BNL-101536-2014-TECH

AD/AP 46;BNL-101536-2013-IR

The Muon Collider (Sandro's Snake)

A. G. Ruggiero

July 1992

Collider Accelerator Department

Brookhaven National Laboratory

U.S. Department of Energy

USDOE Office of Science (SC)

DISCLAIMER

AD/AP-46

Accelerator Development Department
Accelerator Physics Division
BROOKHAVEN NATIONAL LABORATORY
Associated Universities, Inc.
Upton, NY 11973

Accelerator Physics Technical Note No. 46

The Muon Collider
(Sandro's Snake)

A.G. Ruggiero

July 1992

The Muon Collider

(Sandro's Snake)

Alessandro G. Ruggiero
Brookhaven National Laboratory
July 22, 1992

1. Introduction

In the quest for the Higgs bosons, a muon collider may be conceived as the experimental device more affordable and more feasible than electron-positron or very large hadron colliders, like NLC, CLIC, SSC and LHC. Muons have a mass ten times lighter than protons and are therefore easier to be steered on circular trajectories. On the other side their mass is a hundred times heavier than electrons and their motion is considerably less affected by the synchrotron radiation.

Muons are elementary lepton particles, with no internal structure. Like the electrons, they have obvious advantages over the hadron counterpart when they are used as they main projectiles for the production of the Higgs bosons. Moreover, because of their larger mass, they are also better suited than the electrons themselves, due to a considerably larger propagator constant.

Unfortunately, muons do not exist in nature and they have to be produced with the only technique we know these days: impinging an intense beam of protons on a target. This will cause muon production, but with a very large volume of the phase space. Like in the case of the production of antiprotons, in order to make the beam of some use for the subsequent collisions, muons also have to be collected and cooled to a sufficiently high intensity and small beam dimensions, before they can be accelerated and injected in the collider proper.

To make the situation more complicate, there is also the fact that muons are intrinsically unstable particles with a very short lifetime. Accumulation, cooling, acceleration and all other required beam manipulations are then to be executed extremely fast if one requires that a large fraction of the particle beam survives to the collision point.

This paper describes a feasibility study for the design of a muon collider. Recognized the fact that the particle lifetime increases linearly with the energy, we have adopted a scheme where steps of cooling and acceleration are entwined. We have indeed found convenient to accelerate the beam as fast as possible to increase its chances of survival, and necessary to dilute the action of cooling throughout the entire accelerating process to make it more effective and affordable. All acceleration and cooling steps are executed in a single pass essentially along a curvilinear and open path. We do not believe it is possible to handle the beam otherwise in circular and closed rings, as it has been proposed in the past. [1,2]

The example shown in this paper describes a muon collider at the energy of 250 GeV per beam and a luminosity of 4×10^{28} cm^{-2} s^{-1}.We have adopted an extrapolation of the stochastic cooling method for the reduction of the beam emittance.

2. Proposed Scenario

A schematic layout of the muon collider is shown in Figure 1. It is made of three major parts: (1) a high intensity proton source with a target station attached to it for the production of muons; (2) two accelerating sections, one for each beam, with bending dispersed for providing betatron stochastic cooling; and (3) a final collision region which eventually can include a storage and collider ring.

3. The Proton Source and Muon Production

Recent studies of hadron facilities (the EHF, for instance [3]) have demonstrated that it is possible to accelerate proton beams to the energy of 30 GeV at high repetition rates for an average output current of 100 μA. This facility can be made of a 1.2 GeV linac, a 9 GeV booster ring with the circumference of 480 m operating at the repetition rate of 50 Hz, an accumulator ring of the same dimension, operating at constant field, and a main ring of 960 m circumference operating at 25 Hz. If the facility is followed by a Stretcher Ring of the same dimension of the main ring, it is possible to deliver the beam continuously with essentially 100% duty cycle. We shall assume that such proton source is available.

After being slowly extracted from the Stretcher Ring, the proton beam can be rebunched in a sufficiently long travelling wave linac at the frequency f_b = 3 GHz (for preparing the time structure of the muon beam to be generated). By entering a second stage of the linac with considerably larger voltage gradient, it is possible to create a mismatch which will make the proton bunches rotate in their own buckets. After a quarter of the oscillation the bunches will present their narrowest length. At the same time the beam is focussed to a small spot size at the location of a target for the production of the muon particles. All these processes can be accomplished essentially with no beam losses. The dimensions of the proton bunches can be made small enough to have no consequences on the dimensions of the beam of muons. Each proton bunch impinging the target is made of 2×10^5 particles.

Muons are produced in a cascade as the decay product of π mesons in pairs of μ^+ and μ^-. Large production rates are expected [4]; for instance, the following reference values are customarily taken: a yield of 0.1% by accepting a momentum bite of ±5% and a semi-angular aperture of 50 mrad. Since we require the muon beam coming out of the target to have a reasonably small momentum spread for capture and acceleration, we shall take more conservatively a full momentum bite of only 2%. The production rate increases linearly with the momentum spread and about quadratically with the angular acceptance. Thus, with these adjusted values, we can estimate a pair production rate of about 2 x 10^{-4} per proton. There is a continuous streaming of muons of both sign from the target with an average current of approximately 20 nA for each specie. At the same time an optimum production energy can be chosen to be about 1 GeV. Since the mass of the muons is about 100 MeV, this corresponds to $\gamma \sim 10$.

The muon beam has the same rf structure f_b of the proton beam, that is 3 GHz, and the same bunch length. There will be about 40 muons of each sign per bunch. The length of the target should match the range for the muon production; we take here for the following estimate a target length of $l = 1$ cm. The resulting muon beam betatron emittance is then $\varepsilon = l\theta^2 = 8\,\pi$ mm mrad. The normalized emittance is $\varepsilon_{init} = 80\,\pi$ mm mrad.

The two species of muons with opposite electric charge are first separated by a common dipole magnet and then transported by a focussing channel which is to be matched to the beam aspect ratio at the target, that is a value of $\beta_T = 1$ cm. Each of the two beams then undergoes to the same sequence of bending, stochastic cooling and acceleration, until for each of them, the final emittance and energy values are reached.

4. The Accelerating and Cooling Section

This has the shape of a snake (*Sandro's snake*) with convolutions increasing in size toward the large energy end. We can assume that there M_c of such convolutions, each made of a bending arc followed by a straight accelerating section, as shown in Figure 2. The straight sections are made of travelling-wave rf-cavity structures a\` la SLAC for the acceleration of the muon beam; FODO cells for transverse focussing are also provided dispersed. The accelerating rf frequency f_{acc} can also be chosen around 3 GHz, with an effective accelerating gradient W of few tens of MVolt/m. The electric power demand for a continuous mode of operation may be exceedingly too large to be afforded; in this case one might have to resort either to superconducting cavity technology or to the introduction of a duty cycle. As we shall see later, it is indeed possible to re-use the beams over and over in an ultimate large storage ring operating at constant field. During this time the accelerating rf system may be turned off.

If the convolutions are labelled in sequential order, $i = 1, 2, \ldots, M_c$, we can then define L_i to be the overall length of the i-th accelerating section; similarly C_i will denote the arc length of the corresponding convolution. We can then easily estimate the overall length of the accelerating and cooling section. As we shall see, a large contribution to the total length of the section is given by the arcs; thus the amount of accelerating gradient is not necessarily an issue. For completeness we shall also denote with E_i the beam kinetic energy and with ε_{ni} the normalized betatron emittance at the end of the i-th convolution.

The design of the first straight accelerating section may require special care because the muon beam has still a large momentum spread. Just before entering the first convolution, one will apply bunch rotation at the same accelerating frequency f_{acc} to trade momentum spread with length, as it is done in the antiproton sources of Fermilab and Cern. It is not clear at this moment whether this can be accomplished on a fly, along a straight path with one linac, or whether this will require some sort of circular ring at constant energy (as, for instance, the Debuncher Ring at Fermilab).

The arcs are made of several bending and focussing FODO cells. The bending is provided with dipole magnets operating at a constant field which has the same value B throughout the length of the section. The bending will flip direction from one convolution to the next. The bending angle α_i is only a fraction of π. Other geometries are of course possible, provided they allow convergency of the two beams to the collision point. The convenience of this layout, compared to a complete circular ring, is that one can make use of superconducting magnets without having to cycle them at a too large rate. In the arcs the beam energy is constant, but will vary from arc to arc, and the average bending radius and arc length will also vary accordingly.

Since the beam has essentially the speed of light, the bending is also required to provide enough electric delay for the signal processing of the stochastic cooling. This, as shown in Figures 2, 3 and 4, includes several pickup (PU) stations upstream followed by an equal number of kicker (K) stations downstream. The Schottky signal from the beam travels from the PU's to the K's where it also properly amplified and applied to the beam for the stochastic correction. We assume that a total delay of 10 ns is adequate for signal processing. This value sets a minimum that one can calculate for the arc length and bending. As we shall see later, only a little reduction of betatron emittance is required per convolution. We propose a different method of betatron stochastic cooling, described later, which works effectively at very low beam intensity level and for very short beam bunches.

5. The Collision Region

The last convolution of the accelerating and cooling section is of a size large enough to allow the two sides to merge with each other in a straight, head-on collision path as shown in Figure 1. The collision region is the location where the two beams are brought together with a final focus described by β^*. The collision is essentially head-on; if the beam bunches are too close to each other, a small collision angle may be required to avoid beam-beam interaction with subsequent bunches. On the other side the collision angle is to be small enough to avoid any significant reduction of luminosity.

The two beams are *round*, with the same emittance in the two transverse planes; we consider this an advantage since it allows a symmetric focussing arrangement with the same value of β^* in the two planes. Moreover the intensity per bunch, at least in our scenario, is considerably low, so that no serious and disruptive effects are expected from the beam-beam interaction. The large frequency of bunch encounter at the collision point is also an advantage for the experiment setup, which prefers a smoother distribution of the events to be detected in time.

As shown in Figure 1, the last convolution of the two sides can be thought as part of a larger final storage ring. In this case the collision region with the final focus can also be conceived to be an integral, matched part of the storage (and collider) ring. This configuration may be advantageous if the muon beam has acquired enough energy and lifetime to make it survive through several hundreds of revolutions. As the beams circulate and collide until exhaustion in the storage ring, the accelerating section may be temporarily turned off to reduce the electric power demand. This mode of operation is feasible only with a relatively large value of β^*, of few centimeters, if the final focus is to be integral part of the storage ring.

6. Requirements and Goals

The major requirement parameters are the final energy E and the luminosity L. For the round beam configuration at collision we have very simply

$$L = \frac{N_+ N_-}{\beta^* \varepsilon_n} f_b \gamma \tag{1}$$

where N_+ and N_- are the number of particles per bunch of each specie, and ε_n is the final normalized betatron emittance, defined as $\varepsilon_n = 4\pi \gamma \sigma^2/\beta^*$ where σ is the rms beam spot size at the collision point. The requirement on the luminosity versus energy depends eventually on the cross-sections of the μ^+ - μ^- collision according to the following scaling

$$L = L_{ref} \left(\frac{E}{E_{ref}}\right)^2 \tag{2}$$

where, very likely, if $E_{ref} = 1.25$ TeV then $L_{ref} = 1 \times 10^{30}$ cm^{-2} s^{-1}. Combining the two equations above yields the following for the normalized emittance

$$\varepsilon_n = \frac{N_+ N_- f_b \gamma_{ref}^2}{\beta^* \gamma L_{ref}} \tag{3}$$

It is seen then that the required normalized emittance decreases inversely with the energy at collision. It is obvious that this requirement is considerably smaller, by several order of magnitude, than the value ε_{init} of the beam emittance at the point of production. It is proposed here to recover this difference with stochastic cooling.

Muons are unstable particles and they decay rather fast. At the kinetic energy of 1 GeV the lifetime is of only 21 μs. Due to relativistic effects, the lifetime increases linearly with energy; for instance, it is 21 ms at 1 TeV. It is easy to calculate the survival ratio in every convolution of the collider and the fraction of the beam that survives at the large energy end. It is required that this fraction is large enough, so that eventually the beams can be used again in multiple collision mode in the final storage ring.

In order to make some estimate we shall take here a final energy of 250 GeV (that can be fitted easily in the RHIC tunnel at BNL), that is $\gamma = 2500$. The required luminosity is

4×10^{28} cm^{-2} s^{-1} and the normalized beam emittance $\varepsilon_n = 0.1 \times 10^{-8}\ \pi$ mm mrad, that is a reduction of eleven orders of magnitude from the value ε_{init} at the target production.

7. Stochastic Cooling

An optimum luminosity configuration, for a given flux of particles entering the collider, which does not impose too stringent requirements on the beam emittance, is a low repetition rate. Unfortunately in this case the number of particles per bunch is too large and it would make impossible the application of a cooling technique like stochastic cooling. This on the contrary, specially in our case, requires a very low number of particles per bunch and thus a considerably higher repetition rate and a final smaller emittance.

Moreover, the bunches are extremely short; a fact which also makes practically impossible the application of stochastic cooling as ordinarily conceived [5] based on the property of *longitudinal mixing*. We shall deviate here from this and consider a *single pass* cooling method which acts on all particles in the same bunch at the same time.

Let us suppose that at the pickup location (PU) in one convolution we measure the center of mass of a bunch particle distribution in either horizontal or vertical plane (denoted by z). The displacement is caused by statistical fluctuations and given as the average over all the particle position, that is

$$\bar{z} = \frac{1}{N}\sum_m z_m \tag{4}$$

where N is the number of particles in the bunch. At the same time, if we denote with β the amplitude lattice function at any desired location, the beam rms emittance ε can be defined as

$$\beta\varepsilon = \frac{1}{N}\sum_m z_m^2 \tag{5}$$

In average, the relation between these two quantities is the following

$$\bar{z}^2 = \beta\varepsilon/N \tag{6}$$

Suppose now that at the Kicker location (K) the position z_m of each particle is corrected by an amount proportional to the average bunch displacement measured at the previous PU location, by a factor g. It is easily seen that

$$(\beta\varepsilon)_K = \frac{1}{N}\sum_m (z_m - g\bar{z})^2$$

$$= (\beta\varepsilon)_{PU} - (2g - g^2)\bar{z}^2 \quad (7)$$

So that the emittance reduction during a single step occurring in any one of the convolutions is

$$\frac{\Delta\varepsilon}{\varepsilon} = -\frac{(2g - g^2)}{N} \quad (8)$$

The optimum condition, which corresponds to the large emittance reduction, is obtained by setting the gain g = 1. In this case the reduction is just inversely proportional to the number N of particles in the bunch. We shall assume that the system parameters are set for this optimum cooling rate. Nevertheless, this mode of operation works only for a *single pass*. To work again in the successive step, one should regenerate the fluctuation signal at the following PU location. In the case of coasting beams, this is usually done with the longitudinal shear of the particle motion due to the difference in speed among particles, and by the fact that the detecting device measures the position of only a longitudinal section of the beam. We propose here to mix the particle relative order by introducing strong octupoles (or other nonlinear devices) to create enough smear in the betatron motion. The octupoles will be placed in the focussing cells FODO cells of both the accelerating and bending sections.

An interesting feature of this method is that the system can be made to work on a narrow bandwidth. Indeed, it is required that all the particles in the same bunch (and bucket) are observed at the same time. Signal overlapping from different bunches is effectively avoided by choosing an electronic bandwidth which matches the bunching frequency f_b, taken in this paper to be 3 GHz.

Denoting with n_S the number of stochastic cooling *single pass* steps, the final beam emittance is given by

$$\varepsilon_n = \varepsilon_{init}\, e^{-(n_S/N)} \quad (9)$$

It is seen that, with N = 40 particles per bunch, the final required emittance may be obtained with $n_S \sim 1000$ steps.

8. A Possible Solution

There are several ways numbers can be configured together, and the solution we show here is just an example. There may be other more optimal arrangements which need to be found and investigated. In this example we take $M_c = 100$ convolutions.

We begin by taking the same energy gain per convolution (another possibility would be to accelerate faster during the early stages and slower toward the end; one more possibility is just the opposite). This will set the energy gain per convolution to 2.5 GeV which can be achieved over a distance of about 50 meter with an accelerating gradient of 50 MV/m. This alone already require a linear length of 5 kilometers.

We shall take superconducting magnets for the bending arcs. The dipole magnets have a field of 10 Tesla, and we allow for a packing factor of 80%. The arc lengths are adjusted to provide the same total difference of 10 ns between the length of their paths and the length of the associated geometric cords.

We shall assume each arc includes 10 stochastic cooling steps (again, this is just an example of so many possibilities). One possible configuration is sketched in Figure 3 where the steps are entwined with each other in the same arc. Another configuration is shown in Figure 4 where the steps span their function over two consecutive arcs. All these arrangements to work effectively require a considerable amount of transverse mixing from octupole magnets which are placed as often as possible.

The results are shown in Table 1. The total length of the collider is about ten kilometers. At the end, about 98% of the muons have survived. They can then be injected in a storage ring having the dimensions of RHIC, where they can circulate (and collide) for about 400 revolutions, corresponding to their lifetime of 5 ms. Since it takes about 35 μs for the muons to travel the collider, the accelerating rf system can be operated with a duty cycle of less than one percent.

9. Conclusions

We have exposed in this paper the construction in first order approximation of the design of a muon collider. We found this to be a very interesting and appealing project that may be valuable in removing several technical difficulties of an e^+ - e^- linear collider and possibly also of the Super Superconducting Collider.

There are still a lot of questions unanswered, and the concepts exposed still need to be carefully evaluated. For instance, there are some questions concerning the muon production: what is the optimum production energy? This may have an impact on the initial beam betatron emittance together to the production angle; what are really the production rates? These questions can be answered with rather simple experiments, for instance, at the BNL facilities.

It may be possible to upgrade the scenario to larger energies and to larger luminosities. Indeed the collider could be made longer than described here and one can find an optimum configuration of parameters which makes a more efficient use of the beam intensity. Larger luminosities can be obtained by increasing the muon production rate, for instance, by accepting larger momentum bite. What are the limitations here? Still, increasing the intensity is not enough, as one needs to dilute even more the longitudinal particle distribution to accommodate stochastic cooling.

The idea itself of stochastic cooling in a *single pass* needs more study and careful evaluation of hardware limitations. For instance, what are the effects of thermal noise and Schottky noise on the final beam emittance? The optimum gain regime may be limited by the electronic gain toward the high energy end. We find very intriguing (and challenging) the idea of having to deal, and to measure, an intensity as low as few tens of particles per bunch. Some of the extreme technical conditions of the stochastic cooling performance can be experimentally studied at the Fermilab complex. As the beam dimensions get smaller and smaller we may find more and more difficult to generate particle *mixing* with non linear elements as octupole magnets.

Finally, with the Booster soon completely operational, The AGS complex at BNL will be capable of delivering a proton average intensity of 5 to 10 μA. With the addition of a stretcher ring, a high frequency buncher, a target station and a debuncher for the muons we have then an opportunity to demonstrate experimentally several of the concepts exposed here. We can then later expand from there...

10. Acknowledgments

The concepts exposed in this paper were stimulated in the Physics Opportunities section of the Workshop on Advanced Accelerator Concepts held in Port Jefferson, New York, on June 11-15. The author likes to acknowledge and to thank the following persons for their comments, critics and very valuable discussions: P. Chen, D. Cline, K. McDonald, D. Neuffer, R. Noble, R. Palmer and A. Sessler.

11. References

1. A.N. Skrinsky, Proc. of the XX International Conference on High Energy Physics, A.I.P. Conf. Proc. **68** , 1056 (1980)
2. D. Neuffer, Particle Accelerators 14 , **75** (1983)
3. Proposal for a European Hadron Facility, EHF - 87 -18, 18 May 1987. F. Bradamante, University of Trieste, I-34127 Trieste, Italy
4. R. J. Noble, "Particle Production and Survival in Muon Acceleration". Contribution to the Proceedings of the Advanced Accelerator Concepts Workshop, Port Jefferson, Long Island NY, June 14-20, 1992
5. S. van der Meer, "An Introduction to Stochastic Cooling", Physics of Particle Accelerators, AIP Conf. Proc. 153, Volume 2, page 1628 (1987)

Table 1 : Collider Parameters versus the Convolution Number

Convolution no.	Kinetic Energy GeV	Arc Length m	Survival Ratio	Total Length km	Norm. Emittance π mm mrad
5	13.5	13.51	0.996	0.284	2.29E+01
10	26.0	20.71	0.994	0.619	6.57E+00
15	38.5	26.81	0.993	0.988	1.88E+00
20	51.0	32.32	0.991	1.386	5.39E-01
25	63.5	37.33	0.990	1.810	1.54E-01
30	76.0	42.02	0.989	2.258	4.42E-02
35	88.5	46.49	0.988	2.730	1.27E-02
40	101.0	50.79	0.988	3.223	3.63E-03
45	113.5	54.90	0.987	3.737	1.04E-03
50	126.0	58.84	0.986	4.271	2.98E-04
55	138.5	62.60	0.986	4.825	8.54E-05
60	151.0	66.23	0.985	5.397	2.45E-05
65	163.5	69.81	0.984	5.987	7.01E-06
70	176.0	73.38	0.984	6.595	2.01E-06
75	188.5	76.71	0.983	7.220	5.76E-07
80	201.0	80.12	0.983	7.863	1.65E-07
85	213.5	83.32	0.982	8.521	4.72E-08
90	226.0	86.50	0.982	9.196	1.35E-08
95	238.5	89.70	0.981	9.886	3.88E-09
100	251.0	92.93	0.981	10.593	1.11E-09

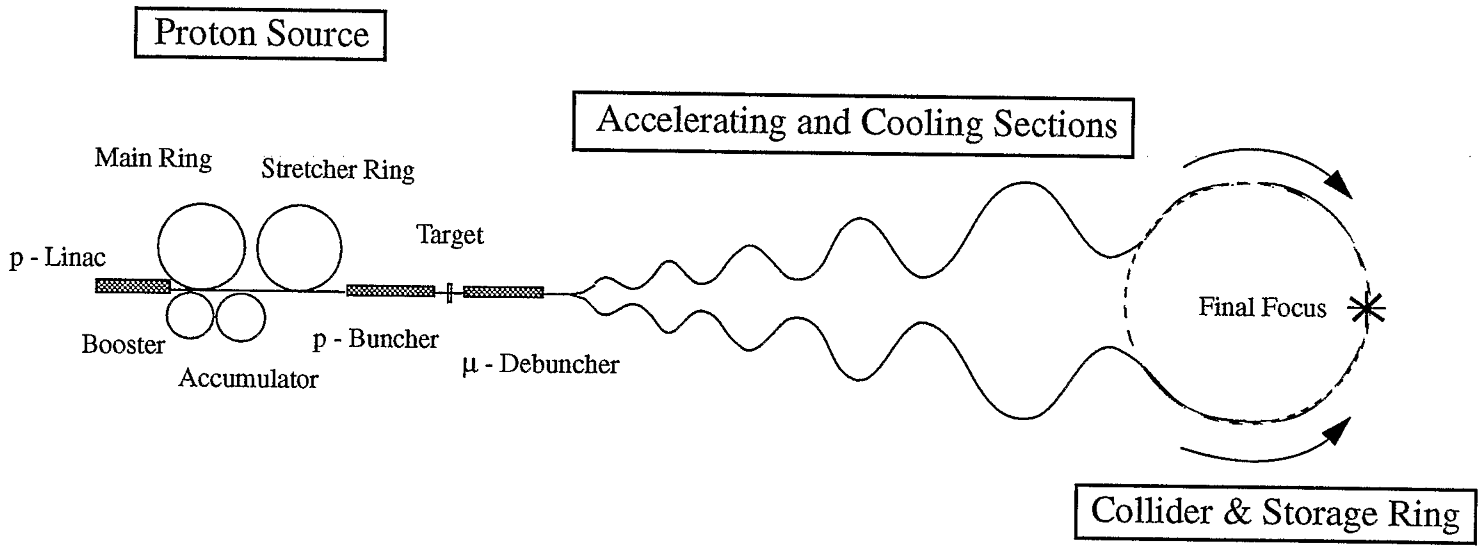

Figure 1: Layout of the Muon Collider (*Sandro's snake*)

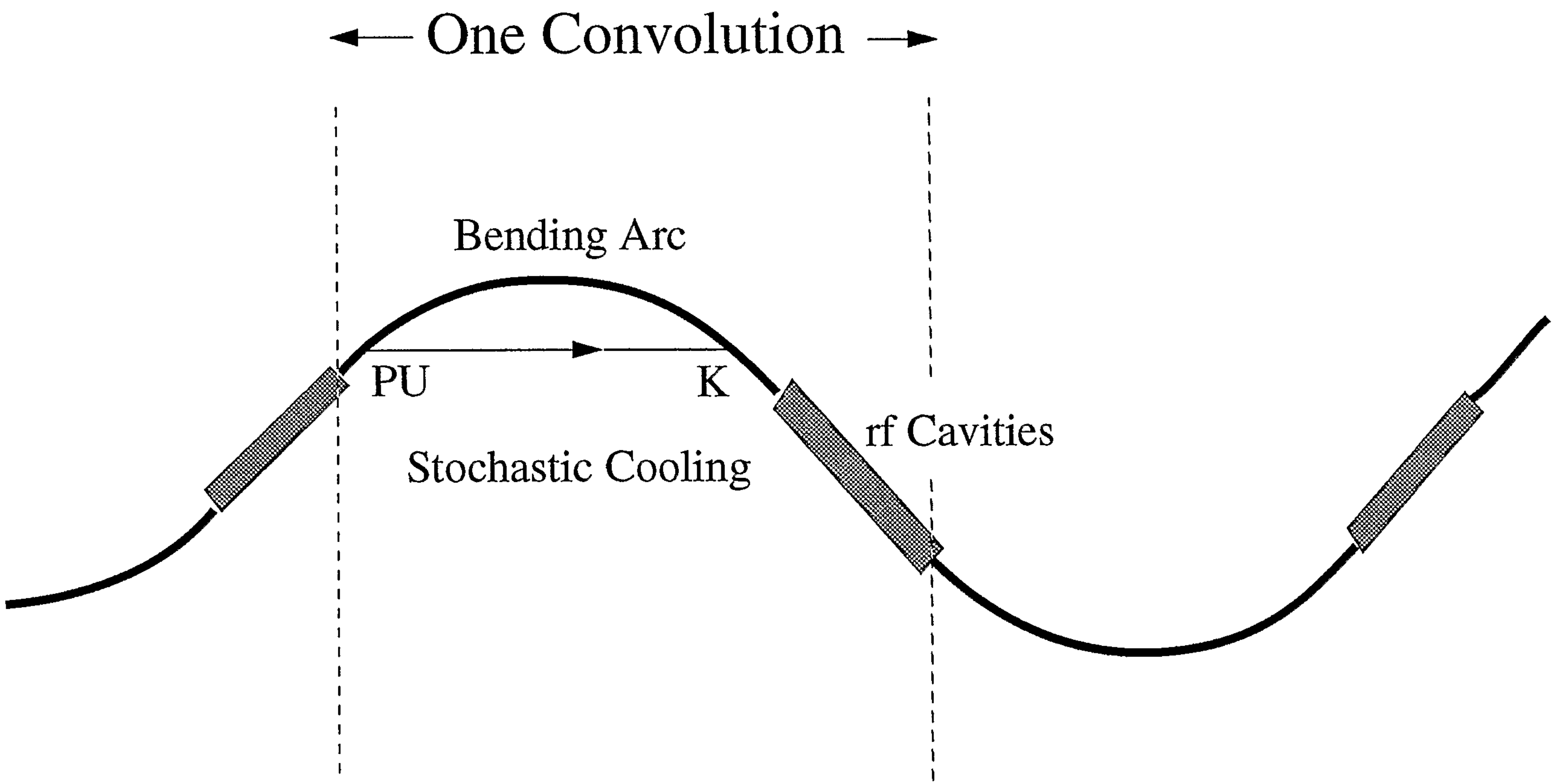

Figure 2 : Details of One Convolution

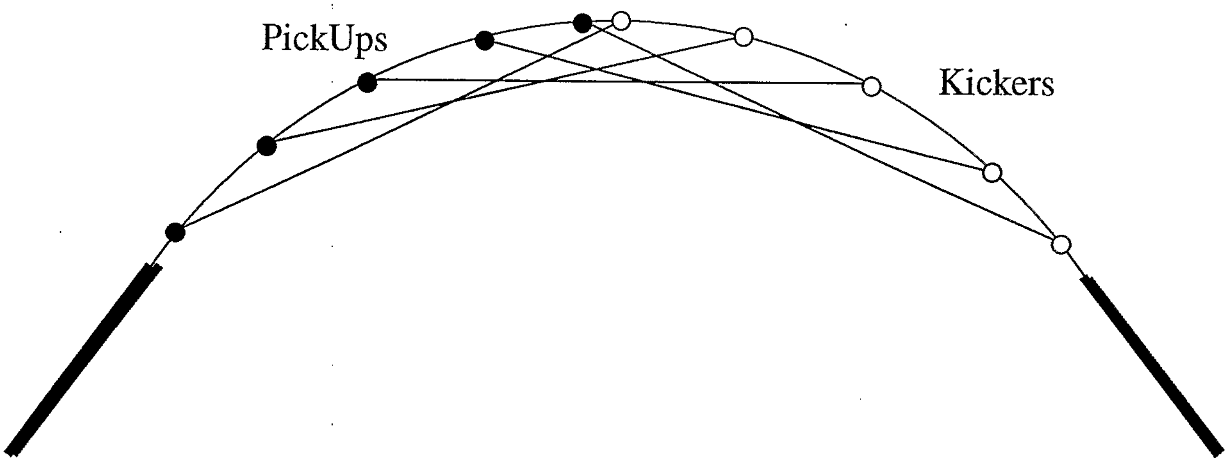

Figure 3 : Multiple Single-Pass Stochastic Cooling Steps
in the same Convolution

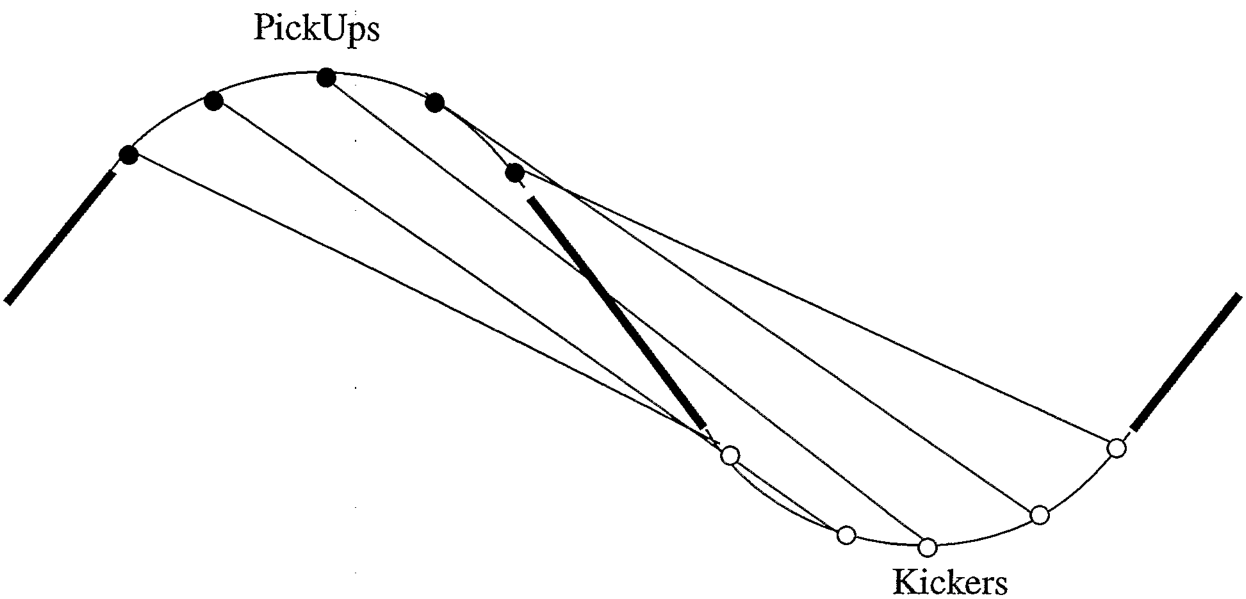

Figure 4 : Multiple Single-Pass Stochastic Cooling Steps
shared by two consecutive Convolutions