EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

SVDM/afm

CERN/PS/84-33 (AA)

AN INTRODUCTION TO STOCHASTIC COOLING

S. van der Meer

ABSTRACT

The principles of stochastic cooling will be discussed with the emphasis on a qualitative explanation of the most interesting points, such as

- how to circumvent Liouville's theorem,
- time domain and frequency domain viewpoint and their correspondence (mixing, beam feedback),
- betatron cooling and momentum cooling; how filters are used to improve momentum cooling,
- stochastic stacking,
- how to make wide-band pick-ups and kickers,
- future developments,
- electron cooling vs stochastic cooling.

The main features of the CERN $\bar{p}$ accumulator (AA) and its future addition (ACOL) also be described.

(Lecture given at the Fermilab Summer School on High-Energy Particle Accelerators, 1984)

Geneva, Switzerland

December 1984

1. INTRODUCTION

Stochastic cooling is a technique that aims at reducing the volume occupied in phase space by beams of particles. Such fields produce conservative forces and Liouville's theorem states that such forces cannot change phase space volume, only distort it. However, one way to reduce the phase volume occupied by the particles is to make use of the fact that particles are points surrounded by empty space. We may distort the phase space in such a way that individual particles are displaced towards the centre of the distribution, while empty space moves outward. This does not violate Liouville's principle, but of course to do it we must a) look where the individual particles are and b) influence them individually. The process is reminiscent of Maxwell's demon, but with the difference that, unlike the demon, we exchange energy with the particles while pushing them around, so that we do not violate the second law of thermodynamics. Incidentally, although we influence individual particles, we are far away from the point where quantum effects become noticeable in all practical applications of stochastic cooling.

If we had just a single particle circulating in a storage ring, we could in principle look where it is in phase space and give it just the right kicks to move it where we want within a single turn. Figure 1 illustrates how this is done for horizontal betatron oscillations (which are usually decoupled to the extent that they involve only two dimensions of phase space). The same can of course be done vertically. For longitudinal cooling we measure the energy somehow and use a longitudinal kicker (Fig. 2).

Figure 1 - Transverse cooling of a single particle.

Figure 2 - Longitudinal cooling of a single particle.

In (a) the pick-up detects the momentum-dependent position at a point with non-zero dispersion. The kicker will suddenly displace the closed orbit and therefore cause horizontal oscillations unless it is placed where the dispersion is zero.

In (b) the dependence of revolution frequency on momentum is used (see the section on filter cooling).

If there is more than one particle in the ring, this approach will still work if the electronics is fast enough to treat each particle separately, up to the point where the system response time is of the same order as the average interval between the passage of the particles. The total cooling time is then equal to the system response time multiplied by the number of particles.

In practice, there are usually far more particles in a storage ring than this, so that they cannot be distinguished individually. However, it turns out as will be seen later that the law given above for the cooling time will remain valid as a fundamental limitation. As an example, for 10^{12} particles in a ring and a system response time of 1 ns, the cooling time would be at least 1000s. Clearly, stochastic cooling is too slow for accelerators where intensities are high and cycle times are measured in seconds. However, for storage rings the method is often useful.

If many particles pass the pick-up within the system's response time, we must reduce the system gain so that we get only a small fraction of the needed correction on each turn. This is necessary because together with the signal from each particle there will be a perturbing signal from all the other particles that cannot be resolved by the system. In other words, each particle is cooled by the effect of its own signal, but simultaneously heated by the uncorrelated signals from the other particles. Now the cooling effect from the particle's own signal is proportional to the system gain, but the random heating by the other particle's signals causes a displacement in phase space whose mean square increases linearly with time, just as e.g. for Brownian motion. The heating correspondingly varies with the square of the system gain (Fig. 3). As a consequence, we may always find an optimum gain value where we get optimum overall cooling; for this gain, the heating rate is equal to half the cooling rate.

Figure 3 - Variation of coherent (cooling) and incoherent (heating) rates with the system gain.

In the following, we shall give a mainly qualitative description of the most important features of stochastic cooling, using as an illustration the CERN antiproton accumulator (AA). A more detailed analysis is given in reference 1.

2. TIME DOMAIN ANALYSIS

The very rough description given above sketches the process in the time domain. It is as if we had discrete samples of particles, corresponding to the number of particles that pass during one response time of the system. We may say, as above, that each particle cools itself and is perturbed by the other particles in the sample. Alternatively, we may consider the average position or momentum of each sample and correct that; this is clearly the best we can do. The two descriptions are, of course, equivalent. For the sake of simplicity we shall analyze the second point of view. We consider betatron cooling, with the arrangement of Fig. 1, but with the difference that we now consider a sample instead of a single particle.

We assume that the system's bandwidth is W . Information theory tells us that the signal transmitted by such a system during one revolution time T can be completely defined in terms of $2WT$ equally spaced samples. Therefore, the number of particles in each sample is

$$N_s = N/2WT \quad (1)$$

if the total number is N .

The pick-up sees the average position x of these N_s particles. The system gain is adjusted so that the kicker will correct this, as in Fig. 1. Therefore, for each particle, we modify the position as follows:

$$x \rightarrow x - \bar{x}.$$

Taking the mean square over many random samples,

$$\overline{x^2} \rightarrow \overline{(x-x)^2} = \overline{x^2} - \overline{x^2},$$

so that per turn the mean square $\overline{x^2}$ decreases by a fraction

$$\frac{\overline{x^2}}{\overline{x^2}} = 1/N_s = 2WT/N. \quad (2)$$

Thus, the cooling rate for the mean square of x is

$$\frac{1}{\overline{x^2}} \frac{d\overline{x^2}}{dt} = 2W/N. \quad (3)$$

We usually define cooling rate for the position x rather than for $\overline{x^2}$. This is a factor 2 lower. Another factor 2 is lost because the particles do not always arrive with the most favourable phase at the pick-up as shown in Fig. 1. Thus, both pick-up and kicker are less effective by the sine of a random phase angle and the cooling rate is multiplied by $\sin^2 = 1/2$. Therefore it is

$$\frac{1}{\tau} = \frac{W}{2N} \quad (4)$$

where τ is the cooling time constant. For longitudinal cooling a similar result is valid except that we do not lose the last factor of 2.

We assumed random samples. This means that from one turn to the next the samples must get new populations. In practice this renewal occurs because different particles have a slightly different energy and therefore different revolution times. If this effect is strong enough to completely renew the samples after each turn, we speak of "good mixing". Bad mixing occurs if the samples are only partially renewed; this reduces the cooling rate.

If we do not entirely correct the observed average sample position $\bar{x}$, but only a fraction $g\bar{x}$, we find in the same way as above

$$\frac{1}{\tau} = \frac{W}{2N} (2g - g^2), \quad (5)$$

which is of course largest for $g = 1$. It now turns out (but we will not prove it here) that the first, linear, term corresponds to the effect of each particle on itself (coherent effect), while the second, incoherent term represents the heating from the other particles in the sample. It is this heating term that is increased by bad mixing (because the perturbation from the other particles remains correlated over several turns). It may also be increased by thermal noise in the system (mainly from the sensitive circuitry near the pick-up). We may define a mixing factor M (>1 for bad mixing) and a factor U that is equal to the ratio of thermal noise power to signal power.

Then

$$\frac{1}{\tau} = \frac{W}{2N} (2g - g^2 (M + U)) \quad (6)$$

and if we optimize g again, we find a lower value

$$\frac{1}{\tau} = \frac{W}{2N(M + U)} \quad (7)$$

Note that this treatment suggests that we correct random fluctuations of samples that are continually renewing themselves. The name "stochastic cooling" stems from this viewpoint. However, if we consider the mechanism in more detail, we see that the cooling is coherent and that only the associated heating is a stochastic process. The name is therefore not really appropriate.

3. FREQUENCY DOMAIN VIEWPOINT

The analysis above is somewhat approximative. If samples mix from one turn to the next, then they will unfortunately also mix to a certain extent between the pick-up and the kicker; this must be taken into account in a more precise description. Also, there is another effect that is difficult to treat in the time domain: the beam feedback.

This is the following. Any signal on the kicker will modulate the beam position or energy. This coherent modulation will persist as the particles turn round and will in turn induce a signal in the pick-up. We therefore have an additional feedback loop via the beam whose influence must be taken into account.

These effects are described more conveniently in terms of frequency response. Moreover, in practice the frequency response of the electrical circuitry and of the kicker and pick-up may be measured with standard equipment (network analysers). Probably the main reason why the frequency domain description is more convenient than the time domain one is that we have two separate parameters (amplitude and phase) to describe the system behaviour instead of only one in the time domain. This allows a clearer understanding, although in principle the two descriptions are equivalent.

The scope of this lecture does not permit a quantitative frequency domain analysis, but it is interesting to consider in some more detail the frequency spectrum of the pick-up signal and the response of the beam to sine-wave signals at the kicker.

4. SCHOTTKY SIGNALS

We shall first consider a so-called sum pick-up. This is a device into which each passing particle induces the same signal, independent of its transverse position. Ideally, the signal would be a delta function; if its duration is finite, we may take this into account by including the corresponding response function into the response of the complete system.

The repetitive delta functions generated by one circulating particle will correspond in the frequency domain to an infinite number of harmonics of the revolution frequency f_0 , each of the same strength (Fig. 4, upper half). In practice, the cooling system only covers a limited frequency range, but this still contains many harmonics.

Fig. 4 Schottky signals

If there are many particles, each will generate its own spectral lines; because of the revolution frequency spread Δf_0 these will form spectral noise bands, called Schottky bands in analogy to the classical Schottky effect due to quantized charge. The width of these

bands is proportional to frequency so that at high frequencies (harmonic number $n > f_0/\Delta f_0$) they will overlap. It may be shown that good mixing of the time domain samples corresponds to overlapping of the Schottky bands inside the system's passband.

For a difference pick-up, whose response is proportional to the transverse position of the particle, the situation is different (Fig. 4, lower half). The position of the particle changes sinusoidally with the betatron frequency, so that the pick-up signal will be like the sum signal, but amplitude-modulated by the betatron frequency. As a result, each Schottky band splits up into two sidebands at frequencies $(n + q)f_0$ and $(n - q)f_0$ where q is the fractional part of the betatron tune Q (Fig. 4, lower half). Again, at high frequencies the bands will overlap.

We can now see how cooling is treated in the frequency domain. For one particle, each Schottky line corresponds to a sine wave that will cool the particle if at the frequency considered its phase at the kicker is correct; each line will also contribute to the heating, as will be discussed in the following section. With optimized gain at all lines, they will all contribute to the overall cooling. This is why a wideband system is needed: we should use as many Schottky lines as possible.

5. THE HEATING EFFECT

We should now consider how the heating by the other particles really occurs. In the time domain, we saw how we could consider the random fluctuations of each sample. It would therefore seem as if only the particles arriving at the pick-up simultaneously (or, rather, within one sample duration) would perturb each other, independent of their revolution frequency.

In the frequency domain, analysis shows that a particle can only be heated by signals with frequencies coinciding with the particle's Schottky lines. If we try to excite a particle with somewhat different frequencies a beat effect will be produced, but on the average (for times much larger than the inverse of the frequency difference) there will be no energy transfer and the particle's oscillations (or momentum change) will not be increased indefinitely. If, on the other hand, we excite with a noise signal, the spectral density at the particle's Schottky frequencies will contribute to the excitation. If we excite during a time Δt , a spectral width of order $1/\Delta t$ will contribute². For times of the order of the cooling time this is always much less than the width of a Schottky band. The picture that emerges is quite different from the time domain picture. Particles with different revolution frequencies will not influence each other at all if the Schottky bands do not overlap. Only for very strong mixing the two viewpoints are equivalent; clearly, our simple time domain analysis neglected correlations from one revolution to the next.

Bad mixing reduces the cooling rate. In the time domain this is explained by the non-renewal of the samples which reduces their fluctuations, thus reducing the amount by which these may be decreased. In the frequency domain we see that bad mixing does not reduce the coherent effect, but that the Schottky bands are narrower, so that their density is higher, increasing the incoherent or heating effect. This would make a lower system gain necessary.

In fact, the fundamental cooling limit turns out to depend on the spectral density rather than the total number of particles. Only for perfect mixing the two are equivalent.

Mixing of the samples between the pick-up and the kicker is treated in the frequency domain as a phase shift depending on revolution frequency. This phase shift occurs within each Schottky band. It does not influence the heating, but reduces the cooling term.

6. THE BEAM FEEDBACK EFFECT

We have seen that a sinusoidal excitation will only lead to heating for those particles that have a Schottky line at the same frequency. However, if the frequency is not identical, there will still be a beating phenomenon, and therefore the position or energy of a beam may be coherently modulated by such a signal even if it does not cause heating.

As a consequence, if we switch on the cooling system, we close a feedback loop: the signals on the kicker will modulate the beam and this will be seen by the pick-up. This will sometimes lead to instabilities, or in any case to an instantaneous modification of the Schottky signals observed by the pick-up. In general, it turns out that if the phase is correctly adjusted for cooling, the feedback is negative and the Schottky signals decrease. In the absence of thermal noise, the open-loop gain for optimum cooling is of the order of unity and the Schottky signals should be approximately halved. This is just an effect of the beam feedback and has nothing to do with cooling; if the cooling circuit is switched off, the original signals reappear. Incidentally, this effect is useful for adjusting the correct gain and phase of the system.

Detailed analysis shows that the beam feedback effect does not have a bad influence on cooling, although it modifies the optimum parameters slightly. However, a consequence is that the gain and phase of the system have to be carefully controlled to avoid instabilities.

7. FILTER COOLING

The use of a filter for longitudinal cooling (i.e. reducing the energy spread), indicated in Fig. 2b, may now be understood. In fact, the filter must provide a phase jump of 180° in the middle of each Schottky band to push high-momentum particles downwards and vice-versa. Of course, this will only work if the Schottky bands do not overlap (bad mixing).

A filter that performs this extraordinary feat was invented by Thorndahl³. In its simplest form it consists of a resistor in parallel with a long transmission line (Fig. 5), shorted at the far end; this arrangement is fed from a current source. The length of the line is chosen so that its resonant frequencies (where it behaves as a short-circuit) coincide with the harmonics of the central revolution frequency f_c at the centre of each band. Just above or below each resonance the line is inductive or capacitive. Thus, around the centre of each Schottky band the filter response shows a notch and the phase jumps by the required 180° .

More complicated filters may be constructed, using several such lines; in some cases active filters incorporating wide-band amplifiers and feedback loops are used.

The use of filters has greatly improved longitudinal cooling for two reasons. Firstly, the filter notches reduce the thermal noise as well as the signal; if we use differential pick-ups as in Fig. 2a, only the signal is reduced in the centre, but the noise (from the preamplifier and matching resistors) remains. In the second place, sum pick-ups are often shorter than differential ones so that more may be placed in the same space; this again

Figure 5 - Simple notch filter.

improves the signal-to-noise ratio. The filters may also help to reduce the thermal noise at frequencies between the Schottky bands. This noise does no harm to the particles, but increases the power needed at the kicker.

Figure 6 shows an illustration of the cooling process; the two curves represent a longitudinal Schottky band, before and after cooling, measured by connecting a sum pick-up to a spectrum analyser that displays voltage vs frequency.

Figure 6 - Longitudinal Schottky scans of a $\bar{p}$ beam before and after precooling. The cooling time is 2 seconds.

Note that it is the power density of the signal and therefore the square of these curves that is proportional to the particle density dN/df_0 . This particular picture shows the 173rd harmonic (near 319 MHz) of the revolution frequency. It was made in the CERN $\bar{p}$ accumulator and represents the precooling (in about 2 seconds) of the $\bar{p}$ beam after injection, by the filter method. This system covers the range from 150 to 500 MHz.

8. MORE ABOUT LONGITUDINAL COOLING - STOCHASTIC STACKING

The main difference between longitudinal and transverse cooling is that the first will change the particle distribution vs. revolution frequency and therefore modify the mixing factor during the process. For transverse cooling this does not happen.

As we have seen, the maximum cooling rate depends on the spectral density dN/df . At the edges of the longitudinal distribution, where the density is low if the Schottky bands are separated, we could in principle allow a higher gain and so obtain faster cooling. For transverse cooling this is hardly worth while; most of the particles are in the centre anyway. Longitudinal cooling is different; whether a notch filter is used or a transverse pick-up as in fig. 2a, the gain will be higher at the edges than in the dense centre although in practice it is not easy to approach the optimum profile. In any case, the distribution is not constant.

A special case of longitudinal cooling, used for accumulating antiprotons in the CERN "AA" machine, is what we call stochastic stacking. A large number of particles is circulating in a ring. At the edge of the distribution (the "stack") we deposit at regular intervals (2.4 s) new, low-density batches of particles. This is done by RF deceleration from the injection orbit (which does not overlap in frequency with the stack) towards the stack edge (fig. 7).

Figure 7 - Stochastic stacking in the AA. The density distribution is shown with the injected $\bar{p}$ beam at the left and the stack at the right.

The newly deposited particles must be pushed up against the density gradient towards the centre of the stack by cooling before the next batch arrives; otherwise Liouville's principle would ensure that they would be pushed away from the stack while the new particles are deposited. Thus, at the stack edge we need fast cooling, but at the momentum corresponding to the stack core (where the density may be typically 30000 times higher) the gain must be correspondingly lower and the cooling will be much slower. This is OK since the particles stay there a long time, typically a day. To shape the gain profile vs frequency (or momentum) is not easy. It is done in practice by using an extreme version of fig. 2a i.e. putting the pick-ups at a place where the dispersion is large so that the orbits of particles with different momentum are separated horizontally, and by using narrow pick-ups that see the tail of the stack, but much less its core. In practice, this is not yet quite good enough and the cooling system is subdivided into several sections with different gain and pick-ups at different radial positions. Transmission-line filters are also used to remove noise from the high-gain sections at the frequencies of the dense core where the cooling is slow.

In contrast to the pick-ups, the kickers must be placed at a point where the dispersion and its derivative are zero. This is because otherwise a longitudinal kick would cause a shift or an angular displacement of closed orbit; this would excite the betatron oscillations.

In the $\bar{p}$ accumulator we precool the particles, also longitudinally (Fig. 7, left hand side) before transferring them to the stack. In addition, we apply transverse cooling to the stack to reduce the emittance. The phase space density is increased by a factor 250000 longitudinally, and 60 for each transverse plane. This gives a factor of 10^9 in 6-dimensional phase space.

9. FOKKER - PLANCK EQUATION

To evaluate the stacking process (with its parameters that vary strongly across the distribution) in a quantitative way, we introduce the flux

$$\phi = dN/dt, \quad (8)$$

i.e. the number of particles per unit time passing a given momentum (or frequency) value. It can be shown that this consists of two components:

$$\phi = F\Psi - Dd\Psi/df_0 \quad (9)$$

where Ψ is the density dN/df_0 , while F and D are approximately constant. The first term is the coherent one, providing cooling if the sign of F is correct, while the second, incoherent term causes the particles to move "down the slope" towards regions of lower density (D is always > 0). F and D of course depend on many things; they vary both with the revolution frequency f_0 and with the harmonic number of the Schottky line considered. To find the flux at a given f_0 , we must sum over all Schottky lines within the passband to obtain F and D . Via the beam feedback mechanism F and D also depend in a complicated way on the density Ψ and its distribution. However, given a certain distribution, F and D may be evaluated numerically.

Then, by combining (8) with the continuity equation

$$\frac{\partial \Psi}{\partial t} + \frac{\partial \phi}{\partial f_0} = 0 \quad (10)$$

we obtain a Fokker - Planck equation

$$\frac{\partial \Psi}{\partial t} = - \frac{\partial}{\partial f_0} (F\Psi) + \frac{\partial}{\partial f_0} \left(D \frac{\partial \Psi}{\partial f_0} \right) \quad (11)$$

that allows us to calculate the development of the distribution vs time. The results of such calculations for the $\bar{p}$ accumulator show a good agreement with experiment.

10. PICK-UPS, KICKERS AND OTHER HARDWARE

The wide-band electronics that is used for stochastic cooling usually consists of commercially available equipment such as amplifiers, splitters, combiners, relays, transmission lines or measuring equipment. Often such equipment is made for an octave bandwidth, e.g. 250 - 500 MHz, 1-2 GHz or 2-4 GHz. A special feature is the small phase error across the band required for cooling purposes; fully matched systems with minimum reflections are needed to obtain this.

Pick-ups often consist of coupling loops, matched at the far end with their characteristic impedance (fig. 8a). If such a loop is a quarter wavelength long at the centre of the band its response is reasonably flat. The signals from opposite loops may be added or subtracted, to get a sum or differential response. Exactly the same elements may be used as kickers. When fed in push-pull, they will provide transversely deflecting fields for beta-tron cooling; with in-phase signals they will become longitudinal kickers that accelerate or decelerate the particles.

Figure 8 - a) Matched quarter-wavelength coupling loops that may be used as pick-ups or kickers, either longitudinal or transverse. For kicker operation, the beam direction is opposite to the one shown.

b) Ferrite-ring longitudinal pick-up or kicker.

If weak beams are cooled (e.g. precooling antiprotons) the signal-to-noise ratio at the pick-ups may be a problem. To improve this, the signals from many pick-ups may be combined, or the matching resistors and/or the preamplifiers may be cryogenically cooled.

At frequencies below 500 MHz, ferrites may be used and a much shorter sum pick-up (or kicker) may consist of a ferrite ring surrounding the beam (Fig. 8b), with a single-turn coupling loop. This acts as a wide-band transformer between the beam and the signal input or output.

Above 1 GHz, sometimes slotted structures are used as pick-ups or kickers (Fig. 9). These are somewhat less sensitive than coupling loops but easier to construct. Essentially they consist of a transmission line running parallel to the beam and coupled to it by transverse slots in its outer conductor. The slots are shorter than half a wavelength so that the coupling is weak and the matching of the line is not disturbed too much; the signals from all slots add up if the dimensions are correctly chosen and if the particles have the same velocity as the wave in the transmission line.

Figure 9 - Slotted structure for use as pick-up or kicker.

For fast cooling systems the most difficult part is often the power amplifier feeding the kickers. Wide-band high-frequency power amplifiers are not much used for other purposes (except military ones such as radar-jamming) and they are surprisingly expensive. Transistor amplifiers have been used below about 1 GHz, travelling wave tubes above; this limit frequency shows signs of increasing.

11. MORE ABOUT THE $\bar{p}$ ACCUMULATOR (AA) AND ITS IMPROVEMENT PROGRAM (ACOL)

The AA ring and its different cooling systems are shown schematically in fig. 10.

Figure 10 - Cooling systems of the AA ring.

The unusual shape of the ring is connected with the conditions needed for efficient cooling. In particular, the two long straight sections have zero dispersion; this is needed for the longitudinal kickers, and it facilitates the injection of the large-emittance beam ($\sim 85\pi$ mm.mrad in both planes, $\Delta p/p = 1.5\%$ full width). The operating momentum of 3.5 GeV/c is chosen for optimum $\bar{p}$ production with a 26 GeV/c primary beam. At this momentum the variation of particle velocity with momentum is small, so that the mixing is mainly obtained by making the orbit circumference depend on momentum. This requires a large dispersion in the circular parts of the ring; the bending magnets are placed so that this is obtained within a regular FODO lattice.

From top to bottom in Fig. 10, we see the following cooling systems:

- a) Longitudinal precooling, as mentioned above. This system works on a small number of particles ($\sim 6 \times 10^6$) and therefore needs 200 pick-ups (of the ferrite-ring type). Because the cooling is fast, the power is high (2 kW), even with 200 kickers of the same type as the pick-ups.

Both pick-ups and kickers surround the injected beam. After precooling, one leg of the ferrite frame of each pick-up and kicker is vertically displaced by a fast moving mechanism to enable the beam to pass to the stack (Fig. 11)

Figure 11 - A vacuum tank containing 100 rectangular ferrite-frame kickers (right hand side). The stack is on the left hand side. One leg of the ferrite frame is moved downward to allow the passage of particles from the injection region on the right towards the stack.

- b) Stack tail cooling. This covers the low-density part of the stack. The main part is longitudinal. Sixteen loop pairs are used as pick-ups, 144 ferrite rings as kickers. The horizontal and vertical stack tail cooling were provided as an insurance against possible blow-up during migration of the particles to the core. This appears to be small; these systems therefore do not contribute much to the stacking rate. The main transverse cooling is done by the core systems.
- c) These core cooling systems (longitudinal, horizontal and vertical) are low-power, high-frequency (1-2 GHz) systems with cooling times of the order of 20 minutes. They use slot-type pick-ups and kickers.

To increase the stacking rate by a factor 10 or more, a second ring surrounding the AA (Fig. 12) is planned (ACOL, first beam foreseen 1987).

Figure 12 - Plan view of the AA ring surrounded by the future ACOL ring.

This ring will accept a 6% momentum spread instead of 1.5% for AA. In addition, the transverse acceptances will be 200π mm.mrad.

The momentum spread will be reduced to about 1.5% by RF bunch rotation. This will be followed by fast transverse and longitudinal cooling before each pulse is transferred to the AA to be stacked. Cooling systems covering the 1-3 GHz band will be used. The AA stack cooling will also be upgraded to higher frequency to accept the higher flux.

A separate ring is needed because it is difficult to do the bunch rotation in the presence of a stack in the same ring. Also, the strong focusing needed for the large acceptance leads to low dispersion and a lower mixing factor, that would make stacking in the same ring more difficult. In fact, the ACOL-AA combination resembles in many respects the p accumulator scheme under construction at Fermilab, where a 2-ring design was adopted for similar reasons.

12. FUTURE DEVELOPMENT

It is clear that the fundamental limitation on the rate of stochastic cooling will lead to the use of higher frequencies. At present, perhaps the main obstacle to this is the slow development in know-how concerning pick-up and kicker design. In particular, one problem is that as the frequencies increase the vacuum chamber cross section is no longer small compared with the wavelength of the signals. This may reduce the efficiency of transverse pick-ups and kickers. It also means that signals can travel down the vacuum tube from the kicker back to the pick-up and cause instability. Ferrite structures surrounding the aperture may be needed to damp these signals.

With increasing frequency (and often, increasing ring diameter) the problem of obtaining the correct phase response is becoming more serious. For instance, with a system working between 8 and 16 GHz, as considered for use in the Sp̄pS collider, the delay between pick-up and kicker would have to be correct to within about 1 mm over a distance of 2 km. A daring new design that might help to solve this problem has been proposed by Boussard⁴ et al.

In the Sp̄pS, the stored particles are bunched. The bunches are about 5 ns long, whereas the time between bunches is several μ s. This makes the cooling difficult. In the first place, the bunching causes each Schottky band to be split up into satellite lines that are separated by the synchrotron frequency. The envelope of these lines is roughly the same as for an unbunched beam, but since the synchrotron frequency is nearly the same for all particles, the satellites are narrow and much too dense for efficient cooling. This may be cured by using a second-harmonic RF system in addition (with about half the voltage of the main system) to spread out the synchrotron frequencies and therefore the satellite lines.

Another problem that remains is that even with a 8 GHz bandwidth the number of samples in a 5 ns bunch is not more than 80. In the frequency domain, the equivalent problem is that neighbouring Schottky bands of a bunched beam are found to be correlated, so that the heating term is increased and we have to use a lower gain.

On the other hand, the long time between bunches will permit the use of the scheme of Fig. 13. Here the pick-up signals are mixed with a carrier, the low-frequency product is sampled, the d.c. component removed (we are only interested in fluctuations), the signal is digitized, transmitted to the kicker, reconverted to analogue and brought back into the high-frequency band by mixing with the same carrier. Forty such channels, with carrier frequencies spaced across the 8-16 GHz band, will provide the Fourier-transformed equivalent of the 80 samples (each channel provides a sine and a cosine term). The advantage is that the transmitted low-frequency data do not have to be precisely synchronised; only the phase of the carriers has to be correct. In fact, these carriers may be generated from the beam itself, both at the pick-ups and at the kickers.

Figure 13 - One channel of the proposed SPS cooling system.

This project, although only at the proposal stage, shows how techniques might develop. Note, however, that for unbunched beams such a scheme would not work; it is the combination of very wide band pick-ups and kickers working at a low duty cycle that makes it interesting

13. ELECTRON COOLING VS STOCHASTIC COOLING

A completely different cooling method is the so-called electron cooling, first developed by Budker et al. in Novosibirsk⁵ and later also tested at CERN and FNAL. This method uses an electron beam coincident with the proton (or antiproton) beam along a certain length. Both types of particles have the same velocity; thus the electrons have a kinetic energy 1836 times lower than the protons. If the electron beam is reasonably cold, the momentum differences between the protons will be transferred to the electrons by collisions; the beams behave more or less as a mixture of two gases.

It is interesting to compare the relative merits of the two methods without going into any details. The following table shows the main differences.

Table I - Comparison between electron and stochastic cooling.

Electron Cooling	Stochastic Cooling
<ul style="list-style-type: none">- cooling time increases strongly with angular spread of particles- not dependent on number of particles N- low energy needed (cooling time increases rapidly with energy)- usually fast cooling (milliseconds to seconds)	<ul style="list-style-type: none">- independent of angular spread- cooling time proportional to N- no energy dependence (except for power required)- not so fast (seconds to hours)

It is mainly the energy dependence of electron cooling (together with the large angular spread of the $\bar{p}$) that has been the reason for adopting stochastic cooling for accumulating antiprotons. With electron cooling, it would have been necessary to decelerate the antiprotons first before they could have been cooled efficiently. For low-energy applications, however, electron cooling may be competitive.

REFERENCES

1. D. Möhl, G. Petrucci, L. Thorndahl, S. van der Meer, Physics and technique of stochastic cooling, Phys. Reports 58 (1980) 73-119. Identical text in CERN/PS/AA 79-23 (1979)
2. H. Hereward, The elementary theory of Landau damping, CERN 65-20 (1965).
3. G. Carron, L. Thorndahl, Stochastic cooling of momentum spread by filter techniques, CERN-ISR-RF/78-12 (1978).
4. D. Boussard, S. Chattopadhyay, G. Dôme, T. Linnekar, Feasibility study of stochastic cooling of bunches in the SPS, CERN/SPS/84-4 (ARF) (1984).
5. G.I. Budker, Proc. Intern. Symp. on Electron and Positron Storage Rings, Saclay 1966, p.11-1-1; Atomn.Energ. 22 (1967).
G.I. Budker et al., Experimental studies of electron cooling, Part. Acc. 7 (1976) 197-266.