STOCHASTIC COOLING WITH NOISE AND GOOD MIXING

A. G. Ruggiero
Fermi National Accelerator Laboratory

I. Introduction

In this note we shall make a few observations and derivations of the stochastic cooling theory. We shall work in the time domain as it was originally proposed by Van der Meer[1] and later amplified by Hereward.[2] This approach is called old-fashioned by some, a term which I do not understand. The new-fashioned method is to carry out the analysis in the frequency domain. This I believe to be a matter of taste and costum, but the two methods are equivalent and ought to give the same result. After all, a system frequency response can be replaced by an equivalent Green's function, and impedances and phase factors can be replaced by amplification and delay coefficients.

The ingredients that are required can be summarized as follows:

(i) A proper definition of the beam signal. This includes a single-particle signal as well as the signal produced by the surrounding ones. Several people like to distinguish the two contributions and call the latter beam or Schottky noise. I believe that this is relevant only up to some point, as we shall see later.

(ii) A proper definition of the noise from the amplification chain. This is a wide-band noise, also called "white" noise. Its spectrum is constant and its effect is completely random. It is quite different from the beam noise, which is not "white," but has a preferential frequency distribution. The integration of the beam-noise spectrum actually leads to a correlated time-dependent signal. To some, the term beam noise could be misleading. Thus one can expect different effects of the system noise and of the beam noise. It is not obvious that they should be simply added to each other.

(iii) Systematic loop errors. We give a few examples: the center of the beam can slowly move from turn to turn, or conversely the pickup device is not centered on the beam center; in the case of the notch filter device, the reference revolution frequency is not accurately determined. These errors would eventually lead to beam "heating" in the same way as the loop noise does.

(iv) Mixing of the signal. This is a crucial issue. Mixing is strongly beam-momentum dependent. Good mixing is achieved in the limit of $\gamma=1$. For large momentum the focusing of the ring is important; one would like to have a transition energy as low as possible. Mixing plays an important role in stochastic-cooling theory and one can draw different conclusions about cooling beams at diferent momenta that at first might sound contradictory. For example, in a bad mixing situation (large momentum) it seems preferable to work in a higher frequency range and momentum cooling seems to be more effective than betatron cooling. At the other end, in the limit $\gamma\to 1$, betatron cooling and momentum cooling are equally effective because the mixing situtation is better.

In all papers on stochastic cooling, one finds the statement that the method does not depend upon beam momentum. I believe this is not correct; not only does the mixing have a strong energy dependence, but also the electronic gain required for a given cooling rate is reduced at least with the first power of the beam momentum. Mixing also enters again in the cooling rate itself, since bad mixing leads to lower rates.

I believe these considerations are very relevant and should be taken into account in designing a large p-$\bar{p}$ colliding device. In this note, we shall look eventually to the case of good mixing, that is, the low momentum case. I believe that partial mixing can also be included in the following time domain, old-fashioned theory, but we shall leave it out for the moment.

Another reason to investigate the low-momentum case is that because we plan to carry out an experiment at Fermilab on the Electron Cooling Ring, we need to become acquainted with the technique.

II. The Stochastic Cooling Loop

The stochastic cooling loop is shown in Fig. 1.

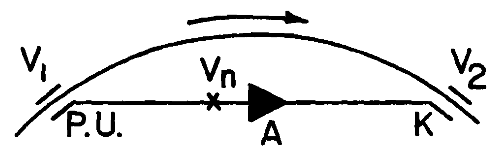

Fig. 1 The cooling loop

A fast beam detector PU is located somewhere around the ring. The pickup and the electronics which supply it have a characteristic rise time τ, so that if there are N particles uniformly spread around the ring, at any given time it is possible to observe a sample of n particles with

$$n = \frac{\tau}{T_0} N = \frac{\tau I}{e}, \tag{1}$$

where T_0 is the revolution period and I the beam current. From the pickup we expect a voltage V_1 which is modulated with time by the beam with resolution τ. We can write:

$$V_1 = sI\tau\bar{z}, \tag{2}$$

where s is the sensitivity and $\bar{z}$ the beam signal, which could be the average displacement from a reference orbit of the n particles simultaneously detected or their off-momentum value as it is measured, for example, by the notch-filter techinque at CERN. Because we are interested in the case of full mixing, we do not have to be specific about the beam signal; the following considerations apply to either betatron or momentum cooling. Nevertheless, the signal could be a combination of stochastic, coherent, and error contributions which are all function of time.

We do not have to specify the nature of the beam detector but we remark that, since it has to be broadband, its response is proportional to the instantaneous number of particles $I\tau$ as shown in (2).

The pickup voltage V_1 is amplified by a chain of amplifiers and applied to a beam kicker K. This could be **either** an electrostatic or magnetic deflector for the transverse cooling or a broadband cavity for momentum cooling. Its effect is to modify the motion of the same sample of beam that was measured at the pickup location by an amount which is proportional to the signal $\bar{z}$,

$$V_2 = \kappa\,(\beta p)(g\bar{z}), \tag{3}$$

where κ is a factor which measures the effectiveness of the kicker and g is the dynamic gain, which is the fraction of the signal which is actually damped with the voltage V_2. On the r. h. side of (3) we explicitly show the dependence on the momentum p and velocity β of the beam.

In the following, we assume that the delay between the pickup and the kicker is properly adjusted to guarantee that one is deflecting the same beam sample that has been detected and by the proper amount. We also assume the bandwidths of the pickup and kicker are matched to each other and that, as a consequence, there is no dependence on either I or τ or any of their conbinations on the r. h. side of (3).

Denoting by A the electronic amplification, we have

$$V_2 = AV_1 . \tag{4}$$

Combining (2), (3) and (4) gives

$$A = g\,\frac{\kappa\beta p}{sI\tau}, \tag{5}$$

which shows the relation between A and g, but also the dependence of the required amplification A for a given gain g on the beam momentum and current and on the system bandwidth. We note, though, that the gain g itself could depend on the beam current and the system bandwidth as we shall see later.

Equation (5) is the result that is crucial to our considerations and we will return to it later.

III. Front-End Noise

In the previous section we have analyzed how the beam signal is handled; in the present one we want to deal with another source of signal: the noise which is generated at the front-end of the amplification chain. As shown in Fig. 1 the noise figure is given by a front-end voltage V_n. This voltage is also amplified and applied to the kicker, and the beam at location K does not have the capability of discriminating between the contribution of the beam (V_2) and the noise contribution (AV_n). The beam will experience a total voltage $V_2 + AV_n$ and the system will interpret it as being caused by an equivalent beam signal $\bar{z} + r$ at the location of the pickup. From (2) we derive

$$V_n = sI\tau r, \tag{6}$$

where r is the equivalent beam displacement induced by the noise voltage V_n - observe that this quantity does not depend on the dynamic gain g or on the system amplification A; Eq. (6) represents the only relation which ties r to V_n. The reader should also note the fact that the beam current I and the risetime τ enter the r. h. side of (6).

There is a crucial difference between $\bar{z}$ and r. In the time domain, this diference can be expressed by observing that $\bar{z}$ has a strong autocorrelation, whereas r, being a white noise, is completely uncorrelated in time. By the frequency domain, this is made even more apparent by noting that r has a frequency -independent spectrum, whereas the frequency contained of $\bar{z}$ is lumped around harmonics of the revolution frequency.

The beam signal $\bar{z}$ ultimately leads to a cooling time T_D which is not expected to depend on the noise signal r. On the other hand, r causes a beam diffusion which is made quite visible, for instance, by opening the circuit on Fig. 1 between the P. U. and the K locations in front of the amplifiers. The two effects will eventually balance off to a minimum size that the beam can reach, which is given by the product of the cooling rate and the diffusion constant due to the noise. The characteristic time required to reach this final value is still given by T_D. We emphasize here the analogy of the two effects of damping and diffusion to synchrotron radiation in electron storage rings.

IV. Beam Dynamics

We take a particle in the beam as reference and follow its motion turn by turn. At one particular turn, the m-th, it will be crossing the beam pickup and will be detected together with n other particles Each particle in the sample gives a signal z_i (i=1, 2, ---, n) and the total signal is

$$\bar{z}_m = \frac{1}{n}\sum_i z_i ,$$

where the index m refers to the m-th turn. This sample at the same time has an emittance which can be described by

$$\sigma_m^2 = \frac{1}{n}\sum_i z_i^2 .$$

We assume that all the n particles travel together between P. U. and K. That is, that no mixing occurs. Then all the particles are kicked by the same amount. When the kick is translated to the location of the pickup (we shall always compare the beam at the same location) the coordinate z_i of each particle is modified as follows

$$z_i \rightarrow z_i - g\,(\bar{z}_m + r),$$

where we have included both the beam signal and the noise signal. This will have caused the emittance of the sample to change to

$$\begin{aligned}\sigma_{m+1}^2 &= \frac{1}{n}\sum_i (z_i - g\bar{z}_m - gr)^2 \\ &= \sigma_m^2 - (2g-g^2)\,\bar{z}_m^2 + g^2r^2 - (2g-g^2)\,r\bar{z}_m\end{aligned} \tag{7}$$

and for beam bary center

$$\bar{z}_{m+1} = \frac{1}{n} \sum_i (z_i - g\bar{z}_m - gr)$$

$$= (1-g)\,\bar{z}_m - gr. \qquad (8)$$

When the reference particle is back to the location of the pickup on the next turn, we assume it has lost its companions during the previous turn and is surrounded by n new, different particles (full mixing).

In this way, a new signal is generated and the cycle is repeated again. Since the reference particle will enter different beam samples, one can assume that (7) applies as an average over several turns to the entire beam. In this approximation one does not expect any correlation between r and $\bar{z}$ and therefore the last turn at the right-hand side of (7) does not give any contribution.

Taking m, the number of turns, as a continuous independent variable, we derive the following differential equations from (7) and (8).

$$\frac{d\sigma^2}{dm} = g^2r^2 - (2g-g^2)\,\bar{z}^2 \qquad (9)$$

$$\frac{d\bar{z}}{dm} = -g\bar{z}, \qquad (10)$$

where, for the last equation, we have assumed that the average value of r is zero.

We have not specified what $\bar{z}$ is. For instance, it could be caused by a coherent beam oscillation with no relation to the beam size σ. In this case, one integrates (10) to get

$$\bar{z} = \bar{z}_o\, e^{-gm}, \qquad (11)$$

which is the usual coherent-oscillation damping formula. The damping rate is given by the gain g, as one would have expected by definition. Insertion of (11) into (9) gives

$$\sigma^2 = \sigma_o^2 + g^2r^2m + \frac{2-g}{2}\,\bar{z}_o^2\,(e^{-2gm} - 1).$$

The second turn at the right-hand side is the diffusion term due to the noise, the last the damping of the apparent emittance due to coherent oscillations (the actual beam emittance does not change).

Note that the turn $g^2\bar{z}^2$ on the right-hand side of (9) has also been called the beam noise turn because it adds positively to the system noise g^2r^2. This definition is arbitrary; this term will also be there when $\bar{z}$ is a coherent oscillation, which we can hardly qualify as noise.

The case of interest is when $\bar{z}$ is a pure stochastic signal due to the finite number n of particles in the beam sample. This signal will change randomly from sample to sample with an expectation value given by

$$\bar{z}^2 = \frac{\sigma^2}{n}. \qquad (12)$$

In this situation, we can disregard Eq. (10) and replace $\bar{z}^2$ on the right-hand side of (9) by its expectation value. This is justified by the approximation that (9) applies in average over several turns. We obtain

$$\frac{d\sigma^2}{dm} = g^2r^2 - \frac{2g-g^2}{n}\,\sigma^2. \qquad (13)$$

More generally, one should have also included errors and have written (12) as

$$\bar{z}^2 = \frac{\sigma^2}{n} + z_r^2.$$

But one can combine the effect of z_r^2 with r^2 and probably ignore it as long $z_r^2 << r^2$. We shall assume in the following that is indeed the case.

Eq. (13) was first derived by Hereward,[2] but he integrates it in a curious way. He introduces the quantity

$$\eta = \frac{r^2}{\sigma^2/n} = \frac{\text{noise power}}{\text{signal power}} \qquad (14)$$

and assumes that η is a constant. This could be an approximation at the beginning of the cooling and for slow cooling, when indeed the beam signal does not change much. But in fast cooling σ^2 would change rapidly, whereas r^2 remains constant. In this regime, η can no longer be regarded as a constant.

Eq. (13) can in fact be integrated to give the general solution

$$\sigma^2 = \sigma_\infty^2 - (\sigma_\infty^2 - \sigma_o^2)\,e^{-\alpha m}, \qquad (14')$$

where

$$\sigma_\infty^2 = \frac{ngr^2}{2-g} \quad \text{and} \quad \alpha = \frac{2g-g^2}{n} \qquad \text{(15a and b)}$$

The solution (14) is also plotted in Fig. 2

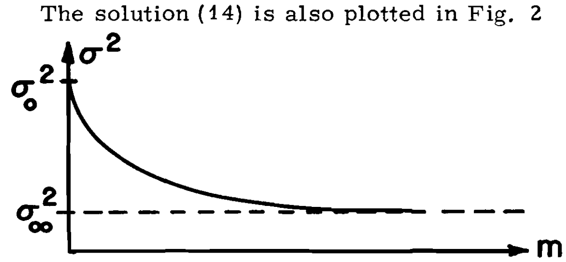

Fig. 2. Stochastic Cooling.

It is still an exponential decay with a cooling rate given by α, which does not depend on the noise figure r. But the cooling saturates at a final value σ_∞ which depends on the noise. When the noise disappears the final size also vanishes.

It is possible that the factor η, Eq. (14), is relatively small at the beginning of cooling, because the beam signal is larger. But toward the end of the cooling, the situation reverses; the noise signal predominates and the factor η cannot be ignored.

Inspection of (15) shows conflicting requirements for the gain g. From one side, one would like to have fast cooling which requires a large g, possibly g=1. On the other side, if cooling has to be effective, the final size σ_∞^2 should be small which require a small g. The case g=1 would work if

$$\sigma_\infty^2 = \frac{nr^2}{2} << \sigma_o^2 .$$

This requires small front-end noise V_n and large bandwidth.

The theory we have outlined above applies only to the case of good mixing. Nevertheless, we may expect that some of the conclusions, at least qualitatively, apply also to the case of bad mixing. For instance, the solution should still have the form of Eq. (14') as sketched in Fig. 2, provided that the cooling rate α and the final beam size σ_∞^2 are properly defined to include a mixing coefficient. One would expect the cooling rate to become smaller and the final beam size to become larger. Thus good mixing represents the optimal situation.

It should be possible to treat the bad mixing case with the approach outlined here by splitting $\bar{z}$ in Eqs. (9) and (10) in two contributions, one from the old particles which still remain with the test particle and one from the new ones which are just refilling the sample under consideration.

V. Consequenses of our Analysis

The conclusion of our analysis can be drawn by combining Eqs. (1), (5), (6) and (15), and

$$T = \text{cooling time} = T_o/\alpha .$$

Several of these equations can be combined to give the following

$$A = g\frac{\kappa\beta p}{sI\tau} \qquad (16)$$

$$T_D = \frac{\tau\, IT_o/e}{g\,(2-g)} \qquad (17)$$

and

$$\sigma_\infty^2 = \frac{gV_n^2}{(2-g)s^2e\tau\, I} . \qquad (18)$$

The next step is to eliminate g from (16) and derive two equations from (17) and (18)

$$\frac{T_D}{T_o} = \frac{\tau\, I_o^2/4e}{I_o - I} \qquad (19)$$

$$\sigma_\infty^2 = \frac{\bar{\sigma}^2 I_o}{I_o - I}, \qquad (20)$$

with

$$I_o = \frac{2\kappa\beta p}{As\tau} \qquad (21)$$

and

$$\bar{\sigma}^2 = \frac{AV_n^2}{2es\kappa\beta p} \qquad (22)$$

These are all the equations that are required to design a cooling loop. They give the cooling time and the final beam size in terms of the machine revolution period T_o, the beam momentum p, the system bandwidth $1/\tau$, the front-end noise V_n, the electronic gain A and the two parameters s and κ which are the sensitivity of the pickup and the effectiviness of the kicker. These equations are quite general.

One has cooling when $s\kappa A > 0$ and $I < I_o$, in which case also $\bar{\sigma}^2 > 0$.

Observe that in the limit of small current, the cooling time is independent of the beam intensity and the system bandwidth

$$\frac{T_D}{T_o} = \frac{\kappa\beta p}{2esA} \qquad (23)$$

and

$$\sigma_\infty^2 = \bar{\sigma}^2 \qquad \text{for } I << I_o .$$

When I approaches I_o the cooling time becomes infinite and there is no more cooling. The final beam size also diverges because of the diffusion caused by the noise. Observe that there are some conflicting requirements on the electronic gain A, as one can see by inspecting (22) and (23): faster cooling is obtained with larger A, which also causes a larger final beam size. In the limit of small current there is also no dependence on the system risetime.

Our result (19) might seem strange and in contradiction with previous results. This was known under the form of Eq. (15b): for a constant dynamic gain g the cooling rate is proportional to the system bandwidth and to the inverse of the beam intensity. One has cooling only if

$$0 < g < 2 .$$

For practical purposes g is given by (5). Since A is usually a large number, it is the quantity that is kept constant, so that g increases with I. When $I > I_o$ then $g > 2$ and one does not have cooling anymore.

VI. The Experiment at Fermilab

An experiment on stochastic cooling has been proposed at Fermilab, to be carried out in the Electron Cooling Ring at a momentum of 644 MeV/c. The three loops, horizontal (H), vertical (V) and momentum-wise (P) are shown in Fig. 3.

The two loops for damping betatron oscillations have the following parameters:

$$s = 200 \text{ V/A}\cdot\text{m}\cdot\text{nsec}$$
$$k = 0.05 \text{ V/m}\cdot(\text{eV/c})$$

and

$$\tau = 2\text{nsec (200 MHz bandwidth)}$$
$$A = 10^6 \text{ (120 db)}$$
$$V_r = 10\ \mu\text{V},$$

which gives

$I_O = 91\text{mA}$
$T_D = 228$ sec for $I << I_O$
$\bar{\sigma} = 4.6$ mm.

The final beam size corresponds to an emittance of $3\pi 10^{-6}$ m (for 95% of the beam). The initial one could be $10\text{-}20 \cdot \pi \cdot 10^{-6}$m. The sensitivity figures given above is for a standard pair of electrodes 6 in. long with a 45° cut. The kicker could also be made of a pair of deflecting electrodes of the same length.

Observe that toward the end of the cooling the ratio of the noise power to the beam signal power, Eq. (14) is given by

$$\eta = \frac{I_O - I}{I},$$

as one can derive from (6) and (20). Thus in the limit of $I << I_O$, one has $\eta >> 1$ and most of the power required is given by the contribution of the noise. If the deflecting plates are matched to an impedance of 50 ohms, with a gain of 120 db and a front end noise of 10μV, the power required is 2W, probably marginal.

References

1. S. Van der Meer, CERN/ISR-PO/72-31.
2. H. G. Hereward, Proc. of the first Course of the International School of Particle Accelerators, Erice, Nov. 1976, page 281. (Hereward made more contributions, but unfortunately they have not all been published).

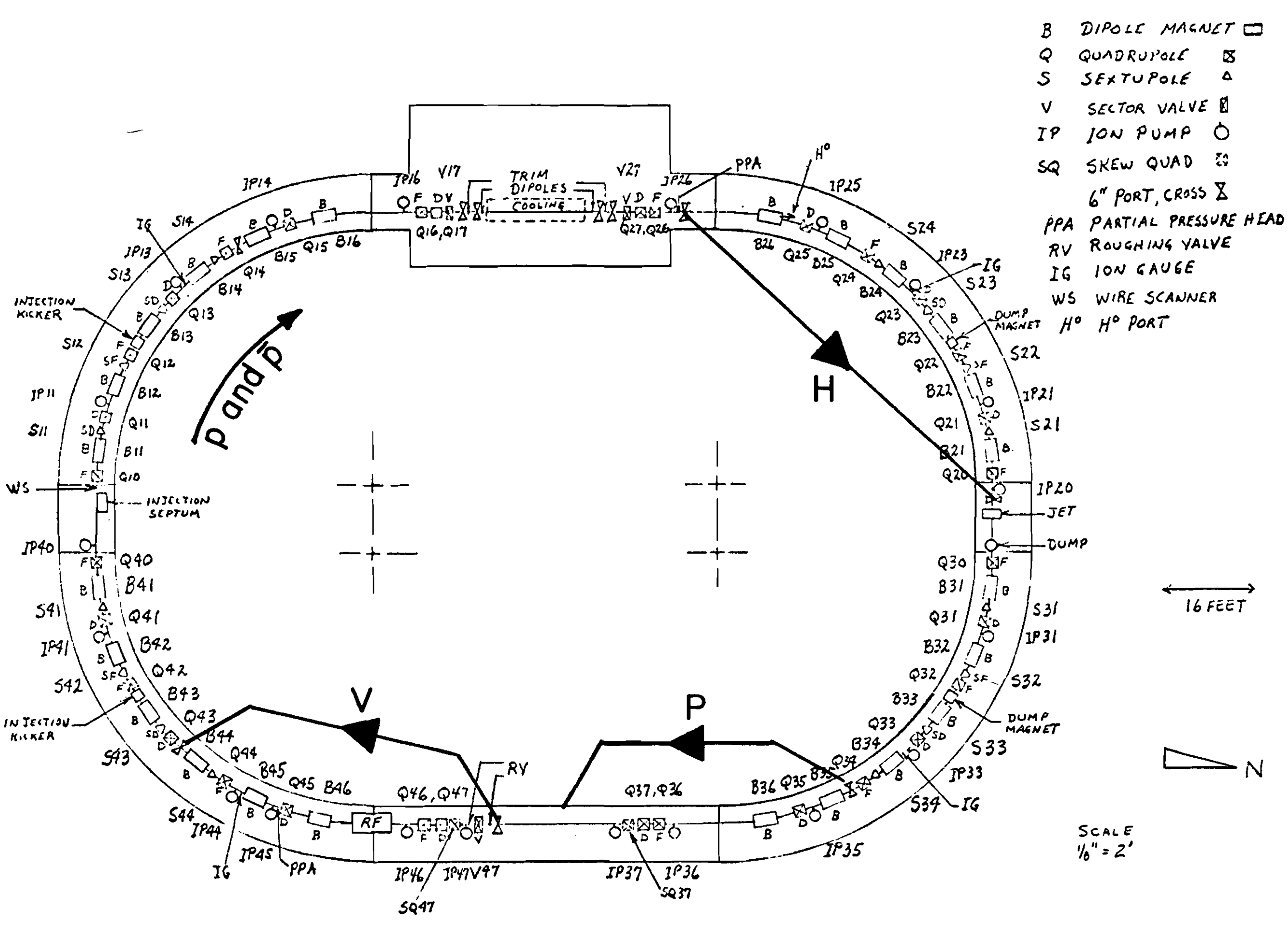

Fig. 3. The stochastic cooling loops for the electron cooling ring.