The status of stochastic cooling

D. Möhl [a,1]

[a] *CERN, 1211 Geneva 23, Switzerland*

Presented at a symposium to celebrate the 30th anniversary of electron cooling, this report is intended to give the status of the companion technique, stochastic cooling, some 28 years after its invention. An overview of past developments reveals the close relationship between the two cooling ideas. Then the report concentrates on the principal ingredients of stochastic cooling in order to discuss the limits encountered and some recent ideas for pushing back these limits.

PACS: 29.20 Dh
Keywords: stochastic cooling, storage rings, phase space density, broadband system, beam Schottky noise, low noise electronics.

Presented at the '11th ICFA Beam Dynamics Workshop on Beam Cooling and Instability Damping' held on board a ship between Moscow and Nizhny Novgorod, 18–26 June 1996

1 Introduction

There are many excellent reviews of stochastic cooling, including the more recent state-of-the art reports given by F. Caspers [1] at the symposium on crystalline beams in Erice 1995 and by J. Marriner [2] at the beam cooling workshop in Montreux 1993. Caspers' papers includes a concise description of existing and planned projects based on stochastic cooling, whilst Marriner focuses on theoretical and technological developments. In the present survey, I will first give a personal view of the evolution of the idea, pointing to the 'cross-fertilization' with electron cooling. I will then concentrate on the principal ingredients such as bandwidth, mixing, and signal to noise ratio, to discuss the limits encountered and recent ideas to push these limits. More details about stochastic cooling in the CERN - Antiproton Collector and Accumulator rings

[1] Tel: +4122 767 3035; Fax: +4122 767 9145; e-mail: MOEHL@PS.MSM.CERN.CH

and in the FERMILAB antiproton machines can be found in the contributions of F. Pedersen [3] and M. Church [4] to this symposium.

2 History

Some dates in the development of stochastic cooling are summarized in Table 1. Here I want to draw attention to the close relationship in which electron and stochastic cooling grew up. In fact, the idea of stochastic cooling is (only) two years younger than the first publication on electron cooling of which we celebrate the 30th anniversary at this symposium. I think we can assume that Budker's idea, which is indeed very suggestive, inspired van der Meer to think of another cooling method that would work at 30 GeV in the ISR. Ever since, the two concepts have evolved in close relationship and with many common features in their development. One can be more precise and observe that virtually all the achievements that were accomplished with stochastic cooling had first been noted in connection with electron cooling. Let me highlight a few points of Table 1 which underline the brotherhood and the teacher-pupil relationship of electron- and stochastic cooling.

For almost a decade following their conception, the principle of beam cooling was regarded as too far-fetched to be practicable. Experimental demonstrations of both electron- and stochastic cooling were finally tried in the same year, 1975, at Novosibirsk and CERN respectively. Inspired by the success of these demonstration experiments, P. Strolin, returning from a visit to Novosibirsk and L. Thorndahl perceived the interest of stochastic cooling, for the purpose of accumulating antiprotons. The same application, but with electron cooling, had already been foreseen by Budker in his earlier papers. Following up on this idea, C. Rubbia and co-workers made their proposal for proton-antiproton collisions in a big synchrotron (e.g. at 300 GeV in the SPS at CERN) with beam preparation in a small Antiproton Accumulator ring (AA). Again a similar scheme, but relying entirely electron cooling, had been dreamt of earlier by Budker and his team. Originally Rubbia's scheme was also based on electron cooling, but - under the influence of a working group including van der Meer and Thorndahl - stochastic cooling was finally adopted.

The construction of the AA began in 1979 and with its commissioning leading to the first p-$\bar{\text{p}}$ collisions in the SPS in 1981, stochastic cooling had grown from an experimental to an operational tool that opened completely new possibilities and led to important discoveries. Best known perhaps are: the observation of the *intermediate vector bosons* at the SPS collider, the evidence for the *top quark* obtained at FERMILAB and the synthesis and detection of *antihydrogen atoms* at the Low Energy Antiproton Ring (LEAR). All this was made possible by stochastic cooling, which works well at the energies, the large angular and momentum spread and the relatively low intensities of the antiproton beam emerging from the production target.

The complementarity of electron- and stochastic cooling is documented by LEAR where both techniques co-exist in a very profitable symbiosis since 1988. Also, since the late 1980s, a number of ion cooling rings (I count 8 up to now) with some resemblance to LEAR came into operation in different laboratories. All of them have electron cooling to prepare light or heavy ion beams with small emittance and small momentum spread for precision experiments at relatively low energy. Thus, the boom of electron cooling started in a somewhat different field, but some of the 'electron cooling rings' also foresee stochastic cooling for higher energy or for large emittance secondary beams e.g. radioactive ions from a target (see the contributions of D. Prasuhn [5] and G. Muenzenberg [6] to this symposium).

Recently, proposals have emerged with electron cooling at energies which were previously 'reserved' for stochastic cooling (the 9 GeV re-cycler planned at FERMILAB, see J. MacLachlan's and S. Nagaitsev's contributions [7], [8] to this symposium). In addition proposals for stochastic cooling with optical bandwith [9] have been put forward which would permit fast cooling at high intensity, a domain previously 'reserved' for electron cooling. Thus, with maturity the two methods tend perhaps to lose some of their specificity.

3 Principal ingredients

3.1 The basic cooling rate equation

The basic set-up for stochastic cooling is sketched in Fig. 1. In a few elementary steps [10] one can derive the following simplified relation for the cooling rate of emittance ($1/\tau = (1/\epsilon)\mathrm{d}\epsilon/\mathrm{d}t$) or momentum deviation ($1/\tau = (1/\Delta p)\mathrm{d}\Delta p/\mathrm{d}t$):

$$\frac{1}{\tau} = \frac{W}{N}\left[\underbrace{2g(1-\widetilde{M}^{-2})}_{\text{coherent effect (cooling)}} - \underbrace{g^2(M+U/Z^2)}_{\text{incoherent effect (heating)}}\right]. \qquad (1)$$

The parameters appearing in (1) have the following significance:

N	Number of particles in the coasting beam	
W	cooling system bandwith	
g	gain parameter	$(g < 1)$
M	desired mixing factor kicker - pick-up	$(M > 1)$
$\widetilde{M}$	undesired mixing factor pick-up - kicker	$(\widetilde{M} > 1)$
U	noise to signal power ratio (for single charged particles)	$(U > 0)$
Z	charge number of beam particles	$(Z \geq 1)$

There is a value g_0 of g for which Eq. (1) has a maximum. With this 'optimum gain' the cooling rate becomes

$$\frac{1}{\tau} = \frac{W}{N}\left[(1 - \widetilde{M}^{-2})^2/(M + U/Z^2)\right] \tag{2}$$

As to the other parameters: N and Z are properties of the beam, W is a property of the cooling system and M, $\widetilde{M}$ and U depend on the interplay of cooling system-, beam- and storage ring characteristics. Achievements in stochastic cooling can be discussed by referring to the values attained for W, M, $\widetilde{M}$ and U.

3.2 Bandwidth limits

Clearly, by virtue of Eq. (1) and (2), it is desirable to work with a large bandwidth in order to maximize the flux of particles N/τ that can be cooled and/or stacked. However, there are severe limitations on W and on the maximum frequency $f_{\max} \geq W$ of the cooling band. One can distinguish two kinds of restrictions: those related to rf-technology to be discussed in this subsection, and those related to the mixing problem to be discussed the next subsection.

A first technological difficulty stems from the fact that beyond cut-off ($f \approx c/4b$ where $2b$ is the diameter of the chamber) the vacuum pipe transmits waveguide modes. Then, given the strong amplification (usually more than 100 db) in the cooling loop, the system of Fig. 1 starts to 'ring' due to signals propagating from kicker to pick-up. A common antidote is to introduce absorbers, i.e. chamber sections containing lossy wall material, which strongly attenuates microwaves. In this way, the frequency limit can be extended but, far above cut-off, the efficiency of the absorbers becomes critical. For 'typical' chamber dimensions, where mode propagation starts around 2 GHz, absorbers working well up to 10 GHz have been devised and absorbers for up to 20 GHz have been contemplated. Ferrite absorbers are most efficient, but they have to be used in regions that are free of magnetic field.

A second limitation comes from the reduction of the pick-up (and kicker) sensitivity. In fact, the field of a moving charge 'reaches out' to a lateral distance $\Delta x \approx c\gamma/\omega$ (where $\gamma = (1 - v^2/c^2)^{-1/2}$). Assuming e.g. a pick-up opening of $b = \pm$ 50 mm, the sensitivity would start to drop in LEAR ($\gamma \approx 1$) at $f \approx 1$ GHz and in the AC ($\gamma = 3.9$) at $f \approx 4$ GHz. Thus (at least for the 'conventional' cooling at microwave frequencies) small aperture and/or high energy are required to be able to profit from a bandwidth far beyond 10 GHz; for high γ, the use of present technology up to about 30 GHz has been contemplated. Various types of 'far-field' pick-ups and kickers have also been analyzed but - apart from the undulators proposed for stochastic cooling at optical frequencies, to be discussed below - no design work is going on to my knowledge.

3.3 The mixing dilemma

Mixing is a central problem and the development of stochastic cooling is intimately linked to the progress in dealing with this enigma. The optics of the storage ring (especially the distance from transition energy $\eta = |\gamma_{\rm tr}^{-2} - \gamma^{-2}|$), the cooling bandwidth (W) and the beam characteristics (especially $\Delta p/p$) all enter into the mixing factors. The designer of a cooling ring has to make a balanced choice of these parameters.

To illustrate the point, let us recall the significance of the mixing factors. In the sampling picture, stochastic cooling is viewed as a procedure, where the coasting beam is subdivided into samples of a time width $T_{\rm s} = 1/2W$ (the response time of the system). The mixing rates $1/M$ and $1/\widetilde{M}$ can then be interpreted as the fraction of the sample length by which a particle width the typical momentum deviation $\Delta p = 2\ \Delta p_{\rm rms}$ slips with respect to the nominal ($\Delta p = 0$) particle on the path kicker to pick-up ('K to P') and pick-up to kicker ('P to K') respectively. Then (omitting numerical factors close to unity)

$$1/M = \Delta T_{\rm kp}/T_{\rm s} = 2\ W\ T_{\rm kp}\ \eta_{\rm kp}\ \Delta p/p,$$
$$1/\widetilde{M} = 2 f_{\rm max}\ T_{\rm pk}\ \eta_{\rm pk}\ \Delta p/p \approx 2W\ T_{\rm pk}\ \eta_{\rm pk}\ \Delta p/p\ . \qquad (3)$$

Here $W = f_{\rm max} - f_{\rm min}$ and, to simplify the discussion, $W \approx f_{\rm max}$ will frequently be assumed in the following. Furthermore $\eta_{\rm kp} = (p/T_{\rm kp})(\partial\ T_{\rm kp}/\partial p)$ is the local off-momentum factor (also called 'phase slip factor') kicker to pick-up and $\eta_{\rm pk}$ the analogous quantity for the path pick-up to kicker.

For a regular lattice one has (usually)

$$\eta_{\mathrm{kp}} = \eta_{\mathrm{pk}} = \eta = |\gamma_{\mathrm{tr}}^{-2} - \gamma^{-2}| \tag{4}$$

i.e. the local η - factors are equal to the off-momentum factor of the whole ring. This is (at least approximately) the case for all existing cooling rings and we will refer to it as the classical design. Then the ratio $\widetilde{M}/M$ is simply given by the corresponding path lengths (L_{pk} and L_{kp})

$$\widetilde{M}/M = T_{\mathrm{kp}}/T_{\mathrm{pk}} = L_{\mathrm{kp}}/L_{\mathrm{pk}} \tag{5}$$

and in the case (e.g. of the AC) where the cooling loop cuts diagonally across the ring; $\widetilde{M} = M$. This reveals the mixing dilemma. To have fast cooling, no mixing P to K (i.e. $\widetilde{M} \gg 1$) and full mixing K to P ($M = 1$) is desired, which is clearly incompatible with Eq. (5).

The usual compromise is to accept imperfect mixing by letting both $\widetilde{M}$ and M be in the range of 2 to 5, say. As a consequence the best possible cooling rate e.g. in case of $\widetilde{M} = M$ is $1/\tau = 0.28\ W/N$ instead of W/N for perfect mixing. More details are given in Table 2, where the optima for a few selected cases are compared to the corresponding situation with perfect mixing (i.e. Eq. (2) with $\widetilde{M} \gg 1,\ M = 1$).

We can now discuss the choice of cooling ring parameters. For the 'classical design' a reasonable choice is $M \approx 2$ to 5 (where, in the case of momentum cooling, the initial $\Delta p/p$ counts for the determination of M). For the bandwidth (and the maximum frequency) one chooses the largest technically feasible value that is compatible with the limits discussed in Section 2.3. This determines the value of η [by virtue of Eqs. (3) and (4)] and thus the transition energy of the ring. An examination of the parameters of existing rings (AA and AC, LEAR...) reveals that the (initial) mixing factors deduced are indeed not far from the theoretical optima. We mention in passing that, for momentum cooling by filter and/or by transit time methods, similar criteria hold although the mixing situation is more involved.

This is not the end of the mixing dilemma. During momentum cooling, as $\Delta p/p$ decreases, the M-factor tends to increase and the mixing situation degrades. One can in principle stay close to the optimum by increasing W ('dynamic bandwidth tuning') and/or η ('transition tuning') as cooling proceeds. Transition tuning has not been foreseen and cannot be tried (at least not easily) in the present generation of cooling rings. Bandwidth adjustment is only meaningful, if the initial momentum spread is 'too large' so that excessive mixing occurs at full bandwidth. The technique has been tried experimentally in the AC and AA but is not used in operation.

More interesting than transition tuning is a 'non-classical' design, where the path P to K is isochronous ($\eta_{\mathrm{pk}} = 0$) whereas the path K to P is strongly flight time dispersive ($\eta_{\mathrm{kp}} \gg 0$). Then Eq. (4) no longer holds and mixing can be ideal i.e. $M = 1$, $\widetilde{M} \gg 1$. More precisely one requires:

$$\begin{aligned} \eta_{\mathrm{pk}} &\ll 1/(2 f_{\max}\, T_{\mathrm{pk}} \Delta p/p) \\ \eta_{\mathrm{kp}} &\gg 1/(2 W T_{\mathrm{kp}} \Delta p/p) \ . \end{aligned} \tag{6}$$

Obviously to meet the conditions of Eq. (6), a special cooling ring lattice is necessary with characteristics which have to be reconciled with the many other requirements of the storage ring. Proposals for such 'ideal mixing lattices' have been worked out e.g. for the 10 GeV SuperLEAR ring [11] (which was, however, not built). Perhaps the next generation of stochastic cooling rings will use such 'semi-isochronous, semi-dispersive' lattices.

3.4 Signal and noise

Assuming an ideal pick-up system, the noise to signal [10] ratio can be expressed as:

$$U = \frac{P'_{\mathrm{n}}/R}{2Ne^2(n_{\mathrm{pu}} K_{\mathrm{pu}}) f_{\mathrm{rev}}} \ . \tag{7}$$

It is given by the spectral power density ($P'_{\mathrm{n}} = \mathrm{d}P_{\mathrm{n}}/\mathrm{d}f$) of the amplifier noise referred to at the entrance (input impedance R) on the one hand, and on the density over the band of the beam Schottky power $2Ne^2(n_{\mathrm{pu}} K_{\mathrm{pu}} R) f_{\mathrm{rev}}$ delivered to the amplifier entrance on the other hand (n_{pu}: number of pick-ups combined in parallel, $K_{\mathrm{pu}} < 1$: pick-up 'sensitivity'). Thus U depends strongly on the amplifier and the pick-up technology.

From Eqs. (1) and (2) it is clear that a balanced design has $U < M$ and more favourably $U < 1$. Then for low intensity ($N < 10^9$, say) both low-noise (cryogenic) amplifiers and an elaborate pick-up system are necessary.

There has been great progress in the design of low-noise broadband systems [12], [13], [14] and the amplifiers developed for the AC and the FERMILAB debuncher are in fact formidable 'HIFI systems' with an unprecedented combination of low noise, large bandwidth, high amplification and ultra-linear phase response characteristics.

The noise power density (P'_{n}) is typically 4×10^{-21} W/Hz for room temperature amplifiers. The cryogenic amplifiers developed for the AC use Gallium arsenide field effect transistors ('GaAs FET' amplifiers). Cooled to a temperature of 30 K, they have a $P'_{\rm n}$ of 5 to 10 times less then the above figure. Then the electronic noise of the entire 'input stage' becomes important and one has also to care about the pick-ups, their terminating resistors, combiners etc. Elaborate cryogenic systems to reduce the noise of these critical components are used in the AC ($N = 5 \times 10^7$) and to some extent also in LEAR ($N = 10^9$). It is remarkable that the GaAs FETs not only permit low noise but also work with extremely large bandwidth and even high power so that they are also used in the final amplifier stage.

On the pick-up front, a large number of electrodes close to the beam is used. To save room, horizontal (H) and vertical (V) electrodes are frequently installed at the same azimuth although the difference signals are then reduced compared to dedicated H and V electrodes in separate locations. The sum signal, used for momentum cooling by the filter or transit time method can be taken from both the H and the V electrodes. In the AC for example the horizontal and vertical electrodes are at different locations but the sum signal from both is used for Δp cooling.

For the difference signal, the sensitivity factor entering into Eq. (7) depends on the square of the ratio of beam size to pick-up aperture ($K_{\rm pu\perp} \propto a^2/h^2$). Several measures are taken to keep this factor large. The aperture is tailored to the local beam size. The closed orbit is carefully corrected, so that the beam passes through the center and room is made for large emittance. This also minimizes the unwanted 'closed orbit signal' on the difference pick-up. The procedure described so far was followed in the first generation of cooling rings (from ICE to AA). A further step was done in the AC: There the position of the plates can be adjusted ('slow movement') to fit the beam at injection and is programmed ('fast movement') to follow the shrinkage of the beam during cooling. During a 2.4 s cycle (or the 4.8 s used in the routine operation of the AC) the aperture goes from typically 10 to 3 cm and back. This works reliably in the AC, but mechanical, vacuum and rf problems which had to be solved are formidable. Other modern cooling rings foresee adjustable pick-ups but no 'fast movement' during a cooling cycle.

We now turn to different pick-up structures. Widely used since the ISR and ICE times is the loop coupler where each plate together with the vacuum tank (or a 'ground plate' in the tank) forms a strip line terminated by its characteristic impedance (typically 50 to 100 Ω) at the downstream end. The signals are coupled-out via matched cables at the upstream end. The loop coupler has the factor

$$(I^2_{\rm output}/I^2_{\rm beam}) \propto \sin^2\{\frac{\omega}{2}(\ell/\beta\ {\rm c} + \ell/v_{\rm line})\}\ .$$

in its frequency response, where βc is the beam velocity and $v_{\rm line} \approx c$ the wave velocity on the pick up (regarded as a strip line). The length (ℓ) of the plates is chosen as small as compatible with a good sensitivity over the band. Usually one chooses

$$\frac{\omega}{2}(\ell/\beta\ c + \ell/v_{\rm line}) = \pi/2$$

near the maximum $\omega = \omega_{\rm max}$ or at mid band $\omega = (\omega_{\rm max} + \omega_{\rm min})/2$.

By combining (i.e. adding in parallel and with the proper delay) the signals from a number ($n_{\rm pu}$) of loops the signal power increases (ideally) by $n_{\rm pu}$, as assumed in Eq. (7). For non relativistic beam velocities it is possible to combine the loops 'in series', injecting, with the proper delay, the signal from the downstream end on to the upstream port of the subsequent loop. Then (ideally) the output current increases with $n_{\rm pu}$ and the signal power [entering into Eq. (7)] with $n_{\rm pu}^2$. In LEAR at low energy ($\beta = 0.3 - 0.1$) loop couplers are used in 'series connection' via variable delays outside the vacuum chamber.

Apart from the loop coupler, experiments with different types of slow-wave structures [15] ('meander lines', helix structures, inductively loaded 'lumped circuits', ferrite structures) have been made for non-relativistic beams ($\beta < 0.6$ say). But for the operation of LEAR loop couplers are preferred because of their excellent performance for large bandwidth and wide range of beam velocities. For β closer to 1 the Faltin type structure (a stripline coupled to the beam via slots in the outer conductor) has proven useful in the AA, but the AC system is entirely based on loop couplers.

For limited bandwidth, it is possible to increase the sensitivity by making the pick-up resonant. In a simple lumped RLC-model, the sensitivity increase linearly with the Q-factor, $Q = f_{\rm res}/\Delta f$, of the resonance and it is clear that a gain in sensitivity is only possible at the expense of the bandwidth (Δf). An example of a resonant pick-up is the resonant cavity with a $Q \approx 10^4$ employed in ICE to observe and to cool in momentum a beam of less than 100 particles (using only a single Schottky 'line'). Another less extreme case is the 'super-electrode' arrangement [13] used in the AC. It may be thought of as 2 loop couplers connected in series by a cable of length $\lambda/2\beta$ at mid-band. The sensitivity is higher by 2 but the bandwidth is reduced beyond one octave ($f_{\rm max} < 2\ f_{\rm min}$) compared to two coupling loops with matched termination and signal combination in parallel.

An impressive technological development has been associated with the construction of the pick-ups, combiners and vacuum feed-throughs. The equipment has to be bakeable at 200–400°C and made of carefully chosen materials to fulfil the strict requirements of the ultra vacuum and GHz electronics. In the ICE-times, pick-ups were just plates installed on isolating spacers with coaxial feeds traversing the vacuum chamber at each end. With the AA, tanks containing many tens of loops came into use. Moving pick-ups were developed for the AC. FERMILAB pioneered printed structures [16]: pick-ups integrated with combiners and terminations, copying, or rather adapting printed circuit technology. In the fixed energy cooling rings, the number of feed-throughs can be kept small by installing a maximum of combiners, terminations etc. inside the vacuum chamber. In rings like LEAR a large number of feeds and adjustable coaxial delays outside the vacuum chamber are necessary to adjust the cooling system for different beam velocities.

Turning briefly to kickers we recall the reciprocity theorem which - accepting some simplification - we quote as: '... each structure that works as a pick-up can also be used as a kicker'. Thus kickers and pick-ups have followed a similar evolution. The signal to noise problem is - to some extent - replaced by the power problem of the kicker. It can become critical when fast cooling and/or high beam energy are at stake. Again the solution is to combine many kicker units.

3.5 Bunched beams

There is an interest in cooling the bunches in hadron colliders in order to counteract the degradation of the coasting beam with time. In large colliders the luminosity half-life is in the range of hours to days and cooling times of this order could greatly improve the situation. High energy and tight bunching of the beam are characteristic of modern colliders which make phase space cooling a difficult task.

Already some time ago, several features of bunched beam stochastic cooling were identified [17]. To make an estimate of the cooling rate, one can replace the bunches by pieces of a coasting beam. Then Eq. (1) can still be used but for N, one has to insert the effective particle number, $N \to N_{\rm b}\ 2\pi R/\ell_{\rm b} = N_{\rm b}/B_{\rm b}$ i.e. the number of particles in the bunch times the bunching ratio (circumference of the ring/length of the bunch). Usually the bunching ratio is a very big number e.g. ≈ 7000 in the SPS collider ($\ell_{\rm b} \approx 1$ m, $R = 1.1$ km). Thus large bandwidth is needed to obtain useful cooling rates ($W = 17$ GHz in the above example to obtain a cooling time of 10 h for 10^{11} particles per bunch).

Another difficulty is due to the strong bunch signal at certain harmonics of the revolution frequency up to a cut-off ($f_{\rm b} \approx \beta c/\ell_{\rm b}$) determined by the bunch length. This signal is proportional to $N_{\rm b}$, (or $k \times N_{\rm b}$ for k bunches). It tends to 'blind' the cooling signals which are given by $\sqrt{N_{\rm b}}$ and thus much weaker. The way out is to use frequencies much higher than $f_{\rm b}$. Then the fall off of the bunch spectrum is important. For a well behaved bunch, this roll off is fast, e.g. for a Gaussian bunch of r.m.s. length σ_ℓ (time width $\sigma_{\rm t} = \sigma_\ell/\beta c$) the frequency spectrum is also Gaussian with

$$\sigma_f = \frac{1}{2\pi\sigma_t} \ .$$

On the other hand, a rectangular bunch of duration $\pm\Delta t$ has a spectrum that only falls off with

$$\frac{\sin(2\pi \ f \ \Delta t)}{2\pi \ f \ \Delta t}$$

i.e. essentially like $1/f$ at high frequency.

This is theory! Observations on various machines (SPS, TEVATRON, AA, LEAR...) all indicate that intense short bunches develop a strong 'rf-activity' which persists at frequencies very much higher than $f_{\rm b}$. A possible explanation is that microwave instabilities develop [18] which - although not leading to observable blow-up or losses - strongly modulate the beam. The increase of the rf-activity with intensity, clearly observed in LEAR, supports this conjecture. The spurious coherent modulation obstructs the observation of Schottky noise and thus the cooling. A way out could be to use the cooling system as a coherent damper to calm the beam prior to cooling, but then the dynamic range of such a system must be very large and the task to cool a beam close to the stability threshold looks acrobatic. Thus, more investigations of the high-frequency signals from short intense bunches are indicated before a system to cool these bunches can be envisaged.

3.6 Specific diagnostics

Since the first cooling experiments the beam Schottky noise has been used to monitor the transverse and longitudinal cooling process. Spectrum analysers are applied, tuned to one or several Schottky bands. Numerous codes, often running on small desk-top computers are now available to acquire and evaluate the information from the spectrum analyzer. Schottky diagnostics is also used, to verify/optimize the setting of the cooling loop. The optimization is done band by band. For this the behaviour of the Schottky signals with the cooling loop 'open' and 'closed' is compared. Based on Sacherer's theory [19] of the 'feedback via the beam', adjustment criteria for optimum gain [20] have been determined. They are summarized by stating; " For $U \ll M$ the optimum gain leads to a signal reduction of 6 db when closing the loop. For $U \gg M$ the optimum gain leads to a reduction of the peak of the closed loop signal to the level of the amplifier noise when the loop is open."

Network analyzers are another precious diagnostic tool. Apart from the task of testing the cooling loop in the absence of the beam, they are used to determine the beam transfer function (beam response to an excitation signal) with the cooling loop open/closed. This provides additional input on the signal suppression due to the 'beam feedback' and permits one to optimize gain and phase.

Although the basic methods have hardly changed during the last decade, the diagnostics and the setting up of the cooling loop have greatly benefitted from the advent of more powerful computers and instruments in recent years.

3.7 New developments

Recently it has been proposed [9] to use optical frequencies and bandwidth ($f_{\max} \approx 2W \approx 10^{14}$ Hz) for stochastic cooling. Such a large bandwidth would permit to cool 10^{13} particles in a few seconds (or 10^{10} in milliseconds) and thus open many new applications.

Special equipment is needed to detect and correct the beam error signals in the optical frequency range and the undesired mixing (or equivalently, unwanted delays in both the beam and the signal path) become very critical. The pick-up proposed is an undulator magnet where a pulse of light is emitted as a particle passes. This pulse is sent through an optical amplifier to a second undulator which acts as a kicker. To solve the mixing problem, special isochronous lattice insertions are proposed similar to those discussed in Section 2.3 above. In the simplest case, they make the flight time from pick-up to kicker slightly dependent on the momentum error of the particle but strictly independent of its betatron amplitude. This is for momentum cooling whereas for betatron cooling an insertion with the opposite properties is required (time of flight weakly dependent on betatron amplitude but independent of momentum). Ideally, each particle then receives a correcting kick proportional to its momentum error or its betatron amplitude.

Tolerances can be established by imposing that the particle and the signal traveling times have to be controlled to a level leading to a phase error $\varphi = 2\pi\ f\ \Delta t < \pi$ up to the highest frequency. This means a precision Δt of the order of 10^{-14} s i.e. $\Delta t/t = 10^{-7}$ for a flight time of 100 ns. Thus, apart from the beam dynamics, the technology of the undulators, light guides and optical amplifiers must be perfectly mastered.

There is a proposal, to test some aspects of this technique at the booster of the Advanced Light Source at Berkeley. The idea is to install two undulators acting as pick-ups several tens of meters apart in a special beam line and to observe and correlate the light signals from a 150 MeV electron beam. Results from the experiment are eagerly awaited. The time of flight methods suggested for optical stochastic cooling are interesting also for the microwave range, to push up the mixing limit.

3.8 Computational tools

The simple Eqs. (1) (2) are good for a first order of magnitude estimate. They can relatively easily be improved by performing the calculation in the frequency domain. In essence one has to replace W by a sum:

$$W \to \; f_{\rm rev} \sum_{n_{\min}}^{n_{\max}}$$

over the cooling band (from $f_{\min} = n_{\min}\ f_{\rm rev}$ to $f_{\max} = n_{\max}\ f_{\rm rev}$) letting g, $\widetilde{M}$, M and U be functions of frequency. A more precise interpretation of these quantities emerges when one analyses the frequency behaviour of the cooling loop [10], [19]. The signal shielding by the 'feedback via the beam' can - in an approximate manner - be included by replacing the gain

$$g(f) \to \frac{g(f)}{1 + g(f)M(f)/2}$$

where $M(f)$ is the mixing factor.

The simple equations, even in their improved form, only describe the evolution of the rms width of the distribution. For the detailed analysis of the cooling and stacking process, a Fokker-Planck type of equation has been used since about 1978 to describe the evolution of the distribution function of a beam subject to stochastic cooling. Again the Novosibirsk workers had used a similar equation earlier (another example supporting our hypothesis that all that was achieved in stochastic was anticipated in electron cooling).

In the Fokker-Planck description, cooling (in one plane) is fully described by the two coefficients F and D. They are given by the average change per unit time $< \Delta x > /\Delta t$, of x (the momentum error or the betatron amplitude), and by the average change of its square respectively: $F =< \Delta x > /\Delta t$ and $2\ D =< (\Delta x)^2 > /\Delta t$. These quantities can be obtained from analysis and/or measurement. In general, both F and D depend on momentum and/or position, as well as on the distribution function itself.

Heating mechanisms, e.g. multiple Coulomb scattering (intra-beam or on the residual gas), can be included in D and particle influx (stacking) and losses due to tails reaching acceptance limits can be taken into account by appropriate boundary conditions in space and time. Computer codes ('Fokker-Planck solvers') are available, at least for the one-dimensional case, where momentum cooling and horizontal and vertical betatron cooling are independent of each other. In many cases, one can derive analytical approximations from the Fokker-Planck equation for instance for the equilibrium distribution or the optimum gain profile for stacking.

More recently tracking codes based on a super-particle approach [21] have also been developed to simulate cooling and stacking.

4 Conclusions

Stochastic cooling of coasting beams has reached a state of maturity. Both theory and technology are adequate to deal with the applications foreseen for the (near) future.

Widening of the scope could be envisaged if larger bandwidth (higher frequencies) could be used. This requires mastering of related mixing, microwave propagation and sensitivity problems.

Bunched beam stochastic cooling in large colliders is hampered by an unexpected strong 'rf-activity' up to the highest frequencies. Its origin is not suffiently clear.

Acknowledgements

The author wishes to thank F. Caspers, S. van der Meer, D. Warner and P. Bryant for their help in preparing this report.

References

(only selected references are given here, a bibliography can be found in the Accelerator School article, in CERN 95–06, see under Ref. [10] below).

[1] F. Caspers: Stochastic Cooling, preprint CERN/PS 96–03, to be published in: Proc. Workshop on Beam Crystallization and Related Topics, Erice, Nov. 1995.

[2] J. Marriner: Theory, Technology and Technique of Stochastic Cooling, Proc. Workshop on Beam Cooling and Related Topics, Montreux, Oct. 1993, CERN Report 94–03 (J. Bosser, ed.) 1994, p. 14.

[3] F. Pedersen: Stochastic Cooling used in the AAC Complex at CERN, these Proceedings.

[4] M. Church (presented by J. MacLachlan): Recent Developments in Stochastic Cooling at FERMILAB, these Proceedings.

[5] D. Prasuhn; Stochastic Cooling for COSY, these Proceedings.

[6] G. Muenzenberg: Exotic Nuclei in Storage Rings, these Proceedings.

[7] S. Nagaitsef: Electron Cooling in the FERMILAB Recycler, these Proceedings.

[8] J. MacLachlan: FERMILAB R&D in Medium Energy Electron Cooling, these Proceedings.

[9] A. Mikhailichenko, M. Zolotorev: Optical Stochastic Cooling, Phys. Rev. Lett. **71** 1993, p. 4146;
M. Zolotorev, A. Zholents: Transit Time Method of Optical Stochastic Cooling, Phys. Rev. **E50** 1994, p. 3084.

[10] D. Möhl, G. Petrucci, L. Thorndahl, S. van der Meer: Physics and Technique of Stochastic Cooling, Phys. Rep. **58** 1980, p. 75;
D. Möhl: Stochastic Cooling, Proc. CERN Acc. School., Rhodes 1993, CERN Report 95–06, (S. Turner, ed.) part II, 1995 p. 587.

[11] G. Cesari, R. Giannini, P. Lefèvre, D. Möhl, G. Vandeplasche: LEAR and SuperLEAR for an Intermediate Energy Antiproton Physics Programme, Proc SuperLEAR Workshop, Zurich 1991, Inst. Phys. Conf. Series 124 (C. Amsler and D. Urner, eds.) 1992, p. 13.

[12] G. Carron, F. Caspers, L. Thorndahl: Development of Power Amplifier Modules for the ACOL Stochastic Cooling System, CERN Report 85–01, 1985.

[13] B. Autin, G. Carron, F. Caspers, L. Thorndahl: Applications of Microwaves to Antiproton Control, Alta Frequenza LVI, 1987, p. 381.

[14] M. Church, J. Marriner: The Antiproton Sources: Design and Operation, Annual Rev. Nucl. Part. Sci. **43** 1993, p. 253.

[15] N. Tokuda: Traveling Wave Pick- ups and Kickers for Stochastic Cooling, Proc. 4th LEAR Workshop, Villars Switzerland 1988, Harwood Academic Publishers, Chur (C. Amsler, ed.) 1988, p. 135.

[16] J. Petter, D. McGinnis, J. Marriner: Novel Stochastic Cooling Pick-ups and Kickers, Proc. Part. Conf . Chicago 1989, p. 636.

[17] H. Herr, D. Möhl: Bunched Beam Stochastic Cooling, Proc. Workshop on Cooling High Energy Beams, Madison 1978, Univ. Wisconsin Madison Report 1979 and internal Reports CERN-PS-Dl Note 79–3 and CERN-EP Note 79–34.

[18] L. Vos: Coherent Signals from Proton Bunches beyond the Stationary Bunch Spectrum, these Proceedings

[19] F. Sacherer: Stochastic Cooling Theory, Report CERN/ISR/TH 78–11, 1978.

[20] S. van der Meer: Optimum Gain for Stochastic Cooling Systems, Report CERN/PS/AA/83–48, 1983, and: Gain Adjustment Criterion for Betatron Cooling, Report CERN/PS/AA/ Note 82–02, 1982.

[21] A. Gerassimov: Macroparticle Simulation of Stochastic Cooling, Part. Acc. **51** 1995, p. 65.

Table 1:
Some history of stochastic cooling

1968:	Idea of stochastic cooling (van der Meer).
1972:	Observation of proton beam Schottky noise (ISR), first publication of cooling idea (van der Meer)
1972–75:	Hardware studies (Thorndhal, Schnell).
1975:	First experimental demonstration (ISR). First proposal for $\bar{p}$ accumulation with stochastic cooling.
1977–83:	Cooling tests at CERN, FNAL, Novosibirsk, INS-Tokyo.
1981–82:	Start-up of the CERN-$\bar{p}$ complex (AA. SPS-Collider. ISR $\bar{p}$-program, LEAR).
1982–84:	Observation of intermediate vector bosons. Nobel prize 1984 to Rubbia and van der Meer.
1986–87:	Start-up of FERMILAB $\bar{p}$-facility based on stochastic cooling.
1990:	Record stacking rate of 6.2×10^{10} $\bar{p}$/h in the CERN AAC.
1992:	Completion of the $\bar{p}$-p program at the SPS.
1993–:	Bunched beam stochastic cooling in large colliders, studies at FNAL, DESY, BNL, LBL.
1993:	Stochastic cooling at optical frequencies, proposal by Mikhailichenko, Zholents and Zolotorev.
1995:	Observation of the top quark at FERMILAB. Observation of nine antihydrogen atoms at LEAR.
1996 (Dec.):	Completion of the LEAR $\bar{p}$-program, stochastic cooling system to be used as wideband damper for ion beams stacked in LEAR.

Table 2:
Best possible cooling rate [Eq. (2)] for different mixing conditions. The ratio $\widetilde{M}/M = 1$ between unwanted and wanted mixing corresponds to the situation in the AC where the cooling loop has to cut diagonally through the ring in order to make up for the delay of the cooling signals with respect to the beam, $\widetilde{M}/M = 2$ is (roughly) the case for some low energy systems in LEAR where the kicker can be closer to the pick-up, $\widetilde{M} \gg M$ is the optimum which can only be achieved with a special cooling ring lattice.

Mixing situation $\widetilde{M}/M$	Noise/signal ratio U	Optimum cooling rate $(1/\tau_o)/(N/W)$	Corresponding M M_o
1	0	0.29	2.2
2	0	0.57	1.1
∞	0	1.0	1
1	10	0.05	3.5
2	10	0.07	2.3
∞	10	0.09	1

Figure Captions

Fig. 1. Set-up for stochastic cooling of horizontal betatron oscillations, schematic.

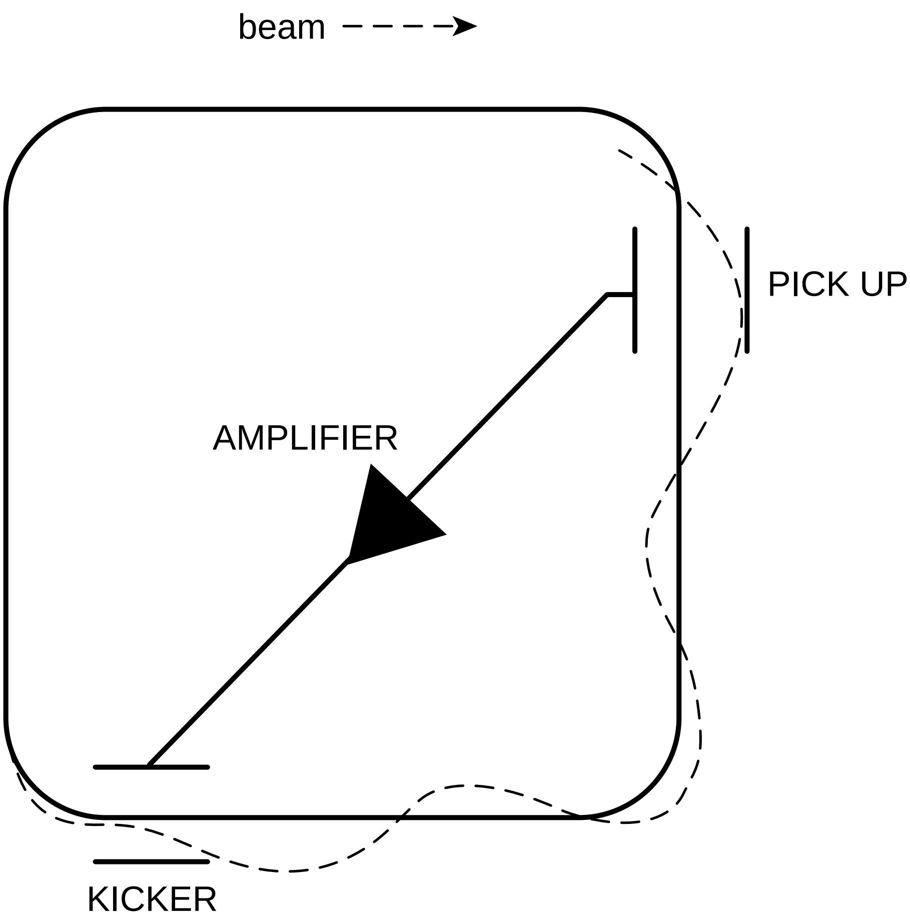

Fig. 1.