Eur Phys J C **33**, s01, s1059–s1063 (2004)
Digital Object Identifier (DOI) 10.1140/epjcd/s2004-03-1738-0

EPJ C direct
electronic only

Muon colliders

From science fiction to real science

Kara Hoffman, on behalf of the Muon Collaboration

The University of Chicago, Enrico Fermi Institute

Received: 8 January 2004 / Accepted: 13 January 2004 /

Abstract. A muon storage ring operating as a Higgs factory holds the promise, unique to a muon collider, of precision physics at the Higgs pole. A precision physics program requires a high luminosity machine, and the task of obtaining the necessary phase space compression and acceleration within the short muon lifetime presents a challenge to the state of the art in accelerator technology. Recent innovations, including ring coolers and gas filled cavities, might just point the way to a feasible design. Here we discuss the newest developments within the context of the physics goals of such a machine.

1 Introduction

With the commissioning of the upgraded Fermilab Tevatron and the construction of the Large Hadron Collider well underway, high energy physics seems poised for a major discovery before the end of the decade. If current theoretical predictions are correct, the high energy hadron collisions produced in these machines will yield the elusive Higgs boson, and perhaps, provide the first glimpse of physics beyond the Standard Model. While these machines will greatly extend the energy frontier, the precision measurements necessary for building a new, more complete model of particle physics will be left to the next generation of accelerators. One intriguing proposal for such a precision machine is a muon collider [1].

Unlike more conventional electron beams, muons, owing to their larger mass, do not suffer significant energy loss to bremsstrahlung, allowing precise tuning of the beam energy. Further, the more massive muon is expected to have a significant coupling to the Higgs boson. These features, unique to muons, would allow measurements of the Higgs mass and lineshape at a level of precision unrivaled by other proposed machines. The phase-space compression (cooling) and acceleration of the beam within the finite muon lifetime, however, provide a daunting challenge, and until recently, the concepts needed to achieve adequate cooling had not been fully developed.

The recent discovery of neutrino oscillations has since invigorated interest in muon accelerators, as it was realized that the decaying muons present a unique opportunity for neutrino studies by providing an intense collimated neutrino beam whose point of origin and flavor composition are known. The path to a muon collider is now envisioned as an incremental, staged program, which, beginning with a proton driver, could produce compelling physics results at each stage. Through the efforts of a vigorous international collaboration, a design for a feasible, cost effective machine is now beginning to emerge.

2 Physics potential of a muon collider

If the muon collider is to serve as a precision machine, its performance requirements will be driven by the physics of the electroweak scale. In the coming years, discoveries at near term machines, such as the LHC, will hopefully provide, in broad strokes, the mass of the Higgs boson, or the structure of the Higgs sector if it proves more complex than the minimal single Higgs doublet prescribed by the Standard Model. With this information in hand, the design of the muon collider can be finalized. Although it is impossible to predict future discoveries, it is instructive to consider a few of the more fully studied scenarios as a benchmark for ongoing research and development.

2.1 Low energy Higgs factory

The minimal Standard Model calls for a single Higgs boson to endow the spectrum of fermions with their mass through the Yukawa couplings. A relatively heavy particle, such as a muon, is therefore expected to have a substantial coupling to the Higgs. In fact, the $h \to \mu\mu$ coupling is an important measurement for which a muon collider is particularly well suited, as it is a direct test of the fermion mass generation mechanism. This large direct coupling would allow the Higgs to be studied as an s-channel resonance at a muon collider, much as the Z was studied at LEP. By contrast, at an e^+e^- linear collider the

---

Send offprint requests to: kara@fnal.gov

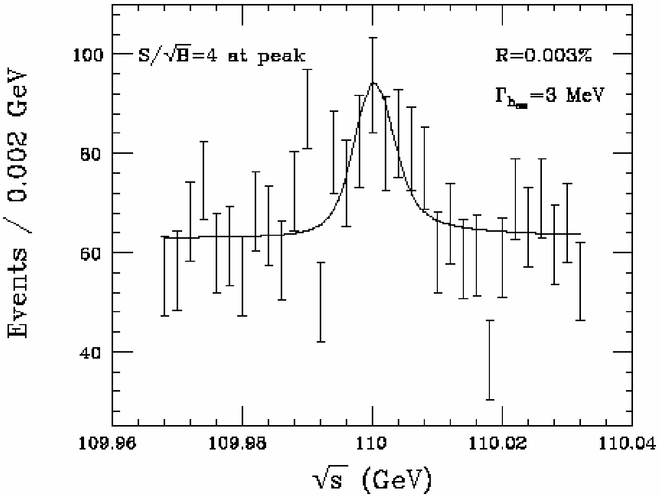

Fig. 1. Number of events and statistical errors in the $b\bar{b}$ final state as a function of center of mass energy. From [2]

Table 1. Comparison of the relative uncertainty for Higgs mass and width measurements achievable at various future and proposed machines assuming a Higgs mass of 110 GeV and Standard Model couplings

	LHC	LC	$\mu^+\mu^-$
$\mathcal{L}$	300 fb^{-1}	500 fb^{-1}	0.2 fb^{-1}
m_h	9×10^{-4}	3×10^{-4}	$1 - 3 \times 10^{-6}$
Γ_h^{total}	> 0.3	0.17	0.2

dominant Higgs production mechanism proceeds through an intermediate Z boson via the Higgstrahlung process, $e^+e^- \to Z^* \to Z^0h^0$.

A muon collider is particularly advantageous if the mass of the Higgs falls below the WW^* threshold. In this scenario, the Higgs would appear as a narrow resonance, and the ability to precisely tune the beam energy could be exploited to preform a scan over the Higgs peak, allowing the width to be measured directly, and a fit to this peak would allow the Higgs mass to be determined to unrivaled accuracy. Figure 1 illustrates such a scan for a hypothetical Higgs with Standard Model couplings and a mass of 110 GeV. In this simulation the beam energy spread, chosen to be smaller than the higgs total width in order to enhance the s-channel production rate, is a mere three parts in one thousand. Although the beam energy could be tuned to this level through continuous spin rotation measurements, it should be noted that there is a trade off between luminosity and beam energy resolution.

2.2 SUSY Higgs factory

If nature happens to be supersymmetric, at least two Higgs doublets would be required, giving rise to a spectrum of five physical Higgs bosons: the neutral scalars H^0 and h^0 (where $m_{H^0} > m_{h^0}$, by definition), the CP-odd A^0, and the charged Higgs pair $H^{\pm}$. The phenomenology of the model is largely determined by these masses, along with $\tan\beta$, defined as the ratio of the expectation values of the two Higgs doublets. In fact, if only one light Higgs is discovered, precision measurements of its properties, as described in Sect. 2.1, might provide the first indication of supersymmetry and determine a range of allowed masses for the heavy Higgs, H^0 and A^0.

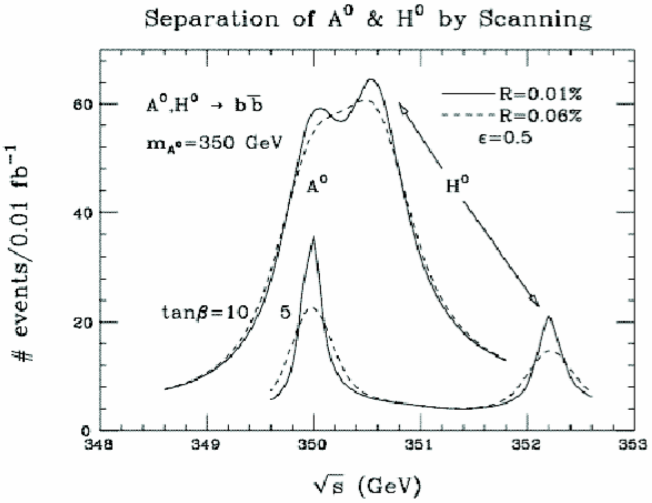

Fig. 2. Resolution of the heavy Higgs, A and H, peaks at $\tan\beta$'s of 5 and 10 for various beam energy resolutions. From [3]

The areas of this parameter space where only one neutral Higgs is within reach of the LHC or an electron linear collider are substantial. At a linear collider, where the dominant production mechanism is $e^+e^- \to Z^* \to H^0A^0$, heavy Higgs' cannot be observed for $m_{H^0} + m_{A^0} > \sqrt{s}$. At the LHC, the discovery channels for the heavy Higgs are $gg \to H^0b\bar{b}, A^0b\bar{b}$ which is dominant at high $\tan\beta$, but a no discovery region exists for an A^0 mass above 250 GeV at intermediate $\tan\beta$. Most of this region could be covered by a muon collider. Further, in the decoupling limit where the H^0 and A^0 are heavy, their masses may be degenerate, particularly at high $\tan\beta$. As shown in Fig. 2, a beam energy scan may be used to distinguish the peaks, but at higher values of $\tan\beta$ the peaks will be smeared out for all but the smallest values of the beam energy spread.

2.3 A machine for the new energy frontier

As the energy frontier is pushed higher, the size of the machines needed to probe the higher energy scales become larger. Muons might offer a significant advantage since, unlike electrons, they can be accelerated in a sharper turn radius and unlike proton colliders, they are not composite particles, and therefore all of their energy is available for collision. A muon collider offers the advantage of being compact, in addition to the advantages already enumerated in the preceding discussion.

Table 2. Beam parameters and luminosity for various operating modes of a muon collider with a six dimensional emittance, $\epsilon_6 = 1.7 \times 10^{-10}$. The number of Higgs per year assumes a cross section of $\sigma = 5 \times 10^4$ fb and a width of $\Gamma = 2.5$ MeV and a "Snowmass" year of 10^7 s. Adapted from [4]

CME (TeV)	0.1			0.4	3
μ/bunch (10^{12})	4			2	2
rms Δp/p %	0.12	0.01	0.003	0.14	0.16
β^* (cm)	4.1	9.4	14.1	2.6	0.3
σ_z (cm)	4.1	9.4	14.1	2.6	0.3
σ_r (μm)	86	196	294	26	3.2
σ_θ (mrad)	2.1	2.1	2.1	1.0	1.1
$\mathcal{L}$ ($cm^{-2}s^{-1}$)	$1.2x10^{32}$	$2.2x10^{31}$	10^{31}	10^{33}	$7x10^{34}$
$\frac{\text{Higgs}}{\text{year}}$ (10^3)	1.9	4	3.9		

3 Design goals

The challenge is defined: in addition to high luminosity, a Higgs factory benefits from a beam energy spread that is small relative to the Higgs width. This is illustrated in Table 2 which tallies the number of Higgs expected for various values of the beam energy resolution and luminosity, assuming a narrow Higgs with Standard Model couplings. The Higgs yield increases in going from Δp/p=0.12% to Δp/p=0.01% in this scenario due to the enhanced cross section even though the luminosity decreases. This decrease in luminosity occurs because the numbers in the table are calculated for a hypothetical fixed value of the six dimensional emittance, $\epsilon_6 = \epsilon_x \epsilon_y \epsilon_z = \sigma_x \sigma_{x'} \sigma_y \sigma_{y'} \sigma_z \sigma_\delta$. The luminosity for a muon collider is given by:

$$\mathcal{L} = \frac{f_0 n_s n_b \gamma_\mu N_\mu^2}{4\pi \epsilon_N \beta^*} \tag{1}$$

where f_0 is the revolution frequency, n_s is the muon lifetime in turns, n_b is the number of bunches, N_μ is the number of muons per bunch, and β^* is the beam envelope parameter such that the beam size at collision is $\sigma^2 = \epsilon_N \beta^* / \gamma_\mu$. Liouville's Theorem tells us that the emittance, which is simply a measure of the phase space enclosed by an ensemble of particles, is a constant of the motion when only acted upon by conservative forces, as it would be in an ideal accelerator. For a fixed value of the emittance, therefore, decreasing the beam energy spread, σ_δ will necessarily increase the transverse emittance, and thus the beam spot size, resulting in a drop in luminosity. The ability to cool a muon beam is clearly critical for the construction of a muon collider, and, in order to achieve the performance in Table 2, the phase space of the muon beam must be decreased by six orders of magnitude from the time the muons are produced in pion decay.

4 Six dimensional cooling and "emittance exchange"

The only technique capable of achieving the amount of cooling required within the finite muon lifetime is ionization cooling. By Liouville's theorem, cooling may only be achieved using dissipative forces. Ionization cooling employs an energy absorbing material to dissipate the energy of the traversing muons. These absorbers are interspersed with rf cavities which replace the lost energy by restoring only the longitudinal component of the momentum. Of course, placing material in the beam also introduces multiple scattering effects which heat the beam. The amount of cooling is given by:

$$\frac{d\epsilon}{dz} = -\frac{\epsilon_N}{\beta^2 E}\frac{dE}{dz} + \frac{\beta_\perp (13.6 \text{ MeV}/c)^2}{2\beta^3 E m_\mu L_R} \tag{2}$$

where the first term is the cooling effect and the second term is the amount of heating. In order to achieve a net cooling effect, the cooling channel must be designed such that first term dominates. It is therefore desirable to choose an energy absorbing material with a large radiation length, L_R, and to place the absorber at a point in the beam where the beam envelope function, $\beta_\perp$, is minimized by strong focusing magnets. While this method can achieve a net cooling effect, the emittance is reduced only in the transverse coordinates at the expense of introducing energy straggling which gradually increases the longitudinal emittance of the beam.

For a neutrino factory, one need only reduce the transverse phase space sufficiently for injection, thus the energy straggling that inevitably results from ionization cooling may be acceptable. If necessary the beam can simply be chopped if the length of the bunch exceeds the acceptance of the cavities. This solution is clearly unacceptable for a muon collider, since, in addition to the desire for a narrow beam energy spread, it is clear from 1 that for a fixed number of muons, the luminosity is maximized by minimizing the number of bunches. It is therefore necessary to either recombine bunches or devise a method to cool the beam longitudinally.

Liouville's theorem suggests a method of longitudinal cooling: since the emittance is constant, increasing the transverse size of the beam will reduce the longitudinal coordinate of the beam in a process known as "emittance exchange". This is, of course, accomplished using a magnetic field, either a bent solenoid or a dipole, which introduces a dispersion causing the drift length to increase proportional to a particle's momentum. A wedge shaped absorber can then be placed in the path of the beam such that the higher momentum particles have a longer path length in the absorber. Since the price paid is an increase in transverse emittance, this longitudinal cooling must be followed with transverse cooling, because if the transverse size of the beam is allowed to grow unchecked, it will eventually lead to beam losses.

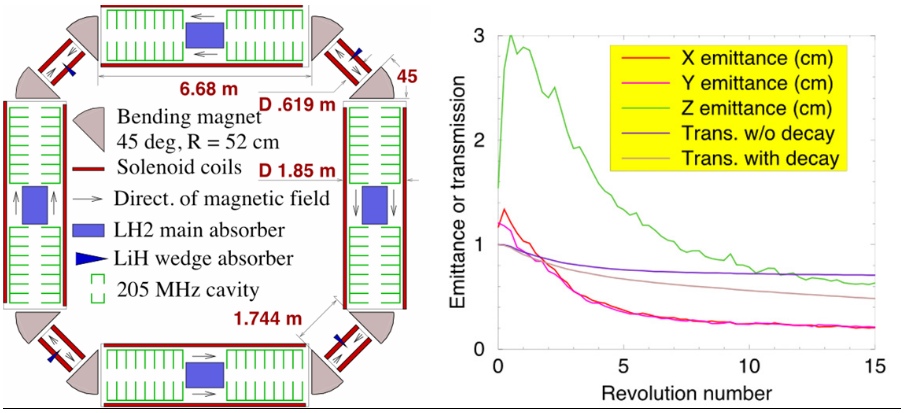

Fig. 3. A schematic of the Balbekov ring cooler and a performance plot showing the emittance and the transmission as a function of the number of turns

4.1 A simple ring cooler

In addition to increasing the transverse emittance, the magnets used to introduce momentum dispersion also necessarily lead to a bend in the beam's trajectory. This fact can be exploited for considerable cost savings by cooling the beam in a ring. Sending the beam through the same channel for multiple passes saves on both real estate and component costs.

Several designs for ring coolers have been proposed, and Fig. 3 illustrates a minimalist version designed by V. Balbekov [6]. In its barest form, a ring cooler includes discrete sections for longitudinal and transverse cooling. This version provides for longitudinal cooling via a short straight section containing a lithium hydride wedge absorber flanked by dipole magnets to produce both bending and the necessary dispersion. Transverse cooling is provided by long straight sections with liquid hydrogen absorbers placed between 205 MHz rf cavities. The longitudinal cooling section is enclosed by a solenoidal field which flips polarity at its center, maximizing the dispersion function at the absorber wedge and preventing a buildup of angular momentum. The transverse cooling sections are surrounded by solenoidal magnets whose field is adiabatically increased from 2.06 T at the edges to 5.15 T at the center in order to minimize the beta function at the absorber so that the heating term in 2 is minimized.

The performance of this solenoidal ring cooler is quantified in the plot on the left of Fig. 3 which shows a significant reduction in all three components of the emittance. Note that the amount by which the emittance is reduced decreases with each successive turn, and the emittance trends toward some equilibrium value,

$$\epsilon_{equilibrium} = \frac{\beta_\perp E_s^2}{2 g_x \beta m_\mu c^2 L_R \frac{dE}{ds}} \ , \tag{3}$$

which can be derived by setting the heating term equal to the cooling term in 2 ($d\epsilon/dz = 0$). It follows that for a cooling channel with a particular focusing parameter, $\beta_\perp$, and a given absorber material with a characteristic $L_R \cdot dE/ds$, there is a minimum emittance which can be obtained. It is therefore convenient to define a merit factor, defined as the final six dimensional emittance divided by the initial six dimensional emittance multiplied by the transmission efficiency to account for lost and decayed muons. For the ring cooler in Fig. 3, simulations predict a merit factor of 38 after 15 turns.

4.2 Bunch compression in a ring cooler

The simultaneous reduction in both the transverse and longitudinal emittance achieved in cooling rings will tend to coalesce a diffuse beam after a number of turns. This suggests that a judiciously designed cooling ring could be used as a bunch compressor. The cooling ring presented in Fig. 3 was optimized for strong transverse cooling which required strong focusing at the absorber and high gradient cavities. A high gradient implies a high frequency and a correspondingly short wavelength, thus a 205 MHz cavity has an acceptance of only $\sim$ 50 cm, an order of magnitude smaller than the bunch length expected as it enters the cooling channel. A ring cooler optimized to efficiently reduce the longitudinal spread by using lower frequency cavities and correspondingly weaker focusing could be used to prepare the beam for injection into downstream strong cooling channels.

The magnetic fields in ring coolers are arranged such that the dispersive effects in the longitudinal cooling sections are ideally canceled in the long straight sections used for transverse cooling. In practice, some parasitic dispersion remains, and, at this early stage of the cooling where there is still a large momentum spread, this can lead to synchrotron oscillations. This parasitic dispersion can be weakened by decreasing the bending angle, and the corresponding decrease in the dispersion can be compensated with a thicker wedge absorber. The most recently studied bunch compressors employ 36 MHz cavities and are octagonal in shape, with eight identical cells, each containing one long and one short section which perform the separate function of transverse and longitudinal cooling respectively.

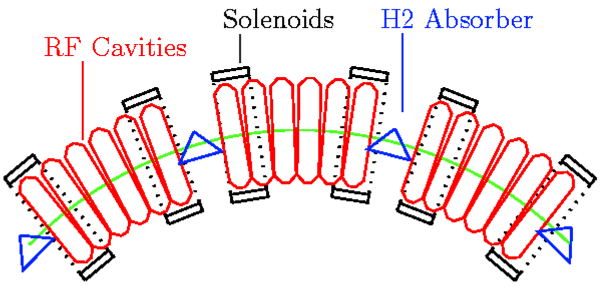

Fig. 4. Three cells of an ionization cooling ring using an alternating solenoid focusing lattice

4.3 Other ring cooler designs

One way to increase the momentum acceptance and eliminate integer betatron resonances is to employ a ring in which all cells are strictly identical, and the functions of transverse and longitudinal cooling are either alternated more rapidly or combined in a single section, so that the phase advance per repeat is small. The desire to simplify the magnet technology motivated the design of cooling lattices which employ only dipoles and/or quadrupoles for dispersion and focusing, and a single wedge absorber for both longitudinal and transverse cooling. Although simulations show that such ring can be used to reduce emittance, with one scenario achieving a merit factor of 16 [7], the overall momentum acceptance was degraded, apparently due to weaker focusing at the absorber compared with solenoid rings. A more promising approach is to use identical solenoid focused cells, a shown in Fig. 4, with a field flip at the center of each cell. In this design the bending and dispersion are introduced by alternately tilting the solenoids by 3^o in the vertical plane. Neglecting problems associated with injection and extraction, alternating solenoid rings with merit factors exceeding 100 have been simulated [7].

4.4 Gas filled rf cavities

The development of high gradient cavities needed for cooling channels is technically challenging. Paschen's Law, which states that the breakdown voltage for discharge between electrodes in gases is a function of the pressure and distance, suggests a way to achieve these gradients that will only work for muons: fill the cavities with gas. In addition to suppressing cavity breakdown, this technique may also cool the beam by using the gas as a homogeneous absorber. Back of the envelope calculations show that, if operated at room temperature, 88 atmospheres of gaseous hydrogen would be needed to achieve the same amount of cooling per unit length as a channel with discrete absorbers. This pressure is well above that needed to suppress breakdown, which has recently been measured as shown in Fig. 5. Decreasing the temperature will increase the density, the figure of merit, and thus decrease the operating pressure needed for cooling with the added bonus of cooling the cavities for greater electrical efficiency. This approach would work best in a beam of constant beta, since the absorber would be continuous through the cooling channel.

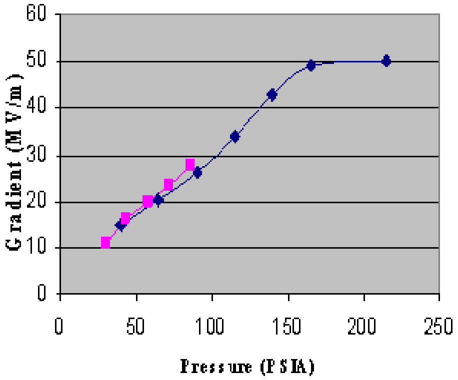

Fig. 5. The Paschen curve for an 805 MHz cavity filled with hydrogen gas at 77K. The diamonds show the recent muon collaboration results

A gas filled channel may even be designed to provide six dimensional cooling. A solenoid plus transverse helical dipole magnetic fields will establish the traversing muons in a helical path through the gas such that higher momentum particles have a longer path length in the ionizing gas. Calculations show that in an idealized 150 m gas filled linear channel, the beam's phase space is reduced by 10^6 [8]. This would be sufficient for a muon collider.

5 Outlook

In the past few years, a number new ideas have been developed which might one day lead to a feasible design for a muon collider, and components such as absorbers and rf cavities are being built and experimentally tested. A Muon International Ionization Cooling Experiment (MICE) is being planned [9] and hopefully, ionization cooling will be experimentally demonstrated before the end of the decade.

References

1. R. Palmer et al.: Nucl. Phys. Proc. Suppl. **51**, A 61 (1996)
2. V. Barger et. al.: *Physics of Higgs Factories*, proceedings of Snowmass 2001, hep-ph/0110340
3. V. Barger et. al.: Phys. Rept. **286**, 1 (1997)
4. M. Alsharo'a et al.: Phys. Rev. ST Accel. Beams **6**, 081001 (2003)
5. D. Neuffer et. al.: *Front End Cooling*, in *Higgs Factory Muon Collider*, eds. D. Cline and G. Hanson, prepared for Snowmass 2001, pgs 45–76
6. V. Balbekov: http://www.mucool.fnal.gov/mcnotes/public/ps/muc0246/muc0246.ps.gz
7. R.B. Palmer: J. Phys. G **29**, 1577 (2003)
8. Y. Derbenev and R.P. Johnson: *Six-Dimensional Muon Beam Cooling in a Continuous, Homogeneous; Hydrogen Absorber*, proceedings of Cool 03, Mt. Fuji
9. E. Radicioni: these proceedings