BNL-65-623
Fermilab-PUB-98/179
LBNL-41935

DRAFT

Status of Muon Collider Research and Development and Future Plans

Charles M. Ankenbrandt,[1] Muzaffer Atac,[1] Bruno Autin,[2] Valeri I. Balbekov,[1] Vernon D. Barger,[3] Odette Benary,[4] J. Scott Berg,[5] Michael S. Berger,[5] Edgar L. Black,[6] Alain Blondel,[7] S. Alex Bogacz,[8] T. Bolton,[9] Shlomo Caspi,[10] Christine Celata,[10] Weiren Chou,[1] David B. Cline,[11] John Corlett,[10] Lucien Cremaldi,[12] H. Thomas Diehl,[1] Alexandr Drozhdin,[1] Richard C. Fernow,[13] David A. Finley,[1] Yasuo Fukui,[14] Miguel A. Furman,[10] Tony Gabriel,[15] Juan C. Gallardo,[13*] Alper A. Garren,[11] Stephen H. Geer,[1] Ilya F. Ginzburg,[16] Michael A. Green,[10] Hulya Guler,[17] John F. Gunion,[18] Ramesh Gupta,[10] Tao Han,[3] Gail G. Hanson,[5] Ahmed Hassanein,[19] Norbert Holtkamp,[1] Colin Johnson,[2] Carol Johnstone,[1] Stephen A. Kahn,[13] Daniel M. Kaplan,[6] Eun San Kim,[10] Bruce J. King,[13] Harold G. Kirk,[13] Yoshitaka Kuno,[14] Paul Lebrun,[1] Kevin Lee,[11] Peter Lee,[10] Derun Li,[10] David Lissauer,[13] Laurence S. Littenberg,[13] Changguo Lu,[17] Alfredo Luccio,[13] Joseph D. Lykken,[1] Kirk T. McDonald,[17] Alfred D. McInturff,[10] John E. Miller,[20] Frederick E. Mills,[1] Nikolai V. Mokhov,[1] Alfred Moretti,[1] Yoshiharu Mori,[14] David V. Neuffer,[1] King-Yuen Ng,[1] Robert J. Noble,[1] James H. Norem,[19,1] Yasar Onel,[21] Robert B. Palmer,[13] Zohreh Parsa,[13] Yuriy Pischalnikov,[11] Milorad Popovic,[1] Eric J. Prebys,[17] Zubao Qian,[1] Rajendran Raja,[1] Claude B. Reed,[19] Pavel Rehak,[13] Thomas Roser,[13] Robert Rossmanith,[22] Ronald M. Scanlan,[10] Andrew M. Sessler,[10] Brad Schadwick,[10] Quan-Sheng Shu,[8] Gregory I. Silvestrov,[23] Alexandr N. Skrinsky,[23] Dale Smith,[19] Panagiotis Spentzouris,[1] Ray Stefanski,[1] Sergei Striganov,[1] Iuliu Stumer,[13] Don Summers,[12] Valeri Tcherniatine,[13] Lee C. Teng,[19] Alvin V. Tollestrup,[1] Yağmur Torun,[13,24] Dejan Trbojevic,[13] William C. Turner,[10] S. Vahsen,[17] Andy Van Ginneken,[1] Tatiana A. Vsevolozhskaya,[23] Weishi Wan,[1] Haipeng Wang,[13] Robert Weggel,[13] Erich H. Willen,[13] Edmund J. N. Wilson,[2] David R. Winn,[25] Jonathan S. Wurtele,[26] Takeichiro Yokoi,[14] Yongxiang Zhao,[13] Max Zolotorev[10]

[1]*Fermi National Laboratory, P. O. Box 500, Batavia, IL 60510*
[2]*CERN, 1211 Geneva 23, Switzerland*
[3]*Department of Physics, University of Wisconsin, Madison, WI 53706*
[4]*Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel*
[5]*Physics Department, Indiana University, Bloomington, IN 47405*
[6]*Illinois Institute of Technology, Physics Div., Chicago IL 60616*
[7]*École Polytechnique, Lab. de Physique Nucléaire et de Hautes Energies, Palaiseau, F-91128, France*
[8]*Jefferson Laboratory, 12000 Jefferson Ave., Newport News, VA 23606*
[9]*Kansas State University, Manhattan, KS 66502-2601*
[10]*Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720*
[11]*University of California Los Angeles, Los Angeles, CA 90095*
[12]*University of Mississippi, Oxford, MS 38677*
[13]*Brookhaven National Laboratory, Upton, NY 11973*
[14]*KEK High Energy Accelerator Research Organization, 1-1 Oho, Tsukuba 305, Japan*
[15]*Oak Ridge National Laboratory, Oak Ridge, TN 37831*
[16]*Institute of Mathematics, Prosp. ac. Koptyug 4, 630090 Novosibirsk, Russia*
[17]*Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544*
[18]*Physics Department, University of California, Davis, CA 95616*
[19]*Argonne National Laboratory, Argonne, IL 60439*
[20]*Magnet Science & Technology, National High Magnetic Field Laboratory, FL 32310*
[21]*Physics Department, Van Allen Hall, University of Iowa, Iowa City, IA 52242*
[22]*DESY, Hamburg, Germany*
[23] *Budker Institute of Nuclear Physics, 630090 Novosibirsk, Russia*
[24]*Department of Physics and Astronomy, SUNY, Stony Brook, NY 11790*
[25]*Fairfield University, Fairfield, CT 06430*
[26]*University of California Berkeley, Berkeley, CA 94720*

(DRAFT, November 9, 1998, DRAFT)

The status of the research on muon colliders is discussed and plans are outlined for future theoretical and experimental studies. Besides continued work on the parameters of a 3-4 and 0.5 TeV center-of-mass (CoM) energy collider, many studies are now concentrating on a machine near 0.1 TeV (CoM) that could be a factory for the s-channel production of Higgs particles. We discuss the research on the various components in such muon colliders, starting from the proton accelerator needed to generate pions from a heavy-Z target and proceeding through the phase rotation and decay ($\pi \to \mu\,\nu_\mu$) channel, muon cooling, acceleration, storage in a collider ring and the collider detector. We also present theoretical and experimental R & D plans for the next several years that should lead to a better understanding of the design and feasibility issues for all of the components. This report is an update of the progress on the R & D since the Feasibility Study of Muon Colliders presented at the Snowmass'96 Workshop [R. B. Palmer, A. Sessler and A. Tollestrup, *Proceedings of the 1996 DPF/DPB Summer Study on High-Energy Physics* (Stanford Linear Accelerator Center, Menlo Park, CA, 1997)].

13.10.+q,14.60.Ef,29.27.-a,29.20.Dh

Contents

---

[*] Corresponding author. Email:gallardo@bnl.gov

List of Figures

List of Tables

I. INTRODUCTION

The Standard Model of electroweak and strong interactions has passed precision experimental tests at the highest energy scale accessible today. Theoretical arguments indicate that new physics *beyond the Standard Model* associated with the electroweak gauge symmetry breaking and fermion mass generation will emerge in parton collisions at or approaching the TeV energy scale. It is likely that both hadron-hadron and lepton-antilepton colliders will be required to discover and make precision measurements of the new phenomena. The next big step forward in advancing the hadron-hadron collider energy frontier will be provided by the CERN Large Hadron Collider (LHC), a proton-proton collider with a center-of-mass (CoM) energy of 14 TeV which is due to come into operation in 2005. Note that in a high energy hadron beam, valence quarks carry momenta which are, approximately, between $\frac{1}{6}to\frac{1}{9}$ of the hadron momentum. The LHC will therefore provide hard parton-parton collisions with typical center of mass energies of $2.3 - 1.5$ TeV.

The route towards TeV-scale lepton-antilepton colliders is less clear. The lepton-antilepton colliders built so far have been e^+e^- colliders, such as the Large Electron Positron collider (LEP) at CERN and the Stanford Linear Collider (SLC) at SLAC. In a circular ring such as LEP the energy lost per revolution in keV is $88.5 \times E^4/\rho$, where the electron energy E is in GeV, and the radius of the orbit ρ is in meters. Hence, the energy loss grows rapidly as E increases. This limits the center-of-mass energy that would be achievable in a LEP-like collider. The problem can be avoided by building a linear machine (the SLC is partially linear), but with current technologies, such a machine must be very long (30-40 km) to attain the TeV energy scale. Even so, radiation during the beam-beam interaction (beamstrahlung) limits the precision of the CoM energy [1].

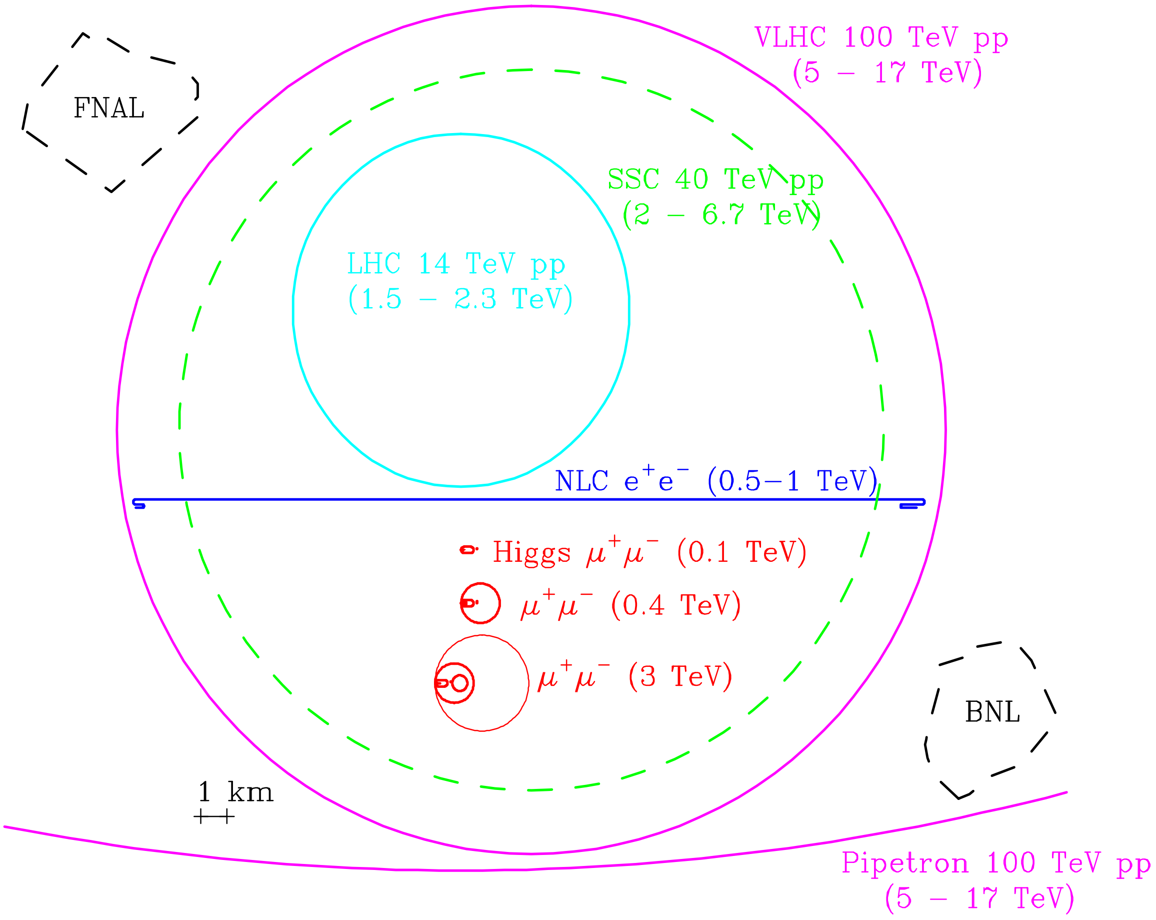


FIG. 1. Comparative sizes of various proposed high energy colliders compared with the FNAL and BNL sites. The energies in parentheses give for lepton colliders their CoM energies and for hadron colliders the approximate range of CoM energies attainable for hard parton-parton collisions.

For a lepton with mass m the radiative energy losses are inversely proportional to m^4. Hence, the energy-loss problem can be solved by using heavy leptons. In practice this means using muons, which have a mass ≈ 207 times that of an electron. The resulting reduction in radiative losses enables higher energies to be reached and smaller collider rings to be used [2,3]. Parameters for 10 to 100 TeV collider have been discussed [4,5]. Estimated sizes of the accelerator complexes required for 0.1-TeV, 0.5-TeV and 4-TeV muon colliders are compared with the sizes of other possible future colliders, and with the FNAL and BNL sites in Fig. 1. Note that muon colliders with CoM energies up to ≈ 4 TeV would fit on these existing laboratory sites. The cost of building a muon collider is not yet known. However, since muon colliders are relatively small, they may be significantly less expensive than alternative machines.

Since muons decay quickly, large numbers of them must be produced to operate a muon collider at high luminosity. Collection of muons from the decay of pions produced in proton-nucleus interactions results in a large initial phase volume for the muons, which must be reduced (cooled) by a factor of 10^6 for a practical collider. This may be compared with the antiproton stochastic cooling achieved in the Tevatron. In this case the 6-dimensional (6-D) phase space is reduced by approximately a factor of 10^6, while with stacking the phase space density [6,7] is increased by a factor of 10^{10}. The technique of ionization cooling is proposed for the $\mu^+\mu^-$ collider [8–11]. This technique is uniquely applicable to muons because of their minimal interaction with matter.

Muon colliders also offer some significant physics advantages. The small radiative losses permit very small beam-energy spreads to be achieved. For example, momentum spreads as low as $\delta P/P = 0.003\%$ are believed to be possible for a low-energy collider. By measuring the time-dependent decay asymmetry resulting from the naturally polarized muons, it has been shown [12] that the beam energy could be determined with a precision of $\delta E/E = 10^{-6}$. The small beam-energy spread, together with the precise energy determination, would facilitate measurements of the masses and widths of any new resonant states scanned by the collider. In addition, since the cross-section for producing a Higgs-like scalar particle in the s-channel (direct lepton-antilepton annihilation) is proportional to m^2, this extremely important process could be studied only at a muon collider and not at an e^+e^- collider [13]. Finally, the decaying muons will produce copious quantities of neutrinos. Even short straight sections in a muon-collider ring will result in neutrino beams several orders of magnitude higher in intensity than presently available, permitting greatly extended studies of neutrino oscillations, nucleon structure functions, the CKM matrix, and precise indirect measurements of the W-boson mass (see section **II.I**).

The concept of muon colliders was introduced by G. I. Budker [2,3], and developed further by A. N. Skrinsky *et al.* [14–21] and D. Neuffer [13,66,22,23]. They pointed out the significant challenges in designing an accelerator complex that can make, accelerate, and collide μ^+ and μ^- bunches all within the muon lifetime of $2.2\,\mu$s ($c\tau = 659$ m). A concerted study of a muon collider design has been underway in the U.S. since 1992 [25–41]. By the Sausalito workshop [29] in 1995 it was realized that with new ideas and modern technology, it may be feasible to make muon bunches containing a few times 10^{12} muons, compress their phase space and accelerate them up to the multi-TeV energy scale before more than about 3/4 of them have decayed. With careful design of the collider ring and shielding it appears possible to reduce to acceptable levels the backgrounds within the detector that arise from the very large flux of electrons produced in muon decays. These realizations led to an intense activity, which resulted in the muon-collider feasibility study report [42,43] prepared for the 1996 DPF/DPB Summer Study on High-Energy Physics (the Snowmass'96 workshop). Since then, the physics prospects at a muon collider have been studied extensively , and the potential physics program at a muon collider facility has been explored in workshops [38] and conferences [39].

Encouraged by further progress in developing the muon-collider concept, together with the growing interest and involvement of the high-energy-physics community, the *Muon Collider Collaboration* became a formal entity in May of 1997. The collaboration is led by an executive board with members from Brookhaven National Laboratory (BNL), Fermi National Accelerator Laboratory (FNAL), Lawrence Berkeley National Laboratory (LBNL), Budker Institute for Nuclear Physics (BINP), University of California at Los Angeles (UCLA), University of Mississippi and Princeton University. The goal of the collaboration is to complete within a few years the R&D needed to determine whether a Muon Collider is technically feasible, and if it is, to design the First Muon Collider.

Table I gives the parameters of the muon colliders under study [47–54], which have CoM energies of 0.1 TeV, 0.4 TeV and 3 TeV and Figs. 2 and 3 show possible outlines of the 0.1 TeV and 3 TeV machines. In the former case, parameters are given in the table for operation with three different beam-energy spreads: $\delta p/p = 0.12$, 0.01, and 0.003%. In all cases, proton bunches containing 2.5-5 $\times\, 10^{13}$ particles are accelerated to energies of 16 GeV. The protons interact in a target to produce $\mathcal{O}(10^{13})$ charged pions of each sign. A large fraction of these pions can be captured in a high-field solenoid. Muons are produced by allowing the pions to decay into a lower-field solenoidal channel. To collect as many particles as possible within a useful energy interval, rf cavities are used to accelerate the lower-energy particles and decelerate the higher-energy particles (so-called phase rotation). With two proton bunches every accelerator cycle, the first used to make and collect positive muons and the second to make and collect negative muons, there are about 10^{13} muons of each charge available at the end of the decay channel per accelerator cycle. If the proton accelerator is cycling at 15 Hz, then in an operational year (10^7 s), about 10^{21} positive and negative muons would be produced and collected.

As stated before, the muons exiting the decay channel populate a very diffuse phase space. The next step in the muon-collider complex is to *cool* the muon bunch, *i.e.*, to turn the diffuse muon cloud into a very *bright* bunch with small longitudinal and transverse dimensions, suitable for accelerating and injecting into a collider. The cooling must be done within a time that is short compared to the muon lifetime. Conventional cooling techniques (stochastic cooling [55] and electron cooling [15]) take too long. The technique proposed for cooling muons is called ionization cooling [8,10,11], and will be discussed in detail in sec. V. Briefly, the muons traverse some material in which they lose both longitudinal and transverse momentum by ionization losses (dE/dx). The longitudinal momentum is then replaced using an rf accelerating cavity, and the process is repeated many times until there is a large reduction in the

TABLE I. Baseline parameters for high- and low-energy muon colliders. Higgs/year assumes a cross section $\sigma = 5 \times 10^4$ fb; a Higgs width $\Gamma = 2.7$ MeV; 1 year $= 10^7$ s.

CoM energy	TeV	3	0.4	0.1		
p energy	GeV	16	16	16		
p's/bunch		2.5×10^{13}	2.5×10^{13}	5×10^{13}		
Bunches/fill		4	4	2		
Rep. rate	Hz	15	15	15		
p power	MW	4	4	4		
μ/bunch		2×10^{12}	2×10^{12}	4×10^{12}		
μ power	MW	28	4	1		
Wall power	MW	204	120	81		
Collider circum.	m	6000	1000	350		
Ave bending field	T	5.2	4.7	3		
Rms $\delta p/p$	%	0.16	0.14	0.12	0.01	0.003
6-D $\epsilon_{6,N}$	$(\pi \mathrm{m})^3$	1.7×10^{-10}	1.7×10^{-10}	1.7×10^{-10}	1.7×10^{-10}	1.7×10^{-10}
Rms ϵ_n	π mm-mrad	50	50	85	195	290
β^*	cm	0.3	2.6	4.1	9.4	14.1
σ_z	cm	0.3	2.6	4.1	9.4	14.1
σ_r spot	μm	3.2	26	86	196	294
σ_θ IP	mrad	1.1	1.0	2.1	2.1	2.1
Tune shift		0.044	0.044	0.051	0.022	0.015
n_{turns} (effective)		785	700	450	450	450
Luminosity	$\mathrm{cm}^{-2}\mathrm{s}^{-1}$	7×10^{34}	10^{33}	1.2×10^{32}	2.2×10^{31}	10^{31}
Higgs/year				1.9×10^3	4×10^3	3.9×10^3

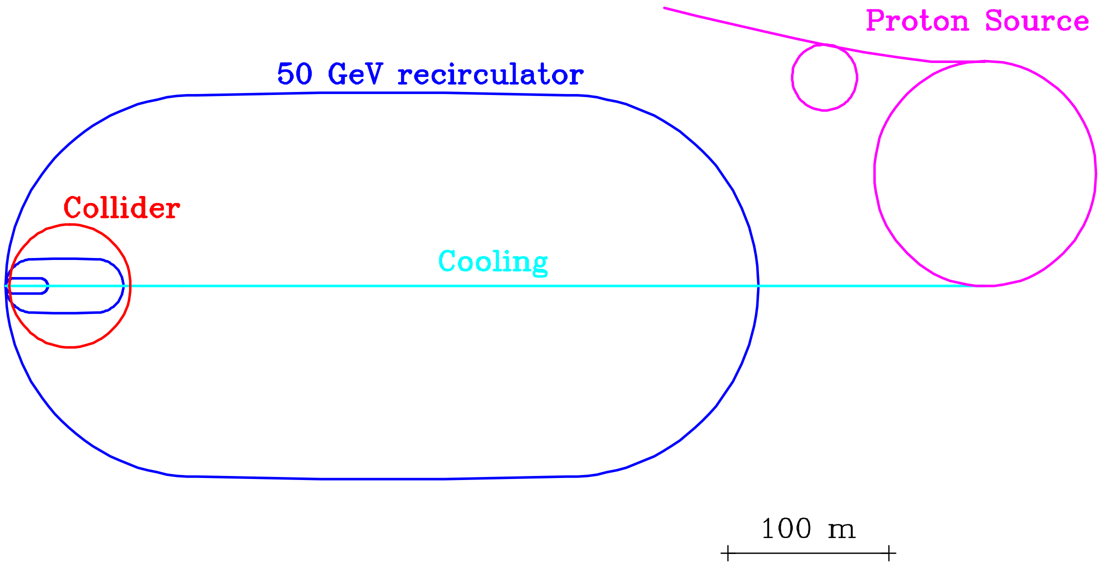


FIG. 2. Plan of a 0.1-TeV-CoM Muon Collider.

transverse phase space occupied by the muons. The energy spread within the muon beam can also be reduced by using a wedge-shaped absorber in a region of dispersion (where the transverse position is momentum dependent). The wedge is arranged so that the higher-energy particles pass through more material than lower-energy particles. Initial calculations suggest that the 6-D phase space occupied by the initial muon bunches can be reduced by a factor of 10^5-10^6 before multiple Coulomb scattering and energy straggling limit further reduction. We reiterate that ionization cooling is uniquely suited to muons because of the absence of strong nuclear interactions and electromagnetic shower production for these particles at energies around 200 MeV/c.

Rapid acceleration to the collider beam energy is needed to avoid excessive particle loss from decay. It can be achieved, initially in a linear accelerator, and later in recirculating linear accelerators, rapid-cycling synchrotron, or fixed-field-alternating-gradient (FFAG) accelerators. Positive and negative muon bunches are then injected in opposite

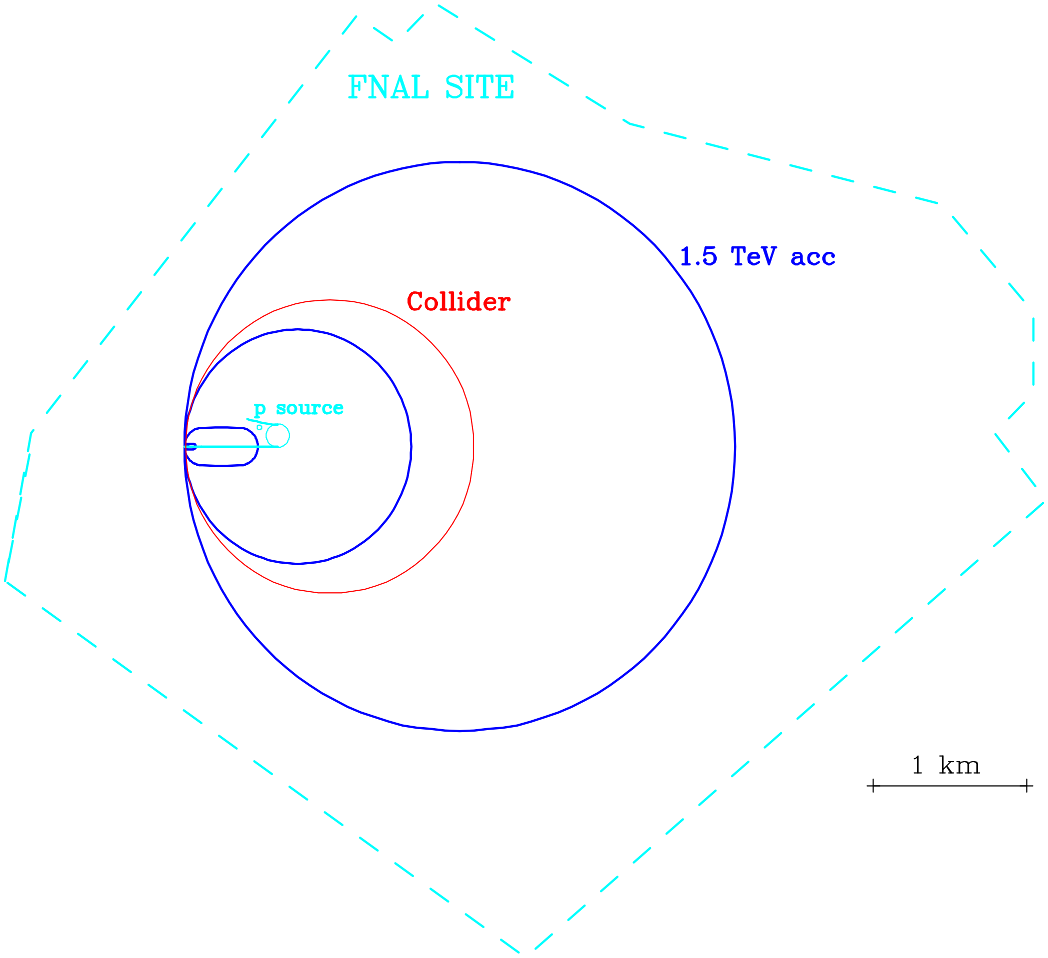


FIG. 3. Plan of a 3-TeV-CoM Muon Collider shown on the Fermi National Laboratory site as an example.

directions into a collider storage ring and brought into collision at the interaction point. The bunches circulate and collide for many revolutions before decay has depleted the beam intensities to an uninteresting level. Useful luminosity can be delivered for about 800 revolutions for the high-energy collider and 450 revolutions for the low-energy one.

There are many interesting and challenging problems that need to be resolved before the feasibility of building a muon collider can be demonstrated. For example, (i) heating from the very intense proton bunches may require the use of of a liquid-jet target, and (ii) attaining the desired cooling factor in the ionization-cooling channel may require the development of rf cavities with thin beryllium windows operating at liquid-nitrogen temperatures in high solenoidal fields. In addition, the development of long liquid-lithium lenses may be desirable to provide stronger radial focusing for the final cooling stages.

This article describes the status of our muon-collider feasibility studies, and is organized as follows. Section II gives a brief summary of the physics potential of muon colliders, including physics at the accelerator complex required for a muon collider. Section III describes the proton-driver specifications for a muon collider, and two site-dependent examples that have been studied in some detail. Section IV presents pion production, capture, and the pion-decay channel, and sec. V discusses the design of the ionization-cooling channel needed to produce an intense muon beam suitable for acceleration and injection into the final collider. Sections VI and VII describe the acceleration scenario and collider ring, respectively. Section VIII discusses backgrounds at the collider interaction point and section IX deals with possible detector scenarios. A summary of the conclusions is given in section X.

II. THE PHYSICS POTENTIAL OF MUON COLLIDERS

A. Brief overview

The physics agenda at a muon collider falls into three categories: First Muon Collider (FMC) physics at a machine with center-of-mass energies of 100 to 500 GeV; Next Muon Collider (NMC) physics at 3–4 TeV center-of-mass energies; and front-end physics with a high-intensity muon source.

The FMC will be a unique facility for neutral Higgs boson (or techni-resonance) studies through s-channel resonance production. Measurements can also be made of the threshold cross sections for production of W^+W^-, $t\bar{t}$, Zh, and pairs of supersymmetry particles — $\chi_1^+\chi_1^-$, $\chi_2^0\chi_1^0$, $\tilde{\ell}^+\tilde{\ell}^-$ and $\tilde{\nu}\bar{\tilde{\nu}}$ — that will determine the corresponding masses to high precision. A $\mu^+\mu^- \to Z^0$ factory, utilizing the partial polarization of the muons, could allow significant improvements in $\sin^2\theta_{\rm w}$ precision and in B-mixing and CP-violating studies. In Fig. 4, we show the cross sections for SM processes versus the CoM energy at the FMC. For the unique s-channel Higgs boson production, where $\sqrt{s}_{\mu\mu} = m_H$, results for three different beam energy resolutions are presented.

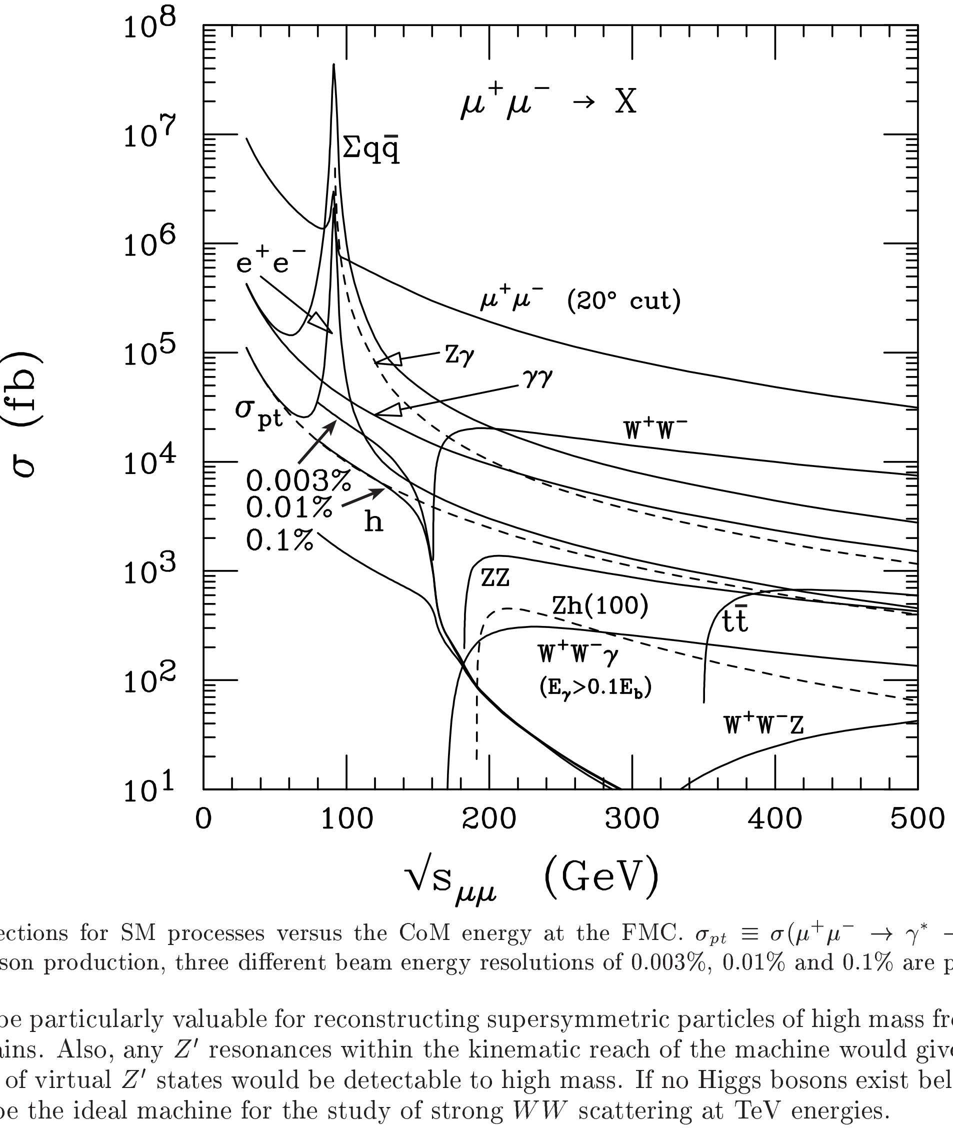


FIG. 4. Cross sections for SM processes versus the CoM energy at the FMC. $\sigma_{pt} \equiv \sigma(\mu^+\mu^- \to \gamma^* \to e^+e^-)$. For the s-channel Higgs boson production, three different beam energy resolutions of 0.003%, 0.01% and 0.1% are presented.

The NMC will be particularly valuable for reconstructing supersymmetric particles of high mass from their complex cascade decay chains. Also, any Z' resonances within the kinematic reach of the machine would give enormous event rates. The effects of virtual Z' states would be detectable to high mass. If no Higgs bosons exist below ∼1 TeV, then the NMC would be the ideal machine for the study of strong WW scattering at TeV energies.

At the front end, a high-intensity muon source will permit searches for rare muon processes sensitive to branching ratios that are orders of magnitude below present upper limits. Also, a high-energy muon-proton collider can be constructed to probe high$-Q^2$ phenomena beyond the reach of the HERA ep collider. In addition, the decaying muons will provide high-intensity neutrino beams for precision neutrino cross-section measurements and for long-baseline experiments [56–63,65,66]. Plus, there are numerous other new physics possibilities for muon facilities [43,38] that we will not discuss in detail in this document.

B. Higgs boson physics

The expectation that there will be a light (mass below $2M_W$) SM-like Higgs boson provides a major motivation for the FMC, since such a Higgs boson can be produced with a very high rate directly in the s-channel. Theoretically, the lightest Higgs boson h^0 of the most general supersymmetric model is predicted to have mass below 150 GeV and to be very SM-like in the usual decoupling limit. Indeed, in the minimal supersymmetric model, which contains the five Higgs bosons h^0, H^0, A^0, $H^\pm$, one finds $m_{h^0} \lesssim 130$ GeV and the h^0 is SM-like if $m_{A^0} \gtrsim 130$ GeV. The light SUSY h^0 is regarded as the jewel in the SUSY crown. Experimentally, global analyses of precision electroweak data now indicate a strong preference for a light SM-like Higgs boson; this could well be the smoking gun for the SUSY Higgs boson. The goals of the FMC for studying the SUSY Higgs sector via s-channel resonance production are: to measure the light Higgs mass, width, and branching fractions with high precision, in particular sufficient to differentiate the MSSM h^0 from the SM $h_{\rm SM}$; and, to find and study the heavier neutral Higgs bosons H^0 and A^0.

The production of Higgs bosons in the s-channel with interesting rates is a unique feature of a muon collider [44,67]. The resonance cross section is

$$\sigma_h(\sqrt{s}) = \frac{4\pi \Gamma(h\to\mu\bar{\mu})\,\Gamma(h\to X)}{\left(s-m_h^2\right)^2 + m_h^2\left(\Gamma^h_{\rm tot}\right)^2}\,. \tag{1}$$

Gaussian beams with root-mean-square (rms) energy resolution down to $R = 0.003\%$ are realizable. The corresponding rms spread $\sigma_{\sqrt{s}}$ in CoM energy is

$$\sigma_{\sqrt{s}} = (2\ {\rm MeV})\left(\frac{R}{0.003\%}\right)\left(\frac{\sqrt{s}}{100\ {\rm GeV}}\right)\,. \tag{2}$$

The effective s-channel Higgs cross section convolved with a Gaussian spread,

$$\bar{\sigma}_h(\sqrt{s}) = \frac{1}{\sqrt{2\pi}\,\sigma_{\sqrt{s}}}\int \sigma_h(\sqrt{\hat{s}})\ \exp\left[\frac{-\left(\sqrt{\hat{s}}-\sqrt{s}\right)^2}{2\sigma^2_{\sqrt{s}}}\right] d\sqrt{\hat{s}}, \tag{3}$$

is illustrated in Fig. 5 for $m_h = 110$ GeV, $\Gamma_h = 2.5$ MeV, and resolutions $R = 0.01\%$, 0.06% and 0.1%. A resolution $\sigma_{\sqrt{s}} \sim \Gamma_h$ is needed to be sensitive to the Higgs width. The light Higgs width is predicted to be

$$\begin{array}{ll} \Gamma \approx 2 \text{ to } 3\ {\rm MeV} & \text{if}\ \ \tan\beta \sim 1.8, \\ \Gamma \approx 2 \text{ to } 800\ {\rm MeV} & \text{if}\ \ \tan\beta \sim 20, \end{array} \tag{4}$$

for 80 GeV $\lesssim m_h \lesssim$ 120 GeV, where the smaller values apply in the decoupling limit of large m_{A^0}. We note that, in the MSSM, m_{A^0} is required to be in the decoupling regime in the context of mSUGRA boundary conditions in order that correct electroweak symmetry breaking arises after evolution of parameters from the unification scale. In particular, decoupling applies in mSUGRA at $\tan\beta \sim 1.8$, corresponding to the infrared fixed point of the top quark Yukawa coupling.

At $\sqrt{s} = m_h$, the effective s-channel Higgs cross section is

$$\bar{\sigma}_h \simeq \frac{4\pi}{m_h^2}\,\frac{{\rm BF}(h\to\mu\bar{\mu})\,{\rm BF}(h\to X)}{\left[1+\frac{8}{\pi}\left(\frac{\sigma_{\sqrt{s}}}{\Gamma^h_{\rm tot}}\right)^2\right]^{1/2}}\,. \tag{5}$$

BF denotes the branching fraction for h decay; also, note that $\bar{\sigma}_h \propto 1/\sigma_{\sqrt{s}}$ for $\sigma_{\sqrt{s}} > \Gamma^h_{\rm tot}$. At $\sqrt{s} = m_h \approx 110$ GeV, the $b\bar{b}$ rates are

$$\text{signal} \approx 10^4\ \text{events} \times {\rm L}({\rm fb}^{-1})\,, \tag{6}$$

$$\text{background} \approx 10^4\ \text{events} \times {\rm L}({\rm fb}^{-1})\,, \tag{7}$$

assuming a b-tagging efficiency $\epsilon \sim 0.5$. The effective on-resonance cross sections for other m_h values and other channels (ZZ^*, WW^*) are shown in Fig. 6 for the SM Higgs. The rates for the MSSM Higgs are nearly the same as the SM rates in the decoupling regime of large m_{A^0}.

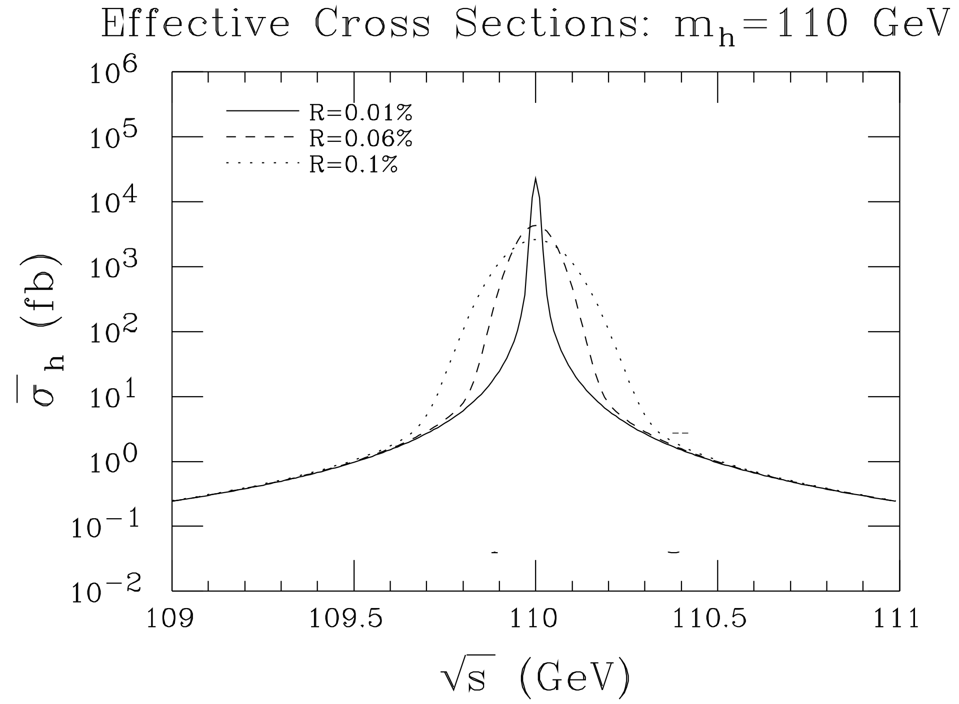


FIG. 5. Effective s-channel Higgs cross section $\bar{\sigma}_h$ obtained by convoluting the Breit-Wigner resonance formula with a Gaussian distribution for resolution R. From Ref. [44].

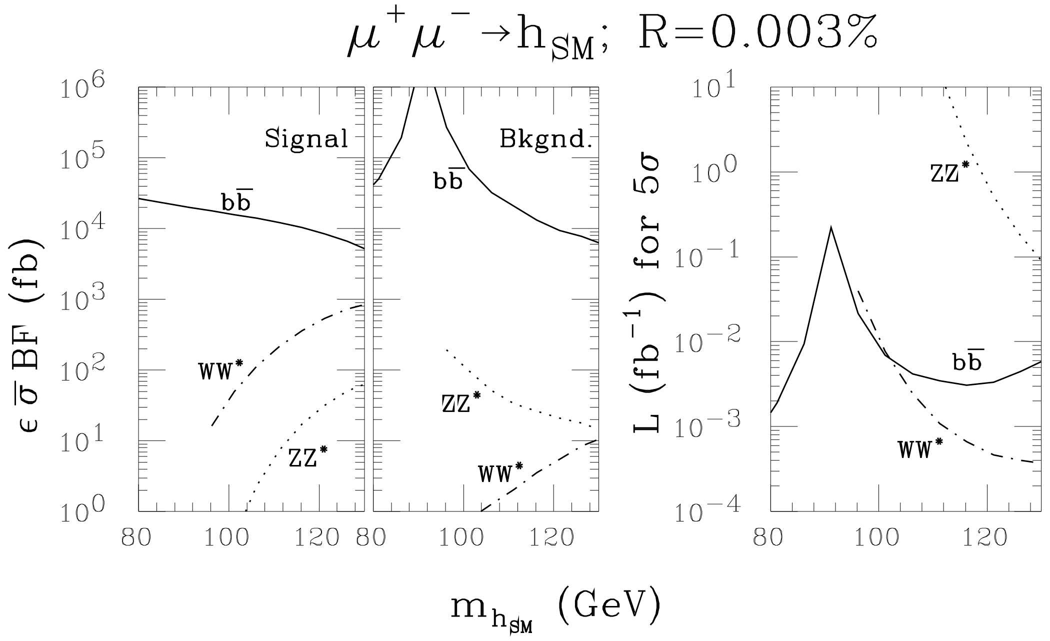


FIG. 6. The SM Higgs cross sections and backgrounds in $b\bar{b}$, WW^* and ZZ^*. Also shown is the luminosity needed for a 5 standard deviation detection in $b\bar{b}$. From Ref. [44].

The important factors that make s-channel Higgs physics studies possible at a muon collider are energy resolutions $\sigma_{\sqrt{s}}$ of order a few MeV, little bremsstrahlung and no beamstrahlung smearing, and precise tuning of the beam energy to an accuracy $\Delta E \sim 10^{-6}E$ through continuous spin-rotation measurements [12]. As a case study, we consider a SM-like Higgs boson with $m_h \approx 110$ GeV. Prior Higgs discovery is assumed at the Tevatron (in $Wh, t\bar{t}h$ production with $h \to b\bar{b}$ decay) or at the LHC (in $gg \to h$ production with $h \to \gamma\gamma, 4\ell$ decays with a mass measurement of

$\Delta m_h \sim 100$ MeV for an integrated luminosity of $L = 300$ fb^{-1}) or possibly at a NLC (in $Z^* \to Zh, h \to b\bar{b}$ giving $\Delta m_h \sim 50$ MeV for $L = 200$ fb^{-1}). A muon collider ring design would be optimized to run at energy $\sqrt{s} = m_h$. For an initial Higgs-mass uncertainty of $\Delta m_h \sim 100$ MeV, the maximum number of scan points required to locate the s-channel resonance peak at the muon collider is

$$n = \frac{2\Delta m_h}{\sigma_{\sqrt{s}}} \approx 100 \tag{8}$$

for a $R = 0.003\%$ resolution of $\sigma_{\sqrt{s}} \approx 2$ MeV. The necessary luminosity per scan point ($L_{\rm s.p.}$) to observe or eliminate the h-resonance at a significance level of $S/\sqrt{B} = 3$ is $L_{\rm s.p.} \sim 1.5 \times 10^{-3}\,{\rm fb}^{-1}$. (The scan luminosity requirements increase for m_h closer to M_Z; at $m_h \sim M_Z$ the $L_{\rm s.p.}$ needed is a factor of 50 higher.) The total luminosity then needed to tune to a Higgs boson with $m_h = 110$ GeV is $L_{\rm tot} = 0.15$ fb^{-1}. If the machine delivers $1.5 \times 10^{31}\,{\rm cm}^{-2}\,{\rm s}^{-1}$ (0.15 fb^{-1}/year), then one year of running would suffice to complete the scan and measure the Higgs mass to an accuracy $\Delta m_h \sim 1$ MeV. Figure 7 illustrates a simulation of such a scan.

Once the h-mass is determined to ~ 1 MeV, a 3-point fine scan [44] can be made across the peak with higher luminosity, distributed with L_1 at the observed peak position in $\sqrt{s}$ and $2.5L_1$ at the wings ($\sqrt{s} = {\rm peak} \pm 2\sigma_{\sqrt{s}}$). Then, with $L_{\rm tot} = 0.4$ fb^{-1} the following accuracies would be achievable: 16% for $\Gamma^h_{\rm tot}$, 1% for σBF(b$\bar{\rm b}$) and 5% for σBF(WW*). The ratio $r = {\rm BF}(WW^*)/{\rm BF}(b\bar{b})$ is sensitive to m_{A^0} for m_{A^0} values below 500 GeV. For example, $r_{\rm MSSM}/r_{\rm SM} = 0.3, 0.5, 0.8$ for $m_{A^0} = 200, 250, 400$ GeV [44]. Thus, using s-channel measurements of the h, it may be possible not only to distinguish the h^0 from the SM h_{SM} but also to infer m_{A^0}.

The study of the other neutral MSSM Higgs bosons at a muon collider via the s-channel is also of major interest. Finding the H^0 and A^0 may not be easy at other colliders. At the LHC the region $m_{A^0} > 200$ GeV is deemed to be inaccessible for $3 \lesssim \tan\beta \lesssim$ 5–10 [68]. At an NLC the $e^+e^- \to H^0A^0$ production process may be kinematically inaccessible if H^0 and A^0 are heavy (mass > 230 GeV for $\sqrt{s} = 500$ GeV). At a $\gamma\gamma$ collider, very high luminosity ($\sim$200 fb^{-1}) would be needed for $\gamma\gamma \to H^0, A^0$ studies.

At a muon collider the resolution requirements for s-channel H^0 and A^0 studies are not as demanding as for the h^0 because the H^0, A^0 widths are broader; typically $\Gamma \sim 30$ MeV for $m_{A^0} < 2m_t$ and $\Gamma \sim 3$ GeV for $m_{A^0} > 2m_t$. Consequently $R \sim 0.1\%$ ($\sigma_{\sqrt{s}} \sim 70$ MeV) is adequate for a scan. This is important, since higher instantaneous luminosities (corresponding to $L \sim 2-10$ fb^{-1}/yr) are possible for $R \sim 0.1\%$ (as contrasted with the $L \sim 0.15$ fb^{-1}/yr for the much smaller $R \sim 0.003\%$ preferred for studies of the h^0). A luminosity per scan point $L_{\rm s.p.} \sim 0.1$ fb^{-1} probes the parameter space with $\tan\beta > 2$. The $\sqrt{s}$-range over which the scan should be made depends on other information available to indicate the A^0 and H^0 mass range of interest. A wide scan would not be necessary if r is measured with the above-described precision to obtain an approximate value of m_{A^0}.

In the MSSM, $m_{A^0} \approx m_{H^0} \approx m_{H^\pm}$ at large m_{A^0} (as expected for mSUGRA boundary conditions), with a very close degeneracy in these masses for large $\tan\beta$. In such a circumstance, only an s-channel scan with the good resolution possible at a muon collider may allow separation of the A^0 and H^0 states; see Fig. 8.

C. Light particles in technicolor models

In most technicolor models, there will be light neutral and colorless technipion resonances, π^0_T and $\pi^{0\prime}_T$, with masses below 500 GeV. Sample models include the recent top-assisted technicolor models [69], in which the technipion masses are typically above 100 GeV, and models [70] in which the masses of the neutral colorless resonances come primarily from the one-loop effective potential and the lightest state typically has mass as low as 10 to 100 GeV. The widths of these light neutral and colorless states in the top-assisted models will be of order 0.1 to 50 GeV [71]. In the one-loop models, the width of the lightest technipion is typically in the range from 3 to 50 MeV. Neutral technirho and techniomega resonances are also a typical feature of technicolor models. In all models, these resonances are predicted to have substantial Yukawa-like couplings to muons and would be produced in the s-channel at a muon collider,

$$\mu^+\mu^- \to \pi^0_T,\, \pi^{0\prime}_T,\, \rho^0_T,\, \omega^0_T, \tag{9}$$

with high event rates. The peak cross sections for these processes are estimated to be $\approx 10^4$–10^7 fb [71]. The dominant decay modes depend on eigenstate composition and other details but typically are [71]

$$\pi^0_T \to gg,\, b\bar{b},\, \tau\bar{\tau},\, c\bar{c},\, t\bar{t}, \tag{10}$$

$$\pi^{0\prime}_T \to gg,\, b\bar{b},\, c\bar{c},\, t\bar{t},\, \tau^+\tau^-, \tag{11}$$

$$\rho^0_T \to \pi_T\pi_T,\, W\pi_T,\, WW, \tag{12}$$

$$\omega^0_T \to c\bar{c},\, b\bar{b},\, \tau\bar{\tau},\, t\bar{t},\, \gamma\pi^0_T,\, Z\pi^0_T. \tag{13}$$

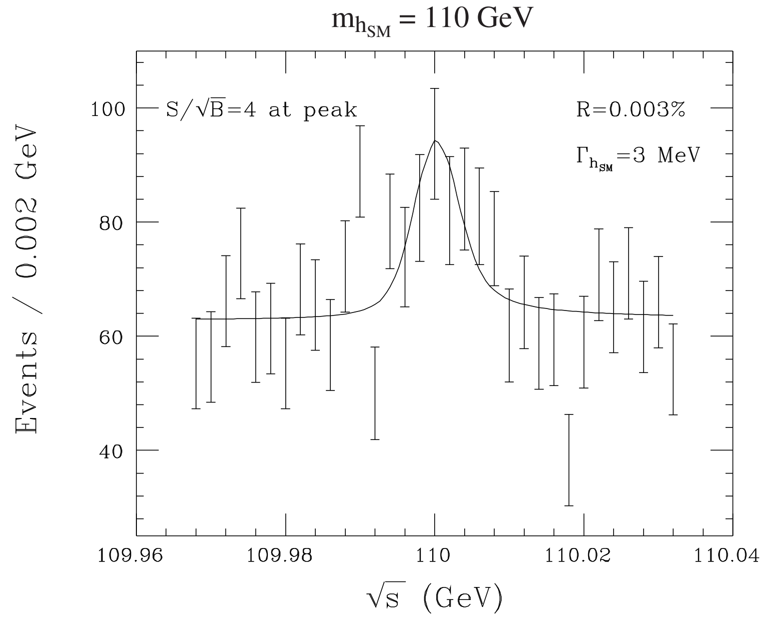


FIG. 7. Number of events and statistical errors in the $b\bar{b}$ final states as a function of $\sqrt{s}$ in the vicinity of $m_{h_{\rm SM}} = 110$ GeV, assuming $R = 0.003\%$. From Ref. [44].

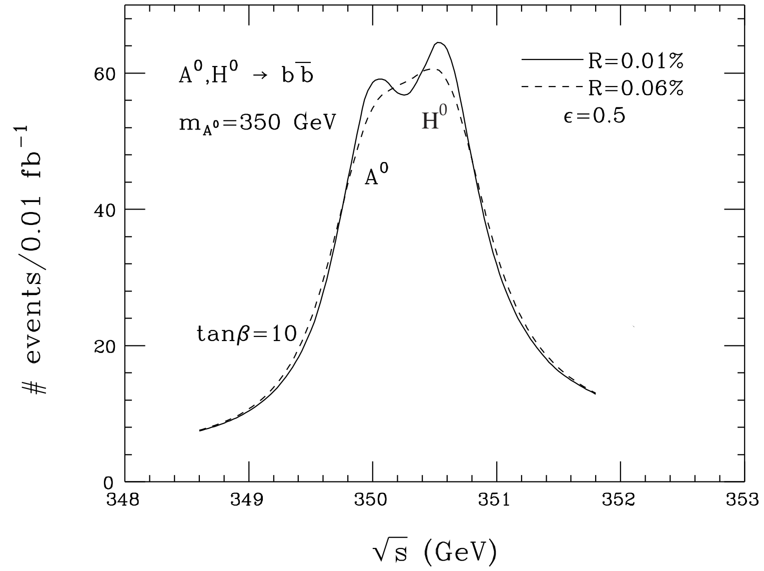


FIG. 8. Separation of A^0 and H^0 signals for $\tan\beta = 10$. From Ref. [44].

Such resonances would be easy to find and study at a muon collider.

D. Exotic narrow resonance possibilities

There are important types of exotic physics that would be best probed in s-channel production of a narrow resonance at a muon collider. Many extended Higgs sector models contain a doubly-charged Higgs boson Δ^{--} (and its Δ^{++} partner) that couples to $\mu^-\mu^-$ via a Majorana coupling. The s-channel process $\mu^-\mu^- \to \Delta^{--}$ has been shown [72] to probe extremely small values of this Majorana coupling, in particular values naturally expected in models where such couplings are responsible for neutrino mass generation. In supersymmetry, it is possible that there is R-parity violation. If R-parity violation is of the purely leptonic type, the coupling $\lambda_{\mu\tau\mu}$ for $\mu^-\mu^+ \to \tilde{\nu}_\tau$ is very possibly the largest such coupling and could be related to neutrino mass generation. This coupling can be probed down to quite small values via s-channel $\tilde{\nu}_\tau$ production at the muon collider [73].

E. Z-factory

A muon collider operating at the Z-boson resonance energy is an interesting option for measurement of polarization asymmetries, B_s^0–$\bar{B}_s^0$ mixing, and of CP violation in the B-meson system [74]. The muon collider advantages are the partial muon beam polarization, and the long B-decay length for B-mesons produced at this $\sqrt{s}$. The left-right asymmetry A_{LR} is the most accurate measure of $\sin^2\theta_{\rm w}$, since the uncertainty is statistics dominated. The present LEP and SLD polarization measurements show standard deviations of 2.4 in A^0_{LR}, 1.9 in $A^{0,b}_{FB}$ and 1.7 in $A^{0,\tau}_{FB}$ [75]. The CP angle β could be measured from $B^0 \to K_s J/\psi$ decays. To achieve significant improvements over existing measurements and those at future B-facilities, a data sample of 10^8 Z-boson events/year would be needed. This corresponds to a luminosity > 0.15 fb^{-1} /year, which is well within the domain of muon collider expectations; $R \sim 0.1\%$ would be more than adequate, given the substantial ~ 2.4 GeV width of the Z.

F. Threshold measurements at a muon collider

With 10 fb^{-1} integrated luminosity devoted to a measurement of a threshold cross-section, the following precisions on particle masses may be achievable [76]:

$$\begin{array}{ll} \mu^+\mu^- \to W^+W^- & \Delta M_W = 20 \text{ MeV}\,, \\ \mu^+\mu^- \to t\bar{t} & \Delta m_t = 0.2 \text{ GeV}\,, \\ \mu^+\mu^- \to Zh & \Delta m_h = 140 \text{ MeV} \end{array} \tag{14}$$

(if $m_h = 100$ GeV) . Precision M_W and m_t measurements allow important tests of electroweak radiative corrections through the relation

$$M_W = M_Z \left[1 - \frac{\pi\alpha}{\sqrt{2}\, G_\mu\, M_W^2 (1-\delta r)} \right]^{1/2} , \tag{15}$$

where δr represents loop corrections. In the SM, δr depends on m_t^2 and $\log m_h$. The optimal precision for tests of this relation is $\Delta M_W \approx \frac{1}{140}\Delta m_t$, so the uncertainty on M_W is the most critical. With $\Delta M_W = 20$ MeV the SM Higgs mass could be inferred to an accuracy

$$\Delta m_{h_{\rm SM}} = 30 \text{ GeV} \left(\frac{m_h}{100 \text{ GeV}} \right) . \tag{16}$$

Alternatively, once m_h is known from direct measurements, SUSY loop contributions can be tested.

In top-quark production at a muon collider above the threshold region, modest muon polarization would allow sensitive tests of anomalous top quark couplings [77].

One of the important physics opportunities for the First Muon Collider is the production of the lighter chargino, $\tilde{\chi}_1^+$ [78]. Fine-tuning arguments in mSUGRA suggest that it should be lighter than 200 GeV. A search at the upgraded Tevatron for the process $q\bar{q} \to \tilde{\chi}_1^+\tilde{\chi}_2^0$ with $\tilde{\chi}_1^+ \to \tilde{\chi}_1^0\ell^+\nu$ and $\tilde{\chi}_2^0 \to \tilde{\chi}_1^0\ell^+\ell^-$ decays can potentially reach masses $m_{\tilde{\chi}_1^+} \simeq m_{\tilde{\chi}_2^0} \sim 170$ GeV with 2 fb^{-1} luminosity and ~ 230 GeV with 10 fb^{-1} [79]. The mass difference $M(\tilde{\chi}_2^0) - M(\tilde{\chi}_1^0)$ can be determined from the $\ell^+\ell^-$ mass distribution.

The two contributing diagrams in the chargino pair production process are shown in Fig. 9; the two amplitudes interfere destructively. The $\tilde{\chi}_1^+$ and $\tilde{\nu}_\mu$ masses can be inferred from the shape of the cross section in the threshold region [80]. The chargino decay is $\tilde{\chi}_1^+ \to f\bar{f}'\tilde{\chi}_1^0$. Selective cuts suppress the background from W^+W^- production and leave $\sim 5\%$ signal efficiency for 4 jets + $\not{E}_T$ events. Measurements at two energies in the threshold region with total luminosity $L = 50$ fb^{-1} and resolution $R = 0.1\%$ can give the accuracies listed in table II on the chargino mass for the specified values of $m_{\tilde{\chi}_1^+}$ and $m_{\tilde{\nu}_\mu}$.

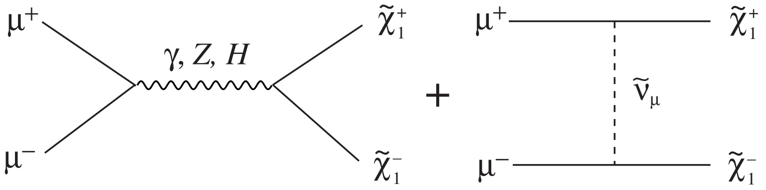


FIG. 9. Diagrams for production of the lighter chargino.

TABLE II. Achievable uncertainties with 50 fb^{-1} luminosity on the mass of the lighter chargino for representative $m_{\tilde{\chi}_1^+}$ and $m_{\tilde{\nu}_\mu}$ masses. From Ref. [80].

$\Delta m_{\tilde{\chi}_1^+}$ (MeV)	$m_{\tilde{\chi}_1^+}$ (GeV)	$m_{\tilde{\nu}_\mu}$ (GeV)
35	100	500
45	100	300
150	200	500
300	200	300

G. Heavy particles of supersymmetry

The requirements of gauge coupling unification can be used to predict the mean SUSY mass scale, given the value of the strong coupling constant at the Z-mass scale. Figure 10 shows the SUSY GUT predictions versus $\alpha_s(M_Z)$. For the value $\alpha_s(M_Z) = 0.1214 \pm 0.0031$ from a new global fit to precision electroweak data [75], a mean SUSY mass of order 1 TeV is expected. Thus, it is likely that some SUSY particles will have masses at the TeV scale. Large masses for the squarks of the first family are perhaps the most likely in that this would provide a simple cure for possible flavor changing neutral current difficulties.

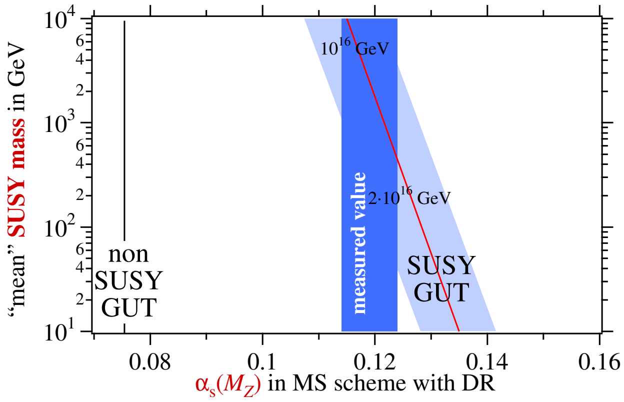


FIG. 10. α_s prediction in supersymmetric GUT with minimal particle content in the Dimensional Regularization scheme.

At the LHC, mainly squarks and gluinos will be produced; these decay to lighter SUSY particles. The LHC will be a great SUSY machine, but some sparticle measurements will be very difficult or impossible there [81,82], namely: (i) the determination of the LSP mass (LHC measurements give SUSY mass differences); (ii) study of sleptons of mass $\gtrsim 200$ GeV because Drell-Yan production becomes too small at these masses; (iii) study of heavy gauginos $\tilde{\chi}_2^\pm$ and $\tilde{\chi}_{3,4}^0$, which are mainly Higgsino and have small direct production rates and small branching fractions to channels usable for detection; (iv) study of heavy Higgs bosons $H^\pm$, H^0, A^0 when the MSSM $\tan\beta$ parameter is not large and their masses are larger than $2m_t$, so that cross sections are small and decays to $t\bar{t}$ are likely to be dominant (their detection is deemed impossible if SUSY decays dominate).

Detection and study of the many scalar particles predicted in supersymmetric models could be a particularly valuable contribution of a high energy lepton collider. However, since pair production of scalar particles at a lepton collider is P-wave suppressed, energies well above threshold are needed for sufficient production rates; see Fig. 11. For scalar particle masses of order 1 TeV a collider energy of 3 to 4 TeV is needed to get past the threshold suppression. A muon collider operating in this energy range with high luminosity ($L \sim 10^2$ to 10^3 fb^{-1}/year) would provide sufficient event rates to reconstruct heavy sparticles from their complex cascade decay chains [82,84].

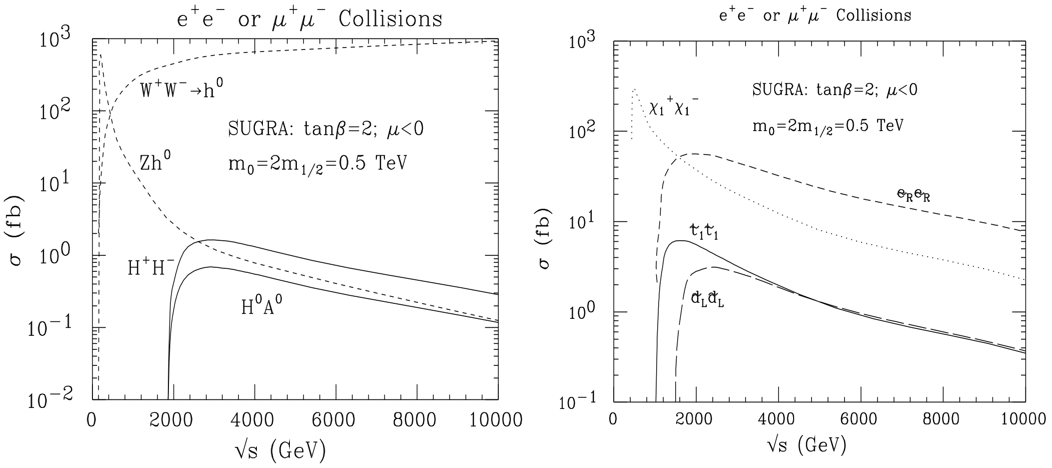


FIG. 11. Cross sections for pair production of Higgs bosons and scalar particles at a high-energy muon collider. From Ref. [83].

In string models, it is very natural to have extra Z bosons in addition to low-energy supersymmetry. The s-channel production of a Z' boson at the resonance energy would give enormous event rates at the NMC. Moreover, the s-channel contributions of Z' bosons with mass far above the kinematic reach of the collider could be revealed as contact interactions [85].

H. Strong scattering of weak bosons

The scattering of weak bosons can be studied at a high-energy muon collider through the process in Fig. 12. The amplitude for the scattering of longitudinally polarized W-bosons behaves like

$$A(W_L W_L \to W_L W_L) \sim m_H^2/v^2 \tag{17}$$

if there is a light Higgs boson, and like

$$A(W_L W_L \to W_L W_L) \sim s_{WW}/v^2 \tag{18}$$

if no light Higgs boson exists; here s_{WW} is the square of the WW CoM energy and $v = 246$ GeV. In the latter scenario, partial-wave unitarity of $W_L W_L \to W_L W_L$ requires that the scattering of weak bosons becomes strong at energy scales of order 1 to 2 TeV. Thus, subprocess energies $\sqrt{s_{WW}} \gtrsim 1.5$ TeV are needed to probe strong WW scattering effects.

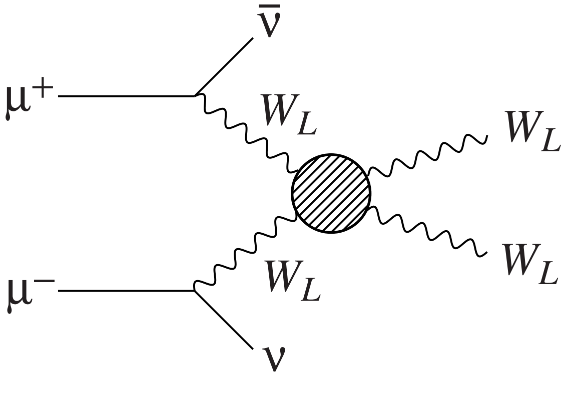


FIG. 12. Symbolic diagram for strong WW scattering.

The nature of the dynamics in the WW sector is unknown. Models for this scattering assume heavy resonant particles (isospin scalar and vector) or a non-resonant amplitude based on a unitarized extrapolation of the low-energy theorem behavior $A \sim s_{WW}/v^2$. In all models, impressive signals of strong WW scattering are obtained at the NMC, with cross sections typically of order 50 fb^{-1} [86]. Event rates are such that the various weak-isospin channels ($I = 0, 1, 2$) could be studied in detail as a function of s_{WW}. After several years of operation, it would even be possible to perform such a study after projecting out the different final polarization states ($W_L W_L$, $W_L W_T$ and $W_T W_T$), thereby enabling one to verify that it is the $W_L W_L$ channel in which the strong scattering is taking place.

I. Front end physics

New physics is likely to have important lepton flavor dependence and may be most apparent for heavier flavors. The intense muon source available at the front end of the muon collider will provide many opportunities for uncovering such physics.

1. Rare muon decays

The planned muon flux of $\sim 10^{14}$ muons/sec for a muon collider dramatically eclipses the flux, $\sim 10^8$ muons/sec, of present sources. With an intense source, the rare muon processes $\mu \to e\gamma$ (for which the current branching fraction limit is 0.49×10^{-12}), $\mu N \to eN$ conversion, and the muon electric dipole moment can be probed at very interesting levels. A generic prediction of supersymmetric grand unified theories is that these lepton flavor violating or CP-violating processes should occur via loops at significant rates, e.g. $\mathrm{BF}(\mu \to e\gamma) \sim 10^{-13}$. Lepton-flavor violation can also occur via Z' bosons, leptoquarks, and heavy neutrinos [87].

2. Neutrino flux

The decay of a muon beam leads to neutrino beams of well defined flavors [88]. A muon collider would yield a neutrino flux 1000 times that presently available. This would result in $\sim 10^6$ νN and $\bar{\nu} N$ events per year, which could be used to measure charm production (~6% of the total cross section) and measure $\sin^2 \theta_{\mathrm{w}}$ (and infer the W-mass to an accuracy $\Delta M_W \simeq 30$–50 MeV in one year) [56–63,65].

3. Neutrino oscillations

A special purpose muon ring has been proposed [57] to store $\sim 10^{21}$ μ^+ or μ^- per year and obtain $\sim 10^{20}$ neutrinos per year from muon decays along ~75-m straight sections of the ring, which would be pointed towards a distant neutrino detector. The neutrino fluxes from $\mu^- \to \nu_\mu \bar{\nu}_e e^-$ or from $\mu^+ \to \bar{\nu}_\mu \nu_e e^+$ decays can be calculated with little systematic error. Then, for example, from the decays of stored μ^-'s, the following neutrino oscillation channels could be studied by detection of the charged leptons from the interactions of neutrinos in the detector:

oscillation	detect
$\nu_\mu \to \nu_e$	e^-
$\nu_\mu \to \nu_\tau$	τ^-
$\bar{\nu}_e \to \bar{\nu}_\mu$	μ^+
$\bar{\nu}_e \to \bar{\nu}_\tau$	τ^+

The detected e^- or μ^+ have the "wrong sign" from the leptons produced by the interactions of the $\bar{\nu}_e$ and ν_μ flux. The known neutrino fluxes from muon decays could be used for long-baseline oscillation experiments at any detector on Earth. The probabilities for vacuum oscillations between two neutrino flavors are given by

$$P(\nu_a \to \nu_b) = \sin^2 2\theta \, \sin^2(1.27 \delta m^2 L/E) \tag{19}$$

with δm^2 in eV^2 and L/E in km/GeV. In a very long baseline experiment from Fermilab to the Gran Sasso laboratory or the Kamioka mine ($L = \mathcal{O}10^4$ km) with ν-energies $E_\nu = 20$ to 50 GeV ($L/E = 500$–200 km/GeV), neutrino charged-current interaction rates of $\sim 10^3$/year would result. In a long baseline experiment from Fermilab to the Soudan mine (L=732 km), the corresponding interaction rate is $\sim 10^4$/year. Such an experiment would have sensitivity to oscillations down to $\delta m^2 \sim 10^{-5}$ eV^2 for $\sin^2 2\theta = 1$ [57].

4. μp collider

The possibility of colliding 200-GeV muons with 1000-GeV protons at Fermilab is under study. This collider would reach a maximum $Q^2 \sim 8 \times 10^4$ GeV^2, which is ∼90 times the reach of the HERA *ep* collider, and deliver a luminosity $\sim 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, which is ∼300 times the HERA luminosity. The μp collider would produce $\sim 10^6$ neutral-current deep-inelastic-scattering events per year at $Q^2 > 5000$ GeV^2, which is more than a factor of 10^3 higher than at HERA. In the new physics realm, leptoquark couplings and contact interactions, if present, are likely to be larger for muons than for electrons. This μp collider would have sufficient sensitivity to probe leptoquarks up to a mass $M_{LQ} \sim 800$ GeV and contact interactions to a scale $\Lambda \sim 6$–9 TeV [89].

J. Summary of the physics potential

The First Muon Collider offers unique probes of supersymmetry (particularly s-channel Higgs boson resonances) and technicolor models (via s-channel production of techni-resonances), high-precision threshold measurements of W, t and SUSY particle masses, tests of SUSY radiative corrections that indirectly probe the existence of high-mass squarks, and a possible Z^0 factory for improved precision in polarization measurements and for B-physics studies of CP violation and mixing.

The Next Muon Collider guarantees access to heavy SUSY scalar particles and Z' states or to strong WW scattering if there are no Higgs bosons and no supersymmetry.

The Front End of a muon collider offers dramatic improvements in sensitivity for flavor-violating transitions (e.g., $\mu \to e\gamma$), access to high-Q^2 phenomena in deep-inelastic muon-proton and neutrino-proton interactions, and the ability to probe very small δm^2 via neutrino-oscillation studies in long-baseline experiments.

The muon collider would be crucial to unraveling the flavor dependence of any type of new physics that is found at the next generation of colliders.

Thus, muon colliders are robust options for probing new physics that may not be accessible at other colliders.

III. PROTON DRIVER

The overview of the required parameters is followed by a description of designs that have been studied in some detail. The section concludes with a discussion of the outstanding open issues.

A. Specifications

The proton-driver requirements are determined by the design luminosity of the collider, and the efficiencies of muon collection, cooling, transport and acceleration. The baseline specification is for a 4-MW, 16-GeV or a 7-MW, 30-GeV proton driver, with a repetition rate of 15 Hz and 10^{14} protons per cycle in 2 bunches (for the 100-GeV machine) or 4 bunches (for the higher energies) of 5×10^{13} or 2.5×10^{13} protons, respectively. Half the bunches are used to make μ^- and the rest for μ^+ [90].

The total beam power is several MW, which is larger than that of existing synchrotrons. However, except for bunch length, these parameters are similar to those of Kaon factories [91] and spallation neutron sources [92]. As in those cases, the proton driver must have very low losses to permit inexpensive maintenance of components.

The rms bunch length for the protons on target has to be about 1 ns to: 1) reduce the initial longitudinal emittance of muons entering the cooling system, and 2) optimize the production of polarized muons. Although bunches of up to 6×10^{13} protons per cycle have been accelerated, the required peak current is 2000 A, which is unprecedented.

Since the collection of highly polarized μ's is inefficient (see section **IV.G**), the proton driver should eventually provide an additional factor of two or more in proton intensity to permit the luminosity to be maintained for polarized muon beams.

B. Possible options

Accelerator designs are site, and to some extent, time dependent, and there have been three studies at three different energies (30 GeV [93], 16 GeV [94] and 24 GeV [95–98]; see also [99]). In general, if the final energy is higher, the required currents are lower, bunch manipulation and apertures are easier, and the final momentum spread and space-charge tune shifts are less. Lowering the final energy gives somewhat more π's/Watt, a lower rf requirement ($V_{rf} \sim E^2$) and perhaps a lower facility cost.

In the low-energy muon collider, where two bunches of protons of 5×10^{13} are required on target, two bunches can be merged outside the driver. These two bunches would be extracted simultaneously from two different extraction ports, and fed by different transmission lines to the same target. By arranging the path lengths of the two lines appropriately, the two bunches can be exactly merged.

1. A generic design

A 7-MW collider-driver design based on parameters originally proposed in the Snowmass Feasibility study [93] consists of a 600-MeV linac, a 3.6-GeV booster and a 30-GeV driver. Both linac and booster are based on the BNL Spallation Neutron Source design [92], using a lower repetition rate and a lower total number of protons per pulse. For the 4-bunch case (2.5×10^{13} protons per bunch), the (95%) bunch area is assumed to be 2 eV-s at injection and < 4.5 eV-s at extraction. The driver lattice is derived from the lattice of the JHF driver using 90° FODO cells with missing dipoles in every third FODO cell, allowing a transition energy that is higher than the maximum energy or, perhaps, imaginary.

2. FNAL study

If a muon collider is built at an existing laboratory, then possibilities abound for symbiotic relationships with the other facilities and programs of that laboratory. For example, the proton driver for a muon collider might result from an upgrade of existing proton-source capabilities, and such an upgrade could then also enhance other future programs that use the proton beams.

Fermilab has conceived such a proton-source development plan [100] with three major components: an upgraded linac and two rapid-cycling (15 Hz) synchrotrons: a prebooster and a new booster. The two synchrotrons operate in series; the four proton bunches for the muon collider are formed in the prebooster and then accelerated sequentially in the prebooster and the booster. The plan could be implemented in stages, and other programs would benefit from each stage, but all three components are required to meet the luminosity goals of the muon colliders that have been considered so far.

TABLE III. Baseline proton-driver parameters of the FNAL study.

	Linac	**Booster**	**Driver**
Energy range (GeV)	1	3	16
Rep. rate (Hz)	15	15	15
RF voltage per turn (MV)		0.15	1.5
Circumference (m)		158	474
Protons per bunch ($\times 10^{13}$)		2.5	2.5
Beam emittance [95%] (π mm-mrad)		200	240
Bunch area [95%] (eV-s)		1.5	< 2.0
Incoherent tune shift @ Inj.		0.39	0.39

Table III presents the major parameters of the two rings. Whenever the needs of the muon collider itself allow some flexibility, the parameters have been chosen to optimize the resulting facility as a proton source for the rest of the future program at Fermilab. For example, the machine circumferences and rf-harmonic numbers result in bunch trains that are compatible with the existing downstream proton machines.

A muon collider requires proton bunches that are both very intense and, at the pion-production target, very short. Strong transverse and longitudinal space-charge forces might disrupt such bunches in the synchrotrons unless measures to alleviate those effects are incorporated in the design. The Laslett incoherent-space-charge tune shift quantifies the severity of the transverse effects. A useful approximation for the space-charge tune shift $\Delta\nu_{sc}$ at the center of a round Gaussian beam is

$$\Delta\nu_{sc} = -\frac{3 r_p N_{\rm tot}}{2\epsilon_n \beta\gamma^2 b} \tag{20}$$

In this expression $r_p = 1.535 \times 10^{-18}$ m is the so-called electromagnetic radius of the proton, $N_{\rm tot}$ is the total number of protons in the ring, ϵ_n is the 95% normalized transverse emittance, β and γ are the usual Lorentz kinematical factors, and $b \leq 1$ is the bunching factor, defined as the ratio of the average beam current to the peak current.

The approximation (20) implies that for a given total number of protons, here 10^{14}, the factors in the denominator are the only ways to reduce the tune shift to a specified maximum tolerable value, taken as 0.4. The bunching factor can be raised somewhat by careful tailoring of beam distributions, but here a typical value of 0.25 is conservatively assumed. Achieving the desired beam intensity then requires a combination of high injection energy, here taken as 1 GeV into the first ring, and large transverse normalized emittances, here assumed to be about 200π mm-mrad. The corresponding required aperture is about 13 cm in the first ring and about 10 cm in the second ring. With such large apertures in rapid-cycling synchrotrons, careful design of the beam pipes for both rings is required to manage eddy-current effects. Two approaches are under consideration. One is a thin Inconel pipe with water cooling and eddy-current coil corrections integrated on the pipe, as in the AGS Booster. The other is a ceramic beam pipe with a conductor inside to carry beam-image currents, as in ISIS.

The Fermilab linac presently delivers a 400-MeV beam, and is capable with modest modifications of accelerating as many as 3×10^{13} protons per cycle at 15 Hz [101]. A significant upgrade is required in order to deliver 10^{14} protons at 1 GeV. The energy can be raised by appending additional side-coupled modules to the downstream end of the linac. Increasing the linac beam intensity probably means increasing both the beam current and the duration of the beam pulse. Injection into the first ring is by charge stripping of the H^- beam; the incoming beam will be chopped and injected into pre-existing buckets to achieve high capture efficiency.

The circumference of the second ring is set equal to that of the existing Fermilab Booster. This choice provides several advantages. First, the new booster could occupy the same tunnel as a relocated Booster; secondly, the beam-batch length from a full second ring matches that of the present Booster, which simplifies matching to downstream machines for other programs. The output energy of 16 GeV then results from an assumed dipole packing fraction of 0.575 and a peak dipole field of 1.3 T, which is the highest dipole field that is consistent with straightforward, nonsaturating design of magnets having thin silicon-steel laminations. Driving such magnets into saturation would cause significant heating of the magnet yoke as well as potential problems with tracking between the dipoles and quadrupoles.

The prebooster also has 1.3-T dipole fields, and its circumference is one third that of the new booster; it operates at an rf harmonic number $h = 4$. The strategy for achieving the required short bunches at the target while alleviating space-charge effects in the rings is to start with four bunches occupying most of the circumference of the first small ring in order to keep the bunching factor large, and to do a bunch-shortening rotation in longitudinal phase space just before extraction from the second synchrotron. The four bunches are accelerated in the first ring to 3 GeV, then transferred bunch-to-bucket into the second ring with its harmonic number $h = 12$. At that energy, the kinematic factor in the tune-shift formula (20) is large enough to compensate for the smaller bunching factor in the second ring. The transfer energy of 3 GeV between the two rings roughly equalizes their space-charge tune shifts.

Both rings employ separated-function lattices with flexible momentum compaction in order to raise their transition energies above their respective extraction energies. This not only avoids having to accelerate beam through transition but also provides other advantages. Intense beams are not subject to certain instabilities such as the negative-mass instability below transition and empirically seem less susceptible to other instabilities such as the microwave instability. Also, the negative natural chromaticity is beneficial for stabilizing the beam below transition, thereby perhaps obviating the need for sextupole correctors, especially in the first ring. Having transition not too far above extraction also provides substantial bucket area in which to accomplish beam-shortening rf manipulations.

Several potential sources of instabilities in the rings have been examined [102], as well as the possibility of compensating the latter effect by inductive inserts in the rings. Space charge is the main factor affecting the stability of the beams; the rings appear to be safe from longitudinal- and transverse-microwave instabilities. Of course, standard stabilizing methods such as active dampers are necessary to counteract some of the instabilities. Flexible momentum-compaction lattices would be useful not only to raise the transition energy above the extraction energy, but also to allow fast changes in the slip factor to facilitate bunch-narrowing manipulations at extraction time. Compensation of longitudinal space-charge effects by means of ferrite-loaded inductive inserts would be useful, especially for the first ring.

The magnet-power-supply circuit for each ring is a 15-Hz resonant system like that of the existing Booster, with dipoles and quadrupoles electrically in series. This implies that the second ring will accelerate only one batch at a time from the first ring, which is all that the muon collider needs. Adding about 15% of second harmonic to the magnet ramp reduces the required peak accelerating voltage by about 25%, which is probably worth doing, especially for the second ring with its large voltage requirement.

One of the advantages of a two-ring system is that the two rings divide the work of accelerating the beam. The rf system of the first ring is relatively modest because of its small circumference and small energy gain; that of the second ring is simplified because its high injection energy means a small rf-frequency swing [103].

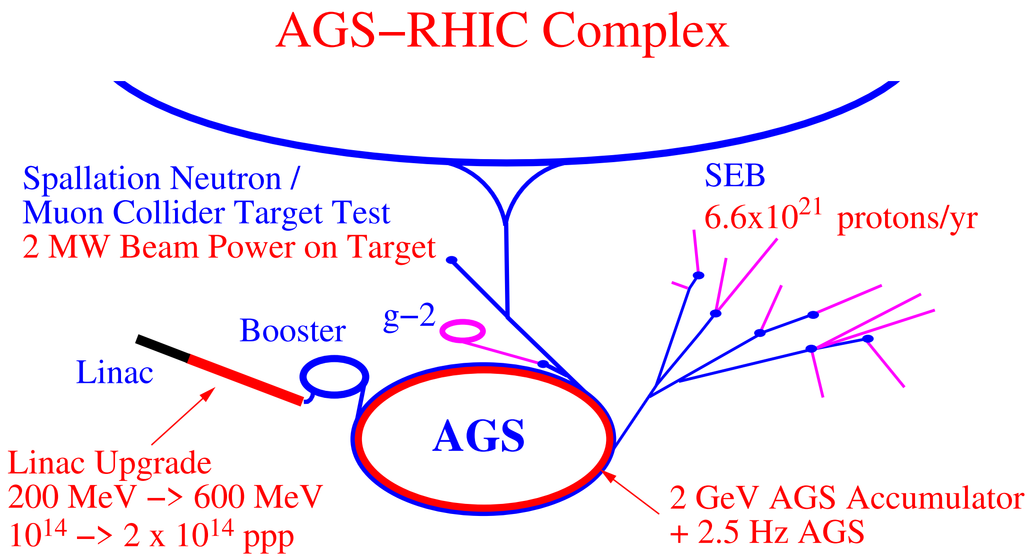


FIG. 13. The AGS–RHIC accelerator complex and a summary of possible intensity upgrades for the AGS.

ESME simulations of longitudinal motion show that the rms bunch length is 2 nsec as desired after the bunch rotation that occurs just before extraction from the second synchrotron.The bunch rotation creates momentum spreads of about 2% with longitudinal emittances of about 2 eV-s per bunch. Such spreads would contribute a few cm in quadrature to the beam size for a short period before extraction. This is thought to be tolerable, given the large apertures that are required in any case. High injection energies help to alleviate these longitudinal effects, which result from space-charge voltages having the same $1/\beta\gamma^2$ kinematic dependence as the transverse tune shifts.

3. AGS upgrade

The third study [95–98] is of an upgrade to the BNL AGS, which should produce bunches larger than those required for the muon collider, but at a lower repetition rate. The AGS presently produces 6×10^{13} protons in eight bunches at 25 GeV and 0.6 Hz. A 2.5-GeV accumulator ring in the AGS tunnel and AGS power-supply upgrade to 2.5-Hz operation would match the repetition rate to the 10-Hz repetition rate of the booster. This would generate 1 MW beam power. With an additional upgrade of the linac energy to 600 MeV, an intensity of 2×10^{14} protons per pulse in four bunches of 5×10^{13} at 25 GeV and 2.5 Hz could be reached, raising the power to 2 MW. The upgrades to the AGS accelerator complex are summarized in Fig.13. Other options are also under consideration, such as the addition of a second booster and 5-Hz operation, that would reach the baseline specification of 4 MW.

The AGS momentum acceptance of ±3% requires that the longitudinal phase space occupied by one bunch be less than 4.5 eV-s. This high bunch density in turn generates stringent demands on the earlier parts of the accelerator cycle. In particular, Landau damping from the beam momentum spread may guard against resistive wall instabilities during injection and longitudinal microwave instabilities after transition. Beam stability can be restored with a more-powerful transverse-damping system and possibly a new low-impedance vacuum chamber. The transverse microwave instability is predicted to occur after transition crossing unless damped by Landau damping from incoherent tune spread or possibly high-frequency quadrupoles.

C. Progress and open issues

Conventional rf manipulations appear able to produce 1- to 2-ns proton bunches if enough rf voltage to overcome the space-charge forces is used, and the beam energy is far enough from transition so the final bunch rotation is fast. Both simulations and experimental work have been directed at demonstrating that a short pulse can be produced easily.

An experiment at the AGS has shown that bunches with $\sigma_z = 2$ ns can be produced near transition from bunches with $\sigma_z \sim 8$ ns by bunch rotation [104,105]. In this experiment, the AGS was flattoped near transition (~ 7 GeV) while the γ_t-jump system was used to bring the transition energy suddenly to the beam energy, letting the bunch-energy spread expand and bunch length contract. The experiment also demonstrated that bunches are stable over periods of 0.1 s. In addition, the data were used to measure the lowest two orders of the momentum compaction factor.

The AGS bunch area, 1.5 eV-s, was comparable to that expected in the proton driver, but the bunch charge (though as large as 3-5 $\times\ 10^{12}$ protons) was only about one tenth of that required by the muon target. The proton driver would use a flexible momentum compaction lattice which would give much better control of the transition energy, permitting a very fast final bunch rotation [106]. In addition, the rf frequency would be higher than that of the AGS so the buckets (and bunches) would initially be only half as long. Thus bunch rotation could be expected to be easier with the new machine, which should compensate for the larger charge.

Simulations with the ESME code have also shown that 1-2 ns bunches of 5×10^{13} can be produced at extraction in a 16-GeV ring with the rf and emittance shown in Table III.

The efficiency of capturing and accelerating beam may be increased by compensation of the space-charge forces in the proton driver. The use of tunable inductive inserts in the ring vacuum chamber may permit active control and compensation of the longitudinal space charge below transition (since the inductive impedance is of the opposite sign to the capacitive space charge). Initial experiments at the KEK proton synchrotron and Los Alamos PSR [107] with short ferrite inserts appear to show a reduction in the synchrotron oscillation frequency shift caused by space charge and a decrease in the necessary rf voltage to maintain a given bunch intensity. Further experiments are needed to demonstrate this technique fully.

The high rf voltage and beam power and the relatively small size of the machine require high-gradient, high-power rf cavities. Fermilab, BNL and KEK are collaborating on research and development of such type of cavities.This work includes the study of magnet alloys and hybrid cavities using both ferrite and new magnet alloys, high-power amplifiers and beam-loading compensation.

The employment of barrier-bucket [108] rf cavities can effectively generate and manipulate a gap in the beam and reduce the space-charge effect. A successful test of this scheme has recently been completed [109], and two $40-$kV barrier cavities have been built by BNL and KEK and are being installed on the AGS.Another high-gradient barrier cavity using magnet alloys is under study at Fermilab.

IV. PION PRODUCTION, CAPTURE AND PHASE ROTATION CHANNEL

This section first discusses the choice of target technology and optimization of the target geometry, and then describes design studies for the pion capture and phase rotation channel. Prospects for polarized muon beams are discussed in detail. The section concludes with an outline of an R&D program for target and phase rotation issues.

Figure 14 gives an overview of the configuration for production of pions by a proton beam impinging on a long, transversely thin target, followed by capture of low-momentum, forward pions in a channel of solenoid magnets with rf cavities to compress the bunch energy while letting the bunch length grow. This arrangement performs the desired rotation of the beam.

A. Pion production

To achieve the luminosities for muon colliders presented in Table I, 2×10^{12} (or 4×10^{12} in the 100 GeV CoM case) muons of each sign must be delivered to the collider ring in each pulse. We estimate that a muon has a probability of only 1/4 of surviving the processes of cooling and acceleration, due to losses in beam apertures or by decay. Thus, 0.8×10^{13} muons ($1.6\ \times 10^{13}$ at 100 GeV) must exit the phase rotation channel each pulse. For pulses of 2.5×10^{13} protons (5×10^{13} for 100 GeV), this requires 0.3 muons per initial proton. Since the efficiency of the phase rotation channel is about 1/2, this is equivalent to a capture of about 0.6 pions per proton: a very high efficiency.

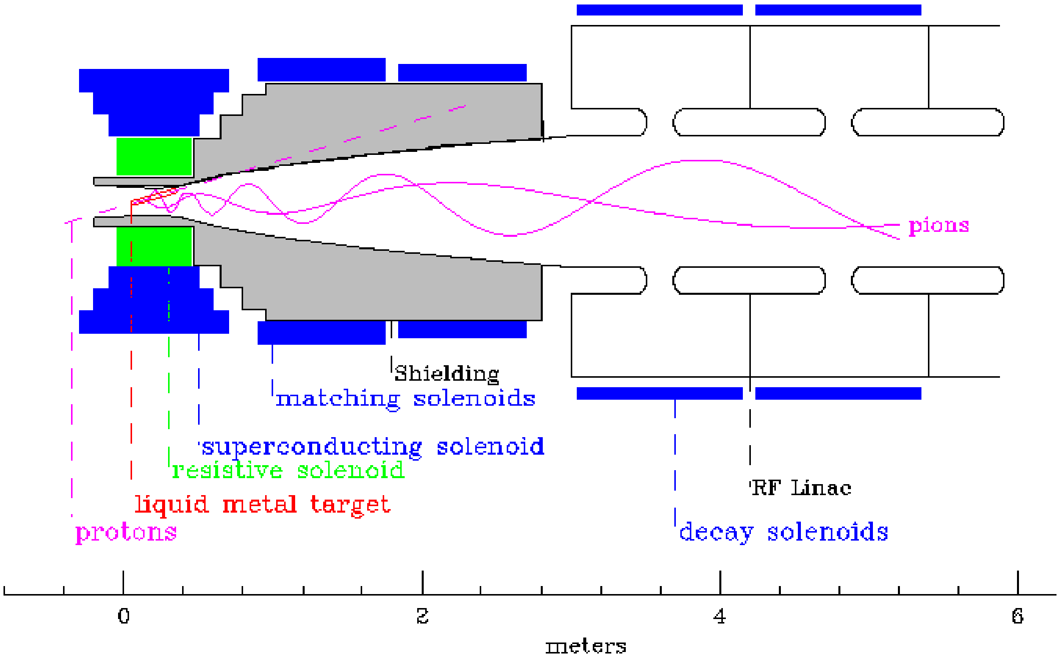


FIG. 14. Schematic view of pion production, capture and initial phase rotation. A pulse of 16-30 GeV protons is incident on a skewed target inside a high-field solenoid magnet followed by a decay and phase rotation channel.

The pions are produced in the interaction of the proton beam with the primary target. Extensive simulations have been performed for pion production from 8-30 GeV proton beams on different target materials in a high-field solenoid [43,110–113]. Three different Monte Carlo codes [114–117] predict similar pion yields despite significant differences in their physics models. Some members of the Collaboration are involved in an AGS experiment BNL E-910 [118] to measure the yield of very low momentum pions, which will validate the codes in the critical kinematic region. This experiment ran for 14 weeks during the Spring of 1996 and has collected over 20 million events, of which about a quarter are minimum bias triggers for inclusive cross section measurements. The targets were varied in material (Be, Cu, Au, U) and thickness (2–100% interaction length (λ_I)) and three different beam momenta were used (6, 12.5, 18 GeV$/c$). Presently, the E910 collaboration is doing a careful analysis of the large data sample obtained. Figure 15 shows the dE/dx energy *vs.* momentum for reconstructed tracks in the TPC; there is clear particle species separation [119].

The pion yield is greater for relatively high Z materials, and for these, the pion yield is maximal for longitudinal momenta of the same order as the average transverse momentum ($\approx$ 200 MeV$/c$). Targets of varying composition ($6 < Z < 82$), radii (0.2-3 cm) and thicknesses (0.5-3 nuclear interaction lengths) have been explored using a Monte Carlo simulation [111]. For a fixed number of interaction lengths, the pion yield per proton rises almost linearly with proton energy, and hence almost proportional to the energy deposited in the target. The yield is higher for medium- and high-Z target materials, with a noticeable gain at $Z > 26$ for 30 GeV proton beams, but with only a minor effect for $E \leq 16$ GeV. This is shown in Fig. 16 where results of detailed MARS13(98) [115] simulations are presented. The curves show the meson yield ($\pi + K$) from the targets in the momentum interval $0.05 \leq P \leq 0.8$ GeV$/c$ (labeled Y) and the number of mesons that are both captured in the high field solenoid and transported into the decay channel (labeled YC). The typical statistical error is a few percent.

B. Target

The target should be 2-3 interaction lengths long to maximize pion production. A high-density material is favored to minimize the size and cost of the capture solenoid magnet. Target radii larger than about 1 cm lead to lower pion rates due to reabsorption, while smaller diameter targets have less production from secondary interactions. Tilting the target by 100-150 mrad minimizes loss of pions by absorption in the target after one turn on their helical trajectory [49,120]. Another advantage of the tilted target geometry is that the high energy and neutral components of the shower can be absorbed in a water-cooled beam dump to the side of the focused beam (see Fig. 14).

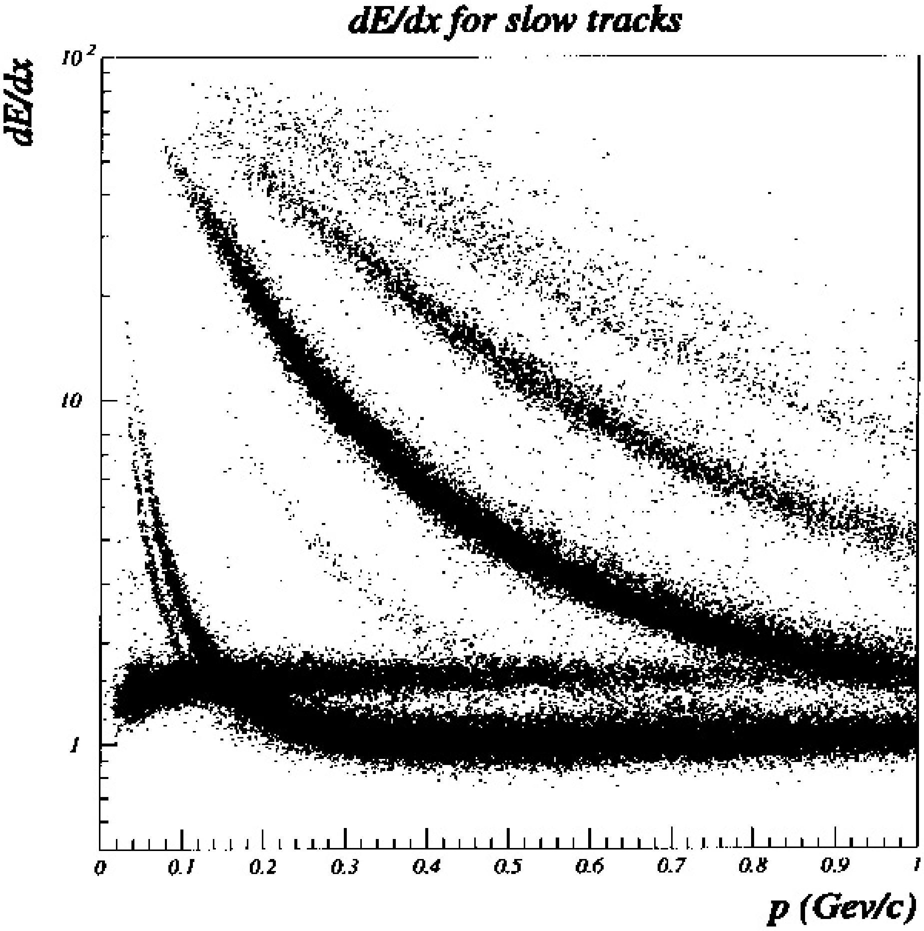


FIG. 15. dE/dx curve in arbitrary units for low momentum tracks; the ionization energy loss is for tracks with 30 or more hits in the TPC. The incident beam momentum is 18 GeV/c. From left to right the bands correspond to muons, pions, kaons, protons and deuterium, respectively. Note the overlap of the (nearly horizontal) electron band with other species.

About 30 kJ of energy is deposited in the target by each proton pulse (10% of the beam energy). Hence, the target absorbs 400 kW of power at the 15-Hz pulse rate. Cooling of the target via contact with a thermal bath would lead to unacceptable absorption of pions, and radiative cooling is inadequate for such high power in a compact target. Therefore, the target must move so as to carry the energy deposited by the proton beam to a heat exchanger outside the solenoid channel.

Both moving solid metal and flowing liquid targets have been considered, with the latter as the currently preferred solution. A liquid is relatively easy to move, easy to cool, can be readily removed and replaced, and is the preferred target material for most spallation neutron sources under study. A liquid flowing in a pipe was considered, but experience at ISOLDE with short proton pulses [121] as well as simulations [122,123] suggest serious problems in shock damage to the pipe. An open liquid jet is thus proposed.

A jet of liquid mercury has been demonstrated [122] but not exposed to a beam. For our application, safety and other considerations favor the use of a low melting point lead alloy rather than mercury. Gallium alloys, though with lower density, are also being considered. Experimental and theoretical studies are underway to determine the consequences of beam shock heating of the liquid. It is expected that the jet will disperse after being exposed to the beam. The target station must survive damage resulting from the violence in this dispersion. This consideration will determine the minimum beam, and thus jet, radius.

For a conducting liquid jet in a strong magnetic field, as proposed, strong eddy currents will be induced in the jet, causing reaction forces that may disrupt its flow [124,125]. The forces induced are proportional to the square of the jet radius, and set a maximum for this radius of order 5-10 mm. If this maximum is smaller than the minimum radius set by shock considerations, then multiple smaller beams and jets could be used; *e.g.*, four jets of 5 mm radius with

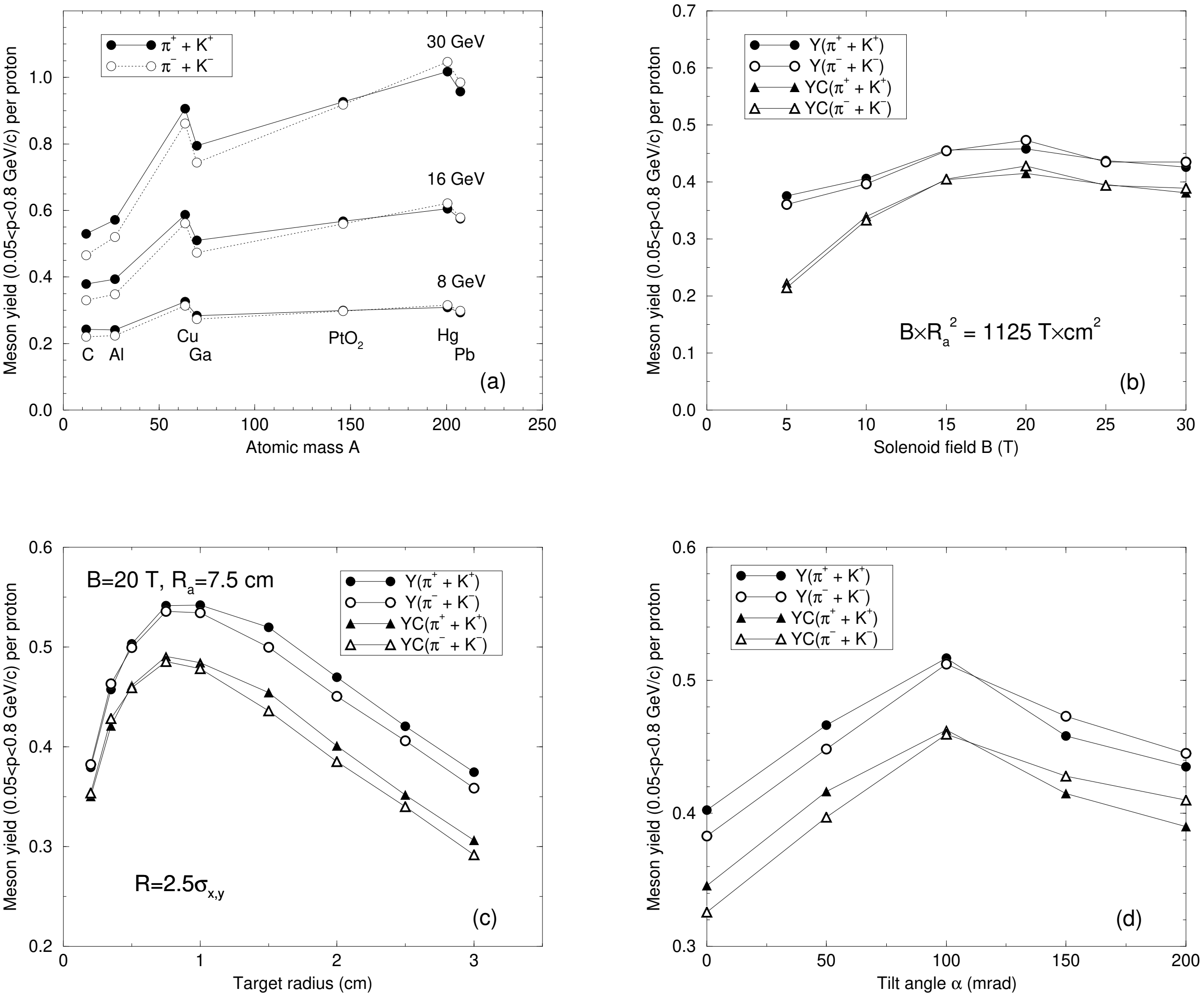


FIG. 16. Meson yield ($\pi + K$) from different targets tilted by angle α in a solenoidal field B of aperture R_a as calculated with the MARS13(98) code. The target is aligned along the beam. The curves labeled YC show mesons that are transported into the decay channel. (a) Yield from a 1.5 λ_I, 1 cm radius target irradiated with 8, 16 and 30 GeV proton beams ($\sigma_x = \sigma_y = 4$ mm) as a function of target atomic mass ($B = 20$ T, $R_a = 7.5$ cm, $\alpha = 0$); (b) Yield from a $3 - \lambda_I$, 1 cm radius gallium target tilted at $\alpha = 150$ mrad in a 16 GeV proton beam ($\sigma_x = \sigma_y = 4$ mm) *vs.* solenoid field for a fixed adiabatic invariant BR_a^2; (c) Yield as a function of radius of a 3 λ_I gallium target in a 16 GeV proton beam ($\sigma_x = \sigma_y = 4$ mm, $B = 20$ T, $R_a = 7.5$ cm, $\alpha = 100$ mrad); (d) Yield from a 3 λ_I, 1 cm radius gallium target *vs.* tilt angle between the axis of the capture solenoid and the proton beam for a 16 GeV proton beam ($\sigma_x = \sigma_y = 4$ mm, $B = 20$ T, $R_a = 7.5$ cm).

four beams with 2.5×10^{13} protons per bunch. Other alternatives include targets made from insulating materials such as liquid PtO_2 or Re_2O_3, slurries (*e.g.*, Pt in water), or powders [126].

A moving solid metal target is not the current baseline solution, but is a serious possibility. In this case, the target could consist of a long flat band or hoop of copper-nickel that moves along its length (as in a band saw) [127]. The band would be many meters in length, would be cooled by gas jets away from the target area, and would be supported and moved by rollers, as shown in Fig. 17.

The choice and parameters of the target are critical issues that need resolution. These can be resolved by experiments with a strong magnetic field and a beam, as discussed in section IV.H.

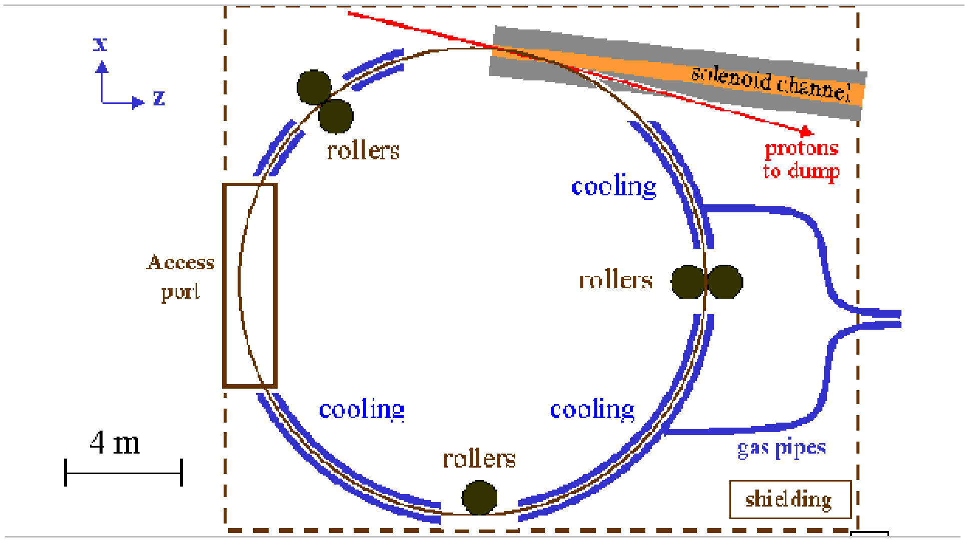


FIG. 17. Alternative concept of a solid metal target in the form of a rotating Cu-Ni band.

C. Capture

To capture all pions with transverse momenta p_T less than their typical value of 200 MeV$/c$, the product of the capture solenoid field B and its radius R_a must be greater than 1.33 T m. The use of a high field and small radius is preferred to minimize the corresponding transverse emittance, which is proportional to BR^2: for a fixed transverse momentum capture, this emittance is thus proportional to R. A field of 20 T and 7.5 cm radius was chosen on the basis of simulations described below. This gives $BR = 1.5$ T m, $BR^2 = 0.1125$ T m^2 and a maximum transverse momentum capture of $p_T = 225$ MeV$/c$.

A preliminary design [128] of the capture solenoid has an inner 6 T, 4 MW, water cooled, hollow conductor magnet with an inside diameter of 24 cm and an outside diameter of 60 cm. There is space for a 4 cm thick, water cooled, heavy metal shield inside the coil. The outer superconducting magnet has three coils, with inside diameters of 60 to 80 cm. It generates an additional 14 T of field at the target and provides the required tapered field to match into the decay channel. Such a hybrid solenoid has parameters compatible with those of existing magnets [129].

The 20 T capture solenoid is matched via a transfer solenoid [110] into a decay channel consisting of a system of superconducting solenoids with the same adiabatic invariant $BR^2 \propto Rp_T$. Thus, for a 1.25 T decay channel, B drops by a factor of 16 between the target and decay channel, R and p_T change by factors of 4 and 1/4, respectively. This permits improved acceptance of transverse momentum within the decay channel, at the cost of an increased spread in longitudinal momentum. Figure 16(b) shows the meson yield as a function of field in the capture solenoid, with the radius of the capture solenoid adjusted to maintain the same BR_a^2 as in the decay channel. The optimum field is 20 T in the capture solenoid.

If the axis of the target is coincident with that of the solenoid field, then there is a relatively high probability that pions re-enter the target after one cycle on their helical trajectory and are lost due to nuclear interactions. When the target and proton beam are set at an angle of 100-150 mrad with respect to the field axis [111], the probability for such pion interactions at the target is reduced, and the overall production rate is increased by 60%, as shown in Fig. 16(d).

In summary, the simulations indicate that a 20 T solenoid of 16 cm inside diameter surrounding a tilted target will capture about half of all produced pions. With target efficiency included, about 0.6 pions per proton will enter the pion decay channel [111].

D. Phase rotation linac

The pions, and the muons into which they decay, have a momentum distribution with an rms spread of approximately 100% and a peak at about 200 MeV$/c$. It would be difficult to handle such a wide spread in any subsequent system. A linac is thus introduced along the decay channel, with frequencies and phases chosen to decelerate the fast particles and accelerate the slow ones; *i.e.*, to phase rotate the muon bunch. Several studies have been made of the design of this system, using differing ranges of rf frequency, delivering different final muon momenta, and differing final bunch lengths. In all cases, muon capture efficiencies of close to 0.3 muons per proton are obtained. Until the early stages of the ionization cooling have been designed, it is not yet possible to choose between them. Independent of the above choices is a question of the location of the focusing solenoid coils and rf cavity design, as discussed below in the section IV.F.

TABLE IV. Parameters of the Lower-Energy Phase Rotation Linacs

Linac	Length (m)	Frequency (MHz)	Gradient (MeV/m)
1	3	60	5
2	29	30	4
3	5	60	4
4	5	37	4

1. Lower energy, longer bunch example

This example captures muons at a mean kinetic energy of 130 MeV. Table IV gives parameters of the linacs used. Monte Carlo simulations [51], with the program MUONMC [130], were done using pion production calculated by ARC [114] for a copper target of 1-cm radius at an angle of 150 mrad. A uniform solenoidal field was assumed in the phase rotation, and the rf was approximated by a series of kicks.

Figure 18 shows the energy *vs.* ct at the end of the decay and phase rotation channel. The abscissa ct is a measure of bunch length at the end of the channel: the total transit time of each π/μ is multiplied by the velocity of light and the total length of the channel is subtracted. Thus a ficticious reference particle at the center of the incident bunch at the target arrives at $ct = 0$ m. A loose final bunch selection was defined with an energy 130 ± 70 MeV and bunch ct from 3 to 11 m. With this selection, the rms energy spread is 16.5%, the rms ct is 1.7 m, and there are 0.39 muons per incident proton. A tighter selection with an energy 130 ± 35 MeV and bunch ct from 4 to 10 m gave an rms energy spread of 11.7%, rms ct of 1.3 m, and contained 0.31 muons per incident proton.

2. Higher energy, shorter bunch example

In this example the captured muons have a mean kinetic energy close to 320 MeV. It is based on a Monte Carlo study which uses the updated MARS pion production model [116] to generate pions created by 16 GeV protons on a 36 cm long, 1 cm radius coaxial gallium target. Figure 19 shows the longitudinal phase space of the muons at the end of an 80 m long, 5 T solenoidal decay channel with cavities of frequency in the 30-90 MHz range and acceleration gradients of 4-18 MeV/m. A total of 0.33 muons per proton fall within the indicated cut (6 m×300 MeV). The rms bunch length inside the cut is 148 cm and rms energy spread is 62 MeV. The normalized 6-D emittance (6-D) is 217 cm^3 and the transverse part is 1.86 cm (the normalized 6-D emittance $\epsilon_{6,N}$ is defined in section **V**).

A sample simulation with lithium hydride absorbers regularly spaced in the last 60 m of a 120 m decay channel and with compensating acceleration captures 0.3 muons with mean kinetic energy of about 380 MeV in a (6 m×300 MeV) window. The longitudinal phase space is about the same as in the previous example but the transverse part shrinks to 0.95 cm due to ionization cooling which reduces the 6-D phase space to 73.5 cm^3.

E. Use of both signs

Protons on the target produce pions of both signs, and a solenoid will capture both, but the subsequent rf systems will have opposite effects on each sign. The proposed baseline approach uses two separate proton bunches to create separate positive and negative pion bunches and accepts the loss of half the pions/muons during phase rotation.

If the pions can be charge separated with limited loss before the phase rotation cavities are reached, then higher luminosity may be obtained. The separation of charged pions in a curved solenoid decay line was studied in [110]. Because of the resulting dispersion in a bent solenoid, an initial beam of radius R with maximum-to-minimum

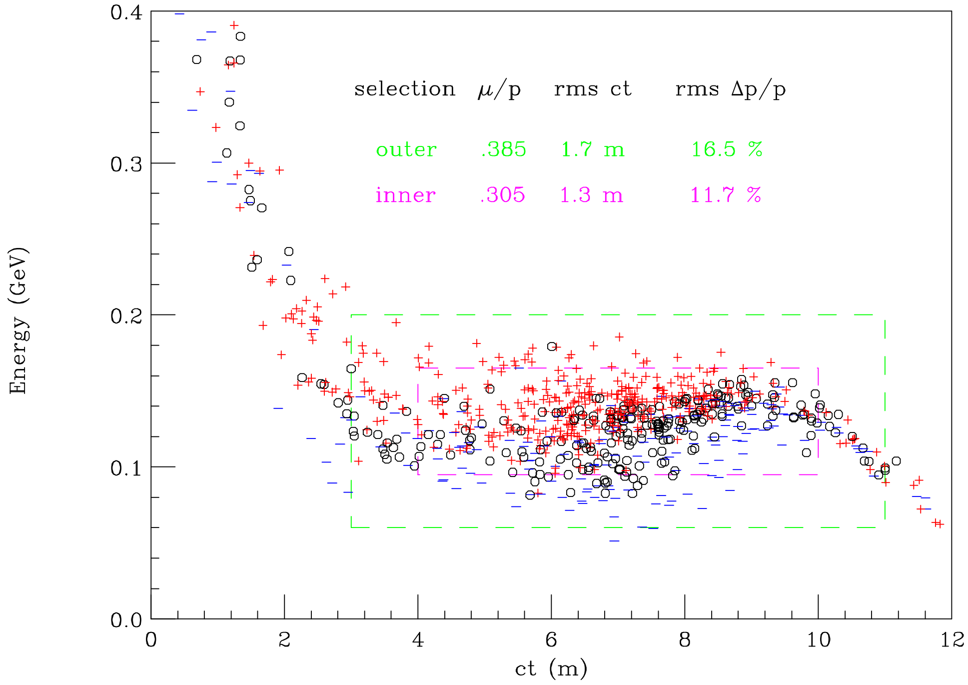


FIG. 18. Energy *vs. ct* of μ's at end of the lower-energy phase rotation channel. The symbols +, o and $-$ denote muons with polarization $P > \frac{1}{3}$, $-\frac{1}{3} < P < \frac{1}{3}$ and $P < -\frac{1}{3}$, respectively.

momentum ratio F will require a large beampipe of radius $(1 + F)R$ downstream to accommodate the separated beams. A septum can then be used to capture the two beams into separate channels. Typically the reduction in yield for a curved solenoid compared to a straight solenoid is about 25% (due to the loss of very low and very high momentum pions to the walls or septum), but this must be weighed against the fact that both charge signs are captured for an overall net gain. A disadvantage is that this charge separation takes place over several meters of length during which time the beam spreads longitudinally. This makes capture in an rf phase rotation system difficult, although a large aperture cavity system could be incorporated in the bent solenoid region to alleviate this. The technique deserves further study and may be useful to consider as an intensity upgrade to a muon collection system.

F. Solenoids and rf

As noted above, capture using higher frequencies appears to be less efficient, and most studies now use frequencies down to 30 MHz. Such cavities, when conventionally designed, are very large (about 6.6 m diameter). In the Snowmass study [131] a reentrant design reduced this diameter to 2.52 m, but this is still large, and it was first assumed that the 5 T focusing solenoids would, for economic reasons, be placed within the irises of the cavities (see Fig. 20).

A study of transmission down a realistic system of iris located coils revealed betatron resonant excitation from the magnetic field periodicities, leading to significant particle loss. This was reduced by the use of more complicated coil shapes [131], smaller gaps, and shorter cavities, but remained a problem.

An alternative is to place continuous focusing coils outside the cavities as shown in Fig. 14. In this case, cost will be minimized with lower magnetic fields (1.25-2.5 T) and correspondingly larger decay channel radii (21-30 cm). Studies are underway to determine the optimal solution.

G. Polarization

Polarization of the muon beams presents a significant physics advantage over the unpolarized case, since signal and background of electroweak processes usually come predominantly from different polarization states.

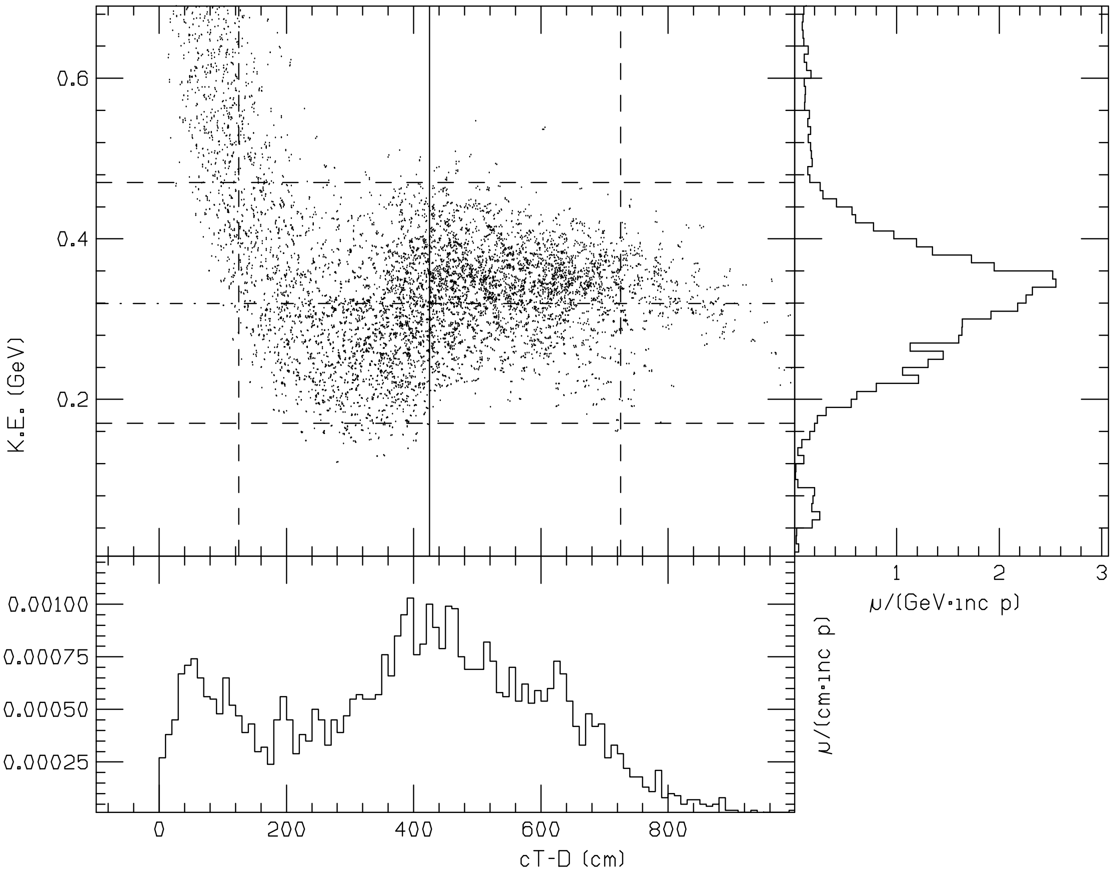


FIG. 19. Longitudinal phase space at the end of decay channel with projections onto time and energy axes. The four dashed lines delineate the region deemed acceptable for the cooling channel.

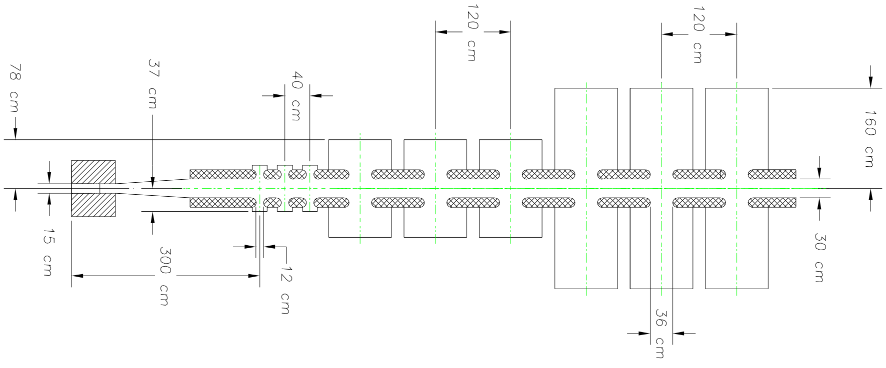


FIG. 20. Schematic of capture and phase rotation using rf cavities with superconducting solenoids (hatched) inside the irises. From left to right, it is shown only three cavities of operating frequencies 90, 50, and 30 MHz respectively.

1. Polarized muon production

In the center of mass of a decaying pion, the outgoing muon is fully polarized ($P = -1$ for μ^+ and $+1$ for μ^-). In the lab system the polarization depends on the decay angle θ_d and initial pion energy [132–134]. For pion kinetic energy larger than the pion mass, the average polarization is about 20%, and if nothing else is done, the polarization of the captured muons after the phase rotation system is approximately this value.

If higher polarization is required, some selection of muons from forward pion decays ($\cos\theta_d \to 1$) is required. Figure 18, above, showed the polarization of the phase rotated muons. The polarization {P$>\frac{1}{3}$, $-\frac{1}{3} < P < \frac{1}{3}$, and P$< -\frac{1}{3}$} is marked by the symbols +, **o** and − respectively. If a selection is made on the minimum energy of the muons, then greater polarization is obtained. The tighter the cut, the higher the polarization, but the less the number of muons that survive. The fraction $F_{\rm loss}$ represents the loss of luminosity of the collider. Figure 21 gives the results of a Monte Carlo study.

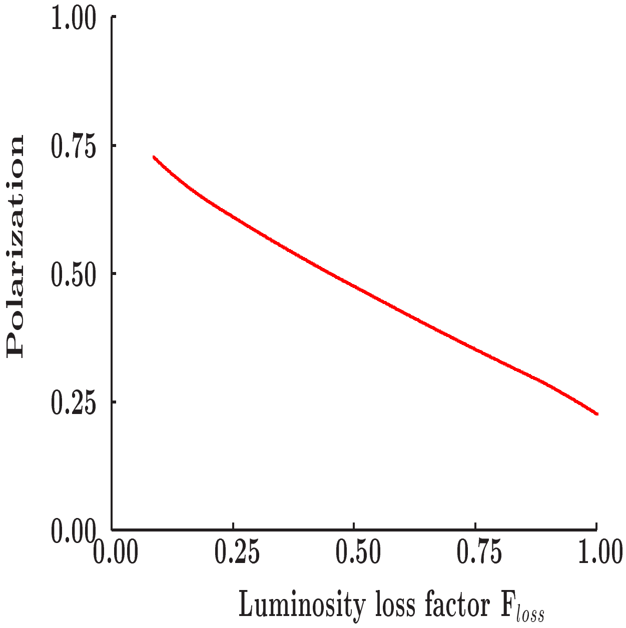


FIG. 21. Polarization *vs.* fraction $F_{\rm loss}$ of μ's accepted.

If this selection is made on both beams, and if the proton bunch intensity is maintained, then each muon bunch is reduced by the factor $F_{\rm loss}$ and the luminosity would fall by $F^2_{\rm loss}$. But if, instead, proton bunches are merged so as to obtain half as many bunches with twice the intensity, then the muon bunch intensity is maintained and the luminosity (and repetition rate) falls only as $F_{\rm loss}$.

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 might 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.

Thus, the goal of high muon beam polarization may shift the parameters of the muon collider towards lower repetition rate and higher peak currents in the proton driver.

2. Polarization preservation

The preservation of muon polarization has been discussed in some detail in [135]. During the ionization-cooling process the muons lose energy in material and have a spin-flip probability $\mathcal{P}$,

$$\mathcal{P} \approx \int \frac{m_e}{m_\mu} \beta_v^2 \frac{\delta E}{E}, \tag{21}$$

where $\delta E/E$ is the fractional loss of energy due to ionization. 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 $\mathcal{P}/\mathcal{P}$ is twice the spin-flip probability, so the reduction in polarization is approximately 20%. This loss is included in Fig. 21.

During circulation in any ring, the muon spin, if initially longitudinal, will precess by $\gamma(g-2)/2$ turns per revolution, where $(g-2)/2$ is 1.166×10^{-3}. A given energy spread $\Delta\gamma/\gamma$ will introduce variations in these precessions and cause dilution of the polarization. But if the particles remain in the ring for an exact integer number of synchrotron oscillations, then their individual average γ's will be the same and no dilution will occur.

In the collider, 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, but long. This might be a preferred solution at high energies but is not practical in the 100 GeV machine. An alternative is to use such a small energy spread, as in the Higgs factory, that although the polarization vector precesses, the beam polarization does not become significantly diluted. In addition, calibration of the Higgs factory collider energy to 1 part in a million [12] requires the spins to precess continuously from turn to turn.

H. R&D program

An R&D program is underway to continue theoretical studies (optimization of pion production and capture) and to clarify several critical issues related to targetry and phase rotation [136]. A jet of the room temperature eutectic liquid alloy of Ga-Sn will be exposed to nanosecond pulses of 1.5×10^{13} 24 GeV protons at the Brookhaven AGS to study the effect of the resulting pressure wave on the liquid. The same jet will also be used in conjunction with a 20 T, 20 cm bore resistive magnet at the National High Magnetic Field Laboratory (Tallahassee, FL) to study the effect of eddy currents on jet propagation. Then, a pulsed, 20 T magnet will be added to the BNL test station to explore the full configuration of jet, magnet and pulsed proton beam. Also, a 70 MHz rf cavity will be exposed to the intense flux of secondary particles downstream of the target and 20 T magnet to determine viable operating parameters for the first phase rotation cavity. The complete configuration of the targetry experiment is sketched in Fig. 22.

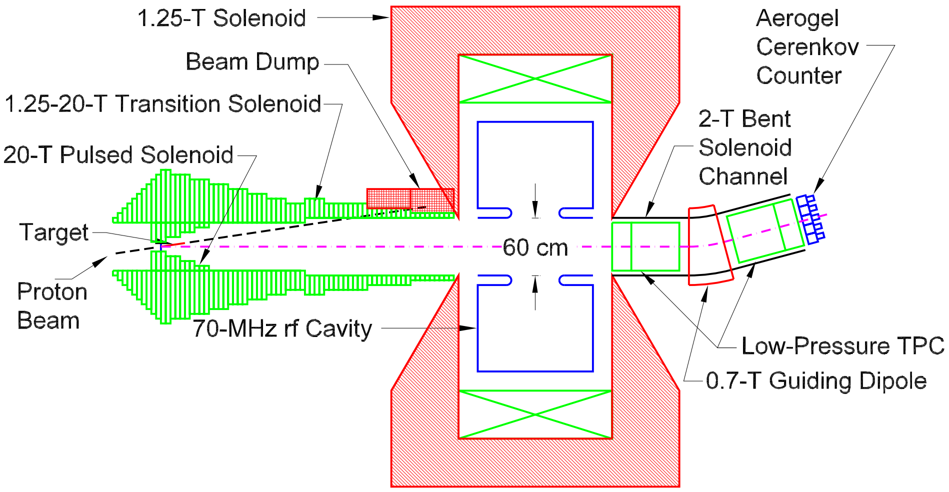


FIG. 22. Plan view of the full configuration of the targetry experiment.

The first two studies should be accomplished during 1999, and the third and fourth in the years 2000/01.

V. IONIZATION COOLING

A. Introduction

The design of an efficient and practical cooling system is one of the major challenges for the muon collider project. For a high luminosity collider, the 6-D phase space volume occupied by the muon beam must be reduced by a factor of $10^5 - 10^6$. Furthermore, this phase space reduction must be done within a time that is short compared to the muon lifetime (μ lifetime $\approx 2\ \mu s$). Cooling by synchrotron radiation, conventional stochastic cooling and conventional electron cooling are all too slow. Optical stochastic cooling [137], electron cooling in a plasma discharge [138], and cooling in a crystal lattice [139,140] are being studied, but appear technologically difficult. The new method proposed for cooling muons is ionization cooling. This technique [15,17,19,141] is uniquely applicable to muons because of their minimal interaction with matter. It is a method that seems relatively straightforward in principle, but has proven quite challenging to implement in practice

Ionization cooling involves passing the beam through some material in which the muons lose both transverse and longitudinal momentum by ionization energy loss, commonly referred to as dE/dx. The longitudinal muon momentum is then restored by reacceleration, leaving a net loss of transverse momentum (transverse cooling). The process is repeated many times to achieve a large cooling factor.

The energy spread can be reduced by introducing a transverse variation in the absorber density or thickness (e.g. a wedge) at a location where there is dispersion (the transverse position is energy dependent). This method results in a corresponding increase of transverse phase space and is thus an exchange of longitudinal and transverse emittances. With transverse coling, this allows cooling in all dimensions.

We define the root mean square rms normalized emittance as

$$\epsilon_{i,N} = \sqrt{\langle \delta r_i^2 \rangle \langle \delta p_i^2 \rangle - \langle \delta r_i \delta p_i \rangle^2} / m_\mu c \tag{22}$$

where r_i and p_i are the beam canonical conjugate variables with $i = 1, 2, 3$ denoting the x, y and z directions, and $\langle ... \rangle$ indicates statistical averaging over the particles. The operator δ denotes the deviation from the average, so that $\delta r_i = r_i - \langle r_i \rangle$ and likewise for δp_i. The appropriate figure of merit for cooling is the final value of the 6-D relativistically invariant emittance $\epsilon_{6,N}$, which is proportional to the area in the 6-D phase space (x, y, z, p_x, p_y, p_z) since, to a fairly good approximation, it is preserved during acceleration and storage in the collider ring. This quantity is the square root of the determinant of a general quadratic moment matrix containing all possible correlations. However, until the nature and practical implications of these correlations are understood, it is more conservative to ignore the correlations and use the following simplified expression for 6-D normalized emittance,

$$\epsilon_{6,N} \approx \epsilon_{x,N} \times \epsilon_{y,N} \times \epsilon_{z,N} \tag{23}$$

Theoretical studies have shown that, assuming realistic parameters for the cooling hardware, ionization cooling can be expected to reduce the phase space volume occupied by the initial muon beam by a factor of $10^5 - 10^6$. A complete cooling channel would consist of $20 - 30$ cooling stages, each stage yielding about a factor of two in 6-D phase space reduction.

It is recognized that the feasibility of constructing a muon ionization cooling channel is on the critical path to understanding the viability of the whole muon collider concept. The muon cooling channel is the most novel part of a muon collider complex. Steady progress has been made both in improving the design of sections of the channel and in adding detail to the computer simulations. A vigorous experimental program is needed to verify and benchmark the computer simulations.

The following parts of this section briefly describe the physics underling the process of ionization cooling. We will show results of simulations for some chosen examples, and outline a six year R&D program to demonstrate the feasibility of using ionization cooling techniques.

B. Cooling theory

In ionization cooling, the beam loses both transverse and longitudinal momentum as it passes through a material. At the same time its emittance is increased due to stochastic multiple scattering and Landau straggling. The longitudinal momentum can be restored by reacceleration, leaving a net loss of transverse momentum.

The approximate equation for transverse cooling in a step ds along the particle's orbit is [13,17,19,23,142,143]

$$\frac{d\epsilon_N}{ds} = -\frac{1}{\beta^2}\frac{dE_\mu}{ds}\,\frac{\epsilon_N}{E_\mu} + \frac{\beta_\perp (0.014GeV)^2}{2\beta^3 E_\mu m_\mu\, L_R}, \tag{24}$$

where β is the normalized velocity, E_μ is the total energy, m_μ is the muon mass, ϵ_N is the normalized transverse emittance, $\beta_\perp$ is the betatron function at the absorber, dE_μ/ds is the energy loss per unit length, and L_R is the radiation length of the material. The betatron function is determined by the strengths of the elements in the focusing lattice [144]. Together with the beam emittance this function determines the local size and divergence of the beam. Note that the energy loss dE_μ/ds is defined here as a positive quantity, unlike the convention often used in particle physics. The first term in this equation is the cooling term, and the second describes the heating due to multiple scattering. The heating term is minimized if $\beta_\perp$ is small (strong-focusing) and L_R is large (a low-Z absorber).

The minimum, normalized transverse emittance that can be achieved for a given absorber in a given focusing field is reached when the cooling rate equals the heating rate in Eq. 24

$$\epsilon_{N,min} = \frac{\beta_\perp (14MeV)^2}{2\beta m_\mu \frac{dE_\mu}{ds} L_R} \tag{25}$$

For a relativistic muon in liquid hydrogen with a betatron focusing value of 8 cm, which corresponds roughly to confinement in a 15 T solenoidal field, the minimum achievable emittance is about 340 mm-mrad.

The equation for energy spread (longitudinal emittance) is [17,23,24].

$$\frac{d(\Delta E_\mu)^2}{ds} = -2\,\frac{d\left(\frac{dE_\mu}{ds}\right)}{dE_\mu}\langle(\Delta E_\mu)^2\rangle + \frac{d(\Delta E_\mu)^2_{\text{stragg.}}}{ds} \tag{26}$$

where the first term describes the cooling (or heating) due to energy loss, and the second term describes the heating due to straggling. ΔE_μ is the rms spread in the energy of the beam.

Ionization cooling of muons seems relatively straightforward in theory, but will require extensive simulation studies and hardware development for its optimization. There are practical problems in designing lattices that can transport and focus the large emittance beam. There will also be effects from space charge and wake fields.

We have developed a number of tools for studying the ionization cooling process. First, the basic theory was used to identify the most promising beam properties, material type and focusing arrangements for cooling. Given the practical limits on magnetic field strengths, this gives an estimate of the minimum achievable emittance for a given configuration. Next, the differential equations for cooling and heating described above have been incorporated into a computer code. Allowance for the shifts in the betatron phase advance due to space charge and aberrations was included. This code was used to develop an overall cooling scenario, which broke the cooling system into a number of stages, and determined the properties of the beam, radio frequency (rf) cavities, and focusing lattice at each stage.

Finally, several tracking codes were either written or modified to study the cooling process in detail. Two new codes (SIMUCOOL [145], and ICOOL [146]) use Monte Carlo techniques to track particles one at a time through the cooling system. All the codes attempt to include all relevant physical processes to some degree, e.g. (energy loss, straggling, multiple scattering) and use Maxwellian models of the focusing fields. They do not yet take into account any space charge or wake field effects. In addition, we have also used a modified version of PARMELA [147] for tracking, which does include space charge effects, and a double precision version of GEANT [148,149].

We have recently developed [150] a model of beam cooling based on a second order moment expansion. A computer code solving the equations for transverse cooling gives results that agree with tracking codes. The code is being extended to include energy spread and bends. It is very fast and is appropriate for preliminary design and optimization of the cooling channel. All of these codes are actively being updated and optimized for studying the cooling problem.

C. Cooling system

The cooling is obtained in a series of cooling stages. Each stage consists of a succession of the following components:

1. Transverse cooling sections using materials in a strong focusing (low $\beta_\perp$) environment alternated with linear accelerators.
2. Emittance exchange in lattices that generate dispersion, with absorbing wedges to reduce momentum spread.
3. Matching sections to optimize the transmission and cooling parameters of the following section.

In the examples that follow it is seen that each such stage lowers the 6-D emittance by a factor of about 2. Since the required total 6-D cooling is $O(10^6)$, about 20 such stages are required. The total length of the system would be of the order of 600 m, and the total acceleration required would be approximately 6 GeV. The fraction of muons remaining at the end of the cooling system is estimated to be $\approx 60\%$.

The baseline solution for transverse cooling involves the use of liquid hydrogen absorbers in strong solenoid focusing fields, interleaved with short linac sections. The solenoidal fields in successive absorbers are reversed to avoid build up of the canonical angular momentum. [151] The focusing magnetic fields are small ($\approx$ 1 T) in the early stages where the emittances are large, but must increase as the emittance falls.

Three transverse cooling examples have been designed and simulated. The first uses 1.25 T solenoids to cool the very large emittance beam coming from the phase rotation channel. The muon beam at the end of the decay channel is very intense, with approximately 7.5×10^{12} muons/bunch, but with a large normalized transverse emittance ($\epsilon_{x,N}(\text{rms}) \approx 15 \times 10^3$ mm-mrad) and a large normalized longitudinal emittance ($\epsilon_{z,N}(\text{rms}) \approx 612$ mm). The second would lie toward the end of a full cooling sequence and uses 15 T solenoids. The third, using 31 T solenoids, meets the requirements for the Higgs factory and could be the final cooling stage for this machine.

The baseline solution for emittance exchange involves the use of bent solenoids to generate dispersion and wedges of hydrogen or LiH to reduce the energy spread. A simulated example is given for exchange that would be needed after the 15 T transverse cooling case.

A lithium lens solution may prove more economical for the final stages, and might allow even lower emittances to be obtained. In this case, the lithium lens serves simultaneously to maintain the low $\beta_\perp$, and provide dE/dx for cooling. Similar lenses, with surface fields of 10 T, were developed at Novosibirsk (BINP) and have been used, at low repetition rates, as focusing elements at FNAL and CERN [7,152–155]. Lenses for the cooling application, which would operate at 15 Hz, would need to employ flowing liquid lithium to provide adequate thermal cooling. Higher surface fields would also be desirable.

Studies have simulated cooling in multiple lithium lenses, and have shown cooling through several orders of magnitude. [11]. But these studies have, so far, used ideal matching and acceleration. Cooling is also being studied in beam recirculators, which could lead to reduction of costs of the cooling section [156,157], but full simulations with all higher order effects have not yet been successfully demonstrated.

D. 15 T solenoid transverse cooling example

The lattice consists of 11 identical 2 m long *cells*. In each cell there is a liquid hydrogen absorber (64 cm long, 10 cm diameter) in the 15 T solenoid focusing magnet (64 cm long, 12 cm diameter). The direction of the fields in the magnets alternates from one cell to the next. Between the 15 T solenoids there are magnetic matching sections (1.3 m long, 32 cm inside diameter) where the field is lowered and then reversed. Inside the matching sections are short, 805 MHz, high gradient (36 MeV/m) linacs. Figure 23 shows the cross section of one cell of such a system, together with the betatron function, and the magnetic field along the axis. For convenience in modeling, the section shown in Fig. 23(a) starts and ends symmetrically in the middle of hydrogen absorber regions at the location of the peaks in the axial magnetic field. In practice each cell would start at the beginning of the hydrogen region and extend to the end of the rf module.

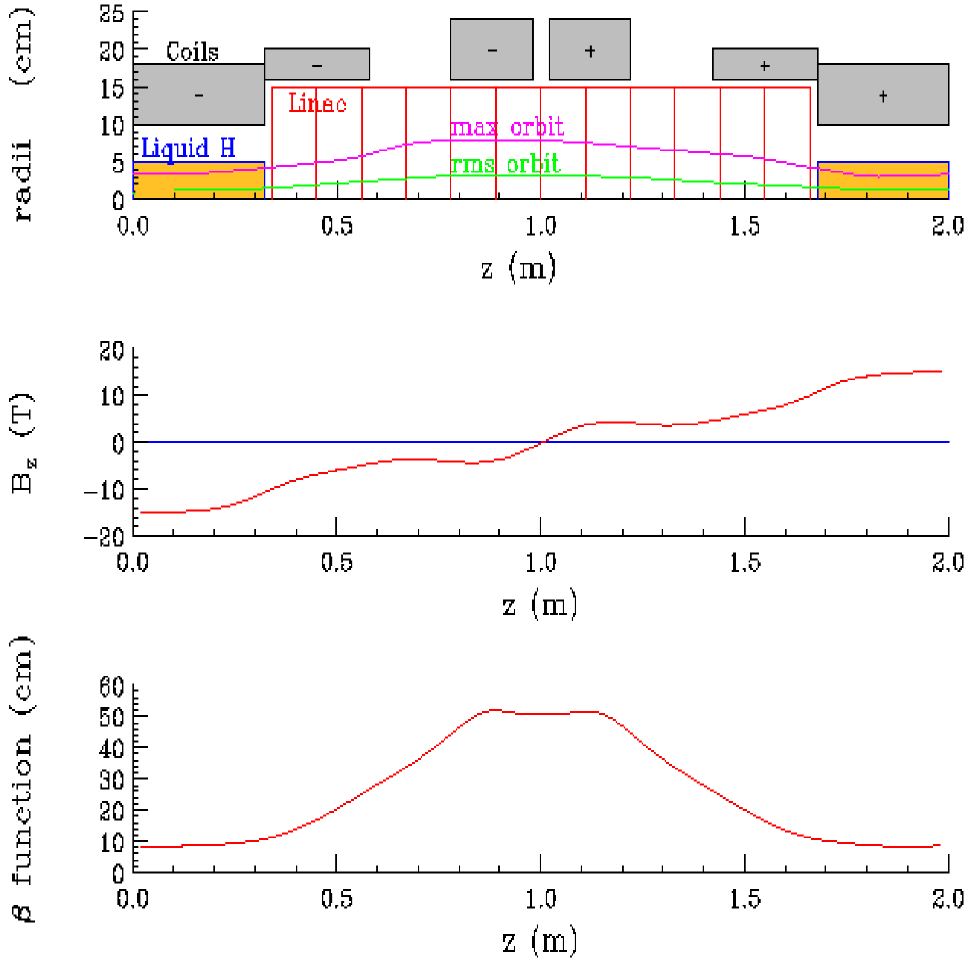


FIG. 23. (a) Cross section of one half period of an alternating solenoid cooling lattice; (b) axial magnetic field *vs.* z; (c) $\beta_\perp$ function *vs.* z.

A GEANT simulation of muons traversing a section of the cooling section is shown in Fig. 24.

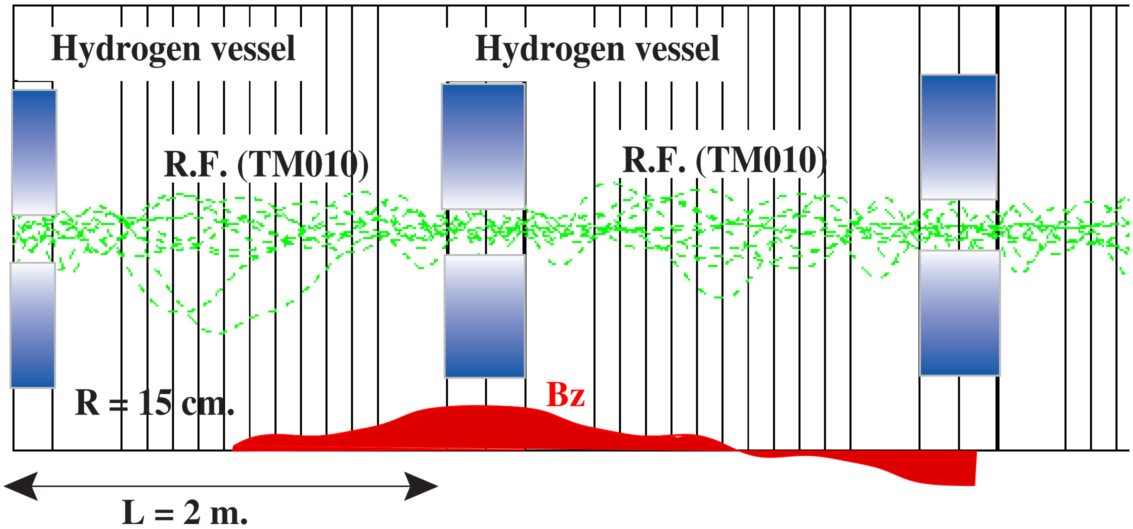


FIG. 24. GEANT simulation of muons traversing a section of the alternating solenoid cooling channel. The variation of the magnetic field B_z is shown for $1\frac{1}{2}$ cells of the figure.

Additional simulations were performed using the program ICOOL [146,151,158]. The only likely significant effects which are not yet included are space charge and wakefields. Analytic calculations for particle bunches in free space indicate that these effects should, for the later stages, be significant but not overwhelming. A full simulation must be done before we are assured that no problems exist. Particles are introduced with transverse and longitudinal emittance (186 MeV/c, 1400 π mm-mrad transverse, and 1100 π mm longitudinal), together with a number of naturally occurring correlations. Firstly, the particles are given the angular momentum appropriate for the starting axial magnetic field. Secondly, particles with large initial radius r_o and/or divergence θ_o have longer pathlengths in a solenoidal field and tend to spread out with time. This can be parameterized by defining an initial transverse amplitude

$$A^2 = \frac{r_o^2}{\beta_\perp^2} + \theta_o^2. \tag{27}$$

The temporal spreading can be minimized by introducing an initial correlation between p_z and A^2 that equalizes the forward velocity of the initial particles. This correlation causes the average momentum of the beam to grow from the reference value of 186 MeV/c to $\approx$ 195 MeV/c. Lastly, a distortion of the longitudinal bunch distribution can be introduced to reflect the asymmetric nature of the "alpha"-shaped rf bucket.

Figure 25(a) shows the average momentum of the beam as a function of distance along the channel. The momentum drops as the beam crosses the liquid hydrogen absorbers. The gradient and phase of the rf cavities have been adjusted so that the reacceleration given to the reference particle equals the mean energy loss. This causes the average momentum of the beam to remain in a narrow band around 195 MeV/c. Figure 25(b) shows the mechanical and canonical angular momenta as a function of distance along the channel. The mechanical angular momentum shows the rotational motion of the beam around the axial solenoidal field. It periodically reverses sign when the solenoids alternate direction. The canonical angular momentum is defined such that it removes the axial field dependence [158]. Without the absorbers, the beam would have a constant (0) value for the canonical angular momentum. However, the presence of absorbers causes the canonical angular momentum to grow and would lead to severe emittance growth by the end of a long channel. This growth is stopped by alternating the direction of the solenoid field, as shown in Fig. 25(b) Simulations have shown that 2 m is a reasonable (half) period for the field, since the net growth in canonical angular momentum is small. In addition synchrobetatron resonances are avoided since the periodicity of the field forces the average betatron wavelength to be 2 m, whereas the synchrotron oscillation wavelength seen in the simulations for this arrangement is $\approx$ 14 m.

Figure 26(a) shows the rms and maximum radius of any particle in the beam distribution as a function of distance along the channel. The rms radius shows that most of the beam is confined to within 2 cm of the axis. The peak rms radius decreases towards the end of the channel as a result of the cooling. The maximum particle radius is about 8 cm, which determines the radius of the windows required in the rf cavities. Figure 26(b) shows the rms momentum

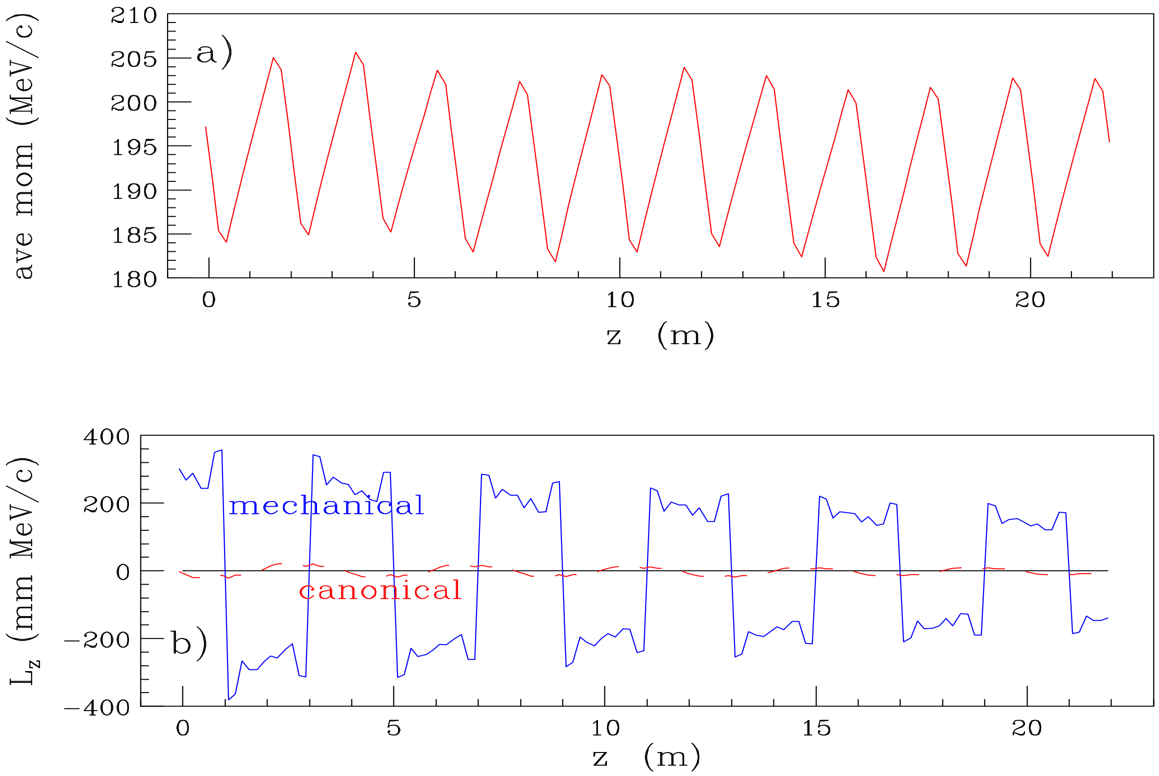


FIG. 25. a) Average momentum *vs.* z; b) Average angular momentum: mechanical (solid curve) and canonical (dashed curve), *vs.* z.

spread corrected for the correlation between p_z and transverse amplitude imposed on the initial particle distribution. The momentum spread grows as a function of distance since the alternating solenoid system only cools the transverse emittance. Figure 26(c) shows the rms bunch length s a function of distance along the channel. Again this grows with distance since this channel does not cool longitudinally.

Figure 27(a) shows the decrease in transverse normalized emittance as a function of distance along the channel. The system provides cooling by a factor of ≈ 2 in both the x and y transverse phase spaces. From the changing slope of the curve we note that the rate of cooling is dropping. This sets ≈ 22 m as the maximum useful length for this type of system. It must be followed by a longitudinal emittance exchange region to reduce the momentum spread and bunch length approximately back to their starting values. Figure 27(b) shows the increase in longitudinal normalized emittance in the channel due to the increase in momentum spread and bunch length. Finally, Fig. 27(c) shows the decrease in the 6-D normalized emittance as a function of distance along the channel. There is a net decrease in 6-D emittance by a factor of ≈ 2 in the channel. Table V gives the initial and final beam parameters.

TABLE V. Initial and final beam parameters in a 15 T transverse cooling section.

		initial	final	final/initial
Particles tracked		1000	980	0.98
Reference momentum	MeV/c	186	186	1.0
Transverse Emittance	π mm-mrad	1400	600	0.43
Longitudinal Emittance	π mm-mrad	1100	2300	2.09
6-D emittance 10^{-12}	$(\pi \text{ m-rad})^3$	2000	800	0.40
rms beam radius in hydrogen	cm	0.8	0.55	0.69
rms beam radius in linac	cm	2.0	1.4	0.70
max beam radius in linac	cm	7.0	7.0	1.0
rms bunch length	cm	1.5	2.2	1.5
max bunch full width	cm	13	19	1.5
rms dp/p	%	3.8	5.6	1.5

This simulation has been confirmed, with minor differences, by double precision GEANT [149] and PARMELA

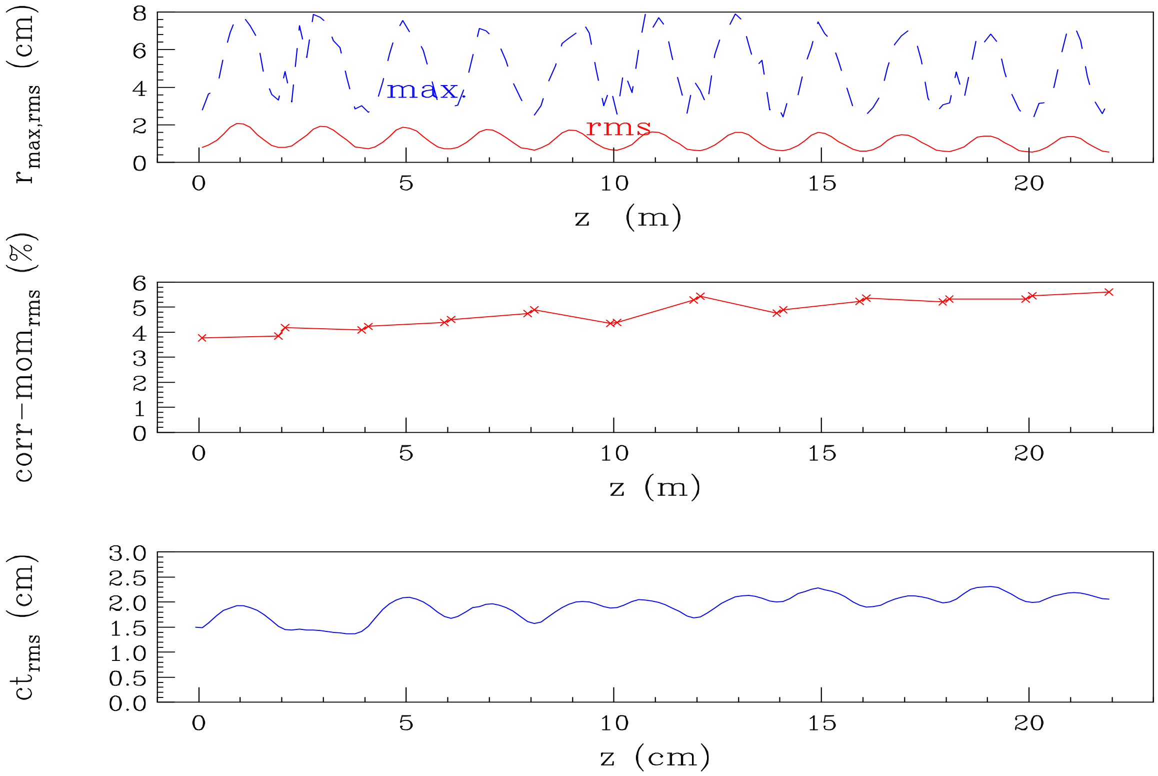


FIG. 26. a) rms and maximum beam radii, b) rms corrected momentum, c) rms bunch length; all *vs.* z.

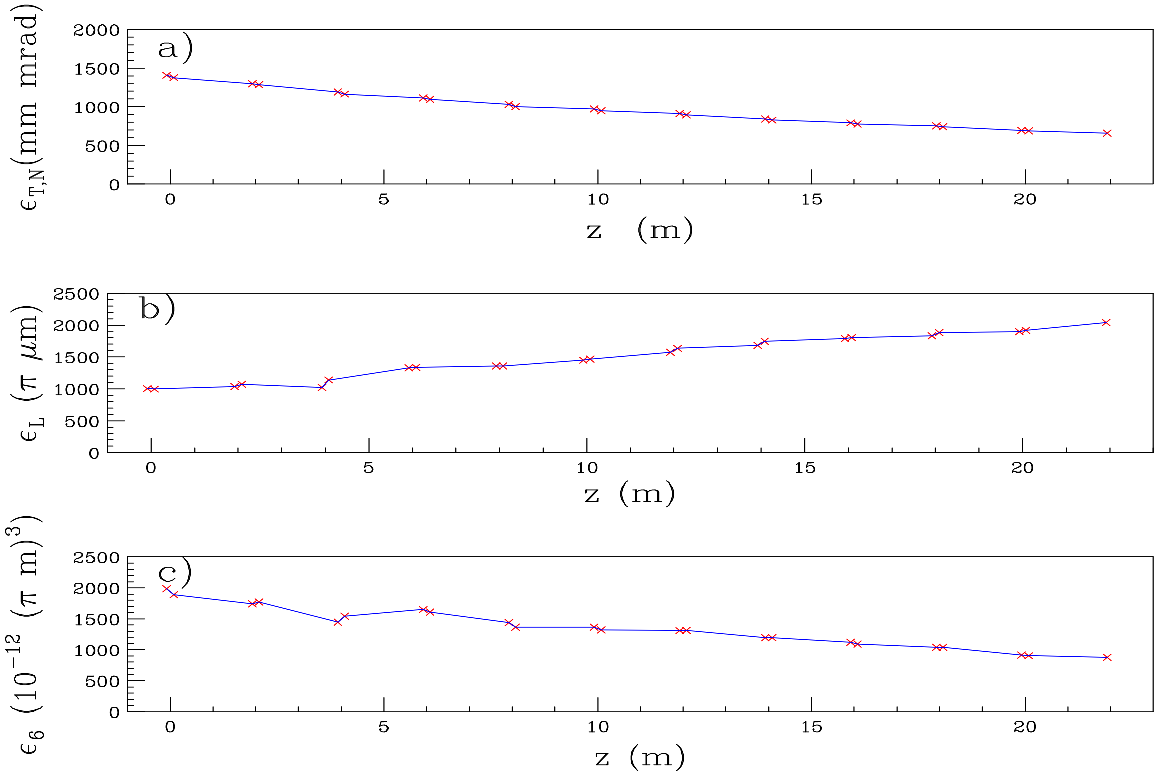


FIG. 27. Emittance *vs.* z: a) transverse emittance; b) longitudinal emittance; and c) 6-D emittance.

[147] codes.

E. 31 T solenoid transverse cooling example

As in the preceding example, the lattice consists of 11 identical 2 m long cells with the direction of the fields in the solenoids alternating from one cell to the next. The maximum solenoidal field is higher (31 T) than in the previous example, but the bore is smaller (8 cm), and the liquid hydrogen absorber has a smaller diameter (6 cm). Between the 31 T solenoids there are 1.3 m long matching sections with an inside diameter of 32 cm, superimposed on a 36 MeV/m reacceleration linac operating at 805 MHz.

Table VI gives the initial and final parameters for the 31 T example, together with the required emittances for a Higgs factory. In setting these requirements a dilution of 20% during acceleration is assumed in each of the three emittances.

TABLE VI. Initial and final beam parameters in a 31 T transverse cooling section.

		initial	final	final/initial	Req.
Particles tracked		4000	3984	0.99	
Reference momentum	MeV/c	186	186		
Transverse Emittance	π mm-mrad	460	240	0.52	240
Longitudinal Emittance	π mm-mrad	850	1600	1.9	
6-D emittance 10^{-12}	$(\pi$ m-rad$)^3$	150	95	0.63	98
rms beam radius in hydrogen	cm	0.44	0.33	0.75	
rms beam radius in linac	cm	.4	1.1	0.80	
max beam radius in linac	cm	6.0	6.0	1.0	
rms bunch length	cm	1.5	1.8	1.2	
max bunch full width	cm	11	19	1.7	
rms dp/p	%	3.5	5.0	1.4	

F. Bent solenoid emittance exchange example

The only practical method of exchanging longitudinal and transverse emittance seems to require dispersion in a large acceptance channel, followed by low-Z wedge absorbers. In this example, the focusing channel consists of a continuous 3.5 T solenoid. The dispersion is generated by bending the solenoid.

In a bent solenoid, in the absence of any dipole field, there is a drift perpendicular to the bend plane of the center of the Larmor circular orbit, which is proportional to the particle's momentum [159]. In our example we have introduced a uniform dipole field over the bend to cancel this drift exactly for particles with the reference momentum. Particles with momenta differing from the reference momentum then spread out spatially, giving the required dispersion (0.4 m). The dispersion is removed, and the momentum spread reduced, by introducing liquid hydrogen wedges [160]. The hydrogen wedges would be contained by thin beryllium or aluminum foils, but these were not included in this simulation.

After one bend and one set of wedges, the beam is asymmetric in cross section. Symmetry is restored by a following bend and wedge system rotated by 90 degrees with respect to the first. Figure 28 shows a representation of the two bends and wedges. The total solenoid length was 8.5 m. The beam tube outside diameter is 20 cm, and the minimum bend radii is 34 cm.

Figure 29(a) shows the magnetic fields (B_z, B_y, and B_x) as a function of the position along the cell. The solenoid bend curvatue is exactly that given by the trajectory of a reference particle (equal in momentum to the average momenta given in Fig. 29(b)) in the given transverse fields. The actual shape of the bend turns out to be very important. Discontinuities in the bend radius can excite perturbations which increase the transverse emittance. We have shown, for example, that the transverse emittance growth in a bent solenoid depends on discontinuities of the bend radius as a function of distance, and its first and second derivatives, the size and tilt of the solenoidal coils, auxiliary fields and the 6-D phase space of the beam. Thus optimization is not straightforward. One solution to this problem is to have long, adiabatic bends. However, this adds undesirable length to the emittance exchange section. We are studying options with coupling sections to tight bends roughly half a Larmor length long, which seems to minimize transverse emittance growth while also minimizing the length of the section. Due to similar problems, the length and longitudinal distribution of the wedge material has also been found to affect emittance growth. For example, the growth can be minimized when the vector sum of the Larmor phases at the absorber elements is small or zero.

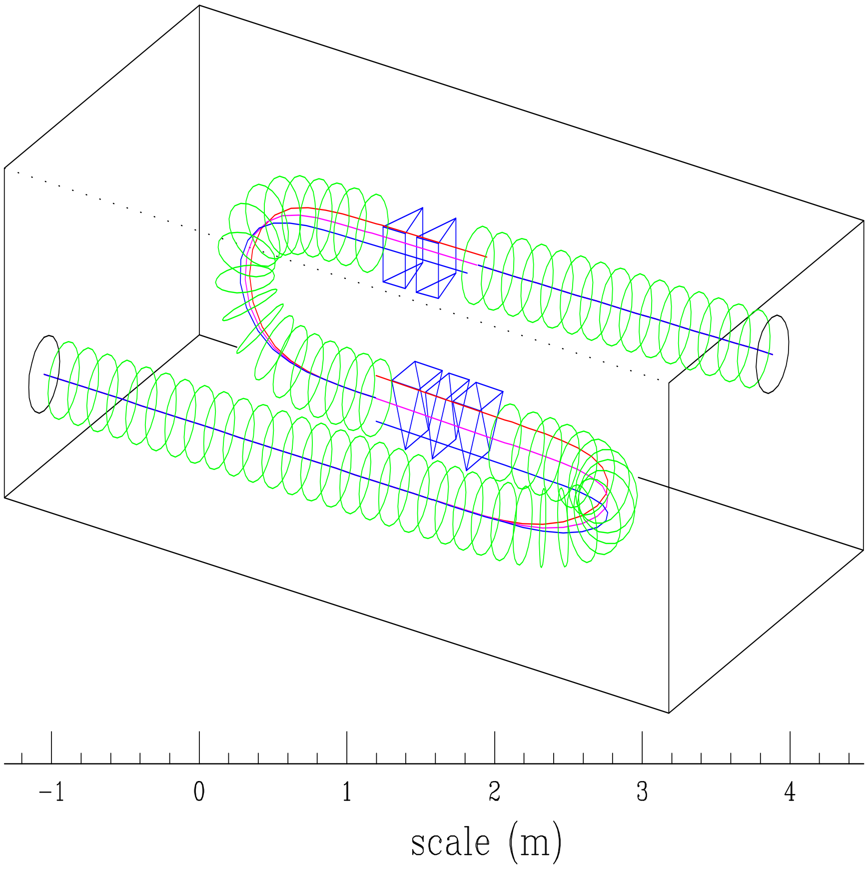


FIG. 28. Representation of a bent solenoid longitudinal emittance exchange section

The simulations were performed using the program ICOOL [146]. The maximum beam radius is 10 cm. Transmission was 100%. Figure 30(a) shows the rms longitudinal momentum spread relative to the reference momentum as a function of the position along the cell. The fractional spread decreases from an initial value of approximately 5%, to a final value of approximately 2.2%. At the same tme, since this is an emittance exchange, the transverse beam area grows, as shown in Fig. 30(b). One notes that the area increases not only in the regions of bends (region 1 and 8), but also in the regions of wedges (2-6 and 9-11). This is probably due to failures in matching that have yet to be understood.

Figure 31 shows scatter plots of the transverse particle positions against their momenta. The dispersion is clearly observed in Fig. 31(b) (after the first bend) and in Fig. 31(e) (after the second). It is seen to be removed, with a corresponding decrease in momentum spread, in Fig. 31(c) (after the first set of wedges) and Fig. 31(f) (after the second set of wedges).

Figure 32 shows a scatterplot of the square of the particle radii vs. their longitudinal momenta, (a) at the start, and (b) at the end of the emittance exchange section . The decrease in momentum spread and rise in beam area are clearly evident.

The initial and final beam parameters are given in table VII. There are significant final beam fluctuations, presumably due to mismatching, and some of the 37% rise in 5-D phase space may be from this cause. The simulations must be extended to include rf so that the 6-D emittance can be studied and the emittance exchange section can be optimized.

Emittance exchange in solid LiH wedges, with ideal dispersion and matching, has also been successfully simulated using SIMUCOOL [161]. Dispersion generation by weak focusing spectrometers [156] and dipoles with solenoids [162] have also been studied.

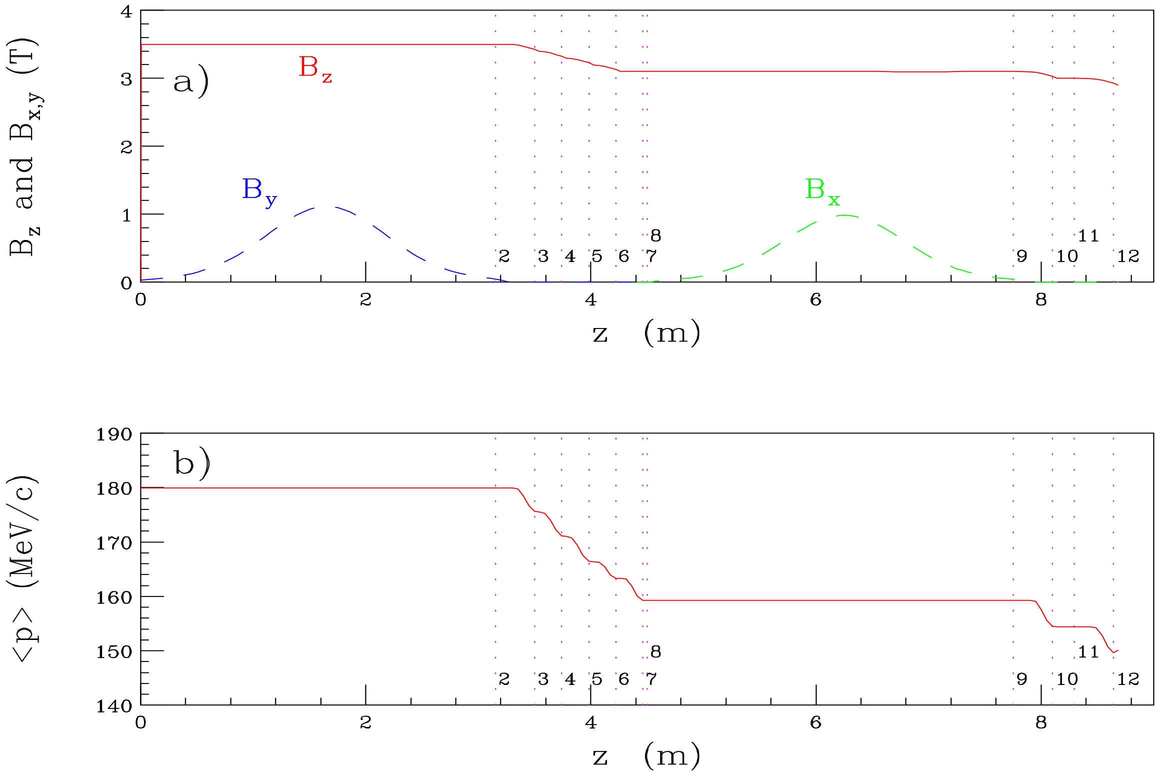


FIG. 29. a) Axial B_z and dipole B_y, B_x, magnetic fields; b) average momentum; both as a function of the position along the cell.

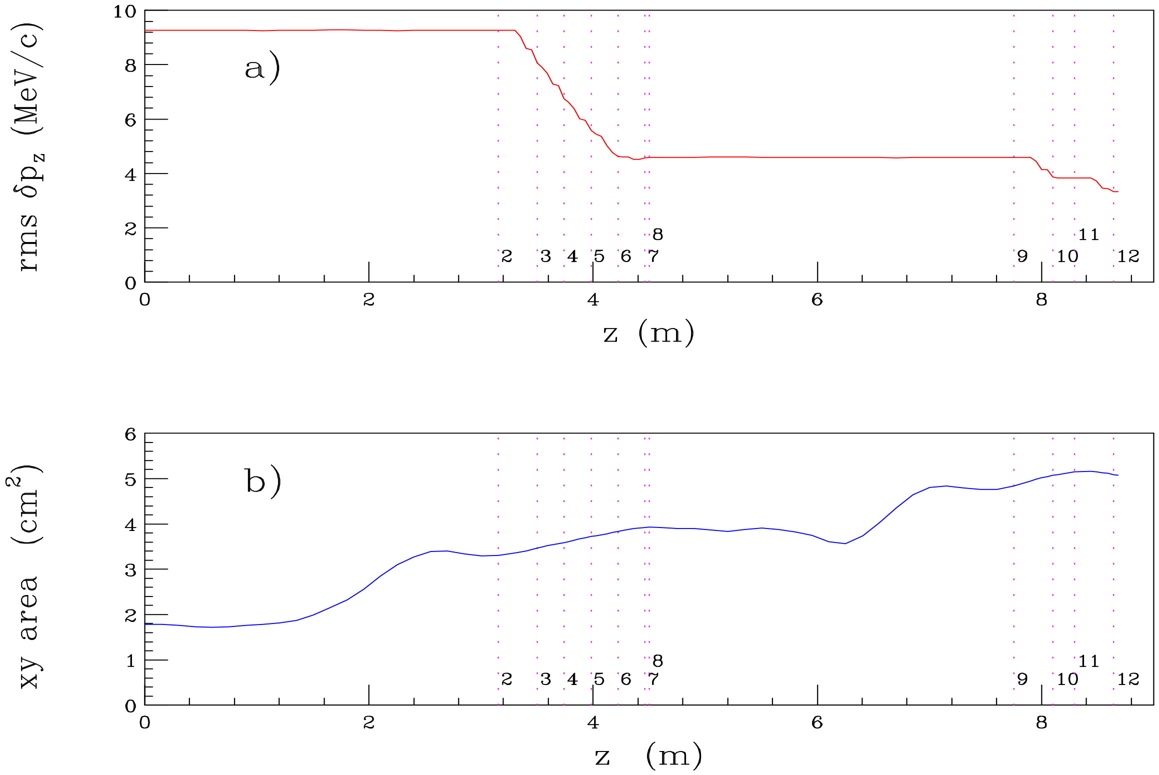


FIG. 30. a) rms longitudinal δp_z with respect to the reference momentum and b) transverse beam area, both as a function of z.

G. RF for the cooling systems

The losses in the longitudinal momentum of the muon beam from the cooling media have to be restored using rf acceleration sections. These rf structures are embedded in solenoidal fields that reverse direction within each section.

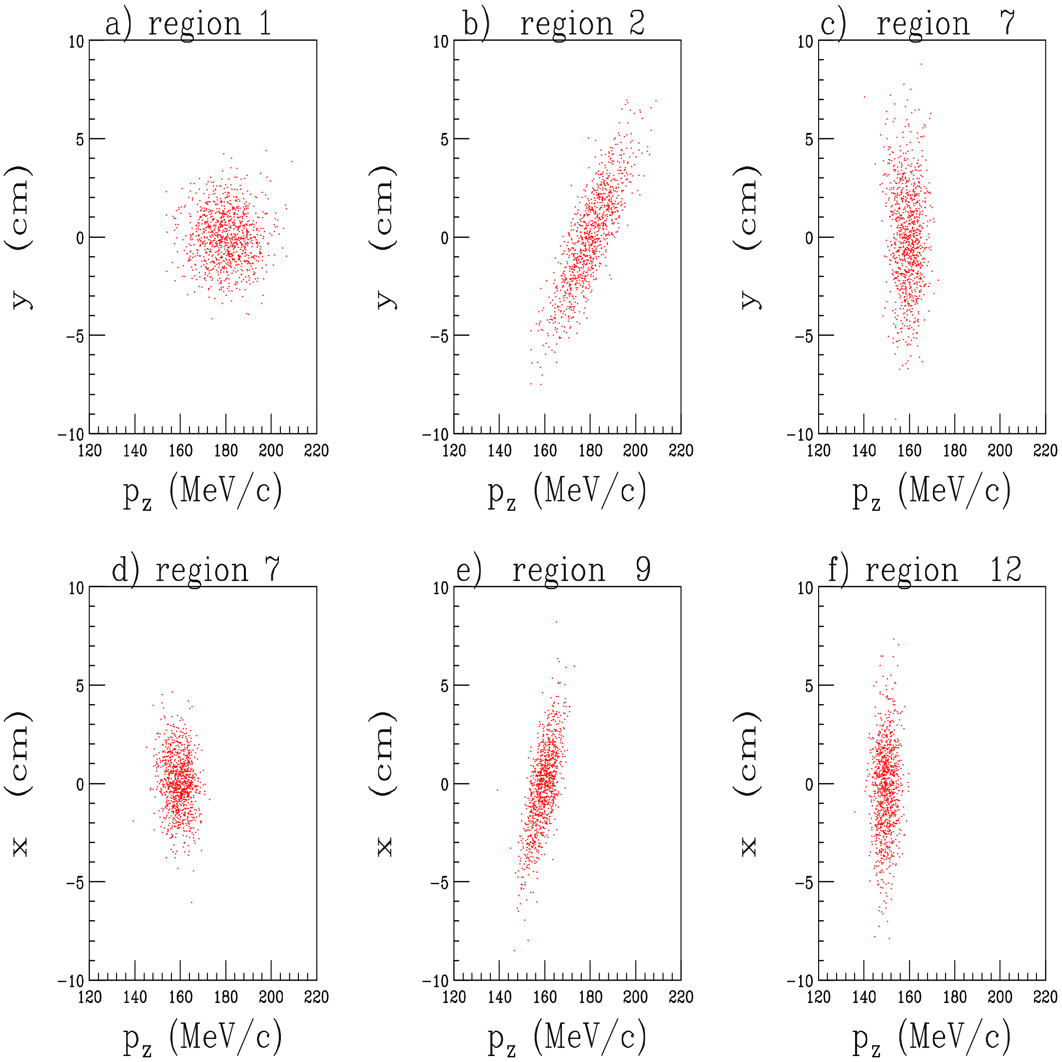


FIG. 31. y *vs.* p_z plots: a) at the start, b) after the first bend, c) after the first set of wedges. x *vs.* p_z plots: d) after the first wedges, e) after the second bend, and f) at the end of the emittance exchange section, following the second set of wedges.

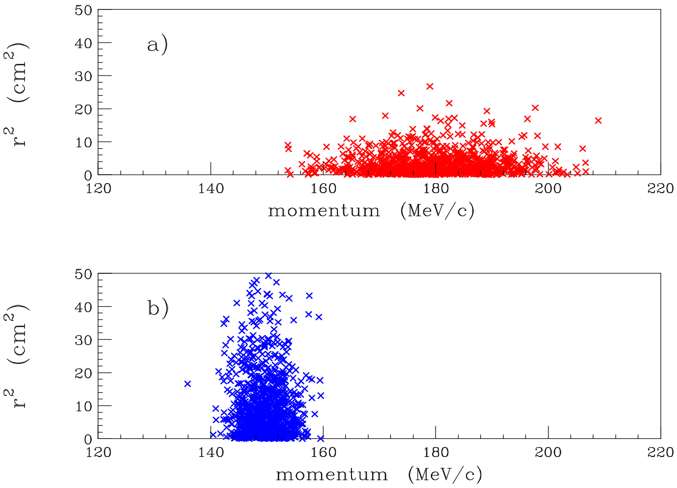


FIG. 32. Scatterplot of squared radii *vs.* longitudinal momentum: a) at the start, and b) at the end of the emittance exchange section.

TABLE VII. Initial and final beam parameters in a longitudinal emittance exchange section.

		initial	final	final/initial
Longitudinal Momentum spread	MeV/c	9.26	3.35	0.36
Ave. Momentum	MeV/c	180	150	0.83
Transverse size	cm	1.33	2.26	1.70
Transverse Momentum spread	MeV/c	6.84	7.84	1.15
Transverse Emittance	π mm-mrad	870	1694	1.95
$\mathrm{Emit}^2_{\mathrm{trans}} \times \Delta p_{\mathrm{long}}$	$(\pi$ m-mrad$)^2$ MeV/c	7.0	9.6	1.37

In the two transverse cooling examples above, the rf frequency is 805 MHz and the peak gradient is 36 MeV/m. The magnetic fields that extend over the cavities vary from 0 to 10 T, reversing in the center. It should be pointed out that in the earlier stages, the bunches are longer, and lower frequencies will be required.

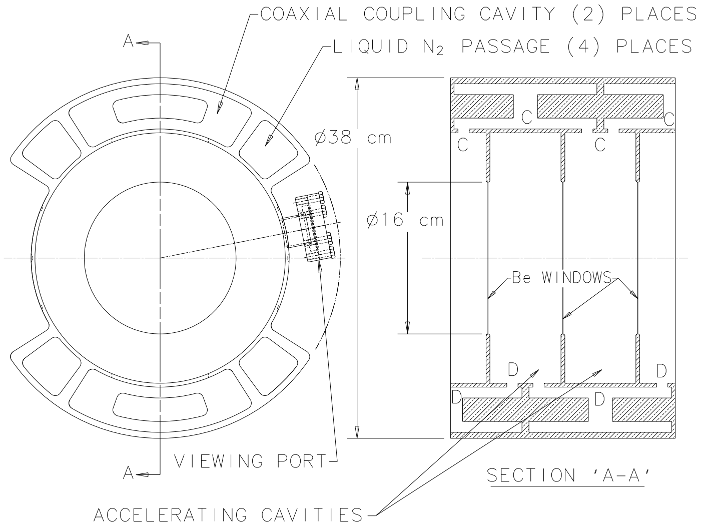


FIG. 33. Two full cell sections plus two half cell sections of the interleaved $\pi/2$ mode accelerating cavities. The volumes labeled **C** are powered separately from the volumes labeled **D**.

In order to realize maximum accelerating gradients within the acceleration cavities, we take advantage of the penetrating properties of a muon beam by placing thin windows between each rf cell, thereby creating an accelerating structure closely approximating the classic pill-box cavity. This permits operating conditions in which the axial accelerating field is equal to the maximum wall field and gives a high shunt impedance.

The windows in the 15 T example are 16 cm diameter, 125 μm thick Be foils. In the 31 T case, they are 10 cm diameter and 50 μm thick.

For these rf structures, we will use an interleaved cavity design in which two parts are independently powered (Fig. 33). The mode of the system will be referred to as $\pi/2$ interleaved. Each section supports a standing wave π mode, with each acceleration cell $\pi/2$ long, giving a good transit time factor. To reduce the peak rf power requirements (by a factor of 2), we are considering operating the cells at liquid nitrogen temperatures.

The characteristics of the rf systems currently being studied are summarized in Table VIII. Figure 34 shows a full 1.3 m section with interleaved cavities. Each cell is 8.1 cm in length and the 1.3 m section consists of 16 cells.

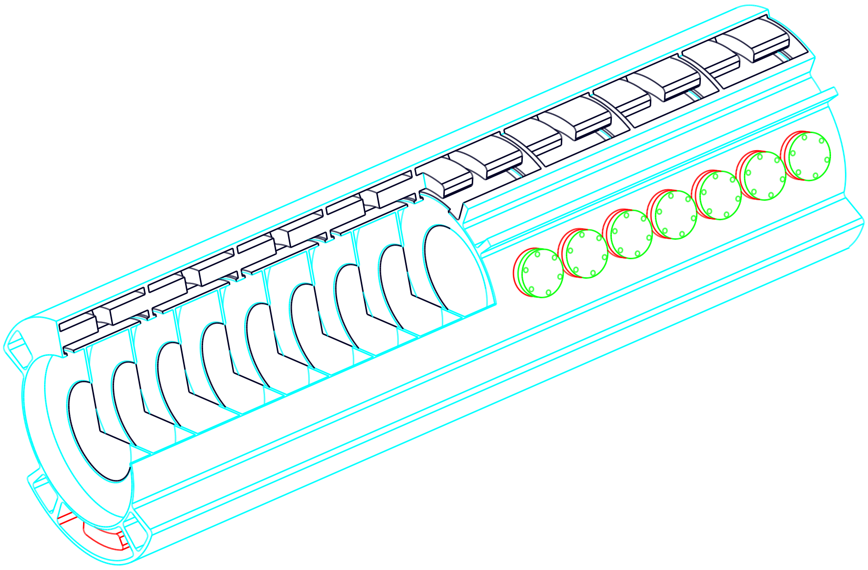


FIG. 34. A 1.3 m acceleration section with quarter section cut away for viewing the inter-cavity windows.

TABLE VIII. Characteristics of the rf system.

RF frequency [MHz]	805
Cavity Length [cm]	8.1
Cavity Inner Radius [cm]	14.6
Cavity Outer Radius [cm]	21
$Q/1000$	2×20
Peak Axial Gradient [MV/m]	36
Shunt Impedance [MΩ/m]	2×44
Zt^2 [MΩ/m]	2×36
Fill Time (3 τ) [μs]	2×12
RF Peak Power [MW/m]	$\frac{1}{2} \times 29$
Ave. Power (15Hz) [KW/m]	5.3
Be window aperture [cm]	16 (10 for 31 T case)
Be window thickness [μm]	127 (50 for 31 T case)

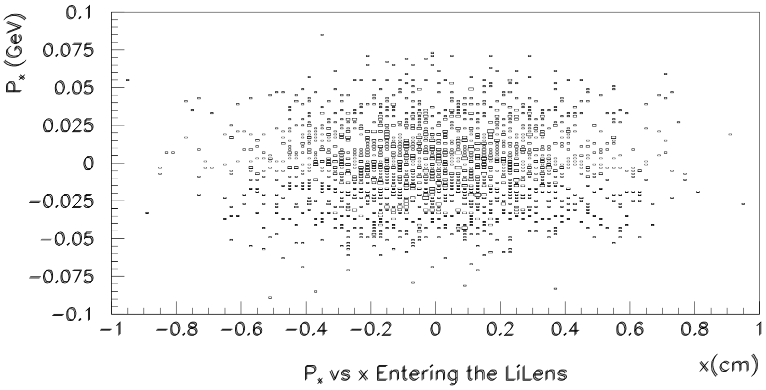
H. The liquid lithium lens

The final cooling element ultimately determines the luminosity of the collider. In order to obtain smaller transverse emittance as the muon beam travels down the cooling channel, the focusing strength must increase, i.e. the $\beta_\perp$'s must decrease. A current within a conductor produces an active lens absorber, which can maintain the beam at small $\beta_\perp$ throughout an extended absorber length, while simultaneously attenuating the beam momentum. An active lens absorber, such as a lithium lens, may prove to be the most efficient cooling element for the final stages.

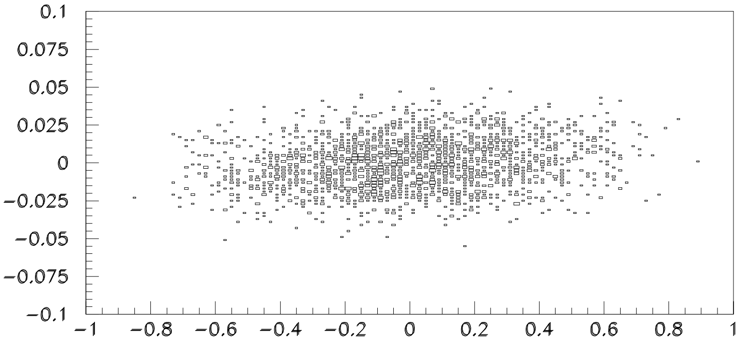
The cooling power of a Li lens is illustrated in Figure 35, where the x *vs.* p_x phase space distributions at the beginning and at the end of the absorber are shown. This example corresponds to a 1 m long lens, with 1 cm radius, and a surface field of 10 T. The beam momentum entering the lens was 267 MeV/c, with Gaussian transverse spatial and momentum distributions: $\sigma_x = \sigma_y = 2.89$ mm, $\sigma_{p_x} = \sigma_{p_y} = 26.7$ MeV/c, and a normalized emittance of $\epsilon_{x,N} = 710$ mm-mrad. The normalized emittance at the end of the absorber was $\epsilon_{x,N} = 450$ mm-mrad (cooling factor ~ 1.57), and the final beam momentum was 159 MeV/c. The results were obtained usig a detailed GEANT simulation of a single stage.

An alternative cooling scheme under study uses a series of Li lenses. The lens parameters would have to vary to match the changing beam emittance along the section and in addition, acceleration of the beam between the lenses has to be included.

Lithium lenses have been used with high reliability as focusing elements at FNAL and CERN [70-74]. Although these lenses have many similar properties to those required for ionization cooling, there are some very crucial differences which will require significant advances in lens technology: ionization cooling requires longer lenses (~ 1 meter), higher



P_x vs x Entering the LiLens



P_x vs x Exiting the LiLe

I. Ionization cooling experimental R&D

An R&D program has been proposed to design and prototype the critical sections of a muon ioization cooling channel. The goal of this experimental R&D program is to develop the muon ionization cooling hardware to the point where a complete ionization cooling channel can be confidently designed for the First Muon Collider. Details can be found in Fermilab proposal P904 [163]. A summary of the R&D program can be found in ref. [164].

The proposed R&D program consists of:

- Developing an appropriate rf re-acceleration structure. It is proposed to construct a 3-cell prototype rf cavity with thin beryllium windows, which will be tested at high power and within a high-field solenoid.
- Prototyping initially a 2 m section, and eventually a 10 m section, of an alternating solenoid transverse cooling stage. It is proposed to test the performance of these sections in a muon beam of the appropriate momentum.
- Prototyping an emittance exchange (wedge) section and measuring its performance in a muon beam of the appropriate momentum.
- Prototyping and bench testing $\sim$ 1 m long liquid lithium lenses, and developing lenses with the highest achievable surface fields, and hence the maximum radial focusing.
- Prototyping a lithium lens–rf–lens system and measuring its performance in a muon beam of the appropriate momentum.
- Developing, prototyping, and testing a hybrid lithium lens/wedge cooling system.

The measurements that are needed to demonstrate the cooling capability and optimize the design of the alternating solenoid, wedge, and lithium lens cooling stages will require the construction and operation of an ionization cooling test facility. This facility will need

1. a muon beam with a central momentum that can be chosen in the range 100-300 MeV/c,
2. an experimental area that can accommodate a cooling and instrumentation setup of initially $\sim$ 30 m in length, and eventually up to $\sim$ 50 m in length, and
3. instrumentation to precisely measure the positions of the incoming and outgoing particles in 6-D phase space and confirm that they are muons.

In the initial design shown in Fig. 36, the instrumentation consists of identical measuring systems before and after the cooling apparatus [165]. Each measuring system consists of (a) an upstream time measuring device to determine the arrival time of the particles to one quarter of an rf cycle ($\sim \pm 300$ ps), (b) an upstream momentum spectrometer in which the track trajectories are measured by low pressure TPC's on either side of a bent solenoid, (c) an accelerating rf cavity to change the particles momentum by an amount that depends on its arrival time, (d) a downstream momentum spectrometer, which is identical to the upstream spectrometer, and together with the rf cavity and the upstream spectrometer forms a precise time measurement system with a precision of a few ps. The measuring systems are 8 m long, and are contained within a high-field solenoidal channel to keep the beam particles within the acceptance of the cooling apparatus.

It is proposed to accomplish this ionization cooling R&D program in a period of about 6 years. At the end of this period we believe that it will be possible to assess the feasibility and cost of constructing an ionization cooling channel for the First Muon Collider, and if it proves feasible, begin a detailed design of the complete cooling channel.

VI. ACCELERATION

A. Introduction

Following cooling and initial bunch compression, the beams must be rapidly accelerated. In this section some of the options in accelerator design will be described and examples of acceleration scenarios presented.

Separate acceleration scenarios are given here for a low momentum spread 100 GeV First Muon Collider (Higgs factory), and for a high luminosity 3 TeV collider. Ideally, though more difficult, this accelerator designed for the low energy machine should be extendable to the $\approx$ 250 GeV beam energy and from there to the $\approx$ 2 TeV beam energy needed for a very high-energy collider.

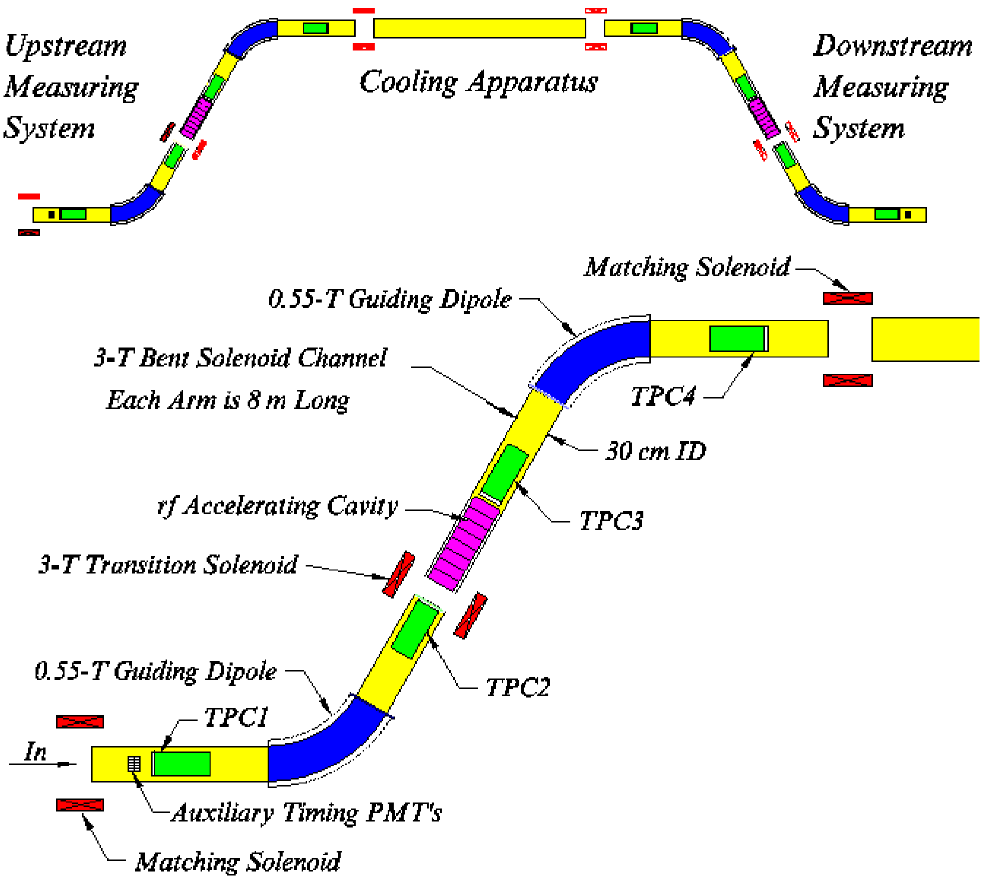


FIG. 36. Schematic of the cooling test apparatus arrangement.

While acceleration of muons to high energy is clearly possible, an optimal and cost-effective acceleration complex is needed. In the scenarios described below, a low-frequency linac would take the beam from the end of cooling to an energy of ≈1 GeV followed by recirculating-linac systems to take the beam to 50-70 GeV. The multi-TeV energy regime can be reached through a series of very rapid cycling synchrotrons. Variations on the acceleration model and potential difficulties are discussed, including the use of Fixed Field Alternating Gradient (FFAG) accelerators in place of, or together with, the recirculating-linacs. Finally, topics for further study and research are described.

B. Accelerator options

The acceleration time is limited by muon decay ($\tau_\mu = 2.2\mu\ s$ at rest) and requires that:

$$eV'_{rf} \ >> \ 0.16\mathrm{MeV/m}$$

where eV'_{rf} is the acceleration rate. An acceleration rate whose value of 0.16 MeV/m is low for a linac, but very high for a conventional synchrotron.

At the lowest energies ($<$ 700 MeV), the momentum spread and beam sizes are so large that only a linac is feasible, and acceleration to full energy in a single-pass linac would be good, but it would be very expensive.

Thus, following the initial linac, some form of recirculating acceleration is preferred. A synchrotron would be possible, in principle, but the acceleration must occur so rapidly that conventional magnet ramping is unlikely to be practical. Two alternative multi pass methods are being considered: recirculating linac accelerators similar to those used at TJNAF and Fixed Field Alternating Gradient (FFAG) accelerators.

In a recirculating linac accelerator, the beam is circulated through the same linac for several passes, with separate, energy matched, fixed-field return paths for each pass. Each return path is optically independent and can be separately designed. In the initial lattice design for the muon recirculating linac accelerator, the return arcs are similar alternating gradient (AG) systems with the same dipole layout, but with differing quadrupole strengths to allow separate tuning and chronicity matching in each arc. Multiple aperture superconducting magnets have also been designed which would reduce the diameter of the recirculating linac, lowering muon loss from decay and possibly being more economical (see Fig 7.12 [43]). In either case, both the linac and return transports must accommodate large transverse emittances (rms) of $\approx 300\,\pi$ mm-mrad. Strong focusing is required not only to keep apertures down, but also to minimize orbit deviations due to the large momentum spreads, which in the initial stages of acceleration, can be as large as 10% rms.

More recently, an adaptation of the FFAG (Fixed-Field Alternating Gradient) accelerator concept has been proposed for $\mu+$ $\mu-$ acceleration [**?**,166]. In this variation, return transports are designed with a very large (factor of 5-10) energy acceptance, so that separate energy turns can pass through the same fixed-field elements. More acceleration turns are possible than with a recirculating linac accelerator which reduces the rf requirements; but the orbits and focusing properties are now energy-dependent. Such FFAG configurations require strong superconducting magnets with large apertures to accommodate the energy-dependent spread in closed orbits. The extra cost associated with increased magnet apertures must be evaluated against potential savings in the number of magnets and reduced rf/turn requirements.

For the higher energy stages, the muon life time is greater and the needed rate of energy increase is less. Thus, above a few hundred GeV, rapid accelerating synchrotrons become possible. In a rapid accelerating synchrotron, the beam is also multi-pass accelerated through an rf system, but the beam returns in a single arc, as the magnetic field is ramped to match the increase in beam energy. As above, more acceleration turns are possible than with a recirculating accelerator, but we now have a single moderate aperture return transport. But rapid cycling synchrotrons have higher power costs and the technical challenges associated with the rapid acceleration time needed to conserve muons.

Thus a complete system would likely include an initial linac followed by a sequence of recirculating linac accelerators and/or Fixed Field Alternating Gradient machines. Depending on the final collider energy, one or more rapid acceleration synchrotrons would follow. Each system increases the beam energy by a factor of 5-10.

C. Scenario examples

Several scenarios have been discussed earlier [167,168], see for instance the parameters (table XI) used in a simulation of longitudinal motion discussed below. The ones given here are more recent, and more detailed, but they should not be taken to be definitive. They are examples that were derived to probe the design problems and to show that solutions should be possible.

1. Acceleration for Higgs collider

Table IX gives an example of a sequence of accelerators for a 100 GeV Higgs Factory, *i.e.* a machine with very low momentum spread (0.003%, see table I) and relatively large rms transverse emittance ($\approx$ 300 π mm-mrad).

Following initial linacs, recirculating accelerators are used. The number of arcs in each recirculating accelerator is about 10. In this example, conventional fixed field 2 T magnets are used, but the effective ramp frequencies that would be needed if pulsed magnets were used are given for reference.

In this example, all the accelerating cavities are room temperature copper structures, and the accelerating gradients are modest ($<$ 10 MeV/m). Nevertheless, the acceleration is rapid enough that the total losses from decay are only 30%. The heating from these decays is also modest ($\approx$ 10 W/m) because of the small number of turns and relatively low energy. Since no superconducting magnets or rf are used in this example, this heating should cause no problem.

In this machine, the transverse emittances are large and strong focusing is thus required, but the maximum momentum spread is moderate (up to 1.37 rms in the first recirculator) and is thus not likely to be a problem.

If the same machine is to also run at a high luminosity, with larger momentum spread, then although the six dimensional emittance is the same as in the Higgs collider discussed above, the transverse emittance is smaller ($\approx 90\pi$ mm-mrad rms), and the longitudinal emittance larger (by about a factor of 4) and the momentum spread in the first recirculating accelerator would be nearly 6% rms; or about 50% full width. This is a very large momentum spread that could only be accepted in FFAG like lattices as discussed below.

Similarly, if the same acceleration is to be usable as the front end for a 250 + 250 GeV or higher energy machine then the transverse emittance will again be less, the longitudinal emittance even larger, and the problem of very large momentum spread will be worse. Clearly, although not absolutely needed for a Higgs Factory, it is desirable to solve this problem even in that First Collider.

A separate parameter set for the high luminosity 50 + 50 GeV collider and a 250 + 250 GeV collider could have been presented, but their parameters are very similar to those of the front end of the 3 TeV machine given below, and are thus omitted here.

TABLE IX. Accelerator parameters for a Higgs Factory (100 GeV)

Acc. type		linac	linac	recirc	recirc	recirc	sums
Magnet type				warm	warm	warm	
rf type		Cu	Cu	Cu	Cu	Cu	
E^{init}	(GeV)	0.10	0.20	0.70	2	7	
E^{final}	(GeV)	0.20	0.70	2	7	50	
Circ.	(km)	0.04	0.07	0.07	0.19	1.74	2.11
Turns		1	1	8	10	11	
Loss	(%)	2.31	3.98	7.27	7.91	13.94	31.06
Decay heat	W/m	0.89	1.98	11.04	12.99	12.44	
B_{fixed}	(T)			2	2	2	
Ramp freq.	(kHz)			281	79.83	8.00	
Disp	m			1	1.50	3	
βmax	m	0.83	1.42	3.00	5.31	21.08	
σ_z^{init}	(cm)	2.71	2.22	1.42	1.64	0.90	
$\Delta p/p^{\text{init}}$	(%)	3.58	2.80	1.64	0.56	0.32	
σ_y	cm	1.09	1.14	1.01	0.85	0.94	
σ_x	cm			1.93	1.19	1.34	
Pipe full height	cm	10.92	11.42	10.14	8.49	9.39	
Pipe full width	cm	10.92	11.42	19.28	11.94	13.35	
rf freq.	(MHz)	200	200	200	200	400	
Acc/turn	(GeV)	0.20	0.40	0.17	0.50	4	
Acc time	(μs)			1	6	62	
η	(%)	5.15	5.36	6.36	2.84	6.92	
Acc. Grad.	(MV/m)	8	8	8	10	10	
Synch. rot's		0.62	0.63	0.62	3.92	23.16	
Cavity rad.	(cm)	54.37	54.88	54.88	60.47	38.26	
Beam time	msec	0.00	0.00	0.00	0.01	0.06	
rf time	msec	0.17	0.17	0.17	0.18	0.13	
Tot. peak rf	(GW)	0.05	0.10	0.05	0.26	4.71	5.17
Ave. rf power	MW	0.14	0.24	0.13	0.68	9.54	10.73
rf wall	MW	0.64	1.16	0.46	2.42	28.06	32.75

2. Acceleration for 3 TeV collider

For a high energy machine, the muon accelerators are physically the largest component and are also probably the most expensive. More work is needed on its design. Table X gives an early example of a possible sequence of accelerators for a 3 TeV collider.

Linacs are used up to 700 MeV, followed by recirculating linac accelerators. In the first of these, because of the very large longitudinal emittance, the momentum spread as the beam enters the first recirculating linac accelerator is 8.5% rms, which is very large. The lattice must have very strong focusing, small dispersion and large aperture. If this is not possible, higher energy linacs or lower frequency rf could relieve the requirement.

For the final three stages, pulsed magnet synchrotrons [169] are used. In the 200 GeV ring, all the magnets in the ring are pulsed, but in the last two rings a superconducting-pulsed hybrid solution is used. In these cases, if only pulsed magnets were used, then the power consumed would be too high, and because only low pulsed fields could be used, the circumferences would also be very large. It is thus proposed to use rings with alternating warm pulsed magnets and superconducting fixed magnets [170] (see figure 37). The fixed magnets are superconducting at 8 T; the pulsed magnets are warm with fields that swing from - 2 T to + 2 T. The effective ramp frequency is given in the table. Both of these rings are in the same tunnel, with the fraction of magnet length pulsed (vs. fixed) being different (73% and 43%).

In all the final three rings superconducting rf is employed to minimize the peak power requirements and to obtain high wall to beam efficiency and thus keep the wall power consumption reasonable. In the final two rings the frequency and cavity designs are chosen to be the same as that in the TESLA [171] proposal.

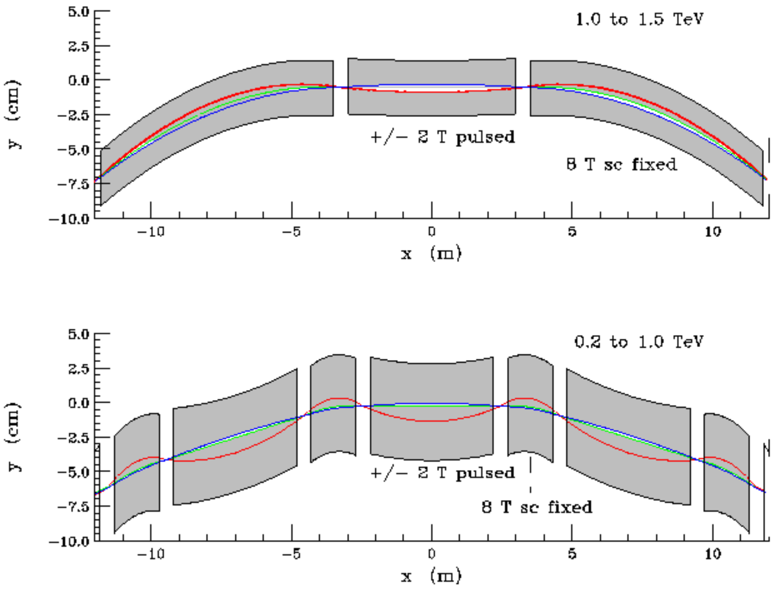


FIG. 37. Schematic of hybrid superconducting-pulse magnet accelerator ring

D. Design issues

1. Recirculating linac accelerator lattice issues

Beam transport R&D for recirculating linac accelerators follows the model of the Thomas Jefferson National Accelerator Facility (TJNAF). The layout is a racetrack with linacs in the straight sections and multi-pass return arcs. At the ends of the linacs the multi-pass beam lines are recombined. A pulse magnet at each separation/recombination point is used to guide the beam into the energy matched return arc. Some initial lattice design concepts for recirculating linac accelerators are being developed. The basic return arc unit would be a FODO lattice, but with the quadrupole strengths varied in order to perturb the arc dispersion function and obtain nearly isochronous motion around the arcs. The arcs are dispersion matched by setting the arc phase advance to a multiple of 2π. Arc designs based upon the flexible momentum compaction (FMC) module can also be used.

In the special case of the very low momentum spread Higgs factory, the transverse emittances are very large ($\approx 300\ \pi$ mm-mrad rms), and will require strong focusing in the lattices. Momentum acceptance in the rings is, in this case, not a problem. But the longitudinal phase space of the muons in the other machines is much larger and requires, at low energies, either long bunches, or large momentum spreads. The requirement of high accelerating gradients argues for high frequencies, and thus short bunches. One therefore needs accelerators with large momentum acceptances.

In the 3 TeV example above, the acceptance at injection into the first recirculating accelerator is 8.5% rms. This is very large by conventional standards, but far less than that in the FFAG lattices being studied [?]. Thus the early return arcs of such a recirculating linac accelerator would have to have very strong focusing, and be FFAG-like. Of course, if a true FFAG accelerator were to be used for its avoidance of the switchyards and multi-aperture magnets, then the specified momentum spread would certainly not be a problem.

Permanent, ferric or superferric ($\approx$2 T), or high field superconducting magnets could be used for recirculating linac accelerators. The lower field magnets may be economic for initial turns, while high field magnets minimize particle travel times, and therefore decay losses. Designs for multi-aperture superconducting magnets suitable for recirculating linac accelerators have been developed [43], and superconducting magnets with as many as 18 apertures with 0.7–7 T fields have been designed. A variety of magnet configurations can be developed; cost/performance optimization will be needed in developing a final choice.

TABLE X. Parameters of Acceleration for 3 TeV Collider

Acc. type		linac	recirc	recirc	recirc	synch	synch	synch	sums
Magnet type			warm	warm	warm	warm	hybrid	hybrid	
rf type		Cu	Cu	Cu	SC Nb	SC Nb	SC Nb	SC Nb	
E^{init}	(GeV)	0.10	0.70	2	7	50	200	1000	
E^{final}	(GeV)	0.70	2	7	50	200	1000	1500	
Circ.	(km)	0.07	0.12	0.26	1.74	4.65	11.30	11.36	29.52
Turns		2	8	10	11	15	27	17	
Loss	(%)	6.11	12.28	10.84	13.94	10.68	10.07	2.65	50.58
Decay heat	W/m	3.67	15.02	16.89	15.91	19.44	30.97	18.09	
B_{pulse}	(T)					2	2	2	
B_{fixed}	(T)		0.70	1.20	2		8	8	
frac pulsed	%						73	43	
Ramp freq.	(kHz)		162	57.34	8.00	2.15	0.50	0.79	
Disp.	m		0.40	0.60	0.80	1	2	4	
$\beta\mathrm{max}$	m	0.89	3.97	8.75	36.29	52.20	108	120	
Mom. compactn	%		1	-0.25	-0.50	-0.50	-0.50	-1	
σ_z^{init}	(cm)	16.34	8.53	5.29	3.57	1.59	0.96	0.78	
$\Delta p/p^{\mathrm{init}}$	(%)	19.27	8.49	5.41	2.47	0.82	0.35	0.09	
σ_y	cm	0.45	0.45	0.42	0.48	0.22	0.16	0.08	
σ_x	cm		3.40	3.25	1.98	0.82	0.71	0.36	
Pipe full height	cm	4.46	4.52	4.22	4.77	2.20	1.62	0.78	
Pipe full width	cm	4.46	33.95	32.49	19.79	8.20	7.06	3.62	
rf Freq	(MHz)	200	100	200	200	800	1300	1300	
Acc./turn	(GeV)	0.40	0.17	0.50	4	10	30	30	
Acc. time	(μs)		3	8	62	232	1004	631	
η	(%)	3.82	0.96	1.97	1.11	10.15	14.37	12.92	
Acc. Grad.	(MV/m)	8	8	10	10	15	25	25	
Synch. rot's		0.81	0.76	1.02	5.82	19.14	54.29	31.30	
Cavity rad.	(cm)	54.88	110	60.47	76.52	19.13	11.77	11.77	
rf time	msec	0.04	0.12	0.05	0.56	0.40	1.25	0.96	
Tot. peak rf	(GW)	0.21	0.14	0.59	1.31	1.06	1.16	1.04	5.51
Ave. rf power	MW	0.14	0.25	0.45	11.04	6.32	21.91	15.07	55.18
rf wall	MW	0.64	0.88	1.62	32.47	18.59	44.72	30.76	130

2. RF peak power requirements

Because of the need for rapid acceleration, the peak rf powers are high, and the resulting numbers of power sources large. For the linacs and early recirculating accelerators, the powers are high because of the high gradients, and low frequencies needed to accelerate the long bunches. At these frequencies (≈200 MHz), currently available sources (triodes and tetrodes) have relatively low maximum output powers and are expensive. Low temperature operation of the cavities, and superconducting or conventional SLED [172] systems, which would reduce the peak power requirements, are being considered.

Study of the example suggests that in the first two recirculators (up to 7 GeV) there is no hope for the rf to keep up with the beam loading. The cavity can only be filled in a suitable filling time (twice the time constant in these examples), and the rf voltage allowed to sag as the beam makes its multiple passes. If excessive sensitivity to beam current is to be avoided, then the stored energy must be large compared to that used, which is somewhat inefficient.

In the final recirculating accelerator, continuous filling (cw) is just possible, but requires yet higher peak power (≈ 5 GW total at 400 MHz) because of the high acceleration rate. The use of superconducting cavities can reduce losses, and thus reduce this peak rf power source requirement, and was included in the above 3 TeV example. At this frequency (400 MHz), klystrons are available with greater power (≈ 20 MW) than that of the sources at the lower frequencies, but a yet higher power klystron (50-100 MW) could probably be developed and would be desirable.

3. Pulsed magnet systems

A pulsed current 4 T magnet has been designed for acceleration to 250 GeV in 360 μs [173], but efficiency favors use of ferric materials in rapid acceleration magnets, although this would limit peak magnetic fields to ≈2 T. The average field can be increased by interleaving magnets swinging from -2 T to 2 T with fixed field 8 T superconducting magnets.

Faster pulsing magnets would require special materials to minimize energy losses from eddy currents. Options include silicon steel, metglass laminations, or finemet laminated tape or powdered solid. A 30μm metglass lamination suitable for several kHz cycling has been developed.A design of suitable pulsed magnets [170](see Fig. 38) has been shown to have sufficiently low losses for this application. The magnets employ cables made of many fine insulated strands (litz cable) and the yokes are made of very thin (0.28 mm), 3% Si-Fe laminations, possibly of Metglas [174,175] for the higher rate cases. Detailed designs must be developed and prototypes constructed and the practical limits of recycling scenarios should be determined.

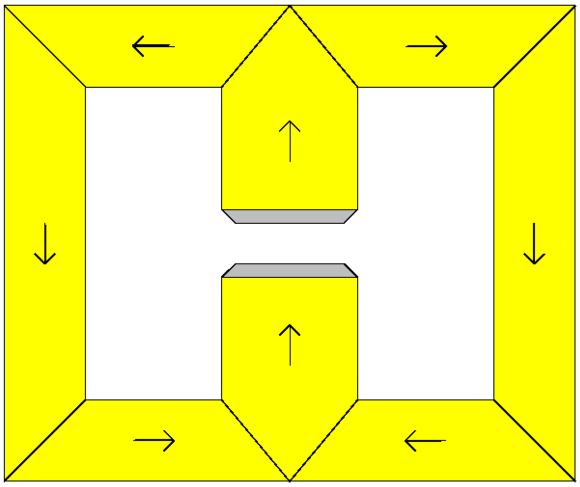


FIG. 38. A 2-d picture of an H frame magnet lamination with grain oriented 3% Si-Fe steel.The arrows show both the magnetic field and the grain direction.

4. Superconducting linacs

While the gradients needed in the acceleration systems are not excessive, they are larger than previous experience at the lower frequencies. The high peak power pulsed operation poses power handling difficulties at lower energies and high peak-current presents collective effect (wakefield) difficulties at higher energies. Higher gradients and efficiencies in all sections would improve performance.

The superconducting rf would operate in pulsed mode, matched to the acceleration time of up to a few ms. This pulse structure is similar to the multibunch acceleration mode planned for TESLA (25 MV/m at 1300 MHz designs), and studies indicate that this design could be adapted to $\mu^+ - \mu^-$ acceleration. At lower frequencies, structures such as the CERN 350 MHz superconducting rf cavities could be used. These cavities have been tested in pulsed mode operation, and tests indicate that pulsed acceleration fields > 10 MV/m are possible [176].

The high single bunch intensities required for high intensities imply large higher order mode losses and large wakefield effects from the short, high-intensity bunches. Higher order mode (HOM) load designs adapted from superconducting rf experience could be used. HOM loads and wakefields are expected to vary as a^{-2} and λ^{-2} and $\sigma^{-1/2}$, where a is the cavity aperture, λ is the acceleration wavelength and σ is the bunch length [177,178]. Calculations indicate that the wakefields would limit bunch intensities to $\approx 2 \times 10^{12}$ with 1300 MHz superconducting rf in a recirculating linac accelerator scenario. The longitudinal dynamics is microtron-like or synchrotron-like and off-crest acceleration enables compensation of the linear part of the wakefields, with synchrotron-like phase stability [167].

E. Simulations

TABLE XI. Parameters of Acceleration for 4 TeV Collider

	Linac	RLA1	RLA2	RCS1	RCS2
E (GeV)	0.1→ 1.5	1.5 → 10	10 → 70	70 → 250	250 → 2000
f_{rf} (MHz)	30 → 100	200	400	800	1300
N_{turns}	1	9	11	33	45
V_{rf}(GV)/turn	1.5	1.0	6	6.5	42
C_{turn}(km)	0.3	0.16	1.1	2.0	11.5
Beam time (ms)	0.0013	0.005	0.04	0.22	1.73
$\sigma_{z,beam}$(cm)	50 → 8	4 → 1.7	1.7 → 0.5	0.5 → 0.25	0.25 → 0.12
$\sigma_{E,beam}$(GeV)	0.005 → 0.033	0.067 → 0.16	0.16 → 0.58	0.58 → 1.14	1.14 → 2.3
Loss (%)	5	7	6	7	10

A study [168] followed the longitudinal motion of particles through a similar sequence of recirculating accelerators (see table XI). Cavities similar to those proposed for TESLA [171] were assumed. Figure 39 shows, after optimization of parameters, the final longitudinal phase space distributions corresponding to wakefields estimated for four different bunch charges: a) very small, b) 0.83×10^{12} muons, c) 2.08×10^{12} muons, and d) 4.17×10^{12} muons.

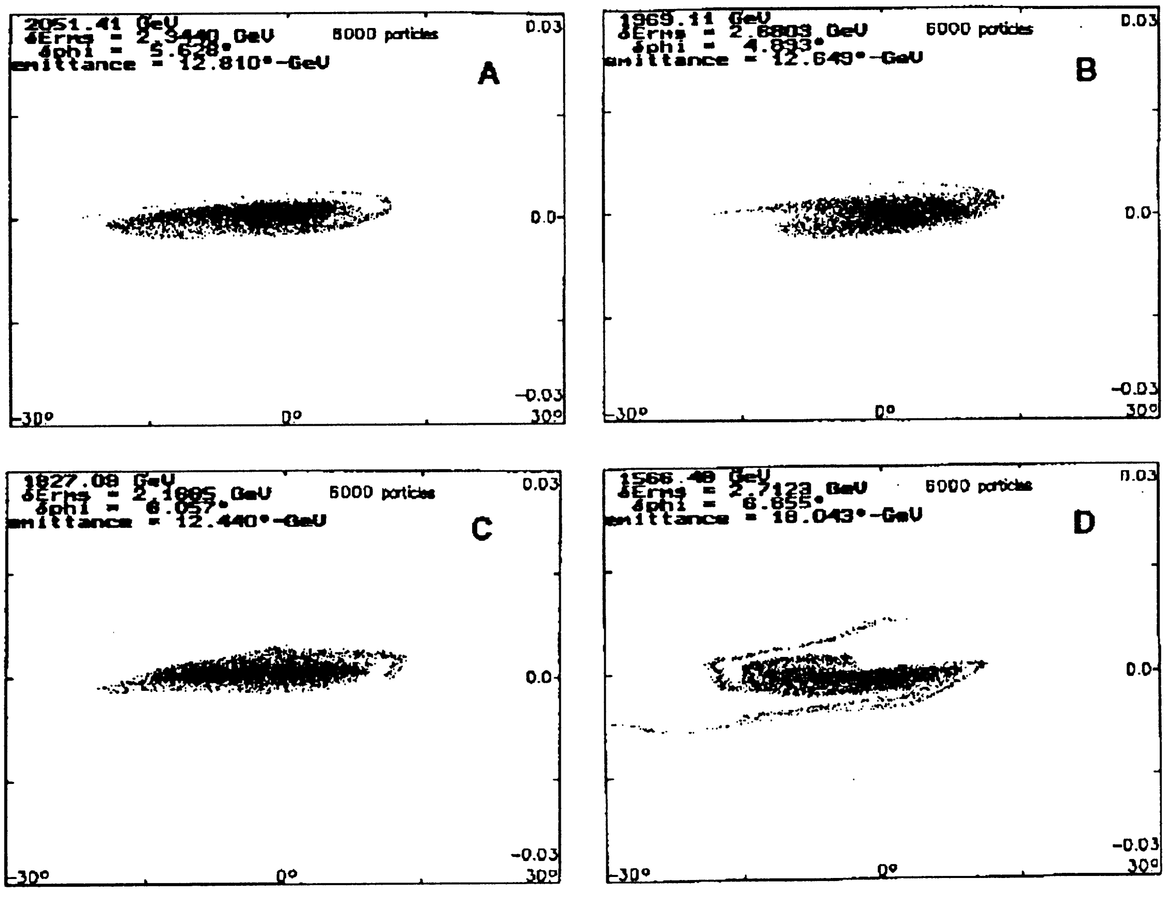


FIG. 39. μ recirculating linac accelerator simulation results with wakefields, with beam accelerated from 200 to 2000 GeV in a 10-turn recirculating linac accelerator. Longitudinal phase space plots for different bunch charges: A) very small number; B) 0.83×10^{12}; C) 2.08×10^{12} and D) 4.17×10^{12} muons in a bunch.

For the design beam charge of 2×10^{12} muons (approximately as for Fig. 39(c)) the wakefield amplitude was estimated to be 2.5 MV/m, the accelerating phase was 35^o, and rf voltage depression 26%. The simulation used an initial longitudinal phase space of 20 eV-s. It gave negligible particle loss, a final longitudinal phase space of 21.6 eV-s, resulting in an increase of longitudinal emittance of only 8%.

F. Acceleration research needed

As discussed above, possible acceleration configurations have been developed, and critical longitudinal motion simulations have been performed. These calculations support the general feasibility of acceleration of muons from cooling to collider energies. However the designs of acceleration systems have not been fully detailed and much work would be needed to obtain a buildable design. Complete transport lattices for linacs and return arcs have not yet been derived, and 6-D phase space-tracking of beams through the accelerators has not been attempted. Also the geometry of combining and separating multi pass beams has not been worked out and optimized.

The rf requirements and systems have been specified at only the rudimentary requirements level, and have not been developed to a constructible level. Optimal configurations and choices of normal or superconducting rf must be developed, as well as more optimal choices in acceleration frequencies. The simple wakefield models used in the initial simulations should be expanded to obtain more realistic systems, and more precise calculations of wakefield effects must be developed.

Rapid accelerating systems have only been outlined at the simplest conceptual level. Prototype magnet design and testing are needed to test the limits of cycling rate and field strengths. Successful magnet concepts must then be specified in terms of stable beam transport configurations, including focusing and transport matching. While beam is stored for only a few turns, the individual bunch intensities are large enough that the possibility of single bunch instabilities must be considered and calculated. The larger number of passes in a recirculating linac accelerator places greater demands on the rf systems and higher order mode (HOM) loads, particularly for superconducting systems.

VII. COLLIDER STORAGE RING

A. Introduction

After acceleration, both μ^+ and μ^- bunches are injected into a separate storage ring. The highest possible average bending field is desirable, to maximize the number of revolutions before decay, and thus maximize the luminosity. Collisions would occur in one, or perhaps two, low β^* interaction areas. Parameters of the rings are given in table I.

B. Lattice design

In order to maintain the required short bunches without excessive rf, the ring must be approximately, or fully, isochronous.

The required β functions at the interaction point are small for a 3 TeV collider, $\beta^* = 3\,\mathrm{mm}$, and the quadrupoles needed to generate them are large (20-30 cm diameter). At 100 GeV CoM, the βs at the IP are more than an order of magnitude larger (see table I), and the quadrupoles are more conventional. In designs of e^+e^- colliders, it has been found that local chromatic correction is essential [179–182] for final focus requirements similar to the cases we consider here.

The rings are racetracks, with two circular arcs separated by an experimental insertion on one side, and a utility insertion for injection, extraction, and beam scraping on the other. The experimental insertion includes the interaction region (IR) followed by a local chromatic correction section (CCS) and a matching section. The chromatic correction section is optimized to correct the ring's linear chromaticity, which is almost completely generated by the IR.

The superconducting bending magnets must be shielded from the electrons emitted by decay of the muon beam. Fields of 8 T have been assumed in the 100 GeV lattices described below, but higher field dipoles would reduce the ring diameters and increase luminosity. Studies of higher field dipoles have been started.

The rf requirements depend on the lattice momentum compaction and bunch parameters. For the very low momentum spread Higgs collider parameters, synchrotron motion does not occur and rf is used solely to correct the impedance generated momentum spread in the bunch. For the higher momentum spread cases, there are two options. If the momentum compaction can be corrected to high order, then synchrotron motion can still be eliminated and the rf is again only used for energy spread correction. Alternatively, if some momentum compaction is retained, then significant rf is needed to maintain the specified short bunches. In either case rf quadrupoles will be required to generate BNS [183,184] damping of the transverse head-tail instability.

C. Lattices

Preliminary lattices have been designed for both 4 TeV and 0.5 TeV machines [185], and several designs now exist for the 100 GeV case. For the 100 GeV CoM collider, two operating cases are being considered: a high-luminosity case with broad momentum acceptance to accommodate a beam with a $\delta p/p$ of $\pm 0.12\%$ (rms), and one with a much narrower momentum acceptance and lower luminosity for beam with $\delta p/p$ of $\pm 0.003\%$ (rms). For the broad momentum acceptance case, β^* must be 4 cm and for the narrow momentum acceptance case, 13 cm. In either case, the bunch length must be held comparable to the value of β^*. It is assumed that a single lattice can be designed that can operate in either mode by adjusting settings.

We describe below two examples of 100 GeV designs. They were generated by two different semi-independent, but converging, studies. The primary difference between the two is in the flexible momentum compaction arc modules.

1. 100 GeV CoM first example

The first lattice design, denoted as Example (a), shown in Fig. 40, has a total circumference of about 350 m with arc modules accounting for only about a quarter of the ring circumference.

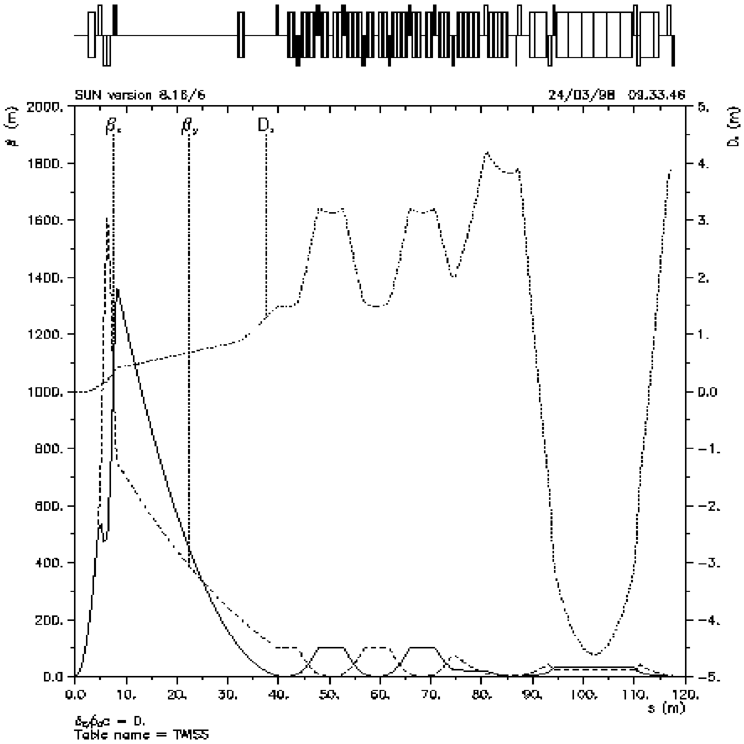


FIG. 40. Example (a): Half of the IR, local chromatic correction, and one of three arc modules.

The low beta function values at the IP are mainly produced by three strong superconducting quadrupoles in the Final Focus Telescope (FFT) with pole-tip fields of 8 T. The full interaction region is symmetric with reflection about the interaction point (IP). Because of significant, large-angle backgrounds from muon decay, a background-sweep dipole must be included in the final focus telescope and placed near the IP to protect the detector and low β quadrupoles [186]. It was found that 2.5 m of an 8 T field is required to attain sufficient background suppression. The first quadrupole is located 5 m away from the interaction point, and the value of the beta function reaches a maximum of 1.5 km in the final focus telescope when the beta functions are equalized in both planes. For this value of beta, quadrupole apertures must be at least 15 cm in radius to accommodate 5σ of a 90 π mm-mrad, 50-GeV muon beam (normalized rms emittance) plus a 2 to 3 cm thick tungsten liner [187]. The natural chromaticity of this interaction region is about -60.

Local chromatic correction of all muon collider interaction regions is required to achieve broad momentum acceptance. The basic approach developed by Brown [180] and others [188], is used. Two pairs of sextupoles, one pair in each plane, are separated by a phase advance of π, all in high dispersion, and each being $(2n+1)\frac{\pi}{2}$ from the IP. The geometric aberrations of each sextupole are canceled by its companion while the chromaticity corrections add.

In our case, this chromatic correction section design (Fig. 41) was optimized to be as short as possible. The β_{max} is only 100 m and the $\beta_{\mathrm{min}} = 0.7$ m, giving a β_{ratio} between planes of about 150 without compromising aperture by a large amplitude dependent tuneshift.

The sextupoles are centered about the opposite plane minimum ($\beta_{min} < 1$), which provides a chromatic correction with minimal cross correlation between planes. A further advantage to locating the opposite plane minimum at the center of the sextupole is that this point is $\frac{\pi}{2}$ away from, or "out of phase" with the chromatic effects from the final focus quadrupoles; *i.e.*the plane not being chromatically corrected is treated like the IP in terms of phase to eliminate a second order chromatic aberration arising from the sextupole.

The large beta ratio combined with the opposite plane phasing allows sextupoles for opposite planes to be interleaved without significantly impacting lattice nonlinearity. In point of fact, interleaving improved lattice performance over a non-interleaved correction scheme, primarily due to a shortening of the chromatic correction section, which is accompanied by a lowering of its contribution to the overall chromaticity [189]. Shallower minima and less variation in beta were also applied to soften chromatic aberrations, but at the expense of an exact π phase advance separation between sextupole partners. The retention of an exact π phase advance difference between sextupoles was found to be less important to the dynamic aperture than elimination of minima with $\beta_{\mathrm{min}} < 0.5$ m.