The final cooling system for a high-energy high-luminosity lepton collider requires reduction of the transverse emittance $\varepsilon_t$ by an order of magnitude to $\sim 0.00003$ m (rms, N), while allowing longitudinal emittance $\varepsilon_L$ to increase to $\sim 0.1$ m. In the present baseline approach, this is obtained by transverse cooling of low-energy muons within a sequence of high field solenoids with low-frequency rf systems. Recent studies of such systems are presented. Since the final cooling steps are mostly emittance exchange, a variant form of that final system can be obtained by a round to flat transform in x-y, with transverse slicing of the enlarged flat transverse dimension followed by longitudinal recombination of the sliced bunchlets. Other variants are discussed. More explicit emittance exchange can greatly reduce the cost of a final cooling system.

**INTRODUCTION**

The P5 report stated that “for e+e- colliders, the primary goals are improving the accelerating gradient and lowering the power consumptions.”[1] Both of these goals are achieved by increasing the mass of the electrons to a level where multturn acceleration to TeV’s is possible, and radiation effects are small. Increasing the mass to 105.66 MeV changes TeV electrons from a radiation source and enables the possibility of multi TeV heavy electron (μ) colliders. Parameters for possible multiTeV Colliders are included in Table 1.

Table 1: High-energy Heavy-lepton Collider Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Higgs (1/8TeV)</th>
<th>3TeV</th>
<th>6TeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beam energy</td>
<td>0.063</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>Heavy e⁻/⁺ bunch</td>
<td>$2 \times 10^{12}$</td>
<td>$2 \times 10^{12}$</td>
<td>$2 \times 10^{12}$</td>
</tr>
<tr>
<td>Circumference (m)</td>
<td>300</td>
<td>2767</td>
<td>6302</td>
</tr>
<tr>
<td>Tune</td>
<td>5.16/4.56</td>
<td>20.1/22.2</td>
<td>38.2/40.1</td>
</tr>
<tr>
<td>Emittance (μ,N)</td>
<td>0.08</td>
<td>-3E-4</td>
<td>-1.2E-3</td>
</tr>
<tr>
<td>Collision β₁ (cm)</td>
<td>3</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>Energy spread</td>
<td>0.003%</td>
<td>0.1%</td>
<td>0.1%</td>
</tr>
<tr>
<td>rep rate</td>
<td>30 Hz</td>
<td>12 Hz</td>
<td>6 Hz</td>
</tr>
<tr>
<td>Luminosity (10^{32}cm⁻²s⁻¹)</td>
<td>0.002</td>
<td>4</td>
<td>12</td>
</tr>
</tbody>
</table>

![Figure 1: Progression of emittances throughout a collider cooling scenario.](image1)

**BASELINE FINAL COOLING**

A baseline approach to final cooling was developed by Palmer et al. This includes transverse ionization cooling of low-energy muons within high field solenoids, with lower energies and higher fields obtaining smaller $\varepsilon_t$ [4, 5] At low-energies, the variation of momentum loss with energy anti-damps the beam longitudinally, increasing $\varepsilon_L$. Figure 1 shows the progression of emittances throughout a collider cooling scenario, with the “final cooling” portion of that displayed as the lines with transverse emittance decrease and longitudinal emittance increase leading to final values at $\varepsilon_t = 25\mu$ and $\varepsilon_L = 30–60$ mm.

For final cooling, the beam momentum is reduced initially to 135 MeV/c and only transverse cooling is used. The final cooling system consists of ~a dozen stages. Each stage consist of a high-field small bore magnet with an H₂ absorber within the magnet, followed by an rf and drift system within lower-field to phase-rotate and reaccelerate the muons. From stage to stage, the muon beam energy is reduced (from 66 MeV toward
5MeV) and the magnet field strength is increased to minimize $\varepsilon_t$. The relevant equations are:

$$\varepsilon_{N, M} = \frac{\beta_t E_t^2}{2 \beta_t c^2 L_R (dE / ds)}$$

With $B=40T$ and $p_0=33$ MeV/c ($E_0=5$ MeV), $\beta_t \approx 0.56cm$ and $\varepsilon_{N, M} \approx 0.00001m$. However, energy loss is strongly antidamping at low energies and the longitudinal emittance increases dramatically, since the final cooling lattices do not include the emittance exchange needed to obtain longitudinal cooling. In the final stages of cooling, this antidamping is as large as the transverse damping; the 6-D emittance $\varepsilon_t^2$ is roughly constant. In the model, the bunches are lengthened and rf rotated between absorbers to keep $dp/p < ~10\%$. This increases the bunch length from 5cm to $\sigma_0 = 4m$ by end of cooling. The rf frequency decreases correspondingly, from ~200 MHz at start to ~4MHz at the end. RF frequencies < 20 MHz were considered unrealistic and the last five stages required induction linacs.

More recently, Sayed et al. [6] have developed a detailed model of the final cooling system with G4Beamline tracking. There are 16 stages with $p_0$ decreasing from ~135 MeV/c to ~55MeV/c (13 MeV). Each stage consists of a Liquid Hydrogen absorber within a high-field solenoid followed by a drift with rf cavities for phase-energy rotation and reacceleration. (see Fig. 2) Peak magnetic fields are limited to < 32T. The rf is simulated by single frequency cavities (325 to 20 MHz). Some of the stages are followed by field-flips to balance the cooling between transverse degrees of freedom. While each stage cools transversely, the longitudinal anti-damping is larger. 6-D emittance is diluted by a factor of ~3 over the full system. The performance is somewhat less than the baseline goals, as may be expected in a first detailed simulation, and more extreme values in B, $f_0$, and $E_p$ may be needed.

![Figure 2: A cell of final cooling.](image)

**Comments on Baseline**

Particularly toward the end of the final cooling, the baseline scenario uses very high fields and induction linacs, which may be expensive and/or impractical. The deceleration to very low energies increases decay loss and makes capture and reacceleration more difficult. We may truncate the cooling system and use beam phase-space manipulations to achieve the desired luminosities.

---

**Alternative Cooling Systems**

The baseline systems use solenoids for focusing. Recently we are also considering using a quadrupole-based final focusing, with $\beta_t < ~1cm$. (See Fig. 3.) Quad focusing is better at higher energies, and a scenario using 0.8 GeV/c $\mu$'s in a storage ring with Be absorbers is being explored. The goal is to obtain $\varepsilon_t < ~10^{-3}$, while $\varepsilon_e < ~0.004m$. [12]

![Figure 3: $\mu$ trajectories (x and y) through a quad doublet for a $\beta_t = 1cm$. cooling channel.](image)

**CIRCULAR MODES IN SOLENOIDAL COOLING**

The 4D transverse emittance is the product of emittance eigenvalues, and in solenoidal fields the eigenmodes (+ and -) are associated with drift (d) and cyclotron (k) modes, respectively; x and y coordinates are not eigenmodes.[7, 8] The k mode coordinates are:

$$\begin{align*}
\kappa_1 &= \frac{c}{eB} (k_x) \\
\kappa_2 &= \frac{c}{eB} (k_y)
\end{align*}$$

and are simply proportional to the kinetic momentum coordinates. The d coordinates are:

$$\begin{align*}
\xi_1 &= \frac{c}{eB} (d_x) \\
\xi_2 &= \frac{c}{eB} (d_y)
\end{align*}$$

and are proportional to the centers of the Larmor motion, associated with the position coordinates. Within a constant B field the k mode is damped, while the d mode is not. Field flips exchange k and d modes, and can balance the emittances.

Without field flips, solenoidal cooling can develop a large asymmetry between modes. The 4-D emittance is

$$\varepsilon_{4D} = \varepsilon_{T}^2 + \varepsilon_{L}^2 = (\varepsilon_p + L)(\varepsilon_p - L)$$

where 2L is the angular momentum and $\varepsilon_p$ is the projected emittance. Edwards et al.[9] have shown that a skew quad transport can translate $\varepsilon_p$ and $\varepsilon_t$ into $\varepsilon_e$ and $\varepsilon_e$ (decoupled). If $\varepsilon_p$ and $\varepsilon_t$ are very different, a “round” beam is transformed to a “flat” beam. The process has been demonstrated in low-mass $e'$ beams.[10] Cooling of heavy $e'$ beams to $\varepsilon_e/\varepsilon_t \gg 10$ has been simulated.
Figure 4: Round to flat skew quad transport at final cooling parameters.

**FINAL COOLING WITH BUNCH SLICING**

Since this “final cooling” is predominantly an emittance exchange between transverse and longitudinal dimensions, it is possible that similar results could be obtained in a final cooling system that explicitly incorporates emittance exchanges, and avoid the extreme parameters required at the end of the baseline.

An alternative approach to final cooling of this type is envisioned as four stages:

1. **Transverse Cooling.** The beam is cooled transversely within magnetic fields and rf systems that are relatively reasonable: $P_\mu \approx -100\text{MeV}/c$, $B < 30\text{T}$, $f_{RF} > \sim 150\text{ MHz}$. This could be much like the first 4—5 stages of the baseline system. Without field-flips between stages, the cyclotron modes can increase, enabling a round to flat transform. The system cools $\varepsilon_x \rightarrow 10^{-4}\text{m}$, while $\varepsilon_y \rightarrow 0.004\text{m}$.

2. **Round to flat beam transform.** Following the technique developed for the ILC injector and other applications,[9] a solenoid transforms a “round” (large $x$, small $y$) emittance: $\varepsilon_x = 0.0004$, $\varepsilon_y = 0.000025$. (See Fig. 4)

3. **Transverse slicing.** The beam is sliced using multiple passes through “slow-extraction–like” septa into a string of bunches ($\sim 16$). The slices are in the thicker transverse plane, obtaining bunches with enlarged longitudinal emittance: $\varepsilon_x = 0.0000025$, $\varepsilon_y = 0.0000025$. (See Fig. 4)

4. **Longitudinal recombination.** The train of bunches is accelerated to an energy ($\sim 10\text{ GeV}$), where a snap coalescence in a storage ring combines these into a single bunch with enlarged longitudinal emittance ($\varepsilon_x = 25\mu$, $\varepsilon_y = 25\mu$, $\varepsilon_z \approx 0.064\text{m}$).[11]

**Variant Without “Round to Flat”**

Similar manipulations are possible without use of the “round to flat” process. The sequence could be:

1. **Transverse Cooling.** A cooling system to minimize emittances within reasonable fields is used. It should cool $\varepsilon_x$ and $\varepsilon_y$ to $10^{-4}\text{m}$, while $\varepsilon_z \rightarrow -0.004\text{m}$.

2. **Transverse slicing.** The beam is sliced using multiple passes through a “slow-extraction–like” septum into a string of bunches ($\sim 10$). The slices are in one plane, obtaining bunches with asymmetric emittances: $\varepsilon_x = 10\mu$, $\varepsilon_y = 100\mu$.

3. **Longitudinal recombination.** The bunches are accelerated into a ring that combines them into a single bunch ($\varepsilon_x = 10\mu$, $\varepsilon_y = 100\mu$, $\varepsilon_z \approx -0.004\text{m}$).

4. **The beams accelerate and collide as flat beams.** Collisions of $\varepsilon_x = 10\mu$, $\varepsilon_y = 100\mu$ could be matched in luminosity to $\varepsilon_x(\varepsilon_y)^{1/2} = -30\mu$ round beams. Flat beam collisions have some advantages. Chromaticity correction is much easier, and detector shielding could be simpler. However, luminosity may be decreased by the “hour glass” effect, if $\beta_z^* \ll$ bunch length.

A thick wedge absorber could also obtain a very small $\varepsilon_x$ with enlarged $\varepsilon_L$ (step 2). The enlarged $\varepsilon_L$ could be single-bunch or multi-bunch in acceleration. [13]

**CONCLUSION**

Within these variations that we have discussed and extensions, we believe R&D will find credible and affordable solutions for the final cooling needed for a high energy, high luminosity next generation lepton collider.

**ACKNOWLEDGMENTS**

We thank R. Palmer, M. Palmer, and Y. Alexahin, for important helpful and original contributions.

**REFERENCES**