# LECTURE 19

## Beam cooling

Stochastic cooling

Electron cooling

Ionization cooling



## Stochastic cooling

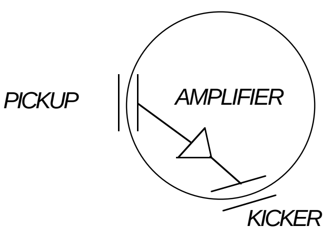
Stochastic cooling is a method for increasing the phase space density of a particle beam. It is usually applied to ion beams, rather than electron beams, as the damping times are relatively long, and cannot compete with radiation damping for high energy electrons.

The most extensive use of this technique has been in the collection and storage of *antiproton beams*. These beams are produced with a very low density, by high-energy proton bombardment of a heavy target. Before they can be used in a proton-antiproton collider, the beam phase space density must be increased by about 6 orders of magnitude. This increase is accomplished through stochastic cooling.

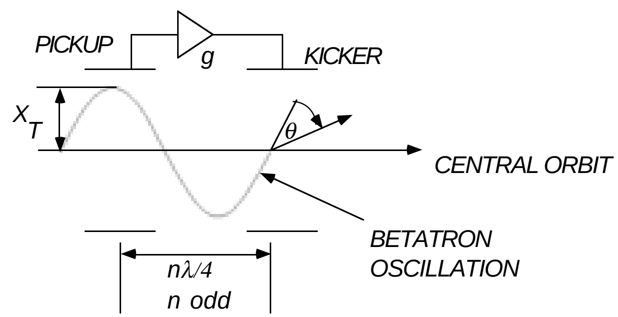


Simon VanderMeer at CERN invented the technique in the late 1970's. The required technology was developed there, and applied to the CERN proton-antiproton collider. Subsequently, Fermilab built its own antiproton source, which further developed and refined this technique. The technique is applied to increase both the transverse and the longitudinal density of the beam.

**Transverse stochastic cooling:** Conceptual system:






"Single-particle" cooling: let there be only one particle in the ring, with a betatron oscillation as shown

$x_T$ = deviation of the particle from the dipole pickup center on turn T

$x_{T+1}$ = deviation of the particle from the dipole pickup center on turn T+1

The stochastic cooling system measures $x_T$ at the pickup, and delivers a kick (angular change θ) to the particle at the kicker such that

$$x_{T+1} = x_T - g x_T$$

where $g x_T$= effect of kicker at pickup; $g$ is adjustable through the amplifier gain.

*g=1 removes the oscillation completely. The transverse momentum is reduced to zero: The particle is "cooled" in one turn*



Now consider an unbunched beam of $N$ particles in the ring

The cooling system treats as a single "sample" all beam particles that pass under the pickup within the system time resolution

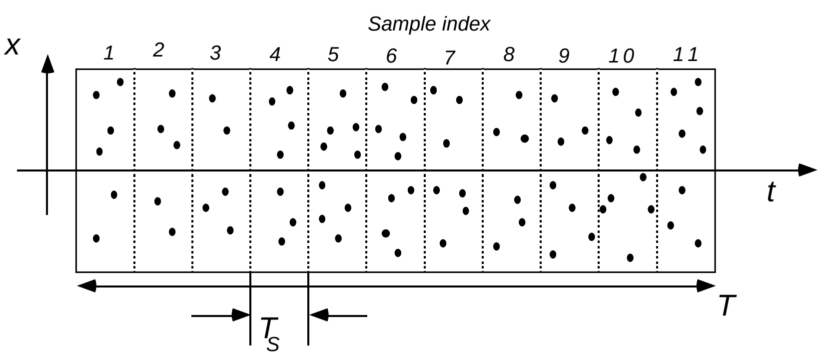
$T_S << T$, the revolution period.

$T_S$ is related to the system bandwidth $W$

$$T_s = \frac{1}{2W}$$

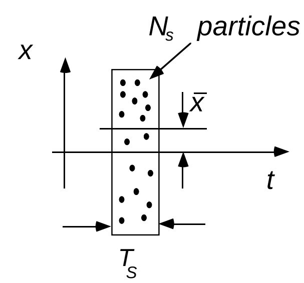
Essentially, the beam is "divided" by the signal processing system into $S$ samples, where $S = \frac{T}{T_S}$. Each sample contains $N_s = \frac{N}{S}$ particles.








A single sample:




Although the entire beam may have a mean position very close to zero, a random subset of the beam containing $N_S$ particles will in general have a mean

$$\bar{x} = \frac{1}{N_s} \sum_{j=1}^{N_s} x_j$$

that may be non-zero, *purely as a result of statistical fluctuations.* Averaged over many samples, the mean is zero $\langle \bar{x} \rangle = 0$, but the variance in the mean is $\langle \bar{x}^2 \rangle = \frac{\sigma_x^2}{N_s}$, where $\sigma_x$ is the rms size of the beam.

The cooling system measures the mean and corrects all particles in the sample:

For particle *i* in the sample, the change in one turn is

$$x_i \to x_i - g\bar{x}$$



$$x_i^2 \to (x_i - g\bar{x})^2 = x_i^2 - 2gx_i\bar{x} + g^2\bar{x}^2$$
$$\Delta x_i^2 = -2gx_i\bar{x} + g^2\bar{x}^2$$

To see what this means, re-write

$$\bar{x} = \frac{1}{N_s}\left( x_i + \sum_{\substack{j=1 \\ j\neq i}}^{N_s-1} x_j \right) = \frac{x_i}{N_s} + \frac{N_s-1}{N_s}\bar{x}^*$$

where $\bar{x}^* = \frac{1}{N_s-1} \sum_{\substack{j=1 \\ j\neq i}}^{N_s-1} x_j$ is the contribution to the mean due to all the particles except the *i*th one. Substitute:



$$\Delta x_i^2 = \frac{-2gx_i}{N_s}\left(x_i + (N_s-1)\bar{x}^*\right) + \frac{g^2}{N_s^2}\left(x_i + (N_s-1)\bar{x}^*\right)^2$$

This is what happens on one turn. On subsequent turns, let us *assume* that the samples containing particle *i* are *statistically independent.* This assumption will not be true in general, and we'll have to go back and correct this analysis later, but making the assumption now makes it easier to understand what's going on.

On each turn, then, a new sample is processed, and averaging over many turns is equivalent to averaging over a collection of statistically independent samples. In doing such an average, we have $\langle \bar{x}^* \rangle = 0$, $\left\langle \left(\bar{x}^*\right)^2 \right\rangle = \langle \bar{x}^2 \rangle = \frac{\sigma_x^2}{N_s}$, so we get



$$\Delta x_i^2 = \frac{-2gx_i^2}{N_s} + \frac{g^2 x_i^2}{N_s^2} + \frac{g^2(N_s-1)^2}{N_s^3}\sigma_x^2$$

The first two terms correspond to "single-particle cooling": the net result is $\Delta x_i^2 = -x_i^2$ for $g$=1 and $N_s$=1 which is the same result we got before.

*The last term is due to the other particles and "heats" particle i*

*This "noise" due to other particles is called "Schottky noise"*

Now average the above result over *all the particles within a given sample*, with the (excellent) assumption that $N_s$>>1

$$\left\langle \Delta x^2 \right\rangle = \frac{-2g}{N_s}\left\langle x^2 \right\rangle + \frac{g^2}{N_s}\sigma_x^2 \Rightarrow$$

$$\Delta\sigma_x^2 = \left(-2g + g^2\right)\frac{\sigma_x^2}{N_s}$$

The beam mean square size changes with a rate

$$\frac{1}{\tau_{x^2}} = -\frac{1}{T}\frac{\Delta\sigma_x^2}{\sigma_x^2} = \frac{1}{N_s T}(2g - g^2)$$

Since

$$\frac{N_s}{N} = \frac{T_s}{T} = \frac{1}{2WT} \Rightarrow \frac{1}{\tau_{x^2}} = \frac{2W}{N}(2g - g^2)$$

The cooling rate for the beam size is



$$\frac{1}{\tau_x} = \frac{1}{2\tau_{x^2}} = \frac{W}{N}(2g - g^2)$$

The maximum cooling rate is obtained for $g$=1:

$$\frac{1}{\tau_{x,\max}} = \frac{W}{N}$$

For example, 1 GHz of bandwidth cools $10^9$ particles at a rate of 1 Hz. The bandwidth dependence of the cooling rate motivates the use of as high a frequency signal processing system as possible. Modern systems operate in the range from 2 to 8 GHz.

Another way of writing the cooling rate

$$N_s = \frac{Nf}{2W},\ \frac{1}{\tau_{x,\max}} = \frac{f}{2N_s}$$



shows that the smaller the number of particles in the sample, and the higher the revolution frequency, the higher the cooling rate.

Two additional effects modify the cooling rate equation:

1. Thermal noise in the pickup/amplifier/kicker system introduces an additional heating term. This is described by adding a term $-g^2\frac{UW}{N}$ to $\frac{1}{\tau}$,

where $U$= thermal noise power/"Schottky noise" power



2. Statistical independence of the samples from turn to turn has been assumed. Sample renewal from turn to turn is called “mixing”.

*If there is no mixing, then after the first correction, cooling stops, since the means of all the samples are zero.*

The mechanism for mixing is the variation per turn $\Delta T$ in revolution periods for different beam particles resulting from the momentum spread within the beam. It takes $M$ turns for complete sample renewal, where

$$M = \frac{T_s}{\Delta T} = \frac{1}{2W\Delta T}$$

To find an expression for $\Delta T$, consider those particles in the sample, which all arrive at the pickup at the same time, but with a momentum spread $\delta$. The revolution period of these particles depends on their momenta through

$$\frac{T}{T_0} = 1 - \eta_C \delta$$

in which $T_0$ is the revolution period for a particle with $\delta$=0. On the next turn, these particles will arrive at the center of the pickup over a time spread

$$\Delta T = T_0 \eta_C \delta.$$

The number of turns required for this effect to cause complete sample renewal is



$$M = \frac{1}{2W\Delta T} = \frac{f_0}{2W\eta_C \delta}$$

This effect increases the "Schottky noise" heating term by the factor $M$.

Inclusion of these effects leads to the following equation for the damping rate

$$\frac{1}{\tau_x} = \frac{W}{N}\left(2g - g^2[M+U]\right)$$

Optimum cooling happens when $g = \frac{1}{M+U}$ and for this condition



$$\frac{1}{\tau_{x,\max}} = \frac{W}{N}\frac{1}{M+U}$$

Cooling is still possible, even for very large $M$ and $U$, although rate is reduced.

Example: Fermilab Debuncher ring.

This is a storage ring with a revolution period of 1.7 μs. It collects 8 GeV antiprotons, produced in high energy proton interactions with a copper target, and stochastically cools them for about 2 s before transferring them to an accumulation ring. The cooling system operates over the 2-4 GHz band, and the ring collects about $2\text{x}10^7$ antiprotons on every cycle, with a momentum spread (after the beam is debunched) of about $\delta$=0.002. The ring has a slip factor of $\eta_C$=0.006.



The mixing parameter $M$ is

$$M = \frac{1}{2WT_0\eta_C\delta} = \frac{1}{2\times 2\times 10^9 \times 1.7\times 10^{-6} \times 0.006 \times 0.002} = 12$$

It thus takes about 12 turns for sample renewal. The microwave pickup system for this ring has a noise figure of $U$=2 (i.e., the amplifier noise power is twice that due to the beam signal). The optimum value for the cooling rate is then

$$\frac{1}{\tau_x} = \frac{W}{N}\left(\frac{1}{M+U}\right) = \frac{2\times 10^9}{2\times 10^7}\left(\frac{1}{12+2}\right) = 1.4 \text{ Hz}$$

In 2 seconds, then, the beam size will be reduced by about a factor of $\exp(2\times 1.4)$~16

Actually, the parameter $U$ is not constant, because as the beam cools the Schottky noise is reduced, so $U$ increases. Eventually, amplifier noise plus Schottky noise heating balances cooling, and the beam reaches equilibrium size.

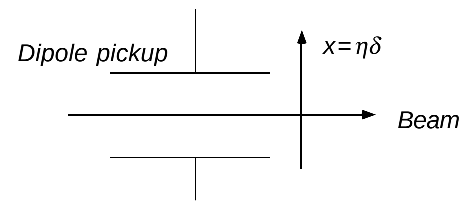
## Longitudinal stochastic cooling

Similar to transverse--except that the pickup definition of "center" is a little trickier:

Techniques to establish a "central energy":

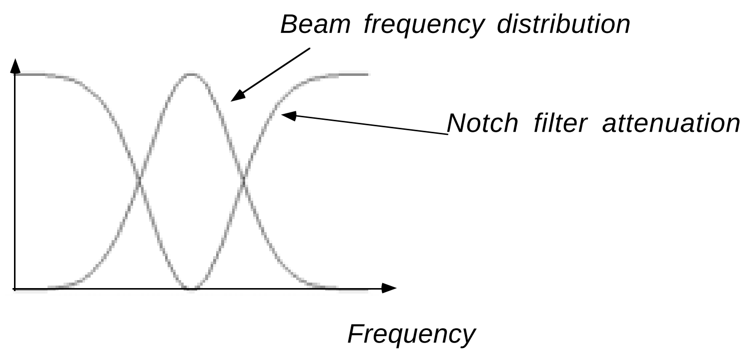
(1) a dipole (position-sensitive) pickup located in a region of the machine in which there is a correlation between the beam's position and its momentum






(2) a longitudinal ("sum") pickup that measures the frequency distribution of the beam, together with an electronic "notch" filter, which provides a frequency-dependent gain.



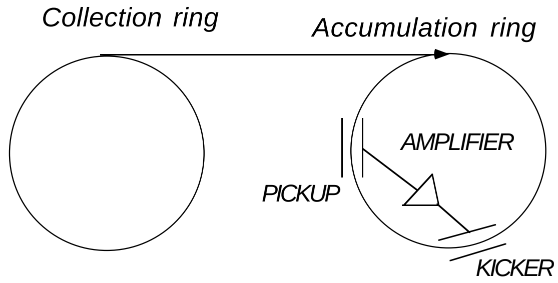



$$\frac{\Delta f}{f} = -\eta_C \frac{\Delta p}{p}$$



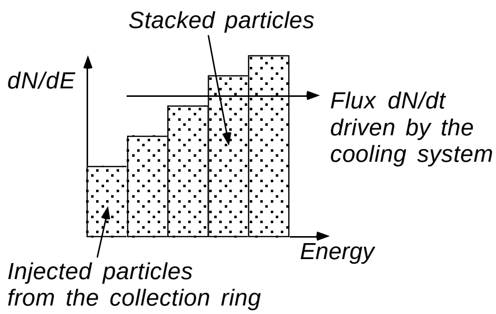
## Stochastic stacking

-a form of longitudinal stochastic cooling. A density gain factor of about $10^5$ is realized by a system like this in the Fermilab Accumulator. Basic idea:



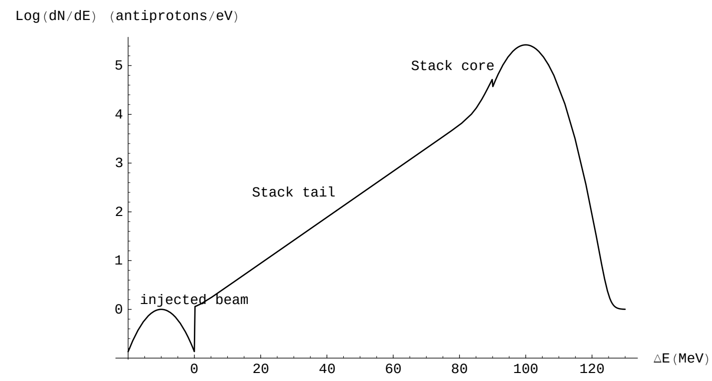
Pickup measures the energy of a sample $\Delta N$ of particles at energy $E - \Delta E_k$;

Kicker changes the sample's energy by $\Delta E_k$; $\Delta E_k$ is adjusted to produce a particle flux $\frac{dN}{dt}$ along the energy axis






*Example: Fermilab Accumulator ring stack-tail system*

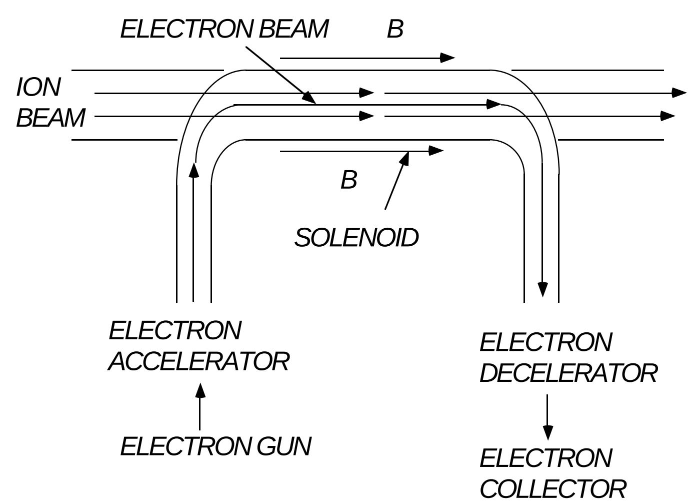





## Electron Cooling

This is a method of cooling of ion beams (such as proton and antiproton beams) in which a “cold”electron beam is brought into contact with a “hot” ion beam. The beams exchange energy through the Coulomb interaction, with the “hot” beam getting colder and the electron beam warming up. The electron beam is then disposed of, and a new “cold” beam supplied to continue the cooling process. The necessary arrangement for overlap of the beams is shown on the next page.



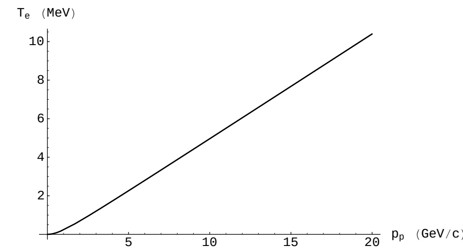


The mechanism for energy exchange between the electrons and the ions is the same as the one responsible for ionization energy loss. The strength of this interaction is maximum when the ions and the electrons are at rest with respect to each other: for two beams, this requires that the beams have equal velocities:

$$\frac{p_e}{m_e c} = (\beta\gamma)_e = \frac{p_{ion}}{m_{ion} c} = (\beta\gamma)_{ion}$$

For example, the following plot shows the kinetic energy ($T_e$) of the electron beam needed to electron cool a proton (or antiproton) beam of momentum $p_p$






Substantial electron energies (MeV) and currents (amps) are required for high energy electron cooling, which is why it is important to recover essentially all of the electron beam energy in the collector.

In the common beam rest frame, the ion and electron beams appear as a plasma that is far from thermal equilibrium. The energy



exchange mechanism between the hot and cold components acts like a drag force on the ions, which, like ionization energy loss, varies as $\frac{1}{(v^*)^2}$ ($v^*$ is the relative electron-ion velocity in the rest frame). Thus, it is not very effective unless there is substantial overlap between the velocity distributions of the two components. This means that electron cooling is least suitable for large phase space beams, or high energy beams, which have large velocity spreads.

The energy spread in the electron beam is determined (ideally) by the temperature of the cathode; typical energy spreads are in the range of 0.5 eV, leading to a velocity spread in the rest frame



$$\delta\beta_e^* = \sqrt{\frac{2\delta T_e^*}{m_e c^2}} = \sqrt{\frac{2\times 0.5}{511{,}000}} \approx 1.4\times 10^{-3}$$

Let’s compare this with the velocity spread of the ion beam.

Longitudinal cooling:

If the ion longitudinal momentum in the lab frame is $p$, this momentum is related to the rest frame momentum by

$$p = \gamma\left(p^* - \frac{\beta}{c}E^*\right) \approx \gamma\left(p^* - \beta mc\right)$$

in which the rest frame momentum is $p^*$, $m$ is the ion mass, and $\beta$ and $\gamma$ refer to the beam velocity in the lab. A momentum spread

$\delta p$ in the lab appears in the rest frame as a momentum spread $\delta p^* = \frac{\delta p}{\gamma}$, leading to a longitudinal velocity spread

$$\delta\beta_{\|}^* = \frac{\delta p^*}{mc} = \frac{\delta p}{\gamma mc} = \frac{p}{\gamma mc}\frac{\delta p}{p} = \beta\frac{\delta p}{p}$$

For efficient cooling, we want

$$\delta\beta_{\|}^* = \beta\frac{\delta p}{p} \approx \delta\beta_e^* \approx 1.4\times10^{-3} \Rightarrow \frac{\delta p}{p} \approx \frac{1.4\times10^{-3}}{\beta}$$

For nonrelativistic ion beams with small $\beta$, large momentum spreads of a few percent or more can be cooled: but for high energy ion beams with $\beta$~1, energy spreads of more than a few tenths of a percent will not be efficiently cooled.



Transverse cooling:

In the laboratory frame, the spread in transverse momentum is related to the beam divergence $z'$ (where $z$ refers to either $x$ or $y$):

$$\frac{\delta p_z}{p} \approx z'$$

The transverse momentum is the same in the rest frame $\delta p_z^* \approx z'p$, so the spread in transverse velocity is

$$\delta\beta_{\perp}^* = \frac{\delta p_z^*}{mc} = \frac{z'p}{mc} = \beta\gamma z'$$

For efficient cooling we need

$$\delta\beta_{\perp}^* = \beta\gamma z' \approx \delta\beta_e^* \approx 1.4\times10^{-3} \Rightarrow z' \approx \frac{1.4\times10^{-3}}{\beta\gamma}$$



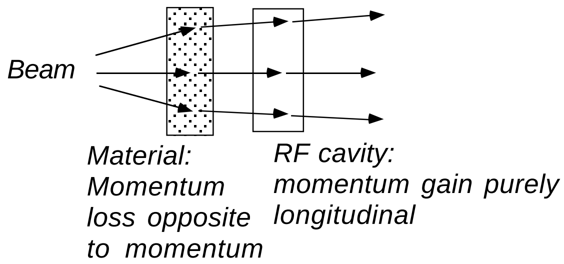
Again, for nonrelativistic ion beams with small $\beta$ and $\gamma$=1, large angular spreads of 10 mrad or more can be cooled: but for high energy ion beams with $\beta$~1 and large $\gamma$, the situation is even worse than in the longitudinal plane. For example, for $\beta\gamma$=10 (i.e., a 10 GeV proton beam), cooling is only effective for $z' \approx 0.14$ mrad, which corresponds to a beam which is already pretty dense.

The cooling rate is proportional to the electron density and is independent of the ion density: hence, electron cooling is most appropriate for enhancing the density of relatively cool, low energy, intense ion beams.



## Ionization Cooling

This is a cooling method that makes use of the ionization energy loss experienced by a particle beam when traversing matter. This energy loss reduces both the transverse and longitudinal components of the momentum of the particle. The longitudinal component is then restored by an rf system; the net result is a reduction of the emittance of the beam.




This scheme is only practical for weakly interacting, high mass particles such as muons, which do not suffer from either nuclear interactions or bremsstrahlung in the material. In addition to the ionization energy loss which provides the cooling mechanism, multiple Coulomb scattering will also take place, which is a random process which produces heating.



The emittance growth due to heating from multiple Coulomb scattering follows from the discussions in Lecture 20, p. 33:

$$\frac{d\varepsilon}{ds} = \frac{\beta_z}{2}\frac{d\langle\theta_{mcs}^2\rangle}{ds}$$

The emittance reduction due to ionization energy loss is a reduction in the normalized emittance:

$$\frac{d\varepsilon_n}{ds} = \frac{d}{ds}(\beta\gamma\varepsilon) = \varepsilon\frac{d}{ds}(\beta\gamma) + \beta\gamma\frac{d\varepsilon}{ds}$$

Using $\frac{d}{ds}(\beta\gamma) = -\frac{\beta\gamma}{\beta^2 E}\frac{dE}{ds}$, and the heating term from above, gives

$$\frac{d\varepsilon_n}{ds} = -\frac{\varepsilon_n}{\beta^2 E}\frac{dE}{ds} + \beta\gamma\frac{\beta_z}{2}\frac{d\langle\theta_{mcs}^2\rangle}{ds}$$



From Lecture 20, pg. 34,

$$\frac{d\langle\theta_{mcs}^2\rangle}{ds} = \frac{1}{X_0}\left(\frac{13.6\ \text{MeV}}{\beta pc}\right)^2$$

So we have

$$\frac{d\varepsilon_n}{ds} = -\frac{\varepsilon_n}{\beta^2 E}\frac{dE}{ds} + \beta\gamma\frac{\beta_z}{2}\frac{1}{X_0}\left(\frac{13.6\ \text{MeV}}{\beta pc}\right)^2$$

$$= -\frac{\varepsilon_n}{\beta^2 E}\frac{dE}{ds} + \frac{\beta_z}{2X_0}\frac{(13.6\ \text{MeV})^2}{\beta^3 Emc^2}$$

A balance between the heating and cooling terms will eventually be reached, resulting in the minimum value of the normalized emittance



$$\varepsilon_{n,\min} = \frac{\beta_z}{2X_0\frac{dE}{ds}}\frac{(13.6\ \text{MeV})^2}{\beta mc^2}$$

To achieve the smallest minimum emittance with an ionization cooling system, we want to have $\beta_z$ (the lattice function) small, $X_0$ (the material's radiation length) large (which means a very low density, low Z material), and *dE/ds* large (which is contradictory to the previous requirement). The best compromise is a low Z material of intermediate density: liquid hydrogen. Very strong focusing is favored to get the smallest possible $\beta_z$.