9 Injection and Extraction

In transferring the beam from one accelerator to another, preservation of the beam properties is essential. Injection should be accomplished with minimum beam loss and often minimal emittance dilution. Single-turn injection, in which a single bunch of particles is injected into a single empty rf bucket, is usually straightforward. In many cases, however, to attain higher bunch currents, one may also wish to accumulate beam in a storage ring by reinjecting different beam pulses into the same rf bucket. This is called multi-turn injection. In addition to conventional schemes, there are several new or more exotic injection techniques, devised to control and improve the properties of the stored beam.

Extraction refers to the removal of beam from an accelerator. It is roughly the reverse process of injection. One difference is that usually at extraction the beam energy is higher. Thus space charge effects are less important, but the hardware requirements for the septa and kicker magnets are more challenging. A high extraction efficiency is necessary to avoid activation of accelerator components and also to make optimum use of the accelerated beam, e.g., to achieve the maximum luminosity. Which extraction procedure is chosen depends on the specific application. Fast one-turn extraction is used for transferring bunches between different circular machines in an accelerator chain. For fixed-target experiments, slow extraction by the controlled excitation of nonlinear betatron resonances is a common technique, which provides a slow uniform depletion of particles in the ring, i.e., 'spill'. Again, several novel extraction schemes are being studied, for example, extraction using a bent crystal.

A good overview of conventional beam injection and extraction can be found in [1] and [2].

9.1 Transverse Single-Turn Injection

For single-turn injection, the beam is brought onto the central orbit using a septum magnet and a fast kicker element, as illustrated in Fig. 9.1. In the following, we assume that the injection is performed horizontally. The expressions derived can be extended easily to the vertical case, or to a combined horizontal and vertical injection.
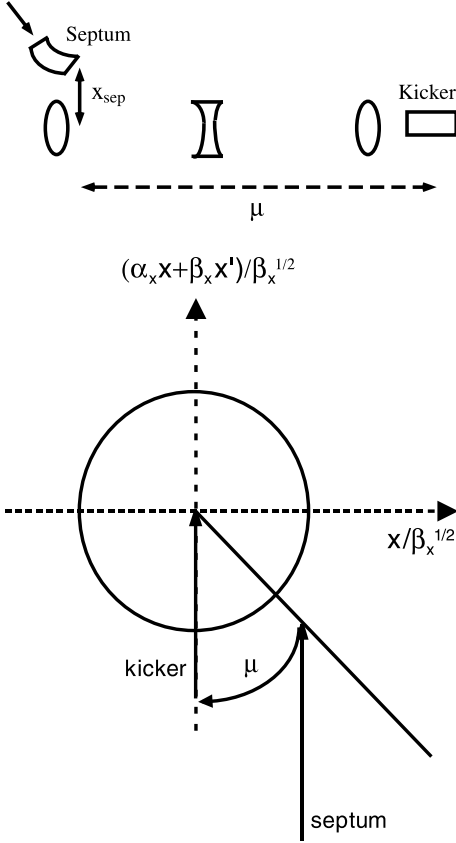
Fig. 9.1. Schematic of single-turn injection with septum and kicker: (*top*) magnet arrangement, (*bottom*) phase-space diagram. The septum strength is adjusted to place the beam onto the line describing an angle μ (the phase advance from septum to kicker) with respect to the vertical [1]

At the exit of the septum, the injected beam must be at a horizontal distance x_{sep} from the center of the machine aperture:

$$x_{\text{sep}} \geq n_{xi}\sigma_{xi} + n_{xs}\sigma_{xs} + D_x \left(\frac{\Delta p}{p}\right)_i + x_{\text{co,rms}} + x_{i,\text{rms}} + d_{\text{sep}} , \qquad (9.1)$$

where σ_{xi}, σ_{xs} are the rms beam sizes of the incoming beam and of the stored beam, respectively, n_{xs} is the beam stay-clear in the ring required for sufficient beam lifetime or negligible injection losses in units of the rms beam size of the stored beam (e.g., reasonable values may be $n_x \geq 8$ for electron rings, and $n_x \geq 4$ for proton rings), n_{xi} is the beam stay-clear for the injected beam also in units of its rms beam size, $(\Delta p/p)_i$ the relative momentum deviation of the injected beam with respect to the ring energy, D_x the dispersion function, $x_{\text{co,rms}}$ the rms closed-orbit offset at the location of the septum, $x_{i,\text{rms}}$ the rms orbit variation of the injected beam, and d_{sep} the thickness of the septum. For simplicity in (9.1) we have assumed that the

injected and the stored beam have the same beam stay-clear, in units of their rms beam size.

The injected beam must be positioned at the center of the aperture when it reaches the kicker. Let R denote the 2×2 transport matrix between the septum and the kicker. Requiring that the beam be on-axis at the kicker, the condition $x_{\mathrm{kic}} = R_{11}x_{\mathrm{sep}} + R_{12}x'_{\mathrm{sep}} = 0$, where x_{kic} denotes the position of the beam with respect to the center of the kicker magnet, determines the correlation of angle x'_{sep} and offset x_{sep} of the injected beam at the exit of the septum:

$$x'_{\mathrm{sep}} = -\frac{R_{11}}{R_{12}}\,x_{\mathrm{sep}} = -\frac{\alpha_{\mathrm{sep}} + \cot\mu}{\beta_{\mathrm{sep}}}\,x_{\mathrm{sep}}\ , \qquad (9.2)$$

where μ denotes the phase advance from septum to kicker, and α_{sep} and β_{sep} are the beta and alpha function at the septum. The angle x'_{sep} can be adjusted by changing the strength of the septum magnet, in order to meet condition (9.2).

Using (9.2) the slope of the injected beam at the kicker is

$$x'_{\mathrm{kic}} = R_{21}x_{\mathrm{sep}} + R_{22}x'_{\mathrm{sep}} = \left(R_{21} - \frac{R_{11}}{R_{12}}R_{22}\right)x_{\mathrm{sep}} = -\frac{1}{R_{12}}x_{\mathrm{sep}}\,. \qquad (9.3)$$

To position the injected beam on the design orbit in the ring, the kicker must apply the opposite angular deflection, which is

$$\theta_{\mathrm{kic}} = -x'_{\mathrm{kic}} = \frac{x_{\mathrm{sep}}}{\sqrt{\beta_{\mathrm{sep}}\beta_{\mathrm{kic}}}\sin\mu}\ . \qquad (9.4)$$

A large value of the beta function at the kicker, β_{kic}, reduces the kicker strength, and a large value of β_{sep} reduces the relative contribution to θ_{kic} which arises from the septum thickness d_{sep} (see (9.1) and recall that σ_{xi}, $\sigma_{xs} \propto \sqrt{\beta_{\mathrm{sep}}}$).

In case of a FODO lattice, the septum and kicker are best placed downstream of a focusing quadrupole, where the beta functions are close to maximum. In the particular case that the phase advance μ is $\pi/2$, the above formulae simplify to $x'_{\mathrm{sep}} = -\alpha_{\mathrm{sep}}x_{\mathrm{sep}}/\beta_{\mathrm{sep}}$, $\theta_{\mathrm{kic}} = x_{\mathrm{sep}}/\sqrt{\beta_{\mathrm{sep}}\beta_{\mathrm{kic}}}$.

The septum may consist of either dc septum magnets or dc electrostatic wires. In either case, the stray or leakage fields of the septum are a concern. These nonlinear fields can affect the quality of the stored beam. The stray fields may be reduced by a magnetic shielding, which extends beyond the septum ends.

The kicker magnets must be fast, since their rise and fall times often determine the minimum size of the gaps between bunches or between bunch trains in the ring, which in turn may depend on the repetition rate of the accelerator. Typical time constants are tens of nanoseconds, and voltage and current levels of 80 kV and 5000 A, with fields of 500 Gauss, are not uncommon. Frequently ferrites are used for field containment, and sometimes a ceramic vacuum chamber is inserted between ferrite and the beam, with a conducting

layer deposited on the inside of the ceramic. The coated ceramic reduces the impedance seen by the beam, as well as the associated heating of the ferrite. It is remarkable that the conducting layer can be much smaller than the skin depth and still provide adequate shielding of the beam fields, since shielding occurs already when the thickness of the metal coating is larger than the square of the skin depth divided by the thickness of the ceramic [3]. The coupling impedance experienced by the beam should be measured prior to installation of the kicker chamber. As an example, measurements for a prototype LHC kicker chamber are documented in [4]. A non-flat shape of the kicker pulse results in unequal deflections for different bunches in the beam. If necessary, a double-kicker system can relax the tolerance on the pulse shape (see the discussion of the KEK/ATF extraction scheme in Sect. 9.7).

Other important injection issues include transient beam loading and phase-space matching. Reference [5] gives a thorough review of beam loading compensation in storage rings, including a discussion of direct rf feedback and of problems that can arise from klystron power limititions. Horizontal and vertical dispersion and beta functions, as well as the ratio of bunch length and energy spread must be matched to the ring (or linac) optics. The transverse optics is matched by varying the strength of quadrupole magnets in the injection line, and, in case of dispersion, possibly also the strength of steering magnets. The longitudinal matching can be achieved by optimizing the amplitude of the storage-ring rf voltage, or, in certain cases, via bunch rotation, bunch compression, or energy compression prior to injection (cf. Sect. 8).

9.2 Multi-Turn Injection

For many applications, the beam is first accumulated in a ring, to increase its intensity, before it is further accelerated or sent towards its final destination. Typically a current-limited cw beam from the injector is thereby converted into a pulsed beam at higher-intensity and energy. The accumulation requires multi-turn injection.

9.2.1 Transverse Multi-Turn Injection

Multi-turn injection usually employs a ramped orbit bump in the vicinity of the septum, i.e., a slow change in the position of the ring closed orbit from turn to turn, instead of a fast kicker. The injection scheme is different for electrons and for protons or heavy ions.

In case of electron rings, radiation damping is utilized. First, a single bunch is injected. Then the orbit bump is reduced over a few revolution periods. After a few damping times, when the beam size has shrunk to its small equilibrium value, the orbit bump is reintroduced, and another bunch

is injected into the same bucket. Similar schemes, though usually in the synchrotron phase space (see below), are employed for accumulation of electron cooled proton beams and stochastically cooled antiproton beams.

For proton or heavy ions beams, the orbit bump is reduced slowly in time, and bunches are injected into different regions of the ring acceptance, so that the early bunches occupy the central region, and the later ones the outer parts of the acceptance. Some emittance dilution is inherent to this scheme. If beam is injected over N_t turns ('N_t-turn injection'), the final emittance can be estimated from the rough formula [1]:

$$\epsilon_f > 1.5 N_t \epsilon_i .\qquad(9.5)$$

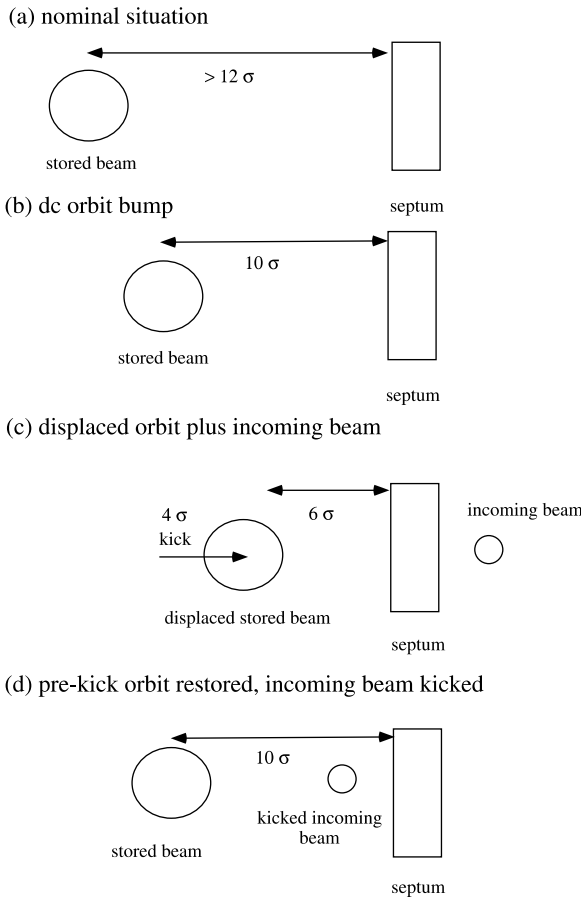Much larger emittance dilutions arise at low beam energy or high intensity, when space charge effects are important.

(a) nominal situation

> 12 σ

stored beam

septum

(b) dc orbit bump

10 σ

stored beam

septum

(c) displaced orbit plus incoming beam

4 σ
kick

6 σ

incoming beam

displaced stored beam

septum

(d) pre-kick orbit restored, incoming beam kicked

10 σ

kicked incoming
beam

stored beam

septum

Fig. 9.2. Injection process for PEP-II [6] (Courtesy M. Donald, 2002)

It is also possible and indeed more elegant, to install two kickers in the ring, powered in parallel by the same pulser. They are arranged such that only the second kicker of the pair deflects the injected beam, while both kickers act on the stored beam. If the phase advance between the two kickers is π or 2π, and the sign of the kick appropriately chosen, the kicker deflections generate a closed bump for the circulating beam. The advantage is that in this case the rise and fall times of the kicker do not have to be smaller than the bunch spacing, but can be on the order of the revolution time. The requirements on the kickers can be further alleviated by a dc orbit bump, which brings the stored beam closer to the septum prior to the injection. Figure 9.2 illustrates such a scheme, which is used at PEP-II [6].

9.2.2 Longitudinal and Transverse Multi-Turn Injection

The accumulation efficiency can be increased by combining transverse and longitudinal injection. This option was studied for LEAR [7], where the injected bunches come from a linac, which allows for an easy variation of their
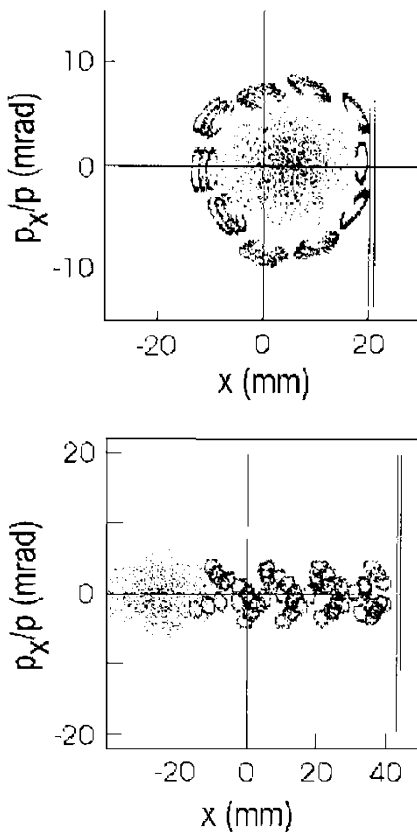

Fig. 9.3. Plots of the simulated horizontal phase space for multiturn injection into LEAR [7]: (*top*) 10 turns after start of purely transverse injection and (*bottom*) 20 turns after start of combined longitudinal transverse injection. Each bunch is represented by three ellipses with slightly different momentum deviations. The *two vertical lines* on the right represent the thickness of the septum (Courtesy Ch. Carli, 2002)

energy. As in the purely transverse multi-turn injection scheme, a local or-
bit bump is created and then decreased during the injection of successive
bunches. At the same time the linac energy is ramped such that, at the injec-
tion septum, the closed orbit corresponding to the instantaneous linac energy
remains constant. In other words, the change in the bump amplitude $x_{\mathrm{bump}}(t)$
and the simultaneous variation of the momentum $\delta(t)$ are related by

$$D_x \delta(t) = -x_{\mathrm{bump}}(t) + x_0 \; , \tag{9.6}$$

where x_0 is a constant and D_x the horizontal dispersion at the injection
septum. In this scheme, the final transverse emittance is smaller than for
the conventional multi-turn injection at the expense of an increased momen-
tum spread. Figure 9.3 compares simulated phase space distributions [7] for
transverse and combined injection into LEAR. Figure 9.4 shows the predicted
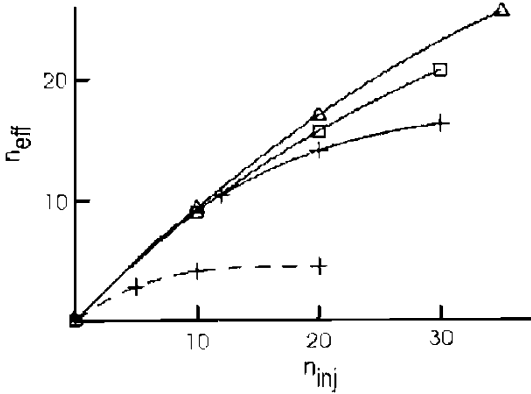improvement in the accumulation efficiency [7].

Fig. 9.4. Effective number of turns n_{eff} stored in LEAR as a function of the
number of injected turns n_{inj} [7]. The *solid lines* represent the combined transverse
longitudinal injection scheme, for three different LEAR optics (*crosses*: standard
optics; *triangles*: improved optics with large beta functions at the electron cooler;
squares: improved optics with intermediate beta functions at the electron cooler).
For comparison, the *dashed line* is for a purely transverse injection and the standard
optics (Courtesy Ch. Carli, 2002)

9.2.3 Longitudinal Multiturn Injection

If for an electron ring the time between subsequent injections is short com-
pared with the radiation damping time, multiturn transverse injection be-
comes difficult. In such a case, longitudinal injection offers a solution. Here
the circulating beam is brought close to the septum with an ac bump and the
incoming beam is injected with a negative energy offset such that the product
of this offset in energy and the dispersion is equal to the distance between

the newly injected and the stored beam. The injected bunches execute slow synchrotron oscillations.

Consider as an example the injection scheme employed at LEP [8]. Half a synchrotron period after the first bunch is injected, the next injection occurs. At this time, the first bunch is at its maximum distance from the septum. The situation is illustrated in Fig. 9.5. Similarily one could conceive injecting every 1/4 oscillation period, thus accumulating 4 injected bunches in one rf bucket. An advantage of longitudinal injection is a factor two faster radiation damping of the injection oscillations, since the longitudinal damping partition number is twice the transverse ($J_z \approx 2J_x$; compare Sect. 4.3). A possible disadvantage is that the acceptable time separation of successive injections is constrained by the synchrotron frequency.
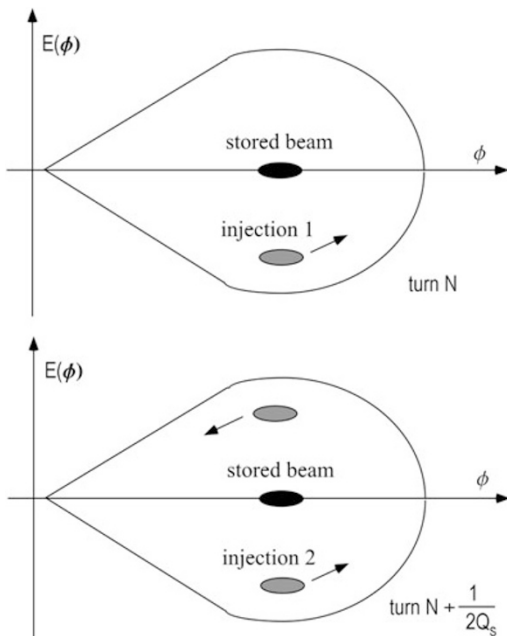

Fig. 9.5. Double injection into the same rf bucket; bunches are injected off-energy at a point with dispersion [8]. Time between the two injections is half a synchrotron oscillation period modulo a full period (Courtesy P. Baudrenghien and P. Collier, 2001)

9.2.4 Phase-Space Painting

For proton and ions beams, the multi-turn injection is often described as "phase-space painting" [9, 10]. This term refers to the injection of many small (linac) bunches into different spots of a 2- or 6-dimensional storage-ring

phase space, so as to generate a desired beam distribution, e.g., an approximately uniform distribution with reduced space-charge effects. The combined longitudinal-transverse injection into LEAR discussed in Sect. 9.2.2 can be considered as an example of phase-space painting. In the simplest case the beam is injected at a fixed position in the longitudinal (or transverse) phase space, and the painting is done automatically by the synchrotron oscillations.

During the injection process, the position of the injected beam in phase space can also be moved adiabatically, i.e., at a speed which is slow compared with the synchrotron oscillations. The injected phase-space density $P(r)$ and its projection $p(x)$ are related via

$$p(x) = 2 \int_x^R \frac{rP(r)\,\mathrm{d}r}{\sqrt{r^2 - x^2}} \; , \tag{9.7}$$

where R denotes the maximum radius in phase space at which bunches are injected. From a desired function $p(x)$, the corresponding radial density $P(r)$ can be computed using (9.7). The radial increment in the injected beam position between two successive bunches is [9]

$$\Delta r \approx \frac{1}{2\pi r P(r) N_{\mathrm{inj}}} \; , \tag{9.8}$$

with N_{inj} the total number of injected bunches.

More complicated schemes are frequently used. Similar to the above horizontal-longitudinal injection for LEAR, one can also combine horizontal and vertical painting. For example, at Rutherford Appleton Laboratory (RAL) a vertical steering magnet in the injection line is ramped, while the guide field in the ring is decreased [10]. Initially, there are small horizontal and large vertical oscillations, while at the end of the injection the situation is reversed. Instead of the vertical steering in the injection line, a programmable vertical orbit bump in the ring could be employed alternatively.

9.3 H^- Charge Exchange Injection

The principle of injection using H^- exchange originated in Novosibirsk [11]. It is now the preferred injection scheme for proton machines [1]. In this scheme H^- ions are accelerated by a linac and are stripped to protons, when they traverse a thin foil during injection into the ring [1, 12], as illustrated in Fig. 9.6.

The stripping of the H^- ions to protons occurs within the ring acceptance. Since during the stripping the particles change their charge, Liouville's theorem on the conservation of the beam density in phase space does not apply, and, thus, in principle, a high proton density could be attained by injecting successive bunches into the same region of phase space. In most practical applications, however, vertical steering in the injection line is combined with a
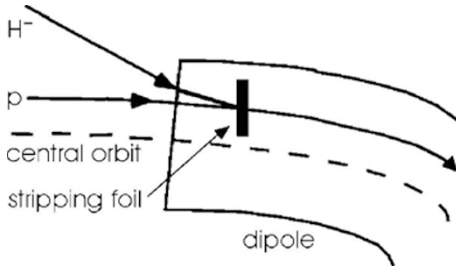
Fig. 9.6. Schematic of H^- stripping injection

ramped horizontal orbit bump in the ring in order to provide a very uniform filling of the phase space and to minimize space-charge effects.

The heating of the stripping foil and stripper scattering effects limit the foil thickness. Typical thicknesses range between 50 and 200 $\mu g\,cm^{-2}$ (less than 1 μm), with stripping efficiencies of 98% for 50 MeV protons [1]. As foil materials, polyparaxylene, carbon and aluminum oxide have been used. The rms scattering angle for a single foil traversal is typically on the order 0.2 mrad. The total scattering angle increases as the square root of the average number of passages through the foil. The stripping foils are supported at three edges, so that vertical beam motion cannot does not reduce the number of foil traversals.

Stacking simultaneously in both betatron and synchrotron phase space reduces the number of foil traversals, however. This can be achieved by changing the magnitude of an orbit bump, while also ramping the frequency and phase of the rf system during the injection cycle. The following lattice parameters at the location of the foil are considered advantageous [1]: $D'_x = 0$, $\alpha_x = 0$, and $D(\Delta p/p) > \sqrt{A\beta}$, where $\Delta p/p$ is the momentum acceptance and A is the transverse acceptance (in emittance units). A location between two symmetric defocusing quadrupoles is suitable for meeting these conditions. The second quadrupole assists in deflecting the unstripped H^- ions.

Stripping foils are used for heavy ions as well. The final charge distribution of the ions depends on the foil thickness and on the particle energy [13]. In the extreme case, the ions can be fully stripped.

Foils are also used for extracting H^0 atoms or protons from H^- storage rings, where the neutral atoms are generated by capturing electrons from the foil. For these applications, also a gas jet or a laser beam [14] can facilitate the extraction in a similar way.

9.4 Resonant Injection

Another injection scheme worth mentioning is a proposal for resonant injection [15]. Here 'bumper' magnets with dipole, quadrupole and octupole fields are excited to produce a separatrix with two stable regions in phase space;

the stored beam is in one region, and a bunch is injected into the other region. Afterwards, the fields are adjusted to merge the two parts of the beam. Then the injection condition is reestablished.

9.5 Continuous Injection

Continuous injection has been proposed as a means for maximizing the luminosity of a circular collider [16]. The motivation is obvious; if the stored beam could be continually replenished so that the current per bunch stays constant, then the average luminosity would roughly equal the peak luminosity. Continuous injection also reduces fill-to-fill variations and avoids transient phenomena, e.g., transient beam loading, thereby establishing quasi-static conditions, which is of interest not only for colliders, but also for light sources. Finally, in a colliding-beam storage ring the beam lifetime τ decreases inversely with the luminosity L,

$$\frac{1}{\tau} = -\frac{1}{N}\frac{dN}{dt} \propto \frac{L}{N} \ . \tag{9.9}$$

Thus, continuous injection supporting a much reduced lifetime could provide a substantial gain in average luminosity.

As an example, taking all these effects together, continuous injection is estimated [16] to potentially increase the average luminosity of the PEP-II B factory by about a factor of 5, assuming that each bunch in both rings can be replenished every 2.1 s. In this example, a 67-ns long orbit bump would move the injected bunches transversely to about 4σ from the stored beam core. This is done so that the injected bunches have an unobstructed passage through the physics detector. The effective minimum beam lifetime which can be supported is given by [16]

$$\tau = \frac{N_b}{\Delta N_b} \ \Delta t \ , \tag{9.10}$$

where N_b is the nominal number of particles per bunch, ΔN_b is the number of particles added into a single bunch per injection, and Δt the time period between successive injections into the same bunch. Extreme parameters for PEP-II are [16] $N_b \approx 1.2 \times 10^{11}$, $\Delta N_b \approx 10^9$, and $\Delta t \approx 2.1$ s, yielding a minimum supportable beam lifetime of $\tau \approx 4.2$ minutes.

9.6 Injection Envelope Matching

At injection into a storage ring, if the incoming beam distribution is not properly matched to the ring optics, the beam envelope in phase space will rotate around the matched design envelope. This oscillation will result in

turn-to-turn beam-size variations, which can be measured using a synchrotron light monitor and a fast-gated camera.

An injection-mismatch measurement from the SLC damping ring [17, 18] is shown in Fig. 9.7. The different pictures correspond to successive turns after injection, at the indicated turn number. Each picture is an average over 8 individual images. Clearly visible is a variation of the bunch shape from turn to turn.
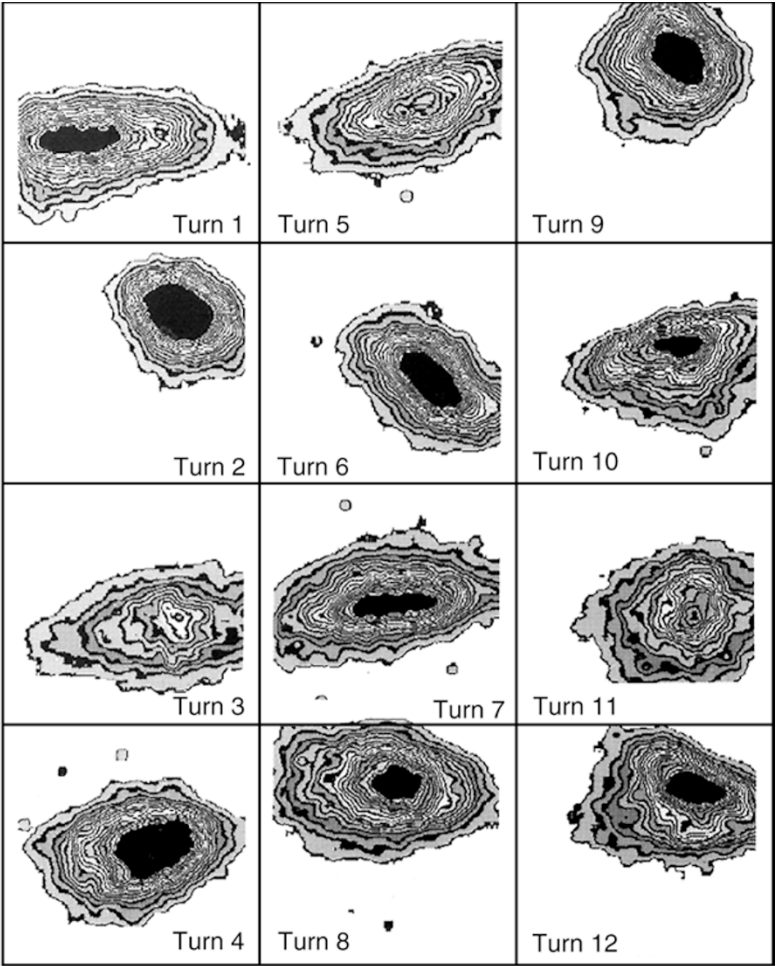

Fig. 9.7. Beam images of the first twelve turns after injection into the SLC damping ring, illustrating the effect of an injection mismatch [17, 18]. These are pictures from a synchrotron light monitor taken with a gated camera. Each image is an average over 8 beam pulses. The beam-size variation from turn to turn is an indication of injection mismatch

At the SLC damping ring, the matching of the injected beam distribution consists of minimizing the measured beam size after 1250 turns, by varying several quadrupoles at the end of the injection beam transport line. A number of 1250 turns was chosen, because at this time the initial beta and dispersion mismatch has completely filamented. Since, on the other hand, the time scale is much shorter than the radiation damping time, the emittance is given directly by $(\mathrm{B_{mag}} \cdot \epsilon)$, where ϵ is the emittance of the injected beam, and $\mathrm{B_{mag}}$ the mismatch factor defined in (4.59).

The beam size variation can be analyzed in the frequency domain by a Fourier transform to determine the amplitude of the mismatch factor and ultimately, for a well calibrated monitor, the beam emittance at injection. A beta mismatch will appear as a frequency line at twice the betatron tune, while a horizontal dispersion mismatch will be evident as a line at the betatron tune itself [18]. If only a beta mismatch is present, the ratio ρ of the dc Fourier component and the component at $2Q_x$ is equal to $(\mathrm{B_{mag}}/\sqrt{\mathrm{B_{mag}^2} - 1})$. From this, $\mathrm{B_{mag}} = 1/\sqrt{1 - \rho^{-2}}$ can be determined [18, 19].

Figure 9.8 shows the beam size squared for the first 100 turns after injection, as well as the FFT (multiplied with γ/β_x where γ is the relativistic Lorentz factor and β_x the beta function). Clearly visible are peaks at $2Q_x$ in the horizontal signal and at $(1 - 2Q_y)$ in the vertical one. The final emittance after filamentation, $(\mathrm{B_{mag}} \cdot \epsilon)$, is given by the dc component of the FFT.
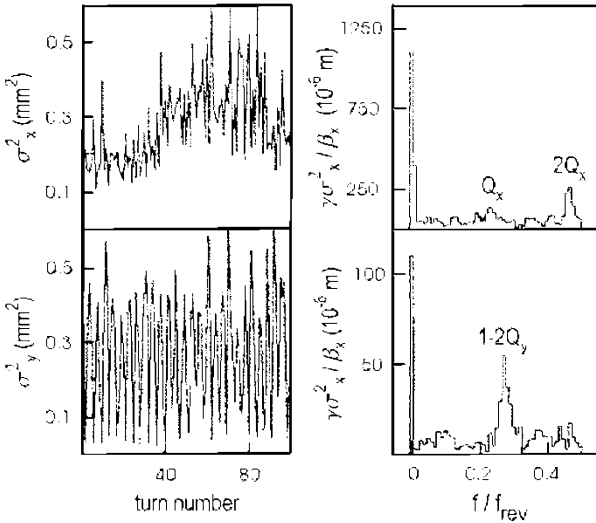

Fig. 9.8. Horizontal (*top*) and vertical (*bottom*) beam sizes for the first 100 turns after injection into the SLC damping ring (*left*) and their FFT (*right*) [18]. Clearly visible in the frequency spectra are lines at $2Q_x$ (*top*) and at $(1 - 2Q_y)$ (*bottom*), whose amplitude is a measure of the amount of beta mismatch
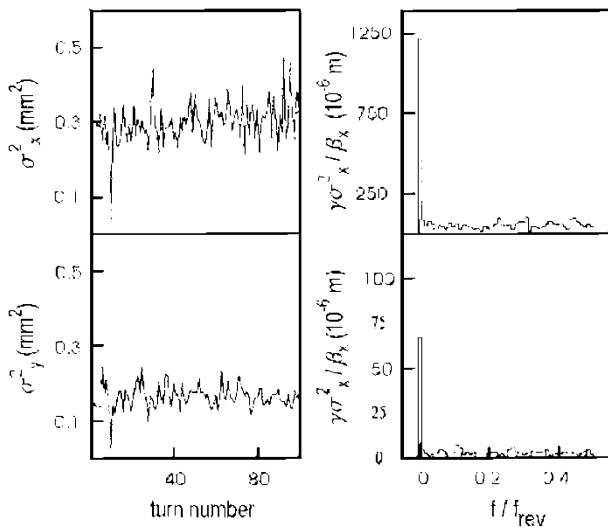
Fig. 9.9. Same as Fig. 9.8, after beta matching [18]. The peaks at twice the betatron tune have disappeared

The standard matching procedure reduces the FFT signals at $2Q_x$, $(1 - 2Q_y)$ and Q_x, as illustrated in Fig. 9.9.

9.7 Fast Extraction

Fast extraction is similar to single-turn injection. Orbit bumps are generated which move the stored beam close to a septum magnet. Then a fast kicker is fired, which deflects the next bunch, or group of bunches, into the extraction channel. If only one kicker is used, the kicker rise time must normally be smaller than the separation between two circulating bunches (or groups of bunches). The pulse length and fall time of the kicker are determined by the number of bunches to be extracted, and by the ring fill pattern. From (9.4), the minimum deflection angle required is

$$\theta_{\mathrm{kic}} = \frac{x_{\mathrm{sep}}}{\sqrt{\beta_{\mathrm{sep}}\beta_{\mathrm{kic}}}\sin\mu} \ , \tag{9.11}$$

where β_{sep} and β_{kic} are the beta functions at septum and kicker, μ is the phase advance between these two elements, and x_{sep} is the minimum displacement at the septum required for clean extraction. The initial orbit bump reduces the value of x_{sep}. In a FODO lattice with a phase advance per cell of, e.g., $90°$, the kicker can be positioned just upstream of the focusing quadrupole, and the septum at the identical position one cell downstream. In this way, the beta functions are maximized, and so the deflection angle required for the kicker is minimized.

For extraction from the damping ring of a linear collider, it is extremely important that the deflection imparted by the kicker has a very small pulse-to-pulse fluctuation ('jitter') and is sufficiently flat over the length of a bunch train. In order to confine the orbit variation at the interaction point (IP) to $0.1\sigma^*$ (σ^* is the IP spot size), the orbit variation at the septum should be smaller than $0.1\sigma_{\rm sep}$. The tolerance on the relative deflection error then is $\Delta\theta_{\rm kic}/\theta_{\rm kic} < 0.1\sigma_{\rm sep}/x_{\rm sep}$, where $\sigma_{\rm sep}$ is the rms beam size at the septum, and $x_{\rm sep}$ the transverse displacement of the kicked beam. Without fast orbit bumps, this can also be rewritten as [20]

$$\frac{\Delta\theta_{\rm kic}}{\theta_{\rm kic}} \leq \frac{1}{10}\,\frac{\sqrt{\epsilon_{\rm ext}\beta_{\rm sep}}}{d_{\rm sep}+n_s\sqrt{\epsilon_{\rm inj}\beta_{\rm sep}}}\,, \tag{9.12}$$

where $\epsilon_{\rm inj}$ and $\epsilon_{\rm ext}$ are the injected and extracted beam emittances, $\beta_{\rm sep}$ the beta function at the septum, and n_s the distance between the closed orbit and the septum plate in units of the injected rms beam size, when the beam is largest (i.e., the injected beam size enters, because the aperture at the septum must be large enough to accommodate the injected beam). For electron rings, one must have $n_s \geq 7$. Using typical parameters for a linear-collider damping ring, the relative jitter tolerance for the kicker, $\Delta\theta_{\rm kic}/\theta_{\rm kic}$, is on the order of a few 10^{-4}, and it is mainly determined by the ratio of the extracted beam emittance to the injected emittance [20]. A possible solution is the use of a double kicker system, separated by a betatron phase advance of π, to cancel the jitter [20, 21]. This compensation scheme is illustrated in Fig. 9.10. One kicker would be placed before the septum and the other in the extraction line. If a pulser feeds both kickers in parallel, with appropriate cable delays, kicker pulse errors in the first magnet are canceled by those in the second
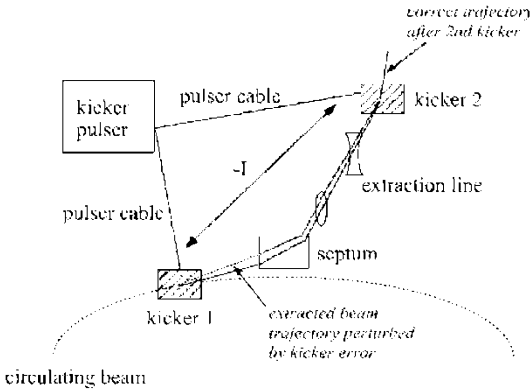

Fig. 9.10. Damping ring extraction with double kicker system, for reducing downstream beam orbit fluctuations. A change in the pulse shape alters the deflection from both kickers equally. The R matrix between the two kickers and cable delay times are chosen such that the effects of the two deflection errors cancel exactly, and the final beam trajectory is unchanged

magnet. A double-kicker system of this type has been built and installed at the KEK ATF damping ring [22].

Similar techniques can be applied to compensate for drifts of the septum field. For example, the NLC design contemplates the use of a compensating bending magnet in the extraction line, which is powered in series with the septum and placed such that field fluctuations will cancel [21].

9.8 Kickers

There are several different types of kicker magnets [6, 23] such as: (1) a current loop inside the vacuum, (2) a terminated transmission line inside the vacuum, (3) a ferrite magnet outside the vacuum, and (4) a multi-cell transmission line with ferrite flux returns [21]. As an example, Fig. 9.11 shows the ferrite kicker and the kicker pulser circuit adopted for PEP-II [6].
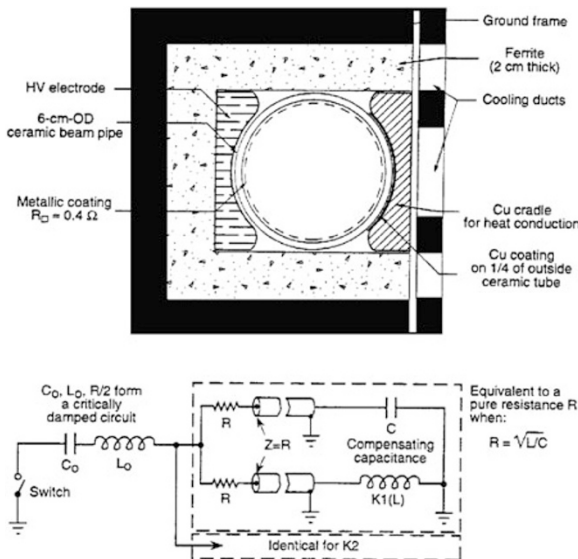

Fig. 9.11. Schematic of PEP-II kickers [6]: (*top*) kicker magnet cross section; (*bottom*) pulsing circuit with FET switch (Courtesy M. Zisman, 2002)

Typical kicker rise and fall times are 50–150 ns (SLC and NLC design: 60 ns, PEP-II: 120 ns). For a fast horizontal kicker with ferrite yoke, the characteristic time constant of the kicker scales as lw/g, where l is the length, w the width, and g the vertical gap of the kicker. This time constant can be reduced by dividing the kicker into several segments of shorter length. The kicker magnets are powered by kicker pulsers, usually based on thyratron cable discharges. The pulse shape can be modified by adding filters and

capacitors in parallel with the charge line. Spark gaps and solid-state FETs, such as in Fig. 9.11, are thyratron alternatives with potentially shorter rise and fall times [21].

For many future applications with closely spaced bunch trains, shorter kicker time constants are desired. A very fast counter-travelling wave kicker was designed and built for the TESLA project [24]. This kicker scheme uses two parallel conducting plates or electrodes. These are excited by short pulses from a generator, generating an electromagnetic wave which travels in a direction opposite to the beam, and produces a horizontal kick. At the end of
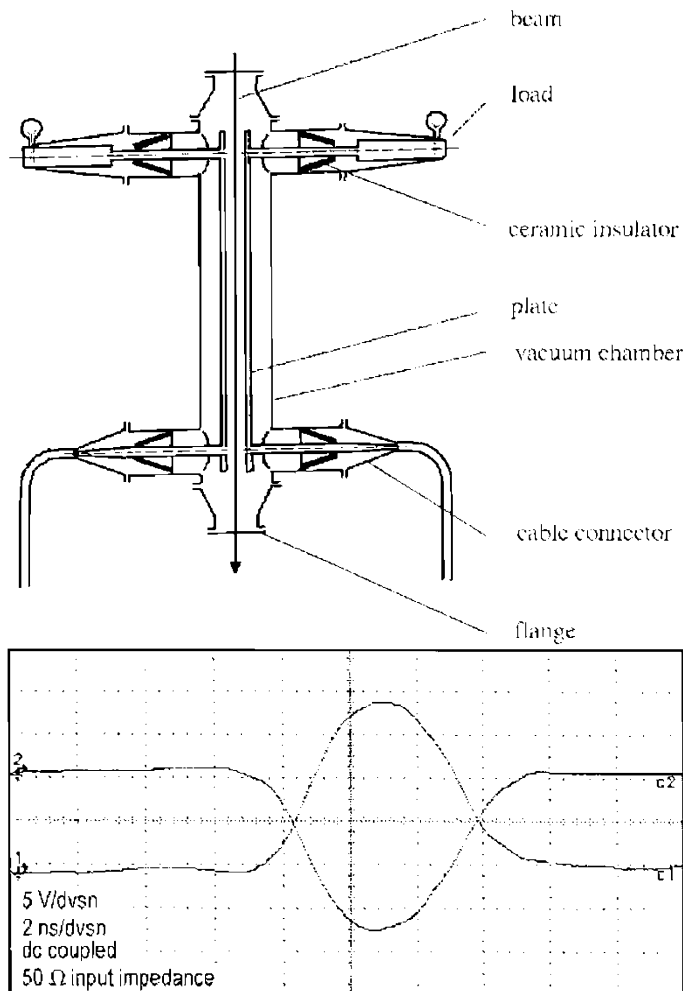

Fig. 9.12. Very fast kicker prototype [24]: (*top*) layout; (*bottom*) measured output rf pulses (Courtesy V. Shiltsev, 1998)

the kicker plates, the wave passes into two ceramic outputs, in which ideally it is fully absorbed without reflection. For a beam travelling opposite to the wave the effects of the magnetic and electric forces add, while they cancel each other for a beam moving in the same direction. The maximum integrated kicker strength in units of voltage is given by

$$S_0 \approx \frac{2U_m l}{a} , \qquad (9.13)$$

where U_m is the maximum pulse voltage at each plate, a is the half aperture, and l the length. A kicker was tested with $U_m = 2$ kV, $a = 25$ mm, and total length l of 0.5 m. Figure 9.12 shows output pulses measured on this kicker prototype, demonstrating a zero-to-zero pulse length of about 6 ns. The maximum pulse height corresponds to the predicted kick strength of
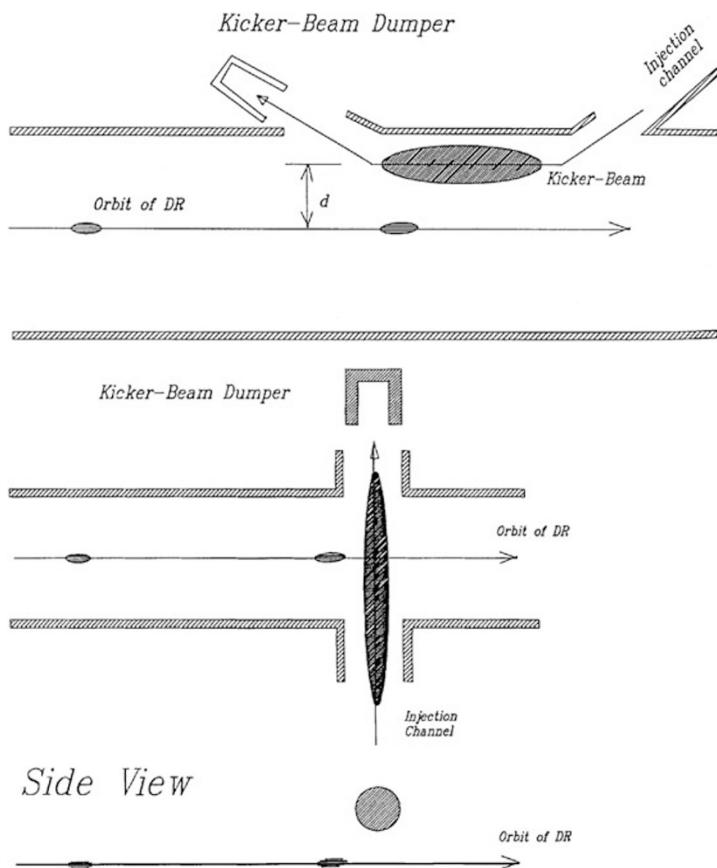

Fig. 9.13. Beam-beam kicker: 'head-on' (*top*) and 'cross' scheme (*bottom*) [25] (Courtesy V. Shiltsev, 1998)

80 kV, or 2.76 Gm. For electron or positron beams with $E = 3.3$ GeV, this would amount to a deflection angle of 24 μrad.

An ultrafast beam-beam kicker was proposed [25], in order to provide even faster kicker pulses. Here, a wide high-charge low-energy bunch traverses the beam pipe either parallel to the beam direction or perpendicular to it. The electro-magnetic or electric field of this bunch is used to deflect (and extract) a bunch circulating in the ring. The pulse length of the beam-beam kicker is determined by the length of the low-energy bunch and can be on the order of 2 ns. Figure 9.13 illustrates two possible geometries.

9.9 Septa

As we have seen in (9.1) and (9.4), a small septum thickness d_{sep} reduces the requirements on the kicker and increases the extraction efficiency. For this reason, electrostatic wire septa have been employed since many years, for example, during fast extraction at the Fermilab Tevatron [26]. The Tevatron electrostatic septum consists of two 354 cm long sections with 86 cm space in between, made from 75% tungsten and 25% rhenium wires of 0.002 inch diameter and 0.1 inch spacing with an angle of 25 μrad between sections. The voltage of 93 kV results in an electric field of 83 kV/cm [26]. Very similar electrostatic deflectors have been proposed for the muon collider [27]. At high energies, the integrated strength of a wire septum often cannot provide a deflection angle large enough for clean extraction, and, in such cases, an additional thin septum magnet is positioned immediately downstream.

In general, two types of septum magnets are widely used [28]: Lambertson iron septum dipoles and current-carrying septum dipoles. A Lambertson magnet is illustrated in Fig. 9.14. The triangular cut-out in the window frame leaves space for the circulating beam. As shown, a kicker deflects the beam horizontally into the septum, by which it is then bent vertically.

Figure 9.15 depicts a current sheet septum. A current carrying septum with thickness d and current density J generates a field $B = \mu_0 J d$. For d of the order of a millimeter, the septum is used in a pulsed mode to provide
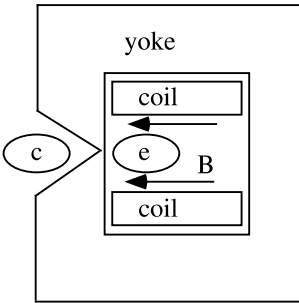

Fig. 9.14. Schematic of Lambertson septum iron magnet [2]; the symbol c represents the circulating beam, the symbol e the extracted beam
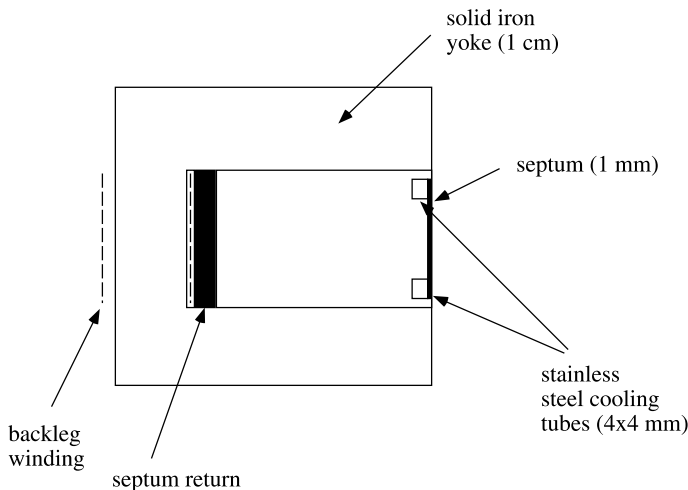
Fig. 9.15. Cross section of current sheet septum [6] (Courtesy M. Zisman, 2002)

enough field strength. For larger thicknesses, dc operation is common. Septum leakage fields which affect the circulating beam are a concern. In addition to normal dipole and higher order fields, the septum stray field may contain a skew quadrupole component.

9.10 Slow Extraction

The beam can be slowly extracted by exciting a third order nonlinear resonance, using sextupoles. Also a second order linear resonance can be used, in combination with octupoles. The extraction efficiency depends on the ratio of the betatron amplitude growth per turn and the septum thickness. It can be improved with a high-beta insertion at the septum.

Figure 9.16 depicts the phase space near the 3rd order resonance, excited by sextupole magnets. Particles inside the inner triangle are stable. Outside the triangle the oscillation amplitude of a particle grows exponentially. Therefore, particles in this region are rapidly lost, along a particular direction in phase space (in this example, towards the right). The size of the triangle depends on the strength of the sextupoles and on the betatron tune.

Near the third-integer resonance, $(3Q-q) \approx 0$, with integer q, the particle motion can be described by a Hamiltonian of the form

$$H(I,\psi,\vartheta) = (Q - q/3)I + \frac{1}{24}(2I)^{3/2}|\tilde{K}_s| \sin(3\psi + \theta_0) , \qquad (9.14)$$

where ϑ is the azimuthal position around the ring, which acts as the time-like variable, and I and ψ are the action-angle variables, which are related to the
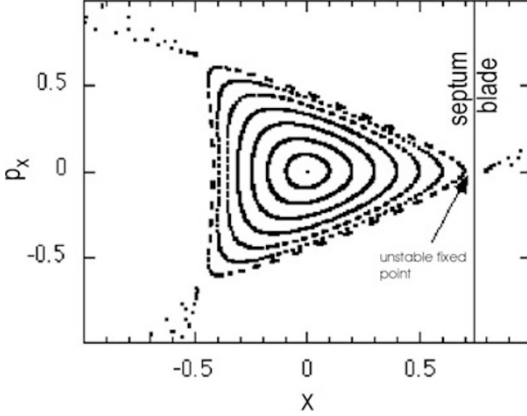
Fig. 9.16. Phase space schematic for slow extraction near the 3rd integer resonance (axes in arbitrary units). The sextupole excitation pattern around the ring is chosen such that the maximum excursion in the horizontal coordinate x occurs at the location of the septum. The position of the septum wire is indicated as a vertical line close to the unstable fixed point

transverse particle coordinates at the septum via $x_{\text{sep}} = \sqrt{2\beta_{\text{sep}}I}\cos\psi$ and $x'_{\text{sep}} = -\sqrt{2I/\beta_{\text{sep}}}\sin\psi - \alpha\sqrt{2I/\beta}\cos\psi$. The term $|\tilde{K}_s|$ is the absolute value and θ_0 the complex phase of the resonant Fourier harmonic of the sextupole distribution around the ring appropriately weighted by the beta function:

$$|\tilde{K}_s|e^{-iq\theta_0} = \frac{1}{2\pi}\int_0^{2\pi} k_s(\theta)\beta^{3/2}(\theta)e^{-iq\theta}\ \mathrm{d}\theta\ . \tag{9.15}$$

Here θ is the azimuthal angle around the ring, and the sextupole strength (in units of m^{-3}) is given by $k_s(\theta) = \partial^2 B_z(\theta)/\partial x^2/(B\rho)$, with $B\rho$ the magnetic rigidity.

Suppose the tune is slightly below the 3rd integer resonance, $(3Q-q) < 0$. Then a corner point of the separatrix coincides with the horizontal position coordinate x_{sep} at the septum, if $\theta_0 = \pi/2$. Above the resonance, $(3Q-q) > 0$, the optimum choice would be $\theta_0 = -\pi/2$. The value of θ_0 can be adjusted by changes to the sextupole configuration, or by changes to the ring optics. The particles arrive at the septum with a large amplitude on every 3rd turn. The amplitude growth over three turns, for a particle near the unstable fixed point (at the asymptotic angle $\psi = \pi/6$), is approximately

$$\Delta x_{\text{sep}} \approx \frac{3\pi x_{\text{sep}}^2 |\tilde{K}_s|}{4\beta_{\text{sep}}^{1/2}}\ . \tag{9.16}$$

This shows that large sextupole strengths and a large beta function at the septum (since $x_{\text{sep}} \sim \sqrt{\beta_{\text{sep}}}$) are advantageous.

A slow spill can be controlled by adjusting either the strength of the sextupoles or the betatron tune. Extraction may also involve beam steering. Also

making use of chromaticity, particles of different momenta can progressively be brought onto the resonance. Extraction starts when the beam particles at one end of the momentum distribution fill the triangular stable area in phase space. The stable area then shrinks to zero for these particles, and subsequently particles of different momenta are extracted.

To achieve a slow extraction efficiency greater than 98%, the thickness of the septum must typically be of the order of 100 μm.

9.11 Extraction via Resonance Islands

A novel method for multi-turn extraction from a circular particle accelerator was explored at the CERN PS [29, 30]. Here, in addition to fast or slow extraction an intermediate extraction mode is needed, which is called multi-turn extraction. The PS serves as injector to the SPS. The latter has an 11 times larger circumference. In order to fill the SPS ring with only two 'shots' from the PS, each PS beam is extracted over 5 turns.

The conventional technique used for this extraction is called the 'continuous transfer' (CT). The principle is illustrated in Fig. 9.17. The tune is moved closed to the quarter integer resonance. Then the beam is deflected so that a fraction of it is shaved off at the electrostatic septum blade. Three other slices are transferred on subsequent turns. The central part is extracted last during the fifth turn, by applying a larger deflection. Since during extraction the beam is cut into 5 transverse pieces, the slices transferred ideally have a five times lower transverse emittance than the original beam.

However, there are various problems with this approach: (1) beam losses at the septum are unavoidable; (2) the extracted slices do not match the natural circular shape of the phase-space trajectories which implies emittance growth
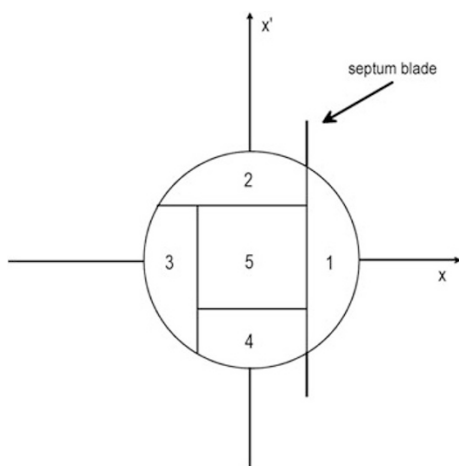

Fig. 9.17. Principle of the conventional 'continuous-transfer' extraction from the CERN PS ring. The beam is shaved by means of an electrostatic septum [29, 30]. The horizontal tune of the PS is set to 6.25 (Courtesy M. Giovannozzi, 2001)

in the downstream SPS; (3) unequal emittances of the extracted slices. For these reasons the CT extraction is not particularly suitable for the CERN neutrino to Gran Sasso (CNGS) proton beam [31].

The alternative novel scheme [29] makes use of stable nonlinear resonance islands. By exciting sextupole and octupole magnets, islands are created in phase space. The position and width of the islands are controlled by moving the betatron tune across the quarter-integer resonance. Initially, the islands are introduced adiabatically near the origin. The beam is thereby split into 5 components, all round in shape, and well matched to the circular phase space structure. Then the tune is shifted away from the resonance, so that the islands separate and approach larger amplitudes. Now the beam can be deflected as in the conventional CT scheme described above.

In this case, however, beam losses can be avoided by deflecting an empty region of phase space (between the resonance islands) onto the septum blade. Each slice is well matched, and, hence, the emittance growth is negligible. Finally, by properly adjusting the island parameters, the slices can be equally populated and be produced such that their emittances are equal. Figure 9.18 shows the proposed tune evolution for this extraction scheme, and Fig. 9.19 the simulated beam distribution at various times of the trapping process. At the end of the process, the islands are well separated. In the simulation, no particles are lost, neither during the island creation ('capture process') nor when shifting the island positions.

Open questions concern the quantitative relation between slice emittance and island parameters, the optimization of the tune change, and the robustness against perturbing effects such as tune modulation, e.g., caused by power-supply ripple.
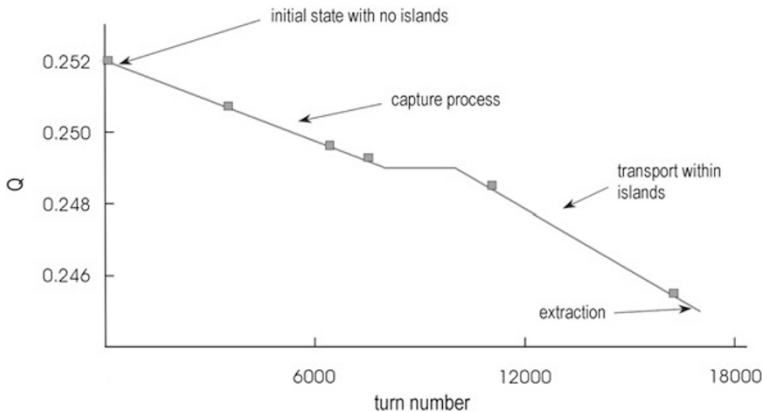

Fig. 9.18. Variation of the small-amplitude tune Q as a function of turn number n during resonant multiturn extraction [29]. The *solid squares* refer to tune values for which phase-space portraits are shown in Fig. 9.19 (Courtesy M. Giovannozzi, 2001)
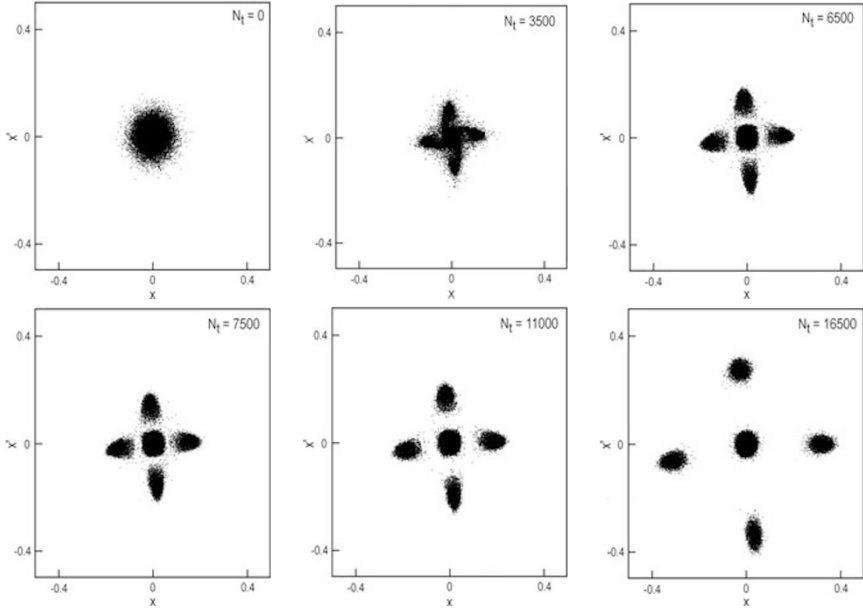
Fig. 9.19. Simulated evolution of the beam distribution during the trapping process of resonant multiturn extraction (axes in arbitrary units) [29]. The different plots correspond to the tune values which are represented by solid squares in Fig. 9.18. Each plot represents 2.25×10^4 points (Courtesy M. Giovannozzi, 2001)

9.12 Beam Separation

A problem similar to injection and extraction is the beam separation near the collision point of a collider, or the beam (re-)combination before and after the arcs of a recirculating linac.

To illustrate the concept and a possible approach, we describe a design example for horizontal beam separation at a Very Large Lepton Collider (VLLC) [32]. The purpose of the beam separation is to feed the two beams into the two separate magnet channels of the collider arcs. The separation is launched in the straight section close to the collision point. The optical lattice in the straight is assumed to be the same FODO lattice as in the arcs, with a cell length L_p and quadrupole focal length f.

An electrostatic separator is placed next to a horizontally focusing quadrupole. Its field is chosen such that the beam are offset by $\pm N_\sigma$ rms beam sizes at the next quadrupole downstream, which is horizontally defocusing. The separation $N_\sigma \sigma_D$ (where σ_D denotes the rms beam size at the quadrupole), the deflection angle ϕ_e and the integrated strength of the septum are related via [32]

$$\phi_e = \frac{2N_\sigma \sigma_D}{L_p} \,, \tag{9.17}$$

where

$$E_x l_e = \phi_e E \ , \tag{9.18}$$

in which E_x denotes the (horizontal) electric field and E the beam energy. A LEP separator consists of 4-m long electrodes, and provides a nominal field of 2.5 MV over a gap of 0.11 m [33]. About two of these separators would be needed for the VLLC application (beam energy 184 GeV) [32]. The defocusing quadrupole enhances the slope between the two beams. Downstream are dc magnetic septum magnets with opposite vertical fields on either side of a current sheet. They add enough slope to the beam that they may be brought into separate channels at the next focusing quadrupole. We denote the half separation of the two channels at that (focusing) quadrupole by d_m, the deflection angle of the magnetic septum by ϕ_m, its integrated magnetic field by Bl_m, and the electrostatic separation at the intermediate defocusing quadrupole with focal length f by x_s. The required deflection by the magnetic septum is then given by

$$\phi_m = \frac{2(d_m - x_s)}{L_p} - (\phi_e + x_s/f) \ . \tag{9.19}$$

The deflection ϕ_m is generated as

$$\phi_m = Bl_m c/E_b \ , \tag{9.20}$$

where B is the septum field and l_m the septum length. For a length of 10 m, a septum field of about 0.25 T is required, which could be produced by a septum of thickness 7 mm, assuming a current density of 60 A/mm^2 in the septum sheet, which is the operation value for the d.c. septum at the SPS. The entire beam separation is illustrated schematically in Fig. 9.20. Dispersion generated by the separation is compensated in the arc dispersion suppressors.
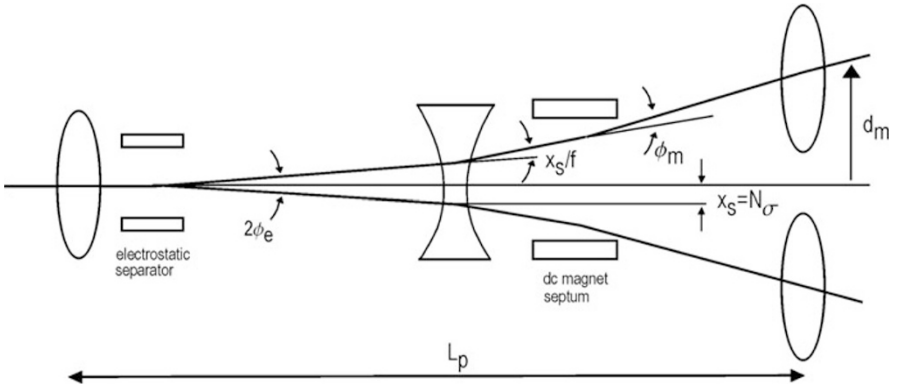

Fig. 9.20. Schematic of beam separation at the VLLC as described in [32]

9.13 Crystal Extraction

Crystal extraction is conceptually quite different from the extraction methods. It was first studied at Dubna and Protvino [34], and later tested extensively at the CERN SPS [35, 36] and at the Fermilab Tevatron [37]. Here, particles in the transverse beam halo, entering a crystal placed close to the beam, are trapped between the crystalline planes [38]. If the crystal is slightly bent, the particles can be deflected outwards, and subsequently be transported to a fixed-target experiment. Figure 9.21 shows a schematic view of crystal extraction.

Crystal extraction is foreseen as an option for the LHC. It would be parasitic to the normal collider operation, and re-utilize the halo particles which do no longer contribute to the collider luminosity.
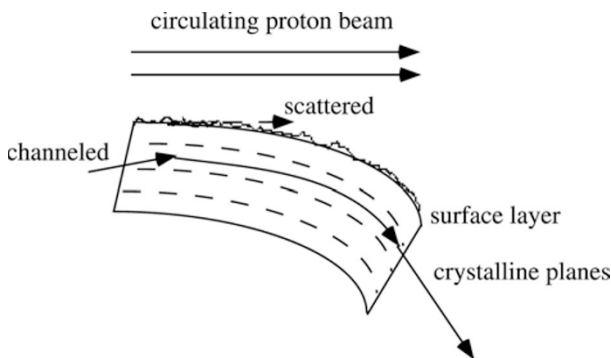

Fig. 9.21. Extraction from the transverse halo of a circulating proton beam by means of a bent crystal. Particles incident with a large impact parameter are channeled and deflected outwards. Particles hitting the inefficient crystal surface layer experience multiple scattering, and may be channeled on a later turn; this is called multi-pass extraction

Channeling occurs if the incident angle of the particles is smaller than the Lindhard critical angle [38]. The critical angle depends on the orientation of the crystal and on the material. The crystal orientation can be defined with respect to an axial direction $[ijk]$ or with respect to a plane (ijk)[1]. For the (110) planar direction in silicon the critical angle is [38]

$$\Psi_{\text{crit}} \approx 5 \ \mu\text{rad} \ \frac{\sqrt{Z}}{\sqrt{p[\text{TeV}/c]}} \ , \tag{9.21}$$

where p is the momentum of the incident particle and Z is its charge in units of the electron charge.

---

[1] In a crystal with cubic symmetry, a vector $[ijk]$ with components along the three orthogonal symmetry axes has a perpendicular atomic plane, which is denoted (ijk).

Thermal vibrations, the discreteness of the crystal lattice, and the presence of the electrons in the target all increase the transverse energy of a channeled particle, and can ultimately lead to dechanneling. This is approximated by an exponential depletion of the number n of channeled particles with the traversed distance z:

$$n = n_0 \exp(-z/L_0) \ . \tag{9.22}$$

The empirical parameter L_0 is called the dechanneling length, and it increases linearly with momentum. For silicon, we have [38] $L_0 \approx 0.9$ m $p[\text{TeV}/c]$. Since scattering on nuclei is an important dechanneling process, the channeling can be improved by cooling of the target.

Another concern are imperfections on the crystal surface. These give rise to an inefficient surface layer, typically a few micrometers thick, in which no channeling takes place. To be extracted in a single pass, a particle must enter the crystal with an impact parameter larger than the thickness of the surface layer. On the other hand, particle passing through the surface area experience multiple scattering, and can re-enter the crystal on subsequent revolutions, this time at a larger impact parameter and under the right conditions to be channeled and extracted.

Channeling is possible only for bending radii larger than a minimum 'critical' radius, whose value depends on the crystal, its orientation, and the beam energy. For a proton beam incident parallel to the (110) plane[2] of a silicon crystal, this critical radius is [38]

$$R_c \approx 0.4 \text{ m } p \ [\text{TeV}/c] \ . \tag{9.23}$$

The efficiency of cystal extraction is defined as the number of particles extracted divided by the number of particles lost. Proton extraction efficiencies up to 18% have been obtained [36]. Using a crystal coated with a 30 μm amorphous SiO layer, pure multi-pass extraction with an efficiency of 4–7% was demonstrated [36]. The importance of multi-pass extraction implies that not only the initial impact parameter, but also machine parameters such as the beta function at the crystal and the betatron tune play an important role for the overall efficiency.

Finally, in addition to protons also heavy ions can be extracted by a bent crystal. For fully stripped Pb ions ($Z = 82$) at 22 TeV, an extraction efficiency of 10% was achieved at the SPS [36]. This value was slightly lower than for protons of equivalent energy per nucleon.

---

[2] In a crystal with cubic symmetry, every vector $[ijk]$ defines a perpendicular atomic plane (ijk).

Exercises

9.1 Septum Fields for Injection and Extraction

Suppose that the minimum beam separation at the septum is $x_{\text{sep}} > n_s \sigma_x$. Derive an expression for the integrated kick strength $B_{\text{kic}} L_{\text{kic}}$, with B_{kic} the kicker magnetic field and L_{kic} its length, as a function of normalized emittance $\epsilon_{x,N}$ and energy for a proton beam.

Assume $n_s = 10$, $\beta = 100$ m, $\mu = \pi/2$, a kicker length of $L_{\text{kic}} = 5$ m, and a normalized emittance $\epsilon_{x,N} = 4$ μm. Compute the magnetic field B_{kic} required at a beam energy of 10 GeV and at 10 TeV.

9.2 Emittance Dilutions due to Injection Errors

Consider injection into a storage ring with a 1 mm orbit error at $\beta = 100$ m, in both betatron phases.

a) Estimate the corresponding (growth in) normalized emittance after complete filamentation for proton, muon and electron beams at 10 GeV and at 1 TeV. Compare this with the design normalized emittances of the LHC (3.75 μm), the multi-TeV Muon Collider (50 μm), and the NLC (3 nm vertically). Note that $m_e c^2 = 511$ keV, $m_p c^2 = 938$ MeV, and $m_\mu c^2 = 105.7$ MeV.

b) In general the orbit error results in emittance growth comparable to the design emittance when it is of the same order as the rms beam size. Calculate the rms beam sizes for a 7-TeV proton beam (LHC), a 500-GeV electron beam (NLC) and a 2-TeV muon beam (MC) at $\beta = 100$ m.

9.3 Filamentation

Consider a point bunch which is injected somewhere in phase space at a radius r. Compute the projected beam density $p(x)$, normalized to unity, after filamentation.

9.4 Particle Impact for Slow Extraction

Derive (9.16).

9.5 Crystal Channeling

What is the maximum bending angle over a length of 3 cm, for the LHC beam energy of 7 TeV?