STOCHASTIC COOLING THEORY AND DEVICES

S. Van der Meer
CERN

I. Principle of Stochastic Cooling

Stochastic cooling uses the spontaneous violation of Liouville's theorem that is always present in a beam with a finite number of particles. The system detects the corresponding fluctuations and acts on the beam (in a strictly Liouvillian way) so that the random density variations do not get a chance of averaging out.

The most efficient cooling would be obtained if each individual particle could be observed separately. In practice, this is quite impossible. Even with the fastest systems proposed up to now, the resolution is of the order of 10^5 particles.

In general, the feedback system detects one parameter of the particle motion (transverse position or phase) and acts on another one (transverse or longitudinal momentum). The analysis is often easiest by considering two effects occurring simultaneously:

a) Coherent effect: each particle is influenced coherently by its own signal.

b) Incoherent effect: blowup is caused by the signals from the other particles ("Schottky noise") or by the amplifier noise.

The coherent effect is proportional with the system gain , the incoherent one with its square. Therefore, it is always possible to choose a gain where the coherent one is predominant. By a proper choice of parameters this will result in cooling.

II. Mixing

The incoherent effect caused by noise depends on the noise spectrum. The particles will only be influenced by the noise frequencies that coincide with harmonics of their revolution frequency (for momentum cooling) or with one of the betatron sidebands (for betatron cooling). This is strictly true only if the particle frequencies are constant; any other frequency will then only cause a beating effect that does not increase with time. In practice, the frequencies change so slowly that it is still true.[1] The noise power density vs. frequency at each of the particle's harmonics is therefore the quantity on which the blowup depends.

The Schottky noise (from the other particles) covers certain frequency regions, the Schottky bands, that also contain the frequencies to which the perturbed particle is sensitive. The power density clearly depends on the frequency spread covered by these bands: the wider this is, the less power density one has. Also, since the width of these bands increases with the harmonic number, higher harmonics contribute less to the incoherent effect.

It often happens that within the bandwidth of the electronic system these bands are separated everywhere and do not overlap. In that case, each of the sensitive frequencies of the perturbed particle is inside a single Schottky band. This situation is often called "bad mixing."

Alternatively, the revolution frequency spread (or the harmonic numbers used) may be so high that the bands overlap and that each particle frequency is inside many different overlapping Schottky bands. This is called "good mixing." Intermediate situations may, of course, also exist.

Seen in the time domain, the signal (or "pulse") caused by a single particle will influence many other particles as well. If this sample of other particles changes its population from one revolution to the next because the revolution time spread is much larger than the pulse duration, we have good mixing. For the opposite case, the sample population changes only slowly. The incoherent effect is then also worse, because the perturbations from the same particle are correlated over more than one turn.

With good mixing, the incoherent effect depends on the total number of particles. Higher harmonic numbers contribute as much as lower ones, because, although the power density is lower, more harmonics overlap there. With bad mixing, the particle density vs. frequency at the revolution frequency of the perturbed particle is important. Higher harmonics are less important than lower ones. Especially in the case of momentum cooling, the resulting equations are then different in character, because the momentum cooling itself increases the density.

In practical cases (e.g., the cooling in the CERN $\bar{p}$ accumulator) the mixing is often bad. In the following analysis of momentum cooling, we shall assume this.

III. Momentum Cooling

We shall first assume that a beam pickup and a longitudinal kicker are used (Fig. 1). Each particle induces a pulse in the pickup that produces a pulse at the kicker . The density distribution is governed by the diffusion equation

$$\frac{\partial\psi}{\partial t} = -\frac{\partial}{\partial E}(F\psi) + \frac{\partial}{\partial E}\left(D\,\frac{\partial\psi}{\partial E}\right), \qquad (1)$$

where ψ = particle density dN/dE
E = particle energy
t = time
F = coherent acceleration rate $d\bar{E}/dt$
D = diffusion constant (incoherent term) $= 1/2\ (\overline{dE^2}/dt)$.

It is convenient to express F and D as a function of the complex system gain G_n at harmonic number n:

$$F = \frac{d\bar{E}}{dt} = 2ef_oR\sum_n \mathrm{Re}(G_n) \qquad (2)$$

$$D = D_1 + D_2\psi \qquad (3)$$

$$D_1 = 2KTf_oR_K \sum_n |G_n|^2 \quad (4)$$

(noise from an amplifier with a 3 dB noise figure)

$$D_2 = 2e^2 f_o{}^3 R^2 \frac{dE}{df} \sum_n \frac{|G_n|^2}{n} \quad (5)$$

(Schottky noise).

The sums are extended over all harmonics of the revolution frequency f_o. The gain G is measured between the amplifier input and output (supposed to have equal impedance).

$$R = \sqrt{n_p R_p n_K R_K}$$

n_p, n_K, R_p, R_K = number and impedance of pickups, kickers

e, K, T = electron charge, Boltzmann constant, room temperature

Of course, around each harmonic number, G, and therefore F and D, may still vary with E. In fact, if F is independent of E, there is no cooling, but only a steady acceleration or deceleration, added to the blowup from the first term. Cooling will result if the coherent effect moves the particles into a direction where F decreases, so that they pile up there. This can be done in two ways:

a) by using a pickup whose sensitivity depends on position and therefore (if placed in a point with non-zero dispersion) on E. This method was first proposed by Palmer.[2]

b) by placing a filter in the feedback loop whose gain depends on frequency in the required way around each harmonic of the revolution frequency. Such a filter was proposed by Thorndahl.[3]

Equation (1) to (5) neglect the effect of feedback from the kicker via the beam towards the pickup. This is usually justified; a more complete theory where this is taken into account is being developed by F. Sacherer.

It is not easy to find solutions for the diffusion equation (1). Even if G depends linearly on E at each harmonic, an analytical solution seems impossible because of the dependence of D on ψ. It is therefore necessary to solve each particular case by numerical integration.

In practice, it is often possible to make the first term of (3) smaller than the second one. This means that the cooling rate is limited by the Schottky noise rather than by the amplifier noise, whose density is then below the Schottky noise density at the Schottky frequencies. Since, however, the amplifier noise is also present between the Schottky bands, it will normally give the largest contribution to the output power required. The available wide-band output power may restrict the cooling rate that can be obtained.

IV. Scaling

It is interesting to compare different cooling systems installed in different rings and starting with different initial distributions. We assume:

a) the initial distributions have the same shape.

b) the amplifier noise is negligible from the point of view of cooling.

c) the cooling is not limited by the available amplifier power.

d) the two systems compared have a similar frequency response (although the frequency scale may be different).

e) in both cases the gain is adjusted to the optimum value.

f) the variation of G with E is the same for both systems if scaled to the width of of the initial particle distribution.

It can then be shown that the time scales as $N/(n_\ell{}^2 \Delta f)$, with

N = total number of particles

n_ℓ = number of revolution frequency harmonics within the passband

Δf = initial spread of revolution frequency.

We may also express this in machine parameters and find then a scaling factor $Nf_o/[W^2|\eta|(\Delta p/p)]$, with

W = system bandwidth

$\eta = (\Delta f/f)/(\Delta p/p)$.

V. Use of Filters

If the dependence of the coherent factor F on the energy E is to be achieved by using the relationship between E and revolution frequency, we need filters that perform in a similar way around each harmonic of the revolution frequency. Such filters may be built using as elements transmission lines with a length equal to half the ring circumference. These lines may be either open or shorted at the far end. They then have an impedance

$$Z = jZ_L \tan(\pi f/f_o) \quad \text{for a shorted line}$$
$$\text{or} \quad Z = -jZ_L \cot(\pi f/f_o) \quad \text{for an open line} \quad (6)$$

Therefore, the response of filters made with such elements is the same around each harmonic of f_o. Shorted or open lines behave like inductances or capacitances, respectively for positive $\Delta f/f_o$.

Since the width of the Schottky bands increases with the harmonic number, this behavior is not quite ideal. However, by combining these lines with lumped elements, filters may be made that give nearly the same characteristic vs. E at each harmonic. An example is given in Ref. 4.

A simple filter may be made as shown in Fig. 2. This filter has zero transmission at each harmonic of the revolution frequency corresponding to a given momentum value; in the neighborhood of these zeros, the transmission varies linearly with frequency. For a limited frequency range, this filter therefore behaves like a linear pickup, except that it is not sensitive to betatron oscillations.

Figure 3 shows Schottky scans obtained with momentum cooling, using such a filter (CERN ICE experiment). The density is proportional to the square of the vertical coordinate; the horizontal scale corresponds to the revolution frequency (or momentum).

The advantage of using filters instead of position-sensitive pickups to make F depend on E is that wide-band sum pickups may be made much shorter than position-sensitive ones. Also, the filter will have minimum gain at the frequencies where the particles will accumulate; it will therefore also diminish the influence of the amplifier noise on the cooled particles.

VI. Momentum Stacking

In the CERN $\bar{p}$ accumulator ring each antiproton pulse will be precooled by a momentum cooling system using the filter method. The particles will then be captured by a normal rf system and deposited at the top of a stack. This stack must be constantly cooled so that space is made free for the next pulse. The total momentum spread of the stack will then remain constant; its density will increase. Particles will migrate towards the bottom of the stack, where they will pile up. Clearly, the system gain at the bottom will have to be much lower than at the top. The optimum gain profile may be found by requiring the steepest possible density increase from top to bottom of the stack, while still maintaining a constant flux of particles migrating against this slope towards the bottom. This flux is

$$\phi = \frac{dN}{dt} = F\psi - D\,\frac{d\psi}{dE}. \qquad (6)$$

For simplicity, we now assume that the gain G is real (i.e., we assume perfect phase at all frequencies of interest). We also neglect the amplifier noise. Then $D = c_1\alpha^2\psi$ and $F = c_2\alpha$, where α is proportional to the system gain. Equating the flux to the required ϕ_o, and adjusting α so that $d\psi/dE$ becomes as steep as possible, we find

$$\alpha = 2\phi_o/(\psi c_2), \qquad (7)$$

i.e., the gain should be inversely proportional with density. The resulting optimum density profile is

$$\psi = \psi_o \exp\,[(E_o - E)/E_d], \qquad (8)$$

where E_o and ψ_o refer to the top of the stack, and

$$E_d = -4c_1\phi_o/c_2^{\,2}. \qquad (9)$$

This quantity determines the density gradient that may be obtained.

In practice, these expressions are modified by many detailed considerations, such as amplifier noise and imperfect phase. Still, the optimum stack profile found for the practical case of the $\bar{p}$ accumulator, where these effects were taken into account, is not dissimilar to Eq. (8). Figure 4 shows this profile and how it develops with time during stacking. The sudden increase in slope near the stack bottom is caused by the use of a feedback system with higher bandwidth in that region.

In fact, the very large density ratio between the top and bottom of the stack necessitates a corresponding gain ratio. This dependence of gain on energy will be obtained by the use of position-sensitive pickups in combination with filters. Three overlapping feedback systems are at present foreseen. Noise filters will be used to prevent that the high-gain systems for the top of the stack will produce too much blowup at frequencies corresponding to the bottom. A more detailed description is given in Ref. 4.

VII. Betatron Cooling

There are two important differences between momentum cooling and betatron cooling:

a) For momentum cooling, the filter method is possible because the frequency of the pickup signal is related to momentum. The dependence of frequency on betatron amplitude, on the other hand, is weak.

b) Mixing is connected with momentum spread. Therefore momentum cooling reduces the mixing, whereas betatron cooling does not.

Because the mixing is constant, Gaussian distributions will remain Gaussian, which simplifies the theory. However, no detailed analysis including the mixing in an exact way is available at present. We shall make the following simplifying assumptions:

a) The mixing is bad,

b) The feedback system has constant gain with zero phase shift over a bandwidth W.

c) The momentum distribution is square, with a total revolution frequency spread $\varepsilon = \Delta f/f_o$.

Because of the last assumption, the bad mixing will cause an increase in Schottky power density[1] by a factor $\Lambda f_o/\varepsilon W$, with

$$\Lambda = \sum_n 1/n,$$

summed over all harmonics of the revolution frequency within the passband.

The cooling rate then is (as in Ref. 5, but corrected for bad mixing)

$$\frac{1}{\tau} = \frac{W}{2N}\left\{2g - g^2\left(\frac{\Lambda f_o}{\varepsilon W} + \eta^*\right)\right\}, \qquad (10)$$

where g is the gain relative to the optimum gain for good mixing and zero amplifier noise, whereas η^* is equal to the ratio of amplifier noise to signal power.

The optimum value for g gives

$$\frac{1}{\tau} = \frac{W}{2N}\cdot\frac{1}{\dfrac{\Lambda f_o}{\varepsilon W} + \eta^*}\,. \qquad (11)$$

As the cooling proceeds, η^* increases because the signal power decreases. Therefore, even if g is continuously adjusted to keep track of this, the cooling rate will decrease.

In the CERN $\bar{p}$ cooling ring, betatron cooling will be done on the stack, so that the cooling rate need not be high. It will be limited mainly by the bad mixing; because of this, amplifier noise will not be a problem.

VIII. Pickups and Kickers

Wideband pickups and kickers used at present for stochastic cooling are of three types:

a) sum pickups or kickers with ferrite rings

b) transverse pickups or kickers formed of $\lambda/4$ directional couplers

c) high frequency devices of traveling wave type with coupling slots.

Sum pickups or kickers with ferrite rings surrounding the beam are used for momentum cooling. It is usually found that for practical momentum cooling systems the output power needed is important. Since it can be decreased by using many kicker gaps, the length of these gaps should be as small as possible. For instance, for the $\bar{p}$ accumulator we plan to use kickers containing 100 or 200 gaps. Since most of the output power is due to amplifier noise, we also must use a large number of pickup gaps, increasing the signal so that the gain may be reduced and the amplifier noise power decreased.

For the same reason, the gap impedance should be high. This is, of course, the reason why a ferrite ring is used. Unfortunately, at high frequencies (a few hundred MHz) the best available ferrites have low permeability and high losses. Therefore, it is doubtful if much more than 50 Ω/gap, as in present structures, may be reached. The power dissipation and cooling of the ferrite in the kickers is also a factor to be taken into account. It may limit the output power even more than the availability and cost of high power wide-band amplifiers.

Transverse pickups and kickers are necessarily much longer. Typically, their length should be about $\lambda/4$ in the middle of the passband, so that they have a reasonable impedance throughout. Because of this length (e.g., 25 cm for a bandwidth of 200-400 MHz), it is usually difficult to find space for a great number of transverse pickups. The signal-to-noise ratio therefore tends to be low. As a consequence, betatron cooling of low intensity beams is slower than momentum cooling.

For frequencies above 1 GHz, where ferrite cannot be used any more, Faltin[6] has developed a wide-band pickup (or kicker) structure that essentially consists of a metal box around the beam with transmission lines arranged above and below it. Slots in the top and bottom of the box couple the beam to the waves traveling along these lines (see Fig. 5). The same structures may be used as a sum or transverse pickup by adding the signals on the lines in phase or with a 180° phase shift. Such structures have been successfully used at CERN both for betatron cooling and for stochastic acceleration.

References

[1] S. van der Meer, Influence of Bad Mixing on Stochastic Acceleration, CERN SPS/DI/PP/Int. Note 77-8.

[2] R. B. Palmer, Brookhaven National Laboratory, private communication (1975).

[3] G. Carron and L. Thorndahl, Stochastic Cooling of Momentum Spread by Filter Techniques in the Cooling Ring, CERN Technical Note ISR-RF/LT/PS, January 1977.

[4] Design Study of a Proton-Antiproton Colliding Beam Facility, CERN/PS/AA 78-3.

[5] H. G. Hereward, Statistical Phenomena-Theory, Proc. of the 1st Course of the International School of Particle Accelerators, Erice, 1976; CERN 77-13, p. 281.

[6] L. Faltin, Dissertation, T-U, Vienna, 1977. L. Faltin, Slot-Type Pick-Up and Kicker for Stochastic Beam Cooling, Nucl. Instr. and Methods 148, 449 (1978).

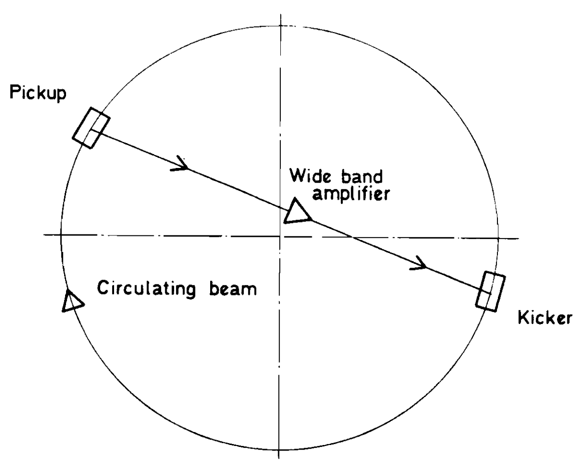

Fig. 1. Principle of stochastic cooling.

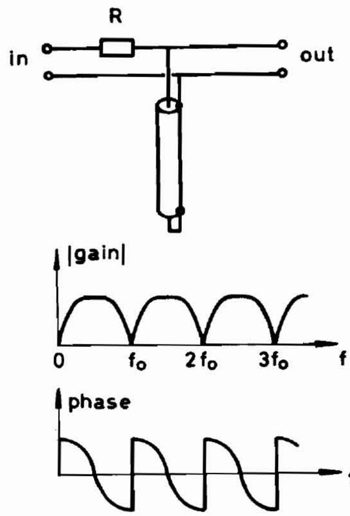

Fig. 2. A simple filter for momentum cooling.

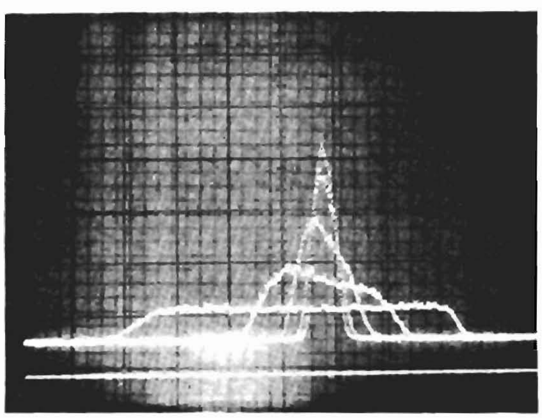

Fig. 3. Momentum cooling as obtained in the ICE experiment at CERN. Number of particles: 10^7. These Schottky scans represent the square toot of the density distributions. Successive scans were made at intervals of 1 minute.

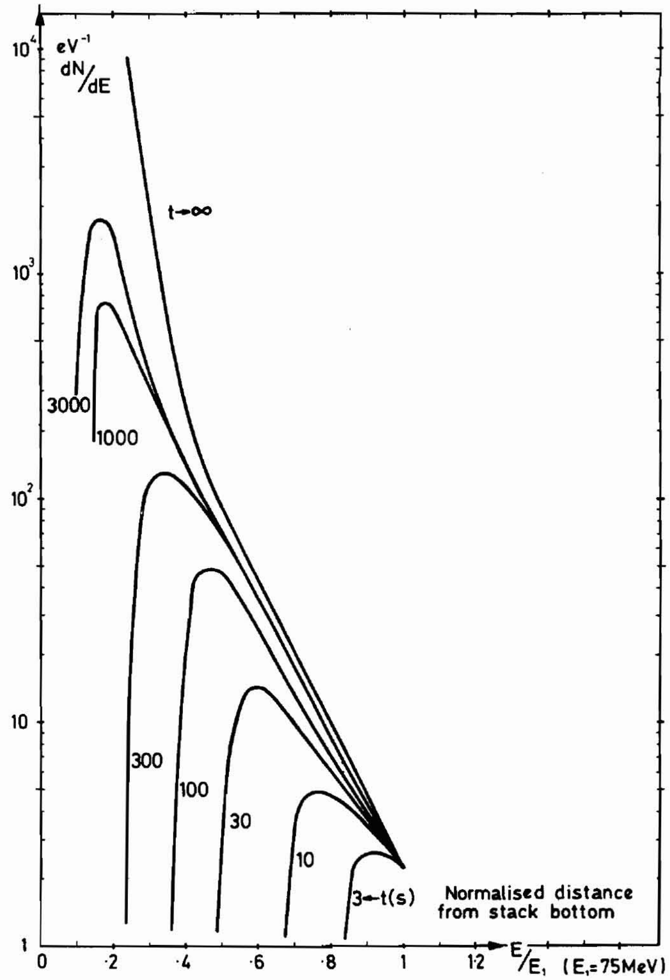

Fig. 4. Density distributions across the antiproton stack (CERN $p\bar{p}$ scheme).

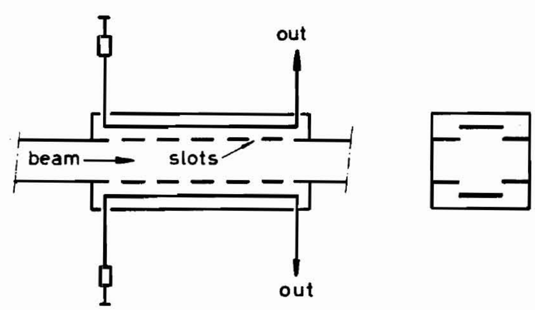

Fig. 5. Wide-band slot-type pickup.