon beams could prove useful for studying CP-violation in the Higgs sector. Muons are produced naturally polarized from π and K decays.

An important consideration for the future design of muon colliders is the extent to which polarization can be maintained through the cooling and acceleration processes.

2.2.3 Precision Threshold Studies

Good beam energy resolution is crucial for the determination of the Higgs width. Another area of physics where the naturally good resolution of a $\mu^+\mu^-$ collider would prove valuable is studies of the $t\bar{t}$ and W^+W^- thresholds, similar to those proposed for the NLC and LEP II. The $t\bar{t}$ threshold shape determines m_t, Γ_t and the strong coupling α_s, while the W^+W^- threshold shape determines m_W and possibly also Γ_W. At a $\mu^+\mu^-$ collider, even a conservative natural beam resolution $R \sim 0.1\%$ would allow substantially increased precision in the measurement of most of these quantities as compared to other machines. Not only is such monochromaticity already greatly superior to e^+e^- collider designs, where typically $R \sim 1\%$, but also at a $\mu^+\mu^-$ collider there is no significant beamstrahlung and the amount of initial state radiation (ISR) is greatly reduced. ISR and, especially, beam smearing cause significant loss of precision in the measurement of the top quark and W masses at e^+e^- colliders.

To illustrate, consider threshold production of the top quark, which has been extensively studied for e^+e^- colliders[24]. Figure 2.8 shows the effects of including beam smearing and ISR for the threshold production of top quarks using a Gaussian beam spread of 1% for the e^+e^- collider[25]. Also shown are our corresponding results for the $\mu^+\mu^-$ collider with $R = 0.1\%$, see [25]. The threshold peak is no longer washed out in the $\mu^+\mu^-$ case. The precision with which one could measure m_t, α_s and Γ_t at various facilities is shown in Table 2.1. Improvements in the determination of m_W should also be possible[23].

The value of such improvements in precision can be substantial. Consider precision electroweak corrections, for example. The prediction for the SM or SM-like Higgs mass m_h depends on m_W and m_t through the one-loop equation

$$m_W^2 = m_Z^2 \left[1 - \frac{\pi\alpha}{\sqrt{2}G_\mu m_W^2(1-\delta r)}\right]^{1/2}, \tag{2.25}$$

where δr depends quadratically on m_t and logarithmically on m_h. Current expectations for LEP II and the Tevatron imply precisions of order

$$\Delta m_W = 40 \text{ MeV}, \tag{2.26}$$

$$\Delta m_t = 4 \text{ GeV}. \tag{2.27}$$

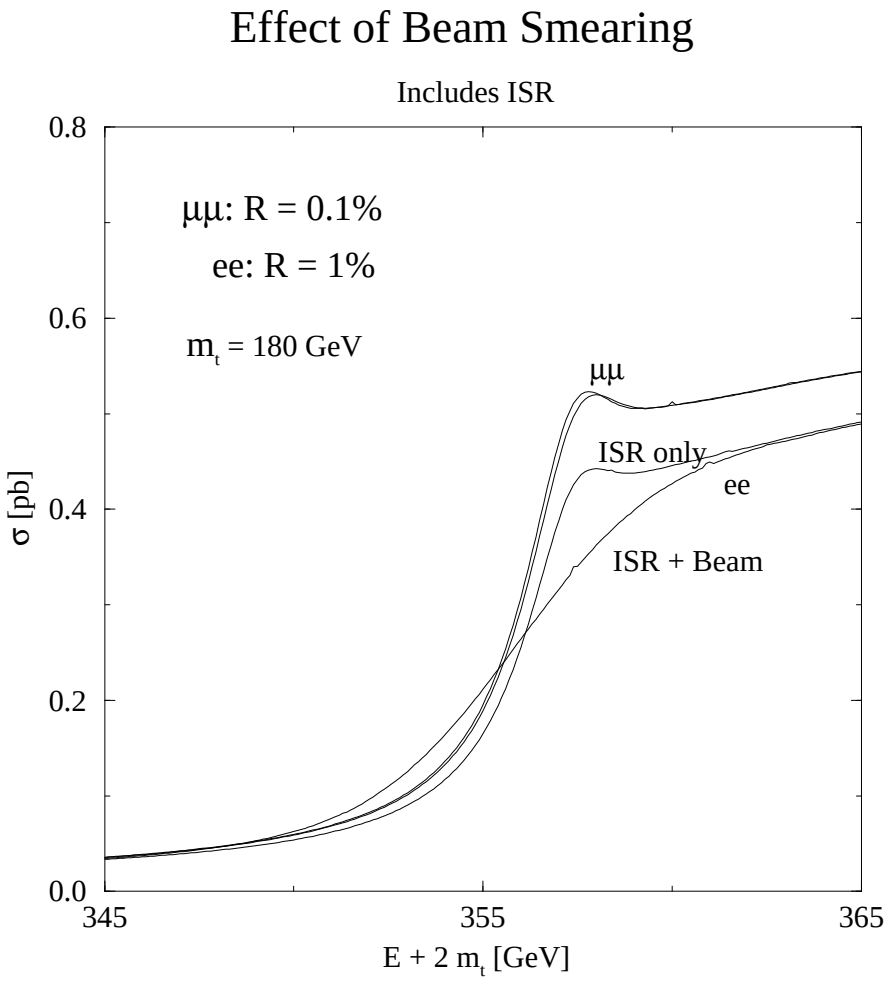


Figure 2.8: The threshold curves are shown for $\mu^+\mu^-$ and e^+e^- machines including ISR and with and without beam smearing. Beam smearing has only a small effect at a muon collider, whereas at an electron collider the threshold region is significantly smeared. The strong coupling is taken to be $\alpha_s(m_Z) = 0.12$.

For the uncertainties of Eq. (2.27) and the current central values of $m_W = 80.4$ GeV and $m_t = 180$ GeV, the Higgs mass would be constrained to the 1σ range

$$50 < m_h < 200 \text{ GeV} . \tag{2.28}$$

In electroweak precision analysis, an error of $\Delta m_W = 40$ MeV is equivalent to an error of $\Delta m_t = 6$ GeV, so increased precision for m_W would be of greatest immediate interest given

Table 2.1: Measurements of the standard model parameters: top mass m_t, strong coupling α_s, and top quark width Γ_t.

	Tevatron ($1000\ pb^{-1}$) ($10\ fb^{-1}$)	LHC ($20\ pb^{-1}$)	NLC ($10\ fb^{-1}$)	FMC ($10\ fb^{-1}$)
Δm_t(GeV)	4 1	2	0.52[a]	0.3
$\Delta\alpha_s$			0.009	0.008
$\Delta\Gamma_t/\Gamma_t$	0.3[b]		0.2	better

[a] See Ref.[26]
[b] See Ref.[27]

the $\Delta m_t = 4$ GeV error quoted above. In order to make full use of the $\Delta m_t \lesssim 0.5$ GeV precision possible at a $\mu^+\mu^-$ collider would require $\Delta m_W \lesssim 4$ MeV. We are currently studying the possibility that the latter can be achieved at a $\mu^+\mu^-$ collider.

Such precisions, combined with the essentially exact determination of m_h possible at a $\mu^+\mu^-$ collider, would allow a consistency test for precision electroweak measurements at a hitherto unimagined level of accuracy. If significant inconsistency is found, new physics could be revealed. For example, inconsistency could arise if the light h is not that of the SM but rather the h^0 of the MSSM and there is a contribution to precision electroweak quantities arising from the H^0 of the MSSM having a non-negligible WW, ZZ coupling. The contributions of stop and chargino states to loops would be another example.

A precise determination of the top quark mass m_t could well be important in its own right. One scenario is that the low-energy spectrum of particles (SUSY or not) has been measured and there is a desert up to the GUT scale. We would then want to extrapolate the low-energy parameters up to the grand unified scale to test in a detailed way the physics at that scale. Then the top quark mass (and the Yukawa coupling) would be crucially important since this parameter determines to a large extent the evolution of all the other Yukawas, including flavor mixings. These considerations become especially important if the top quark Yukawa coupling is determined by an infrared quasi-fixed point for which very small changes in the top quark mass translate into very large changes in the renormalized values of many other parameters in the theory.

2.2.4 CP Violation and FCNC in the Higgs Sector

A nonstandard Higgs sector could have sizable CP-violating effects as well as new flavor changing neutral current (FCNC) effects that could be probed with a $\mu^+\mu^-$ collider. A general two Higgs doublet model has been studied in Refs. [28, 29, 30]. There one would either (i) measure correlations in the final state, or (ii) transversely polarize the muon beams to observe an asymmetry in the production rate as a function of spin orientation. For the second option, the ability to achieve transverse polarization with the necessary luminosity is a crucial consideration.

New FCNC effects could be studied as well[31]. For example a Higgs in the s-channel could exhibit the decay $\mu^+\mu^- \to H^0 \to t\overline{c}$. This decay would have to compete against the $WW^\star$ decays.

2.2.5 Exotic Higgs Bosons/Scalars

In general, a muon collider can probe any type of scalar that has significant fermionic couplings. Interesting new physics could be revealed. To give one example, consider the possibility that a doubly-charged Higgs boson with lepton-number-violating coupling $\Delta^{--} \to \ell^-\ell^-$ exists, as required in left-right symmetric models where the neutrino mass is generated by the see-saw mechanism through a vacuum expectation value of a neutral Higgs triplet field. Such a Δ^{--} could be produced in $\ell^-\ell^-$ collisions. This scenario was studied in Ref. [32] for an e^-e^- collider, but a $\mu^-\mu^-$ collider would be even better due to the much finer energy resolution (which enhances cross sections) and the fact that the $\Delta^{--} \to \mu^-\mu^-$ coupling should be larger than the $\Delta^{--} \to e^-e^-$ coupling.

Most likely, a Δ^{--} in the $\lesssim 500$ GeV region would already be observed at the LHC by the time the muon collider begins operation. In some scenarios, it would even be observed to decay to $\mu^-\mu^-$ so that the required s-channel coupling would be known to be non-zero. However, the magnitude of the coupling would not be determined; for this we would need the $\mu^-\mu^-$ collider. In the likely limit where $\Gamma_{\Delta^{--}} \ll \sigma_{\sqrt{s}}$, the number of Δ^{--} events for $L = 50$ fb^{-1} is given by

$$N(\Delta^{--}) = 6 \times 10^{11} \left(\frac{c_{\mu\mu}}{10^{-5}}\right) \left(\frac{0.01\%}{R(\%)}\right) , \tag{2.29}$$

where the standard Majorana-like coupling-squared is parameterized as

$$|h_{\mu\mu}|^2 = c_{\mu\mu} m^2_{\Delta^{--}}(\text{ GeV}) . \tag{2.30}$$

Current limits on the coupling correspond to $c_{\mu\mu} \lesssim 5 \times 10^{-5}$. Assuming that 30 to 300 events would provide a distinct signal (the larger number probably required if the dominant Δ^{--} decay channel is into $\mu^-\mu^-$, for which there is a significant $\mu^-\mu^- \to \mu^-\mu^-$ background), the muon collider would probe some 11 to 10 orders of magnitude more deeply in the coupling-squared than presently possible. This is a level of sensitivity that would almost certainly be adequate for observing a Δ^{--} that is associated with the triplet Higgs boson fields that give rise to see-saw neutrino mass generation in the left-right symmetric models.

2.2.6 Physics at a 2⊗2 TeV $\mu^+\mu^-$ Collider

Bremsstrahlung radiation scales like m^{-4}, so a circular storage ring can be used for muons at high energies. A high energy lepton collider with center-of-mass energy of 4 TeV would provide new physics reach beyond that contemplated at the LHC or NLC (with $\sqrt{s} \lesssim 1.5$ TeV). We concentrate primarily on the following scenarios for physics at these energies: (1) heavy supersymmetric (SUSY) particles, (2) strong scattering of longitudinal gauge bosons (generically denoted W_L) in the electroweak symmetry breaking (EWSB) sector, and (3) heavy vector resonance production, like a Z'.

SUSY Factory

Low-energy supersymmetry is a theoretically attractive extension of the Standard Model. Not only does it solve the naturalness problem, but also the physics remains essentially perturbative up to the grand unification scale, and gravity can be included by making the supersymmetry local. Since the SUSY-breaking scale and, hence, sparticle masses are required by naturalness to be no larger than $1 - 2$ TeV, a high energy $\mu^+\mu^-$ collider with $\sqrt{s} = 4$ TeV is guaranteed to be a SUSY factory if SUSY is nature's choice. Indeed, it may be the only machine that would guarantee our ability to study the full spectrum of SUSY particles. The LHC has sufficient energy to produce supersymmetric particles but disentangling the spectrum and measuring the masses will be a challenge due to the complex cascade decays and QCD backgrounds. The NLC would be a cleaner environment than the LHC to study the supersymmetric particle decays, but the problem here may be insufficient energy to completely explore the full particle spectrum.

Most supersymmetric models have a symmetry known as an R-parity that requires that supersymmetric particles be created or destroyed in pairs. This means that the energy

required to find and study heavy scalars is more than twice their mass. (If R-parity is violated, then sparticles can also be produced singly; the single sparticle production rate would depend on the magnitude of the violation, which is model- and generation-dependent.) Further, a p-wave suppression is operative for the production of scalars (in this case the superpartners to the ordinary quarks and leptons), and energies well above the kinematic threshold might be required to produce the scalar pairs at an observable rate, as illustrated in Fig. 2.9. In addition, a large lever arm for exploring the different threshold behavior of spin-0 and spin-1/2 SUSY sparticles could prove useful in mass determinations.

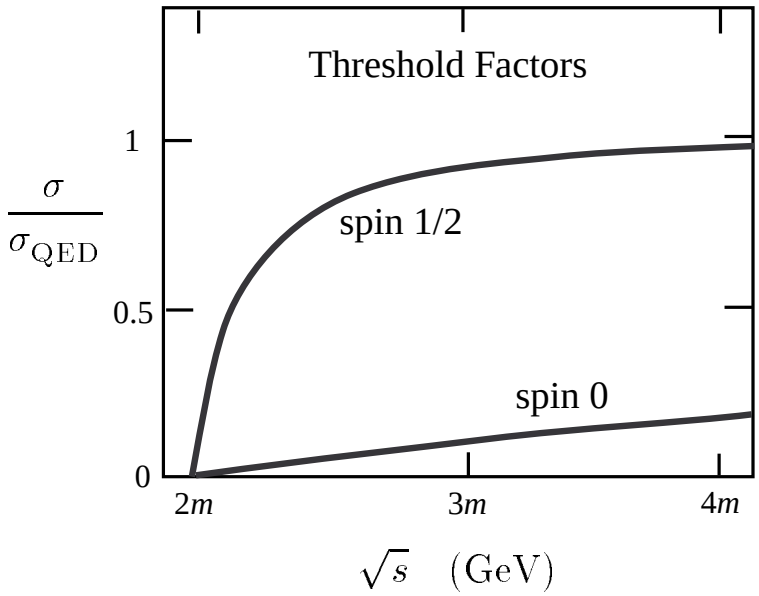


Figure 2.9: Comparison of kinematic suppression for fermion pairs and squark pair production at e^+e^- or $\mu^+\mu^-$ colliders.

To be more specific, it is useful to constrain the parameter space by employing a supergravity (SUGRA) model. Such models are particularly attractive in that the breaking of the electroweak symmetry is accomplished radiatively by the large top quark Yukawa coupling driving one of the Higgs doublet masses negative through renormalization group evolution. The simplest SUGRA models contain the following parameters:

- a universal scalar mass m_0;
- a universal gaugino mass $m_{1/2}$;

- the ratio of the electroweak scale Higgs v's, $\tan\beta = v_2/v_1$;
- a universal trilinear term A_0;
- the sign of the Higgs mixing: sign(μ).

The parameters above are constrained by various means. Experimental bounds on the superpartner masses put a lower bound on $m_{1/2}$. Naturalness considerations yield upper bounds on both $m_{1/2}$ and m_0, which, in turn, imply upper limits on the superparticle masses. If one supposes that the LSP is the cold dark matter of the universe, then there is an upper limit on m_0 so that the annihilation channels for the LSP are not suppressed by the heavy scalar masses. The A_0 parameter is limited by the requirement of an acceptable vacuum state; $1 \lesssim \tan\beta \lesssim 50 - 60$ is required for perturbativity of the Yukawa couplings. A representative choice of parameters that is consistent with all these constraints, but at the same time illustrates the power of a $\mu^+\mu^-$ collider is:

$$\begin{aligned} m_0 = 2m_{1/2} = 500 \text{ GeV} , \\ \tan\beta = 2, \quad A_0 = 0, \quad \mu < 0 . \end{aligned} \tag{2.31}$$

By adopting a large ratio of $m_0/m_{1/2} = 2$ the scalars become heavy (with the exception of the lightest Higgs boson) compared to the gauginos. The particle and sparticle masses obtained from renormalization group evolution are:

$$m_{h^0} = 88 \text{ GeV}, \quad m_{A^0} = 921 \text{ GeV} , \tag{2.32}$$

$$m_{H^\pm} = m_{H^0} = 924 \text{ GeV} , \tag{2.33}$$

$$m_{\tilde{q}_L} \simeq 752 \text{ GeV}, \quad m_{\tilde{q}_R} \simeq 735 \text{ GeV} , \tag{2.34}$$

$$m_{\tilde{b}_1} = 643 \text{ GeV}, \quad m_{\tilde{b}_2} = 735 \text{ GeV} , \tag{2.35}$$

$$m_{\tilde{t}_1} = 510 \text{ GeV}, \quad m_{\tilde{t}_2} = 666 \text{ GeV} , \tag{2.36}$$

$$m_{\tilde{\nu}} \sim m_{\tilde{\ell}} \sim 510 - 530 \text{ GeV} , \tag{2.37}$$

$$m_{\tilde{\chi}^0_{1,2,3,4}} = 107, 217, 605, 613 \text{ GeV} , \tag{2.38}$$

$$m_{\tilde{\chi}^+_{1,2}} = 217, 612 \text{ GeV} . \tag{2.39}$$

Thus, the choice of GUT parameters, Eq. (2.31), leads, as desired, to a scenario such that pair production of heavy scalars is only accessible at a high energy machine like the NMC.

First, we consider the pair production of the heavy Higgs bosons

$$\mu^+\mu^- \;\to\; Z \to H^0 A^0 \,, \tag{2.40}$$

$$\mu^+\mu^- \;\to\; \gamma, Z \to H^+ H^- \,. \tag{2.41}$$

The cross sections are shown in Fig. 2.10 versus $\sqrt{s}$. A $\mu^+\mu^-$ collider with $\sqrt{s} \gtrsim 2$ TeV is needed and well above the threshold the cross section is $\mathcal{O}(1$ fb). In the scenario of Eq. (2.31), the decays of these heavy Higgs bosons are predominantly into top quark modes ($t\bar{t}$ for the neutral Higgs and $t\bar{b}$ for the charged Higgs), with branching fractions near 90%. Observation of the H^0, A^0, and $H^\pm$ would be straightforward even for a pessimistic luminosity of $L = 100$ fb^{-1}. Backgrounds would be negligible once the requirement of roughly equal masses for two back-to-back particles is imposed.

In other scenarios the decays may be more complex and include multiple decay modes into supersymmetric particles, in which case the overall event rate might prove crucial to establishing a signal. In some scenarios investigated in Ref. [33] complex decays are important, but the $\mu^+\mu^-$ collider has sufficient production rate that one or more of the modes

$$(H^0 \to b\bar{b}) + (A^0 \to b\bar{b}) \,, \tag{2.42}$$

$$(H^0 \to h^0 h^0 \to b\bar{b}b\bar{b}) + (A^0 \to X) \,, \tag{2.43}$$

$$(H^0 \to t\bar{t}) + (A^0 \to t\bar{t}) \,, \tag{2.44}$$

are still visible above the backgrounds for $L \gtrsim 500$ fb^{-1}. Despite the significant dilution of the signal by the additional SUSY decay modes (which is most important at low $\tan\beta$), one can observe a signal of $\gtrsim 50$ events in one channel or another.

The high energy $\mu^+\mu^-$ collider will yield a large number of the light SM-like h^0 via $\mu^+\mu^- \to Z^\star \to Zh^0$ and WW fusion, $\mu^+\mu^- \to \nu\overline{\nu}h^0$. In contrast to a machine running at FMC energies ($\sqrt{s} \sim 500$ GeV), where the cross sections for these two processes are comparable, at higher energies, $\sqrt{s} \gtrsim 1$ TeV, the WW fusion process dominates as shown in Fig. 2.10.

Any assessment of the physics signals in the pair production of the supersymmetric partners of the quarks and leptons is model-dependent. However, as illustrated by the specific SUGRA scenario masses of Eq. (2.39), squarks are expected to be somewhat heavier than the sleptons due to their QCD interactions which affect the running of their associated 'soft' masses away from the universal mass m_0 in the evolution from the GUT scale to low

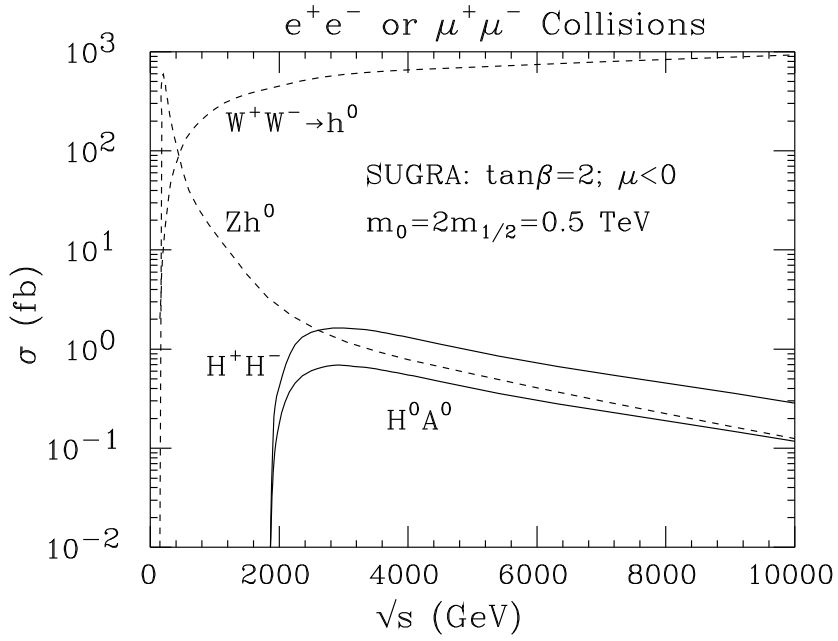


Figure 2.10: Pair production of heavy Higgs bosons at a high energy lepton collider. For comparison, cross sections for the lightest Higgs boson production via the Bjorken process $\mu^+\mu^- \to Z^\star \to Zh^0$ and via the WW fusion are also presented.

energies. Except for the LSP, the lightest superpartner of each type decays to a gaugino (or gluino) and an ordinary fermion, and the gaugino will decay if it is not the LSP. Since the particles are generally too short-lived to be observed, we must infer everything about their production from their decay products.

We illustrate the production cross sections for several important sparticle pairs in Fig. 2.11 for the SUGRA model of Eq. (2.31). For a collider with $\sqrt{s} \sim 4$ TeV, cross sections of ~ 2–30 fb are expected.

The final states of interest are determined by the dominant decay modes, which in this model are $\widetilde{e}_R \to e\widetilde{\chi}_1^0$ ($BF = 0.999$), $\widetilde{\chi}_1^+ \to W^+\widetilde{\chi}_1^0$ ($BF = 0.999$), $\widetilde{d}_L \to \widetilde{\chi}_1^- u, \widetilde{\chi}_2^0 d, \widetilde{g}d$ ($BF = 0.52, 0.27, 0.20$), and $\widetilde{t}_1 \to \widetilde{\chi}_1^+ t$. Thus, for example, with a luminosity of $L = 200$ fb^{-1} at $\sqrt{s} = 4$ TeV, $\widetilde{d}_L$ pair production would result in $200 \times 2 \times (0.52)^2 = 100$ events containing two u-quark jets, two energetic leptons (not necessarily of the same type), and substantial missing energy. The SM background should be small, and the signal would be clearly visible.

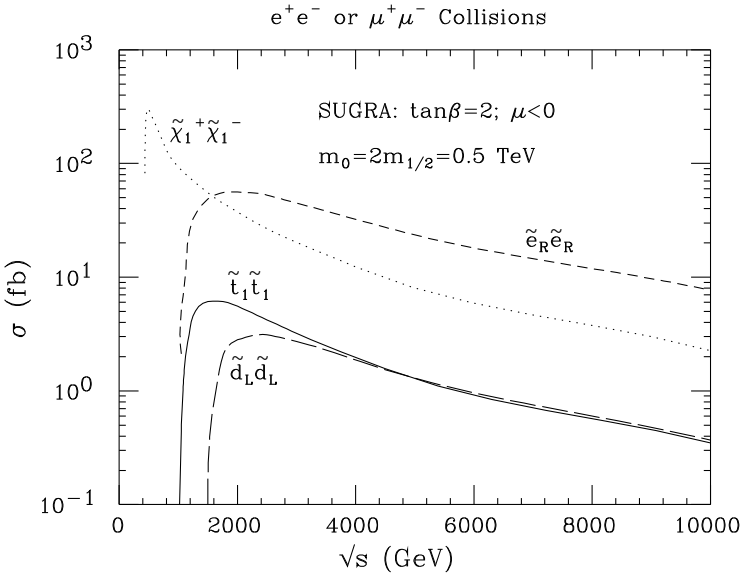


Figure 2.11: The production cross sections for SUSY particles in a supergravity model with heavy scalars.

The energy spectra of the quark jets would allow a determination of $m_{\widetilde{d}_L} - m_{\widetilde{\chi}_1^+}$ while the lepton energy spectra would fix $m_{\widetilde{\chi}_1^+} - m_{\widetilde{\chi}_1^0}$. If the machine energy can be varied, then the turn-on of such events would fix the $\widetilde{d}_L$ mass. The $\widetilde{\chi}_1^+$ and $\widetilde{\chi}_1^0$ masses would presumably already be known from studying the $\ell^+\ell^-$+missing-energy signal from $\widetilde{\chi}_1^+\widetilde{\chi}_1^-$ pair production, best performed at much lower energies. Thus, cross checks on the gaugino masses are possible, while at the same time two determinations of the $\widetilde{d}_L$ mass become available (one from threshold location and the other via the quark jet spectra combined with a known mass for the $\widetilde{\chi}_1^+$).

This example illustrates the power of a $\mu^+\mu^-$ collider, especially one whose energy can be varied over a broad range. Maintaining high luminosity over a broad energy range may require the construction of several (relatively inexpensive) final storage rings.

The $W_LW_L \to W_LW_L$ probe of EWSB

A compelling motivation for building any new machine is to discover the mechanism behind EWSB. This may involve directly producing the Higgs particle of the Standard Model or supersymmetric particles. Alternatively it could be that no light Higgs bosons exist; then

general arguments based on partial wave unitarity require that the interactions of the longitudinal gauge bosons (W and Z) become strong and nonperturbative. The energy scale where this happens is about 1–2 TeV, implying that a collider needs to probe vector boson scattering at energies at least this high. The LHC energy and the currently envisioned NLC energies (up to ~ 1.5 TeV) are marginally able to do this. In contrast, a 4 TeV muon collider is in the optimal energy range for a study of strong vector boson scattering. The construction of a multi-TeV e^+e^- collider is also a possibility[35].)

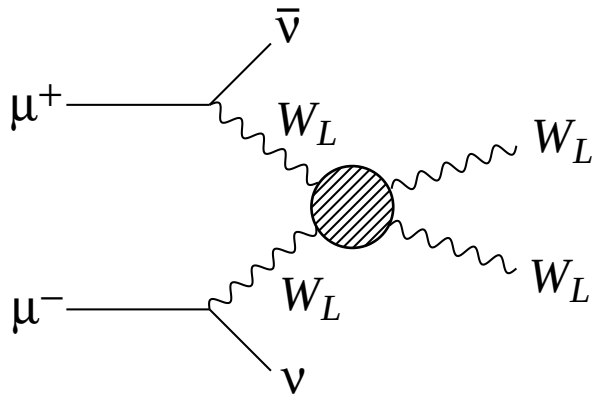


Figure 2.12: Symbolic diagram for strong WW scattering.

Strong electroweak scattering (SEWS) effects can be estimated by using the Standard Model with a heavy Higgs as a prototype of the strong scattering sector. The SM with a light Higgs is an appropriate definition of the electroweak background since only transversely polarized W's contribute to vector boson scattering when the Higgs has a small mass. For a 1 TeV SM Higgs boson, the signal is thus defined as

$$\Delta\sigma = \sigma(m_{h_{SM}} = 1 \text{ TeV}) - \sigma(m_{h_{SM}} = 10 \text{ GeV}) \; . \tag{2.45}$$

Results for $\Delta\sigma$ are shown in Table 2.2 for $\sqrt{s} = 1.5$ TeV (possibly the upper limit for a first e^+e^- collider) and 4 TeV. The strong scattering signal is relatively small at energies of order 1 TeV, but grows substantially as multi-TeV energies are reached. Thus, the highest energies in $\sqrt{s}$ that can be reached at a muon collider could be critically important.

Many other models for the strongly interacting gauge sector have been constructed in addition to the SM, including[36]:

Table 2.2: Strong electroweak scattering signals in $W^+W^- \to W^+W^-$ and $W^+W^- \to ZZ$ at future lepton colliders.

$\sqrt{s}$	$\Delta\sigma(W^+W^-)$	$\Delta\sigma(ZZ)$
1.5 TeV	8 fb	6 fb
4 TeV	80 fb	50 fb

- a ("Scalar") model in which there is a scalar Higgs resonance with $M_S = 1$ TeV but non-SM width of $\Gamma_S = 350$ GeV;
- a ("Vector") model in which there is no scalar resonance, but rather a vector resonance with $M_V = 1$ TeV and $\Gamma_V = 35$ GeV;
- a model, denoted by "LET" or "$m_{h_{SM}} = \infty$", in which the SM Higgs is taken to have infinite mass and the partial waves simply follow the behavior predicted by the low-energy theorems;
- a model (denoted by "LET-K") in which the LET behavior is unitarized via K-matrix techniques.

To differentiate among models, a complete study of the physics of strongly interacting gauge bosons would be required. In particular, all the following vector-boson scattering channels must be studied:

$$W^+W^- \to W^+W^-, ZZ\ , \tag{2.46}$$
$$W^\pm Z \to W^\pm Z\ , \tag{2.47}$$
$$W^\pm W^\pm \to W^\pm W^\pm\ . \tag{2.48}$$

Partial exploration of the three isospin channels can be made at the LHC. The signal and background for gold-plated (purely leptonic) events is shown in Table 2.3 for the LHC operating at 14 TeV with $L = 100$ fb^{-1}, for several of the above models. These channels have also been studied for a 1.5 TeV NLC[37], and, again, event rates are at a level that first signals of the strongly interacting vector boson sector would emerge, but the ability to discriminate between models and actually study these strong interactions would be limited.

For a $\mu^+\mu^-$ collider operating at 4 TeV the statistical significances markedly improve. Table 2.4 summarizes the total signal S and background B event numbers, summing over

Table 2.3: Total numbers of $W_LW_L \to$ 4-lepton signal S and background B events calculated for the LHC[36], assuming $L = 100$ fb^{-1}.

	Bkgd	Scalar	Vector	LET-K
$ZZ(4\ell)$	1	5	1.5	1.5
$(2\ell 2\nu)$	2	17	5	4.5
W^+W^-	12	18	6	5
W^+Z	22	2	70	3
$W^\pm W^\pm$	4	7	12	13

Table 2.4: Total numbers of $W^+W^-, ZZ \to$ 4-jet signal S and background B events calculated for a 4 TeV $\mu^+\mu^-$ collider with integrated luminosity 200 fb^{-1}. Events are summed over the mass range $0.5 < M_{WW} < 1.5$ TeV except for the W^+W^- channel with a narrow vector resonance for which $0.9 < M_{WW} < 1.1$ TeV. The statistical significance $S/\sqrt{B}$ is also given. The hadronic branching fractions of WW decays and the $W^\pm/Z$ identification/misidentification are included.

channels	SM $m_{h_{SM}} = 1$ TeV	Scalar $M_S = 1$ TeV	Vector $M_V = 1$ TeV	SM $m_{h_{SM}} = \infty$
$S(\mu^+\mu^- \to \bar{\nu}\nu W^+W^-)$	1900	1400	370	230
B(backgrounds)	1100	1100	110	1100
$S/\sqrt{B}$	57	42	35	6.9
$S(\mu^+\mu^- \to \bar{\nu}\nu ZZ)$	970	700	220	350
B(backgrounds)	160	160	160	160
$S/\sqrt{B}$	77	55	17	28

diboson invariant mass bins, together with the statistical significance $S/\sqrt{B}$ for different models of the strongly-interacting physics. A broad Higgs-like scalar will enhance both W^+W^- and ZZ channels with $\sigma(W^+W^-) > \sigma(ZZ)$; a ρ-like vector resonance will manifest itself through W^+W^- but not ZZ; while the $m_{h_{SM}} = \infty$ (LET) amplitude will enhance ZZ more than W^+W^-. The $m_{h_{SM}} = \infty$ signal for W^+W^- is visible, although still far from robust; the ratio S/B can be enhanced by making a higher mass cut (*e.g.* $M_{WW} > 0.7$ TeV), but the significance $S/\sqrt{B}$ is not improved.

Signals and the irreducible electroweak background for the W^+W^- and ZZ modes are shown in Fig. 2.13. The complementarity of these two modes is clear from the figure.

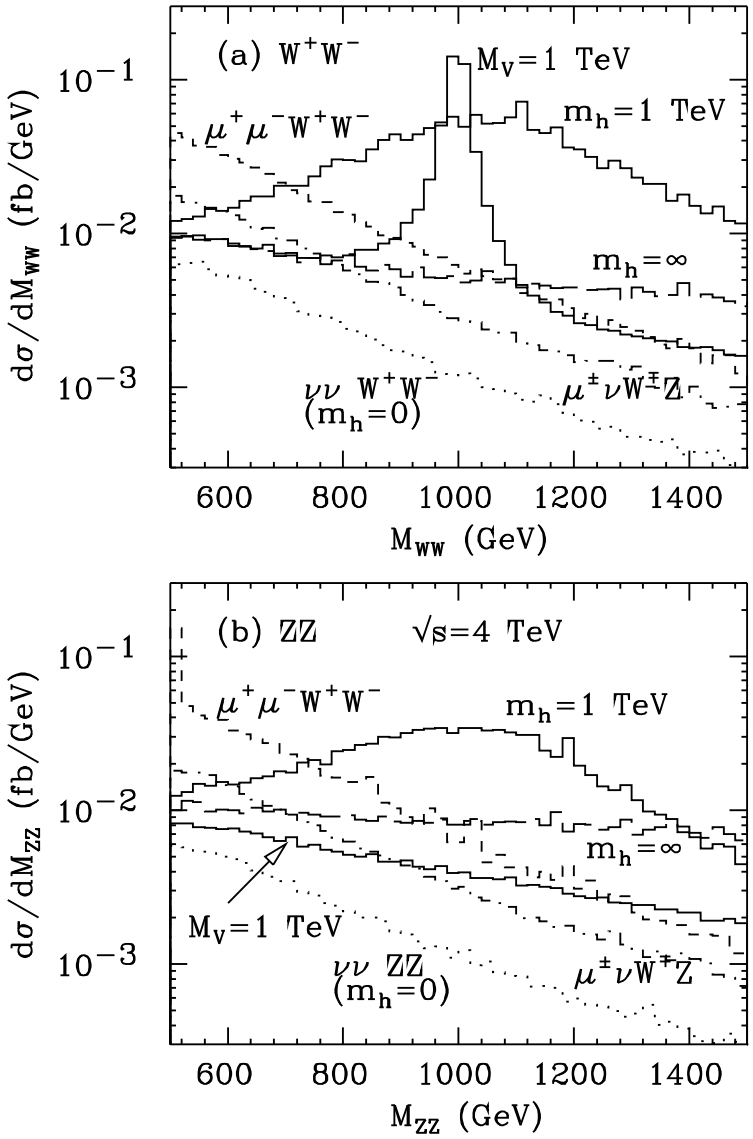


Figure 2.13: Histograms for the signals and backgrounds in strong vector boson scattering in the (a) W^+W^- and (b) ZZ final states. The background is given by the strictly electroweak $m_{h_{SM}} = 0$ limit of the Standard Model. The three signals shown are (I) a vector resonance with $M_V = 1$ TeV, $\Gamma_V = 35$ GeV, (II) the SM Higgs with $m_{h_{SM}} = 1$ TeV, and (III) the SM with $m_{h_{SM}} = \infty$ (LET model). In the figure the shorthand notation h is used for h_{SM}.

However, to make use of this complementarity it is crucial to be able to distinguish final state W and Z bosons using the dijet invariant masses. This is possible provided there is sufficient jet energy resolution, as discussed in Ref. [37].

Finally, we note that event numbers in the 1 TeV SM Higgs and Vector resonance cases, and possibly even in the $m_{h_{SM}} = \infty$ (LET) case, are such that not only could a substantial overall signal be observed, but also at high L the shape of the excess, due to strong interactions, in the distribution in vector boson pair mass could be measured over a broad interval in the 1 TeV range. For instance, from Fig. 2.13a in the case of $m_{h_{SM}} = \infty$, a 100 GeV interval from 1.4 TeV to 1.5 TeV would contain $L \times 100$ GeV $\times (4 \times 10^{-3}$ fb/ GeV$) = 400$ signal events for $L = 1000$ fb^{-1}, thereby allowing a 5% measurement of the $m_{W^+W^-}$ signal distribution in this bin. The level of accuracy in this one bin alone would distinguish this model from the Vector or $m_{h_{SM}} = 1$ TeV models. The difference between the three different distributions plotted in Fig. 2.13 could be tracked in both channels. The ability to measure the distributions with reasonable precision would allow detailed insight into the dynamics of the strongly interacting electroweak sector when the collider achieves energies substantially above 1 TeV. Thus, if some signals for a strongly interacting sector emerge at the LHC, a $\sqrt{s} = 3 - 4$ TeV $\mu^+\mu^-$ (or e^+e^-, if possible) collider will be essential.

Exotic Heavy States

The very high energy of a 4 TeV collider would open up the possibility of directly producing many new particles outside of the Standard Model. Some exotic heavy particles that could be discovered and studied at a muon collider are (1) sequential fermions, $Q\overline{Q}$, $L\overline{L}$[38], (2) lepto-quarks, (3) vector-like fermions[39], and (4) new gauge bosons like a Z' or W_R[40].

A new vector resonance such as a Z' or a technirho, $\rho_{\rm TC}$, is a particularly interesting possibility. The collider could be designed to sit on the resonance $\sqrt{s} \sim M_V$ in which case it would function as a Z' or $\rho_{\rm TC}$ factory as illustrated in Fig. 2.14. Alternatively, if the mass of the resonance is not known a priori, then the collider operating at an energy above the resonance mass could discover it via the bremsstrahlung tail shown in Fig. 2.7. Figure 2.15 shows the differential cross section in the reconstructed final state mass M_V for a muon collider operating at 4 TeV for two cases where the vector resonance has mass 1.5 TeV and 2 TeV. Dramatic and unmistakable signals would appear even for integrated luminosity as low as $L \gtrsim 50 - 100$ fb^{-1}.

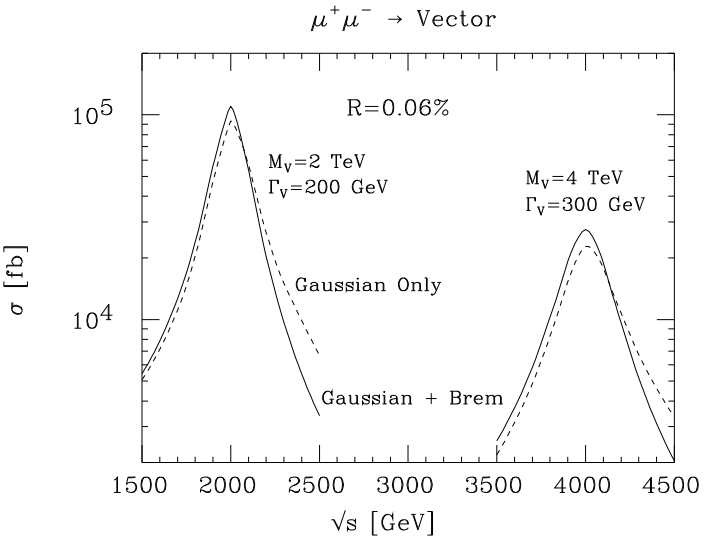


Figure 2.14: High event rates are possible if the muon collider energy is set equal to the vector resonance (Z' or $\rho_{\rm TC}$) mass. Two examples are shown here with $R = 0.06\%$.

2.2.7 Conclusions

A muon collider is very likely to add substantially to our knowledge of physics in the coming decades. A machine with energy in the range $\sqrt{s} = 100$–500 GeV is comparable to the NLC and provides valuable additional features. The most notable of these is the possibility of creating a Higgs boson in the s-channel and measuring its mass and decay widths directly and precisely. Even if a light Higgs does not exist, studies of the $t\bar{t}$ and W^+W^- thresholds at such a low-energy machine would yield higher precision in determining m_t and m_W than possible at other colliders. A $\mu^+\mu^-$ collider with energy as high as $\sqrt{s} \sim 4$ TeV appears to be entirely feasible and is ideally suited for studying a strongly-interacting symmetry breaking sector, since the center-of-mass energy is well above the energy range at which vector boson interactions must become strong. Many other types of exotic physics beyond the Standard Model could be probed at such a high machine energy. For example, if supersymmetry exists, a 4 TeV $\mu^+\mu^-$ collider would be a factory for sparticle pair production. Observation of a heavy Z' in the bremsstrahlung luminosity tail would be straightforward and the machine energy could later be reset to provide a Z' factory. All the issues presented in this paper will be discussed in greater detail in a forthcoming review article[9].

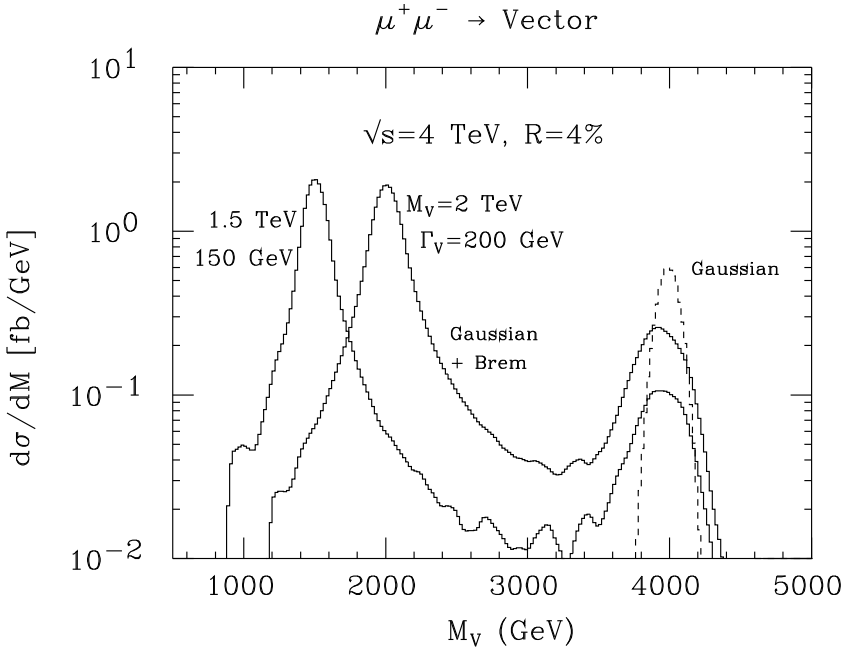


Figure 2.15: A heavy vector resonance can be visible in the bremsstrahlung tail of a high energy collider. Here a $\mu^+\mu^-$ collider operating at 4 TeV is shown for $M_V = 1.5$ TeV and 2 TeV.

2.3 Higgs Boson Physics in the s-Channel at $\mu^+\mu^-$ Colliders

2.3.1 Introduction

Despite the extraordinary success of the Standard Model (SM) in describing particle physics up to the highest energy available today, the mechanism responsible for electroweak symmetry-breaking (EWSB) has yet to be determined. In particular, the Higgs bosons predicted in the minimal Standard Model and the theoretically attractive Supersymmetric (SUSY) Grand Unified Theory (GUT) extensions thereof have yet to be observed. If EWSB does indeed derive from non-zero vacuum expectation values for elementary scalar Higgs fields, then one of the primary goals of constructing future colliders must be to *completely* delineate the associated Higgs boson sector. In particular, it will be crucial to discover all of the physical Higgs bosons and determine their masses, widths and couplings.

The remainder of the introduction is divided into two subsections. In the first, we briefly review crucial properties of the Standard Model and MSSM Higgs bosons. In the second, we outline basic features and parameters of the proposed $\mu^+\mu^-$ colliders, and give a first description of how they relate to our ability to discover and study the SM and MSSM Higgs bosons in s-channel $\mu^+\mu^-$ collisions.

Higgs Bosons in the SM and the MSSM

The EWSB mechanism in the Standard Model is phenomenologically characterized by a single Higgs boson (h_{SM}) in the physical particle spectrum. The mass of the h_{SM} is undetermined by the theory, but its couplings to fermions and vector bosons are completely determined, being given by $gm_f/(2m_W)$, gm_W and $gm_Z/\cos\theta_W$ for a fermion f, the W and the Z, respectively. Although the SM Higgs sector is very simple, it leads to problems associated with naturalness and mass hierarchies which suggest that the SM is simply an effective low-energy theory. Recent summaries of the phenomenology of the SM Higgs sector can be found in Refs. [41, 42].

The most attractive extensions of the SM that solve the naturalness and hierarchy problems are those based on supersymmetry. The Higgs sector of a supersymmetric model must contain at least two Higgs doublet fields in order to give masses to both up and down quarks and to be free of anomalies. If it contains two, and only two, Higgs doublet fields, then the strong and electroweak coupling constants all unify reasonably well at a GUT scale of order 10^{16} GeV. Thus, the minimal supersymmetric Standard Model, defined as having exactly two Higgs doublets, is especially attractive. The resulting spectrum of physical Higgs fields includes three neutral Higgs bosons, the CP-even h^0 and H^0 and the CP-odd A^0. At tree-level the entire Higgs sector is completely determined by choosing values for the parameters $\tan\beta = v_2/v_1$ (where v_2 and v_1 are the vacuum expectation values of the neutral members of the Higgs doublets responsible for up-type and down-type fermion masses, respectively) and m_{A^0} (the mass of the CP-odd A^0). For a summary, see Refs. [41, 42].

In the MSSM there is a theoretical upper bound on the mass of the lightest state h^0 [43, 44] which is approached at large m_{A^0} and large $\tan\beta$. After including two-loop/RGE-improved radiative corrections [45, 46] the bound depends upon the top quark (t) and top squark ($\tilde{t}$) masses and upon parameters associated with squark mixing. Assuming $m_t = 175$ GeV and

$m_{\tilde{t}} \lesssim 1$ TeV, the maximal mass is

$$m_{h^0}^{\max} \sim 113 \text{ to } 130 \text{ GeV}\,, \tag{2.49}$$

depending upon the amount of squark mixing. The 113 GeV value is obtained in the absence of squark mixing. Figure 2.16 illustrates the mass of the h^0 versus the parameter $\tan\beta$ for $m_{A^0} = 100$, 200 and 1000 GeV. Mass contours for the MSSM Higgs bosons are illustrated in Fig. 2.17 in the conventional $m_{A^0}, \tan\beta$ parameter plane. Both these figures include two-loop/RGE-improved radiative corrections to the Higgs masses computed for $m_t = 175$ GeV, $m_{\tilde{t}} = 1$ TeV and neglecting squark mixing.

The Higgs sector of the MSSM can be extended to include extra singlet fields without affecting any of its attractive features. A general supersymmetric model bound of

$$m_{h^0} \lesssim 130 \sim 150 \text{ GeV} \tag{2.50}$$

applies for such non-minimal extensions of the MSSM, assuming a perturbative renormalization group (RGE) evolved grand unified theory (GUT) framework.

The couplings of the MSSM Higgs bosons to fermions and vector bosons are generally proportional to the couplings of the SM Higgs boson, with the constant of proportionality being determined by the angle β (from $\tan\beta$) and the mixing angle α between the neutral Higgs states (α is determined by m_{A^0}, $\tan\beta$, m_t, $m_{\tilde{t}}$, and the amount of stop mixing). Those couplings of interest in this report are [47]

$$\begin{array}{cccccc} & \mu^+\mu^-, b\overline{b} & & t\overline{t} & ZZ, W^+W^- & ZA^0 \\ h^0 & -\sin\alpha/\cos\beta & & \cos\alpha/\sin\beta & \sin(\beta-\alpha) & \cos(\beta-\alpha) \\ H^0 & \cos\alpha/\cos\beta & & \sin\alpha/\sin\beta & \cos(\beta-\alpha) & -\sin(\beta-\alpha) \\ A^0 & -i\gamma_5\tan\beta & & -i\gamma_5/\tan\beta & 0 & 0 \end{array} \tag{2.51}$$

times the Standard-Model factor of $gm_f/(2m_W)$ in the case of fermions (where m_f is the relevant fermion mass), or $gm_W, gm_Z/\cos\theta_W$ in the case of the W, Z, and $g(p_A - p_h)^\mu/2\cos\theta_W$ in the case of ZA^0, where $p_A(p_h)$ is the outgoing momentum of $A^0(h^0, H^0)$.

An important illustrative limit is $m_{A^0} \gtrsim 2m_Z$, since this is typical of SUSY GUT models [48]. In this limit, $\alpha \approx \beta - \pi/2$, $m_{A^0} \sim m_{H^0}$, m_{h^0} approaches its upper limit for the given

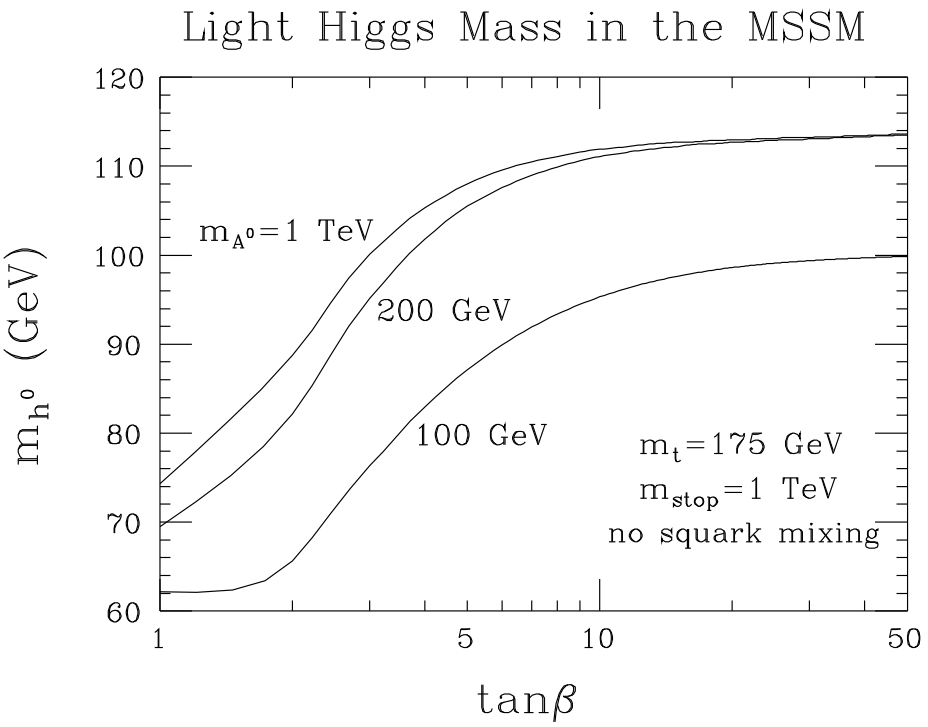


Figure 2.16: m_{h^0} vs $\tan\beta$ for $m_{A^0} = 100$, 200 and 1000 GeV. Two-loop/RGE-improved radiative corrections are included, see Refs. [45, 46], taking $m_t = 175$ GeV, $m_{\tilde{t}} = 1$ TeV and neglecting squark mixing.

value of $\tan\beta$, and the coupling factors of the Higgs bosons are approximately

$$\begin{array}{cccccc} & \mu^+\mu^-, b\bar{b} & & t\bar{t} & ZZ, W^+W^- & ZA^0 \\ h^0 & 1 & & 1 & 1 & 0 \\ H^0 & \tan\beta & & -1/\tan\beta & 0 & -1 \\ A^0 & -i\gamma_5 \tan\beta & & -i\gamma_5/\tan\beta & 0 & 0 \end{array} \tag{2.52}$$

times the Standard-Model factors as given below Eq. (2.51). Thus at large m_{A^0} it is the h^0 which is SM-like, while the H^0, A^0 have similar fermion couplings and small, zero (respectively) tree-level WW, ZZ couplings. Note that the H^0 and A^0 couplings to $\mu^+\mu^-$ and $b\bar{b}$ are enhanced in the (preferred) $\tan\beta > 1$ portion of parameter space.

For $m_{A^0} \lesssim m_Z$, the roles of the h^0 and H^0 are reversed: in this mass range the H^0 becomes roughly SM-like, while the h^0 has couplings (up to a possible overall sign) roughly like those given for H^0 in Eq. (2.52). (See Refs. [42, 49, 41] for details; Ref. [41] gives

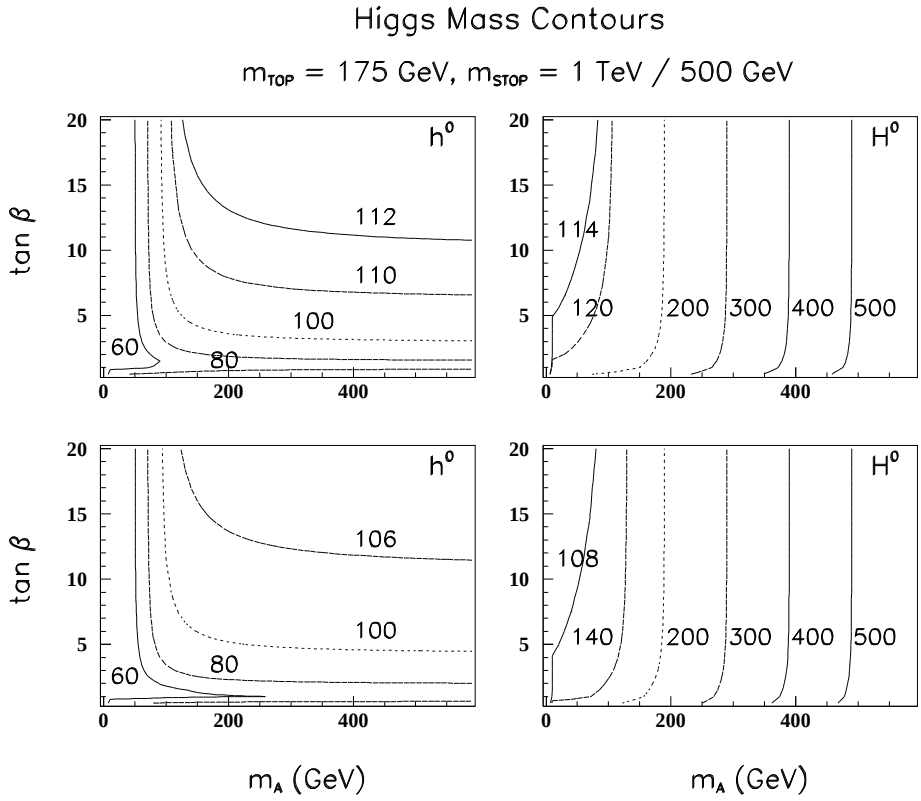


Figure 2.17: Contours for the h^0 and H^0 masses in $(m_{A^0}, \tan\beta)$ parameter space. Results include two-loop/RGE-improved radiative corrections computed for $m_t = 175$ GeV, with $m_{\tilde{t}} = 1$ TeV (upper plots) and $m_{\tilde{t}} = 500$ GeV (lower plots), neglecting squark mixing.

the corrections that imply that the simple rules are only roughly correct after including radiative corrections.) It is also useful to recall [47, 49] that the ZA^0H^0 (ZA^0h^0) coupling is maximal (~ 0) at large m_{A^0}, while at small m_{A^0} the reverse is true. The following discussions emphasize the case of large m_{A^0}.

The Higgs boson widths are crucial parameters for the searches and studies. In particular, we shall see that the width compared to the resolution in $\sqrt{s}$ of the machine is a crucial issue. Widths for the Standard Model Higgs h_{SM} and the three neutral Higgs bosons h^0, H^0, A^0 of the MSSM are illustrated in Fig. 2.3; for the MSSM Higgs bosons, results at $\tan\beta = 2$ and 20 are shown. As a function of $\tan\beta$, the total width of h^0 is plotted in Fig. 2.3.1 for $m_{h^0} = 100$, 110 and 120 GeV. We note that for masses below ~ 130 GeV, both the h_{SM} and a SM-like h^0 have very small widths (in the few MeV range); we will discover

that these widths are often smaller than the expected resolution in $\sqrt{s}$. At high $\tan\beta$ and large $m_{A^0} \sim m_{H^0}$, the $\mu^+\mu^-$, $\tau^+\tau^-$ and $b\bar{b}$ couplings of the H^0 and A^0 are greatly enhanced (being proportional to $\tan\beta$). Consequently, $\Gamma^{\rm tot}_{H^0}$ and $\Gamma^{\rm tot}_{A^0}$ are generally large compared to the expected $\sqrt{s}$ resolution.

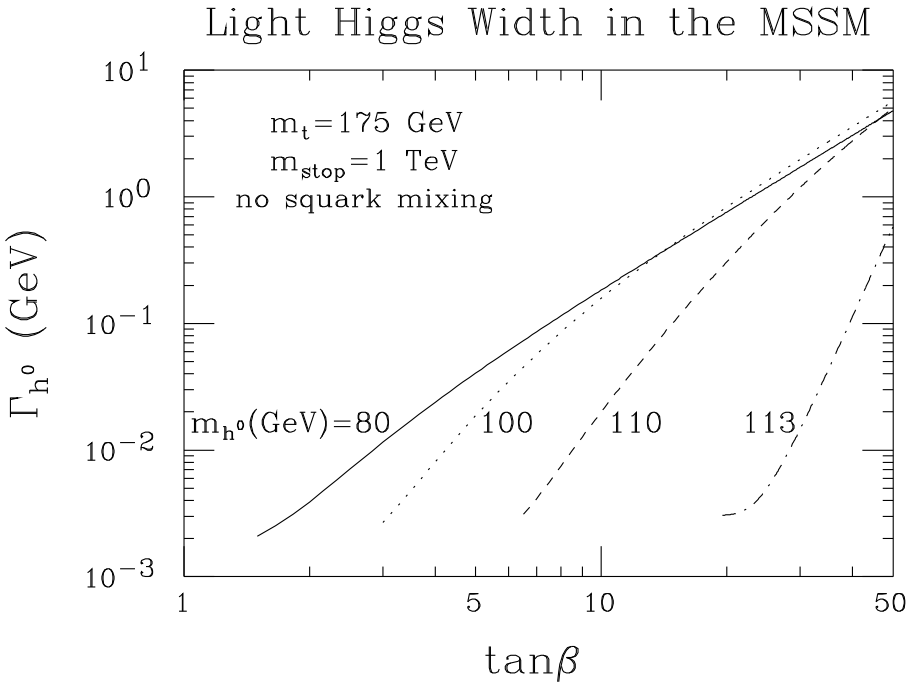


Figure 2.18: $\Gamma^{\rm tot}_{h^0}$ vs $\tan\beta$ for $m_{h^0} = 80$, 100, 110 and 113 GeV, assuming $m_t = 175$ GeV. Two-loop/RGE-improved radiative corrections to Higgs masses, mixing angles and self-couplings have been included, taking $m_{\tilde{t}} = 1$ TeV and neglecting squark mixing. SUSY decay channels are assumed to be absent.

Figure 2.3.1 illustrates the h_{SM} branching fractions for the $\mu^+\mu^-$, $b\bar{b}$, $WW^{(\star)}$ and $ZZ^{(\star)}$ decay modes. For an h_{SM} with $m_{h_{SM}} \lesssim 130$ GeV, the $b\bar{b}$ branching fraction is of order 0.8–0.9, implying that this will be the most useful discovery channel. Once the $WW^{(\star)}$ and $ZZ^{(\star)}$ modes turn on ($m_{h_{SM}} \gtrsim 2m_W$), the h_{SM} becomes broad and the branching fraction $BF(h_{SM} \to \mu^+\mu^-)$, which governs s-channel production, declines precipitously. Branching fractions for the h^0 of the MSSM are similar to those of h_{SM} for $m_{h_{SM}} = m_{h^0}$ when m_{A^0} is large. At high $\tan\beta$ and large $m_{A^0} \sim m_{H^0}$, the enhancement of the $\mu^+\mu^-$, $\tau^+\tau^-$ and $b\bar{b}$ couplings implies that the $b\bar{b}$, $\tau^+\tau^-$ and $\mu^+\mu^-$ branching fractions of the H^0 and A^0 are

the only important ones, and are not unlike those of a light h_{SM}, with relative magnitudes determined by $m_b^2 : m_\tau^2 : m_\mu^2$.

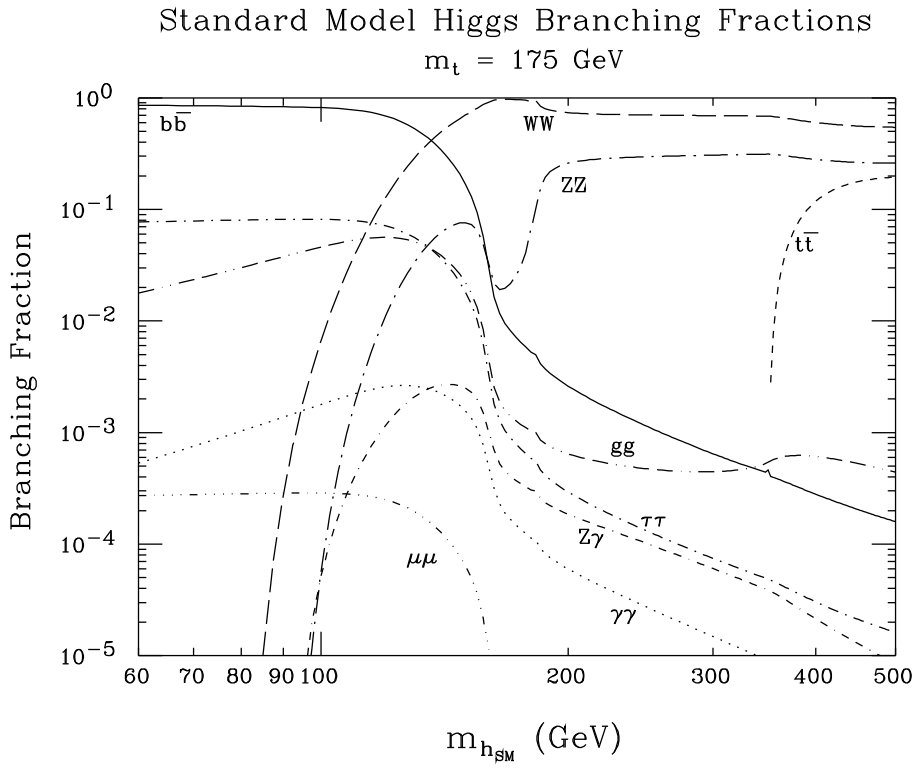


Figure 2.19: Branching fractions for the Standard Model h_{SM}.

Finally, it is relevant to note that in non-minimal extensions of the MSSM, parameter choices are possible such that the lightest Higgs boson to which the bound of Eq. (2.50) applies has very weak coupling to ZZ. This has been demonstrated [50] in the case of the minimal non-minimal supersymmetric model (MNMSSM), which contains one extra singlet Higgs representation, yielding three neutral Higgs bosons in all. However, for parameter choices such that the lightest Higgs decouples from ZZ, there is a strong upper bound on the mass of the least massive Higgs boson *with significant ZZ coupling*. The proof of this fact in the MNMSSM case relies on the observation that as the lighter Higgs bosons decouple from ZZ, the upper bound on the next heaviest Higgs boson moves down. This result may generalize to the case of more singlets.

s-Channel Higgs Boson Physics at $\mu^+\mu^-$ Colliders

The ability of a new accelerator to fully explore EWSB physics weighs heavily in its justification. Recently, there has been much interest in the possibility of constructing a $\mu^+\mu^-$ collider [51, 52, 53, 54], and a survey of the physics opportunities at such a collider has been made [55]. It is currently anticipated that a $\mu^+\mu^-$ collider can, at a minimum, achieve the same integrated luminosities and energies as an e^+e^- collider [56, 57, 58]. Further, with adequate detector segmentation the extra backgrounds resulting from muon decays can be tamed [59]. It then follows that a $\mu^+\mu^-$ collider can essentially explore all the same physics that is accessible at an e^+e^- collider of the same energy. In particular, all the established techniques for probing EWSB at e^+e^- colliders are applicable at a $\mu^+\mu^-$ collider. In addition, should one or more Higgs boson(s) (generically denoted by h) with substantial $\mu^+\mu^-$ coupling(s) exist, a $\mu^+\mu^-$ collider opens up the particularly interesting possibility of direct s-channel $\mu^+\mu^- \to h$ production. The SM Higgs boson, h_{SM}, is a prototypic example. Direct s-channel h_{SM} production is greatly enhanced at a $\mu^+\mu^-$ collider compared to an e^+e^- collider because its coupling to the incoming $\mu^+\mu^-$ is proportional to the lepton mass. Quantitative studies of s-channel Higgs production have been presented in Refs. [55, 60]. With the machine energy set to the Higgs mass ($\sqrt{s} = m_h$) the $\mu^+\mu^- \to h_{SM}$ rate is sufficiently large to allow detection of the h_{SM}, provided that $m_{h_{SM}} \lesssim 2m_W$ (the so-called intermediate Higgs mass region). In addition, *all* the Higgs bosons of the minimal supersymmetric model (MSSM) are produced in sufficient abundance in s-channel $\mu^+\mu^-$ collisions to allow their detection for most of the model parameter space.

In the present report, we expand on these results and provide the documentation underlying the discussion of Ref. [60] on precision studies of both the SM h_{SM} and the MSSM Higgs bosons. We find that the basic properties of the h_{SM} can be determined with remarkable accuracy in $\mu^+\mu^-$ s-channel production, and that the properties of MSSM Higgs bosons can be detailed over a larger fraction of model parameter space than at any other proposed accelerator. One particularly important conclusion is that s-channel Higgs production at a $\mu^+\mu^-$ collider of appropriate design has greater potential for distinguishing between a light SM h_{SM} and the SM-like h^0 of the MSSM than other processes/machines. The techniques and strategies for attaining the above results, and the associated requirements for the machine and detector, are discussed at length.

Two possible $\mu^+\mu^-$ machines are being actively studied [52, 53, 54]:

- A first muon collider (FMC, for short) with low c. m. energy ($\sqrt{s}$) between 100 and 500 GeV and $\mathcal{L} \sim 2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ delivering an annual integrated integrated luminosity $L \sim 20$ fb^{-1}.

- A next muon collider (NMC) with high $\sqrt{s} \gtrsim 4$ TeV and $\mathcal{L} \sim 10^{35}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ giving $L \sim 1000$ fb^{-1} yearly; the extent to which such a machine could be run at high luminosity for $\sqrt{s}$ values starting at 500 GeV remains to be determined.

One of our goals will be to quantify the amount of integrated luminosity that is required to detect and study the various Higgs bosons via s-channel production as the Higgs mass is varied. For s-channel study of a SM-like Higgs boson, only the lower energy machine is relevant because a SM-like Higgs can only be detected in s-channel collisions if it has mass $\lesssim 2m_W$, given the anticipated luminosity. However, higher $\sqrt{s}$ will be important if the MSSM is the correct theory. The expected luminosity will allow detection and study of the heavier MSSM Higgs bosons (the CP-odd A^0 and the CP-even H^0) via s-channel production at the FMC for m_{A^0}, m_{H^0} up to the maximal $\sqrt{s}$. If the NMC can be run with high luminosity at $\sqrt{s}$ values starting at the maximal FMC energy (~ 500 GeV) and above, then the ability to discover the A^0 and H^0 via s-channel production would extend to correspondingly higher masses.

For s-channel Higgs studies, it will be important to deliver the maximum possible luminosity at c.m. energies where Higgs bosons are either expected or observed. Fortunately, this should be possible for the proposed FMC designs due to the fact that the final muon storage ring(s) would comprise a modest fraction of the overall cost [61]. (The most costly component of a muon collider is the muon source — decays of pions produced by proton collisions.) It is thus envisioned that multiple storage rings could eventually be tailor-made for c.m. energies spanning the desired range. This approach could presumably also be used to allow the high energy NMC to run with high luminosity at $\sqrt{s}$ values starting at ~ 500 GeV, where the FMC leaves off.

A crucial machine parameter for s-channel studies of Higgs bosons is the energy resolution of the colliding beams. A Gaussian shape for the energy spectrum of each beam is expected to be a good approximation, with an rms deviation, R, most naturally in the range [62]

$$R = 0.04\% \text{ to } 0.08\%$$

which could be decreased to as low as

$$R = 0.01\%$$

via additional cooling. Excellent energy resolution is mandatory to detect and study a Higgs boson with a very narrow width, which is the case for the h_{SM} with $m_{h_{SM}} \lesssim 2m_W$ and the lightest MSSM Higgs boson. The large value of the muon mass compared to the electron mass makes possible the required energy resolution in three ways:

i) it is possible (albeit, probably expensive) to achieve $R = 0.01\%$;

ii) bremsstrahlung smearing, while non-negligible, leaves a large portion of the narrow central Gaussian beam energy peak intact.

iii) designs with small beamstrahlung are naturally achieved;

Henceforth, we neglect beamstrahlung since quantitative calculations of this are unavailable.

The rms spread in $\sqrt{s}$ (denoted by $\sigma_{\sqrt{s}}$) prior to including bremsstrahlung is given by

$$\sigma_{\sqrt{s}} = R\sqrt{s}/\sqrt{2}\,, \tag{2.53}$$

where R is the resolution in the energy of each beam. A convenient formula for $\sigma_{\sqrt{s}}$ is

$$\sigma_{\sqrt{s}} = (7\text{ MeV})\left(\frac{R}{0.01\%}\right)\left(\frac{\sqrt{s}}{100\text{ GeV}}\right)\,. \tag{2.54}$$

The critical issue is how this resolution compares to the calculated total widths of Higgs bosons when $\sqrt{s} = m_h$. For $R \lesssim 0.01\%$, the energy resolution in Eq. (2.54) is smaller than the Higgs widths in Fig. 2.3 for all but a light SM-like Higgs. We shall demonstrate that the smallest possible R allows the best measurement of a narrow Higgs width, and that the total luminosity required for discovery by energy scanning when $\Gamma_h^{\rm tot} \lesssim \sigma_{\sqrt{s}}$ is minimized by employing the smallest possible R. For a Higgs boson with width larger than $\sigma_{\sqrt{s}}$, results from a fine scan with small R can be combined without any increase in the luminosity required for discovery and width measurement.

The Feynman diagram for s-channel Higgs production is illustrated in Fig. 2.2. The s-channel Higgs resonance cross section is

$$\sigma_h(\sqrt{\hat{s}}) = \frac{4\pi\Gamma(h\to\mu\mu)\,\Gamma(h\to X)}{\left(\hat{s}-m_h^2\right)^2 + m_h^2[\Gamma_h^{\rm tot}]^2}\,, \tag{2.55}$$

where $\hat{s} = (p_{\mu^+} + p_{\mu^-})^2$ is the c. m. energy squared of a given $\mu^+\mu^-$ annihilation, X denotes a final state and $\Gamma_h^{\rm tot}$ is the total width.[1] The sharpness of the resonance peak is determined by $\Gamma_h^{\rm tot}$. Neglecting bremsstrahlung for the moment, the effective signal cross section is obtained by convoluting $\sigma_h(\hat{s})$ with the Gaussian distribution in $\sqrt{\hat{s}}$ centered at $\sqrt{\hat{s}} = \sqrt{s}$:

$$\overline{\sigma}_h(\sqrt{s}) = \int \sigma_h(\sqrt{\hat{s}})\, \frac{\exp\left[-(\sqrt{\hat{s}} - \sqrt{s})^2/(2\sigma^2_{\sqrt{s}})\right]}{\sqrt{2\pi}\sigma_{\sqrt{s}}}\, d\sqrt{\hat{s}}\,. \tag{2.56}$$

Figure 2.4 illustrates the effective cross section, $\overline{\sigma}_h(\sqrt{s})$, as a function of $\sqrt{s}$ for $m_h = 110$ GeV and beam energy resolutions of $R = 0.01\%$, $R = 0.06\%$, and $R = 0.1\%$. Results are given for the cases: h_{SM}, h^0 with $\tan\beta = 10$, and h^0 with $\tan\beta = 20$. All channels X are summed over.

In the case where the Higgs width is much smaller than the Gaussian width $\sigma_{\sqrt{s}}$, the effective signal cross section result for $\sqrt{s} = m_h$, denoted by $\overline{\sigma}_h$, is

$$\overline{\sigma}_h = \frac{2\pi^2\Gamma(h \to \mu\mu)\, BF(h \to X)}{m_h^2} \times \frac{1}{\sigma_{\sqrt{s}}\sqrt{2\pi}} \qquad (\Gamma_h^{\rm tot} \ll \sigma_{\sqrt{s}})\,. \tag{2.57}$$

Henceforth, we adopt the shorthand notation

$$G(X) = \Gamma(H \to \mu\mu)\, BF(h \to X) \tag{2.58}$$

for the numerator of Eq. (2.57). The increase of $\overline{\sigma}_h(\sqrt{s} = m_h)$ with decreasing $\sigma_{\sqrt{s}}$ when $\Gamma_h^{\rm tot} \ll \sigma_{\sqrt{s}}$ is apparent from the h_{SM} curves of Fig. 2.4. In the other extreme where the Higgs width is much broader than $\sigma_{\sqrt{s}}$, then at $\sqrt{s} = m_h$ we obtain

$$\overline{\sigma}_h = \frac{4\pi BF(h \to \mu\mu) BF(h \to X)}{m_h^2} \qquad (\Gamma_h^{\rm tot} \gg \sigma_{\sqrt{s}})\,. \tag{2.59}$$

Note that this equation implies that if there is a large contribution to the Higgs width from some channel other than $\mu\mu$, we will get a correspondingly smaller total event rate due to the small size of $BF(h \to \mu\mu)$. That $\overline{\sigma}_h(\sqrt{s} = m_h)$ is independent of the value of $\sigma_{\sqrt{s}}$ when $\Gamma_h^{\rm tot} \gg \sigma_{\sqrt{s}}$ is illustrated by the $\tan\beta = 20$ curves for the h^0 in Fig. 2.4. Raw signal rates (*i.e.* before applying cuts and including other efficiency factors) are computed by multiplying $\overline{\sigma}_h$ by the total integrated luminosity L.

[1] Effects arising from implementing an energy-dependent generalization of the $m_h\Gamma_h^{\rm tot}$ denominator component of this simple resonance form are of negligible importance for our studies, especially for a Higgs boson with $\Gamma_h^{\rm tot} \ll m_h$.

The basic results of Eqs. (2.57) and (2.59) are modified by the effects of photon bremsstrahlung from the colliding muon beams. In the case of a narrow Higgs boson, the primary modification for $\sqrt{s} = m_h$ is due to the fact that not all of the integrated luminosity remains in the central Gaussian peak. These modifications are discussed in sec. 2.12.1; to a good approximation, the resulting signal rate is obtained by multiplying $\overline{\sigma}_h$ of Eq. (2.57) by the total luminosity L times the fraction f of the peak luminosity in the Gaussian after including bremsstrahlung relative to that before (typically $f \approx 0.6$). For a broad Higgs resonance, the lower energy tail in the luminosity distribution due to bremsstrahlung makes some contribution as well. In the results to follow, we avoid any approximation and numerically convolute the full effective luminosity distribution (including bremsstrahlung) with the Higgs cross section of Eq. (2.55). In performing this convolution, we require that the effective $\mu^+\mu^-$ c.m. energy be within 10 GeV of the nominal value. Such a requirement can be implemented by reconstructing the mass of the final state as seen in the detector; planned detectors would have the necessary resolution to impose the above fairly loose limit. This invariant mass selection is imposed in order to reduce continuum (non-resonant) backgrounds that would otherwise accumulate from the entire low-energy bremsstrahlung tail of the luminosity distribution.

As is apparent from Fig. 2.4, discovery and study of a Higgs boson with a very narrow width at the $\mu^+\mu^-$ collider will require that the machine energy $\sqrt{s}$ be within $\sigma_{\sqrt{s}}$ of m_h. The amount of scanning required to find the correct $\sqrt{s}$ depends upon R. From Fig. 2.4 it is apparent that the larger R is, the less the accuracy with which the machine energy needs to be set at each scan point and the fewer the number of scan points needed. But, small R results in much greater event rate for $\sqrt{s} \simeq m_h$. *If $\sqrt{s}$ can be rapidly changed with an accuracy that is a small fraction of R*, then we shall find that smaller R implies that less total time (and, hence, luminosity) will be required for the scan. Further, we find that $R \sim 0.01\%$ and the ability to set $\sqrt{s}$ with an accuracy of order 1 part in 10^6 are both required if we are to be able to measure the Higgs width with sufficient precision to distinguish between the SM h_{SM} and the MSSM h^0 when the latter is SM-like. Thus, for a $\mu^+\mu^-$ collider to reach its full potential, it should be designed so that $R \sim 0.01\%$ and so that it is possible to vary $\sqrt{s}$ rapidly and with great precision. These are not insurmountable tasks [61], but careful planning is certainly required. For Higgs bosons with a large width, the design demands upon the $\mu^+\mu^-$ collider are clearly less.

Due to the bremsstrahlung tail, it is also possible to search for a Higgs boson by running

the $\mu^+\mu^-$ collider at an energy well above the mass of the Higgs boson itself. In some collisions, one (or both) of the muons will have radiated enough of its initial energy that the effective $\sqrt{\hat{s}}$ of the collision is much lower than $\sqrt{s}$. In this circumstance, detection of the Higgs boson requires reconstruction with good resolution of the effective $\sqrt{\hat{s}}$ of each collision from the final state momenta. For a final state mass bin centered at $\sqrt{\hat{s}} = m_h$, if $d\mathcal{L}/d\sqrt{\hat{s}}$ is slowly varying in the vicinity of $\sqrt{\hat{s}} = m_h$ over an interval several times the Higgs total width $\Gamma_h^{\rm tot}$, the effective cross section is

$$\overline{\sigma}_h = \frac{2\pi^2 \Gamma(h \to \mu\mu)\, BF(h \to X)}{m_h^2} \times \left.\frac{d\mathcal{L}}{d\sqrt{\hat{s}}}\right|_{\sqrt{\hat{s}}=m_h} . \tag{2.60}$$

In exploring the possible utility of this bremsstrahlung tail for Higgs detection, we have performed our explicit calculations using the spectrum obtained for $R = 0.1\%$. However, we note that the bremsstrahlung tail well away from the central Gaussian peak is essentially independent of the beam energy resolution R. If a mass resolution in the final state of ± 5 GeV is possible in the $b\overline{b}$ final state, then even when running the FMC at full nominal energy of $\sqrt{s} = 500$ GeV we find that it will be possible to detect a Higgs boson with m_h in a broad range below $\sqrt{s}$ (but not near m_Z) provided that the $h \to \mu^+\mu^-$ coupling is significantly enhanced with respect to the SM $h_{SM} \to \mu^+\mu^-$ coupling. The total integrated luminosity required for Higgs discovery using the bremsstrahlung tail will be compared to that needed for discovery by scanning using a large number of $\sqrt{s}$ machine energy settings.

Highly polarized beams may be possible since the muons are naturally polarized from $\pi^\pm$ ($K^\pm$) decays in the parent rest-frame. However, the luminosity for polarized beams may be significantly reduced during the cooling and acceleration process. If a degree of polarization P is possible for *both* beams, then, relative to the unpolarized case, the s-channel Higgs signal is enhanced by the factor $(1 + P^2)$ while the background is suppressed by $(1 - P^2)$. High polarization P of both beams would be useful if the luminosity reduction is less than a factor of $(1 + P^2)^2 / (1 - P^2)$, *i.e.* the factor which would leave the significance of the signal unchanged. For example, $P = 0.84$ would compensate a factor of 10 reduction in luminosity [63]. We mainly present our results without assuming high polarization beams, but we comment on improvements with beam polarization.

With this introduction, we now proceed with a detailed description of the capability of a $\mu^+\mu^-$ collider to detect and study different types of Higgs bosons. In the next section, we begin with SM-like Higgs bosons. The following section explores the non-SM-like Higgs bosons of the MSSM. The final section gives our conclusions.

2.3.2 A SM-like Higgs Boson

We first review the prospects for discovering and studying a SM-like Higgs boson without s-channel production at a $\mu^+\mu^-$ collider. We then turn to the role of s-channel $\mu^+\mu^- \to h$ production, emphasizing the prospects for precision studies of the Higgs mass and width.

Discovery and Study Without s-Channel Production

Neutral Higgs bosons that are coupled to ZZ with roughly SM-like strength can be discovered via $Z^\star \to Zh$ production for $m_h \lesssim 0.7\sqrt{s}$ at either an e^+e^- collider or a $\mu^+\mu^-$ collider [64]. This discovery reach applies to both the h_{SM} and to the h^0 of the MSSM in the large-m_{A^0} portion of parameter space where it is SM-like in its couplings. The stringent upper bound on m_{h^0}, Eq. (2.49), in the MSSM implies that even a $\sqrt{s} = 300$ GeV machine is guaranteed to find the h^0 if it exists.

As described in the Introduction, we can also consider adding extra singlets to the MSSM two-doublet Higgs sector. In the MNMSSM model, containing one singlet Higgs field, we noted that even if the lightest Higgs boson has small ZZ coupling, there is always a CP-even Higgs boson with substantial ZZ coupling and modest mass. Refs. [50] demonstrate that at least one of the CP-even Higgs bosons of the MNMSSM model will be detected in the Zh mode at a machine with c.m. energy $\sqrt{s} = 500$ GeV. Since it appears that this result may generalize to the case of more than one additional singlet, we regard it as relatively certain that any supersymmetric theory in the SUSY GUT context will contain at least one CP-even Higgs boson that will be discovered in the Zh mode at a machine with $\sqrt{s} = 500$ GeV, and its mass will be in the intermediate mass range ($\lesssim 2m_W$).

Assuming that a SM-like h is discovered in the Zh mode, an important question for s-channel production and study of the h in $\mu^+\mu^-$ collisions is the accuracy with which its mass can be measured á priori via Zh production. The better this accuracy, the easier it will be to set $\sqrt{s}$ of the $\mu^+\mu^-$ collider to a value centered on m_h within the rms spread $\sigma_{\sqrt{s}}$. Another critical question bearing on the importance of the s-channel $\mu^+\mu^- \to h$ production mode is whether the Zh mode is useful for measurement of the h width. We find that it is not.

Generally speaking, the accuracy of the Higgs boson mass measurements depends on the detector performance and the signal statistics. As a general guide, we consider two examples for the uncertainty on m_h in the mass range $m_h < 2m_W$ (*i.e.* below where W-pair decays

become important)

$$\Delta m_h \simeq 4.0 \text{ GeV}/\sqrt{N} \quad \text{(SLD)}, \tag{2.61}$$

$$\simeq 0.3 \text{ GeV}/\sqrt{N} \quad \text{(super} - \text{LC)}. \tag{2.62}$$

where our notation will always be that ΔX represents the absolute magnitude of the 1σ error on the quantity X; that is the 1σ limits on X are $X \pm \Delta X$. Equation (2.61) results for performance typified by the SLD detector [65], where 4 GeV is the single event resolution and N is the number of events in the $Z(\to q\overline{q})h(\to b\overline{b})$, $Z(\to q\overline{q})h(\to \tau\overline{\tau})$, plus $Z(\to \ell^+\ell^-)h(\to$ any) modes. For a SM-like Higgs, these modes have an effective final state branching fraction that varies between about 70% and 50% as m_h varies from low masses up to 140 GeV. We plot Δm_h in Fig. 2.1 according to Eqs. (2.61) and (2.62), with $N = \epsilon L\sigma(Zh)BF(\text{effective})$, assuming detection efficiencies of $\epsilon = 0.9$ [$\epsilon = 0.5$] for the $Z(\to \ell^+\ell^-)h(\to$ any) [$Z(\to q\overline{q})h(\to b\overline{b})$, $Z(\to q\overline{q})h(\to \tau\overline{\tau})$] modes and assuming a fixed $\sqrt{s} = 500$ GeV. For SLD detector performance, results for luminosities of $L = 1$, 10, and 50 fb^{-1} are shown; with these integrated luminosities, m_h (for $m_h \lesssim 150$ GeV) will be determined to an accuracy of at least 1.4, 0.5, 0.21 GeV (respectively).

Equation (2.62) is applicable for a "super" performance Linear Collider detector (hereafter referred to as the super-LC detector) [66, 67], the special features of which include excellent momentum resolutions and high b-tagging efficiency. For this detector, the best determination of $m_{h_{SM}}$ is obtained by examining the recoil mass peak in Zh_{SM} production. For $Z \to \ell^+\ell^-$ events, the resolution for the recoil mass is expected to be of order 0.3 GeV per event. A measurement of $m_{h_{SM}}$ to ± 0.3 GeV$/\sqrt{N} \sim \pm 20$ MeV would be possible for $m_{h_{SM}} \lesssim 140$ GeV and $L = 50$ fb^{-1}, as illustrated in Fig. 2.1, assuming detection efficiency of $\epsilon = 0.9$ for the $Z(\to \ell^+\ell^-)h(\to$ any) mode. The total width $\Gamma^{\rm tot}_{h_{SM}}$ could also be measured down to $\sim$200 MeV using the Zh_{SM} recoil mass distribution. However, this latter sensitivity is not likely to be useful since $\Gamma_{h_{SM}} \lesssim 10$ MeV for $m_{h_{SM}} \lesssim 140$ GeV (see Fig. 2.3).

It could happen that there is no e^+e^- collider at the time the $\mu^+\mu^-$ collider is built but that the LHC has been operational for several years. One of the primary modes for discovery of a SM-like Higgs boson at the LHC is the $\gamma\gamma$ mode. Simulations by the LHC collaborations indicate that this mode is detectable for $50 \lesssim m_h \lesssim 150$ GeV. For $m_h \gtrsim 130$ GeV, discovery will be possible in the 4ℓ mode. Both modes, but especially the $\gamma\gamma$ mode, offer the possibility of a very accurate determination of the Higgs mass. Resolution will be 1% or better in the $\gamma\gamma$ mode, and probably not much worse than 1% in the 4ℓ mode. Thus, even in the absence of an

e^+e^- collider, the LHC can reasonably be expected to provide us with a $\lesssim 1\%$ determination of m_h in the mass region where the Higgs total width is small.

s-Channel Production of a SM-like *h*

Once a SM-like Higgs boson is found in the Zh mode at either an e^+e^- collider or the $\mu^+\mu^-$ collider itself,[2] or at the LHC, it will generally be easy to also produce and detect it via direct s-channel production at a $\mu^+\mu^-$ collider [60] if $m_h \lesssim 2m_W$. Should there be no e^+e^- collider in operation, an important question at a $\mu^+\mu^-$ collider will then be whether to concentrate subsequent running on s-channel production or on Zh production, as the best means for studying the properties of the h in detail. Generally speaking, these two different processes provide complementary information and it would be very valuable to accumulate substantial integrated luminosity in both modes.

The potential importance of s-channel production of a SM-like h is illustrated by two facts pertaining to distinguishing between the MSSM h^0 and the SM h_{SM}.

(1) Expected experimental errors imply that the ability to discriminate between the SM h_{SM} and the MSSM h^0 on the basis of the branching fractions and production rates that can be measured in the Zh channel is limited to m_{A^0} values below about 300 GeV [41].

(2) Both the total width and the production rate (proportional to $\Gamma(h \to \mu^+\mu^-)$) of a SM-like h could be measured at a muon collider with sufficient accuracy so as to distinguish the h^0 from the h_{SM} in the large-m_{A^0} region 300 GeV $\lesssim m_A \lesssim$ 600 GeV where the h^0 is approximately SM-like.

A quantitative discussion of the MSSM parameter space region for which deviations of the total width and production rate from SM expectations are measurable will be given later. For now we emphasize that (2) requires the excellent $R = 0.01\%$ beam energy resolution.

Choosing the right $\sqrt{s}$ Our proposed strategy is to first discover the SM-like h via $\ell^+\ell^- \to Zh$ or in hadron collisions in order to determine the $\sqrt{s}$ region in which $\mu^+\mu^- \to h$ s-channel production should be explored. If $\Gamma_h^{\rm tot}$ is smaller than the rms spread $\sigma_{\sqrt{s}}$ in $\sqrt{s}$

[2] While discovery at a $\mu^+\mu^-$ collider is also possible by scanning in s, the Zh mode is more luminosity efficient for discovery.

(as is the case for the SM when $m_{h_{SM}} \lesssim 140$ GeV), then to obtain the maximum $\mu^+\mu^- \to h$ production rate it is necessary to set $\sqrt{s}$ equal to m_h within $\lesssim \sigma_{\sqrt{s}}$. The ability to do this is assessed by comparing the errors on m_h from Zh production to both the $\sqrt{s}$ spread $\sigma_{\sqrt{s}}$ at a $\mu^+\mu^-$ collider and to $\Gamma_h^{\rm tot}$. As an illustration, consider $h = h_{SM}$. With the super-LC $L = 50$ fb^{-1} determination of $m_{h_{SM}}$ to ± 20 MeV, $\sigma_{\sqrt{s}}$ for $R = 0.01\%$ will be at worst a factor of 2 or 3 smaller than the uncertainty in $m_{h_{SM}}$ and only two or three tries will be needed to set the $\mu^+\mu^-$ collider energy to a value equal to $m_{h_{SM}}$ within the rms spread in $\sqrt{s}$. If the SLD $L = 50$ fb^{-1} determination of $m_{h_{SM}}$ to 210 MeV is all that is available, then for $m_{h_{SM}} \lesssim 2m_W$ two or three tries would be adequate to set $\sqrt{s} \simeq m_{h_{SM}}$ within $\sigma_{\sqrt{s}}$ only if $R = 0.06\%$. The number of settings required in the case of $R = 0.01\%$ would be a factor of 6 larger. If only SLD performance and $L = 1$ fb^{-1} is available in the Zh_{SM} mode, or if only a $\sim 1\%$ determination of $m_{h_{SM}}$ from the LHC is provided, both of which imply errors on $m_{h_{SM}}$ that are $\gtrsim 1$ GeV, then even with $R = 0.06\%$ one must scan over 10 to 20 $\sqrt{s}$ values to determine the central $\sqrt{s} \simeq m_{h_{SM}}$ value within the rms $\sqrt{s}$ error, $\sigma_{\sqrt{s}}$. Later, we will compute the amount of luminosity that must be invested at each $\sqrt{s} = m_h$ choice in order to detect a SM-like Higgs signal.

In contrast to the above narrow width situation, for $m_{h_{SM}} \gtrsim 200$ GeV one finds $\Gamma_{h_{SM}}^{\rm tot} \gtrsim \sigma_{\sqrt{s}}$ for $R \leq 0.06\%$. Then, even if $m_{h_{SM}}$ is only known to within $\Gamma_{h_{SM}}^{\rm tot}$, we can immediately set $\sqrt{s}$ for the $\mu^+\mu^-$ collider to be within the Higgs peak. Unfortunately, we find that the event rate in s-channel collisions is too low to allow detection of the h_{SM} in this case. This situation does not arise in the case of the h^0 of the MSSM, which is guaranteed to have $m_{h^0} \lesssim 130$ GeV.

Detecting a SM-like h in the s-channel The effective cross section, $\overline{\sigma}_{h_{SM}}(\sqrt{s} = m_{h_{SM}})$ for inclusive SM Higgs production is given in Fig. 2.3.2 versus $\sqrt{s} = m_{h_{SM}}$ for resolutions of $R = 0.01\%$, 0.06%, 0.1% and 0.6%. These results include Gaussian and bremsstrahlung smearing effects. For comparison, the $\mu^+\mu^- \to Z^\star \to Zh_{SM}$ cross section is also shown, evaluated at the energy $\sqrt{s} = m_Z + \sqrt{2}m_{h_{SM}}$ for which it is a maximum. The s-channel $\mu^+\mu^- \to h_{SM}$ cross sections for small R and $m_{h_{SM}} \lesssim 2m_W$ are much larger than the corresponding Zh_{SM} cross section. The increase in the $\mu^+\mu^- \to h_{SM}$ cross section that results if bremsstrahlung smearing is removed is illustrated in the most sensitive case ($R = 0.01\%$).

For a SM-like Higgs boson, the only potentially useful final state modes X are $b\overline{b}$, $WW^{(\star)}$

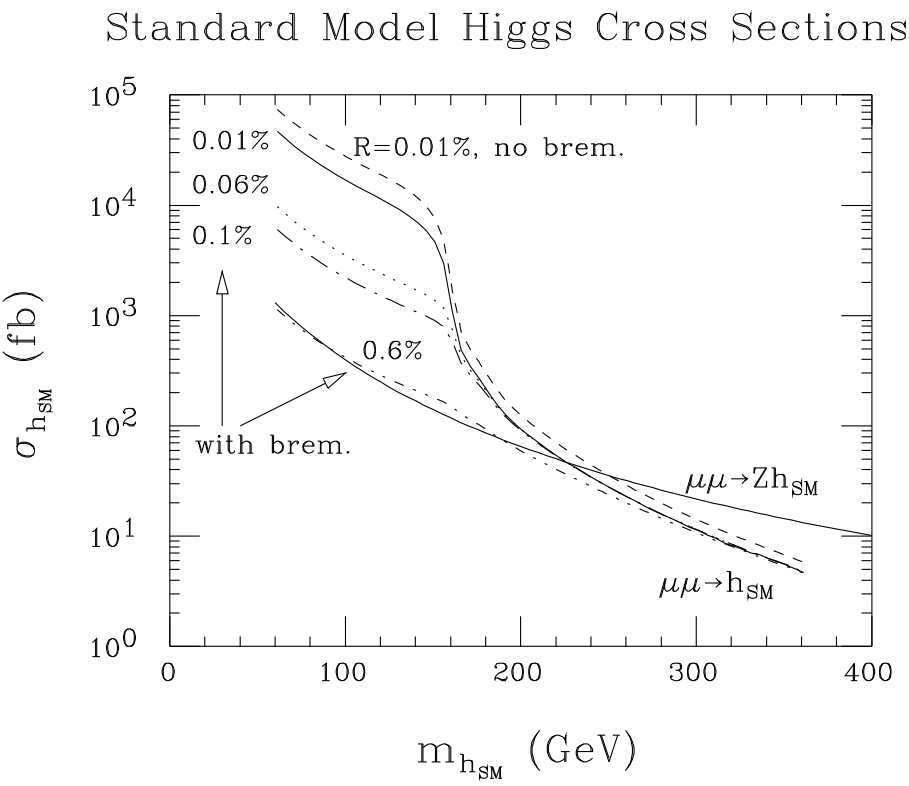


Figure 2.20: Cross sections vs $m_{h_{SM}}$ for inclusive SM Higgs production: (i) the s-channel $\overline{\sigma}_h$ for $\mu^+\mu^- \to h_{SM}$ with $R = 0.01\%$, 0.06%, 0.1% and 0.6%, and (ii) $\sigma(\mu^+\mu^- \to Zh_{SM})$ at $\sqrt{s} = m_Z + \sqrt{2}m_{h_{SM}}$. Also shown is the result for $R = 0.01\%$ if bremsstrahlung effects are not included.

and $ZZ^{(\star)}$, where the $^{(\star)}$ indicates the possibility that the weak boson is virtual. The $t\bar{t}$ channel does not give a viable signal for the range of luminosity that we consider. All these channels have irreducible backgrounds from $\mu^+\mu^-$ continuum production processes. We note that

(a) The light-quark backgrounds to the $b\bar{b}$ channel can be rejected using b-tagging. We assume a 50% efficiency for isolating the $2b$ final state (via tagging one of the b's); this efficiency is to include cuts and detector efficiencies.

(b) For the $b\bar{b}$ final state, we have checked that interference between the s-channel signal and the backgrounds is never of importance. This is because the Higgs signal contributes to RR and LL helicity amplitudes for the incoming muons, whereas the

backgrounds come almost entirely from RL and LR helicity combinations (the RR and LL background contributions are suppressed by a factor of m_μ/E at the amplitude level).

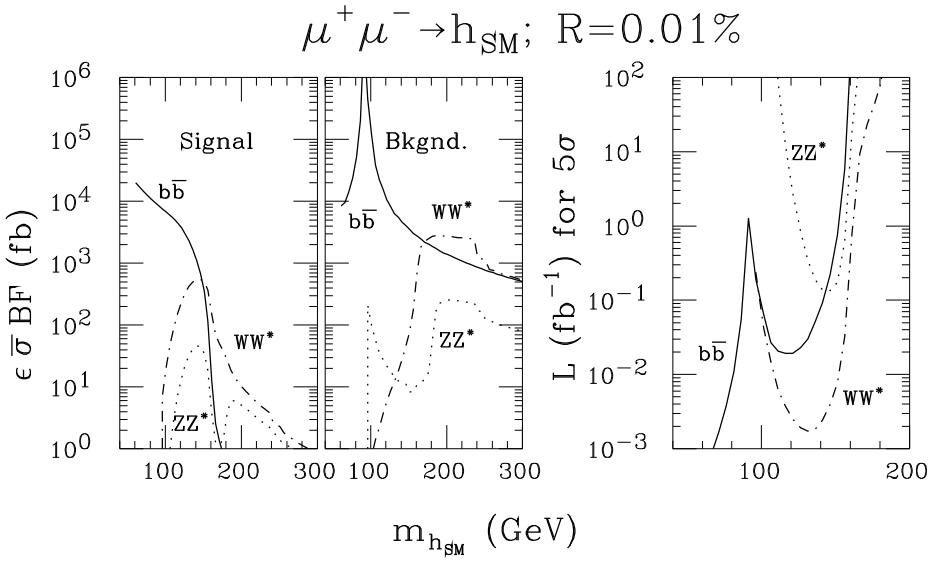
(c) For the $WW^{(\star)}$ and $ZZ^{(\star)}$ final states the useful channels depend upon whether or not the $W^{(\star)}$ or $Z^{(\star)}$ is virtual. We shall find that discovery in these channels is only possible for $m_h \lesssim 2m_W$, in which case the final states of interest are $WW^\star \to \ell\nu 2j$ with $BF^{\rm eff}_{WW} \sim 0.3$ and $ZZ^\star \to 2\ell 2j, 2\nu 2j, 4\ell, 2\ell 2\nu$ with $BF^{\rm eff}_{ZZ} \sim 0.42$, $4j$ final states having too large a QCD background and mass reconstruction of the real W or Z being impossible in the $2\ell 2\nu$ or 4ν final states, respectively. (Here, we consider only $\ell = e$ or μ.) In our analysis, we assume an overall efficiency of 50% for isolating these channels. For the $ZZ^\star$, a cut requiring that $M^\star$ (the invariant mass of the virtual $Z^\star$) be greater than a given value $M^{\star\,\rm min}$ is imposed. Full details regarding our procedures in the $WW^{(\star)}$ and $ZZ^{(\star)}$ channels are presented in sec. 2.12.2.

The h_{SM} signal and background cross sections, $\epsilon\overline{\sigma}BF(X)$, for $X = b\overline{b}$, and the above $WW^{(\star)}$ and $ZZ^{(\star)}$ final states are presented in Fig. 2.3.2 (including a channel-isolation efficiency of $\epsilon = 0.5$) as a function of $m_{h_{SM}}$ for SM Higgs s-channel production with resolution $R = 0.01\%$ and $R = 0.06\%$. For both resolutions, we also plot the luminosity required for a $S/\sqrt{B} = 5\sigma$ signal in the $b\overline{b}$, $WW^{(\star)}$ and $ZZ^{(\star)}$ channels. In the case of the $WW^{(\star)}$ final state, we give event rates only for the mixed leptonic/hadronic final state modes; in the case of the $ZZ^{(\star)}$ final state we include the mixed hadronic/leptonic and (visible) purely leptonic final state modes listed earlier.

From Fig. 2.3.2 we see that:

- $R = 0.01\%$, $L = 0.1$ fb^{-1} would yield a detectable s-channel Higgs signal for all $m_{h_{SM}}$ values between the current LEP I limit of 63 GeV and $2m_W$ except in the region of the Z peak; a luminosity $L \sim 1$ fb^{-1} at $\sqrt{s} = m_{h_{SM}}$ is needed for $m_{h_{SM}} \sim m_Z$.
- For $R = 0.06\%$, 5σ signals typically require about 20–30 times the luminosity needed for $R = 0.01\%$; $L = 30$ fb^{-1} would be required for a 5σ signal if $m_{h_{SM}} \sim m_Z$.

This argues for a $\mu^+\mu^-$ collider design with R near the 0.01% level. A search for the h_{SM} (or any Higgs with width smaller than the achievable resolution) by scanning would be most efficient for the smallest possible R. For a specific illustration, let us consider $m_{h_{SM}} \sim 110$ GeV and assume that just $L = 1$ fb^{-1} has been accumulated in the Zh_{SM} mode



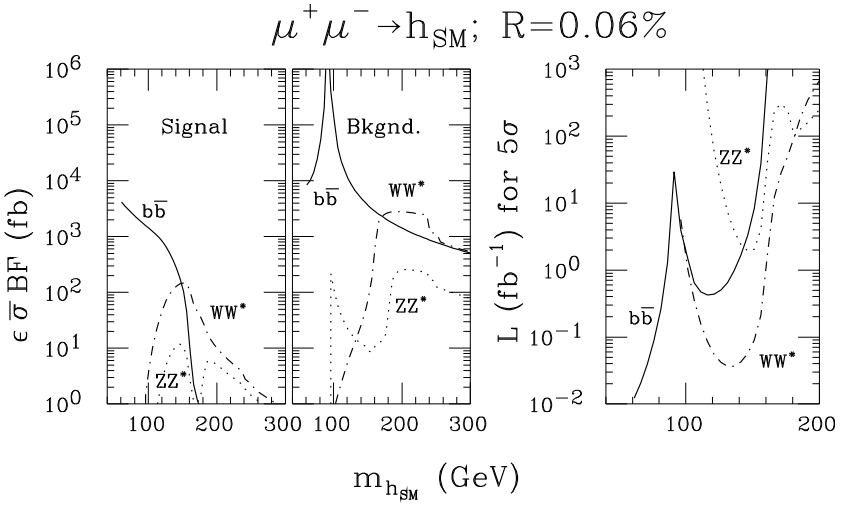


Figure 2.21: The (a) h_{SM} signal and (b) background cross sections, $\epsilon\bar{\sigma}BF(X)$, for $X = b\bar{b}$, and useful (reconstructable, non-$4j$) $WW^{(\star)}$ and $ZZ^{(\star)}$ final states (including a channel-isolation efficiency of $\epsilon = 0.5$) versus $m_{h_{SM}}$ for SM Higgs s-channel production. Also shown: (c) the corresponding luminosity required for a $S/\sqrt{B} = 5$ standard deviations signal in each of the three channels. Results for $R = 0.01\%$ and $R = 0.06\%$ are given.

(at either an e^+e^- collider or at the $\mu^+\mu^-$ collider itself). Fig. 2.1 shows that the error in the determination of $m_{h_{SM}}$ will be of order ± 0.8 GeV (assuming an SLD-type detector). How much luminosity will be required to observe the h_{SM} in the s-channel by zeroing in on $m_{h_{SM}}$ within the rms resolution $\sigma_{\sqrt{s}}$? The number of scan points required to cover the 1.6 GeV mass zone at intervals of $\sigma_{\sqrt{s}}$, the luminosity required to observe (or exclude) the Higgs at each point, and the total luminosity required to zero-in on the Higgs using the scan is given in Eq. (2.63), for resolutions of $R = 0.01\%$ and 0.06%.

R	$\sigma_{\sqrt{s}}$	#points	L/point	$L_{\rm tot}$
0.01%	7 MeV	230	0.01 fb^{-1}	2.3 fb^{-1}
0.06%	45 MeV	34	0.3 fb^{-1}	10.2 fb^{-1}

(2.63)

More generally, the L required at each scan point decreases as (roughly) $R^{1.7}$, whereas the number of scan points only grows like $1/R$, implying that the total L required for the scan decreases as $\sim R^{0.7}$. Thus, the $\mu^+\mu^-$ collider should be constructed with the smallest possible R value. (Note that if the Higgs resonance is broad, using small R, although not necessary, is not harmful since the data from a fine scan can be rebinned to test for its presence.) In the case of a narrow Higgs, a by-product of the above zeroing-in scan will be to ascertain if the Higgs width is in the $\lesssim \sigma_{\sqrt{s}}$ range. However, the large number of $\sqrt{s}$ settings required when conducting a scan with small R implies that it must be possible to quickly and precisely adjust the energy of the $\mu^+\mu^-$ collider. For example, if the machine can deliver 50 fb^{-1} per year and $R = 0.01\%$, so that only $L \sim 0.01$ fb^{-1} should be devoted to each point, we must be able to step the machine energy in units of ~ 7 MeV once every hour or so.

Let us compare the above procedure, where the Zh mode at low luminosity is used to find the SM-like h and then s-channel collisions are used to zero-in on m_h, to the possibility of searching directly for the h by s-channel scanning without the benefit of Zh data. The latter would be a possible alternative if the $\mu^+\mu^-$ collider were to be built before the light Higgs boson is observed at either the LHC or an e^+e^- collider. The question is whether it is most useful to employ the Zh mode or direct s-channel production for initial discovery. We shall suppose that precision radiative corrections pin down the mass of the SM-like Higgs boson to a 20 GeV interval, although this may be way too optimistic. Let us again focus on $m_h = 110$ GeV. The number of scan points required to cover the 20 GeV mass zone at intervals of $\sigma_{\sqrt{s}}$, the luminosity required to observe (or exclude) the Higgs at each point, and

the total luminosity required to zero-in on the Higgs using the scan is given in Eq. (2.64), for resolutions of $R = 0.01\%$ and 0.06%.

R	$\sigma_{\sqrt{s}}$	#points	L/point	$L_{\rm tot}$
0.01%	7 MeV	2857	0.01 fb^{-1}	29 fb^{-1}
0.06%	45 MeV	426	0.3 fb^{-1}	128 fb^{-1}

(2.64)

Thus, much greater luminosity would be required (not to mention the much greater demands upon the machine for performing efficiently such a broad scan) than if the Zh mode is employed for the initial h discovery. Note that it is not useful to expend more than $L \sim 1$ fb^{-1} in the Zh mode simply to pin down the mass; however, precision studies with $L = 50$ fb^{-1} in this mode would be useful for determining $\sigma(Zh) \times BF(h \to X)$ for various different final states, X [41].

For $m_{h_{SM}}$ above $2m_W$, $\Gamma^{\rm tot}_{h_{SM}}$ rises dramatically, $BF(h_{SM} \to \mu^+\mu^-)$ falls rapidly and, thus [see Eq. (2.59) and Fig. 2.3.2], $\overline{\sigma}_h$ declines precipitously. Even after combining all channels, the luminosity requirements in the double-on-shell WW and ZZ final states are such that Higgs detection in s-channel production will be difficult. How severe a drawback is this? One of the unique and most important features of s-channel Higgs production is the ability to scan with sufficient statistics to determine the width of a narrow Higgs boson. In the case of the h_{SM}, only below WW threshold is the Higgs so narrow that this is the only possible measurement technique. The h_{SM} can be detected straightforwardly in the standard Zh_{SM} mode and, at the super-LC detector, its width can be measured down to 0.2 GeV via the recoil mass spectrum in Zh_{SM} events with $Z \to \ell^+\ell^-$. Since $\Gamma^{\rm tot}_{h_{SM}} \gtrsim 0.2$ GeV for $m_{h_{SM}} \gtrsim 2m_W$, this Zh_{SM} technique becomes viable just as s-channel detection becomes difficult. Without the super-LC detector there could, however, be a gap between the $m_{h_{SM}} \lesssim 2m_W$ region where s-channel measurement of $\Gamma^{\rm tot}_{h_{SM}}$ will be possible at a muon collider and the region $m_{h_{SM}} \gtrsim 200$ GeV where $\Gamma^{\rm tot}_{h_{SM}}$ becomes comparable to the event by event mass resolution of ~ 4 GeV (see earlier discussion and Fig. 2.3) and would become measurable at a linear e^+e^- collider. The high resolution for lepton momenta of the super-LC detector could thus prove critical in avoiding a gap in the region between about 150 GeV and 200 GeV where $\Gamma^{\rm tot}_{h_{SM}}$ measurement might not be possible using either s-channel scanning or the Zh_{SM} mode.

The most important conclusions of this subsection are two:

(1) Excellent beam energy resolution is absolutely critical to guaranteeing success in detecting a SM-like h in $\mu^+\mu^- \to h$ s-channel collisions and to our ability to perform

detailed studies once the Higgs boson mass is known. Every effort should therefore be made to achieve excellent resolution. (It is only if $m_h > 2m_W$ where the SM-like Higgs boson begins to become broad that the advantage of having small R declines. But, for such masses s-channel discovery of the SM Higgs will be very difficult in any case, as we have discussed.)

(2) The scanning required when R is small implies that the machine design must be such that $\sqrt{s}$ can be *quickly* reset with a precision that is a small fraction of $\sigma_{\sqrt{s}}$.

Precision Measurements: m_h and $\Gamma_h^{\rm tot}$

Once the machine is set to the central value of $\sqrt{s} = m_h$, one can proceed to precisely measure the mass m_h and the total width $\Gamma_h^{\rm tot}$. A precision determination of the total width $\Gamma_h^{\rm tot}$ is of particular interest to differentiate between the h_{SM} and the h^0 of the MSSM. Knowledge of the total width will also allow extraction of the partial width (and associated Higgs couplings) for any channel in which the Higgs can be observed.

A precise measurement of the Higgs mass is possible via s-channel collisions. We initially focus our discussion on $m_{h_{SM}} \lesssim 2m_W$, for which $\Gamma_{h_{SM}}^{\rm tot}$ is quite likely to be smaller, perhaps much smaller, than the rms $\sqrt{s}$ resolution, $\sigma_{\sqrt{s}}$. Despite this, a highly accurate determination of $m_{h_{SM}}$ is still possible via a straightforward scan in the vicinity of $\sqrt{s} = m_{h_{SM}}$. In Fig. 2.3.2 we illustrate sample data points (statistically fluctuated) in the case of $m_{h_{SM}} = 110$ GeV, assuming $L = 0.5$ fb^{-1} is accumulated at each $\sqrt{s}$ setting. A resolution of $R = 0.01\%$ is assumed. The solid curve is the theoretical prediction. A visual inspection reveals that $m_{h_{SM}}$ can be pinned down to within about 4 MeV using seven scan points centered around $\sqrt{s} = m_{h_{SM}}$ (involving a combined luminosity of 3.5 fb^{-1}). Using somewhat more sophisticated techniques, to be described shortly, we will find that with this same total luminosity we can do better. These latter techniques are those needed for a direct measurement of the total Higgs width $\Gamma_{h_{SM}}^{\rm tot}$.

If the partial widths for $h_{SM} \to \mu^+\mu^-$ and $h_{SM} \to b\overline{b}$ are regarded as theoretically computable with no systematic uncertainties (not a valid assumption in the case of the MSSM h^0), then determination of $\Gamma_{h_{SM}}^{\rm tot}$ is straightforward based on Eq. (2.57). We have plotted the theoretical predictions for $m_{h_{SM}} = 110$ GeV in Fig. 2.3.2 corresponding to keeping the above partial widths constant while varying only $\Gamma_{h_{SM}}^{\rm tot}$ by $\pm 10\%$. Assuming that the background can be absolutely normalized by a combination of theory and experiment, the height of the

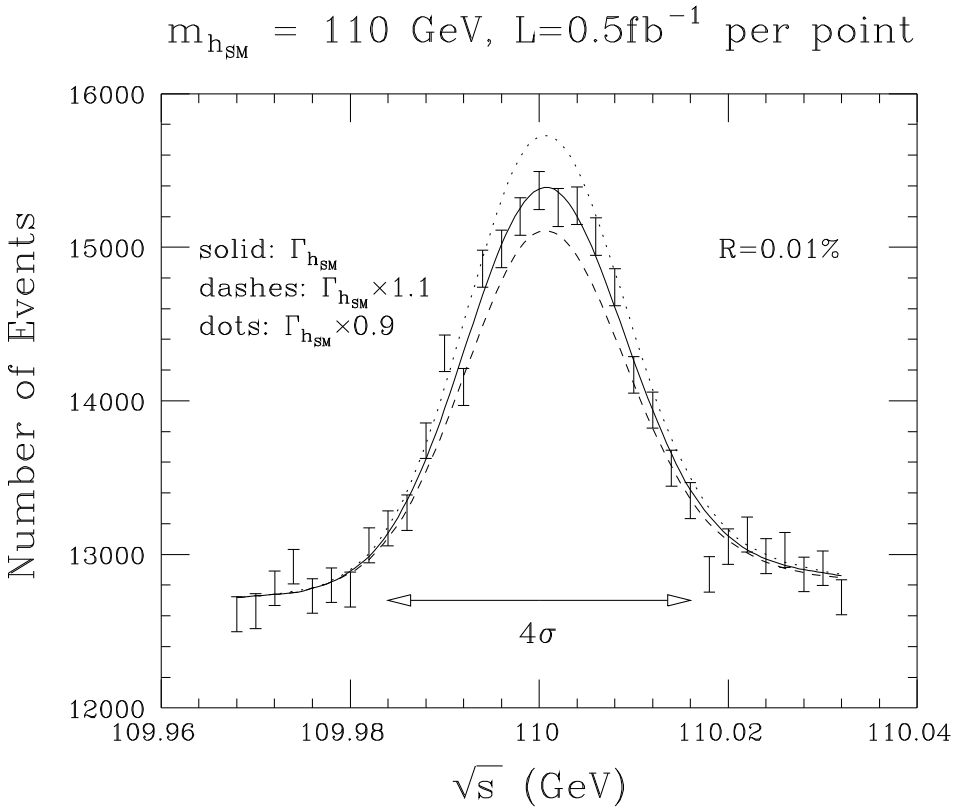


Figure 2.22: Number of events and statistical errors in the $b\bar{b}$ final state as a function of $\sqrt{s}$ in the vicinity of $m_{h_{SM}} = 110$ GeV, assuming $R = 0.01\%$, and $L = 0.5$ fb^{-1} at each data point. The precise theoretical prediction is given by the solid line. The dotted (dashed) curve is the theoretical prediction if $\Gamma^{\rm tot}_{h_{SM}}$ is decreased (increased) by 10%, *keeping the* $\Gamma(h_{SM} \to \mu^+\mu^-)$ *and* $\Gamma(h_{SM} \to b\bar{b})$ *partial widths fixed at the predicted SM value.*

peak is a measure of $\Gamma^{\rm tot}_{h_{SM}}$. The seven central points would determine $\Gamma^{\rm tot}_{h_{SM}}$ to better than 10%.

Since in practice we are not able to accurately pre-determine the partial widths, a *model-independent* technique for discriminating between the total width of the SM h_{SM} and that of some other SM-like h must be devised that does not involve a theoretical computation of the partial widths. Such a determination of the total width requires measurements sensitive to the breadth of the spectrum illustrated in Fig. 2.3.2. We outline below a procedure by which roughly $L \sim 3$ fb^{-1} of total luminosity will allow a $\pm 33\%$ determination of $\Gamma^{\rm tot}_{h_{SM}}$ (for

$m_{h_{SM}} = 110$ GeV) without any assumption regarding the partial widths.

The key observation is that if one adjusts the partial widths so that the normalization of the theoretical curve at $\sqrt{s} = m_{h_{SM}}$ agrees with experiment, then the normalization of the wings of the theoretical curve will be correspondingly increased or decreased in the case that $\Gamma_h^{\rm tot}$ is larger or smaller, respectively. Experimental measurements of sufficient precision both at a central $\sqrt{s}$ value and on the wings would thus allow a direct measurement of $\Gamma_{h_{SM}}^{\rm tot}$ via the ratio of the central peak cross section to the cross sections on the wings (the partial widths cancel out in the ratio). With this in mind, we define the quantity

$$d \equiv |\sqrt{s} - m_{h_{SM}}|/\sigma_{\sqrt{s}} \tag{2.65}$$

and propose the following procedure:

(1) Perform a rough scan to determine $m_{h_{SM}}$ to a precision $\sigma_{\sqrt{s}} d$, with $d \lesssim 0.3$; d will not be known ahead of time, but the value of d, and hence of $m_{h_{SM}}$ will be determined by the procedure.

(2) Then perform three measurements. At $\sqrt{s}_1 = m_{h_{SM}} + \sigma_{\sqrt{s}} d$ we employ a luminosity of L_1 and measure the total rate $N_1 = S_1 + B_1$. Then perform two additional measurements at

$$\sqrt{s}_2 = \sqrt{s}_1 - n_{\sigma_{\sqrt{s}}} \sigma_{\sqrt{s}} \tag{2.66}$$

and one at

$$\sqrt{s}_3 = \sqrt{s}_1 + n_{\sigma_{\sqrt{s}}} \sigma_{\sqrt{s}} \tag{2.67}$$

yielding $N_2 = S_2 + B_2$ and $N_3 = S_3 + B_3$ events, respectively, employing luminosities of $L_2 = \rho_2 L_1$ and $L_3 = \rho_3 L_1$. We find that $n_{\sigma_{\sqrt{s}}} \sim 2$ and $\rho_2 = \rho_3 \sim 2.5$ are optimal for maximizing sensitivity and minimizing the error in determining d (*i.e.* $m_{h_{SM}}$) and $\Gamma_{h_{SM}}^{\rm tot}$.

(3) To determine $m_{h_{SM}}$ and $\Gamma_{h_{SM}}^{\rm tot}$ consider the ratios

$$\begin{aligned} r_2 &\equiv (S_2/\rho_2)/S_1 \quad = (S_2/L_2)/(S_1/L_1) \\ r_3 &\equiv (S_3/\rho_3)/S_1 \quad = (S_3/L_3)/(S_1/L_1)\,. \end{aligned} \tag{2.68}$$

The ratios r_2 and r_3 are governed by d and $\Gamma_{h_{SM}}^{\rm tot}$. Conversely, we have implicitly $d = d(r_2, r_3)$ and $\Gamma_{h_{SM}}^{\rm tot} = \Gamma_{h_{SM}}^{\rm tot}(r_2, r_3)$. Determining the statistical errors $\Delta m_{h_{SM}}$ and

$\Delta\Gamma^{\rm tot}_{h_{SM}}$ is then simply a matter of computing the partial derivatives of d and $\Gamma^{\rm tot}_{h_{SM}}$ with respect to the $r_{2,3}$ (we do this numerically) and using errors on the ratios $r_{2,3}$ implied by statistics. The procedure is detailed in sec. 2.12.3, as is the cross check on its accuracy that we have used.

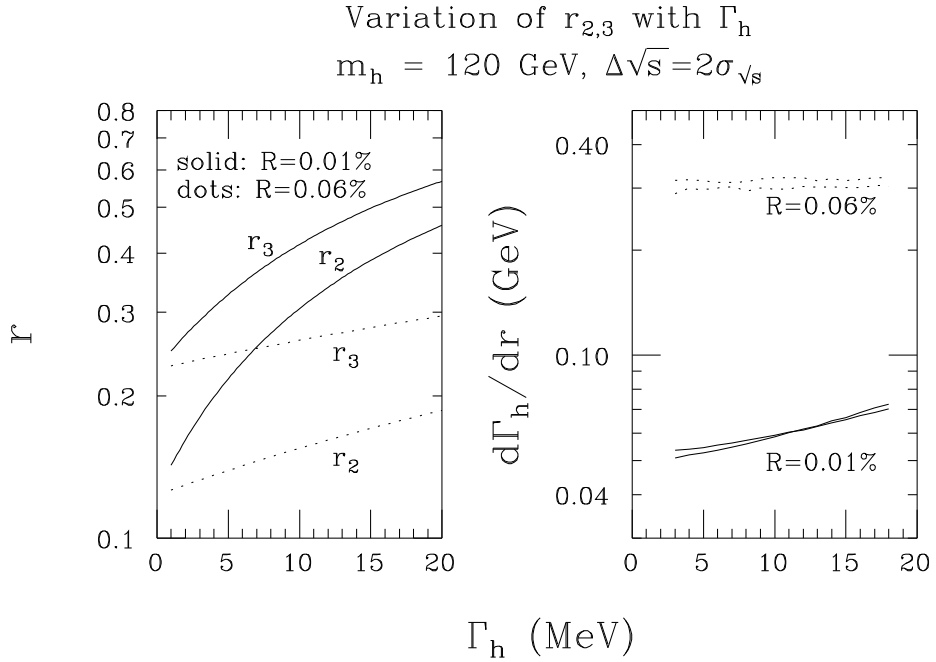


Figure 2.23: We plot r_2 and r_3 as a function of Higgs width, $\Gamma^{\rm tot}_h$, for resolutions of $R = 0.01\%$ and $R = 0.06\%$, assuming that $\sqrt{s} = m_h = 120$ GeV. Also shown are the derivatives $d\Gamma^{\rm tot}_h/dr$ as a function of $\Gamma^{\rm tot}_h$. We have taken $n_{\sigma_{\sqrt{s}}} = 2$ corresponding to a shift in $\sqrt{s}$ of $\mp 2\sigma_{\sqrt{s}}$ in computing r_2 and r_3, respectively.

The utility of the ratios r_2 and r_3 is basically governed by how rapidly they vary as d and $\Gamma^{\rm tot}_h$ are varied in the ranges of interest. Since we are most interested in $\Gamma^{\rm tot}_h$ here, we illustrate the sensitivity of $r_{2,3}$ to $\Gamma^{\rm tot}_h$ in Fig. 2.3.2 taking $\sqrt{s} = m_h = 120$ GeV. For this figure we employ $n_{\sigma_{\sqrt{s}}} = 2$ for computing r_2 and r_3, respectively. Results are shown for resolutions $R = 0.01\%$ and $R = 0.06\%$. Because of the bremsstrahlung tail, r_2 is substantially larger than r_3. Nonetheless, both r_2 and r_3 show rapid variation as $\Gamma^{\rm tot}_h$ varies in the vicinity of $\Gamma^{\rm tot}_{h_{SM}}$ in the case of $R = 0.01\%$, but much less variation if $R = 0.06\%$. The error in the determination of $\Gamma^{\rm tot}_h$ is basically determined by $d\Gamma^{\rm tot}_h/dr_{2,3}$. Figure 2.3.2 shows that these

derivatives are almost the same and quite small for $R = 0.01\%$. The much larger values of these derivatives for $R = 0.06\%$ imply that determining $\Gamma_h^{\rm tot}$ accurately would be very difficult in this case.

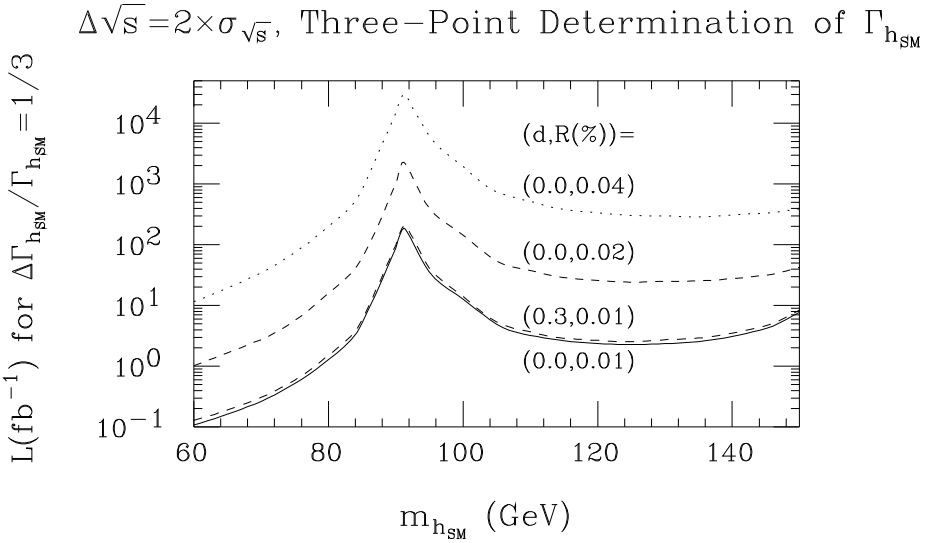


Figure 2.24: Luminosity required for a $\Delta\Gamma^{\rm tot}_{h_{SM}}/\Gamma^{\rm tot}_{h_{SM}} = 1/3$ measurement in the $b\overline{b}$ final state using the three point technique described in the text. Results for resolutions of $R = 0.01\%$, 0.02% and 0.04% are shown for $d = 0$, where $d = |\sqrt{s} - m_{h_{SM}}|/\sigma_{\sqrt{s}}$. The result for $d = 0.3$ and $R = 0.01\%$ is also shown.

In Fig. 2.3.2, we plot the total luminosity $L = L_1 + L_2 + L_3 = 6L_1$ required to achieve $\Delta\Gamma^{\rm tot}_{h_{SM}}/\Gamma^{\rm tot}_{h_{SM}} = 1/3$ in the $b\overline{b}$ final state as a function of $m_{h_{SM}}$ for several beam resolutions. (The error scales statistically; *e.g.* to achieve a 10% measurement would require $(10/3)^2$ as much luminosity.) We also illustrate the fact that the total luminosity required is rather insensitive to the initial choice of d for $d \lesssim 0.3$; $d = 0.3$ results in no more than a 20% increase in the luminosity needed relative to $d = 0$.

In Fig. 2.3.2, we plot the 1σ error $\Delta m_{h_{SM}}$ that results using our three-point technique after accumulating the luminosity required for a $\Delta\Gamma^{\rm tot}_{h_{SM}}/\Gamma^{\rm tot}_{h_{SM}} = 1/3$ measurement in the $b\overline{b}$ final state. The specific result plotted is for $R = 0.01\%$ and $d = 0$, but is essentially independent of R and d given the stated luminosity. Also shown, for comparison, is $\Gamma^{\rm tot}_{h_{SM}}$ itself. We see that $\Delta m_{h_{SM}}$ is of order 1.5–2 times $\Gamma^{\rm tot}_{h_{SM}}/10$, *i.e.* a fraction of an MeV for $m_{h_{SM}} \lesssim 130$ GeV. (Again, $\Delta m_{h_{SM}}$ scales as $1/\sqrt{L}$.)

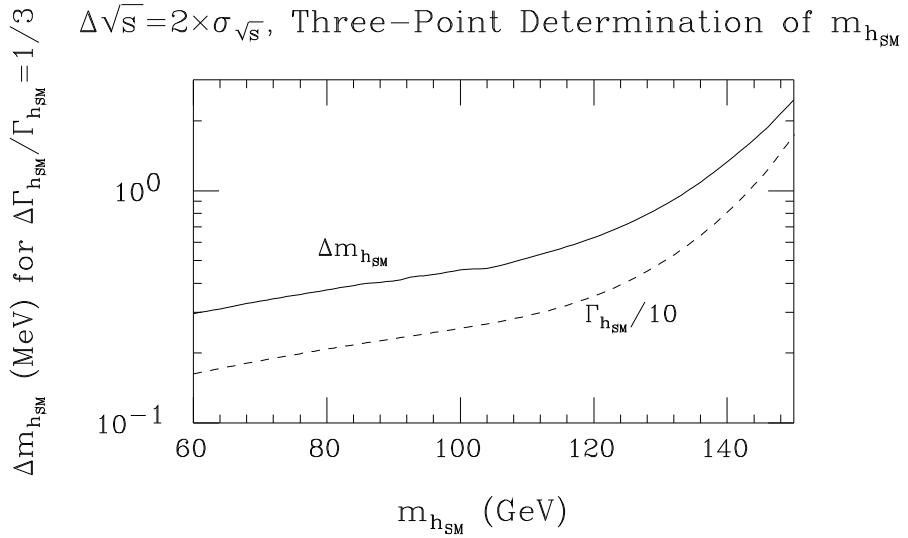


Figure 2.25: We plot the 1σ error, $\Delta m_{h_{SM}}$, in the determination of $m_{h_{SM}}$ using the three point technique described in the text with $R = 0.01\%$ and $d = 0$. The error given is that achieved for the luminosity that allows a $\Delta\Gamma^{\rm tot}_{h_{SM}}/\Gamma^{\rm tot}_{h_{SM}} = 1/3$ measurement in the $b\bar{b}$ final state. For such luminosity, $\Delta m_{h_{SM}}$ is essentially independent of R and d. Also shown, for comparison, is $\Gamma^{\rm tot}_{h_{SM}}/10$.

It should be stressed that the ability to precisely set the energy of the machine when the three measurements are taken is crucial for the success of the three-point technique. A misdetermination of the spacing of the measurements in Eqs. (2.66) and (2.67) by just 3% (*i.e.* $\sqrt{s}$ uncertainty of order 0.25 MeV for any one setting near $m_{h_{SM}} \sim 120$ GeV) would result in an error in $\Gamma^{\rm tot}_{h_{SM}}$ of 30%. For a measurement of $\Gamma^{\rm tot}_{h_{SM}}$ at the 10% level the $\sqrt{s}$ settings must be precise at a level of better than one part in 10^6. This is possible [61] provided the beam can be partially polarized so that the precession of the spin of the muon as it circulates in the final storage ring can be measured. From the precession and the rotation rate the energy can be determined. The ability to perform this critical measurement needed for the determination of the total width of a narrow Higgs must be incorporated in the machine design.

Precision Measurements: $\Gamma(h \to \mu^+\mu^-) \times BF(h \to X)$

Assuming that the Higgs width is much narrower than the rms uncertainty in $\sqrt{s}$,

Eq. (2.57) shows that the event rate in a given channel measures $G(X) = \Gamma(h \to \mu^+\mu^-) \times BF(h \to X)$. If the background can be determined precisely (either by off-resonance measurements or theory plus Monte Carlo calculation), the error in the determination of this product is $\sqrt{N}/S$, where $N = S + B$ and S, B are the number of signal, background events, respectively. The results for $\sqrt{N}/S$ in the case of $P = 0$ and $L = 50$ fb^{-1} in the $b\bar{b}$, $WW^{(\star)}$ and $ZZ^{(\star)}$ modes are shown in Fig. 2.3.2 for $h = h_{SM}$. For each final state, the efficiencies and procedures employed are precisely those discussed with regard to Fig. 2.3.2. Good accuracy in this measurement is possible for $m_{h_{SM}} \lesssim 2m_W$ even if $m_{h_{SM}}$ is near m_Z.

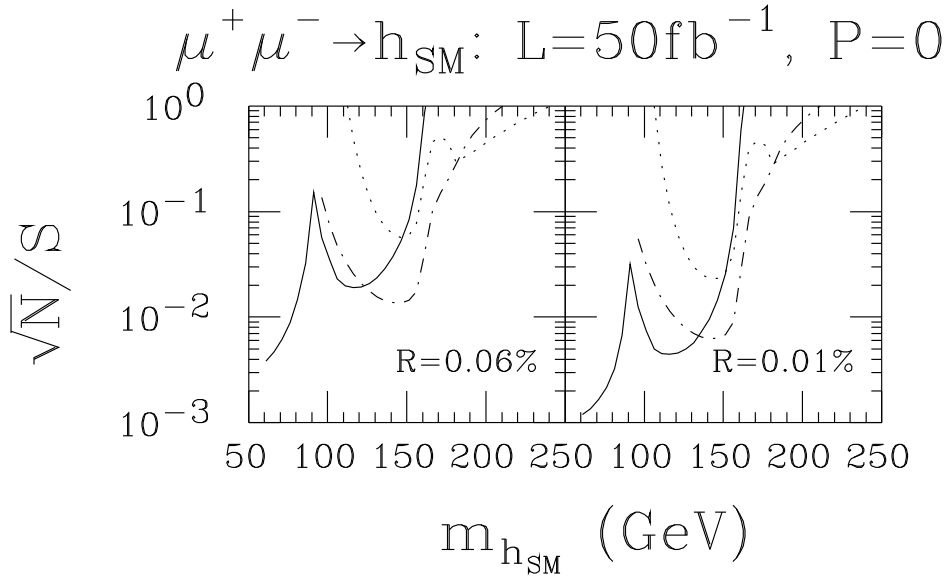


Figure 2.26: Fractional error in determining $\Gamma(h_{SM} \to \mu\mu) \times BF(h_{SM} \to X)$ for $X = b\bar{b}$ (solid), $WW^{(\star)}$ (dotdash) and $ZZ^{(\star)}$ (dots), assuming $L = 50$ fb^{-1}. (See text for $WW^\star$ and $ZZ^\star$ final states employed.)

h^0 or h_{SM}?

We now discuss the possibility of distinguishing the MSSM h^0 from the SM h_{SM} using precision measurements of $\Gamma_h^{\rm tot}$ and $G(b\bar{b}) \equiv \Gamma(h \to \mu\mu) \times BF(h \to b\bar{b})$. The accuracy to which $\Gamma_h^{\rm tot}$ and $G(b\bar{b})$ need to be determined can be gauged by the ratio of the h^0 predictions to the h_{SM} predictions for these quantities at $m_{h^0} = m_{h_{SM}}$. Contours for various fixed values of these ratios are plotted in Fig. 2.3.2 in the standard $(m_{A^0}, \tan\beta)$ parameter space [68]. In computing results for $\Gamma_h^{\rm tot}$ and $G(b\bar{b})$ for h^0 we have taken $m_{\tilde{t}} = 1$ TeV, $m_t = 175$ GeV,

and included two-loop/RGE-improved radiative corrections to the Higgs masses, mixing angles and self-couplings, neglecting squark mixing. The ratios for both $\Gamma_h^{\rm tot}$ and $G(b\overline{b})$ are substantially bigger than 1, even out to fairly large m_{A^0} values. This is because the h^0 retains somewhat enhanced $b\overline{b}$, $\tau^+\tau^-$ and $\mu^+\mu^-$ couplings until quite large m_{A^0} values. Two facts are of particular importance:

- $\Gamma_{h^0}^{\rm tot}$ is enhanced relative to $\Gamma_{h_{SM}}^{\rm tot}$ by virtue of the enhanced partial widths into its dominant decay channels, $b\overline{b}$ and $\tau^+\tau^-$.
- The enhancement in $G(b\overline{b})$ derives mainly from $\Gamma(h \to \mu\mu)$, as can be deduced by comparing Fig. 2.3.2(b) and 2.3.2(c).

This latter point is also apparent in Fig. hltohsmratios(d), where we observe that the MSSM to SM ratio of $BF(h \to b\overline{b})$'s is very close to 1 along the 1.1 contour of the MSSM/SM $G(b\overline{b})$. This is because the enhanced $b\overline{b}$ partial width in the numerator of $BF(h \to b\overline{b})$ is largely compensated by the extra contribution to the total width from this same channel. Thus, in comparing the MSSM to the SM, a measurement of $G(b\overline{b})$ is most sensitive to deviations of $\Gamma(h \to \mu\mu)$ from SM expectations. As seen numerically in Fig. 2.3.2(e), $\Gamma(h \to \mu\mu)$ grows rapidly at lower m_{A^0} or higher $\tan\beta$. For small squark mixing, a deviation in $G(b\overline{b})$ from the SM value implies almost the same percentage deviation of $\Gamma(h \to \mu\mu)$ from its SM value. However, when squark mixing is large, this equality breaks down. In general, one must separately determine $\Gamma(h \to \mu\mu)$ in order to probe MSSM vs. SM differences. The procedure for this will be discussed shortly.

The measured value of m_h provides a further constraint. For example, suppose that a Higgs boson is observed with $m_h = 110$ GeV. A fixed value for m_h implies that the parameters which determine the radiative corrections to m_{h^0} must change as m_{A^0} and $\tan\beta$ are varied. For example, if squark mixing is neglected, then the appropriate value of $m_{\tilde{t}}$ is a function of m_{A^0} and $\tan\beta$. Given the assumption of no squark mixing and the fixed value of $m_h = 110$ GeV, results for the same ratios as plotted in Fig. 2.3.2 are given in Fig. 2.3.2. Also shown are contours of fixed $\Gamma(h^0 \to \mu\mu)$ and contours of fixed $m_{\tilde{t}}$ (as required to achieve $m_{h^0} = 110$ GeV). The vertical nature of the $\mu\mu$ ratio and partial width contours implies that a measurement of any of these quantities could provide a determination of m_{A^0} (but would yield little information about $\tan\beta$).

Contours for other mixing assumptions, can also be plotted. The only contours that remain essentially unaltered as the amount of squark mixing is varied (keeping $m_h = 110$ GeV)

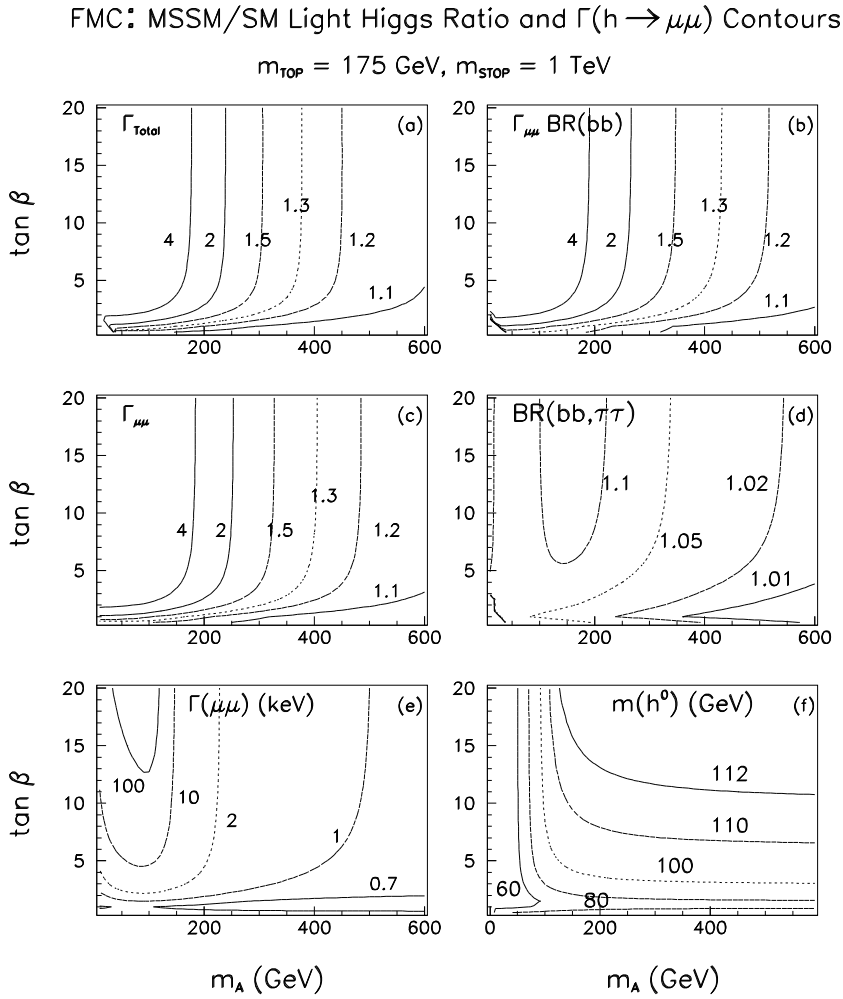


Figure 2.27: Contours of constant MSSM/SM ratios for $\Gamma_h^{\rm tot}$, $\Gamma(h \to \mu\mu) \times BF(h \to b\bar{b})$, $\Gamma(h \to \mu\mu)$ and $BF(h \to b\bar{b}, \tau\tau)$ in $(m_{A^0}, \tan\beta)$ parameter space. We have taken $m_t = 175$ GeV, $m_{\tilde{t}} = 1$ TeV, and included two-loop/RGE-improved radiative corrections, neglecting squark mixing, for Higgs masses, mixing angles and self-couplings. Also shown are contours for fixed values of $\Gamma(h^0 \to \mu\mu)$ using units of keV, and contours of fixed m_{h^0}. This graph was obtained using the programs developed for the work of Ref. [68].

are those for the ratio $\Gamma(h^0 \to \mu\mu)/\Gamma(h_{SM} \to \mu\mu)$ and for the $\Gamma(h^0 \to \mu\mu)$ partial width itself. Once $m_{h^0} \lesssim 100$ GeV, even these contours show substantial variation as a function

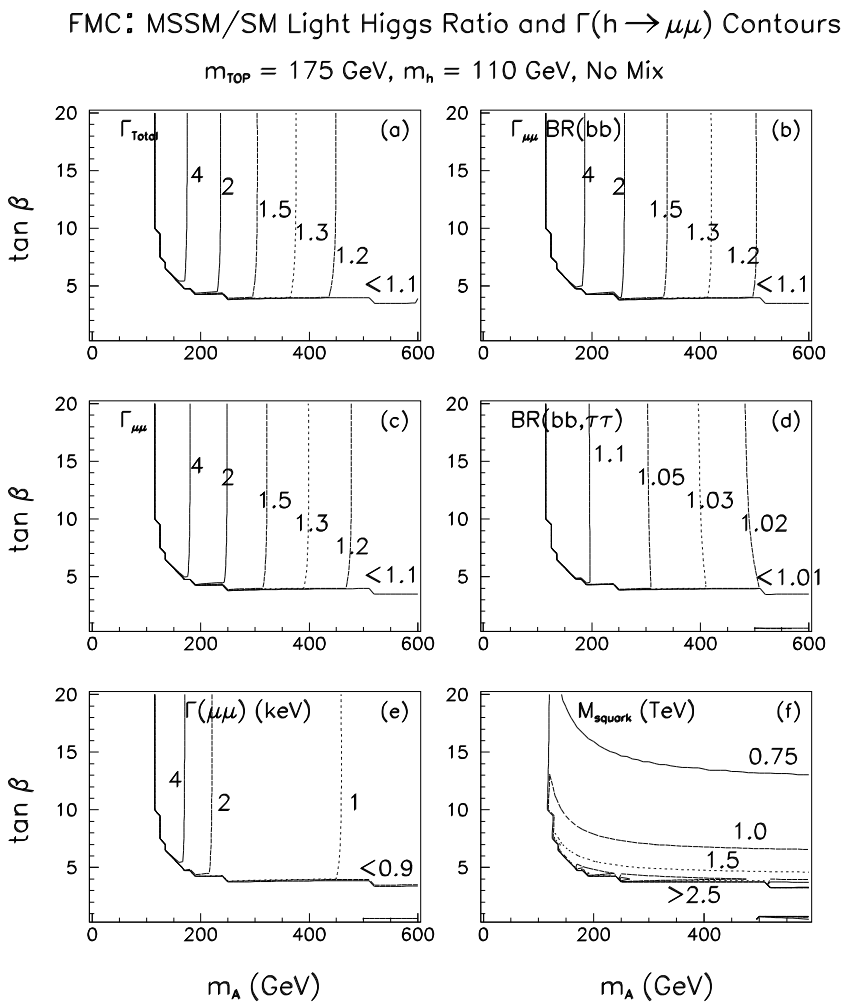


Figure 2.28: Contours of constant MSSM/SM ratios for $\Gamma_h^{\rm tot}$, $\Gamma(h \to \mu\mu) \times BF(h \to b\overline{b})$, $\Gamma(h \to \mu\mu)$ and $BF(h \to b\overline{b}, \tau\tau)$ in $(m_{A^0}, \tan\beta)$ parameter space. We have taken $m_t = 175$ GeV, and we adjust $m_{\tilde{t}}$ so as to keep a fixed value of $m_{h^0} = 110$ GeV after including two-loop/RGE-improved radiative corrections for Higgs masses, mixing angles and self-couplings, neglecting squark mixing. Also shown are contours for fixed values of $\Gamma(h^0 \to \mu\mu)$ in keV units, and contours for fixed values of $m_{\tilde{t}}$ in TeV units. This graph was obtained using the programs developed for the work of Ref. [68].

of the squark mixing parameters. However, it remains true that a determination of the $\mu\mu$ partial width or partial width ratio provides at least a rough determination of m_{A^0}.

In order to assess the observability of the differences between predictions for $\Gamma_h^{\rm tot}$, $G(b\overline{b})$, and $\Gamma(\mu\mu)$ for the h^0 compared to the h_{SM}, we must examine more closely the error in the experimental determination of these quantities, and consider the theoretical uncertainties in our predictions for them.

Interpreting a measurement of $\boldsymbol{\Gamma_h^{\rm tot}}$ Consider first the total width measurement. Here, the experimental error is the key issue. The h^0 may have a mass of order 110 GeV in the large-m_{A^0} region where it is SM-like, provided $\tan\beta$ is not near 1 (see Fig. 2.17). According to Fig. 2.3.2, $L \sim 3$ fb^{-1} is required to measure $\Gamma_h^{\rm tot}$ to $\pm 33\%$, provided $R = 0.01\%$. A $\pm 10\%$ measurement would require $L \sim 33$ fb^{-1} (using $\Delta\Gamma_{h_{SM}}^{\rm tot} \propto 1/\sqrt{L}$). As seen most clearly from Fig. 2.3.2, this accuracy would probe MSSM/SM differences at the 3σ level for $m_{A^0} \lesssim 400$ GeV if squark mixing is small.

Detecting a difference between the h^0 and h_{SM} using $\Gamma_h^{\rm tot}$ could prove either somewhat easier or much more difficult than outlined above, because the $\tan\beta$, $m_{\tilde{t}}$ values and the degree of squark mixing could very well be different from those assumed above. For example, if $m_{h^0} = 110$ GeV, $\tan\beta \gtrsim 5$ and squark mixing is large, m_{A^0} values above 400 GeV would be probed at the 3σ level by a 10% measurement of $\Gamma_h^{\rm tot}$. On the other hand, the radiative corrections could yield a smaller m_{h^0} value, *e.g.* $m_{h^0} \lesssim 100$ GeV is quite likely if $\tan\beta$ is near 1 or $m_{\tilde{t}}$ is small. In this range, predicted deviations from predictions for the h_{SM} with $m_{h_{SM}} = m_{h^0}$ are not dissimilar to those obtained discussed above. However, a luminosity $L \gtrsim 100$ fb^{-1} would be required for a $\pm 10\%$ measurement of $\Gamma_h^{\rm tot}$ for 80 GeV $\lesssim m_{h^0} \lesssim$ 100 GeV.

Other theoretical uncertainties include: i) extra contributions to $\Gamma_{h^0}^{\rm tot}$ in the MSSM model from SUSY decay modes; ii) the gg decay width of the h^0 could be altered by the presence of light colored sparticles; iii) the h_{SM} could have enhanced gg decay width due to heavy colored fermions (*e.g.* from a fourth family).

Nonetheless, a $\mu^+\mu^-$ collider determination of $\Gamma_h^{\rm tot}$ will be a crucial component in a model-independent determination of all the properties of a SM-like h, and could provide the first circumstantial evidence for a MSSM Higgs sector prior to direct discovery of the non-SM-like MSSM Higgs bosons.

Interpreting a measurement of $\Gamma(h \to \mu\mu) \times BF(h \to b\bar{b})$ How does the h^0–h_{SM} discrimination power of the total width measurement compare to that associated with a measurement of $G(b\bar{b}) \equiv \Gamma(h \to \mu\mu) \times BF(h \to b\bar{b})$? Figure 2.3.2 shows that $\pm 0.4\%$ accuracy in the determination of $G(b\bar{b})$ is possible for $L = 50$ fb^{-1} and $R = 0.01\%$ in the $m_{h^0} \sim 110$–115 GeV mass range predicted for $m_{A^0} \gtrsim 2m_Z$ and larger $\tan\beta$ values, assuming $m_{\widetilde{t}} \gtrsim 0.75$ TeV and no squark mixing.

An uncertainty in $BF(h \to b\bar{b})$ arises from $\Gamma(h \to b\bar{b}) \propto m_b^2$ due to the uncertainty in m_b. Writing $BF(h \to b\bar{b}) = \Gamma_b/(\Gamma_b + \Gamma_{\rm non-}{}_b)$, the error in $BF(h \to b\bar{b})$ is given by

$$\Delta BF(h \to b\bar{b}) = \frac{2\Delta m_b}{m_b} BF(h \to b\bar{b}) BF(h \to \text{non} - b) \; . \tag{2.69}$$

Since $BF(h \to$ non-$b)$ is not very large (0.1 to 0.2 in the mass range in question for either the h_{SM} or h^0), even a 10% uncertainty in m_b only leads to $\Delta BF(h \to b\bar{b}) \lesssim 0.05$. Eventually m_b may be known to the 5% level, leading to $\lesssim 2.5\%$ uncertainty in the branching fraction. Comparison to Fig. 2.3.2 shows that a 2.5% uncertainty from m_b, in combination with a still smaller statistical error, has the potential for h^0–h_{SM} discrimination at the 3σ statistical level out to large m_{A^0} for $m_h = 110$ GeV, if squark mixing is small. However, as squark mixing is increased, it turns out that the maximum m_{A^0} that can potentially be probed decreases if $\tan\beta$ is large.

$BF(h \to b\bar{b})$ is also subject to an uncertainty from the total width. For example, in the MSSM $BF(h^0 \to b\bar{b})$ could be smaller than the SM prediction if $\Gamma^{\rm tot}_{h^0}$ is enhanced due to channels other than the $b\bar{b}$ channel itself (*e.g.* by supersymmetric decay modes, or a larger than expected gg decay width due to loops containing supersymmetric colored sparticle or heavy colored fermions). Thus, a measurement of $G(b\bar{b})$ alone is not subject to unambiguous interpretation.

We note that the $L = 50$ fb^{-1} $\mu^+\mu^-$ collider measurement of $G(b\bar{b})$ is substantially more powerful than a $L = 50$ fb^{-1} precision measurement of $\sigma(e^+e^- \to Zh) \times BF(h \to b\bar{b})$ at an e^+e^- collider [41]. The ratio of the h^0 prediction to the h_{SM} prediction is essentially equal to the h^0 to h_{SM} $BF(h \to b\bar{b})$ ratio and is predicted to be within 1% (2%) of unity along a contour very close to the 1.1 (1.2) contour of $\Gamma(h \to \mu^+\mu^-)BF(h \to b\bar{b})$; see panels (b) and (d) in Figs. 2.3.2 and 2.3.2. Since at best 5% deviations in $G(b\bar{b})$ and $BF(h \to b\bar{b})$ can be detected at the 1σ level (after combining a possibly small statistical error with a large theoretical error), we see from the 1.05 ratio contour for $BF(h \to b\bar{b})$ in Figs. 2.3.2 and 2.3.2 that the $\sigma(Zh)BF(h \to b\bar{b})$ and $BF(h \to b\bar{b})$ ratios, that can be determined experimentally

at an e^+e^- collider, only probe as far as $m_{A^0} \lesssim 250$–300 GeV at the 1σ significance level, with even less reach at the 3σ level.

We must again caution that if m_h is close to m_Z, there could be substantially worse experimental uncertainty in the $G(b\overline{b})$ measurement than taken above. Pre-knowledge of m_h is necessary to determine the level of precision that could be expected for this measurement.

Combining measurements We now discuss how the independent measurements of $\Gamma_h^{\rm tot}$ and $G(b\overline{b})$ can be combined with one another and other experimental inputs to provide a model-independent determination of the properties of the h. We consider three complementary approaches.

(1) A model-independent determination of $\Gamma(h \to \mu\mu)$ can be made by combining the s-channel $\mu^+\mu^-$ collider measurement of $G(b\overline{b})$ with the value of $BF(h \to b\overline{b})$ measured in the Zh mode at an e^+e^- collider or the $\mu^+\mu^-$ collider. With $L = 50$ fb^{-1} of luminosity, $BF(h \to b\overline{b})$ can potentially be measured to $\pm 7\%$ [41]. From our earlier discussion, the error on $G(b\overline{b})$ will be much smaller than this if $m_h \gtrsim 100$ GeV, and $\Gamma(h \to \mu\mu)$ would be determined to roughly ± 8–10%. Figures 2.3.2 and 2.3.2 show that this procedure would probe the h^0 versus h_{SM} differences at the 3σ level out to $m_{A^0} \sim 400$ GeV if $\tan\beta$ is not close to 1 (see the 1.3 ratio contour in the figures). This is a far superior reach to that possible at the 3σ level at either the LHC, NLC and/or $\gamma\gamma$ collider. Further, we note that the $\mu\mu$ partial width at fixed $m_h \gtrsim 100$ GeV is relatively independent of the squark mixing scenario and provides a rather precise determination of m_{A^0} [41].

(2) A model-independent determination of $\Gamma(h \to b\overline{b})$ is possible by computing $\Gamma_h^{\rm tot} BF(h \to b\overline{b})$ using the value of $\Gamma_h^{\rm tot}$ measured at the $\mu^+\mu^-$ collider and the value of $BF(h \to b\overline{b})$ measured in the Zh mode. Taking 10% accuracy for the former and 7% accuracy for the latter, we see that the error on $\Gamma(h \to b\overline{b})$ would be of order 12%. The ratio contours for $\Gamma(h \to b\overline{b})$ are the same as the ratio contours for $\Gamma(h \to \mu\mu)$. Thus, ignoring systematics, this measurement could also probe out to $m_{A^0} \gtrsim 400$ GeV at the 3σ level if $m_h \sim 110$ GeV, see Fig. 2.3.2. However, the $2\Delta m_b/m_b$ systematic uncertainty in the partial width is also of order 10% for 5% uncertainty in m_b, implying a total statistical plus theoretical error of order 16%. This would restrict 3σ sensitivity to h^0 vs. h_{SM} differences to $m_{A^0} \lesssim 300$ GeV.

(3) A third approach uses only the $\mu^+\mu^-$ collider measurements. We note that

$$W \equiv \Gamma(h \to \mu\mu)\Gamma(h \to b\overline{b}) = [\Gamma_h^{\rm tot}] \times [\Gamma(h \to \mu\mu) BF(h \to b\overline{b})]\,. \tag{2.70}$$

In the MSSM (or any other type-II two-Higgs-doublet model) the $\mu\mu$ and $b\bar{b}$ squared couplings have exactly the same factor, call it f, multiplying the square of the SM coupling strength. Thus,

$$W = \Gamma(h \to \mu\mu)\Gamma(h \to b\bar{b}) \propto f^2 \left(\frac{g}{2m_W}\right)^4 m_\mu^2 m_b^2 \,. \tag{2.71}$$

Following our earlier discussion, in the MSSM f^2 would be $(1.3)^2 \sim 1.7$ along the 1.3 ratio contours for $\Gamma(h \to \mu\mu)$ in Figs. 2.3.2 and 2.3.2. For $m_h \gtrsim 100$ GeV, experimental errors in W of Eq. (2.70) would be dominated by the $\pm 10\%$ error on $\Gamma_h^{\rm tot}$. The dominant systematic error would be that from not knowing the value of m_b: $\Delta W/W = 2\Delta m_b/m_b$. Thus, a combined statistical and theoretical 1σ error for W below 20% is entirely possible for $m_h \gtrsim 100$ GeV, in which case deviations in f^2 from unity can be probed at the 3σ level for m_{A^0} values at least as large as $m_{A^0} \sim 400$ GeV. Since both $\Gamma(h^0 \to \mu\mu)$ and $\Gamma(h^0 \to b\bar{b})$ are relatively independent of the squark mixing scenario for fixed m_{h^0} and fixed m_{A^0}, a fairly reliable value of m_{A^0} would result from the determination of f^2.

By combining the strategies just discussed, one can do even better. Thus, a $\mu^+\mu^-$ collider has great promise for allowing us to measure the crucial $b\bar{b}$ and $\mu^+\mu^-$ couplings of a SM-like h, provided m_h is not within 10 GeV of m_Z (nor $\gtrsim 2m_W$) and that $m_A \lesssim 400$ GeV. In particular, for such masses we can distinguish the h^0 from the h_{SM} in a model-independent fashion out to larger m_{A^0} than at any other accelerator or combination of accelerators.

The $WW^\star$ and $ZZ^\star$ channels Precision measurements of $\Gamma(h \to \mu^+\mu^-)BF(h \to X)$ are also possible for $X = WW^\star$ and, to a lesser extent, $ZZ^\star$, see Fig. 2.3.2. Once again, $\Gamma(h \to \mu^+\mu^-)$ can be determined in a model-independent fashion using $BF(h \to X)$ measured in the Zh mode, and $\Gamma(h \to X)$ can be computed in a model-independent fashion as the product $BF(h \to X)\Gamma_h^{\rm tot}$. We will not go through the error analysis in detail for these cases, but clearly determination of both the WW and ZZ couplings will be possible at a reasonable statistical level. Unfortunately, the h^0WW, h^0ZZ couplings are very close to the SM values for $m_{A^0} \gtrsim 2m_W$ and the expected statistical errors would not allow h^0 vs. h_{SM} discrimination.

2.3.3 Non-SM-like Higgs Bosons in the MSSM

In what follows, we shall demonstrate that it is possible to observe the H^0 and A^0 in s-channel Higgs production for $m_{A^0} \sim m_{H^0} > \sqrt{s}/2$ over much of $(m_{A^0}, \tan\beta)$ parameter

space. It is this fact that again sets the $\mu^+\mu^-$ collider apart from other machines.

1. The LHC can only detect the H^0 and A^0 for masses above 200–250 GeV if $\tan\beta$ is either large or $\lesssim$ 3–5; a wedge of unobservability develops beginning at $m_{A^0} \gtrsim 200$ GeV, covering an increasingly wide range of $\tan\beta$ as m_{A^0} increases [69]. This is illustrated in Fig. 2.3.3 from Ref. [70].

2. At an e^+e^- collider, $Z^\star \to ZA^0, ZH^0$ production will be negligible when $m_{A^0} > 2m_Z$.

3. $e^+e^- \to Z^\star \to A^0H^0$ could easily be kinematically disallowed, especially for e^+e^- machine energies in the $\sqrt{s} \sim 500$ GeV range — GUT scenarios often give $m_{A^0} \sim m_{H^0} \gtrsim 300$ GeV.

4. If an e^+e^- collider is run in the photon-photon collider mode, discovery of the H^0 and A^0 in the $m_{A^0}, m_{H^0} \gtrsim 200$ GeV region via $\gamma\gamma \to A^0, H^0$ requires extremely high luminosity ($\gtrsim 200$ fb^{-1}) [71].

5. s-channel production of the A^0 and H^0 will not be significant in e^+e^- collisions due to the small size of the electron mass.

A $\mu^+\mu^-$ collider can overcome the limitations 3 and 5 of an e^+e^- collider, though not simultaneously. If the $\mu^+\mu^-$ collider is run at energies of $\sqrt{s} = m_{A^0} \sim m_{H^0}$, then we shall find that s-channel production will allow discovery of the A^0 and H^0 if $\tan\beta \gtrsim 3-4$. Here, the kinematical Higgs mass reach is limited only by the maximum $\sqrt{s}$ of the machine. Alternatively, the $\mu^+\mu^-$ collider can be designed to have $\sqrt{s} \sim 4$ TeV in which case $m_{A^0} \sim m_{H^0}$ values up to nearly 2 TeV can be probed via the $Z^\star \to A^0H^0$ process, a mass range that encompasses all natural GUT scenarios. We focus in this report on s-channel production and detection. In our analysis, we will assume that more or less full luminosity can be maintained for all $\sqrt{s}$ values over the mass range of interest (using multiple storage rings, as discussed in the introduction).

MSSM Higgs Bosons in the s-Channel: $\sqrt{s} = m_h$

Here we investigate the potential of a $\mu^+\mu^-$ collider for probing those Higgs bosons whose couplings to ZZ, WW are either suppressed or absent at tree-level — that is the A^0, the H^0 (at larger m_{A^0}), or the h^0 (at small m_{A^0}). The $WW^{(\star)}$ and $ZZ^{(\star)}$ final states in s-channel production are then not relevant. We consider first the $b\bar{b}$ and $t\bar{t}$ decay modes, although

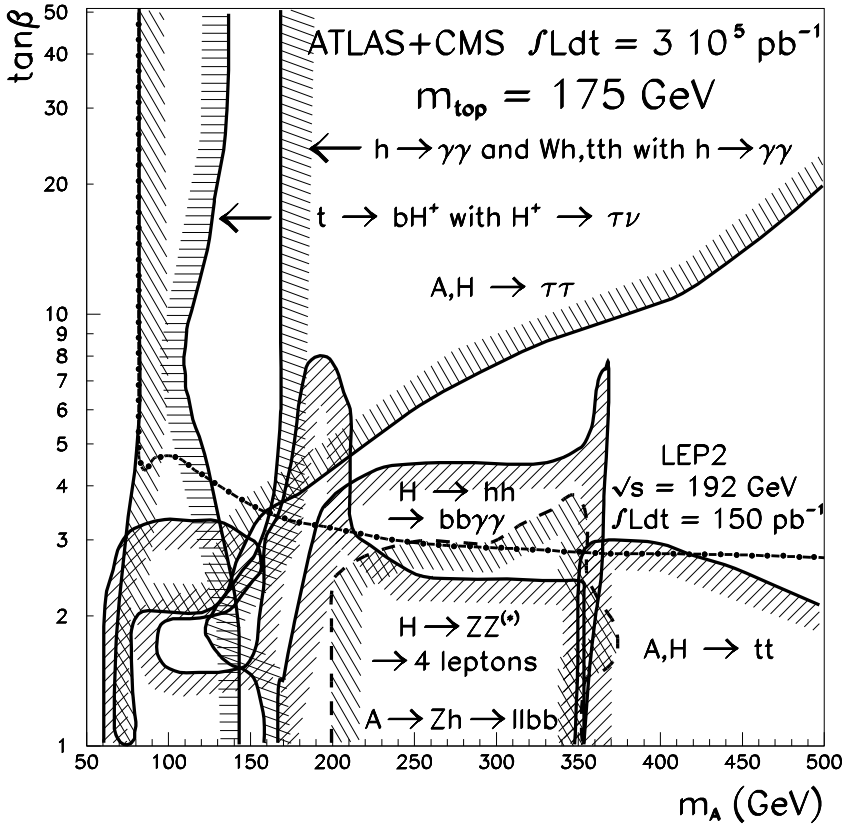


Figure 2.29: MSSM Higgs discovery contours (5σ) in the parameter space of the minimal supersymmetric model for ATLAS+CMS at the LHC: $L =$ 300 fb^{-1} per detector. Figure from Ref. [70]. Two-loop/RGE-improved radiative corrections are included for m_{h^0} and m_{H^0} assuming $m_{\tilde{t}} = 1$ TeV and no squark mixing.

we shall later demonstrate that the relatively background free $H^0 \to h^0 h^0$ or $A^0 A^0 \to b\bar{b}b\bar{b}$, $H^0 \to ZA^0 \to Zb\bar{b}$ and $A^0 \to Zh^0 \to Zb\bar{b}$ modes might also be useful.

Figure 2.3.3 shows the dominant branching fractions to $b\bar{b}$ and $t\bar{t}$ of Higgs bosons of mass $m_{A^0} = 400$ GeV $\approx m_{H^0}$ versus $\tan\beta$, taking $m_t = 170$ GeV. The $b\bar{b}$ decay mode is dominant for $\tan\beta > 5$, which is the region where observable signal rates are most easily obtained. From the figure we see that $BF(H^0, A^0 \to b\bar{b})$ grows rapidly with increasing $\tan\beta$ for $\tan\beta \lesssim 5$, while $BF(H^0, A^0 \to t\bar{t})$ falls slowly.

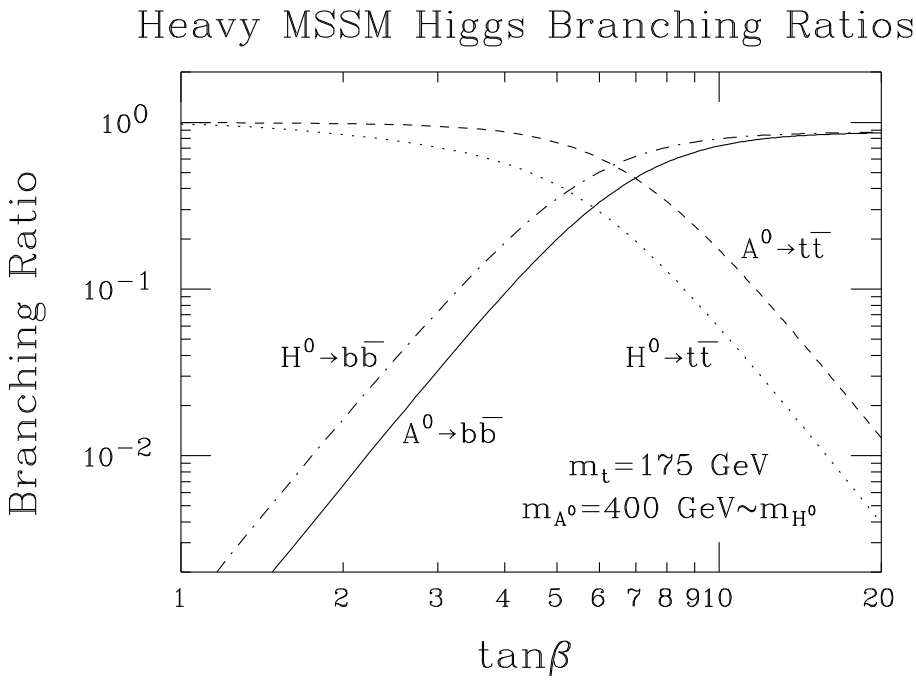


Figure 2.30: Dependence of the $b\bar{b}$ and $t\bar{t}$ branching fractions of the heavy supersymmetric Higgs bosons on $\tan\beta$. Results are for $m_t = 175$ GeV and include two-loop/RGE-improved radiative corrections to Higgs masses, mixing angles, and self-couplings, computed with $m_{\tilde{t}} = 1$ TeV neglecting squark mixing.

Resolution compared to Higgs widths

The first critical question is how the resolution in $\sqrt{s}$ compares to the H^0 and A^0 total widths. The calculated H^0 and A^0 widths are shown in Fig. 2.3 versus m_{H^0}, m_{A^0} for $\tan\beta = 2$ and 20. In Fig. 2.3.3 we give contours of constant total widths for the H^0 and A^0 in the $(m_{A^0}, \tan\beta)$ parameter space. For $m_{A^0}, m_{H^0} \lesssim 500$ GeV, the H^0 and A^0 are typically moderately narrow resonances ($\Gamma_{H^0,A^0} \sim 0.1$ to 6 GeV), unless $\tan\beta$ is larger than 20. For a machine energy resolution of $R = 0.06\%$, and Higgs masses in the 100 GeV to 1 TeV range, the resolution $\sigma_{\sqrt{s}}$ in $\sqrt{s}$ will range from roughly 0.04 GeV to to 0.4 GeV, see Eq. (2.54). Thus, Figs. 2.3 and 2.3.3 indicate that the H^0 and A^0 widths are likely to be somewhat larger than this resolution in $\sqrt{s}$. For $R = 0.01\%$, this is always the dominant situation.

When the $\sqrt{s}$ resolution is smaller than the Higgs width, then Eq. (2.55), with $\sqrt{s} \sim m_h$ shows that the cross section will behave as the product of the $\mu\mu$ and final state branching

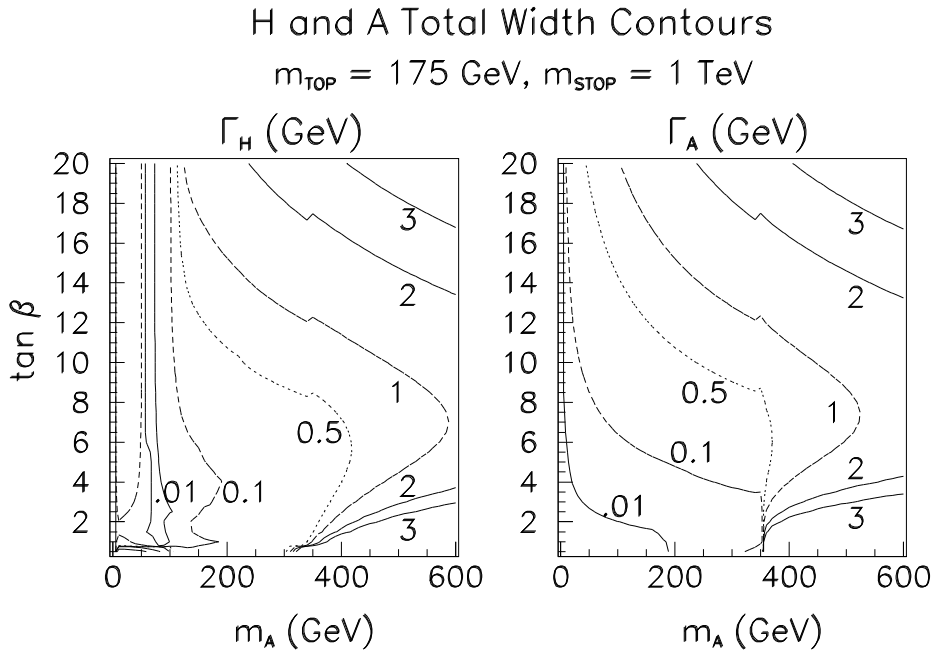


Figure 2.31: Contours of H^0 and A^0 total widths (in GeV) in the $(m_{A^0}, \tan\beta)$ parameter space. We have taken $m_t = 175$ GeV and included two-loop/RGE-improved radiative corrections using $m_{\tilde{t}} = 1$ TeV and neglecting squark mixing. SUSY decay channels are assumed to be absent.

fractions. For low to moderate $\tan\beta$ values, $BF(H^0, A^0 \to \mu\mu)$ and $BF(H^0, A^0 \to b\bar{b})$ grow with increasing $\tan\beta$, while $BF(H^0, A^0 \to t\bar{t})$ falls slowly. Thus, the number of H^0 and A^0 events in both the $b\bar{b}$ and $t\bar{t}$ channels increases with increasing $\tan\beta$. It is this growth with $\tan\beta$ that makes H^0, A^0 discovery possible for relatively modest values of $\tan\beta$ larger than 1. For higher $\tan\beta$ values, the $\mu\mu$ and $b\bar{b}$ branching fractions asymptote to constant values, while that for $t\bar{t}$ falls as $1/(\tan\beta)^4$. Thus, observability in the $t\bar{t}$ channel does not survive to large $\tan\beta$ values.

Overlapping Higgs resonances

The Higgs widths are a factor in the observability of a signal in that approximate Higgs mass degeneracies are not unlikely. For larger m_{A^0}, $m_{A^0} \sim m_{H^0}$, while at smaller m_{A^0} values, $m_{h^0} \sim m_{A^0}$ at larger $\tan\beta$, as illustrated in Fig. 2.3.3, where the plotted mass difference should be compared to the Higgs widths in Figs. 2.3 and 2.3.3. At large m_{A^0} and $\tan\beta$, there can be significant overlap of the A^0 and H^0 resonances. To illustrate the possibilities, we show

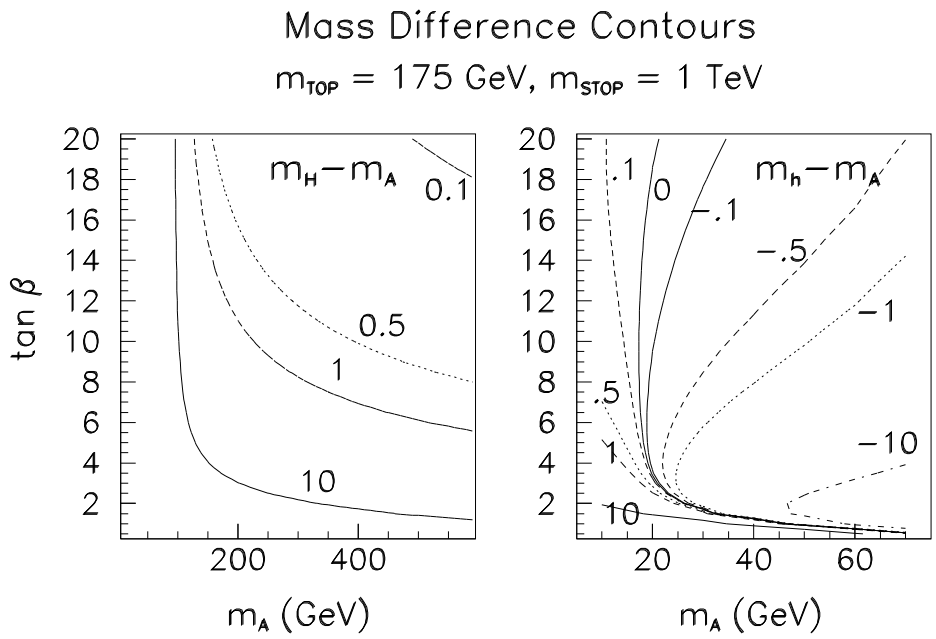


Figure 2.32: Contours of $m_{H^0} - m_{A^0}$ (in GeV) in the $(m_{A^0}, \tan\beta)$ parameter space. Two-loop/RGE-improved radiative corrections are included taking $m_t = 175$ GeV, $m_{\tilde{t}} = 1$ TeV, and neglecting squark mixing.

in Fig. 2.3.3 the event rate in the $b\bar{b}$ channel as a function of $\sqrt{s}$ (assuming $L = 0.01$ fb^{-1} and event detection/isolation efficiency $\epsilon = 0.5$) taking $m_{A^0} = 350$ GeV in the cases $\tan\beta = 5$ and 10. Continuum $b\bar{b}$ background is included. Results are plotted for the two different resolutions, $R = 0.01\%$ and $R = 0.06\%$. For $R = 0.01\%$, at $\tan\beta = 5$ the resonances are clearly separated and quite narrow, whereas at $\tan\beta = 10$ the resonances have become much broader and much more degenerate, resulting in substantial overlap; but, distinct resonance peaks are still visible. For $R = 0.06\%$, at $\tan\beta = 5$ the resonances are still separated, but have been somewhat smeared out, while at $\tan\beta = 10$ the H^0 and A^0 peaks are no longer separately visible. The $R = 0.06\%$ smearing does not greatly affect the observation of a signal, but would clearly make separation of the H^0 and A^0 peaks and precise determination of their individual widths much more difficult.

In the following section, we perform our signal calculations by centering $\sqrt{s}$ on m_{A^0}, but including any H^0 signal tail, and vice versa. At small m_{A^0}, there is generally only small overlap between the A^0 and h^0 since their widths are small, but we follow a similar procedure there. We also mainly employ the optimistic $R = 0.01\%$ resolution that is highly

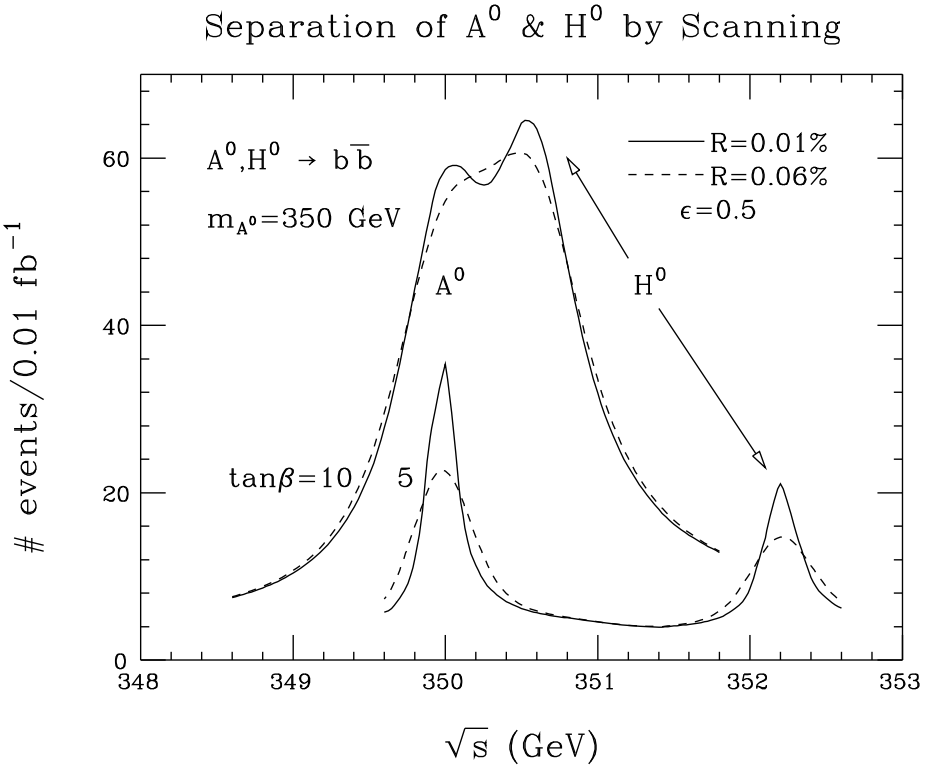


Figure 2.33: Plot of $b\overline{b}$ final state event rate as a function of $\sqrt{s}$ for $m_{A^0} = 350$ GeV, in the cases $\tan\beta = 5$ and 10, resulting from the H^0, A^0 resonances and the $b\overline{b}$ continuum background. We have taken $L = 0.01$ fb^{-1} (at any given $\sqrt{s}$), $\epsilon = 0.5$, $m_t = 175$ GeV, and included two-loop/RGE-improved radiative corrections to Higgs masses, mixing angles and self-couplings using $m_{\tilde{t}} = 1$ TeV and neglecting squark mixing. SUSY decays are assumed to be absent. Curves are given for two resolution choices: $R = 0.01\%$ and $R = 0.06\%$

preferred for a SM-like Higgs boson. Since the MSSM Higgs bosons do not have especially small widths, results for $R = 0.06\%$ are generally quite similar.

Observability for h^0, H^0 and A^0

We first consider fixed $\tan\beta$ values of 2, 5, and 20, and compute $\epsilon\overline{\sigma}_h BF(h \to b\overline{b}, t\overline{t})$ for $h = h^0, H^0, A^0$ as a function of m_{A^0}. (The corresponding h^0 and H^0 masses can be found in Fig. 2.17.) Our results for $R = 0.01\%$ appear in Figs. 2.3.3, 2.3.3, and 2.3.3. Also shown

in these figures are the corresponding $S/\sqrt{B}$ values assuming an integrated luminosity of $L = 0.1$ fb^{-1}; results for other L possibilities are easily obtained by using $S/\sqrt{B} \propto 1/\sqrt{L}$. These figures also include (dot-dashed) curves for $R = 0.06\%$ in the $b\bar{b}$ channel at $\tan\beta = 2$.

Figure 2.3.3 shows that the h^0 can be detected at the 5σ statistical level with just $L = 0.1$ fb^{-1} for essentially all of parameter space, if $R = 0.01\%$. Only for $\tan\beta \lesssim 2$ is m_{h^0} sufficiently near m_Z at large m_{A^0} (for which its $\mu^+\mu^-$ coupling is not enhanced) that more luminosity may be required. At low m_{A^0}, the h^0 is not SM-like and has highly enhanced $\mu^+\mu^-$ and $b\bar{b}$ couplings. It is also no longer extremely narrow, and is produced with a very high rate implying that high statistics studies of its properties would be possible. The $R = 0.06\%$ $\tan\beta = 2$ curve illustrates the large loss in observability that occurs for non-optimal resolution when the h^0 is SM-like at large m_{A^0} and has a very small width.

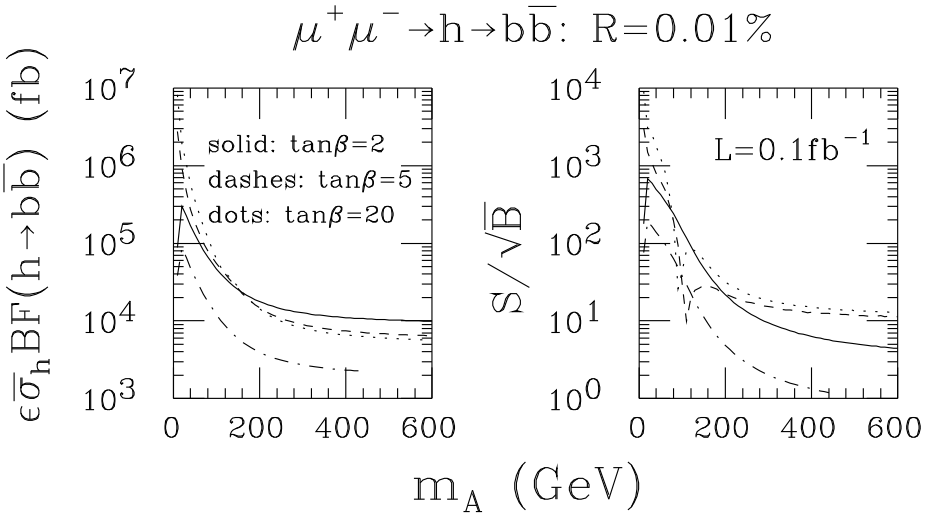


Figure 2.34: Plot of $\epsilon\overline{\sigma}_{h^0} BF(h^0 \to b\bar{b})$ vs m_{A^0} for $\tan\beta = 2$, 5 and 20. Also shown is the corresponding $S/\sqrt{B}$ for $L = 0.1$ fb^{-1}. We have taken $R = 0.01\%$, $\epsilon = 0.5$, $m_t = 175$ GeV, and included two-loop/RGE-improved radiative corrections to Higgs masses, mixing angles and self-couplings using $m_{\tilde{t}} = 1$ TeV and neglecting squark mixing. SUSY decays are assumed to be absent in computing BF. Also shown as the dot-dashed curve are the $R = 0.06\%$ results at $\tan\beta = 2$ in the $b\bar{b}$ channel.

Results for $\epsilon\overline{\sigma}_h BF(h \to b\bar{b}, t\bar{t})$ for $h = H^0$ and $h = A^0$ are displayed in Figs. 2.3.3 and 2.3.3, respectively, along with the corresponding $L = 0.1$ fb^{-1} $S/\sqrt{B}$ values. For a luminosity

of $L = 0.01$ fb^{-1}, the $S/\sqrt{B}$ values of the figures should be reduced by a factor of 0.32. For $L = 0.3$, multiply by 1.7. This range of luminosities will be that which arises when we consider searching for the H^0 and A^0 by scanning in $\sqrt{s}$. The dot-dashed curves illustrate the fact that $R = 0.06\%$ resolution does not cause a large loss in observability relative to $R = 0.01\%$ in the case of the A^0 and, especially, the H^0; the largest effect is for the $\tan\beta = 2$ case in the $b\bar{b}$ channel. For $\tan\beta = 5$ and 20, and for all $t\bar{t}$ curves, the results for $R = 0.06\%$ are virtually indistinguishable from those for $R = 0.01\%$.

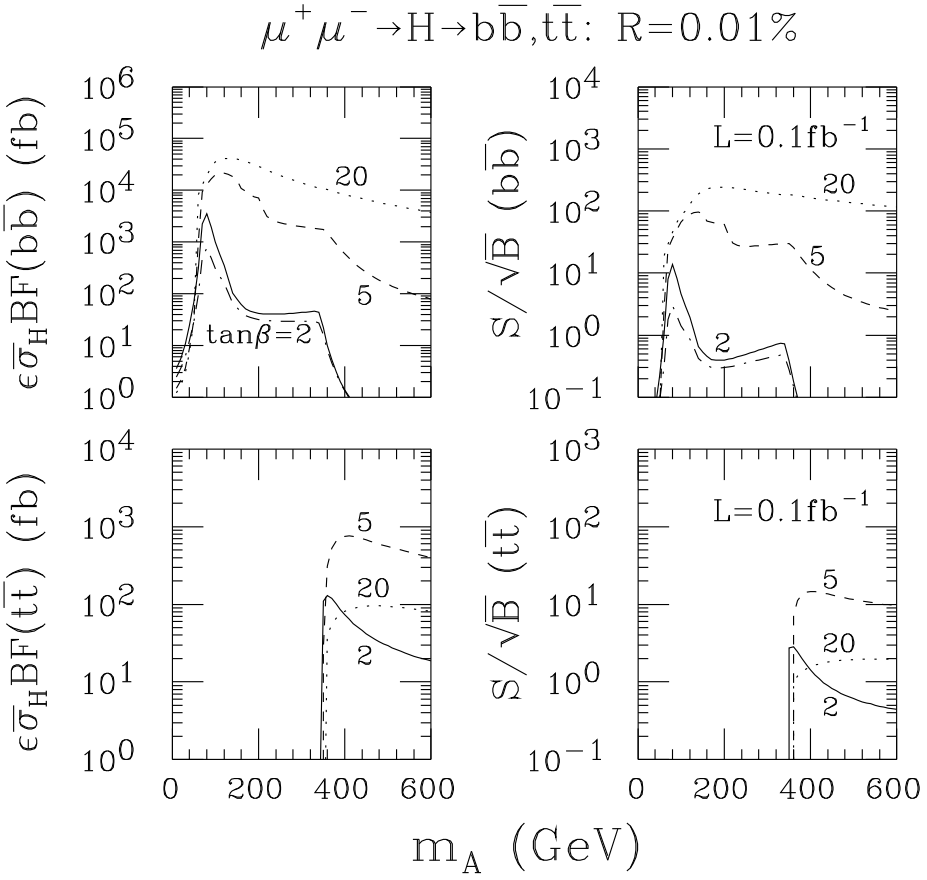


Figure 2.35: Plot of $\epsilon\overline{\sigma}_{H^0}BF(H^0 \to b\bar{b}, t\bar{t})$ vs m_{A^0} for $\tan\beta = 2$, 5 and 20. Also shown are the corresponding $S/\sqrt{B}$ values for $L = 0.1$ fb^{-1}. The inputs are specified in the caption of Fig. 2.3.3. Also shown as the dot-dashed curve are the $R = 0.06\%$ results at $\tan\beta = 2$ in the $b\bar{b}$ channel.

An alternative picture that is especially useful for assessing the parameter space region over which h^0, A^0 and/or H^0 discovery will be possible at the $\mu^+\mu^-$ collider is that given

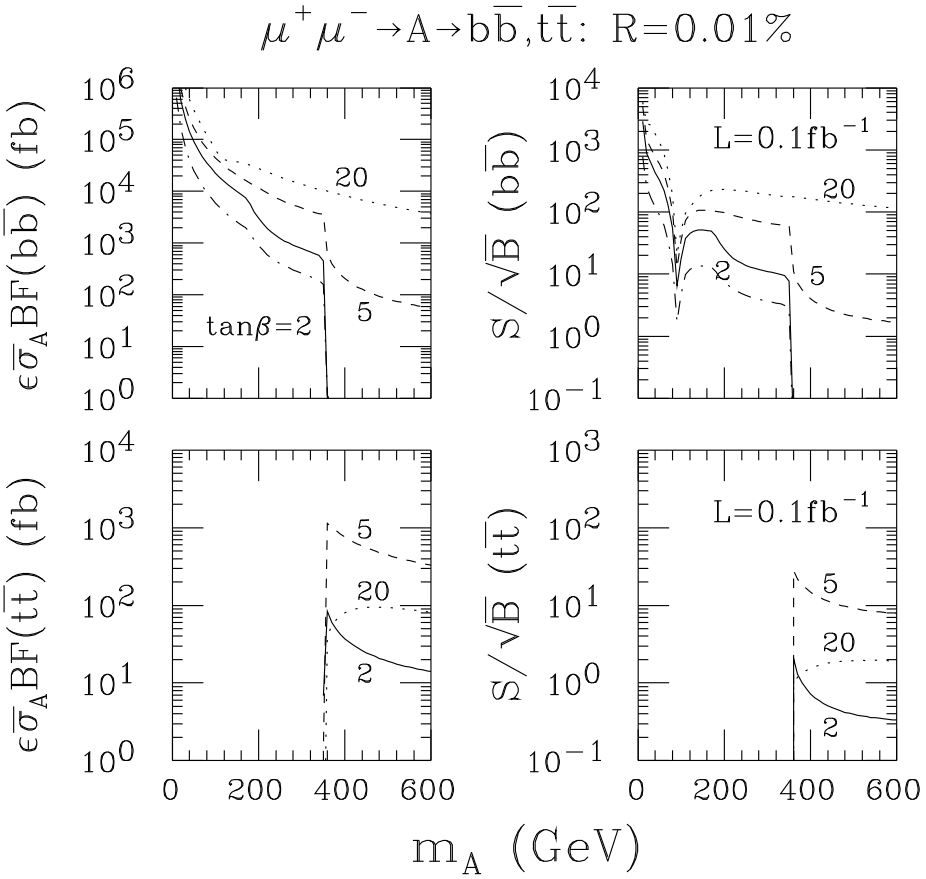


Figure 2.36: Plot of $\epsilon\overline{\sigma}_{A^0} BF(A^0 \to b\bar{b}, t\bar{t})$ vs m_{A^0} for $\tan\beta = 2$, 5 and 20. Also shown are the corresponding $S/\sqrt{B}$ values for $L = 0.1$ fb^{-1}. The inputs are specified in the caption of Fig. 2.3.3. Also shown as the dot-dashed curve are the $R = 0.06\%$ results at $\tan\beta = 2$ in the $b\bar{b}$ channel.

in Fig. 2.3.3, for which we have taken $R = 0.06\%$. The contours in $(m_{A^0}, \tan\beta)$ parameter space denote the luminosity required for a 5σ signal when $\sqrt{s}$ is taken equal to the Higgs mass in question. For the window labelled $H^0 \to b\bar{b}$ we take $\sqrt{s} = m_{H^0}$, for the $h^0 \to b\bar{b}$ window we take $\sqrt{s} = m_{h^0}$, while $\sqrt{s} = m_{A^0}$ for the $A^0 \to b\bar{b}$ and $A^0 \to t\bar{t}$ contours. The 5σ contours are for luminosities of $L = 0.001$, 0.01, 0.1, 1, and 10 fb^{-1}. The larger the L the larger the discovery region. In the case of $A^0 \to t\bar{t}$, 5σ is only achieved for the four luminosities $L = 0.01, 0.1, 1, 10$ fb^{-1}. In the case of the h^0, $L = 10$ fb^{-1} always yields a 5σ signal within the parameter space region shown.

With regard to the h^0, Fig. 2.3.3 shows that for $R = 0.06\%$ and luminosities somewhat

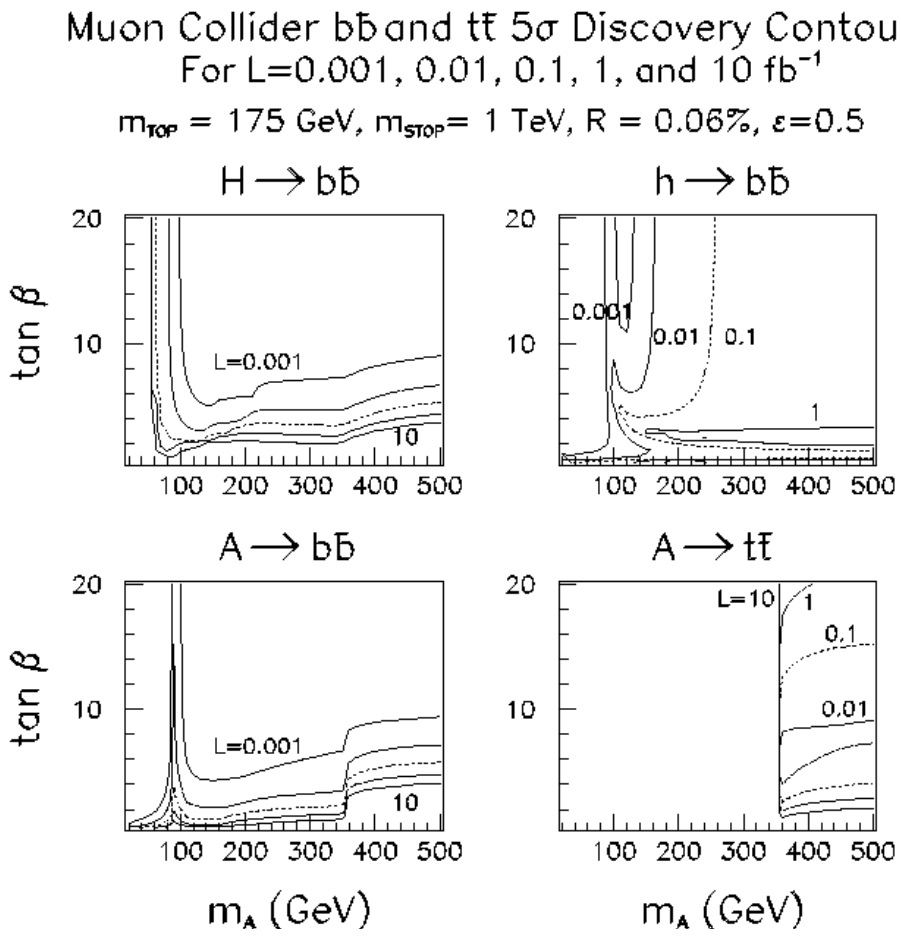


Figure 2.37: Contours in $(m_{A^0}, \tan\beta)$ parameter space of the luminosity required for 5σ Higgs signals. Contours for $L = 0.001$, 0.01, 0.1, 1, and 10 fb^{-1} are given. For $A^0 \to t\bar{t}$, $L = 0.001$ fb^{-1} does not yield a 5σ signal and no corresponding contour appears. For $h^0 \to b\bar{b}$, $L = 10$ fb^{-1} yields a 5σ signal for all of parameter space, and so only $L = 0.001 - 1$ fb^{-1} contours appear. The inputs are specified in the caption of Fig. 2.3.3.

less than 1 fb^{-1}, h^0 could only be detected in the $b\bar{b}$ mode at large m_{A^0} if $\tan\beta$ is sufficiently far from 1 that m_{h^0} is not near m_Z. In contrast, when m_{A^0} is sufficiently small that m_{h^0} is small and the h^0 is no longer SM-like, and has enhanced $\mu\mu$ and $b\bar{b}$ couplings, rather modest luminosity is required for a 5σ signal at $\sqrt{s} = m_{h^0}$; for instance, $L \lesssim 0.001$ fb^{-1} will allow detection of a signal from the h^0 (and the possibly overlapping A^0) over most of the $m_{A^0} \lesssim 100$ GeV portion of parameter space even for $R = 0.06\%$. However, we have noted

that it is theoretically quite likely that m_{A^0} is large and that the h^0 is SM-like. Detection of the H^0 and A^0 then becomes of paramount interest.

Detecting the H^0 and A^0 by scanning in $\sqrt{s}$

In order to discover the H^0 or A^0 in the $\gtrsim 250$ GeV region, we must scan over $\sqrt{s}$ values between 250 GeV and 500 GeV (the presumed upper limit for the FMC). The separation between scan points is determined by the larger of the expected widths and the $\sqrt{s}$ resolution, $\sigma_{\sqrt{s}}$. If $\tan\beta \gtrsim 2$, then for m_{H^0} and m_{A^0} near 250 GeV, the A^0 and H^0 widths are of order $0.05 - 0.1$ GeV. For masses near 500 GeV, their widths are at least 1 GeV (cf. Fig. 2.3.3). Meanwhile, for $R = 0.01\%$ ($R = 0.06\%$), $\sigma_{\sqrt{s}}$ ranges from ~ 0.018 GeV (~ 0.11 GeV) to ~ 0.035 GeV ($\sim .21$ GeV) as $\sqrt{s}$ ranges from 250 GeV to 500 GeV. Thus, it is reasonable to imagine using scan points separated by 0.1 GeV for $m_{A^0} \sim m_{H^0}$ near 250 GeV, rising to 1 GeV by $\sqrt{s} = 500$ GeV. It will also be important to note that the luminosity required per point for detection of the A^0 and H^0 is less for masses below $2m_t$ than above. In assessing the detectability of the H^0 and A^0 by scanning we devote

- $L = 0.01$ fb^{-1} to each of 1000 points separated by 0.1 GeV between 250 and 350 GeV,
- $L = 0.1$ fb^{-1} to each of 100 points separated by 0.5 GeV between 350 and 400 GeV,
- and $L = 0.3$ fb^{-1} to each of 100 points separated by 1 GeV between 400 and 500 GeV.

This selection of points more or less ensures that if the H^0 and A^0 are present then one of the scan points would have $\sqrt{s} \sim m_{H^0}, m_{A^0}$ within either the $\sigma_{\sqrt{s}}$ resolution or the Higgs width. The total luminosity required for this scan would be 50 fb^{-1}.

We now employ the 5σ contours of Fig. 2.3.3 to assess the portion of $(m_{A^0}, \tan\beta)$ parameter space over which the above scan will allow us to detect the H^0 and A^0 in the $b\bar{b}$ and $t\bar{t}$ channels. The 5σ luminosity contours of interest will be the curves corresponding to $L = 0.01$ fb^{-1}, $L = 0.1$ fb^{-1} and $L = 1$ fb^{-1}. The 5σ contour for $L = 0.3$ fb^{-1} luminosity per point, as employed in our scan procedure from 400 to 500 GeV, is midway between these last two curves. Fig. 2.3.3 shows that, by performing the scan in the manner outlined earlier, one can detect the H^0, A^0 in the $b\bar{b}$ mode for all $\tan\beta$ values above about $2 - 4$ for $m_{H^0}, m_{A^0} \lesssim 2m_t$ and above about $3 - 5$ for $2m_t \lesssim m_{H^0}, m_{A^0} \lesssim 500$ GeV. Meanwhile, in the $t\bar{t}$ mode, the $A^0 \to t\bar{t}$ signal can be seen for $m_{A^0} \gtrsim 2m_t$ provided $\tan\beta \gtrsim 3$. Together, the $b\bar{b}$ and $t\bar{t}$ signals are viable for a remarkably large portion of parameter space, which

includes, in particular, essentially all of the wedge region where the LHC lacks sensitivity (see Fig. 2.3.3). At worst, there would be a very small $\tan\beta$ window for $m_{A^0} \gtrsim 2m_t$ between $\tan\beta = 3$ and $\tan\beta = 4$, for which the signal might be missed during the above described scan and also no signal seen at the LHC. In practice, it might be desirable to simply devote several years of running to the scan in order to ensure that the A^0 and H^0 are detected if present.

The implementation of the above scan is very demanding upon the machine design because:

- several rings may be needed to have high luminosities over a broad range of $\sqrt{s}$;
- it must be possible *over this broad range of energies* to quickly (for example, once every hour or so in the 250–350 GeV range) reset $\sqrt{s}$ with an accuracy that is a small fraction of the proposed step sizes.

It is too early to say if these demands can both be met.

Finally, we note the obvious conflict between this scan and the desirable $\sqrt{s} = m_{h^0}$, $L = 50$ fb^{-1} study of the SM-like h^0. A multi-year program will be required to accomplish both tasks.

Non-$b\bar{b}$ final state modes for heavy Higgs detection

The reader may note that $\sqrt{s} = m_{H^0}$ does not yield an observable s-channel signal in the $b\bar{b}$ mode for $m_{A^0} \lesssim 100$ GeV. Although the H^0 is SM-like in this parameter region in that it does not have enhanced coupling to $\mu\mu$ and $b\bar{b}$, its decays are dominated by h^0h^0 and, for $m_{A^0} \lesssim 60$ GeV, A^0A^0 pairs; ZA^0 decays also enter for small enough m_{A^0}. This means that the H^0 total width is quite large, in particular much larger than the $\sqrt{s}$ spread. The large total width also implies that $BF(H^0 \to \mu\mu)$ is small. Equation (2.59) then shows that the production rate for the H^0 will be small, and that the rate in the $b\bar{b}$ final state will be further suppressed by the small value of $BF(H^0 \to b\bar{b})$. The only possible channels for observation of the H^0 in the $m_{A^0} \lesssim 100$ GeV region are h^0h^0, A^0A^0, ZA^0. As we discuss below, these could prove to be viable.

The full set of channels to be considered are

$$H^0 \to h^0h^0, \qquad H^0 \to A^0A^0, \qquad H^0 \to ZA^0, \qquad A^0 \to Zh^0. \tag{2.72}$$

The h^0h^0, A^0A^0 final states primarily ($\sim 80\%$ of the time) yield $4b$'s. The ZA^0, Zh^0 final states yield $2j2b$ about 60% of the time. In either case, we can demand that there be two pairs of jets, each pair falling within narrow mass intervals. In addition, two b-tags can be required. Thus, these channels will have small background. To illustrate the size of the signal in these channels, we present in Fig. 2.3.3 the $L = 10$ fb^{-1} signal rates for the above four modes, assuming a net 50% efficiency (including branching fractions and tagging efficiencies, as well as double mass-binning). In the $H^0 \to h^0h^0$ case, at least 50 events are obtained in essentially all but the $m_{A^0} = 60 - 230, \tan\beta \gtrsim 2.5$ region; the 5000 event contour is confined to a narrow region around $m_{A^0} = 65 - 70, \tan\beta \gtrsim 2$ and to the (disjoint) teardrop region labelled; the 50 and 500 event contours are as labelled. At least 500 events are predicted in the $m_{A^0} \lesssim 60$ region for all $\tan\beta$. In the $H^0 \to A^0A^0$ case, at least 500 events are obtained in the $m_{A^0} \lesssim 60$ and $\tan\beta \gtrsim 2$ region. In the $H^0 \to ZA^0$ case, only the 5 event level is achieved over even the small piece of parameter space shown. Finally, in the $A^0 \to Zh^0$ case all contours are easily identified by the labeling. No events are expected for m_{A^0} below about 200 GeV, where the $A^0 \to Zh^0$ decay mode is no longer kinematically allowed. It is kinematics that also dictates the rather restricted regions at low m_{A^0} for which $H^0 \to A^0A^0$ and $H^0 \to ZA^0$ events occur.

In order to discuss the observability of the above signals, we need to compute the background level, which we do not do in this report. After b-tagging and mass reconstruction we believe that backgrounds should be modest. In the absence of any explicit calculation we can only make the following guesstimates. Based on the event rates of Fig. 2.3.3 it should be possible to study the $H^0 \to h^0h^0$ channel over a significant fraction of parameter space with $L \sim 1$ fb^{-1}. In particular, luminosities at and above this level could open up the $m_{A^0} \lesssim 60$ GeV region for both this mode and the $H^0 \to A^0A^0$ mode. In contrast, it will obviously require very substantial luminosity to detect $H^0 \to ZA^0$, even when not kinematically suppressed. A viable $A^0 \to Zh^0$ signal may be possible, when kinematically allowed, only so long as m_{A^0} and $\tan\beta$ are not large; when m_{A^0} is large the tree-level coupling is suppressed (which suppression occurs most rapidly at large $\tan\beta$) and there are too few events for a useful signal.

Although these modes provide somewhat more challenging signals than the $b\bar{b}$ channel signal, their observation would provide tests of important Higgs couplings. In particular, detection of the $H^0 \to h^0h^0$ and $H^0 \to A^0A^0$ modes would allow a direct probe of these very interesting Higgs boson self-couplings. The procedure will be outlined in a later section. In

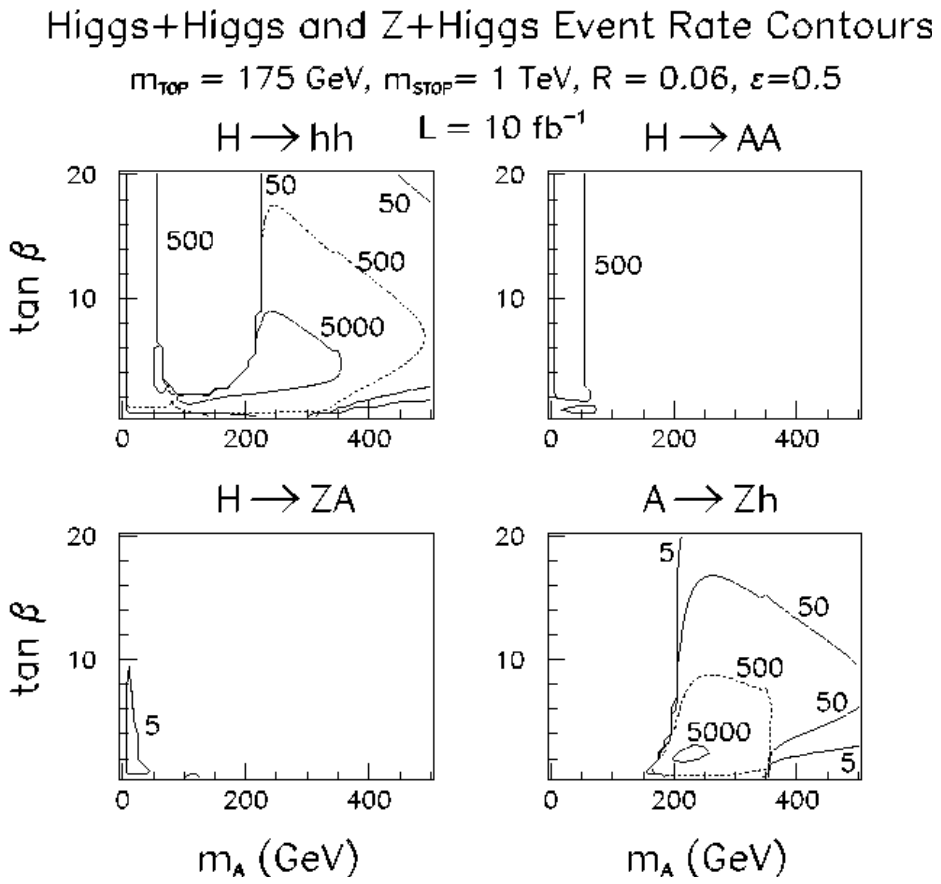


Figure 2.38: Event rate contours for $H^0 \to h^0 h^0$, $H^0 \to A^0 A^0$, $H^0 \to Z A^0$ and $A^0 \to Z h^0$ in $(m_{A^0}, \tan\beta)$ parameter space for integrated luminosity $L = 10$ fb^{-1}. Contours for 5, 50, 500 and 5000 events are shown in the first and last cases. There are 500 or more $H^0 \to A^0 A^0$ events if $m_{A^0} \lesssim 60$ GeV and $\tan\beta \gtrsim 2$, but $H^0 \to Z A^0$ barely reaches the 5 event level. Two-loop/RGE-improved radiative corrections to Higgs masses, mixing angles and self-couplings are included, taking $m_t = 175$ GeV, $m_{\tilde{t}} = 1$ TeV and neglecting squark mixing.

general, determination of the Higgs boson self-couplings is quite difficult at other machines. In particular, even when a relevant branching fraction can be measured, knowledge of the total width is required in order to extract the partial width and coupling. Without a $\mu^+\mu^-$ collider, measurement of the total width is only possible if the width is substantially larger than the resolution implied by final state mass reconstruction at the Higgs mass. This is not

the case for the H^0 and A^0 unless $\tan\beta$ is very large.

MSSM Higgs Boson Detection Using the Bremsstrahlung Tail Spectrum

In this section, we discuss an alternative way of searching for the A^0 and H^0 by running the $\mu^+\mu^-$ collider at full energy but looking for excess events arising from the luminosity on the low-energy end of the bremsstrahlung tail (see sec. 2.12.3). This latter technique proves to be somewhat competitive with the scan technique just described, provided that excellent resolution in reconstructing the $b\overline{b}$ final state mass can be achieved and provided that large total integrated luminosity is devoted to such running. It would have two distinct advantages over the scanning approach.

- It would not require the construction of multiple rings in order to maintain high luminosity over a broad range of $\sqrt{s}$ collision energies.
- A large number of events in the Zh mode for the SM-like h^0 could be simultaneously accumulated.

As for the scan procedure, the bremsstrahlung tail technique is viable only if the $h \to \mu^+\mu^-$ coupling is significantly enhanced relative to the SM $h_{SM} \to \mu^+\mu^-$ coupling; only then is a Higgs boson with mass substantially below $\sqrt{s}$ produced at a large rate by virtue of the bremsstrahlung tail. Of course, once the H^0 and/or A^0 is found using the bremsstrahlung technique, it would then be highly desirable to run the machine with $\sqrt{s} \sim m_{H^0}, m_{A^0}$ in order to study in detail the widths and other properties of the H^0, A^0.

For our study of the bremsstrahlung tail possibility, we shall assume that the $b\bar{b}$ final state mass can be reconstructed to within ± 5 GeV. A full study of this mode of detection should generate events, smear the b jets using expected resolutions, allow for semi-leptonic b decays, and incorporate tagging efficiencies. The reconstructed mass of the $b\overline{b}$ final state for each event should then be binned and one would then look for a peak over the expected background level. We will not perform this detailed simulation here. Instead, we compute as a function of $m_{b\overline{b}}$ (the central value of the $b\overline{b}$ final state mass) the number of events in the interval $[m_{b\overline{b}} - 5 \text{ GeV}, m_{b\overline{b}} + 5 \text{ GeV}]$. In estimating the significance of any peak seen in the spectrum, we will choose $m_{b\overline{b}}$ at the center of the peak, and compare the excess of events in the above interval (the signal S) to the number of events expected if there is no Higgs boson present (the background B). The statistical significance will be computed as $S/\sqrt{B}$.

In computing the number of events we assume an integrated luminosity of $L = 50$ fb^{-1} and assume an event reconstruction and tagging efficiency of $\epsilon = 0.5$. Correspondingly, only the continuum $b\bar{b}$ final states from $\gamma^\star, Z^\star$ processes will be included in B (using also $\epsilon = 0.5$). These latter assumptions are the same ones employed in our other analysis.

Mass peaks

It will be useful to first display some typical mass peaks. In Fig. 2.3.3, we plot the number of events in the interval $[m_{b\bar{b}}-5 \text{ GeV}, m_{b\bar{b}}+5 \text{ GeV}]$ as a function of $m_{b\bar{b}}$ for three m_{A^0} choices: $m_{A^0} = 120$, 300 and 480 GeV. In each case, results for $\tan\beta = 5$ and 20 are shown. The event enhancements derive from the presence of the H^0 and A^0 Higgs bosons. There would be no visible effect for the choice of $m_{A^0} = 100$ GeV for any $\tan\beta$ value below 20. This is because all the Higgs masses are sitting on the very large Z peak and, in addition, none of the $\mu^+\mu^-$ couplings are fully enhanced. For the three m_{A^0} values considered in Fig. 2.3.3, we observe event excesses for $\tan\beta = 20$ in all cases. For $\tan\beta = 5$, the $m_{A^0} = 300$ GeV peak is clear, while $m_{A^0} = 480$ GeV yields a shoulder of excess events (that is statistically significant); nothing is visible for $m_{A^0} = 120$ GeV. For $\tan\beta \lesssim 2$, no peaks or excesses would be visible for any of the above m_{A^0} choices. Finally, we note that enhancements due to the h^0 resonance would not be visible, regardless of $\tan\beta$, for $m_{A^0} \gtrsim 100$ GeV.

Significance of signals

We will now proceed to survey the $S/\sqrt{B}$ expectations. We do this as a function of location in the $(m_{A^0}, \tan\beta)$ parameter space as follows. For each choice of $(m_{A^0}, \tan\beta)$ we determine m_{h^0} and m_{H^0}. We then compute $S/\sqrt{B}$ for the three locations $m_{b\bar{b}} = m_{h^0}$, $m_{b\bar{b}} = m_{H^0}$ and $m_{b\bar{b}} = m_{A^0}$, where S and B are computed by counting events in the $m_{b\bar{b}} \pm 5$ GeV window. Effects from overlapping Higgs resonances are included. The 5σ discovery contours for each of these three window locations are plotted in $(m_{A^0}, \tan\beta)$ parameter space for integrated luminosities of $L = 0.5$, 5, 50 and 200 fb^{-1} in Fig. 2.3.3, taking $\sqrt{s} = 500$ GeV and $R = 0.1\%$.

As expected from Fig. 2.3.3, the window centered at $m_{b\bar{b}} = m_{h^0}$ only yields a statistically significant excess if $\tan\beta$ is large and m_{h^0} is not near m_Z. (m_{h^0} near m_Z at high $\tan\beta$ corresponds to $m_{A^0} \sim 95$ GeV.) Since the Zh^0 mode will yield an observable signal regardless of the $(m_{A^0}, \tan\beta)$ values, the bremsstrahlung tail excess would mainly be of interest as a probe of the $\Gamma(h^0 \to \mu^+\mu^-)$ partial width prior to running at $\sqrt{s} = m_{h^0}$.

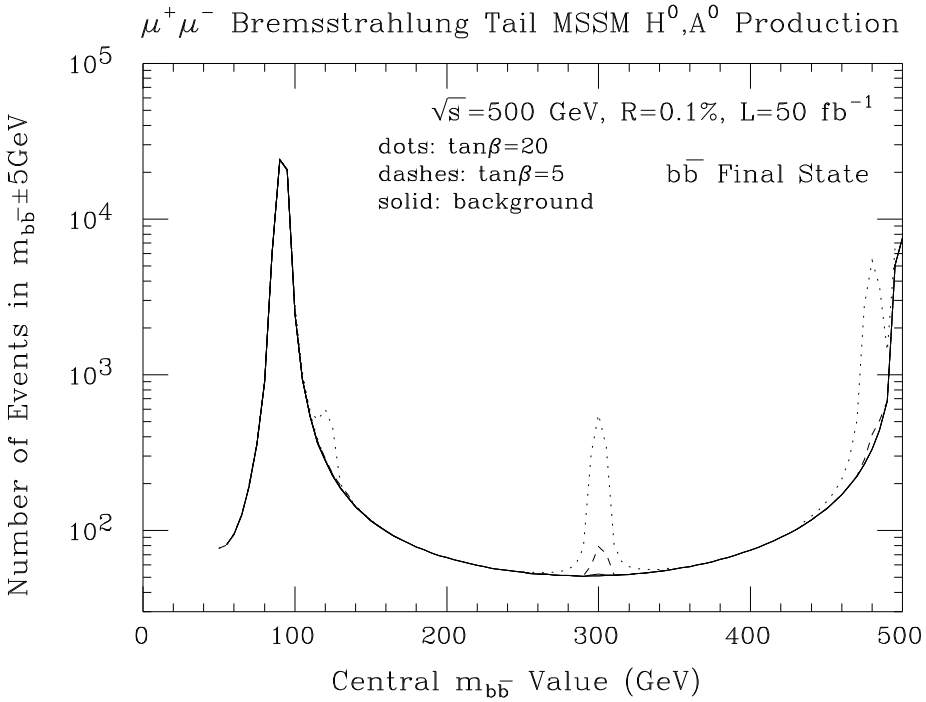


Figure 2.39: Taking $\sqrt{s} = 500$ GeV, integrated luminosity $L = 50$ fb^{-1}, and $R = 0.1\%$, we consider the $b\bar{b}$ final state and plot the number of events in the interval $[m_{b\bar{b}} - 5 \text{ GeV}, m_{b\bar{b}} + 5 \text{ GeV}]$, as a function of the location of the central $m_{b\bar{b}}$ value, resulting from the low $\sqrt{\hat{s}}$ bremsstrahlung tail of the luminosity distribution. MSSM Higgs boson H^0 and A^0 resonances are present for the parameter choices of $m_{A^0} = 120$, 300 and 480 GeV, with $\tan\beta = 5$ and 20 in each case. Enhancements for $m_{A^0} = 120$, 300 and 480 GeV are visible for $\tan\beta = 20$; $\tan\beta = 5$ yields visible enhancements only for $m_{A^0} = 300$ and 480 GeV. Two-loop/RGE-improved radiative corrections are included, taking $m_t = 175$ GeV, $m_{\tilde{t}} = 1$ TeV and neglecting squark mixing. SUSY decay channels are assumed to be absent.

However, the ± 5 GeV intervals centered at $m_{b\bar{b}} = m_{H^0}$ and $m_{b\bar{b}} = m_{A^0}$ (which, include events from the overlapping A^0 and H^0 resonances, respectively) yield 5σ statistical signals for a substantial portion of parameter space if L is large. With $L = 50$ fb^{-1}, a 5 sigma discovery of the H^0 and A^0 using the $\sqrt{s} = 500$ GeV bremsstrahlung tail is viable down to $\tan\beta \gtrsim 6.5$ at $m_{A^0} = 250$ GeV improving to $\tan\beta \gtrsim 5$ at 480 GeV. This is not quite as far down in $\tan\beta$ as can be probed for $250 \lesssim m_{A^0} \lesssim 500$ GeV by the previously described scan

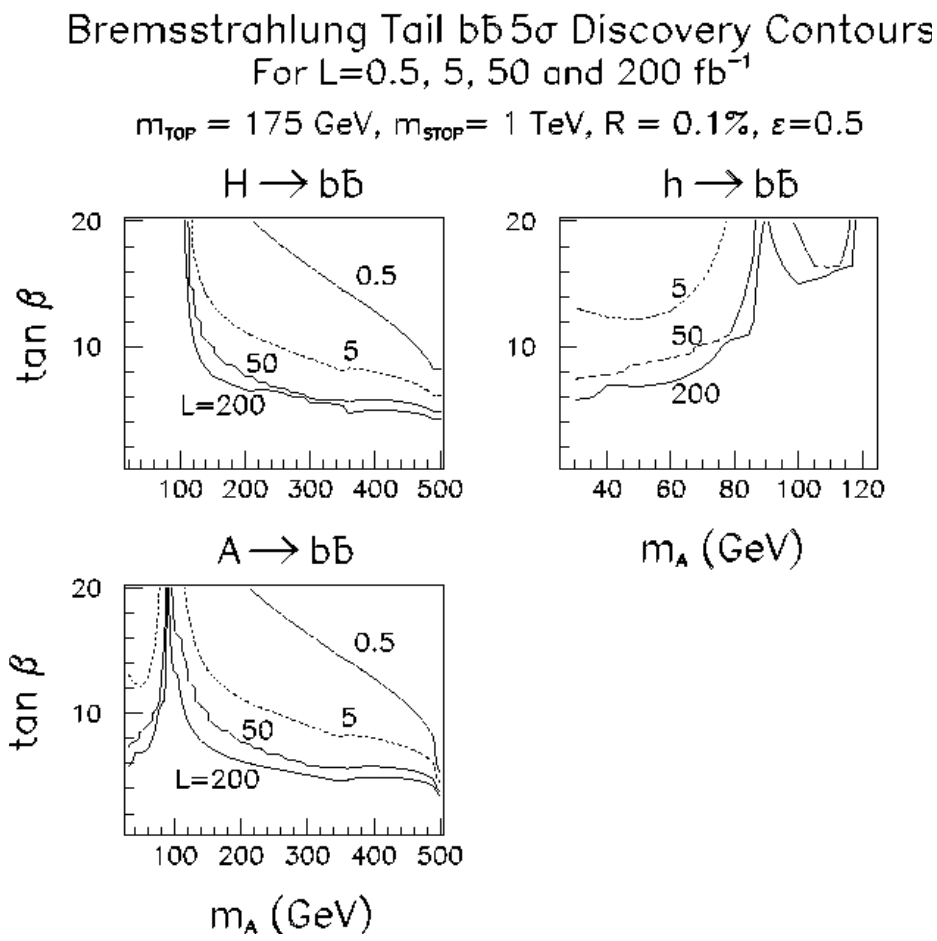


Figure 2.40: Taking $\sqrt{s} = 500$ GeV and $R = 0.1\%$, we consider the $b\bar{b}$ final state and compute the Higgs signal (S) and background (B) rates in the mass interval $[m_{b\bar{b}} - 5 \text{ GeV}, m_{b\bar{b}} + 5 \text{ GeV}]$, with $m_{b\bar{b}} = m_{H^0}$, $m_{b\bar{b}} = m_{h^0}$, and $m_{b\bar{b}} = m_{A^0}$, resulting from the low $\sqrt{\hat{s}}$ bremsstrahlung tail of the luminosity distribution. $S/\sqrt{B} = 5$ contours are shown for integrated luminosities of $L = 0.5$, 5, 50, and 200 fb^{-1}. Two-loop/RGE-improved radiative corrections are included, taking $m_t = 175$ GeV, $m_{\tilde{t}} = 1$ TeV and neglecting squark mixing. SUSY decay channels are assumed to be absent.

over a series of $\sqrt{s}$ values using $0.01 - 0.3$ fb^{-1} of luminosity at each scan point. As m_{H^0}, m_{A^0} move closer to m_Z, the 5σ discovery contours move to much larger $\tan\beta$ values, whereas the scanning technique would yield 5σ signals for $\tan\beta$ values as low as $\tan\beta \sim 3 - 4$ all the way down to $m_{A^0} \gtrsim 60$ GeV.

Strategy: scan vs. maximum energy

If $Z^\star \to H^0A^0$ is not observed at a $\sqrt{s} = 500$ GeV e^+e^- machine and if discovery of the H^0 and A^0 in the $250 - 500$ GeV mass range is the primary goal, at the $\mu^+\mu^-$ collider it would be a close call as to whether it would be better to immediately embark on the $\sqrt{s}$ scan or accumulate luminosity at the maximum machine energy. The $\sqrt{s}$ scan probes $\tan\beta$ values that are lower by only 1 or 2 units than the bremsstrahlung tail search. This statement assumes that a final state mass resolution of order ± 5 GeV can be achieved (even after including all semi-leptonic decay effects and so forth) in the $b\overline{b}$ final state for the latter search. If not, the $\sqrt{s}$ scan is the preferred technique. Thus, resolution and missing energy could become critical issues for the detector(s) in deciding the best approach.

If an e^+e^- collider is not operational at the time a $\mu^+\mu^-$ collider begins running, then the decision as to which approach to choose for H^0 and A^0 discovery becomes even more delicate unless the LHC has clearly ruled out $m_{A^0}, m_{H^0} \lesssim 250$ GeV (which it probably can do — see Fig. 2.3.3). Without a lower bound on m_{A^0}, m_{H^0}, the $\sqrt{s}$ scan would have to be extended to lower $\sqrt{s}$, requiring more luminosity. In contrast, by accumulating $L = 50$ fb^{-1} at full energy, $\sqrt{s} = 500$ GeV, it would be possible to simultaneously either discover or rule out $m_{A^0}, m_{H^0} \lesssim \sqrt{s}/2$ for all $\tan\beta$ and $\sqrt{s}/2 \lesssim m_{H^0}, m_{A^0}, \lesssim \sqrt{s}$ for $\tan\beta \gtrsim 5 - 7$. Note that $m_{A^0}, m_{H^0} \lesssim \sqrt{s}/2 - 20$ GeV can be ruled out in the $Z^\star \to H^0h$ mode with perhaps as little as $5 - 10$ fb^{-1}. For luminosities of order 10 fb^{-1} the bremsstrahlung tail technique would probe $\tan\beta \gtrsim 11$ for $m_{A^0} \sim 250$ GeV improving to $\tan\beta \gtrsim 6$ for $m_{A^0} \sim 500$ GeV. After accumulating the $L = 5 - 10$ fb^{-1}, the $\mu^+\mu^-$ collider could then be switched to the scan mode of operation if no signal has been found.

Detailed Studies of the $\boldsymbol{H^0}$ and $\boldsymbol{A^0}$

However the H^0 and A^0 are first detected, one will wish to measure the total and partial widths of the H^0 and A^0. Once again, the $\mu^+\mu^-$ collider can play a crucial role. We will not give detailed estimates of what can be accomplished, but rather confine ourselves to outlining the procedures and strategies. The time scale and available luminosity for implementing these procedures depends dramatically upon whether or not one must first discover the H^0 and A^0 by scanning or in the bremsstrahlung tail (either of which would require a luminosity expenditure of $L \sim 50$ fb^{-1}), as opposed to observing them at the LHC (typically possible for $\tan\beta \lesssim 3 - 4$ at high m_{A^0}) or at an e^+e^- collider (requiring $m_{A^0}, m_{H^0} \lesssim \sqrt{s}/2$).

One might presume that once a Higgs boson with $\Gamma_h^{\rm tot}$ larger than the rms $\sqrt{s}$ spread is

discovered, direct measurement of the Higgs width would be quite straightforward with a simple scan over several $\sqrt{s}$ settings. This is indeed the case unless there is a second nearby Higgs boson. As it happens, the A^0 and H^0 are sufficiently degenerate in some regions of parameter space (large m_{A^0} and large $\tan\beta$), see Figs. 2.3.3 and 2.3.3, that a measurement of the widths of the A^0 and H^0 separately will require sorting out two overlapping resonance bumps, which, in turn, necessitates an appropriate scan. Two sample possibilities were illustrated earlier in Fig. 2.3.3, where the H^0 and A^0 resonance bumps that would appear as a function of $\sqrt{s}$ are illustrated for $m_{A^0} = 350$ GeV in the cases $\tan\beta = 5$ and 10. As noted earlier, separation of the peaks and precision width measurements are both much easier if we have excellent beam energy resolution; we assume $R = 0.01\%$. At $\tan\beta = 5$, we estimate that by accumulating roughly 0.01 fb^{-1} at each of 3 appropriately placed $\sqrt{s}$ choices near the center and on either side of each of the two separated peaks, the widths of the H^0 and A^0 could be measured to about 33%; 10% width determination would require about 0.1 fb^{-1} per point. At the higher $\tan\beta = 10$ value, one would clearly have to accumulate data in the dip between the overlapping peaks, near both peaks, below the double peak and above the double peak, and perform a fit to the two Higgs resonances simultaneously. A minimum of 5 data points would be required. Again, roughly 0.01 fb^{-1} per point would be needed to determine $\Gamma^{\rm tot}_{H^0}$ and $\Gamma^{\rm tot}_{A^0}$ to the 33% level, or 0.1 fb^{-1} per point for a 10% determination. Very large $\tan\beta$ values yield the worst scenarios since the H^0 and A^0 peaks are, then, simultaneously broad and very degenerate. Determination of the individual widths would become extremely difficult.

The production rate in a given channel is proportional to $BF(h \to \mu^+\mu^-)BF(h \to X)$ (for $\sigma_{\sqrt{s}} \ll \Gamma^{\rm tot}_h$), see Eq. (2.59). We then proceed as follows:

- $BF(h \to \mu^+\mu^-)$ and $BF(h \to b\overline{b})$ can be obtained individually if we use the type-II doublet prejudice that the $\mu^+\mu^-$ and $b\overline{b}$ couplings squared are modified relative to the SM coupling by the same factor, f. (A value of m_b must be specified.)

- Given the individual branching fractions, the partial widths can then be computed:

$$\Gamma(h \to \mu^+\mu^-, b\overline{b}) = \Gamma^{\rm tot}_h BF(h \to \mu^+\mu^-, b\overline{b}) \tag{2.73}$$

- One can use event rates in other observable channels, coupled with the $BF(h \to \mu^+\mu^-)$ determination, to obtain results for $BF(h \to X)$.

- $\Gamma_h^{\rm tot} \times BF(h \to X)$ then yields the partial width and coupling for any observable channel X. For example, if the $H^0 \to h^0 h^0$ channel can be detected we could determine the very interesting associated partial width (and, thence, coupling) via $\Gamma(H^0 \to h^0 h^0) = \Gamma_{H^0}^{\rm tot} BF(H^0 \to h^0 h^0)$ or, equivalently,

$$\Gamma(H^0 \to h^0 h^0) = \frac{[\Gamma_{H^0}^{\rm tot}]^2 BF(H^0 \to \mu\mu) BF(H^0 \to h^0 h^0)}{\Gamma(H^0 \to \mu\mu)} \,. \tag{2.74}$$

Of course, if $\Gamma_h^{\rm tot}$ and $\sigma_{\sqrt{s}}$ are close in size, one must avoid the approximation of Eq. (2.59), but determination of f and the partial widths and branching fractions would nevertheless be straightforward.

Determining a Higgs Boson's CP Properties

A $\mu^+\mu^-$ collider might well prove to be the best machine for directly probing the CP properties of a Higgs boson that can be produced and detected in the s-channel mode. This issue has been explored in Refs. [72, 73] in the case of a general two-Higgs-doublet model.

The first possibility is to measure correlations in the $\tau^+\tau^-$ or $t\bar{t}$ final states. Via such measurements, a $\mu^+\mu^-$ collider is likely to have greater sensitivity to the Higgs boson CP properties for $L = 20$ fb^{-1} than will the e^+e^- collider for $L = 85$ fb^{-1} (using correlation measurements in the Zh production mode) if $\tan\beta \gtrsim 10$ or $2m_W \lesssim m_h \lesssim 2m_t$. Indeed, there is a tendency for the $\mu^+\mu^-$ CP-sensitivity to be best precisely for parameter choices such that CP-sensitivity in the $e^+e^- \to Zh$ mode is worst. Somewhat higher total luminosity ($L \sim 50$ fb^{-1}) is generally needed in order to use these correlations to distinguish a pure CP-odd state from a pure CP-even state.

The second possibility arises if it is possible to transversely polarize the muon beams. Assume that we can have 100% transverse polarization and that the μ^+ transverse polarization is rotated with respect to the μ^- transverse polarization by an angle ϕ. The production cross section for a h with coupling $a + ib\gamma_5$ then behaves as

$$\sigma(\phi) \propto 1 - \frac{a^2 - b^2}{a^2 + b^2} \cos\phi + \frac{2ab}{a^2 + b^2} \sin\phi \,. \tag{2.75}$$

To prove that the h is a CP admixture, use the asymmetry

$$A_1 \equiv \frac{\sigma(\pi/2) - \sigma(-\pi/2)}{\sigma(\pi/2) + \sigma(-\pi/2)} = \frac{2ab}{a^2 + b^2} \,. \tag{2.76}$$

For a pure CP eigenstate, either a or b is zero. For example, in the MSSM the Higgs sector is CP-conserving; $b = 0$ for the CP-even h^0 and H^0, while $a = 0$ for the CP-odd A^0. In such cases, it is necessary to employ a different asymmetry than that discussed in Ref. [73]. The quantity

$$A_2 \equiv \frac{\sigma(\pi) - \sigma(-\pi)}{\sigma(\pi) + \sigma(-\pi)} = \frac{a^2 - b^2}{a^2 + b^2} \tag{2.77}$$

is $+1$ or -1 for a CP-even or CP-odd h, respectively. Background processes in the final states where a Higgs boson can be most easily observed (*e.g.* $b\overline{b}$) can dilute these asymmetries substantially. Whether or not they will prove useful depends even more upon the very uncertain ability to transversely polarize the muon beams, especially while maintaining high luminosity.

Note that longitudinally polarized beams are not useful for studying the CP properties of a Higgs produced in the s-channel. Regardless of the values of a and b in the h coupling, the cross section is simply proportional to $1 - \lambda_{\mu^+}\lambda_{\mu^-}$ (the λ's being the helicities), and is only non-zero for LR or RL transitions, up to corrections of order m_μ^2/m_h^2.

2.3.4 Summary and Conclusion

A $\mu^+\mu^-$ collider would be a remarkably powerful machine for probing Higgs physics using direct s-channel production, and thus ultimately for finding the underlying theory of the scalar sector. In this report we have concentrated on the procedures and machine requirements for direct measurement of the properties of a Higgs boson.

SM-like Higgs Boson

We expect that a SM-like h (which nominally includes the h^0 of the MSSM) will first be detected either at the LHC or in the Zh mode at an e^+e^- collider. If not, it would be most advantageous to expend a small amount of luminosity at full machine energy to discover it in the Zh mode at the $\mu^+\mu^-$ collider. Once m_h is approximately known, a $\mu^+\mu^-$ collider can zero-in on $\sqrt{s} \simeq m_h$ for detailed studies of a SM-like Higgs boson provided $m_h \lesssim 2m_W$ (as is the case for the h^0 of the MSSM). The mass can be measured to a fraction of an MeV for $m_{h_{SM}} \lesssim 130$ GeV.

Crucial to a model-independent determination of all the properties of the Higgs boson at the $\mu^+\mu^-$ collider is the ability to make a direct precision measurement of its total width,

which is very narrow for a SM-like h when $m_h < 2m_W$. The proposed method (described in sec. 2.12.3) relies on measuring the ratio of the central peak cross section to the cross section on the wings of the peak, a ratio that is determined by $\Gamma_h^{\rm tot}$ alone. Once $\Gamma_h^{\rm tot}$ is measured, determinations of the crucial $\mu^+\mu^-$ and $b\overline{b}$ couplings are possible. The precision for $\Gamma_h^{\rm tot}$ and the $\mu^+\mu^-$ and $b\overline{b}$ partial widths/couplings achieved for total integrated luminosity of $L = 50$ fb^{-1} and an excellent beam resolution of $R = 0.01\%$ would be sufficient to distinguish the MSSM h^0 from the SM h_{SM} at the 3σ statistical level for values of the parameter m_{A^0} as large as ~ 400 GeV provided that m_{h^0} is not in the range $80 \lesssim m_{h^0} \lesssim 100$ GeV (*i.e.* near m_Z). No other accelerator or combination of accelerators has the potential of seeing the h^0 vs. h_{SM} differences at this level of precision out to such large m_{A^0} values. For a SM-like Higgs with $m_h \gtrsim 200$ GeV, the event rate is too low for detection in the s-channel.

Machine requirements for the precision studies are:

- High luminosity $\mathcal{L} \gtrsim 2 \times 10^{33}{\rm cm}^{-2}{\rm s}^{-1}$ at $\sqrt{s} \sim m_h$.
- Excellent beam energy resolution of $R = 0.01\%$.
- Ability to adjust the machine energy $\sqrt{s}$ accurately (to one part in a million) and quickly (once an hour in the initial scan to precisely determine m_h) over a $\sqrt{s}$ interval of several GeV.

Non-SM-like Higgs Bosons

For other Higgs bosons with weak WW, ZZ couplings (such as the H^0 and A^0 of the MSSM), but enhanced $\mu^+\mu^-$ and $b\overline{b}$ couplings, discovery in s-channel collisions at the $\mu^+\mu^-$ collider is typically possible. There are three possible techniques. In order to compare these techniques it is reasonable to suppose that the H^0 and A^0 have been excluded for $m_{H^0}, m_{A^0} \lesssim \sqrt{s}/2$ via the $Z^\star \to H^0A^0$ mode at an e^+e^- collider running with $\sqrt{s} \sim 500$ GeV.

a) Scan method
In this approach, a scan for the H^0 and A^0 of the MSSM would be made over a sequence of $\sqrt{s}$ values all the way out to the maximal $\sqrt{s}$ value achievable at the $\mu^+\mu^-$ collider. Assuming that $L = 50$ fb^{-1} is devoted to the scan and that both the e^+e^- and the $\mu^+\mu^-$ colliders have maximal energies of order 500 GeV, discovery via the scan would be robust for $250 \lesssim m_{H^0,A^0} \lesssim 500$ GeV if $\tan\beta \gtrsim 3$ to 4. Fortuitously, the domain

$250 \lesssim m_{H^0}, m_{A^0} \lesssim 500$ GeV, $\tan\beta \lesssim 3$, in which much more luminosity would clearly be required for discovery at the $\mu^+\mu^-$ collider, is a parameter region where the H^0 and A^0 are likely to be accessible at the LHC for accumulated luminosity of 300 fb^{-1} per detector (ATLAS+CMS), as illustrated in Fig. 2.3.3. There is, nonetheless, a small window, $3 \lesssim \tan\beta \lesssim 4$, at large m_{A^0} (between about 400 and 500 GeV) for which the LHC and the $\mu^+\mu^-$ collider might both miss seeing the H^0 and A^0 unless higher luminosities are accumulated.

In order that the required $L = 50$ fb^{-1} can be optimally distributed over the full $250 - 500$ GeV scan range in the course of a year or two of running, it would be necessary to design the storage ring or rings so that it would be possible to adjust $\sqrt{s}$ quickly and accurately (to within a small fraction of the step size, which must be $\lesssim 0.1$ GeV in some mass ranges) while maintaining the full luminosity.

b) Bremsstrahlung tail method

In this technique, the A^0 and H^0 search is made while running the $\mu^+\mu^-$ collider at full energy, looking for excess events arising from the luminosity at the low-energy end of the bremsstrahlung tail. This approach is competitive with the scan technique if the $b\overline{b}$ final state mass can be reconstructed with excellent resolution (roughly ± 5 GeV, including all detector effects and semi-leptonic b decays). The lower $\tan\beta$ limits for 5σ signals are about one to two units higher than for the scan technique in the $m_{A^0} = 250 - 480$ GeV range. Thus the bremsstrahlung search leaves a larger gap between the upper limit in $\tan\beta$ for which H^0, A^0 discovery would be possible at the LHC ($\tan\beta \lesssim 3 - 4$ at high m_{A^0}) and the lower limit for which the H^0, A^0 would be detected at the $\mu^+\mu^-$ collider ($\tan\beta \gtrsim 5 - 7$) than would the scan technique.

The bremsstrahlung technique has the advantage of not requiring that high luminosity be maintained over a broad range of $\sqrt{s}$ collision energies while being able to step quickly and accurately in $\sqrt{s}$, but detector costs associated with the very demanding resolution in the $b\overline{b}$ invariant mass might be high.

c) Pair production

It may well be possible to build a $\mu^+\mu^-$ collider with $\sqrt{s}$ substantially above 500 GeV. If a $\sqrt{s} \geq 1$ TeV machine with high luminosity were built instead of a 500 GeV collider, it could discover the H^0, A^0 for $m_{H^0}, m_{A^0} \geq 500$ GeV in the pair production mode.

If the H^0, A^0 have already been discovered, either

- with $m_{H^0}, m_{A^0} \lesssim 250$ GeV in the $Z^\star \to H^0 A^0$ mode at an e^+e^- collider, or
- with $m_{H^0}, m_{A^0} \lesssim 2$ TeV in the $Z^\star \to H^0 A^0$ mode at a 4 TeV $\mu^+\mu^-$ collider, or
- with $m_{H^0}, m_{A^0} \lesssim 500$ GeV at the LHC (if $\tan\beta \lesssim 3-4$ or $\tan\beta \gtrsim 8-20$),

scanning over a broad energy range would not be necessary at the $\mu^+\mu^-$ collider. By constructing a single appropriate storage ring and devoting full luminosity to accumulating events at $\sqrt{s} \simeq m_{A^0}, m_{H^0}$, detailed studies of the total widths and partial widths of the A^0 and H^0 would then be possible at the $\mu^+\mu^-$ collider *for all* $\tan\beta$ *values above 1.*

Summary of Machine and Detector Requirements

We re-emphasize the crucial machine and detector characteristics for detection and study of both SM-like Higgs bosons and non-SM-like Higgs bosons.

- High luminosity, $\mathcal{L} \gtrsim 2 \times 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, is required at any $\sqrt{s}$ where a Higgs boson is known to exist and throughout any range of energy over which we must scan to detect a Higgs boson.
- A machine design such that beamstrahlung is small compared to the effects of bremsstrahlung (included in our studies) is highly desirable for scan searches and precision studies. However, significant beamstrahlung might improve the ability to discover Higgs bosons using the low-energy tail of the luminosity spectrum.
- An extremely precise beam energy, $R \sim 0.01\%$, will be needed for precision studies of a narrow-width SM-like Higgs boson. Such precise resolution is also extremely helpful in the zeroing-in scan for a very narrow SM-like and is not harmful for discovering a Higgs boson with broad width. Precision measurements of the non-SM-like H^0 and A^0 widths and separation of these two resonances when they overlap becomes difficult if R is substantially larger than 0.01%.
- To zero-in on $\sqrt{s} \simeq m_h$ for a narrow-width SM-like Higgs boson requires being able to rapidly set $\sqrt{s}$ with an accuracy that is small compared to the beam resolution R, for $\sqrt{s}$ values within about a few GeV of the (approximately known) value of m_h. To

discover the H^0 and A^0 by scanning requires being able to rapidly set $\sqrt{s}$ with an accuracy that is small compared to their widths over a $\sqrt{s}$ interval of order several hundred GeV.

- To measure $\Gamma_h^{\rm tot}$ for a SM-like h to $\pm 10\%$, it must be possible to set $\sqrt{s}$ with an accuracy of order 1 part in 10^6 over $\sqrt{s}$ values in an interval several times Rm_h, *i.e.* over an interval of tens of MeV. This (and the accuracy for the mass measurements) requires a machine design that allows quick spin rotation measurements of a polarized muon in the storage ring.

- If both muon beams can be polarized and the polarization (P) maintained through the cooling and acceleration process, the significance of the s-channel Higgs signal can be significantly enhanced provided the factor by which the luminosity is decreased is less than $(1+P^2)/(1-P^2)$.

- To detect non-SM-like Higgs bosons with enhanced $\mu^+\mu^-$ couplings in the bremsstrahlung luminosity tail when the machine is run at full energy, one needs excellent mass resolution ($\sim \pm 5$ GeV) in the $b\overline{b}$ final state mass as reconstructed in the detector.

In conclusion, if a Higgs bosons is discovered at the LHC and/or an e^+e^- collider, construction of a $\mu^+\mu^-$ collider with $\sqrt{s}$ covering the range of masses observed will become almost mandatory purely on the basis of s-channel Higgs physics. There are many other motivations for building a $\mu^+\mu^-$ collider, especially one with $\sqrt{s} \gtrsim 2$ TeV, based on other types of new physics that could be probed. The physics motivations for a high-energy $\mu^+\mu^-$ collider will be treated elsewhere [74].

Bibliography

[1] *Proceedings of the First Workshop on the Physics Potential and Development of $\mu^+\mu^-$ Colliders*, Napa, California (1992), Nucl. Instru. and Meth. **A350**, 24 (1994).

[2] *Proceedings of the Second Workshop on the Physics Potential and Development of $\mu^+\mu^-$ Colliders*, Sausalito, California (1994), ed. by D. Cline, American Institute of Physics Conference Proceedings 352.

[3] *Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop*, Ed. J. C. Gallardo, AIP Conference Proceedings 372, 1996.

[4] *Proceedings of the Symposium on Physics Potential and Development of $\mu^+\mu^-$ Colliders*, San Francisco, CA (1995); Supplement to Nuclear Physics B. .

[5] R.B. Palmer and A. Tollestrup, unpublished report.

[6] D.V. Neuffer, Ref. [2], p. 22.

[7] D.V. Neuffer and R.B. Palmer, Ref. [2], p. 70.

[8] V. Barger, M.S. Berger, K. Fujii, J.F. Gunion, T. Han, C. Heusch, W. Hong, S.K. Oh, Z. Parsa, S. Rajpoot, R. Thun and B. Willis, *Physics Goals of a $\mu^+\mu^-$ Collider*, appearing in Ref. [2], p. 55, hep-ph/9503258.

[9] V. Barger, M. Berger, J.F. Gunion, T. Han, and R. Phillips, in preparation.

[10] G.P. Jackson and D. Neuffer, private communications.

[11] P. Janot, *Proceedings of the 2nd International Workshop on "Physics and Experiments with Linear e^+e^- Colliders",* eds. F. Harris, S. Olsen, S. Pakvasa and X. Tata, Waikoloa,

HI (1993), World Scientific Publishing, p. 192, and references therein; T. Barklow and D. Burke, private communication.

[12] See "JLC-I", KEK-92-16, December 1992.

[13] K. Kawagoe, *Proceedings of the 2nd International Workshop on "Physics and Experiments with Linear e^+e^- Colliders"*, Eds. F. Harris, S. Olsen, S. Pakvasa and X. Tata, Waikoloa, HI (1993), World Scientific Publishing, p. 660.

[14] J.F. Gunion, A. Stange, and S. Willenbrock, preprint UCD-95-28 (1995), hep-ph/9602238, to be published in *Electroweak Physics and Beyond the Standard Model*, World Scientific Publishing Co., eds. T. Barklow, S. Dawson, H. Haber, and J. Siegrist.

[15] V. Barger, M. Berger, J.F. Gunion, and T. Han, Phys. Rev. Lett. **75**, 1462 (1995).

[16] V. Barger, M. Berger, J.F. Gunion, and T. Han, UCD-96-6, hep-ph/9602415.

[17] R.B. Palmer, private communication.

[18] H. Haber, R. Hempfling and A. Hoang, CERN-TH/95-216.

[19] M. Carena, J.R. Espinosa, M. Quiros and C.E.M. Wagner, Phys. Lett. **B355**, 209 (1995); J.A. Casas, J.R. Espinosa, M. Quiros and A. Riotto, Nucl. Phys. **B436**, 3 (1995).

[20] J.F. Gunion and H.E. Haber, Phys. Rev. **D48**, 5109 (1993).

[21] See, for example, R. Arnowitt and P. Nath, Phys. Rev. **69**, 725 (1992); Phys. Lett. **B289**, 368 (1992); G. G. Ross and R. G. Roberts, Nucl. Phys. **B377**, 571 (1992); S. Kelley, J. L. Lopez, D. V. Nanopoulos, H. Pois, and K. Yuan, Nucl. Phys. **B398**, 3 (1993); M. Drees and M. M. Nojiri, Phys. Rev. **D47**, 376 (1993); M. Olechowski and S. Pokorski, Nucl. Phys. **B404**, 590 (1993); D. J. Castaño, E. J. Piard, and P. Ramond, Phys. Rev. **D49**, 4882 (1994); V. Barger, M.S. Berger, and P. Ohmann, Phys. Rev. **D49**, 4908 (1994); M. Carena, M. Olechowski, S. Pokorski, and C.E.M. Wagner, Nucl. Phys. **B419**, 213 (1994); G. Kane, C. Kolda, and J. Wells, Phys. Rev. **D49**, 6173 (1994); W. de Boer, R. Ehret, and D.I. Kazakov, Z. Phys. **C67**, 647 (1995); M. Carena and C.E.M. Wagner, Nucl. Phys. **B452**, 45 (1995); S.F. King and P. White, Phys. Rev. **D52**, 4183 (1995).

[22] Z. Parsa (unpublished).

[23] S. Dawson, appearing in Ref. [3], hep-ph/9512260.

[24] For discussions of the $t\bar{t}$ threshold behavior at e^+e^- colliders, see V.S. Fadin and V.A. Khoze, JETP Lett. **46** 525 (1987); Sov. J. Nucl. Phys. **48** 309 (1988); M. Peskin and M. Strassler, Phys. Rev. **D43**, 1500 (1991); G. Bagliesi, *et al.*, CERN Orange Book Report CERN-PPE/92-05; Y. Sumino, *et al.*, Phys. Rev. **D47**, 56 (1992); M. Jezabek, J.H. Kühn, T. Teubner, Z. Phys. **C56**, 653 (1992).

[25] M.S. Berger, talk presented at the *Workshop on Particle Theory and Phenomenology: Physics of the Top Quark*, Iowa State University, May 25–26, 1995, hep-ph/9508209.

[26] P. Igo-Kemenes, M. Martinez, R. Miquel and S. Orteu, CERN-PPE/93-200, Contribution to the *Workshop on Physics with Linear e^+e^- Colliders at 500 GeV*, K. Fujii, T. Matsui, and Y. Sumino, Phys. Rev. **D50**, 4341 (1994).

[27] C.P. Yuan, talk presented at *Int. Workshop on Elementary Particle Physics: Present and Future*, Valencia, Spain, June 5–9, 1995, hep-ph/9509209.

[28] B. Grzadkowski and J.F. Gunion, Phys. Lett. **B350**, 218 (1995).

[29] D. Atwood and A. Soni, Phys. Rev. **D52**, 6271 (1995).

[30] A. Pilaftsis, Rutherford preprint RAL-TR/96-021.

[31] D. Atwood, L Reina and A. Soni, Phys. Rev. Lett. **75**, 3800 (1995).

[32] J.F. Gunion, preprint UCD-95-36, hep-ph/9510350, to be published in the Proceedings of the Santa Cruz e^-e^- Workshop, Santa Cruz, CA, Sept. 4-5, 1995.

[33] J.F. Gunion and J. Kelly, Phys. Rev. **D51**, 2101 (1995); and in preparation.

[34] M.S. Chanowitz and M.K. Gaillard, Nucl. Phys. **B261**, 379 (1985).

[35] D. Burke, *Proceedings of the Symposium on Physics Potential and Development of $\mu^+\mu^-$ Colliders*, San Francisco, CA (1995); Supplement to Nuclear Physics B.

[36] J. Bagger, V. Barger, K. Cheung, J.F. Gunion, T. Han, G.A. Ladinsky, R. Rosenfeld, and C.-P. Yuan, Phys. Rev. **D52**, 2878 (1995).

[37] V. Barger, K. Cheung, T. Han, and R.J.N. Phillips, Phys. Rev. **D52**, 3815 (1995).

[38] The constraints on a fourth generation in the context of supergravity models can be found in J.F. Gunion, D.W. McKay, and H. Pois, Phys. Rev. **D53**, 1616 (1996).

[39] For a recent survey, see V. Barger, M.S. Berger, and R.J.N. Phillips, Phys. Rev. **D52**, 1663 (1995).

[40] J.L. Hewett and T.G. Rizzo, Phys. Rep. **183**, 193 (1989).

[41] J.F. Gunion, A. Stange, and S. Willenbrock, Weakly Coupled Higgs Bosons, preprint UCD-95-28, to appear in *Electroweak Symmetry Breaking and New Physics at the TeV Scale*, to be published by World Scientific.

[42] See J.F. Gunion, H.E. Haber, G.L. Kane and S. Dawson, *The Higgs Hunters Guide*, Addison-Wesley Publishing, and references therein.

[43] M. Drees, Int. J. Mod. Phys. **A4**, 3635 (1989); J. Ellis *et al.*, Phys. Rev. **D39**, 844 (1989); L. Durand and J.L. Lopez, Phys. Lett. **B217**, 463 (1989); J.R. Espinosa and M. Quirós, Phys. Lett. **B279**, 92 (1992); P. Binétruy and C.A. Savoy, Phys. Lett. **B277**, 453 (1992); T. Morori and Y. Okada, Phys. Lett. **B295**, 73 (1992); G. Kane, *et al.*, Phys. Rev. Lett. **70**, 2686 (1993); J.R. Espinosa and M. Quirós, Phys. Lett. **B302**, 271 (1993); U. Ellwanger, Phys. Lett. **B303**, 271 (1993); J. Kamoshita, Y. Okada, M. Tanaka *et al.*, Phys. Lett. **B328**, 67 (1994).

[44] V. Barger, *et al.*, Phys. Lett. **B314**, 351 (1993); P. Langacker and N. Polonsky, Phys. Rev. **D50**, 2199 (1994).

[45] H. Haber, R. Hempfling and A. Hoang, **CERN-TH/95-216** (1995).

[46] M. Carena, J.R. Espinosa, M. Quiros and C.E.M. Wagner, Phys. Lett. **B355**, 209 (1995); J.A. Casas, J.R. Espinosa, M. Quiros and A. Riotto, Nucl. Phys. **B436**, 3 (1995).

[47] For a review, see *e. g.*, J.F. Gunion and H.E. Haber, Nucl. Phys. **B272**, 1 (1986).

[48] See, for example, R. Arnowitt and P. Nath, Phys. Rev. **69**, 725 (1992); Phys. Lett. **B289**, 368 (1992); G. G. Ross and R. G. Roberts, Nucl. Phys. **B377**, 571 (1992); S. Kelley, J. L. Lopez, D. V. Nanopoulos, H. Pois, and K. Yuan, Nucl. Phys. **B398**, 3

(1993); M. Drees and M. M. Nojiri, Phys. Rev. **D47**, 376 (1993); M. Olechowski and S. Pokorski, Nucl. Phys. **B404**, 590 (1993); D. J. Castaño, E. J. Piard, and P. Ramond, Phys. Rev. **D49**, 4882 (1994); V. Barger, M.S. Berger, and P. Ohmann, Phys. Rev. **D49**, 4908 (1994); M. Carena, M. Olechowski, S. Pokorski, and C.E.M. Wagner, Nucl. Phys. **B419**, 213 (1994); G. Kane, C. Kolda, and J. Wells, Phys. Rev. **D49**, 6173 (1994); W. de Boer, R. Ehret, and D.I. Kazakov, Z. Phys. **C67**, 647 (1995); M. Carena and C.E.M. Wagner, Nucl. Phys. **B452**, 45 (1995); S.F. King and P. White, Phys. Rev. **D52**, 4183 (1995).

[49] V. Barger, M.S. Berger, R.J.N. Phillips, and A.L. Stange, Phys. Rev. **D45**, 4128 (1992); V. Barger, Kingman Cheung, R.J.N. Phillips, and A.L. Stange, Phys. Rev. **D46**, 4914 (1992).

[50] J. Ellis, J.F. Gunion, H.E. Haber, L. Roszkowski and F. Zwirner, Phys. Rev. **D39**, 844 (1989); B.R. Kim, S.K. Oh and A. Stephan, *Proceedings of the 2nd International Workshop on "Physics and Experiments with Linear e^+e^- Colliders"*, eds. F. Harris, S. Olsen, S. Pakvasa and X. Tata, Waikoloa, HI (1993), World Scientific Publishing, p. 860: J. Kamoshita, Y. Okada and M. Tanaka, Phys. Lett. **B328**, 67 (1994); S.F. King and P.L. White, preprint SHEP-95-27 (1995), hep-ph 9508346; U. Ellwanger, M.R. de Traubenberg and C.A. Savoy, Z. Phys. **C67**, 665 (1995).

[51] *Proceedings of the First Workshop on the Physics Potential and Development of $\mu^+\mu^-$ Colliders*, Napa, California (1992), Nucl. Instru. and Meth. **A350**, 24 (1994).

[52] *Proceedings of the Second Workshop on the Physics Potential and Development of $\mu^+\mu^-$ Colliders*, Sausalito, California (1994), ed. by D. Cline, American Institute of Physics Conference Proceedings 352.

[53] *Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop*, Ed. J. C. Gallardo, AIP Conference Proceedings 372, 1996.

[54] *Proceedings of the Third Workshop on the Physics Potential and Development of $\mu^+\mu^-$ Colliders*, San Francisco, California (1995), ed. by D. Cline, to be published.

[55] V. Barger, M.S. Berger, K. Fujii, J.F. Gunion, T. Han, C. Heusch, W. Hong, S.K. Oh, Z. Parsa, S. Rajpoot, R. Thun and B. Willis, Physics Goals of a $\mu^+\mu^-$ Collider, appearing in Ref. [52], p. 55, hep-ph 9503258.

[56] R.B. Palmer and A. Tollestrup, unpublished report.

[57] D.V. Neuffer, Ref. [52], p. 22.

[58] D.V. Neuffer and R.B. Palmer, Ref. [52], p. 70.

[59] D.J. Miller, Ref. [52], p. 191.

[60] V. Barger, M. Berger, J.F. Gunion, and T. Han, Phys. Rev. Lett. **75**, 1462 (1995).

[61] R.B. Palmer, private communication.

[62] G.P. Jackson and D. Neuffer, private communications.

[63] See, for example, Z. Parsa, $\mu^+\mu^-$ *Collider and Physics Possibilities*, unpublished; K. Hagiwara and D. Zeppenfeld, Nucl. Phys. **B313**, 560 (1989), Appendix B.

[64] For references to in-depth studies of physics at future e^+e^- colliders, see e.g. *Proceedings of the Workshop on Physics and Experiments with Linear Colliders*, ed. F.A. Harris, *et al.*, World Scientific (1993); *Proceedings of the Workshop on Physics and Experiments with Linear* e^+e^- *Colliders*, Waikoloa, Hawaii (April 1993), ed. F. Harris *et al.* (World Scientific, 1993); JLC Group, KEK Report 92-16 (1992); *Proceedings of the Workshop on Physics and Experiments with Linear Colliders*, Saariselkä, Finland (Sept. 1991), ed. R. Orava *et al.*, World Scientific (1992).

[65] P. Janot, *Proceedings of the 2nd International Workshop on "Physics and Experiments with Linear* e^+e^- *Colliders"*, eds. F. Harris, S. Olsen, S. Pakvasa and X. Tata, Waikoloa, HI (1993), World Scientific Publishing, p. 192, and references therein; T. Barklow and D. Burke, private communication.

[66] See "JLC-I", KEK-92-16, December 1992.

[67] K. Kawagoe, *Proceedings of the 2nd International Workshop on "Physics and Experiments with Linear* e^+e^- *Colliders"*, eds. F. Harris, S. Olsen, S. Pakvasa and X. Tata, Waikoloa, HI (1993), World Scientific Publishing, p. 660.

[68] J.F. Gunion and H.E. Haber (unpublished).

[69] Z. Kunszt and F. Zwirner, Nucl. Phys. **B385**, 3 (1992); H. Baer, M. Bisset, C. Kao, and X. Tata, Phys. Rev. **D46**, 1067 (1992); H. Baer, M. Bisset, D. Dicus, C. Kao, and X. Tata, Phys. Rev. **D47**, 1062 (1993); J.F. Gunion *et al.*, Phys. Rev. **D46**, 2040 (1992); J. Gunion and L. Orr, Phys. Rev. **D46**, 2052 (1992); J. Gunion, H. Haber, and C. Kao, Phys. Rev. **D46**, 2907 (1992); V. Barger, K. Cheung, R. Phillips, and A. Stange, Phys. Rev. **D46**, 4914 (1992); J. Dai, J. Gunion, and R. Vega, preprint UCD-95-25 (1995); ATLAS Technical Proposal, CERN/LHCC/94-43, LHCC/P2 (1994); CMS Technical Proposal, CERN/LHCC 94-38, LHCC/P1 (1994); D. Froidevaux, F. Gianotti, and E. Richter-Was, ATLAS Internal Note PHYS-No-64 (1995); F. Gianotti, to appear in the Proceedings of the European Physical Society International Europhysics Conference on High Energy Physics, Brussels, Belgium, July 27 - August 2, 1995.

[70] D. Froidevaux, F. Gianotti, L. Poggioli, E. Richter-Was, D. Cavalli, and S. Resconi, ATLAS Internal Note, PHYS-No-74 (1995).

[71] J.F. Gunion and H.E. Haber, *Proceedings of the 1990 DPF Summer Study on High Energy Physics: "Research Directions for the Decade"*, editor E. Berger, Snowmass (1990), p. 206; Phys. Rev. **D48**, 5109 (1993).

[72] B. Grzadkowski and J.F. Gunion, Phys. Lett. **B350**, 218 (1995).

[73] D. Atwood and A. Soni, Phys. Rev. **D52**, 6271 (1995).

[74] V. Barger, M. Berger, J.F. Gunion, T. Han, and R. Phillips, in preparation.

[75] E.A. Kuraev and V.S. Fadin, Sov. J. Nucl. Phys. **41**, 466 (1985); R.N. Cahn, Phys. Rev. **D36**, 266 (1987); M. Peskin, SLAC Summer Institute: 1989, p. 71.

Contributors

- V. Barger (Univ. of Wisconsin), Editor
- M.S. Berger (Indiana Univ.)
- J.F. Gunion (UC, Davis)
- T. Han (UC, Davis)

List of Figures

List of Tables

Chapter 3

PROTON SOURCE

Contents

3.1 Requirements

The proton driver requirements are determined by the design luminosity of the collider, and the efficiencies of muon collection, cooling, transport and acceleration. These numbers are shown in Table 3.1. In addition to accelerating a large charge, the machines must operate at high repetition rates, which are determined by the μ lifetime at high energies and the overall power minimization in the accelerator systems. The rms bunch length for protons on target has been set at 1 ns to: 1) minimize the initial longitudinal emittance of μ's entering the cooling system, and 2) optimize the separation of the populations of + and

– polarizations off the target, see Section 11.2.1, and Figure 11.7. Since the collection of polarized μ's is inefficient, we assume the proton driver must eventually provide an additional factor of approximately two to compensate for the inefficiency in producing these beams. An additional requirement is that the proton driver system must have low losses, to permit inexpensive maintenance of components.

Table 3.1: Proton Driver Requirements

	30 GeV	**10 GeV**	
Rep. Rate	15	30	Hz
Protons	10^{14}	10^{14}	/pulse
Bunches	4	2	at target
Protons	2.5×10^{13}	5×10^{13}	/bunch

The proton driver needs to deliver very narrow, high intensity proton bunches for the pion production target. The main requirements for the driver were listed in Tb. 1 in the Introduction. Note that this amounts to 7 MW of beam power in the proton beam. This level of beam power is much higher than what is currently available at proton accelerators. However, many detailed proposals have been worked out for multi-GeV hadron facilities or neutron spallation sources that can achieve similar levels of beam power. For our requirements, the designs for the KAON facility, which is a 30 GeV, 3 MW machine and a combination of the BNL 5 MW spallation neutron source (SNS), and the ANL 10 GeV SNS design are most appropriately used as the starting point for the design of the muon collider proton driver.

Table 3.2: Linac parameters

	$\mu\mu$ **Collider**	**BNL-SNS**	
Max. Energy	0.6	0.6	GeV
Rep. Rate	15	60	Hz
Trans Emittance [95 %]	2.4	2.4	$\pi\ mm\,mrad$
Tot. Energy Spread	2.4	2.4	MeV

There are many ways of achieving the required beam intensities and power, (see Fig 3.1). Tbs. 3.2, 3.3, 3.4 show one possible set of parameters for a proton driver consisting of a 600 MeV linac, a 3.6 GeV Booster and a 30 GeV Driver. Both the Booster and Driver would operate at a repetition rate of 15 Hz with a total of four bunches with 0.25×10^{14} protons per bunch. The relatively low repetition rate of both Booster and Driver makes it possible to use a metallic vacuum chamber with eddy current correction coils.[1] Both Linac

Table 3.3: Booster parameters

	$\mu\mu$ **Collider**	**BNL-SNS**	
Injection Energy	0.6	0.6	GeV
Max. Energy	3.6	3.6	GeV
Rep. Rate	15	30	Hz
Protons per Pulse	1×10^{14}	1.45×10^{14}	
Number of Bunches	4	2	
Circumference	360	360	m
Trans. Emittance [95 %]	260	260	$\pi\ mm\,mrad$
Inc. Tune Shift Inj	0.25	0.25	
rf Voltage per Turn	400	400	kV
rf Frequency (h=4)	2.62-3.24	1.31-1.62	MHz
Long. Emittance [95 %]	2	4	eVs

Table 3.4: Driver Parameters

	$\mu\mu$ **Collider**	**AGS**	**KAON**	
Injection Energy	3.6	1.5	3	GeV
Max. Energy	30	24	30	GeV
Rep. Rate	15	1	10	Hz
Protons per Pulse	1×10^{14}	0.6×10^{14}	0.6×10^{14}	
Number of Bunches	4	8	225	
Circumference	1080	800	1078	m
Transition Gamma	38	8.8	30 i	
Max. Dispersion	2.3 m	2.2 m	7.4 m	
Trans. Emittance [95%]	260	100	100	π mm mrad
Inc. Tune Shift Inj	.10	.10	.10	
RF Voltage per Turn	4	0.4	2.6	MV
Harmonic Number	12	8	225	
RF Frequency	3.24-3.33	2.77-3.00	60.8-62.5	MHz
Long. Emittance [95 %]	< 4.5	4.5	0.2	eVs

and Booster designs are copied from the BNL SNS design[2] with the only difference being a lower repetition rate (15 Hz instead of 30 Hz), and a lower number of protons per pulse (1×10^{14} instead of 1.45×10^{14}). The Driver design is based on the experience with the AGS and on the Japanese hadron Project (JHP)[3] and KAON[4] Driver design. The Driver

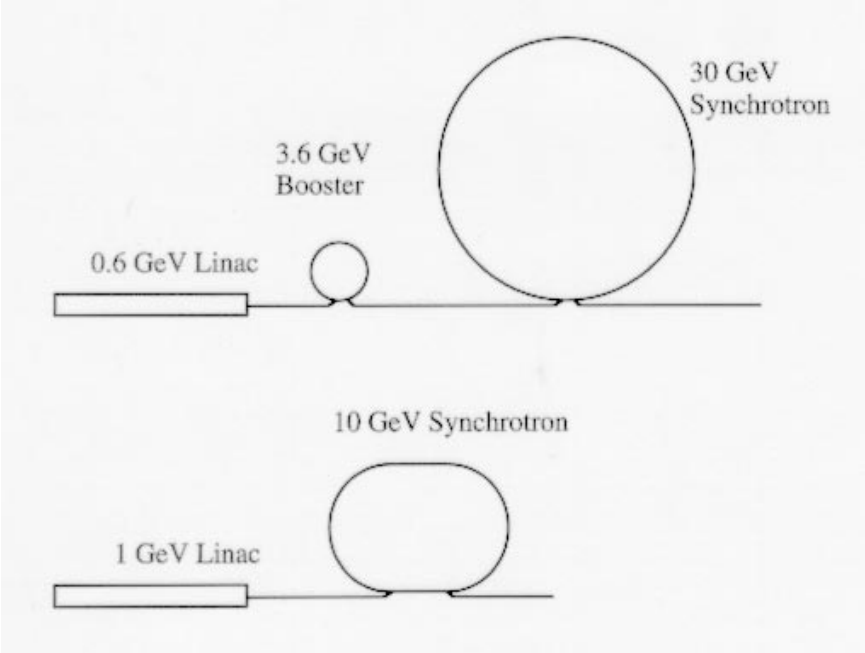


Figure 3.1: Options considered were primarily the simple alternatives of: a) a synchrotron/booster operating at a maximum energy of 30 GeV and, b) a single ring operating at 10 GeV.

lattice is derived from the lattice of the JHP driver using 90 degree FODO cells with missing dipoles in every third FODO cell. Such a lattice allows one to easily achieve a transition energy that is higher than the maximum Driver energy of 30 GeV, or is imaginary, which eliminates the need to cross transition energy but also avoids space charge driven microwave instabilities.

3.2 Production of Short Bunches

There are a number of methods that can be used to generate the short proton pulses required at the pion production target. We list a few that have been considered, either alone or in combination.

1. If the bunches are extracted near transition energy, γ_t, they will have a large $\delta p/p$ and small bunch length. Effects such as longitudinal space charge could be used to compress bunches above transition.

2. RF could be used in a number of ways, for example increasing the voltage will shorten the bunch length, which is proportional to $V^{-1/4}$. Quadrupole modes can be slowly excited in the bunches, and higher frequencies can be used to shorten the overall bucket

length. Bunch rotation should be simpler and faster however. Bunches can be flattened by slowly decreasing the voltage or placing the bunch in the unstable fixed point and then rotated with increased voltage. Bunch rotation can be either a single process, or taken in two steps, with energy shear near γ_t and a separate time shear done further from γ_t.

3. Bunch shortening instabilities, driven by an inductive wall, could be excited by changing the wall impedance, perhaps by unbiasing ferrite.

4. Large numbers of bunches could be coalesced in an internal ring, or low frequency rf linacs, probably induction linacs, can be used to generate a long energy ramp which will coalesce in an external ring or long beam line.

5. Kickers and chicane systems could be used to take a finite number of equal energy proton bunches over different paths to meet at the target.

6. Many short μ bunches could be combined to form a single intense bunch in the μ cooling system.

The simplest option seems to be to extract near γ_t, However it is not clear that the extracted bunches could be made sufficiently short to provide a 1 ns bunch length. Experiments with bunches of 10^{13} protons have been kept circulating for periods of 100 ms with values of $\eta = 0.0005$, with no losses, no negative mass instability, and good agreement with theoretical predictions[5] The negative mass instability was seen in that part of the beam which was above transition, but not in the part of the bucket below transition energy. The stability of unbunched beams near transition has also been studied near transition at the FermiLab antiproton accumulator. [6] Bunching near transition would require that the momentum spread at extraction would be large, however if the final energy of the accelerator is large enough, the fractional momentum spread $\delta p/p$ can, in principle, be accommodated fairly easily at 30 GeV and with some difficulty, at 10 GeV.

Chicane systems can be used to generate a single pulse from a finite number of pulses, however conservation of phase space requires that the bunches cannot be exactly in time and collinear at the exit of such a system. In addition, the total length of chicane beamlines must be roughly $(n - 1/2)$ times the initial maximum separation of bunches.

Bunch Rotation Bunch rotation seems to offer the most reliable procedure for producing short bunches and these methods have been studied in some detail. The longitudinal emittance of the beam, ϵ_L, seems to be more or less independent of injection energy, repetition

rate and other parameters for a variety of accelerators, with $\epsilon_L = 1$ eV-s/10^{13} protons, see Table 3.5. In general, the charge/pulse is more closely related to the transverse admittance, and ϵ_L should not be directly related to factors which determine the maximum current limit of the machine.

Table 3.5: Longitudinal Phase Space

	protons/b	ϵ_L **[eV-s]**	**p/eV-s**
IPNS II	1.0×10^{14}	7.5	1.3×10^{13}
BNL-SNS	7.5×10^{13}	4.0	1.8×10^{13}
ISIS(inj)	2.0×10^{13}	2.0	1.0×10^{13}
FNAL	5.0×10^{12}	2.0	2.5×10^{12}
IPNS	3.0×10^{12}	0.4	7.5×10^{12}
KAON	1.5×10^{12}	0.06	2.5×10^{13}
BNL-Booster	1.4×10^{13}	2.0	9.6×10^{12}

Assuming 2.5×10^{13} protons/bunch, this would imply $\epsilon_L = 2.5$ eV-s at injection. When the beam had reached the extraction energy, we require that the bunch length for 4 σ would be 4 ns, however that would imply a momentum spread of 0.06 at 10 GeV and 0.02 at 30 GeV. Although both these numbers are larger than the momentum admittance of most synchrotrons, the debuncher ring of the antiproton source at FermiLab, which operates at 8 GeV, accepts a $\Delta p/p > 0.05$ and contains this beam for a much longer time than the few turns the short, large momentum spread bunch will circulate in the driver synchrotron.

A common feature of many methods is that the bunch compression is a function of the momentum spread $\delta p/p$ and the momentum dependence of path lengths. The time required for this compression is

$$t_b = \frac{\phi_{rf}}{2\pi\, f_{rf}\, \eta\, \delta p/p}, \tag{3.1}$$

and is proportional to the required rf phase change, ϕ_{rf}, and inversely proportional to the rf frequency, f_{rf}, slip factor, η, and the momentum spread, $\delta p/p$. Because of the large currents involved, it is desirable to bunch as quickly as possible to avoid problems with instabilities. Nevertheless coalescence of bunches spread around the circumference might require on the order of 10 - 20 ms in a typical ring. Thus small compressor rings, which could accommodate large η and $\delta p/p$ could be used with induction linacs which would produce a large and linear spread in the energies from front to back in a bunch, or train of bunches.

The bunching time is a function both of the beam energy γ and the difference between the beam and transition energies, $\gamma_t - \gamma$. Because the machine circumference, η and $\delta p/p$

are dependent on the beam momentum, the bunching time goes like p^n, with the exponent n close to 4, depending on the assumptions used to determine the machine circumference and rf frequency. Thus lower energy rings will have much faster bunching times than high energy rings.

Two methods of bunch rotation have been considered. Decreasing the rf voltage to spread the bunch out in time, followed by rotation, can be done for either a single bunch or a number of smaller bunches using a subharmonic, (see Fig. 3.2)[7]. Since the process is nonlinear for large amplitudes, the primary limitation is the initial phase angle which can be rotated into the required bunch length. For an rf frequency of 3 MHz, the maximum rf phase angle is $\sim 45° - 50°$. An alternative method, which requires control of γ_t, is to flat-top the machine at about 1 unit below transition, where synchrotron rotation is slow, then vertically shear the bunch with the linear part of the rf waveform. This is followed by a horizontal shear, done with the transition energy moved further above the beam energy, so the bunch can rotate quickly to a vertical position, (see Figure 3.3)[8]. Nonlinearities also limit this method, since energy variations within the bunch near transition produce variation in η. Nevertheless it seems possible to compensate some nonlinearities by distorting the bunch shape before bunch rotation.

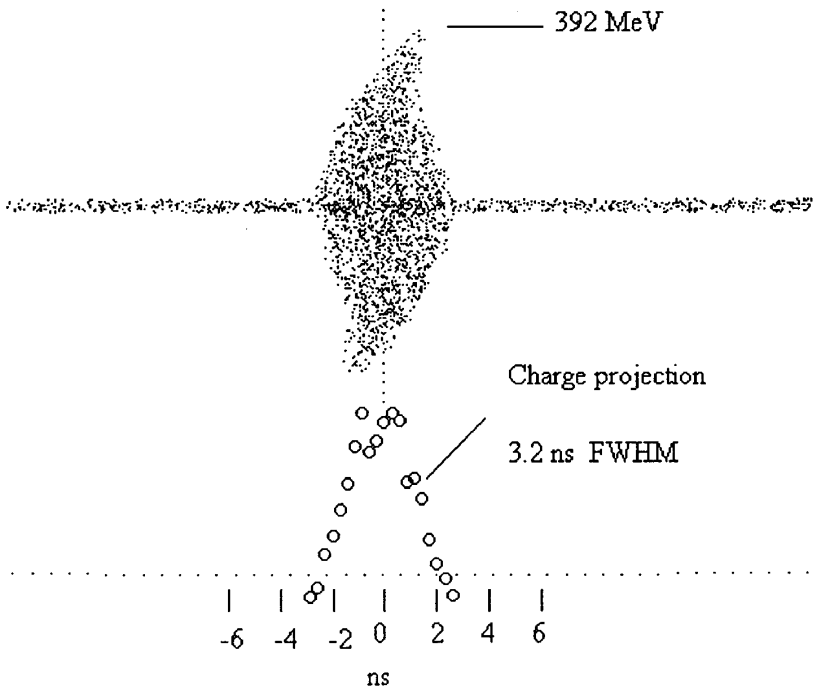


Figure 3.2: Bunch rotation using a 1/4 turn in synchrotron space.

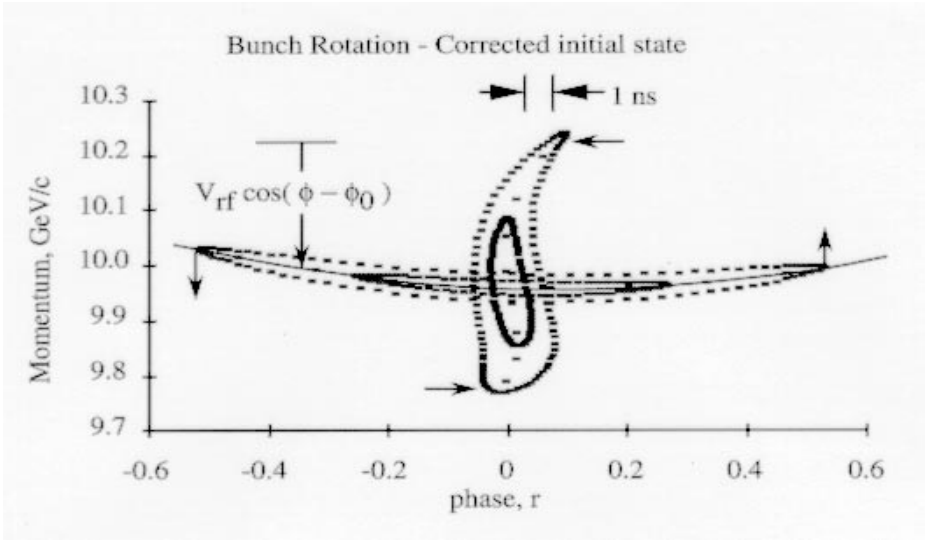


Figure 3.3: Bunch rotation with an energy sheer followed by a rotation in synchrotron space.

Bunch rotation techniques have been demonstrated by Cappi et al. [9] on the CERN PS. In these tests the bunch was rotated by π in longitudinal phase space, giving 2 - 3 times the normal beam current. Since the rotation in longitudinal phase space was by π radians, it was possible to compare the longitudinal emittance before and after the bunch rotation and measure a small, $\sim$1.2, emittance increase. The emittance increase just from bunching alone would presumably be the square root of this value. In order to have control of the transition energy without using high tunes and many quadrupoles and dipoles, we have considered a version of the Flexible Momentum Compaction lattice, suggested by Lee, Ng and Trbojevic.[10] While this lattice can be used to produce imaginary γ_t's, it seems most useful when tuned to produce γ_t values several GeV above the extraction energy. A benefit of this tune seems to be that matching to the zero dispersion straight sections is simple, since the dispersion is naturally close to zero at the ends of the periods. The lattice is also fairly efficient, as it can accommodate a large number of dipoles.

3.3 Stability During Acceleration

Beam in the synchrotron will be subject to instabilities from a number of causes[11]. In general it seems desirable to produce the short bunch for the shortest possible time interval to minimize instabilities. Space charge tune shifts at injection and extraction, structure resonances, the microwave instability, transverse resistive wall and head tail thresholds must all be avoided as much as possible. Multibunch instabilities will probably require damping. In this context it is useful to remember that: 1) The charge/bunch is only a factor of three

beyond the Brookhaven AGS, the charge/pulse is only 50% more than is regularly achieved, and the bunch would be in the ring for only $\sim$ 2 % of the AGS acceleration time. 2) The accelerator would be operating entirely below transition, where beams are more stable. Nevertheless, every increase in machine performance has been accompanied by the discovery of new types of instabilities.[12]

Structure resonances In general one would like to minimize the driving terms and the growth rates to give the best opportunity of extracting the beam before significant beam loss. The CERN booster has operated with very large space charge tune shifts at injection by tuning out structure resonances.

Space charge Our goal is to create a bunch hitting a target with $2.5 \cdot 10^{13}$ protons with an rms length of 1 nsec at 30 GeV, or $5 \cdot 10^{13}$ protons at 10 GeV. It seems difficult to create such a bunch in equilibrium in a ring. For example if the initial bunch at injection is space charge limited, the tune shift will be reduced by the ratio of $\beta\gamma^2$ but increased by the ratio of the bunching factor. For the case developed below, 1 GeV injection and 8 GeV extraction in a ring of length 1600 nsec, the tune shift at extraction is higher than at injection by a factor of two or three. On the other hand, large transient tune shifts have been observed.[13]

Microwave instability Short intense bunches could be expected to produce microwave instabilities, since the threshold is inversely proportional to the peak current, I_{pk},

$$\frac{Z_{\|}}{n} = \frac{F \mid \eta \mid \beta^2 E/e}{I_{pk}} \left(\frac{\Delta p}{p}\right)^2, \tag{3.2}$$

with $F = 1$. This is the "Keil-Schnell" criterion, ignoring niceties of the dispersion equation. This threshold is apparently exceeded by a factor of ten in coasting beam, and a factor of three in bunched beam in ISIS.[14] There is some disagreement about the reason for this.

Transverse Resistive Wall instability The growth times are dominated by the impedance of the kicker magnets, while the thresholds are determined by the space charge impedance. There is a relationship between space charge tune shift and transverse impedance which sets a limit to the ability to stabilize the motion with Landau damping.[15] If the space charge tune shift is at its maximum value, Landau damping will cause this limit to be exceeded. Then a fed back kicker will be needed to stabilize the lower modes

3.4 Stability of Short Bunches

The bunch hitting the target will have a peak current of 1600 A (at 30 GeV) or 3200 A (at 10 GeV) which is significantly larger than any current seen in a proton synchrotron. Although we expect instabilities, they will be moderated by three effects: 1) the large current will exist in the ring for a very short time, perhaps only a few turns, 2) the intense bunch is only required at the target, and 3) the short bunch is in many respects a more stable configuration than the long bunch that produced it.

We consider a number of instability mechanisms and their effect on an intense, one ns proton bunch. Although there are a number of options being considered, it has been necessary to look primarily at one example. We have chosen the 10 GeV option with 2.5×10^{13} protons/bunch, assuming two of these bunches could be combined at the target. In general beams are more stable at higher energies, however bunching times are also longer. Higher currents are probably more troublesome so we have not looked at the bunches with 5×10^{13} /bunch.

Structure Resonances The large incoherent space charge tune shift will require the beam to cross a number of resonance lines and will be the cause of some emittance growth, however the short bunch will last only a few turns and the growth times of these effects has been fairly long. This effect will be a more serious problem at injection.

Transverse Space Charge The incoherent space charge tune shift given by

$$\Delta\nu_{inc} = \frac{3\, r_p\, N_t}{2\, A\, B\, \beta\, \gamma^2} \approx 0.2, \tag{3.3}$$

where r_p is the classical proton radius, N_t is the number of protons per bunch, A is the phase space area of the bunch, B is the bunching factor, β and γ are the relativistic velocity and mass factors. The coherent tune shift is

$$\Delta\nu_{coh} = \frac{r_p\, N_t\, \beta_{av}\, \epsilon_1}{\pi\, \gamma\, h^2} \approx 0.0004, \tag{3.4}$$

where β_{av} refers to the average beta function around the ring, ϵ_1 is the Laslett coefficient for the vacuum chamber, and h is the vacuum chamber height. The large incoherent tune spread will tend to stabilize the beam by introducing Landau damping. The coherent tune shift is a function of the vacuum chamber shape and can be reduced by going to a circular shape where $\epsilon_1 = 0$.

Longitudinal Space Charge Space charge will cause the beam to be effected by a longitudinal voltage per turn

$$V(z) = \left(\frac{\beta^2 c^2 L}{2} - \frac{g_0 \, R}{2 \, \epsilon_0 \, \gamma^2} \right) \lambda'(z), \tag{3.5}$$

where the first term in parenthesis is the inductive term, g_0 is a function of the beam and vacuum chamber dimensions, $\lambda'(z)$ is the derivative of the longitudinal charge density, R is the radius of the machine, ϵ_0 is the permittivity of free space, and z is the position along the bunch. This effect will tend to lengthen the beam below transition and shorten the beam above transition (negative mass instability). For the shortest bunches, the voltage produced will be equal to ~ 1 MV/turn. While very large, this voltage is much smaller then the ± 200 MeV/c momentum slewing required to bunch the proton beam, and even small compared to the momentum spread / number of turns required to bunch the beam, ± 200 MeV/50 turns $= \pm 4$ MeV/turn. In fact the perturbation on the production of a short bunch, while significant in slow bunching, is almost negligible if the bunching takes place over less than a few hundred turns. In this context it is interesting to note that the longitudinal space charge does cause an increase in the final bunch length, however the contribution to the bunch length increase is independent of the degree of bunching because as the bunch gets shorter and the voltage becomes larger, the projection onto the time axis becomes smaller, thus each turn contributes roughly the same (small) increase in bunch length. The negative mass instability can be avoided by operating below transition, as is planned. It should also be noted that the bunch shape can be controlled at injection to some extent so one can assume either a Gaussian or a parabola, which would have a linear longitudinal voltage profile.

Transverse Resistive Wall The large space charge tune spread ($\nu_{inc} \sim 0.2$) will tend to damp the beam with a time constant

$$\frac{1}{\tau_d} = \frac{\omega \, \Delta\nu_{inc}}{2 \, \pi}, \tag{3.6}$$

ω being the rotational frequency of the synchrotron. With a tune spread of 0.2 this is essentially 5 turns, which is very roughly the number of turns that the short bunch would exist in the machine before extraction in the 10 GeV option. This would mean that any excitation must occur almost in a single turn, a time that is very short compared to the excitation of this effect in existing machines.

Head-Tail The bunching process involves a huge momentum slewing, and space charge induced damping. Adding chromaticity with sextupoles would produce a large tune shift between the front and rear of the bunch and would permit considerable Landau damping.

Longitudinal Microwave The Keil Schnell criterion gives the allowable range of longitudinal impedance as $Z_\parallel/n < F \mid \eta \mid \beta^2 E/e(\Delta p/p)^2/I_{pk}$, where F is a numerical factor (~ 1), η is the slip factor for dispersed beams, β is the velocity E/e is the beam energy, $(\Delta p/p)$ is the momentum spread for a given longitudinal position in the bunch, and I_{pk} is the maximum beam current. As has been pointed out by Schnell [16], bunch rotation to a shorter overall bunch length gives a more stable configuration because the momentum spread is proportional to I_{pk}, but the term in the numerator is squared, thus the allowable $Z_\parallel/n$ increases as the bunch becomes shorter. Two other points can be made: 1) The growth time of longitudinal oscillations would be roughly 1/4 of the synchrotron period for synchrotron oscillations excited by a voltage of $V = I_{pk} Z_\parallel/n$, which would be comparatively slow. 2) The CERN PS has run with beams near γ_t and found them to be stable.

High Frequency Cavity Beam Loading A rough estimate of the allowable wall impedance $Z_\parallel/n$ for high frequency loading in the rf cavities can be obtained by requiring the voltage induced to be small relative to the voltage provided by the cavities for acceleration or bunch rotation. This constraint gives the relation $(\, V = I_{pk}\, Z_\parallel/n \,) << (\, V_{rf} \approx 2$ MV/turn). This relation can then be used to produce limits on the high frequency behavior of the cavities.

Robinson Instability The rf cavity tuning can be adjusted to mitigate this. The cavity gap impedances may have to be actively adjusted using a high degree of local rf feedback.

Multibunch Modes Although the bunches are short, the rf frequency would be in the range of 3 - 5 MHz, so feedback and active damping should be comparatively easy to do.

Intra-Beam Scattering The intra beam scattering growth rate has been estimated and found to be quite long ($\sim$ sec) so this does not seem to be a concern for the short time the beam will be bunched.

Charge Neutralization by Residual Gas Although the bunches will be dense, normal accelerator vacuums should be able to insure that focusing by trapped electrons should be minimal, either in the accelerator or in a single purpose compressor ring.

3.5 Components

Lattice Issues The lattice has not been determined at this time. Two features may be desirable: 1) efficient use of circumference by bending magnets and RF and, 2) control of

γ_t. Since the acceleration gradients in these rapid cycling machines are on the order of 1 TeV/sec, it is desirable to have efficient use of rf, and a higher circulation frequency (smaller circumference) aids this. Control of γ_t is desirable to insure that one does not have to operate above transition, even at $\sim$ 30 GeV. It is also desirable to be able to control transition during bunch rotation.

We have considered several options for the proton driver lattice. The 30 GeV option could use a variant of the lattice proposed for the Japanese Hadron Project [3]. At 10 GeV one possible choice is a FODO lattice with eight super-periods and six cells in a super-period. The half cell length is 4.9 m and there are two long straight sections with zero dispersion per super-period and two dipoles per cell. The tunes would be $\sim$14 and γ_t would be about 12. A γ_t jump system based on a system proposed by Visnjic[17] can move the γ_t by one or more units during the bunch rotation and extraction.

We have also considered a Flexible Momentum Compaction lattice[10] for both options. This lattice can be tuned for large or imaginary γ_t, is very efficient but requires tuning for zero dispersion straight sections(see Figure 3.4). Both this lattice and that proposed for the Japanese Hadron Project seem quite sensitive to γ_t, in that quad changes of roughly 1% can move γ_t by $\sim$ 10% without significantly changing the tunes. This makes them desirable for this application.

RF System The RF system could be modeled after the cavities designed for the IPNS-II synchrotron[18]. These cavities produce 18 kV/gap over a frequency range from 1.12 to 1.50 MHz, a swing of 33%. The options considered here require higher frequencies ($\sim$ 3 MHz) but smaller frequency range (3% 30 GeV and 14% 10 GeV). A detailed design, with a larger inner radius for the ferrite rings and the smaller frequency swing, should give an acceleration gradient of greater than 15 KV/m, (see Fig 3.5). Beam loading at high intensities has been discussed by J. Griffin[19].

Injection Minimizing losses during the acceleration process will require precise control of the initial phase space distribution of the beam in both the longitudinal and transverse dimensions. The KAON Factory Study [4] described painting algorithms which will produce the desired distributions using charge exchange injection. It will also be necessary to capture any remaining neutral beam to minimize local activation.

Vacuum The large magnetic field swings required by the high repetition rate will not penetrate thick metallic vacuum chambers. The ISIS [20] synchrotron solved this problem by constructing a ceramic vacuum chamber with wires parallel to the beam on the inside of

the chamber to carry the image charge. Capacitors which would pass beam frequencies and block magnet frequencies permitted the magnetic field to penetrate the wire screen. This solution works well at ISIS, but is more expensive and uses magnet apertures less efficiently than a metallic chamber.

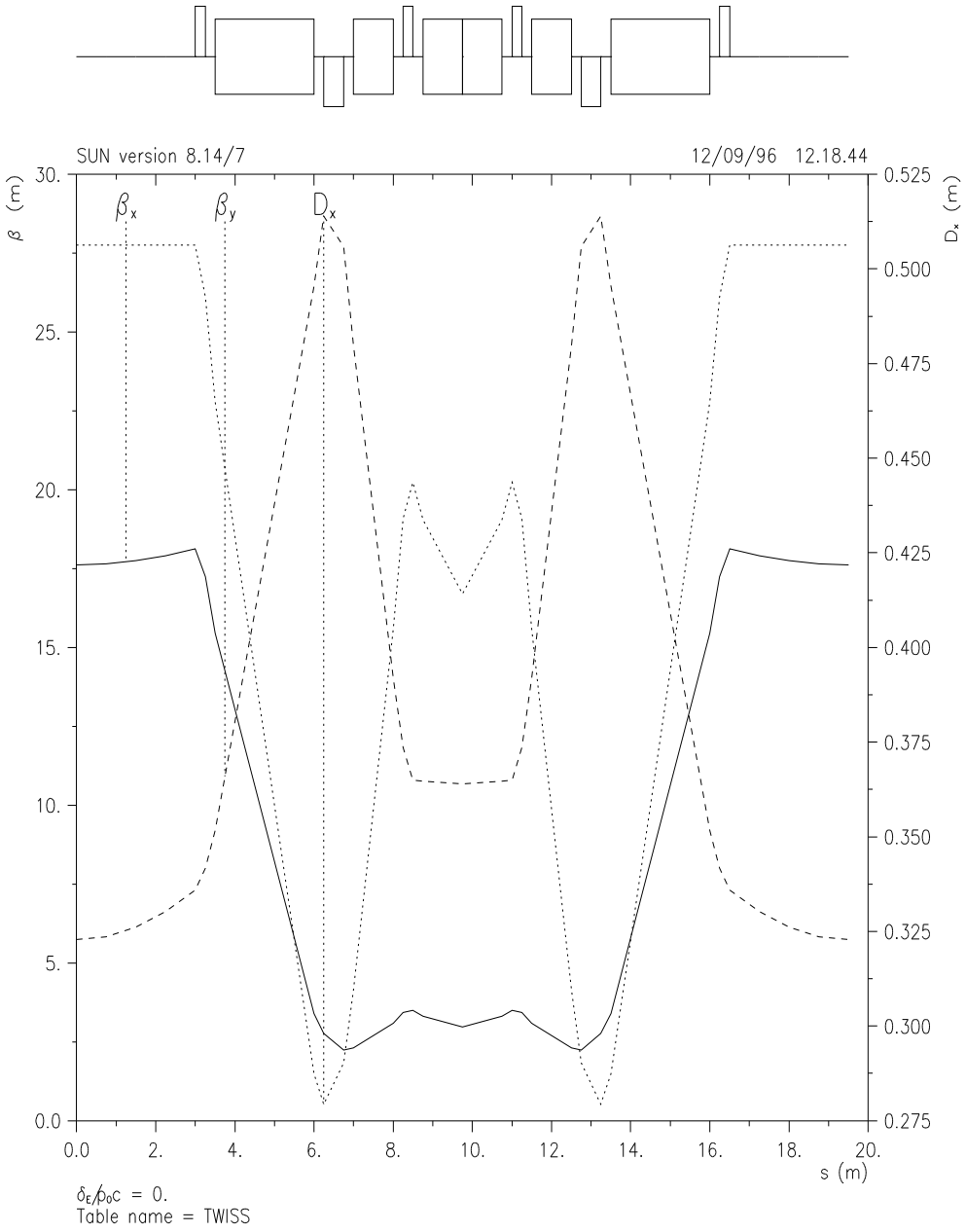


Figure 3.4: One cell of a FMC lattice.

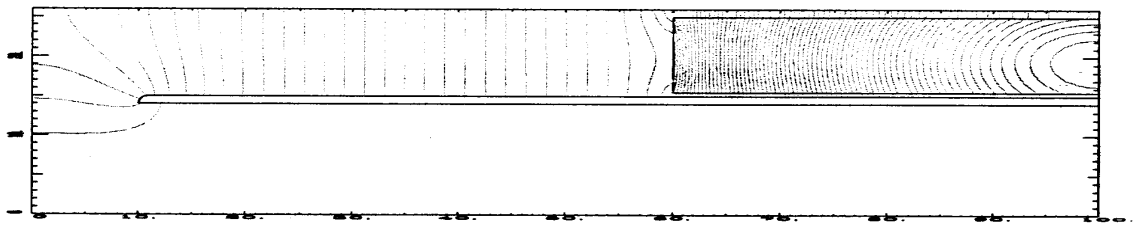


Figure 3.5: A candidate 3 - 4 MHz RF cavity. This cavity has a beam aperture of 28 cm, a length of 100 cm and would generate 15 kV/m when powered by 12 kW.

Magnets/PS Two options exist: resonant power supplies and driving the magnets directly, perhaps with some load leveling. Resonant power supplies require less load from the grid, and a two frequency system has been proposed which increases the acceleration time, while keeping the overall rate constant[21], however both systems require an uneven acceleration profile, which requires additional rf. Direct excitation of magnets would minimize the rf requirements.

Extraction The primary problem would be to avoid losses, since even a small fraction of the 5 MW of beam power would cause considerable activation of the extraction septa and downstream components. The problem has been considered for neutron spallation sources [2] [18].

3.6 Examples

It is too early to fix any parameters of the design of the proton driver. We provide some details on options which have been studied.

3.6.1 30 GeV

A proton driver operating at 30 GeV would closely follow the designs of spallation neutron sources and KAON as discussed above. The high proton energy permits transition to a short bunch using a normal bunch rotation. Compared to the 10 GeV option, the fractional momentum spread is smaller and the required charge per bunch is also smaller. On the other hand, the longer bunching time requires that the large peak current I_{pk} circulates in the synchrotron for a longer time.

3.6.2 10 GeV

A 10 GeV, 30 Hz synchrotron would operate at higher repetition rates and could be smaller, simpler and cheaper. We have considered a design with a 1 GeV linac and overall circumference of about 580 m. By eliminating one beam transfer, this system might have lower losses than a booster / driver combination accelerating to higher energy. Two bunches would be combined at the target with chicanes to give the required 5×10^{13} protons, keeping the bunches in the ring smaller. The larger momentum spread would be more difficult to confine and extract.

Because of the large fractional momentum acceptance required, we have assumed that this option would operate with the two stage bunch rotation described above. The first stage would involve running near transition and the second stage would be quick bunching with the transition energy moved moved perhaps 3 GeV above the beam energy.

Table 3.6: 10 GeV Option Parameters

	Driver	
Injection Energy	1	GeV
Max. Energy	10	GeV
Rep. Rate	30	Hz
Protons per Pulse	1×10^{14}	
Number of Bunches	4	
Circumference	580	m
Transition Gamma	11.9	
Max. Dispersion	1.6	m
Trans. Emittance [95%]	300	π mm m rad
Inc. Tune Shift Inj	.32	
RF Voltage per Turn	2	MV
Harmonic Number	6	
RF Frequency	3.3-3.7	MHz
Long. Emittance [95 %]	2.5	eVs

One possible parameter set, based loosely on designs for pulsed neutron sources, would use a FODO lattice with short bending magnets to produce a racetrack shaped ring with two long straight sections and a transition energy of about 12. An injection energy of 1 GeV would require a normalized emittance of 300 π mmmr in both x and y, and magnets with half apertures of 0.08 m.

Roughly 2 MV / turn of RF would be required. For a small number of final bunches this

frequency might be in the range of 2 - 4 MHz. Assuming cavities giving 10 kV/m, these cavities might require ∼200 m of straight section space.

The 1 GeV linac for this option would be based on the FermiLab 400 MeV injector, a drift-tube plus coupled-cavity, room temperature linac. This linac presently can accelerate up to 50 mA of H^- beam at 15 Hz with a maximum 125 μsec pulse length (4×10^{13} protons per pulse or 6×10^{14} protons per sec). For a muon collider, this linac design can be upgraded to 30 Hz, 65 mA current and a 250 μsec pulse length. This would provide the needed 3×10^{15} protons per sec. The duty cycle of 7.5×10^{-3} is still comfortably low for a room temperature linac. The low energy part of this linac would consist of a 30 keV, 75 mA H^- ion source, a 2 MeV, 200 MHz RFQ and five 200 MHz DTL tanks to accelerate the beam to 116 MeV. These tanks would be nearly identical to the existing FermiLab DTL ($E_0 = 2.5$ MV/m) and could be powered by the standard 5 MW triodes.

Following the DTL would be an 800 MHz coupled-cavity linac for acceleration to 1 GeV. An average gradient $E_0 = 6$ MV/m would keep the cavity spark rate below 10^{-3} per pulse for the entire linac based on FermiLab experience. The 800 MHz linac would be 233 meters in length. This linac would be segmented into nineteen, 9 MW modules so proven Litton 12 MW klystrons could be used. Seven such klystrons power the FermiLab 400 MeV side-coupled linac. One expects the normalized emittances to be nearly the same at 400 MeV and 1 GeV. Based on present 400 MeV beam parameters, the 95% normalized transverse emittance should be 7 π mm-mrad, and the full longitudinal emittance 10^{-4} eV-sec, or 30 MeV-degrees (805 MHz), at the end of the 1 GeV linac. Scaling from 400 MeV to 1 GeV with a gradient of 6 MV/m, the full width energy and phase spreads are expected to be 3.7 MeV and 8.1 degrees. This beam will need to be debunched to reduce the energy spread for injection into the synchrotron.

3.6.3 2.5 Hz

A high energy, low rep-rate driver is also being considered. There are advantages in operating a 30 Hz driver at a 6 times lower rep-rate but with 6 times more protons / pulse. The lower rep-rate would permit much less accelerating voltage, simpler magnets, cheaper power supplies, smaller eddy current effects in metal vacuum chambers and better matching to the filling requirements of the super-conducting linacs used in the muon accelerator. The additional charge could be accommodated around the circumference of the driver without raising the peak current and the beam pulses from the Booster could be accumulated in an additional 3.6 GeV storage ring with the Driver circumference.

3.6.4 Polarized μ Production

Polarized beams can be produced from both π^+ and π^- by capturing only one polarization, although this process is inefficient. More protons are required on the target to make up for the increased losses. An additional factor of two in proton intensity at the target can be provided by adding another synchrotron in parallel. The cost of this method would be less than double the cost of a single synchrotron because many components would be used in common. The beams could be combined at the target with septum magnets in a similar manner to that proposed for combining bunches with chicanes.

3.7 R & D Issues

The proton driver described above is similar to existing synchrotrons and designs. Some R & D would be useful to evaluate bunching methods, examine instabilities that might be driven by high currents and study operating modes which minimize losses.

Bunching tests which can be done in the Brookhaven AGS can look at $I_{pk} \sim 50 - 150$ A, which approaches the range at which the acceleration would take place. This would also provide data on the nonlinearities of bunch rotation.

Both theoretical and experimental studies of instabilities in rings with high I_{pk} would be useful. Since this current would be present for a short time, during which the bunch properties would be changing rapidly, the environment would be different from that usually encountered in synchrotrons.

With a 5 MW beam it will be desirable to minimize losses to permit simple maintenance of accelerator components. There are a number of techniques which have been developed to minimize losses in high current machines, such as more efficient disposal of the linac beam in charge exchange injection, painting the phase spaces to insure minimal losses during capture and acceleration, sufficient rf to insure protons do not escape from buckets.

Bibliography

[1] G.T. Danby and J.W. Jackson, *Part. Accel.* **27** 33 (1990).

[2] 5 MW Pulsed Spallation Neutron Source, *Preconceptual Design Study* (June 1994) BNL-60678

[3] Y. Mori et al., *Outline of JHP Synchrotron Design*, Institute for Nuclear Study, University of Tokyo, Oct. 1995, INS-Rep-117.

[4] *KAON Factory Study, Accelerator Design Report*, TRIUMF, Vancouver, B. C.

[5] Cappi, Delahaye, and Reich, *IEEE Trans Nucl Sci.* **NS-28** 2389, 1981

[6] X-Q Wang, *A Study of Longitudinal Coherent Effects of Unbunched Beams Near Transition in the FermiLab Accumulator*, PhD Thesis, Ill. Inst. of Tech. (1991)

[7] J. Griffin, *Proton Driver Proposal*, unpublished note, FermiLab 1996

[8] J. Norem, *Production of Short Bunches near* γ_t, Unpublished Note, Argonne, 1996

[9] Cappi, Garoby, Mohl, Vallet and Wildner, European Particle Accelerator Conference, EPAC 94 London, June 1994

[10] Lee, Ng and Trbojevic, Phys. Rev. **E48** (1993), 3040

[11] F. Mills *Problems in the Design of a Proton Driver for a Muon Collider*, FNAL TM-1983, 1996

[12] Garayte, in Adv. of Acc Phys. and Tech., H. Schopper ed., World Scientific. Singapore 1993, p67

[13] Private communication, Alfred W. Maschke, 1972. The tune shift of several was demonstrated at injection in the AGS using the RF to bunch the beam transiently

[14] Grahame Rees, private communication, 1995.

[15] Baartman, Picacho Plaza Workshop, Santa Fe, 1993.

[16] W. Schnell, CERN Report 77-13, 1977, p231.

[17] V. Visnjic, Phys. Rev. Lett. **73** (1994), 2860

[18] IPNS Upgrade, A Feasibility Study, Argonne National Laboratory Report ANL-95/13, April 1995

[19] J. E. Griffin, *Aspects of Operation of the FermiLab Booster RF System at Very High Intensity*, Unpublished Note, FermiLab, Nov. 1995

[20] *Spallation Neutron Source: Description of Accelerator and Target*, B. Boardman ed. Rutherford Appleton Laboratory Report RL-82-006 (1982)

[21] W. F. Praeg, IEEE, Trans. Nucl. Sci. **NS-30**, (1983), 4

Contributors

- J. Norem, (ANL)
- J. Griffin, (FermiLab)
- C. Johnstone (FermiLab)
- F. Mills, (FermiLab)
- A. Moretti, (FermiLab)
- M. Popovic, (FermiLab)
- T. Roser, (BNL)

List of Figures

List of Tables

Chapter 4

TARGETRY AND PION PRODUCTION

Contents

4.1 Target and Collection Optimization for Muon Colliders

4.1.1 Basic Description

To achieve adequate luminosity in a muon collider it is necessary to produce and collect large numbers of muons. The basic method starts with a proton beam impinging on a thick target

(one to two interaction lengths) followed by a long solenoid which collects muons resulting mainly from pion decay. Because of their short lifetime muons must be generated by a single proton pulse for each new acceleration cycle. Production and collection of pions and their decay muons must be optimized while keeping in mind limitations of target integrity and of the technology of magnets and debuncher cavities.

Early estimates of muon yield, based on conventional lithium lens and quadrupole magnet collection methods, indicated that roughly 1000 protons are needed for every muon delivered to the collider rings [1]. This results from inherent limitations in the momentum acceptance of these systems (typically less than ± 5 percent) which causes most (potential) muons produced to be wasted. Motivated by neutrino beamline experience, a solenoid collection scheme for pions has been suggested [2]. Cursory simulations indicated significant improvement in muon yields to about 0.3 muon per proton for proton energies below 100 GeV while above this energy a collection system with two lithium lenses could surpass a solenoid.

Luminosity estimates indicate that 2×10^{12} muons per bunch delivered at 15 to 30 Hz are required for a 2+2 TeV muon collider. Assuming a yield of roughly one charged pion per proton and the efficiencies for pion to muon conversion, muon cooling and muon acceleration are each about 0.5, the approximate number of protons needed at the target is 10^{14} per pulse. However the power required for a 15 to 30 Hz rapid-cycling proton synchrotron with 10^{14} protons per pulse becomes relatively expensive above 30 GeV. Also multiplicities and pion yields increase less rapidly above 30 GeV. For energies below 3 GeV pion yields drop off, and there is an increasing asymmetry in the π^+ to π^- ratio due to Δ resonance production. Since the proton bunch must be a few nanoseconds long at the target to facilitate pion debunching and momentum spread reduction, the proton energy must also be high enough to reduce space-charge effects. These considerations suggest that the kinetic energy of the proton driver should be between 3 and 30 GeV. Actual yields and target energy densities may well depend considerably on incident energy. This has motivated detailed pion production studies at 8 and 30 GeV. Inter-comparisons and conclusions derived from these studies, such as in the optimization of target size or solenoid field with respect to muon yield, are expected to be much less sensitive to incident energy. Hence the results at these two energies are viewed as complimentary providing a firm understanding of pion and muon production and target integrity in this energy range.

The basic collection scheme, as outlined by Palmer et al [2] is illustrated in Fig. 4.1 and forms the starting point for the simulations described in this chapter. A very high-field hybrid solenoid extends the length of a target upon which a proton beam impinges. Based on near term technology, fields of 20 to 28 T appear to be achievable for this purpose. This target solenoid collects pions with a large momentum spread and with large angles and guides

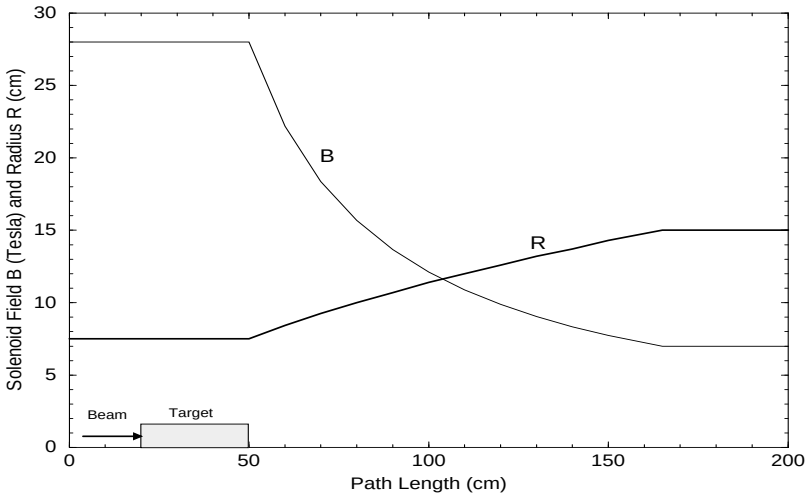


Figure 4.1: Capture solenoid field and inner radius as a function of distance.

them downstream into a long solenoid channel (5 to 7 T) where they decay to muons. Most studies described in this chapter were done for a 28 T solenoid since this was the field of choice until recently. Present scenarios prefer a more conservative 20 T solenoid. Yields are expected to be reduced by 15 to 20% compared to the 28 T case, but basic conclusions will not be changed.

The high-field solenoid aperture of 7.5 cm is chosen to give a large transverse phase space acceptance adequate for a transverse momentum

$$p_{\perp}^{max} = qBa/2 \tag{4.1}$$

where B is the magnetic field, q the particle charge, and a the solenoid radius. The normalized acceptance of this solenoid for pions is

$$A_n = ap_{\perp}^{max}/\, m_{\pi}c = qBa^2/2m_{\pi}c. \tag{4.2}$$

For a 28 T solenoid the momentum acceptance is 0.314 GeV/c, and the phase space acceptance is 0.17 m·rad. This is much larger than the intrinsic pion beam emittance at the target, $r_p p_{\perp}^{max}/m_{\pi}c = 0.02$ m·rad for a proton beam radius $r_p = 1$ cm. Hence there is no reason to further reduce proton beam size—which may thus be set by considerations of yield and target heating rather than pion emittance. For a 20 T solenoid the acceptance is reduced by 30%, but most of the pion beam is still within the central phase space region, and yields do not suffer proportionally.

The target region is followed immediately by a roughly one meter long matching section which reduces the field via a $B_0/(1+\alpha z)$ dependence. In this region the pipe radius increases

to a value which corresponds to the radius of the lower field solenoid serving as pion decay region. This keeps the product Ba^2 constant and the acceptance unchanged. The parameter $\alpha = (qB_0/2p_\pi)(d\beta_f/dz)$ is chosen such that for a characteristic pion momentum the rate of change of the beta focusing function ($\beta_f = 2p_\pi/qB$) with distance is less than 0.5, which might still be considered an adiabatic change of the field.

One of the outstanding problems associated with the solenoid collection of secondary beams immediately after a target is the disposition of the remnant high energy proton beam and the various non-muon producing particles. These unwanted particles form an intense swarm which outnumber the pions several to one and can potentially induce significant radioactivity far downstream. Although no final solution to this problem exists, a possible solution was discussed in [3]. There a *curved* solenoid field is introduced to separate positive and negative pions within a few meters of the target. This permits each to be placed in separate RF buckets for acceleration which effectively doubles the number of muons per bunch available for collisions and increases the luminosity fourfold. The proton beam and neutrals impinge on the curved walls in this limited region, and few manage to travel far downstream with the pions. This option is not part of the reference design for the 2+2 TeV collider, but it is discussed later in this chapter to encourage further thought on the matter of beam cleanup and charge separation.

An extensive number of simulations [3, 4] have been performed for pion production from 8 and 30 GeV proton beams on different target materials in a 28 T solenoid. Solid (graphite, aluminum, copper, tungsten) and liquid (gallium, mercury, lead) targets of different radii ($0.4\,\mathrm{cm} \leq \mathrm{r} \leq 2\,\mathrm{cm}$) and thicknesses ($0.5\lambda_I \leq \mathrm{L} \leq 3\lambda_I$), where λ_I is nuclear interaction length, have been explored. Values of λ_I are taken from [8].

Table 4.1 presents an overview of target parameters which have resulted from these studies. Of the many materials studied, only three are presented here based on their practicality and potentially useful properties for high-power targetry. Details of this work are found in the cited references and the essential results will be discussed in the following sections. In brief this work has demonstrated by simulation that yields of 1 to 2 charged pions per proton are achievable from any of the targets considered if collection occurs in a large aperture solenoid and that pion to muon decay efficiencies can exceed 0.75 in the downstream decay channel. Although an adequate supply of muons has been demonstrated by simulation, the beam phase space is, of course, tremendous and requires substantial cooling.

High–Z targets are preferred for producing secondary beams when collection is done with a thin lens to reduce depth of focus problems (*e.g.*, antiproton collection with a lithium lens). This concern is absent with solenoid collection, so other constraints determine the choice of target material. At the proton intensities required for a muon collider, high–Z

targets suffer from extremely high heat loads due to electromagnetic showering with only a marginal increase in pion yield compared to low–Z targets. In going from 8 to 30 GeV and doubling the average beam power, the average dissipated power in high–Z targets increases by a factor of 2 with a doubling of the yield. For low–Z targets of the same length ($1.5\lambda_I$) the average power changes very little with a doubling of the yield in this energy range. For high–Z materials the power densities suggest the need for much larger diameter targets, microchannel cooling or circulating liquid targets. Low–Z targets may then be preferred for the muon collider application. High energy proton beams are not necessarily preferred on the basis of pion yield, although they may be required for making very short bunches.

An issue not addressed in this chapter is the tendency of high–Z targets to produce many radionuclides, especially alpha emitters, which are then transported about by a circulating liquid target system or would produce accumulated helium and embrittlement in a solid target. Difficulties in the containment of a hot radioactive liquid in the target station during an accident may also argue against the use of high–Z recirculating targets.

4.1.2 Target Region Studies

Overall Optimization

For the collection geometry described in the previous section, target composition, length and radius are varied and pion yield is studied using particle production and transport simulation codes. The proton beam is assumed to have an emittance of $\epsilon_N^{rms} = 4\times10^{-5}$ m·rad consistent with a value expected from a high-intensity proton source. The focusing function at the target is conservatively chosen to result in a relatively wide beam with σ_x=σ_y=0.4 cm.

Several computer codes for particle production are used in this study due to preference and availability at different laboratories. This also allowed consistency checks of the codes. The MARS code [5], developed over many years at IHEP and FermiLab for particle–matter interaction simulations, is used for simulating particle production and transport in thick targets within the solenoid field. The DPMJET [6] and ARC [7] codes are used to compare particle production and spectra with MARS. The ARC code continues to be improved and recently was upgraded to simulate thick targets with re-absorption. The MARS is also used to study energy deposition in the target and surrounding solenoid. Calculated pion, kaon, and proton spectra from MARS at the target exit for a representative case are shown in Fig. 4.2.

All codes agree remarkably well for the total pion yield from all nuclei in the middle of the studied range of incident proton energies (see Fig. 4.3–Fig. 4.6).

However, it appears that ARC underestimates pion yield at low energies, and at 30 GeV the ARC code predicts about 40% more pions than MARS and DPMJET for all target materials

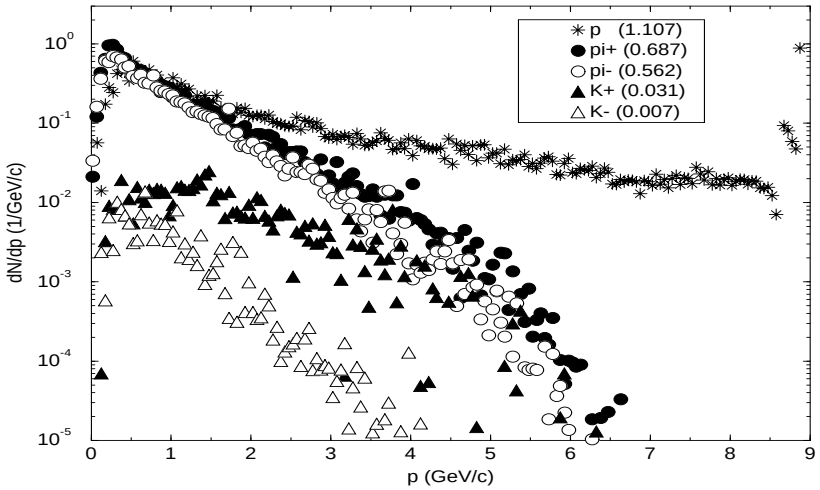


Figure 4.2: Proton, pion and kaon spectra for 8 GeV protons incident on a copper target ($1.5\lambda_I$ length, 1 cm radius) in the 28-T solenoid. Total particle yields are shown in parentheses.

(See Tb. 4.1). The discrepancy for proton energies $E \geq 20\,\text{GeV}$ appears as an excess of pions below 100 MeV kinetic energy (200 MeV/c momentum) (see Fig. 4.7–Fig. 4.8).

This region is not well measured experimentally. A BNL experiment (E910) is in progress to clarify this situation. If an appreciable fraction of these low energy pions escape the target, their presence may affect the optimization of the phase rotation cavities. The cavity system however is designed to collect pions up to at least 700 MeV so the presence of extra low energy pions should not drastically change overall muon yields.

The MARS code describes all the physics processes, so particle decay, interaction, and transport down the solenoid channel can be simulated within a single run as well. It is found preferable in this case to write a special, fast code for tracking particles after the target, using as input a particle file generated by MARS at the end of the target. The tracking code is used as an aid in optimizing target performance, and does not include debuncher cavities. The detailed description of the decay channel with rotation cavities following the target is left to the next chapter.

The special code keeps track of vectorial positions and momenta of each particle as it traverses the beamline as well as time elapsed since the arrival of the incident proton at the target. In addition the code performs $\pi/K \rightarrow \mu$ decay Monte Carlo selection and full kinematics. Muons are progressively downweighted by their decay probability as they traverse the channel. Pions and muons intercepting the beampipe are considered lost. In

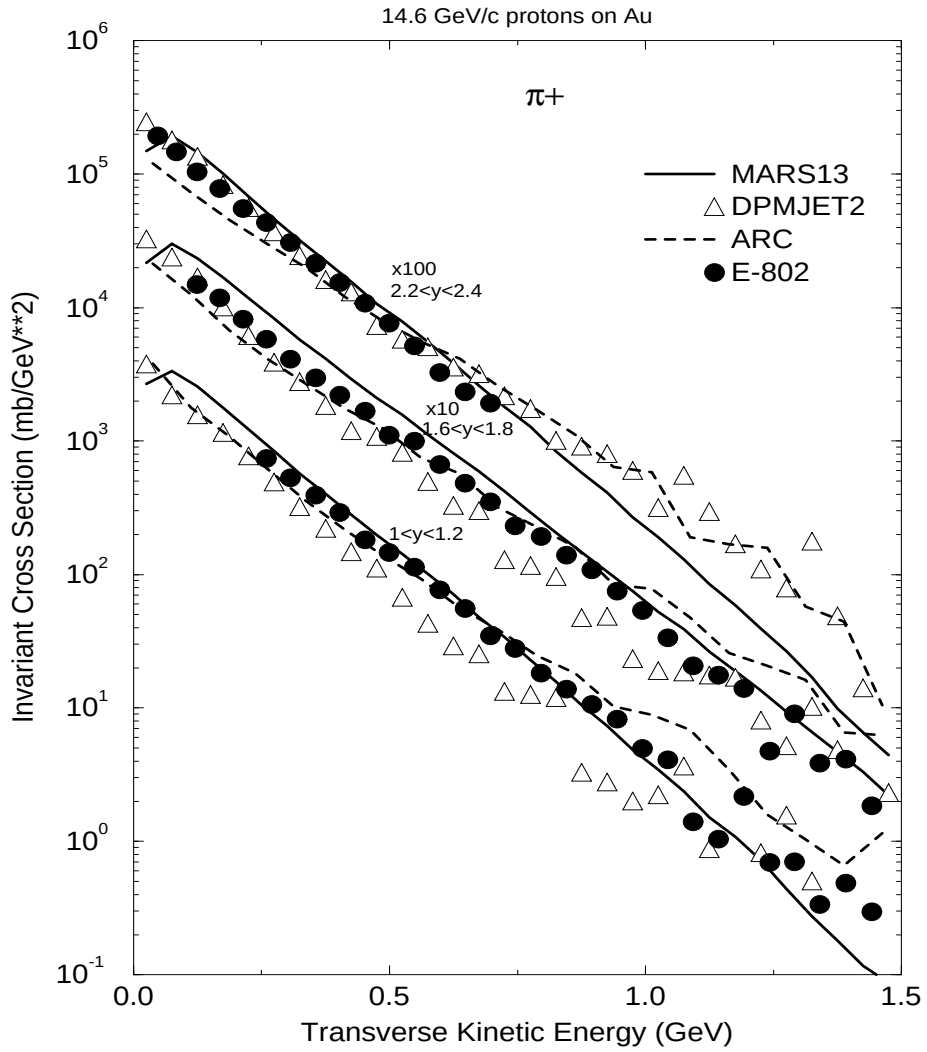


Figure 4.3: Positive pion production cross section for 14.6 GeV/c protons incident on a gold thin target as calculated with MARS, DPMJET and ARC and measured in E-802 experiment at BNL [7].

principle there is a small (but presumably negligible) fraction which may scatter back out of the wall or—in the case of a pion—produce a secondary pion which may rejoin the beam.

A large variety of particles is produced by primary protons and subsequently by secondary and higher generation particles. For 8 GeV p–p interactions the average charged particle multiplicity is about three [8] with a modest increase expected for p–nucleus colli-

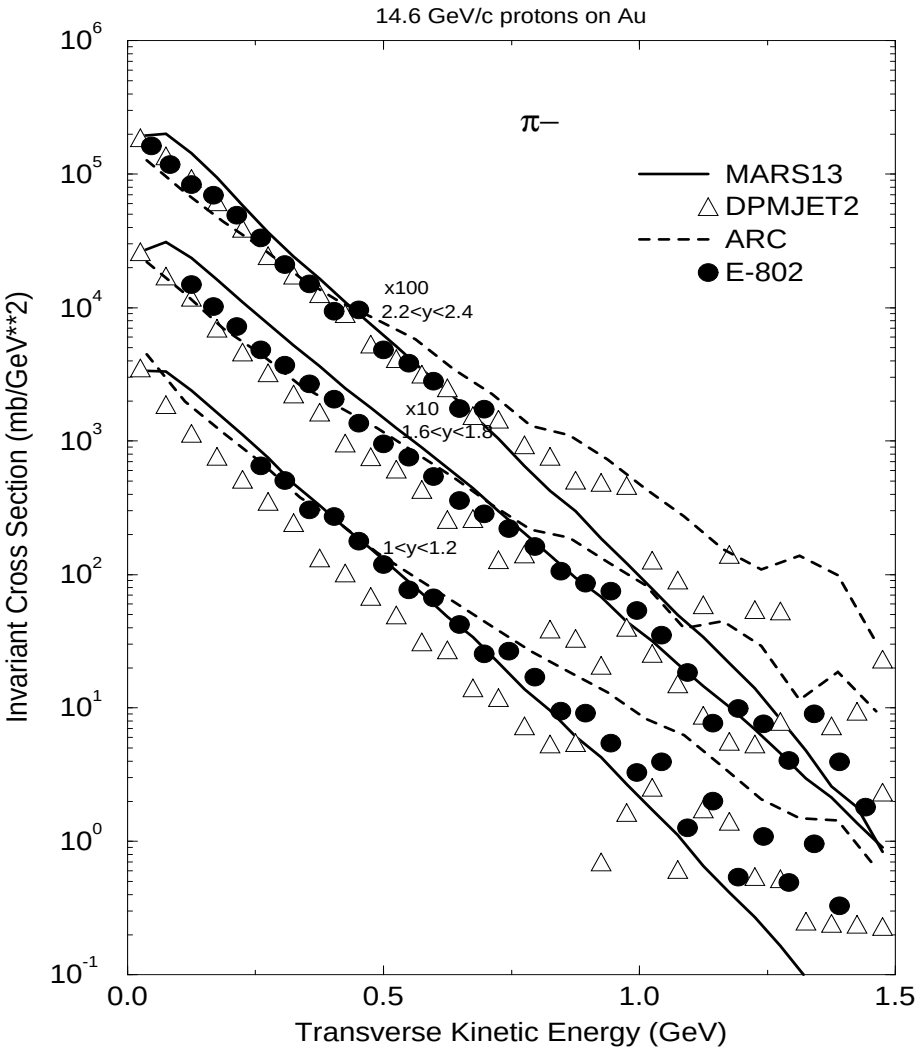


Figure 4.4: Negative pion production cross section for 14.6 GeV/c protons incident on a gold thin target as calculated with MARS, DPMJET and ARC and measured in E-802 experiment at BNL [7].

sions. Excluding the incident protons this leaves an average of somewhat in excess of one charged particle produced per interaction—mostly as pions. For 30 GeV the average charge multiplicity varies from 5.5 to 7 for low–Z to high–Z materials. Charged pion multiplicities range from 3.5 to 5 accordingly. Much of this added multiplicity is due to low momentum pions which suffer significant absorption in thick targets. Hence actual pion yields outside

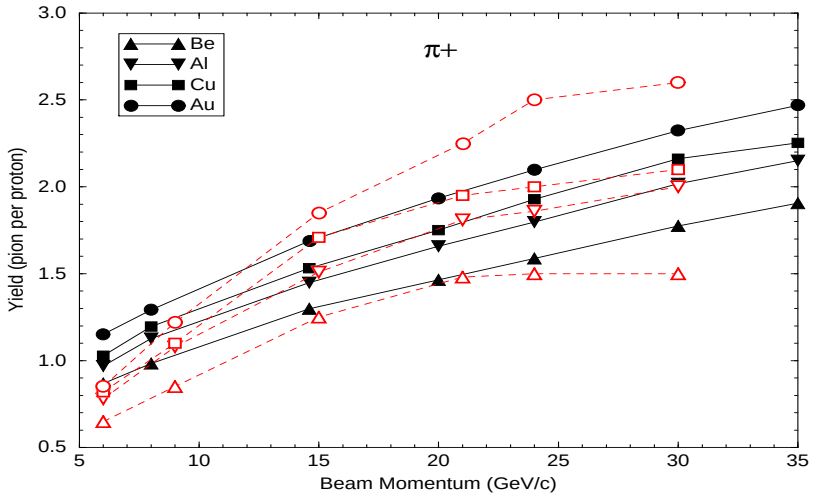


Figure 4.5: Forward π^+ yield from various nuclei *vs* incident proton momentum as calculated with MARS (filled symbols) and ARC (opaque symbols).

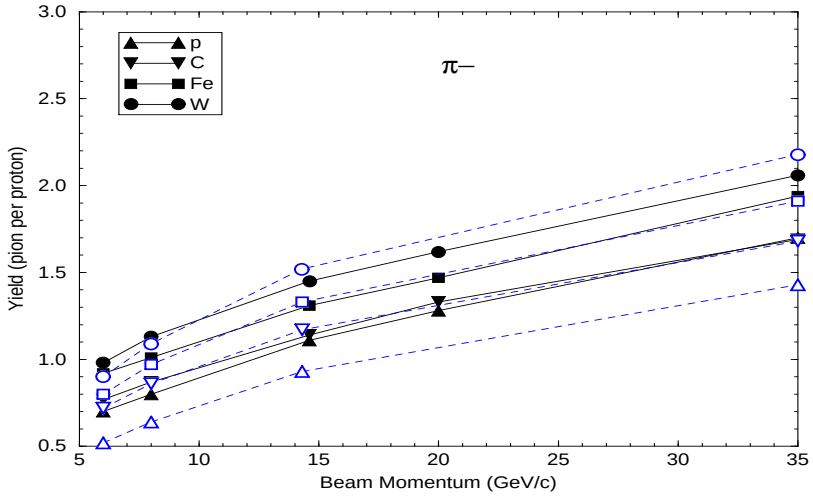


Figure 4.6: Total π^- yield from various nuclei *vs* incident proton momentum as calculated with MARS (filled symbols) and DPMJET (opaque symbols).

of thick targets do not increase as much as the basic multiplicities. Of all produced pions one expects roughly one third to be π^0 which decay quickly into photons leading to electromagnetic cascades in the target. For heavier targets the shorter radiation length permits considerable growth of these cascades leading to many low energy electrons and photons. Among the outgoing particles there will also be some nucleons and nuclear fragments which

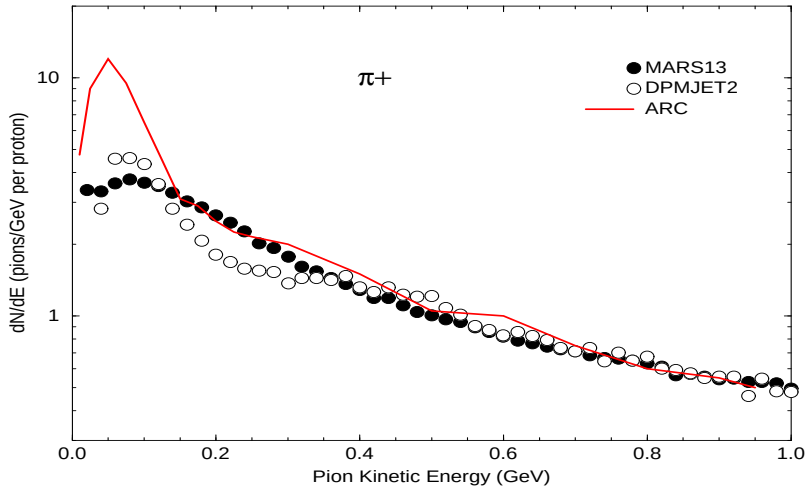


Figure 4.7: Energy spectra of π^+ for 24 GeV/c protons on Hg nuclei as calculated with MARS, DPMJET and ARC.

are dislodged from the target nuclei. All these processes are represented in the MARS code along with elastic and quasi elastic scattering of incident and produced particles.

Simulation of π/μ transport in constant solenoidal fields is readily performed using exact helical trajectories. In the matching region, where the field is more complicated, the simulation proceeds by taking small steps (0.1–0.5 cm) and sampling the field along the

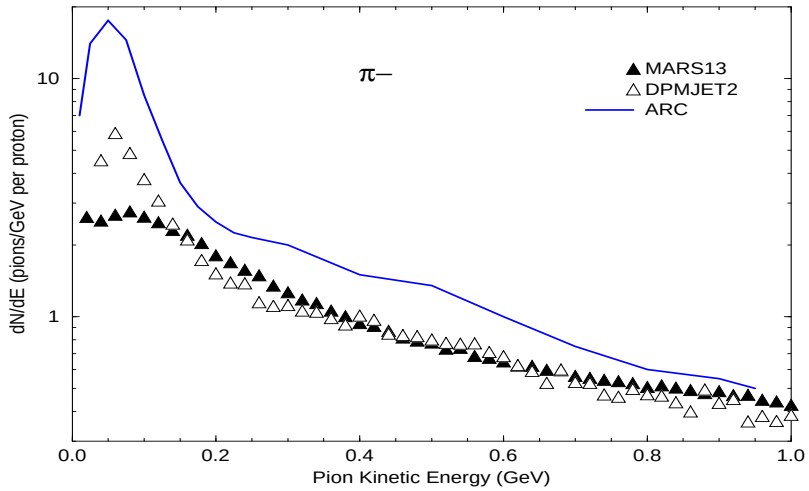


Figure 4.8: Energy spectra of π^- for 24 GeV/c protons on Hg nuclei as calculated with MARS, DPMJET and ARC.

trajectory. The declining field in the matching region means that according to the $\nabla\cdot\boldsymbol{B}=0$ condition the field has a radial component: $B_r \simeq -\frac{1}{2}r\partial B_z/\partial z$. For the above z-dependence $\partial B_r/\partial z = -\frac{1}{2}r\partial^2 B_z/\partial z^2 \neq 0$ and it follows [9] from the $\nabla\times\boldsymbol{B}=0$ condition that B_z must depend on r. This requires that an extra term be present in B_z which—in turn—requires (via $\nabla\cdot\boldsymbol{B}=0$) an extra term in B_r, etc. For the present simulations the iteration is pursued up to quadratic correction terms:

$$\begin{aligned} B_z &= \frac{B_0}{1+\alpha z}\left[1-\frac{1}{2}\left(\frac{\alpha r}{1+\alpha z}\right)^2\right] \\ B_r &= \frac{B_0\alpha r}{2(1+\alpha z)^2}\left[1-\frac{3}{4}\left(\frac{\alpha r}{1+\alpha z}\right)^2\right]. \end{aligned} \tag{4.3}$$

It should be remarked that the analysis simplifies considerably if B_z is made to decline *linearly* with distance in the matching region: $B_z = B_0(1-az)$. Then $B_r = \frac{1}{2}raB_0$ independent of z and both $\nabla\cdot\boldsymbol{B}=0$ and $\nabla\times\boldsymbol{B}=0$ are satisfied. Results of simulations performed with a linear field do not differ significantly from those obtained with the $(1+\alpha z)^{-1}$–dependence.

The results presented in the rest of this section on optimization of pion yield and on energy deposition in the target and solenoid material are obtained with the current version of the MARS code [5].

Pion Yield

A crude target optimization with respect to yield starts by 'tagging' those pions (and kaons) which result in an acceptable muon deep in the decay channel for different targets followed by the 'standard' geometry as described in Section 4.1 (see Fig. 4.1). In excess of 90% of all accepted muons are thus shown from MARS and DPMJET to be the progeny of pions in the momentum range 0.2–2.5 GeV/c for both 8 and 30 GeV protons. As mentioned earlier, the ARC code predicts an excess of pions below 200 MeV/c and the effect of these extra pions on decay channel optimization is discussed in the next chapter. Then for a series of MARS runs, *pion* yield at the target exit and in the above momentum range is determined for various target parameters—without simulation of the collection channel. In addition to contributing little to the muon yield outside the target, pions with momenta less than 0.2 GeV/c have velocities below 0.82 c and thus will quickly drop far behind the main pulse of faster particles unless the debuncher cavities are placed very near the target. Fig. 4.9 show momentum versus time scatter plots of pions, kaons and muons for an 8 GeV proton beam with $\sigma_t = 3$ nsec incident on a 22.5 cm copper target. In all plots $t=0$ refers to the center of the proton bunch at the target entrance. Materials investigated as target candidates are carbon, aluminum, copper, gallium, tungsten, iridium, mercury and lead. This set spans the

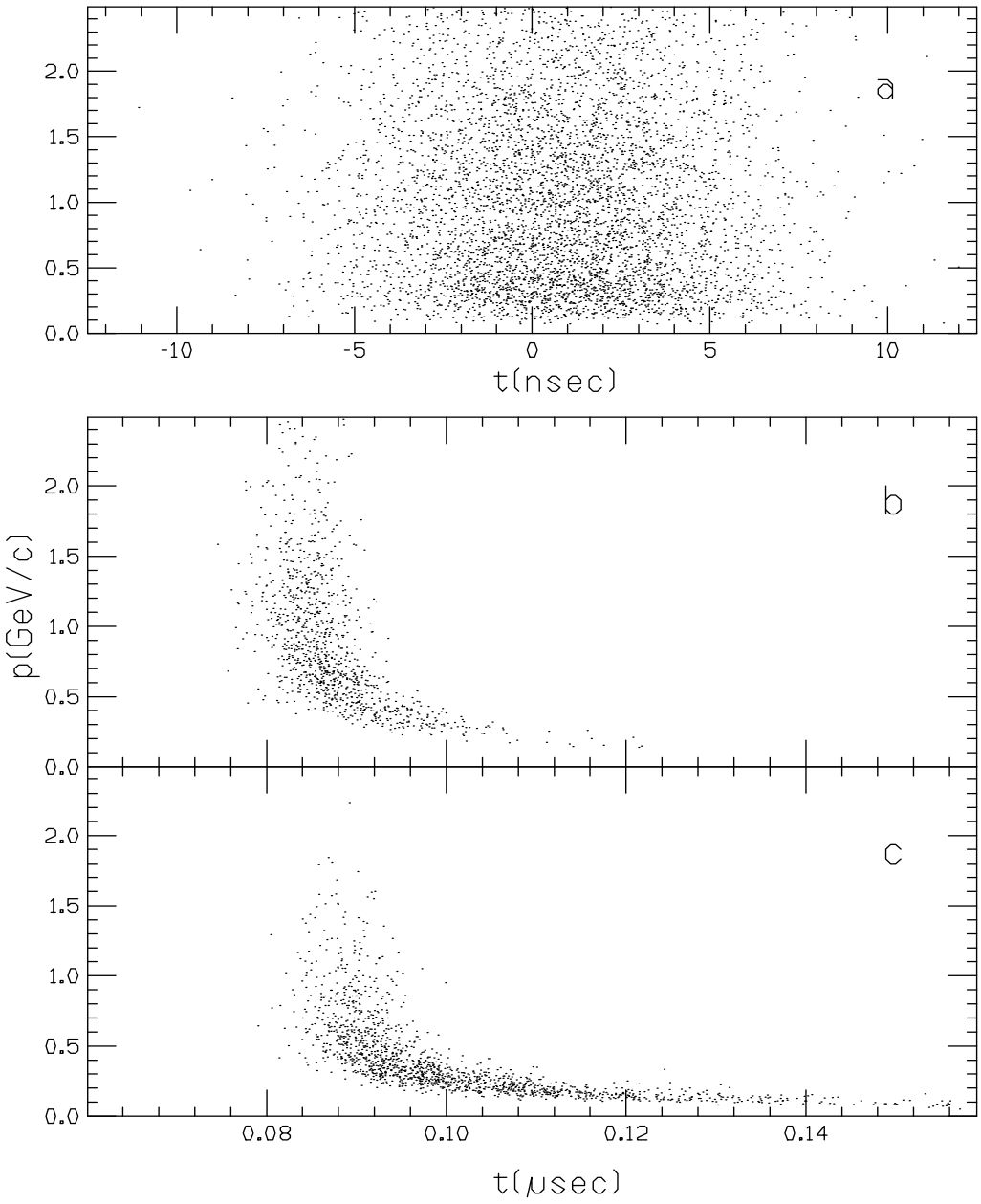


Figure 4.9: (a) π and K momentum *vs* time distribution immediately after the target for 8 GeV proton beam with $\sigma_t = 3$ nsec. (b) π and K distributions 25 meters downstream of target. (c) μ distribution 25 meters downstream of target.

Periodic Table and ranges in density from 1.8 to 22.4 g/cm^3. It is found that the optimal target radius needed to maximize the pion yield is about 2.5 times the rms beam size for all target materials and lengths, at both 8 and 30 GeV. This corresponds to a 1 cm radius target for the beam used in this study. Almost all studies reported here are carried out with this target radius.

Target length is varied from 0.5 to 3.0 nuclear interaction lengths (λ_I) for the different target materials. Fig. 4.10 shows the pion yield at the target exit for 8 and 30 GeV protons

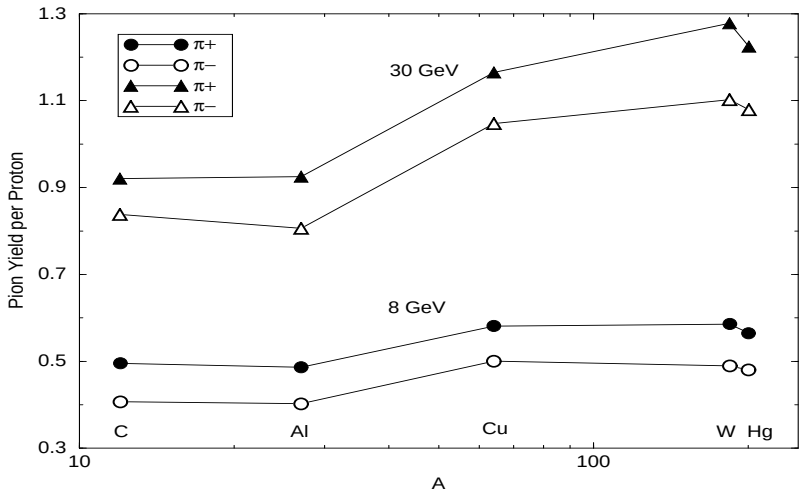


Figure 4.10: Pion yield from 1.5 λ_I targets of various materials irradiated with 8 and 30 GeV protons *vs* target atomic weight for π momenta of $0.2 \leq \mathrm{p} \leq 2.5\,\mathrm{GeV/c}$. Target radius r=1 cm, beam *rms* spot size σ_x=σ_y=4 mm.

as a function of target atomic weight. The π^+ and π^- yields versus target length for light (C), medium (Cu) and heavy (Hg) materials are presented in Fig. 4.11 for 30 GeV proton beam. The optimal target length is about 1.5 λ_I at 8 GeV, but yields vary by no more than 10% over a range of 1 to 2.5 λ_I. Yields are rather insensitive with respect to target composition at 8 GeV. At 30 GeV the optimal length for high–Z materials is about 2 λ_I but yields are only 10% lower for a 1.5 λ_I length. The yield for carbon is roughly constant in the range 1.5 to 3.0 λ_I. Note that pion yields for all materials approximately double from 8 to 30 GeV. Similar behavior is found with ARC though the yields increase by a factor of 2.5 from 8 to 30 GeV.

An alternate way to increase pion yield is to use projectiles heavier than protons [10]. Pion yield is proportional to nucleon number at a given momentum per nucleon. Since yield rises less than linearly with momentum, a gain occurs by using a heavy projectile at the same momentum as for the original proton. Deuterons and tritons are prime candidates since they have the same electric charge as the proton, making energy loss in the target about the same. Comparing 30 GeV/c tritons and protons, the increase in pion yield (for kinetic energies of 0.05 to 0.75 GeV) is about a factor of two. Although tritium is a low-energy beta emitter with a 12.3 year half-life, the total amount needed in a year of muon collider operation is

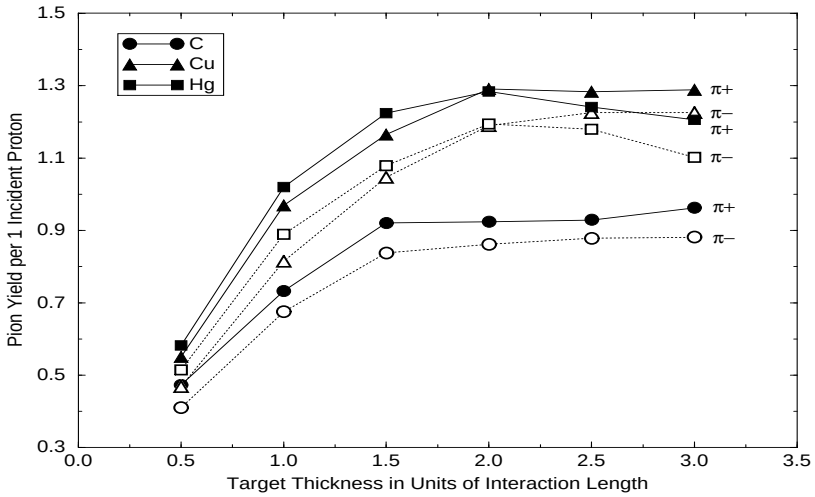


Figure 4.11: Pion yield per 30 GeV proton for 1 cm radius targets of various materials *vs* target length for π momenta of $0.2 \leq \mathrm{p} \leq 2.5\,\mathrm{GeV/c}$.

less than 0.25 gram.

Target Heating

Beam power deposited in the target varies greatly with composition due mainly to increased electromagnetic shower development in high–Z materials. With 5×10^{13} protons at 8 GeV on a 1.5 λ_I, 1 cm radius target, the average power dissipation at 30 Hz ranges from $0.39\,\mathrm{kW/cm^3}$ in carbon to $7.6\,\mathrm{kW/cm^3}$ in iridium (see Fig. 4.12). A single pulse peak energy deposition (on axis) in the target ranges from 20 J/g (C) to 35 J/g (Ir) at 8 GeV. This is at least a factor of ten below the shock damage limit. These values would double for the reference design of 10^{14} protons per pulse (3×10^{15} protons per second). For forced water cooling of solid targets, the maximum surface heat flux (ϕ_{max}) that can be practically removed is about 200 $\mathrm{W/cm^2}$. This implies a maximum target radius r=$2\phi_{max}/P$ where P is the average power density in $\mathrm{W/cm^3}$. Hence at 3×10^{15} protons per second, a 1 to 2 cm radius carbon target appears a viable candidate with adequate cooling. Heavier targets probably need to have a much larger radius at this beam intensity to lower the power density. Alternatively, at high power densities one may resort to 'microchannel' cooling wherein target wires are interspersed with small diameter cooling channels or recirculating liquid targets (mercury, lead or gallium).

Target heating becomes much worse in high–Z targets at 30 GeV and 1.5×10^{15} protons per second. Fig. 4.13 shows the power dissipation in carbon, copper and mercury targets, which

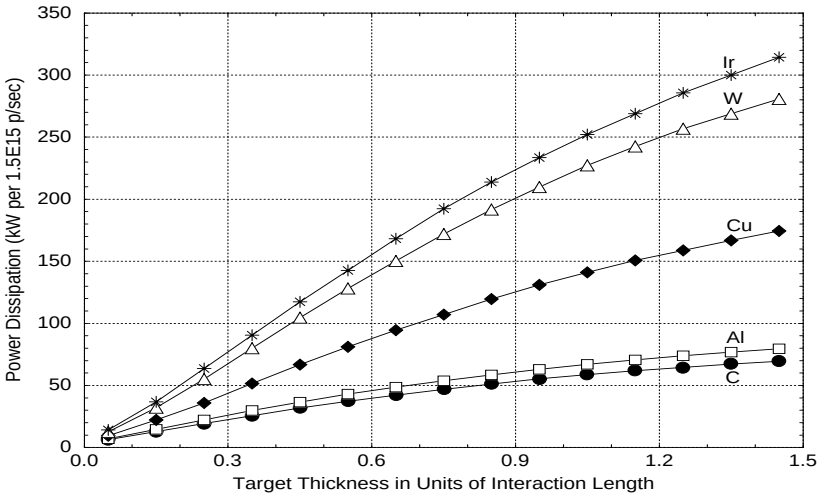


Figure 4.12: Average power dissipation in different 1 cm radius targets due to 8 GeV incident beam of 5×10^{13} protons at 30 Hz. Beam *rms* spot size σ_x=σ_y=4 mm.

is significantly higher compared to the 8 GeV case. A single pulse peak energy deposition ranges from 60 J/g (carbon) to 200–240 J/g (heavy targets) and rapidly reaches and exceeds the melting point in a sequence of pulses for any solid target except carbon. This suggests that a wider beam or recirculating liquid targets are necessary. Solid carbon targets of radius 1 to 2 cm are still acceptable at this energy.

Starting with energy deposition distributions in targets generated by MARS, thermal and stress analysis have been performed with the ANSYS code [11]. Ideal cooling with Δ T=0 at r=1 cm is assumed. When irradiated with 1.5×10^{15} protons per second at 8 GeV, equilibrium is approached in about one second in copper and carbon targets 1.5 λ_I long and 1 cm in radius. A steady-state temperature rise Δ T=T-T_0 relative to room temperature T_0=27°C reaches maximum of 347°C in copper and 186°C in carbon. These temperatures will roughly double for 3×10^{15} protons per second. Fig. 4.14 shows the temporal evolution of the maximum temperature rise Δ T=T-T_0 for 30 GeV protons in a copper target, while Fig. 4.15 is for a carbon target. One sees that even with an ideal cooling the core of a copper target will be melt in about one second. The situation is similar with all the studied solid targets (Al, Cu, W, Ir), except graphite which easily survives at the same beam parameters with a maximum steady-state temperature of $\leq$ 250°C.

The calculated equivalent stress map in a copper target at 30 GeV after the first pulse of 5×10^{13} protons is shown in Fig. 4.16. The maximum equivalent pressure is 163 MPa (6894 Pa=1 psi). In one second the peak stress reaches $\sim$1.5 GPa, which according to Fer-

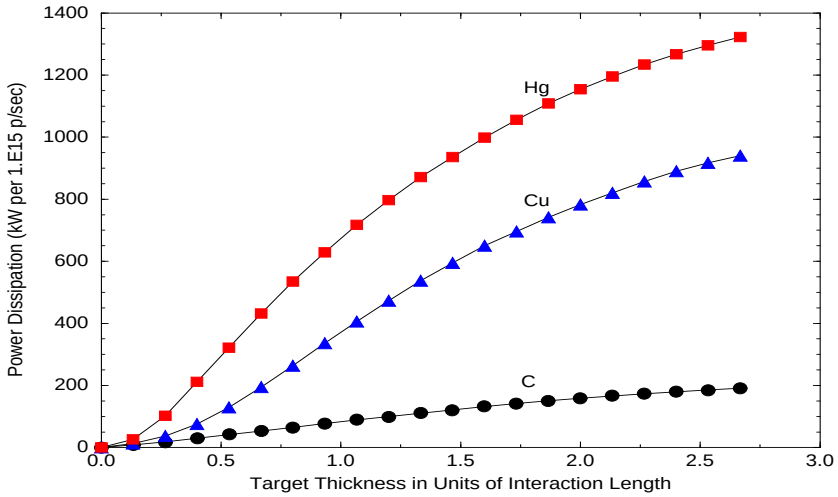


Figure 4.13: Average power dissipation in different 1 cm radius targets due to 30 GeV incident beam of 5×10^{13} protons at 30 Hz. Beam *rms* spot size σ_x=σ_y=4 mm.

miLab antiproton source experience is right at the limit of what can be accommodated in a copper target. For tungsten alloys and all other solid targets except carbon, the situation is even worse. Another problem is the expansion of solid targets. For 30 GeV protons on a $2\lambda_I$ (30 cm) long, 1 cm radius copper target, Fig. 4.17 shows the maximal deformation along the target axis and in the radial direction at shower maximum as a function of irradiation time.

Maximal longitudinal and radial displacements after one second are $\sim$1.3 mm and $\sim$ 0.05 mm, respectively.

There are then three possibilities for targetry at 30 GeV:

- medium or heavy solid targets – copper, tungsten alloys, iridium – using special techniques to reduce a peak power density (beam sweeping, target rotation, larger beam spot size);
- graphite target with a very low power dissipation (Fig. 4.13) and power density/temperature rise (Fig. 4.15) and very good thermal properties; pion yield is lower by only 40% compared to copper and heavier targets;
- liquid metal targets (gallium, mercury or lead) which recirculate to remove heat.

Target Options

In order to compare performance of different target options, a scoping survey was made of a variety of possibilities. Solid targets ranged from high–Z (high melting point tungsten

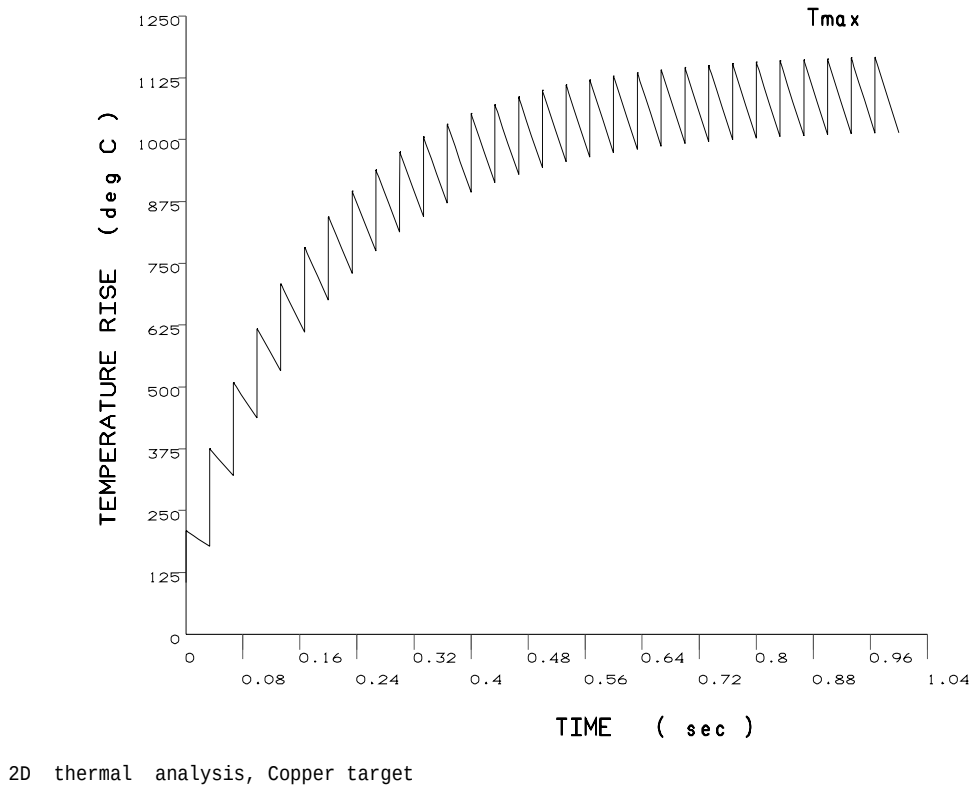


Figure 4.14: Maximum temperature rise Δ T relative to room temperature T_0=27°C in a 1 cm radius $2\lambda_I$ long copper target when irradiated by 30 GeV beam of 5×10^{13} protons at 30 Hz, as calculated with MARS-ANSYS.

(W) and, more easily machined, tantalum (Ta)) to medium–Z (copper (Cu) and low thermal expansion iron-nickel alloys (SuperInvar)) and to low–Z (nuclear grade graphite (C)). A broad set of liquids was considered for liquid targets as well as forced cooling of solid targets. Many coolants are rejected because their melting points are too high (*e.g.*, Sn70Pb30) or because of high chemical reactivity (NaK, molten salts). The best coolants include water and low-melting point alloys of gallium (*e.g.*, GaInSn). The main disadvantages of water are its high vapor pressure since it can permeate the vacuum system if there is a leak and the need for pressurization if boiling is a concern. Gallium's main drawback is the possibility of a large MHD pressure drop if it is forced at high speed across the strong magnetic field ($\geq$ 20 T) of the collector solenoid. Combinations of these materials were studied in six distinct target configurations, described below. The calculations are analytic with simplifying assumptions (typically, uniform power density of heating over a specified volume). This initial survey is only intended to elucidate the relative strengths and weaknesses of target options.

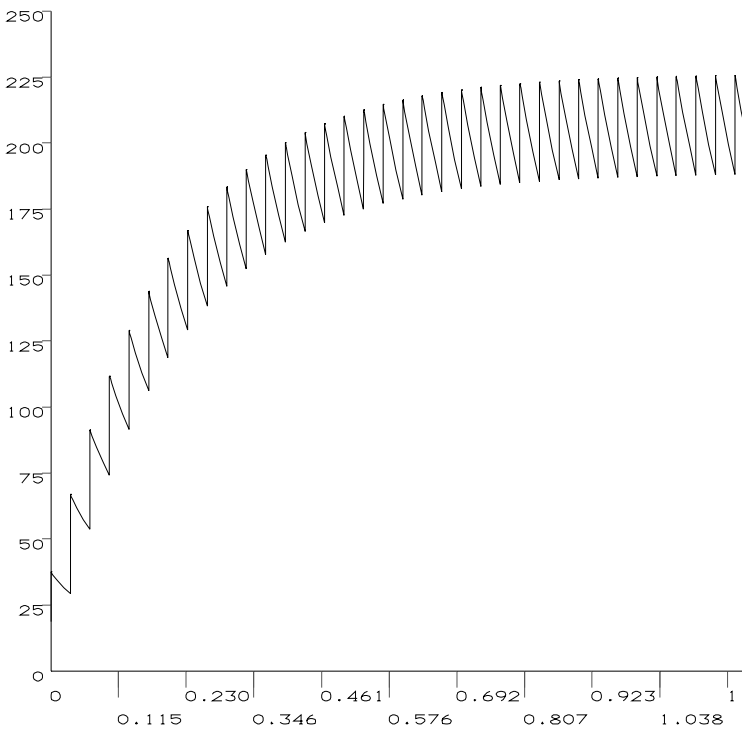


Figure 4.15: Maximum temperature rise Δ T relative to room temperature T_0=27°C in a 1 cm radius $2\lambda_I$ long graphite target when irradiated by 30 GeV beam of 5×10^{13} protons at 30 Hz, as calculated with MARS-ANSYS.

The target length for various materials was assumed to be $1.5\lambda_I$ and the (uniform) linear power density was specified as: C - 188 kW/0.57 m; Cu, SuperInvar, Ga - 900 kW/0.23 m; Ta, W - 1430 kW/0.22 m. These densities are roughly 50% higher than the values for the 30 GeV case in Table 4.1 to allow for an added design margin. Unirradiated material properties are assumed for simplicity. The following target configurations have been considered:

1. Solid cylinder target – surface cooled. Analysis confirms that graphite is the only viable solid option in this geometry, which is due to the relatively low heating power density. However, a large radius graphite cylinder of 12 cm is suggested which is readily cooled by water flowing through the solenoid bore at 3 m/s. This radius is larger than the present solenoid design. There is a film drop $\Delta T_f = 44°C$ across the target-coolant interface, so for a coolant inlet temperature of 40°C the water temperature will approach 85°C at some locations; pressurization may be needed to avoid local boiling/hot spots. An additional concern in this geometry is that the center of the graphite may rapidly exceed 350°C if there is a loss of flow accident (LOFA) while the beam is still operating.

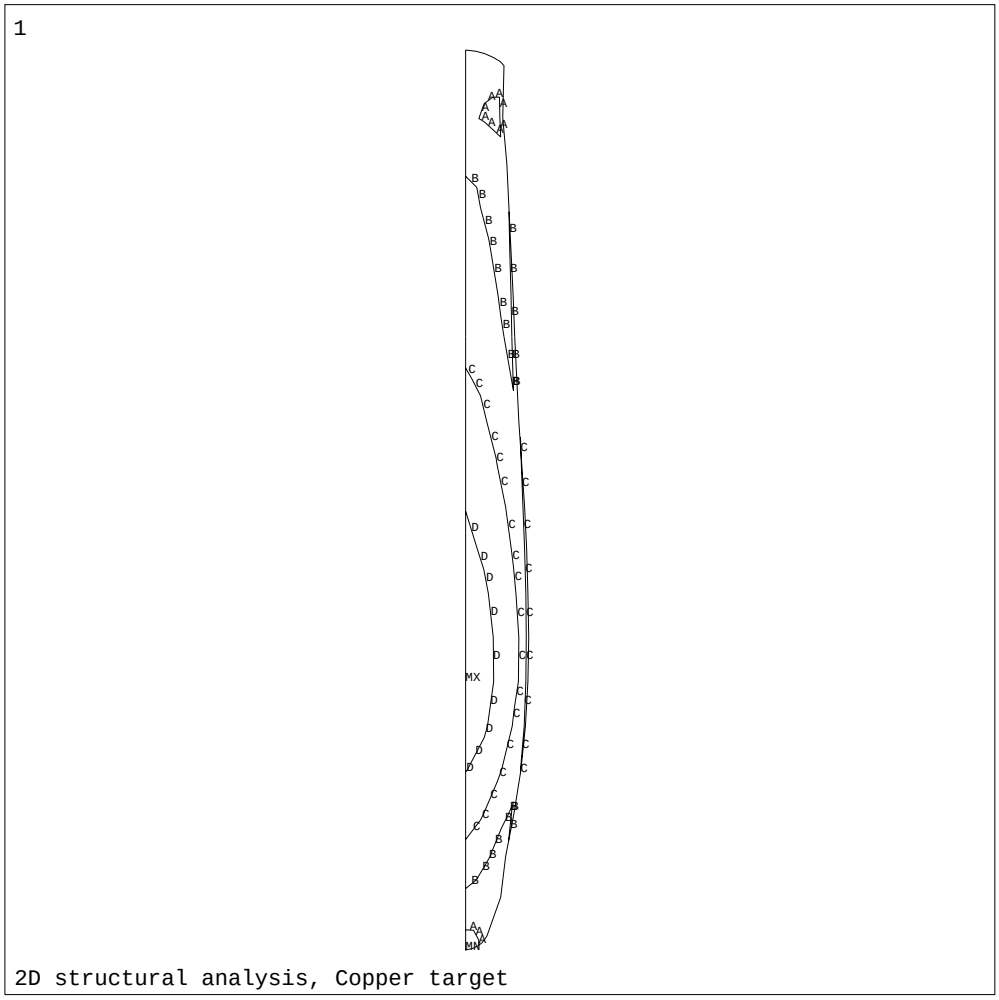


Figure 4.16: Stress isocontours (Pa) in 1 cm radius 30 cm long copper target after a single pulse of 5×10^{13} protons at 30 GeV, as calculated with MARS-ANSYS. Maximum value is 163 MPa.

2. Solid target – core cooled with narrow channels. Again graphite performs well, but even high–Z materials are satisfactory for steady-state heat removal. In this geometry a long bar (*e.g.*, graphite, 57 cm long) is the target, and the power density is kept modest by using a large cross-section (14 cm $\times$ 14 cm). Coolant channels (1 mm radius, spaced 7.5 mm apart) are drilled across the bar. This close spacing of coolant channels guarantees a very small temperature variation through the target ($\Delta T \approx 1^\circ C$ for graphite). Such small ΔT values minimize thermal stress and increase the lifetime against thermal fatigue. Water supplied at a volumetric flow rate of 710 gpm (41 liter/s) is adequate to provide 10 m/s flow through the channels, with a small film drop (1.6°C) and a small inlet/outlet temperature rise ($\Delta T_{io} = 1^\circ C$) for the water. There is evidently little need to consider pressurized water for this application. Hole-plugging from various causes is a potential concern.

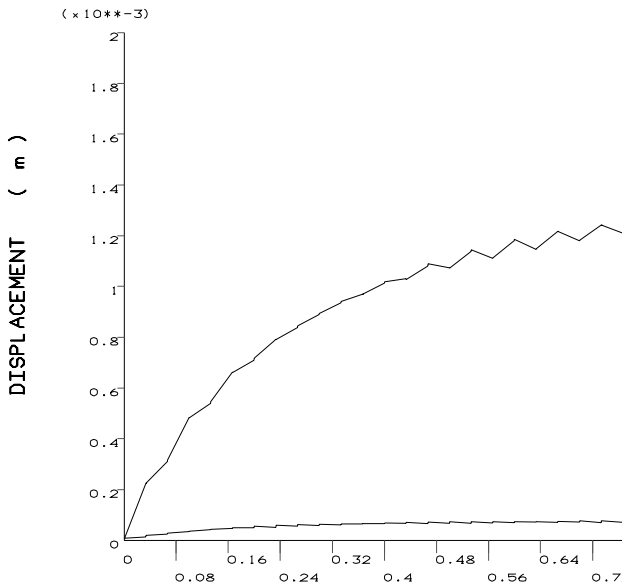


Figure 4.17: Time dependence of maximum longitudinal (upper curve) and transverse (lower curve) expansions in 1 cm radius 30 cm long copper target irradiated with 30 GeV protons of 5×10^{13} per pulse at 30 Hz, as calculated with MARS-ANSYS.

3. Solid target – cross-cut disks. An alternative core-cooled design has the target sliced into disks (*e.g.*, 1 cm thick with a 1 mm coolant channel between adjacent slices), and the proton beam passes through the whole stack of these disks. Such a design was proposed for IPNS-U, a target with specifications similar to the muon collider. This option performs similarly to the narrow-channel drilled design.

4. Liquid target – water-cooled metal in tank. A cylindrical tank (8 cm radius and 23 cm long) of stagnant gallium is inserted into the solenoid bore and cooled by concentric loops of flowing water, immersed in the liquid metal. The liquid target circumvents the fatigue life limits associated with solid targets and also simplifies maintenance, as the target and coolant lines may be readily drained and purged. Even without convection the Ga peak temperature remains less than 980°C, far below its boiling point (near 2070°C). A gas pocket and vent are provided to allow free expansion of the Ga in the event of shock generation from the beam pulses. A disadvantage of this design is the complexity. Inner and outer coolant loop headers are needed to provide single-pass parallel water flow paths as the film drop is high (53°C) and the single pass $\Delta T_{io} = 7°C$, with a total of thirty coolant loops (5 mm inner radius) carrying 13 m/s water. An additional concern is a possible requirement to pressurize the water lines to avoid boiling/burn-out.

5. Liquid target – self-cooled metal in tank. The target is simplified if a single liquid serves both as the beam target and as the heat removal agent. Water cannot be ruled out as a target option, but, like organic coolants, it has a relatively low boiling point and suffers decomposition from radiolysis. Flowing liquid Ga does appear to be attractive. In this design inlet and outlet headers flow the liquid parallel to the solenoid's magnetic field in order to minimize the MHD pressure drops. Within the steel tank (10 cm radius, so a larger solenoid is again required) the Ga flows a short distance across the magnetic field at a slow speed (0.041 m/s) calculated to minimize the MHD pressure drop (0.044 MPa), yet adequate to keep the Ga outlet temperature less than 380°C. Further optimization should aim towards 200°C as a goal, to minimize liquid metal corrosion concerns. The volumetric flow rate of Ga, 17 gpm, is roughly that achieved for the silicon diffraction crystals of the APS at Argonne National Laboratory. A thin ceramic coating (*e.g.*, 10 μm Al_2O_3) inside the steel tank will essentially eliminate the MHD pressure drop, permitting larger flow rates and an even smaller temperature rise in the coolant.

6. Liquid target – free-falling metal curtain. The previous designs all require the proton beam to pass through a window (*e.g.*, Be or Ti) which offers additional concerns regarding mechanical integrity and activation. Liquid metals, like gallium, have such low vapor pressure that they might be left open to vacuum. In this scheme the inlet header provides liquid Ga to an upper trough which supplies a falling curtain of liquid or droplets to a collecting trough leading to the outlet header. The height and width of the falling shower (6 cm × 1 cm) is an adequate target cross-section area for a proton beam with σ_x=σ_y=0.4 cm. Issues here include MHD effects on falling Ga in a strong field and the possible generation of high velocity droplets from beam-generated shocks. The gravity head in the lower trough is adequate to allow flow rates of 17 gpm, with associated outlet temperatures of 380°C.

Shock and Fatigue Damage to Target

Fatigue is not an issue for liquid target designs (except for the windows) but is a problem for solid components. Graphite is certainly one of the best solid target material choices, considering that the target receives up to a billion high energy density pulses per year.

Shock resistance is a different property - the ability to survive a single pulse. A figure of merit which measures shock resistance is $R = (S_m C_p)/(E\alpha)$, where S_m is the yield stress, C_p is the heat capacity, E is Young's modulus, and α is the thermal expansion coefficient. If the peak energy density deposited in a short pulse is ϵ, then shock resistance is likely whenever $R \gg \epsilon$. Metals like Ta, W, and Cu have R in the range 20–40 kJ/kg, whereas graphite has $R = 900$ kJ/kg and SuperInvar has $R = 4500$ kJ/kg, showing the superiority of

these latter two candidates. By spreading the beam size to about 10 cm radius the peak beam energy density might be reduced nearly two orders of magnitude to values of $\epsilon = 2.8$ kJ/kg for SuperInvar and $\epsilon = 0.9$ kJ/kg for graphite. For these two materials $R \gg \epsilon$, as desired.

Liquid targets offer different shock-generation concerns – cavitation, erosion, and conduit cracking. Such concerns, if they are serious problems, might be mitigated by adding gas bubbles to the liquid. Further work is needed to analyze this as well as other constraints: target lifetime limitations (fatigue, shocks, corrosion, radiation swelling, etc.), target and solenoid activation, off-normal events (LOFA, leaks, crack response), and reliability and ease of maintenance. Whereas solid targets (C or SuperInvar) cooled with water appear viable solutions, it would appear that the simplest approach is to use the coolant itself as the target. If the MHD effects of liquid metals in strong magnetic fields can be easily handled, then Ga appears superior to water as the liquid target choice. Window lifetime is an issue for any closed-conduit liquid system; such concerns are minimized with a flowing Ga curtain open to the vacuum – a design which, however, presents its own unique problems.

Solenoid Heating

Energy deposition in the primary 28 Tesla solenoid resulting from the intense radiation environment around the target might cause quenching. Based on hybrid designs reported in the literature, the solenoid is nominally assumed to consist of a normal-conducting 'insert', starting at 7.5 cm radius, and a superconducting 'outsert' starting at 30 cm. The latter will quench if the heat load becomes excessive.

Fig. 4.18 shows average power density as a function of radius for 5×10^{13}, 8 GeV protons at 30 Hz on a 1.5 λ_I Cu target. The end of the target coincides with the end of the primary solenoid. Instantaneous energy deposition and power density are highest at the downstream end of the solenoid (see Fig. 4.19) and are lower everywhere for lower–Z targets.

For proton energies of 8 and 30 GeV, at 30 cm radius—where the superconducting solenoid starts—both peak instantaneous energy deposition and the power density are close but below the experimentally determined quench limits for FermiLab Tevatron dipoles (0.5 mJ/g and 8 mW/g, respectively). For 10^{14} protons per pulse the power density will approximately reach this quench limit for a copper target but not for carbon.

4.1.3 Pion and Muon Collection

Particles produced in the target are transported along a beamline which forms the first stage of muon collection and acceleration IN route to the collider. Attention must also be paid at this point to disposal of the other non-μ-producing particles, mostly nucleons, $e^{\pm}$, and

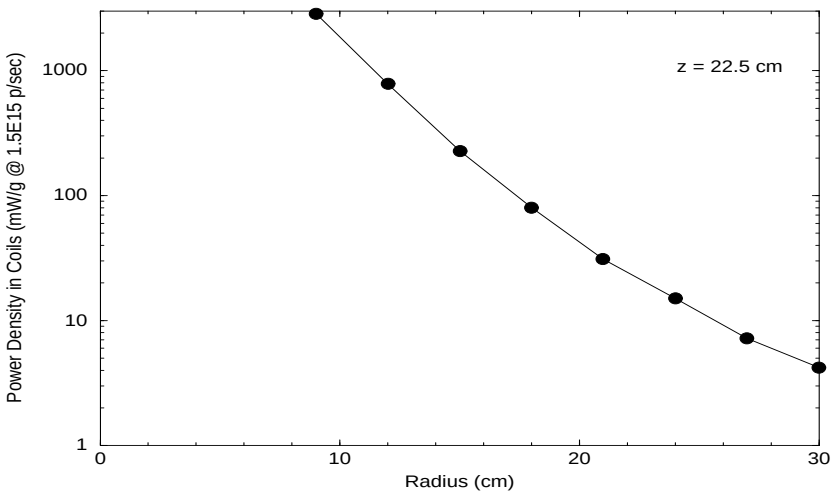


Figure 4.18: Power density in 28 T hybrid solenoid coils as a function of radius due to particle debris from 1.5 λ_I copper target irradiated by 8 GeV beam of 5×10^{13} protons at 30 Hz.

photons. At a minimum such a beamline involves some focusing of the produced pions and their muon progeny just to keep them from being lost on the walls. Then, when an optimum population is reached, the muons are cooled and accelerated. More ambitious schemes may begin cooling and/or acceleration earlier. This section concentrates only on the general pion/muon collection aspects of the post-target beamline.

For each proton bunch on target an intense pulse of mostly protons, electrons, pions, kaons and muons starts down the 7 T solenoid channel. Neutrals like photons and neutrons are unaffected by the magnetic field and are lost onto the walls according to their initial trajectory.

Fig. 4.20 shows particle densities as a function of time at the beginning of the decay channel. Total number of particles per proton of each species are indicated in parentheses. These distributions do not include the time spread of the proton beam. The latter—which depends on the design of the proton driver—is readily folded into the results of Fig. 4.20 at any stage in the simulation (prior to the RF cavities). When this pulse arrives at a debuncher cavity (proposed to reduce particle momentum spread) particles of the wrong sign relative to the electromagnetic wave would actually become bunched, with an increase in their momentum spread. Most such particles would quickly be lost downstream in any magnetic bend.

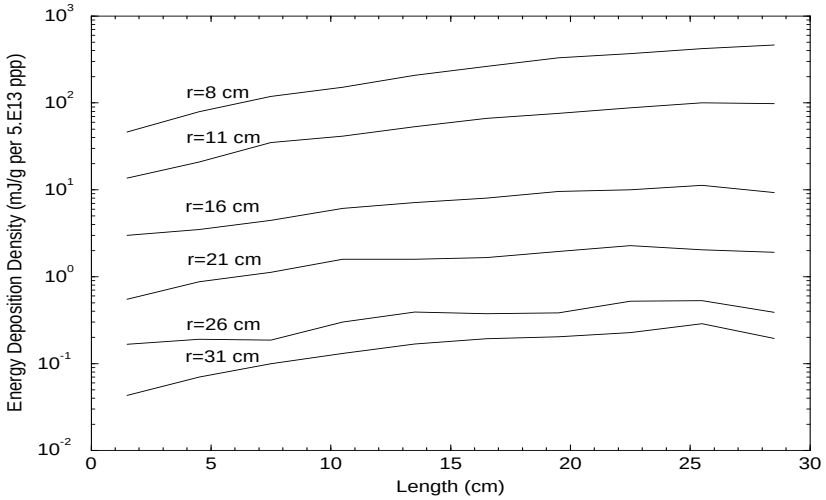


Figure 4.19: Energy deposition density at various radii of a 28 T hybrid solenoid as a function of a longitudinal position in case of 2 λ_I copper target irradiated by a single pulse of 5×10^{13} protons at 30 GeV.

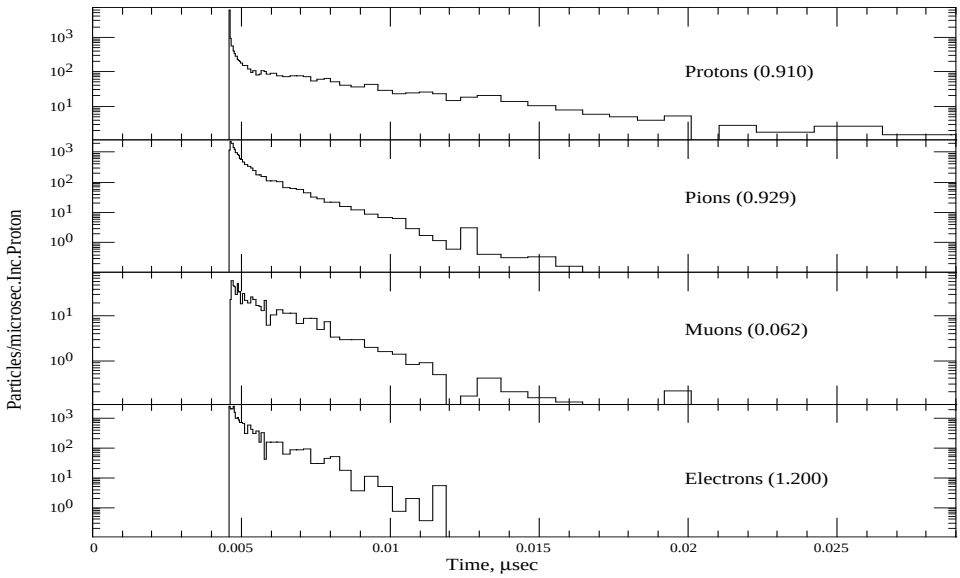


Figure 4.20: Particle densities as a function of time at end of matching region (1.15 m after end of target).

Particle Decay

Pions and kaons immediately begin decaying into muons downstream of the target ($\lambda_\pi = 56p$, $\lambda_K = 7.5\,p$, and $\lambda_\mu = 6233\,p$ where λ is in meters and p in GeV/c). Particles that do not intercept the walls in their first Larmor gyration typically are transported down the entire

7 T channel. The vast majority of lost particles are wiped out in the first 15 meters. This straight collection channel (without RF debuncher cavities in the simulation) is quite efficient with only 40% of all muon-producing particles lost on the walls and close to 60% yielding transported muons.

While decay is fully incorporated into the simulations, a few qualitative remarks may help interpret results. Only $\pi \to \mu\nu$ and $\mu \to e\nu\overline{\nu}$ decays are of real importance to this problem. Kaons are practically negligible as a source of muons in the present context: (1) their total yield is only about a tenth that of pions, (2) their branching ratios to muons are somewhat less favorable and (3) the decay kinematics produces muons typically with much larger $p_\perp$ than do pions. When they are included in a full simulation it is seen that only about 1% of all muons in the accepted phase space are due to kaons.

As a function of distance traversed along the pipe, z, pions decay to muons at a rate

$$dN_\pi/dz = \frac{1}{\lambda_\pi} e^{-z/\lambda_\pi} \tag{4.4}$$

where $\lambda_\pi = p_z^\pi \tau_\pi / m_\pi$ and m_π, τ_π, and p_z^π are pion mass, lifetime, and momentum along the pipe axis. There is a similar equation for muons. From the decay laws of radioactive chains, the fraction of muons at z is given by

$$N_\mu/N_\pi = \frac{\lambda_\mu}{\lambda_\mu - \lambda_\pi} \left(e^{-z/\lambda_\mu} - e^{-z/\lambda_\pi} \right). \tag{4.5}$$

From Eq. (4.5) the maximum muon yield is realized at

$$z_{opt} = \frac{1}{\lambda_\pi - \lambda_\mu} \ln \frac{\lambda_\pi}{\lambda_\mu}. \tag{4.6}$$

To arrive at a more concrete (but approximate) estimate of z_{opt}, p_z^μ is replaced by its average value

$$\overline{p_z^\mu} = \frac{m_\pi^2 + m_\mu^2}{2m_\pi^2} p_z^\pi \simeq 0.785 p_z^\pi. \tag{4.7}$$

When inserted into Eq. (4.6) this results in

$$z_{opt} \simeq 251 p_z^\pi \tag{4.8}$$

in meters with p_z^π in GeV/c. Eq. (4.5) then indicates that at z_{opt} the number of muons per pion produced at the target is about 0.95.

For a *spectrum* of pions, optimization of z requires folding Eq. (4.7) with the p_z^π of the spectrum. But even without such a folding, a rough knowledge of the spectrum establishes a distance scale for the decay channel. It also follows that for a spectrum 0.95 μ/π must be regarded as an upper limit. Since at distances of order z_{opt} and beyond most pions have

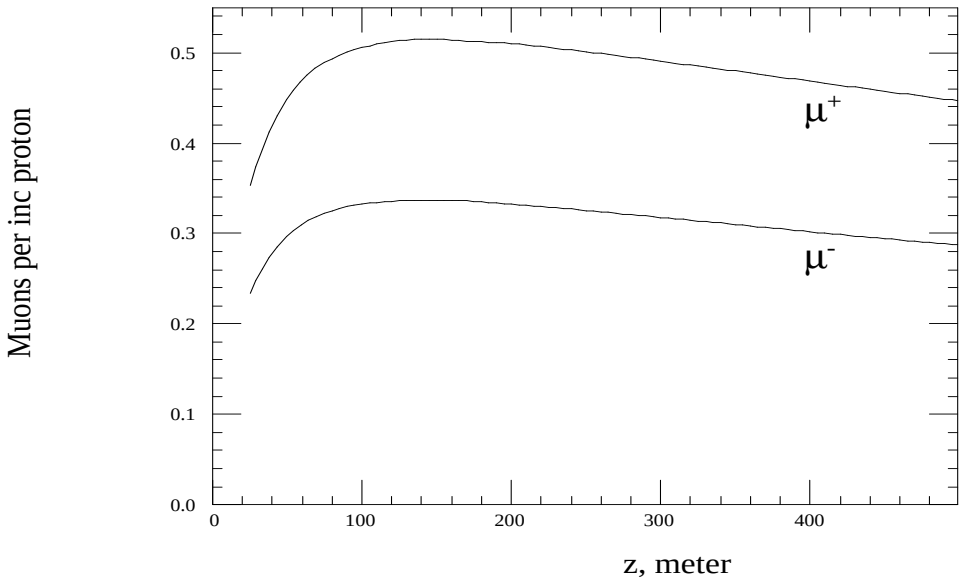


Figure 4.21: Muon yield vs distance from target for 22.5 cm copper target and standard straight decay solenoid.

decayed, the muon yield is governed by the muon decay length and one expects a broad maximum (at a z_{opt} corresponding roughly to the peak p_z^π of the spectrum produced at the target) where the theoretical maximum of 0.95 μ/π should be close to being realized. Taking 0.5 GeV/c as a characteristic pion momentum for the distribution, one expects the muon population to reach a maximum about 125 meters downstream of the target and fall off slowly after that. For a uniform distribution in the range 0.25–0.75 GeV/c a maximum 0.94 μ/π is attained at 130 m. Fig. 4.21 shows muon yield per proton versus distance from a simulation with a 22.5 cm long copper target.

In this case the maximum yield is 0.52 μ^+ and 0.34 μ^- per 8 GeV proton.

Curved Solenoid for Beam Separation

The proposed straight-solenoid plan uses two separate 8 GeV proton bunches (or four bunches at 30 GeV) to create separate positive and negative muon bunches accepting the loss of half the muons after phase rotation. In addition the debuncher cavities have to contend with a large population of protons, electrons and positrons that will tend to mask the desired π/μ bunches. Great advantage may be gained if the pions can be charge-separated as well as isolated from the bulk of protons and other charged debris before reaching the cavities. For the same number of proton bunches on target, a scheme which permits charge separation would produce a gain of a factor of two in luminosity. By coalescing the two proton bunches, this becomes a factor of four with little effect on target heating or integrity.

The solenoid causes all charged particles to execute Larmor gyrations as they travel down the decay line. As is well known from plasma physics, a gradient in the magnetic field or a curvature in the field produces drifts of the particle guide centers. Drift directions for this case are opposite for oppositely charged particles. Drift velocities depend quadratically on particle velocity components. This is exploited here by introducing a gentle curvature to the 7 T decay solenoid [3].

In the decay line, most particles moving in the curved solenoid field have a large velocity parallel to the magnetic field (v_s of order c) and a smaller perpendicular velocity ($v_\perp \simeq 0.3\,\mathrm{c}$ or less) associated with their Larmor gyration. In the curved solenoid the v_s motion gives rise to a centrifugal force and an associated 'curvature drift' perpendicular to both this force and the magnetic field. The field in the curved solenoid also has a gradient (field lines are closer near the inner radius than near the outer radius) resulting in an added 'gradient drift' in the *same* direction as the curvature drift. Averaged over a Larmor gyration, the combined drift velocities can be written as [12]

$$\vec{v}_R + \vec{v}_{\nabla B} = \frac{m\gamma}{q}\frac{\mathbf{R}\times\mathbf{B}}{R^2B^2}(v_s^2 + \frac{1}{2}v_\perp^2), \tag{4.9}$$

where $m\gamma$ is the relativistic particle mass, q the particle charge, and R is the radius of curvature of the solenoid with central field B. Note that in the present application the curvature drift ($\propto v_s^2$) is typically much larger than the gradient drift ($\propto v_\perp^2/2$). This is in contrast to a plasma where these contributions are comparable.

The drift velocity changes sign according to charge so positive and negative pions become transversely separated. For unit charge and for $\mathbf{R} \perp \mathbf{B}$ the magnitude of the drift velocity can be written in convenient units as

$$\beta_d = \frac{E\left(\beta_s^2 + \frac{1}{2}\beta_\perp^2\right)}{0.3RB}, \tag{4.10}$$

where E is particle energy in GeV, R is in meters and B in Tesla. The total drift displacement, D, experienced by a particle moving for a distance, s, along the field follows immediately from Eq. (4.10)

$$D = \frac{1}{0.3B}\frac{s}{R}\frac{p_s^2 + \frac{1}{2}p_\perp^2}{p_s} \tag{4.11}$$

with D in meters, B in Tesla and momenta in GeV/c. Note that only the *ratio* s/R appears in Eq. (4.11) which corresponds to the angle traversed along the curved solenoid. A typical 0.5 GeV pion ($p_\perp \ll p_s$) in a 7 T solenoid with R = 25 m has a drift velocity of about $10^{-2}c$. After moving 20 meters downstream in the solenoid, a 0.5 GeV positive and negative pion should be separated by about 35 cm.

The present study considers only *circularly* curved solenoids. Here the curvature and the $\nabla \times \boldsymbol{B} = 0$ condition requires the field, which is nonzero only along ϕ (i.e. along the axis of the curved beampipe) to have a $\frac{1}{R}$-dependence. This is readily incorporated in the detailed step-by-step simulations. Fig. 4.22 shows the pion and muon distributions 20 m

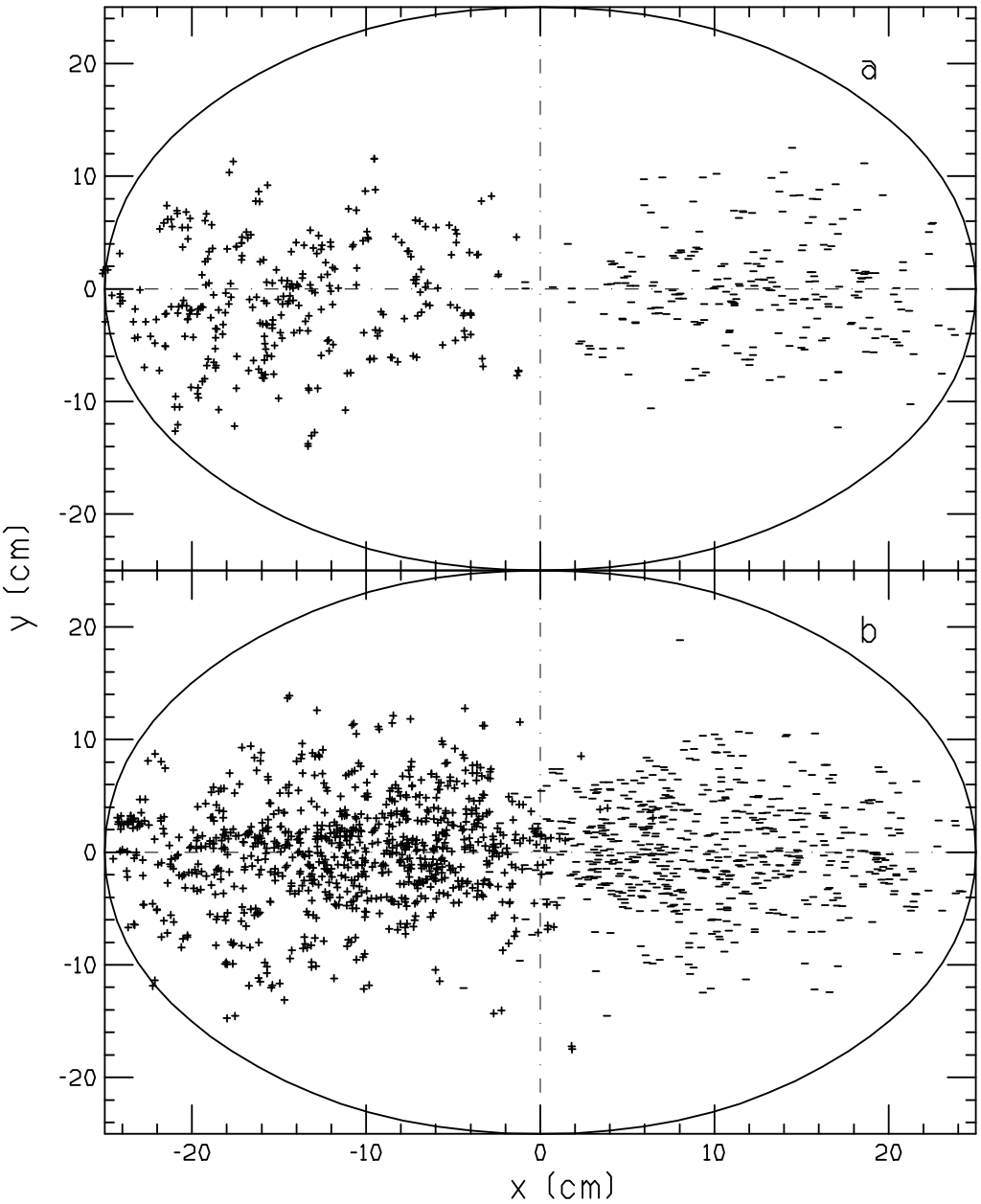


Figure 4.22: Position of (a) π^+ and π^-, and (b) μ^+ and μ^- 20 m downstream along curved solenoid (inner radius a=25 cm, R=25 m, B=7 T).

downstream of the target (which is in a 28 T field) calculated for a curved solenoid geometry. The centroid separation agrees well with what is expected from the drift formula. Also as

expected, higher energy pions are shifted farther and low energy pions less. Decay muons created up to this point are separated by a comparable margin. At this point one could place a septum in the solenoid channel and send the two beams down separate lines to their own debunching cavities. Note the pion and muon transverse beam sizes are not excessively large after charge separation.

The curved solenoid also serves well to rid the beamline of neutral particles and most of the remnant protons after the target. Neutrals like photons and neutrons intercept the curved solenoid with their straight trajectories and deposit their energy over a large area. Beam protons which underwent little or no interaction in the target have such large forward momenta that they are unable to complete one full gyration before intercepting the curved wall downstream. Hence Larmor-averaged drift formulae cannot be applied.

Fig. 4.23 provides some snapshots of this tight proton bunch (energy above 5 GeV) moving away from the lower energy protons at successive downstream locations. At four meters downstream all beam-like protons have intercepted the wall. Protons remaining in the pipe for long distances have momenta similar to the positive pions and thus will accompany them downstream. Roughly 0.7 protons per positive pion/muon are still in the pipe at 10 meters which should not overburden the debuncher cavities with extraneous beam. Electrons and positrons have typically much lower momenta than pions and muons. Simulations indicate that they do not drift far from the curved solenoid axis and most would be lost at the septum.

For the same diameter pipe the broadening associated with charge separation in the curved solenoid leads to increased particle losses on the walls compared with the straight case. An increase in pipe diameter is thus very desirable. To optimize the geometry with respect to yield would require many simulation runs. Since the curved regions may extend over long distances, this leads to much longer computation times for detailed step-by-step simulations to the point of becoming prohibitive when exploring a large parameter space. For survey type calculations a simplified procedure was therefore adopted.

The procedure adopted is then that for each pion encountered while reading a MARS file:

- the position vector of the Larmor guide center is determined;
- the drift displacement vector $\mathbf{D}$ is added to this vector, with appropriate sign, along the direction perpendicular to $\mathbf{B}$ and $\mathbf{R}$;
- it is then determined whether the entire Larmor circle fits inside the *half*-aperture appropriate to its charge, i.e., the side in the direction of the drift.

More precisely, this last condition is $r_g < a - r_L$ and $\pm x_g > r_L$ where r_L is the Larmor gyration radius, a is the pipe radius and subscript $_g$ refers to the final guide center position

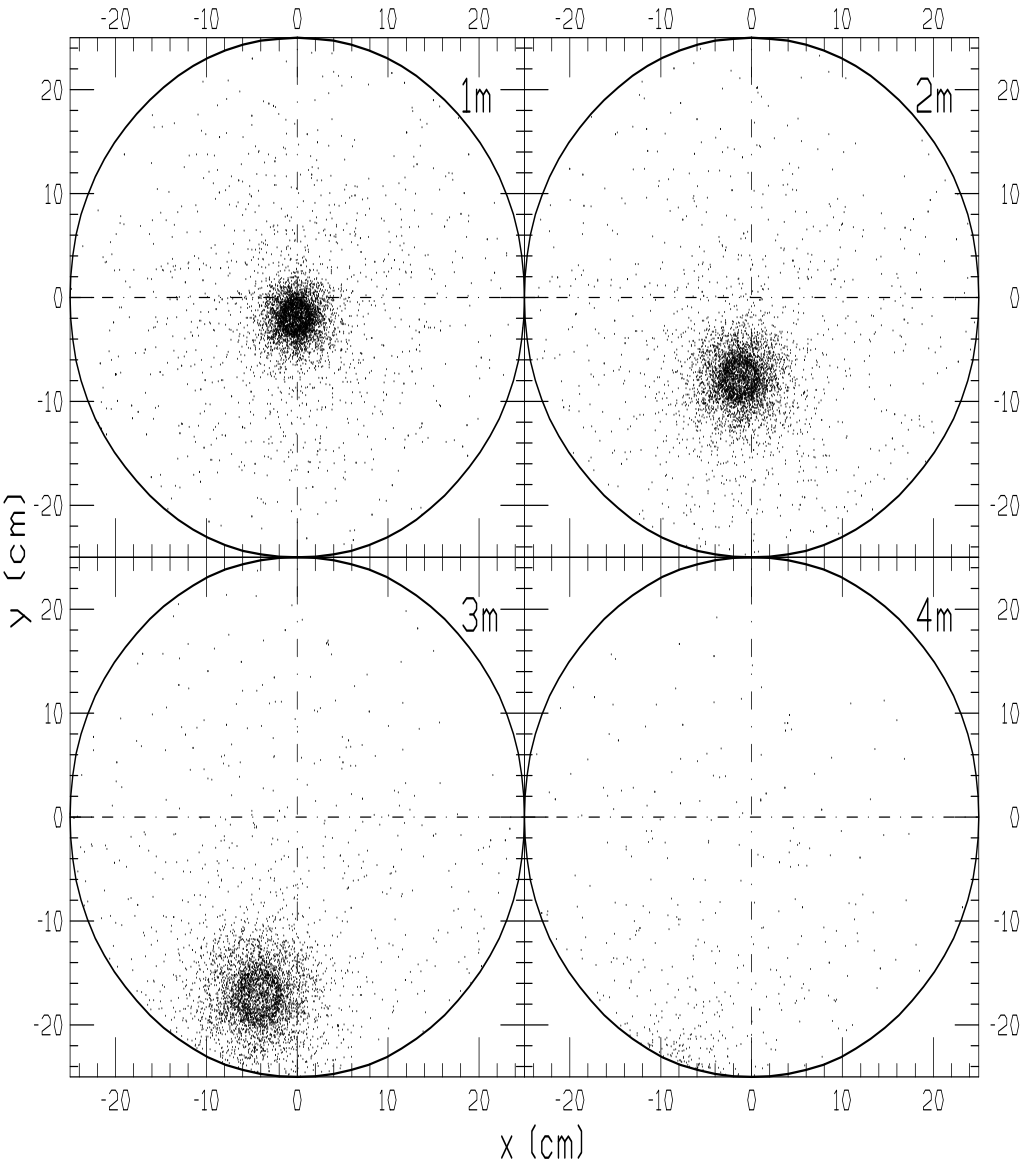


Figure 4.23: Scatter plot of x,y coordinates of protons above 5 GeV while traversing curved solenoid. Distance along center of curved solenoid is in upper right hand corner.

with the sign of x_g dependent on particle charge. If a pion meets these criteria it is assumed to contribute to the yield.

The study of pion/muon yield in a circularly curved beampipe (starting immediately after the target) with *constant* central field can thus be reduced to a problem with just three parameters B, s/R, and a. Some sensible range of values can thus be readily explored over a reasonably dense grid. An optimization based on yields alone is perhaps somewhat unrealistic. A measure of how effectively one can separate the two components into different beamlines is provided by computing the centroid of each distribution as well as its *rms* radius.

Computation is very fast and readily repeated for different sets of parameters to perform a more complete optimization.

A more realistic scenario starts with a 28 T field surrounding the target followed by a *curved matching region* which accomplishes simultaneously both transition to lower field and charge separation. The changing field causes an adiabatic decline in $p_\perp$ according to

$$p_{\perp f}^2 = p_{\perp i}^2 \frac{B_f}{B_i} \tag{4.12}$$

along with a corresponding change in p_s so as to conserve total p. Subscript $_i$ refers to initial and $_f$ to final values of B and $p_\perp$, i.e., those prevailing at S, the total distance along the central field line. Ignoring the other field components due to the declining field (cf. Eq. (4.3)), the total drift becomes

$$D = \int v_D dt = \int v_D \frac{ds}{v_s} = \int_0^S \frac{1}{300BR} \frac{p_s^2 + \frac{1}{2}p_\perp^2}{p_s} ds, \tag{4.13}$$

where now B, p_s and $p_\perp$ all depend on s. Assuming a *linear* decline of the central field $B = B_i(1 - as)$, and the dependence of p_s and $p_\perp$ on s this entails, one obtains

$$D = \frac{S}{300(B_i - B_f)R} p_0 \left[\ln \frac{(p_0 - p_{sf})(p_0 + p_{si})}{(p_0 + p_{sf})(p_0 - p_{si})} + p_{sf} - p_{si} \right], \tag{4.14}$$

where p_0 is the total momentum of the pion.

Thus for fixed B_i the problem remains confined to three parameters: B_f, S/R, and a. Note also that the Larmor radius changes with s here. Fig. 4.24(a) shows $\pi^+\mu^+$ yield in a curved solenoid with a constant 50 cm pipe radius for different values of the final magnetic field at the end of the matching region for a 22.5 cm long copper target. Eq. (4.13) can also be applied to a field having the $B_0/(1 + \alpha s)$ dependence. Again an expression for D, though somewhat lengthier than Eq. (4.14), is readily obtained and again the problem remains one of the same three parameters. For comparison, Fig. 4.24(b) shows $\pi^+\mu^+$ yield for the $1/(1 + \alpha s)$ field dependence with everything else as in Fig. 4.24(a). Note that the yields peak at somewhat smaller s/R.

Because of the advanced magnet technology required for very high field (> 15T) solenoids, it is of interest to investigate yields obtained when lower magnetic fields prevail throughout the entire geometry including the target. To keep matters simple a constant 7 T field and 50 cm solenoid radius is assumed—which might be considered state-of-the-art [13]. The solenoid is straight for the target portion, then curves to affect the desired charge separation, then straightens again to form the decay channel. In this last portion separation of plus and minus beams at a septum is to take place but details of this implementation are not

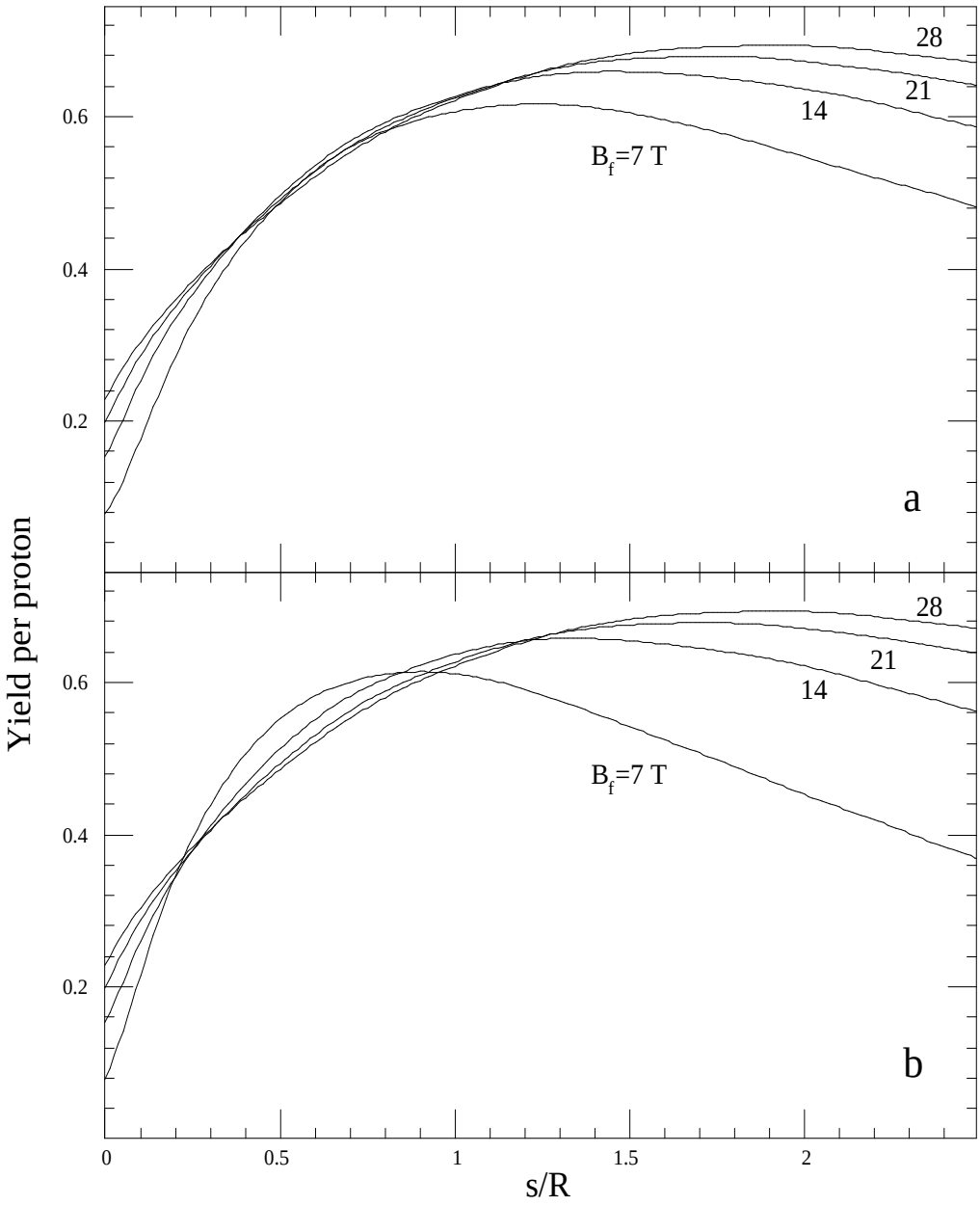


Figure 4.24: (a) Yield of positive pions and muons vs s/R for 22.5 cm copper target in straight solenoid with $B_0 = 28$ T followed by curved solenoid with $B = B_0(1 - \alpha s)$. Labels indicate final B reached at s/R. (b) Same for $B = B_0/(1 + \alpha s)$ in curved solenoid.

considered here. Such a solenoid has a transverse momentum acceptance of 0.52 GeV/c and a normalized phase space acceptance of 1.87 m·rad.

Fig. 4.25(a,b) show $\pi\mu$ yields for this type solenoid and for $1.5\lambda_I$ copper and carbon targets, respectively. Fig. 4.25(c) presents the yield curves for the copper target when the length is increased to $2\lambda_I$, suggesting longer targets to be better for this geometry. Yields

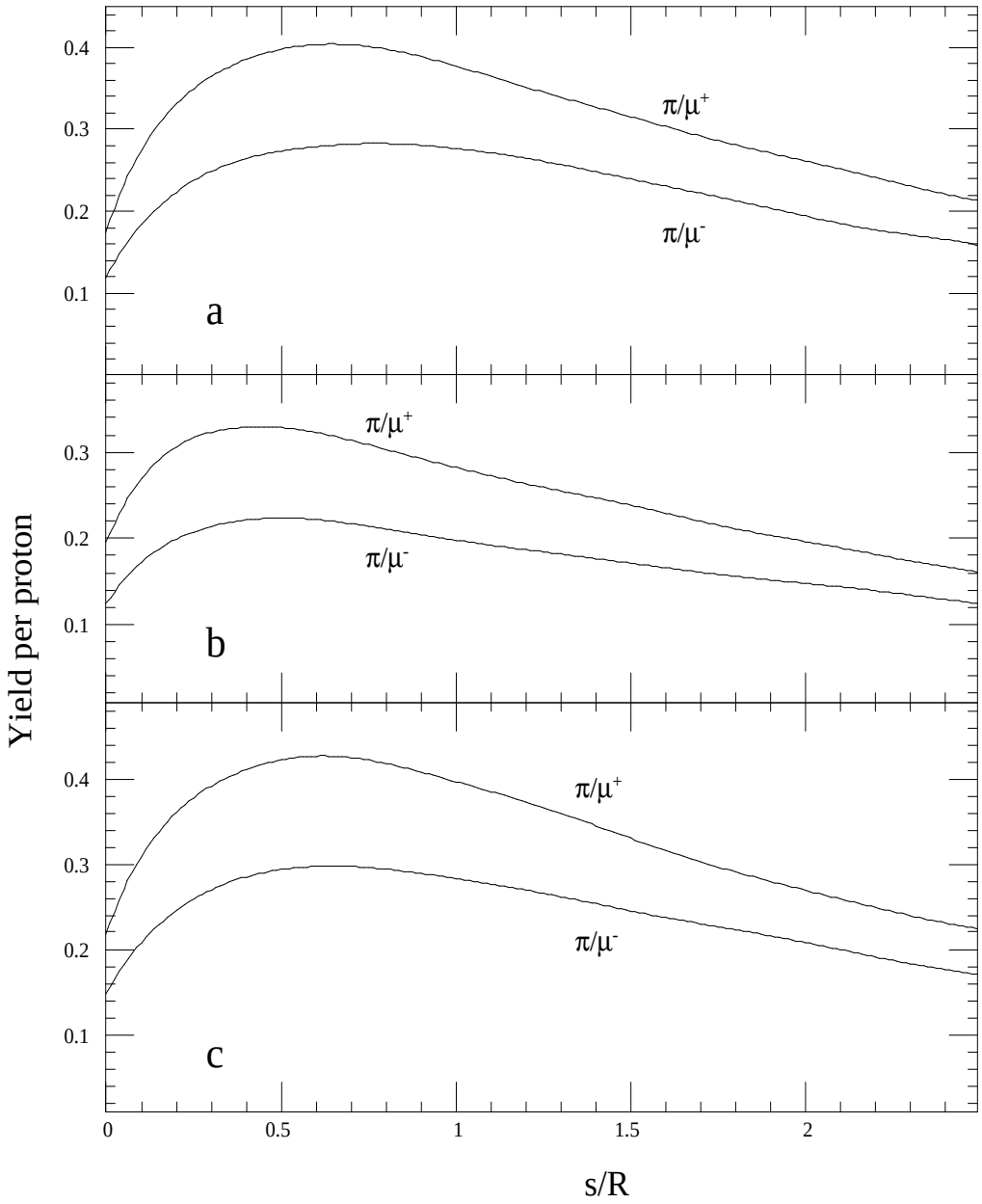


Figure 4.25: (a) Yield vs s/R for 22.5 cm copper target in solenoid of 7 T throughout; (b) same for 57 cm carbon target; (c) same for 30 cm copper target.

are presented as a function of s/R and it is thus advantageous to begin the straight (decay) portion of the pipe in the region near the maxima. Fig. 4.26(a,b) show respectively the centroid position of the plus and minus beams within the beampipe and their *rms* size for the standard copper target. The latter refers only to the distribution of the guide centers and excludes the spread due to the Larmor motion. Recall that both centroid and *rms* size refer only to those particles for which the entire Larmor circle fits inside the proper half-aperture.

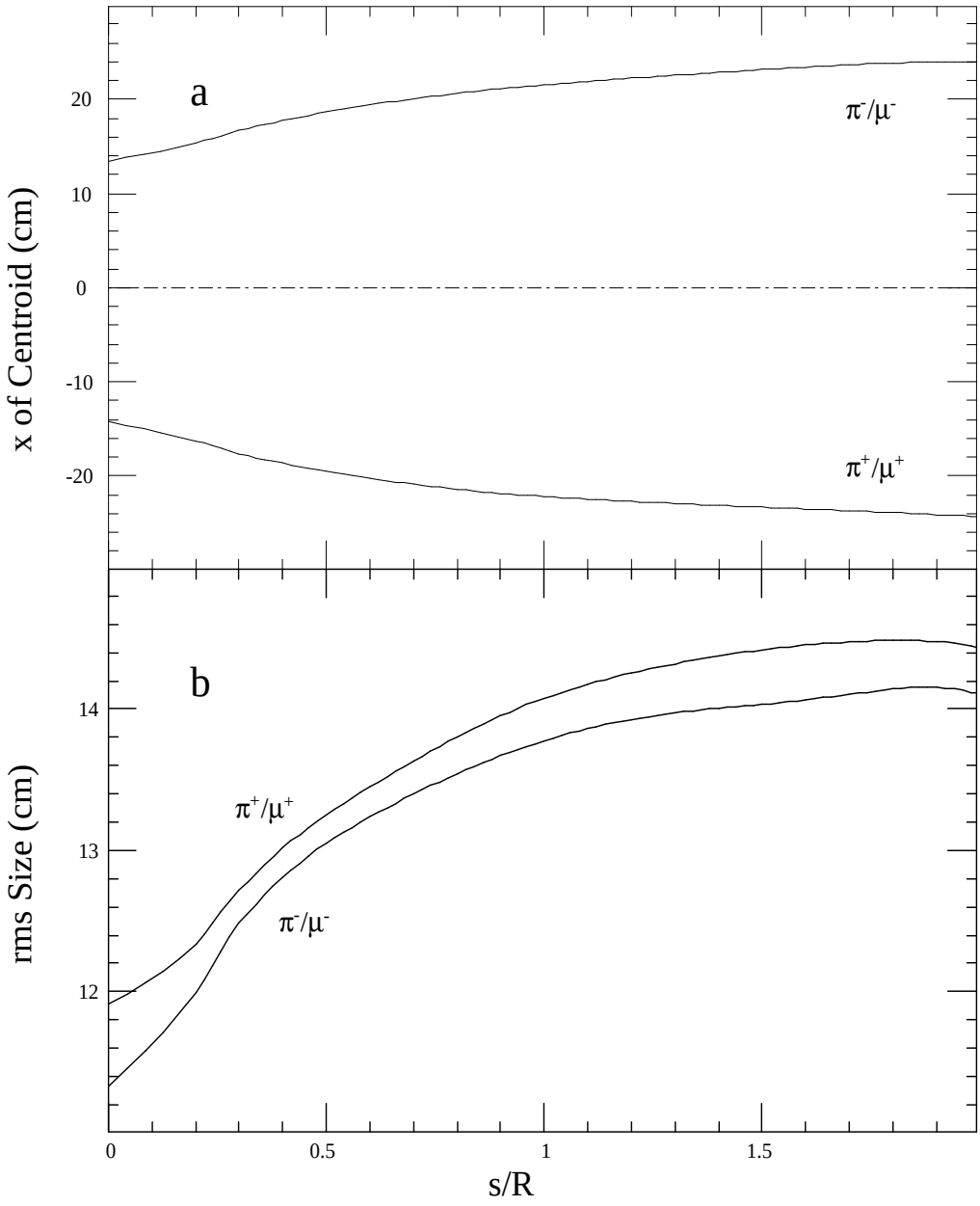


Figure 4.26: (a) Distance of centroid to magnet center vs s/R for 22.5 cm copper target, 7 T, r=50 cm solenoid throughout. Note beampipe extends beyond graph; (b) rms size of each beam vs s/R.

This accounts for the non-zero centroid positions at z=0.

4.1.4 Conclusions

The target studies show that while a 1 to 2 λ_I copper target is optimum for yield, lower-Z targets are not much worse—about 20% depending on the collection geometry for 8 GeV protons. Hence lower-Z targets, because of the lower energy deposition associated with them, may still be the targets of choice. A 30 GeV proton beam is not preferred on the basis of yield per megawatt of power deposited in the target, but may be needed to make short (1 nsec) bunches. The use of tritons instead of protons at the same momentum can increase pion yield per projectile on target by up to a factor of two.

Target heating is very severe in high Z materials at 30 GeV. Spreading the beam diameter to a large fraction of the solenoid bore generally helps by lowering the average heating power density and the shock energy density deposition. A variety of configurations appear to satisfy the steady-state heat removal target requirements. Microchannel cooling, large diameter beams and targets or recirculating liquid targets may be used to deal with the severe target heating problems in high Z targets. Solid carbon targets however are still workable with adequate cooling. In some of the considered target configurations the transverse size of the target assembly exceeds the 15 cm inner solenoid diameter. In subsequent designs some compromise is then required between the solenoid and target parameters. Lower energy proton beams at lower repetition rates (*e.g.*, 8 GeV, 15 Hz) would also help reduce target powers substantially.

Quenching due to energy deposition in superconducting solenoids near the target is a problem only for high-field/small-diameter magnets and high Z targets. Lower field solenoids with larger diameter are much less likely to quench and also pose less technological difficulties. While the yields associated with them are somewhat lower there certainly exists some reasonable trade-off.

The simulations confirm the superiority of muon collection with the solenoid scheme in this energy regime. Total yields of about 0.5 muons or more per proton of either charge appear to be obtainable. Considerations of π/μ decay indicate a collection limit of about 0.95 muons per pion. Kaons appear to contribute far less than their numbers to the usable muon flux and are practically negligible in this application. The pion momentum spectrum after the target generated by either 8 GeV or 30 GeV protons peaks in the range 0.2 to 0.3 GeV/c in MARS and DPMJET. The collection system with phase rotation tends to favor the lower energies and most muons are expected to be in the 0.2 to 0.5 GeV/c range as discussed in the next chapter.

Charge separation by *curved* solenoids practically doubles the number of muons collected and appears to be beneficial in disposing of the host of unwanted particles generated in the

target along with the through-going proton beam. Yields are sensitive to field strengths and solenoid diameter as well as to the s/R parameter—which indicates where to end the curvature and send the beams their separate ways. The curved solenoid is not part of the reference design, but this study indicates there certainly are ways to perform charge separation and beam disposal after the target.

4.2 Design Studies of Capture Magnet Systems

4.2.1 Introduction

The capture solenoid system includes the following components: 1) the 20 Tesla solenoid that captures the various particles that result from the proton beam interaction with the target; and 2) a transition solenoid system that transfers the captured particles from the 20 Tesla capture induction to a 5 Tesla transport induction that will be used through out the phase rotation process.

4.2.2 The Capture Solenoid System

The basic capture system consists of a high field solenoid to capture all pions with a transverse momentum of less than 225 MeV. The high field in the capture region changes to a lower field so that the cloud of particles, which comes from the target can be transported to and through the phase rotation system. Capture of the pions from the target can occur in a solenoid of any induction provided the solenoid bore is large enough[14, 15]. In order to make the capture bore and the bore of the phase rotation solenoids small enough, pion capture has to occur at a magnetic induction in the range of 20 - 30 T. The induction of solenoids used to transport the pions and muons in the phase rotation system has should be in the 3 to 7 T range.

The capture induction was set at 20 T. The selection of 20 T for the capture induction allows one to use a Bitter type water cooled solenoid alone or as part of a hybrid magnet system in conjunction with a superconducting outsert solenoid using niobium titanium. The hybrid magnet option was selected for the following reasons: 1) The operating power for the hybrid solenoid is substantially lower if a superconducting outsert is used. 2) The current density in the water cooled Bitter solenoid can be low enough to insure that its life time will be long (A reasonable life time goal might be 25000 hours.) 3) Additional space inside the Bitter solenoid can be made available for a heavy metal water cooled shield. This reduces the incident energy from the target into the water cooled solenoid and the surrounding

superconducting solenoid.

The inside radius of the capture solenoid is inversely proportional to the induction of the capture solenoid and is directly proportional to the transverse momentum of the particles captured from the target. The length of the capture solenoid is approximately proportional to the length of the target (in our case, about 1.4 to 1.6 times the length of the target). The most efficient production of pions comes from a target that is about 1.5 interaction lengths long. A carbon target would be about 600 mm long while a copper target would be about 225 mm long. High Z targets such as those made from mercury or gold would be even shorter. High Z targets will produce more pions, but they also absorb more of the incident energy from the proton beam. There is a trade off between pion production and the ability of the target to survive being struck by the incident proton beam.

The inner clear bore radius of the capture solenoid was set at 75 mm. At 20 T, particles with a transverse momentum of about 225 MeV/c can be captured. With proper target selection a large number of low energy pions will be created. Depending on the model used for particle production from the target, the number of pions produced by the target can range from 1.0 to 1.5 pions per incident 30 GeV proton on the target. Many of these pions will be at energies around 200 MeV. Most of these low energy pions will have a transverse momentum less than 200 MeV/c and thus will be captured by a 20 T solenoid with a clear bore radius of 75 mm. The length of the capture region for the solenoid magnet system described here is based on a 225 mm long copper target.

4.2.3 Water-Cooled Bitter Solenoid Insert

The water cooled Bitter solenoid insert shown in Fig. 4.27 has an inner radius of 120 mm. The extra inner bore radius allows a heavy metal (tungsten or some other high Z, high density metal) shield that is 30 mm thick to be inserted around the target region. An extra 15 mm on the inside of the solenoid is allotted for water cooling channels for the Bitter magnet. The Bitter solenoid insert outer radius has been set at 345 mm. The superconducting outsert solenoid cryostat starts at a radius of 370 mm. The 25 mm between the Bitter solenoid and the cryostat wall is allocated for water cooling. Much of the space at the ends of the Bitter insert solenoid should be filled with a water cooled heavy metal shield to protect the superconducting solenoids on the outside from uncaptured particles coming from the target.

The Bitter solenoid insert is divided along its length into three regions, each with a different current density[16]. The gradation of the current density in the windings allows for the magnetic field to be shaped along the solenoid axis, in the high field region. The current density in the Bitter solenoid section can be changed by changing the thickness of the plates

in the solenoid. The thick plates in a turn mean a lower the current density in that turn. Along the boundary where the current density changes, there is a change in coil strain. This may affect the overall dimensions of the solenoid as it is powered to full current while it is in the outsert coil field. The location of the Bitter solenoid sections is shown in Figures 4.27 and 4.30. The Bitter solenoid system is the same in both of these figures.

Tb.4.2 shows the dimensions and design current densities for the three sections of the Bitter solenoid. The highest current density section has an overall current density of 69.3 A mm^{-2}, which is a factor of two lower than the highest current density sections of some of the high field hybrid solenoids[17, 18]. Most of the magnet power is dissipated in the highest current density section. The highest current density sections are most prone to failure due to fatigues and strain of the Bitter coil conductor.

The cooling water must be properly treated to reduce corrosion as a factor in the reduction of magnet life time. The cooling system must handle about 12 MW (7.2 MW of resistive heating and 4.8 MW of beam power).

There appear to be a number options for the Bitter solenoid to increase its life time. These options include: 1) reducing the current density in the plates (Up to a point this will reduce the stress in the magnet.); 2) increasing the induction generated by the outsert solenoid (This would require the outsert solenoid to have niobium tin sections.); and 3) changing the alloy of the bitter coil plates to make them more resistant to corrosion and to make them stronger. Decreasing the current density of the Bitter solenoid will make it thicker (for a given induction produced by the Bitter solenoid). This has the advantage of decreasing the heating in the superconducting outsert solenoids from the target.

Liquid nitrogen cooling of the Bitter solenoid section could be considered. The potential advantages of a nitrogen cooled magnet are: 1) lower direct power consumption (by as much as a factor of six) with a smaller power supply for the magnet, 2) higher conductor strength and greater fatigue resistance, and 3) virtually no corrosion in the cooling channel. Liquid nitrogen cooling has several disadvantages: 1) More space is required for cooling because one is dealing with a boiling cryogenic fluid; 2) The input power requirements for the nitrogen refrigerator and the magnet power supply will be larger than the input power required for the water cooled magnet (about a factor of 2 higher); 3) The radial space required for a liquid nitrogen cooled solenoid would be larger for a given coil current density because there is a cryostat; and 4) the insert solenoid will be more costly. 5) At least some of the beam power will be absorbed in the Bitter solenoid, further increasing the input power to the refrigerator. It is probable that the disadvantages of a nitrogen cooled Bitter solenoid outweigh its advantages, so a water cooled insert solenoid has been chosen for the preliminary design.

4.2.4 Superconducting Outsert Solenoid

The water cooled Bitter insert solenoid is surrounded by a superconducting outsert solenoid system. The combination of the Bitter magnet insert and the superconducting outsert will generate a magnetic induction of 20 Tesla in the region of the target. For our purposes, the superconducting outsert is defined as the three coils that are immediately outside of the Bitter magnet insert. Other superconducting solenoids downstream from the outsert coils form the transition region that shapes the magnetic field between the target and the phase rotation system. The transition region solenoid can share a common cryostat with the superconducting outsert solenoids. The space inside the outsert solenoids at the ends of the Bitter insert solenoid must be filled with water cooled energy absorbers to absorb energetic particles that come from the target but are not captured by the capture solenoid system. The energy absorber must be thick enough to prevent heating of the superconductor in the outsert and transition region coils. The energy absorber must extend some distance downstream from the target. As a result, the outsert solenoid cryostat can start at a radius of 370 mm.

The outsert coil alone is designed to produce a central induction of over 6 Tesla. Operating under this condition, the outsert solenoid can be made from niobium titanium in a copper matrix operating at 4.2 K. The outsert solenoid has three coils with an inner radius of 400 mm. The outer radius of these coils is about 540 mm. The peak magnetic induction in the outsert coil region is about 6.9 Tesla when the insert coils are fully powered. The peak field in the outsert coil system at full design current goes down to 6.4 Tesla when the Bitter insert coils are not powered. Tb.4.3 shows the coil dimensions, average coil current density and peak magnetic induction in the coil for the three outsert coils labeled coils 4, 5 and 6.

In Tb.4.3, the current density in coils is kept constant. The outer radius of the outsert coils could be kept constant and the current density can be varied. Correction coils can be used to correct the field leaving the target region into the pion transfer section. It should be noted that the current density given for the coils is the average current density including helium and insulation. The matrix current density in the winding is about 1.5 times the average coil current density given in Tb. 4.3.

The superconducting outsert solenoids are shown in Figures 4.27 and 4.29. The outsert solenoid system consists of the three superconducting solenoids mounted outside the Bitter solenoid insert system. In both figures the outsert superconducting solenoids are the same. A change in the current density in the third outsert solenoid (the one down stream from the target) may be used to help shape the field downstream from the target where the solenoid induction drops from 20 T to 5 T before the pions are transported to the phase rotation

system.

4.2.5 The Transfer Solenoid System

The transfer solenoid system guides the magnetic field so that the magnetic induction decreases in a smooth fashion from 20 to 5 T at the start of the phase rotation channel. A smooth transition of the magnetic induction can be achieved if the magnetic induction B(z) as a function of the distance from the end of the target z has the following relationship;

$$B\left(z\right) = \frac{B_0}{z \ + \ A} \tag{4.15}$$

where B_0 and A are fitting parameters. If one wants to get the magnetic field down from 20 T to 5 T in 1.2 m, B_0 = 8 Tm and A = 0.4 m. Once the magnetic induction has reached the design value for phase rotation or pion beam transport, it is desirable for the induction to remain at a constant value.

The inside radius of the beam pipe increases as the induction decreases. The minimum radius for the beam pipe r(z) at a distance z from the end of the target, can be calculated using the following relationship;

$$r\left(z\right) = \sqrt{\frac{B\left(0\right)}{B\left(z\right)}} r\left(0\right) \tag{4.16}$$

where r(0) is the radius of the beam pipe at the end of the target and B(0) is the magnetic induction at the end of the target. Equation 4.16 recognizes that transverse pion momentum has been transferred to forward pion momentum as the magnetic induction is decreased. In our case, the radius for pion capture around the target r(0) = 75 mm when B(0) = 20 T. A transfer of the captured pions from the 20 Tesla target region to the 5 Tesla beam transport and phase rotation region requires an increase in the minimum beam pipe radius to 150 mm.

Once the solenoid magnetic induction has been brought down to the field in the pion decay channel or phase rotation channel, the diameter of the solenoid coils must be changed to match the diameter of the coils in the decay channel solenoids or the solenoids in the phase rotation cavities. The change in solenoid diameter must be accomplished without altering the desired induction along the solenoid axis. Two options for a transfer solenoid system were studied. The option that is to be used depends on the location of the 5 Tesla solenoid within the first set of RF cavities in the phase rotation system.

The first option has the final 5 Tesla solenoid coil diameter set at 400 mm. The system is suitable for transfer into phase rotation cavities that have their superconducting solenoids inside the cavities. Superconducting solenoids inside the cavities is suitable option if a large

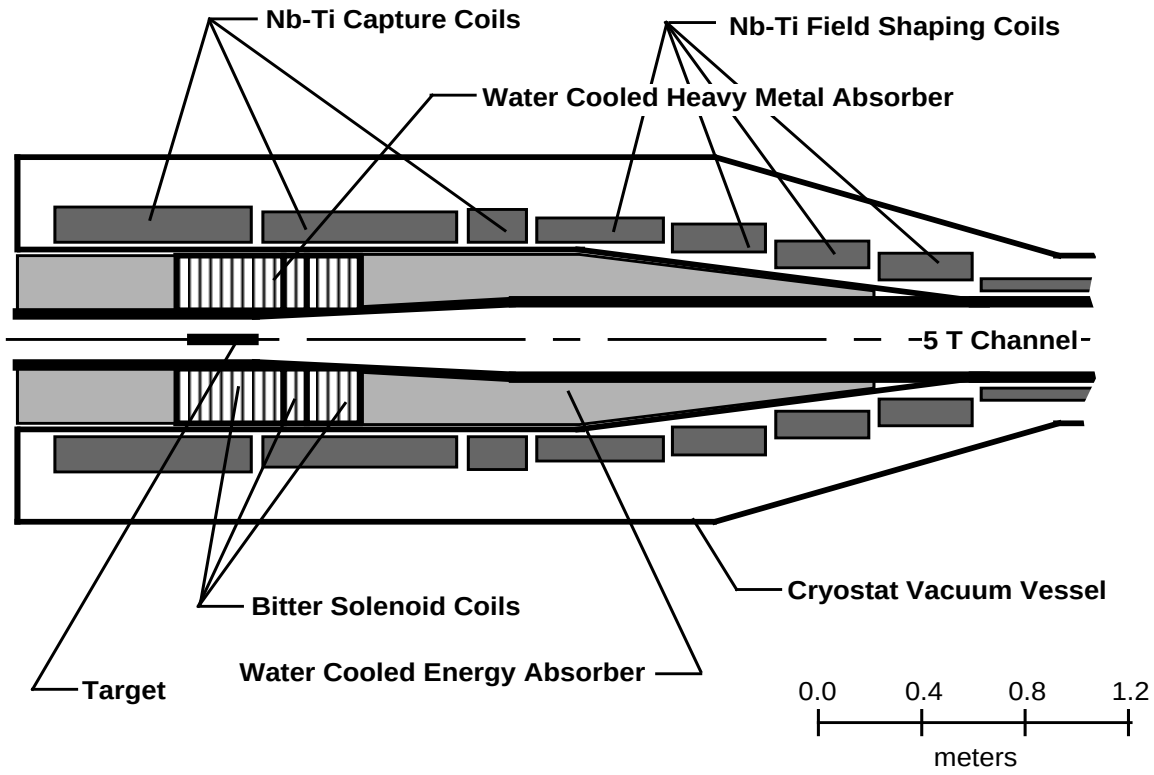


Figure 4.27: Option one: Capture and transfer solenoid system

field variation along the cavity string is acceptable. This option would also be used if a long transfer channel is used between the capture system and the phase rotation system. Figure 4.27 shows the pion capture solenoids and four transfer solenoids that grade down to a final coil inside diameter of 400 mm. The water cooled Bitter magnet and the three niobium titanium outsert solenoids for the capture system are also shown in Figure 4.27.

Also shown in Figure 4.27 are water cooled energy absorbers to absorb the energy that comes from particles generated in the target that are not captured by the capture solenoid system. Most of the incident beam power put into the target will be absorbed in the target, in the Bitter solenoids or in the water cooled absorber system. Less than 0.01 percent of the incident beam power is expected to end up in the superconducting magnets. A little over one percent of the beam incident energy will be in the pion beam that goes to the phase rotation system.

A plot of magnetic induction along the axis of the capture and transfer solenoids as a function of distance from the end of the target is shown in Figure 4.28. Figure 4.28 shows that there is a smooth transition of the magnetic induction from the end of the target to the pion transfer system or the first phase rotation cavity (with the solenoid within it). The transition from 20 Tesla to 5 Tesla occurs in about 1.2 meters. This transition does not quite

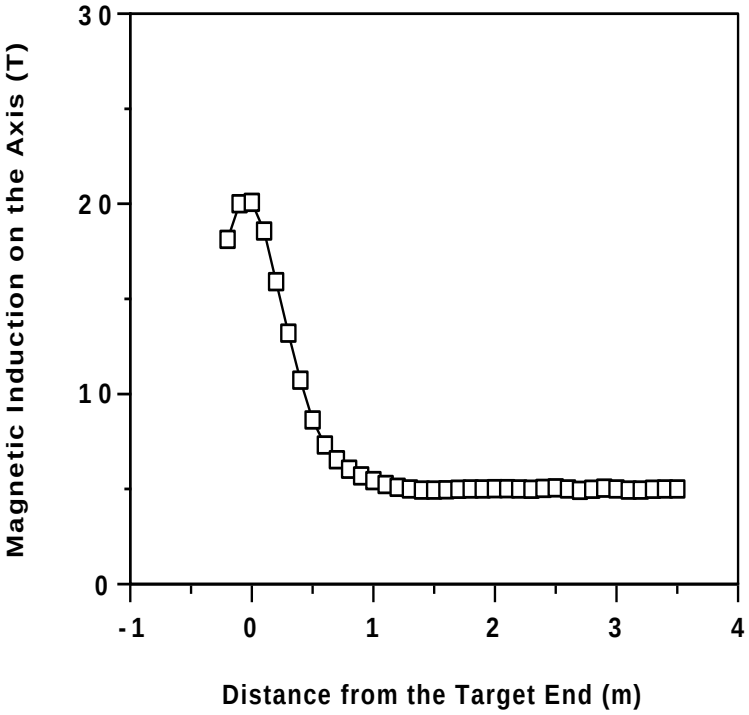


Figure 4.28: The on axis magnetic induction as a function of distance from the target in the option one capture solenoid system

fit Equation 4.15, but the transition is probably smooth enough to allow for good efficiency of pion transfer from the capture solenoid to the phase rotation channel. The magnetic field on axis can be shaped using correction coils in order to insure efficient pion transfer.

The second option chosen has the size of the solenoids growing as the field drops to 5 Tesla for transfer into the phase rotation system. This option would be used if the superconducting solenoid is outside the phase rotation RF cavities. When the solenoid lies outside the cavity, the field along the phase rotation channel is relatively uniform (to one or two percent). The obvious disadvantage of the phase rotation solenoids on the outside is the cavity is the large stored energy of the solenoid system on the outside of the cavities.

The solenoid inside coil diameter chosen for option two is 1600 mm. This solenoid is large enough to surround the first two sets of cavities of the high energy (200 to 700 MeV) phase rotation cavities. For the low energy (50 to 250 MeV) phase rotation cavities, the first stage solenoid coil inside diameter would be about 1900 mm. Figure 4.29 shows the pion capture solenoids and the transfer solenoids that grade up to a final coil inside diameter of 1600 mm. The water cooled Bitter magnet and the three niobium titanium outsert solenoids for the capture system are also shown in Figure 4.29. Figure 4.29 shows water cooled energy

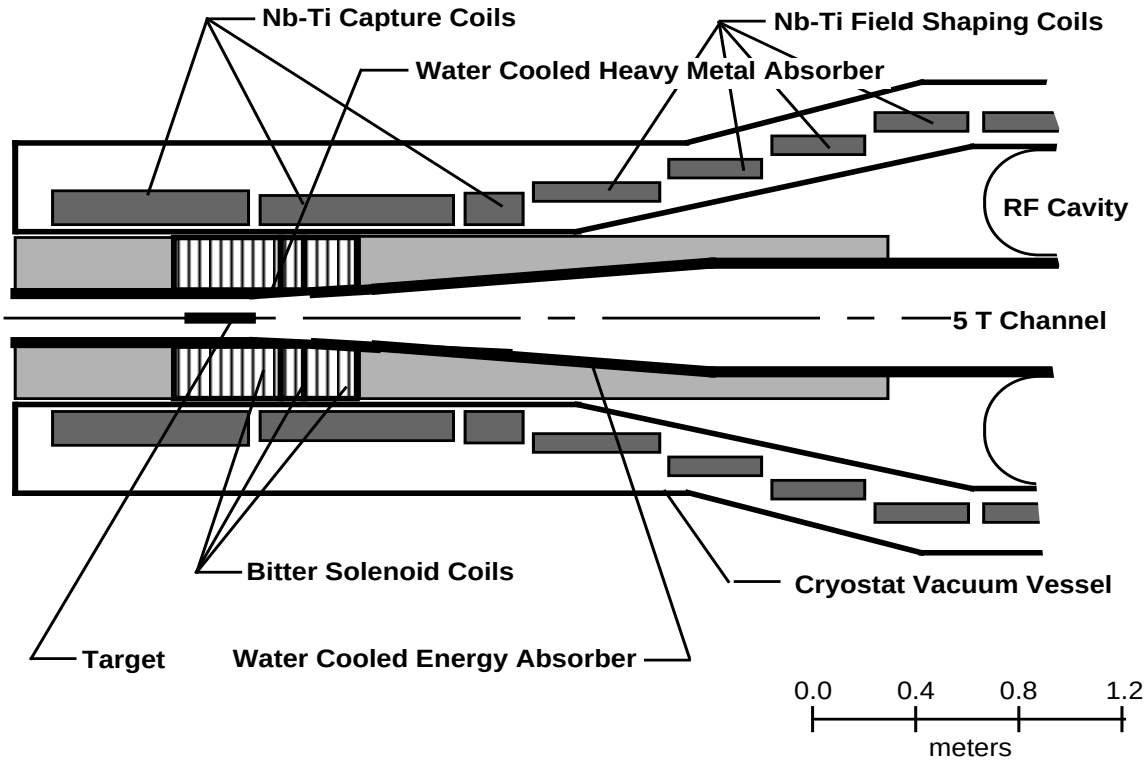


Figure 4.29: Option two: Capture and transfer solenoid system

absorbers to absorb the beam power that comes from particles generated in the target.

A plot of magnetic induction along the axis of the option two capture and transfer solenoids as a function of distance from the end of the target is shown in Figure 4.30. Figure 4.30 shows that there is a smooth transition of the magnetic induction from the end of the target to the pion transfer system or the first phase rotation cavity (with the solenoid within it). An effort was made to fit the on axis field to the relationship given by Equation4.15. The transition from 20 Tesla to 5 Tesla occurs in about 1.1 meters. The magnetic field on axis can be shaped using correction coils in order to insure efficient pion transfer. The on axis induction versus distance profile shown in Figure 4.30 is similar to Figure 4.28.

The two options for the transfer solenoid system yield a similar field shape from x = 0 to x = 3 meters. The maximum outside diameter and the stored energy of the two options are quite different. Table 4.4 shows some parameters for the capture and transfer solenoid options. About 15.5 MJ of the system stored magnetic energy is due to the current that flows in the water cooled Bitter insert coils. The Bitter solenoid is inductively coupled with the superconducting solenoids. A shut down of the Bitter solenoid will increase the current in the superconducting magnets. This effect should be tolerable provided the superconducting

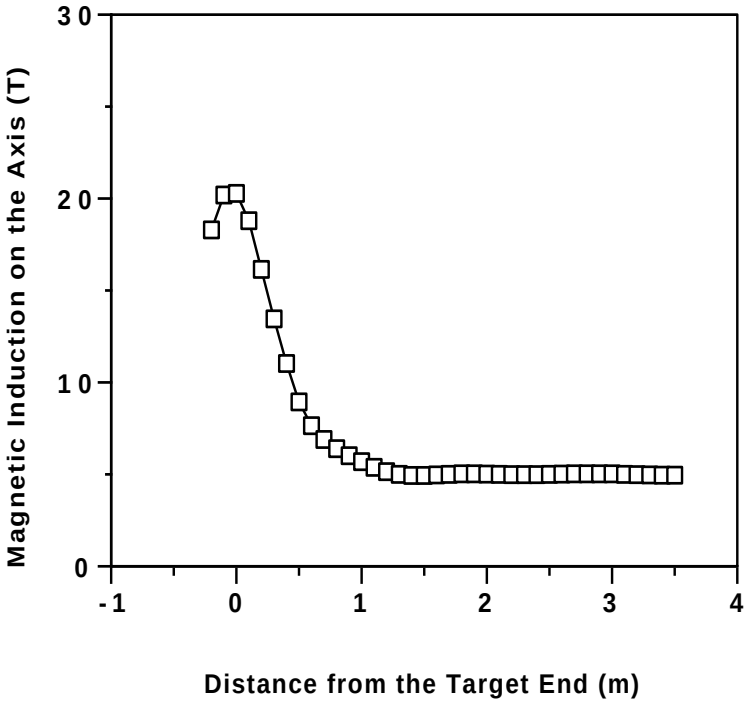


Figure 4.30: The on axis magnetic induction as a function of distance from the target in the option two capture solenoid system

solenoids are well cooled. A loss of current in the Bitter solenoids system will probably cause the superconducting solenoids to turn normal.

Two types of superconducting solenoids could be used in the superconducting part of the capture solenoid system shown in Figures 4.28 and 4.30. One could use either a cable in conduit conductor (CICC) or a magnet with well-cooled copper stabilized conductor in a helium bath. Either approach can be used in the stored energy range shown in Table 4.4. The CICC system has the advantage of having the helium inside the conductor. The shell around the cable and the helium forms part of the magnet support structure. It is probable that no other support structure is needed. Therefore, CICC conductors appear to be an attractive option. In either case, the heating from the target must be minimized. CICC magnets have been designed for use in an environment where heating is an important factor.

A fully stabilized conductor magnet would likely be made from a copper stabilized conductor in the half hard state. A pure aluminum matrix conductor is also a possibility, but additional support is needed to support the hoop forces in the magnet. Therefore, the pure aluminum conductor option is probably not attractive.

If energy deposition in the superconductor is a factor, the superconducting magnet di-

ameter can be increased to accommodate more material to absorb energy from particles that come from the target. If the field shape is not good enough to insure efficient transfer of the pions from the target to the phase rotation system, the field shape in the capture and transport solenoids can be improved by using additional water cooled correction coils between the water-cooled beam pipe and the superconducting solenoids.

Table 4.1: Target and particle production parameters appropriate to 8 GeV and 30 GeV proton beams for different target materials.

Parameter List	8 GeV C	8 GeV Cu	8 GeV Hg	30 GeV C	30 GeV Cu	30 GeV Hg
N_p/bunch, 10^{13}	5	5	5	2.5	2.5	2.5
$N_{bunches}$	2	2	2	4	4	4
σ_t (nsec)	1	1	1	1	1	1
f_{rep} (Hz)	30	30	30	15	15	15
P_{beam} (MW)	3.8	3.8	3.8	7.2	7.2	7.2
ϵ_{Nrms}, 10^{-5} (m-rad)	4	4	4	4	4	4
ϵ_{rms}, 10^{-6} (m-rad)	4.2	4.2	4.2	1.3	1.3	1.3
β_{target} (m)	3.8	3.8	3.8	12	12	12
$\sigma(x)$ (mm)	4	4	4	4	4	4
$\sigma(x')$ (mrad)	1	1	1	0.3	0.3	0.3
B_{sol} (T)	28	28	28	28	28	28
a_{sol} (cm)	7.5	7.5	7.5	7.5	7.5	7.5
$p_{\perp max}$ (GeV/c)	0.314	0.314	0.314	0.314	0.314	0.314
A_N (m-rad)	0.17	0.17	0.17	0.17	0.17	0.17
L_{target} (cm)	57	22.5	21	57	22.5	21
r_{target} (cm)	1	1	1	1	1	1
π^+/p (MARS)	0.50	0.58	0.56	0.91	1.16	1.22
π^-/p (MARS)	0.41	0.50	0.48	0.83	1.05	1.08
π^+/p (ARC)	0.52	0.62	0.59	1.31	1.62	1.56
π^-/p (ARC)	0.37	0.51	0.52	1.15	1.62	1.53
ρ (g/cc)	2.26	8.96	13.55	2.26	8.96	13.55
C_v (J/g/K)	0.7	0.39	0.14	0.7	0.39	0.14
α, 10^{-6} (/°C)	2.9	17	-	2.9	17	-
E_Y, 10^6 (psi)	1.6	17	-	1.6	17	-
Avg. P_{target} (kW)	140	360	550	125	600	950
Max. ρ_E (J/g)	40	52	70	60	180	200
Max. ΔT (°C)	60	140	500	80	420	1400
Max. Pres. (kpsi)	0.436	116	-	0.581	348	-
Pow. Den. (kW/cc)	0.782	5.0	8.2	0.625	8.5	14.4
Heat Flux (kW/cm^2)	0.392	2.54	4.2	0.353	4.24	7.2

Table 4.2: Insert Bitter solenoid design parameters

Parameters	Coil 1	Coil 2	Coil 3
Inner Radius (mm)	120.0	120.0	120.0
Outer Radius (mm)	300.0	300	300.0
Left End Location (mm)	-294.0	123.0	215.0
Right End Location (mm)	123.0	215.0	418.0
Ave. Current Density (A mm^{-2})	69.28	43.28	24.08
Estimated Power (MW)	6.18	0.66	0.30

Table 4.3: Outsert solenoid design parameters

Parameters	Coil 4	Coil 5	Coil 6
Inner Radius (mm)	400.0	400.0	400.0
Outer Radius (mm)	543.0	525.0	539.0
Left End Location (mm)	-763.0	40.0	835.0
Right End Location (mm)	-3.0	790.0	1061.0
Ave. Current Density (A mm^{-2})*	40.00	40.00	40.00
Coil Peak Induction (T)	∼6.8	∼6.9	∼6.0

Table 4.4: Parameters for the capture and transfer solenoid system

Parameters	Option 1	Option 2
Magnet Section Length to x = 3 meters (m)	3.92	3.92
Cryostat Outside Diameter at x = 0 (m)	1.50	1.50
Cryostat Outside Diameter at x = 3 meters (m)	0.70	2.00
Warm Bore Diameter at x = 3 meters (m)	0.30	1.50
Capture Magnetic Induction at Target (T)	20.0	20.0
Length of the Target Region (m)	0.23	0.23
Length for Transfer to Transport Induction (m)	1.20	1.10
Nominal Transport Magnetic Induction (T)	5.0	5.0
Stored Magnetic Energy to x = 3 meters (MJ)	37.9	53.6
Stored Energy S/C Magnet to x = 3 meters (MJ)	22.4	38.1
Stored Energy for x > 3.0 meters (MJ/m)	1.58	22.1

Bibliography

[1] R. J. Noble, Proc. 3rd Int. Workshop on Advanced Accelerator Concepts, Port Jefferson, NY, AIP Conf. Proc. No. 279, 949 (1993).

[2] R. Palmer et al., BNL–61581 (1995).

[3] N. V. Mokhov, R. J. Noble and A. Van Ginneken,"Targetry and Collection Optimization for Muon Colliders", FermiLab–Conf–96/006, and Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop, Ed. J. C. Gallardo, AIP Conference Proceedings 372, 1996.

[4] R. Palmer et al., "Muon Collider Design", BNL–62949 (1996).

[5] N. V. Mokhov, "The MARS Code System User's Guide, version 13 (95)", FNAL–FN–628 (1995).

[6] J. Ranft, "The DPMJET Code System" (1995).

[7] D. Kahana and Y. Torun, "Analysis of Pion Production Data from E-802 at 14.6 GeV/c using ARC", BNL–61983 (1995).

[8] *Review of Particle Properties*, Phys. Rev. *D50*, 1177 (1994).

[9] J. D. Jackson, *Classical Electrodynamics*, 2nd Ed., J. Wiley, New York, 1975.

[10] A. Van Ginneken, "Deuterons or Tritons for Muon Collider Driver", FermiLab–FN–647 (1996).

[11] ANSYS (rev. 5.1), Swanson Analysis System, Inc., SASI/DN–P511:51, Houston (1994).

[12] F. F. Chen, *Introduction to Plasma Physics*, Plenum, New York, pp. 23–26 (1974).

[13] M. Green, "Pion Capture Magnet System",Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop, Ed. J. C. Gallardo, AIP Conference Proceedings 372, 1996.

[14] N. V. Mokhov, R. J. Noble and A. Van Ginneken, *Targetry and Collection Optimization for Muon Colliders* FermiLab-Conf-96/006 to appear in the Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop, October 1995, Montauk, NY .

[15] M. A. Green, *Some Options for the Muon Collider Capture and Decay Solenoids*, Lawrence Berkeley National Laboratory Report LBL-37995, to appear in the Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop, October 1995, Montauk, NY.

[16] J. Weggel, *Collector Magnets for $\mu^+\mu^-$ Colliders*, an unpublished report concerning preliminary design studies of capture magnets system for 2 TeV on 2 TeV and demonstration colliders.

[17] Miller, J. R., Bird, M. D., Bole, S., et al, *An Overview of the 45T Hybrid Magnet System for he National High Field Magnet Laboratory*, IEEE Transactions on Magnetics 30, No. 4, p 1563, (1994).

[18] Iwasa,Y., Leupold, M. J., Weggel, R. J., and Williams J. E. C., *Hybrid III: The System, Test Results, the Next Step*, IEEE Transactions on Applied Superconductivity 3, No. 1, p 58, (1993).

Contributors

- Robert J. Noble, (FermiLab), Editor
- David Ehst, (ANL)
- Michael Green, (LBNL)
- David Kahana, (BNL)
- Nikolai Mokhov, (FermiLab)
- Igor Novitski, (FermiLab)
- Yagmur Torun, (SUNY, Stony Brook)
- Andy Van Ginneken, (FermiLab)

List of Figures

List of Tables

Chapter 5

MUON PHASE ROTATION CHANNEL

Contents

5.1 Introduction

The basic concept is to collect as large an amount of pions as possible, channel them along a decay channel while using acceleration cells to collect as many particles as possible within a usable energy spread. The general strategy is to boost low-energy particles and decelerate higher energy particles by means of appropriately place acceleration cavities.

Two alternatives for the capture and phase rotation of pions issuing from the target and their muon decay products are being considered: namely an rf approach and an induction approach.

The rf approach is characterized by requiring that the rf cavities be placed close to the pion production target (i.e. within a few meters) in order that the frequencies for phase rotation be high enough that the cavities not become unreasonable in size. The induction linac approach is characterized by requiring a significant drift between the production target and the first induction cell (i.e. of the order of 25 m) in order to match the beam pulse length to acceleration wave lengths that can be reasonably achieved without incurring excessive core losses (i.e. 50 to 100 ns). The output muon pulse from the induction linac then needs to be compressed to $\sim$ 20 ns for injection into the cooling channel whereas the output pulse length of the rf approach to phase rotation is roughly matched to the cooling channel. Since the induction cells produce a factor of two or more lower accelerating gradient than the rf cells, the induction approach requires more axial length. On the other hand there are uncertainties about operating rf cavities in the high radiation environment within a few meters of the pion production target so discussion of both approaches is presented.

5.2 rf Approach

5.2.1 Introduction

The design of the phase rotation system is greatly influenced by the energy range chosen for the particle collection. We develop two solutions to accommodate a low-energy collection (50-250 MeV) and a high-energy collection (200-700 MeV). The pion production spectrum at the proton target is the most important factor in determining the best choice for the collection system.

The rf approach is characterized by designing a linac system which can begin immediately after the targeting of the main proton beam. We wish to begin the process as soon as possible so as to avoid the necessity of high-gradient, ultra-low frequency cavities. Figures 5.1 and 5.2 show the longitudinal structure of the beam after drift distances of 2 m and 24 m. Notice that lower frequency rf cavities are required for the cases in which either low-kinetic energy capture or long drifts are required.

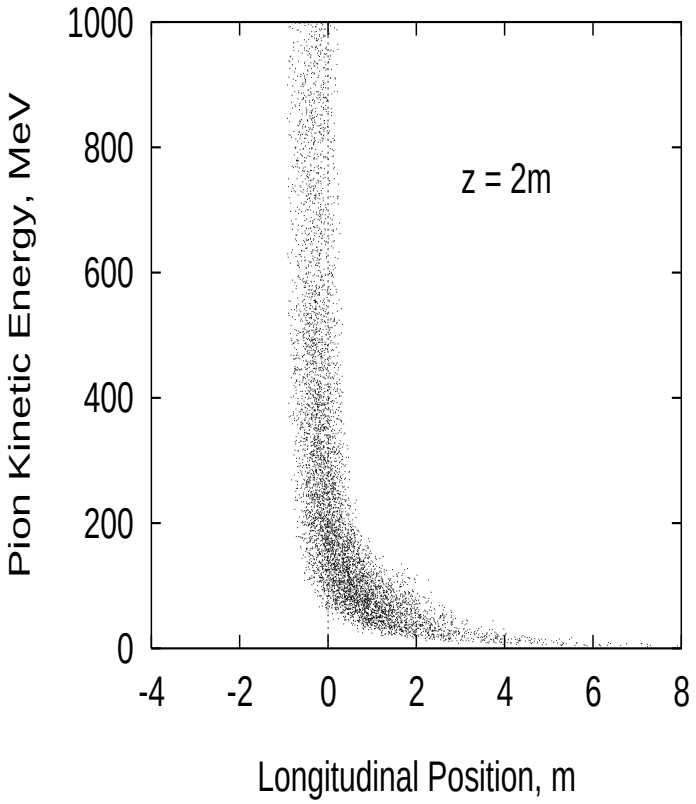


Figure 5.1: Collected particles after 2 m drift from the proton target

5.2.2 rf Cavities

The basic approach is to begin the phase rotation process as soon as possible in order to minimize the phase spread of the captured beam. By initiating the capture and phase rotation early we can take advantage of the higher accelerating gradients available with higher frequency cavities. Empirical evidence (see Figure 5.3) implies that pulsed rf systems in the frequency range below 500 MHz can be reliably operated at two Kilpatricks[1]. Also, the kinetic energy of the particles to be captured is of fundamental importance as it dictates the frequency of the cavities to be used. Notice that for particles with KE $>$ 200 MeV (See Figures 5.1 and 5.2), the phase spread of the particle beam is largely determined by the initial spread of the initiating proton beam (1 ns for this example).

Our choice of frequencies is determined by the phase spread of the beam. We begin with a high-frequency cavity which can deliver higher accelerating gradients and then proceed to lower frequency cavities as the beam continues it's longitudinal spread along the beam channel. Another important factor in choosing the rf cavity frequencies is the desire to have the same rf channel work for both positive and negative pulses. This can most easily be achieved by insuring that all cavities operate at an odd harmonic of some convenient fundamental frequency. Hence, for example, cavities of 300 MHz, 100MHz and 60 MHz

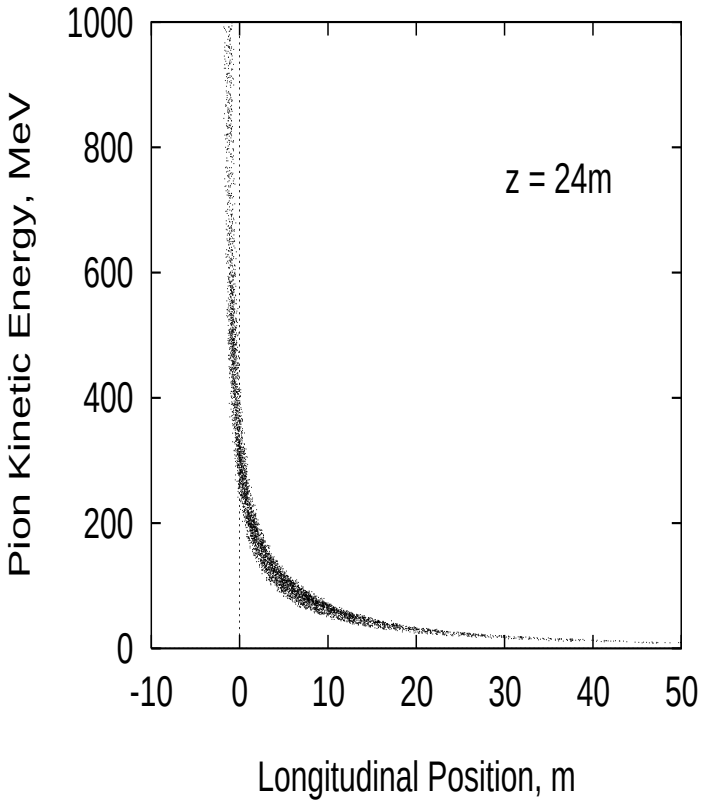


Figure 5.2: Collected particles after 24 m drift from the proton target

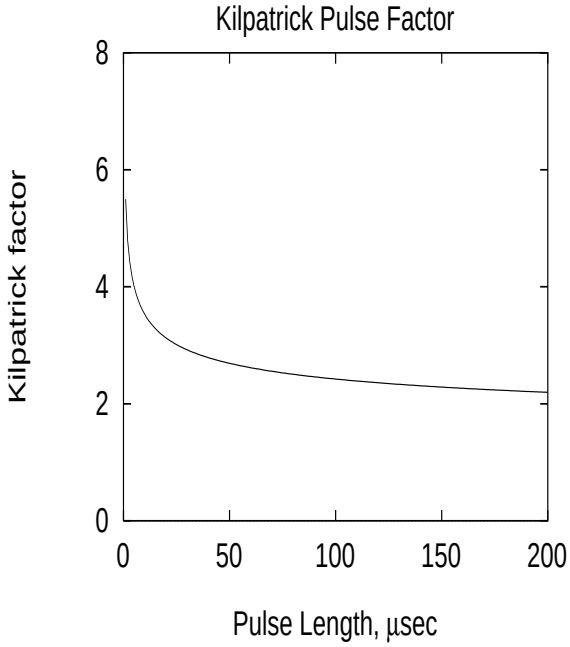


Figure 5.3: Kilpatrick Factor Limits for Pulsed rf Systems

(see Figure 5.4) satisfy this requirement since they are respectively 15×, 5×, and 3× the fundamental frequency of 20 MHz thus guaranteeing that a shift of 180 degrees at 20 MHz

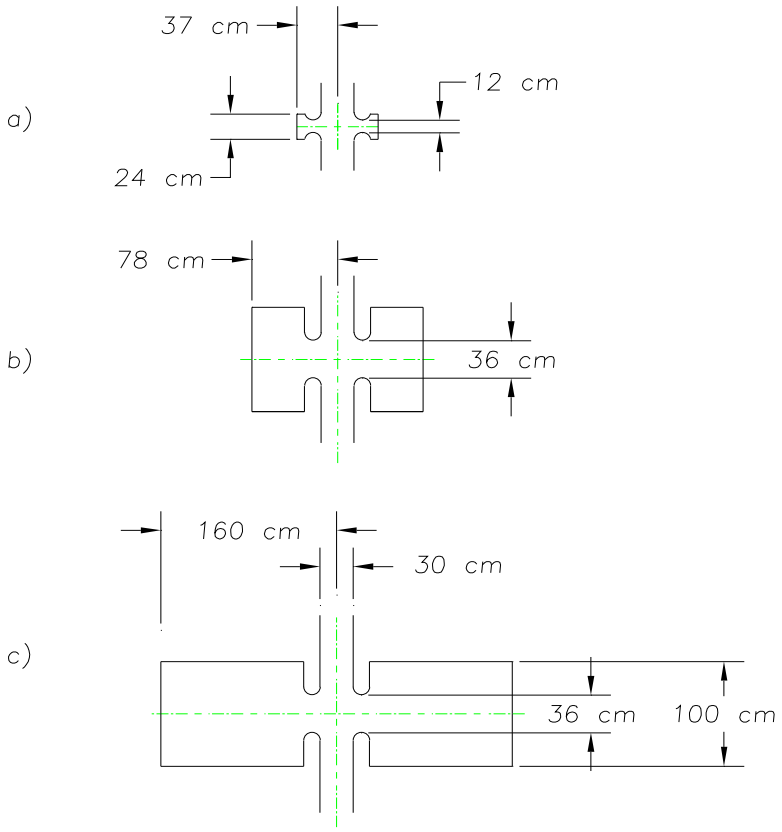


Figure 5.4: The high-energy collection rf cavities: a) 300 MHz; b) 100 MHz; c) 60 MHz. The fundamental harmonic is 20 MHz.

will result in a shift of an odd multipole of π for all frequencies.

We show in Figure 5.5, a scheme in which particles are collected and acceleration begun with the gap of the first cell placed 3 m beyond the beginning of the proton target. This solution for the high-energy collection starts with 300 MHz cavities and then proceeds down in frequency. For convenience of display, only three cavities of each frequency type is shown in Figure 5.5; however, the rf system parameters for the full solution are given in Tb.5.1. The average rf power is given assuming a 15 Hz repetition rate.

For low-energy collection (50 MeV-250 MeV) we need to consider lower frequency rf cavities. In this case we use a fundamental frequency of 10 MHz and choose 90 MHz, 50 MHz and 30 MHz cavities. Figure 5.6 shows a schematic of the rf cavities. The collection system is depicted in Figure 5.7 and the full rf system parameters are given in Tb.5.2.

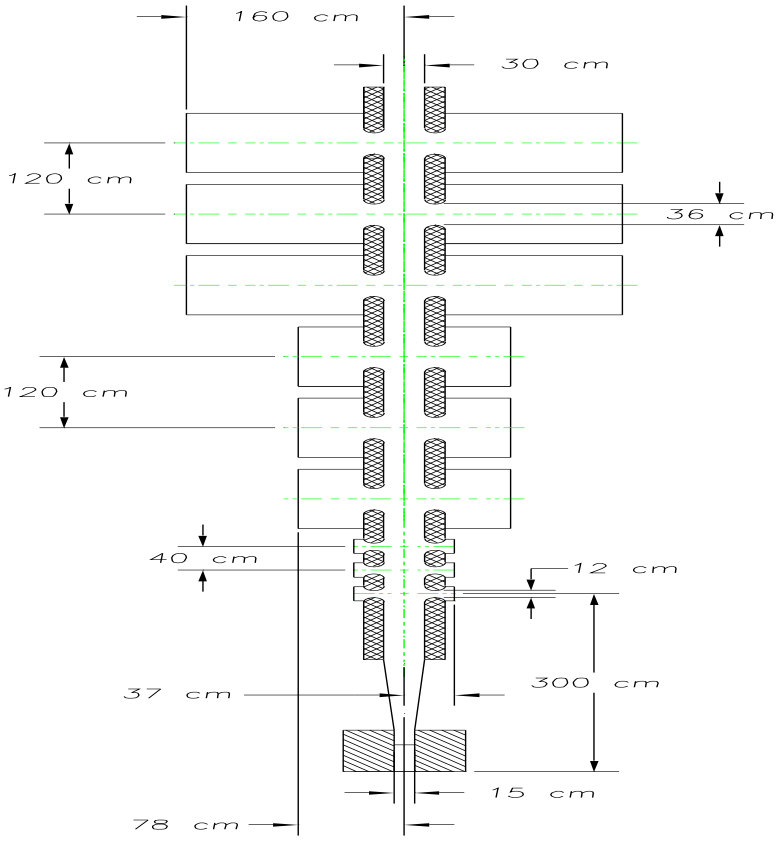


Figure 5.5: Schematic of the high-energy collection linac. Only three sections each are depicted of the 300, 100 and 60 MHz cavities. Hatched area is 20 T Bitter Solenoid, cross-hatched areas are 5 T superconducting solenoids.

5.2.3 Decay Channel Solenoids

The phase rotation magnet system is a continuation of the magnet system that captures and transports the pions that have been generated off the proton target. The solenoid field at the exit of the capture system must be continued in the phase rotation system. The pions in the phase rotation system still have a transverse momentum as high as 112 MeV/c. An induction of 5 T is required to contain the pion transverse momentum as the pions decay to muons as they are being bunched in the phase rotation system. Two approaches have been studied for the phase rotation rf cavity transport solenoid system. The first approach is to put the superconducting solenoids within the rf cavity so that they surround the cloud of pions coming from the target. The second approach surrounds the entire rf cavity with a large superconducting solenoid. In both cases the average magnetic induction along the channel that carries the pion cloud is 5 T. The channel that carries the pions to the phase

rotation system will have a warm bore diameter of 30 cm. The two approaches for the phase rotation rf cavity solenoids are illustrated in Figure 5.8 for rf cavities that have an outside diameter of 150 cm, a channel diameter of 30 cm and a cell length of 120 cm. In both cases, the acceleration gap is 36 cm.

Solenoids that are inside the rf cavity are less expensive, but the field must span the gap between the coils. If the superconducting solenoids are inside of the cavities, the gap between the coils is the cavity acceleration gap plus the allowance for the cryostat, the

Table 5.1: High-energy collection linac parameters

rf frequency [MHz]	300	100	60
Cavity Length [cm]	40	120	120
Full Gap length [cm]	12	36	36
Cavity Radius [cm]	37	78	160
Beam Pipe Aperture [cm]	30	30	30
Q/1000 (from SFISH)	21.5	54.8	44.6
Avg Gradient [MV/m]	12.5	4.5	3.6
rf Peak Power [MW]	8.8	2.2	1.1
Avg Power (15Hz) [KW]	10	19	13
Stored Energy [J]	101	166	208
Linac Segment [m]	12	12	54
Total Power (15Hz) [KW]	300	192	585

Table 5.2: Low-energy collection linac parameters

rf frequency [MHz]	90	50	30
Cavity Length [cm]	120	120	120
Full Gap length [cm]	36	36	36
Cavity Radius [cm]	90	206	126
Beam Pipe Aperture [cm]	30	30	30
Q/1000 (from SFISH)	53.4	71.1	16.8
Avg Gradient [MV/m]	4.2	3.3	2.1
rf Peak Power [MW]	1.8	1.1	4.8
Avg Power (15Hz) [KW]	17	26	43
Stored Energy [J]	165	261	423
Linac Segment [m]	6	18	18
Total Power (15Hz) [KW]	85	390	640

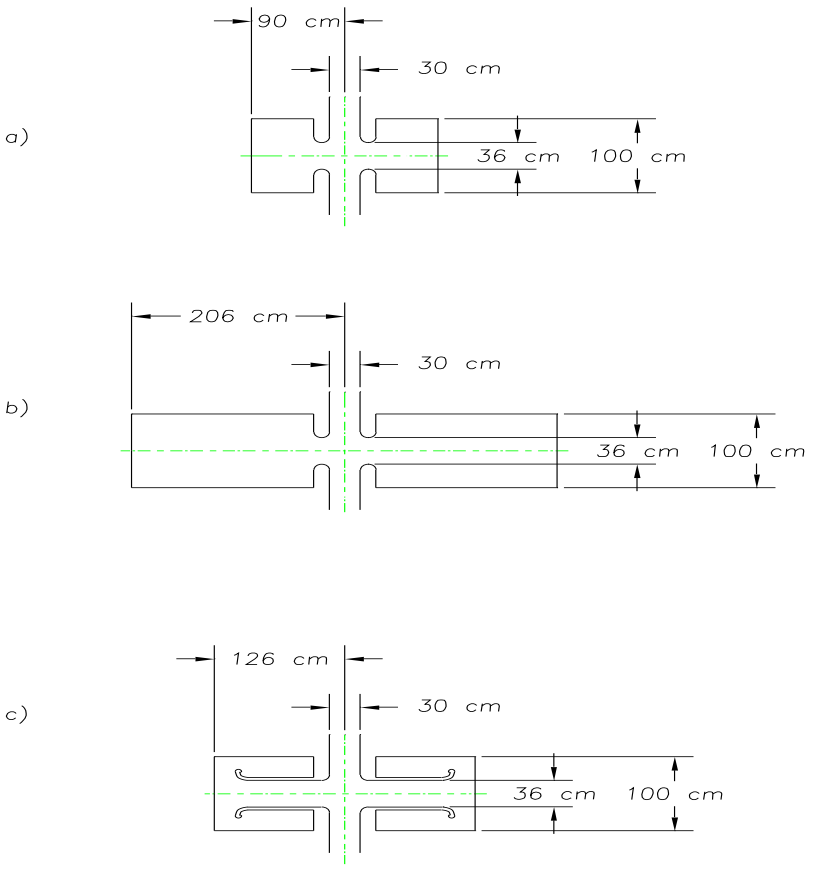


Figure 5.6: The low-energy collection rf cavities: a) 90 MHz; b) 50 MHz; c) 30 MHz. The fundamental harmonic is 10 MHz.

insulation and the end of the bobbin. The best that can reasonably be expected is a gap between coils that is only 4 cm larger than the accelerator gap. The gap between the coil ends shown in Figure 5.8a is about 40 cm. In addition, a 3 cm gap is shown in the middle of the superconducting coil. This gap allows the coil bobbin to be attached to the cryostat cold mass support system, which is not shown in Figure 5.8. The ends of the coils are built up to make up for the current that is lost in the gap. In order for there to be an average induction of 5 T, there must be 4.775 MA turns in each 120 cm long cell. The magnetic induction along the axis of the solenoid will not be uniform. The 3 cm gap in the middle of the coil has almost no effect on field uniformity, but the 40 cm gap between the coils will have a large effect on the field uniformity. The reason for the lack of field uniformity is that the 40 cm gap between the coils is larger than the inside diameter of the coils (about 34 cm). The 3 cm gap has only a small effect on the field uniformity because this gap is much smaller than the coil diameter. The effect of coil diameter and the gap between coils is shown in

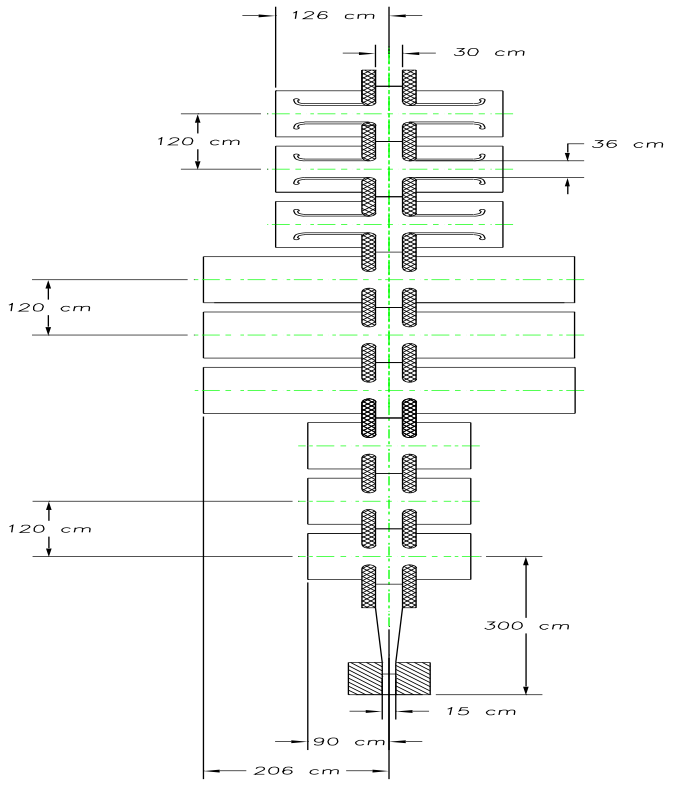


Figure 5.7: Schematic of the low-energy collection linac. Only three sections each are depicted of the 90, 50 and 30 MHz cavities. Hatched area is $20\,T$ Bitter Solenoid, cross-hatched areas are $5\,T$ superconducting solenoids.

Figure 5.9 and Tb.5.5.

We consider the possibility that the solenoidal axial field uniformity in the decay channel may have to be better than ± 5 % to avoid resonant particle losses. In order to improve the field uniformity, we have developed additional rf cavity solutions in which all cavities have axial lengths of 40 cm and accelerating gaps of 12 cm while retaining the 30 cm beam aperture for the decay channel. These rf solutions are given in Tbs.5.3 and 5.4.

The superconducting solenoid outside of the cavity is illustrated in Figure 5.8b. The gap between the coils in governed by the distance needed between coils to get services into the rf cavity. This gap will allow for a 10 cm warm diameter pipe to penetrate through the solenoid cryostat. If holes are provided for rf cavity services, the cryostat for the superconducting solenoid coils can be made quite long (say eight or ten rf cavity modules long). The current density in the larger diameter coils must be lower because the magnet stored energy is higher.

As a result, the coils are shown to be thicker. The field uniformity along the magnet axis is much better for the solenoids located outside the rf cavity compared to the solenoids located inside the rf cavity. The gap between the coils for the large coil case is small compared to the solenoid diameter. The effect of the solenoid diameter on field uniformity and other parameters is shown in Figure 5.9 and Tb.5.5.

From Figure 5.9 and Tb.5.5, it is clear that cases 1 and 2 with the coil inside the cavity have a far less uniform field along the axis than does case 3 where the solenoid is outside the rf

Table 5.3: High-energy collection linac parameters

rf frequency [MHz]	300	100	60
Cavity Length [cm]	40	40	40
Full Gap length [cm]	12	12	12
Cavity Radius [cm]	37	87	87
Beam Pipe Aperture [cm]	30	30	30
Q/1000 (from SFISH)	21.5	33.5	14.4
Avg Gradient [MV/m]	12.5	4.5	3.6
rf Peak Power [MW]	8.8	1.0	3.4
Avg Power (15Hz) [KW]	10	5	13
Stored Energy [J]	101	53	131
Linac Segment [m]	12	12	54
Total Power (15Hz) [KW]	300	158	1760

Table 5.4: Low-energy collection linac parameters

rf frequency [MHz]	90	50	30
Cavity Length [cm]	40	40	40
Full Gap length [cm]	12	12	12
Cavity Radius [cm]	100	98	140
Beam Pipe Aperture [cm]	30	30	30
Q/1000 (from SFISH)	34.3	13.5	11.1
Avg Gradient [MV/m]	4.2	3.3	2.1
rf Peak Power [MW]	0.9	3.5	2.5
Avg Power (15Hz) [KW]	5	15	14
Stored Energy [J]	52	149	145
Linac Segment [m]	6	18	18
Total Power (15Hz) [KW]	78	670	652

Table 5.5: Parameters for three solenoid magnet configurations in or around a 150 cm outside diameter rf cavity

Parameter	CASE 1	CASE 2	CASE 3
rf Cell Length (cm)	120	120	120
Cavity Outside Diameter (cm)	150	150	150
Solenoid Coil Inside Diameter (cm)	34	34	160
Solenoid Coil Outside Diameter (cm)	44	44	180
Cavity Acceleration Gap (cm)	36	11	36
Gap Between the Coils (cm)	40	15	15
Coil Length per Cell (cm)	80	105	105
Average Induction along Axis (T)	5.00	5.00	5.00
Minimum Induction on Axis, B_{min} (T)	2.324	4.381	4.986
Maximum Induction on Axis, B_{max} (T)	6.520	5.276	5.103
Induction Ratio B_{max}/B_{min}	2.80	1.20	1.02
Pion Chamber Diameter (cm)	44	32	30
Peak Induction in the Coil (T)	7.24	5.82	5.31
Stored Energy (MJ per meter)	1.35	1.14	21.2

cavity. Reducing the acceleration gap improves the field uniformity along the axis (compare cases 1 and 2), but the price for increased field uniformity on the axis is an increased length of the phase rotation system.

A non-uniform solenoid field along the axis of the phase rotation linac has two consequences. First the diameter of the pion decay channel must be increased to accommodate the transverse momentum in the pions at the lowest field along the axis. Second a periodic field may induce oscillations in the beam that could lead to pion losses. The increase in beam radius due to the induction decrease on axis may be estimated using the following expression:

$$r\left(z\right) = \sqrt{\frac{B_{ave}}{\mathrm{B\left(z\right)}}}\; r\left(B_{ave}\right) \tag{5.1}$$

where r(z) is the radius of the beam pipe at location z and B(z) is the magnetic induction at the location z. r(B_{ave}) is the radius of the beam pipe when the induction is at the average induction B_{ave}. For a transverse momentum of 112 MeV/c, B_{ave} = 5 T and r(B_{ave}) = 15 cm.

From Tb.5.5, one can see that the peak induction in the winding is considerably higher than the average induction for the cases 1 and 2 where the solenoid is inside the cavity. This

increase in magnetic induction affects the performance of the superconductor.

The stored energy per unit length for solenoids surrounding the cavity is much larger than for the case when the solenoids are inside the cavity. The cost of the superconducting solenoid is proportional to the stored energy to some power (from 0.6 to 0.7). From a cost standpoint, it can be argued that decreasing the field on the solenoid axis is desirable, but the diameter of the beam tube must be increased in order to contain a pion beam with the desired transverse momentum.

When the superconducting solenoids are installed outside the rf cavities, the cavity design should minimize the cavity diameter and the solenoid coils should be as close to the cavity as possible. The gaps between solenoid coils are far less important than when the coil is outside the cavity, but these gaps should be minimized to improve field uniformity.

5.2.4 Beam Dynamics

The beam dynamics of the particle collection system is modeled with the computer program PARMELA. Particle are generated within a target surrounded by a high-field solenoid (20 to 28 T are considered) and subsequently transported down a lower field solenoid (5 to 7 T) channel with rf cavities placed at appropriate intervals. Rf phases are adjusted for each cell to optimize the longitudinal phase of the particle bunch exiting the complete system. Figure 5.10 shows the results of capturing and phase rotating particles for two collection systems corresponding to the collection regions described earlier. The final number of particles collected for each system is dependent on the assumed initial particle spectra.

5.3 Induction Linac Approach

5.3.1 Introduction

In this section we describe an induction linac approach to phase rotation of the muon bunch prior to entering the cooling channel. The instantaneous energy spread is reduced from $\frac{dE}{E} \approx \pm 100\,\%$ to $< \pm 10\,\%$ by allowing the muons to drift and spread longitudinally and then using induction cells to phase rotate the beam. The head-to-tail sweep can be reduced from an energy spread of $\approx 200\,\%$ to a few per cent. The equations describing longitudinal pulse dynamics, voltage waveforms and induction cell parameters have been given in a previous report[2]. That report applied the equations to a muon production spectrum from early considerations of the muon collider which had high collection energy (0.3 to 1.0 GeV). Results for the decay of muons from low-energy captured pions have been examined. Specifically

we consider two cases: (1) $0.15 < E_\mu < 0.3\,\mathrm{GeV}$ and $0.25 < E_\mu < 0.7\,\mathrm{GeV}$. In this section muon energy is total energy including the muon rest mass since that is what appears in the dynamical equations.

5.3.2 Input Muon Pulse Energy and Intensity

A typical distribution of mean energy and beam current in a muon pulse is shown in Figures 5.11. This case corresponds to the lower energy of the two cases we consider here. The time behavior of energy and current are well characterized by simple exponential functions given in the caption to Figures 5.11. We refer to the Proton Source and Targetry and Pion Production chapters for detailed discussion of pion production, capture and decay into muons. The calculations in Figures 5.11 were done with 24 GeV/c protons incident on a Hg target, 5×10^{13} protons per pulse, rms pulse width 3 ns. The Hg target is in a 20 T solenoid field, tapering to 5 T in the pion decay channel. The spectra in Figures 5.11 are 24 m from the target and the integrated yield from 0 to 50 ns is approximately 0.4 muons per incident proton. Because of the exponential attenuation it is relatively inefficient to accelerate the lowest energy muons. We will somewhat arbitrarily truncate the muon distribution at 50 ns. There is also a high energy group of muons with energy extending above 1 GeV and arriving in time bins before T = 0 in Figures 5.11. These highest energy muons have a high instantaneous energy spread $\frac{dE}{E} \approx 0.5$ and are less amenable to phase rotation. They will also be truncated. Some of these muons are treated in the higher energy case with phase rotation beginning at z = 171 m instead of 24 m. The instantaneous rms energy width of muons from 0 to 50 ns is shown in Figure 5.12 and falls from $\frac{dE}{E} = 0.08$ at the head to $\frac{dE}{E} = 0.02$ at the tail.

The Monte Carlo calculations for the high-energy case, $0.25 < E_\mu < 0.7\,\mathrm{GeV}$, can also be fit with simple exponential functions. At z = 171 m from the proton target $\mathrm{E(GeV)} = 0.213 + 0.516\,e^{-\mathrm{T(ns)}/25.9}$ and $\mathrm{I(A)} = 87\,e^{-\mathrm{T(ns)}/41.1}$. For phase rotation we will consider the pulse extending from 0 to 70 ns. The integrated yield from 0 to 70 ns is approximately 0.4 muons per incident proton.

5.3.3 Acceleration Wave Shape, Voltage Gradient and Mean Muon Energy

The time dependence of acceleration wave shapes and the resultant flattening of mean energy as the muons propagate through the accelerator are shown in Figures 5.13. The rate of spreading of the wave shape depends on the energy distribution, transverse momentum

distribution and accelerating gradient. The initial energy distribution in Figure 5.13b at z = 0 is the same as Figure 5.11a. The mean transverse momentum is 25 MeV/c and the rms transverse momentum width is 43 MeV/c. The acceleration wave shapes shown in Figure 5.13a produce a flat energy profile at z = 170 m, which is simply the initial head-to-tail energy spread divided by the average accelerating gradient that is applied to the tail of the pulse, here assumed to be 1 MV/m. The pulse width of the flat distribution at z = 170 m is 130 ns. The gradient in an induction accelerator is ultimately limited by the increasing radial size of the magnetic cores and of course also by the fraction of accelerator length that can be occupied by induction cells. Assuming 40 % of the axial accelerator length is available for induction cells a gradient of 1 MV/m seems to be practical and is well below the theoretical maximum.[2] A factor of two increase in gradient and reduction in axial length may be possible. The higher energy case is qualitatively similar to Figures 5.13 but the overall accelerator length is 480 m due to the larger initial head to tail energy spread. The accelerating gradient is again taken to be 1 MV/m. Parameters for both cases will be summarized in the following section.

The width of the output muon pulse from the phase rotation induction linac needs to be reduced from $\sim$ 130 ns to $\sim$ 20 ns before injection into the muon cooling channel. This can be accomplished by continuing the acceleration until the energy of the tail slightly exceeds the head (i.e. $\sim$ 10%) and then allowing the bunch to ballistically compress as it drifts. The drift distance and fractional muon loss due to decay are decreased if the compression is carried out at low energy. This could be done by decelerating the muons once the phase rotation is accomplished. Alternatively, the phase rotation could be done by simultaneously decelerating the head and accelerating the tail of the pulse so the output energy matches what is desired for compression. The optimum induction accelerator strategy for matching to the cooling channel is a subject for further study.

5.3.4 Accelerator Parameters

A schematic of an induction cell is shown in Figure 5.14 which also defines some of the geometric quantities that appear in Tb.5.6. There are four components to the cell: (1) the high voltage pulsed power feed, (2) the magnetic core, (3) the vacuum insulator and (4) the acceleration gap. The induction core has axial length w and radial width ΔR. The inside and outside radii of the magnetic core are R_1 and R_2. The high voltage pulsed power lead enters along one side of the core, encircles it and returns to ground. The magnetic core volume behind the vacuum insulator is filled with insulating dielectric fluid. There are three insulating gap widths indicated: the dielectric gap g_d, the vacuum insulator gap g_s and

the acceleration gap g_v. These must withstand dielectric breakdown, vacuum surface flash over and vacuum breakdown respectively. The angle between the insulator surface and the metal electrode surfaces on the vacuum side is shown to be $\approx 25^\circ$ to 45° to maximize the breakdown limiting field strength. The re-entrant acceleration gap prevents radiation from the beam channel reaching the insulator surface and possibly initiating breakdown. The primary concern is shielding the synchrotron radiation produced by muons and by electrons from muon decay. The total length of the induction cell is L_c. The values of these geometrical parameters as well as certain electrical parameters for the two cases we are considering are given in Tb.5.6.

The entire accelerator consists of axially stacked arrays of acceleration cells as in Figure 5.14 interleaved with 5 T superconducting solenoids for focusing the beam. The inside and outside radii of the solenoid cryostats are assumed to be 15 and 30 cm respectively. The inside radius coincides with the beam tube radius. For the examples in Tb.5.6 we have assumed that the solenoids occupy 60 % of the axial length and the acceleration cells 40 %. The inside radius of the magnetic induction core is assumed to be $R_1 = 30\,\mathrm{cm}$. Increased accelerating gradient could be achieved by increasing the inner magnetic core radius and allowing the core volume to extend over the outside of the solenoids. The integrated design of the acceleration cells and solenoids must avoid saturation of a significant fraction of the induction core by solenoid flux leaking out of the beam channel. The cell voltage $V_0 = 50\,\mathrm{kV}$ was chosen with consideration of how the pulse power system might be configured. In principle one could choose spark gaps, thyratrons or saturating magnetic cores for the final stage switch into the induction cores since they can all be configured to handle the required power level. Spark gaps could allow operation up to $V_0 = 250\,\mathrm{kV}$. However electrode erosion would require replacement after a few million pulses, or a day of operation at 30 Hz, so they are ruled out. Saturating magnetic cores have been shown to switch coaxial lines charged up to $\approx 200\,\mathrm{kV}$ at kHz rep rates and high power thyratrons are commercially available for switching up to $\approx 50\,\mathrm{kV}$ so either of these seems possible. 50 kV was chosen as a compromise between these two possibilities; a saturating magnetic core discharging a coaxial line charged to $2V_0 = 100\,\mathrm{kV}$ or a thyratron discharging a Blumlein line charged to $V_0 = 50\,\mathrm{kV}$. Induction linac parameters for the two cases (1) $0.15 < E_\mu < 0.3\,\mathrm{GeV}$ and $0.25 < E_\mu < 0.7\,\mathrm{GeV}$ are given in Tb.5.6. They are determined by four primary relationships for; (1) the cell axial length, (2) the volt-seconds of the magnetic core, (3) the voltage rise time and (4) the voltage insulation and shielding of the vacuum insulator.[2] For the magnetic core material we have chosen *Metglas 2605SC* which has high saturation magnetic flux and the dynamic hysteresis loss has been measured in the switching times of interest.[3, 4] We assume a packing fraction $p_f = 0.75$. Reasonable allowances and considerations have also been made for mechanical

structure, voltage insulation and induction core cooling. Where it makes a difference, parameters given in Tb.5.6 are for the last cell in the accelerator which has the highest volt seconds requirement. Values for the entire accelerator are then obtaining by multiplying values for the last cell by the number of cells and ignoring the $\approx 20\,\%$ corrections that arise if axial averages are taken.

The overriding difference between the two cases in Tb.5.6 is the factor of 2.8 in axial length due to the head to tail energy spread at z = 0 and assumed accelerating gradient V'_m = 1 MV/m in both cases. The acceleration cells for the two cases are rather similar and the differences in detail are traceable to the somewhat different acceleration wave shapes imposed by the two muon energy spectra. Each cell consumes about 38 J of electrical energy per pulse, or at 30 Hz rep rate (2 pulses at 15 Hz to accommodate each muon charge state) about 1.1 kW of electrical power. The electrical energy is about evenly divided amongst hysteresis loss (39 %), charging the cell capacitance (25 %) and reflected energy due to impedance mismatch (36 %). The magnetic core leakage current is 3.6 to 4 kA so beam loading with a maximum current ≈ 100 A is relatively light and produces negligible pulse distortion. Overall the first case consumes 4.1 MW of pulsed electrical power and the second case 12 MW. The total weight of *Metglas 2605SC* is 104 tonne for the first case and 229 tonne for the second. To get a feeling for the scale of induction accelerator involved, the total weight of magnetic core material involved is ten to twenty times less than has been contemplated for the drive beam of a relativistic klystron version of a 1 TeV e^+e^- collider.[5]

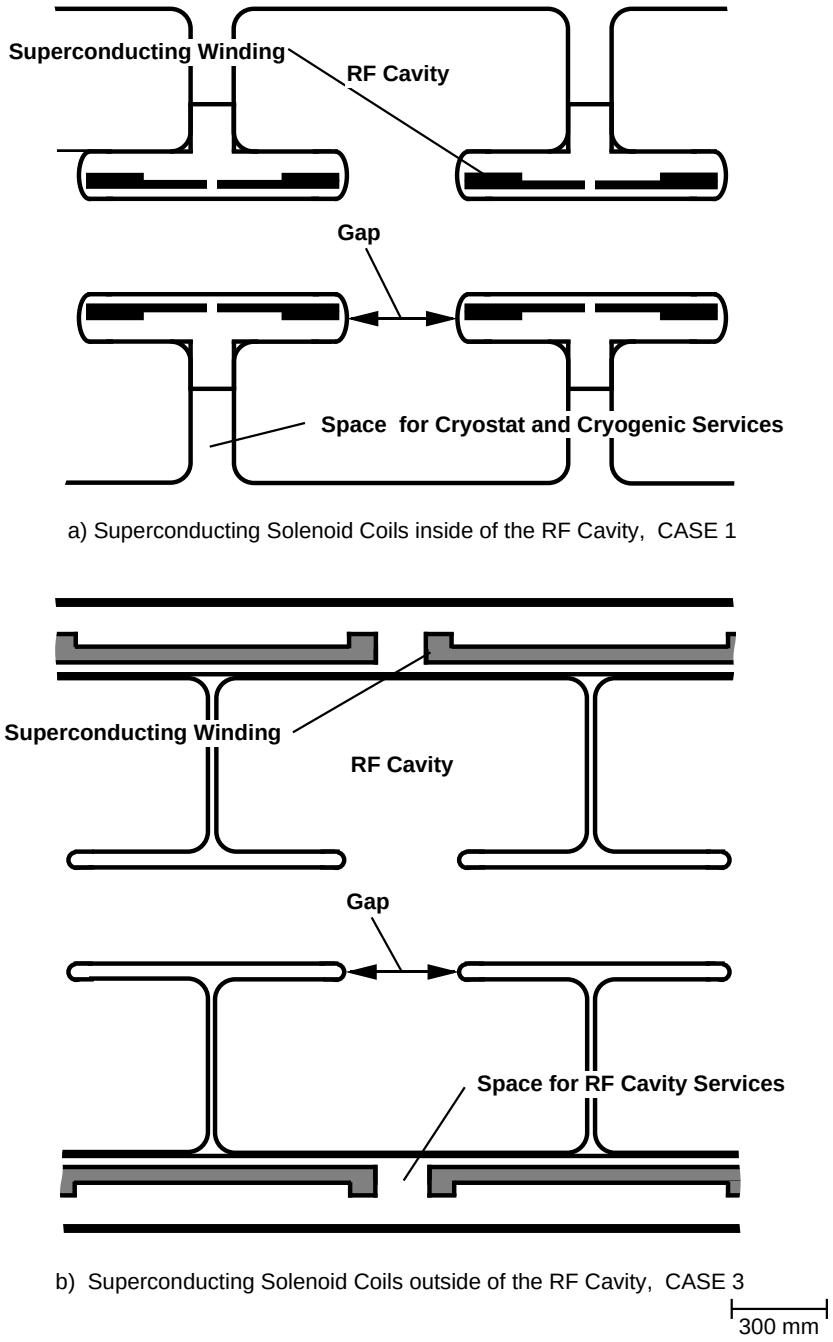


Figure 5.8: A schematic representation of superconducting solenoid scenarios for the phase rotation cavities

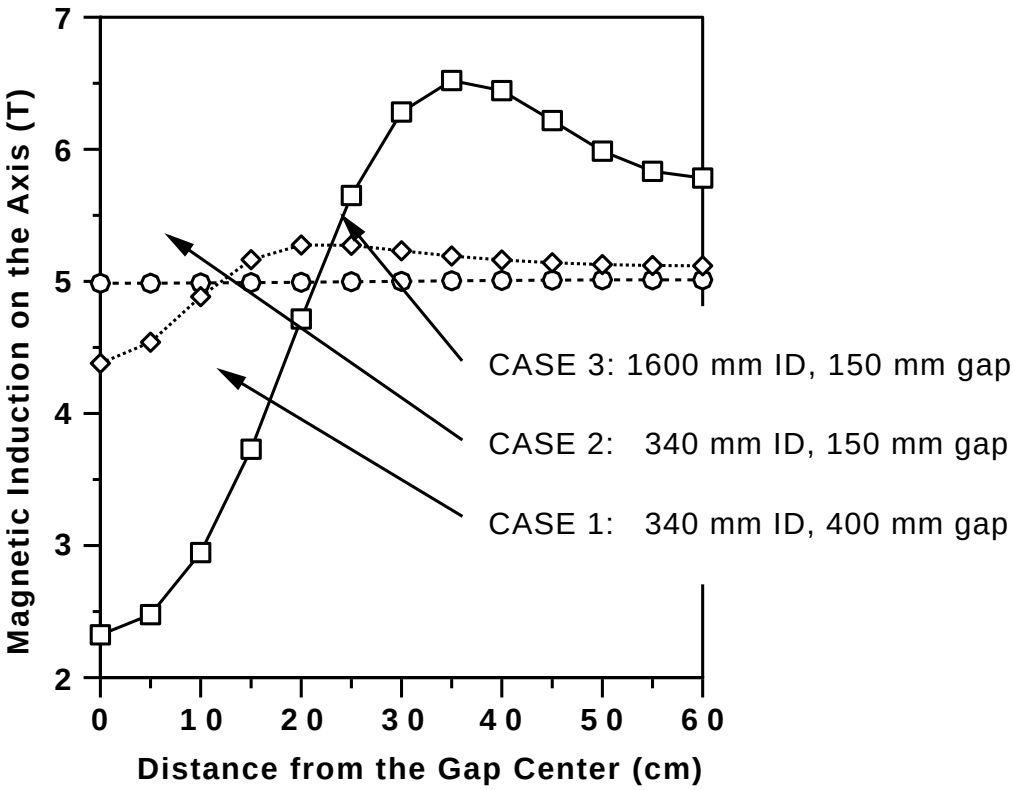


Figure 5.9: Magnetic induction along the rf cavity axis for three solenoid magnet and rf cavity configurations

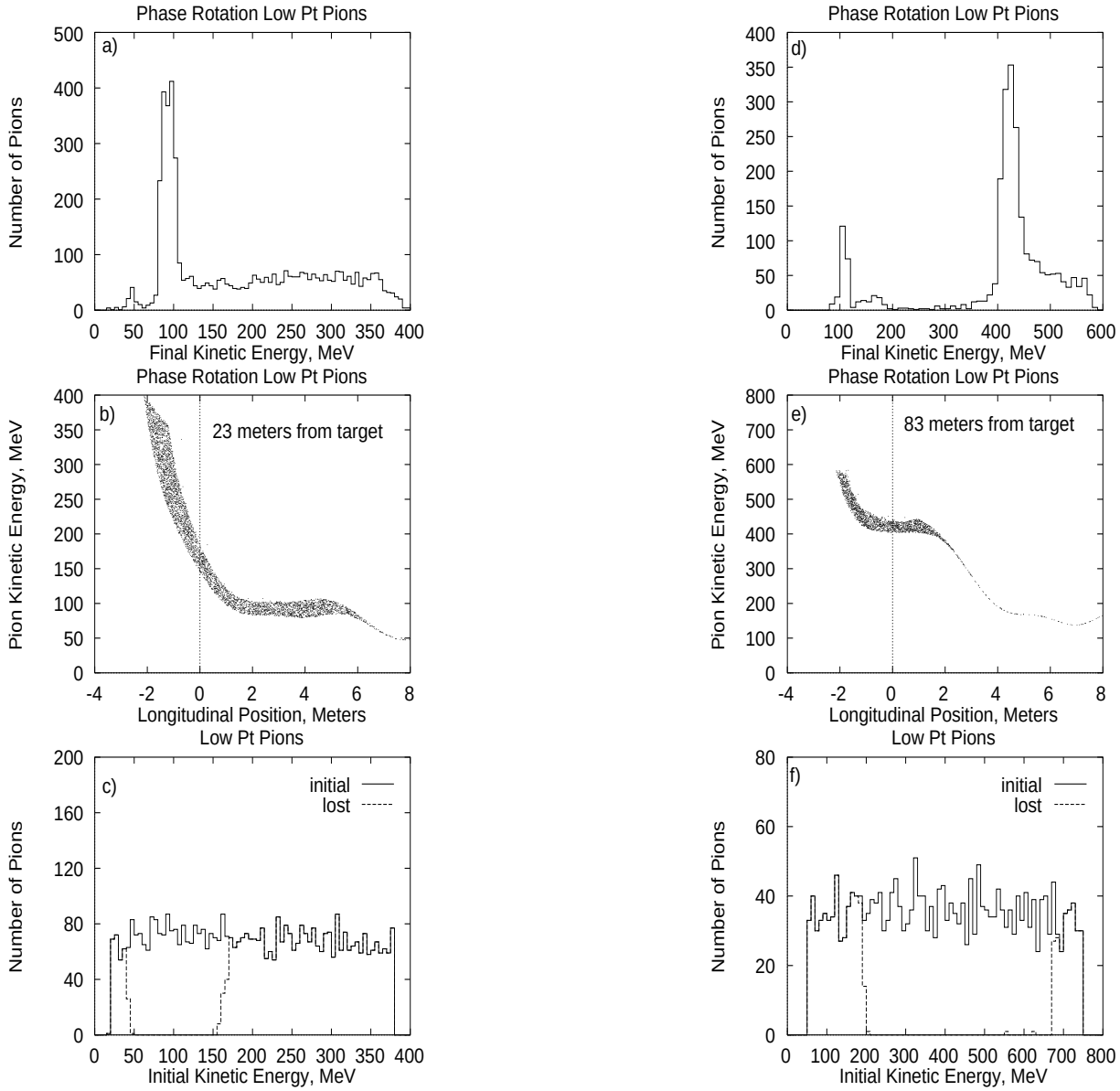


Figure 5.10: The results of particle dynamics modeling of pion capture immediately following the proton target. a), b), and c) depict results in which low-kinetic energy pions are collected, while d), e) and f) depict the corresponding results for the high-kinetic energy case.

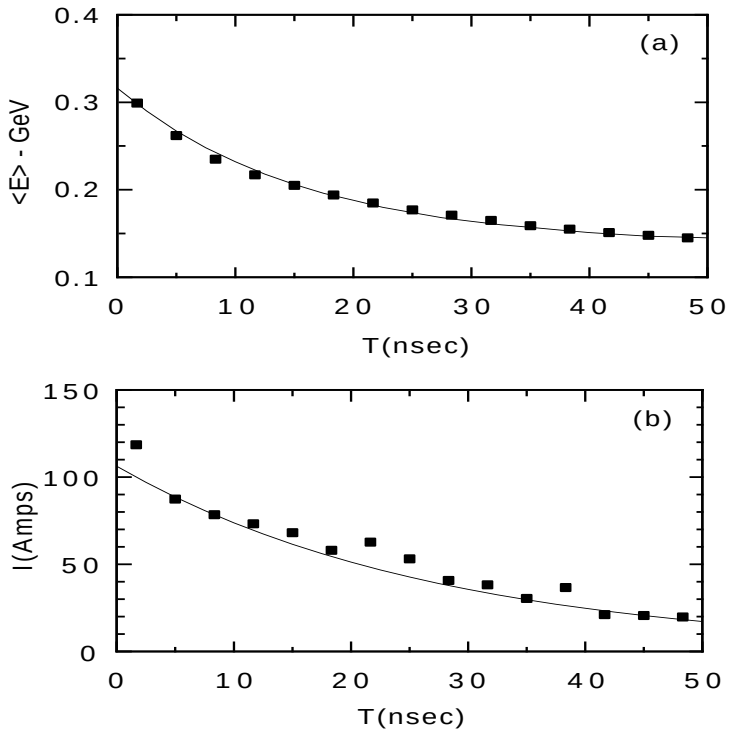


Figure 5.11: (a) Mean muon energy and (b) muon current z = 24 m from the production target. Solid square symbols are Monte Carlo results, the solid lines are exponential fits; $\mathrm{E(GeV)} = 0.137 + 0.179\, e^{-\mathrm{T(ns)}/15.8}$ and $\mathrm{I(A)} = 106.2\, e^{-\mathrm{T(ns)}/27.5}$.

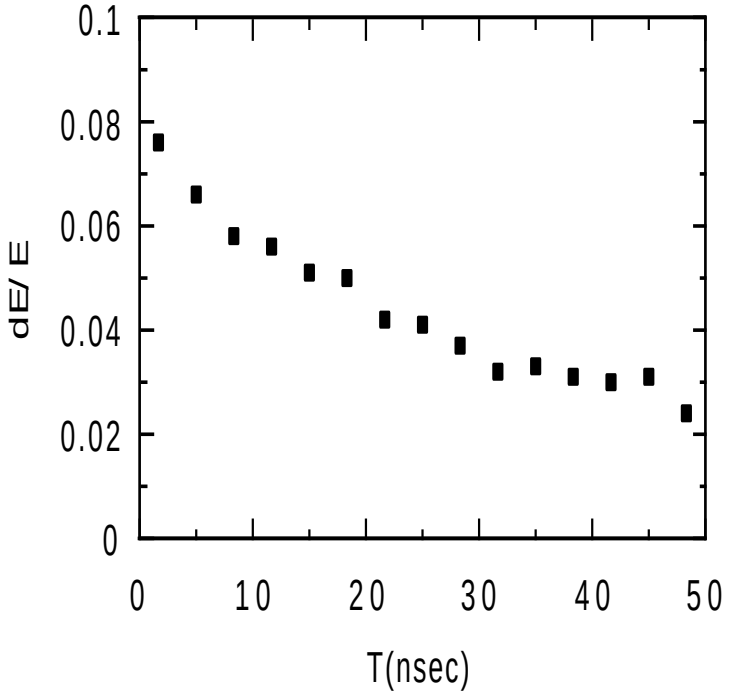


Figure 5.12: Instantaneous rms energy width of muons z = 24 m from the target.

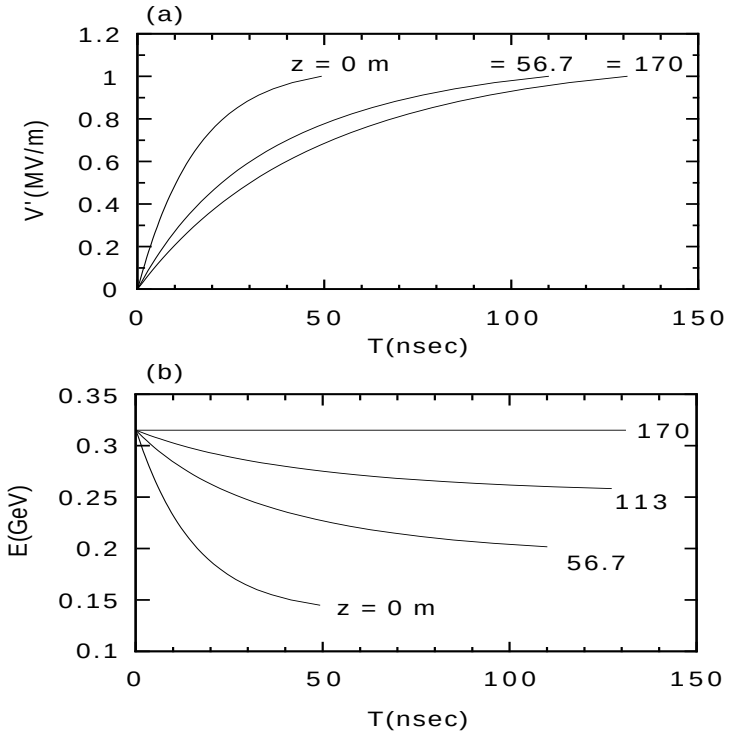


Figure 5.13: (a) Accelerator gap voltage wave shape at z = 0, 56.7 and 170 m and (b) mean muon energy at z = 0, 56.7, 113 and 170 m. z = 0 is 24 m from the production target.

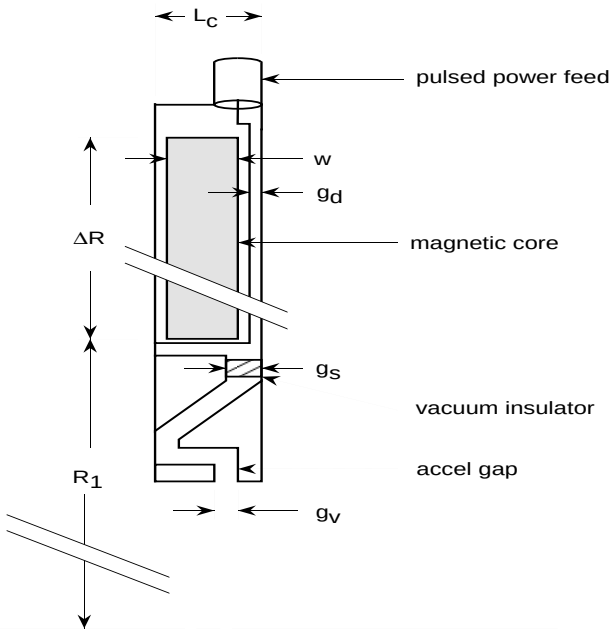


Figure 5.14: A schematic of an induction acceleration cell.

Table 5.6: Induction accelerator parameters for two cases; (a) input muon spectrum 0.145 to 0.316 GeV and (b) 0.244 to 0.720 GeV.

Parameter	Units	Value	
		(1)	(2)
Voltage gradient, V′	MV/m	1	
Cell voltage, V	kV	50	
Cell length, Lc	cm	1.88	1.85
E_{min}	GeV	.145	.244
E_{max}	GeV	.316	0.720
Accel length	m	171	476
Number of cells		3642	10302
Cell voltage rise time	ns	46.8	44.1
Pulse length	ns	131	114
Volt seconds per cell	$10^{-3}V \times s$	4.35	3.65
Rep rate	Hz	30	
Maximum core flux swing	T	2.5	
Average Bdot	$T/\mu s$	19.1	22.0
Core loss	kJ/m^3	3.77	4.34
Core axial width, w	cm	1.34	1.38
Inside radius of magnetic cores, R_1	m	0.30	
Core outer radius, R_2	m	0.462	0.428
Dielectric gap width, g_d	mm	2.73	1.89
Dielectric gap field stress	MV/cm	0.183	0.264
Vacuum insulator width, g_s	cm	1.0	
Accel gap width, g_v		0.5	
Vacuum insulator field stress limit	kV/cm	50	
Accel gap field stress limit	kV/cm	100	
Dielectric gap field stress limit	MV/cm	0.5	0.55
Insulator shielding margin	mm	1.11	0.717
Cell capacitance	nF	8.56	9.30
Core leakage resistance	Ohms	8.60	7.34
Core leakage current	kA	3.61	3.97
Vol. of core matl.(2605SC) per cell	$10^{-3}m^3$	3.90	3.03
Core weight per cell	kgm	28.5	22.2

Parameter	Units	Value		
		(1)		(2)
Core energy loss per cell per pulse	J	14.7		13.2
Gap capacitance energy per cell per pulse	J		9.43	9.93
Mismatch energy per cell per pulse	J	13.9		15.7
Total energy per cell per pulse	J	38.1		38.7
Total power per cell	kW	1.14		1.16
Core power loss	MW	1.60		4.07
Capacitance charging power	MW	1.03		3.07
Mismatch power	MW	1.52		4.84
Total power	MW	4.16		12.0
Dielectric fluid temp increase	K		10.0	
Flow rate of dielectric fluid	gms/sec-cell	24.8		22.2
Max core temp increase	K	10.7		10.9

Bibliography

[1] P. Wilson, SLAC, Private Communication

[2] W.C. Turner, *An Induction Linac Approach to Phase Rotation of a Muon Bunch in the Production Region of a $\mu^+\mu^-$ Collider*, LBL-38009(1995). Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop, Ed. J.C. Gallardo, AIP Press, to be published.

[3] Allied-Signal, Parsippany, New Jersey.

[4] C.H. Smith, *Applications of Amorphous Magnetic Materials at Very-High Magnetization Rates*, Proc. of Magnetism and Magnetic Materials Conference, Boston (1989).

[5] S. Yu et al., *Relativistic-Klystron Two-Beam Accelerator Based Power Source for a 1 TeV Center-of-Mass Next Linear Collider Preliminary Design Report*, UCRL-ID-119906 (1995).

Collaborators

- H. G. Kirk, (BNL) Editor
- M. Green, (LBNL)
- A. Moretti, (FermiLab)
- W. C. Turner, (LBNL)
- Y. Zhao, (BNL)

List of Figures

List of Tables

Chapter 6

IONIZATION COOLING

Contents

6.1 Theory of Ionization Cooling

6.1.1 Introduction

In order to generate sufficient muons for the collider, it is necessary to capture a very large fraction of the pions created at the target. These pions, and the muons into which they decay, are then necessarily very diffuse (i.e. they have a very large emittance). In order to achieve the required luminosity of $10^{35} cm^{-2}\, s^{-1}$ at 2+2 TeV, it is necessary to reduce the transverse emittance in both the x and y dimensions by a factor of ≈300 and the longitudinal emittance by a factor of ≈10. This represents a reduction of the overall 6-dimensional phase space by a factor of $\approx 10^6$. Therefore, it is essential to provide some means for cooling the muon beams.

The large mass of the muon compared to that of the electron prevents cooling by radiation damping, while the short lifetime of the muon prevents conventional stochastic or electron cooling. Fortunately, the process of ionization cooling can be used[1]-[4]. Because of their long interaction length this is possible only for muons. In this process the muon loses transverse and longitudinal momentum by dE/dx in a material and then has the longitudinal momentum (but not the transverse momentum) restored in a subsequent *rf* cavity. The combined effect is to reduce the beam divergence and thus the emittance of the beam. The process is complicated by the simultaneous presence of multiple scattering in the material, which acts as a source of *heat* and increases the emittance. The cooling effect can dominate for low Z materials in the presence of strong focussing fields. One solution being considered for the collider is to use absorbers made of lithium, beryllium, or liquid hydrogen inside a lattice of solenoid magnets. The absorber provides the energy loss, while the large aperture solenoids provide the required focussing.

The muons considered here lose energy primarily because of electromagnetic interactions with the atomic electrons. The rate of energy loss -dE/dx is shown as a function of muon energy for several materials in Fig. 6.1. The energy loss falls dramatically as the particle energy increases from very small values. It reaches a minimum value for muons with energy around 300 MeV. Above this is the region of relativistic rise, where the energy loss increases very gently.

The basic principle illustrating ionization cooling of the transverse emittance is shown in Fig. 6.2. A diffuse beam of muons is focused onto a block of material. Muons travelling at an angle through the material lose both transverse and longitudinal components of momentum.

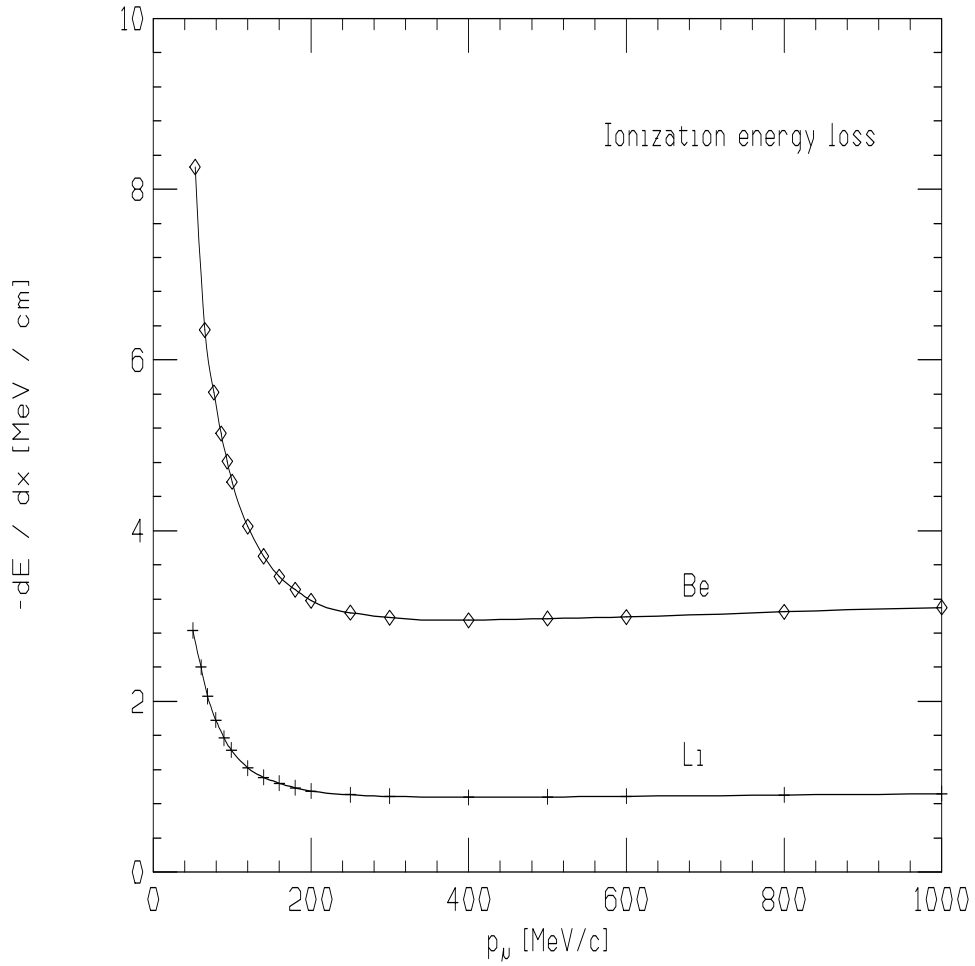


Figure 6.1: dE/dx as a function of muon momentum for Li and Be.

Note that any part of the dE/dx curve shown in Fig. 6.1 could be used for transverse cooling, although the focussing will be weaker at higher momentum. The beam then enters an accelerator cavity where the longitudinal momentum is returned to its starting value. However, since the transverse momentum is not replaced, the divergence of the beam is reduced. If the focusing and scattering properties of the material are such that the beam size does not increase, the geometric emittance of the beam is reduced by this process.

In order to cool the longitudinal emittance it must be arranged that the higher energy particles in the beam lose more energy than the lower energy particles. On the dE/dx curve shown in Fig. 6.1 this only occurs for muon energies greater than $\approx 400\,\mathrm{MeV}$. This natural longitudinal cooling is straightforward, but inefficient. A more practical idea, shown in Fig. 6.3, is to introduce dispersion into the beam, so that the muons receive a transverse displacement proportional to their deviation from the mean momentum. Then a wedge shaped absorber can be used to cause the higher momentum muons to lose more energy, and

thus reduce the momentum spread in the beam.

6.1.2 Transverse Emittance Cooling

It will be useful to develop a simple model of emittance cooling in order to understand its dependence on various physical parameters, the optimum materials to use, and to estimate limits on the minimum achievable emittance. Let us adopt a coordinate system where z is along the direction of particle motion, and x and y are the orthogonal transverse directions. In the general case the geometric emittance is defined statistically as

$$\epsilon_x^2 = < x^2 >< \theta^2 > - < x\,\theta >^2 \tag{6.1}$$

where θ is the divergence angle of the particle trajectory projected onto the x-z plane, and the expectation values are taken over all the particles in the beam. We define the normalized

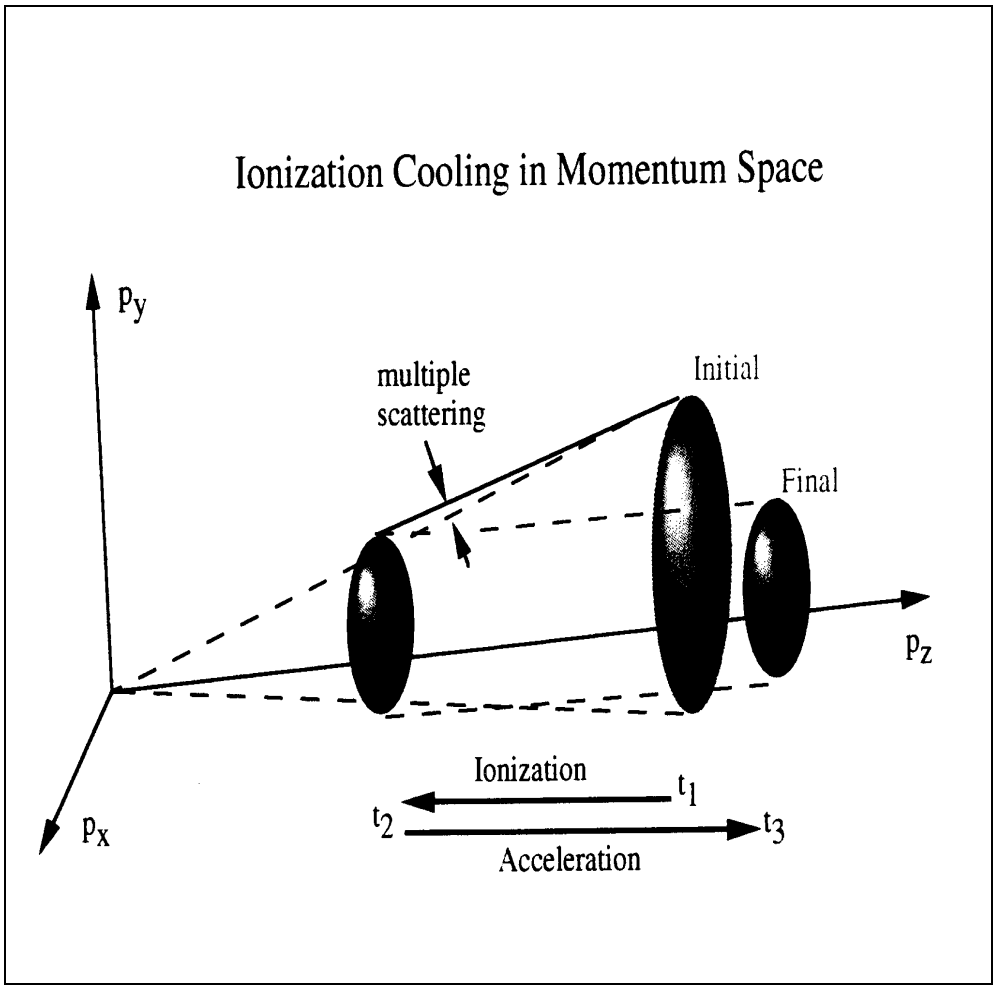


Figure 6.2: Basic principle of ionization cooling of transverse emittance.

transverse emittance

$$\epsilon_{xN} = \beta\,\gamma\,\epsilon_x \tag{6.2}$$

where β, γ are the usual relativistic velocity and energy expressions. It is convenient to consider the change in the normalized emittance as the particles proceed through the absorbing material. Cooling of the normalized emittance takes place entirely in the absorber. The cooling arises because of the decrease of the $\beta\gamma$ factor in Eq. 6.2. Heating also occurs in the absorber, primarily because of multiple scattering. This will enter as an increase in the factor ϵ_x in Eq. 6.2. The subsequent accelerator section plays no role in this analysis, since the normalized emittance is unchanged by the acceleration process.

Now consider the change in ϵ_{xN} as the beam travels along the z direction into the material.

$$\frac{d\epsilon_{xN}}{dz} = \epsilon_x \frac{d(\beta\,\gamma)}{dz} + \beta\,\gamma \frac{d\epsilon_x}{dz} \tag{6.3}$$

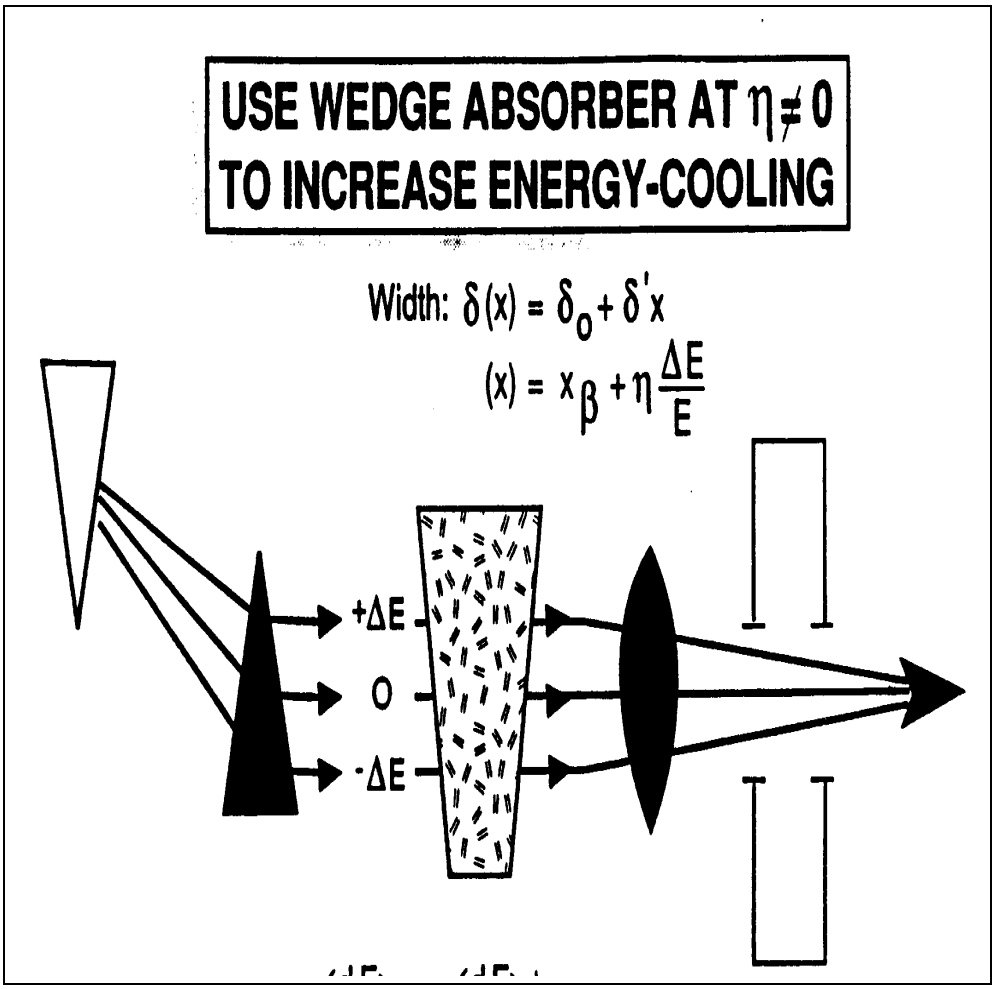


Figure 6.3: Basic principle of longitudinal cooling using a wedge absorber.

It is possible to relate the first term in Eq. 6.3 to the ionization cooling

$$\frac{d\epsilon_{xN}}{dz}(cool) = -\frac{1}{\beta^2}\frac{\epsilon_{xN}}{E}\left|\frac{dE}{dz}\right| \tag{6.4}$$

where E is the total energy of the muons. The second term in Eq. 6.3 gives rise to heating, which acts to increase the emittance as the beam proceeds through the material.

$$\begin{aligned}\frac{d\epsilon_{xN}}{dz}(heat) &= \frac{\beta\gamma}{2\epsilon_x}\Big[< x^2 > \frac{d}{dz} < \theta^2 > + < \theta^2 > \frac{d}{dz} < x^2 > \\ &\quad -2 < x\theta > \frac{d}{dz} < x\theta >\Big]\end{aligned} \tag{6.5}$$

At this point the simple theory derived above is exact, except for the neglect of additional processes such as muon bremsstrahlung and muon nuclear interactions that we assume are not important in the energy range under consideration.

Now let us assume that the cooling is taking place near a beam waist, so that we can neglect the effects of correlations in the beam parameters. If the focusing is sufficiently strong, we may also assume that the growth in the transverse size of the beam is negligible. It can be shown [5] that this will be the case, provided that the conditions

$$\sigma_{xo}^2 \gg \frac{\theta_c^2 L}{2\,\omega^2} \tag{6.6}$$

and

$$\sigma_{xo}^2 \gg \frac{\theta_c^2}{4\,\omega^3} \tag{6.7}$$

are satisfied, where L is the length of the absorber material, ω is the focusing strength parameter [m^{-1}] , and σ_{xo} is the size of the beam entering the absorber. θ_C^2 is a parameter [m^{-1}] related to the strength of the multiple scattering

$$\theta_C = \frac{E_S}{p\,c\,\beta}\frac{1}{\sqrt{L_R}} \tag{6.8}$$

where $\mathrm{E}_S = 15$ MeV, p is the particle's momentum, and L_R is the radiation length for the scattering medium. Thus, neglecting the second and third terms in Eq. 6.5, we find that

$$\frac{d\epsilon_{xN}}{dz}(heat) \approx \frac{\beta\,\gamma}{2\,\epsilon_x} < x^2 > \frac{d}{dz} < \theta^2 > \tag{6.9}$$

If we use the relation

$$< x^2 >= \beta_\perp\,\epsilon_x \tag{6.10}$$

from betatron focusing theory, we can write Eq. 6.9 in the form

$$\frac{d\epsilon_{xN}}{dz}(heat) \approx \beta\,\gamma\,\frac{\beta_\perp}{2}\frac{d}{dz} < \theta^2 > \tag{6.11}$$

where $\beta_\perp$ is the betatron focusing parameter. The simplest expression for the change in θ^2 can be found by using the Rossi-Greisen model[6] of multiple scattering

$$\theta \approx \frac{E_S}{p\,c\,\beta}\sqrt{\frac{z}{L_R}} \tag{6.12}$$

Although this expression is adequate for developing the simple scaling model, we use more accurate Gaussian models and Moliere scattering theory in the Monte Carlo modelling described in section 2. With this approximation we can rewrite Eq. 6.11 as

$$\frac{d\epsilon_{xN}}{dz}(heat) \approx \frac{\beta_\perp}{2}\frac{E_S^2}{\beta^3\,E\,m\,c^2}\frac{1}{L_R} \tag{6.13}$$

where m is the mass of the muon.

We notice from Eqs. 6.4 and 6.13 that the rate of cooling decreases as the beam proceeds through the absorber, while the rate of heating increases. At some point the rates become equal and the absorber must end or the emittance will begin to increase again. If we consider the case when the two rates are equal, we find that the corresponding value of the emittance is

$$min\,\epsilon_{xN} \approx \frac{\beta_\perp\,E_S^2}{2\,\beta\,m\,c^2\,L_R|\frac{dE}{dz}|} \tag{6.14}$$

This is the minimum achievable emittance for the given material and focusing conditions. Note that the dependence on the focusing enters through the parameter $\beta_\perp$ in the numerator, while the dependence on the absorbing material enters through the product $L_R\,dE/dz$ in the denominator.

Table 6.1 gives the relevant parameters for some materials that might be useful for ionization cooling. The third column gives the energy loss for a minimum ionizing particle. The last column gives the coefficient of $\beta_\perp$ in Eq. 6.14 for a relativistic particle ($\beta = 1$). Thus, for example, if the focusing produces a value of $\beta_\perp = 1$ cm, the estimated minimum emittance using Li would be ≈ 80 mm mr. The minimum achievable emittance would be a factor of 2 smaller if similar focusing could be used with liquid hydrogen.

It should be pointed out that transverse ionization cooling is always accompanied by some heating of the longitudinal emittance. This arises because ionization energy loss is a statistical process and there is a spread of energy losses (straggling) around the mean value. The design of the optimized cooling system reduces 6-dimensional phase space, as discussed in section 6.3.

In an alternate approach to estimating the equilibrium emittance, Skrinsky and Parkhomchuk [1],[3] have considered the transverse damping rate arising from the "frictional force" due to the ionization energy loss dE/dz. They derived the expression

$$min\,\epsilon_{xN} \approx\ Z\frac{L_C}{L_I}\frac{m_e}{m}\beta_\perp \tag{6.15}$$

Table 6.1: Materials for ionization cooling

Material	ρ	dE/dx	L_R	cof. of $\beta_\perp$
	[g/cm^3]	[MeV/cm]	[cm]	[mm mr/cm]
liq. H_2	0.071	0.286	890.	42
liq. He	0.125	0.242	756.	59
LiH	0.82	1.34	102.	78
Li	0.534	0.875	155.	79
Be	1.848	2.95	35.3	103
:CH_2:	0.93	1.93	47.9	116
C	2.265	3.95	18.8	144
liq. N_2	0.807	1.47	47.	155
Al	2.70	4.36	8.9	275

where L_C is the Coulomb logarithm, L_I is the Bethe-Bloch logarithm, and m_e is the mass of the electron. This expression gives an equilibrium emittance of 120 mm mr for Li with $\beta_\perp = 1$ cm , compared with the value of 80 that we found with the previous method. This indicates the level of uncertainty in this type of analysis. Accurate predictions for the amount of emittance reduction are derived using the Monte Carlo calculations described in section 6.2.

6.1.3 Longitudinal Emittance Cooling

We define the normalized longitudinal emittance as

$$\epsilon_{zN} = \beta_z \, \gamma \, \delta \, \sigma_z \tag{6.16}$$

where σ_z is the bunch length,

$$\delta = \frac{\sigma_{p_z}}{p_z} \tag{6.17}$$

is the fractional momentum spread, and σ_{p_z} is the r.m.s. momentum spread in the beam. Now consider the change in the emittance as we take a step dz into an absorber material

$$\frac{d}{dz}\epsilon_{zN} = \beta\,\gamma\,\delta\,\frac{d}{dz}\sigma_z + \beta\,\gamma\,\sigma_z\,\frac{d}{dz}\delta + \delta\,\sigma_z\,\frac{d}{dz}(\beta\,\gamma) \tag{6.18}$$

where we have assumed the motion is predominantly along z. For relativistic beams changes in the bunch length will be small, so we drop the first term in Eq. 6.18. Then, using relations among the relativistic variables, it is possible to rewrite Eq. 6.18 in the form

$$\frac{d}{dz}\epsilon_{zN} \approx \frac{\beta\,\gamma\,\sigma_z}{p_z}\,\frac{d}{dz}\sigma_{p_z} \tag{6.19}$$

We identify three effects that modify the energy spread in the beam. Firstly, because of the curvature of the dE/dx curve shown in Fig. 6.1, particles with different energies lose different amounts of energy. The fractional loss in energy in the step dz gives

$$\frac{d}{dz}\sigma_{p_z} = \frac{\sigma_E}{\beta\, c}\,\frac{d}{dE}\left(\frac{dE}{dz}\right) \tag{6.20}$$

This term increases the energy spread for particle energies below minimum ionization and decreases it (cools) for energies above it. However, the rate of cooling is very small since the slope of the dE/dx curve in the region of relativistic rise is small. For the materials listed in Table 6.1 the rate of change of dE/dx over the energy range 600 - 800 MeV varies from 0.41 x 10^{-4}/cm for liquid hydrogen to 4.46 x 10^{-4}/cm for aluminum, in roughly the reverse order as their efficiency for transverse cooling. A longitudinal cooling scheme based on this effect, which achieves the requirements of the muon collider, would be at least an order of magnitude longer than the scheme presented in section 3.

Statistical fluctuations in the energy loss over a fixed distance of material is known as straggling. This leads to a second effect, which increases the amount of energy spread[7]

$$\frac{d}{dz}\sigma_{p_z} = \frac{K_s}{2\,\beta\, c\,\sigma_E}\,\gamma^2\,\left(1-\frac{1}{2}\beta^2\right) \tag{6.21}$$

The constant

$$K_s = 4\,\pi\,(r_e\, m_e\, c^2)^2\,\frac{N_A\, Z\,\rho}{A} \tag{6.22}$$

where r_e is the classical radius of the electron, m_e is the mass of the electron, c is the speed of light, N_A is Avogadro's number, and $\{Z, \rho, A\}$ are the {atomic number, density, atomic weight} of the material. Note that the growth in energy spread is proportional to γ^2, so cooling at low energies is preferred.

It is also possible to cool the beam longitudinally by placing a transverse variation in absorber density or thickness in a region of non-zero dispersion, as shown in Fig. 6.3. We refer to this case as using a *wedge absorber*. We can write the fractional change in momentum spread as

$$\frac{d}{dz}\sigma_{p_z} \approx \frac{1}{\beta\, c}\,\frac{dE}{dz}\,\frac{\eta\,\delta}{\alpha\, L_o} \tag{6.23}$$

where the dispersion $\eta = dx/d\delta$, the wedge angle $\alpha = dx/dz$, and L_o is the thickness of the wedge at x = 0. The longitudinal cooling will be associated with heating in the transverse phase space due to multiple scattering in the absorber.

6.1.4 Transport Equation Approach

In the transport equation approach a differential equation is set up that describes the evolution of a distribution function for the variables of the theory, such as angle, position, and

energy. Fernow and Gallardo[8] considered the distribution $W(y, \theta; z)$ in transverse position y and deflection angle θ for the case of constant energy and an external focusing force. They confirmed that the beam size remains constant for the case of a strong magnetic field and derived conditions for the validity of Eq. 6.13 for the increase in emittance due to heating.

Vsevolozhskaya[9] considered the transport equation for the more general case where energy loss and reacceleration are also considered. This approach combines changes in transverse and longitudinal phase space into a unified formalism. Consider a beam of muons traversing a cooling section consisting of blocks of absorber, accelerator sections, and external focusing. The function $P(r, \theta, E; z)\, d^2r\, d\theta\, dE$ gives the distribution of particles in the beam that have transverse radius between r and r+dr, polar angle between θ and $\theta + d\theta$, and energy between E and E+dE at a distance z along the section. Then P satisfies the transport equation

$$\frac{\partial P}{\partial z} + \theta\frac{\partial P}{\partial r} + (\theta\frac{\xi_o}{pv} + kr)\frac{\partial P}{\partial \theta} - 2\xi_o\frac{P}{pv} + (\xi_o - \xi)\frac{\partial P}{\partial E} = \frac{E_k^2}{4p^2v^2L_R}\frac{\partial^2 P}{\partial \theta^2} \tag{6.24}$$

where ξ_o is dE/dx due to acceleration, ξ is -dE/dx due to ionization energy loss, k is the strength of the external focusing, and $E_k \approx 20$ MeV is a characteristic scattering energy.

6.1.5 Comparison of Focusing Methods

Ionization cooling requires some form of focusing to limit emittance growth due to an increase in the size of the beam. Particles in the beam undergo betatron oscillations given in a region of constant focusing strength by

$$x(z) = \sqrt{\epsilon\,\beta_\perp}\, cos\left(\frac{z}{\beta_\perp} - \psi_o\right) \tag{6.25}$$

where ψ_o is the phase advance. The focusing strength is defined as

$$k\,[m^{-2}] = \frac{0.3\ g\,[T/m]}{p\,[GeV/c]} \tag{6.26}$$

where g is the field gradient. The betatron parameter is related to the focusing strength by

$$\beta_\perp = \frac{1}{\sqrt{k}} \tag{6.27}$$

while the focal length of an element of length L is

$$f \approx \frac{1}{kL} \tag{6.28}$$

We have considered four possible focusing elements for a linear cooling section.

Quadrupole FODO cell

Quadrupoles can be made with gradients g≈ 2 T/cm. By its nature a quadrupole focuses in one transverse dimension and defocuses in the other. The FODO lattice consists of equal strength, horizontally and vertically focusing quadrupoles separated by a drift distance. A cooling channel inside a FODO (or FDO) lattice could be a suitable arrangement for a higher momentum muon beam, although this has not been studied in detail. Some disadvantages of a quadrupole focusing system is that at least a triplet of elements are required for symmetric focusing and the defocusing action of the first quadrupole reduces the angular acceptance of the lattice.

Solenoid

A solenoid produces a central longitudinal magnetic field given by

$$B_z = \mu_o \, n \, I \tag{6.29}$$

where I is the current and n is the number of turns per unit length. A particle travelling parallel to the axis of a solenoid at a radius r_o receives an azimuthal momentum kick

$$p_\phi = \frac{e \, B_z \, r_o}{2} \tag{6.30}$$

while crossing the fringe field region at the end of the solenoid. This produces a radial force in the central part of the solenoid, causing the particle to follow a helical trajectory with radius of curvature

$$R = \frac{r_o}{2} \tag{6.31}$$

Thus the radius of a single particle oscillates between r_o and the axis of the solenoid. The envelope of a matched beam, however, is constant inside the solenoid. A short solenoid can be used as a lens with focal length

$$f \approx \frac{4 \, p^2}{e^2 \, B^2 \, L_s} \tag{6.32}$$

It is also possible to confine the beam in a long solenoidal channel. The angle of rotation around the axis of the solenoid of a particle at a distance z along the central portion of the solenoid is given by

$$\theta = \frac{e \, B_z \, z}{p_z} \tag{6.33}$$

The betatron amplitude is given by

$$\beta_\perp = \frac{2 \, p_z}{e \, B_z} \tag{6.34}$$

These relations can be significantly altered if space charge effects in the beam are important[10].

Solenoid FOFO cell

The solenoid FOFO cell consists of a series of short solenoids, separated by a distance d. The system focuses both transverse planes simultaneously. We can determine the basic properties of the cell in the thin lens approximation in the usual way by comparing the transfer matrix for the cell with the standard form of the Courant-Snyder matrix[11]-[12]. The phase advance per cell is

$$cos\,\psi = 1 - \frac{d}{2\,f_s} \tag{6.35}$$

where f_s is the focal length of an individual solenoid. The maximum value of the betatron amplitude occurs at the midplane of the solenoid

$$\beta_{max} = d\frac{\kappa}{\sqrt{\kappa - \frac{1}{4}}} \tag{6.36}$$

where $\kappa = f_s/d$. For stable motion we must have $\kappa > 1/4$, or $f_s > d/4$. The minimum value occurs at the midpoint between the solenoids where

$$\beta_{min} = d\sqrt{\kappa - \frac{1}{4}} \tag{6.37}$$

The minimum momentum that can be stably transmitted through the FOFO lattice is

$$p_{min} = \frac{e\,B\sqrt{L_s\,d}}{4} \tag{6.38}$$

Lithium lens

In the lithium lens an axial current produces an azimuthal magnetic field B at a radius r inside the lens given by

$$B = \frac{\mu_o\,J\,r}{2} \tag{6.39}$$

where J is the current density inside the lithium. Gradients ≈ 7 T/cm have been achieved at FNAL[13]-[14] and the possibility of reaching gradients as high as 20 T/cm has been considered[15]. The particle follow a helical orbit around the lens axis. This lens focuses both transverse planes simultaneously. The focusing strength is

$$k = \frac{e\,\mu_o\,J}{2\,p} \tag{6.40}$$

The betatron amplitude is

$$\beta_\perp = \sqrt{\frac{p\,R}{0.3\,B_R}} \tag{6.41}$$

where R is the radius of the lithium rod and B_R is the value of the magnetic field at R. The largest incident angle of a trajectory that will be confined in the rod is

$$\theta_o = \sqrt{\frac{0.3\,R\,B_R}{p}} \tag{6.42}$$

6.1.6 Emittance Exchange

Since transverse cooling of a moving beam is always associated with longitudinal heating and longitudinal cooling is always associated with transverse heating, it is useful to consider each of the steps in the cooling scenario from the point of view of emittance exchange. However, we should emphasize from the start that there is no Liouville theorem for 6-dimensional phase space conservation in the presence of materials and dissipative forces. The reduction in one phase space dimension and the increase in another one depend quantitatively on the actual configuration that is used. Fig. 6.4 shows a mixed phase space diagram where the transverse size of the beam is plotted along the horizontal axis and the relative momentum spread is plotted along the vertical axis. Consider an initial elliptical beam (A) in this phase space. If the beam enters a dispersive region, to lowest order, the beam size grows while

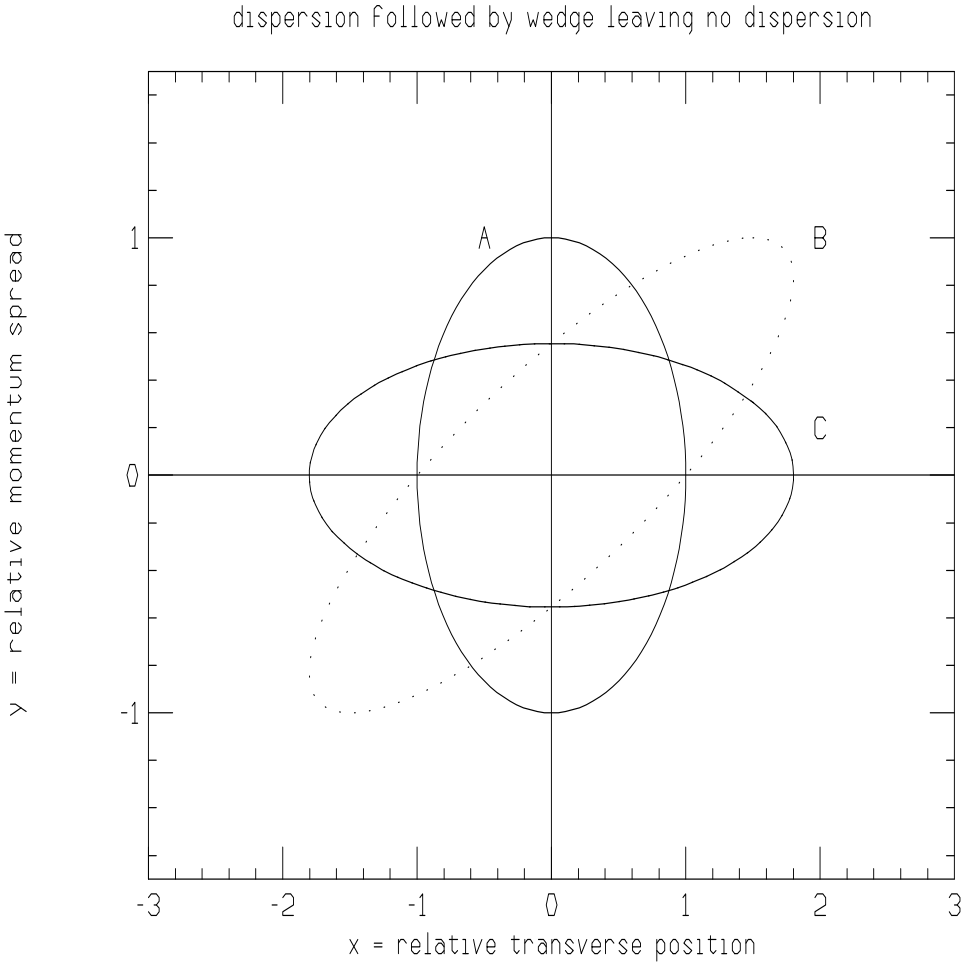


Figure 6.4: Emittance exchange diagram. Horizontal axis is transverse position; vertical axis is momentum spread

the momentum spread remains constant. This produces the tilted ellipse (B) in the figure. If the beam enters a wedge at this point, the momentum spread decreases while the size remains approximately constant. This produces the flattened ellipse (C) in the figure. The important point is that this mechanism allows transverse and longitudinal emittance to be exchanged. For example, it is possible to cool more than necessary in one dimension and then exchange the cooling with another dimension, if that turns out to easier than directly cooling to the desired value.

6.2 Monte Carlo Modelling of Ionization Cooling

6.2.1 Introduction

The analytic expressions developed in the previous section show the dependence of emittances on various quantities, such as the muon momentum and magnetic field strengths. This is useful for rough optimization of cooling system parameters. The expressions should also be suitable for estimating achievable emittances, hopefully to within a factor of 3. However, this simple differential equation approach has a number of shortcomings:

- it only describes the average change in emittance
- it cannot give the distribution of particles in phase space
- it cannot show any possible correlations between the transverse and longitudinal phase space variables
- it doesn't describe the loss of beam particles
- it doesn't include the effects of additional processes, such as incoherent nuclear scattering, bremsstrahlung, direct pair production, and nuclear interactions
- it does not include errors in the cooling elements, such as magnet strengths and positions, etc.
- it does not include space charge effects

For this reason, accurate numerical estimates of the achievable transverse and longitudinal emittances and the particle losses for a given cooling scenario must be obtained using Monte Carlo simulations. An excellent overview of important issues in the simulation of ionization cooling has been given by Van Ginneken[16].

6.2.2 Energy Loss and Multiple Scattering

The most important processes for ionization cooling in the 10-300 MeV muon kinetic energy range are ionization energy loss off the atomic electrons and multiple Coulomb scattering off the nucleus of the absorber material. The mean rate of energy loss is described using the Bethe-Bloch formula. However, significant fluctuations (straggling) of the energy loss occur for each step through the material. The distribution of these fluctuations (Landau, Vavilov, or Gaussian) are determined by the muon velocity and a parameter (described below) that is proportional to the step size. Fluctuations in the scattering angle after any step is given by the Moliere distribution. The actual distributions in energy loss and scattering angle are modified by the presence of the focusing magnetic field in the absorber material. In addition, the rarer processes described in the following section can introduce energy-angle correlations in the distributions.

The mean rate of ionization energy loss is given by the Bethe-Bloch equation[17]. A charged particle (z = 1) traversing matter loses energy at a rate

$$-\frac{dE}{dz} = \frac{K\,Z\,\rho}{A\,\beta^2}\left[\frac{1}{2}ln\,\frac{2\,m_e\,c^2\,\beta^2\,\gamma^2\,T_{max}}{I^2} - \beta^2 - \frac{\delta}{2}\right] \tag{6.43}$$

where {Z,A,ρ ,I} are the {atomic number, atomic weight, density, ionization potential} of the material, $\{\beta, \gamma$,c$\}$ are the usual relativistic factors for the incident particle, and m_e is the electron rest mass. The constant K is given by

$$K = 4\,\pi\,N_A\,r_e^2\,m_e\,c^2 \tag{6.44}$$

where N_A is Avogadro's number and r_e is the classical radius of the electron. The maximum possible energy transfer to the atomic electron is given by

$$T_{max} = \frac{2\,m_e\,c^2\,\beta^2\,\gamma^2}{1 + 2\,\gamma\,\frac{m_e}{M} + (\frac{m_e}{M})^2} \tag{6.45}$$

where M is the mass of the incident particle. Fluctuations in the amount of energy loss are determined using the dimensionless ratio

$$\kappa = \frac{\xi}{T_{max}} \tag{6.46}$$

where the characteristic energy

$$\xi = \frac{2\,\pi\,e^4\,N_A\,Z\,\rho}{m_e\,\beta^2\,c^2\,A}\;t \tag{6.47}$$

is directly proportional to the spatial step size t. For $\kappa \leq 0.01$ the fluctuations can be described by the Landau distribution[18], provided[19] that the step is not allowed to be so

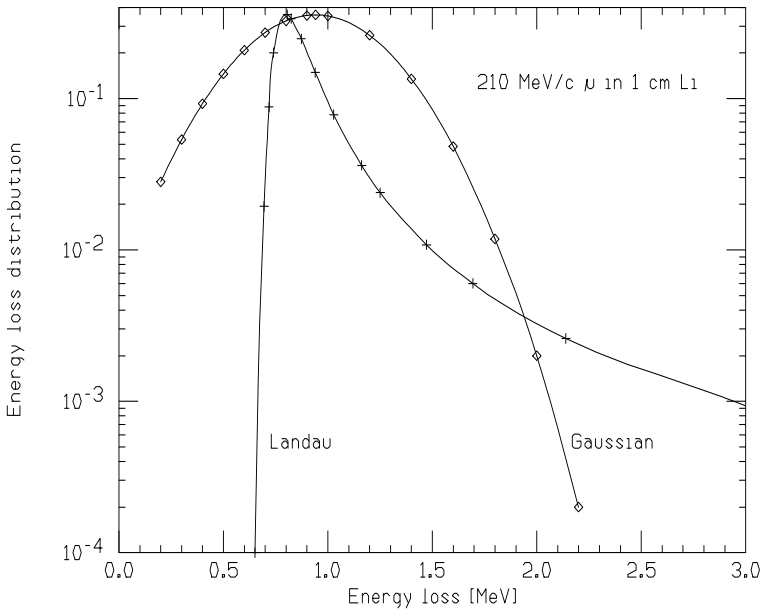


Figure 6.5: Landau distribution of energy loss in Li

small that the energy loss becomes comparable with the mean ionization potential I. Fig. 6.5 shows the Landau distribution for muons in 1 cm of Li. The mean value of the energy loss is the same as that given by the Bethe-Bloch formula, although the most probable energy loss is smaller than the mean value. A Gaussian distribution is shown on the same figure with a mean value given by the Bethe-Bloch formula and a standard deviation given by the Bohr formula (below).

For $0.01 < \kappa < 10$ energy loss is accurately described by the Vavilov distribution[20]. In addition to the value of κ , this distribution depends on the muon velocity . The lower limit of validity corresponds to characteristic energy losses on the order of the ionization potential. At very low values of κ the Vavilov and Landau distributions coincide. Fig. 6.6 shows the Vavilov distribution for muons in a 9 cm step of Li. The Gaussian distribution is obviously a better approximation here than for the 1 cm case. The Vavilov distribution was derived for a spin J = 0 particle. Van Ginneken[16] has calculated a correction to the distribution for a spin J = 1/2 particle, such as the muon, that causes a small increase in the width of the distributions.

A Gaussian distribution accurately describes the energy loss when $\kappa \geq 10$. The variance of the distribution, first calculated by Bohr[7], is

$$\sigma_E^2 = 4\,\pi\,(r_e\,m_e\,c^2)^2\,N_A \frac{Z\,\rho}{A}\,\gamma^2\,\left(1 - \frac{\beta^2}{2}\right)\,t \tag{6.48}$$

Note that the variance grows quadratically with the energy γ and linearly with the thickness

t.

Coherent multiple Coulomb scattering of the muon from the nucleus of the absorber material is described by the Moliere distribution[21]. The characteristic scattering angle is

$$\theta_c = \chi_c \sqrt{B} \tag{6.49}$$

The parameter χ_c is determined from

$$\chi_c^2 = 4\,\pi\,(r_e\,m_e\,c^2)^2 \frac{1}{(p\,c\,\beta)^2} \frac{N_A\,\rho\,Z\,(Z+1)}{A}\,t \tag{6.50}$$

and is directly proportional to the thickness. The dimensionless quantity B is determined from the equation

$$B - ln\,B = ln\,\Omega_o \tag{6.51}$$

where Ω_o is the mean number of collisions. Fig. 6.7 shows the Moliere distribution of scattering angles for muons crossing a 1 cm thickness of Li. The figure also shows a Gaussian distribution with standard deviation given by the Rossi formula, Eq. 6.12, with $E_S = 21.2$ MeV. The actual distribution of scattering angles is considerably wider than the Gaussian approximation for angles greater than θ_c.

The simulation code SIMUCOOL of Van Ginneken[16] introduces cut-off parameters for the energy loss and scattering angle. The corrected Vavilov distribution is used to simulate events below the cutoff energy loss on a statistical basis. Events with larger energy losses are treated as individual μe scattering events using the Bhabha formula. The binding energy of

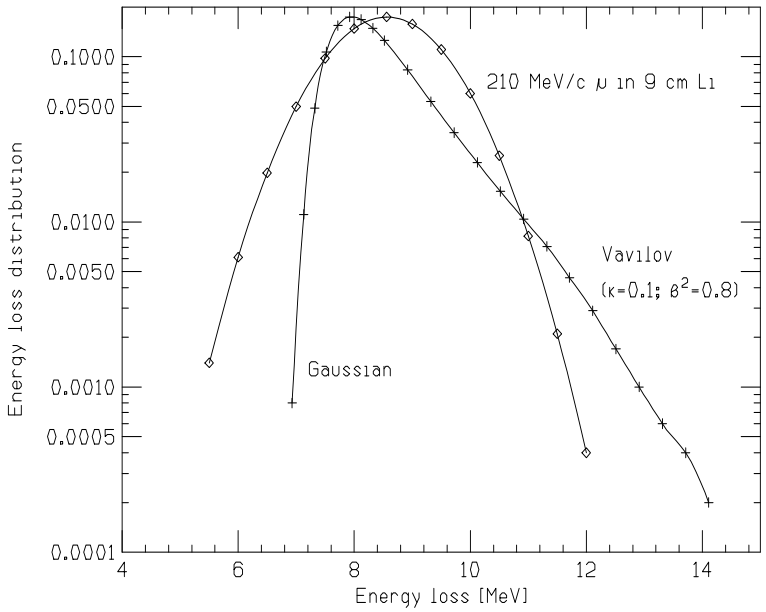


Figure 6.6: Vavilov distribution of energy loss in Li

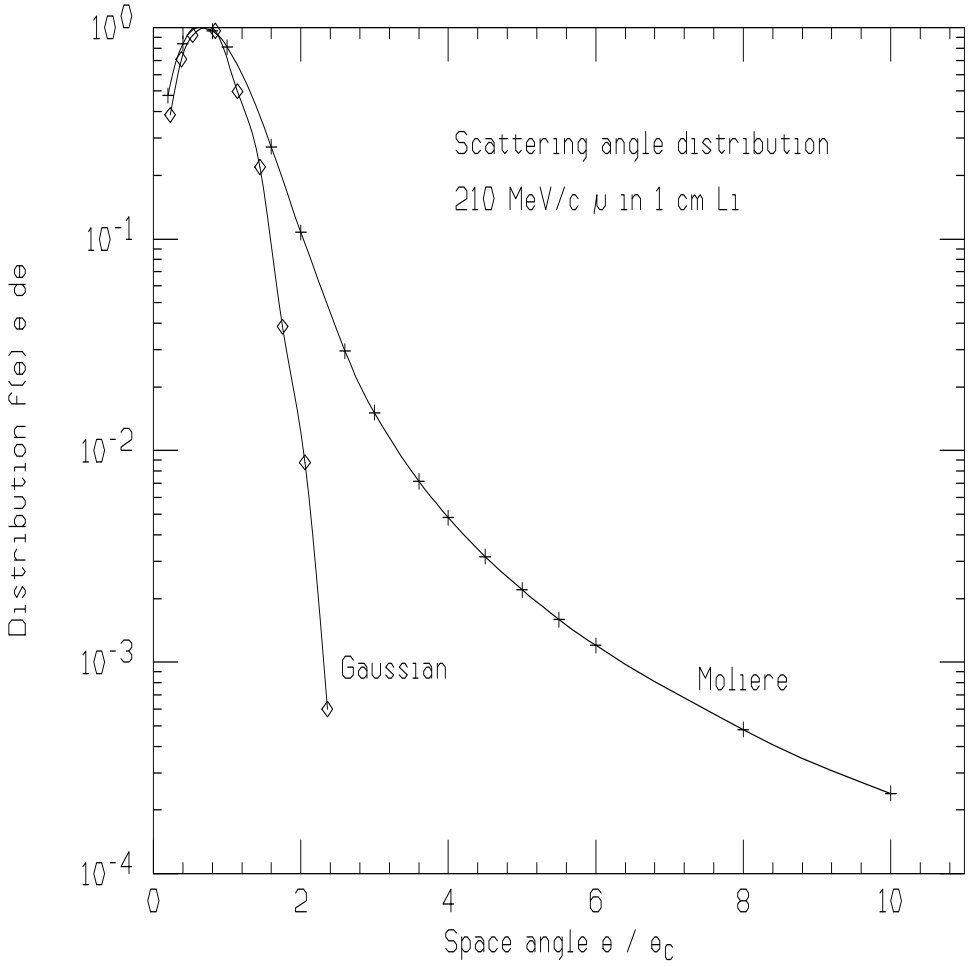


Figure 6.7: Moliere scattering angle distribution in Li

the electron can be neglected for these relatively large losses. A given energy loss implies a unique scattering angle for the muon, thereby introducing energy-angle correlations into the distributions.

For Coulomb scattering between muons and nuclei the SIMUCOOL code introduces a cutoff angle at $2\sigma_\theta$. Below this angle events are simulated statistically using a Gaussian distribution. Above it the Rutherford formula with a nuclear form factor is used to simulate individual events. Provided the cutoff is below the maximum angle as calculated on the basis of nuclear size, this should be a good approximation[22]. Note that here the Gaussian σ is smaller than the unrestricted σ_θ given by Eq. 6.12. Again energy-angle correlations are introduced by applying the energy loss that accompanies the large angle events. Alternatively, codes based on the GEANT scattering algorithm (e.g. MUMC and ICOOL) select a scattering angle from the Moliere scattering distribution whenever the mean number of expected scatters in a simulation step is larger than 20. Otherwise, the actual number N

of scatters is calculated from a Poisson distribution, and the scattering angle is determined from a series of N Rutherford scatterings.

6.2.3 Other Relevant Effects

We have seen in the previous section that the most important effects to simulate, as far as ionization cooling is concerned, are ionization energy loss and coherent multiple Coulomb scattering. Other effects play only a minor role in the ionization cooling process.

Muon decays are an important source of particle losses. However, since the decays are uncorrelated with the energy loss and scattering effects, they do not have a direct influence on the achievable muon emittances.

Van Ginneken has considered a number of other processes that only make a small contribution for 10 - 300 MeV muons. These include bremsstrahlung, incoherent Coulomb scattering off individual protons in the nucleus, direct pair production, and deep inelastic nuclear interactions. Of these incoherent scattering may cause considerable angular deviation, whereas the others tend to produce relatively large energy loss. At low energy (where $\Delta E > T_{max}$) μe scattering becomes identically zero. For the low energy muons used in the cooling section of the muon collider, these processes only occur at very low levels.

6.2.4 Tracking Simulations of Cooling

Confirmation of ionization cooling using simulations of particle tracking has been made using at least five independent codes[23]. We present several examples here, as way of illustration.

Fig. 6.8 shows a SIMUCOOL simulation of transverse emittance cooling of 1 GeV muons in a storage ring. The beam repeatedly crosses a 1 cm thick Be absorber in the ring. Each frame shows the transverse phase space at the traversal number shown in the lower right corner. The number in the upper left of each frame shows the number of muons remaining in the beam. Muons were removed from the simulation when x or x$'$ exceeded 3σ. The fractional reduction in phase space area is shown in the upper right corner. Thus in this example the transverse emittance is reduced to its asymptotic value after about 1500 turns. The emittance was reduced by a factor of 5.

Modelling of cooling in a 1.5 m long Li rod has also been done using SIMUCOOL[24]. The rod had a radius of 10 cm and a 10 T surface field. The incident muon momentum was 400 MeV/c. The initial normalized transverse emittance was 1900 mm mr. The cooling achieved was 18%, compared to 20% predicted by the rms equations. The cooling was found to be insensitive to the precise betatron function matching over the range from 10-12 cm. The

energy spread produced by the absorber had strong correlations with the particles amplitude and time.

A PARMELA simulation[25] of the minimum achievable transverse emittance using a 60 cm long rod of passive Li absorber inside a 7 T solenoid is shown in Fig. 6.9. The rod occupies longitudinal distances from 70 to 130 cm. Cooling of the normalized emittance occurs for large values of the initial emittance. However, as the initial emittance is reduced, eventually the cooling just balances the heating and the emittance remains constant throughout the rod. This represents the case of minimum achievable emittance. If the initial emittance is reduced further, heating dominates and the normalized emittance increases in the Li.

6.3 Scheme for Ionization Cooling

6.3.1 Overview

For the muon collider we require a reduction of the normalized x and y transverse emittances by almost three orders of magnitude (from ≈15,000 to 40 mm mr), and a reduction of the normalized longitudinal emittance by one order of magnitude [26]. This cooling is obtained in a series of cooling sections. Each of these sections contains some combination of the

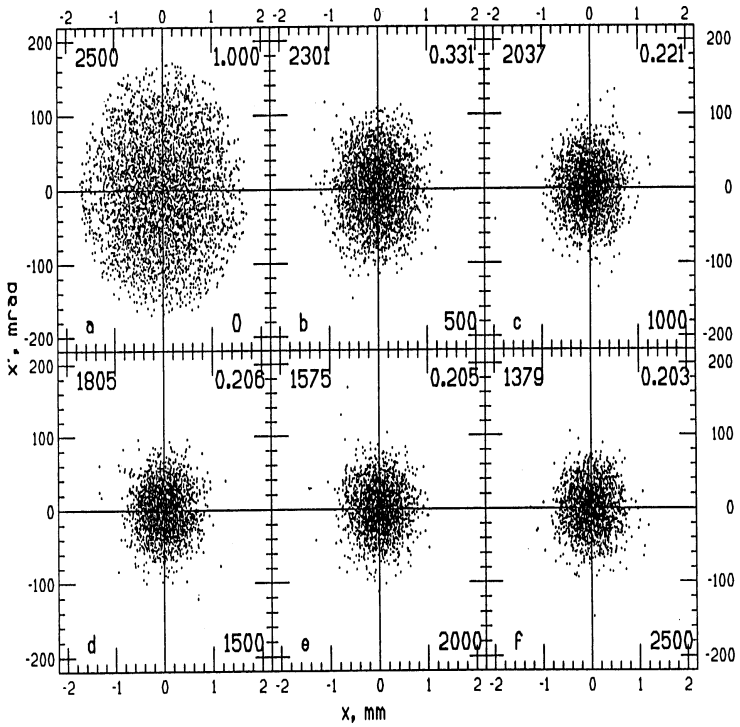


Figure 6.8: SIMUCOOL simulation of transverse emittance cooling in a storage ring

following elements:

- transverse cooling cells
- longitudinal cooling cells
- linear accelerators
- matching cells

The cooling section consists of alternating parts for transverse cooling, longitudinal cooling, and reacceleration. The cooling parameters are adjusted to provide the most efficient cooling for the beam conditions at that location. Matching cells must be provided to accept the beam from the phase rotation part of the accelerator. If induction linacs are used there, then it will probably be necessary to also use induction linacs in the early part of the cooling system. A matching section will be required to deliver the cooled beam to the

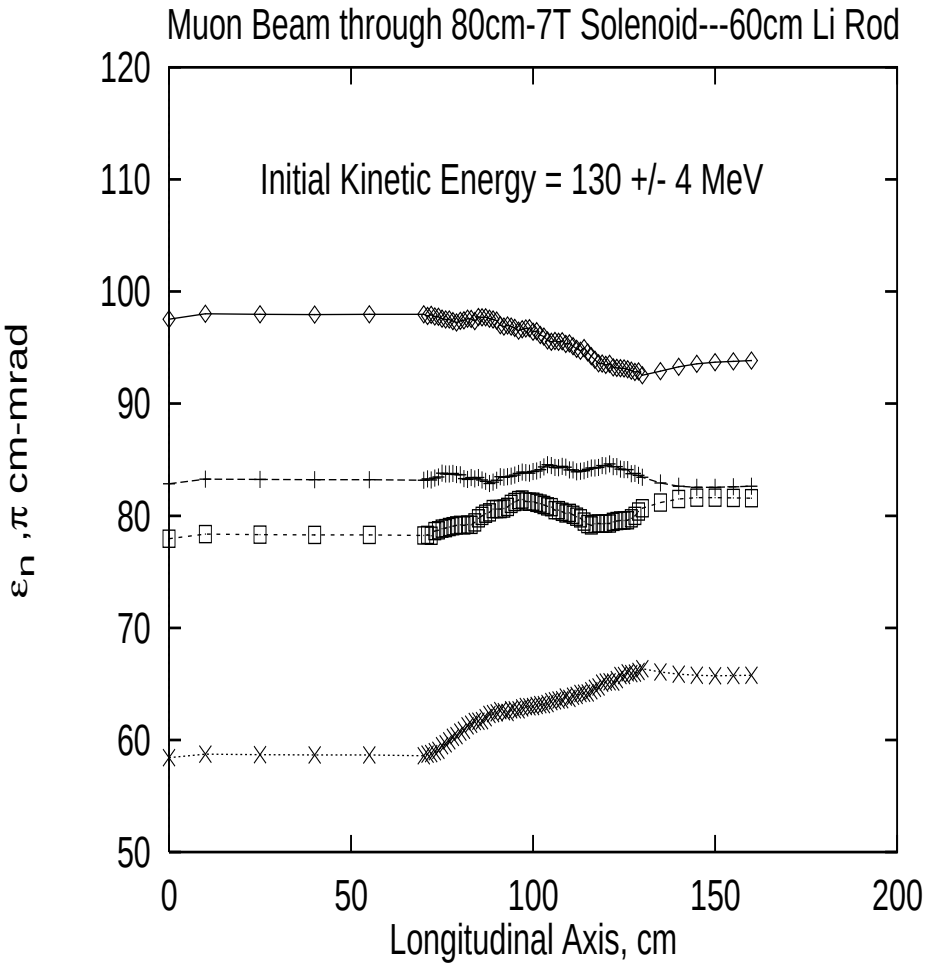


Figure 6.9: PARMELA simulation of transverse emittance cooling with a passive Li rod inside a solenoid

first accelerator booster ring. In addition, matching cells (consisting of a pair of solenoids, for example) must be provided for the transition between the different parts of the cooling section. Throughout this process appropriate momentum compaction and *rf* fields must be used to control the bunch in the presence of space charge, wake field, and resistive wall effects.

Transverse emittance cooling

Transverse cooling of the initial, large emittance muon beam is accomplished using FOFO cooling cells. These consist of a lattice of solenoids with alternating field directions and a short piece of absorber in the space between them. The solenoids provide a small $\beta_\perp$ in the absorbers. The requirement for efficient cooling is that the angle of the beam trajectory with the system axis inside the absorber caused by the focussing magnets must be large compared to the characteristic scattering angle in the material. Thus any practical design must be able to transmit a beam with a large angular acceptance. The preferred absorber material is LiH because it's higher density allows shorter lengths of absorber and better overlap with the regions of small $\beta_\perp$.

Near the end of the cooling sections, where the beam emittance is small, the FOFO solenoids may not provide sufficiently strong focusing. At this point current-carrying lithium rods might be necessary to produce the final transverse cooling. In this case the rod serves simultaneously to maintain the low $\beta_\perp$ and provide the required energy loss. Similar lithium rods, with surface fields of 10 T , were developed at Novosibirsk and have been used as focusing elements at FNAL and CERN. It might be possible to use liquid lithium columns to raise the surface field to 20 T and improve the resultant cooling [27]. Another possibility is to cool the longitudinal emittance more than the required amount and then use reverse wedges in a dispersive region to exchange longitudinal for transverse emittance.

Longitudinal emittance cooling

Longitudinal cooling is provided by using wedge-shaped absorbers in a dispersive region of the beamline. The lattice here consists of more widely spaced alternating solenoids with the dispersion introduced by using chicanes (series of dipoles). Dispersion must be introduced and removed by wedges in both the x and y dimensions.

Cooling section accelerators

Linear accelerators are used initially for bunch compression in order to minimize the amount of low frequency *rf* required. In the cooling channel accelerators are used to restore the beam energy lost in the absorbers and wedges. It would be desirable, if practicable, to

economize on linac sections by forming groups of stages into recirculating loops.

6.3.2 Example Cooling Scenario

Model examples of the cooling lattice have been designed that make use of a linear array of the cooling elements discussed in the previous section. The model discussed here contains all the essential components for achieving the emittance reduction required for the muon collider described in this document. However, it is incomplete since all the matching sections have not been specified, some higher order corrections and aberrations have not been considered and the performance has not been optimized. Computer codes for the analytic design of cooling scenarios and for particle tracking through all the elements of the cooling system are under active development. Tracking studies to check the performance of the analytic model are currently in progress. Given sufficient time and resources, it should be possible to produce a fully optimized cooling design.

The model design uses a FOFO lattice for initial focusing of the muon beam. Spherical aberrations due to solenoid end effects, wake fields, and second order *rf* effects have not yet been included. The phase advance in each cell of the FOFO lattice is made as close to $\pi/2$ as possible in order to minimize the $\beta_\perp$ at the location of the absorber, consistent with the phase spread caused by:

- the maximum space charge transverse defocusing
- a 3σ fluctuation of momentum
- a 3σ fluctuation in amplitude

Bending magnets are introduced to generate dispersion, but the dispersion is kept equal to zero at the center of all solenoids. The maximum allowed beam angle with respect to the axis, due to dispersion, is 67^o. In the early FOFO stages, the solenoids are relatively large and their fields are limited to 13 T. In later stages the transverse emittance is reduced using Li lenses with a surface field of 13 T. The maximum bending fields used in the emittance exchange regions are 9.2 T. Except for the dispersive sections and the Li lenses, it is assumed that both charges will use the same channel.

The transverse and longitudinal emittances are shown in Fig. 6.10 as a function of section number, together with the beam energy. In the early sections, relatively strong wedges are used inside a FOFO lattice to rapidly reduce the longitudinal emittance, while the transverse emittance is reduced relatively slowly. The object here is to reduce the bunch length, thus allowing the use of higher frequency and higher gradient *rf* in the reacceleration linacs. In the later sections, the emittances are reduced close to their asymptotic limits. The charges

are separated for the Li lens sections at the end. During the final cooling the energy is allowed to fall to 15 MeV in order to minimize the value of $\beta_\perp$ Some important properties of the cooling system based on the analytic calculations of the expected beam properties are summarized in Table 6.2.

6.4 Magnets for the Muon Cooling System

The muon cooling system consists of hundreds of superconducting solenoids that have central inductions ranging from 2 to 13 tesla. Solenoids can be found in both the cooling sections and the emittance exchange sections of the muon cooling system. Solenoids of two types are presented here; the cooling section solenoids and the emittance exchange solenoids. Both types of solenoids are relatively short compared to their coil inside diameter. Both types

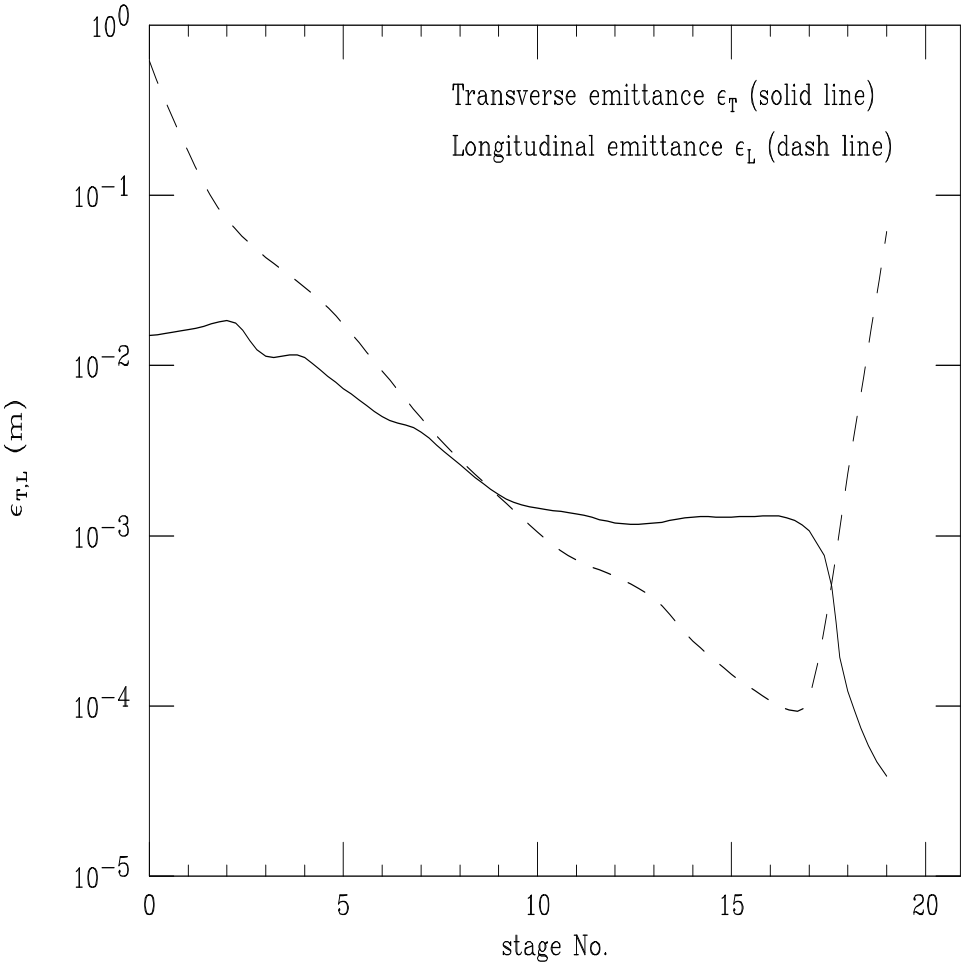


Figure 6.10: Normalized transverse and longitudinal emittances as a function of section number in the model cooling system

Table 6.2: Cooling section summary

total length		743	m
sections		19	
total acceleration		4.8	GeV
accelerator length		690	m
μ decay loss		45	%
contingency loss		20	%
	Entrance	Exit	
KE	300	15	MeV
p	392	58	MeV/c
β	0.966	0.481	
ϵ_{xN}(rms)	15000	39	mm mr
ϵ_{zN}(rms)	61.2	6.0	m %
σ_z	1.50	0.35	m
$\frac{\delta p}{p}$	11.0	31.7	%
μ intensity	7.5	3.0	10^{12} / bunch

of solenoids in the muon cooling system will have a warm bore. The bore tube should be cooled to remove the energy deposited from muon decay products. Muon decay products can deposit around a watt per meter into the warm bore tube that travels down through all of the magnets in the muon cooling channel.

The muon cooling system also has dipoles that are part of the emittance exchange sections of the cooling channel. We believe that the dipoles can be modified picture frame superconducting dipoles. This type of design was chosen because the length of the dipole is relatively short in comparison to the horizontal and vertical aperture. (In some cases the dipole length is shorter than the dipole gap.) The central induction of the dipoles is high enough to preclude the use of conventional water cooled dipoles.

6.4.1 Muon Cooling Solenoids

The cooling solenoids are close together, and they are powered with alternating polarity. This arrangement produces a zero field (on axis) at a point between the solenoids. In the low field region between the solenoids are pieces of lithium or beryllium to cool the muons. The arrangement of a representative cooling solenoid is shown in Fig. 6.11. A parameter set for the cooling section solenoids is presented in Table 6.3 for a representative set of solenoids.

The cooling section solenoids get smaller as the muon beam emittance decreases. The central induction at the center of the solenoids goes up as one goes down the muon cooling channel.

Table 6.3: Muon cooling solenoid magnet parameters for three sections in the muon cooling channel

Parameter			
Solenoid Induction* (T)	5.0	7.0	11.0
Solenoid Length (mm)	531	386	131
Warm Bore Radial Aperture(mm)	310	65	39
Solenoid Coil Inside Radius (mm)	325	80	54
Solenoid Coil Outside Radius (mm)	475	160	108
Solenoid Current Density** (A mm^{-2})	53.2	83.2	274.1
Peak Induction in Coil** (T)	6.5	7.2	12.4
Solenoid Stored Energy** (kJ)	4752	290	163
Coil Superconductor	Nb-Ti	Nb-Ti	Nb_3Sn
Superconductor Temperature (K)	4.2	4.2	4.2

* at the center of the solenoid magnet

** at the design induction given above

The 5 and the 7 tesla solenoids can be fabricated from niobium titanium conductor that operates at 4.2 K. At some point down the cooling channel, the solenoid induction becomes too large to be generated by niobium titanium at 4.2 K. When the solenoid central induction reaches 11 tesla, the solenoid must be fabricated from an A-15 superconductor (multifilamentary niobium tin is the most commonly used A-15 conductor). Some of the solenoids that have a central induction between 7 tesla and 9.5 tesla can be fabricated from niobium titanium, but they must be cooled to 1.8 K. The high field solenoids can be hybrid solenoids with small niobium tin solenoids inside larger niobium titanium outsert solenoids.

The 11 tesla solenoid described in Table 6.3 was assumed to be made from multifilamentary niobium tin in a copper based matrix. A conductor of this type would have up to 30 percent copper within the matrix. The coil current density appears to be reasonable, given the stored energy of the solenoid.

Since the solenoids operate with opposite polarity, there are large forces pushing the solenoids apart. It has been assumed that the solenoids can be separated by cold supports. Room temperature ports between the solenoid coils can be used to provide water cooling to the lithium or beryllium rods used to cool the muons. The solenoid sections can be fabricated so that several solenoid pairs are in a group. These solenoids would share common cryogenic

cooling and perhaps a common power supply and leads

6.4.2 Muon Emittance Exchange Solenoids

The emittance exchange solenoids are further apart and they have the same polarity (in terms of solenoid performance, the polarity of the adjacent solenoid does not seem to make much difference). The parameter set for some representative emittance exchange solenoids is presented in Table 6.4. Like the cooling section solenoids, the emittance exchange solenoids get smaller as the muon beam emittance decreases. The central induction of the solenoids goes up as one goes down the cooling channel.

The 5 and the 7 tesla solenoids can be fabricated from niobium titanium conductor that operates at 4.2 K. When the solenoid central induction reaches 11 tesla, the solenoid must be fabricated from an A-15 superconductor, such as niobium tin. The 11 tesla solenoid is quite

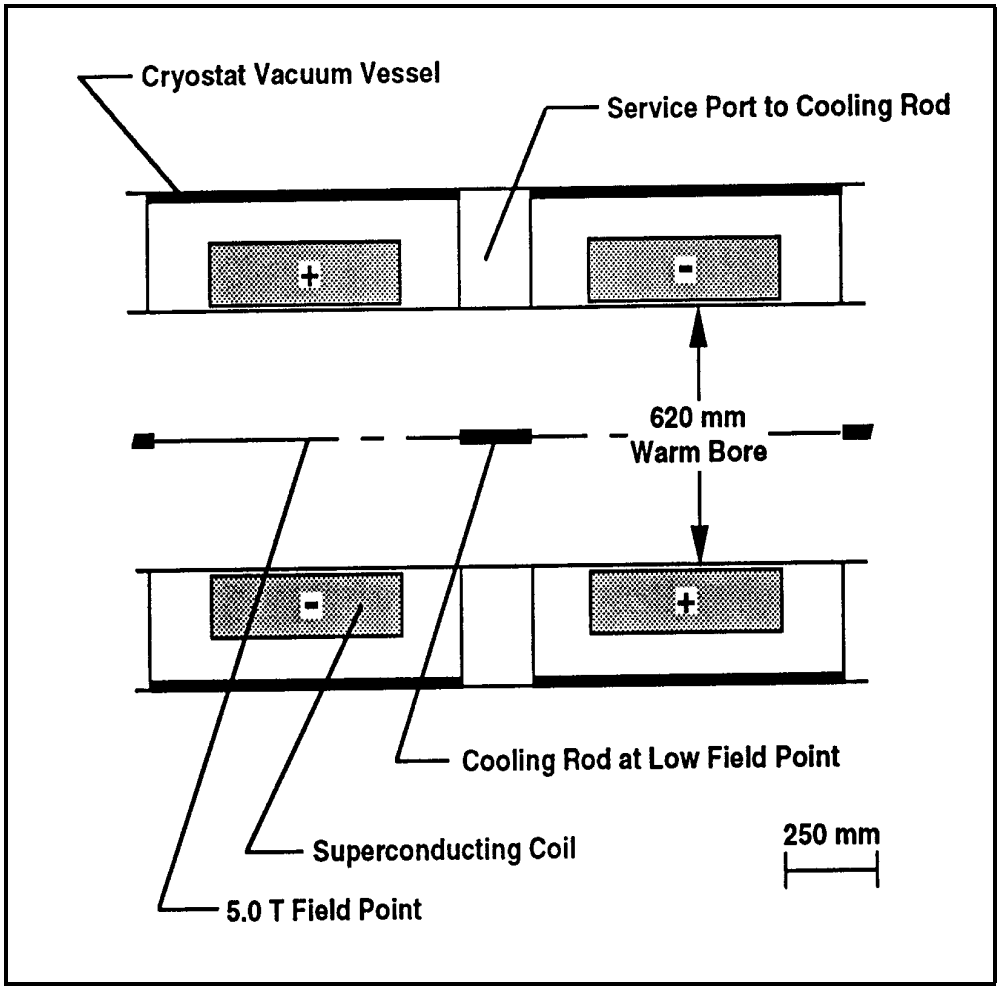


Figure 6.11: Cross-section of the superconducting solenoids of the muon cooling channel

Table 6.4: Muon emittance exchange solenoid magnet parameters for three sections in the muon cooling channel

Parameter			
Solenoid Induction* (T)	5.0	7.0	11.0
Solenoid Length (mm)	586	281	62
Warm Bore Radial Aperture(mm)	309	75	19
Solenoid Coil Inside Radius (mm)	325	90	34
Solenoid Coil Outside Radius (mm)	400	170	88
Solenoid Current Density** (A mm^{-2})	43.3	93.4	374.8
Peak Induction in Coil** (T)	6.0	7.5	13.9
Solenoid Stored Energy/Cell** (kJ)	4130	290	120
Coil Superconductor	Nb-Ti	Nb-Ti	Nb_3Sn
Superconductor Temperature (K)	4.2	4.2	4.2

* at the center of the solenoid magnet

** at the design induction given above

short compared to its cold bore diameter. As a result, more current is needed in the coil to generate 11 tesla at the magnet center. The peak magnetic induction in the coil is quite high for this magnet. Operation of this magnet at 4.2 K may be marginal, but operation appears to be possible at 1.8 K. Other A-15 materials, such as niobium titanium tin, may improve the margin at 4.2 K. The high field solenoids can be hybrid solenoids with small niobium tin solenoids inside niobium titanium outsert solenoids.

The major problem with the muon emittance exchange system is the dipoles that are mixed with the solenoids. In order for the dipole field to be sharply defined, the dipoles must return their flux with an iron return path. The dipole iron return path may interfere with the solenoids that are near the dipoles. Individually, the solenoids can be built. The combination of dipoles and solenoids may be a problem. Considerable design work and a three dimensional magnetic field analysis are required before this problem can be completely solved. In some sections, quadrupole focusing of the muon beam may be desirable. Assuming that the dipoles and solenoids can be mixed, the forces between the two types of magnets can be taken up by a cold support between them.

6.4.3 Muon Emittance Exchange Dipoles

The emittance exchange dipoles bend the muon beam and spread it according to momentum. The dipoles that are in this part of the muon cooling system must have an aperture large enough to transport the muon beam and the sagitta of the bent beam. Early in the muon cooling process, these dipoles are large. They bend the muon beam over angles approaching 60 degrees. As the muon emittance is reduced, the vertical aperture of the dipoles is reduced. The horizontal aperture still has the bent beam sagitta, but it is also reduced to a great extent. The bend angle does go down somewhat as one proceeds down the cooling channel.

Table 6.5 presents the design parameters for a representative set of emittance exchange dipoles. At the start of muon cooling the vertical beam diameter is enormous. As a result, the dipole length to vertical aperture ratio is less than one. This means that the effective length of the dipole is quite a bit longer than its physical length. As the muon emittance is reduced, the dipole gap gets smaller. Unfortunately, the dipole length also gets shorter, so the length to gap ratio is still around one. Because, the length to aperture ratio is around one, conventional superconducting dipole designs cannot be employed. A modified picture frame dipole design is proposed because this type of dipole will behave like a short conventional room temperature copper and iron dipole. The proposed magnet would be superconducting and it would be similar to magnets that have been proposed for compact light sources and small electron storage rings. Figures 6.12 shows a cross-sectional view of a possible dipole.

Table 6.5: Muon emittance exchange dipole magnet parameters for three sections in the muon cooling channel

Parameter			
Dipole Average Induction (T)	1.75	2.13	3.25
Dipole Bend Angle (deg)	59	50	40
Dipole Iron Length (mm)	600	290	66
Dipole Horizontal Half Aperture (mm)	400	119	36.5
Dipole Vertical Half Aperture (mm)	319	85	29
Dipole Cold Iron Height (mm)	1440	460	190
Dipole Cold Iron Width (mm)	1610	560	250
Dipole Stored Energy (kJ)	~ 370	~ 23	~ 1.9
Dipole Coil Superconductor	Nb-Ti	Nb-Ti	Nb-Ti

The dipoles shown in Table 6.5 appear to be feasible in and of themselves. The magnetic field will extend beyond the ends of the magnet by about a quarter of a gap (in the way conventional copper and iron dipoles behave). The iron in the dipole would be cold. The

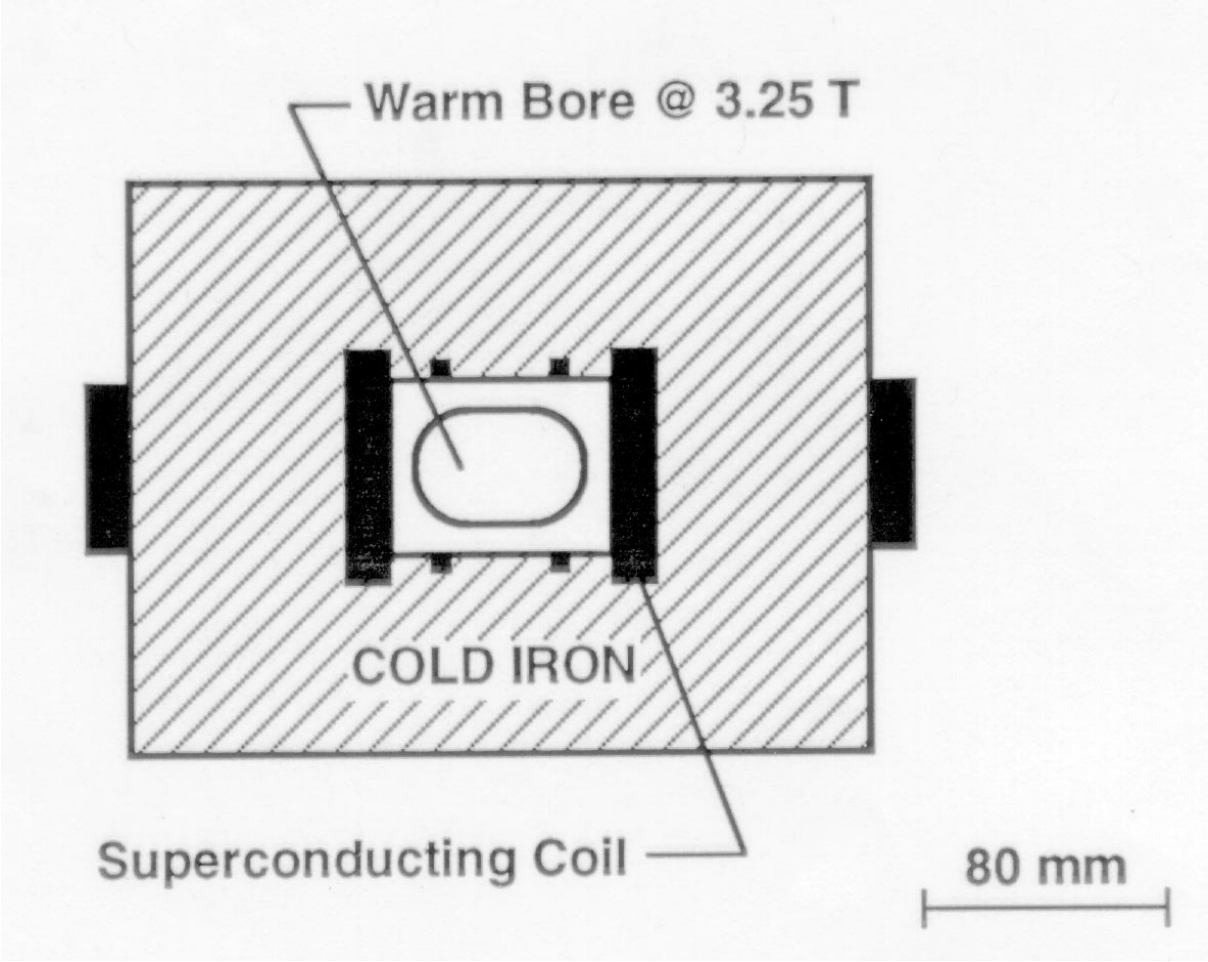


Figure 6.12: Cross-section of muon emittance exchange dipole of the muon cooling system

dipole might share a common cryostat with other elements in the emittance exchange magnet string. The dipoles and the adjacent solenoids would all have a warm bore. The elements for reducing the energy of high momentum muons would also be at room temperature. The problem with the dipoles is the fact that they share space with some high field superconducting solenoids that have an aperture comparable to the dipoles. The space between the dipoles and adjacent solenoids is an important issue. This space may have to be increased or a magnetic shield may have to be used.

6.5 A Possible Experimental Demonstration

A BNL-CEBAF-Fairfield-Michigan-Stony Brook-Yale collaboration has submitted a letter of intent[28] to the AGS to build a prototype cooling cell with properties as close as possible

to those required in a muon collider. The goal would be to measure the performance of the device and compare it with predictions from the Monte Carlo programs that are being used to design the collider. Since the performance of the cooling section plays an essential role in achieving the required luminosity and the cooling section may represent a significant fraction of the total cost of a muon collider, it is essential that the design Monte Carlo accurately model the achievable cooling. The experiment would also check that muon losses in the focusing fields and absorbing materials do not exceed the predicted amounts.

The initial phase of the proposed experiment would use a cylindrical piece of lithium, 12 cm in radius and 60 cm in length as the cooling rod. A solenoid with a maximum central field strength of 7 T would focus the muons in the interior of the rod. The absorber material would be replaced at some point with different lengths of absorber and with Be or LiH absorbers in order to further check the accuracy of the Monte Carlo predictions.

In order to simulate the conditions in the muon collider as closely as possible, the momentum of the incident muon should be around 200 MeV/c. For a 0.6 m long lithium rod the final momentum would be around 140 MeV. The focusing in the rod should be as strong as possible, since Eq. 6.14 shows that the minimum emittance that can be reached is proportional to $\beta_\perp$.

The layout of the proposed experiment is shown in Fig. 6.13. A large, diffuse beam of muons passes through a hadron absorber and enters an approximately 6 m long experimental area. The dipole D1 together with the position measuring detectors P1-5 are used to measure the momentum of each incoming track. A fraction of the beam will be captured by the solenoid magnet S1 and focused into an absorber. The momentum of the muons leaving the solenoid is measured using the spectrometer consisting of dipole magnet D2 and detectors P6-10. Trigger counters would tag tracks entering the absorber. A particle identification system would be necessary to identify pion and electron background in the muon beam. The dipole D1 would have to be adjusted to examine a series of different central values for the incident momentum. The momentum spread could be determined by software selection of the tracks used in the analysis.

In the following we consider an example case with an initial normalized emittance of 4000 mm-mr. The beam has a 1σ radius of about 3 cm inside the lithium. The dipoles have central field values around 0.3 T and have widths and lengths appropriate to existing 72D18 magnets at the AGS. We assume a bend angle in the dipoles of 0.2 rad.

PARMELA results for the change in transverse normalized emittance as a function of axial distance z are shown in Fig. 6.14 and summarized in Table 6.6 Under the conditions listed in Table 6.6, the cooling rate is always larger than the heating rate and the **normalized** emittance of the beam leaving the rod is smaller than the normalized emittance entering it.

However, the **geometric** emittance increases after traversing the rod. In an actual muon collider this stage must be followed by a reacceleration section that restores the starting momentum in order to get cooling of the geometric emittance.

One of the design requirements for the experiment is the ability to measure individual muon tracks. This will permit software "control" of the effective cooling rate. Software selection of initial beam tracks will allow us to reconstruct final beam emittance as a function of initial beam emittance. The experiment can be reanalyzed for smaller initial transverse emittances. As the initial emittances decrease, the cooling in the lithium becomes less effective relative to the heating and the net decrease in normalized emittance also drops. As we saw in Fig. 6.9, this configuration should still produce net cooling until the initial emittance reaches about 820 mm-mr. These measurements are important since they demonstrate the minimum emittance achievable for a given field strength and absorber length.

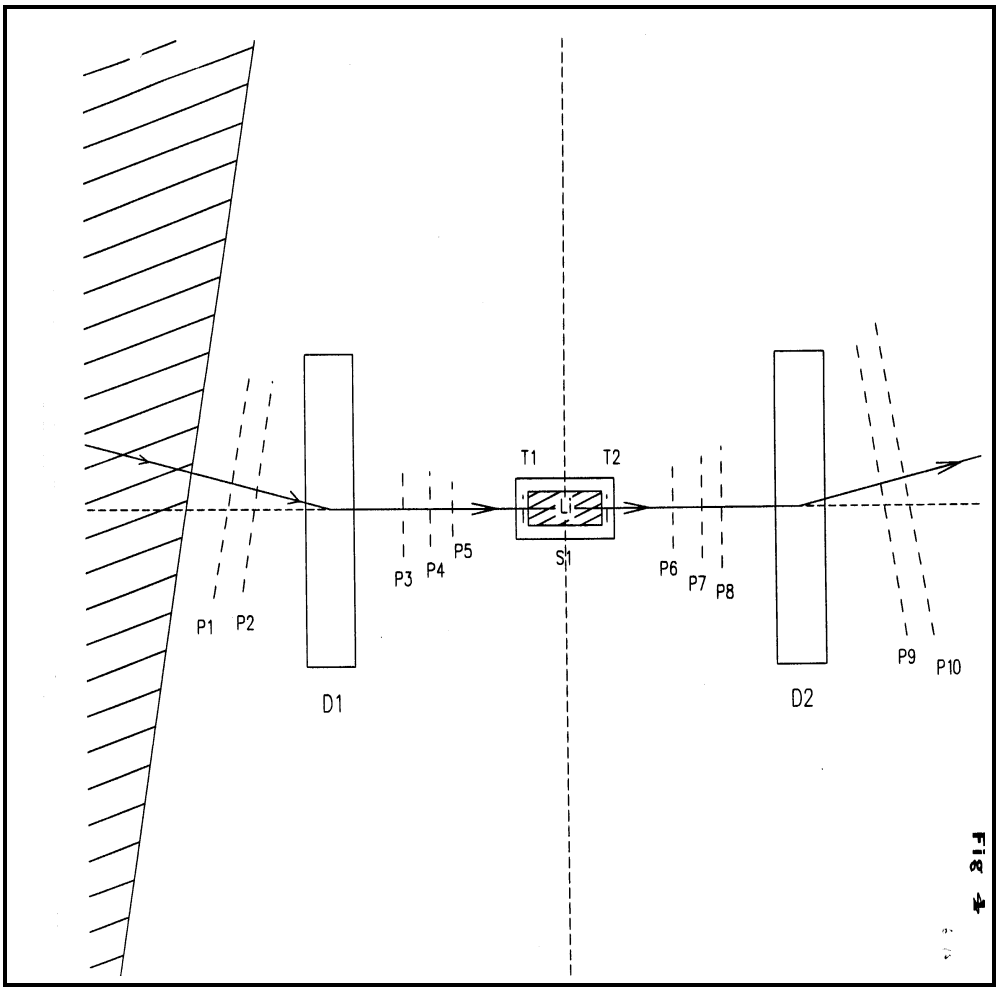


Figure 6.13: Layout of possible AGS experiment to test ionization cooling of transverse emittance

Table 6.6: Transverse emittance cooling experiment

	IN	OUT	
KE	130	70 MeV	
p	211	140	MeV/c
β	0.894	0.798	
γ	2.23	1.66	
ϵ_{xN}	4000	3100	mm mr
ϵ_x	2006	2340	mm mr
$\beta_\perp$	20.1	13.3	cm
σ_x	20	18	mm
$\sigma_{x'}$	100	133	mr

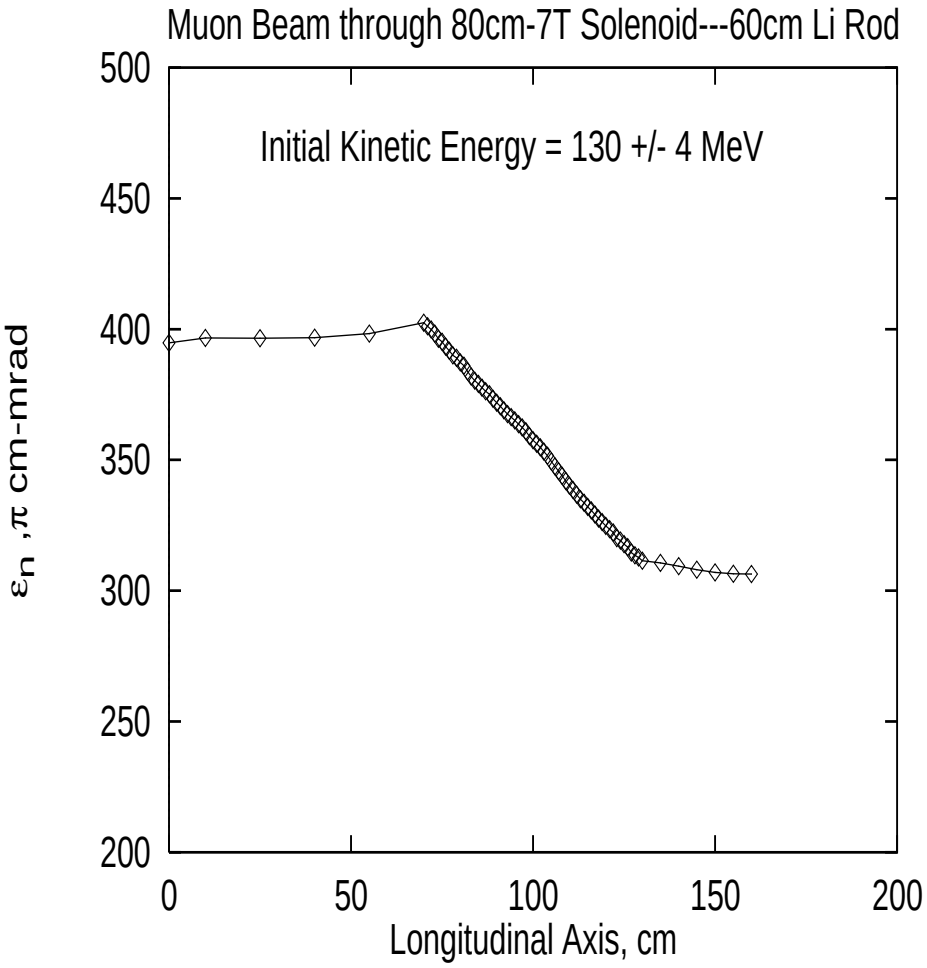


Figure 6.14: PARMELA simulation of transverse emittance in the AGS cooling experiment

The experimental test of cooling would require accurately measuring the position, angle, and momentum of individual muon tracks before and after the rod in order to measure the initial and final emittances. Since the chambers are likely to be in the strong fringe field of the solenoid, a good field map will be required. We need at least three x and y measuring planes on each side of the rod to measure the curvature of the track. The six chambers closest to the rod require measurements with moderate resolution ($\approx 300\,\mu m$). The expected fractional change in normalized emittance for the case described in Table 6.6 is about 23%. In order to measure the smaller effects expected with lower initial emittances, we would like to determine $\Delta\epsilon_{xN}/\epsilon_{xN}$ to an accuracy of $\approx 0.1\%$.

In order to measure individual tracks we propose using the slow extracted beam from the AGS. The useful muon rate should be larger than $\approx 10^2$ per spill so that a complete measurement can be made in about an hour, in order to minimize the effects of systematic errors. If the useful rate is smaller than $\approx 10^4$ per spill, it should be possible to use a fairly simple data acquisition system.

6.6 Summary

The process of ionization cooling offers a method for reducing the 6-dimensional normalized emittance of the muon beam by a factor of $\approx 10^6$. A simple analytic theory has been developed that demonstrates the dependence of the net cooling on various experimental parameters. The simple theory has been checked and realistic arrangements have been examined using Monte Carlo simulations. Transverse cooling of the initial beam can be achieved using passive Li or Be absorbers in a FOFO lattice or a solenoid channel. The last factor of 10 in transverse cooling probably requires the use of current-carrying Li lenses. Efficient longitudinal cooling requires the use of wedge shaped absorbers in a dispersive section of the beam line. An example, multi-stage cooling scenario has been developed that meets the requirements of the muon collider using a mixture of FOFO lattice, Li lenses, wedge absorbers, *rf* accelerators, and matching sections. Preliminary designs have been made of solenoids for use in the FOFO lattice and of solenoids and dipoles for use in the emittance exchange sections. Detailed simulation work, further optimization, and preparations for experimental demonstrations of critical components are currently in progress.

Bibliography

[1] A.N. Skrinskii & V.V. Parkhomchuk, *Methods of cooling beams of charged particles*, Sov. J. Part. Nucl., **12**, 223 (1981); V.V. Parkhomchuk & A.N. Skrinsky, *Ionization cooling: physics and applications*, Proc. 12th Int. Conf. High Energy Phys., 1983, p. 485.

[2] D. Neuffer, *Principles and applications of muon cooling*, Proc. 12th Int. Conf. High Energy Phys., 1983, p. 481.

[3] A.N. Skrinsky, *Ionization cooling and muon collider*, Proc. Workshop on Beam Dynamics and Technology Issues for $\mu^+\mu^-$ Colliders, J. Gallardo, Ed., AIP Conf. Proc. 372, 1996, p. 133.

[4] D. Neuffer, *Principles and applications of muon cooling*, Part. Acc. 14, 75, (1983).

[5] R.C. Fernow & J.C. Gallardo, *Validity of the differential equations for ionization cooling*, Physics Potential and Development of $\mu^+\mu^-$ Colliders, Second workshop, D. Cline, Ed, AIP Conf. Proc. 352, 1995, p. 170.

[6] B. Rossi, *High Energy Particles*, Prentice-Hall, 1952.

[7] U. Fano, *Penetration of protons, alpha particles, and mesons*, Ann. Rev. Nuc. Sci. 13, 1 (1963), eqn. 72.

[8] R.C. Fernow & J.C. Gallardo, *Muon transverse ionization cooling: stochastic approach*, Phys. Rev., **E 52**, 1039 (1995).

[9] T. Vsevolozhskaya, *Kinetics of ionization cooling of muons*, Proc. Workshop on Beam dynamics and technology issues for $\mu^+\mu^-$ colliders, in J. Gallardo, Ed., AIP Conf. Proc. 372, 1996, p. 159; T. Vsevolozhskaya, *Ionization cooling of muons by strong focusing with a field of longitudinal current*, Nucl. Inst. Meth.,**A 374**, 281, (1996).

[10] J. Gallardo, *Useful formulae for uniform solenoid channel*, Center for Accelerator Physics, BNL, $\mu\mu$ Tech. note 7-95.

[11] H. Wiedemann, *Particle Accelerator Physics: basic principles and linear beam dynamics*, Springer-Verlag, 1993, p. 187.

[12] D. Carey, *The optics of charged particle beams*, Harwood, 1987, p. 210.

[13] B. Bayanov et al., *A lithium lens for axially symmetric focusing of high energy particle beams*, Nucl. Instr. Meth. 190, 9 (1981).

[14] G. Dugan et al, *Mechanical and electrical design of the Fermilab lithium lens and transformer system*, IEEE Trans. N 30, 3660, (1983).

[15] V. Balbekov, *Achievable transverse emittance of beam in muon collider*, Proc. Workshop on Beam dynamics and technology issues for $\mu^+\mu^-$ colliders, J. Gallardo, Ed., AIP Conf. Proc. 372, 1996, p. 140.

[16] A. Van Ginneken, *Fluctuations of muon energy loss and simulation of ionization cooling*, Nucl. Instr. Meth., **A 362**, 213 (1995).

[17] *Particle Data Group, Review of Particle Properties*, Phys. Rev., **D50**, 1173 (1994).

[18] L. Landau, *On the energy loss of fast particles by ionization*, L. D. Landau, Collected Papers, D. ter Haar, Ed., Pergamon Press, 1965.

[19] *GEANT manual*, v. 3.2.1, CERN program library W5013, 1994.

[20] P. Vavilov, *Ionization losses of high energy heavy particles*, Sov. Phys. JETP, 5, 749 (1957).

[21] H. Bethe, *Moliere's theory of multiple scattering*, Phys. Rev. 89, 1256 (1953).

[22] J.D. Jackson, *Classical Electrodynamics*, Wiley, 1962.

[23] Simulation work with tracking codes that has demonstrated ionization cooling include A. Van Ginneken (SIMUCOOL), R. Thun (private communication), R. Palmer & J. Gallardo (MUMC), H. Kirk (PARMELA), and R. Fernow (ICOOL).

[24] A. Van Ginneken & D. Neuffer, private communications, February 1996.

[25] Los Alamos Accelerator Code Group, PARMELA; the code was modified by H. Kirk to include ionization loss and multiple scattering.

[26] R. Palmer et al, *Muon Collider Design*, Proc. of the Symposium on Physics Potential and Development of Muon-Muon Colliders, San Francisco, CA, 1995.

[27] G.I. Silvestrov, *Lithium lenses for muon colliders*, Proc. Workshop on Beam dynamics and technology issues for $\mu^+\mu^-$ colliders, J. Gallardo, Ed. AIP Conf. Proc. 372, 1996, p. 168.

[28] R. Fernow et al., *Possible demonstration of ionization cooling using absorbers in a solenoidal field*, Proc. Workshop on Beam dynamics and technology issues for $\mu^+\mu^-$ colliders, J. Gallardo, Ed., AIP Conf. Proc. 372, 1996, p. 159. Some earlier ideas on a demonstration experiment are described in R.C. Fernow et al., *A possible ionization cooling experiment at the AGS*, Physics Potential and Development of $\mu^+\mu^-$ Colliders, Second workshop, D. Cline, Ed., AIP Conf. Proc. 352, 1995, p.155.

Contributors

- R. C. Fernow, (BNL) Editor
- J. C. Gallardo, (BNL)
- M. Green, (LBNL)
- H. G. Kirk, (BNL)
- D. V. Neuffer, (FermiLab)
- J. Norem, (ANL)
- R. B. Palmer, (BNL/SLAC)
- A. Skrinsky, (BINP)
- A. Van Ginneken, (FermiLab)
- D. Winn, (Fairfield Univ.)

List of Figures

List of Tables

Chapter 7

THE ACCELERATION SYSTEMS

Contents

7.1 Introduction

In this chapter we discuss the systems for acceleration of muons from the low-energy output of the cooling system ($\approx 0.1\,\mathrm{GeV}$) to full energy at $2\,\mathrm{TeV}$, where the muons are transferred to storage in the collider ring. We first describe the requirements of the acceleration system. The most critical of these is that the muons must be accelerated before they decay, which sets limitations on the acceleration rate, and constrains the potential form of acceleration systems, although the lifetime is sufficiently long to enable multipass systems. Other requirements include phase-space matching from the output of the cooling system to the collider, and beam acceptances. We then describe potential acceleration systems, which include full-energy linac, rapid-cycling synchrotron and recirculating-linac options. Because it appears most readily obtainable within existing technology, our current baseline acceleration choice is a sequence or cascade of recirculating linacs, each of which increases beam energy by about an order of magnitude and accommodates bunch length reductions by almost as much. We present a candidate scenario. Results of particle tracking in this candidate choice provide a *proof of principle* of the general approach. We then describe in some detail the required rf systems and beam transports. Optimization considerations and possible variations are discussed. Difficulties associated with injection from the cooling system are also discussed; a single-pass $\approx 1\,\mathrm{GeV}$ linac tailored for matching from cooling into the first recirculating linac will be required. Other acceleration scenarios, which include rapid-cycling magnetic fields, are also described and their potential advantages are discussed, as well as the technology developments needed for their implementation. It is probable that an optimal system will incorporate rapid-cycling in some portion of its acceleration system.

7.2 Constraints and Requirements

The central difficulty in a $\mu^+\mu^-$ collider is that muons decay, and the muons must be collected, cooled, accelerated, and collided within their limited lifetimes. The lifetime is sufficient to permit multipass acceleration and collisions, but that lifetime must be carefully budgeted to avoid excessive decay. An accompanying problem is that muons are created within a fairly diffuse phase space, and even after ionization cooling the beam phase-space volume is relatively large. The acceleration system must accommodate that complete phase-space volume and compress it within the acceleration cycle to match the requirements of the collider. In this section we describe the constraints resulting from these requirements.

7.3 Lifetime Constraints

The muon lifetime is $\tau_\mu = 2.2\,\mu s$ in the muon rest frame. In the lab frame the lifetime is increased by the relativistic factor $\gamma = E_\mu/m_\mu$, where E_μ is the energy and m_μ is the mass ($m_\mu = 0.10566\,\mathrm{MeV}$) of the muon. The muon decay rate along the beam path length s is

$$\frac{dN}{ds} = -\frac{1}{L_\mu \gamma}, \tag{7.1}$$

where $L_\mu = c\tau_\mu \approx 660\,\mathrm{m}$, and where we have used the relativistic approximation $v/c \approx 1$. In a non-accelerating transport, this implies the usual exponential beam loss:

$$N = N_o e^{-\frac{s}{L_\mu \gamma}}. \tag{7.2}$$

In an accelerating section, γ is not constant:

$$\gamma = \gamma_o + \gamma' s = \gamma_o + \frac{eV'_{rf}}{m_\mu c^2} s, \tag{7.3}$$

where eV'_{rf} is the accelerating gradient. Using this in the decay equation we obtain the solution

$$N(s) = N_o \left(\frac{\gamma_o}{\gamma_o + \gamma' s}\right)^{\frac{1}{L_\mu \gamma'}} \qquad \text{or} \qquad \frac{N(s)}{N_o} = \left(\frac{E_o}{E_{\mathrm{final}}}\right)^{\frac{1}{L_\mu \gamma'}} \tag{7.4}$$

where E_o and E_{final} are the initial and final energies within the accelerating section. Low losses require that the exponential factor must be small, which means that $L_\mu \gamma' >> 1$. This can be written as $L_\mu \frac{eV'_{\mathrm{rf}}}{m_\mu c^2} >> 1$, which means $\mathbf{eV'_{rf} \gg 0.16\,MV/m}$ is required. This general rule must be followed throughout the entire muon system. For example, beam-cooling and reacceleration must also occur in systems whose averaged accelerating gradients (including loss and transport elements) are much greater than $0.16\,\mathrm{MeV/m}$ to avoid large decay losses. For a multiturn muon accelerator, the gradient criterion can be rewritten as

$$E' \to \frac{E_{\mathrm{final}}}{N_{\mathrm{turns}} 2\pi R} \gg 0.16\,\mathrm{MeV/m}, \tag{7.5}$$

where R can be written in terms of the mean bending field B of a complete turn at full energy and the magnetic rigidity $B\rho$ as $R = B\rho/B \approx 0.00334\,E_{\mathrm{final}}(\mathrm{MeV})/B(T)$, and N_{turns} is the total number of acceleration turns. Inserting this into the previous equation obtains a criterion for any multi-turn accelerator:

$$\frac{\mathbf{N_{turns}}}{\mathbf{B(T)}} \ll \mathbf{300}. \tag{7.6}$$

Thus, multiturn $\mu-$acceleration with up to hundreds of turns is possible. The gradient criterion requires somewhat faster acceleration than has typically been obtained in previous multi-pass systems, and that constrains our choices in acceleration systems (see below).

7.4 Phase Space and Intensity Constraints

The acceleration system must take beam from the cooling system to full energy. In the cooling section, the initial muons are collected, cooled, and pre-accelerated into moderately compact μ^+ and μ^- bunches at $E_\mu \approx 1\,\mathrm{GeV}$. Studies of the cooling system indicate that an rms energy spread of $\approx 1.5\,\%$ with a bunch length of $\approx 25\,\mathrm{cm}$ at 1 GeV are reasonable design goals, and we use these as reference initial parameters. The accelerator must accelerate these bunches to 2 TeV and transfer them into the collider, which requires a final energy spread of $\approx 0.1\,\%$ and a bunch length reduced to $\approx 0.3\,\mathrm{cm}$. These collision requirements set the longitudinal phase-space area of the beam at collisions at $\approx 3\,\mathrm{mm} \times 2\,\mathrm{GeV}$ ($\frac{\Delta E}{E} = 0.001$ for 2 TeV), or 0.02 eV-s, which is not much larger than the beam emittance at the beginning of the acceleration. The cooling system also reduces the normalized transverse emittance to a design value of $\epsilon_N \approx 0.25 \times 10^{-4}$ m-rad. The acceleration system must accelerate this beam to full energy while maintaining an emittance of $\epsilon_N < 0.5 \times 10^{-4}$ m-rad. Thus relatively little emittance dilution can be allowed in the acceleration and beam transfers, both transversely and longitudinally. Also transport dynamic and physical apertures must be sufficient to accept the beam throughout the accelerator, with low beam loss. The design intensity is 2×10^{12} μ's per bunch, which is a relatively high charge per bunch (larger than existing accelerators). The acceleration system must also accommodate these intense bunches. Wakefield and beam loading effects can become important, particularly in the higher-energy end of the accelerator, where bunch-lengths are reduced toward 0.3 cm, obtaining high-peak currents. Another intensity-dependent limitation is μ-decay in the transport, which will deposit electrons with an average of 1/3 of the μ energy throughout the system. Since the decay rate decreases as the energy increases, the mean beam energy deposition per meter and per μ is a constant

$$\frac{dE}{ds} = \frac{m_\mu c^2}{3L_\mu} \quad (\text{per } \mu). \tag{7.7}$$

This comes to $\approx 0.5\,\mathrm{watts/m/turn}$ from a beam of 2×10^{12} μ's at 30 Hz. This level could become a significant problem with multiturn passes through superconducting elements.

7.5 Acceleration Options

From these constraints, we can develop acceleration scenarios, which include full-energy linac, rapid-cycling synchrotron and recirculating-linac options, as well as hybrid approaches.

7.5.1 Single Pass Linac

A single-pass linac can easily meet the gradient constraint. For instance, the SLAC linac has a gradient of 17 MV/m, roughly 100× the gradient criterion. However single-pass rf structures are prohibitively expensive and do not exploit a primary advantage in muons: our ability to bend them into multipass devices, enabling multipass use of the accelerating structures. As discussed above, the μ lifetime is sufficiently long to permit multipass acceleration.

7.5.2 Recirculating Linacs

A multiturn approach is the use of recirculating linacs. A schematic view of a recirculating linac is shown in Fig.7.1. In a recirculating linac (RLA), the beam is accelerated and returned for several passes of acceleration in the same linac, but a separate return path is provided for each pass. At the end of the linac, the beam passes through beam-separation dipoles, which sort the beam by energy, directing it to an energy-matched return arc (A pulsed kicker magnet system may also be used for the beam separations). The various energy transports are then recombined at the end of the arc for further acceleration in the following linac. The beam passes through both arcs and linacs until full energy is reached, and the beam is then transferred from the highest energy line to the next RLA or the collider.

The RLA permits economic reuse of an expensive accelerating structure for several turns of acceleration. Since the beam passes through a separate transport on each turn, the magnets can be at fixed-field, allowing superconducting magnets, and simplified designs. However since each turn does require a separate return transport, cost and complexity considerations limit the number of turns to a finite number ($\approx 10 - 20$). This is very compatible with the lifetime constraint: ($N_{\rm turns} << 300B(T)$), which then can be met with relatively modest field magnets, and typically beam-decay survivals of $\approx 95\,\%$ are obtained in μ RLA's. High-field magnets are not required. RLA's are rather ideally matched to the μ lifetime constraint.

The same RLA system can be used to accelerate both μ^+ and μ^- bunches. The oppositely charged bunches would propagate around the RLA's in opposite directions. If the bunches are injected into opposite sides of each RLA at the beginning of the separate linacs, then energy match of the beams in each arc is obtained, as well as phase matching across the arcs. Separate (but symmetric) transport lines into the higher-energy RLA's and into the collider would be needed.

Because of the independence of each return transport, there is a broad flexibility in RLA design, with only the rf acceleration frequency and voltage remaining constant from linac pass to linac pass. Since return paths are independent, the synchronous phase ϕ_s and the chronicity, $M_{56} = \frac{\partial z}{\partial(\delta p/p)}$, where z is the particle position within the bunch, can

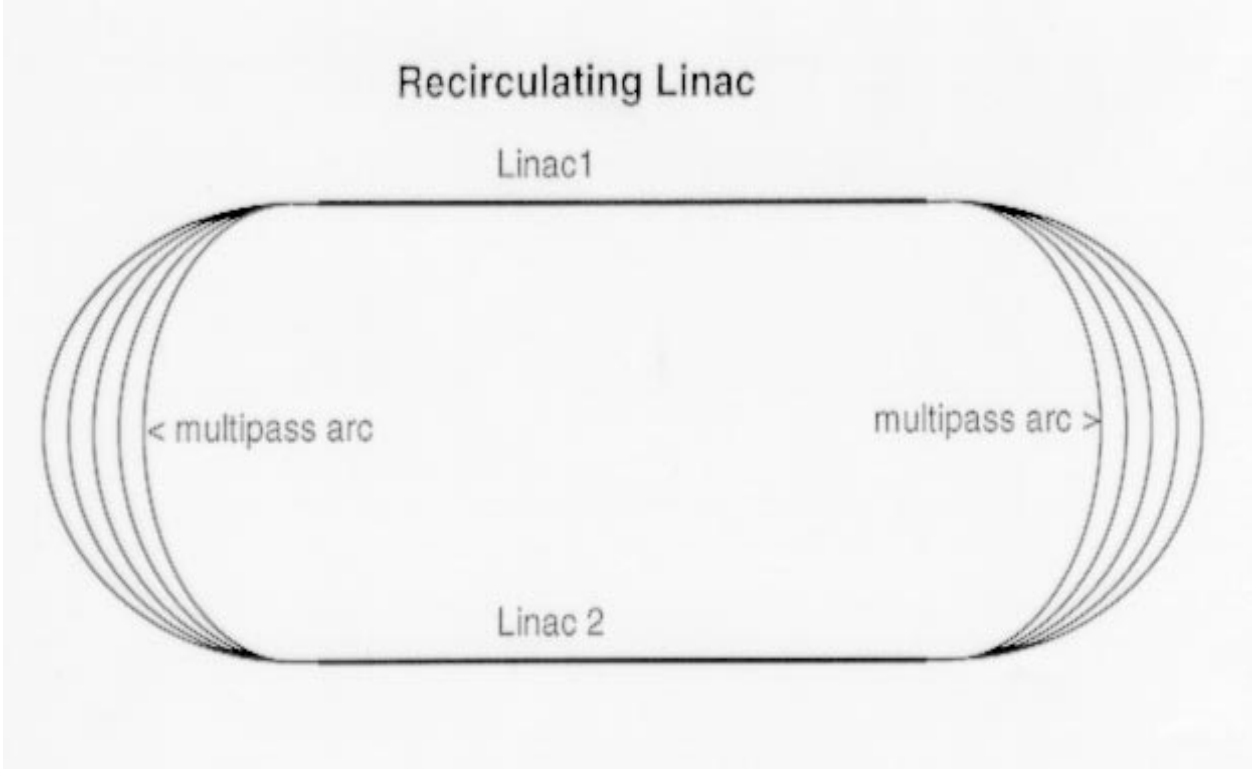


Figure 7.1: Schematic view of a recirculating linac (RLA). The beam is accelerated through several passes of the linacs. On each return arc, the beam passes through a different transport path, matched to the increasing beam energy. Magnetic fields are fixed and the number of return transports (per arc) equals the number of linac passes.

be changed arbitrarily from pass to pass to fit the beam-bunching requirements. Higher-harmonic rf and/or additional compressor arcs can also be added, if needed. (Our prototype RLA acceleration scenario does use a bunch length compressor at the beginning of each RLA.)

The major disadvantage in the RLA is the large cost associated with providing a separate transport for each return arc. Cost-saving transport systems or multi-pass design modifications are desired.

7.5.3 Rapid-Cycling Synchrotrons

A rapid-cycling synchrotron consists of rf accelerating structures within a circular magnetic beam transport, and the magnetic fields are increased from low-field to high-field while the beam is accelerated from low to high energies, passing many times through the same transport system. in Fig.7.2 The magnetic fields must change rapidly to follow the beam accel-

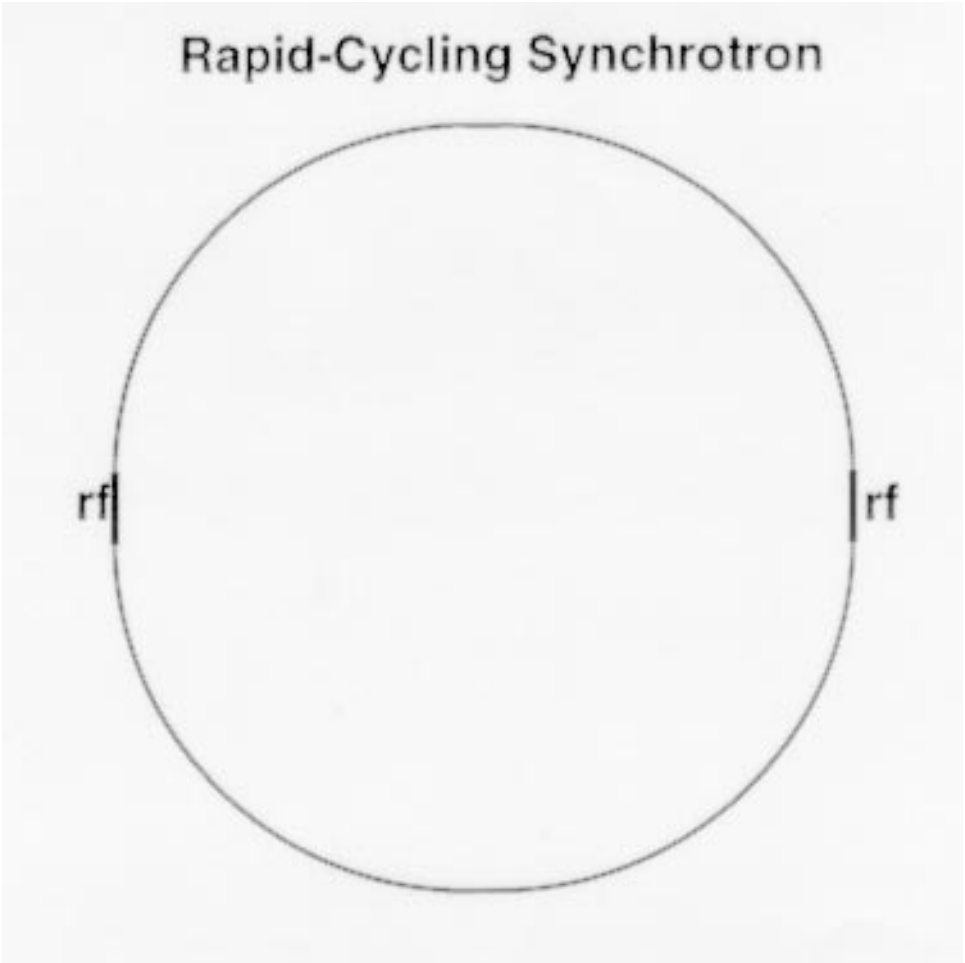


Figure 7.2: Schematic view of a rapid-cycling synchrotron (RCS). The beam is accelerated for many turns through the rf, while the magnetic fields in the ring cycle from low-field to high-field following the beam energy.

eration. In currently established technology, rapid-cycling synchrotrons use iron-dominated conventional magnets ($B < 2\,T$) and cycle at AC rates ($f < 60\,\mathrm{Hz}$). Only high-energy muons have sufficiently long lifetimes for these cycle times. As an example, the beam could be accelerated from 100 to 2000 GeV in a ring with $R = 5\,\mathrm{km}$ ($B = 1.33\,T$) using a 19 GV/turn rf system (1 km of 19 MV/m rf) in a 100-turn cycle, and this would have an acceleration cycle of $\approx 12\,\mathrm{ms}$ with a decay survival of 46.2 %. This would still be a large and rather expensive system. Higher-frequency cycling magnets could be developed, but would require very thin laminations (see below).

Instead of iron-dominated magnets, we are also investigating the recently-developed possibility of using pulsed conductor magnets, and these could cycle at much faster rates and reach higher fields than more established technology. These would more easily meet the

acceleration criterion ($N_{\rm turns} \ll 300B(T)$), and could be used for lower energies. These are discussed in more detail below.

Other rapid-cycling scenarios, using such possibilities as transports containing both high-field and rapid- cycling elements, or obtaining rapid-cycling by rotating fixed-field magnets, are being developed, and are also discussed below. While a practical technology is not yet well established, rapid-cycling elements will probably be incorporated into an eventual μ-accelerator.

7.6 Acceleration Scenario

Within currently developed technology, recirculating linacs appear clearly possible, and are reasonably well matched to μ-accelerator requirements. We will therefore consider RLA acceleration in more detail. As a proof of principle, we present and discuss an explicit acceleration scenario for the 2 TeV collider.

7.6.1 Baseline Scenario Description

In this baseline scenario, the beams are accelerated from 1 GeV to full energy (2 TeV). In this process, the $\mu-$bunch lengths are compressed from $\approx 0.25\,$m at 1 GeV to a length of $\approx 0.003\,$m at full energy. It is also important to obtain the acceleration and bunch compression with minimal phase space dilution, in order to avoid energy-spread blowup and beam losses. These factors of a 1000 in energy increase and 100 in bunch-length compression are not practical in a single RLA. A sequence of RLA's, with rf frequency increasing as the bunch length decreases, are used. A schematic view of such a multi-stage RLA accelerator is displayed in Fig.7.3 and parameters for our initial prototype scenario are displayed in Tb. 7.1.

This scenario is a modularized 4-stage RLA system, with parameters based on discussions in the collaboration meetings. In each stage the energy is increased by a factor of < 10 (1 to 9 to 70 to 250 to 2000 GeV).[1] The rf frequency is also increased from RLA to RLA, from 100 to 350 to 800 to 1300 MHz, as the bunch length decreases. Each of the 4 RLA's consists of two linacs (with 0.5, 3, 8, and 56 GV of rf per linac for RLA 1, 2, 3, and 4, respectively) with recirculating arcs connecting them, and a total of $9 - 16$ turns in each stage. This prototype scenario also includes buncher rf and transport systems (B1, B2, B3, B4) at the entrance of each RLA. Most of the bunch-length compression occurs within the buncher systems, which use the same rf as the RLA acceleration systems but are much shorter.

We have simulated this initial scenario using the program μRLA, and some results are

Table 7.1: Parameters of a 4-RLA scenario

	Buncher 1	RLA 1	Buncher 2	RLA 2	Buncher 3	RLA 3	Buncher 4	RLA 4
Energy in (GeV)	1	1	9.6	9.6	70	70	250	250
Energy out (GeV)	1	9.6	9.6	70	70	250	250	2000
Nturns		9		11		12		16
V_{rf} per linac (GV)	0.08	0.5	0.86	3	2.64	8	10	56
ϕ_s (degree)	90	20	90	15	90	16	90	14
rf frequency (MHz)	100	100	350	350	800	800	1300	1300
gradient (MV/m)	5	5	10	10	15	15	20	20
L(linac) (m)	16	100	86	300.0	176.0	533.3	500	2800
Arc length (m)		30		175		520		3500
B_{arc} (T)		3.4		4.2		5.2		6.0
M_{56} per arc (m)	6.0	$0.4 \rightarrow 1.9$	1.5	$0.1 \rightarrow 0.6$	1.5	$0.15 \rightarrow 0.6$	0.8	$0.3 \rightarrow 2.3$
Time in module (μs)		7.8		35		84.2		672
Decay Losses(%)		9.0		5.2		2.4		3.6
Bunch Length (cm)	$25 \rightarrow 8.3$	4.8	1.4	1.3	0.72	0.59	0.30	0.29
ΔE_{rms} (GeV)	0.05	0.09	0.31	0.34	0.61	0.80	1.5	1.5
Emittance (eV-ms)	13.6	14.0	14.0	14.1	14.0	15.1	15.0	14.2

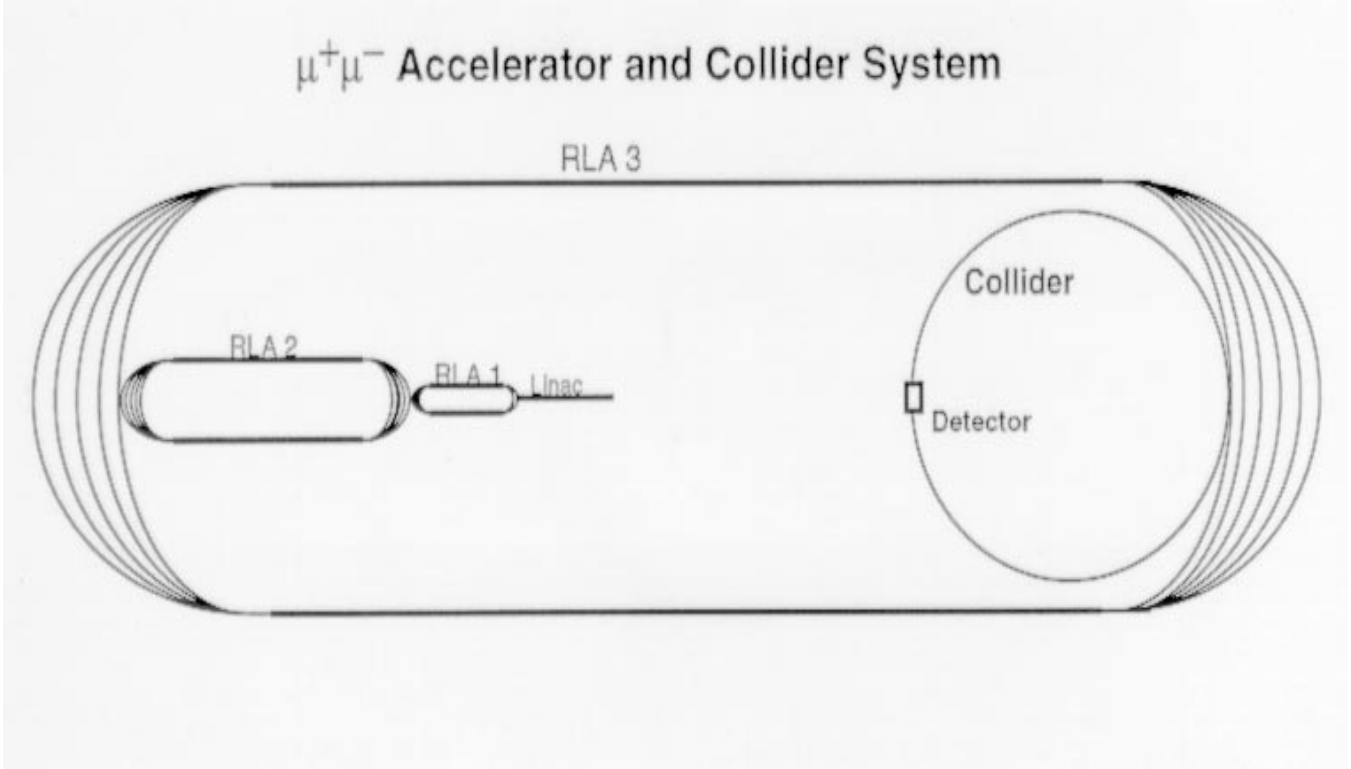


Figure 7.3: Conceptual view of an multi-stage RLA-based accelerator, showing a linac feeding beams into a sequence of 3 recirculating linacs (RLA1, RLA2, RLA3) followed by a collider ring. Note that the drawing is not to scale.

included in Tb. 7.1, and displayed in Fig.7.4. Some phase-space dilution and mismatch does occur, particularly in transfers between RLAs. However the rms emittance dilution is $< 5\,\%$ per RLA or $10\,\%$ over the entire system. Particle loss through the beam dynamics is less than $1\,\%$. Particle loss through μ-decay is somewhat larger, but less than $\approx 5\,\%$ per RLA or $\approx 20\,\%$ over the entire system. (We have assumed mean gradients of up to $20\,\mathrm{MV/m}$ in the linacs, and mean bending fields of up to $\approx 6\,T$ in the arcs.) Bunch compression to $\sigma < 0.003\,\mathrm{m}$ is obtained through rebunching and matching with the frequency increase from RLA to RLA, and is acceptable.

The simulations demonstrate that a cascade of RLAs can provide acceptable acceleration with bunching for a $\mu^+\mu^-$ collider, with minimal dynamic and decay beam loss and emittance dilution. This scenario sets a *proof of principle* baseline for the exploration of acceleration scenarios. It is certainly not fully optimized, and does not exploit the full degrees of freedom possible in RLA scenarios.

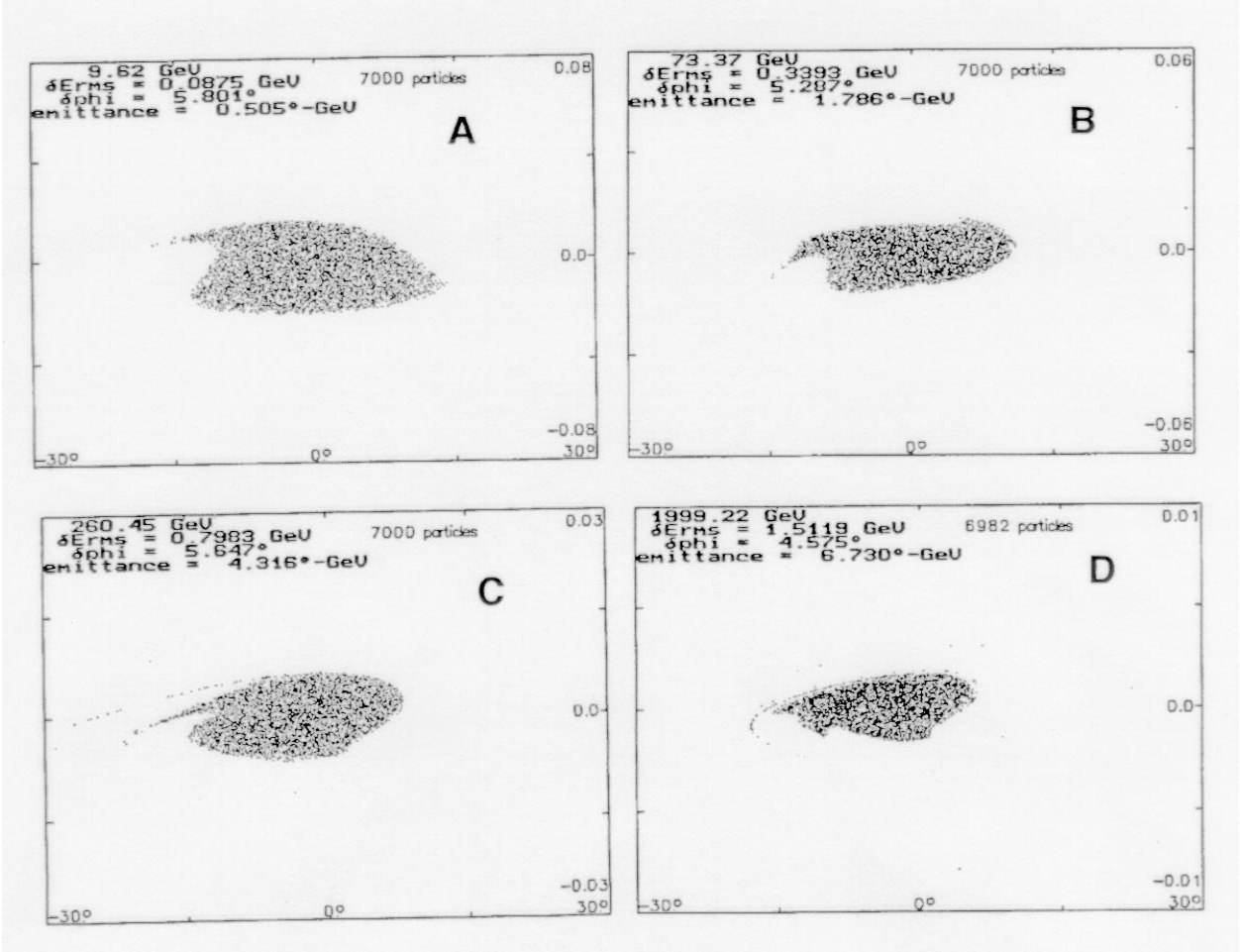


Figure 7.4: Some simulation results from the code μRLA.

7.6.2 Bunching/Acceleration Considerations

The acceleration and bunching must occur with minimal phase space dilution. The flexibility inherent in RLA scenarios permits many variations in compression scenarios, but it is also quite easy to obtain very badly matched schemes within that broad flexibility. Phase-space dilution can be avoided by careful phase-space matching of the beam in each stage of the acceleration scenario.

Within an RLA, phase space dilution can be minimized by matching the bunch in both energy and phase to the stability region or "rf bucket" associated with the longitudinal motion parameters. That defines a region in phase space that extends in phase from $-\phi_s$ to $\approx 2\phi_s$, where ϕ_s is the stable acceleration phase, and in energy spread to

$$\frac{\Delta E}{E} = \pm\sqrt{\frac{eV_{rf}\lambda}{EM_{56}}}\sqrt{\frac{2(\sin\phi_s - \phi_s\cos\phi_s)}{\pi}} \approx \pm\sqrt{\frac{2eV_{rf}\lambda}{3\pi EM_{56}}\phi_s^3} \tag{7.8}$$

where V_{rf}, λ are the rf voltage (per linac) and wavelength, and M_{56} is the chronicity of an arc:

$M_{56} = \frac{dz}{d(\Delta E/E)}$, where $z = \frac{\lambda\phi}{2\pi}$. Efficient acceleration requires keeping the phase ϕ_s as close to crest (0°) as possible. Minimal nonlinear acceleration spread also requires minimal bunch length. In initial scenarios, we have minimized phase-space dilution by minimizing bunch lengths within the RLA and maintaining matched energy spread using the above equation. For fixed ϕ_s this implies that ΔE should remain constant, which implies that M_{56} should increase linearly with energy E. (A similar condition occurs naturally in microtrons.)

Bunch-length compression with minimal phase space dilution is obtained through a separate buncher system in which the beam passes through an rf system at zero crossing ($\phi_s = -90°$) followed by a compressor arc. At zero crossing, the nonlinear content of the rf acceleration is minimized and a very long bunch (up to $\Delta\phi = \pm 60°$) can be compressed with minimal distortion.

Thus in our initial scenario, we have chosen a separated-function approach in which most of the bunching occurs in bunchers at the beginning of each RLA, and bunch-lengths are kept nearly constant within the body of the RLA. This separation minimizes nonlinear acceleration and consequent phase-space dilution. This requires a separate rf system for bunching at the entrance of each RLA (see Tb. 7.1); however these systems are a small fraction of the acceleration rf, and these separate rf systems are possible.

However, this separated function approach is not required, but is an initial simplifying approximation for this first proof of principle example. Modification of phase and M_{56} in the last and first few acceleration half-turns of each RLA can provide the same bunching effect. In future scenario development, the bunchers will be more integrated with the acceleration rf, reducing the need for separate systems and enabling more gradual bunching.

7.6.3 Injector and Matching from Cooling

For injection, we have used idealized ≈ 1 GeV beams with an energy spread and bunch length within reasonable reach of ionization cooling systems. However we do not yet have complete conditions for matching from the cooling system. An initial acceleration and matching system is needed to take the beam from the low-energy of final cooling energy-loss into the RLA system. This will be an ≈ 1 GeV low- frequency linac, with initial bunching, and parameters for such a system are described in Tb.7.2.

Beam production and cooling studies should more precisely define the μ-source, which in turn will specify the injector linac requirements, and the revised matching conditions may affect the initial specifications of the RLA system.

Table 7.2: Injection Linac parameters

Energy in	(GeV)	0.10
Energy out	(GeV)	1.0
Bunch length in	(m)	1.0
Bunch length out	(m)	0.25
$\Delta E_{\rm rms}$ in	(MeV)	5
$\Delta E_{\rm rms}$ out	(MeV)	20
frequency	(MHz)	$30 - 100$
Linac length	(m)	200
decay losses	(%)	8

7.6.4 Scenario Variations

In this baseline design, we have used a 4-RLA system. There is considerable variation allowable in the multiple-RLA concept, and we have previously presented other 3- and 4-RLA designs. The initial and final energies of each RLA, the number of passes, and the rf frequency in each RLA can be changed. In Ref.[2] we described a 3-RLA scenario with 2 to 20 to 200 to 2000 GeV acceleration using 100, 400 and 1600 MHz rf. and in Ref.[3] we also developed a 4-RLA scenario (1 to 9.6 to 80 to 250 to 2000 GeV using 100, 400, 1300 and 2000 MHz). In general, using fewer RLA's and fewer passes simplifies matching and reduces decay losses but increases rf requirements. Phase space acceptance into lower-frequency rf is also somewhat easier. In the present example we have chosen 4-RLA's with frequencies of 100, 350, 800, and 1300 MHz. The 250 GeV end-point of the third RLA is chosen to match a possible energy for a first-stage demonstration collider (250×250 GeV). The rf frequencies of 350, 800 and 1300 MHz are chosen because they match existing and planned SRF systems at CERN (350), LANL (800), and DESY/TESLA (1300). This allows the direct extension of previous experience to this new application, possibly even using some of the same components. In particular, the intensive SRF **R&D** effort at DESY has goals which are very close to our requirements (high gradient, high power, moderate cost, etc.) and could result in large SRF systems that can be adapted to our accelerator, and we have used this frequency for most of our acceleration. The peak frequency of 1300 MHz is also compatible with our high-intensity constraints. Wakefields at the DESY/TESLA level appear tolerable (see below), and a higher SRF frequency should obtain larger wake-fields. In this first scenario, we have used separate bunching rf and transports between RLA's.

This is not entirely necessary, and much of the bunching can be incorporated in the final and initial passes of the RLA's, by changing ϕ_s and M_{56} in those passes. For example, we have found a variant solution with similar performance, in which the bunching at 250 GeV between RLA's 3 and 4 is incorporated into a last half- pass of 3-RLA and the first full pass of 4-RLA. This more gradual bunching also reduces the peak relative energy spreads ($\Delta E/E$) in the initial RLA passes. Multi-harmonic rf systems could also be used to improve linearity, particularly in bunching, but have also not been included in this first scenario.

7.7 Acceleration System Components

In a multiturn RLA system there is a balance between rf acceleration and beam transport costs and requirements. Increasing the number of turns per RLA directly reduces the linac lengths and therefore linac costs, but it also increases the total amount of beam transport, adding cost and complexity. We have not yet developed sufficiently detailed cost estimates that can determine an accurate optimum. In this section we discuss some of the considerations which must be included in developing an optimum design.

7.7.1 rf Considerations

We need a separate rf linac system for each RLA, with lower frequencies for the initial lower-energy RLAs, where the beam has a relatively long bunch length, and higher frequencies for the high energy end, where the bunches are shortened. Very high-gradient is not essential in the acceleration, but rather minimal cost is. The Tb. 7.1 scenario requires ≈ 150 GV of rf cavities; this would require ≈ 7.5 km of accelerating structures at 20 MV/m. This includes $3-4$ separate rf systems, which we have labeled as low frequency (≈ 100 MHz), medium frequency ($300-800$ MHz) and high frequency ($1300-2000$ MHz) systems. We will discuss the various requirements of these sections next.

Low-frequency (≈ 100 MHz)

The first RLA in our scenario uses 100 MHz rf to accelerate beam from ≈ 1 to 9.6 GeV using two 0.5 GeV Linacs in a 9-turn cycle. Tb.7.3 shows parameters for the 100 MHz rf system and Fig.7.5a shows possible cavity cross-sections, with field lines.[4, 5] The 100 MHz cavities are designed to provide an average accelerating gradient of ≈ 5 MV/m through the linac channel.

In order to achieve this gradient, peak surface fields within the cavity will be 18 MV/m while the cavity-to-cavity displacement along the linac line will be 120 cm. The cavities will

Table 7.3: rf parameters for 100, 350, 800 and 1300 MHz systems

Frequency (MHz)	100	350	800	1300
rf type	Cu 300K	4.5K SRF	2K SRF	2K SRF
Cavity Design Source	FNAL-BNL	CERN	LANL	DESY
Cells/cavity	3	4	7	9
Cavity active length 9cm)	3×90	170	127.2	103.5
Cavity total length (cm)	360	240	186	135
Cavity beam aperture (cm)	22.5	37.7	17	10.3
cavity outer radius (cm)	90	37.7	17	10.3
Q	54000	3.2×10^9	3×10^9	3×10^9
Peak field/accel. field	3.6	2.1	2.09	2.0
Average design gradient (MV/m)	5	10	15	20
Reference gradient (MV/m)	5	6	12.5	25
rf peak power (MW)	3×1.9	0.1	0.15	0.21

run at an rf power level corresponding to 1.9 Kilpatrick limit, which is not excessive for this pulsed rf system. The peak and average rf power characteristics (see Tb.7.3) are well within capabilities of standard rf sources. The cavity has a fill ratio of 33 % for the gap length to linac length. This allows additional components to be installed along the beam line such as focusing elements, beam monitoring devices, etc. $100\,m$ of linac will be necessary in order to achieve an acceleration of 500 MeV. This will require a total rf power consumption rate of 360 KW.

Medium-frequency ($\approx 300 - 800$ MHz)

In our scenario, we have used 350 MHz rf in our second RLA. This RLA accelerates beam from 9.6 to 70 GeV in 11 turns and requires two 3 GeV linacs. The rf system must be active for $\approx$ 40 ms at 30 Hz in the present scenario (0.12 % duty cycle). While it is not certain that these must be superconducting, the frequency is quite close to the CERN LEP II SRF system (350 MHz), [6, 7, 8] which operates at 6 MV/m with high cw currents. That technology could be adapted to our application. We need 6 GV of 350 MHz acceleration in the present scenario, and this is roughly three times that installed in CERN LEP II. It is reasonable that improvements (such as adding high-power pulsed processing) plus low-

duty cycle operation would permit gradients of 10 MV/m or (hopefully) more. Experimental studies are under consideration to determine the effectiveness of pulsed operation and of pulsed power processing in these cavities.

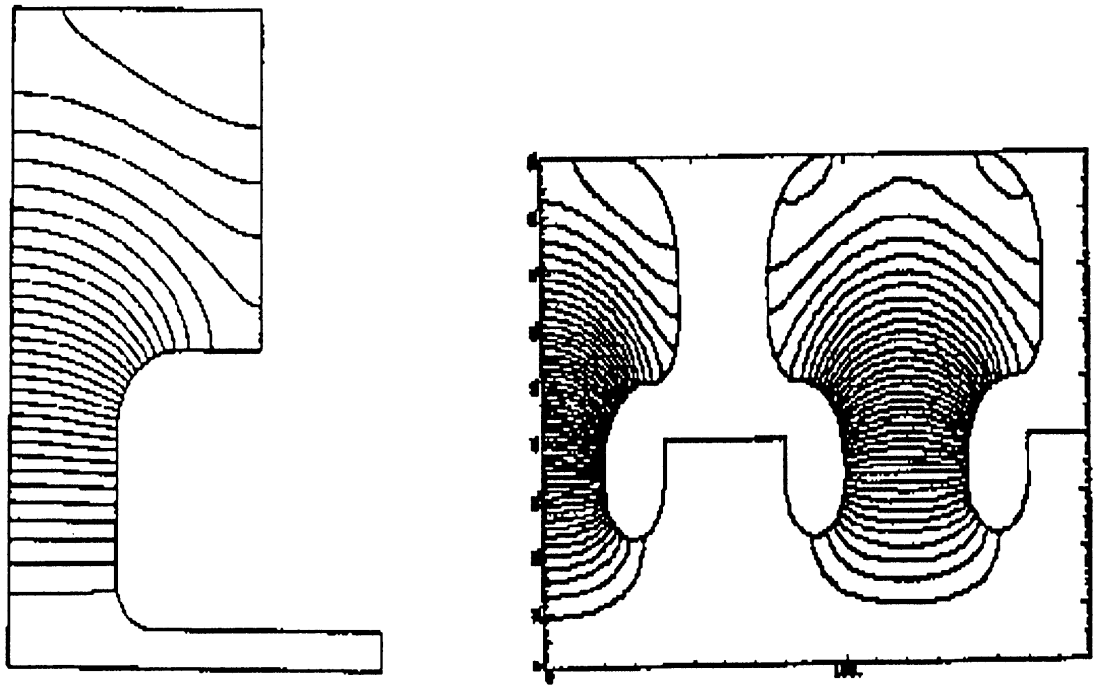


Figure 7.5: 100 MHz rf cavity systems with field lines. Figure shows half of a single cell cavity and half of a 3-cell cavity.

We have used 10 MV/m in the present scenario. Fig.7.6 shows a cross-section of CERN 350 MHz SRF cavities. These cavities are made of copper with a $1\,\mu$ inner layer of niobium, which is magnetron sputtered on the interior. The copper provides good electrical and thermal conductivity for the bulk of the cavity while the niobium layer provided the superconducting cavity surface. Solid niobium cavities were also built; niobium-sputtered copper cavities were preferred because of better thermal stability, lower surface resistance, reduced external magnetic- field effects, and lower cost.

The present scenario uses 800 MHz rf in the third RLA (An alternative scenario using 1300 MHz has also been developed.). This RLA accelerates from 70 to 250 GeV in 12 turns, and requires two 8 GeV linacs. The rf system must be active for ≈ 100 ms (0.3 % duty cycle). The choice of 800 MHz is based on Los Alamos experience with 800 MHz SRF; 800 MHz was the PILAC design frequency.[9] The Los Alamos APT program also plans to develop SRF systems at 800 MHz, and it is likely that the technology will be adaptable to high-gradient μ-

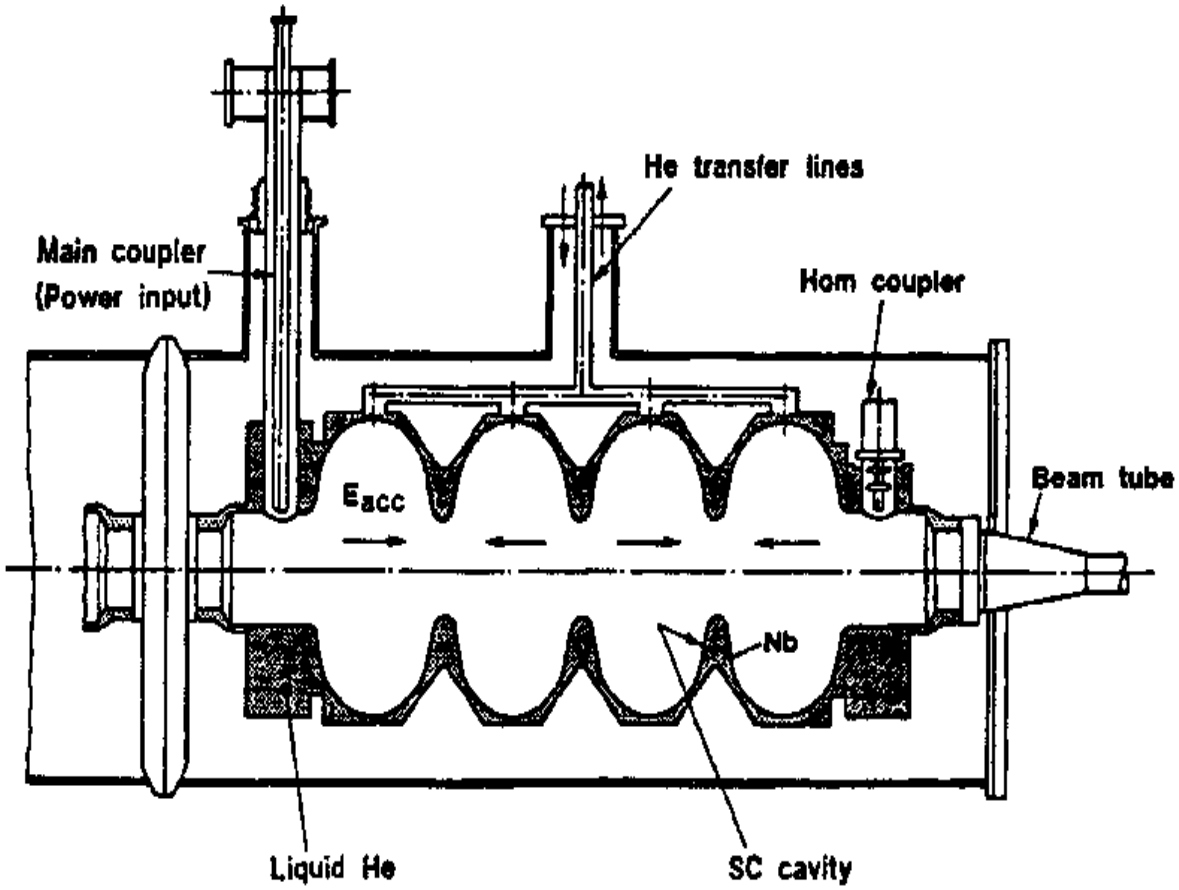


Figure 7.6: 350 MHz rf cavity system. Cross section of the superconducting CERN cavity.

acceleration, and could be included in our scenario. From Los Alamos experience a gradient of 20 MV/m seems possible,[10] and we have used 15 MV/m in Tb. 7.1. PILAC was designed for 12.5MV/mand 16MV/m was achieved. An advantage of 800 MHz systems are that they are (approximately) the highest frequency Nb rf systems that can operate at 4K liquid He temperatures, which simplifies cryogenic requirements and improves efficiency.

High-frequency ($\approx 1300 - 2000$ MHz)

In the present scenario, a 1300 MHz accelerator is used for the $250 - 2000$ GeV RLA. (We have also considered 2000 MHz for the $250 - 2000$ GeV RLA.) The accelerator uses two 56 GV linacs to accelerate beam to full energy in 16 turns. For economy, we require high gradient ($\approx$ 25 MV/m) acceleration and the rf cavities must sustain field throughout the multipass acceleration time, which is 0.8 ms in the 2 TeV RLA, which implies a 2.4 % duty cycle. This is the largest acceleration system and it dominates the total SRF requirements.

The parameters imply that SRF cavities should be used, and SRF technology does promise high-gradient at these duty cycles. The TESLA SRF is designed at parameters extremely close to ours: up to 25 MV/m gradient at 1300 MHz, active for 1 ms cycles repeating at $5 - 15$ Hz,[11, 12] and we plan to extend the results of that **R& D** effort to obtain the SRF system for our main acceleration system.

The TESLA SRF cavity is a 9-cell 1300MHz structure and is displayed in Fig.7.7.

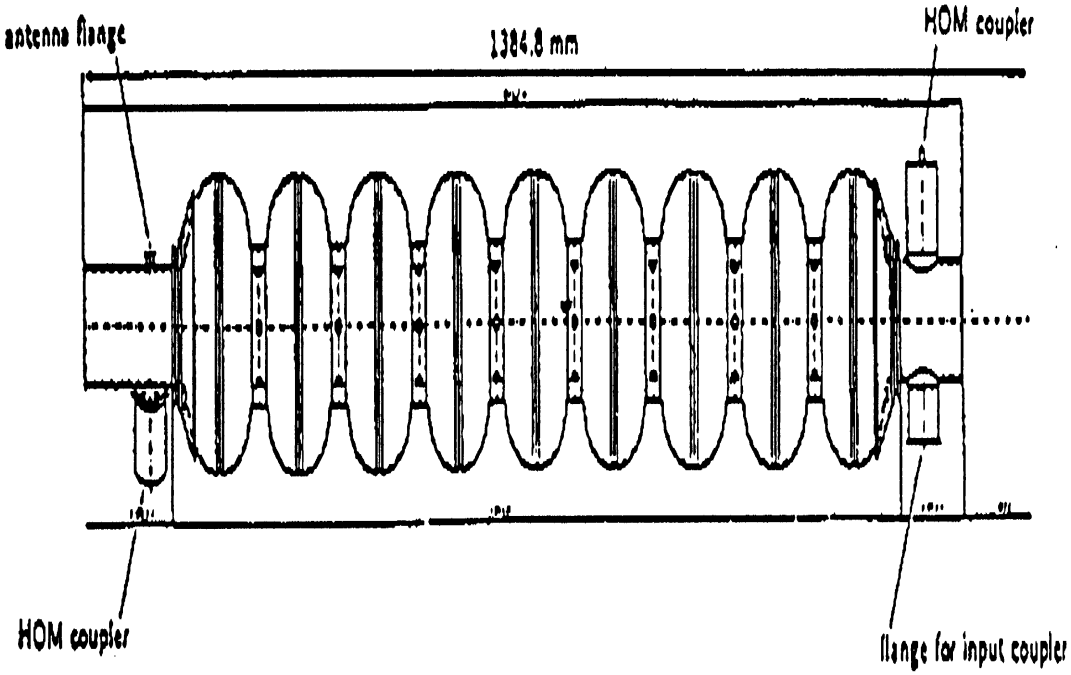


Figure 7.7: A TESLA 9 cell 1300 MHz cavity.

These cavities are made of high purity niobium sheets (2.8 mm thick) by standard fabrication methods (deep drawing and electron beam welding). It is critical in the cavity shape design to minimize the ratio of the peak to accelerating electric fields E_{peak}/E_{acc} (2) and the ratio of peak magnetic field to accelerating field H_{peak}/E_{acc} (4.2 MT/MV/m). The number of cells per cavity was limited to 9 because of higher order mode (HOM) damping requirements. There are one input power coupler and two HOM couplers (with $Q_{ext} = 10^4 \quad 10^5$) per cavity.

Because of the higher peak currents the HOM power requirements should be significantly larger in our application than in the TESLA case, so the couplers would need to be modified to handle the higher power. However, because the bunches are more widely spaced, the HOM damping rate need not be increased.

A TESLA cryomodule consists of 8 SRF cavities (a cost-effective choice) and is 12.2 m long. The total rf heat load to 2K in a cryomodule is estimated to be 21.4 W. The static heat load budget is 2.8 W at 2K, 14 W at 4.5K and 77 W at 70 K. We may need some additional cryogenic capacity to handle heating from muon decay products, although most of the electrons from decay will simply continue through to the end of the linac. HOM load cryogenics will also be modified. Magnetic shielding is included to improve cavity Q and the shielding reduces the residual external magnet field to less than 20 mgauss. 4 cryomodules (50m) share a klystron with 10 MW peak-power.

To obtain the highest possible gradients and cavity quality factors (Q), the TESLA **R&D** program is exploring methods to reduce and eliminate field emission (FE) and thermal breakdown in TESLA cavities. Key technologies which they are studying include : 1) Semiconductor industry standard cleaning techniques to remove FE particles from the rf surfaces, 2) Material removal from the cavity surfaces by chemical etching, 3) Increasing the thermal conductivity by employing Ti solid-state gettering in a pre-purification treatment in a ultra-high vacuum oven, 4) Rinsing cavities with high pressure deionized water to remove surface particles, 5)high-power rf pulse processing to remove the remaining FE sites. Significant progress has been made in these efforts: A recent prototype cavity has reached $E_{acc} = 26\text{MV/m}$ with $Q = 3 \times 10^{10}$ in CW operation and $E_{acc} = 31\text{MV/m}$ in pulsed operation.

In other scenarios, we have also considered 1600 MHz and 2000 MHz SRF systems for the final μ-accelerator, and these frequencies could be used in a final configuration. However we expect wakefield limitations to be worse for higher frequencies (see below). Also we expect that the intensive research in 1300 MHz cavities for TESLA will produce optimal high-gradient configurations which can be extended to our case, and therefore minimize the subsequent **R&D** requirements. [13]

Wake-field Considerations

A high-luminosity $\mu^+\mu^-$ collider will have very high-density bunches, with 10^{12} or more particles per bunch. At these high-intensities, collective effects can be important. In the short bunches prepared for the collider, the dominant effect is expected to be the short-range wakefield.

Mosnier and Napoly[14] have evaluated wakefields in the TESLA 9-cell structure (1300 MHz, $1\,m$ length), which is designed to accelerate at 25 MV/m and is thus very similar to the rf system for the high-energy high-frequency RLA(s). They obtain a maximum wakefield across an electron bunch of 1 mm length of ≈ 15 V/pC. For 10^{12} μ's (1.6×10^5 pC), this is 2.4 MV or almost 10% of the accelerating voltage. Our bunches are 4 times longer than 1mm, and the short-range longitudinal wakefield is reduced in proportion to the square root of that length, so the wakefield would be a factor of 2 smaller. The wakefield is also proportional to $1/a^2$, where a is the cavity aperture. This aperture is naturally larger for longer wavelength (lower frequency) rf. The TESLA cavity aperture could be somewhat increased, by up to a factor of ≈ 2 to reduced wake-fields, but with some degradation of other cavity parameters. We may also have more than 10^{12} μ's per bunch for high luminosity, and scenarios with up to 4×10^{12} have been generated. For our recirculating linac scenarios, we expect that the largest wake-field effects will occur in the highest- energy (2 TeV) recirculating linac, since that linac has the highest frequency rf and the shortest bunches. We have studied these effects by simulations of particle motion which include wake-fields in the final 200 to 2000 GeV linac of the 3RLA scenario of ref. 3. To include wakefield nonlinearity effects in our simulations, we have used a simplified short-range model in which the longitudinal wakefield deceleration on each particle is proportional to the charge in front of the particle, with the full bunch charge giving the total wakefield. (This model was used in the CEBAF FEL design. [15]) Following the TESLA values (scaled to 4mm bunch length) we estimate a total wakefield of 7.5 V/pC, or 1.2 to 4.8 MV wakefield per 25 MV acceleration for $1-4\times10^{12}$ μ's. The first-order and second-order wakefield effects (magnitude and slope) can be compensated by increasing the rf voltage and changing the accelerating phase. Higher order effects are not compensated, and can give nonlinear distortion to the motion, causing emittance dilution and eventual beam loss. Some simulation results are displayed in Tb.7.4 and in Fig.7.8. For 1.0 and 2.5 MV (per 25 MV) cases we can increase the rf voltage (by 4% and 12.4%, respectively) and shift the rf phases from 12° at zero wakefield to 18° and 25°, respectively. We then obtain similar performance to the zero wakefield case, with similar distortion and phase space dilution. For 5 MV, the rf voltage would need to be increased by 30% and the rf phase ϕ_ss moved to 35°. Significant orbit distortion is seen (emittance dilution of 30%). Although no beam loss occurs, the phase space distortion is at the limit of acceptability. Thus for moderate size bunches ($1-2\times10^{12}$) the wake-fields can be compensated, but much larger charges could lead to significant distortion and beam loss. In further studies, it has also been found that the wake-field effects are reduced as the number of acceleration turns increases, since the beam develops more synchrotron oscillations which average the effects over the beam.

Table 7.4: μRLA simulation results with wakefields for a 2 TeV recirculating linac. In these simulations we used initially Gaussian beams with 20 eV-ms normalized rms emittance at 200 GeV.

Case	Wakefield amplitude	Accelerating phase	rf voltage depression	Bunch Length	$\delta E_{\rm rms}$ (GeV)	Final Emittance
1	0	13°	0%	5.58°	2.41	22.5 eV-ms
2	1 MV/m	18°	4.5%	4.89°	2.68	22.0
3	2.5	25°	12%	6.06°	2.17	21.6
4	5	35°	26%	6.66°	2.71	31.3

In summary, the μ-collider design intensities are close to the intensities at which wake-field effects can become a limitation. Significant monitoring of the effects and accurate evaluations of the wake-fields are needed; it is important to ensure that the adverse effects are truly minimized.

7.7.2 Transport Considerations

The beam transports for the recirculation arcs are relatively straightforward, but are nontrivial, since they require good transverse matching throughout the system to avoid emittance dilution. High field is not required, and even conventional fields (B< 2T) can be adequate. Since the beam passes through a different return arc on each turn, the total amount of beam transport is relatively large (≈ 160 km of arcs in the baseline scenario). The transport can easily become very expensive, so cost-saving designs are needed, such as multiple-aperture or rapid-cycling hybrid designs, and these are presented below.

Transport lattices

Each transport must be achromatic (matched to zero dispersion), and also must have a chronicity M_{56} matched to the bunching requirements. A transport modeled on the CEBAF RLA could be used. The M_{56} values are small compared to the natural chronicities. The average dispersion in an arc (given by $\eta = M_{56}/\pi$) varies from $\approx$ 0.1 to 1 m in these cases. Flexible momentum compaction lattices, where the average η is reduced by including perturbations to negative η, could be used for some of these arcs.[16] Note that at the beginning and end of the arcs beam-separation and beam-recombination transports for all passes must be inserted, and this adds considerable complication. CEBAF[17] has a 5-pass

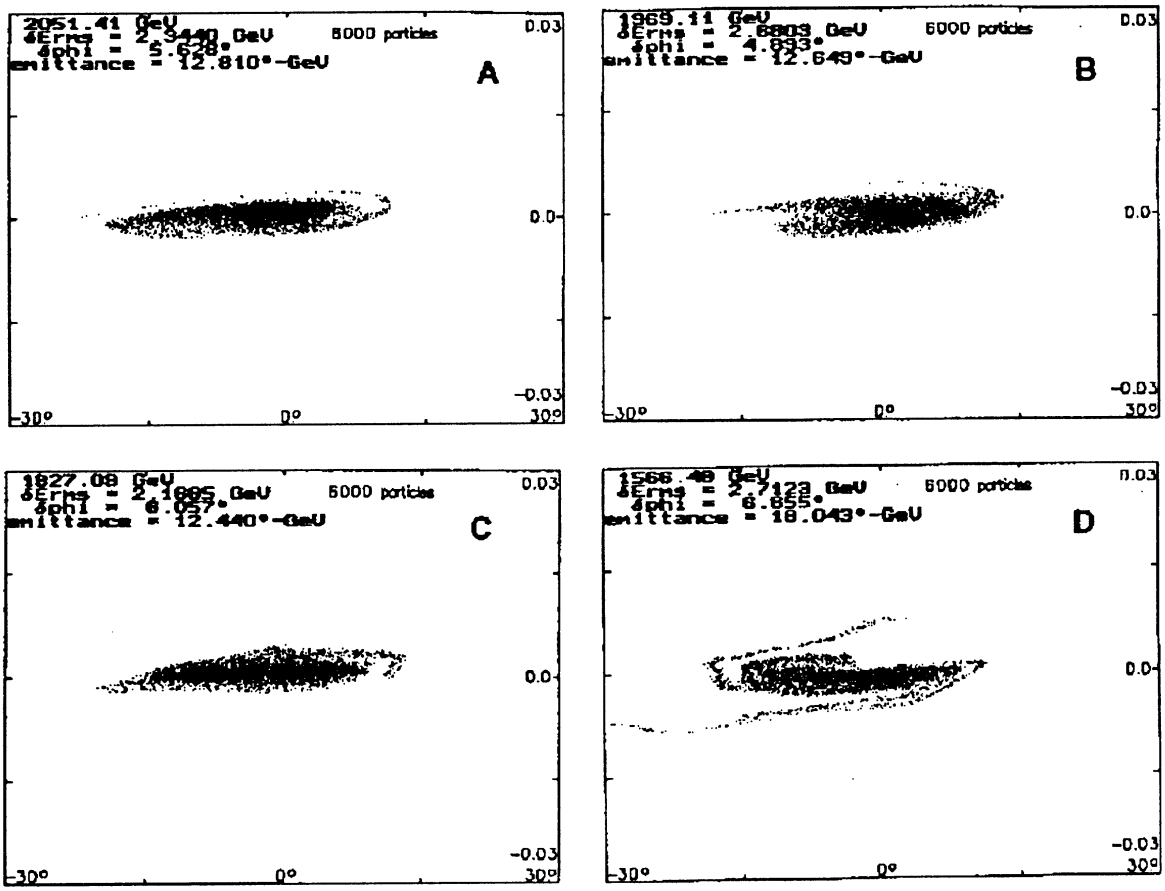


Figure 7.8: μRLA simulation results with wake fields, with beam accelerated from 200 to 2000 GeV in a 10-turn RLA.

separation and recombination system with carefully matched transports, and it is easy to imagine an $\approx$ 10-pass extrapolation of that system to our case, and that is what we have used. However many more passes (> 20?) may lead to impractically congested designs.

A significant concern is the relatively large energy spreads (up to 5% rms) which occur in the initial turns of the lower energy RLA's. Detailed design of arc transports which can accommodate these, without losses or emittance dilution, is a challenge. It is likely that the scenarios should be modified to reduce the energy spread requirements, possibly by keeping the bunches longer, which may require lower-frequency or multi-harmonic rf systems.

Recirculating Arc Magnets

Each recirculating linac (RLA) has two long, parallel linacs with a large energy gain per pass, and semi-circular arcs of fixed-field superconducting magnets at each end. Fig.7.9 is a

sketch of a RLA, showing the two arcs and two linacs and, enlarged, the separation, phasing, and steering magnets which direct the muons between the linacs and the recirculating arcs.

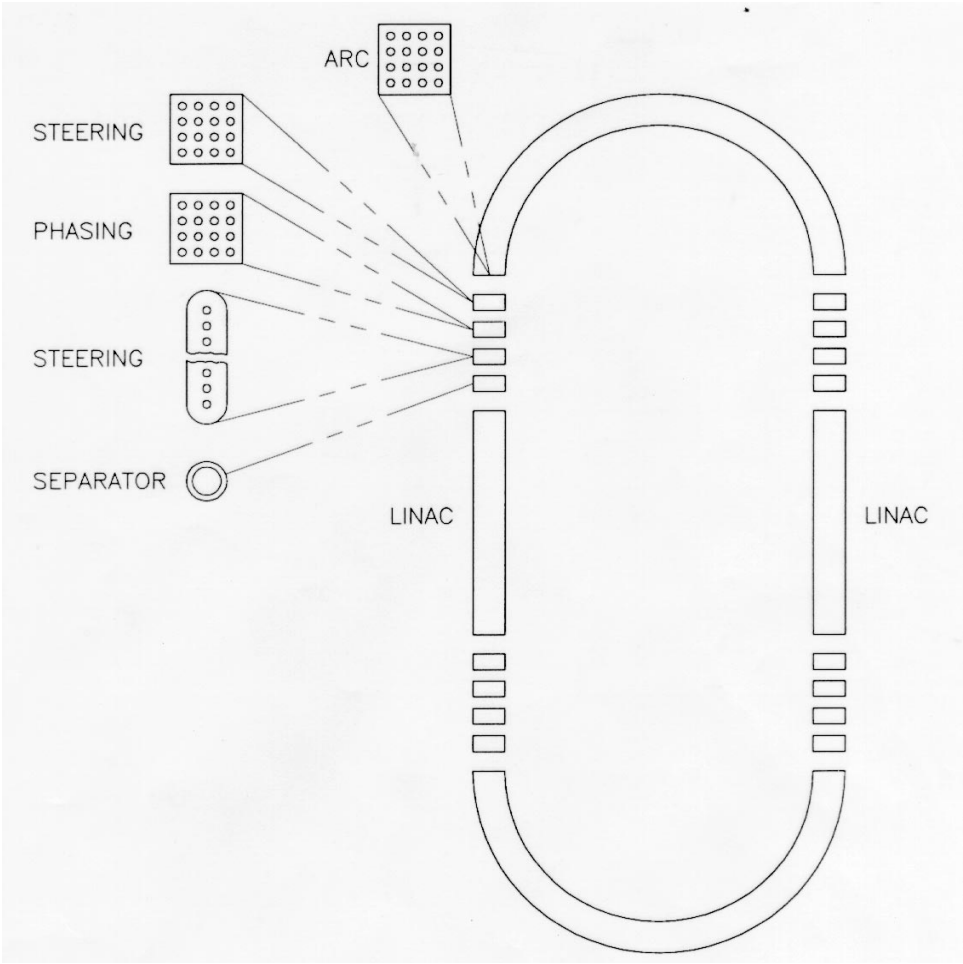


Figure 7.9: Overview of a μ recirculating linac, showing the arcs with beam separation and multiple-aperture magnetic transport.

The number of turns in each RLA depends upon a balance among cost, complexity and performance considerations of rf and beam transport components. These considerations may require as many as 20 recirculating passes; we present a magnet design for 16-passes in the highest energy RLA, corresponding to the current acceleration scenario. We use a multiple-aperture magnet design, in which the passes go through separate (different field) apertures in the magnetic structure.[18]

The design is presented in Fig.7.10, which displays a cross-section of a set of four stacked magnets, each containing 4 apertures, all within the same pressure vessel (cryomodule). The aperture at the top left has a field of 7 T, and there is an 0.406 T difference between adjacent apertures in a stack. With superconducting cable at SSC parameters, two layers of cable are

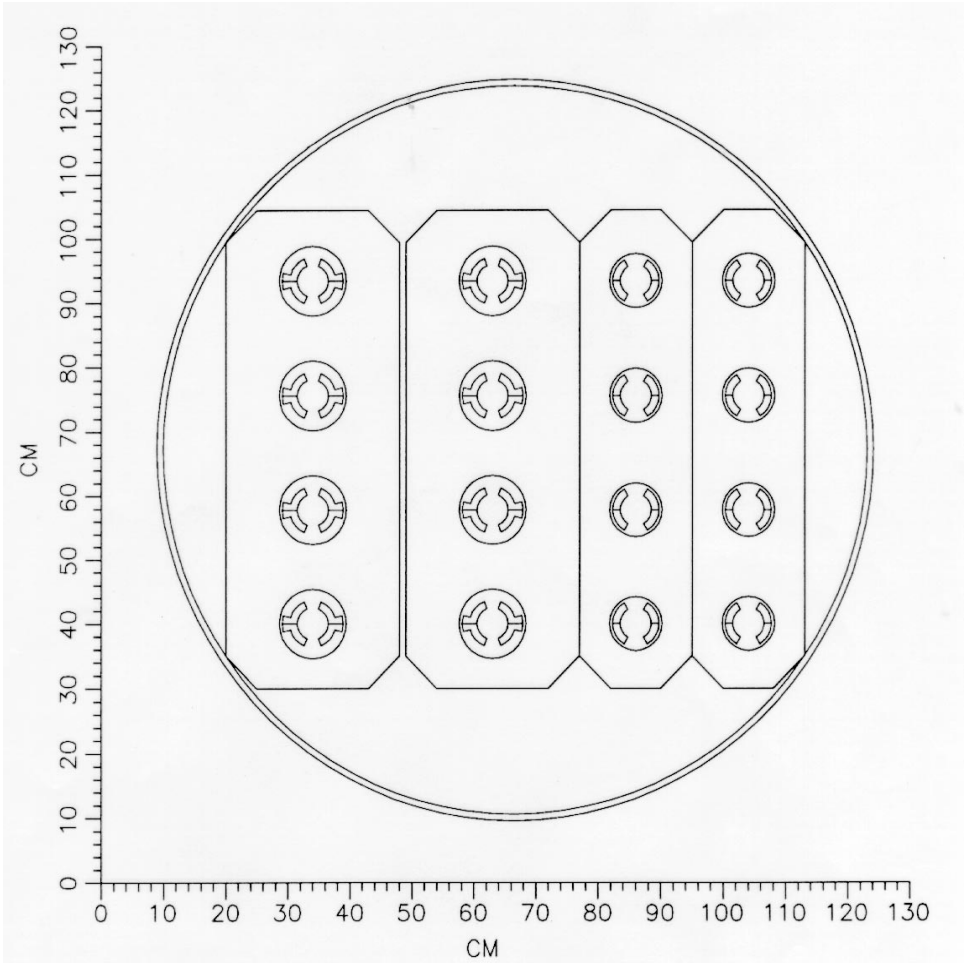


Figure 7.10: A 16-aperture dipole, composed of four stacks of four apertures. The highest field (7 T) aperture would be in the lower corner.

required for fields above ≈ 4 T, and we have arranged that all apertures in a stack have the same configuration. Thus there are 8 apertures with two-layer coils and 8 with single-layer coils. The two-layer coils eliminate sextupole and decapole components; the single-layer coils eliminate sextupole components. The two high-field stacks have two layers of superconductor in the apertures, each 12 mm thick. The highest current density is $31,000\,\mathrm{A/cm^2}$ in the 60 coils of the 7 T aperture (similar to the $29,000\,\mathrm{A/cm^2}$ in the inner layer of the 6.6 T SSC dipole). The lower currents for the lower field apertures would be obtained by using fewer turns.

The stacks in Fig.7.10 result in reduced reluctance for the flux return path (compared to individual magnets), thus reducing the amp-turn requirement in each aperture. However the stack arrangement would naturally have a vertical (skew) gradient in the field. This is eliminated by putting different currents in the upper and lower coils in an aperture. There

is also sufficient flux leakage out of the highest field stack to influence the field in the stack adjacent to it. However, the field in the space between these stacks is low enough that a thin foil layer of pure Nb, which has an H_c of $\approx 1.4 \times 10^5$ A/m (1800 Oe), could be used as an inexpensive shield, eliminating the flux leakage.

In this design, the pressure vessel also serves to constrain the transverse forces on the iron, which include the Lorentz force and the coil pre-stress forces. The wall of the pressure vessel could easily contain the assumed 20 atmosphere He pressure. However, the end plates must withstand Lorentz force on the coil ends and the He pressure, and must have penetrations for 16 beam tubes; these end plates could therefore be difficult to design and assemble.

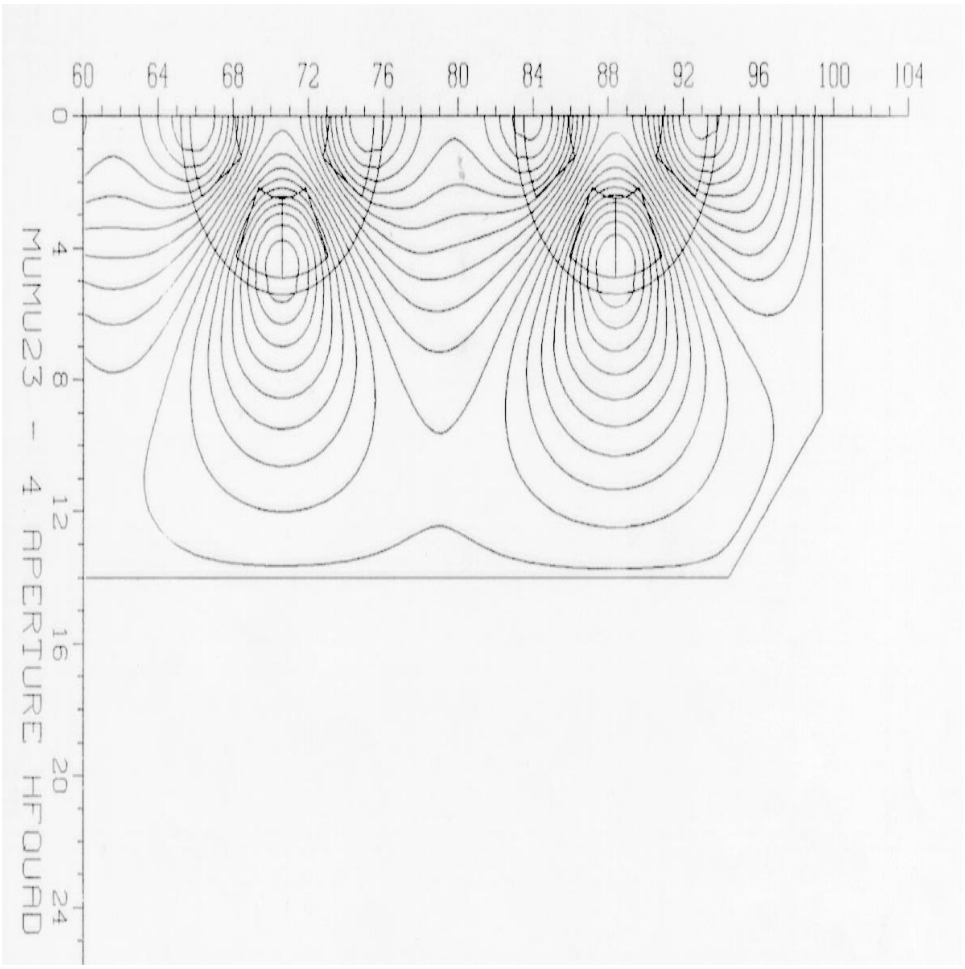


Figure 7.11: Field lines in a quadrant of the highest field dipole and quadrupole stack of the 16 aperture dipole.

Fig.7.11a shows the field lines in the high field dipole stack. A similar stacked-aperture design is possible for the quadrupoles. Fig.7.11b shows the right half of the topmost quad aperture, and half of the next lower aperture, in a 5-aperture quadrupole stack with dimen-

sions matched to the dipole stack. The field lines are those resulting from readily achievable current densities in the 30 current blocks. In the highest gradient aperture, the gradient is $175\,\mathrm{Tm}^{-1}$.

An alternative scenario would be to have a separate pressure vessel for each aperture, i.e., up to 16 separate magnets. This would require more power but have less inter-magnet interference, but could still be economic, if the magnets are modularized for minimal cost. Magnet designs for fewer passes are also possible and should be proportionately easier; in Fig.7.12 we show a cross-section of a 3-magnet stack for a 9-pass arc.[18]

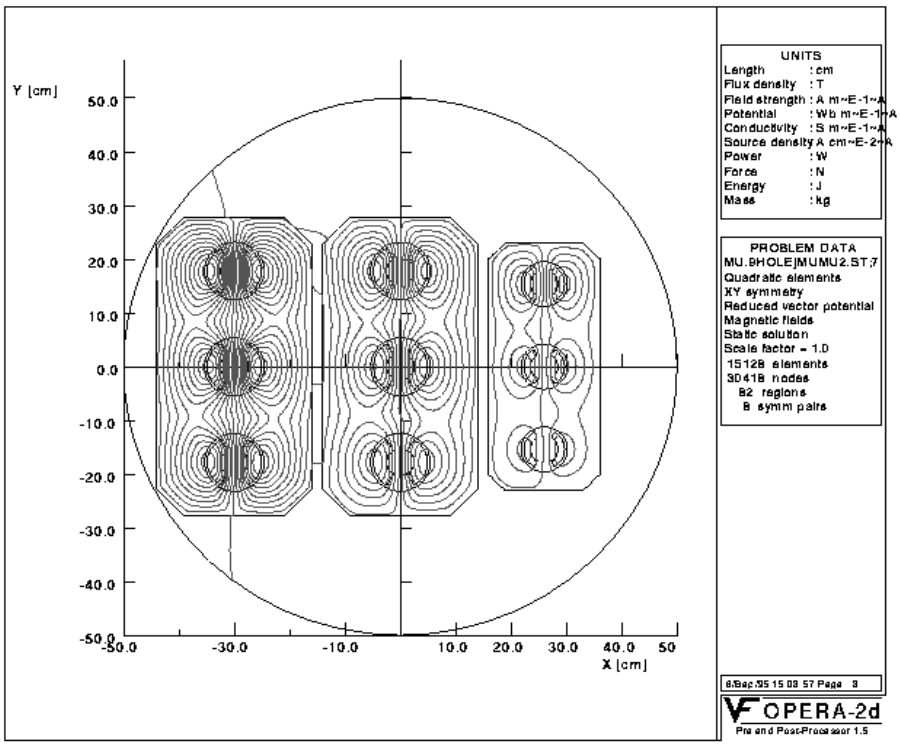


Figure 7.12: A 9 aperture dipole, composed of three stacks of three apertures.

We have used superconducting high-field magnets in this initial design. Other *low-cost* technologies could be used (permanent magnets, super-ferric, etc.), either in combined or separated elements. High field is not required, if the total number of turns remains modest.

Hybrid Magnets

The total length of beam transport, and therefore the cost, could be reduced by cycling the magnetic fields in the return arcs so that the same transport line can be used for multiple

acceleration passes. (In the limit where the same transport line is used for all passes, we obtain the case of a rapid-cycling synchrotron.) However, superconducting magnets cannot cycle at the high rates needed for μ acceleration.

One approach to multipass transport is the use of hybrid magnets, in which rapid-cycling and superconducting high-field magnetic elements are mixed and pulsed so that several passes can go through the same transport.

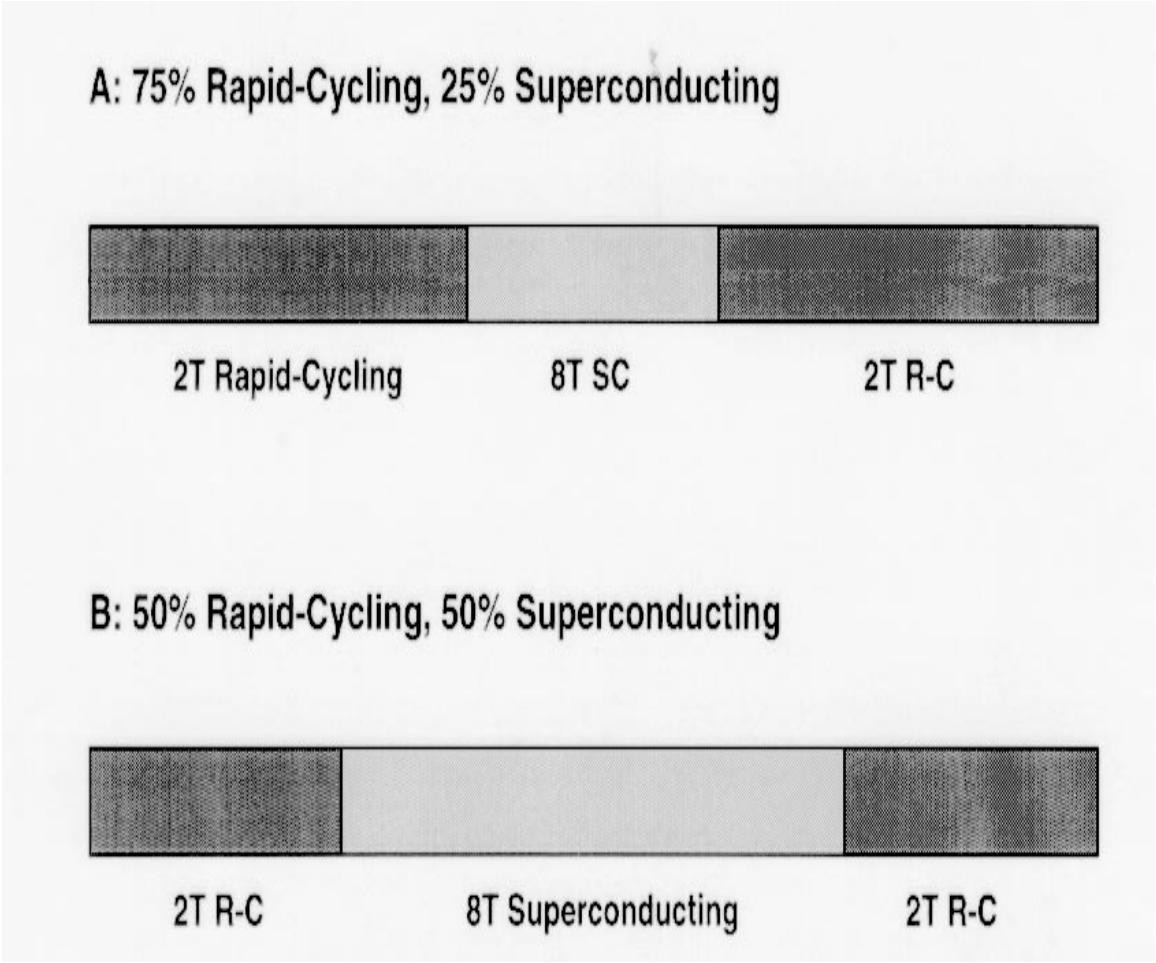


Figure 7.13: Configurations for hybrid rapid-cycling dipoles.

Figure 7.13 shows potential layouts for hybrid magnets, in which high-field magnets are surrounded by rapid-cycling magnets. In this case the rapid-cycling magnets are iron-dominated, and can cycle from $-2\,T$ to $+2\,T$, while the superconducting section is fixed at $8\,T$. Fig.7.13b shows a design with 50% high-field and 50% rapid-cycling, so the mean field can increase from $3.0\,T$ to $5.0\,T$, Fig.7.13a shows a design with 25% $8\,T$ and 75% $12.0\,T$ rapid-cycling; it would take the mean field from $0.5\,T$ to $3.5\,T$. Thus a sequence of two arcs of compound magnets can accommodate acceleration by almost an order of magnitude, which

is what is needed in the accelerators. One can also design compound magnets to obtain 3 or 4 arc versions, with extended capabilities.

A significant problem is the large degree of orbit variation within a magnet unit as the matched beam energy changes, because of the differential bending between the outer cycling magnets and the fixed inner magnets. In the first of the above two cases, we obtain a variation of the central orbit of 5 mm for a 10 m compound magnet with a total bend of 0.0075 radians (2 TeV). In the second case, which has a factor of 5 energy change, the same bend produces an orbit change of 3.3 cm within the cycle. While somewhat extended, these orbit changes are not impossibly large, and magnets can be designed to accommodate them. In the following section these magnet-cycling possibilities are extended to obtain rapid-cycling magnets for rapid-cycling synchrotron scenarios.[20]

7.8 Other Acceleration Scenarios

In this section, we discuss other methods of acceleration than the recirculating linac scenarios and their extensions discussed above. These are all various approaches to obtaining a rapid-cycling synchrotron acceleration. Since all multiturn methods are constrained by the lifetime to $N_{\rm turns} << 300\,B$, relatively large rf systems are still required. However, major cost savings could be obtained in the elimination of multi-pass transport systems by using transports with changing magnetic fields. Cycling magnetic fields at the rate needed for μ acceleration is a significant challenge and requires untested approaches. However we have identified some possible methods, which we describe below.

7.8.1 Pulsed Dipoles

Magnets that are ramped up quickly in field as the muons gain energy in multi-pass linacs can be used in rapid-cycling acceleration of the beam.[21] A pulsed magnet scheme appears feasible for a 250 GeV machine, but the parameters appear less favorable for a 2 TeV machine. A particular design for a pulsed dipole magnet with parameters chosen to be practical is shown in Figs.7.14 and 7.15, and some parameters are displayed in Tb.7.5. Fig.7.14 shows a quarter of the magnet, in which four turns of multistrand copper formed into a trapezoidal shape are used in each of two required coils. The turns are placed in an approximate cos θ current distribution to maximize the field and minimize unwanted harmonics. The iron yoke contributes substantially to the field and provides mechanical support. Cooling requirements are modest and can be satisfied by circulating water in pipes (not shown) passing through the yoke. The magnet shown in Fig.7.14 produces a field of 4 T at a current of 29.5 kA.

Table 7.5: Parameters for pulsed conductor-dominated accelerator and storage ring dipoles

Parameter	unit	Accelerator	Storage Ring
		Dipole	Dipole
Coil inner radius	cm	2	2
Magnet length	m	10	10
Field	T	4	6
Current	kA	29.5	24.9
Stored energy	kJ	160	360
Inductance	mH	0.37	1.2
Coil resistance	mW	19	44
Ramp time, 10% to 90%	μs	360	
Store Time (for 250 GeV)	μs		5000
Power supply voltage	kV	31.2	1.1
Power into magnet 2 Hz	kW	19	452
Power into 250 GeV ring	MW	2.7	39.4

The conductor cross sectional area is 1.275 cm^2. For a magnet length of 10 m, the inductance is 0.37 mH and the resistance is 19 mW. If 250 GeV corresponds to 90% of full field, a total of 146 such dipoles is required.

Assuming two linacs with a length of 313 m each and an accelerating gradient of 9 MeV/m, and a dipole filling factor of 70% in the arcs, the time required to accelerate the muons from 25 GeV to 250 GeV is 360 μs (40 turns). The power to drive the magnet during this acceleration time can come from the discharge of the stored energy in a capacitor bank. A design with such an LRC power supply, producing the waveform shown in Fig.7.16, has a quarter period of 550 μs. Note that the required acceleration is not linear but follows the sinusoidal field increase in the magnets. The voltage required is 31.2 kV and the required storage capacitance is 340 μF. Each magnet has its own power supply and it is triggered to discharge in synchronization with the acceleration cycle. After a quarter cycle, the energy is recovered by the power supply in the next quarter cycle, with an efficiency of $\approx 80\%$.

The required voltage (31.2 kV) is uncomfortably high. It can be reduced by connecting some or all of the turns in the magnet in parallel, fed from separate subsections in the power supply. Power supplies with power output similar to that required here operating

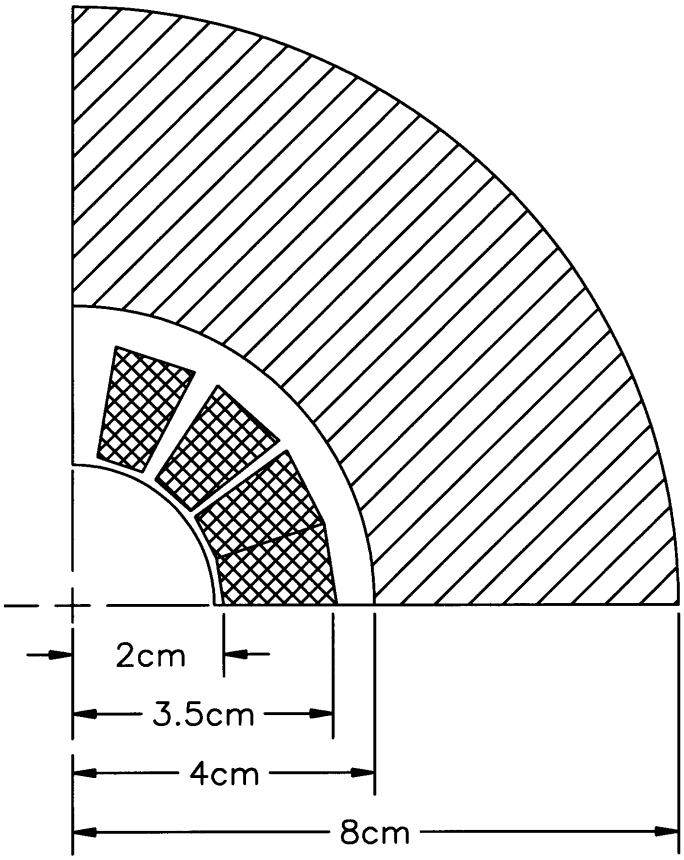


Figure 7.14: Cross sections of pulsed current dipoles for a μ rapid-cycling accelerator dipole (4 T).

at voltages 5 kV are used at accelerators to inject beam or to capture antiprotons in fast kickers.[21] During a half-cycle, the heat deposited in the coils is 9400 J. This gives an estimated temperature rise per cycle in the coils of 0.13 C. At a repetition rate of 2 Hz, the average power dissipated in each magnet due to this resistive heating is 19 kW. For the entire machine (144 dipole magnets), the power dissipated in these magnets at 2 Hz is 2.7 MW. Tb.7.5 includes the parameters for this particular design. The calculations used to obtain these values are approximations and further work will be needed to refine the results.

A similar design approach using pulsed magnets can also be considered for the collider ring. Here, the magnet current need not rise as quickly, but the magnet must have a constant flat-top current for 5 ms for 250 GeV collisions. A possible design is also shown in Fig.7.15. With two layers of turns, a field of 6 T can be achieved. With a dipole filling factor of 0.7, the time for 1000 orbits is 4.2 ms. A current of 24.6 kA and a voltage of 1.1 kV is required

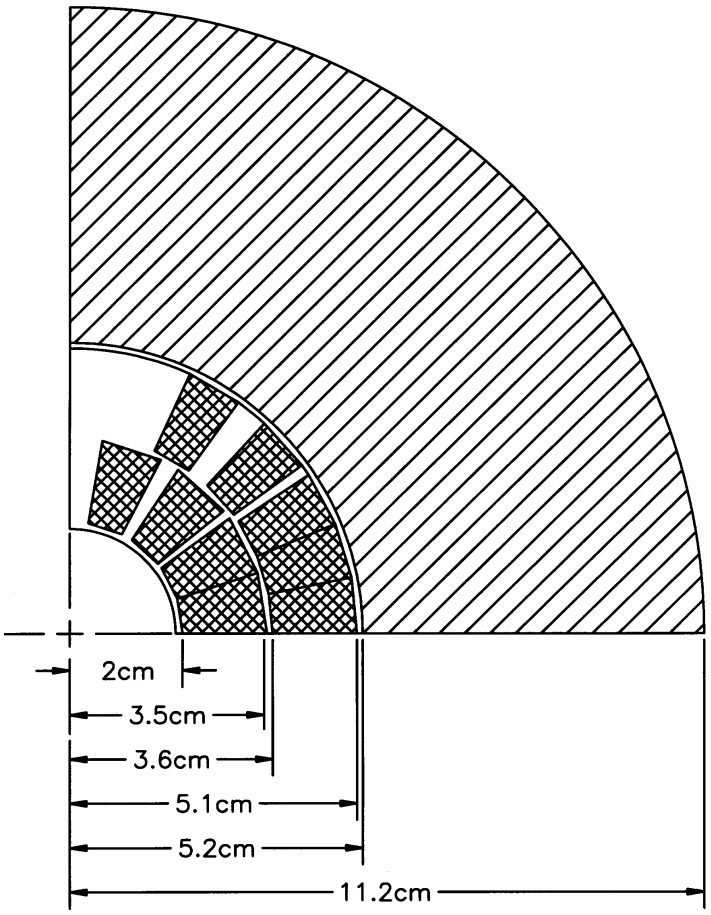


Figure 7.15: Cross sections of pulsed current dipoles for a collider dipole (6 T).

to maintain the current in the magnet. The average power in each magnet at 2 Hz is 452 kW. For the entire ring, with 144 10 m long magnets, the average power at 2 Hz is 39.4 MW. Tb.7.5 includes the parameters of such a design. Note that a pulsed storage ring is not required; the rapid-cycling accelerator could instead fill a fixed-field superconducting ring.

Rapid-cycling dipoles and hybrid rapid-cycling/high-field

Rapid-cycling iron dominated magnets are an established technology but are limited to 60 Hz repetition rates and 2T fields, and both of these parameters are somewhat lower than those desired for μ acceleration. We propose some possible extensions of existing technology to meet this challenge.

An increase in average B-field can be obtained by interlacing fixed-high-field dipoles with ramping magnets, and the range of magnetic field change can be extended by ramping the dipoles fully from -2T to +2T. The ramp rate must also be increased. Scaling from

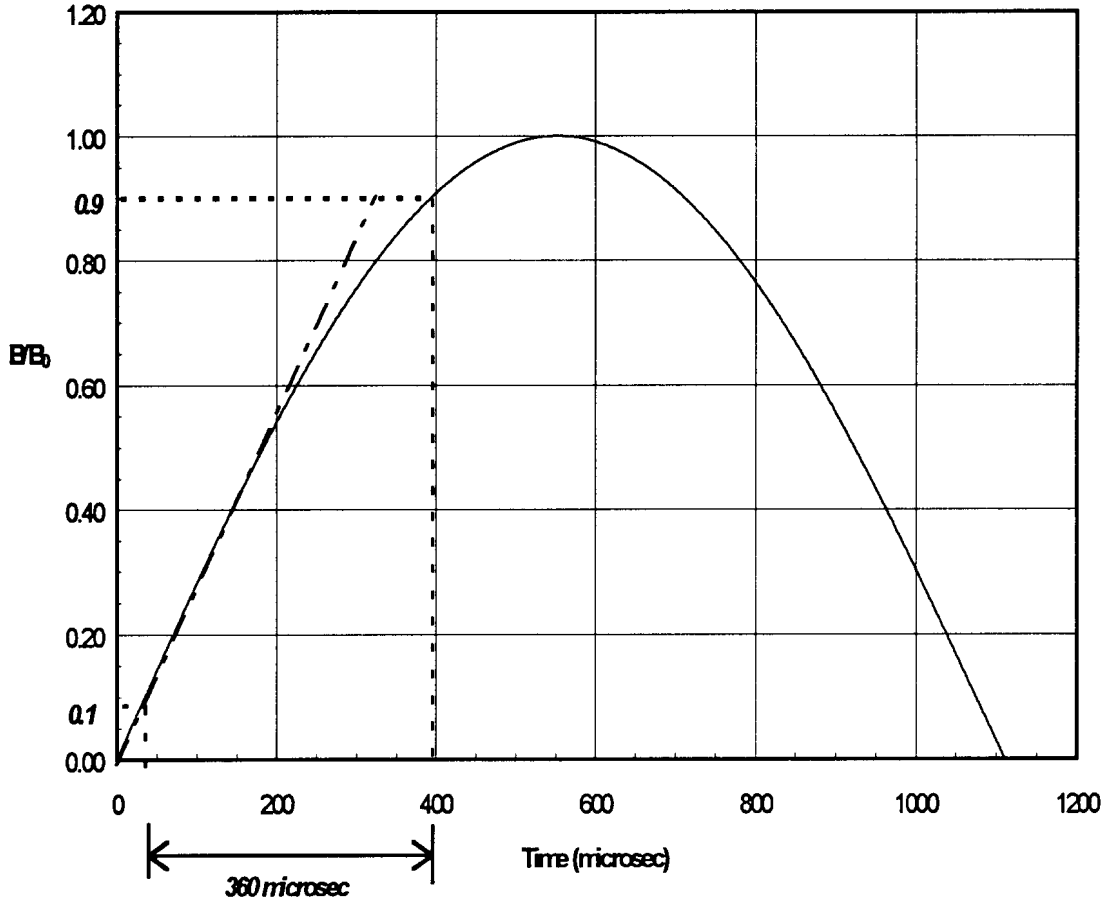


Figure 7.16: Ramp for rapid-cycling pulsed-dipoles for acceleration to 250 GeV.

the KAON booster dipole (50 Hz)[22], design concepts for 250 and 125 Hz μ accelerator dipoles have been developed. In this design concept, the gap within the dipoles is a $8 \times 3\,\mathrm{cm}$ rectangle and a high-B ferrite such as supermendur (2%V 49%Fe 49%Co) is used on the pole tips to get the peak field up to $2T$. Eddy currents are reduced by the use of 0.1 mm thick laminations (instead of 0.47 mm), and hysteresis losses are reduced by using grain-oriented silicon steel. A 6-turn $4 \times 10\,\mathrm{cm}$ Copper conductor (with 0.5 mm square conductor wires) is used, with stainless steel cooling tubes. Power requirements are formidable in these rapid-cycling magnets; a $6\,m$ dipole requires 35/60 kW for 125/250 Hz, respectively.

These rapid-cycling magnets have been combined with $8\,T$ high-field superconducting magnets, in configurations similar to those of Fig.7.13, to form the basis for the rapid-cycling μ accelerator design presented in Tb.7.5, and shown in Fig.7.17. That design consists of two hybrid recirculating linacs, which accelerate the beam from 105 to 730 to 2000 GeV using 250 and 125 Hz dipoles respectively. With an overall dipole occupancy of 70% (split between 2T

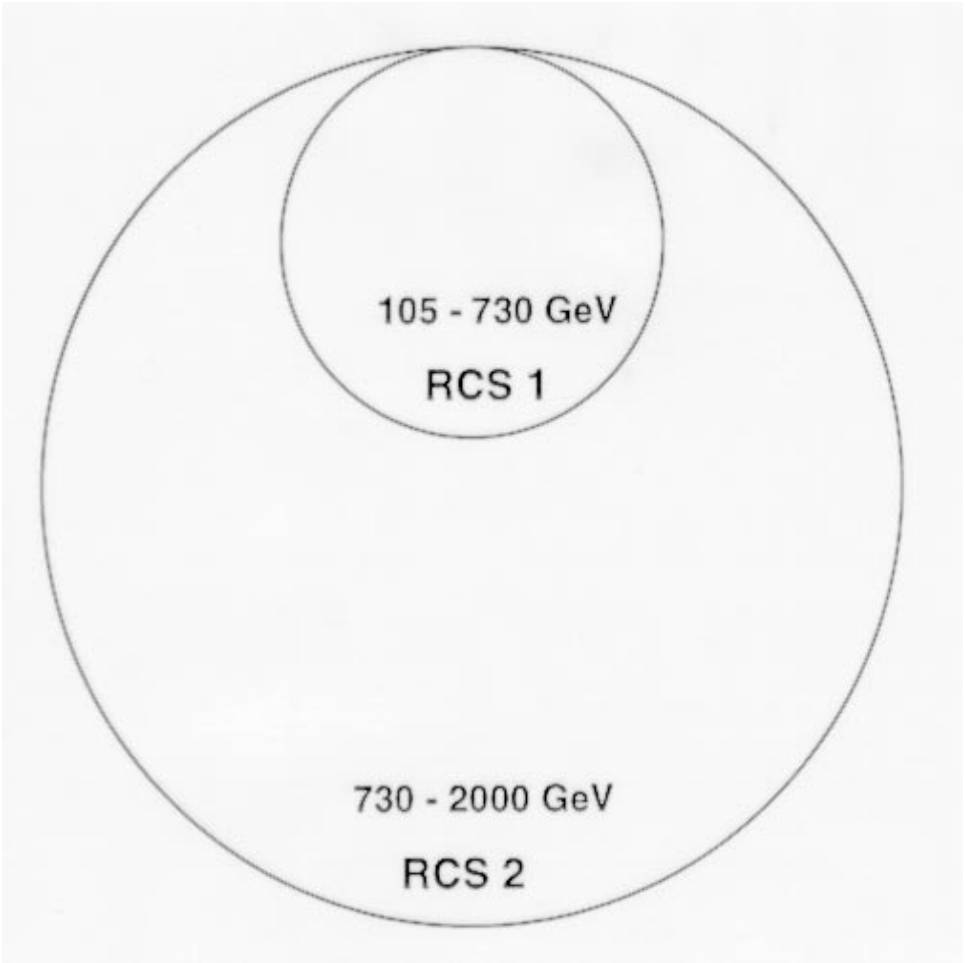


Figure 7.17: Sequence of two rapid-cycling synchrotrons for acceleration from 100 GeV to 2 TeV.

RC and 8T SC units), we obtain 1000 m and 2300 m mean radii for the two rings. (Focusing could be obtained by rapid-cycling (non-hybrid) quadrupoles, which would cycle from low-field to 2T pole-tip fields.) Each accelerator reaches full energy in 100 turns, requiring 6.5 and 15.3 GV rf, respectively. Decay during acceleration is less than 20% per ring. Thus the rf structure requirements are dramatically less than for RLA systems, and only one transport path per ring is required. These advantages are balanced by the very large power requirements of the continuously rapid-cycling magnets (almost 100 MW). This power could be reduced by an order of magnitude, if the power could be gated to pulses matched to the μ collider cycle time (30 Hz), and could be further reduced if the cycle rate was reduced, as was discussed for the pulsed-conductor magnets of the previous section.

Table 7.6: Parameters for a 2 TeV Rapid-Cycling Hybrid Accelerator. The accelerator consists of two rapid-cycling synchrotrons; one takes the beam from 105 to 730 GeV, and a second that takes the beam up to 2 TeV.

Parameter	Symbol	RCS 1	RCS 2
Radius	R	1000	2300 m
Initial Energy	E_i	105	730 GeV
Final Energy	E_f	730	2000 GeV
Ramp frequency	f_F	250	125 Hz
Acceleration orbits	N_t	96	83
rf Voltage	V_{rf}	6.5	15.3 GV
Dipole B field (ave.)	B	$0.5-3.5$	$1.5-4.1$ T
Dipole composition	(2T RC/8T SC)	75%/25%	65%/35%
RCS Dipole power (cw)	P_{cw}	35 MW	32 MW
RCS Dipole power(30 Hz)	$P_{\rm pulsed}$	4.5 MW	8 MW
Beam Decay Loss	η_L	26%	14%

7.8.2 Rotating Dipoles/Magnets

The disadvantage of high power rapid-cycling magnets can be reduced, if instead of cycling currents, we rotate the dipoles to obtain a rapid-cycling field. Two adjacent permanent-magnet 2T dipoles counter- rotating at 250 Hz (15,000 rpm) produce a sum field which cycles from -2T to +2 T . Mechanical rotations of this rate, or even faster, are readily obtained using current technology.

The dipoles would be constructed of a radiation-hard material, such as samarium-cobalt, placed in a Halbach *magic-ring* configuration.[23] The interior field is uniform and can exceed the remanent field of the rare earth material: $B = B_r \ln(OD/ID)$, where B_r is the remanent field (1.15 T for samarium cobalt, but 1.3 T for Alnico). 2 T is obtained for Sm_2Co_{17} with outer diameter (OD) of 9 cm and inner diameter (ID) of 1.5 cm. Counter-rotating magic-ring dipoles would have an average field which would cycle between - 2 T and +2 T. These counter-rotating dipoles could be used in place of the rapid-cycling magnets in the rapid-cycling synchrotrons of Tb.7.6, and would greatly reduce power requirements. Substantial problems may exist in this untested configuration. Radiation damage of permanent magnet materials is a concern. The rotating magnets have localized vertical bends in the cycle which

may cause emittance dilution, coupling or even instability. We have not included focusing elements, but counter-rotating permanent magnet quadrupoles could be used. To avoid air friction the magnets will rotate within a vacuum; which would then also be the beam vacuum system. One is concerned whether the rotation might demagnetize the material. Considerable study and tests are needed for this unconventional possibility.

7.9 Comments and Conclusions

We have presented a candidate scenario for a high-energy $\mu^+ - \mu^-$ accelerator. That scenario is based on a multi-stage RLA approach, and it includes a proof-of-principle calculation of the design concept. Much further optimization and design development is needed, and we can identify some of these development goals:

- The bunch-compression and acceleration scenario should be further optimized and simulated.
- Complete lattices are needed, with designs for the transport arcs, including beam separation and recombination.
- An accurate cost algorithm for rf and beam transport components is needed.
- Matching from the output of the μ cooling system to the initial acceleration should be more clearly defined with reoptimization of final cooling and initial acceleration. A consistent injector linac should be specified.
- rf acceleration development is needed, both in the low-frequency rf systems needed in the first stages and in the high- frequency SRF needed in the high-energy accelerators.
- Variations using rapid-cycling should be considered and studied. Prototype rapid-cycling magnet components could be designed and built.

Bibliography

[1] R. Palmer et al.,*Muon Colliders*, Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop, Ed. J. C. Gallardo, AIP Conference Proceedings 372 (1996).

[2] D. Neuffer, *Acceleration to Collisions for the $\mu^+\mu^-$ Collider*, ibid.

[3] D. Neuffer, *Analysis and Simulations of Longitudinal Motion in μ-Recirculating Linacs*, to be published in Nucl. Inst.and Meth. A, 1996.

[4] H. Kirk, presentation to the 9th Advanced ICFA beam Dynamics Workshop 1996, unpublished.

[5] Z. Qian, unpublished (1996).

[6] H. Lengeler, in Proc. 5th General Accelerator Physics Course, S. Turner, ed. CERN 94-01, p. 791 (1994).

[7] G. Cavallari et al., in Proc. 5th Workshop on rf Superconductivity, DESY M92-01, p. 23 (1992).

[8] C. Benvenuti, in Proc. 5th Workshop on rf Superconductivity, DESY M92-01, p. 189 (1992).

[9] G. Spalek and H. A. Thiessen, Proc. 5th Workshop on rf Superconductivity, DESY M92-01, p. 95 (1992).

[10] H. A. Thiessen, private communication (1996).

[11] Q-S Shu for the TESLA collaboration, presented at the CEC/ICMC Conf, Columbus, July 1995.

[12] TESLA-Collaboration - D. Edwards, ed., TESLA Test Facility Linac - Design Report, TESLA 95-01 (1995).

[13] Q.-S. Shu et al., *Experimental Investigation of Quenches in Superfluid He of TESLA 9-Cell Superconducting Cavities*, Proc. 7th RF Superconductivity Workshop, Saclay, France, Oct 16-20, 1995.

[14] A. Mosnier and O. Napoly, Proc. XV Int. Conf. on High Energy Accelerators, J. Rossbach, ed., p. 963 (1992)

[15] Laser Processing Consortium, Free Electron Lasers for Industry Volume 2: UV Demo Conceptual Design, CEBAF, May 1995.

[16] S. Y. Lee, K. Y. Ng, and D. Trbojevic, Phys. Rev. **E 48**, 3040 (1993).

[17] CEBAF Design Report, CEBAF, Newport News VA (1986), unpublished.

[18] S. Kahn, G. Morgan, and E. Willen, *Recirculating arc dipole for the* $2+2\,\mathrm{TeV}$ $\mu^+\mu^-$ *Collider*, Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop, Ed. J. C. Gallardo, AIP Conference Proceedings 372 (1996).

[19] D. Summers, to be published in Proceedings of the Symposium on Physical Potential and Development of Muon-Muon Colliders, San Francisco, CA, Dec. 1995.

[20] E. Willen, *Pulsed dipole Magnets for the Muon Collider*, Magnet Note 555-31, BNL, May 1996.

[21] R. Winje and K. Bourkland, IEEE Trans. Nuc. Sci., NS-22, No.3 (1975), 1242

[22] TRIUMF-KAON FACTORY STUDY Accelerator Design Report, unpublished (1990).

[23] K. Halbach, *Design of Permanent Multipole Magnets with Oriented Rare Earth Cobalt Material*, Nucl. Instr. Meth. 169 (1980) 1.

Contributors

- D. Neuffer, (FermiLab) Editor
- C. Ankenbrandt, (FermiLab)
- A. McInturff, (FermiLab)
- G. Morgan, (BNL)
- Q-S Shu, (CEBAF)
- S. Simrock, (CEBAF)
- D. Summers, (Univ. of Mississippi)

List of Figures

List of Tables

Chapter 8

MUON COLLIDER RING

Contents

8.1 Introduction

The collider ring of the Muon Complex allows for (about) 1000 collisions per bunch, rather than the single collision that is possible in a linear collider geometry. If the transverse beam size at the collision point in a muon collider were the same as that in an electron-positron linear collider, there would be a full increase in luminosity of order 1000. This is not the case. The muon bunch is cooled as much as possible, but still has an emittance that is significantly larger than the extremely low emittances required in an electron-positron linear collider. The luminosity scales as $L = fN^2/((\epsilon_x\beta_x^*)(\epsilon_y\beta_y^*))^{1/2}$, where $\epsilon_{x(y)}$ is the beam emittance in the $x(y)$ phase plane, $\beta_{x(y)}^*$ the corresponding beta-function at the interaction point (which is limited, by the hour-glass effect, to be no less than the bunch length), and f the collision frequency.

While the NLC electron-positron linear collider and the muon collider are quite different, in particular this muon collider feasibility study is for a 4 TeV (center of mass) machine and the NLC energy range is .5-1.5 TeV, it is instructive to compare the parameters that are required for achieving the design luminosity. For typical NLC parameters: $f = 18,000 = 100 \times 180$ (bunches per pulse $\times$ pulses per second), $N = 10^{10}$, $\epsilon_{nx} = 5 \times 10^{-6}$m, $\beta_x^* = 10^{-2}$m, $\epsilon_{ny} = 5 \times 10^{-8}$m, $\beta_y^* = 10^{-4}$m for a luminosity of $7 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$ at $250\,\text{GeV} \times 250\,\text{GeV}$.

In contrast, the muon collider has $f = 30,000Hz = 2 \times 1000 \times 15$ (bunch pairs/pulse $\times$ collisions/bunch-pair $\times$ pulses/second), $N = 2\times10^{12}$, a round beam with $\epsilon_n = 5\times10^{-5}$mrad, $\beta^* = 3\times10^{-3}$ m, for a luminosity of $10^{35}\text{cm}^{-2}\text{s}^{-1}$ at 2 TeV $\times$ 2 TeV. Thus, the muon collider achieves its luminosity primarily with an increased number of particles (and from an increased number of collisions per bunch-pair). The muon collider has a much larger emittance and beam size at the IP ($\sigma_x\sigma_y|_{NLC} \approx 10^{-15}\text{m}^2$, while $\sigma_x\sigma_y|_{muon} \approx 10^{-12}\text{m}^2$). The large emittance may relax some tolerances on component alignment compared to electron-positron linear colliders, but these bunch parameters, coupled with the need to store ~ 1000 turns, present many difficulties.

The ring design is challenging, with its high current, low beta, and isochronicity requirements. The low beta and high current are required to obtain the luminosity, and the isochronicity is required to maintain the short bunch length without excessive rf. Luminosity cannot be further increased by reducing the muon bunch length, for the bunch length is determined by longitudinal cooling, which is, already, as much as can be done. For the very high charge one must be concerned with collective instabilities and wakefields.

The ring brings many technical complications to the project (as well as cost, not addressed herein), but we we have an ideas on how each can be handled, and advocate further investigation of a collider ring that achieves the very high luminosity requisite for operation at high energy.

This chapter examines the design and analysis of various systems envisioned for the muon collider main ring. This work includes magnet designs, lattice design, vacuum requirements, analysis of collective instabilities, and beam-beam interaction. First lattice studies are presented (Section 8.2), and then the superconducting magnets, both dipoles and quadrupoles (Section 8.3). In Section 8.4 we consider the ring rf system. Section 8.5 is devoted to the vacuum chamber which require special care as it must absorb the energy of decay electrons and their synchrotron radiation. The classical beam-beam interaction, incoherent and coherent, are covered in the next section (Section 8.6). Incoherent pair production is covered in Sec. 8.7. Finally, in Section 8.8, collective effects are discussed. The major parameters of the muon collider are presented in Table 8.1.

8.2 Lattice

The lattice for a 2-TeV on 2-TeV muon collider must satisfy three major design constraints. The first and most difficult of these is provision of an Interaction Region (IR) with an extremely low β^* (~ 3 mm) consistent with an acceptable dynamic aperture. Second, the ring must exhibit a high degree of isochronicity in order to preserve short 3 mm long bunches with a modest rf system. Lastly, there must be small corrected chromaticity, so that the momentum-dependent tune spread of the beam fits between resonances. Technically, considerable shielding must be incorporated into the design to protect the superconducting magnets from the high muon-decay backgrounds. The following sections describe a preliminary lattice, Ref. [1], which is intended to meet the above requirements.

Table 8.1: High energy-high luminosity $\mu^+\,\mu^-$ collider

Maximum c-m Energy [TeV]	4
Luminosity $\mathcal{L}[10^{35}\mathrm{cm}^{-2}\mathrm{s}^{-1}]$	1.0
Circumference [km]	8.08
Time Between Collisions [μs]	12
Energy Spread σ_E[units 10^{-3}]	2
Pulse length σ_z[mm]	3
Free space at the IP [m]	6.25
Luminosity life time [No.turns]	900
Horizontal betatron tune, ν_x	55.79
Vertical betatron tune, ν_y	38.82
rms emittance, $\epsilon_{x,y}$ [μm-rad]	0.0026
rms normalized emittance, $\gamma\epsilon_{x,y}$ [μm-rad]	50.0
Beta-function values at IP, $\beta^*_{x,y}$ [mm]	3
rms Beam size at IP [μm]	2.8
Quadrupole pole fields near IP [T]	6.0
Peak beta-function, $\beta_{x\mathrm{max}}$ [km]	284
Peak beta-function, $\beta_{y\mathrm{max}}$ [km]	373
Magnet Aperture closest to IP [cm]	12
Beam-Beam tune shift per crossing	0.05
Repetition Rate [Hz]	15
rf frequency [GHz]	1.3
rf voltage [MeV]	130
Particles per Bunch [units 10^{12}]	2
No. of Bunches of each sign	2
Peak current $\mathcal{I} = eNc/\sqrt{2\pi}\sigma_z$ [kA]	12.8
Average current $\mathcal{I} = eNc/$Circum [A]	0.032

8.2.1 Overview

To obtain the design luminosity, considering the projected muon production rate, requires very low β-function values at the IP, $\beta^* = 3\,\mathrm{mm}$. Additionally, the *hour-glass* effect requires bunch lengths comparable to β^*. Maintaining such short bunches with moderate rf voltages requires a momentum compaction factor of 10^{-6} or less. (Just to preserve the bunch length for one turn requires $|\alpha| \ll \sigma_z/C\delta \simeq 10^{-4}$, where C is the circumference.) These are the major challenges of the muon collider lattice design. Studies addressing these issues have

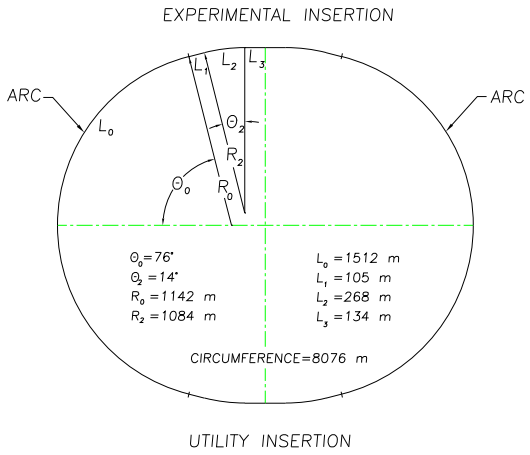


Figure 8.1: The complete collider ring layout.

been recently published [2].

The very small β^* requirement is particularly difficult to achieve because substantial space must be reserved for the detector about the IP and also because the superconducting magnets must be shielded from the decay products of the muon beam. In the initial design, superconducting quadrupoles nearest the IP had to accommodate a 6 cm thick tungsten liner in order to dissipate heat generated by the beam. Recent work has been successful in reducing the required shielding to 2 cm using combinations of sweep dipoles and collimators. With the thicker liner, quadrupole gradients were reduced by about a factor of 4 near the IP. However, with the thinner liner, the strength reduction will be factor of 2, since the beam size and liner thickness will be comparable. Additionally, a magnet-free length of ± 6.5 m must be reserved about the IP to provide the ± 150 mr acceptance angle needed for the detector. As a result, peak beta-function values as high as several hundred kilometers cannot be avoided; these high values produce extremely large chromaticities (several thousand before correction).

One preliminary design of an entire collider ring has been developed using the experimental insertion and arc modules described in Ref.[2]-[4] plus a utility insertion. A schematic layout of this ring design is shown in Fig. 8.1. A second collider ring design, has been developed by K. Oide[4]. This design incorporates a more sophisticated nonlinear correction scheme in the IR. Unfortunately, because of time constraints, it was impossible to include an equal discussion of the two designs here. Therefore, we will proceed with a detailed presentation of the first design, to be followed by a section comparing the two designs, and, finally, a discussion of recent progress.

8.2.2 Components of the Lattice

The ring has an oval shape, with reflection symmetry about two perpendicular axes, see Fig. 8.1. The lattice has two nearly circular 152° arcs joined by two insertions, each containing two 14° bending sections. These are called the experimental and utility insertions.

The two arcs are identical; each contains 22 periods and one dispersion suppressor at each end. The insertions are geometrically identical; each is symmetric about its center. Each half insertion has three parts: two straight sections separated by a bending section. The bending sections are identical in the experimental and utility insertions, except for the sextupole strengths. In contrast, the straight parts have different quadrupole lengths and gradients in the two insertions. Thus, as a focusing structure the ring has one superperiod with reflection symmetry about the line joining the centers of the two insertions.

Arc module In order to have very short 3 mm bunches in the 2 TeV muon collider, the storage ring must be quasi-isochronous, which requires that the momentum compaction α be very close to zero[9], where α is defined in terms of offsets of the momentum p and equilibrium orbit circumference C by

$$\alpha(p) = \frac{p}{C}\frac{dC}{dp}\,. \tag{8.1}$$

The lattice must be designed so that over the momentum range, the momentum compaction remains small.

The particle motion in longitudinal phase space depends on its arrival time at the RF cavities, which varies as a function of circumference and velocity v. To first order the time difference with respect to the reference particle is:

$$\frac{\Delta T}{T_0} = \eta\frac{\Delta p}{p_0} = (\alpha_0 - \frac{1}{\gamma^2})\frac{\Delta p}{p_0}, \tag{8.2}$$

where T_0 and α_0 are the revolution period and momentum compaction of the reference particle; ΔT and Δp are the time and momentum deviations, respectively, of the off-momentum particle relative to the synchronous particle with momentum p_0; η is the *phase slip* factor; γ is the Lorentz relativistic factor, and $\alpha_0 = \alpha(p_0)$. The transition energy γ_t is defined by $\alpha = 1/\gamma_t^2$.

In an isochronous ring $\eta = 0$, so to first order the arrival time is independent of the momentum; i.e., $\gamma_t = \gamma$. For 2 TeV muons $\gamma \approx 2 \times 10^4$, so $\alpha \approx 2.5 \times 10^{-9}$. In a regular FODO lattice, α is much larger. To bring the first order value of α to zero requires that the $\langle D/\rho \rangle$ through all of the dipoles be equal to zero, where D is the dispersion and ρ the radius of curvature.

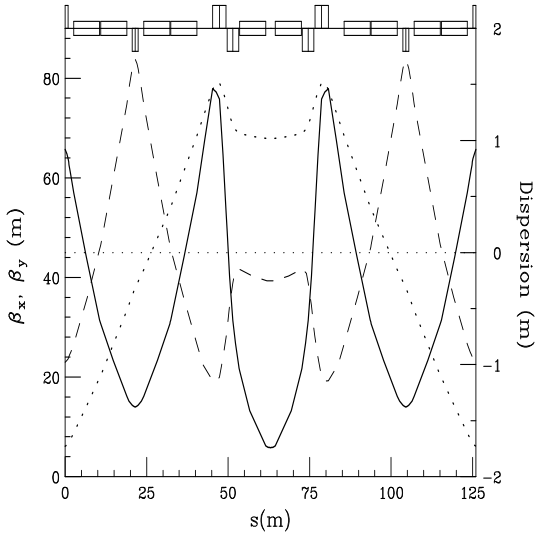


Figure 8.2: Betatron (β_x solid-line; β_y dash-line) and dispersion (dot-line) functions of an arc-module.

In a FODO lattice α is positive. This muon collider ring design has bending regions in the insertions with a FODO structure whose contributions to α are positive, so the contributions of the arcs must be negative with nearly the same magnitude as those of the insertions. For the present design, the value needed for each arc is $\alpha_{\rm arc} = -1.15 \times 10^{-4}$.

This value of $\alpha_{\rm arc}$ can be obtained by building an arc whose periods are FMC modules. An FMC module[4] is a structure composed of two FODO cells separated by a matching insertion which transforms $(\beta_x, \alpha_x, \beta_y, \alpha_y, D, D')$ to $(\beta_x, -\alpha_x, \beta_y, -\alpha_y, D, -D'.)$

The contribution to α of the module can be adjusted by choosing the appropriate value of D with $D' = 0$ at the end of the module. For the module design used here (see Fig.8.2), the matching insertion contains two quadrupole doublets and two dipoles. The two quadrupole gradients and drift lengths are adjusted to bring α_x, α_y and D' to zero at the center of the module. The number of modules and the bending angles of the dipoles are chosen to give the entire arc the bending angle of 152° needed to close the ring.

The arc modules also contain sextupoles; there are two families adjusted to bring the chromaticities of the arc to zero.

Dispersion suppressor A dispersion suppressor module is located at each end of the arc. The purpose of these modules is to bring the dispersion and its slope to zero values in the adjacent insertions.

The suppressor on the downstream end just before an insertion is shown in Fig.8.3; the upstream suppressor is obtained by reflection. This suppressor module is identical to a

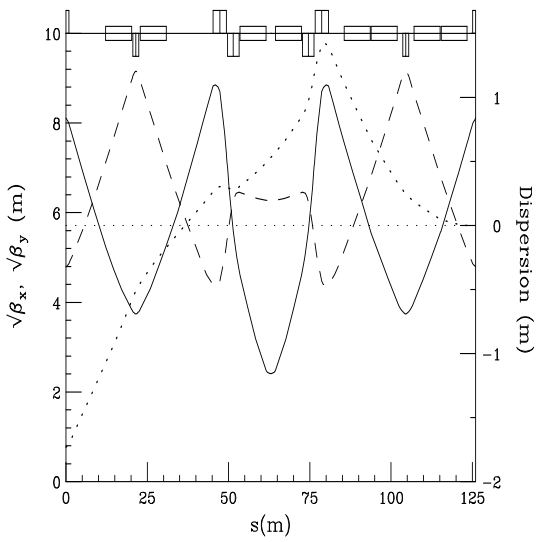


Figure 8.3: Betatron (β_x solid-line; β_y dash-line) and dispersion (dot-line) functions of a dispersion suppressor module.

regular module except that the first four dipoles have been replaced by two dipoles with normal length and different field values. The missing dipoles have been replaced by drift spaces so that the quadrupoles and sextupoles are not changed.

Experimental insertion The design of an insertion with an extremely low-beta interaction region for a muon collider[5] presents a challenge similar to that encountered for the Next Linear Collider (NLC)[6]. The design used here for each half of the symmetric low-beta insertion follows the prescription proposed by Brown[7]; it consists of two telescopes with a chromatic correction section between. Fig.8.4 shows the left half of the insertion, starting at the end of the arc dispersion suppressor and ending at the IP.

The first telescope, called the Matching Telescope (MT), on the left of the figure, brings the beta functions from the arc to a focus of about 3 cm. To the right of the MT lies the Chromatic Correction Section (CCS), which contains two pairs of non-interleaved sextupoles. One pair, situated at positions of maximum β_x and large dispersion D, corrects horizontal chromaticity; the other pair, at maximum β_y positions, corrects vertical chromaticity. The horizontal-correcting pair is farthest from the IP, and the vertical-correcting pair is closest. The sextupoles of each pair are separated by betatron-phase intervals of $\phi = \pi$, and they are located at positions where the phase interval from the IP is an odd multiple of $\pi/2$. To the right of the CCS, the Final Focus Telescope (FFT) transports the beta functions from a focus of a few centimeters to a 3 mm focus at the IP.

The low beta-function values at the IP are obtained with four strong quadrupoles in

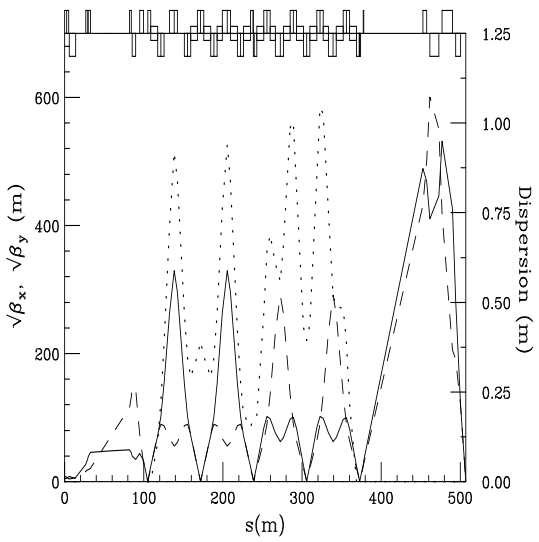


Figure 8.4: Experimental insert (half) with extremely small beta function at the IP.

the FFT with high beta values; these generate large chromaticities, which are corrected locally with the two sextupole pairs in the CCS. This sextupole arrangement cancels the second-order geometric aberrations of the sextupoles, which reduces the second order tune shift by several orders of magnitude. The momentum bandwidth of the system is limited by third-order aberrations and residual second-order amplitude-dependent tune shifts. These aberrations arise from: a) small phase errors between the sextupoles and the final quadruplet; b) finite length of the sextupoles.

The residual chromaticities could be reduced with additional sextupoles at locations with nonzero dispersion, as suggested by Brinkmann[8]. Finally, a system of octupoles could be designed to correct third-order aberrations. Overall, it is believed possible to construct a system with a bandwidth of $\approx 1\,\%$.

The most complex part of the insertion is the CCS. A somewhat oversimplified description follows. The CCS consists of eight FODO cells, each with $\pi/2$ phase advances. The first four cells from the left begin at the center of a QF quadrupole and contain the two horizontal S_x sextupoles, which are next to QFs; the next four cells begin at the center of a QD quadrupole and contain the vertical S_y sextupoles, which are next to QDs. The low-beta focus at the beginning of the CCS repeats itself every two cells and produces the high beta values needed in the sextupoles. The dipoles are placed in a way to cancel the dispersion and its slope at both ends of the CCS and to produce dispersion maxima in the sextupoles.

The strengths of the sextupoles S_x and S_y are adjusted to produce zero first-order chromaticity values for the insertion, while trim quadrupoles are used to minimize the second order chromaticity ($\partial^2 Q/\partial\delta^2$). The complete insertion has very small residual chromaticity,

and is nearly transparent when attached to the arc lattice.

The total length of the half-insertion is 507 m. It contains 44 quadrupoles, 14 dipoles and 4 sextupoles.

Utility insertion The utility insertion closely resembles the experimental insertion, except that the low-beta foci are relaxed in order to lower the beta-function maxima by a factor of about 1000.

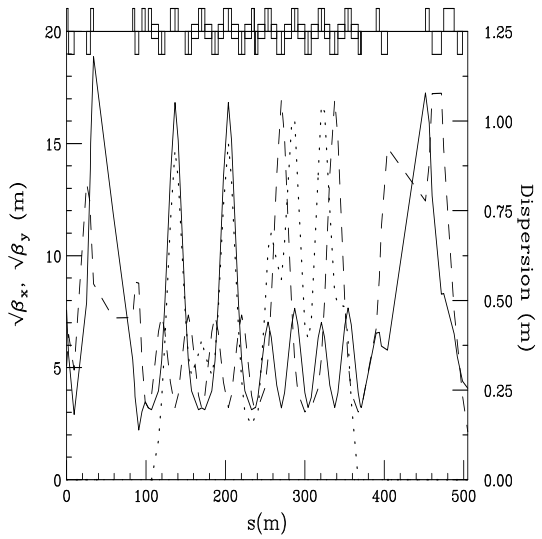


Figure 8.5: Utility insertion (half)

The CCS section for the utility insertion is the same as the one in the experimental insertion, except that the sextupoles are adjusted to cancel the chromaticities of the utility insertion (which are much smaller than those of the experimental insertion). Further changes will probably be needed to accommodate systems for injection, RF, and scraping.

8.2.3 Performance

The variations of the fractional part of the tunes $Q_{x,y}$ as functions of δ are shown in Fig.8.6. Q_y is essentially flat over a bandwidth of $\pm 0.4\%$; Q_x has obvious non-linear components, although the variation of tune, peak to peak is less than 0.04 within a bandwidth of $-0.15\,\%$ to $0.3\,\%$. Likewise, the β^* variation (Fig.8.7) is negligible within a bandwidth of $\pm 0.3\,\%$. The remaining figures show the chromaticity (Fig.8.8), the momentum compaction α (Fig.8.9) and the amplitude dependent tunes $dQ/d\epsilon$ (Fig. 8.10), versus δ. ¿From the results shown in

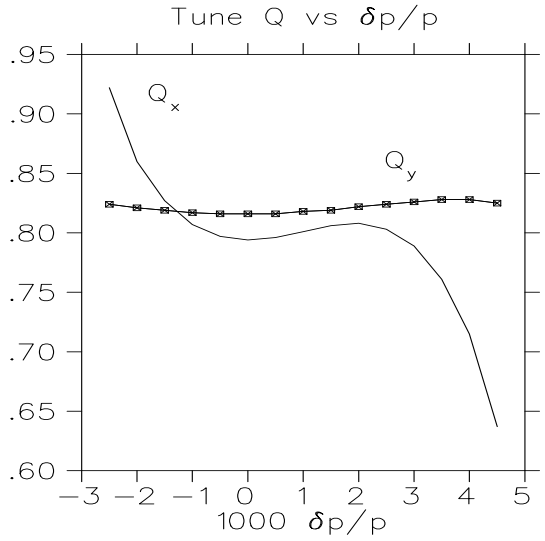


Figure 8.6: Fractional tunes $Q_{x,y}$ vs $\frac{\Delta p}{p}$

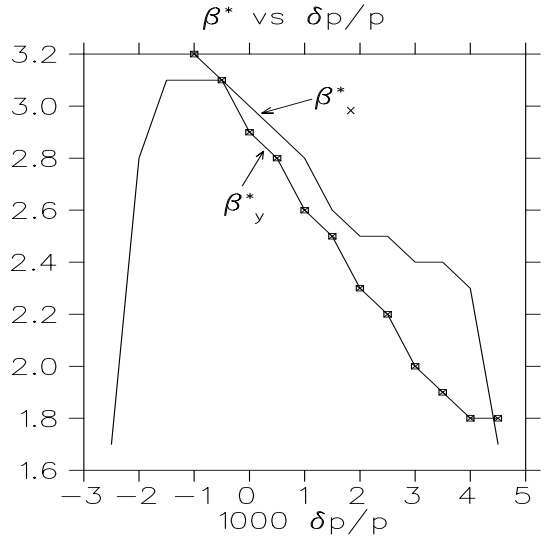


Figure 8.7: Beta function β^* vs $\frac{\Delta p}{p}$

the above figures, it appears that the momentum aperture of the CD has a range $\delta = \Delta p/p$ of 0.007.

The variation of the momentum compaction factor α versus δ, shown in Figure 8.9, is too large to preserve short bunches with a moderate rf voltage. Methods for coping with this problem are discussed in the following section.

Tracking runs [11] using TEAPOT indicate that the dynamic aperture (about one sigma) is too small by roughly a factor of four. The lattice is presently being tuned to increase its dynamic aperture and, along with recent improvements, the aperture has been increased to

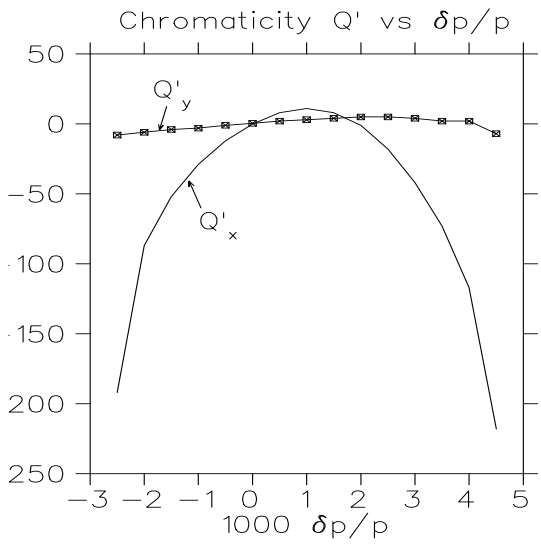


Figure 8.8: Chromaticity vs $\frac{\Delta p}{p}$

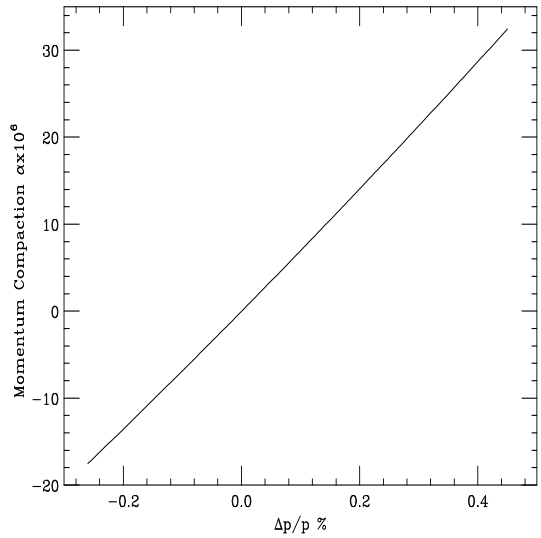


Figure 8.9: Momentum compaction α vs $\frac{\Delta p}{p}$

about five sigma without changing its basic structure. The improvements and adjustments to the collider lattice are outlined in Section 8.2.5.

8.2.4 Control of the Momentum Compaction

The collider ring lattice has been adjusted to be approximately isochronous for the reference particle. That is, the lattice has been designed so that the momentum-compaction factor,

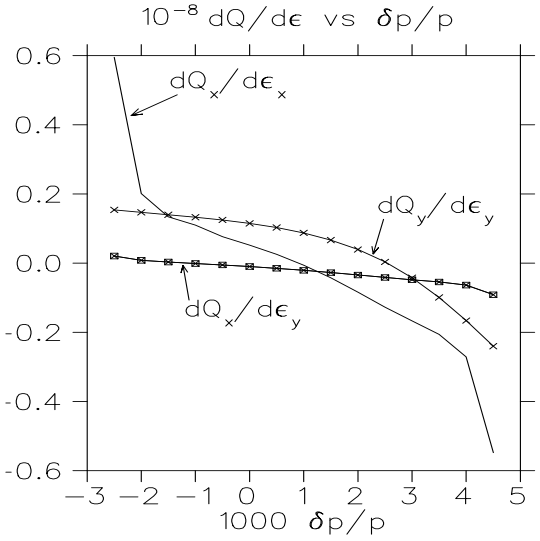


Figure 8.10: Amplitude dependent tune shift $\frac{dQ}{d\epsilon}$ vs $\frac{\Delta p}{p}$

$\alpha(p)$, defined by

$$\alpha(p) = \frac{p}{C}\frac{dC}{dp}\,. \tag{8.3}$$

is approximately zero. In practice, in order to maintain a 3 mm bunch and a modest rf, $\alpha(p_0)$ must be about 10^{-6}.

However, over the desired momentum aperture, $\alpha(p)$ varies, as shown in Figure 8.8, so that over the momentum range of $\pm.004$, $\alpha(p)$ exceeds 10^{-5}. If we expand $\alpha(p)$ in powers of $\delta = p/p_0 - 1$:

$$\alpha(p) = \alpha_1 + \alpha_2\delta + \alpha_3\delta^2 + \mathcal{O}(\delta^3) \tag{8.4}$$

we see from Figure 8.9 that $\alpha_1 = 0$, and $\alpha_2 = 0.006$. This value of α_2 leads to an unacceptable longitudinal head-tail instability. Since α_2 has a contribution from each sextupole of $-2SD^3$, where S is the strength and D is the dispersion in the sextupole, it is possible to correct α_2 with one or more sextupole families.

Initially, horizontal and vertical chromaticities, but not the α_2, of the arcs, experimental insertion and utility insertion were canceled using three independent pairs of sextupole families. Alternatively, the chromaticities of the ring, most of which arise from the experimental insertion, can be canceled by using only the insertion sextupoles. This frees the arc sextupoles to control α_2. Specifically, by inserting a horizontal sextupole next to each of the central F quadrupoles in the arc modules, the α_2 term can be eliminated.

It is possible to control both the linear α_2 and the quadratic α_3 coefficients using two sextupoles pairs in each module, one pair next to the two center QFs and the other pair next to the QF at the module ends–see Figure 8.2. By choosing appropriate strengths for

the different sextupole families, one can make $\alpha_2 \simeq \alpha_3 \simeq 0$, which reduces the variation of α over a range in δ of $\pm.004$ to only about 10^{-7} (for the arcs). Unfortunately, this method halved the total stable momentum range of the ring.

Another way to control α_2 without generating a contribution to α_3 is to pair sextupoles separated by phase intervals of π in the arcs. Thus we conclude that control over both α_2 and α_3 can be achieved to a precision of 10^{-7}; however, this degree of correction may not be necessary (especially for α_3).

In summary, the isochronicity of the ring can be controlled precisely. The momentum-compaction coefficients chosen for the ring will be based on rf bucket and collective instability calculations.

8.2.5 Collider lattice comparisons

As mentioned previously, a different design for the muon collider has been developed by K. Oide[2]. For discussion purposes, the detailed design presented in this document will be referred to as the CD (collaboration design) and Oide's design as the OD. Since tracking results show the OD to have larger dynamic aperture than the CD, the following discussion of the salient differences between the two lattices may be useful.

For the arcs, the CD uses FMC modules, while the OD adopted the KEK B factory modules. Since the dynamic aperture problems arise from the IR, not from the arcs, this difference is probably unimportant.

Both the CD and the OD are similar to linear collider IRs, but there are important differences between these two designs. In the CD the maximum horizontal and vertical beta-function values, and consequently the chromaticities, are equalized. On the other hand, in the OD, the horizontal β_{xmax} is less than β_{ymax} by more than an order of magnitude. As can been seen from the table, the OD ring has a substantially larger vertical high-beta value (900 vs. 350 km) than the CD, and its apertures are correspondingly larger. The vertical chromaticity comparison is less dramatic, since the increased β_{ymax} is offset by the shortened length of the high-beta quadrupoles.

The unequal β_{max} values in the low-beta quadrupoles of the OD compensates for the chromatic correction scheme which favors the vertical plane, due to the fact that the vertical sextupoles are closer to the IP than the horizontal ones. (Increased nonlinearities in the x plane are caused by the intervening y-plane sextupoles.) Thus, for a 2-TeV muon collider, it appears that an asymmetry in chromaticity may be important for chromatic correction.

The strength of the chromatic correction sextupoles is another contributing factor to nonlinearites. In the OD, increased dispersion and much higher β functions allowed the

correction sextupole strengths to be reduced by almost an order of magnitude compared with those of the CD.

The last important feature of the OD to be mentioned here is the addition of octupole and decapole components in the IR quadrupole fields; this addition appeared to increase the dynamic aperture dramatically.

Modifications to the present CD currently under investigation include use of unequal maximum $\beta_{x,y}$ values, optimization of the chromatic correction section, use of higher multipole corrections, and addition of permanent-magnet quadrupoles between the IP and the first superconducting quadrupole.

8.2.6 Summary

Studies have been underway to improve the experimental insertion. The approach of placing permanent magnets within 2 m of the IP to reduce β_{max} and chromaticity has been discarded. Exposure to the high radiation environment in a muon collider ring raised concern over the magnetic field lifetime of a permanent magnet. Instead, a "Bitter" quadrupole is being considered as an alternative magnet to place near the IP. In a preliminary design by B. Weggel during the Snowmass conference, a small "Bitter" quadrupole was introduced with a pole-tip field of just under 4 T. Shielding modifications included placing tungsten collimators between the superconducting IR quadrupoles to shadow and thus protect them from the high backgrounds. To be effective, the collimators had to be 15 cm long with a 4-sigma aperture (the quadrupoles have a 5-sigma aperture). Additionally, sweep dipoles were placed about a meter upstream of the final-focus quadrupoles to eliminate the muon-decay products generated in the preceding long drift. In the previous IR design, most of the decay products struck the IR quadrupoles, creating the unacceptable heat loads.

The combined effect of adding dipoles and collimators to the IR allowed the protective tungsten liner of the superconducting elements to be reduced from 6 cm to 2 cm. This allowed the effective gradient of the final-focus quadrupoles to be increased. When the Bitter quadrupole was also included and placed 4 m away from the IP, peak β functions decreased by almost a factor of two, and chromaticities by a factor of 3 in the horizontal plane and 2 in the vertical. Higher order aberrations were reduced by about two orders of magnitude.

Initially, the dynamic aperture did not increase as a consequence of the IR improvements. The reason for this proved to be the CCS. Optimization work on the CCS proved to be as important as the improvements made to the IR. To maximize momentum aperture, the peak β functions in the chromatic correction sextupoles were deliberately large, which had the

intended effect of reducing their strength and therefore their contribution to higher-order aberrations. However, the large beta functions in the sextupoles increased significantly their contribution to amplitude-dependent tuneshifts. A better way to decrease sextupole strength (and length) is to increase the dispersion function at their locations. Increasing the dispersion and reducing the peak β functions at the sextupoles was the approach used to minimize their higher-order contributions to the amplitude-dependent tuneshifts (without increasing aberrations significantly). When peak β functions in the CCS were reduced from 100 km to 50 km, tracking showed the on-momentum aperture to be 5 sigma. The full momentum acceptance was .3%. Results were found to be strongly tune-dependent and a phase trombone was introduced into the collider ring to adjust tunes independently and without disturbing the lattice. Presently a 10−km version of the CCS with same final focus structure is being tested.

After the FT and CCS optimization is complete using only sextupoles, the addition of octupoles and perhaps decapoles will be studied to further reduce the amplitude-dependent and aberration terms. Also, in future, it is hoped that the "Bitter" quadrupole, which has a high power consumption, can be removed if high T_c superconductor research indicates that we can employ stronger quadrupole gradients in the final focus.

8.3 Superconducting Magnets

The number of collisions during a storage time is inversely proportional to the ring diameter, since the muon decay time constant is fixed by the muon energy. The dipoles in the ring should therefore have a very high field. A reasonable value is 8.5T, supplied by superconducting magnets. As in other rings, the superconductor must be shielded from heat generated by the beam. In the muon collider there are several heat sources: 1) Muon synchrotron radiation, 2) Muon decay (electrons, positron and their synchrotron radiation; neutrinos pass through the walls and do not deposit energy in the accelerator components), 3) Muons that escape from the bunch and hit the vacuum chamber. Estimates show that muon synchrotron radiation is not significant at 2 TeV and that muons that escape from the bunch can be held to a low level. Thus the main contribution to the heating comes from the decay electrons and positrons hitting the inner wall of the vacuum chamber and the synchrotron radiation they emit hitting the outer wall. Secondary radiation from these impacts in turn deposits energy elsewhere in the aperture. The magnet requirements for a muon collider are thus strongly influenced by the decay of the muons. The muons decay into electrons and positrons with around 35% of the muon energy. In the main ring, this amounts to a significant power deposition in the walls. Either a thick absorber will need to surround the

vacuum chamber, increasing the bore of the magnets, or a magnet design that moves the superconducting coils off the midplane (so that absorbing material in the midplane of the ring can be some distance from the coil) needs to be developed.

This section describes design options for dipoles and quadrupoles for the collider ring. It does not include any discussion about superconducting magnets needed in other subsystems of the muon collider nor those that are part of the detector system around the collision point.

8.3.1 Energy Deposition Due to Muon Decay

The collider will be a single separated function ring of superconducting magnets that guides both the negative and the positive muons. The number of muons that decay in a given length L_T can be estimated using the expression

$$N_d = \frac{N_\mu L_T E_0}{\tau_0 E_T c} \tag{8.5}$$

where N_μ is the number of muons transported through a structure per second, N_d is the number of muons that decay in the structure per second, L_T is the length of the structure, E_T is the muon energy, E_0 is the muon rest energy, c is the speed of light, and τ_0 is the muon decay time constant at rest ($\tau_0 = 2.197 \times 10^{-6}$ s). Equation 8.5 is applicable when the transit time for the muon through length L is less than the decay time constant of the muon at energy E_T. The power deposited into the magnet structure from muon decay can be estimated using the expression

$$P \simeq 0.35 N_d E_{ave} \tag{8.6}$$

where N_d is the number of muons that decay per second (See Equation 8.5), and E_{ave} is the average energy of the muon in the structure. The factor 0.35 in Equation 8.6 is the portion of the muon energy that ends up in the decay electrons or positrons. The remainder of the muon energy is transported out of the ring by the decay neutrinos. Table 8.2 presents calculations for muon decay in each of the accelerator components and the collider ring. Included in the Table 8.2 is the number of turns through the component and the total transit length L_T through the structure. Table 8.2 gives an estimate of the decayed muon power that is transferred to electrons and positrons. This is the portion of the decayed muon power that can end up in the superconducting magnet system. The beam flux of muons that enters the accelerator section is assumed to be 3×10^{12} muons per bunch.

The size of the region where the decay electrons, positrons and synchrotron radiation strike the wall of the vacuum chamber is determined primarily by the kinematics of the decay process. For a $3\,\sigma$ vertical beam size of 4 mm, this region is only about 5 or 6

Table 8.2: Muon decay parameters for various parts of a muon collider

Component	Peak Energy (GeV)	Number of Turns	L_T(km)	Total Muon Decay Rate $10^{13}s^{-1}$	Heating Power (kW)	Peak Heat per unit L (Wm^{-1})
Linac	1.0	-NA-	0.12	1.9	0.6	-NA-
First Ring	9.6	9	2.17	1.2	3.6	1.64
Second Ring	79	12	11.3	0.8	19.7	1.75
Third Ring	250	18	29.2	0.4	36.8	1.26
Fourth Ring	2000	18	227	0.6	378	1.66
Collider Ring	2000	1000	7.9	13.1	14600	1840

mm high even at muon energies as low as 20 GeV. As stated earlier, however, secondary interactions cause considerable energy deposition to occur throughout the aperture, so all of the aperture must be considered in the design of any energy-absorbing system. Calculations of energy deposition in the magnet structure are given in section 8.3.4.

8.3.2 Collider Ring Dipoles

The design of the dipoles and quadrupoles is dependent on the percentage of the muon decay product energy that can be removed from within the 4 K mass of the dipole before degradation of magnet performance. Several design approaches can be considered for the superconducting dipole magnets in the collider ring. One approach is to use a conventional cosine theta type of dipole provided the heavy (tungsten) radiation shield is thick enough to reduce the energy into the superconducting coils by about three orders of magnitude. The collider dipole warm bore is about 20 mm. The thickness of tungsten needed to reduce the heating from muon decay by three orders of magnitude is about 65 mm. Thus, the cold bore of the superconducting dipole coils must be 160 mm. The heavy radiation shield could be cooled with a fluid at room temperature. A collider ring dipole with a tungsten liner is illustrated in Fig. 8.11. The energy from the decay products that escapes this liner is well diffused in the coils (see section 8.3.4). Another approach to building collider ring dipole magnets is to have the coils completely separated on the mid plane. The iron return yoke would probably be cold because of the heavy supports needed between the yoke and the coils to restrain the attractive forces between coils. The coils must be separated so that less than 0.1% of the energy from the muon decay products ends up in the superconducting coils or

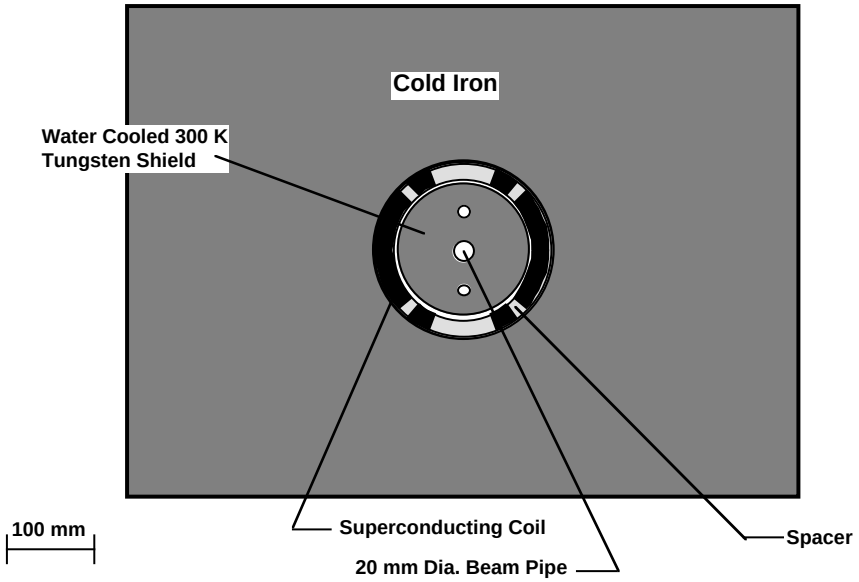


Figure 8.11: A cold iron 8.5 T cosine theta dipole with a 65 mm thick tungsten liner at 300 K

its surrounding support structure that is at 4 K. The rest of the muon decay product energy ends up in the separately-cooled radiation shield. Figure 8.12 shows cold iron dipoles with separation between the coils to allow for a warm vacuum chamber. The dipoles shown in Fig. 8.12 have reasonably good field quality.

Calculations by I. Stumer (private communication) suggest that almost all of the muon decay energy could be captured in absorbers placed at the ends of the magnets. The absorbers would have narrow apertures to capture as much of the radiation as possible yet allow passage of the beam; these narrow apertures may present beam impedance problems. Further work is necessary to confirm the viability of this promising approach.

8.3.3 Collider Ring Quadrupoles and Sextupoles

Several designs can be considered for the collider ring quadrupoles. 1) A design with a thick tungsten liner, similar to the dipole of Fig. 8.11, can be made. With the same 160 mm cold bore aperture, a gradient of 100 T/m could be achieved. 2) A figure-of-eight conventional quadrupole with a pole radius of 12 mm can be used. Conventional quadrupoles of this size can achieve a gradient of 100 to 120 T/m. Most of the muon decay energy can be absorbed by a cooled absorber outside the quadrupole. 3) A quadrupole can be designed that has its coils off the mid plane, Fig. 8.13. This quadrupole design is similar to the dipole design shown in Fig. 8.12. Depending on the field allowed in the superconductor, quadrupole gradients of 140 to 180 T/m can be achieved.

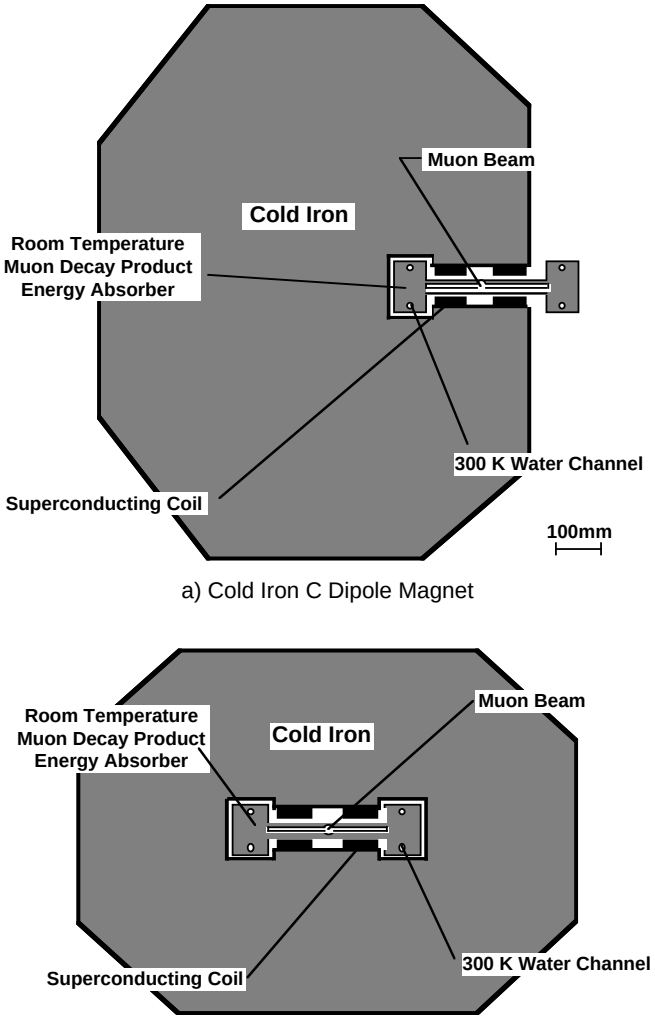


Figure 8.12: Two versions of an 8.5 T cold iron split dipole that would have less than 0.1% of the muon decay power deposited within the superconducting coils

The tuning sextupoles for the collider ring can be conventional. If these sextupoles have the pole at a radius of 12 mm, gradients as high as $3000\mathrm{T/m}^2$ may be achieved. There appear to be no superconducting sextupole designs able to absorb the muon decay products that can achieve this gradient.

8.3.4 Reduction of Heat Load in SC magnets

Due to muon decays, about 2 kW of power are deposited every meter along the collider ring.This results in a heat load that significantly exceeds levels that can be tolerated by existing SC magnets. The energy-deposition distributions in the storage ring components

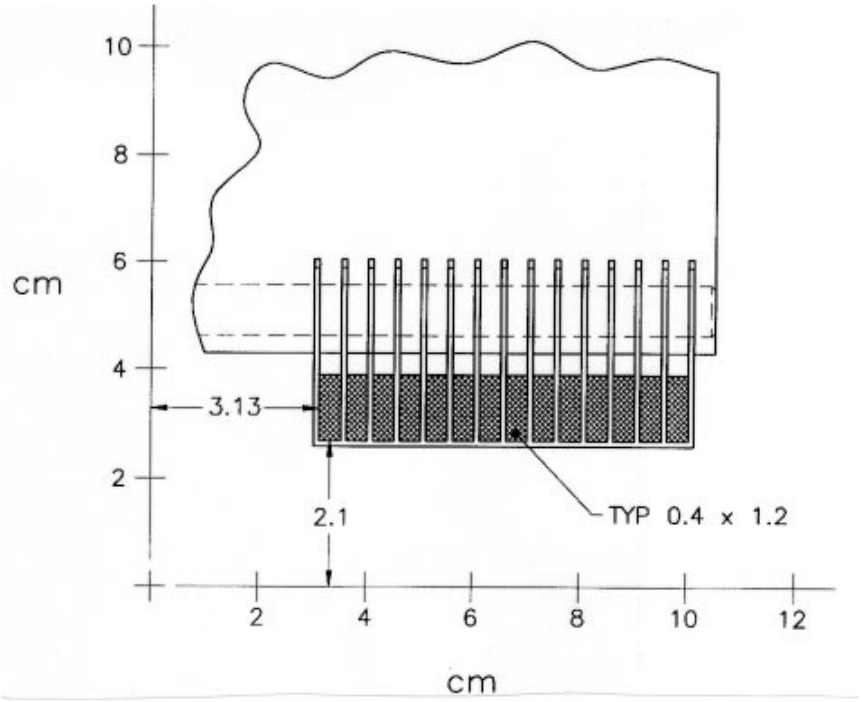


Figure 8.13: A quadrupole design that avoids superconductor on the midplane so that muon decay particles can escape. The drawing shows one pole of the design

from muon decay have been calculated for a 2 TeV muon beam with the MARS code [12]. Even with a longitudinally uniform source, there is an increased background at the high-β_{peak} locations near the IR. Fig. 8.14 shows the azimuthal distribution of power deposited in the first cable shell of the arc SC magnets with tungsten liners of different thicknesses. There is a significant azimuthal dependence of power density due to the effect of the strong magnetic field. The lateral gradient of energy deposited in the SC coil is very strong both with and without a liner. The peak power density exceeds the expected quench limits for the magnet of the assumed type by more than an order of magnitude. A 4 cm tungsten liner provides a considerable reduction of the maximum power deposition density from the quench stability standpoint (see Fig. 8.15). Another concern is the power dissipation in the cold magnet components. As mentioned above, up to 1 kW of power per each beam would be deposited in every meter of the lattice, which is about 1000 times above a possible limit for such extended systems as a collider ring. Fig. 8.16 shows power dissipation in tungsten liner, liquid helium, SC coils, yoke and cryostat components as a function of the liner thickness per one meter of the arc lattice per one beam. One sees that 6 cm liner is required in the arcs in a cosine theta approach. The studies show that in the interaction region two quadrupoles nearest to the interaction point also need at least 6 cm of tungsten shield in front of the SC

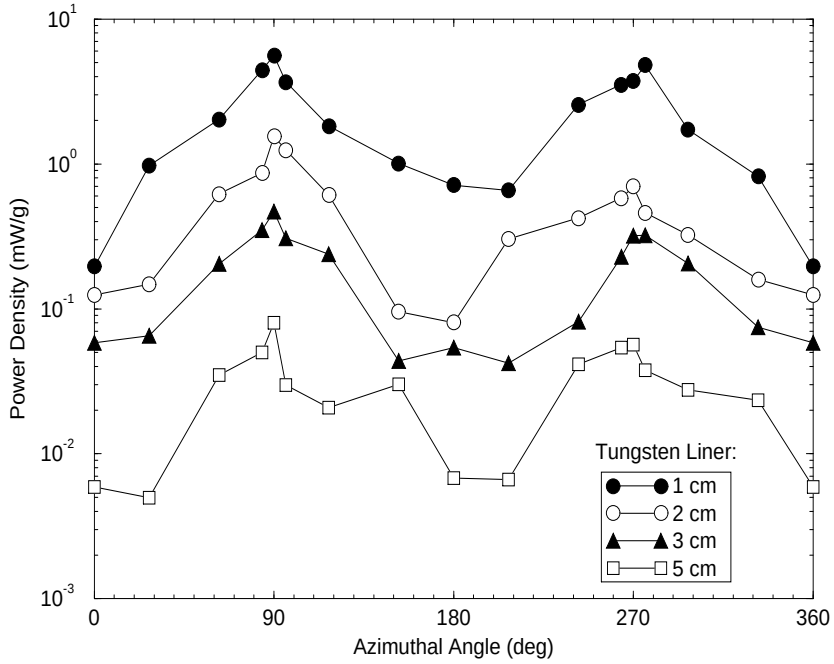


Figure 8.14: Azimuthal distribution of power density in the first SC cable shell in the collider arc for different tungsten liners inside the aperture for 2 TeV muon beam decays

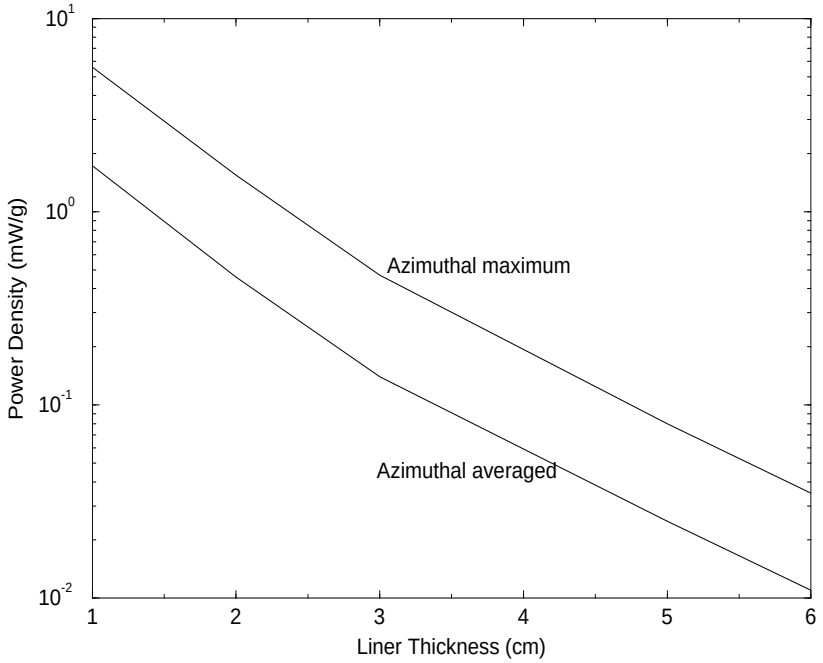


Figure 8.15: Maximum and azimuthal averaged power density in the first SC cable shell in the collider arc *vs* tungsten liner thickness for 2 TeV muon beam decays.

coils in order to reduce the power deposited in the cold mass to the level of ≈ 1 W/m. Even with that, the heat load would be higher compared to that in the arc magnets. This can be tolerated locally, but preferably another solution should be found. A 3 cm tungsten liner is fine for the rest of the IR.

8.4 Radio-Frequency System for the Collider Ring

A radio-frequency system is required for the muon collider ring principally to maintain the tight longitudinal focusing of the muon bunch. It is needed also to make up the energy lost by the beams to higher order mode (HOM) losses in the vacuum system, synchrotron radiation and resistive losses in the walls of the beam pipe.

The beam and collider-ring parameters used in this subsection, which may differ slightly from those in Table 8.1 are in Table 8.3. Two factors influence the choice of rf frequency. The short bunch can be maintained with less voltage at the higher frequencies, while the wakefields and losses are smaller at lower frequencies. The 10-cm wavelength SLAC linac structure was considered as a prototype, but it had to be rejected because of the very large

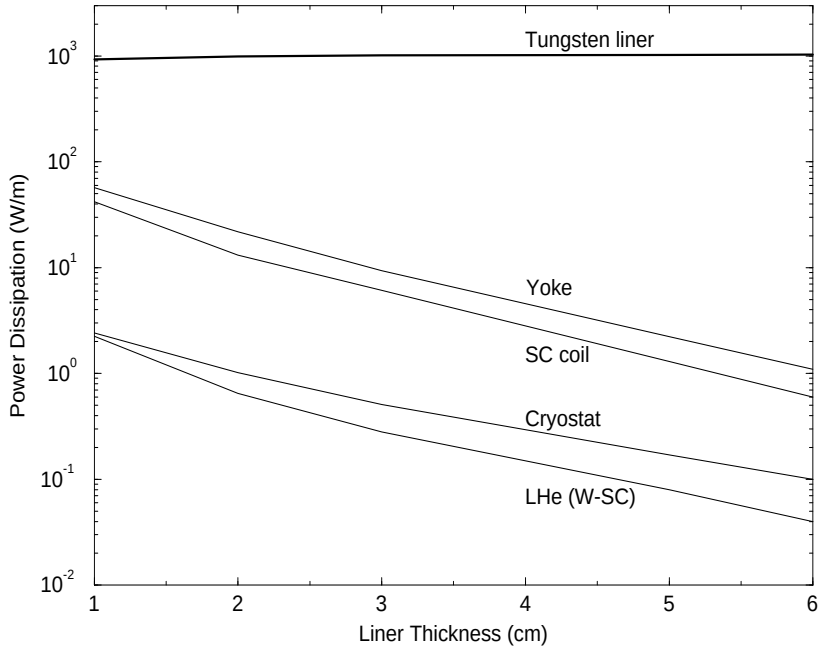


Figure 8.16: Power dissipation in the arc magnet components *vs* tungsten liner thickness for 2 TeV muon beam decays.

distortion of the rf wave shape by the higher order mode losses in the rf cavities by the high-intensity muon bunch. In the SLAC-type structure, in which an average accelerating field strength of 20 MV/m can be utilized, the peak wakefield voltage is estimated to be 15 MV/m, which would cause an intolerable distortion of the rf accelerating field [14].

Table 8.3: Collider parameters used for the rf system

rms bunch length	σ_z	3 mm
rms energy spread	σ_E/E	1.5×10^{-3}
longitudinal emittance	$\epsilon_L = \sigma_E \sigma_z / c$	3×10^{-2}eV-sec
collider circumference	2πR	7 km
revolution frequency	f_0	43.8 kHz
compaction factor	α	10^{-6}
muon storage time		1,000 turns

This led to our considering a TESLA-type rf system–a cryogenic, 1.3 GHz (23 cm wavelength), standing-wave system, which can be operated at 25 MV/m average accelerating field strength and in which the peak wakefield distortion due to the muon bunch is estimated to

be in the range of 3 or 4 MV/m. An rf bucket that is matched to the longitudinal emittance of the muon beam requires 130 MV peak voltage. The corresponding, nominal synchrotron tune ν_s is 5.6×10^{-4} synchrotron oscillations per turn, so that in the 1,000-turn storage time of the muon it undergoes only about 0.56 of a synchrotron oscillation.

Since the loading time of the TESLA structure (800 microseconds) is much longer than the revolution time of the muon beams (23 microseconds), the rf pulse length in the muon collider must extend over the storage time on the muon beam, which is 23 milliseconds or greater. The nominal TESLA operation is with 1.3 millisecond pulses at 10 pps, so that operation with much longer pulse lengths needs to be ascertained.

The dynamics of the muon beam in longitudinal phase space is being investigated with a self-consistent particle-tracking program. We approximate the single-particle wake potential for the TESLA structure, with the SLAC wake potential scaled to the TESLA frequency [15] and adjusted to fit the amplitude of a wakefield voltage distribution as calculated by Mosnier [16] for the TESLA structure. The computer program is self-consistent in that the wake fields are continually recalculated, taking into account the evolving beam shape and distribution in longitudinal phase space.

The wakefield simulation, which includes muon decay, has been used to study the longitudinal dynamics in the ring. The simulations assumed a constant slippage factor of 10^{-6}. The potential well distortion can generate large motion of the bunch center. With a nominal bunch offset of 0.082 radians oscillations of the bunch center, which have a peak amplitude of order .6cm at zero offset, are reduced to less than 0.05 cm.

Even with an offset, the bunch shape evolves during the storage time. The rms bunch length σ_z initially drops from 3 mm to a minimum of about 2.5 mm at about turn number 340 , then rises to a maximum of 3.2 mm at turn 760, and then falls again, reaching about 2.8 mm at turn 1,000. The rms energy spread σ_E/E initially rises from 1.5×10^{-3} to about 2.0×10^{-3} at turn 350, then descends to a minimum of 1.5×10^{-3} at turn 750, and then rises again, reaching about 1.85×10^{-3} at turn 1,000. The longitudinal emittance slowly increases by about 7 percent over the 1,000 turns. These oscillations can be seen in Fig. 8.21 in Sec. 8.8. Further studies exploring methods to reduce the bunch shape oscillations are underway.

The peak wakefield voltage gradient at the first turn was about 4.4 MV/m and the beam-loading factor, defined as the average energy loss per muon divided by the peak rf energy gain is 10%.

The useful bucket area is very sensitive to the slope of the lattice compaction factor with energy. The "design" lattice in Sec. 8.2 envisions an α for which the contributions from nonlinear (in energy deviation) terms is negligible. Nevertheless, it is worthwhile to

determine the limits longitudinal dynamics place on the maximum allowable values of the nonlinear momentum compaction terms (i.e., α_2, and α_3 as discussed in Sec. 8.2).

The bunch has resistive-wall losses [17] of be about 15 MeV/turn/muon in a room-temperature copper beam pipe of radius 1.7 cm and an intensity of 2×10^{12} muons and an rms length of 3 mm. This loss is comparable to the higher order mode losses in the rf cavities, so that the voltage-wave distortion due to these resistive losses must be considered as well. These resistive losses must be replenished by the rf system, but they do not change the parameters significantly as bunch length effects are dominant. Studies of higher order mode losses in the rest of the ring are underway.

8.5 Ring Vacuum Chamber

The choice of ring parameters requires a careful study of beam tube power fluxes and beam tube vacuum issues. These include calculations of the flux, energy and power of particles that escape from the bunches in the collider ring and estimation of where they will hit the walls. Sources of such power at the walls are synchrotron radiation of muons and decay electron-positrons, as well as the decay electrons and positrons themselves and any muons that escape from the bucket. Significant effort will need to be invested in determining, and controlling phenomena that may lead to the escape of particles from the beam core into a halo as well as in the design of a beam halo scraping system which will reduce the flux of lost muons into the detectors and superconducting coils. Some possible mechanisms are the classical beam-beam interaction, hard scattering with background gas or in the beam-beam interaction, incoherent and coherent pair production at the IP, and nonlinear dynamics. The consequences of a muon leaving the bucket need to be carefully studied since a muon that leaves the ring will not be stopped by the shielding, and may deposit energy in the superconducting magnets and create detector background. The current thinking is that only 10^{-6} of the muons can be allowed to escape per turn.

8.5.1 Particle Fluxes

Excluding escaping muons, the three sources of particle flux that strike the beam tube wall are:

- synchrotron radiation from the circulating muons,
- electrons and positrons from the decay of the muons and

- synchrotron radiation emitted by the decay electrons and positrons before they strike the beam tube.

The particle flux, characteristic energy and power flux of each of these is summarized in Table 8.4. For these estimates we have assumed there are two bunches for each muon charge with 2×10^{12} muons per bunch at injection. The intensities and power levels are average values taking account of muon decay and sum the contributions from both signs of muons. The muon energy is assumed to be 2 TeV so the muon lifetime is $\tau_\mu = 41.6$ msec and fresh bunches are injected at 15 Hz, $T_{rep} = 66.7$ msec. At the present time it is undecided whether or not surviving muons will be extracted prior to each injection. For the estimates in Table 8.4 we assume they are not. If surviving muons are extracted just prior to injection of fresh bunches then the intensities and powers are multiplied by the fraction 1 - $exp(-T_{rep}/\tau_\mu) = 0.80$. The bend field is taken to be B = 9 T ($\rho^B = 741$ m) over two thirds of the ring so the total circumference including straight sections is 7.0 km ($T_{rev} = 23.3$ μsec). We have omitted muons diffusing past the dynamic aperture from Table 8.4 since the rate hasn't yet been estimated from tracking calculations. The goal for treating these halo muons would be to intercept as many as possible with a beam scraping and collimation system at a few isolated locations far removed from the detector - e.g. in the utility region on the opposite side of the ring. The particle fluxes in Table 8.4 are distributed uniformly around the ring with synchrotron radiation of course only in the bends. For comparison with the fluxes in Table 8.4 we will give the halo muon numbers that correspond to the provisional 10^{-6} loss probability per turn, all time averaged and without extraction of surviving muons and summed over both signs of muons: 2.14×10^{11} lost $\mu^\pm/s$, 68.5 kW lost $\mu^\pm$ beam power. The particle loss rate is more than three orders of magnitude less than the muon decay rate and the power level is about twice the $\mu^\pm$ synchrotron radiation power. The $\mu^\pm$ however are much more penetrating than $e^\pm$ and photons so they present special problems for shielding the detector and superconductor.

The most important thing to notice in Table 8.4 is that the largest particle flux is synchrotron radiation from muons while the largest power flux on the beam tube is due to $e^\pm$ from muon decay and the associated $e^\pm$ synchrotron radiation. Synchrotron radiation from $\mu^\pm$ is all absorbed on the outside of the ring. The synchrotron radiation loss per turn is 1.05 MeV. The radiation damping time is $\sim 2 \times 10^6$ turns, or a factor $\sim 10^3$ times greater than the muon lifetime. The $e^\pm$ are all absorbed on the inside of the ring and the synchrotron radiation from $e^\pm$ is divided with some falling on the inside and some on the outside of the ring. The amount of energy each $e^\pm$ radiates before it hits the beam tube and the fractions of $e^\pm$ synchrotron radiation absorbed on the inside and outside of the ring depend on the

Table 8.4: Time average $\mu^{\pm}$ particle fluxes, characteristic energies and power incident on the beam tube assuming no extraction of surviving muons prior to each injection. All rates are summed over both signs of muons.

Source	Intensity	Characteristic energy	Power
	particles/m/sec		W/m
$\mu^{\pm}$ syn. rad.	5.7×10^{16}	$E_c = 2.7$ keV	7.7
$e^{\pm}$ from $\mu^{\pm}$. decay	1.7×10^{10}	$< E > = 700$ GeV	1500
			(=1900-400)
$e^{\pm}$ syn. rad.	2.3×10^{12}	$< E_c > = 2.1$ GeV	400

detailed geometry and dimensions of the beam tube. So far two general situations have been discussed: (1) a warm cylindrical beam tube surrounded by an absorber with $cos(\theta)$ coils at cryogenic temperatures on the outside and (2) a warm beam tube with a slot on the inside radius to allow $e^{\pm}$ to escape and strike an absorber which is placed outside the cryogenic windings of C - magnet coils. In the second case an additional smaller absorber at the outside radius would be inside the C-coils and absorb synchrotron radiation. For estimates in this preliminary report we will assume a simple warm cylindrical beam tube with radius $r_w = 1.0$ cm. In that case $\mu^{\pm}$ synchrotron radiation photons travel a sagittal distance 3.85 m before striking the outside radius of the beam tube at an angle 5.2 mrad. Looking at the particle energies in Table 8.4, the $\mu^{\pm}$ synchrotron radiation is relatively soft with critical energy 2.7 keV. These photons will be absorbed on the surface of the beam tube and photodesorb a significant amount of gas estimated in Sec. 8.5.3 below. Photodesorbtion by $\mu^{\pm}$ synchrotron radiation will determine the gas pressure in the beam tube, thermal desorption is relatively unimportant even for an unbaked beam tube. The $e^{\pm}$ and accompanying synchrotron radiation are enormously more energetic than the $\mu^{\pm}$ synchrotron radiation and will produce electromagnetic showers that penetrate deeply into the shielding and magnet structure surrounding the beam tube. The $e^{\pm}$ have a broad energy spectrum that extends from essentially zero to the full 2 TeV $\mu^{\pm}$ energy. The mean $e^{\pm}$ energy is equal to $0.35 \times E_{\mu} = 700$ GeV. The $e^{\pm}$ synchrotron radiation parameters given in Table 8.4 have been averaged over the $e^{\pm}$ energy spectrum. The $e^{\pm}$ synchrotron radiation in the 9 T bend field is large enough that its effect in reducing the $e^{\pm}$ energy before intercepting the beam tube must be taken into account or the magnitude of synchrotron radiation will be overestimated. For example in a 9 T bend field the mean critical energy averaged over the $e^{\pm}$ spectrum from 2 TeV muon decay is 4.2 GeV; this is reduced to 2.1 GeV when account is taken of synchrotron radiation reducing the $e^{\pm}$ energy and the average is taken over the path length from the

point of emission to the point of intercept with the beam tube at $r_w = 1.0\,\text{cm}$. Similarly the power radiated by $e^{\pm}$ has been integrated over the $e^{\pm}$ spectrum and corrected for emission of synchrotron radiation. The $e^{\pm}$ radiate about 20% of their energy before striking the inside radius of the $r_w = 1\,\text{cm}$ beam tube. This radiation energy has been subtracted from the decay energy received by $e^{\pm}$ in Table 8.4 to avoid double counting the energy reaching the beam tube. Absorbing the $e^{\pm}$ and accompanying synchrotron radiation and shielding the superconducting cable and detectors will be serious design challenges. These last two particle fluxes however should not have a significant impact on the beam tube vacuum.

8.5.2 Beam Gas Scattering and Beam Tube Gas Density Requirement

Circulating muons undergo two types of collisions with gas molecules in the beam tube that can cause deleterious effects: (1) multiple small angle Coulomb collisions and (2) single large angle nuclear Coulomb collisions. In the absence of significant radiation damping multiple small angle collisions cause a steady increase in beam emittance and loss of luminosity. Setting the emittance growth time equal to 10^4 turns, approximately ten times the luminosity lifetime, sets an upper bound on average beam tube gas pressure of 3.1 mTorr CO scattering equivalent. Single large angle nuclear Coulomb collisions can lead to a betatron amplitude that exceeds the dynamic aperture of the collider ring or intercepts a physical aperture. Here we assume the dynamic aperture exceeds the physical aperture. The characteristic physical apertures in the collider ring are summarized in Table 8.5. The IR quads have the smallest aperture with $H/\sigma = 3.1$ normalized to the beam rms beam size in one transverse dimension. In reality there will be a halo scraper yet to be designed with normalized aperture somewhat smaller then the IR quadrupoles. For estimating single large angle scattering loss we will assume a limiting aperture $H/\sigma = 3.1$. A loss probability of 10^{-6} per turn has been provisionally adopted as a limit for beam halo losses that can be tolerated by the detectors. Allowing a large angle scattering loss probability of this same magnitude leads to a beam tube gas pressure requirement of less than 4.2 mTorr CO scattering equivalent, not too different than the multiple scattering requirement. Because of the relatively short storage times in a muon collider and different collision characteristics compared to electrons and protons, the beam tube vacuum requirement is enormously more relaxed than in conventional electron and proton storage rings. This is a fortunate and perhaps even necessary circumstance since dealing with the large particle and power fluxes of $e^{\pm}$ from muon decay may restrict the possibilities for pumping.

Table 8.5: Characteristic apertures and rms beam dimensions in the collider ring. Note that $\epsilon = \sigma^2/\beta = 2.65 \times 10^{-9} m - rad$

Location	Lattice β (m)	σ (mm)	aperture radius H (cm)	H/σ
Arc	50	0.28	1.0	35.7
IP	3×10^{-3}	2.2×10^{-3}	.025	114
IR quad	360,000	24.0	7.5	3.1

8.5.3 Sources of Beam Tube Gas and Estimate of Beam Tube Gas Density

The two most significant sources of gas molecules in the collider beam tube are:

- photodesorbtion by $\mu^{\pm}$ synchrotron radiation and
- thermal desorption.

To estimate the beam tube pressure due to photodesorbtion we use experimental measurements of $E_c = 3$ keV photodesorbtion coefficients on a LEP aluminum vacuum chamber [18]. Photodesorbtion coefficients decrease steadily with photon exposure as molecules in the surface oxide layer are desorbed and pumped out. For our numerical estimates we take the magnitudes of desorption coefficients after exposure to 10^{21} photons/m. This corresponds to a few hours of operation of the muon collider ring at design intensity, a modest conditioning time. Similarly the outgassing rate of unbaked vacuum chambers decreases steadily with time and to be specific we take outgassing rates characteristic of chemically cleaned aluminum after 100 hours of pumping [19]. Accurately calculating the pressure from the gas sources depends on the detailed dimensions of the beam vacuum chamber and location and sizes of vacuum pumps. Here we are only interested in an order of magnitude estimate so we take what is probably a worst case - a beam tube with radius $r_w = 1.0$ cm and pumped only at 20 m intervals, a reasonable allowance for the length of superconducting dipole magnets. We also assume the beam tube is near room temperature owing to the requirements of removing the large power fluxes due to muon decay. In this situation it is trivial to choose vacuum pumps so the effective pumping speed is limited by and essentially equal to the conductance of the beam tube. With these assumptions the gas loads and beam tube axially averaged partial pressures in Table 8.6 have been calculated. The beam tube pressure is effectively dominated by the photodesorbtion of CO and CO_2 and taken together amount to about 10^{-4} Torr CO scattering equivalent, more than an order of magnitude below the requirement discussed in the previous section. The partial pressure of the dominant gas due

Table 8.6: Estimated sources of gas and average partial pressures in the collider beam tube.

gas	Q/photo	$\langle p \rangle$	Q/thermal	$\langle p \rangle$
	Torr cm^2 sec^{-1}	Torr	Torr cm^2 sec^{-1}	Torr
H_2	2×10^{-5}	2×10^{-5}	4×10^{-8}	3.6×10^{-8}
CH_4	3×10^{-7}	8×10^{-7}	3×10^{-9}	8×10^{-9}
H_2O	-	-	2×10^{-6}	5.4×10^{-6}
CO	10^{-5}	3.4×10^{-5}	3×10^{-8}	1×10^{-7}
CO_2	10^{-5}	4.3×10^{-5}	3×10^{-9}	1.3×10^{-8}

Table 8.7: Resistive wall dissipation for Cu, Al and SS beam tubes.

material	$\langle p \rangle$(MW)	$\langle p/L \rangle$(W/m)	$\langle \Delta W \rangle$(MeV/turn)
Cu	.59	83.6	17.1
Al	.74	106	21.5
SS	3.8	543	111

to thermal outgassing - H_2O is approximately 20 times less than the sum of photodesorbed CO and CO_2. Eventually, as photodesorbtion cleans up the CO and CO_2, thermal outgassing of H_2O may become the dominant beam tube gas component during operation.

8.5.4 Beam Tube Resistivity and Image Current Losses

The image current of the bunched muons will be carried on the surface of the beam tube in a thin layer with thickness of the order of a skin depth. This current will drive an Ohmic loss that is estimated in Table 8.7 for copper, aluminum and stainless steel, assuming a beam tube with radius $r_w = 1.0$ cm, two bunches for each sign of muon with 2×10^{12} muons per bunch injected, no extraction of surviving muons prior to each injection and an rms bunch length $\sigma_z = 3$ mm. The results in Table 8.7 are time averaged, taking account of muon decay. The average muon energy loss per turn in Table 8.7 is an intensity weighted average. For copper and aluminum the resistive wall energy loss exceeds the synchrotron radiation loss by a factor of approximately twenty.

8.6 Classical Beam-Beam Interaction

A preliminary look at the classical beam-beam interaction for the muon collider has been carried out [20]. The rest of this Section is taken from Ref. [20]. The beam-beam simulations

for the muon collider in the incoherent classical approximation, taking into account muon decay, show that for the nominal intensity of 2×10^{12} per bunch the beam-beam interaction is reasonably weak and allows room for upgrading the luminosity performance. The analysis shows that classical coherent beam-beam effects will almost certainly not spoil the beam luminosity lifetime. The beam-beam effects become noticeable if the number of particles is increased by a factor of three or more. These conclusions are based on the assumption that the dynamical aperture is sufficiently large (say, 5σ or larger). For a smaller aperture nonlinear lattice dynamics will need to be included in the calculation.

8.6.1 Introduction.

The classical beam-beam effects arise from the interaction of the particles in one beam with the classical electromagnetic fields of the opposing beam. The fundamental dynamics is the electromagnetic deflection of the particles. Quantum beam-beam effects deal with particle annihilation and creation as described by relativistic quantum mechanics.

Incoherent effects are those that are well described by the interaction of a single particle in one beam with the other beam (or by the simple superposition of such interactions), while coherent effects are those that can only be explained by the interaction of the beams with each other as whole.

A basic example of an incoherent classical effect is the blowup of the beam core (emittance blowup) as the beams collide turn after turn; as a consequence of this blowup the luminosity degrades, at least to some extent. In this case the phase space of the core particles remains essentially structureless (approximately gaussian in the case of e^+e^- machines). Another example is the development of large-amplitude tails in the particle distribution, which leads to a decrease of the beam lifetime as the particles are gradually lost to the machine aperture. In this case, the phase space of these large amplitude particles has a characteristic structure that is dominated by one or more resonances arising from the combined dynamics of the beam-beam force and the nonlinear magnetic fields of the machine. These two phenomena dominate the beam-beam dynamics of essentially all hadron and lepton colliders built so far. For well-tuned e^+e^- colliders with good dynamic aperture, these effects have vastly different time scales: the core blowup always happens over a few damping times, while the development of significant tails can be arranged to happen over thousands of damping times or even longer [21].

The signature for classical coherent effects is a nontrivial structure of the phase-space of the core particles. This space structure can arise when the tune is close to a low-order resonance. An example of this kind of effect is the flip-flop state in e^+e^- colliders: in this

case, for sufficiently high bunch current, the two beams reach an equilibrium situation in which one of them is blown up while the other is not. This effect has been observed in most colliders. Other coherent states that have been predicted in simulations, and perhaps observed experimentally, are period-2 or -higher fixed points, in which the sizes of the two beams fluctuate from turn to turn in a periodic pattern. Simulations generally show that the time scale for these effects is of the order of 10-20 damping times [22, 23].

An example of an incoherent quantum effect for the muon collider is the reaction $\mu^+ + \mu^- \to e^+ + e^-$ that can happen during the beam-beam collision. A muon can also interact with the collective electromagnetic field of the opposing bunch to produce e^+e^- pairs; this is a coherent quantum effect. The latter effect is believed to be negligible for this ring, while the former is addressed in Sec. 8.7.

In this section, however, we will show that incoherent classical effects are weak, at least for nominal parameter values, and that coherent classical effects are very unlikely to materialize. We will also provide rough criteria for the tolerances for the ratio β^*/σ_z, and for the longitudinal displacement of off-IP collisions which can arise from RF phasing errors or from timing errors in the injection process.

8.6.2 Physics of the Incoherent Simulation.

We carry out a simulation with the code TRS [24]. This is a "strong-strong" simulation in which both beams are dynamical, and their emittances evolve according to their mutual interaction. The simulation is fully six-dimensional, and the beam-beam interaction is represented as a thick lens by dividing up the bunches into 5 "slices." We assume one bunch per beam, and a single interaction point (IP). The beams are represented by "macroparticles" (1024 per bunch in this case), and the machine lattice is assumed to be strictly linear, so that it is represented by a simple phase advance matrix. This linear approximation is valid provided that the dynamical aperture is sufficiently large (say, 5σ or larger). The three tunes are taken as input quantities to the simulation, and we set the chromaticity to 0. From other work, we know that the values we have chosen for the number of slices and macroparticles are adequate for the nominal muon collider specifications [25].

The beams are described at time $t = 0$ by six-dimensional gaussian distributions whose σ's are determined by the specified nominal parameters of the collider (see Table 8.8). We then let the bunches collide for 1000 turns, keeping track of the six-dimensional coordinates of all the macroparticles, and measure from these the beam sizes and the luminosity at every turn as they evolve according to the beam-beam dynamics. The code uses the so-called "soft-gaussian approximation" by virtue of which, for the purposes of computing the beam-

beam kick, the opposing bunch is *assumed* to have a gaussian shape in the two transverse dimensions, albeit with time-dependent σ's. This approximation is generally reliable provided no coherent effects appear, which is almost certainly the case for the muon collider, as we shall discuss below. We take into account the muon decay by simply multiplying the number of particles per bunch N in each beam by the exponential decay factor $\exp(-t/\tau)$, where τ is the Lorentz-dilated muon decay constant.

8.6.3 Beam-beam Simulation.

Simulation Conditions.

For the purposes of this simulation, we assume parameters as listed on Table 8.8 (both beams have the same parameters). In this table β^* is the common value of the horizontal

Table 8.8: Muon collider parameters.

C [km]	7
E [TeV]	2
N	2×10^{12}
β^* [mm]	3
ϵ_N [mm–mrad]	50
f_c [kHz]	42.86
ν_x	0.57
ν_y	0.64
ν_s	1/160
σ_z [mm]	3

and vertical beta-functions and the same is true for the normalized emittance ϵ_N. The values for the horizontal and vertical fractional tunes ν_x and ν_y were picked arbitrarily (the integral part of the tune does not enter the simulation).

With these values, the beam size at the IP is

$$\sigma^* = \sqrt{\beta^* \epsilon_N / \gamma} = 2.74 \ \mu\text{m} \tag{8.7}$$

where $\gamma \approx 18,900$ is the usual relativistic factor. The nominal value for the peak luminosity is

$$\mathcal{L} = \frac{f_c N^2}{4\pi\sigma^{*2}} = 1.82 \times 10^{35} \ \text{cm}^{-2} \ \text{s}^{-1} \tag{8.8}$$

which is not exactly the time averaged luminosity $\times 10^{35}$ cm^{-2} s^{-1}, as listed elsewhere in these report.

It is worthwhile to note that the beam-beam parameter,

$$\xi = \frac{r_0 N}{4\pi\epsilon_N} = 0.046 \tag{8.9}$$

has a fairly typical value: In fact, beam-beam parameter values like this have been attained or exceeded in several e^+e^- colliders (here r_0 is the classical radius of the muon). Actually, it is intriguing that the values of γ and ξ are similar to those in the former PEP collider, so certain aspects of the incoherent beam-beam interaction can be expected to be similar to those in PEP.

An important parameter in colliding rings is the damping time. Assuming that the synchrotron energy loss per turn in the muon collider is 1 MeV, the transverse damping time is

$$\tau_x = \frac{2 \text{ TeV}}{1 \text{ MeV}} = 2 \times 10^6 \text{ turns} \tag{8.10}$$

which is much larger than the 1000 turns' duration of a cycle. The large difference between these time scales is crucial in explaining some beam-beam effects.

Simulation Results.

Fig. 8.17 shows the luminosity as a function of turn number obtained under the assumption that the muon is a stable particle. One can see that it decreases by $\sim 4\%$ during the course of the 1000 turns due to the incoherent emittance blowup. This fractional decrease is small because the beam-beam parameter is modest, and because the cycle time is so small relative to the damping time.

Fig. 8.18 shows the luminosity taking into account the finite muon lifetime. As expected from the previous result, the curve is essentially determined by the exponential decay factor of the muons.

Fig. 8.19 shows the luminosity *vs.* turn number for three values of the number of particles per bunch N. For each value of N we carried out the simulation for three random number seeds; thus the spread in the curves for each case gives an idea of the statistical errors of the calculation. The bottom curves, corresponding to the nominal value of 2×10^{12}, are the same as in Fig. 8.18. The middle curves, for $N = 4 \times 10^{12}$, still behave quite nominally. However, it is clear that the curves for $N = 6 \times 10^{12}$ decay faster than exponentially due to substantial emittance blowup. In addition, when we included a 10σ physical aperture in the simulation, we observed that there were no particle losses for the first two cases, but there was a $\sim 2\%$ integrated beam loss for the case where $N = 6 \times 10^{12}$. Although this is a small

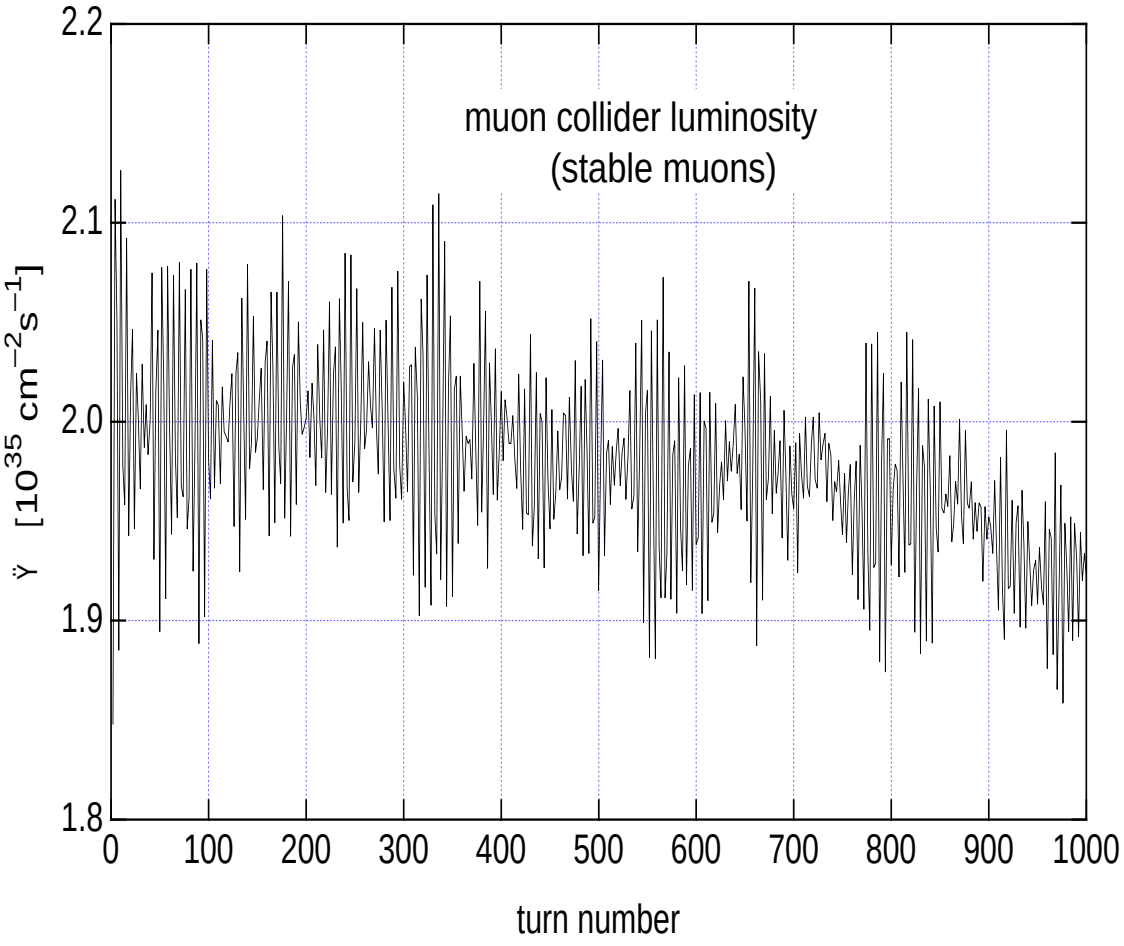


Figure 8.17: Luminosity as a function of turn number assuming that the muons are stable particles. The luminosity decreases by $\sim 4\%$ over 1000 turns due to incoherent emittance blowup.

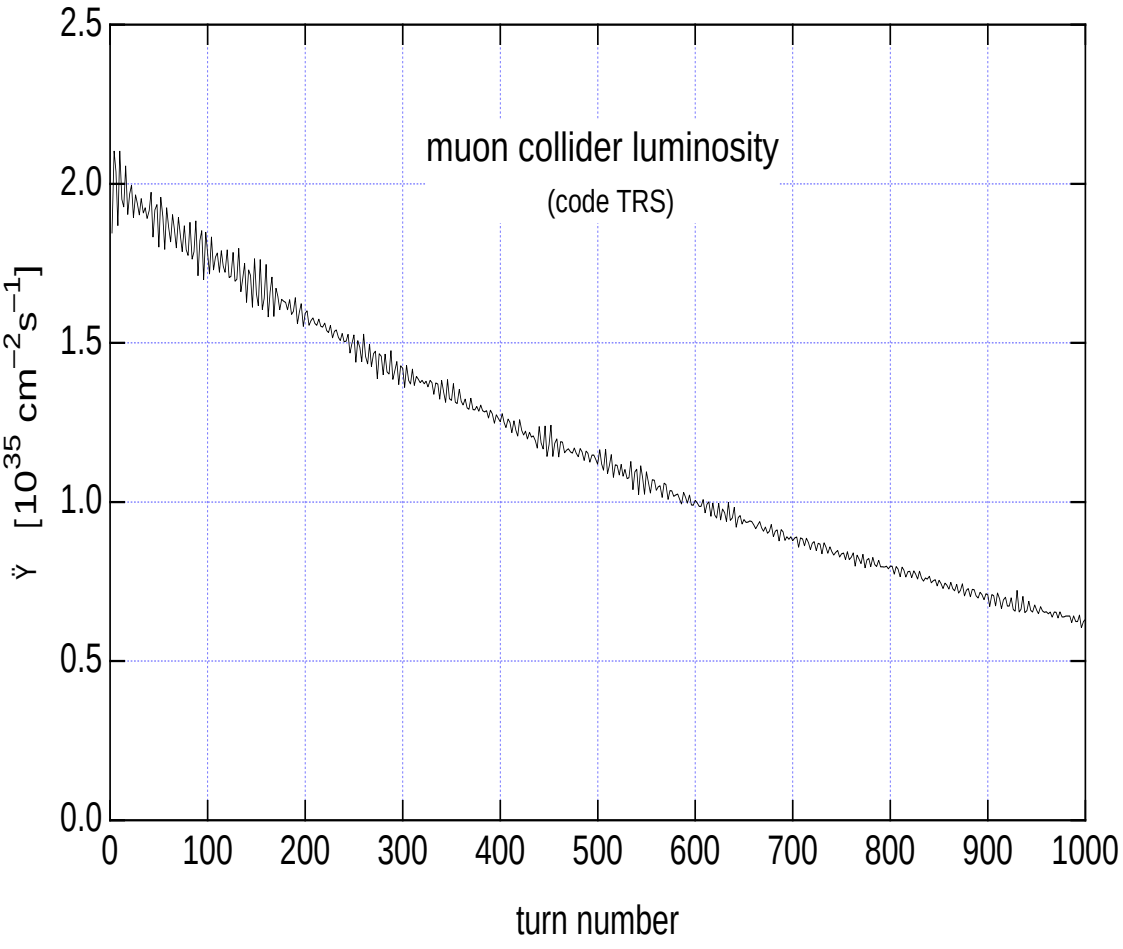


Figure 8.18: Luminosity as a function of turn number, taking into account the finite muon lifetime. The curve follows closely the expected exponential decay dependence.

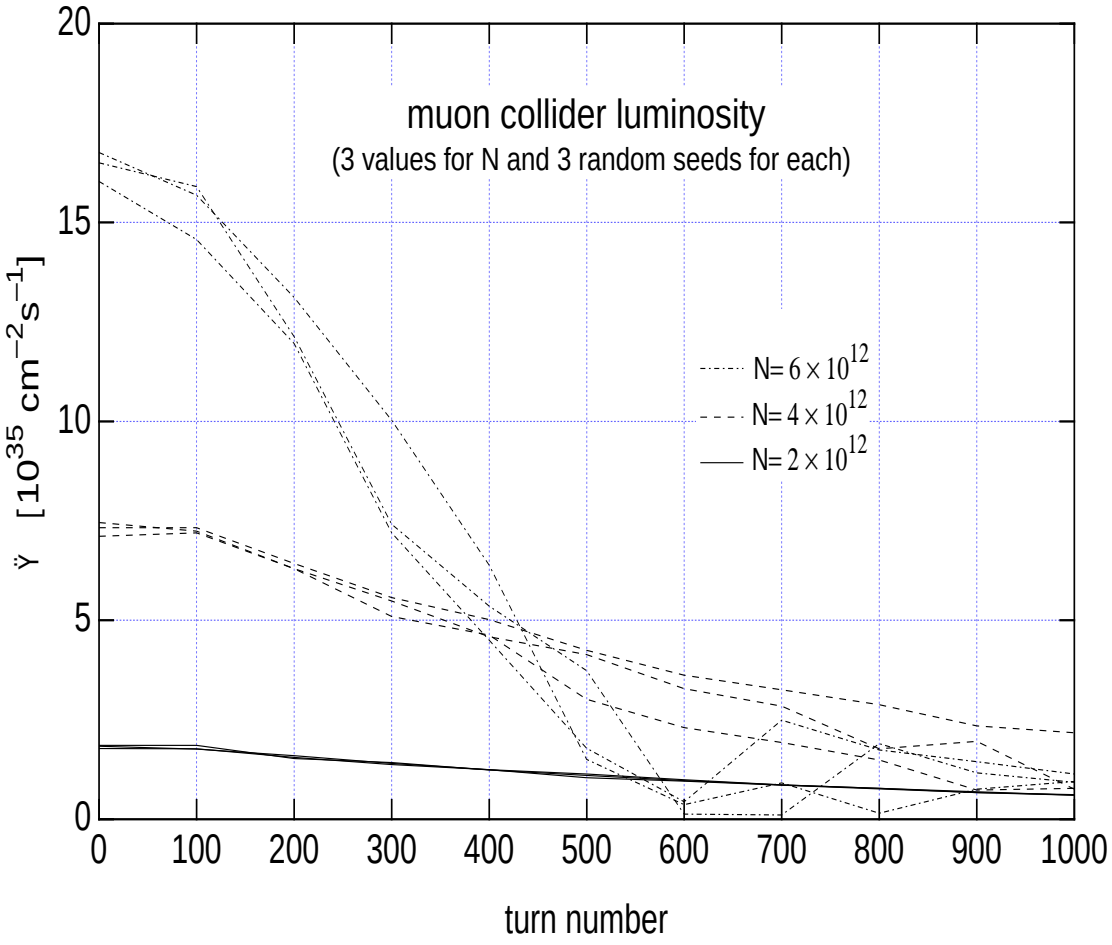


Figure 8.19: Luminosity as a function of turn number for three different values of the number of particles per bunch N. For each case we show three runs, each corresponding to a different random number seed; the spread of the curves for each case gives an idea of the statistical accuracy of the calculation.

fraction of particles, it is reasonable to interpret this as a symptom that the beam-beam strength is being pushed beyond a prudent limit, and the results of this simulation cannot be taken as a reliable guide. When this kind of behavior is seen, it is likely that other detrimental effects, not included in this simulation, will become important and will lead to even more unfavorable behavior. We conclude from this calculation that the incoherent beam-beam effect is weak for the nominal current and that there is some room for upgrading the luminosity by increasing the bunch current by a factor of ~ 2 but not more than this.

8.6.4 Other Classical Beam-Beam Issues.

Coherent Effects.

Classical coherent effects significantly distort the phase space of the beam core away from the gaussian shape. This distortion may be static or time dependent, and leads to luminosity degradation; thus, despite the theoretical interest of these effects, in practice one wants to identify the conditions under which they appear in order to avoid them.

Simulation studies for e^+e^- machines [22] show that these effects materialize for beam-beam parameter values $\gtrsim 0.05$ and for isolated values of the fractional tune where certain low-order resonances dominate the dynamics. More importantly, these results also show that coherent effects take a long time to develop, on the order of 10 damping times or more, simply because it takes a long time for the particles to redistribute in phase space in order to give rise to a clear structure. At the beginning of the simulation (the time scale being set by the damping time), there is no hint of structure, and the phase space distribution is essentially gaussian. This is the situation relevant to the muon collider, since the cycle time is a tiny fraction of the damping time. Furthermore, these results are obtained in the zero-bunch-length approximation, and there are indications [26] that a nonzero bunch length strongly suppresses coherent effects. Although more research is needed, and experimental work under controlled conditions needs to be carried out to confirm the simulation results, we can safely conclude from presently available information that these effects are unlikely to appear in the muon collider.

Beam Tails and Beam-Beam Lifetime.

While the beam core determines the luminosity of a collider, the beam tails determine the beam lifetime. The beam core, typically understood to be the phase space region within $\sim 3\sigma$ of the center, is not very sensitive to nonlinear machine resonances because the lattice magnetic fields are typically quite linear in this region. On the other hand, the beam tail

extends out to sufficiently large amplitudes so that its dynamics is sensitive to an interplay [27] of beam-beam and lattice nonlinearities (beam-gas scattering can also come into play in subtle ways, although typically it has a clearer effect on the beam core).

There has been much recent progress in understanding and properly simulating the beam tails in e^+e^- colliders. These new tools make use of a clever algorithm by which the brute-force tracking is "accelerated" by 2-3 orders of magnitude in order to determine the particle density and flux at large amplitudes (up to $\sim 20\sigma$ or so) [28, 29]. From the particle flux one can then determine the lifetime, given the physical aperture. For the purposes of this article, one can roughly summarize the conclusions as follows: for a lattice with larger dynamic aperture (meaning 5σ or more), and for reasonable values of the beam-beam parameter (meaning 0.05 or less), it is not difficult to find working points for which the beam-beam lifetime is of the order of $10^7 - 10^9$ turns (however, the lifetime can degrade by several orders of magnitude by relatively small changes in these parameters). In any case, the instability of the muon will almost certainly dominate the beam lifetime, so at least from this perspective, the beam tails will not be important.

Thus the beam tails might be much more important for other reasons such as background and radiation. The important thing, therefore, is to specify the maximum acceptable number of muons that can hit the vacuum chamber during the 1000 turns of a cycle. Such a criterion is closely related to that of the dynamic aperture. In the above-mentioned e^+e^- simulations, the damping time, typically of order $10^3 - 10^4$ turns, also plays an important role. The muon collider, as mentioned earlier, essentially has no damping, so in this respect it is akin to proton colliders. It seems therefore that the tracking tools used to determine the dynamic aperture of such machines are the right ones for this case, provided they are augmented to include a beam-beam element. Such a code development should be relatively simple, although the analysis will likely involve many iterations.

Hourglass Effect for Centered Collisions.

Because of the geometrical divergence of the beams at the IP, the luminosity is actually smaller than the nominal value given by Eq. (8.8), which represents the limiting value as the bunch length $\sigma_z \to 0$. As σ_z grows at fixed β^*, the luminosity decreases due to this "hourglass effect." Neglecting all dynamical effects, this purely geometrical reduction factor is given, for symmetric round gaussian beams, by the formula [30]

$$\frac{\mathcal{L}(\sigma_z)}{\mathcal{L}(0)} = \int_{-\infty}^{\infty} \frac{dt}{\sqrt{\pi}} \frac{e^{-t^2}}{1 + (t/t_x)^2} = \sqrt{\pi}\, t_x\, e^{t_x^2}\, \mathrm{erfc}(t_x) \tag{8.11}$$

where $t_x \equiv \beta^*/\sigma_z$. For the muon collider, t_x has been chosen to be unity; β^*/σ_z] It can be shown that the luminosity degrades quickly as σ_z increases.

Hourglass Effect for Longitudinally-Displaced Collisions.

By virtue of the hourglass effect, the luminosity also degrades if the bunches collide at a point away from the optical IP. If the central collision is longitudinally displaced by a distance s_c from the IP (but the bunches still collide transversely head-on), the luminosity reduction factor is given by [30]

$$\frac{\mathcal{L}(s_c,\sigma_z)}{\mathcal{L}(0,0)} = \int_{-\infty}^{\infty} \frac{dt}{\sqrt{\pi}} \frac{e^{-(t-t_z)^2}}{1+(t/t_x)^2} \tag{8.12}$$

where $t_z \equiv s_c/\sigma_z$ and t_x is the same as above. In Fig. 8.20 we show the luminosity reduction

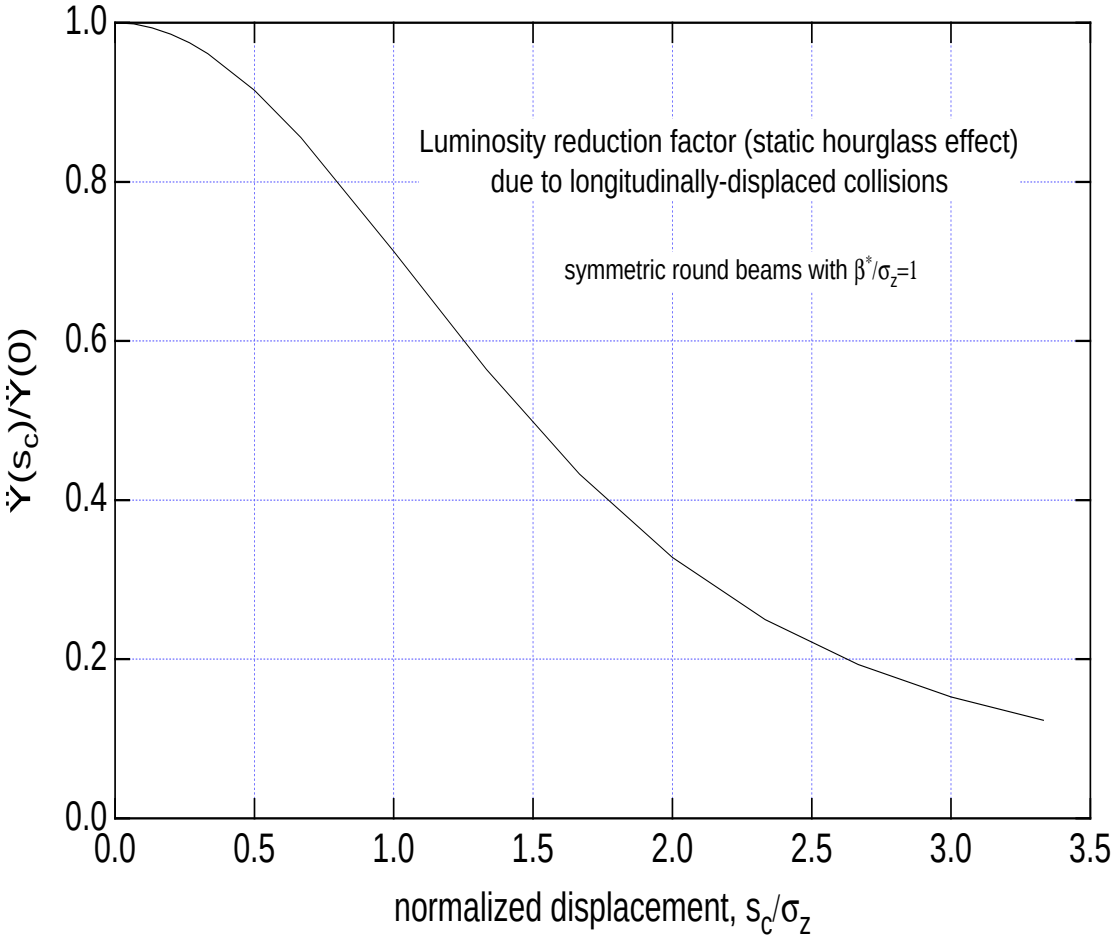


Figure 8.20: The hourglass luminosity reduction factor when the collisions are longitudinally displaced from the IP by a distance s_c, plotted as a function of s_c/σ_z. Note that the normalization is $\mathcal{L}(s_c = 0, \sigma_z = 3$ mm), *i.e.* the nominal value.

factor as a function of t_z (please note that in this figure we have normalized the reduction factor to $\mathcal{L}(0, \sigma_z)$ and *not* to $\mathcal{L}(0, 0)$ as in Eq. (8.12)). One can see that the luminosity degrades quickly when the collision point is farther away than $\sim 1\sigma_z$ from the optical IP. This gives an idea of the RF phasing errors or injection timing errors that can be tolerated.

8.6.5 Summary

We have carried out some basic beam-beam simulations for the muon collider in the incoherent classical approximation, taking into account the muon decay. We conclude that the beam-beam interaction is reasonably weak and allows room for upgrading the luminosity performance. We have argued that classical coherent beam-beam effects will almost certainly not materialize, and neither will long tails that might spoil the beam lifetime. We have provided some very basic constraints on the ratio β^*/σ_z and on off-center collisions arising from the hourglass effect. Finally, we will attempt to prioritize work that remains to be done.

From the perspective of the classical beam-beam dynamics, the four key features that distinguish the muon collider are:

1. A relatively modest beam-beam parameter, $\xi = 0.046$.

2. A short cycle of 1000 turns.

3. A long damping time, $\tau_x = 2 \times 10^6$ turns.

4. Unstable muons.

The first feature is shared with many e^+e^- colliders; the second makes this collider not too different from single-pass colliders; the third one makes it resemble a hadron collider; and the fourth, of course, is unique to this machine. It is fair to say that one can understand all features of the classical beam-beam interaction from the interplay of these four characteristics.

We have shown by means of beam-beam simulations that the classical incoherent beam-beam effect is quite weak for the muon collider in its present design. From this perspective, there is room for upgrading the luminosity, if necessary, by increasing the bunch current by a factor of 2 or so but not more than this.

We have argued that coherent classical beam-beam effects are very unlikely to materialize.

We have also argued that beam tails are unlikely to affect the luminosity lifetime. Undoubtedly there will be a certain number of large-amplitude muons that will hit the chamber,

and it seems important to establish this number. This issue is closely related to the determination of the dynamic aperture, and single-particle tracking tools used for hadron colliders, duly augmented to include the beam-beam interaction, seem appropriate to address this issue.

From purely geometrical considerations, we have provided a rough estimate (probably a lower bound) of the sensitivity of the luminosity to the ratio β^*/σ_z and to the longitudinal displacement of the collision point from the IP. These estimates yield fairly standard results: one should not choose the ratio β^*/σ_z below ~ 1 or so, and one should not allow collisions to be displaced from the optical IP by more than $\sim 1\sigma_z$ in either direction.

Much work remains to be done to firm up the limits imposed by the beam-beam interaction. Here is a brief suggested list, roughly in order of priority:

1. Develop a dynamic aperture tool by augmenting a single-particle tracking code to include a "beam-beam lens."

2. Track specific lattices, including all nonlinearities, and estimate from the results the number of muons that hit the vacuum chamber during 1000 turns; iterate this process to determine tolerances on machine nonlinearities.

3. Establish the sensitivity of the beam-beam interaction to longitudinal and transverse alignment errors and jitter.

8.7 QED Effects at the Interaction Region

A significant source of e^+e^- pairs in the $\mu^+\mu^-$ collider is the *incoherent* process $\mu^+\mu^- \to \mu^+\mu^-e^+e^-$ which has a cross section about 10 mb (at a muon energy 2 TeV). An analysis of the energy distribution of the pairs shows that they may generate a significant and undesirable background in the detector. The bulk of this Section is taken from [32]. The detector design chapter addresses this in further detail.

Coherent pair production in the muon collider occurs as an individual muon interacts with the magnetic field of the opposing bunch at the IP. It is now believed to be an insignificant effect in the parameter regime of the collider. [32].

Since the heat deposition in the collider creates a limit on the number of muons that can hit the wall each turn, we must investigate the energy spectrum of those muons that create pairs, and follow their trajectories. This must be done for muons that pair produce at the IP and for those that decay naturally during the ≈ 1000 turn storage. The characteristic

magnetic field of bunch at the collision point is, for the parameters of Table 8.1,

$$B \sim \frac{eN}{2\sigma_x\sigma_z} \sim 0.5 \times 10^3\ T. \tag{8.13}$$

The asymptotic cross section for the production of e^+e^- pairs at the collision of charged particles was first obtained 60 years ago [33]. A precise treatment of this process, along with numerous references, are can be found in Ref. [34].

A detailed analysis of the main features of the pair production process $pp \to ppe^+e^-$, with numerical estimates, has been considered in detail [35]. This process is similar to the pair production in the muon ring. The equations in Ref. [35] are valid for the muon collider ring, with the evident changes in notation, and some numerical estimates should be reconsidered due to a much higher Lorentz factor in the muon collider.

Some rough estimates of the process are:

1) The total cross section of the process is [36]:

$$\sigma \approx 10\ mb. \tag{8.14}$$

2) The main mechanism for pair creation is the two–photon production of e^+e^- pairs via the collision of two virtual photons which are emitted by the muons. The main features of process can be obtained with the equivalent photon (Weizsäcker–Williams) approximation (EPA). The spectra of virtual photons are found from an analysis of Feynman diagrams. Their dependence on the energy ω and virtuality Q^2 has the form

$$dn(\omega, Q^2) = \frac{\alpha}{\pi}\frac{dx}{x}\frac{dQ^2}{Q^2}\left[1 - x + \frac{x^2}{2} - (1-x)\frac{Q^2_{min}}{Q^2}\right]; \tag{8.15}$$

$$Q^2 \geq Q^2_{min} = m_\mu^2\frac{x^2}{1-x}. \tag{8.16}$$

where $x = \omega/E$ and the last inequality is obtained easily from kinematics.

The cross section for the subprocess $\gamma\gamma \to e^+e^-$ has a maximum when $W^2 \approx 8m_e^2$, where W is the effective mass of the produced system. Therefore, the effective mass of the produced e^+e^- system is near threshold, and the transverse momenta of produced particles is $\sim m_e$. Furthermore, this cross section decreases quickly with the growth of the virtuality Q^2 above m_e^2. In other words, the main contribution to the cross section is in the region

$$m_e^2 > Q^2 > Q^2_{min}.$$

Now, with Eq. (8.16), we obtain the limit for the energy of the photons:

$$x = \frac{\omega}{E} < \frac{m_e}{m_\mu} \Rightarrow \omega < \gamma m_e. \tag{8.17}$$

Therefore, — in accordance with the *naive* expectations — the Lorentz–factor of the produced e^+e^- pair cannot be higher than that of the initial muon.

An integration of Eq. (8.15) over virtuality gives the number of equivalent photons that are produced:

$$dn(\omega) = \frac{2\alpha}{\pi}\frac{dx}{x}\left[\ln\left(\frac{m_e}{m_\mu x}\right) - \frac{1}{2}\right]. \tag{8.18}$$

Note, that this quantity is much lower than that for the two-photon production of muons or hadrons, which is

$$\sim (2\alpha/\pi)(dx/x)\ln(1/x). \tag{8.19}$$

The source of this difference is the much higher upper limit of effective virtualities for these processes, which determined by the much higher scale of the Q^2 dependence for these subprocess. (For more detail discussion see Ref. [34].)

3) The produced pairs are distributed uniformly in the rapidity scale. The distribution over the total energy of pair ϵ is

$$d\sigma = \begin{cases} \frac{56\alpha^4}{9\pi m_e^2}\frac{dk_z}{\epsilon}\left[\ln^2\gamma - \ln^2\frac{\epsilon}{m_e}\right] & \text{at} \quad \epsilon < \gamma m_e; \\ \sim \frac{\alpha^4}{m_e^2}\frac{d\epsilon}{\epsilon}\left(\frac{\gamma m_e}{\epsilon}\right)^2 & \text{at} \quad \epsilon > \gamma m_e. \end{cases} \tag{8.20}$$

Here k_z is the longitudinal momentum of the pair, $|k_z| \approx \epsilon$.

The mean energy of a pair is $\sim 2m_e\gamma/\ln\gamma \sim 2$ GeV. In accordance with Eq. (8.14), the number of pairs produced is about 10^5 per bunch collision, i.e. about 10^8 during the luminosity lifetime in the ring. Therefore, the entire energy losses due to the discussed process are about $2 \times 10^{-6}\%$ of the muon energy, i.e. negligible.

4) The distribution over the energy of one electron ϵ_1, emitted along the motion of initial μ^+, has the form

$$d\sigma = \frac{56\alpha^4}{9\pi m_e^2}\frac{dk_{1z}}{\epsilon_1}\left[\ln^2\gamma - \ln^2\frac{\epsilon_1}{m_e}\right] \quad (\epsilon_1 \gg m_e). \tag{8.21}$$

Here k_{1z} is the longitudinal momentum of the electron. The effective mass of produced pairs is near the threshold and their total transverse momentum is very low. Therefore, the bulk of the electrons and positrons move initially in the same direction as the beam.

However, as was pointed out by Palmer [37], the created electrons are deflected by the magnetic field of the opposing muon bunch and may fly into the detector. Detailed studies of this potentially serious problem are being carried out by the detector group.

8.8 Single Bunch Collective Effects

Introduction

The design of a high luminosity ($2.5 \times 10^{30}\text{cm}^{-2}$ per collision) muon collider ring, from the perspective of the physics of collective effects, has some unique features which need to be examined. (1) Muons have a very short life time: $\tau_\mu \simeq 41.6$ ms at 2 TeV, corresponding to a thousand "effective" turns in a ring with the circumference of 7 kilometers. (2) The bunch is short: $\sigma_z = 3$ mm. (3) The bunch has a large charge: $N = 2 \times 10^{12}$. (4) The momentum compaction α is very small: $\alpha \leq 10^{-6}$. These features lead us to some unusual aspects of the ring operation: The intense bunch required for the high luminosity makes instabilities likely and very small α requires careful estimations of nonlinear corrections to the particle orbit and to the collective dynamics.

The longitudinal equations motion of a particle in a circular machine are

$$z' = -\eta\delta, \tag{8.22}$$

$$\delta' = K(z), \tag{8.23}$$

where z is the oscillation amplitude with respect to the bunch center, $' = d/ds$, s measures distance around the ring, $\delta = dp/p$, $\eta = \alpha - 1/\gamma^2$, $\alpha = pdC/Cdp$. The force $K(z)$ that a particle experiences can be modeled as having two parts, one is due to the radio frequency (rf) cavities, and the other is from the wake fields generated by the interaction between beam and cavities or other components of its electromagnetic environment,

$$K(z) = K_{rf}(z) + K_{wake}(z), \tag{8.24}$$

where

$$K_{rf}(z) = \frac{eV_{rf}(z)}{CE}, \tag{8.25}$$

and

$$\begin{aligned} K_{wake}(z) &= \frac{F_{\parallel wake}(z)}{E} \\ &= -\frac{r_0}{\gamma C}\int_z^\infty dz'\rho(z')W_0'(z-z'). \end{aligned} \tag{8.26}$$

In Eq. 8.26, $T_0 = 2\pi/\omega_0 = C/c$, $E = \gamma m_\mu c^2$, $r_0 = e^2/m_\mu c^2$, $N = \int dz'\rho(z')$, and $C = 2\pi R$ is the circumference of the collider ring.

In Eq. 8.25, when the amplitude of synchrotron motion is small compared with the rf wavelength such that the rf voltage is linearized as $V_{rf}(z) = \hat{V}\sin(\omega_{rf}z/c) \approx \hat{V}\omega_{rf}z/c$ and

the momentum compaction is expanded as $\eta = \eta_1 + \eta_2\delta + \eta_3\delta^2$ (with the contributions from η_2 and η_3 negligible), the synchrotron oscillation frequency is $\omega_{s0} = \left(e\eta_1\, c\,\omega_{rf}\hat{V}/CE\right)^{1/2}$.

The transverse equation of motion is

$$y''(z,s) + \frac{\omega_\beta^2}{c^2}y(z,s) = -\frac{Nr_0}{\gamma C}\int_z^\infty dz'\rho(z')W_1(z-z')y(z',s). \tag{8.27}$$

A simulation code that solves these equations has been developed to study collective phenomena in the ring.

Static Effects

As a starting point we consider a TESLA-like rf system, and a quasi-isochronous lattice which has $\eta_1 = 10^{-6}$, with the contributions from η_2 and η_3 negligible. With a bunch length $\sigma_z = 3$ mm, $\sigma_\delta = 1.5 \times 10^{-3}$ rms energy spread and an 130MV of L-band rf, the muon bunch is matched to the rf and a muon would undergo $\sim .56$ synchrotron oscillations in one thousand turns. The beam-loading factor, defined as $\triangle\mathcal{E}$/particle/V_{rf}, is 10% when only the cavity losses are included. Resistive losses are estimated to roughly equal the rf losses, but have not been explicitly included in our calculation. Other losses have not been calculated, and may lead to an increase in the rf voltage. These loses will need to be replenished even if the momentum compaction is reduced to $\eta_1 = 10^{-7}$, as may be required because of microwave instabilities.

Parasitic Loss The beam loses energy when it experiences the impedance of the rf cavities. The rf impedance is modeled by using Wilson's scaling formula for the longitudinal wake function [15]. Explicitly, choosing as starting point TESLA's rf structure, we use

$$W_0'(z<0) = 226(\frac{f_{TESLA}}{f_{SLAC}})^2 \times \exp[-(|z|\frac{f_{TESLA}}{0.1839 f_{SLAC}})^{0.605}], \tag{8.28}$$

where $f_{TESLA} = 1.3$GHz and $f_{SLAC} = 2.856$GHz. Causality requires: $W_0'(z) = 0$ for $z > 0$. The simulation code computes the wake voltage $V_{wake}(z) = \int_z^\infty dz'\rho(z')W_0'(z-z')$, and the energy loss

$\triangle\mathcal{E} = -\int_{-\infty}^\infty dz\rho(z)\int_{-\infty}^\infty dz'\rho(z')W_0'(z-z')$.

The peak wake voltage is further scaled from Wilson's formula to give parasitic cavity losses 17V/pC/m at $1\sigma_z$ for a Gaussian beam. This choice of wake amplitude makes it consistent with the TESLA rf cavity studies[16].

The effects of radiation and diffusion of muons are small in a muon lifetime (radiation damping time $\sim 10^6$ turns), so that, unlike in electron rings, equilibrium is not achieved by radiation damping. The intense muon bunch generates significant wakes, and these wakes

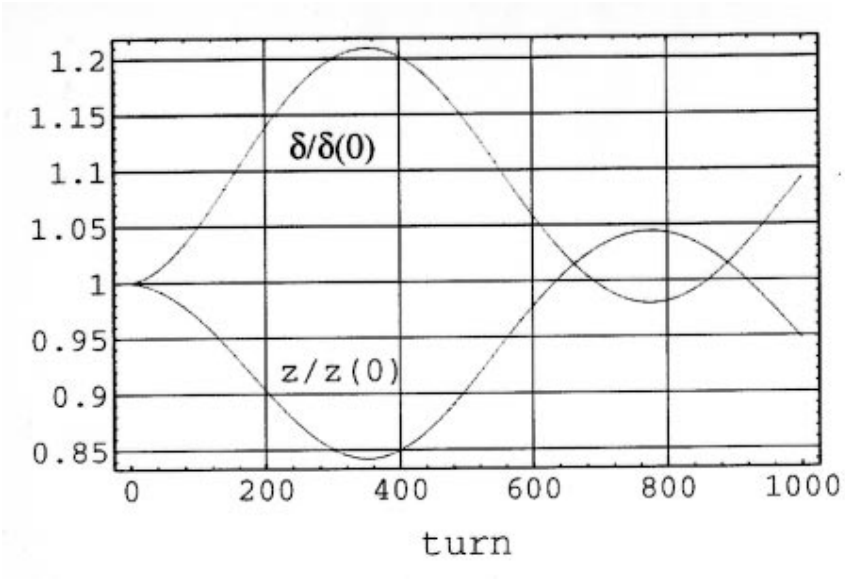


Figure 8.21: Scaled rms bunch size and rms energy-spread vs. turn, where at injection: $z(0) = 3\text{mm}$, $\delta(0) = 0.15\%$.

in turn cause significant changes in the vacuum rf potential. This potential-well distortion causes oscillation of the bunch center, bunch size, and distribution function in the rf bucket. Fig. (8.21) shows the oscillations of the rms bunch size and bunch energy spread. The bunch centroid tends to move forward to a higher rf voltage, so that the energy loss can be compensated. As a result, it makes a counter-clockwise rotation in $\delta - z$ phase space, as shown in Fig. (8.22). The parasitic losses from the rf and motion of the bunch centroid are compensated for by injecting the beam with an rf phase offset of 0.082 radians with respect to the bunch center, as shown by the dashed line in Fig. (8.22).

Coherent Effects

Microwave Instability The microwave instability is presently considered the most serious challenge to maintaining a short bunch. Presently studies are underway to examine the limits this instability places upon Z/n. The ring parameters obviously will not satisfy the Keil-Schnell criterion for stability, but rather hope to reduce the growth rate to an acceptable amount during the ~ 1000 turns of beam storage. The microwave instability growth rate is weaker at smaller η values, and the instability may require the lattice to operate at $\eta = 10^{-7}$. At this value of η the particles barely move longitudinally, and the possibility and consequences of compensating for the wake potential with rf are being considered. In the absence of longitudinal motion the main problem is to maintain an energy spread within the longitudinal acceptance of the ring.

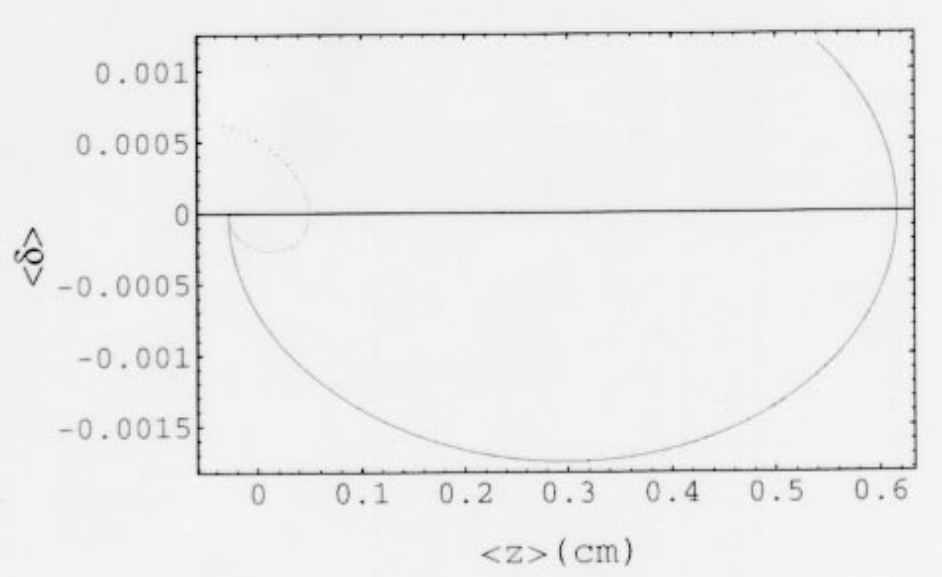


Figure 8.22: Center of energy-spread vs. center of bunch size. The rf phase offsets are, $\phi = 0, 0.082$ radian, for the solid line and dashed line, respectively.

Beam-Break-Up For times much shorter than the synchrotron oscillation period, particles are almost frozen longitudinally in the bunch, and the transverse wakefield dynamics has many similarities with that in a linac [10, 40]. In a linac, the transverse wake field generated by the head of the bunch drives the tail, causing Beam-Break-Up (BBU). A dimensionless parameter that characterizes the BBU strength is [17]

$$\Upsilon(z) = Nr_0|\langle W_1(z)\rangle|c/4\omega_\beta\gamma, \tag{8.29}$$

where ω_β is the betatron angular frequency, and $\langle W_1(z)\rangle = \int_z^\infty dz' W_1(z-z')\rho(z')$ is the convoluted total transverse wake function. The tail of an off-axis bunch doubles its offset in a number of turns $n \simeq 1/\Upsilon$, so long as $n \ll 1/2\nu_s$, i.e., when the particle's synchrotron motion can be ignored.Here ν_s is the synchrotron tune.

Simulation results for the BBU-like instability using a resonator model are shown in Fig. (8.23). The main point is that while the motion is unstable, it is easily cured with only a small amount of BNS damping, as discussed below.

Head-Tail Instabilities When the transverse oscillation frequency is modulated by the energy oscillation, the chromaticity, which is the slope of the frequency to the energy, builds up a head-tail phase that bootstrap from the first half synchrotron period to the next, and drives the system into instability without threshold. This head-tail instabilities occurs in both transverse and longitudinal motion [41]. The effect of transverse head-tail (THT) instability is small when $\eta_1 \leq 10^{-6}$.

For the longitudinal motion, the longitudinal chromaticity involves the non-linear part of slip factor: η_2. The bucket height and the growth time of the longitudinal head-tail (LHT) instability are both proportional to η_1/η_2. Different design of the lattice lead to very different results for the geometry of bucket and the collective effects. It is assumed here that the contributions of η_2 and η_3 to the dynamics are sufficiently small, even if $\eta_1 = 10^{-7}$, that they can be neglected. Simulations indicate that the longitudinal head-tail instability can be controlled by not allowing η_2/η_1 to become too large. Detailed studies of the acceptable parameter ranges are underway.

BNS Damping

Since the synchrotron radiation radiation damping is negligible and the ring is quasi-isochronous (so that the effect of Landau damping is very small), neither of these are likely to may damp collective instabilities. "External" mechanisms, such as BNS damping, may be needed to stabilize the system. The BNS damping can be achieved by a radio frequency quadrupole (RFQ), which introduces a betatron tune spread across the bunch such that the bunch tail experiences a larger betatron focusing than the bunch head [17]. Fig. (8.23) shows that the BBU-like instability is stabilized when a small BNS tune spread is applied to the beam. One should note that, the BNS damping works for the ring only when the potential-well distortion is compensated by rf phase offset, such that the bunch shape remains approximately stationary. This is because the amount of BNS tune spread obtained from the prescribed formula $\triangle\nu_\beta(z)/\nu_\beta = \Upsilon(z)/\pi\nu_\beta$, involves the bunch's density profile. To maintain the correct BNS detuning condition, the bunch shape should not seriously deviate from its initial state; otherwise, one needs to adjust the BNS tune spread accordingly. Investigations are underway to determine if such an rf quadrupole is feasible. In addition, the transverse chromaticity, which causes betatron tune spread, may provide some Landau damping of the instability.

Conclusions

Various single bunch collective effects have been examined. The longitudinal microwave instability is, at present, seen to be the greatest threat to maintaining the bunch length. Operation at $\eta_1 = 10^{-7}$ is being considered, along with ideas for compensating the energy variation induced by the longitudinal wake. The transverse strong head-tail instability with the small η is seen to be BBU-like and can be stabilized by BNS damping. Other instabilities are not believed to be severe over the short storage times.

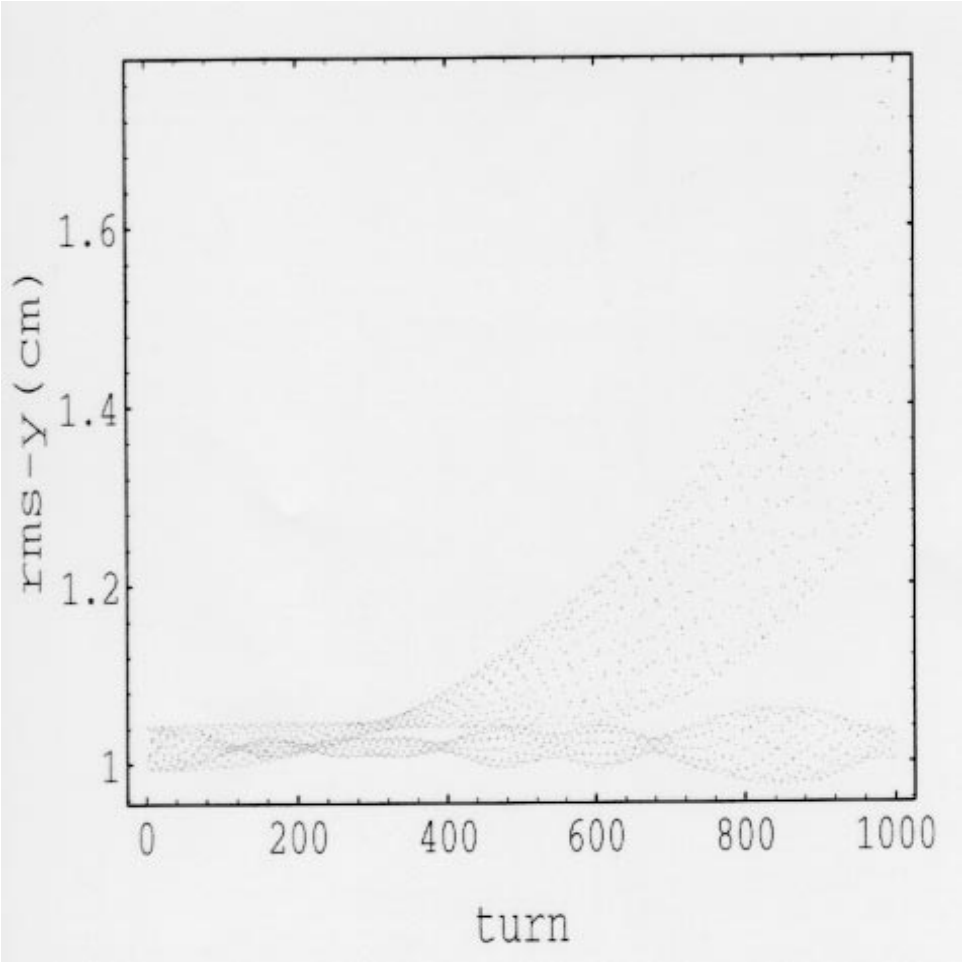


Figure 8.23: Blow up of the rms beam size, due to the BBU-like effect, where $1/\nu_s = 1784$. Note that $R_s/Q = 18225(\Omega)$, $b_{cavity} = 1.3$cm are used for the resonator model, injection error $\Delta y = 0.2$cm, and $\Upsilon(1\sigma_z) \simeq 0.017$. After BNS damping is applied, $\triangle\nu_\beta(1\sigma_z)/\nu_\beta \simeq 6 \times 10^{-5}$, the beam size fluctuates only slightly around 1 cm, a nominal injection beamsize.

Bibliography

[1] A. Garren, et al., *Design of the Muon Collider Lattice: Present Status*, to appear in Proc. of the San Fransisco Workshop on the Physics of Muon Colliders, 1996.

[2] N.M. Gelfand, *A Prototype Lattice Design for a 2 TeV $\mu^+\mu^-$ Collider*, Fermilab Report TM-1933, 1995; King-Yuen Ng, presentation at the 9th Advanced ICFA Beam Dynamics Workshop, Montauk, New York, Nov. 1995; D. Trbojevic, et al., *Design of the Muon Collider Isochronous Storage Ring Lattice*, Proceedings of the Micro Bunches Workshop, Eds. E. Blum, M. Dienes and J. Murphy, Conference Proceedings 367 (1996) ; C. Johnstone and K.-Y. Ng, *Interaction Regions for a Muon Collider* , ibid.

[3] Juan C. Gallardo and Robert B. Palmer, *Final Focus System for a Muon Collider A Test Model*, to appear in Proc. of the San Fransisco Workshop on the Physics of Muon Colliders, 1996.

[4] S. Y. Lee, K. Y. Ng and D. Trbojevic, *Minimizing dispersion in flexible-momentum-compaction lattices*, Phys. Rev. E **48**, 3040 (1993)

[5] R. B. Palmer, *Beam Dynamics in a Muon Collider* Beam Dynamics Newsletter, No.8 (1994) 27, Eds. K. Hirata, S. Y. Lee and F. Willeke.

[6] F. Zimmermann, et al., *A Final Focus System for the Next Linear Collider*, SLAC-PUB-95-6789, June 1995, presented at PAC95, Dallas, Texas, May 1-5, 1995; to be published; O. Napoly, *CLIC Final Focus System: Upgraded version with increased bandwidth and error analysis*, DAPNIA/SEA 94 10, CLIC Note 227, 1994.

[7] K. Brown, *A conceptual Design of Final Focus Systems for Linear Colliders*, SLAC-PUB-4159, (1987).

[8] R. Brinkmann, *Optimization of a Final Focus System for Large Momentum Bandwidth*, DESY-M-90-14, 1990.

[9] K. Johnsen, *Effects of Nonlinearities on Phase Transition*, Proc. CERN Symposium on High Energy Accelerators and Pion Physics, Geneva, Vol.**1**, 106 (1956).

[10] K.Y. Ng, *Higher-Order Momentum Compaction for a Simplified FODO Lattice and Comparison with SYNCH*, Fermilab Internal Report FN-578, 1991.

[11] L. Schachinger, private communication.

[12] N. V. Mokhov, "The MARS Code System User's Guide, version 13 (95)", FNAL–FN–628 (1995).

[13] N. V. Mokhov and S. I. Striganov, *Simulation of Backgrounds in Detectors and Energy Deposition in Superconducting Magnets at $\mu^+\mu^-$ Colliders*, Proc. of the 9th Advanced ICFA Beam Dynamics Workshop, Ed. J C. Gallardo, Conference Proceedings 372 (1996); also Fermilab–Conf–96/011 (1996).

[14] Perry Wilson, SLAC, private communication).

[15] Perry Wilson, SLAC-PUB-2884 (Rev.), "High Energy Electron Linacs", 1991, p. 77)

[16] A. Mosnier, DAPNIA-SEA-92-06, CE Saclay.

[17] Alex Chao, "Physics of Collective Beam Instabilities in High Energy Accelerators", Eqn. 2.194, p. 118.

[18] A. Mathewson, E. Alge, O. Grobner, R. Souchet and P. Strubin, J. Vac. Sci. Technol. A5, 2512 (1987).

[19] A. Mathewson, C. Reymermier and S. Zhang, CERN Vac. Tech. Note 95-22 (1995).

[20] M. Furman, *The Classical Beam-Beam Interaction for the Muon Collider: a First Look* , BF-19/CBP-Note-169/LBL-38563, April 1996.

[21] D. Rice, *Observations of the Beam-Beam Effect in PEP, SPEAR and CESR*, Proc. Third Advanced ICFA Beam Dynamics Workshop (Beam-Beam Effects in Circular Colliders), I. Koop and G. Tumaikin, eds., Novosibirsk, May 29–June 3, 1989, p. 17.

[22] S. Krishnagopal and R. Siemann, *Coherent Beam-Beam Interaction in Electron-Positron Colliders*, Phys. Rev. Lett. **67**, pp. 2461–2464 (1991).

[23] R. Siemann, *The Beam-Beam Interaction in e^+e^- Storage Rings*, SLAC-PUB-6073, March 1993; Proc. Joint US-CERN School on Particle Accelerators: *Frontiers of Particle*

Beams, Factories with e^+e^- rings, Benalmádena, Spain, October, 1992, pp. 327–363 (Springer Verlag Lecture Notes in Physics no. 425, M. Dienes, M. Month, B. Strasser and S. Turner, eds.).

[24] J. L. Tennyson, undocumented code *TRS*, 1989.

[25] M. A. Furman, *Beam-Beam Issues in Asymmetric Colliders*, LBL-32561/UC-414/ESG-205/ABC-77, July 1992, invited talk presented at the Conference on B Factories: The State of the Art in Accelerators, Detectors and Physics, Stanford, CA, April 6–10, 1992 (proceedings, p. 109); invited talk presented at the Washington APS Meeting, Washington, DC, April 18–23, 1992.

[26] R. H. Siemann, private communication.

[27] D. Shatilov and A. Zholents, *Lifetime and Tail Simulations for Beam-Beam Effects in PEP-II B Factory*, to be published in the Proc. 1995 Particle Accelerator Conference and International Conference on High Energy Accelerators (Dallas, Texas, May 1–6, 1995); LBL-36484.

[28] J. Irwin, *Simulation of Tail Distributions in Electron-Positron Circular Colliders*, SLAC-PUB-5743, February 1992; Proc. Third Advanced ICFA Beam Dynamics Workshop (Beam-Beam Effects in Circular Colliders), I. Koop and G. Tumaikin, eds., Novosibirsk, May 29–June 3, 1989, p. 123.

[29] T. Chen, J. Irwin and R. Siemann, *Simulation of the Beam Halo from the Beam-Beam Interaction*, SLAC-PUB-6379, October, 1993; Phys. Rev. **E49**, March 1994, p. 2323.

[30] This is an old subject; a relatively recent reference is: M. A. Furman, *Hourglass Effects for Asymmetric Colliders*, LBL-30833, Proc. 1991 Particle Accelerator Conf., San Francisco, May 6–9, 1991, p. 422.

[31] R.B. Palmer. Beam Dynamics Newsletter. # 8 (1995); Report at 3-rd Int. Conf. *Physics Potential & Developments of $\mu^+\mu^-$ Colliders*, San Francisco, December 13–15, 1995.

[32] I. Ginzburg, *The e^+e^- pair production at $\mu^+\mu^-$ collider.*, submitted for publication.

[33] L.D. Landau, E.M. Lifshitz. Phys. Zs. Sovjet **6** (1934) 244.

[34] V.M. Budnev, I.F. Ginzburg, G.V. Meledin and V.G. Serbo. Phys.Rep. **15C** (1975) 181.

[35] V.M. Budnev, I.F. Ginzburg, G.V. Meledin, V.G. Serbo. Phys. Lett. **39B** (1972) 526; Nucl.Phys. **B63** (1973) 519.

[36] C. Racah. Nuovo Cim. **14** (1937) 93.

[37] R.B. Palmer. Private communication.

[38] A.I. Nikishov, V.I. Ritus. Sov. ZhETF **46** (1964) 776.

[39] K.Y. Ng, *Beam Stability Issues in a Quasi-Isochronous Muon Collider*, Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop, Ed. J. C. Gallardo, AIP Press, Conference Proceedings 372 (1996).

[40] W. H. Cheng, A. M. Sessler, and J. S. Wurtele, *Studies of Collective Instabilities, in Muon Collider Rings*, Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop, Ed. J.C. Gallardo, AIP Press, to be published.

[41] C. Pellegrini, Nuovo Cimento 64A, 447 (1969). B. Chen and A. W. Chao, Particle Accelerators, vol. 43(1-2), pp77-91,1993.

[42] W.-H.Cheng, A.M. Sessler and J.S. Wurtele, "Alternating chromaticity: a damping mechanism for the transverse head-tail instability", in prepariation.

Contributors

- J. S. Wurtele, (Univ. California Berkeley) Editor
- W-H Cheng (LBNL)
- E. Courant (BNL)
- M. Furman (LBNL)
- J. Gallardo (BNL)
- A. Garren (LBNL)
- I. Ginzburg (Inst. of Math., Novosibirsk)
- M. Green (LBNL)
- C. Johnstone (FermiLab)
- N. Mokhov (FermiLab)
- K-Y. Ng (FermiLab)
- K. Oide (KEK)
- R. Palmer (BNL)
- J. Peterson (LBNL)
- A. M.Sessler (LBNL)
- D. Trbojevic (BNL)
- W. Turner (LBNL)
- E. Willen (BNL)

List of Figures

List of Tables

Chapter 9

DETECTOR BACKGROUND AND DETECTOR DESIGN

Contents

9.1 Introduction

In this chapter issues relating to the conceptual design of a detector for a 2 TeV $\times$ 2 TeV Muon Collider are discussed. The physics justification for a lepton collider in such an energy range has been extensively studied [1] ,[2]. Here design considerations specific to a muon

collider are considered. From the experimental point of view there are advantages and new challenges that must be faced in the design of the detector and the experimental area.

The detector is designed for a muon collider with $\sqrt{s} = 4$ TeV, an average luminosity of 10^{35} cm^{-2}s^{-1} and an integrated luminosity of 1000 fb^{-1}; beam crossings occur every 10 μs and each muon bunch has 2×10^{12} muons.

The physics benchmark used in determining the criteria for the detector was strong WW scattering resulting in W pairs, Z pairs and top quark pairs. The generic question of tagging b-decays in this environment is also being investigated. The list of physics topics being addressed will be expanded as the design matures but these two topics are potentially significant for this energy regime and provide significant constraints on the detector requirements.

A major advantage of a muon collider is the ability to recirculate the muons without an overwhelming radiative energy loss. The present assessment is that the muons will survive for about 1000 turns before the losses, primarily from muon decays, significantly reduce the luminosity. From a purely experimental point of view there are two notable advantages of muon colliders compared to electron colliders: the center of mass energy of the collisions is better defined, and at high energy the background process arising from double photon interactions is suppressed.

The experimental environment of a muon collider offers a new design challenge as the level of background, arising mostly from the interactions of electrons originating from muon decay to $e\nu_\mu\nu_e$, needs to be better understood and controlled.

In section 9.2 physics processes are discussed and the quantities that need to be measured outlined. In Section 9.3, the particular machine related background problems associated with muon colliders and the present conceptual design of an intersection region are discussed. In Section 9.4 the detector specifications and the conceptual design of a detector are presented. Some very preliminary results on machine backgrounds for a lower energy collider (250 GeV x 250 GeV) are presented in Section 9.5 while in Section 9.6 the present status of the current effort is summarized.

9.2 Physics Aspects of a 4 TeV $\mu^+\mu^-$ Collider

The physics goals of a $\mu^+\mu^-$ collider are discussed in Chapter 2. The larger muon mass does provide some possible advantages:

- s-channel Higgs production is enhanced by a factor of $(m_\mu/m_e)^2 \approx 40000$.
- QED radiation is reduced by a factor of $[\ln(\sqrt{s}/m_\mu)/\ln(\sqrt{s}/m_e)]^2$, leading to smaller $\gamma\gamma$ backgrounds and a smaller beam energy spread.

There are also disadvantages:

- Muon decays give large backgrounds and hence a more difficult environment for the experiments. These problems can be mitigated by careful shielding design.

The physics capabilities of $\mu^+\mu^-$ and e^+e^- colliders with the same energy and luminosity are similar, so that the choice between them will depend mainly on the feasibility and cost of the accelerators.

Possible strongly interacting W_LW_L cross sections from Barger[4] are shown in Fig. 2.13. For the present study a 1 TeV Higgs boson has been used. The signals and backgrounds were calculated using PYTHIA. The studies so far have used a toy detector and have not included machine related backgrounds. As these backgrounds are better understood, they will be included.

Fig. 9.1 shows the mass distribution for the 1 TeV Higgs signals and physics backgrounds from PYTHIA in a toy detector, which includes segmentation of $\Delta\eta = \Delta\phi = 0.05$ and the angular coverage, $20^\circ < \theta < 160^\circ$, assumed in the machine background calculations. Since the nominal luminosity is $1000\,\mathrm{fb}^{-1}$, there are $\gtrsim 1000$ events per bin at the peak. The loss in signal from the 20° cone is larger for this process than for s-channel processes but is still fairly small, as can be seen in Fig. 9.2. The dead cone has a larger effect on $\gamma\gamma \to WW$ and thus the accepted region has a better signal to background ratio.

It would be desirable to separate the WW and ZZ final states in purely hadronic modes by reconstructing the masses. Whether this is possible or not will depend on the details of the calorimeter performance and the level of the machine backgrounds. If it is not, then one can use the $\sim 12\%$ of events in which one $Z \to ee$ or $\mu\mu$ to determine the Z rate. Clearly there is a real challenge to try to measure the hadronic modes.

The background from $\gamma\gamma$ and γZ processes is smaller at a muon collider than at an electron collider but not negligible. Since the p_T of the photons is usually very small while the WW fusion process typically gives a p_T of order M_W, these backgrounds can be reduced by making a cut $p_{T,WW} > 50\,\mathrm{GeV}$, as shown in Fig. 9.3. This cut keeps most of the signal while significantly reducing the physics background. The present calculation is undoubtedly optimistic, since it neglects the degradation in resolution from the machine background, but an effective cut with an acceptable signal loss should be possible.

Tagging the outgoing muons would give an additional handle on separating $WW \to WW$ from $\gamma\gamma \to WW$, $\gamma Z \to WZ$, etc. Even after the $p_T > 50\,\mathrm{GeV}$ cut is made on the final state, most of these muons will be inside the 20° cone, as can be seen in Fig. 9.4. Tagging the muons inside the detector will be difficult. However, the muons will have $\gtrsim 1\,\mathrm{TeV}$, so

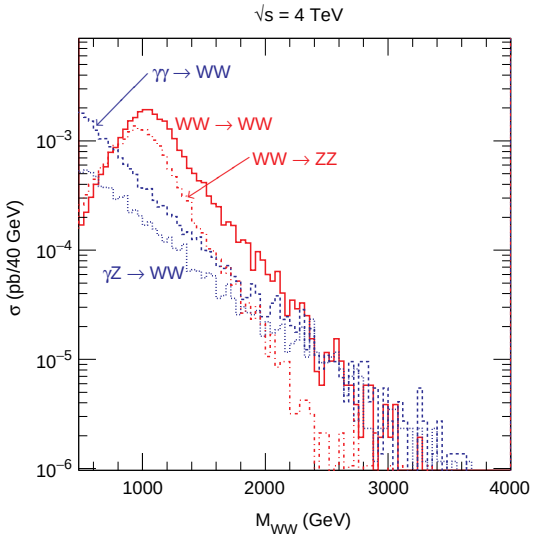


Figure 9.1: Signals and physics backgrounds for a 1 TeV Higgs boson at a $\mu\mu$ collider, including the effect of a 20° dead cone around the beamline.

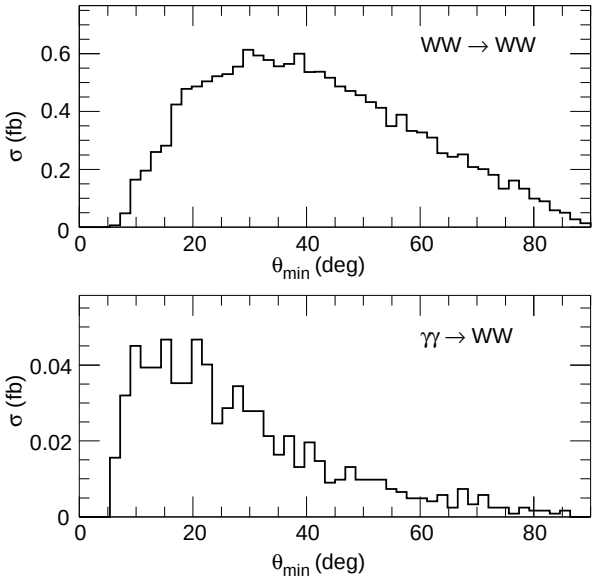


Figure 9.2: $WW \to WW$ signal and $\gamma\gamma \to WW$ background vs. the minimum angle, $\theta_{\min}$, of the W.

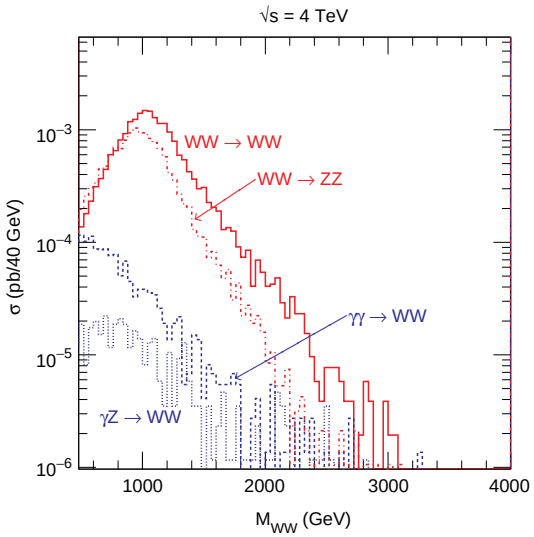


Figure 9.3: Signals and physics backgrounds for a 1 TeV Higgs boson with a cut $p_{T,WW} > 50\,\text{GeV}$.

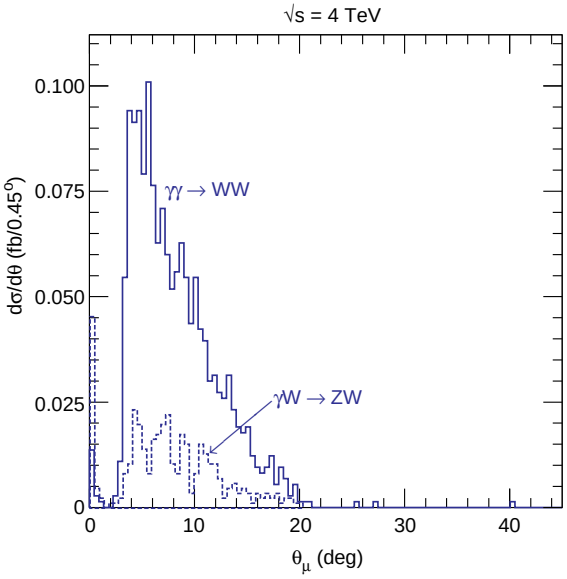


Figure 9.4: Angle of outgoing muon in $\gamma\gamma \to WW$ and $\gamma Z \to WZ$ processes after making the $p_{T,WW} > 50\,\text{GeV}$ cut described in the text.

they penetrate the cone easily and can in principle be tagged behind it. Whether this is possible in practice will, of course, depend on the muon halo, which is still being studied.

This physics simulation work is still in its infancy. As the machine related backgrounds are better understood they will be added to the simulation to see if the analysis strategy discussed above is still reasonable.

9.3 Machine Induced Backgrounds

9.3.1 Overview

A primary concern for a muon collider detector design is the level of the machine related backgrounds. There are three logically distinguishable sources; the muon halo, muon decays giving rise to high energy electrons and beam-beam interactions.

The term **muon halo** refers to those muons which are lost from the beam bunches as the muons traverse the whole collider ring. In conventional lower energy accelerators beam particles which are lost away from the intersection region are of little concern as they can be locally shielded. However 2 TeV muons can traverse kilometers of shielding so beam particle losses, almost anywhere around the machine, have the potential to generate background in the detector. The background from this source depends on a detailed knowledge of the beam profile and a credible model for beam halo and beam losses. This is not available at present. What is discussed below is a tool which is being developed to trace muons not only inside the beam pipe but also to continue this tracking outside the beam into the magnet fringe fields and shielding material. This tool will be very useful when a more detailed knowledge of the beam is known. It will eventually be necessary to devise muon spoilers and shielding for the muon halo.

With 2×10^{12} muons per bunch in the machine there are approximately 2×10^5 **muon decays** per meter giving rise to high energy electrons. The momentum distribution of these electrons follows the usual three body decay kinematics. These off-energy, off-axis electrons undergo bremsstrahlung when they traverse magnetic fields and when exiting the beam pipe interact and produce high energy electromagnetic showers and to a lesser extent hadrons and muons. Much of this debris can be locally shielded so the primary concern is muon decays near the intersection region. In the current lattice design there is a pre-focus for the beam about 130 m from the intersection point. In the discussion below it is assumed that all debris from muon decays outside this final 130 m "straight section" can be locally shielded and only decays occurring in the straight section will contribute to the machine related detector background. Much of the detector related work has been to optimize the

shielding of this final focus section and this will be discussed below. Detailed simulations of electromagnetic showers and neutron and hadron production were carried out to study the effects of various shielding strategies.

In an e^+e^- collider there is a substantial background from **beam-beam** interactions. Due to the larger muon mass, this contribution is expected to be less for a muon collider. This background contribution is considered in Chapter 8.5. Recent calculations by I.F.Ginzburg[5] and by P. Chen[6] have shown that coherent effects are not important. The classical beam-beam effects provide a smaller background than the other topics discussed here.

Since this is the first attempt to design a high energy muon collider there is a real question of how to verify the results on the backgrounds discussed here. This is compounded by the fact that there is not enough computer power to follow all the muon decays and their interactions. In an attempt to mitigate concern two very different approaches are being taken to study these problems.

One approach uses the MARS code [10] which performs fast inclusive simulations of hadronic and electromagnetic cascades, muon and low energy neutron transport in accelerator, detector and shielding components of arbitrary complexity. It has been used extensively for two decades in numerous applications at the Tevatron, UNK, SSC and LHC colliders, and for SDC, GEM, D0 and CMS detectors (for recent work see [11], [12]). The MARS code has been used from the very beginning of the muon collider study to analyze backgrounds and detector performance [11],[13],[14] as well as for calculating energy deposition in superconducting magnets, pion production, targetry and radiation issues (see other chapters). In MARS simulations all the particle interaction processes in the lattice with the detailed 3-D dipole and quadrupole geometry and magnetic field maps, in the 1.45 m radius tunnel surrounded by the soil/rock (ρ=2.24 g/cm^3), in the 26 m long and 10 m radius experimental hall and in the detector are taken into account. For background studies a 275 m section of the interaction region is considered. A single MARS run includes:

- forced $\mu \to e\nu\bar{\nu}$ decays in the beam pipe (beam muon decay studies) or beam halo interactions with the limiting aperture beam pipe;

- tracking of created electrons in the beam pipe under the influence of the magnetic field with emission of the synchrotron photons along the track;

- simulation of electromagnetic showers in the collider and detector components induced by electrons and synchrotron photons hitting the beam pipe, with appropriate hadron and prompt muon (Bethe–Heitler pairs and direct positron annihilation) production;

- simulation of muon interactions (bremsstrahlung, direct e^+e^- pair production, ionization, deep inelastic nuclear interactions and decays) along the tracks in the lattice, detector, tunnel and experimental hall components and air, and in the surrounding soil/rock;

- simulation of electromagnetic showers created in the above muon interaction vertices;

- simulation of hadronic cascades generated in muon and photon interactions, with daughter electromagnetic showers, with muon production (π and K decays, prompt muons in hadronic and electromagnetic interactions), and with low–energy neutron transport;

- histogramming and analysis of particle energy spectra, fluence and energy deposition in various detector and collider regions.

Energy thresholds are 1 MeV for muons and charged hadrons, 0.3 MeV for electrons and photons, and 0.5 eV (0.00215 eV in some cases) for neutrons.

The second approach provides a more detailed look at shower development and sources. It uses GEANT, the standard CERN high energy physics simulation tool to do all the particle tracking and it has been augmented, where necessary, with the appropriate physics processes for this muon collider study. EGS[7] was used to give full shower development for electromagnetic showers. GEANT-3.21 was used with FLUKA[8] to propagate hadronic showers and high energy γ hadron interactions and with the MICAP option[9] to transport low energy neutrons. This approach allowed for detailed studies of specific problems and a step by step approach to developing a satisfactory shielding strategy. Much of the discussion below follows the logic of the latter approach with the MARS conclusions provided for comparison purposes.

9.3.2 The Muon Halo

Beam particles which stray from the muon bunch anywhere around the collider ring, may propagate through the accelerating chamber walls, magnet yokes and other surrounding structures, and eventually reach the detector to contribute to the background. Halo muons decay and produce synchrotron radiation and secondaries, part of the overall detector background. The only practical way to control stray muons is by deflection with magnetic fields, scraping and cleanup will be done on the opposite side of the ring from the detector. Clearly to get as clean a machine as possible the bunches will have to be prepared carefully.

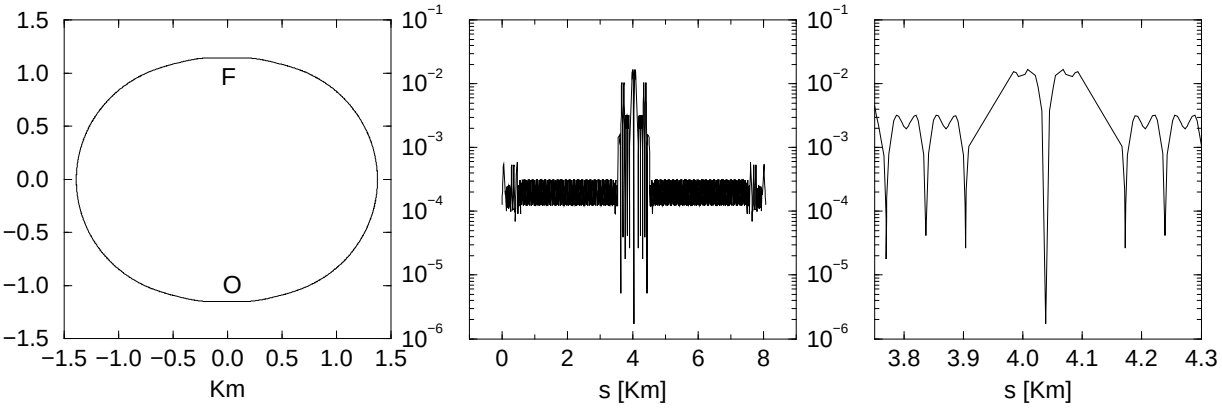


Figure 9.5: a: reference orbit for tracking from O to the final focus F in the Garren lattice. b: horizontal envelope [m] of the muon beam calculated with *Mad* for an emittance of10π $mm - mrad$. c: beam envelope [m] near the final focus.

First results on numerical tracking of the muon halo are presented here. No attempt was made to track secondaries, to optimize magnetic shields, or represent a realistic beam halo. The tracking was done by numerical integration of the equations of motion of the muons through the magnets of the collider, represented by field maps. The machine lattice is based on the work of D. Trbojevic, J. Gallardo and finally Al Garren. It will be referred to as the Garren lattice. The collider is a ring 8 km in circumference, Fig. 9.5a, with beta functions varying in wide limits in the final focus (FFC) region, F in the figure. Fig. 9.5b,c show the horizontal envelope of the beam for a normalized emittance of 10π $mm - mrad$. The lattice optics were calculated with the *Mad* code[3].

A distribution of particles in phase space is propagated starting at O in Fig. 9.5a. The algorithm used is a quasi-symplectic Predictor-Corrector with variable step; the starting step was 1 mm. The integration proceeded through field maps of dipoles and quadrupoles (super conducting RHIC type), both in the gap and in the coil and yoke regions. A third type of magnet, toroids, to be used as magnetic shield, was included for future use. This calculation assumed a field strength of $\sim$9 T.

There is a tracked reference particle, a muon starting on axis at O. The angle of deflection of this particle through the dipoles, compared with the dipole angle of *Mad*, provided the field calibration. Two reference systems of coordinates were used: system (i) relative to the reference orbit and system (ii) relative to the laboratory. System (i) shows tracked particles in the vicinity of the machine, while system (ii) was mainly used to see whether a stray particle falling out of the machine would eventually reenter it. System (i) was rotated at each dipole end, and successive dipole map slices were staggered to follow the reference

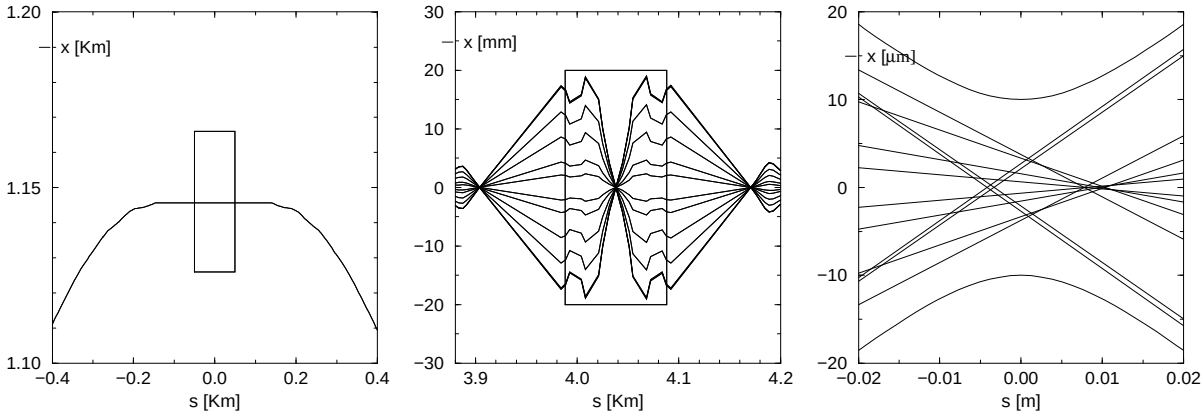


Figure 9.6: a: Reference orbit detail for the final focus showing a 20 m × 20 m detector box. b: Colliding muons (accepted) in the box. Some orbits are shown ; step every 0.15 μrad in the starting angle c: Blow-up of the central part of the previous figure showing the final focus.

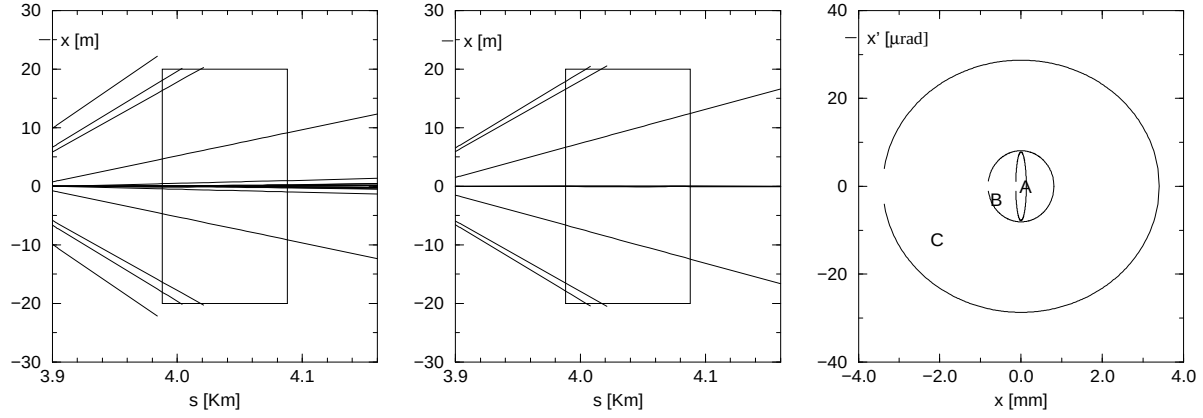


Figure 9.7: a,b: Final focus orbits for background muons relative to the detector box, for various initial angles and positions. c: Phase space acceptance (and halo). B are accepted, C are *not* accepted.

trajectory. It is an unusual tracking. Generally, one tracks particles *inside* an accelerator. Here, the tracking is mostly on the *outside*.

The Final Focus (FFC) region with a nominal 20 m $\times$ 20 m detector box is shown in Fig. 9.6a. Fig. 9.6b shows the trajectories of muons that are accepted by the lattice and will propagate in the collider. These well behaved muons produce the beam at the intersection point for which the machine is being designed. Fig. 9.6c is a blow-up of the FFC, with trajectories and their envelope shown.

Fig. 9.7a,b shows some badly behaved muons, i.e. muons that propagate through magnet yokes and coils and escape the lattice in the FFC, where the fast oscillations of the machine functions shake them off. These muons impinge on the detector box. They are the background halo.

The initial muon distribution used in the calculations is shown in the diagram of Fig. 9.7c, in the horizontal plane (a similar one holds for the vertical). The A ellipse represents the machine acceptance calculated from *Mad* beta for a normalized emittance of $10\pi\ mm-mrad$. The B ellipse contains the well behaved muon beam of Fig. 9.6b. The region between the contours of B and C contains the bad halo muons of Fig. 9.7a,b. Muons outside C are soon lost for good, generally when they encounter the first difficulties with the lattice bends.

Clearly the work discussed here has the potential to significantly contribute to the detailed design of the muon collider. With the addition of muon energy loss and muon decay probabilities, this tool will help not only to predict detector backgrounds from halo muons but also to aid in the design of muon spoilers to reduce the potential halo background in the detector.

9.3.3 Design of the Intersection Region

The current design of the Intersection Region is driven by the desire to reduce the background from muon decays in the detector as much as possible. For this study the 130 m final focus section (Fig. 9.8) which includes the final four quadrupoles, three toroids, a 2 Tesla solenoidal field for the detector and the connecting beam pipe and shielding was modeled in GEANT (Fig. 9.9) with all the appropriate magnetic fields and shielding materials. Studies of the effects of high energy electrons hitting specific edges and surfaces were carried out and the shielding adjusted or augmented to mitigate the apparent effects of particular background problems. Effects due to electrons, photons, neutrons and charged hadrons and muons were considered in turn to try to optimized the design. While the current design is not fully optimized, it is a marked improvement over a much simpler design which had been used in the past. More importantly, it helped develop the tools and strategy to do such an

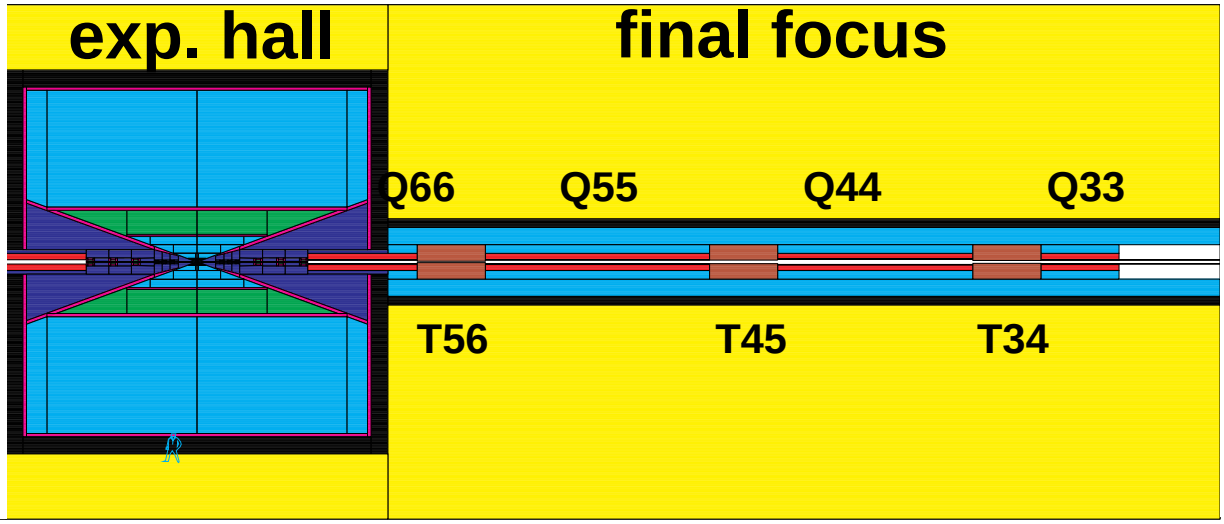


Figure 9.8: Region around the Intersection Region modeled in GEANT. The black regions represent tungsten shielding. The final quadrupoles (Q) and toroids (T) on one side of the detector enclosure are shown. The shaded areas around the intersection point represent the various detector volumes used in calculating particle fluences.

optimization as the lattice is further developed.

The final focus may be thought to be composed of 3 separate regions. The longest of these, from 130 m to approximately 6.5 m contains the quadrupole magnets which bring the beam to the final focus in the intersection region. The space available between the four quadrupoles was used to install toroids. They fulfill a double role: first they are used as scrapers for the electromagnetic debris; secondly, they serve as magnetic deflectors for the Bethe–Heitler(BH) muons generated upstream. The effect of the toroids on the BH muons will be discussed later. In order to optimize the inner aperture of the toroids, the σ_x and σ_y envelope of the muon bunch at every exit of the quadrupoles has been estimated. The inner aperture of each toroid was chosen to match the 4 σ ellipse of the muon bunch at that point. For example, Fig. 9.10 represents the the xy envelope of the muon bunch at the exit of Q55. Similarly, Fig. 9.11 shows the same distribution for the decay electrons. Fig. 9.12 and Fig. 9.13 show the y projection of the above envelopes.

The second region, from 6.5 m to 1.1 m contains tungsten plus additional shielding boxes to help contain neutrons produced by photons in the electromagnetic showers (Fig. 9.14). A shielding box consists of a block of Cu surrounded by polyboron. The shielding here is designed with inverted cones to reduce the probability of electrons hitting the edges of collimators or glancing off shielding surfaces (Fig. 9.15). The beam aperture at the entrance to this section is reduced to 2.5 cm and by the exit of the section to 4.5 mm. This profile

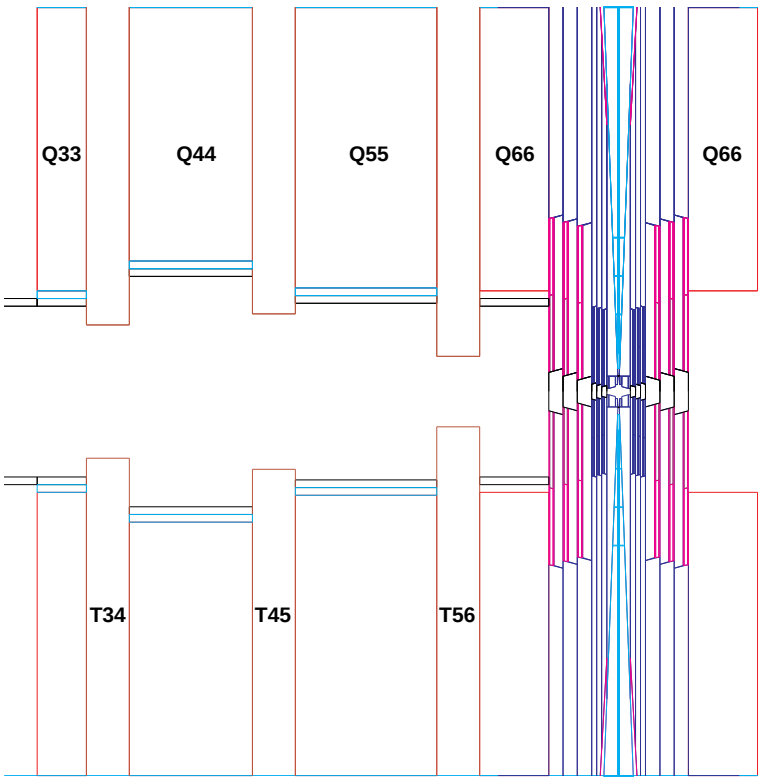


Figure 9.9: GEANT description of the Intersection Region. The horizontal scale is 100 m while the vertical is only about 5 cm. The shapes in the figure are the actual GEANT surfaces used in the simulation. Q represents a quadrupole and T a toroid magnet.

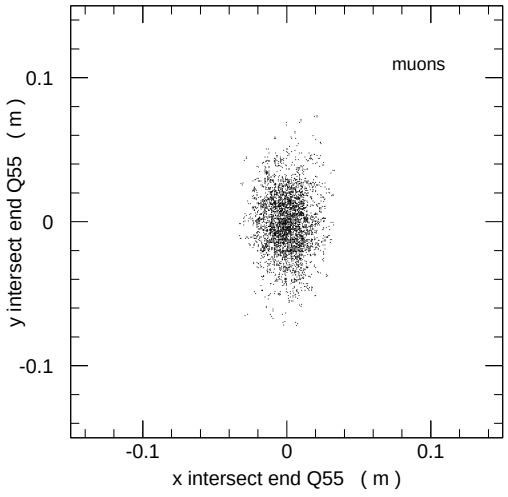


Figure 9.10: The xy envelope of the muon bunch at the exit of Q55.

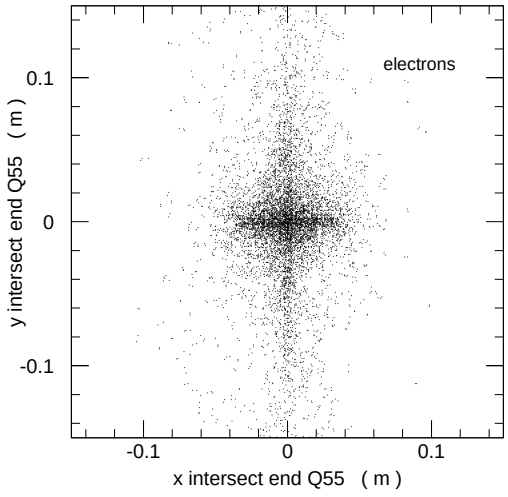


Figure 9.11: The xy envelope of the decay electrons at the exit of Q55.

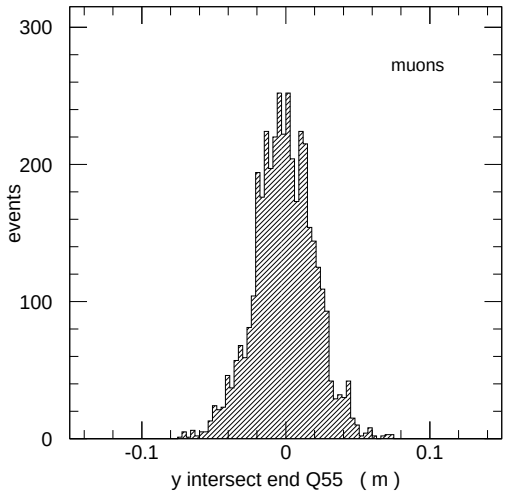


Figure 9.12: The y projection of the muon bunch at the exit of Q55.

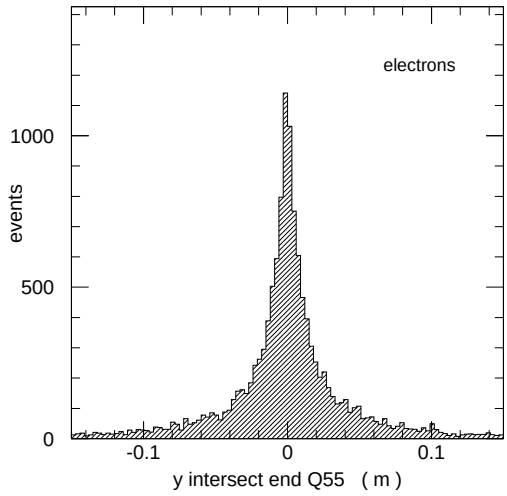


Figure 9.13: The y projection of the decay electrons at the exit of Q55.

follows the beam envelope as the particles approach the intersection region. The intersection region itself (Fig. 9.16) is designed as an inverse cone to prevent electrons which reach this region from hitting any shielding as this region is directly viewed by the detector. Approximately 2% of the electrons from muon decays in the final 130 m around the intersection point interact with shielding in the intersection region, 30% in the adjacent region, 58% in the outermost region where the final quadrupoles exist and the other 10% pass right through the region without hitting any shielding.

A 20° tungsten cone around the intersection region is required for the reduction of the electromagnetic component of the background. The cone is lined, except very near the intersection region with polyboron to reduce the slow neutron flux. In the shielding calculations it is also assumed that there is a polyboron layer before the calorimeter and surrounding the muon system. In earlier designs this cone was only 9°. Whether or not the full 20° is required is still under study and work is ongoing to evaluate the physics impact of this choice of the shielding cone angle. It is likely that, after optimization is completed, the cone angle will be reduced.

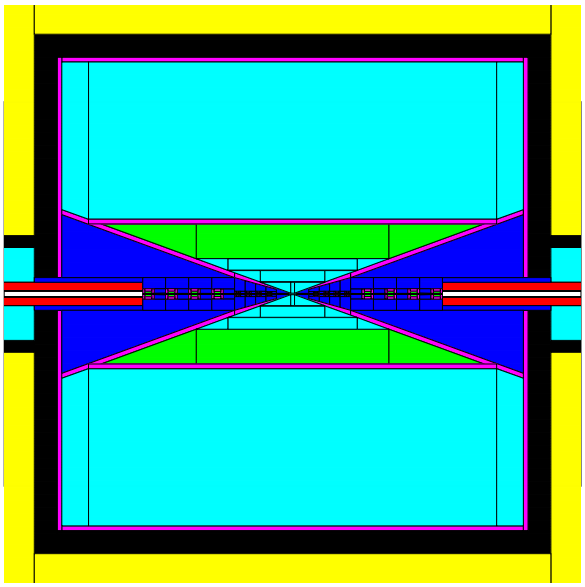


Figure 9.14: Detail near Intersection region. The final 10 meters on either side of the intersection point are shown. The dark regions are tungsten while the detector volumes are the shaded regions around the intersection point. The lighter region around the tungsten represent the polyboron for neutron absorption. The sections to reduce the neutrons can be seen as the boxes along the beamline decreasing in size as they approach the intersection point.

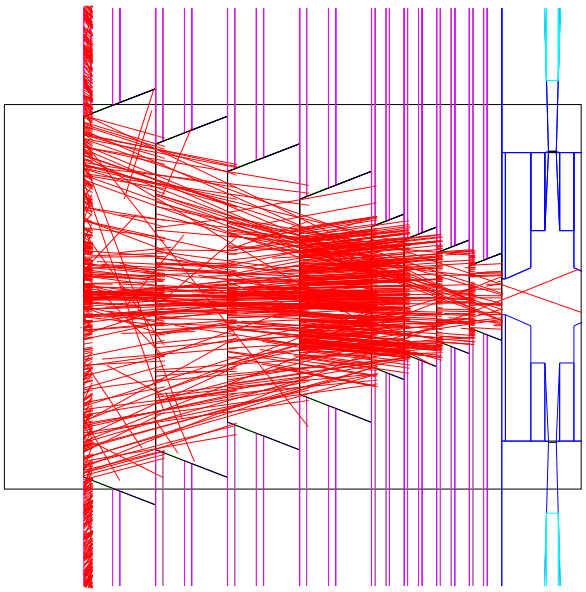


Figure 9.15: Expanded View of Region (2) near the Intersection point. The lines represent electrons from a random sample of muon decays.

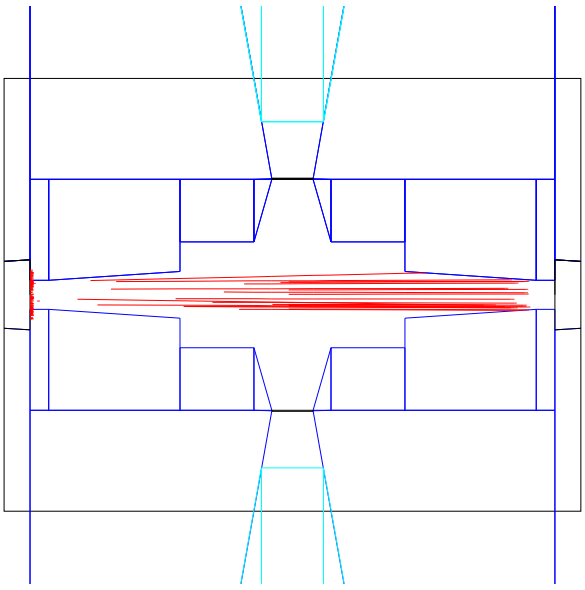


Figure 9.16: Detailed View of Region (1), the Intersection Region. The lines represent electrons from a random sample of muon decays.

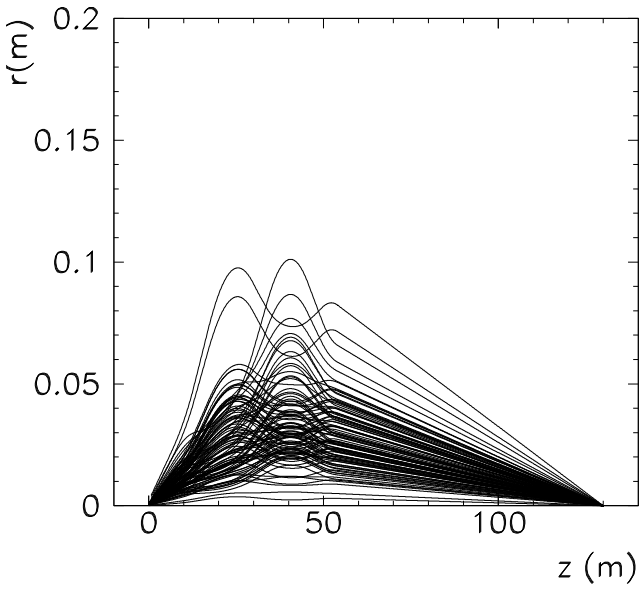


Figure 9.17: Muon Trajectories in the Final Focus Region with Muon decays turned off

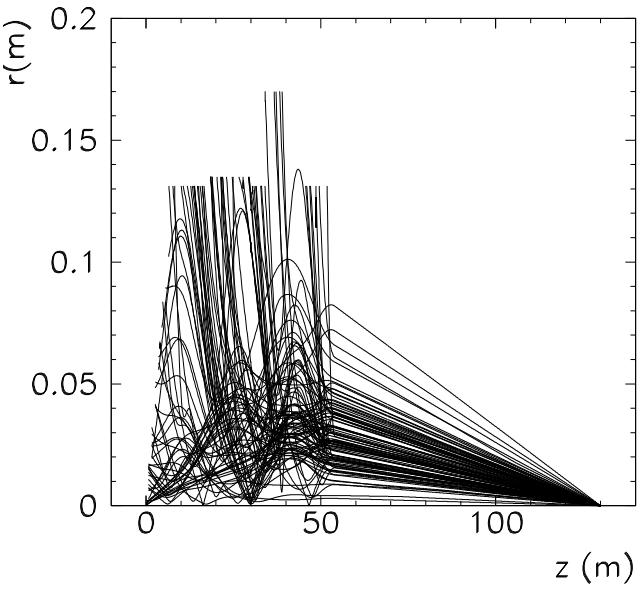


Figure 9.18: Muon Trajectories in the Final Focus Region with Muon Decays allowed. The decay electrons are tracked until they reach either a magnet or shielding.

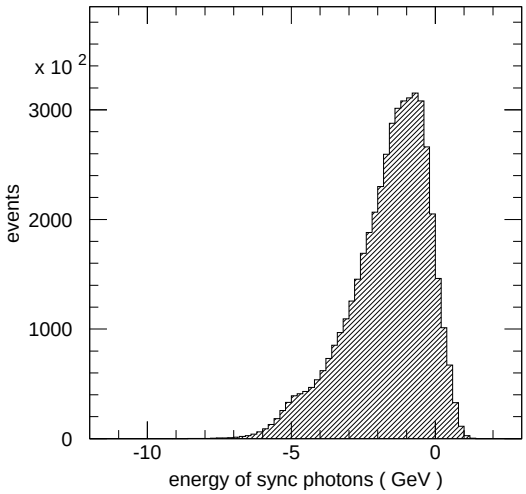


Figure 9.19: Log of the energy of synchrotron radiation photons.

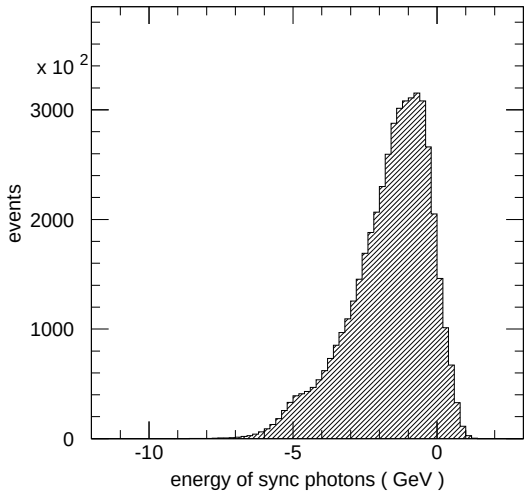


Figure 9.20: Log of the critical energy.

9.3.4 Muon Decay Backgrounds

Results using GEANT Simulation

The primary source of machine related background in the detector is muon decays in the straight section near the intersection region. The trajectory of muons, assuming they do not decay, approaching the final focus is shown in Fig. 9.17. When decays are included (Fig. 9.18) the decay electrons essentially follow the parent muon trajectory in the field free region but when they enter the final focus quadrupoles they are sprayed off-axis and eventually hit the magnets, beam pipe and other shielding.

These high energy electrons shower to produce not only electrons and photons but to eventually produce to a much lesser extent neutrons and other charged and neutral hadrons and even muons.

The backgrounds in the detector are defined as the fluence of particles (number of particles per cm^2 per beam crossing) across surfaces which are representative of the various kinds of detectors which might be considered. For this study the calorimeter was assumed to be a composition of copper and liquid argon in equal parts by volume which represents a good resolution calorimeter with approximately 20% sampling fraction. The other volumes of the detector were vacuum. The calorimeter starts from a radius of 150 cm and is 150 cm deep. The tracker volume is defined from 50 to 150 cm . An array of horizontal and vertical planes were placed in the detector volumes. These planes were used for flux calculations; their

positions are evident in the tables of rates below.

A number of specific backgrounds have been considered at this point. **Bremsstrahlung** from the primary muons has been calculate and, as one would suspect, makes little contribution to the final background in the detector.

The **electromagnetic showers** generated by the decay electrons are fully tracked through GEANT. The present simulation has an electron and a photon cutoff at 25 keV. The mean photon energy is 1 MeV. Due to the 2 T field, all electrons are confined to a very small radius. The soft photon flux has a small probability of producing uncorrelated hits in the tracking chambers, but mostly deposits energy in the electromagnetic calorimeter. As one can see from Fig. 9.18, the electrons can reach rather large distances from the beam axis, resulting in substantial synchrotron radiation in the high field regions of the quadrupoles. Every electron radiates on average 300 photons. The average critical energy is 700 MeV and the average energy of the photons is 500 MeV. The respective distributions can be seen in Fig. 9.19 and Fig. 9.20. The total energy carried by the synchrotron radiation amounts to 20% of the initial average energy of the electron. Because of the small energy carried by the synchrotron photons and the fact that they do not point well toward the small opening of the shield at the intersection region, the contribution of the synchrotron radiation photons is small.

One electron generated background of concern is **Bethe-Heitler muon pairs**. Even though the pairs are in general created near the initial electron impact point, the muons can penetrate the shielding to reach the detector. The photo-pair production of muons by electrons impinging on heavy targets were simulated according to Y.S. Tsai[15]. The probability of an electron to generate a muon on a thick tungsten target is shown in Fig. 9.21 as a function of electron energy. The average muon momentum produced by a 500 GeV electron is 17 GeV and the distribution is shown in Fig. 9.22. The momentum distribution of the generated Bethe Heitler muons in the whole final focus region is shown in Fig. 9.23. The average momentum is 27 GeV and the average probability for an electron to produce such a muon is $5.6{\times}10^{-4}$.

Fig. 9.24 shows the GEANT tracing of 50 Bethe–Heitler muons. All the magnetic fields were active: the quadrupoles, the field in the return yoke of the quadrupoles, the three toroids with peak B-ϕ fields of 4, -4, and 4 T. The experimental cavity had a solenoidal field of 2 T. Ionization losses occur in all the materials (metal, the ground, concrete walls, etc...) and decays are allowed. While the beneficial role of the toroids is clear, lower energy muons are bent away from the vertex and the tracker volumes, optimization will be needed.

Many Bethe–Heitler muons will cross the calorimeter and catastrophic bremsstrahlung losses could cause spikes in the energy distribution. Fig. 9.25 shows the energy distribution

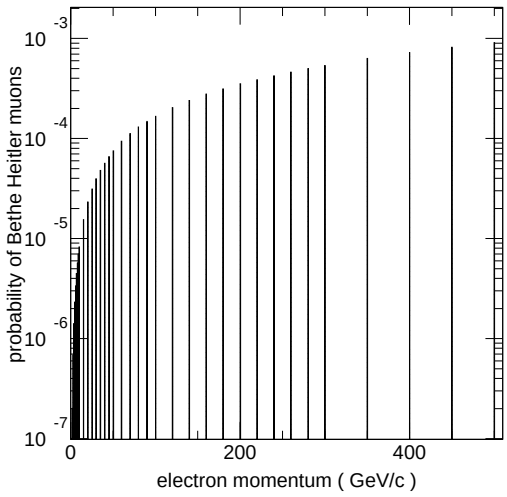


Figure 9.21: Probability of an electron to generate a muon on a thick tungsten target.

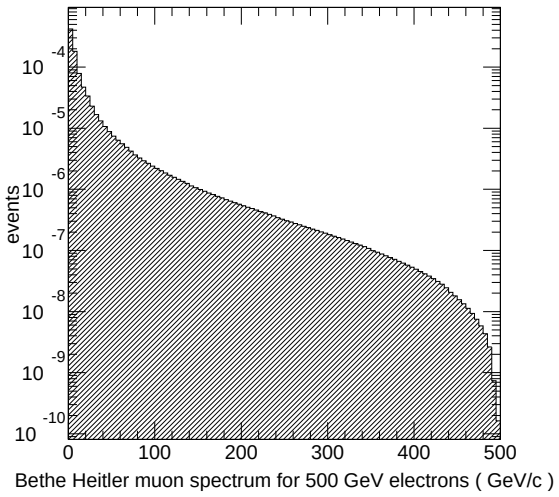


Figure 9.22: Muon momentum spectrum produced by a 500 GeV electron.

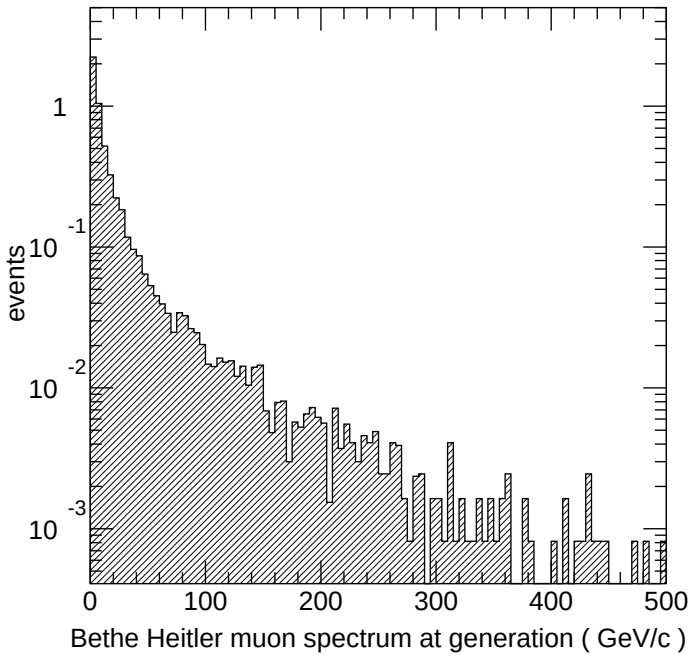


Figure 9.23: Bethe-Heitler muon spectrum in the final focus region.

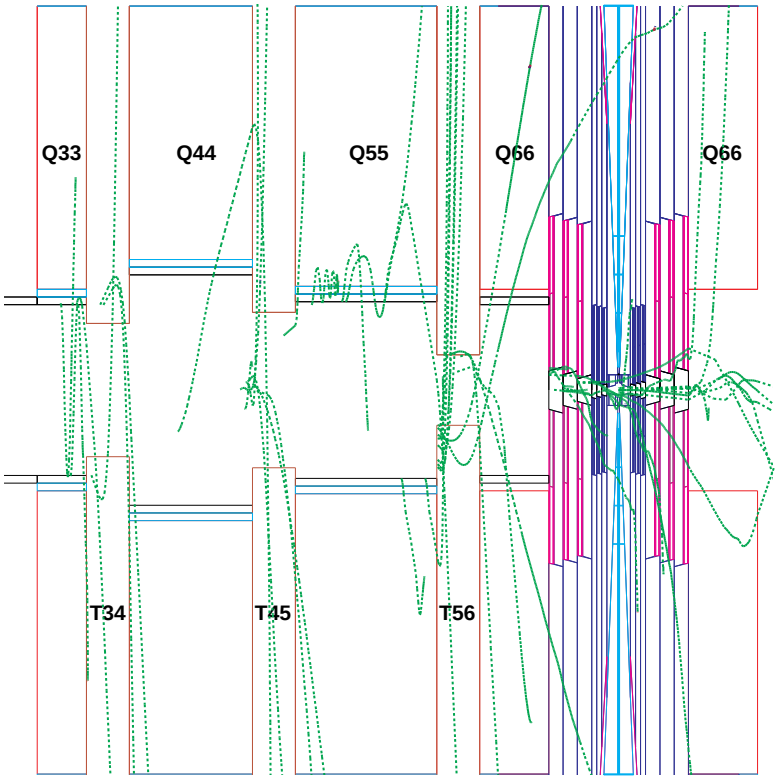


Figure 9.24: Tracing of Bethe-Heitler muons by GEANT.

in the calorimeter, as a function , integrated over r, of z and ϕ, due to Bethe–Heitler muons in a single crossing of 2 bunches of 10^{12} muons each. The energy spikes can be easily identified by the abnormal longitudinal profile of the shower. Nevertheless, it will impose constraints on the calorimetric performance which can be achieved. It is clear that good longitudinal segmentation will be required.

A major effort has been directed towards understanding the **hadron background** in the detector. The hadrons come primarily from photon interactions in the shielding. While the probability of photoproduction of hadrons is quite small relative to other processes the large number of photons released per crossing and their high initial energies make this an important background issue. The relevant photon energy range and the relative cross sections for different processes are illustrated in Fig. 9.26. In the few MeV region (5 MeV - 150 MeV) the preferred models have the photon interacting with the nucleus as a whole (Giant Dipole Resonance) or with a component of the nucleus (Quasi-Deuteron Region). In both these regions the photon shakes free one or more (more than one about 20% of the time) neutrons. In the present model only one neutron is released with the appropriate energy and angular distribution. In practice, just as many protons are released, but since their kinetic energy is low, they are ignored at present. The region from 150 MeV to about

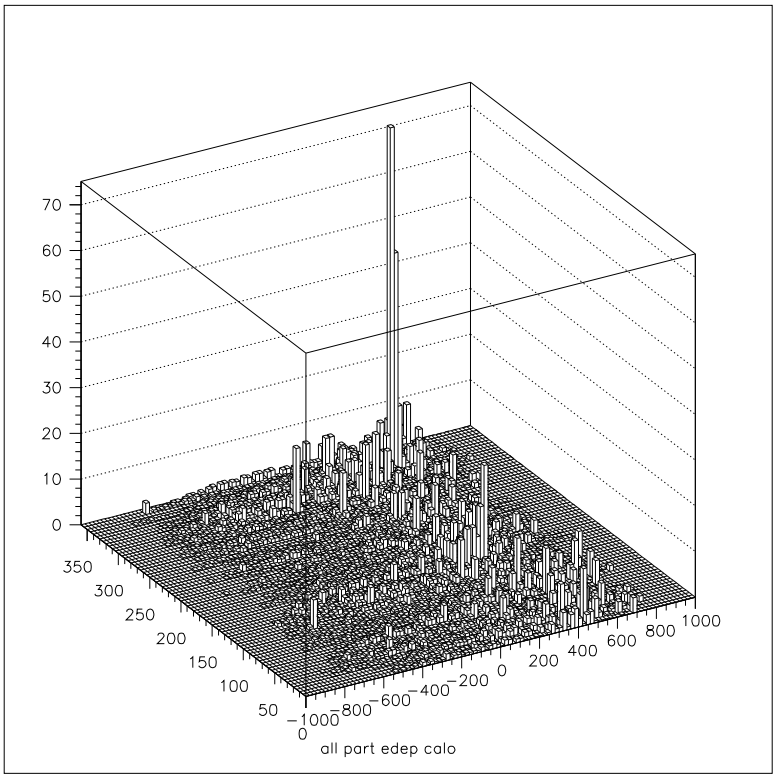


Figure 9.25: Energy distribution of Bethe–Heitler muons as a function of ϕ and z in the calorimeter. Only muons from interactions from one beam are shown.

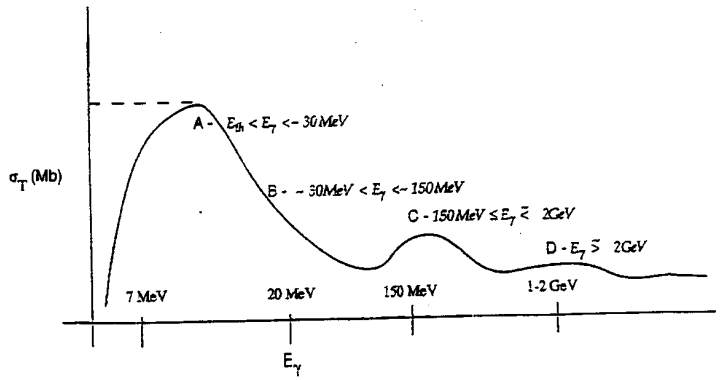


Figure 9.26: Cross section for photoproduction of hadrons.

2 GeV is the resonance region which is presently modeled as the production and decay of an N*. At higher energies vector dominance is assumed and a ρ meson decaying into pions is modeled. The present approximations are thought to give a reasonable representation of the number and spectrum of hadrons from the calculated photon flux. All the photons resulting from nuclear de-excitation were ignored.

A typical high energy (650 GeV) electron hitting a tungsten collimator produces initially, on average, 266 neutrons of average kinetic energy 2.9 MeV, 2.6 N* and 1.1 ρ mesons. Few of the neutrons reach the detector. However due to rescattering, the number of neutrons in the detector region per crossing is in the thousands. The momentum distributions of neutrons at generation and in the detector region per crossing are shown in Fig. 9.27 and Fig. 9.28 respectively. The initial kinetic energy of neutrons from the giant dipole source is .003 GeV, while the initial kinetic energy of neutrons and protons from the resonance region is .09 GeV and the average momentum of the generated pions is .75 GeV/c.

In figure Fig. 9.29 the neutron background is shown in the xz plane, and in Fig. 9.30 in the plane perpendicular to the z axis at the intersection point. These pictures are for 5 separate 1 Tev electrons impinging on the tungsten shield at a distance of 110 cm from the IP. Only the neutrons are shown in the figures.

In figure Fig. 9.31 and Fig. 9.32 the charged hadron backgrounds in the same two projections are shown. In this case 500 separate 1 Tev electrons were generated. The pictures are dominated by the large number of recoil protons kicked off the polyboron shield by energetic neutrons. The average momentum of the protons is only 200 MeV and this particular background can probably be suppressed by a layer of metal covering the polyboron shield. The flux of recoil protons is reduced by a factor of 3 when a 20 ns gate is applied. The average momentum of the pions is 240 MeV.

Some of the background particles are non-relativistic and/or are generated far away from the intersection point. Fig. 9.33 shows the time of generation of the slowest background particles: the neutrons. The remaining background from the previous bunch crossings is at the level of 0.1% for the whole final focus region and is at the level of 1% for the tracker/vertex volume, Fig. 9.34. (assuming a bunch crossing every 10-20 μ)

A summary of the backgrounds from muon decay is given in Table 9.1, Table 9.2 and Fig. 9.35, where the fluences for the various backgrounds as a function of detector radius are shown. In order to facilitate the positioning and the orientation of the detectors, two types of fluences have been calculated: radial where the particle traverses a cylinder parallel to the beam axis and longitudinal, where the particles traverse a plane perpendicular to the beam axis at the interaction point (z=0). The fluences at a given radius are similar in both cases reflecting the fact that most of the particles are very soft. All the fluxes are integrated

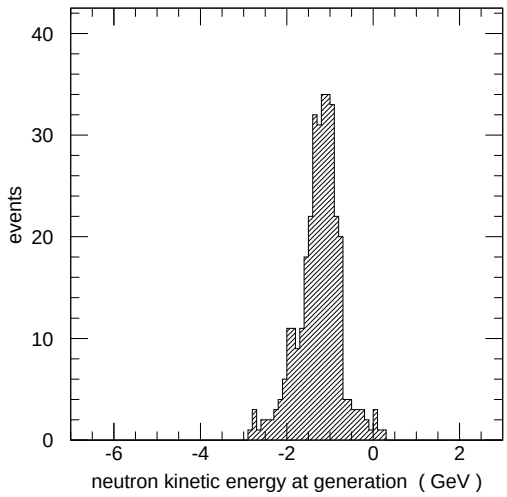


Figure 9.27: Log of generated neutron energy spectrum.

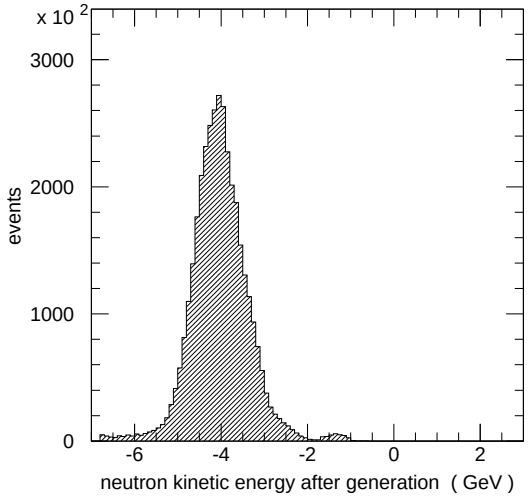


Figure 9.28: Log of neutron energy in the detector region.

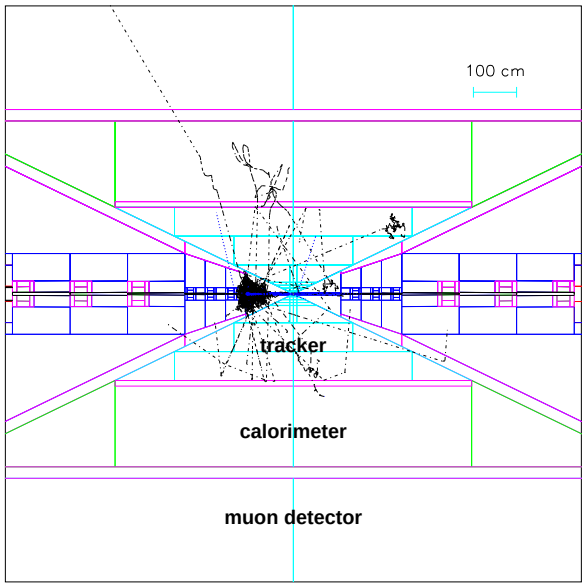


Figure 9.29: Neutron distribution in xz plane.

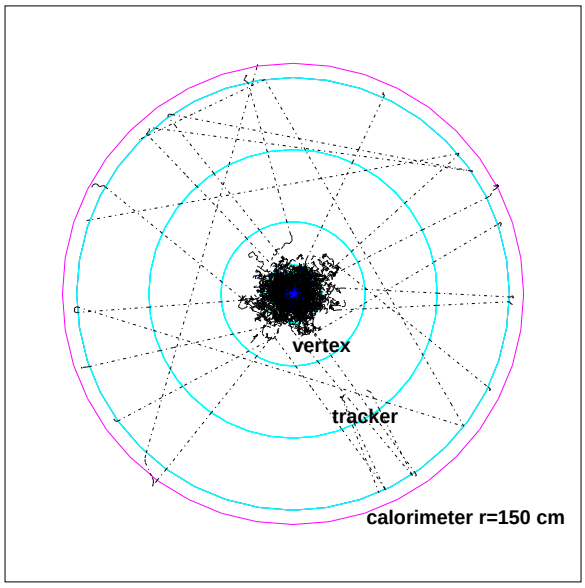


Figure 9.30: Neutron distribution normal to beams at z=0.

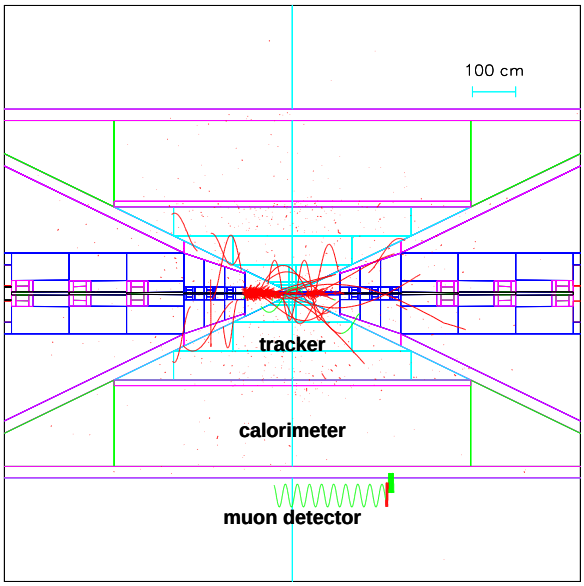


Figure 9.31: Charged hadron distribution in xz plane.

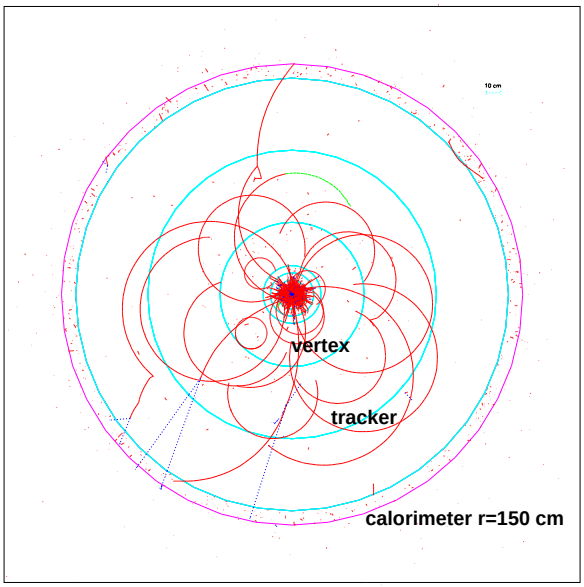


Figure 9.32: Charged hadron distribution normal to beams at z=0.

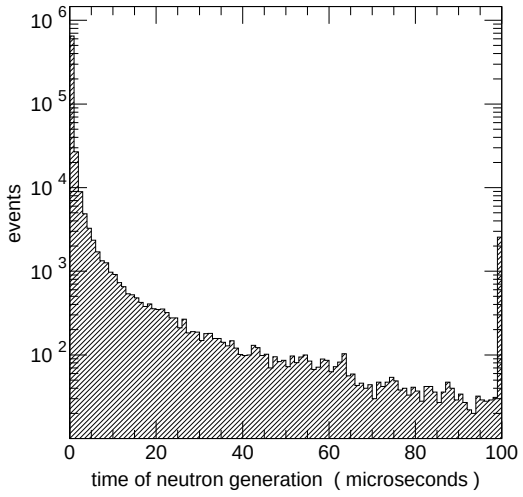


Figure 9.33: Time spectrum of the neutron background.

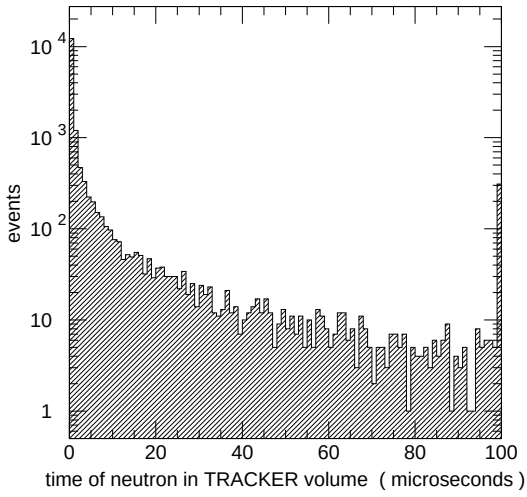


Figure 9.34: Time spectrum of neutrons into the tracker.

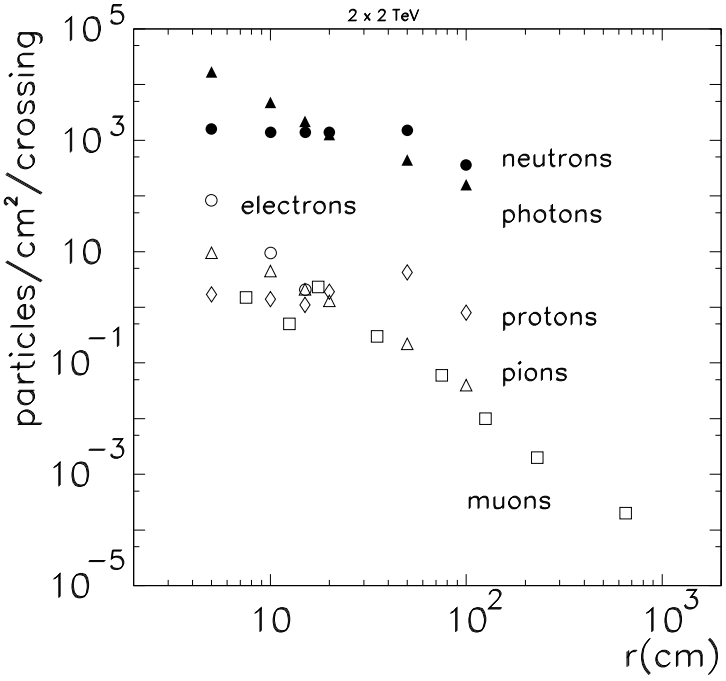


Figure 9.35: Particle fluences as a function of raduius.

over 20 μs. Only the proton and neutron fluxes are affected by timing cuts of the order of 20 ns. No significant differences were observed for integration times of 2 μs and 10 μs.A word of caution : in order to be able to compare our backgrounds with the ones calculated by MARS code or the ones predicted for NLC , all the values in Table 9.1, Table 9.2 and Fig. 9.35 apply for the crossing of two bunches of $10^{12}\mu$'s each . The average energies of the particles are given in Table 9.3.

With the present design the total energy deposited in the calorimeter is 427 TeV. The main contributors are the soft electromagnetic component (13 Tev), the hadronic byproducts of the electromagnetic shower (411 TeV) and the energy deposited by the Bethe– Heitler muons with their occasional catastrophic bremsstrahlung (3.4 TeV for a 4 Tesla toroidal field). The last component is not uniformly distributed in the calorimeter and imposes limits on jets measurements . In the absence of toroids, the muons deposit only 2.7 TeV in the calorimeter while for a toroidal field of 8 Tesla , the amount of energy deposited grows to 5.9 TeV .

Table 9.1: Longitudinal Particle Fluences from Muon Decays and Interactions from the GEANT Calculation. Fluence = particles/cm^2/crossing for two bunches of $10^{12}\mu$'s each.

Detector	Radius(cm)	γ's	neutrons	$e^{\pm}$	$\pi^{\pm}$	protons	$\mu^{\pm}$
Vertex	5-10	7900	1100	69	14.4	0.8	1.5
	10-15	3100	1200		3.7	0.05	0.5
	15-20	1600	1000		4.6	4.0	2.3
Tracker	20-50	450	870		0.8	3.9	0.3
	50-100	120	520		0.1	2.2	0.06
	100-150	130	330		0.003	0.4	0.01
Calorimeter	160-310						0.002
Muon	310-10000						0.0002

Table 9.2: Radial Particle Fluences from Muon Decays and Interactions from the GEANT Calculation. Fluence = particles/cm^2/crossing for two bunches of $10^{12}\mu$'s each.

Detector	Radius(cm)	γ's	neutrons	$e^{\pm}$	$\pi^{\pm}$	protons	$\mu^{\pm}$
Vertex	5	16900	1600	84.0	9.5	1.7	.35
	10	4800	1400	9.4	4.5	1.4	0.43
	15	2200	1400	2.1	2.1	1.1	0.33
	20	1250	1400		1.3	1.9	0.20
Tracker	50	440	1500		0.22	4.2	0.032
	100	160	360		0.04	0.8	0.008

Results from MARS

Similar calculations have been carried out using the MARS code. At present the intersection region being used here is quite different from the one used in the GEANT simulation discussed

Table 9.3: Mean kinetic energies and momenta of particles as calculated by GEANT.

Particle	γ	p	$\pi^{\pm}$	n
$\langle$Kinetic E$\rangle$, MeV	1	30	240	10

Detector	Radius	μ momentum from Bethe–Heitler(GeV)	μ momentum from π decay
Vertex	10-20	24	
Tracker	50-100	66	0.13
	100-150	31	
Calorimeter	160-310	19	

above. In particular the shielding cone is only 9°.

Fig. 9.36 shows the muon flux entering the interaction region toward the detector. There is a noticeable rate across the entire detector. Fig. 9.37 shows the formation of long range particle fluxes in the 275 m interaction region: lattice components, tunnel and experimental hall air, surrounding soil/rock, and detector components are simulated with the interaction point at 10 m.

Particle fluxes in the central detector are shown in Fig. 9.38. There is a rather uniform distribution of neutrals in the cavity with charged fluxes almost three orders of magnitude lower. Mean energies of those particles in the inner tracker are given in Table 9.4

Table 9.4: Mean energies of particles in inner tracker for 2 TeV μ^+ decays in the interaction region as calculated with MARS.

Particle	γ	$e^\pm$	$h^\pm$	n	μ
$\langle$E$\rangle$, MeV	2.5	80	249	0.2	3630

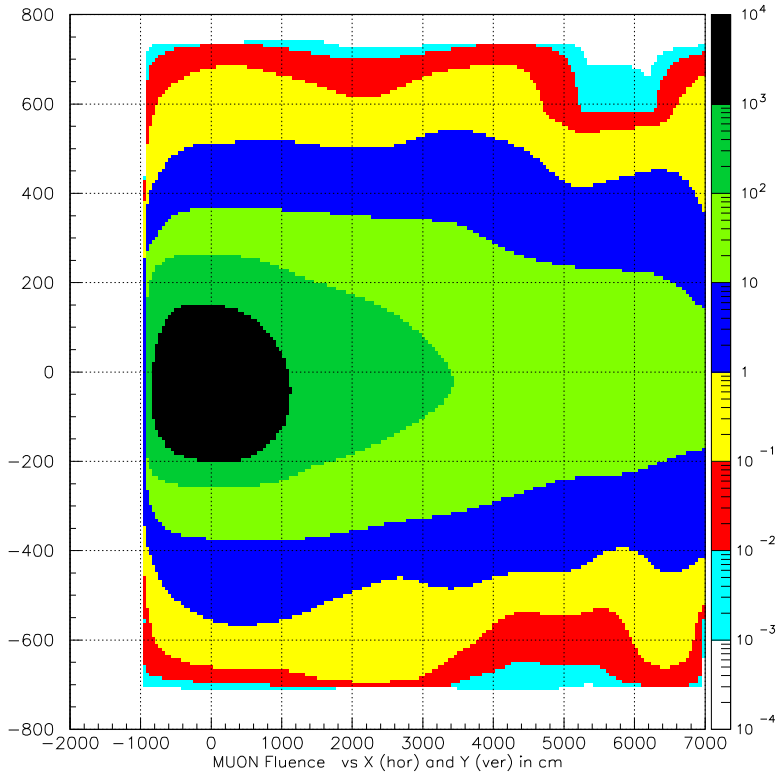


Figure 9.36: Muon flux contours ($cm^{-2}s^{-1}$) in a vertical plane of the collider tunnel and surrounding soil/rock at the entrance to the interaction region for 2 TeV muon beam decays as calculated with MARS. Beam axis is at x=y=0.

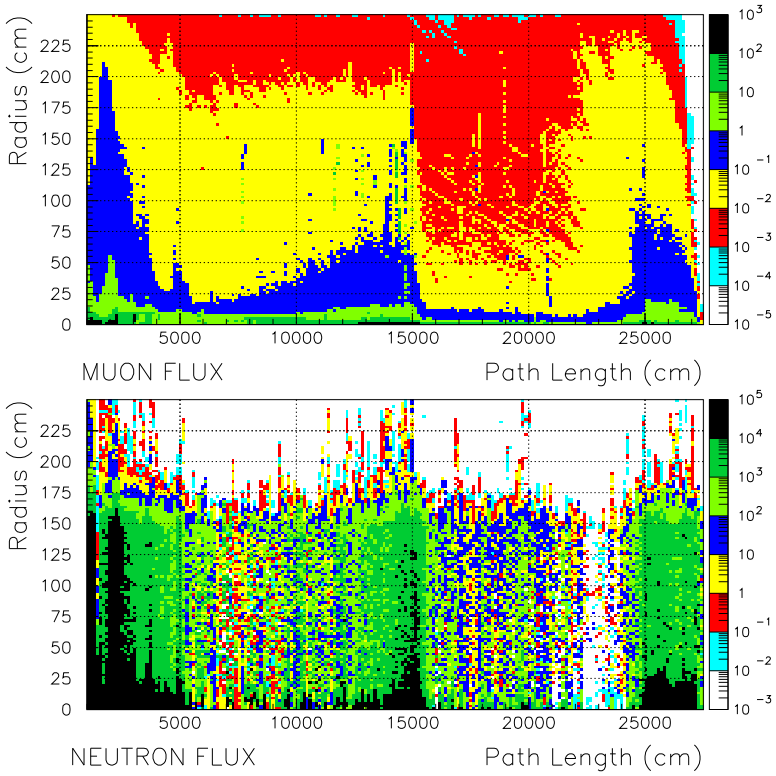


Figure 9.37: Muon (top) and neutron (bottom) flux contours ($cm^{-2}s^{-1}$) in tunnel, detector and surrounding soil/rock for 2 TeV muon beam decays as calculated with MARS. IP is at L=10 m.

9.4 Detector Specifications and Design

The physics requirements of a muon collider detector are similar to those of an electron collider. The main difference has to do with the machine related backgrounds and the added shielding that is needed near the beam pipe.

At this time little detailed work has been done on the design of a detector. Most of the discussion has centered around the types of detectors which might function well in this environment. The background levels detailed in the previous section are much higher than the comparable levels calculated for the SSC detectors and appear to be in excess of the levels expected at the LHC. Clearly segmentation is the key to successfully dealing with this environment. One major advantage of this muon collider over high energy hadron colliders is the long time between beam crossings; the LHC will have crossings every 25 ns compared to the 10 μs expected for the 4 TeV μ-collider. Much of the detector discussion has focused on ways to exploit this time between crossings to increase the segmentation while holding the number of readout elements to manageable levels.

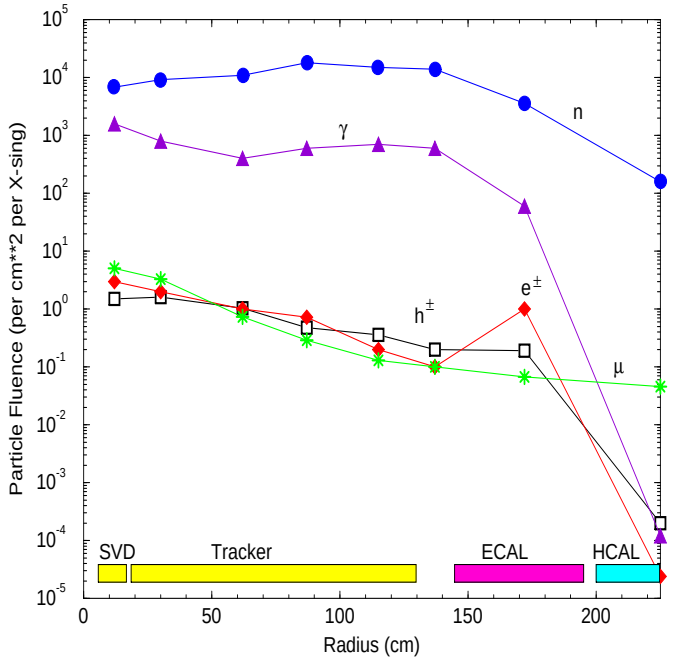


Figure 9.38: Particle flux radial distributions in a ± 1.2 m central detector region per bunch crossing due to 2 TeV muon decays in both beams as calculated with the MARS code.

While there is some detailed discussion of specific technologies below, the conceptual state of the detector design can be summarized as follows. The machine related backgrounds in the muon system behind the calorimeter are minimal. The issue for muon measurements in therefore to what extent track information from the inner tracking is required to obtain the required momentum precision. To the extent that the vertex point plus a track in the outside chambers is adequate, there is no real background issue to deal with. Much the same conclusion applies to the calorimeter system. Concerns about radiation damage and the high granularity that seems to be an advantage might necessitate a liquid electromagnetic system. The study of strong WW scattering requires very good hadron resolution to separate W and Z particles decaying into jets and reasonably precise p_T cuts on the WW system.

The real impact of the backgrounds will be felt in the inner tracking and vertex systems. One attractive possibility for a tracking system is a Time Projection Chamber[16]. This is an example of a low density, high precision device which takes advantage of the long time between crossings to provide low background and high segmentation with credible readout capability. A more detailed discussion of the use of such a device in a muon collider environment is given below. Silicon appears to be an adequate option for vertex detection.

Table 9.5: Detector Performance Requirements.

Detector Component	Minimum Resolution/Characteristics
Magnetic Field	Solenoid; B$\geq$2 T
Vertex Detector	b-tagging, small pixels
Tracking	$\Delta p/p^2 \sim 1\times10^{-3}(GeV)^{-1}$ at large p High granularity
EM Calorimeter	$\Delta E/E\sim 10\%/\sqrt{E}\oplus 0.7\%$ Granularity: longitudinal and transverse Active depth: 24 X_0
Hadron Calorimeter	Δ E/E$\sim 50\%/\sqrt{E}\oplus 2\%$ Granularity: longitudinal and transverse Total depth (EM + HAD)$\sim 7\lambda$
Muon Spectrometer	$\Delta p/p \sim 20\%$ 1 TeV

Again, because of the time between beam crossings, an attractive option here is the Silicon Drift Detector[17] and this is also discussed in detail below. It should be noted that the present backgrounds are quoted for an inner vertex detector radius of 10 cm. Work is ongoing to decide how close to the intersection point one can place this detector and still have relatively low occupancy. Given the large low energy photon flux in this region it is possible that tracking with the vertex detector is not practical because of the resulting large combinatorial problem. In this case the vertex detector would not be used for track finding but rather to project back tracks found in the tracking system (TPC) to determine their origin.

An interesting question which has yet to be addressed is whether or not it is possible to tag high energy muons which penetrate the tungsten shielding which in the present design extends to 20° from the beam axis. For example, in the case of $\mu\mu \to \nu\nu W^+W^-$ the primary physics background is due to $\mu\mu \to \mu\mu W^+W^-$. To reduce the background, in addition to a high p_T cut on the WW pair, in might be advantageous to tag forward going muons. These μ's would penetrate the shielding.

9.4.1 Detector Performance Requirements

The detector performance criteria that are used for the design of the detector are summarized in Table 9.5.

The rationale for these particular choices is as follows. A solenoid is the natural field and also has the virtue of curling up soft charged particles. The strength of the field may

need to be increased to trap more particles to a smaller radius. The vertex detector with the tracking will be used for tagging b's from top. The tracking detector requirements are rather modest in terms of momentum resolution, however, there is a large background and so it must have a large number of effective pixels.

The Electromagnetic(EM) calorimeter is used to identify electrons, photons and the core of jets. It is crucial to have high granularity to deal with the background. The low energy photons will interact at the beginning of the EM calorimeter, so longitudinal segmentation will allow it to be reduced in the analysis. The transverse and longitudinal segmentation will also allow the determination of the photon angle to separate photons from the interaction point from background.

The hadron calorimeter must measure jets well enough to separate W's from Z's. Another important measurement will be the missing transverse energy. A study needs to be done to determine the parameters to achieve this goal: energy resolution, e/π response, segmentation, total depth. The parameters listed here are taken from common hadron calorimeters being proposed for current experiments. The segmentation again will help to distinguish the particles from the interaction region from background which will come from upstream.

The muon spectrometer needs to identify and measure muons, but the momentum resolution need not be extraordinary. Again since there will be a large number of muons from sources other than the interaction region, some redundancy will be required to reject the background.

9.4.2 Strawman Detector

The object of this present exercise is to see if a relatively conventional detector can be built using state-of-the-art (or not far beyond) technology to satisfy the physics needs of the muon collider.

A layout of the detector is shown in Fig. 9.39. The main features of the detector are: A large cone (20°) that is probably not instrumented and is used to shield the detector from the machine induced background. The element nearest to the intersection region is the vertex detector located at as small a radius as possible. A number of technologies including Silicon Drift Detectors(SDD), Silicon Pixels [18], and CCD detectors have been considered. Both SDD and pixels will probably work due to their inherent large granularity, but SDD seem especially attractive because of the reduced number of readout channels and potentially easier construction. Inner tracking technologies considered were cathode pad chambers, silicon strips and TPCs. The use of a TPC is interesting as the amount of material is minimized and thus the detector does not suffer as much from low energy photon

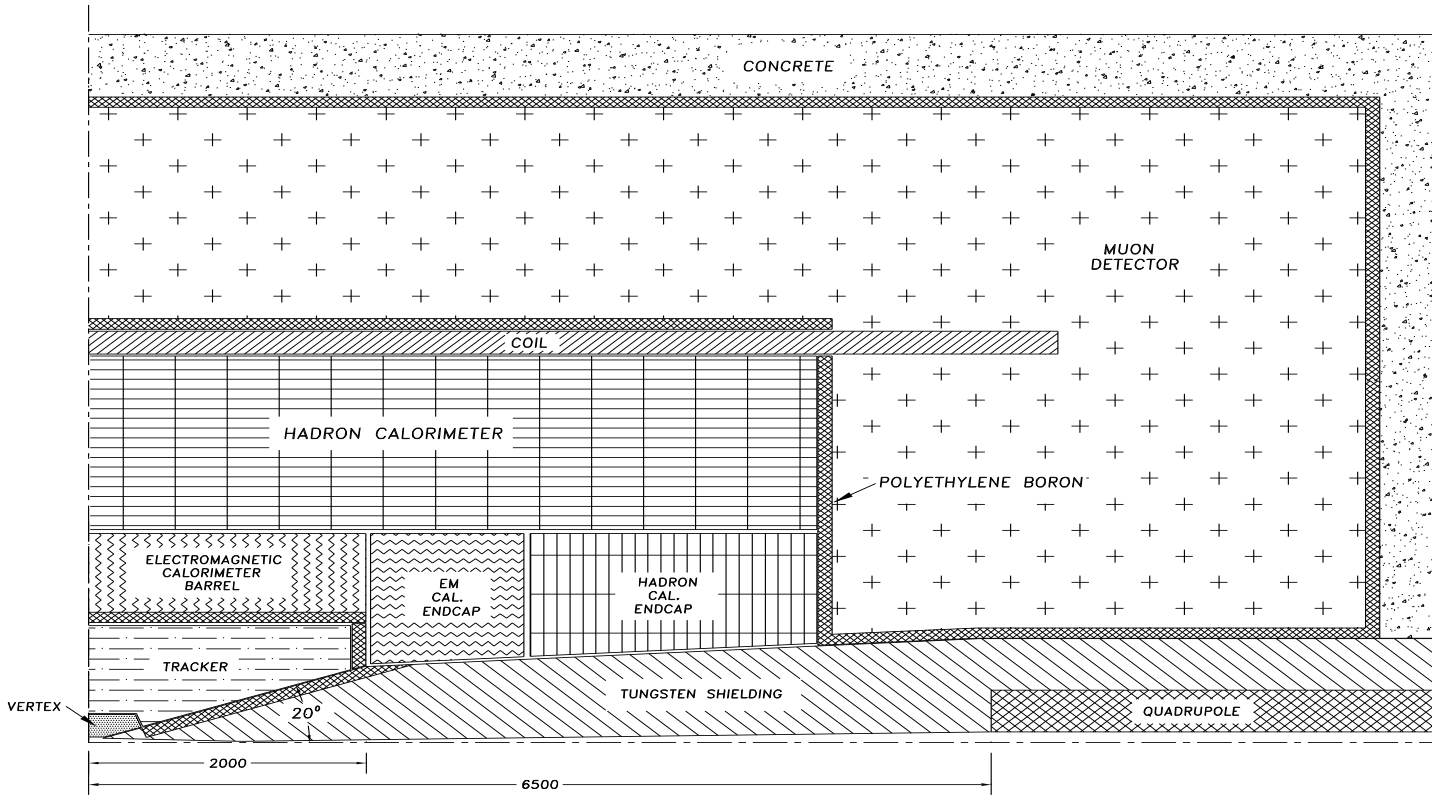


Figure 9.39: Strawman Detector

and neutron backgrounds.

For the calorimeter system there are many options. A liquid argon accordion calorimeter for the EM part [18], [19] and a scintillator hadronic calorimeter appear appropriate. This combination gives a hadronic resolution that is of order $50\%/\sqrt{E}$ which may be good enough. The high granularity of the EM section allows good electron identification which will be of help in tagging b-jets. In addition the longitudinal and transverse granularity allow for corrections on an event by event basis to compensate for the fact that the calorimeter is non-compensating as well as to allow the identification of catastrophic muon bremsstrahlung.

There is a single magnet, with a field of 2 Tesla in the tracking region. The magnet is located behind the hadron calorimeter.

The muon system is a stand-alone system. The chambers in the muon system are Cathode Strip Chambers(CSC) that can be used for both a two dimensional readout as well as a trigger. These chambers have good timing resolution and relatively short drift time which minimizes neutron background problems.

Silicon Drift Vertex Detector

Semiconductor Drift Detectors appear to be an attractive solution for vertexing and inner tracking for an experiment at a $\mu^+\mu^-$ collider. Silicon drift detectors are ideal detector elements for finding secondary vertices and for charged particle tracking close to the interaction region. In the present conceptual design the silicon drift detectors are followed by a large volume Time Projection Chamber (TPC) filled with a fast, low Z gas. The combination of the precise position information provided by the silicon drift detectors and the very good track recognition of the TPC will accomplish all the required tracking tasks, even in the presence of the $\mu^+\mu^-$ collider background.

The $\mu^+\mu^-$ bunches collide once every 10 μs. The background particles produced by the decay of circulating muons and the subsequent cascade processes arrive at the detector at the same time as the particles produced by the studied $\mu^+\mu^-$ interactions. To reconstruct the trajectories of particles produced in the interactions and to reject the background, the highest number of position resolution elements is required. Both tracking technologies use the $10\mu s$ of "clean" time between the bunch crossings to read the position information produced by the particles at the crossing time. One can think of the silicon drift detectors as having an intrinsic signal pipeline within the active volume of the detector. This feature allows a very high detector granularity with a minimum of non-active material in the passage of particles and with a relatively modest number of read–out channels.

Silicon drift detectors are able to provide an unambiguous position resolution of several μm in two perpendicular directions. However, the number of silicon layers is limited and it may be difficult to connect hits in several planes to reconstruct particle tracks in the presence of the low energy photon background. The pattern recognition of the vertex detector will be supplemented by the track information from the TPC.

Silicon drift detectors are able to provide very precise position and ionization measurements with a relatively modest amount of electronics. A perspective view of the drift detector[17] is shown in Fig. 9.40. In principle, the electric field of the drift detector forces electrons liberated by an ionizing particle to drift parallel to the large semiconductor surface to the anode. The transit time of electrons inside the detector measures the distance of an incident particle from the anode. The charge sharing among anodes gives the coordinate perpendicular to the drift direction.

Silicon drift detectors are slowly gaining acceptance in experimental particle physics. CERN experiment NA45 has already published several results obtained with a doublet of cylindrical silicon drift detectors.[20] A similar detector is being used in WA98 and a vertex detector based on a p-type silicon drift detector was approved as an upgrade of experiment

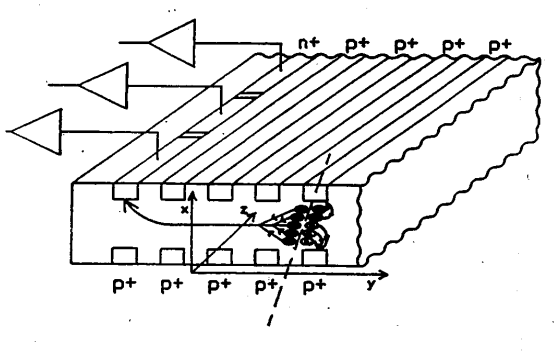


Figure 9.40: Perspective view (not to scale) of a semiconductor drift detector. Electrons created by an ionizing particle are transported long distances parallel to the detector surface. The anode is divided into short segments to measure the coordinate perpendicular to the drift direction.

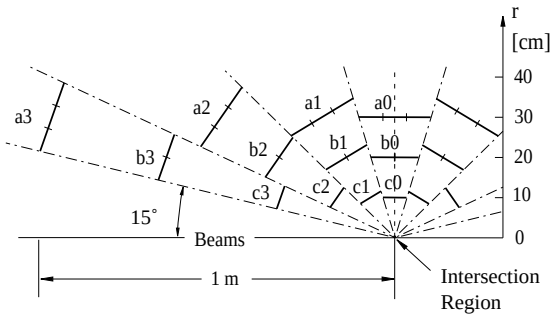


Figure 9.41: Cross section along the beam axis of the vertex and inner tracking detector. Detector has full azimuthal coverage and extends in the polar angle to the shields of the intersection region. Heavy lines indicate the rings of detectors centered around the beam axis.

NA49. At BNL a vertex detector based on silicon drift detectors is being constructed by the STAR collaboration at RHIC.[21]

The best silicon detectors are capable of achieving a position resolution of 4 μm with a detector 300 μm thick. However, these results were obtained only with normally incident charged particles. For other angles of incidence, the resolution can be degraded by the fluctuations in the density of the ionization (Landau) in the silicon. When a fast particle crosses the detector at an inclined angle the detector measures the center of gravity of the charge produced by the particle in a given direction. This center of gravity of the produced charge is on average located in the middle plane of the detector. Due to the fluctuations in the linear density of the ionization, the position of the center of gravity fluctuates in the measured direction. These fluctuations may be larger than the intrinsic resolution of the detector.

The presence of a magnetic field modifies the trajectories of electrons in a silicon drift detector and normal incidence is no longer the ideal one. There is, however, a proper incidence angle which does not give any degradation of the position resolution.

Fig. 9.41 shows the cross section of the proposed silicon drift detector along the beam axis. There is a uniform magnetic field along the beam direction. Different sectors and different layers are visible in Fig. 9.41. Sectors are labelled $0, 1, 2, 3$ and each sector covers a certain region in rapidity. Layers are indicated by letters a, b and c and are placed roughly

30, 20 and 10 cm from the beam. The most important requirement is to maintain a perpendicular incidence of fast particles onto the detector surface to retain the best possible position resolution. There is only a very small overlap between the neighboring sectors. The position of the individual sectors along the beam direction is defined by this non-overlapping requirement, by the desire to maintain normal incidence and by the average distance from the beam axis which is kept constant for all sectors.

The proposed geometry is a possible compromise between the size of the detector in the forward direction and the increase of the density of background particles closer to the intersection point. The geometry minimizes the amount of silicon to cover a given solid angle at any distance. Last but not least the silicon drift detectors in this geometry are less likely to be crossed by low energy spiraling tracks.

The granularity of the proposed detector seems be reasonable for the rates of background particles. Based on rates in Table 9.1 the number of hits, the occupancy and the radiation dose in individual silicon detectors can be estimated. Relevant fluxes are radial for sectors number 0 and longitudinal for sectors number 3. The layer" c0" located 10 cm from the beam has the highest number of hits per cm^2: 11 from the interaction of neutral particles and 16 from crossings of charged particles. For a pixel size of 316×316 μm^2 the number of pixel per cm^2 is 1000. In this case the occupancy of background hits is less than 3% in the layer "c0". Layer c1,c2 and c3 have progressively lower occupancies, the estimated occupancy in the layer c3 being about 2%. The density of background hits decreases almost with the square of the distance from the beams and the occupancies in layers "b" and "a" are less than 1% and 0.4% respectively.

One could even think about locating one more layer, called"d", at about 5 cm from the beams. Detectors with a finer anode pitch than that in the layers farther away would be placed in layer "d". The number of channel would increase only slightly. By also implementing a more sophisticated algorithm for detecting two close hits, the pixel size of the detectors in the "d" layer can be reduced down to 150×150 μm^2. The occupancy of background hits would only be about 3% even in this layer.

The damage due to the radiation dose may be the most annoying problem. If detectors are produced from an n–type silicon with a bulk doping level of $1.5 \times 10^{12}/cm^3$ the detectors have to be replaced after a year of operation. The use of p–type material seems to be more appropriate for this application. P–type silicon drift detectors are being developed in LBNL. These detectors are supposed to be much more radiation resistant. Some R&D may be required.

The decrease of the maximum drift in p–type silicon speed by a factor of 3 is not a problem; the longest drift time can still be shorter than the time between beam crossings.

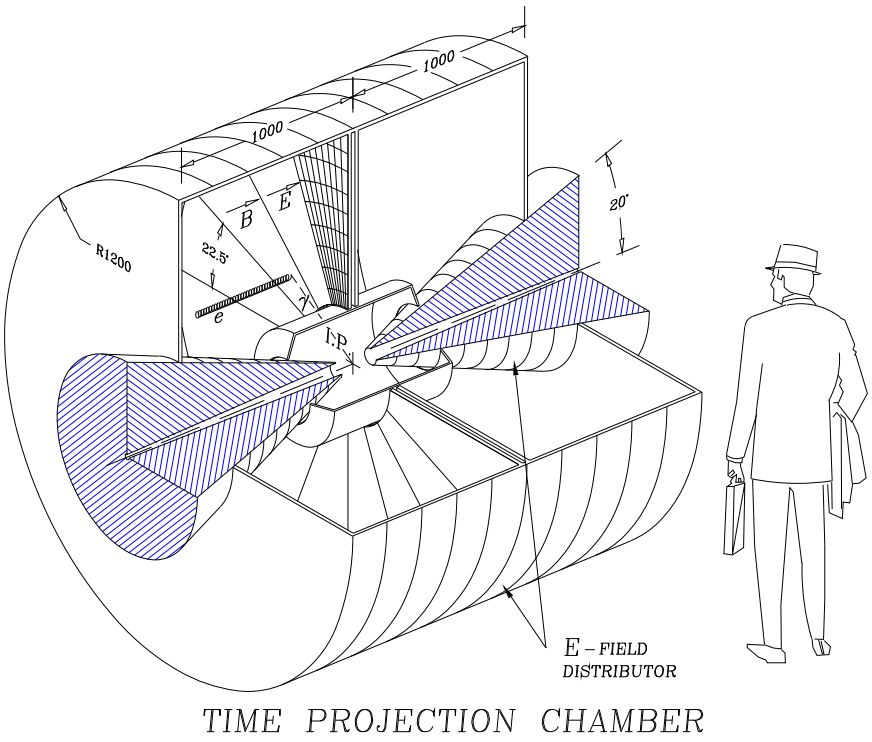


Figure 9.42: Sketch of TPC for Muon Collider Detector.

This is clearly only an initial outline of an inner tracking detector and the real details still remain to be worked out. The number of sectors and number of layers as well as their positions will likely change with a better understanding of backgrounds and with the evolution of the design of the shielding. However it is clear that it should be possible to build such a device for this environment which, at least at 10 cm radius, has a relative low occupancy and which can serve as an excellent vertex detector.

Time Projection Chamber (TPC)

An interesting candidate for tracking at a muon collider is a Time Projection Chamber (TPC). This device has good track reconstruction capabilities in a high density environment, good 3-dimensional imaging and provides excellent momentum resolution and track pointing to the vertex region. It is perhaps particularly well suited to this environment as the long time between bunch crossings ($\sim$ 10 μs) permits drifts of $\sim$1 m and the average density of the device is low compared to more conventional trackers which helps to reduce the measured background rates in the device. In the present detector considerations the TPC would occupy the region between the conic tungsten absorber and electromagnetic calorimeter in the region from 35 cm to 120 cm. (Fig. 9.42)

The TPC is a cylinder, filled with gas at one atmosphere pressure, in parallel axial magnetic and electric fields. The drift region of the TPC is bounded by a field distributor which creates a uniform electric field. The amplification region collects ionization charge and may be a standard cathode pad chamber or any other detector which provides fine readout granularity. The r-ϕ coordinate is measured by interpolating signals on the cathode pads; the r coordinate is given by the pad number and the z coordinate is obtained from the drift time. A special grid is used as a gate to remove undesired events and to significantly reduce the penetration of positive ions into the drift region. The gate is open just before the beam crossing and locked after ~ 2 μs if first level trigger does not appeared.

There are three basic requirements in deciding on a suitable chamber gas for a TPC at a muon collider. To reduce background gamma and neutron interactions in the detector volume, a low density gas mixture should be chosen as the detection medium of the TPC. Another important parameter is the electron drift velocity. Since the time between beam crossings is fixed (10 μs in the present design) the drift velocity should be high enough to collect all the ionization deposited in the drift region. Finally the detection medium should not contain low atomic number gases to help reduce the transfer energy to the recoil nucleus and in this way to reduce its range in the gas. The gas mixture 90% He + 10% CF_4 satisfies all these requirements and it could be an excellent candidate for the TPC. It does not contain hydrogen which would cause a deleterious effect from the neutrons, has a density 1.2 mg/cm^3 and a drift velocity of 9.4 cm/μs. The single electron longitudinal diffusion for this gas is $\sigma_l = 0.15$ mm/$\sqrt{cm}$. The transverse diffusion, which is strongly suppressed by the 2 Tesla magnetic field is given by,

$$\sigma_t = \frac{\sigma_t(B=0)}{\sqrt{1+(\omega\tau)^2}} = 0.03\text{mm}/\sqrt{\text{cm}} \tag{9.1}$$

Each time slice will contain about 25 ionization electrons, and the expected precision in r- ϕ and z coordinates is,

$$\sigma_\phi = \sqrt{\frac{Z\sigma_t^2}{25} + (0.05)^2} \qquad (\text{mm}) \qquad \sigma_Z = \sqrt{\frac{Z\sigma_l^2}{25} + (0.15)^2} \qquad (\text{mm}) \tag{9.2}$$

Z - drift length (cm). The precision of r-coordinate is defined by the anode wire pitch - 3mm.

Low energy photons, neutrons and charged particles produce the main background for the inner tracker. Photons in the MeV region interact with matter mainly by Compton scattering. For a 1 MeV photon the probability of producing a Compton electron in 1 cm of gas is $\xi_\gamma = 4.5 \cdot 10^{-5}$. For an average photon fluence $h_\gamma = 100$ cm^{-2} about $N_\gamma = 4 \cdot 10^4$ electron tracks are created in the chamber volume. Because the transverse momentum of Compton electrons is rather small the electrons are strongly curled by the magnetic field

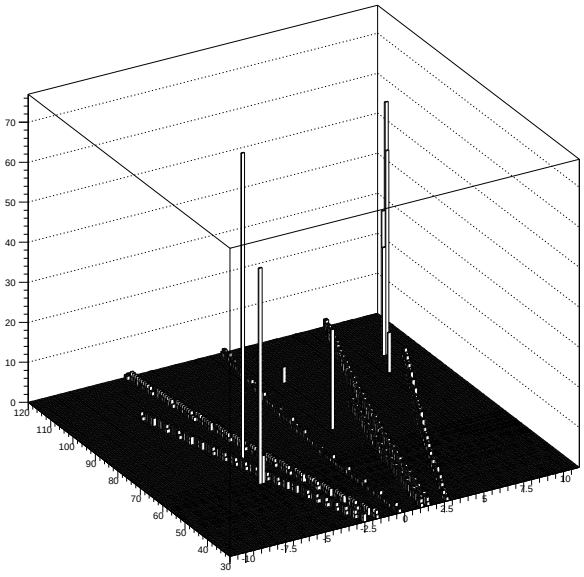


Figure 9.43: Number of digitized points in an xy cell from a signal event and background Compton electrons.

and move along the magnetic field lines. Most of the electrons have a radius less than one millimeter and their projection on the readout plane covers not more than one readout pitch, 0.3×0.4 cm^2. The average length of the Compton electron tracks in the TPC is one meter and therefore, the volume occupied by electron tracks is $v_{comp.e} = 4.8 \cdot 10^5$ cm^3. Since the total chamber volume is 10^7 cm^3, the average occupancy due to background photon interactions is equal to,

$$< occupancy >_\gamma = \frac{V_{comp.e}}{V_{total}} = 4.4 \cdot 10^{-2} \tag{9.3}$$

For neutrons in the MeV region the primary interaction with matter is elastic collisions. In this case the energy transfer to the nucleus has a flat distribution and the maximum transfer energy is given by $4E_nA/(A+1)^2$ or $4E_n/A$ when $A \gg 1$. The maximum transfer momentum is $P_{max} = \sqrt{2m_pE_n}$ for H and $P_{max} = \sqrt{8m_pE_n}$ for $A \gg 1$. The ionization energy loss for slow heavy particles is,

$$dE/dx \propto (Z/\beta)^2 = (ZE_{total}/P)^2 \tag{9.4}$$

Z - nucleus charge, E_{total} - total nucleus energy, P - nucleus momentum. Taking into account that $E_{total} \approx A \cdot m_p$,

$$dE/dx \propto \frac{1}{2} \cdot \frac{m_p}{E_n} \qquad (forH) \qquad dE/dx \propto \frac{A^4}{32} \cdot \frac{m_p}{E_n} \qquad (forA \gg 1) \tag{9.5}$$

The calculated mean energy of background neutrons is $E_n = 27\ MeV$. In this case, for hydrogen, the ionization loss of recoil protons is about 20 times higher than for minimum

ionizing particles, that is $dE/dx \approx 20\ keV/cm$ and their mean range in the gas is several meters. For carbon, for example, the ionization loss is much higher and $dE/dx \approx 24\ MeV/cm$. This implies that for the gas chosen, the mean length of the recoil nucleus tracks will only be a few millimeters. This simple estimation shows that the operating gas of the TPC should not contain hydrogen to exclude long range tracks in the chamber volume.

The occupancy due to background neutron interactions can be estimated as follows. The calculated neutron fluence is $< n >= 10^3\ cm^{-2}$. As discussed above the track of the recoil nucleus occupies, typically, not more than one volume cell of the TPC, $v_n = 0.3{\times}0.4{\times}1.0$ cm^3. The probability of a background neutron interacting in 1 cm of the gas is $\xi_n = 2 \cdot 10^{-5}$, the number of recoil tracks $N_n =< n > \cdot \xi_n \cdot V_{total} = 2 \cdot 10^5$ and therefore the neutron occupancy is,

$$< occupancy >_n = \frac{N_n \cdot v_n}{V_{total}} = 0.24 \cdot 10^{-2} \tag{9.6}$$

The overall background occupancy is not very high. However, given the number of background tracks, it may be difficult to extract the event tracks without any cleaning procedure. To optimize chamber performance and to study the efficiency of 'cleaning' algorithms a chamber simulation program was developed. This simulation of high momentum event tracks and background tracks includes ionization, drift and diffusion electrons in the gas, multiplication and other details of the detection process. The volume of the TPC was divided into 1 cm z slices and each piece of track crossing such a slice was digitized for each pad row on which it induces charge. In this way each piece of track transfers to a three dimensional point.

It is easy to clean out the recoil tracks owing to their large ionization density per cell. To do this only a simple cut to remove all volume cells which contain a charge in excess of some preset threshold is required. This cut will only eliminate about 1% of the TPC volume.

The Compton tracks can be removed in a similar way. Again the volume of the TPC is divided into 1 cm z slices and each piece of track crossing such a slice was digitized for each pad row on which it induces charge. The result is that each piece of track transfers to a three dimensional point. Then all the digitized points from the TPC volume are projected on the readout plane which is divided into 0.3×0.3 cm^3 cells. After that, the number of points contained in each cell is computed. Because almost all points of a Compton track lie along the z-axis most of them will be projected into one cell and therefore the number of points in this cell will be very different from hit cells from non-background tracks. Fig. 9.43 clearly demonstrates the difference between the high momentum event tracks and the low energy Compton electron tracks. It is obvious that to remove low momentum electron tracks, all cells containing more than some threshold number of points should be excluded. Applying this procedure a few percent of volume is lost but the quality of the high momentum tracks

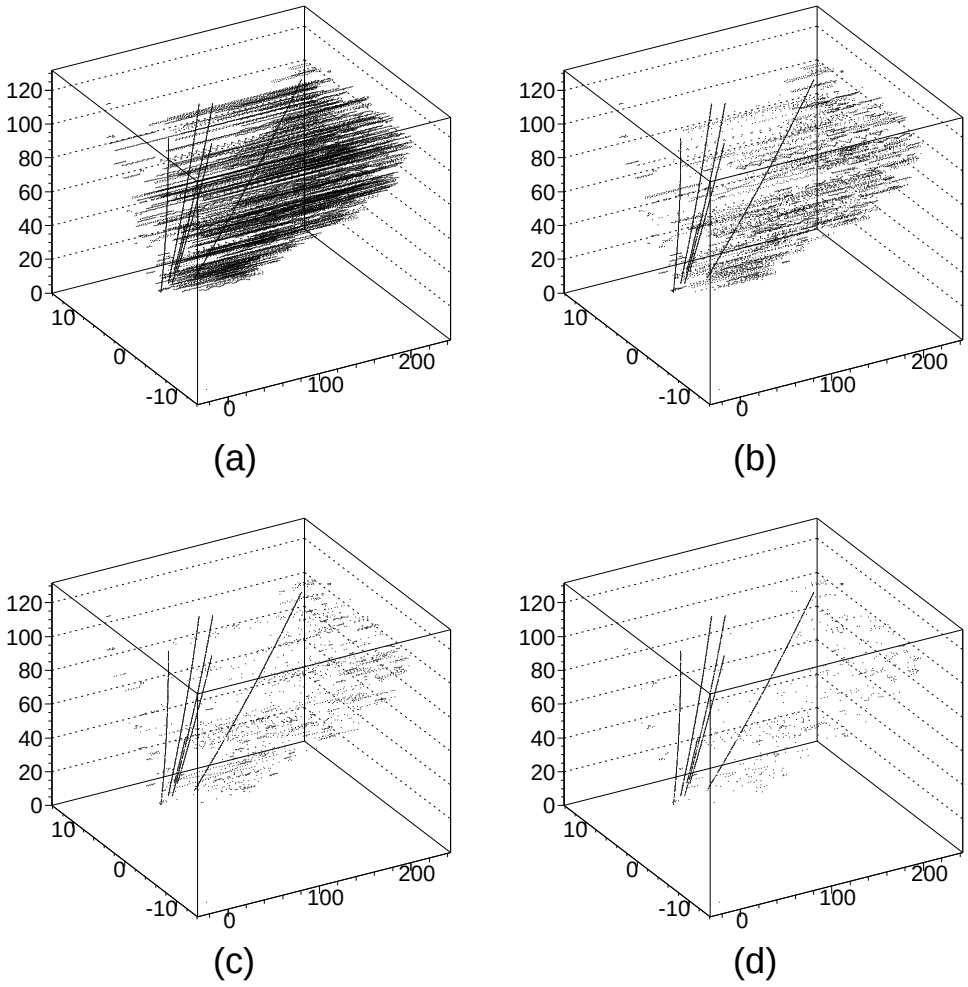


Figure 9.44: Charge distribution from a signal event and background Compton electrons with threshold cuts; (a) has no cut applied while (d) has the most stringent cut.

is not substantially changed. This is clearly illustrated in Fig. 9.44 where one sector of the TPC is shown after the application of different value threshold cuts.

Once the clean up from neutron and Compton background is carried out, some information about TPC performance can be obtained. Fig. 9.45a shows the transverse momentum resolution for tracks with $P_t = 50\ GeV/c$. The momentum resolution is about 1.2% in this case. In Fig. 9.45b the difference along the x-axis between the real vertex and the vertex interpolated from track intersections is shown. For these distributions tracks were generated uniformly in η and $P_t \geq 10\ GeV/c$. In neither case has information on the bunch crossing point been used.

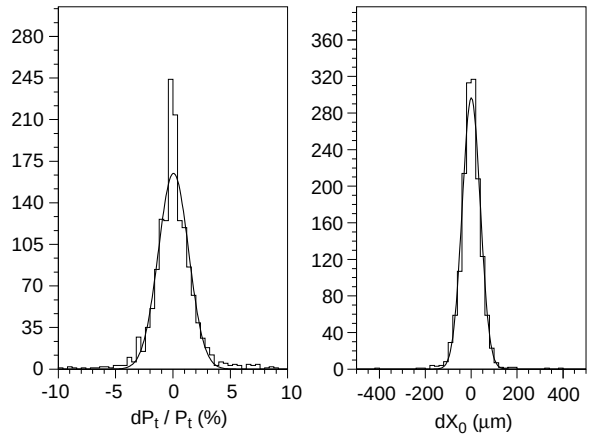


Figure 9.45: Momentum (a) and position (b) resolution for the TPC.

Electromagnetic Calorimeter

An accordion liquid argon calorimeter is being developed for the ATLAS collaboration[18]. A similar calorimeter designed for the GEM Collaboration at the SSC is shown in Fig. 9.46[19].

The calorimeter is made of a sandwich of stainless steel clad lead absorber plates interleaved with Kapton electrodes. Both are bent into a corrugated shape - hence the name accordion. The plates are arranged roughly radially for the barrel calorimeter so that there are no through going cracks. The Kapton electrodes can be subdivided to measure θ and depth according to the requirements of the experiment. The electrodes are ganged together in ϕ again according to the design. In this way, almost arbitrary segmentation can be achieved in all three directions: θ, ϕ and depth. Liquid argon calorimeters have shown high radiation resistance and excellent timing resolution[22]. In ATLAS a design for the end cap geometry called the "Spanish Fan" has also been successfully tested.

The large amount of hadronic energy deposited in the calorimeter is mostly in the form of low energy particles. Most of this energy will be deposited in the first section of the EM calorimeter and to handle it will require both transverse and longitudinal segmentation in the calorimeter. This will allow one to reconstruct the energy of the particle coming from the intersection region by using the added information on its direction as well as longitudinal shower development.

From the GEANT background calculations, the total energy deposited from the electromagnetic debris is $\sim$ 13 TeV. However, if one divides the calorimeter into $\sim 2\times10^5$ cells, the mean energy would be about 65 MeV/cell. Certainly, energetic electromagnetic showers from γ's or electrons or the core of jets will stand out above this uniform noise. Since the readout is every 10 μs, multiplexing is possible to reduce costs compared to the LHC where collisions occur every 25 ns.

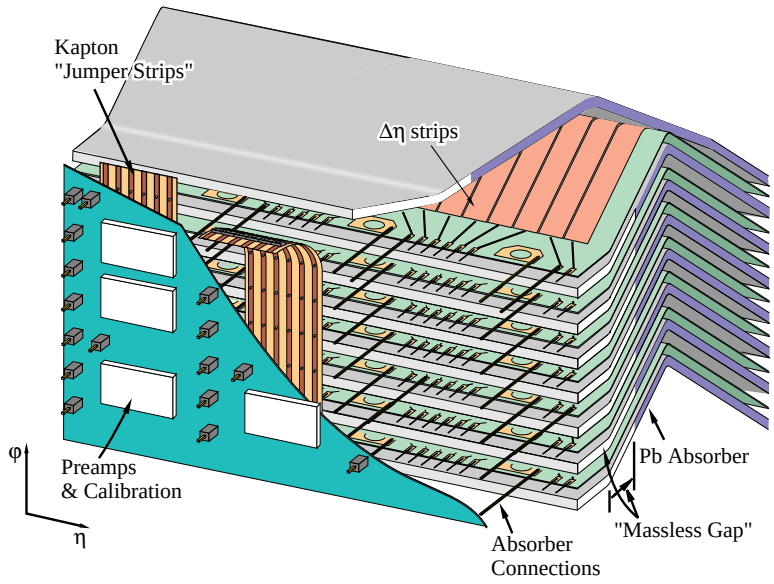


Figure 9.46: Liquid argon accordion calorimeter designed for the GEM experiment. The strips measure the azimuthal angle.

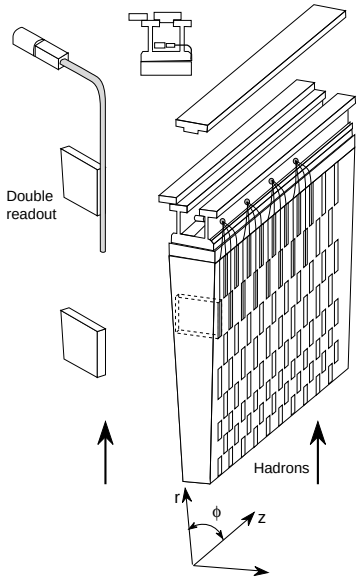


Figure 9.47: Scintillator tile calorimeter from the ATLAS experiment.

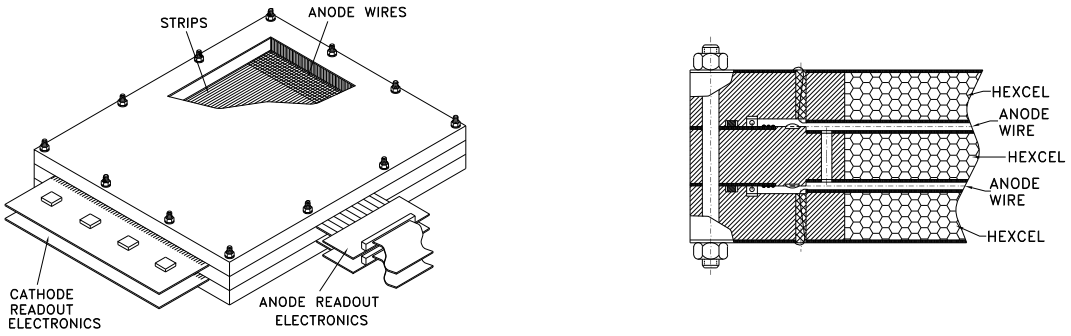


Figure 9.48: A two-layer module of the Cathode Strip Chamber.

Hadron Calorimeter

A good choice for the hadron calorimeter is a scintillator tile device being designed for ATLAS[18]. It uses a novel approach where the tiles are arranged perpendicular to the beam direction to allow easy coupling to wave-length shifting fibers[23]Fig. 9.47

From the GEANT background calculations, the total energy deposited in the calorimeter from electromagnetic and hadronic showers and muons is about 100 TeV. Again, this is rather uniform with a low mean energy per cell. It should be thought of as a sort of pedestal noise and should not affect the measurement of jets. With a tile calorimeter of the type discussed here it should be possible to achieve a resolution of Δ E/E$\sim$ 50%/$\sqrt{\mathrm{E}}$

Muon Spectrometer

Triggering is probably the most difficult aspect of muon spectrometers in large, 4π detectors in both lepton and hadron colliders. In addition, a muon system should be able to cope with the larger than usual muon backgrounds that would be encountered in a muon collider. Segmentation is, again, the key to handling these high background rates. Cathode Strip Chambers (CSC) are an example of a detector that could be used in the muon system of a muon collider experiment. CSCs are multi-wire proportional chambers (MWPC) with segmented cathodes so they provide measurements of both coordinates. The coordinate along the anode wires can be determined with high precision (of order 50 μm) by determining the center of gravity of the charge induced on the cathodes by the electron avalanche. The resolution in the transverse coordinate is limited by the anode wire spacing (just as in traditional MWPCs) and is of the order of one mm. Coarser strips can be used if such resolution is not needed. The short maximum drift time (of order 35 ns) allows bunch crossing identification as well as prompt signals that can be used for triggering. Thus one

detector performs all functions necessary for a muon system:

- Precision coordinate(50 to 70 μm)
- Transverse coordinate(of order mm or coarser as needed)
- Timing (to a few ns)
- Trigger primitives

In addition, the cathodes can be lithographically segmented almost arbitrarily resulting in pixel detectors the size of which is limited only by the density and signal routing of the readout electronics.

Cathode strip chambers were first developed as pixel (or pad) chambers for an experiment at the AGS (E814) to handle the high secondary particle densities in heavy ion collisions. They were further developed for muon identification and measurement for the GEM experiment at the SSC[25]. They have been adopted for use in parts of the muon systems of both CMS and ATLAS experiments at the LHC as well as the muon arms of the PHENIX experiment at RHIC. An effort to develop a monolithic front end is well under way at BNL with fully functional prototypes already fabricated and tested.

A detector is usually built as a multi-layer module. The CSC design, shown in Fig. 9.48 has two layers formed by three flat, rigid panels, each made of a 23 mm thick sheet of nomex honeycomb (hexcel) and two 1.19 mm thick copper-clad FR4 laminates (Fig. 9.48b), the 17 μm thick copper forming the cathodes. The panel frames are made of machined zelux (fiberglass reinforced lexan). They provide the 2.54 mm step for the anode plane of gold-plated tungsten wires 30 μm in diameter. The frames of the outer panels have a milled cavity with enough room for epoxy beads for the wire attachment as well as the anode blocking capacitors. A rubber gasket just outside this cavity provides the gas seal for the assembly. In this manner no components under high voltage are outside the seal, thus minimizing the risk of high voltage breakdowns.

The position sensing cathode strips are lithographically etched on either side of the central panel. These cathodes are precisely positioned with respect to each other with the aid of locating pins. The strips are oriented at 90° with respect to the anode wires, providing precision position measurement in the direction along the anode wires. The outer cathodes in each layer are segmented in orthogonal strips or left as continuous (un-etched) copper planes if the wires are read out to provide the transverse coordinate. The cathode readout pitch, W, is 5.08 mm. With a 1% charge interpolation this readout spacing provides a resolution of $\approx 50\mu m$. The interpolation precision is limited by the signal to noise ratio and,

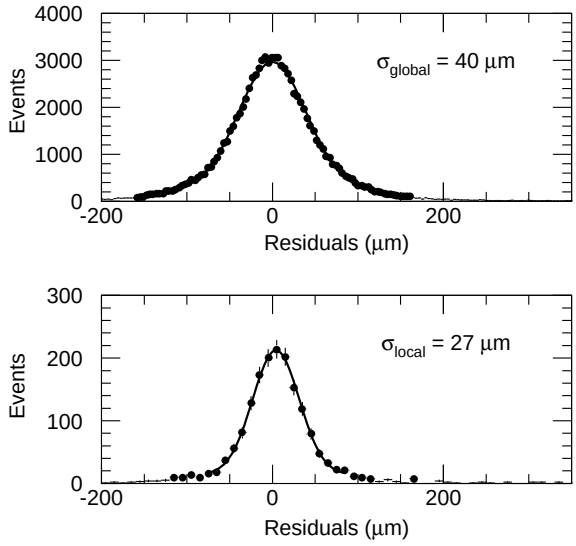


Figure 9.49: Typical spatial resolution of the Cathode Strip Chamber.

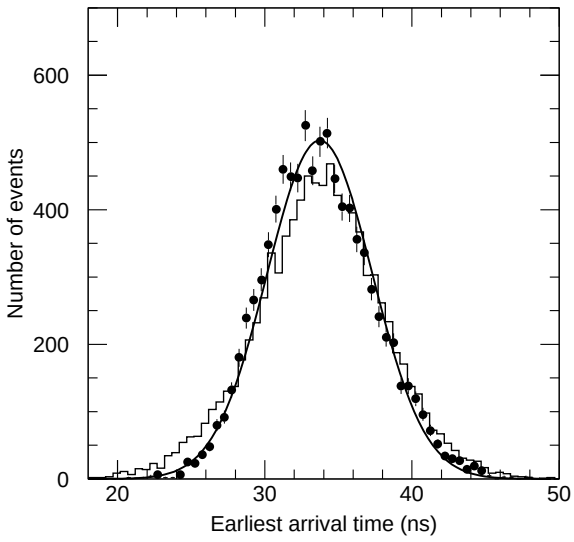


Figure 9.50: Typical timing resolution of the Cathode Strip Chamber.

in principle, can be improved for a given electronic noise by increasing the gain. In practice inter-calibration of neighboring channels and geometric deviations eventually set the limit in the achievable resolution. Prototype CSC detectors have been constructed for both the GEM and ATLAS experiments. Fig. 9.49 shows the measured resolution from such a prototype.

Although the maximum drift time for typical gases is of order 35 ns, event timing with an rms of a few ns can be achieved in a multilayer module by determining the earliest time of arrival from all layers. Fig. 9.50 shows the measured timing resolution in the same prototype.

9.5 Backgrounds for a 250 GeV x 250 GeV Collider.

Work has begun on estimating the various machine related backgrounds for a 250 GeV x 250 GeV muon collider. Since no real final focus or lattice design for such a machine exists at this time the dimensions for the lower energy machine are just scaled from the higher energy machine: the longitudinal dimensions were divided by the square root of 8 while the radial dimensions and the fields were identical to the 4 Tev case. It is also assumed that the machine will have 8 Tesla dipoles resulting in a ring 650 meters in circumference, that there will be only one bunch per fill and consequently a bunch crossing every 2 microseconds. The initial luminosity will be reduced by 50 percent after 1000 loops, and an initial bunch of 10^{12} muons will generate 1.5×10^6 decay electrons every meter. The final focus extends 46 meters from the intersection point and contains only quadrupoles. The last quadrupole ends

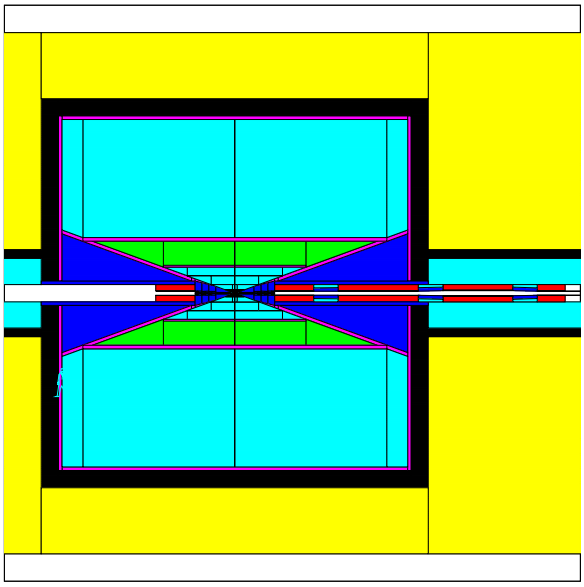


Figure 9.51: Intersection region for 250 GeV x 250 GeV Machine.

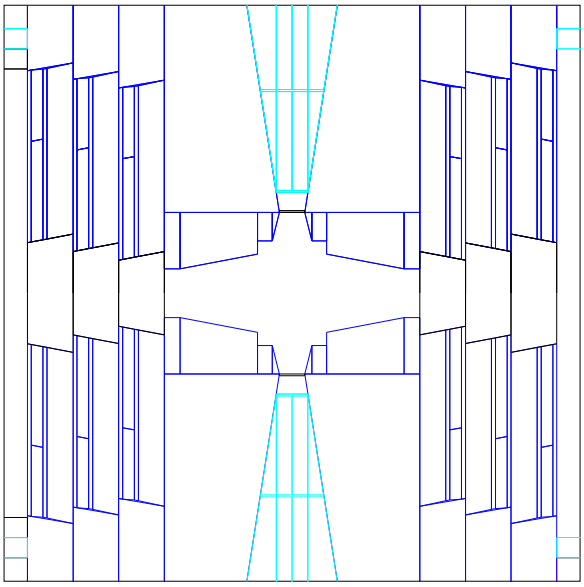


Figure 9.52: GEANT description of Intersection region for 250 GeV x 250 GeV machine.

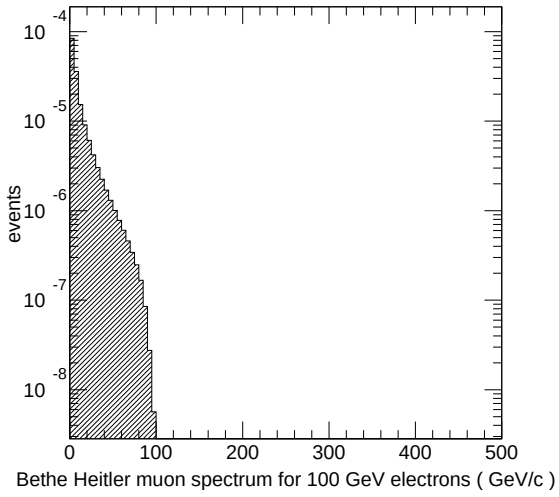


Figure 9.53: Bethe–Heitler muon spectrum from a 100 GeV electron.

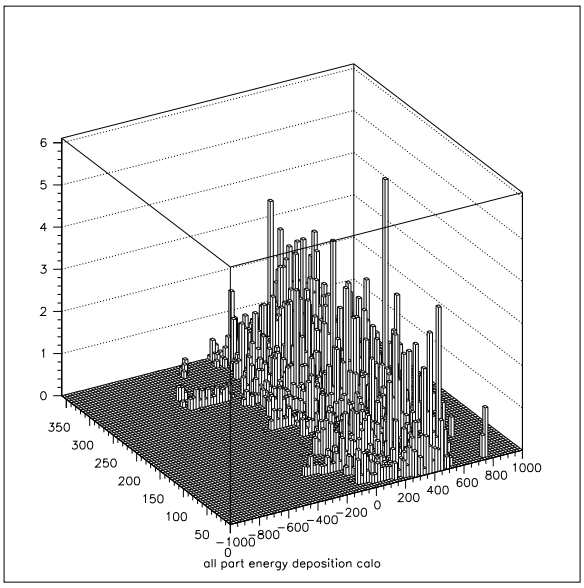


Figure 9.54: Energy deposited by muons in the calorimeter.

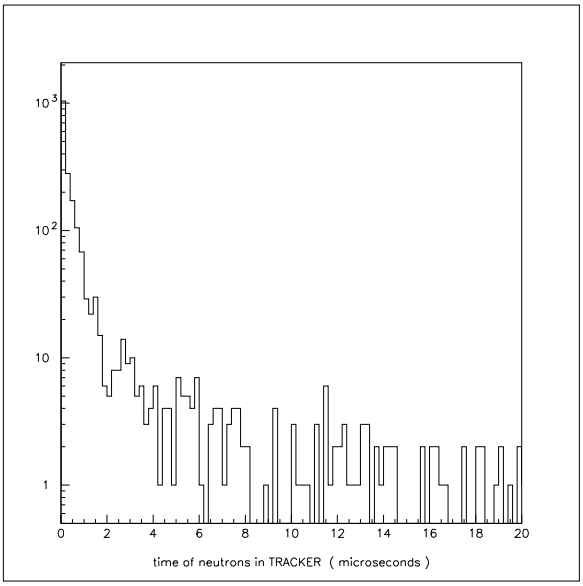


Figure 9.55: Neutron time of flight.

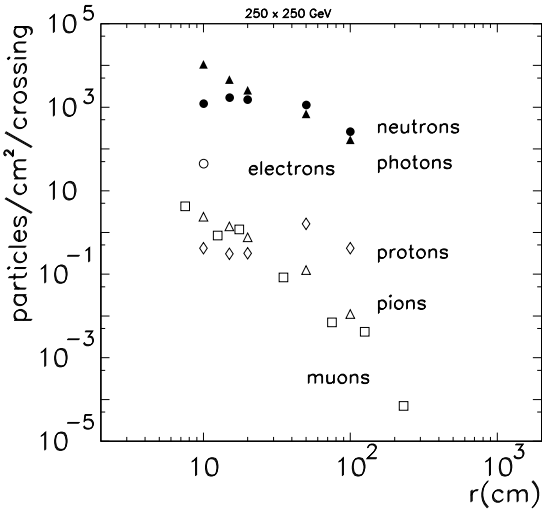


Figure 9.56: Fluences for the 250 GeV X 250 GeV machine.

2.3 meters from the IP and as in the 4 Tev case all the backgrounds generated outside the final focus region are ignored.

The average energy of the electrons from muon decay is 88 GeV. The critical energy of the synchrotron radiation has an average of 21.4 MeV, resulting in an average photon energy of 16.5 MeV. The fraction of electron energy carried by synchrotron radiation photons is only 1.65% compared to 20% for the 4 Tev case.

The same methods used for the 4 Tev study were applied in the present study. The experimental area is again a cylindrical cavity of 10 meter radius and 20 meters length Fig. 9.51. The final focus geometry here is much more compact and there are no toroids between the final focus quadrupoles. The σ_x and σ_y envelope of the muon bunch in the spaces between the quadrupoles was calculated and non-magnetic scrapers with ellipsoidal apertures of 4σ were assumed in transporting the decay electrons. The beam aperture at the exit of the last quad is 2.5 cm and the shielding material used is tungsten. The same inverted cones method for the shield design was used. In this case the angle is 11 mrad compared to 4 mrad for the 4 Tev machine. The final shield around the IP starts at 1.1 m and has an opening of 1.2 cm radius compared to 0.42 cm for the 4 Tev machine. This leaves the intersection region more exposed to electromagnetic debris than was the case for the 4 Tev machine. The details of the 250 GeV design can be seen in Fig. 9.52.

The spectrum of Bethe–Heitler muons for a 100 GeV electron hitting the tungsten shielding is shown in Fig. 9.53. The probability for Bethe–Heitler muon production, as a function

of electron momentum is shown in Fig. 9.21. The average momentum of the Bethe–Heitler muons generated in the whole final focus region is 10.1 GeV and they have an average relative angle to the electron direction of 18 mrad. The average probability of a decay electron to produce a Bethe–Heitler muon is 1.2×10^{-4}.

The total energy deposited in the calorimeter (333 Tev) has three main contributions : 35 Tev is pure electromagnetic, 297 Tev is from hadronic sources and 1.2 Tev is deposited by the Bethe–Heitler muons. The energy deposited by muons is shown in Fig. 9.54.

Another difference from the 4 Tev collider is the fact that the bunch crossings here occur every 2 microseconds. Since there is a high flux of slow neutrons, the neutrons surviving from the previous crossings were estimated. In Fig. 9.55 the timing of neutrons relative to the bunch crossing is shown. It is evident that for a 2 microsecond gate, the contribution from previous crossings is only at the 1% level.

The fluences for different particles are shown in Fig. 9.56. They are all radial fluxes except for the muons where a longitudinal flux is preferred. All the backgrounds are comparable to those for the 4 Tev collider. In other words, the lowering of the energy by a factor of 8 compensates for the order of magnitude increase in the rate of muon decay.

This is only a very preliminary analysis of the machine related backgrounds for the 250 GeV x 250 GeV collider. The positions and strengths of the magnets were not optimized. From the point of view of background reduction, one would prefer a much bigger distance between the end of the last quadrupole and the IP. This would make the hadronic shield deeper and will reduce the 1.2 cm radius of the iris which controls the photon flux .

9.6 Summary and Conclusions.

Two independent background calculations have been used for a preliminary study of the expected background level at a 4 TeV muon collider. The optimization of the intersection region is still at its infancy, but the results of both studies show that the level of background while still large, can be managed with proper design of the intersection region and choice of detector technologies. This is in large part due to the fact that the background is composed of many very soft particles which behave like a pedestal shift in the calorimeter. The tracking and vertexing systems will have to be highly segmented to handle this flux of background particles.

A large amount of work is still needed in order to optimize the intersection region and the final focus. In particular a better understanding of the trade off between the different backgrounds is required.

The strawman detector presented here is meant only to show that the muon collider

detector has unique problems and advantages. An optimized detector needs to be developed taking these problems into consideration. Present state-of-the-art technologies seem to be sufficient to build a detector which will meet the physics requirements of a 4 TeV collider. There is still a large amount of uncertainty and more studies are needed. In particular, the vertex detector and calorimeter performance in this level of background needs to be studied.

Some preliminary calculations for machine related backgrounds for a lower energy collider (250 GeV x 250 GeV) have also been carried out. It appears at this time that the backgrounds in this case are comparable to the levels at the 4 Tev machine. Since little attention has yet been paid to the details of the final focus for this lower energy machine it is possible that reductions in the machine related backgrounds will be achievable in the future.

Bibliography

[1] For references to in-depth studies of physics at future e^+e^- colliders, see e.g. *Proceedings of the Workshop on Physics and Experiments with Linear Colliders*, ed. F.A.Harris, *et al.*, World Scientific(1993); *Proceedings of the Workshop on Physics and Experiments with Linear e^+e^- Colliders*, Wailkoloa, Hawaii(April 1993), ed. F. Harris *et al.*,(World Scientific, 1993); JLC Group, KEK Report 92-16(1992); *Proceedings of the Workshop on Physics with Linear Colliders*, Saariselkä, Finland(Sept. 1991), ed. R. Orava *et al.*, World Scientific (1992).

[2] See e.g. *Proceedings of the First Workshop on the Physics Potential and Development of $\mu^+\mu^-$ Colliders*, Napa, California, 1992, Nucl. Instr. and Meth. **A350**, 24(1994).

[3] 13) THE MAD PROGRAM. By H. Grote, F.C. Iselin, E. Keil (CERN), J. Niederer (Brookhaven). CERN-LEP-TH-89-13, Mar 1989. 4pp. Presented at 1989 Particle Accelerator Conf., Chicago, IL, Mar 20-23, 1989. Published in IEEE Part.Accel.1989:1292-1294.

[4] V. Barger, *et al.*, Phys.Rev. **D52**, 3815(1995).

[5] I. J. Ginzburg, *The e^+e^- pair production at $\mu^+\mu^-$ collider*, Preprint: hep-ph/9601273(1996).

[6] P. Chen, *Beam-Beam Interaction in Muon Colliders*, SLAC-PUB-7161(April, 1996).

[7] R. L. Ford and W. R. Nelson, *The EGS Code System: Computer Programs for the Monte Carlo Simulation of Electromagnetic Cascade Showers*, SLAC-0210, Jun 1978.

[8] A. Fasso, A. Ferrari, J. Ranft and P. R. Sala, *FLUKA: present status and future developments*, Proc. of the IV Int. Conf. on Calorimetry in High Energy Physics, La Biodola (Is. d'Elba), Italy(Sept. 20-25 1993), Ed. A. Menzione and A. Scribano, World Scientific, p. 493(1993).

[9] C. Zeitnitz and T. Gabriel, Nucl. Instr. Meth., **A349**, p. 106 (1994).

[10] N. V. Mokhov, *The* MARS *Code System User's Guide, version 13 (95)*, FNAL–FN–628 (1995).

[11] N. V. Mokhov, *Comparison of Backgrounds in Detectors for LHC, NLC and $\mu^+\mu^-$ Colliders*, in Proceedins of the Symposium on Physics Potential and Development of $\mu^+\mu^-$ Colliders, San Francisco, CA, December 1995; Fermilab–Conf–96/062.

[12] A. I. Drozhdin, M. Huhtinen and N. V. Mokhov, *Accelerator Related Background in the CMS Detector at LHC*, Fermilab–Pub–96/113 and CERN CMS TN/96-056 (1996).

[13] G. W. Foster and N. V. Mokhov, *Backgrounds and Detector Performance at a 2+2 TeV $\mu^+\mu^-$ Collider*, in *AIP Conference Proceedings*, **352**, pp. 178-190 (1995); also Fermilab–Conf–95/037 (1995).

[14] N. V. Mokhov and S. I. Striganov, *Simulation of Backgrounds in Detectors and Energy Deposition in Superconducting Magnets at $\mu^+\mu^-$ Colliders*, Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop, Ed. J. C. Gallardo, AIP Conference Proceedings 372, 1996.

[15] Yung-Su Tsai, Review of Modern Phys., **46**, 4(1974).

[16] *The Time Projection Chamber: A New 4π Detector for Charged Particles*, D.R. Nygren (SLAC). PEP-0144, (Received Dec 1976). 21pp. In Berkeley 1974, Proceedings, Pep Summer Study, Berkeley 1975, 58-78.

[17] E. Gatti and P. Rehak, Nucl. Instr. and Meth. **225**, 608 (1984).

[18] *ATLAS Technical Proposal for a General-Purpose pp Experiment at the Large Hadron Collider at CERN*, CERN/LHCC/94-43, LHCC/P2 (15 December 1994).

[19] *GEM Technical Design Report Submitted by Gammas, Electrons, and Muons Collaboration to the Superconducting Super Collider Laboratory*, GEM-TN-93-262; SSCL-SR-1219 (July 31, 1993).

[20] G. Agakichev, *et al.*, Phys. Rev. Lett., **75**, 1272(1995).

[21] *The STAR Experiment at the Relativistic Heavy Ion Collider*, Nucl. Phys., **A566**,1994.

[22] O. Benary, *et al.*, Nucl. Inst. and Meth., **A332**, 78(1993).

[23] O. Guidemeister, F. Nessi-Tadaldi and M. Nessi, Proc. 2nd Int. Conf. on Calorimetry in HEP, Capri, 1991.

[24] P. Rehak and E. Gatti, *Intersection between Particle and Nuclear Physics*, AIP Conference proceedings 338, 320 (1995).

[25] G. Bencze *et al.*, Nucl. Inst. and Meth., **A357**, 40(1995).

Contributors

- D. Lissauer, (BNL) Editor
- A. Baltz, (BNL)
- O. Benary, (Tel-Aviv Univ.)
- S. Geer, (FermiLab)
- H. Gordon, (BNL)
- S. Kahn, (BNL)
- P. Lebrun, (FermiLab)
- A. Luccio, (BNL)
- H. Ma, (BNL)
- N. Mokhov, (FermiLab)
- M. Murtagh, (BNL)
- F. Paige, (BNL)
- R. B. Palmer, (BNL)
- V. Polychronakos, (BNL)
- S. Protopopescu, (BNL)
- P. Rehak, (BNL)
- I. Stumer, (BNL)
- H. Takai, (BNL)
- V. Tchernatine, (BNL)
- A. Tollestrup, (FermiLab)
- W. Willis, (Columbia Univ.)

List of Figures

List of Tables

Chapter 10

RADIATION ISSUES

Contents

10.1 Introduction

All aspects of radiation control at a $\mu^+\mu^-$ collider complex will be folded into the design to insure that compliance with applicable regulations is achieved, and that the accelerators and detectors are operated in a reliable and safe manner. Radiological impact on the work place and on the environment will be kept as low as reasonably achievable (ALARA). This

includes the establishment of a stringent set of radiation limits and design goals for off- and on-site radiation levels, quantification of radiation source terms, specification of shielding design criteria, installation of appropriate radiation instrumentation, provision for access control, and control of residual activation. The entire $\mu^+\mu^-$ collider facility is assumed to be located on the existing Fermilab or BNL sites. This greatly simplifies considerations for monitoring and controlling beam loss induced radiation fields from facility operations. Most of these beam facilities will be within tunneled enclosures (within the dolomite layer) which avoids many potential difficulties.

The basic $\mu^+\mu^-$ collider facility is outlined in [1]. The radiation source terms at this facility are mainly connected with an intense rapid-cycling proton driver, a target station, pion decays in a decay channel, unavoidable muon decays in the accelerator chain and in the final collider and spent muon beam absorption. The first analysis of the radiation environment at muon colliders [2] has shown that the spectrum of radiation issues is wide and challenging. Some of the problems appear to be severe, but can be mitigated with the proposed measures. A series of dedicated simulations have been performed with the MARS code [3] to understand formation of radiation fields in the complex and to assure that there are ways to meet the stringent regulation requirements. Considered in detail are the main collider arcs, the interaction region and absorption of spent muon beam. Although attention is paid mainly to the 2×2 TeV collider, both 2 TeV and 250 GeV muon beams are considered.

The parameters used are from [1]. The radiation levels (per second) around the storage ring are calculated assuming 1000 turns as a beam lifetime and should be multiplied by 10^7 s (collider operational year) when compared to the annual limits. The on-site annual dose limit is taken as 100 mrem/year. The Fermilab off-site limit is 10 mrem/year.

10.2 Proton Source and Muon Production

The proton driver generates $\sim$3 MW of 8 or 30 GeV protons which are directed onto a π-production target. This will require an isolated tunnel with target-vault for π-production, transport line for $\pi \to \mu\nu$ decays and a beam absorber for the spent proton beam. This is similar to the present p-bar production line at Fermilab, except that overall beam power is somewhat higher while the proton energy is lower. A relatively straightforward extension of the p-bar experience should lead to a satisfactory beam handling and shielding solution for the production facility. The entire production region will be well within the existing site and separated from accidental personnel exposure. Isolation and control of the facility would be eased by placing it underground.

10.3 Accelerator Chain

Compared to other parts of the facility, there is nothing serious here, but certainly dedicated studies are required.

10.4 Collider Arc and IR

10.4.1 Source Term

In contrast to hadron colliders, where the interaction points are a serious source of radiation, almost 100% of the prompt radiation in muon collider detectors and in muon storage ring arises in the lattice. The decay length for 2 TeV muons is $\lambda_D \approx 10^7$ m. With 10^{12} muons in a bunch one expects 2×10^5 decays per meter of lattice in a single pass of two bunches. Under the assumed scenario this becomes 2×10^8 decays per meter per store or 6×10^9 decays per meter per second. Electrons from $\mu \to e\nu\overline{\nu}$ decay have a mean energy of approximately 1/3 of that of the muons. These ~ 700 GeV electrons, generated at the rate of 6×10^9 per meter per second, travel to the inside of the ring magnets while radiating many energetic 0.1–1 GeV synchrotron photons towards the outside of the ring [4]. Electromagnetic showers induced by these electrons and photons in the collider components create high radiation levels both in a detector and in the storage ring. Another source of radiation in the muon storage ring is beam halo interactions at limiting apertures, located primarily in the interaction region (IR) optics.

Simulations with the MARS code are done for the realistic lattice described in the previous chapter. All the particle interaction processes are simulated in (1) the lattice with detailed 3-D dipole and quadrupole geometry and magnetic field maps, (2) the 1.45 m radius tunnel surrounded by soil/rock (ρ=2.24 g/cm^3), (3) a 26 m long and 10 m radius experimental hall and (4) the detector are taken into account [4], [5].

A single MARS run includes:

- forced $\mu \to e\nu\overline{\nu}$ decays in the beam pipe (beam muon decay studies) or beam halo interactions with the limiting aperture beam pipe;
- tracking of created electrons in the beam pipe under influence of the magnetic field with emission of synchrotron photons along the track;
- simulation of electromagnetic showers in collider and detector components induced by electrons and synchrotron photons hitting the beam pipe, with production of hadrons and prompt muons via Bethe–Heitler pairs and direct positron annihilation;

- simulation of muon interactions (bremsstrahlung, direct e^+e^- pair production, ionization, deep inelastic nuclear interactions and decays) along the tracks in the lattice, detector, tunnel and experimental hall components and air, and in the surrounding soil/rock;

- simulation of electromagnetic showers initiated at the above muon interaction vertices;

- simulation of hadronic cascades generated in muon and photon interactions, with daughter electromagnetic showers, with muon production (π and K decays, prompt muons in hadronic and electromagnetic interactions), and with low–energy neutron transport;

- histogramming and analysis of particle energy spectra, fluence and energy deposition in various detector and collider regions.

Energy thresholds are 1 MeV for muons and charged hadrons, 0.3 MeV for electrons and photons, and 0.5 eV (0.00215 eV in some cases) for neutrons.

Fig. 10.1 shows particle energy spectra inside the 1.5 cm radius arc aperture for 2 TeV muon decays, while Fig. 10.2 is for particles outside the aperture, averaged over the tungsten liner, magnet components, tunnel air and a few meters of the surrounding soil/rock. In the aperture, one can see a pronounced 700 GeV peak in the decay positron spectrum and a significant number of $\sim$1 GeV photons, whereas most of the particles are rather low energy. Overall mean particle energies and relative multiplicities are given in Table 10.1

Table 10.1: Mean energies of particles and their relative multiplicities in showers induced by 2 TeV μ^+ decays in the arcs averaged over the aperture, magnet components, tunnel air and surrounding soil/rock.

Particle	γ	e^+	e^-	$h^\pm$	n	μ
$\langle$E$\rangle$, MeV	380	12800	63.7	195	0.142	21300
$\langle$N$\rangle$	6120	284	335	0.027	203	0.156

10.4.2 Prompt Radiation

Radiation fields in the lattice components are dominated by electromagnetic showers induced by $\sim$700 GeV decay electrons and positrons and by $\sim$1 GeV synchrotron photons. In the tunnel, experimental hall and in the first meters of the surrounding soil/rock, the field is composed of low energy photons and neutrons. Farther from the tunnel the only significant

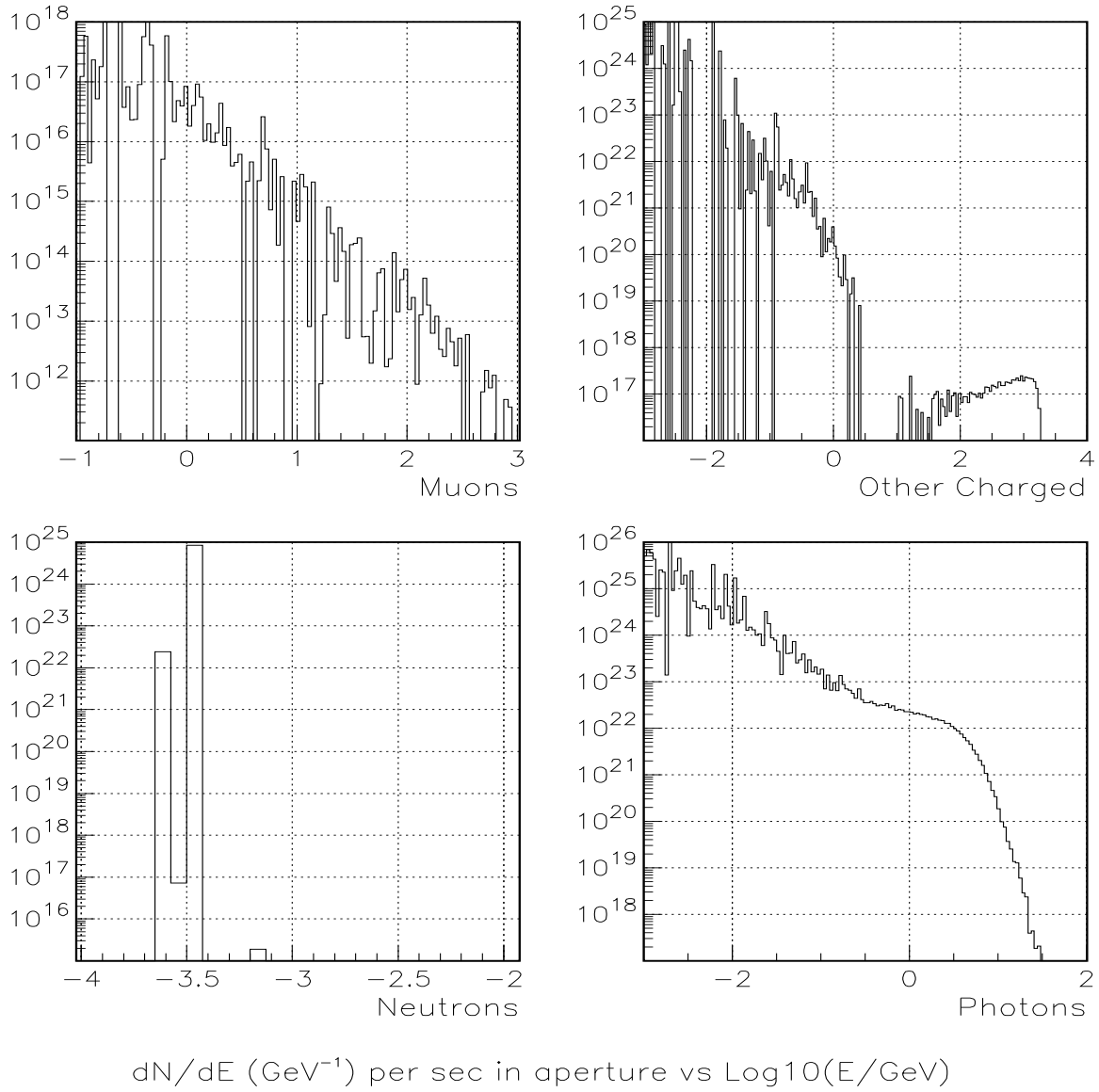


Figure 10.1: Energy spectra of muons, $h^{\pm}$ and $e^{\pm}$, neutrons and photons in the aperture of the arc magnets induced by 2 TeV muon beam decays.

component is muons generated in electromagnetic and hadronic cascades in the magnets. Fig. 10.3 shows isodose contours around the collider tunnel. The distributions are asymmetric in the horizontal plane because of lattice and tunnel curvature and effects of the magnetic field. With 10^7 s as a collider operational year, the tolerable on-site limit in the soil/rock is reached at about 6 m above the orbit plane, 10 m toward the ring center and ~ 75 m outward in the horizontal plane. With 7 m above the ring plane the surface area can even be accessible to the public. All underground facilities (electronics rooms etc.) have to be inside the ring at ≥ 10 m distance from the beam axis. Prompt radiation levels in the experimental hall and

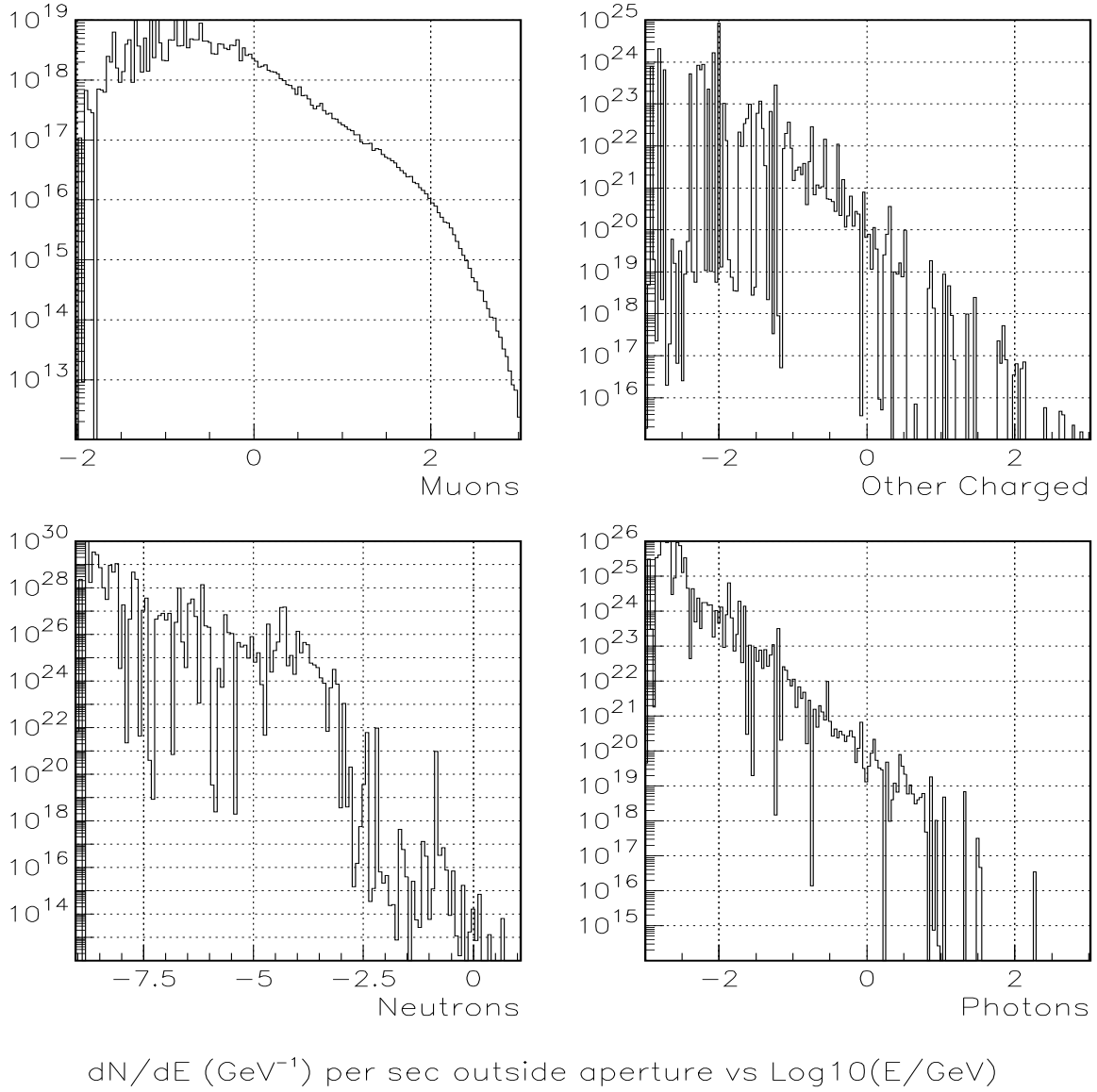


Figure 10.2: Energy spectra of muons, $h^{\pm}$ and $e^{\pm}$, neutrons and photons averaged over the arc magnets, tunnel air and a few meters of the surrounding soil/rock due to 2 TeV muon beam decays.

detector are considered in a separate chapter.

10.4.3 Radioactivation Around Tunnel

In estimating induced radioactivity, the standard approach is based on the number of inelastic nuclear interactions of hadrons with energy ≥ 50 MeV (stars). Star densities in the collider, detector and shielding components can be directly converted into a residual dose rate. Pho-

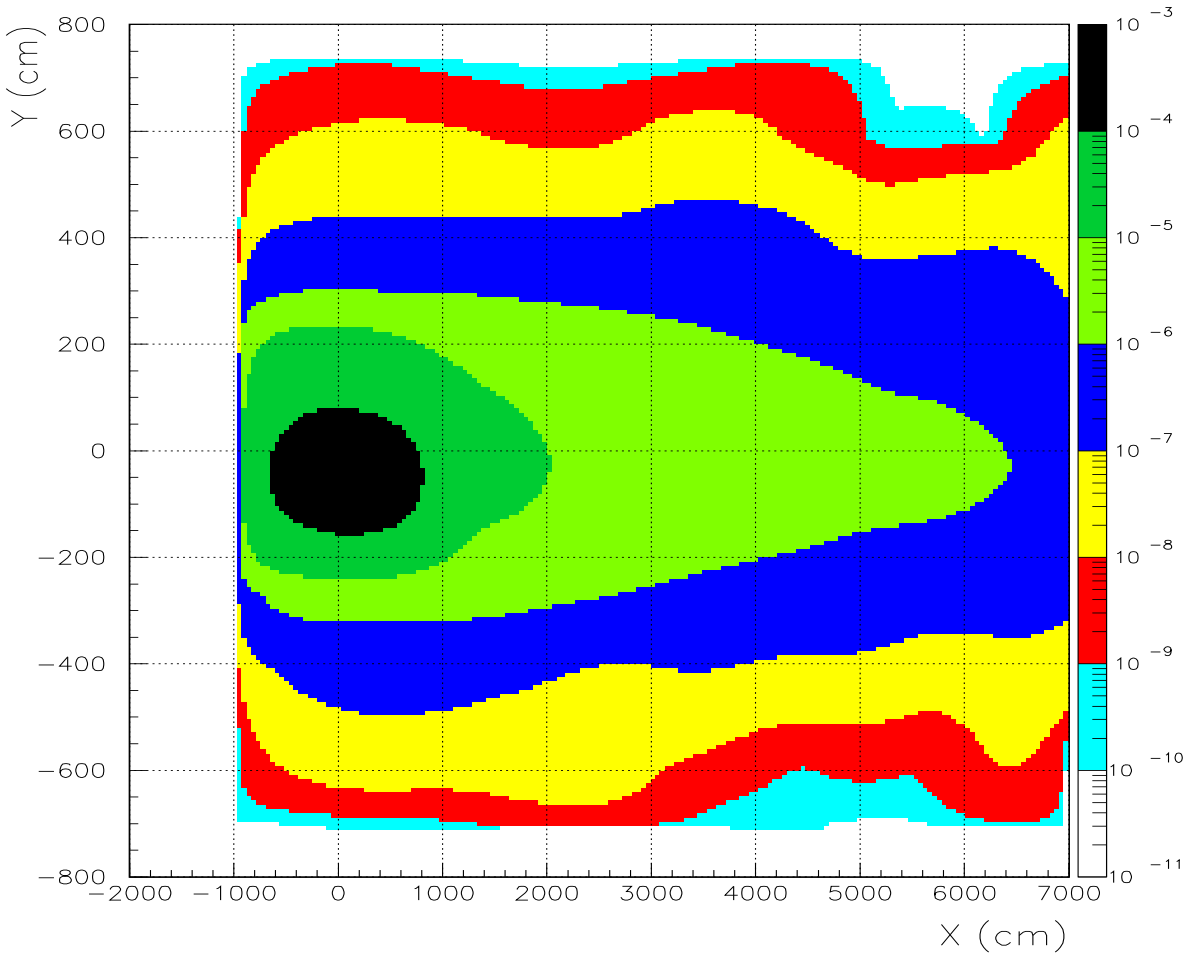


Figure 10.3: Isodose contours in the vertical plane across the collider tunnel and surrounding soil/rock for 2 TeV muon beam decays. y axis is up and x axis points outward along the ring radius. Beam axis is at x=y=0. Right scale is dose rate in rem/sec.

tohadrons in the first layers of the tunnel shielding and hadrons produced along muon tracks farther from the tunnel are a source of soil/rock and groundwater activation around the collider [6]. Two radionuclides, ^{3}H and ^{22}Na, produced in the soil/rock, completely determine the activity concentration that can be found in groundwater. In calculations, isotope production is observed in the first meters around the tunnel, which would require insulation or drainage of that region. The dolomite stratum at Fermilab may naturally satisfy this condition. Further studies are needed.

10.4.4 Radioactivation of Lattice and Detector Components

Due to unavoidable $\mu \to e\nu\overline{\nu}$ decays, about 2 kW of power is deposited in every meter of the collider ring. Generated hadrons induce radioactivation of magnet components. As shown in [2], [5] and in the previous chapter there should be a thick tungsten liner inside the cosine theta magnets to reduce heat load to cryogenics and avoid quench of the superconducting coils. The required thickness of such a liner is 6 cm in the arcs and in the two quadrupoles nearest to the interaction point and about 3 to 4 cm for the most of the interaction region quadrupoles. There is a significant azimuthal dependence of power density and hadron production due to the strong magnetic field. Fig. 10.4 shows the azimuthal distribution of power density in the innermost layer of the tungsten liner in the arc dipoles for a 2 TeV muon beam. The higher peak is due to showers induced by decay positrons (for μ^+ beam) and the

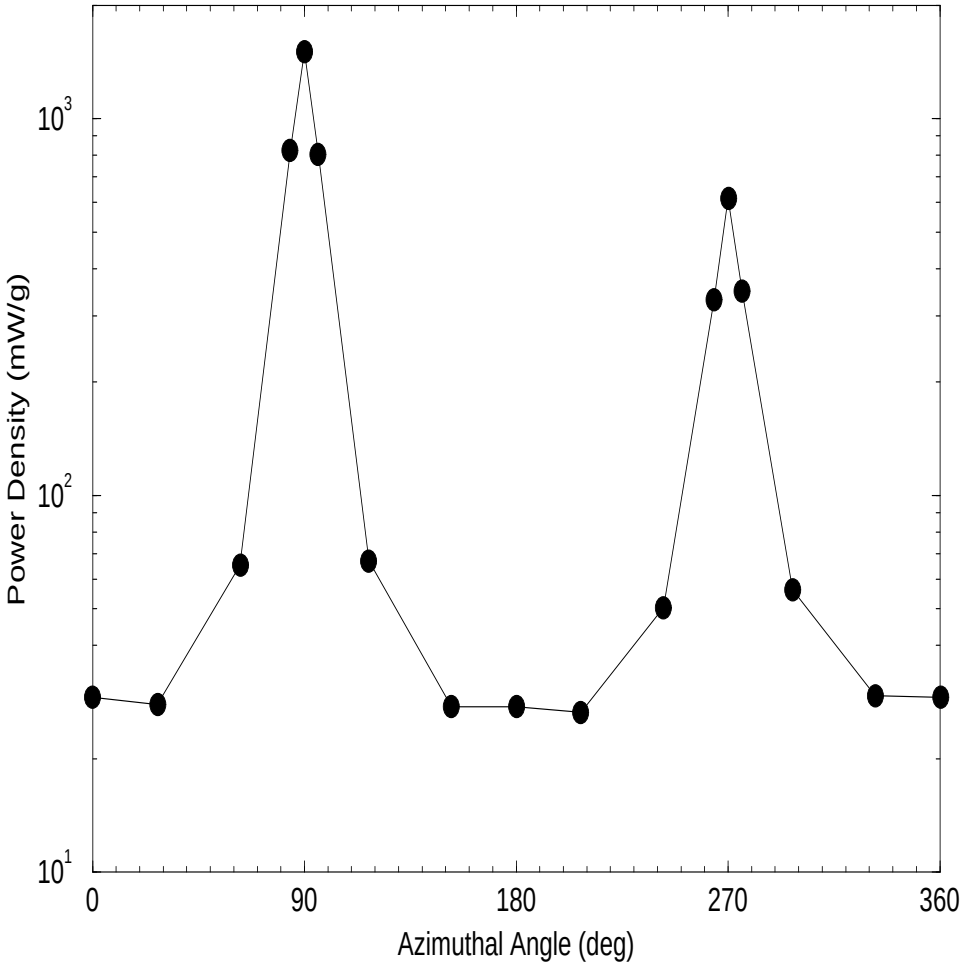


Figure 10.4: Azimuthal distribution of power density in the innermost layer of the tungsten liner inside the arc dipole aperture for 2 TeV muon beam decays.

second peak at the opposite side in the orbit plane is due to synchrotron photons emitted by

positrons. With two beams in the ring, the results presented are simply doubled: electrons hit the inward wall (ϕ=90 deg) and photons hit the outward wall (ϕ=270 deg). Hadron dose and radionuclide production also follow this pattern.

Residual dose rates in magnet components have been estimated assuming a few years of the collider operation and measured immediately after shutdown. With a 6 cm tungsten liner in the arcs, contact dose rate at the inner liner surface is $P_\gamma \sim 9$ rad/hr, at the outer surface $P_\gamma \sim 0.2$ rad/hr, for SC coils and yoke it is $P_\gamma \sim 0.03$ rad/hr, and at the magnet outer shell it is $P_\gamma \sim 0.003$ rad/hr. For the two first quadrupoles in the IR, the tungsten liner is much hotter. The dose rate drops logarithmically with a cooling time (time after shutdown) and at least inversely with the distance from the extended object. For example, the above numbers are 2.3 times smaller 1 day after shutdown.

In the detector region residual dose rate even in the near beam components (tungsten nose) is rather low compared to the machine. Dose rates in the detector drop rapidly with distance from beam axis.

10.5 Spent Beam Absorption

10.5.1 Prompt Radiation

In operating scenarios considered, muons are extracted after about 1000 turns and sent to a beam absorber. Contrary to hadron machines, energy losses extend over a few kilometers (2 TeV case), the absorber needs not be cooled, and spent muon beam can be sent to the soil/rock directly. Fig. 10.5 shows the particle flux attenuation in the soil/rock with a density of ρ=2.24 g/cm^3 for 2 TeV muons. As shown in [6], a characteristic of high-energy muons is the intensive production of electrons, photons and hadrons along the muon vector. The energies of these particles accompanying the muon tracks are sufficient to affect overall flux and dose distributions, and become a source of radionuclide production deep in the rock.

Fig. 10.6 shows isodose contours in the soil/rock for 2 TeV muons. The outer contour coincides with the tolerable on-site dose limit. This is also not very different from the off-site limits. It is 3.55 km long with a maximum width of 18 m at 2.6 km. At small distances, the required radial thickness of soil/rock shielding above the muon beam is 6–7 m which is the same as shielding required for the collider ring. Thus, there is a simple solution to solve the problem: deflect the extracted beam down by 4.5 mrad. With that, muon fluxes are completely confined beneath the ground.

For the 250×250 GeV collider the situation is even easier (see Fig. 10.7): the contour of the allowable on-site dose is only 810 m long with a maximum width of 14.6 m at 700 m.

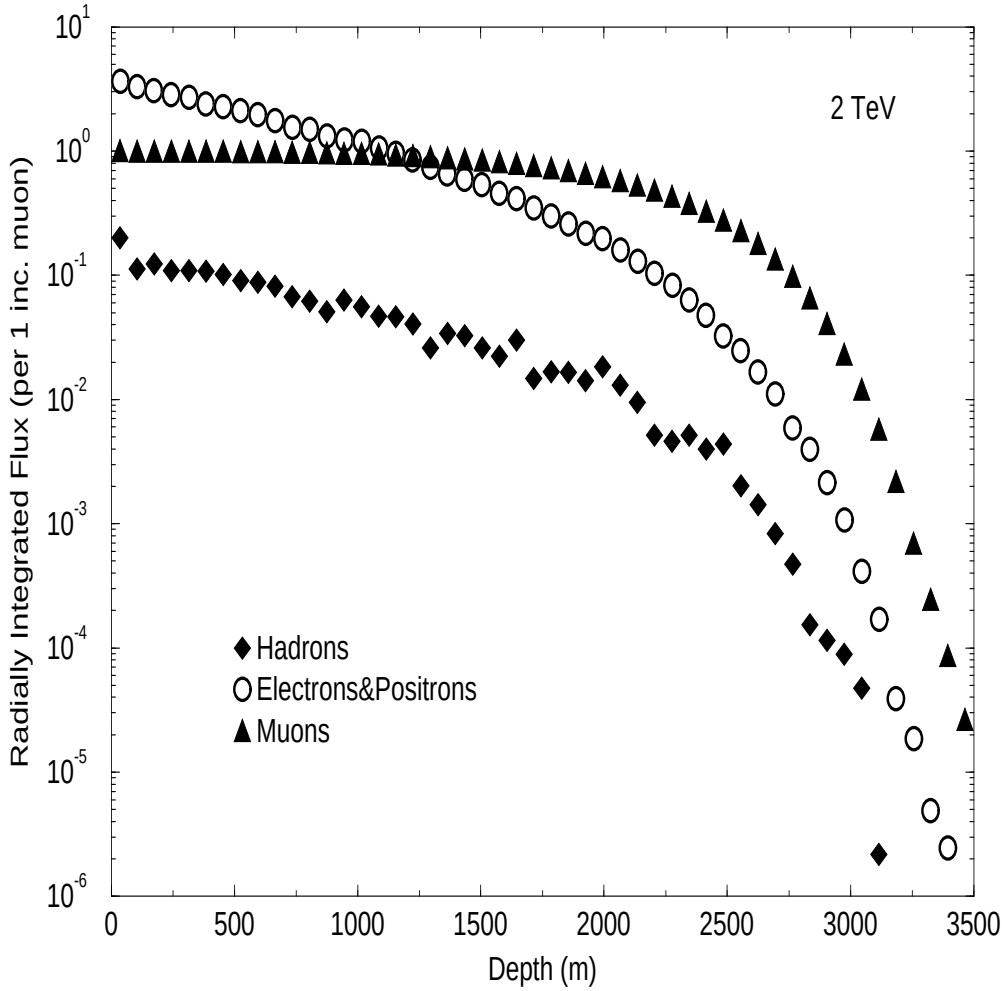


Figure 10.5: Transversely integrated flux of muons, e^+e^- and hadrons in the soil/rock (ρ=2.24 g/cm^3) per one 2 TeV muon beam.

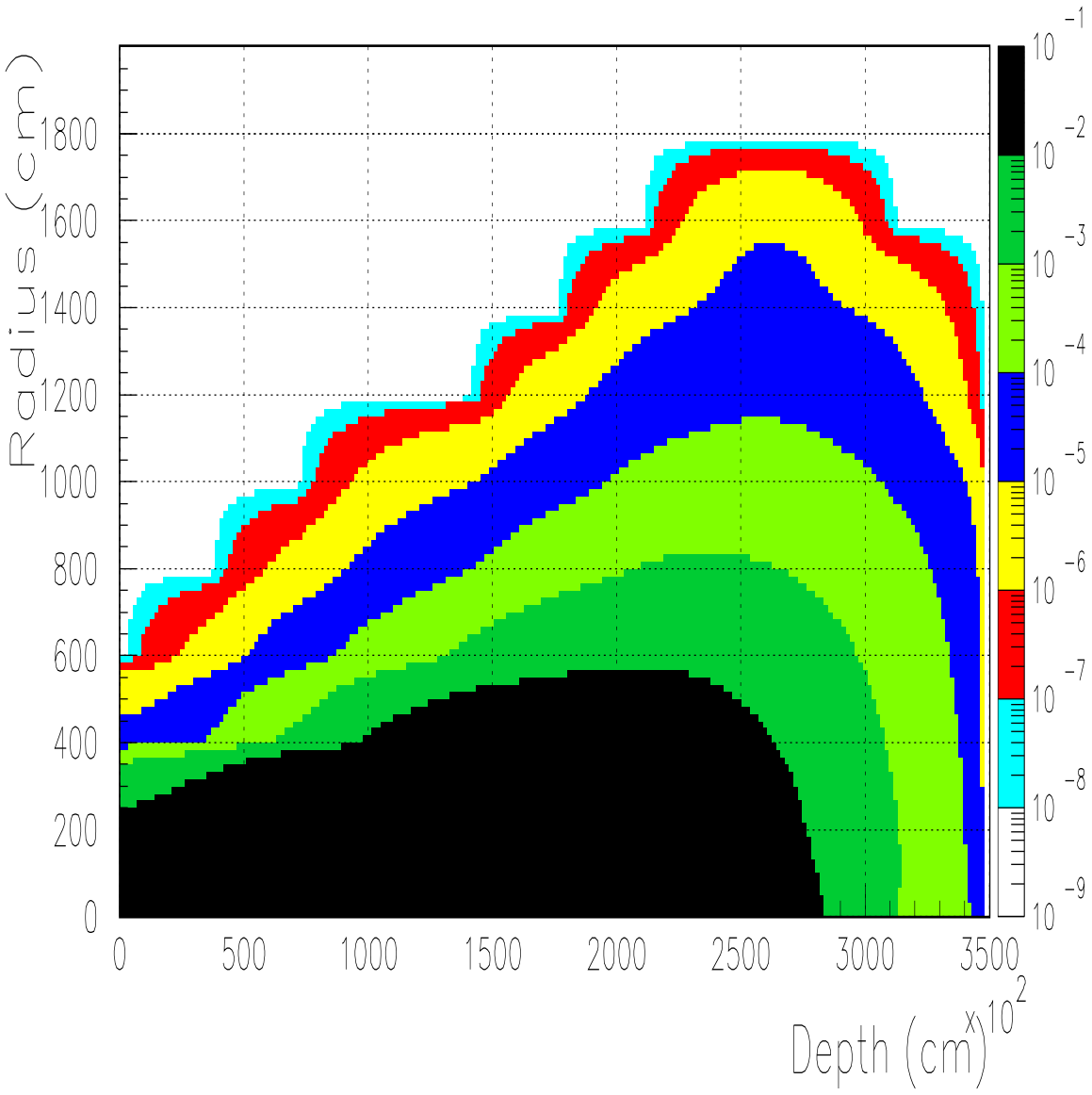


Figure 10.6: Isodose contours in the soil/rock (ρ=2.24 g/cm^3) for 3×10^{13} extracted 2 TeV muons per second. Right scale is dose rate in rem/sec.

Depending on the depth of the ring, $\sim$10 mrad vertical kick down would comply with the regulation requirements.

10.5.2 Radioactivation

As mentioned above, hadrons produced in the interactions along the extracted beam accompany the muon tracks and result in radionuclide production in the soil/rock. Fig. 10.8 and Fig. 10.9 show star density distributions for 2 TeV and 250 GeV muon beams, respectively.

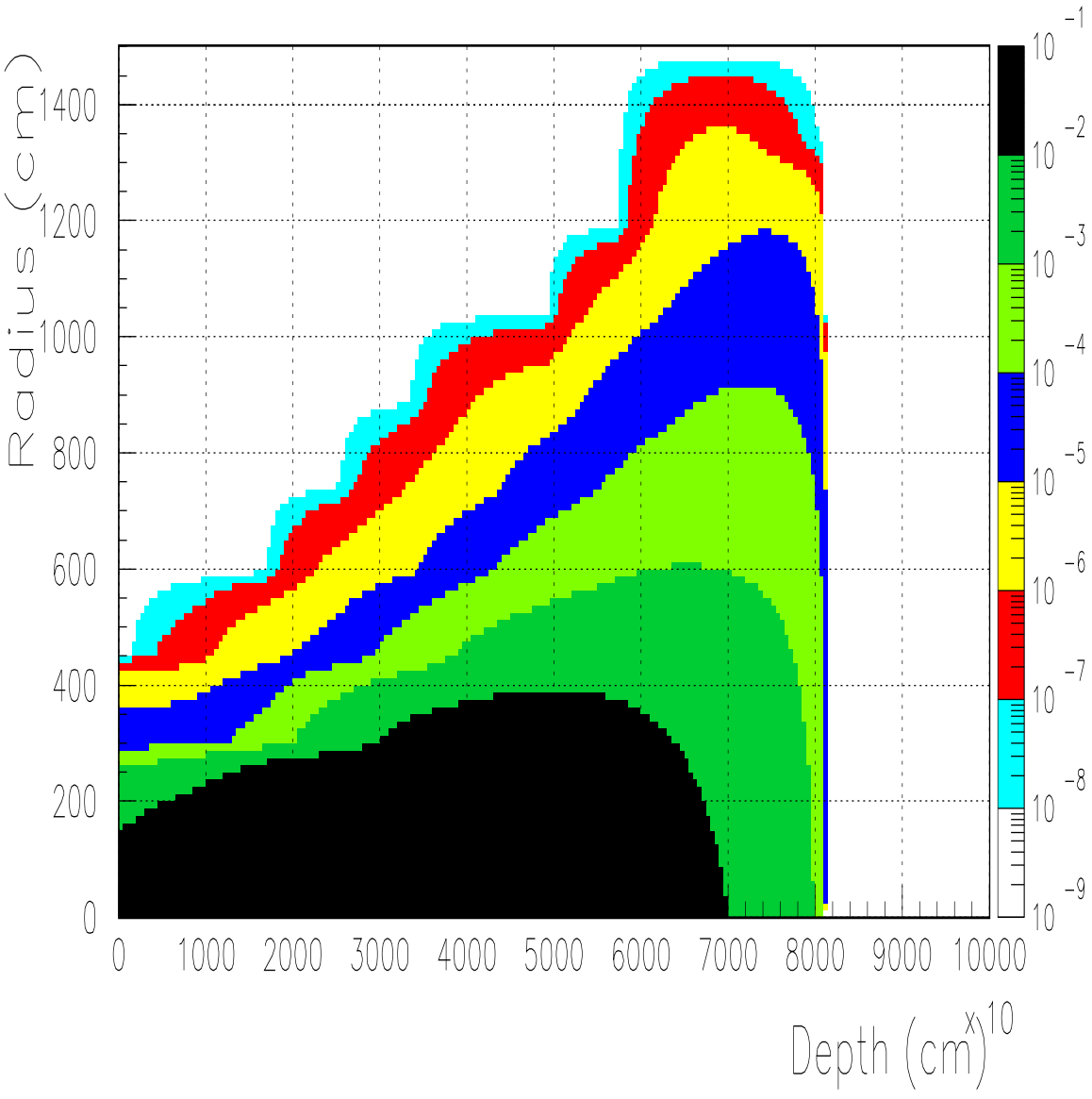


Figure 10.7: Isodose contours in the soil/rock (ρ=2.24 g/cm^3) for 3×10^{13} extracted 250 GeV muons per second. Right scale is dose rate in rem/sec.

Estimates show that at the design parameter operation of a 2×2 TeV collider, the absorp-

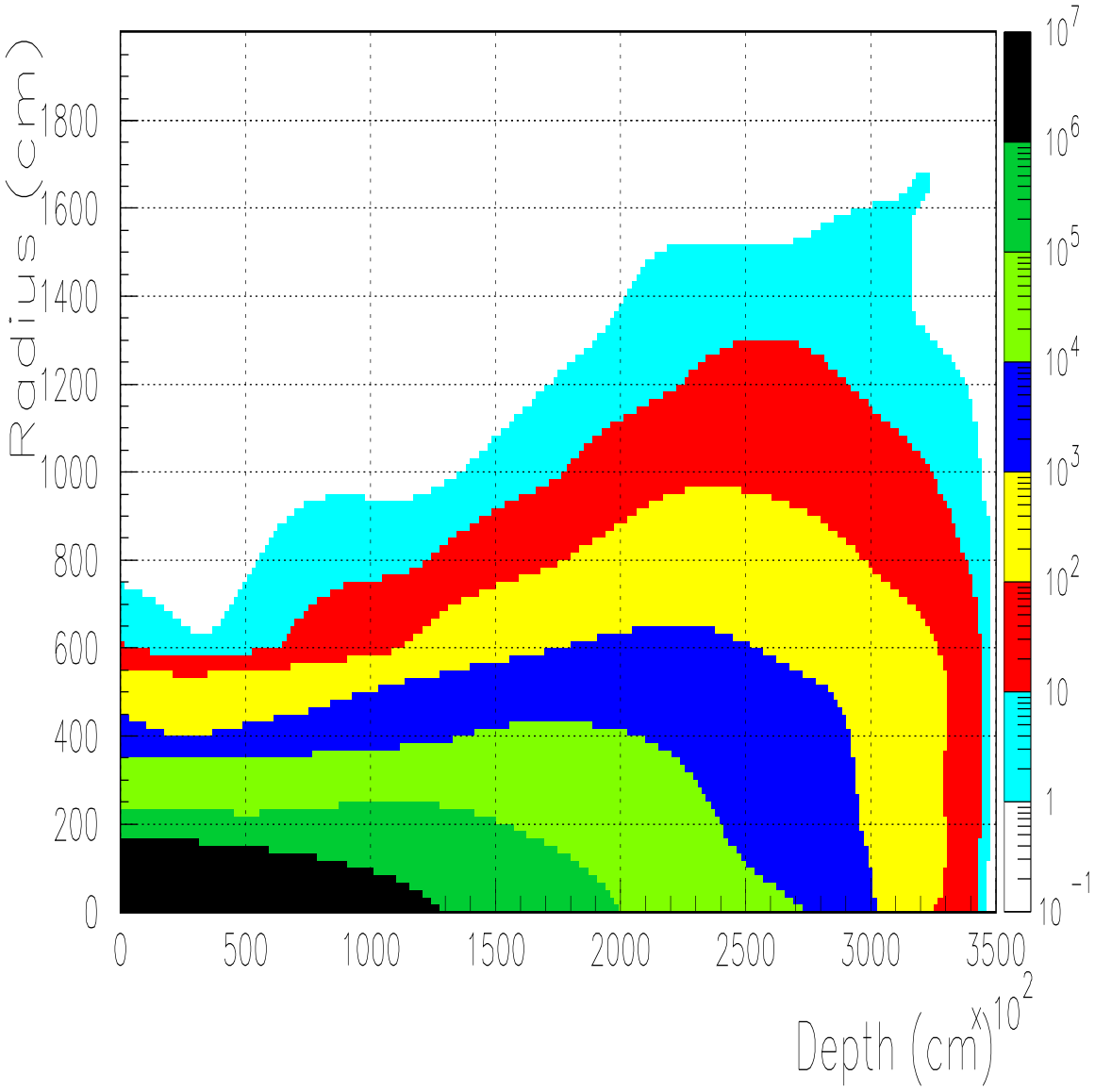


Figure 10.8: Star density contours in the soil/rock (ρ=2.24 g/cm^3) for 3×10^{13} extracted 2 TeV muons per second. Right scale is star density in cm^{-3}s^{-1}.

tion of the spent beam can result in annual activity concentration which may exceed the stringent limits for ^{3}H and ^{22}Na radionuclides, 20 pCi/cm^3 and 0.2 pCi/cm^3 respectively, if the beam disposal lines are in aquiferous layers. The problem is solved if the beam is directed into the impervious dolomite layer (Fermilab) or to an isolated 2.5 km long 2 m radius rock or concrete plug (2 TeV beam). For 250 GeV beam this plug is about 550 m long and 1 m in radius. A steel plug can shorten the indicated dimensions.

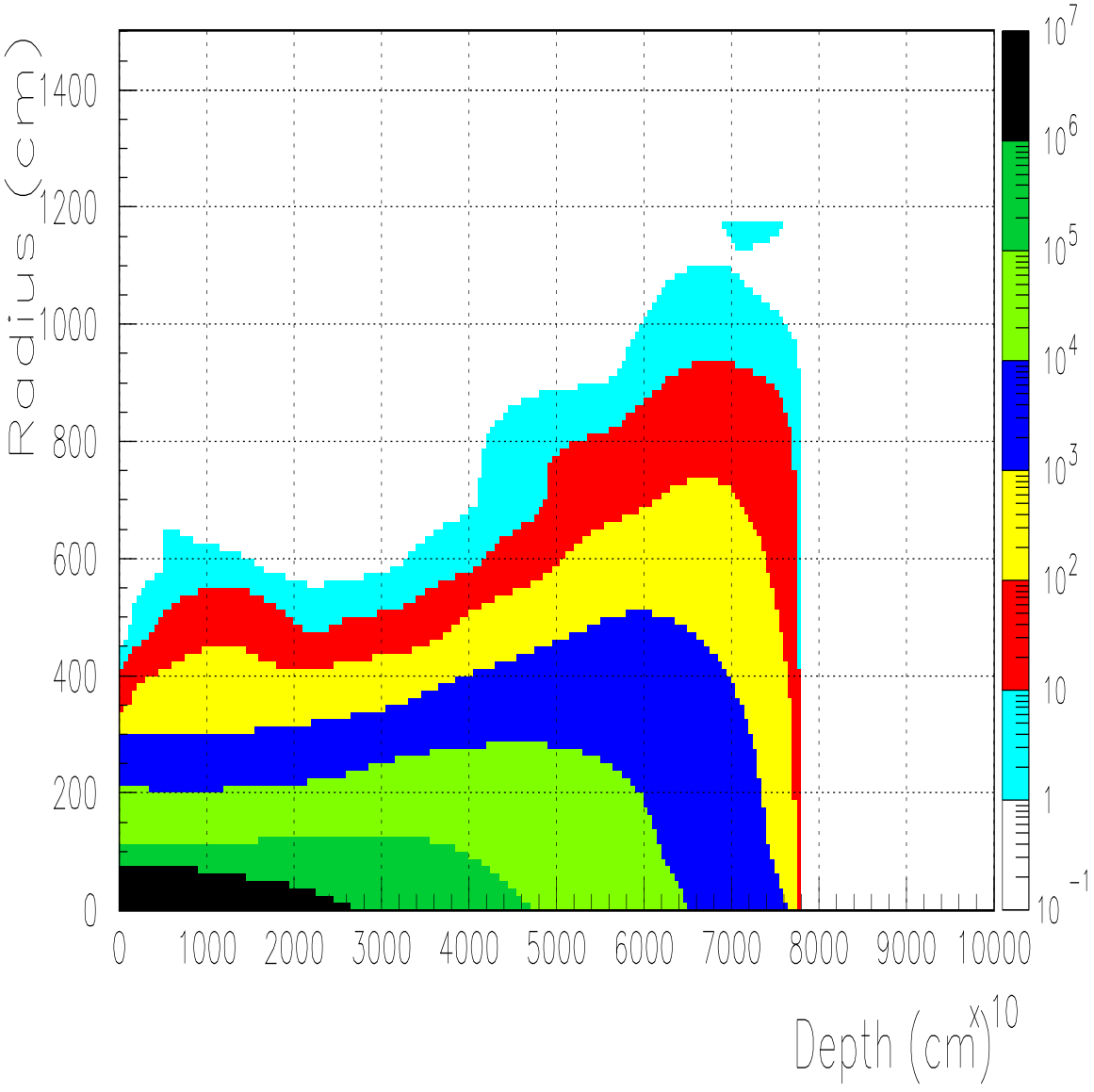


Figure 10.9: Star density contours in the soil/rock (ρ=2.24 g/cm^3) for 3×10^{13} extracted 250 GeV muons per second. Right scale is star density in cm^{-3}s^{-1}.

10.6 Accidental Beam Loss

As at any accelerator facility there is always a non-zero probability for uncontrolled beam loss in a case of a system failure. While the averaged beam power is large, the actual beam energy within the beam pipe in any pulse is relatively small, $\leq 1.4\,\mathrm{MJ}$. This is the largest beam energy that may be lost in a single accident. This is similar to the energy stored in the Tevatron, so similar constraints apply. The big difference is that in all the complex parts except the proton source, the lost particles are the weakly interacting muons, so, contrary to hadron machines, the only concern is a prompt radiation along direction of the lost muon beam. As a first approach the results of the two previous sections are directly applicable to a failure case with a single pulse of $\sim 2 \times 10^{12}$ muons of up to 2 TeV energy. At any location in the facility a single beam accident at 2 TeV and full intensity creates an admissible dose contour in the soil/rock about 2 km long and $\leq 2\,\mathrm{m}$ radius (tangent to the ring if it happens in the collider ring) confined deep beneath the ground within the site boundaries. For the $250 \times 250\,\mathrm{GeV}$ collider, the corresponding contour is much smaller, $\sim 450\,\mathrm{m}$ by $1.5\,\mathrm{m}$.

10.7 Conclusions

In operational mode all radiation problems both for 2 TeV and 250 GeV appear quite solvable. Similarly, no great problems seem to arise from accidental beam loss.

Bibliography

[1] R. Palmer et al., *Muon Collider Design*, BNL–62949, submitted to the Proceedings of the Symposium on Physics Potential and Development of Muon-Muon Colliders, San Francisco, CA, Dec. 1995.

[2] N. V. Mokhov, *The Radiation Environment at Muon Colliders*, in *Proc. of the 2nd Worshop on Simulating Accelerator Radiation Environments (SARE2)*, CERN, October 1995.

[3] N. V. Mokhov, *The* MARS *Code System User's Guide, version 13 (95)*, FNAL–FN–628 (1995).

[4] G. W. Foster and N. V. Mokhov, *Backgrounds and Detector Performance at a 2+2 TeV $\mu^+\mu^-$ Collider*, in *AIP Conference Proceedings*, **352**, pp. 178-190 (1995); also Fermilab–Conf–95/037 (1995).

[5] N. V. Mokhov and S. I. Striganov, *Simulation of Backgrounds in Detectors and Energy Deposition in Superconducting Magnets at $\mu^+\mu^-$ Colliders*, in Proceedings of the 9th Advanced ICFA Beam Dynamics Workshop, Ed. J. C. Gallardo, AIP Conference Proceedings 372, 1996; also Fermilab–Conf–96/011 (1996).

[6] N. V. Mokhov, *Muons at LHC. Part 1: Beam Dumps*, CERN/TIS–RP/TM/95–27 (1995).

Contributors

- N. Mokhov, (FermiLab) Editor

List of Figures

List of Tables

Chapter 11

OPTIONS

Contents

11.1 Introduction

Up to this point, this report has discussed the design of a muon collider with:

1. beam energies of 2 + 2 TeV
2. operating at its maximum energy
3. with a fixed rms energy spread of 0.12%

4. with no attention to maximizing polarization

In this section we discuss modifications to enhance the muon polarization's, operating parameters with very small momentum spreads, operations at energies other than the maximum for which a machine is designed, and designs of machines for different maximum energies. In particular we will give parameters of machines with center of mass energy of 0.5 TeV.

11.2 Polarization

11.2.1 Polarized Muon Production

The specifications and components in the baseline design have not been optimized for polarization. Nevertheless, simple manipulations of parameters and the addition of momentum selection after phase rotation does generate significant polarization with relatively modest loss of luminosity. The only other changes required to give polarization at the interaction point are rotators in the transfer lines, and a solenoid in the collider opposite the IP.

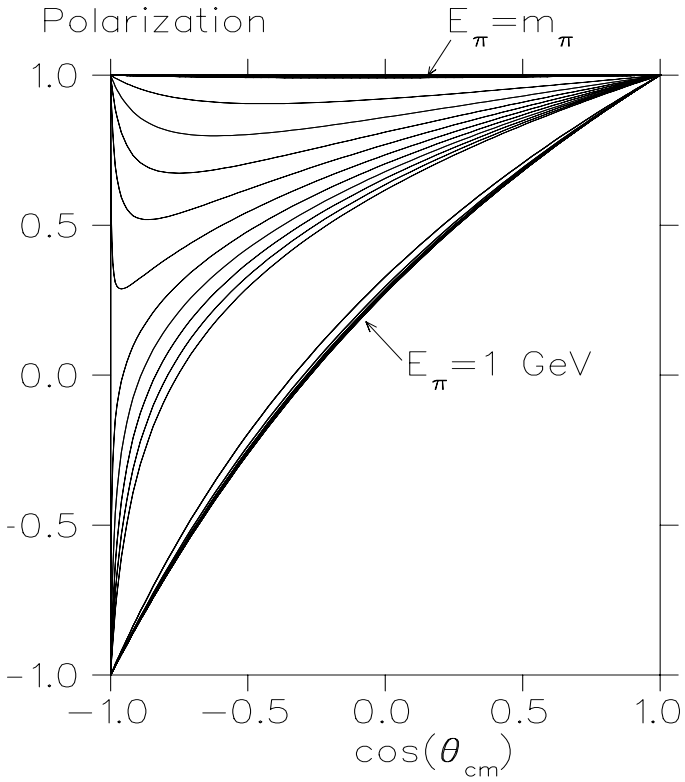


Figure 11.1: Muon polarization in the lab system vs the cosine of the center-of-mass decay angle, for a number of pion energies.

In the center of mass of a decaying pion, the outgoing muon is fully polarized (-1 for μ^+ and +1 for μ^-). In the lab system the polarization depends[1] on the decay angle θ_d and initial pion energy E_π. Fig. 11.1 shows this polarization as a function of the cosine of the center of mass decay angle, for a number of pion energies. It is seen that for pion kinetic energy larger than the pion mass, the dependence on pion energy becomes negligible. The polarization is given by $\mathcal{P} = \cos\omega$ where the Wigner angle ω satisfy the relations

$$\sin\omega = \sin\theta_{\rm cm}\frac{p_\pi^{\rm lab}}{p_\mu^{\rm lab}}\frac{m_\mu}{m_\pi} \tag{11.1}$$

$$\cos\omega = \left\{E_\pi^{\rm lab}p_\mu^{\rm cm} + \cos\theta_{\rm cm}E_\mu^{\rm cm}p_\pi^{\rm lab}\right\}/p_\pi^{\rm lab}m_\pi \tag{11.2}$$

with $p_\pi^{\rm lab}$, $E_\pi^{\rm lab}$ the pion momentum and energy in the laboratory frame and $p_\mu^{\rm lab}$, $E_\mu^{\rm cm}$ the muon momentum in the laboratory frame and energy in the center of mass. A Monte Carlo calculation[2] of the capture, decay and phase rotation gave muon polarization of approximately 0.22.

If higher polarization is required, some deliberate selection of muons from forward pion decays ($\cos\theta_d \to 1$) is required. This could be done by selecting pions within a narrow energy range and then selecting only those muons with energy close to that of the selected pions. But such a procedure would collect a very small fraction of all possible muons and would yield a very small luminosity. Instead we wish, as in the unpolarized case, to capture pions over a wide energy range, allow them to decay, and to use rf to phase rotate the resulting distribution.

Consider the distributions in velocity vs ct at the end of a decay channel. If the source bunch of protons is very short and if the pions were generated in the forward direction, then the pions, if they did not decay, would all be found on a single curved line. Muons from forward decays would have gained velocity and would lie above that line. Muons from backward decays would have lost velocity and would fall below the line. A real distribution will be diluted by the length of the proton bunch, and by differences in forward velocity due to the finite angles of particles propagating in the solenoid fields. In order to reduce the latter, it is found desirable to lower the solenoid field in the decay channel from 5 to 3 T. When this is done, and in the absence of phase rotation, one obtains the distribution shown in Fig. 11.2, where the polarization P$> \frac{1}{3}$, $-\frac{1}{3} < P < \frac{1}{3}$, and P$< -\frac{1}{3}$ is marked by the symbols '+', '.' and '-' respectively. One sees that the +'s are high, and the -'s are low, all along the distribution.

It is found that phase rotation does not remove this correlation: see Fig. 11.3. Now, after a time cut to eliminate decays from high energy pions, there is a simple correlation of polarization with the energy of the muons.

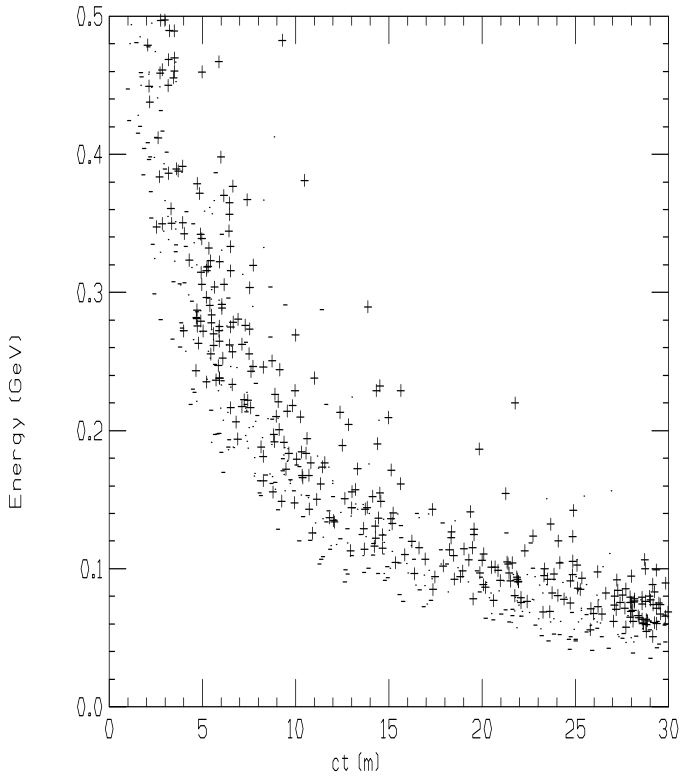


Figure 11.2: Energy vs ct of muons at end of decay channel without phase rotation; muons with polarization P$> \frac{1}{3}$, $-\frac{1}{3} < P < \frac{1}{3}$, and P$< -\frac{1}{3}$ are marked by the symbols '+', '.' and '-' respectively.

If a selection is made on the minimum energy of the muons, then for muons after the required time cut, net polarization is obtained. The higher the cut on energy, the greater the polarization, but the less the fraction $F_{loss} = N_\mu^{\rm OUT}/N_\mu^{\rm IN}$ of muons that are selected. The cut in time can probably be obtained from the phasing of the rf used to capture the bunch. Alternatively, it could be provided by a second energy cut applied after a 90 degree longitudinal phase rotation.

In order to provide the required cut on energy, one needs to generate dispersion that is significantly larger than the beam size. Collimation from one side can then select the higher energy muons. After collimation, the remaining dispersion should be removed. The generation of sufficient dispersion, in the presence of the very large emittance, is non-trivial. The only practical method appears to be the use of a bent solenoid (as discussed above in the target section). First the solenoid is bent in one direction to generate the dispersion; the collimator is introduced; then the solenoid is bend in the other direction to remove the dispersion. The complete system thus looks like an "S" or "snake".

Particles with momentum p_μ in a magnetic field B have a bending radius of R_B, given

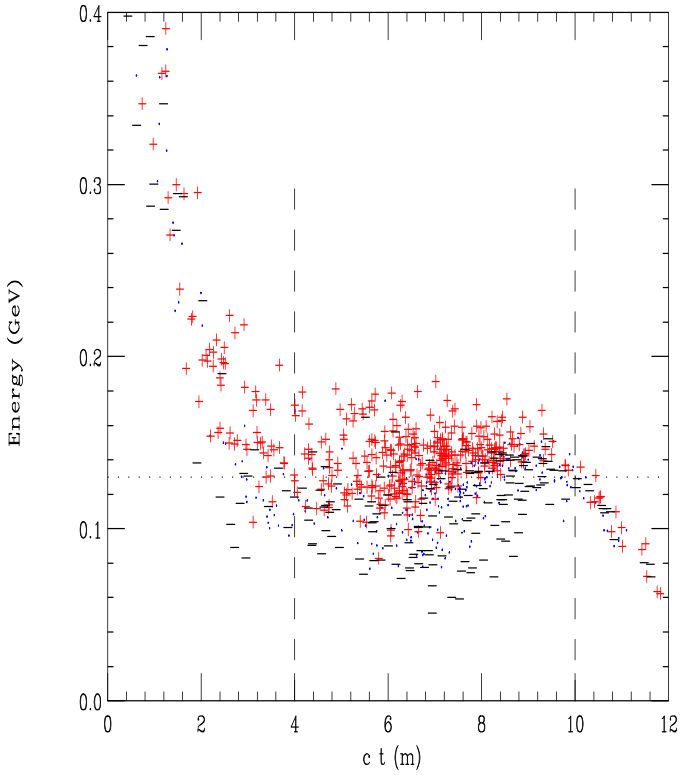


Figure 11.3: Energy vs ct of muons at end of decay channel with phase rotation; muons with polarization P> $\frac{1}{3}$, $-\frac{1}{3} < P < \frac{1}{3}$, and P< $-\frac{1}{3}$ are marked by the symbols '+', '.' and '-' respectively.

by:

$$R_B = \frac{(ep_\mu/mc)}{c\ B}. \tag{11.3}$$

If the particles are trapped in a solenoid with this field, and the solenoid is bent with a radius $R_{\rm bend}$, where

$$R_{\rm bend} >> R_B, \tag{11.4}$$

then those particles, besides their normal helical motion in the solenoid, will drift in a direction (z) perpendicular to the bend, with a drift angle ($\theta_{\rm drift} = dz/ds$) given by:

$$\theta_{\rm drift} \approx \frac{R_B}{R_{\rm bend}} \tag{11.5}$$

The integrated displacement in z, ie. the dispersion D, is then:

$$D = \theta_{drift}\ s \approx \phi\ R_B, \tag{11.6}$$

where ϕ is the total angle of solenoid bend.

As an example, we have traced typical particles with momenta of 150 and 300 MeV/c through a snake with $\phi = \pi$, $B = 1\,T$, and $R_{\rm bend} = 6\ m$. Fig. 11.4a shows the trajectories of muons as viewed from the z direction. No significant dispersion is seen. Fig. 11.4b shows the same trajectories, where the vertical positions z are plotted against s, the distance along the snake. The two momenta are seen to be dispersed during the right hand turn and recombined by the left hand turn. Negligible emittance growth is observed during the bends, but a small growth of emittance is observed at the sudden (and, in this simulation, unphysical) field gradient changes at the start and end of the bends. Even these emittance growths are acceptable providing $R_{\rm bend} > 10\ R_B$.

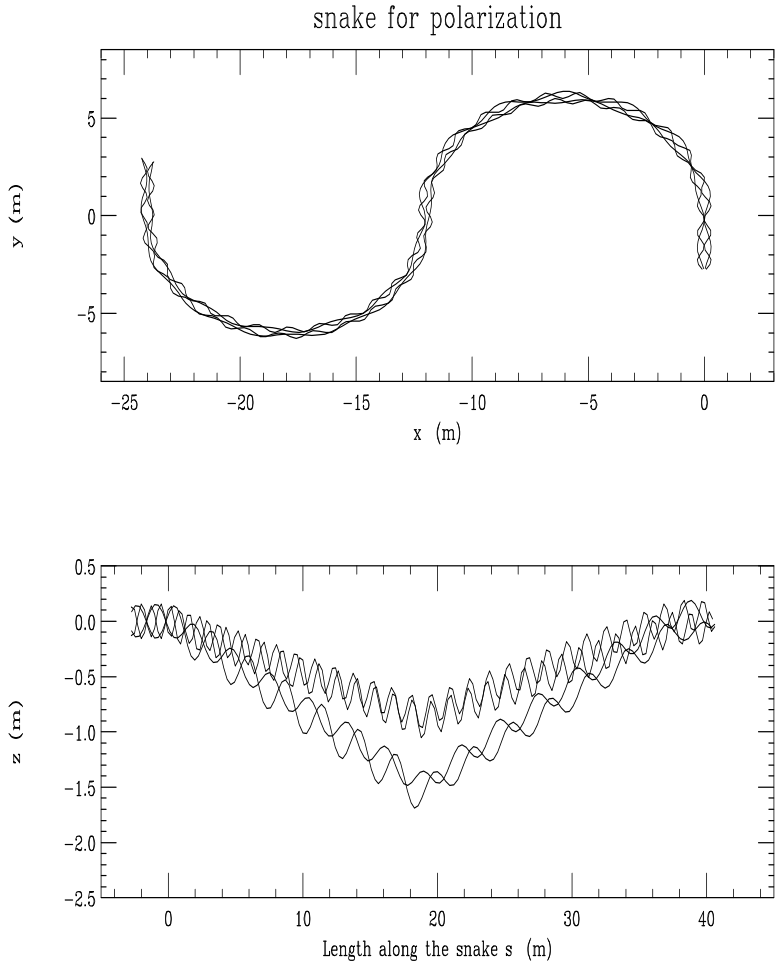


Figure 11.4: Dispersion snake: trajectories in the bending plane as seen from the perpendicular direction z (upper plot); trajectories in the vertical plane, z plotted against length along the snake s (lower plot).

Fig. 11.5 and Tb. 11.1 give the results of a Monte Carlo study in which dispersion is introduced, and progressive cuts applied, to the muons at the end of the phase rotation.

In this calculation, in order to shorten the computation time, the trajectories were not actually traced through a snake. Instead, the particles were propagated through 20 m of straight solenoid, followed by the application of dispersion equal to 6 times the momentum in GeV/c. A snake that would give such dispersion could have the parameters: solenoid field: 3 T, giving $R_B = 0.25\ m$ at the average momentum of 230 MeV/c. The diameter of the snake bends should be greater than 5 m. The bend angle required is 320 degrees, which would require some variations in bend curvature to avoid the solenoid crossing its self, but is not impractical.

Tb. 11.1 gives results for two fields in the decay channel solenoids: $5\,T$, the field in the point design; and $3\,T$, chosen to increase the polarization. It is seen that for weak cuts and small polarization, it is better to avoid the loss of muons from the lower, $3\,T$, field, but with stronger cuts the lower field gives greater polarization. In Fig. 11.5, and subsequent plots, only data from the preferred fields are shown beyond the cross over.

Table 11.1: Production polarization vs collimator position.

B (T)	cut (m)	$F_{\rm loss}$	$P_{\rm init}$	$P_{\rm final}$	$P_{\rm vec}$	$R_{v/s}$	$H_{\rm vec}$	$E_{\rm ave}$ (MeV)	δE (MeV)
5	0.00	1.000	0.23	0.18	0.36	1.45	1.03	130	23
5	1.00	0.960	0.27	0.21	0.41	1.54	1.05	144	23
5	1.12	0.890	0.30	0.24	0.46	1.64	1.06	147	20
5	1.24	0.759	0.36	0.29	0.53	1.80	1.08	151	18
5	1.30	0.614	0.41	0.33	0.60	1.99	1.11	157	17
5	1.40	0.360	0.48	0.39	0.67	2.26	1.15	166	15
5	1.50	0.163	0.56	0.45	0.75	2.64	1.20	177	15
3	0.00	0.801	0.22	0.18	0.34	1.43	1.03	130	22
3	1.06	0.735	0.29	0.23	0.44	1.61	1.05	133	22
3	1.16	0.673	0.35	0.28	0.52	1.77	1.08	137	19
3	1.26	0.568	0.41	0.33	0.59	1.98	1.11	141	17
3	1.32	0.417	0.50	0.40	0.69	2.32	1.16	147	15
3	1.40	0.264	0.59	0.47	0.77	2.78	1.22	151	13
3	1.48	0.126	0.70	0.56	0.86	3.58	1.32	159	13
3	1.56	0.055	0.77	0.62	0.90	4.25	1.38	168	12

It is seen from Tb. 11.1 that the energy cut not only increases the polarization, but also decreases the energy spread δE of the remaining muons. In Fig. 11.6 the fractional energy spread $\delta E/E$ is plotted against the loss factor $F_{\rm loss}$. The energy spread is reduced almost

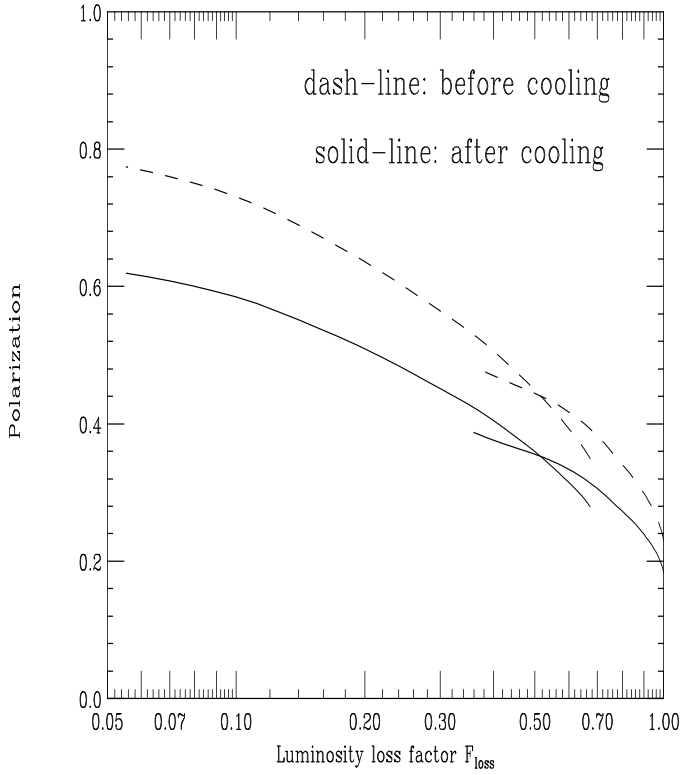


Figure 11.5: Polarization vs $F_{\rm loss}$ of muons accepted; the dashed line shows polarization as selected before cooling; the solid line gives polarization after cooling.

a factor of two for reasonable collimator positions. This reduction in energy spread would eliminate the need for the first stage of emittance cooling.

A Monte Carlo study has also been done on the effect of variations of the proton bunch length σ_t. In this study, the dispersion was not specifically introduced. Instead, polarization was generated by imposing an idealized cut on muon energies. The results of this, compared with using dispersion and a position cut, are qualitatively the same. Figure 11.7a shows the polarization before cooling as a function of σ_t for three values of the loss factor F_{loss}. It is seen that serious loss of polarization occurs when the rms width is more than 1 ns. Figure 11.7b shows the muon rms energy spread after the polarization cut. Again it is shown as a function of σ_t for three values of the loss factor F_{loss}. With no cut, the rise in energy spread would be serious ($\delta E > 20$ MeV is difficult to cool) for an rms width more than 1 ns. But with polarization cuts, the energy spread is so reduced that a larger proton bunch length would not be a problem.

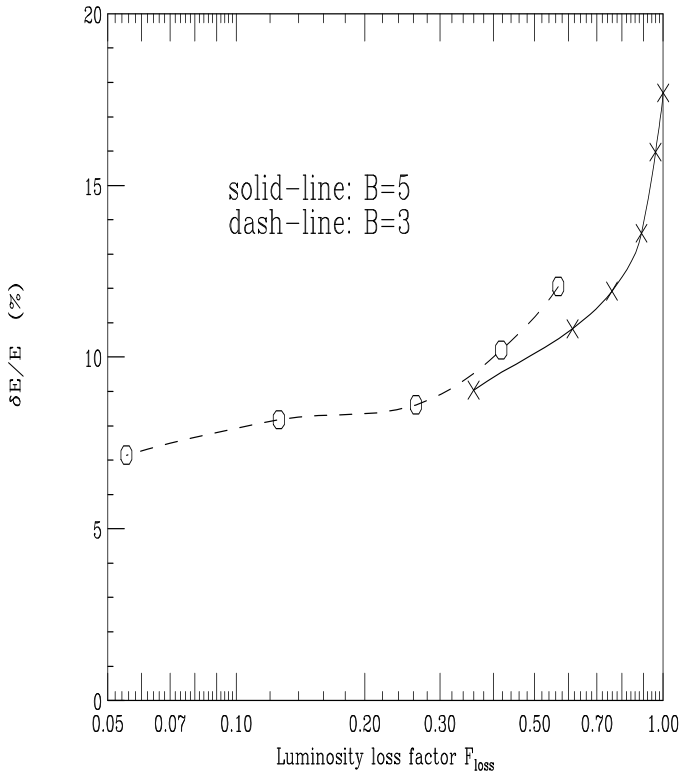


Figure 11.6: The fractional energy spread $\Delta E/E$ is plotted against the loss factor $F_{\rm loss}$ for different magnetic fields.

11.2.2 Polarization Preservation

A recent paper[3] has discussed the preservation of muon polarization in some detail. During the ionization cooling process the muons lose energy in the material and have a spin flip probability $\mathcal{P}$, where

$$\mathcal{P} \approx \int \frac{m_e}{m_\mu}\beta_v^2 \, \frac{dE}{E} \tag{11.7}$$

where β_v is the muon velocity divided by c, and dE/E is the fractional loss of energy due to ionization loss. In our case the integrated energy loss is approximately 3 GeV and the typical energy is 150 MeV, so the integrated spin flip probability is close to 10%. The change in polarization dP/P is twice the spin flip probability, so the reduction in polarization is approximately 20 %. This dilution is included in the $P_{\rm final}$ column in Tb.11.1 and is plotted as the line in Fig. 11.5.

During circulation in any ring, the muon spins, if initially longitudinal, will precess by $\gamma(g-2)/2$ turns per revolution in the ring; where $(g-2)/2$ is 1.166 10^{-3}. An energy spread $\delta\gamma/\gamma$ will introduce variations in these precession and cause dilution of the polarization. But if the particles remain in the ring for an exact integer number of synchrotron oscillations,

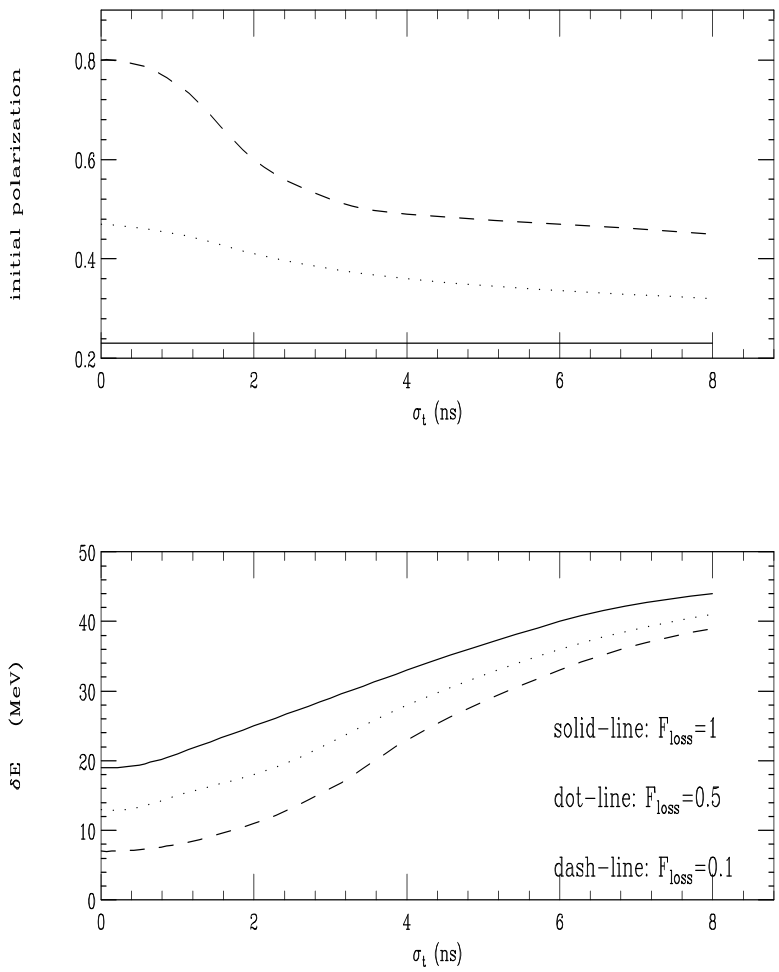


Figure 11.7: Polarization vs σ_t, the proton bunch length (upper plot). Muon *rms* energy spread vs σ_t (lower plot). Both plots for three values of the loss factor $F_{\rm loss}$.

then their individual average γ's will be the same and no dilution will occur. It appears reasonable to use this *synchrotron spin matching*[3] to avoid dilution during acceleration. In the collider, however, the synchrotron frequency will be too slow to use *synchrotron spin matching*, so one of two methods must be used:

- Bending can be performed with the spin orientation in the vertical direction, and the spin rotated into the longitudinal direction only for the interaction region. The design of such spin rotators appears relatively straightforward. The example given in the above reference would only add 120 m of additional arc length, but no design has yet been incorporated into the lattice.

- The alternative is to install a 120 m 10 T solenoid (Siberian snake) at a location exactly opposite to the intersection point. Such a solenoid reverses the sign of the horizontal

polarization and generates a cancellation of the precession in the two halves of the ring.

Provision must also be made to allow changes in the relative spins of the two opposing bunches. This could be done, prior to acceleration, by switching one of the two beams into one or the other of two alternative injection lines.

11.2.3 Benefits of Polarization of Both Beams

We consider two examples of the general advantage of having polarization in both beams. Individual physics experiments would have to be considered to determine how important such advantages are.

Consider the polarization of a vector spin state generated by the annihilation of the two muons.

$$P_{\rm vec} \;=\; \frac{F^{++} - F^{--}}{F^{++} + F^{--}} \tag{11.8}$$

When only one beam has polarization P_1, then $P_{\rm vec} = P_1$. But if both beams have polarization P in the same direction (ie. with opposite helicities), then

$$P_{\rm vec} \;=\; \frac{(P+1)^2 - (P-1)^2}{P+1)^2 + (P-1)^2} \tag{11.9}$$

In Fig. 11.8 both the polarization of each beam P, and the resulting polarization of a vector state $P_{\rm vec}$ are plotted against the loss factor F_{loss}.

A second advantage is that the ratio $R_{\rm vec/sc}$ of vector to scalar cross section can be manipulated to enhance either the vector or the scalar state. If the polarization directions have been chosen to enhance the ratio of vector to scalar states, then:

$$R_{v/s} \;=\; \frac{1+P}{1-P}. \tag{11.10}$$

Tb. 11.1 and Fig. 11.9 show this ratio as a function of the loss factor $F_{\rm loss}$.

Tb. 11.1 also shows that the fraction of total luminosity in a given state can be enhanced. If polarizations are chosen to enhance the vector state, then the fraction of vector luminosity is increased from 1/2 to $(1+P)/2$, ie. the enhancement factor $H_{\rm vec} = (1+P)$, but this is seen to be only a modest effect.

11.2.4 Luminosity Loss

If nothing else is done, then the luminosity will drop as $F^2_{\rm loss}$; where $F_{\rm loss}$ is the fraction muons lost by the muon momentum cut. At the same time, however, the space charge, wakefield, and loading during the cooling and acceleration will all be reduced; as will the beam beam

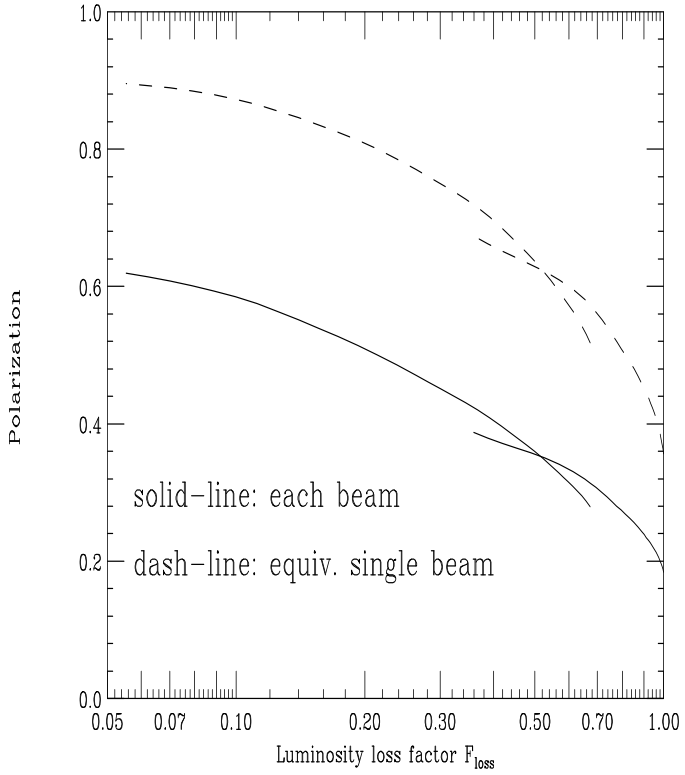


Figure 11.8: Polarization of each beam P, and the resulting polarization of a vector state $P_{\rm vec}$ vs. the loss factor $F_{\rm loss}$.

tune shift in the collider. Clearly, the cooling could now be reoptimized and some part of the lost luminosity recovered.

An alternative way to recover the luminosity would be to increase the proton bunch intensity by the factor $F_{\rm loss}$. If this were done, then the original number of muons per bunch would be generated; all the wake field, loading and space charge effects would be the same; and the luminosity per bunch crossing would be the same. If we assume that the total proton current is determined by the driver, then such an increase in proton intensity per bunch will necessitate a reduction in the number of bunches or repetition rate, by the same factor $F_{\rm loss}$. The luminosity will then fall by this factor, without the square.

For instance, in the unpolarized case of the 4 TeV collider, there were two bunches of each sign. If the momentum cut is chosen to give a value of $F_{\rm loss} = 1/2$, and the proton beam is distributed into 2 instead of 4 initial bunches, then the final number of muons per bunch, the loading, beam beam tune shift etc, would all be the same as in the unpolarized case. The luminosity would be down by a factor of only two, for polarization of 34 % in both beams.

For higher polarization at good luminosity it would be desirable to have a proton source

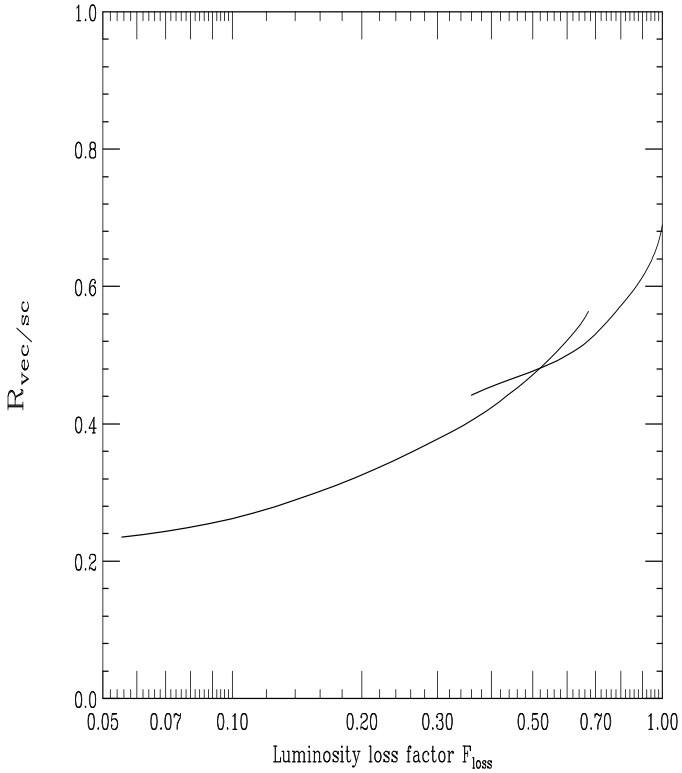


Figure 11.9: The ratio of vector to scalar states, $R_{v/s}$ vs the loss factor $F_{\rm loss}$

with the option of a lower repetition rate, but even larger numbers of protons per spill. For example 4×10^{14} protons per pulse at 4 Hz. It should then be possible to extend this method to an operation with $F_{loss} = 1/8$, and polarization of both beams of 57 %.

One also notes that the luminosity could be maintained at the full unpolarized value if the proton source intensity could be increased. Such an increase in proton source intensity in the unpolarized case would be impractical because of the resultant excessive high energy muon beam power, but this restriction does not apply if the increase is used to offset losses in generating polarization. If, for instance, the driver repetition rate were increased from 15 to 30 Hz, the fractions $F_{\rm loss}$ set at 0.5, and the number of bunches reduced to one, then the full luminosity of 10^{35} ($\rm cm^{-2}s^{-1}$) would be maintained with polarization of both beams of 34 %.

11.2.5 The Case for Polarized $\mu^+\mu^-$ Colliders

Higgs Physics

The most interesting question in particle physics now is associated with the origin of mass. It is generally assumed that the exchange of fundamental scalar particles, called the "scalar

sector" is somehow responsible for this. For super-symmetry modes, this scalar sector is even more complex and interesting (see Tb.11.2)[4]-[6].

In this section, we highlight one of the most interesting goals of a $\mu^+\mu^-$ collider: the discovery of a Higgs boson in the mass range beyond that to be covered by LEP I & II ($\sim$ 80–90 GeV) and the natural range of the supercolliders.

With a high-mass t quark, precision LEP/SLD data and the theorists' dreams of a SUSY world, the scalar (pseudoscalar sector) is possibly very complex and may require several types of colliders[7]. Consider:

- If the low-mass Higgs has $m > 130$ GeV, MSSM is not allowed.
- If $m > 200$ GeV, there are constraints from the requirement that perturbation theory be useful up to very high energy and from the stability of the vacuum.
- If $m < 130$ GeV, MSSM is possibly allright, but we may expect other particles (H, A) and the width of the low mass Higgs may change.
- The scalar sector may be extremely complex, requiring pp (LHC) and $\mu^+\mu^-$ colliders (and possibly NLC and $\gamma\gamma$ colliders).
- In high energy collisions, vector states are allowed unless a special method is used. Consider $\mu^+\mu^-$ colliders with polarized $\mu^\pm$

Table 11.2: The scalar sector

	(100–500) GeV	scalars $(H, A, ...)$
	$\nearrow$	
$\mu^+\mu^-$		
	$\searrow$	
	≥ 2 TeV	W^+W^-
		Z^0Z^0 production in scalars

This cannot be done for pp or e^+e^- colliders.

- A $\mu^+\mu^-$ collider is complimentary to the LHC/CMS detector.

There are several ways to determine the approximate mass of the Higgs boson in the future[7]. Suppose it is expected to be at a mass of 135 $\pm$ 2 GeV, the energy spread of a

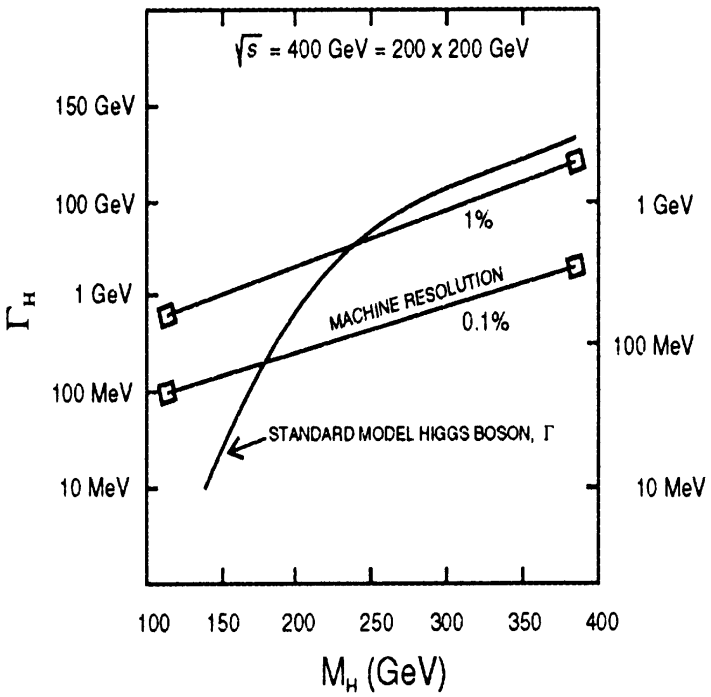


Figure 11.10: Higgs search at a $\mu^+\mu^-$ collider (required machine resolution and the expected Higgs width).

$\mu^+\mu^-$ collider can be matched to the expected width (see Fig.11.10). An energy scan could yield a strong signal to background especially with polarized $\mu^+\mu^-$ in the scalar configuration [5],[7]. Once the Higgs is found, the following could be carried out:

1. Measurement of width, to separate Standard Model Higgs from SUSY or other Higgs models[5],

2. Measurement of the Branching fractions, the rare decay will involve loop effects that can sample very high energies.

Polarization will play an essential role for any $\mu^+\mu^-$ collider [5],[6]

Production of Polarized $\mu^\pm$ Beams

Polarization is natural for $\mu^\pm$, since they are produced in weak decays and are initially fully polarized because of the V–A interaction. There are three proposed methods for producing intense polarized $\mu^\pm$ beams:

- Accelerate polarization and cool the $\pi^\pm$ (A. Skrinsky *et al.*)[8]

Table 11.3: Depolarization processes in the $\mu^+\mu^-$ complex (Norum Rossmanith scheme)

	Plan	Comments
1.	Decay channel (some acceleration also)	Use synchrotron spin matching; Small effect on polarization.
2.	Cooling channel ($P \sim P_0\ e^{(-K/a)}$, where $a \sim 200$ m for Be).	Depolarization $\sim (m_e/m_\mu)\beta^2$; For $\beta \sim 0$, the effect is very small.
3.	During acceleration to 250 GeV or 2 TeV (CEBAF-type recirculation does not cause severe depolarization)	Cross integer resonances (3 for 250GeV, 21 for 2 TeV); Two effects cause depolarization during acceleration: energy spread and resonances.
4.	Depolarization in collider	Possible large depolarization; spin rotation will be needed to keep small

- Use $K^\pm$ decays and "narrow-band neutrino-like beam" (D. Cline)
- Use pion decays and a short proton bunch (R. Palmer *et al.*)[9].

Fig.11.5 shows the tradeoff between intensity and polarization in one of these schemes[9]. This is one of the major areas of research for $\mu^+\mu^-$ colliders.

Polarization Preservation in the $\mu^+\mu^-$ Collider Complex

R. Rossmanith[3] has presented a scheme to ensure polarization preservation. Because of the value of the $(g-2)_\mu$ for the $\mu^\pm$, it should be much easier to maintain large polarization provided certain steps are taken in the collider complex[3]. The preferred polarization state up to the high energy collider will be longitudinal. These steps are outlined here in Tb.11.3 and in Fig.11.11[3]. It is extremely good news that a $\mu^+\mu^-$ collider with high polarization may be feasible, provided the initial problem of achieving high polarization at the source is solved. This is one of the major problems of a $\mu^+\mu^-$ collider.

In Fig.11.11a it is shown a possible scenario for arranging the spin rotators in order to obtain varying helicity directions from interaction to interaction. After the particles have passed the two spin rotators surrounding the interaction region, the spin aims in the opposite direction and changes its direction the next time it passes the interaction region. This means that polarized interactions with a low systematic error can be obtained. In Fig.11.11b it is

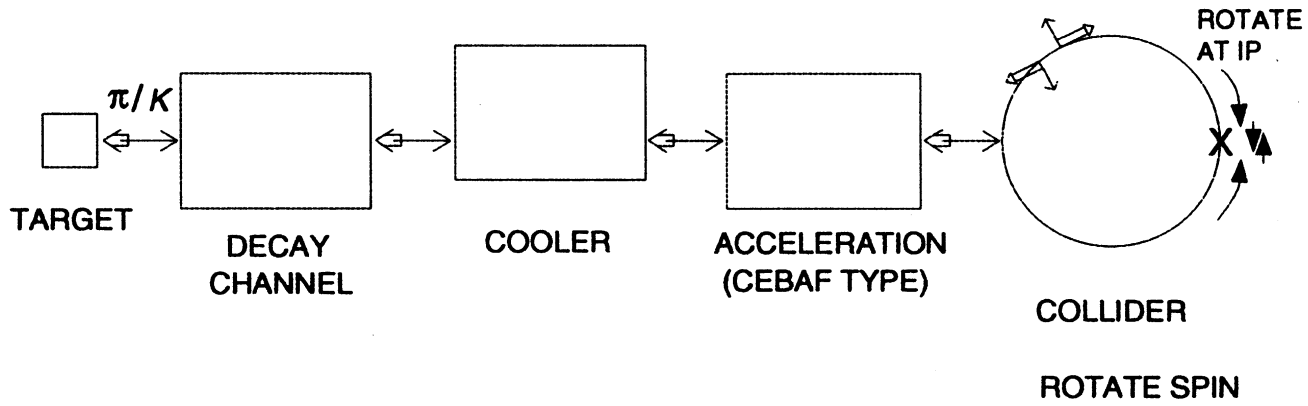


Figure 11.11: (A) A possible scenario for arranging the spin rotators; (B) Possible spin rotator for muons in the main ring

shown a possible spin rotator for the main ring. The spins are rotated 45 degrees from the vertical towards the momentum axis by the first three ca. 10-T, 10-m-long vertically deflecting magnets. The spin is afterwards rotated by 180 degrees around the vertical axis by 12 normal bending magnets and finally into the longitudinal direction by the last three vertically deflecting magnets. The additional space requirement for the spin rotators is 120 m on each side of the interaction region. H and V denote, respectively, horizontally and vertically deflecting magnets

t Quark Physics

The status of the t quark study from FNAL for the CDF detector was reviewed by D. Amidei[10]. In Fig.11.12, we show that Feynman diagrams for the production of $t\bar{t}$ for both e^+e^- and $\mu^+\mu^-$ collisions[6]. Because of the larger mass of the μ compared to the e, the diagram with a scalar intermediate state can be important (see process depicted in Fig.11.12b). If we fully polarize the $\mu^+\mu^-$ system to give a net zero scalar state, we believe the scalar sector will be enhanced to the point that a measurable asymmetry will be generated. Thus, one could search for evidence of a scalar particles far from the central mass. This is a unique feature of polarized $\mu^+\mu^-$ colliders[6]. The detector design will play a crucial role in such studies[11],[12].

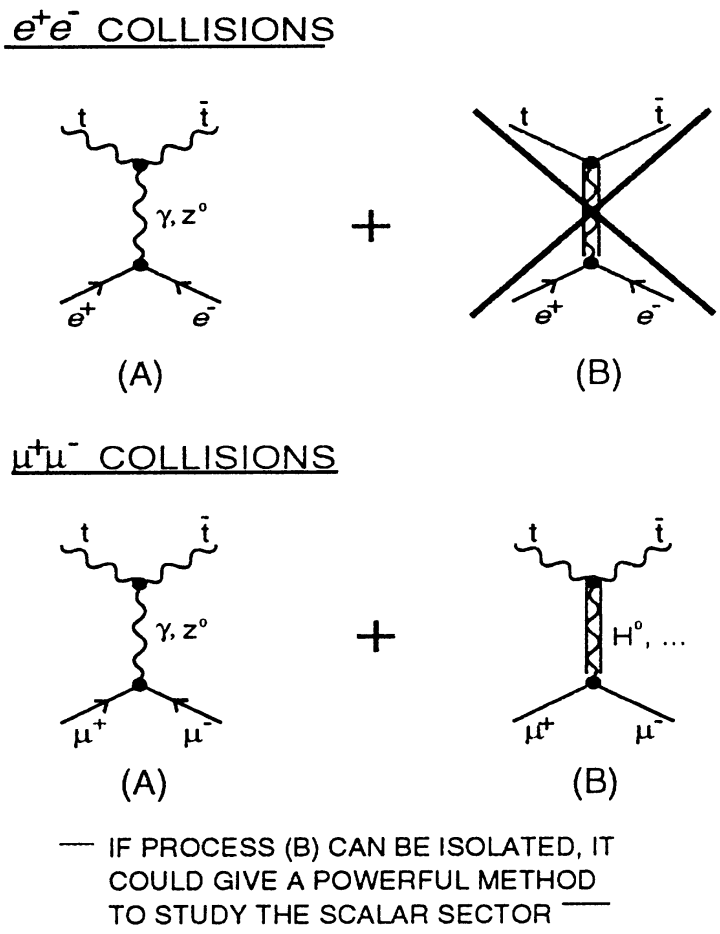


Figure 11.12: $t\bar{t}$ production at $\mu^+\mu^-$ colliders.

11.3 Scaling of Luminosity vs Energy and Momentum Spread

The bunch populations decay exponentially, yielding an integrated luminosity equal to its initial value multiplied by an *effective* number of turns $n_{\rm eff} = 150\ B$, where B is the mean bending field in T.

The luminosity is given by:

$$\mathcal{L} \;=\; \frac{\mathrm{N}_\mu^2\ \mathrm{f_{rep}}\ \mathrm{n_{eff}} \mathrm{n_b} \gamma}{4\pi\ \beta^*\ \epsilon_\mathrm{n}} \mathrm{H(A,D)} \tag{11.11}$$

where N_μ is the number of muons in one bunch, n_b is the number of bunches and the enhancement factor $H(A,D)$ is[13]

$$H(A,D) \approx 1 + D^{1/4}\left[\frac{D^3}{1+D^3}\right]\left\{\ln\left(\sqrt{D}+1\right) + 2\ln\left(\frac{0.8}{A}\right)\right\}, \tag{11.12}$$

with $A = \sigma_z/\beta^*$, and $D = \frac{\sigma_z n_\mu}{\gamma\sigma_t^2} r_e(\frac{m_e}{m_\mu})$.

In the cases we are considering: $A \approx 1$, $D \approx 0.5$ and $H(A, D) \approx 1$.

11.3.1 Luminosity vs Energy, for a Given Ring

For a fixed collider lattice, operating at energies lower than the design value, the luminosity will fall as γ^3. One power comes from the γ in Eq.11.11; a second comes from n_e, the effective number of turns, that is proportional to γ; the third factor comes from β^*, which must be increased proportional to γ in order to keep the beam size constant within the focusing magnets. The bunch length σ_z must also be increased proportional to γ so that the required longitudinal phase space is not decreased; so A $= \sigma_z/\beta^*$ remains constant.

11.3.2 Scaling for Collider Rings for Different Energies

As noted above, the luminosity in a given ring will fall as the third power of the energy at which it is operated. Such a drop is more rapid than the gain in typical cross sections, and, as we shall see, it is more rapid than the drop in luminosity obtained with rings designed for the lower energies. It would thus be reasonable, having invested in a muon source and accelerator, to build a sequence of collider rings at spacings of factors of 2-3 in maximum energy. We will now derive scaling rules for such collider rings.

The luminosity

$$\mathcal{L} \;=\; \frac{N_\mu^2\; n_{eff} n_b f_{rep}\; \gamma}{4\;\pi\;\epsilon_n\;\beta^*} \;\propto\; \frac{N_\mu\; I_\mu\; \gamma}{\epsilon_n\;\beta^*} \tag{11.13}$$

which, since $\Delta\nu_{bb}$, the beam beam tune shift is given by:

$$\Delta\nu_{bb} \;\propto\; \frac{N_\mu}{\epsilon_n}, \tag{11.14}$$

gives:

$$\mathcal{L} \;\propto\; \frac{I_\mu\;\Delta\nu_{bb}\;\gamma}{\beta^*} \tag{11.15}$$

where $I_\mu = N_\mu n_b f_{rep}$ is the muon flux.

If a final focus multiplet is scaled keeping the relative component lengths and the pole tip fields constant, then one obtains:

$$\ell^* \;\propto\; \sqrt{a_{max}\;\gamma} \tag{11.16}$$

$$\theta^* \;\propto\; \sqrt{\frac{a_{max}}{\gamma}} \propto \sqrt{\frac{\epsilon_n}{\beta^*\;\gamma}} \tag{11.17}$$

$$\beta^* \;\propto\; \frac{\epsilon_n}{a_{max}} \tag{11.18}$$

where θ^* is the rms angle of muons diverging from the focus. ℓ^* is the free space from the target to the first quadrupole (proportional to all quadrupole lengths in the multiplet), and a_{max} is the maximum aperture of any quadrupole (proportional to all apertures in the multiplet).

The normalized emittance ϵ_n is constrained by the ionization cooling, but since one can exchange transverse and longitudinal emittance, it is, in principle, the six dimensional emittance ϵ_6 that is constrained. Extending the lepton emittance conventions, we define:

$$\epsilon_6 \;=\; (\epsilon_n)^2 \; \delta\sigma_z\gamma\beta_v. \tag{11.19}$$

where $\delta = \frac{\delta p}{p}$. With this definition, the area of the six dimensional phase space is given by, $\Phi_6 \;=\; \pi^3 \; m_\mu^3 \; \epsilon_6$. σ_z cannot be large compared with the focus parameter β^*, so, taking them to be proportional to one another, and taking the normalized velocity $\beta_v = 1$, then:

$$\epsilon_6 \;\propto\; (\epsilon_n)^2 \; \delta\beta^*\gamma \tag{11.20}$$

and from the above:

$$(\epsilon_n)^3 \;\propto\; \frac{\epsilon_6 \; a_{max}}{\gamma \; \frac{dp}{p}} \tag{11.21}$$

$$(\beta^*)^3 \;\propto \frac{\epsilon_6}{\gamma \; \delta \; a_{max}^2} \tag{11.22}$$

11.3.3 Six Dimensional Emittance Dependence on N_μ and ϵ_n

The six dimensional emittance ϵ_6 obtained from the cooling will, because of more detailed constraints, depend to some extent on the number of muons n_μ, and on the final transverse emittance ϵ_n.

The dependence on the number of muons is relatively straightforward. As the number of muons per bunch rises, the longitudinal space charge forces increase and it becomes impossible, without changing the rf systems, to maintain the same bunch lengths. As a result the bunch lengths must be increased by the square root of the number of muons.

A study, using the analytic formulae, was used to derive cooling sequences with differing parameters. First, sequences were calculated with numbers of initial muons per bunch of 1, 2, 3.75, 7.5, and 15 x 10^{12} (corresponding to muons in the collider of .1, .2, 1, 2, and 4 x 10^{12}). The final transverse emittance at the end of the cooling was required to be 4 10^{-5} m, (corresponding to an emittance in the collider of 5 10^{-5} m). The six dimensional emittances obtained are plotted in Fig. 11.13a. It is seen that for $N_\mu > 10^{12}$ the six dimensional emittances are indeed approximately proportional to the root of the number of muons (the line shows this dependence).

The study also obtained cooling sequences giving six dimensional emittances for a range of final transverse emittances. The dependence here is more complicated. If emittance exchange between longitudinal and transverse emittances could be achieved without material then the six dimensional emittance should be independent of the final transverse emittance chosen. But the exchange does require material wedges, and Coulomb scattering in those wedges increases the six dimensional emittances; and it does so to a greater extent if the transverse emittance is small. In Fig. 11.13b, we show the six dimensional emittances obtained for 5 representative transverse emittances. Over the range of interest the dependence of ϵ_6 is approximately the inverse root of ϵ_n (the line shows this dependence).

For the purposes of this study, we may thus assume that:

$$\epsilon_6 \ \propto \ \sqrt{\frac{N_\mu}{\epsilon_n}} \tag{11.23}$$

11.3.4 Energy Scaling, Allowing the Emittances to Vary

If N_μ is limited by the beam beam tune shift:

$$N_\mu \ \propto \ \epsilon_n \ \Delta\nu_{bb} \tag{11.24}$$

substituting this in equation 11.23:

$$\epsilon_6 \ \propto \ \sqrt{\Delta\nu_{bb}} \tag{11.25}$$

giving:

$$\epsilon_n \ \propto \ \Delta\nu_{bb}^{1/6} \ \left(\frac{a_{\max}}{\gamma \ \delta}\right)^{1/3} \tag{11.26}$$

$$\beta \ \propto \ \frac{\epsilon_n}{a_{\max}} \tag{11.27}$$

$$n_\mu \ \propto \ (\Delta\nu_{bb})^{1\frac{1}{6}} \ \left(\frac{a_{\max}}{\gamma \ \delta}\right)^{1/3} \tag{11.28}$$

$$\tag{11.29}$$

so:

$$\mathcal{L}|_{\Delta\nu} \ \propto \ I_\mu \ \gamma^{4/3} \ \Delta\nu_{bb}^{5/6} \ a_{\max}^{2/3} \ (\delta)^{1/3} \tag{11.30}$$

One notes however that as γ or δ decreases the required number of muons N_μ rises, and will at some point become unreasonable. If we impose a maximum number of muons $N_{\max}$, then, when this bound is reached,

$$\epsilon_n \propto \ N_{\max}^{1/7} \ \left(\frac{a_{\max}}{\gamma \ \delta}\right)^{2/7} \tag{11.31}$$

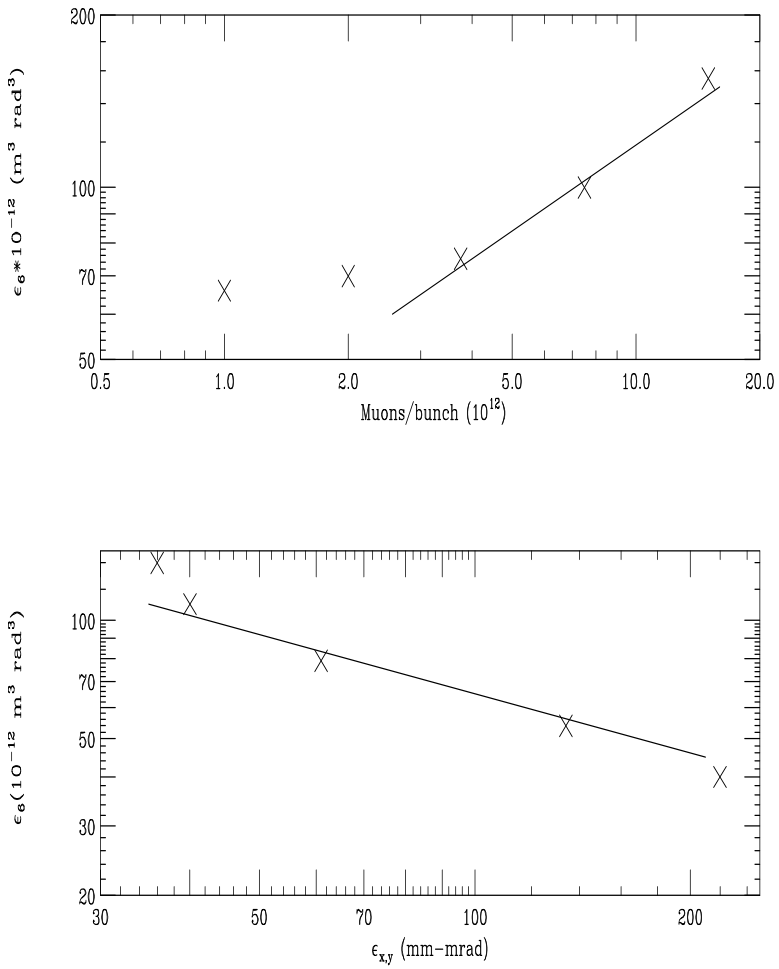


Figure 11.13: Six-dimensional emittance ϵ_6 vs a) muon intensity N_μ entering the cooling section; b) the transverse emittance ϵ_n at the end of the cooling section.

$$\beta \propto \frac{\epsilon_n}{a_{\max}} \tag{11.32}$$

and:

$$\mathcal{L}|_{N_\mu} \propto I_\mu N_{\max}^{12/7} \gamma^{11/7} a_{\max}^{3/7} \left(\frac{\delta p}{p}\right)^{4/7} \tag{11.33}$$

Using the above relationships. and assuming a constant value of $a_{\max}$ we obtain the scaled parameters for a sequence of colliding rings given in Tb. 11.4. Fig.11.14 shows the luminosities that would be available at all energies, including those requiring the use of rings at energies less than their maximum. The lines and dashed lines indicate the luminosities with a bound on N_μ of 4 10^{12}. The line gives luminosities for the nominal rms $\delta p/p$ of 0.12 %, while the dashed line is for a $\delta p/p$ of 0.01 %.

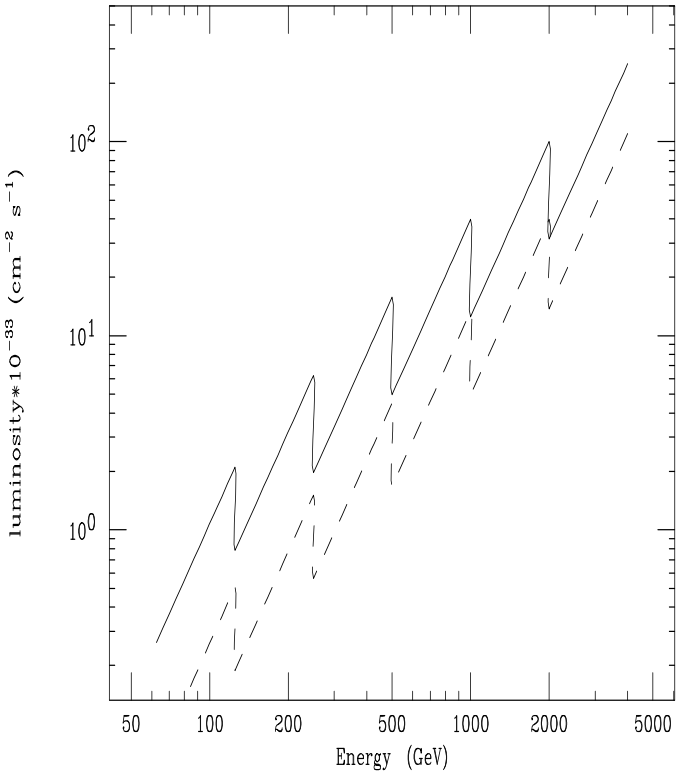


Figure 11.14: Luminosity vs energy assuming rings spaced by factors of two in energy; the line is for $\delta E/E = 0.12\,\%$, the dashed line is for $\delta E/E = 0.01\,\%$.

Table 11.4: Scaling of parameters with energy and momentum spread

E (GeV)	Luminosity ($cm^{-2}s^{-1}$)	emittance (m-rad)	$\#\mu$ (10^{12})	$\delta\nu$	β (mm)	len* (m)	$\beta_{\max}$ (km)	chrom	$\delta E/E$ (%)
4000	2.5E+35	4.0E-05	1.6	0.040	2.4	9.2	882	12829	0.12
2000	1.0E+35	5.0E-05	2.0	0.040	3.0	6.5	350	3600	0.12
1000	4.0E+34	6.3E-05	2.5	0.040	3.8	4.6	139	1010	0.12
500	1.6E+34	7.9E-05	3.2	0.040	4.8	3.3	55	283	0.12
250	6.3E+33	1.0E-04	4.0	0.040	6.0	2.3	22	80	0.12
125	2.1E+33	1.2E-04	4.0	0.033	7.3	1.6	9	23	0.12
4000	1.1E+35	9.1E-05	3.6	0.040	5.5	9.2	385	5604	0.01
2000	4.0E+34	1.1E-04	4.0	0.036	6.7	6.5	156	1603	0.01
1000	1.3E+34	1.4E-04	4.0	0.029	8.2	4.6	64	465	0.01
500	4.5E+33	1.7E-04	4.0	0.024	10.0	3.3	26	135	0.01
250	1.5E+33	2.0E-04	4.0	0.020	12.2	2.3	11	39	0.01
125	5.1E+32	2.5E-04	4.0	0.016	14.9	1.6	4	11	0.01

Bibliography

[1] K. Assamagan, et al., Phys Lett. **B**335, 231 (1994); E. P. Wigner, Ann. Math. **40**, 194 (1939) and Rev. Mod. Phys., **29**, 255 (1957).

[2] R. Palmer et al., *Muon Collider Design*, *Proceedings of the Symposium on Physics Potential and Development of* $\mu^+\mu^-$ *Colliders*, San Francisco, CA (1995); Supplement to Nuclear Physics B, to be published.

[3] B. Norum and R. Rossmanith, *Polarized Beams in a Muon Collider*, *Proceedings of the Symposium on Physics Potential and Development of* $\mu^+\mu^-$ *Colliders*, San Francisco, CA (1995); Supplement to Nuclear Physics B, to be published.

[4] D. Cline, Nucl. Instrum./Meth., A 350 (1994) 24, and the following four papers (pp. 27-56), which constitute a mini-conf. proceeding of the Napa meeting.

[5] V. Barger *et al.*, "Particle Physics Opportunities at $\mu^+\mu^-$ Colliders," in *Physics Potential & Development of* $\mu^+\mu^-$ *Colliders* (Proc., 3rd Int. Conf., San Francisco, Dec. 1995), Elsevier, in press.

[6] D. B. Cline, "Physics Potential of Polarized $\mu^+\mu^-$ Colliders," to be published in the Proceedings of the Montauk Workshop on $\mu^+\mu^-$ Colliders, Nov. 1995.

[7] D. Cline, "Physics Potential and Development of $\mu^+\mu^-$ Colliders," UCLA preprint CAA-115-12/94 (1994).

[8] A. Skrinsky, "Polarized Muon Beams for a Muon Collider," in *Physics Potential & Development of* $\mu^+\mu^-$ *Colliders* (Proc., 3rd Int. Conf., San Francisco, Dec. 1995), Elsevier, in press.

[9] R. Palmer *et al.*, "Muon Collider Design," in *Physics Potential & Development of* $\mu^+\mu^-$ *Colliders* (Proc., 3rd Int. Conf., San Francisco, Dec. 1995), Elsevier, in press.

[10] D. Amidei, "Top Physics at the Fermilab Tevatron," in *Physics Potential & Development of $\mu^+\mu^-$ Colliders* (Proc., 3rd Int. Conf., San Francisco, Dec. 1995), Elsevier, in press.

[11] M. Atac, "Detector Possibilities for a $\mu^+\mu^-$ Collider in *Physics Potential & Development of $\mu^+\mu^-$ Colliders* (Proc., 3rd Int. Conf., San Francisco, Dec. 1995), Elsevier, in press.

[12] V. Polychronakos, "$\mu^+\mu^-$ Detector Studies," in *Physics Potential & Development of $\mu^+\mu^-$ Colliders* (Proc., 3rd Int. Conf., San Francisco, Dec. 1995), Elsevier, in press.

[13] P. Chen and K. Yokoya, Phys. Rev. **D38** 987 (1988); P. Chen, *SLAC-PUB-4823* (1987); Proc. Part. Accel. School, Batavia, IL, (1987); AIP Conf. Proc. **184**, 633 (1987).

Contributors

- R. B. Palmer, (BNL/SLAC) Editor
- D. Cline, (UC, Los Angeles)
- R. Rossmanith, (DESY)
- B. Norum, (Univ. Virginia)

List of Figures

List of Tables

Chapter 12

APPENDIX

Contents

12.1 Final Stage Electron Cooling for Muons

A preliminary examination of final stage electron cooling of negative muons with intense electron beams is presented here. This scheme seems feasible with the present day cutting edge technology. Furthermore, a low density Z pinch combined with an intense relativistic electron beam may be a viable cooling scheme for positive muons as well.

12.1.1 High Intensity Electron Beams

Electron beam currents that are in the mega-Ampere range have been generated by diodes. Although most of these diodes operate with pulses that are in the nsec range, some diodes have operated with pulse lengths of up to 2 microseconds. More conventional electron guns (some with plasma cathodes) can also be stacked up (or even scaled up) to yield hundreds of kA to 1 MA of current. Although in most practical applications (pulse lengths of 100 microseconds or longer), current densities in beam forming gun structures are limited to $100A/cm^2$ due to voltage breakdown effects, much higher current densities are possible in devices operating with pulse lengths that are sufficiently short (no more than a few microseconds, i.e., shorter than an arc propagation time). A hybrid system in which an electron beam is launched and is propagated through a plasma channel can be a very attractive option, since it is possible that neither technique may need to be "pushed" to its technological limit to reach resultant axial currents exceeding 10 MA that are about 1 meter long. Hollow-beam electron guns may be particularly suitable for such an application due to their larger perveance and enhanced stability in addition to the obvious advantage of their hollow structure. Another component of the presented scheme – the so called Z pinch – involves a sudden compression of a low-density plasma by means of a large discharge current that lasts for a few micro-seconds. Its fill pressure is below a milli-Torr. First, a low-density, low-temperature plasma is created by rf, lasers or exploding wires. Second, a large voltage is applied to the end plates that drives a very large axial current that compresses the plasma due to an inward acceleration of a surface current shell. At first glance a Z pinch seems to be a poor option due to its minuscule radial dimension, nevertheless, discharge currents of 10 MA over a few centimeters have been reached in a rather expensive system [1]. In a series of experiments with magnetized Z pinches, 2 MA were reached for a length of 0.8 meters with an axial magnetic field of 1.5 Tesla [2]. Various pion and muon focusing options with spark channels, Z-pinches and electron beams were explored [3]. These devices are described in detail elsewhere [3].

12.1.2 Electron Cooling Effects

A low thermal spread electron beam moving at the same velocity with a hotter charged particle beam will have a cooling effect on that beam. Electron beams, low density Z pinches, or hybrid systems can be designed to have electrons moving at the same velocity as pions and muons during the discharge. Pions and muons focused into such a discharge channel will be cooled by the electrons. At first glance, this idea does not seem very feasible since pions and muons are not trapped; consequently, cooling must be on a time scale much shorter than a

second (which is typical for electron beam coolers). However, if the parameters from LEAR are scaled up, this idea seems more interesting. Using calculations from Poth's CERN report, [4] a 1 Ampere electron beam will cool antiprotons in 0.03 seconds if exposed continuously to the electron beam (since they are subjected to the cooling effects of the electrons for only 1/50 of their orbit, cooling occurs in 1.5 seconds). Theoretically, the thermal equilibration time is given by [5]

$$\tau = 5.56 \times 10^{18} \frac{(m_h T_c + m_c T_h)^{3/2}}{(m_c m_h)^{1/2} Z_h^2 Z_c^2 \lambda n_c} [sec.] \tag{12.1}$$

where subscripts c and h refer to cold and hot particles, respectively and λ is the Coulomb logarithm. It is clear from Eq. 12.1 that equilibration time is proportional to the density of the lower temperature particles and for electron beams with equal cross section (and velocity), the electron density is proportional to the current. To scale up from LEAR, consider a 3 meter long 1 MA electron beam channel. The transit time of a pion or a muon through that distance is 10 nsec; therefore, to compensate for this, shorter cooling period, the electron current (density) must be raised, 1 MA to make up six orders of magnitude. An additional gain is made by the fact that electrons equilibrate faster with lighter particles (pions and muons here versus antiprotons in LEAR). Since the energy equipartion time is proportional to the square root of the mass ratio, cooling time is reduced by a factor of 2.6 for pions and about 3 for muons. Thus, the cooling properties of such an electron beam channel for pions and muons and the electron cooler for antiprotons are not too far apart. Furthermore, these cooling channels can be stacked. However, Eq. 12.1 shows a very strong temperature dependence. Those pions and muons whose temperature is not too far off the electron temperature can indeed be cooled in such a channel. At LEAR, electron cooling of 308.6 MeV/c antiprotons with an initial momentum spread of 2×10^{-3} was performed [6]. To cool 2 GeV/c pions or muons with a thermal spread of about 200 MeV, cooler parameters need to be increased by close to four orders of magnitude. This can be accomplished by increasing the total current to 10 MA and by stacking channels to a total length of more than 3 km. Such a cooler is not very feasible due to its cost. However, if pions are initially cooled by other means, electron cooling can be used as a final stage cooler. Consider a muon beam that was cooled and slowed down to a momentum p = 300 MeV/c and a momentum spread of

$$\frac{\Delta p}{p} = 0.04 \tag{12.2}$$

To calculate the cooling time, Poth's formula [4] is used

$$\tau = \frac{\sigma_\mu^3 + \Delta_e^3}{6\pi Z^2 R_e^2 R_\mu^2 n_c L_c} \tag{12.3}$$

where L_c is the Coulomb logarithm, R is the classical radius, σ_μ and Δ_e are velocity spreads of muons and electrons, respectively (in MKS units). For the electron velocity to match that of 300 MeV/c muons, 0.516 MeV electron are needed. Therefore, a 10 MA electron beam or in a very low density Z pinch with a 1 cm radius (to match the radius of the muon beam) will have a density of $n_e = 4.77 \times 10^{15}$electrons/$cm^3$ and we choose for the electrons to have a thermal spread of 3.48 keV to match the velocity spread of the muons (and $L_c = 15$). Using Eq. 12.3 to calculate the cooling time, yields 6.48×10^{-9} sec. Hence, since 300 MeV/c muons travel a distance of 1.68 meters during this time (their velocity is 2.6×10^8 m/sec), a cooling channel of 1.68 meters is needed. At the end of this cooling channel, $\Delta p/p = 2.86 \times 10^{-3}$. An additional stage with much colder electrons $T_e = 0.1$ eV can be added. In this stage, cooling occurs according to Eq. 12.3 in 2.16 nsec (in a 10 MA channel). At this current, a cooling channel length of 56 cm is needed (or the current can be reduced in a longer channel) and $\Delta p/p$ can be reduced to

$$\frac{\Delta p}{p} = 1.53 \times 10^{-5} \tag{12.4}$$

These preliminary calculations indicate that two channels, containing 10 MA 516 keV electron beams, with a total length of 2.24 meters can serve as a very effective final cooling stage for the muons. Furthermore, it can reduce their momentum spread by more than three orders of magnitude. We notice in passing that, 10 MA Z pinches and the electron beams, discussed above, have been achieved in practice. In cooling μ^-, the magnetic field generated by the co-moving electrons focuses μ^-. But, such a magnetic field defocuses μ^+. To cool μ^+ particles, no net axial current should exist in a channel. One possibility is to shoot the 10 MA electron beam through a 10 MA Z pinch such that the two currents cancel each other. Confinement can then be provided by a multiple magnetic field. For a Z pinch with parameters achieved earlier [1], the confining magnetic field can be calculated from $B^2/2\mu_0 = nkT$ to be 6.34 Tesla, or the required confinement can be achieved if the Z pinch current is greater than 10 MA, with an axial magnetic field of about 2 Tesla. In such a channel scattering by the Z pinch particles can be shown not to be a problem, since electron-electron scattering time is 320 nsec, while electron-muon scattering time is 3.6 msec, both of which are much longer than the total cooling time of 8.48 nsec. This configuration is expected to be stable for the time scale of interest (nsec.) to muon colliders. Although such a velocity space configuration is potentially micro-unstable, the resulting instabilities, if they occur, will have growth rates that are slower than the cooling time. This configuration resembles that of reverse field pinches, which are more stable than conventional Z pinches (since a number of macro-instabilities have a slower growth rate).

12.1.3 Conclusions

A hybrid system of electron beams with Z pinches, is an interesting idea to pursue, as a final stage cooler for the muon collider. The scheme proposed for cooling positive muons involves a more complex configuration, which is also worth looking into.

12.2 Ionization Energy Loss in a Crystal Channel

A possible method for high energy muon cooling via ionization energy loss in a focusing crystal channel is outlined here. We show that starting with an initially 'cool' muon beam, e.g. coming from a 25 GeV photo-production source, with the normalized emittance of $\epsilon_N = 10^{-5}$ m rad, one can decrease the transverse emittance to less than $\epsilon_N = 10^{-7}$ m rad by passing muons through total of 280 meters of the crystal absorber. For a practical implementation, we suggest a storage ring configuration, where sections of crystal absorber are alternated with a conventional high gradient re-acceleration (20 MeV/m) inserts. The necessary circular confinement (bending) would be provided by additional sections of bent crystals – employing powerful steering features of bent Silicon crystal demonstrated by recent experiments. Required cooling length of 280 meters constitutes only about 2×10^{-3} of the muon life-time in the laboratory frame (about 90 turns in our model 'cooling ring'). Dominant heating process (due to multiple scattering on electron gas inside a crystal channel) limits the minimum achievable emittance to about $\epsilon_N = 10^{-9}$ m rad, while the characteristic ionization cooling damping length is about 62.5 m. Feasibility of effective ionization cooling rests on the ultra high fundamental crystal fields available in a solid state environment.

12.2.1 Introduction

We outline here a possible method for high energy muon cooling via ionization energy loss in a focusing crystal channel. Recent experiments in the U.S., Europe and Russia have shown impressive progress in high-efficiency steering of charged particle beams by means of the bent crystal channeling [11], [12]. The scope of these experimental studies has been focused on variety of possible applications of crystal components for high energy accelerators. A model calculation, presented here, shows that starting with an initially 'cool' muon beam (e.g. coming from a photo-production source, providing 25 GeV $\mu^{\pm}$ pairs with the normalized emittance of $\epsilon_N = 10^{-5}$ m rad) one can decrease emittance to less than $\epsilon_N = 10^{-7}$ m rad by multiple passage of the muon beam through a sequence of Silicon crystal absorbers followed by a conventional rf re-acceleration module. Appropriate bending and circular confinement of the muon beams should allow multiple passage through the crystal absorber to decrease the

transverse emittance below $\epsilon_N = 10^{-9}$ m rad. Experimental demonstration of high efficiency muon channeling promised by the high focusing fields in a crystal [13], [14] is under way; e.g. an experiment at TRIUMF [15]. Dominant heating process (due to multiple scattering on electron gas inside a crystal channel) is also taken into account. Derived here, transverse emittance 'cooling equation' shows that the minimum achievable emittance (equilibrium cooling limit) is of the order of $\epsilon_N^{min} = 10^{-9}$ m rad, while the characteristic transverse emittance damping length is about 62.5 m. To reach the final emittance of $\epsilon_N = 10^{-7}$ m rad one would have to pass muons through total of 62.5×2 log10 meters = 280 meters of the focusing crystal channel. Both processes, ionization cooling and re-acceleration, require ultra high fundamental crystal fields available in a solid state environment. There are at least two effective ways to provide $\mu^{\pm}$ for a muon collider. First, through π production with hadron beams and subsequent π decay [7]. Second, by photo-production, or electro-production [16]. The advantage of the later is that the bunches are very short, compared to any hadronic source, due to the bunch structure of the linac. Furthermore, the $\mu^{\pm}$ can be produced with very low transverse emittance [16]. The disadvantage is a relatively low intensity. Nevertheless, the final luminosity of a μ^+ μ^- collider will depend on the emittance – a method that produces very low $\mu^{\pm}$ emittance might be comparable with a hadronic source [8].

12.2.2 Crystal Channeling

We explore unique properties of relativistic channeling of charged particles in a bent crystal as a technique for particle beam steering [11], [12]. Particularly we are interested in the circular confinement of muon beams in a crystal with a strain imposed curvature as a possible functional element of a high energy storage ring. The transverse motion of a relativistic ($\gamma >> 1$) muon of mas m_μ, channeling through a crystal is described, in linear approximation, by the following equation [13]

$$m_\mu \gamma \frac{d^2x}{dt^2} + U'(x) = 0, \tag{12.5}$$

where x is the distance from the centerline between the atomic planes and U(x) is the averaged planar continuum electrostatic potential energy at the distance x. For a positive particle the continuum potential well is given to a good approximation by

$$U(x) = \frac{1}{2}\phi x^2. \tag{12.6}$$

The focusing strength of a crystal channel for [110] planar channeling in Silicon has been experimentally measured and has a value of [14] $\phi = 6 \times 10^{12}\mathrm{GeVm}^{-2}$. As one can see from Eq. 12.5 and Eq. 12.6 the beam dynamics of charged particles channeling through a straight

crystal channel corresponds to a transverse harmonic oscillator moving relativistically in the longitudinal direction – the crystal channel plays the role a strongly focusing transfer line characterized by the beta function, β, (or alternatively by the betatron wavelength, λ) both expressed by the following formula

$$\beta = \frac{\lambda}{2\pi} = \sqrt{\frac{E_\mu}{\phi}}. \tag{12.7}$$

Assuming 25 GeV muons channeling in Silicon the corresponding beta function has the following value $\beta = 2 \times 10^{-6} m$. Following Tsyganov [17], we consider motion of planar channeled particles in a crystal, which is bent elastically in a direction perpendicular to the particle velocity and to the channeling planes. The effect of bending introduces a centrifugal force into the equation of transverse motion, Eq. 12.5. The modification of the crystal continuum potential due to the bending curvature, ρ, may be described as follows [13]

$$U(x) \rightarrow \frac{1}{2}\phi x^2 - \frac{m_\mu \gamma c^2}{\rho} x, \quad m_\mu \gamma c^2 = E_\mu. \tag{12.8}$$

Adding a linear (centrifugal) piece to the crystal potential is equivalent to lowering one side of the continuum potential well and raising the other. One can see from Eq. 12.5 and Eq. 12.8 that the equilibrium planar trajectory will move away from the midpoint of the planar channel toward the plane on the convex side of the curved planar channel. However, such shift would cause some fraction of the channeled particles to leave the potential well (dechannel). The critical curvature at which no particle can remain channeled is reached when the equilibrium point of planar channeled motion is shifted to the position of the planar wall on the outside of the curve. This critical radius of curvature, known as the Tsyganov radius [17] is defined by the following equilibrium condition, $U'(a/2) = 0$, where a $= 2.2 \times 10^{-10}$ m is the distance between adjacent atomic planes and U(x) is given by Eq. 12.8. This translates into the following explicit expression for the Tsyganov radius, ρ_T

$$\rho_T = \frac{2E_\mu}{\phi a}. \tag{12.9}$$

Using simple formula, which links the equivalent magnetic bending field, B, with the particle trajectory's curvature, ρ, one can get the maximum available equivalent bending field corresponding to the Tsyganov curvature, expressed as follows

$$B_T[Tesla] = 3.34 \times \frac{1}{2}\phi a. \tag{12.10}$$

Its numerical value for Silicon is evaluated as follows: $B_T = 2 \times 10^3$ Tesla. We notice in passing, that the maximum bending field is energy independent. Assuming 25 GeV muons channeling through a 2 cm- long section of a Silicon crystal the maximum bending angle, θ_T, derived from Eq. 12.9, is equal to enormous value of about 0.5 rad. This value of the critical bending angle will be used later in the paper.

12.2.3 Ionization Energy Loss

Here we propose a fast muon cooling scheme based on the ionization energy loss [18] experienced by high energy muons (25 GeV) channeling through a Silicon crystal. Applying classical theory of ionization energy loss [19] to a relativistic muons passing through a Silicon crystal of length, ΔL, yields the total energy loss, ΔE_μ, experienced by muons, which is expressed in the following simple form

$$\Delta E_\mu[MeV] = 4 \times 10^2 \times \Delta L[m]. \tag{12.11}$$

The following useful quantity, Λ

$$\frac{1}{\Lambda} = \frac{1}{E_\mu}\frac{\Delta E_\mu}{\Delta L}, \tag{12.12}$$

describes a characteristic damping length – over which particle looses all its energy. Relativistic muons passing though the crystal loose energy uniformly in both the transverse and longitudinal directions according to Eq. 12.11. After passing through a short section of a crystal ($\Delta L << \Lambda$) muons are re-accelerated longitudinally to compensate for the lost longitudinal energy. Combining both processes (ionization energy loss and re-acceleration) leads to the transverse emittance shrinkage. Introducing the normalized transverse emittance, ϵ_N, in the following standard way

$$\epsilon_N = \gamma\sigma_x\sigma_{x'}, \tag{12.13}$$

one can write down the normalized emittance budget in the form of the following cooling/heating equation

$$\frac{d\epsilon_N}{dL} = -\frac{\epsilon_N}{\Lambda} + \left(\frac{\Delta\epsilon_N}{\Delta L}\right)_{scatt}. \tag{12.14}$$

The last term in the above equation accounts for the transverse heating processes (Coulomb scattering) increasing the beam divergence according to the following relationship

$$\left(\frac{\Delta\epsilon_N}{\Delta L}\right)_{scatt} = \frac{1}{2}\gamma\beta\frac{\Delta\langle\theta^2\rangle_{scatt}}{\Delta L}. \tag{12.15}$$

Here β is the beta function of a focusing crystal channel, defined by Eq. 12.7, which has enormously small value of 2×10^{-6} m, for 25 GeV muons channeling through a Silicon crystal.

One has to distinguish between the emittance of the macroscopic beam outside the crystal and the emittance of 'beamlets' in the individual channels. Assuming channeling condition – the beam divergence is equal or smaller than the critical channeling angle, $\theta_c = 34$ micro radians – a macroscopic muon beam with energy of 25 GeV and normalized emittance of 10^{-5} m rad will have a spot size of about 1 mm. Only part of the entering beam will be accepted into channels, since particles that impinge too near crystal planes will be scattered

to large angles even if their initial pitch angles are less than the critical angle, θ_c. The emittance cooling equation, Eq. 12.14, applies to the channeled 'beamlets'. Crystal channel cooling, discussed here, occurs only for those muons that enter channels successfully stage after stage. All other unchanneled muons simply experience normal ionization cooling with heating due to Coulomb scattering off atomic nuclei. The macroscopic emittance of the final beam is determined by all muons including the micro-emittance of the channeled beam mixed in with empty phase-space due to crystal plane blocking.

For muon channeling in a dielectric crystal the dominant scattering process comes from the elastic (Rutherford) muon scattering off the conduction electrons, which are present in the channel. One can integrate the Rutherford cross section over the solid angle, which yields the following formula

$$\frac{\Delta\langle\theta^2\rangle_{scatt}}{\Delta L} = 16\pi n_c \frac{r_\mu^2}{\gamma^2} log\left(\frac{\theta_{max}}{\theta_{min}}\right), \tag{12.16}$$

where

$$log\left(\frac{\theta_{max}}{\theta_{min}}\right) \approx 5. \tag{12.17}$$

Here, $r_\mu = 1.4 \times 10^{-17}$ m, n_c is the concentration of the conduction electron gas in a crystal channel. The average electron gas concentration in Silicon, n can be estimated as follows

$$n = \frac{S}{a^3} = 6 \times 10^{29} m^{-3}, \tag{12.18}$$

where S = 6 is the coordination number for the basic crystallographic cell for Silicon crystal (cubic face centered) and $a = 2.2 \times 10^{-10}$ m is the distance between two neighboring [110] planes. One expects the channel electron density, n_c, to be less then the average density, n, since electrons tend to be concentrated around nuclei. From the critical angle, $\theta_c =$ 34 micro radians, and the dechanneling length, $l_d = 2$ cm, for 25 GeV/c momentum, one can estimate the electron channel density, n_c, as $6 \times 10^{28} m^{-3}$, with the multiple scattering formula, Eq. 12.16 and the following identity

$$\frac{d\langle\theta^2\rangle_{scatt}}{dx} \approx \frac{\theta_c^2}{l_d}. \tag{12.19}$$

Now, one can summarize balance between ionization energy loss (cooling) and multiple scattering (heating) in the derived cooling/heating equation, Eq. 12.14, in terms of the following two quantities

$$\Lambda = 62.5\,\mathrm{m}, \tag{12.20}$$

and

$$\alpha = \left(\frac{\Delta\epsilon_N}{\Delta L}\right)_{scatt} = 40\pi n_c \frac{r_\mu^2}{\gamma}\beta = 1.2 \times 10^{-11}\,\mathrm{rad}. \tag{12.21}$$

Integrating the cooling equation, Eq. 12.14, one obtains the following compact solution in terms of the normalized transverse emittance evolution

$$\epsilon_N = \epsilon_N^0 e^{-\frac{L}{\Lambda}} + \Lambda\alpha\left(1 - e^{-\frac{L}{\Lambda}}\right). \tag{12.22}$$

The second term in Eq. 12.22 sets the equilibrium cooling limit of

$$\epsilon_N^{min} = \Lambda\alpha,\ L \to \infty. \tag{12.23}$$

Assuming 25 GeV muons one gets the equilibrium limit of the normalized emittance of

$$\epsilon_N^{min} = 0.75 \times 10^{-9}\mathrm{m\,rad}. \tag{12.24}$$

Practical realization of muon cooling at 25 GeV could be done in a compact 'cooling ring', where one would employ powerful steering properties of bent crystals (see previous subsection) to provide circular confinement of the muon beam. Projecting experimental results for proton channeling in a bent Silicon crystal, one can assume that 25 GeV muons channeling through a 2 cm-long crystal should follow (without significant dechanneling effects) a bend of $\theta = 2\pi \times 10^{-2}$ rad (compare with the critical bending angle of $\theta_T = 5 \times 10^{-1}$ rad. calculated in the previous section). Assuming bending angle per cell of, $\theta = 4\pi \times 10^{-2}$ rad, only fifty ($50 \times \theta = 2\pi$) of the functional bending cells would be needed to complete the entire cooling ring. Its effective circumference would be equivalent to 3 meters of Silicon crystal. Assuming characteristic damping length, L, of 62.5 meters, the energy loss suffered by the muon beam after passing through a 2 cm - long section of a Silicon crystal, is equal to 8 MeV. In principle, a conventional high gradient (20 MeV/m) acceleration inserts (40 cm - long rf insert following every 2 cm - long crystal absorber) could be used to replenish the suffered energy loss ($0.4m \times 20MeV/m = 8MeV$). The proposed cooling ring of fifty-fold symmetry is illustrated schematically in Fig. 12.1 It has a nominal circumference of 63 meters! Our goal is to start with the initial muon phase-space of the normalized emittance of 10^{-5} m rad and cool it down to the final emittance of 10^{-7} m rad. One can see from Eq. 12.21 that to achieve this goal muons have to pass through the total Silicon crystal length of $L = 2\log 10 \times \Lambda = 280$ m. In the proposed cooling cell architecture the total cooling medium (Silicon) length of L = 280 m is equivalent to about 90 turns of the beam circulation in the ring. The lost energy is replenished every $\Delta L = 2$ cm, which easily satisfies the adiabatic re-acceleration condition ($\Delta L << \Lambda = 62.5$ m). To go beyond the above simple analytic calculation, we are planning to carry out a realistic computer simulations of planar channeling in bent crystals. One should tracks a charged particle through the distorted crystal lattice with the use of a realistic continuous potential approximation and taking into account the processes of both single and multiple scattering on electrons, nuclei as well as on various defects and imperfections of the crystal lattice.

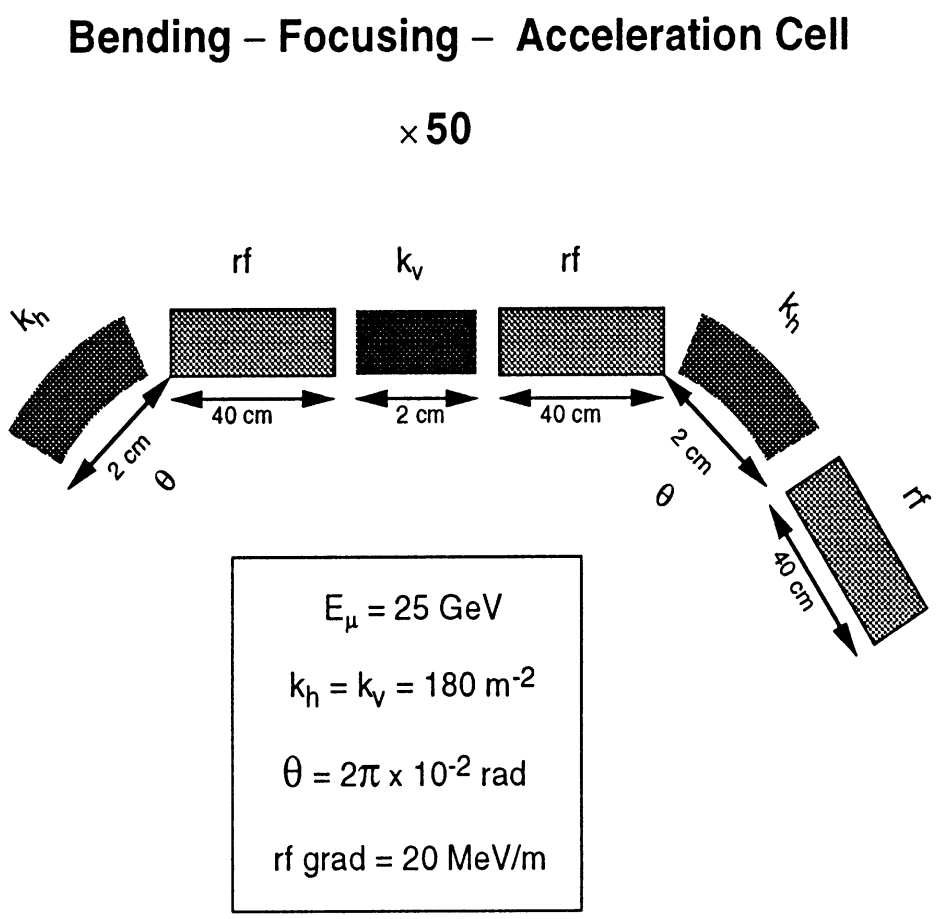


Figure 12.1: Layout of a 'cooler ring' consisting of fifty bending-focussing-acceleration multi-functional cells. A straight piece of Silicon crystal rotated by 90° separating two sections of bent crystals provides vertical focusing which, maintains betatron phase stability in the proposed lattice. A conventional rf, 40 cm - long inserts (20 MeV/m) follow every 2 cm - long section of Silicon crystal absorber.

12.2.4 Conclusions

We pointed out that initially cool muons obtained from a photo-production source could be used as a starting point for a high energy $\mu^+\mu^-$ collider complex, providing that an effective cooling scheme is available. We suggest employing ionization energy loss in an alternating focusing crystal channel as a cooling mechanism, since initially small muon phase space allows for efficient channeling through long sections of Silicon crystal. The ultra-strong focusing in a crystal channel results in the ultra small beta function. Derived here cooling equation shows that it is quite feasible to decrease the transverse emittance of a muon beam by two orders of magnitude. Our model calculation done for 25 GeV muons shows that final emittances as

low as 10^{-9} m rad could be achieved, limited only by multiple scattering off the conduction electrons in the crystal. We conclude our study with the following observation: the proposed ionization crystal cooling could be used at some later stages of the collider scheme, e.g. for the final cooling, due to a 'favorable' energy scaling of the relevant cooling characteristics, ϵ_N^{min} and Λ. Their energy scaling can be summarized as follows

$$\epsilon_N^{min} \sim \gamma^{-3/2}. \tag{12.25}$$

$$\Lambda \sim \log\gamma. \tag{12.26}$$

Therefore, the proposed cooling mechanism scaled to higher energies looks even more attractive.

12.3 Frictional Cooling – Recent Experimental Results

Frictional cooling – that is cooling a beam of very low energetic charged particles by moderation in matter and simultaneous acceleration in an electrostatic field – has been shown to be feasible during our experiments in 1994-1995 at PSI. In agreement with our previous closed form and Monte-Carlo calculations we found a significant increase in spectral density and a decrease in the angular spread in the case of a beam of negative muons.

12.3.1 Introduction

Without any doubt many experiments in muon physics become feasible only when intense sources of muons with low energy and, sometimes even more important, with small energy spread and divergence are available. This includes experiments where slow muons, both positive and negative, are used as probes in surface and thin film physics and experiments with gas targets, e.g. in muon-catalyzed fusion research [20]. High quality muon beams are also needed to set-up a high luminosity $\mu^+\mu^-$ collider [7]. Slow muons are usually produced by moderation of the high energetic muons from pion decay. New developments are the extraction of slowed down muons from an anticyclotron [21], [22] and the conversion of muons via muon-catalyzed dt-fusion [23]. Unfortunately all these methods yield a divergent muon beam with a wide energy distribution and poor density in phase space. One way to enhance the the quality of the beam is the method of frictional cooling [24]. It relies on the fact that at very low energies (for muons below roughly 10 keV) the stopping power increases with increasing energy, as shown in Fig. 12.1. The application of an electric field along the flight path of the muons in a moderator lets muons of a certain (equilibrium) energy [25], [26] T_{eq} unaffected in velocity and accelerates lower energy muons as they gain more

energy from the electrostatic field than they lose due to their interaction with the moderator. Muons with higher energy are decelerated as long as they lose more energy than they gain, this, naturally, only up to the point where energy loss and gain are equal again (higher energy muons are accelerated and lost for cooling). In addition the frictional force acts in a direction opposite to the muon motion while the electrostatic force accelerates the muons in beam direction only. Therefore the beam divergence diminishes too. The two cooling

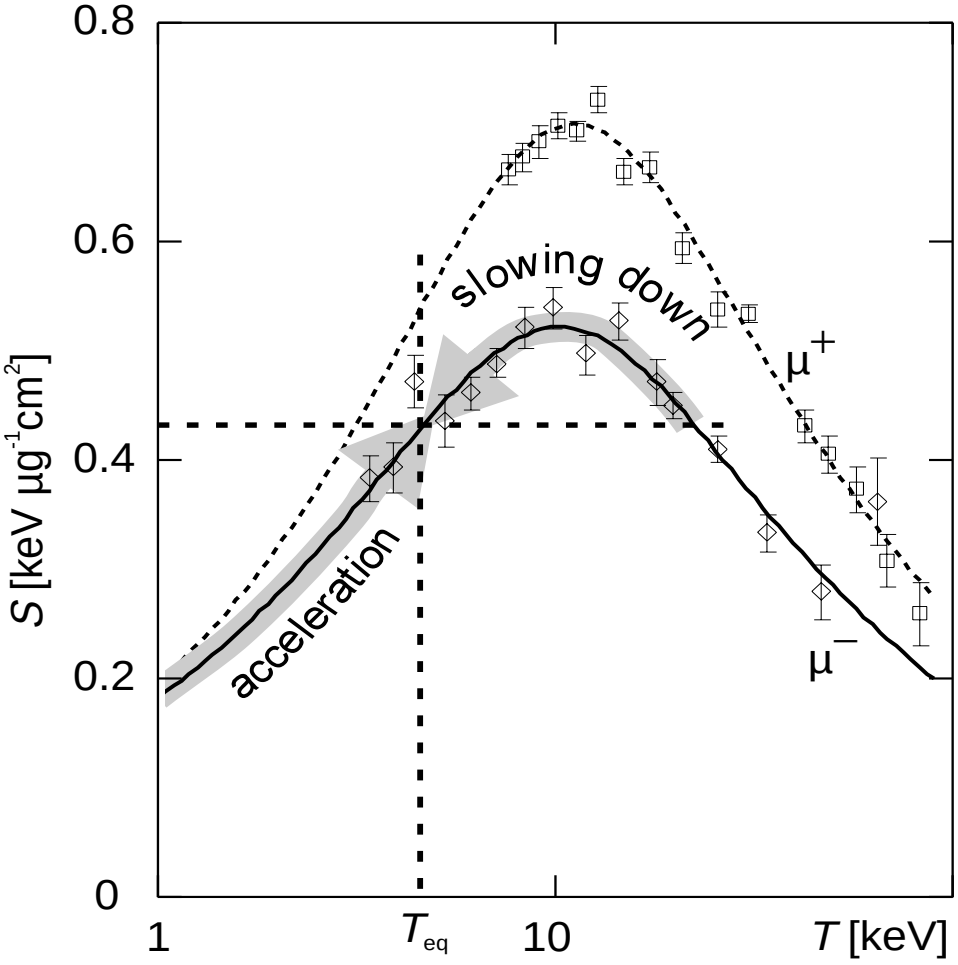


Figure 12.2: Average stopping power S(T) of carbon for both positive muons and negative muons as a function of energy T; T_{eq} denotes the equilibrium energy; the energy region where cooling takes place is marked by the arrows labeled "acceleration" and "slowing down".

effects just described are limited by multiple scattering and straggling. Nevertheless our closed-form and Monte-Carlo calculations [28], including straggling and scattering, verified the method of frictional cooling to be feasible and efficient. The experimental proof was found during our experiment at Paul Scherrer Institute (PSI) in the Summer of 1994. More detailed studies on the frictional cooling have been done within our beam time in the Spring

of 1995.

12.3.2 Experimental Arrangement

The experiments take place in the pE5 Area at PSI which provides the most intense beam for experiments using slow muons. As the momentum of the muon beam has to be kept low $p_\mu \sim 10\,\mathrm{MeV/c}$ the electron contamination is quite large. Therefore a Wien filter is used to clean the beam. The experimental set-up shown in Fig. 12.3 is placed inside the superconducting solenoid which is part of the existing phase space compression apparatus [29]. Frictional cooling is achieved with a stack of thin graphite foils mounted on stainless steel rings. In front and behind the foils we placed some additional rings not covered with foils. These extra rings allow a smooth high voltage variation. To this end the first and last rings are kept on ground voltage whereas the respective upstream and downstream foils are put on a voltage of U_{up} = -10 to -20 kV and U_{down} = -3 kV, respectively. All rings are connected to a resistive voltage-divider chain. This gives a voltage difference ΔU between two adjacent foils of about 1 to 2 kV. With this set-up we are able to run with negative muons only. In the case of positive muons we have to change the sign of the upstream voltage. This will build up a trap for the secondary electrons knocked out from the foils by the muons and some charge will be accumulated in the stack region leading to a high-voltage break through. A more elaborate arrangement might make frictional cooling feasable also for μ^+. To measure the effect of the frictional cooling a time-of-flight (TOF) technique is used. A muon entering the apparatus is slowed down from approximately 0.5 MeV to energies in the 10 keV range and detected at time t_1 in the entrance detector, a parallel plate avalanche counter (PPAC). A strong magnetic field guides the muon escaping from the PPAC on a spiral orbit along the field lines through the experimental arrangement. The muon loses energy according to the stack voltage U_{up} and enters the foil stack. While crossing one of the graphite foils the muon knocks out secondary electrons. These electrons are accelerated downstream as well as the muon and are able to produce secondary electrons from the next foils. This leads to a pulse of near to one hundred secondary electrons moving downstream with energies up to approximately 10 keV. These electrons are detected at time t_1 in the microchannel plate (MCP) detector at the very end of our apparatus. While the muon passes the stack it is being slowed down in the foils and accelerated in between. Finally the muon leaves the stack as a particle within a cooled beam, follows the magnetic field lines and hits the MCP detector at time t_3. The times the muon enters or leaves the stack needs to be calculated from t_2. Therefore we have to make some corrections. The main contributions are the energy spread of the electrons released from the stack and the time the muon spends

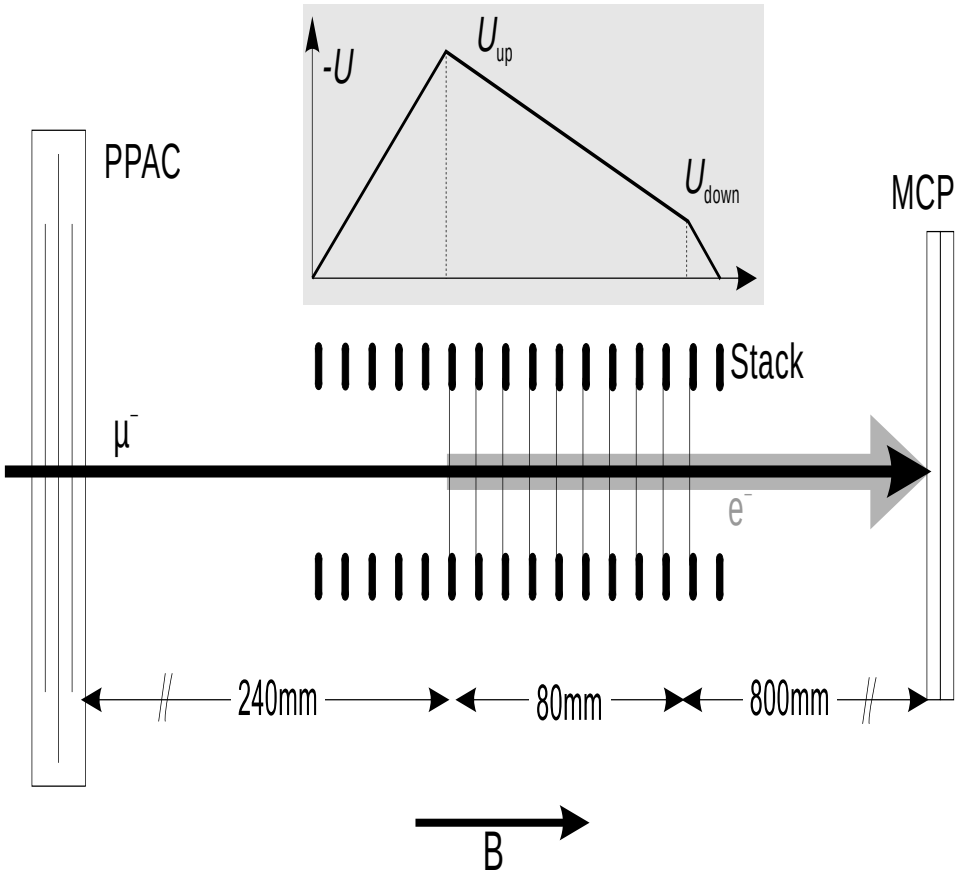


Figure 12.3: Schematic view of the experimental setup as it is housed by the superconducting solenoid. This solenoid generating a high magnetic field in beam direction is not shown in this sketch. The small insert shows the voltage distribution in the region of the foil stack.

inside the stack. These effects will be studied in detail in the near future but at the moment we have only a preliminary knowledge of these numbers. From the TOF values and the length of the flight paths we obtain the energies T_1 and T_2 of the muon in front and behind the stack, respectively. As mentioned before the strong magnetic field of approximately 5 Tesla of the superconducting solenoid surrounding the set-up makes the diverging muons spiral around the field lines and guides them from the PPAC to the MCP. The spiral radius is a function of the transverse momentum of the muon. In order to study the influence of the frictional cooling on the beam divergence we placed a collimator behind the stack. This collimator consists of a stack of parallel plastic foils (thickness of 0.2 mm) with a spacing of 1.5 mm or 1 mm and a length of 50 mm in beam direction. The comparison of the count rates under different conditions gives a measure of the beam divergence.

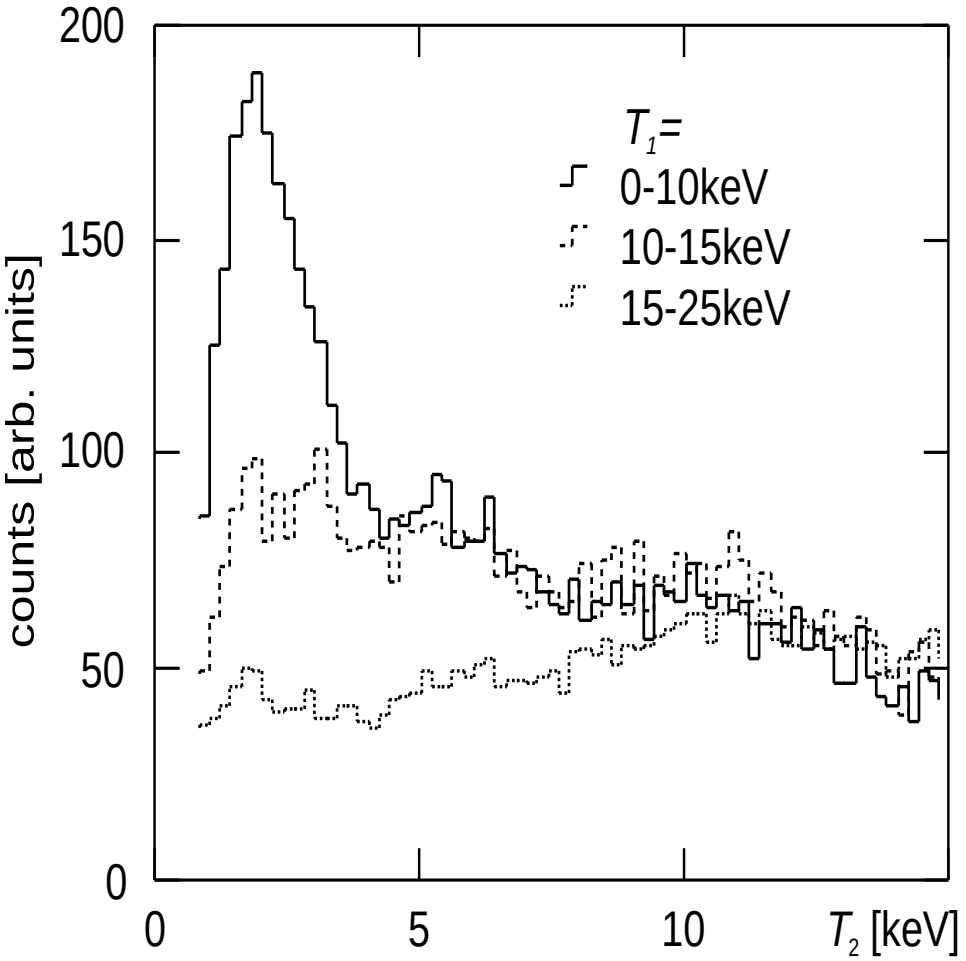


Figure 12.4: Energy spectra of the outgoing muons as a function of the energy T_1 of the incident muons 10 foils 4.3 mg/cm^2 (Carbon), each, U_{up} = -18 kV, U_{down} = -3 kV and ΔU = 1.7 kV. The energy calibration is preliminary.

12.3.3 Results

Fig. 12.4 shows the energy spectra of the outgoing muons as a function of the energy of the incident muons. The energy distribution of the incident muons is not flat at all. Therefore each spectrum is multiplied by a certain factor to correct the differences in count rate. No background subtraction is done so far. If we concentrate on the incident muons at low energies we can see a clear peak of cooled muons with a width of less than 2 keV. Position and width of the peak are in good agreement with our previous calculations. This peak vanishes when we select higher incident energies T_1. We find no such peak when we turn off the cooling by setting $\Delta U = 0$ kV.

12.3.4 Summary and Outlook

This first evaluation of part of our data shows the feasibility of frictional cooling in practice. We see a significant increase in spectral density and a decrease in the angular spread in the case of a beam of negative low-energy muons according to the predictions of our closed-form and Monte-Carlo calculations. In addition frictional cooling gives a sharp pulse of a large number of secondary electrons at energies up to 10 keV. These electrons provide a 100 % efficient muon trigger and allows the muon to be detected even with a scintillation counter. Frictional cooling found its first application in our measurement of the pm kinetic energy in a low-pressure hydrogen gas target 1 (mbar and below) [26]. The frictional cooling is used together with other techniques to stop the muons in this low density target and to provide a muon trigger with help of the secondary electrons created. A further development is the so-called frictional accumulation [30]. With this technique based on frictional cooling we should be able to convert intermediate-energy negative muons into a low-energy muon beam. First results from a Monte-Carlo simulation show that for a divergent muon beam entering the accumulation stage at energies up to 150 keV the conversion efficiency into a beam of a few keV is about 30 %.

Bibliography

[1] Work by McDaniel, et al., D.L. Cook, private communication, 1994; Rahman, Wessel, and Rostoker, reported current levels of 5 MA in *Phys. Rev. Lett.* **74** 714 (1995).

[2] A series of papers (by Golovin, et al., Komelkov, et al., and Andrianov, et al.), presented at the second U.N. Conf. on Peaceful Uses of Atomic Energy, Vol. **31** and **32**, Geneva, 1958.

[3] A. Hershcovitch, Some Options for Pion and Muon Focusing in the AGS g-2 and Neutrino Oscillation Experiments (and Possible New Focusing and Cooling Schemes for a Muon Collider), AGS/AD Tech. Note 413 (1995).

[4] H. Poth, CERN- EP/90-04.

[5] B.A. Trubnikov, Particle Interactions in a Fully Ionized Plasma *Rev. of Plasma Phys.* **1** (Consultants Bureau, N.Y.) 105 (1965).

[6] H. Poth, et al., *Nuclear Instruments and Methods in Physics Research A* **287** 328 (1990).

[7] D.V. Neuffer, *Nuclear Instruments and Methods in Physics Research A* **350** 27 (1994).

[8] D.B. Cline, *Nuclear Instruments and Methods in Physics Research A* **350** 24 (1994).

[9] S.A. Bogacz, D.B. Cline and D.A. Sanders, *Nuclear Instruments and Methods in Physics Research B*, Proceedings of Relativistic Channeling Workshop, Aarhus, Denmark, July, 1995, in press.

[10] S.A. Bogacz and D.B. Cline, *International Journal of Modern Physics A*, in press.

[11] V.M. Biryukov, *Phys. Rev. E* **51** 3522 (1995).

[12] S.A. Bogacz, D.B. Cline and S. Ramachandran, *Nuclear Instruments and Methods in Physics Research B* **882** 122 (1996).

[13] W.M. Gibson, in *Relativistic Channeling*, edited by R.A. Carrigan, Jr. and J.A. Ellison, Vol. **165B**, NATO ASI Series, Plenum Press 1986.

[14] R. Wedell, in *Relativistic Channeling*, edited by R.A. Carrigan, Jr. and J.A. Ellison, Vol. **165B**, NATO ASI Series, Plenum Press 1986.

[15] S.A. Bogacz, D.B. Cline and D.A. Sanders, *Nuclear Instruments and Methods in Physics Research B*, Proceedings of Relativistic Channeling Workshop, Aarhus, Denmark, July, 1995, in press.

[16] W. Barletta and A. Sessler, *Nuclear Instruments and Methods in Physics Research A* **350** 57 (1994).

[17] E.N. Tsyganov, FERMILAB internal report, TM-682, (1976).

[18] A.N. Skrinsky and V.V. Parakhomchuk, *Sov. Journal of Nuclear Physics* **3** 12 (1981).

[19] E.A. Perevedentsev and A.N. Skrinsky, Proc. 12-th Int. Conf. on High Energy Accel. 485 (1983).

[20] H. Daniel, *Nuclear Instruments and Methods in Physics Research B* **68** 459 (1992).

[21] H. Daniel, Los Alamos Report, LA-UR 84-3248, 1984, unpublished.

[22] L.M.Simons, *Die Zyklotronfalle, Phys. Bl.* **48** 261 (1992).

[23] P. Strasser, K. Ishida, S. Sakamoto, K. Shimomura, N. Kawamura, E. Torikai and K. Nagamine, Proceedings of the Conference on Hyperfine Interactions (1995).

[24] H. Daniel, *Muon Catalyzed Fusion* **4** 425 (1989).

[25] P. Wojciechowski, Ph. D. Thesis, Tech. University of Munich, 1995, unpublished.

[26] P. Hauser, F. Kottmann, C. Maierl, V. Markushin, M. Muhlbauer, C. Petitjean, W. Schott, D.Taqqu, P. Wojciechowski, Proceedings of the Conference on Hyperfine Interactions (1995).

[27] M. Muhlbauer, Diploma thesis, Tech. University of Munich, 1992, unpublished.

[28] M. Muhlbauer, H. Daniel, F.J. Hartmann, *Hyp. Int.* bf 82 459 (1993).

[29] D.Taqqu, *Nuclear Instruments and Methods in Physics Research A* **274** 288 (1986).

[30] D. Taqqu, Proceedings of the Conference on Hyperfine Interactions (1995).

Contributors

- A. Bogacz, (UC Los Angeles) Editor
- D. Cline, (UC Los Angeles)

List of Figures

Chapter 13

SUMMARY OF PARAMETERS

Contents

13.1 Introduction

This Chapter contains a summary of the muon collider parameters. It consolidates the tables scattered throughout this report. It is intended to give a self-consistent snapshot of the design parameters of the several components of the muon collider complex.

Table 13.1: Proton driver requirements; target and particle production parameters; capture and transfer solenoid system

Energy [GeV]	30
Rep. Rate [Hz]	15
Protons [/pulse]	10^{14}
Bunches [@ target]	4
Protons [/bunch]	2.5×10^{13}
σ_t [ns]	1
P_{beam} [MW]	7.2
ϵ_{Nrms}, 10^{-6} [m-rad]	40
β_{target} [m]	12
$\sigma(x)$ [mm]	4
$\sigma(x')$ [mrad]	0.3
B_{sol} [T]	20
a_{sol} [cm]	7.5
$p_{\perp max}$ [GeV/c]	0.225
A_N [m-rad]	0.12
L_{target} [cm]	22.5
r_{target} [cm]	1
$\pi^{\pm}/p$	1.2
P on target [kW]	600
Nominal Transport Magnetic Induction [T]	5.0
Stored Magnetic Energy to x = 3 meters [MJ]	37.9
Stored Energy S/C Magnet to x = 3 meters [MJ]	22.4
Stored Energy for x > 3.0 meters [MJ/m]	1.58

Table 13.2: Low-energy pion collection linac parameters

rf frequency [MHz]	90	50	30
Cavity Radius [cm]	90	206	126
Beam Pipe Aperture [cm]	30	30	30
Avg Gradient [MV/m]	4.2	3.3	2.1
rf Peak Power [MW]	1.8	1.1	4.8
Avg Power (15Hz) [KW]	17	26	43
Stored Energy [J]	165	261	423
Linac Length [m]	6	18	18
Total Power (15Hz) [KW]	85	390	640

Table 13.3: Cooling section summary

total length		743	m
sections		19	
total acceleration		4.8	GeV
accelerator length		690	m
μ decay loss		45	%
contingency loss		20	%
	Entrance	Exit	
KE	300	15	MeV
p	392	58	MeV/c
ϵ_{xN}(rms)	15000	39	mm mr
ϵ_{zN}(rms)	61.2	6.0	m %
σ_z	1.50	0.35	m
$\frac{\delta p}{p}$	11.0	31.7	%
μ intensity	7.5	3.0	10^{12} / bunch

Table 13.4: Parameters of a 4-RLA scenario

	RLA 1	RLA 2	RLA 3	RLA 4
Energy in [GeV]	1	9.6	70	250
Energy out [GeV]	9.6	70	250	2000
Nturns	9	11	12	16
V_{rf} per linac [GV]	0.5	3	8	56
rf frequency [MHz]	100	350	800	1300
gradient [MV/m]	5	10	15	20
L(linac) [m]	100	300.0	533.3	2800
Arc length [m]	30	175	520	3500
B_{arc} [T]	3.4	4.2	5.2	6.0
Decay Losses[%]	9.0	5.2	2.4	3.6
rms Bunch Length [cm]	4.8	1.3	0.59	0.29
rms ΔE_{rms} [GeV]	0.09	0.34	0.80	1.5

Table 13.5: High energy-high luminosity $\mu^+\,\mu^-$ collider

Maximum c-m Energy [TeV]	4
Luminosity $\mathcal{L}[10^{35}\mathrm{cm}^{-2}\mathrm{s}^{-1}]$	1.0
Circumference [km]	8.08
Time Between Collisions [μs]	12
Energy Spread σ_E[units 10^{-3}]	2
Pulse length σ_z[mm]	3
Free space at the IP [m]	$\pm$6.25
Luminosity lifetime [No.turns]	900
Horizontal betatron tune, ν_x	55.79
Vertical betatron tune, ν_y	38.82
rms emittance, $\epsilon_{x,y}$ [$10^{-6}\pi$ m-rad]	0.0026
rms normalized emittance, $\gamma\epsilon_{x,y}$ [$10^{-6}\pi$ m-rad]	50.0
Beta-function values at IP, $\beta^*_{x,y}$ [mm]	3
rms Beam size at IP [μm]	2.8
Quadrupole pole fields near IP [T]	6.0
Peak beta-function, $\beta_{x\mathrm{max}}$ [km]	284
Peak beta-function, $\beta_{y\mathrm{max}}$ [km]	373
Magnet Aperture closest to IP [cm]	12
Beam-Beam tune shift per crossing	0.05
Repetition Rate [Hz]	15
rf frequency [GHz]	1.3
rf voltage [MeV]	130
Particles per Bunch [units 10^{12}]	2
No. of Bunches of each sign	2
Peak current $\mathcal{I} = eNc/\sqrt{2\pi}\sigma_z$ [kA]	12.8
Average current $\mathcal{I} = eNc/$Circum [A]	0.032
Bending Field [T]	8.5

List of Tables