Journal of
Synchrotron
Radiation

ISSN 1600-5775

Received 6 March 2014
Accepted 28 June 2014

Lattice design challenges for fourth-generation storage-ring light sources

Michael Borland,* Glenn Decker, Louis Emery, Vadim Sajaev, Yipeng Sun and Aimin Xiao

Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA.
*E-mail: borland@aps.anl.gov

Third-generation low-emittance storage-ring light sources based on double- and triple-bend cells and undulator magnets have been in operation around the world for more than two decades. On the horizon is a new generation based on the multi-bend achromat (MBA) lattice concept promising two to three orders of magnitude higher brightness than is available in today's sources. In this paper, the challenges inherent in designing MBA lattices, as well as potential solutions, are described. Topics covered include lattice concepts, scaling of storage-ring performance, brightness optimization, nonlinear dynamics, beam lifetime and injection schemes.

Keywords: lattice; storage ring; beam dynamics; brightness.

1. Introduction

Utilization of synchrotron radiation from electron storage rings began with parasitic experiments on machines dedicated to high-energy physics. In time, these sources were replaced by a second generation that were dedicated to synchrotron radiation, but not well optimized for that purpose. Bending and wiggler magnets served as the primary radiation-generating devices, emitting radiation over a broad spectrum. The invention of the undulator magnet and development of the technology for producing much smaller electron beam size and divergence brought about a third generation and a rapid increase in the number of sources in operation. These new sources deliver radiation of vastly higher quality than their predecessors and collectively serve tens of thousands of users annually.

Not so recently, it was considered common knowledge that, with this third generation, storage rings had essentially reached their ultimate performance. This was in spite of the fact that several groups (*e.g.* Einfeld & Plesko, 1996; Kaltchev *et al.*, 1996) had explored a possible path to dramatically better sources. That path, the multi-bend achromat (MBA) lattice, has been adopted by several facilities now under construction and is being seriously contemplated as a means of upgrading existing facilities. For large hard X-ray facilities in particular, it promises two to three orders of magnitude improvement in brightness, justifying the designation of these machines as a fourth generation of storage-ring light sources.

In this paper we review the promise and challenges of these machines. We begin with a review of storage-ring physics, including basic discussions of single-particle dynamics, collective effects and the source of emittance. We then present a study of storage-ring scaling using a model that illustrates both the potential and difficulties of MBA lattices. We describe the various methods that are used to counter some of these difficulties, including methods of optimizing nonlinear dynamics. Finally, we look at an exploratory design for a possible upgrade of the Advanced Photon Source (APS), using this as an illustration of possible solutions and performance.

2. Measures of radiation quality

Modern storage-ring light sources are valued for their ability to produce intense highly collimated radiation from X-rays to infrared. Brightness, an important measure of beam quality, is expressed in simplified form (for untilted beams) as

$$B = \frac{N_\gamma}{4\pi^2(\Delta\lambda/\lambda)\Delta t\Sigma_x\Sigma_{x'}\Sigma_y\Sigma_{y'}}, \qquad (1)$$

where N_γ is the number of photons within a fractional energy or wavelength band $\Delta\lambda/\lambda$ arriving during time Δt. The factor $F_s = N_\gamma/(\Delta\lambda/\lambda)\Delta t$ is the spectral flux. $\Sigma_{x,y}$ and $\Sigma_{x',y'}$ represent the r.m.s. size and divergence of the photon beam in the horizontal and vertical planes. Hence, the brightness is simply the spectral flux per unit volume in transverse phase space.

The properties of the X-ray beam depend on the properties of the parent electron beam, the properties of the device that causes the radiation, and the photon energy of interest. We may determine the properties of the X-ray beam by convolving the radiation distribution for a single electron, which depends on electron energy, the properties of the radiator and the photon energy, with the electron beam distribution. For undulator radiation, an approximate description of the single-

electron transverse distribution is given by Elleaume (2003), who characterizes the r.m.s. source size as

$$\sigma_r \simeq (2\lambda L_\mathrm{u})^{1/2}/(2\pi) \tag{2}$$

and the divergence as

$$\sigma_{r'} \simeq (\lambda/2L_\mathrm{u})^{1/2}, \tag{3}$$

where L_u is the length of the undulator. It is convenient to define the intrinsic radiation emittance as

$$\varepsilon_r = \sigma_r \sigma_{r'} = \lambda/2\pi \tag{4}$$

and the beta function as

$$\beta_r = \sigma_r/\sigma_{r'} \simeq L_\mathrm{u}/\pi. \tag{5}$$

The emittance is an invariant quantity that is proportional to the phase-space area occupied by the radiation, while the beta function characterizes the partitioning of the emittance between size and divergence.

Again taking the simplest case, the electron beam can also be described with Gaussian parameters. However, in this case we consider the emittance and beta function to be the fundamental quantities, since these are under direct control of the accelerator designer. Letting q stand for x or y and taking the simplest case [compare equation (16)], we have an r.m.s. size

$$\sigma_q = \left(\varepsilon_q \beta_q\right)^{1/2} \tag{6}$$

and an r.m.s. divergence

$$\sigma_{q'} = \left(\varepsilon_q/\beta_q\right)^{1/2}. \tag{7}$$

Given that we have characterized both the intrinsic radiation distribution and the electron beam in terms of upright Gaussian, we can easily compute the total photon beam parameters by simple quadrature addition, obtaining

$$\Sigma_q \Sigma_{q'} = \left(\varepsilon_q \beta_q + \varepsilon_r \beta_r\right)^{1/2} \left(\frac{\varepsilon_q}{\beta_q} + \frac{\varepsilon_r}{\beta_r}\right)^{1/2}. \tag{8}$$

To increase the brightness, we must reduce this product for both planes. Clearly, if $\varepsilon_q \ll \varepsilon_r$, the electron beam transverse properties will typically be unimportant as long as β_q and β_r are of similar magnitude. In the case when ε_q and ε_r are comparable, the optimum condition requires $\beta_q \approx L_\mathrm{u}/\pi$, in which case $\Sigma_q \Sigma_{q'} = \varepsilon_q + \varepsilon_r$.

Another measure of beam quality that is independent of intensity is the coherent fraction, given by

$$f_\mathrm{c} = \frac{\varepsilon_r^2}{\Sigma_x \Sigma_{x'} \Sigma_y \Sigma_{y'}}. \tag{9}$$

This is simply the ratio of the intrinsic (*i.e.* minimum possible) phase-space volume to the actual volume. It is customary to speak of being 'diffraction-limited' when

$$\varepsilon_q < (1/2)\varepsilon_r \simeq \lambda/4\pi. \tag{10}$$

In this case, the coherent fraction is approximately 44%. The coherent flux is readily computed as $F_\mathrm{c} = f_\mathrm{c} F_\mathrm{s} = B(\lambda/2)^2$.

When the coherent fraction is close to unity, we may reasonably say that we have an 'ultimate' light source. It is interesting, then, to understand how far existing light sources are from this ideal.

For an undulator that fills a typical 5 m-long straight section, the ideal beta function is $\beta_r = 1.6$ m, which is achieved in some existing machines. Emittance, however, is another matter. In practical units, we may rewrite (10) as

$$\begin{aligned} \varepsilon_q\,[\mathrm{pm}] &\lesssim 100/E_p\,[\mathrm{keV}] \\ \varepsilon_q\,[\mathrm{pm}] &\lesssim 8\lambda\,[\mathrm{\AA}]. \end{aligned} \tag{11}$$

Hence for 'typical' X-ray energies of 1 and 10 keV, we would need electron beam emittances of 100 and 10 pm, respectively, to achieve the diffraction-limited condition.

At third-generation sources, typical vertical emittances ε_y are 1 to 40 pm, which is in the diffraction-limited regime, depending on the photon energy. In the horizontal plane, however, ε_x is typically 1 to 5 nm, which is more than an order of magnitude larger than the diffraction limit. Hence, there is great potential to improve the brightness, coherent fraction and coherent flux from storage rings provided we can reduce the horizontal emittance significantly.

3. Review of storage ring physics

3.1. Basic concepts

A typical third-generation storage ring is a highly periodic configuration of magnets arrayed around a long narrow vacuum chamber that forms a closed loop. The electron beam circulates within this chamber, its path guided by 'bending' magnets (also known as dipole magnets). Bending magnets were the primary source of radiation in earlier-generation sources and are often still used for that purpose. However, the bending magnets occupy only a modest fraction (perhaps 20–30%) of the circumference. The remainder of the circumference is filled with other necessary components, such as quadrupole magnets, which provide focusing, and sextupole magnets, which correct chromatic and other focusing aberrations. Dipole magnets are characterized by having a nominally constant value of vertical magnetic field $B_y = B_0$ within some arc of the circumference, while quadrupole magnets nominally provide a simple gradient in magnetic field,

$$\begin{aligned} B_y &= B_1 x, \\ B_x &= B_1 y. \end{aligned} \tag{12}$$

Sextupole magnets, the highest-order magnets generally found in light sources, are described by

$$\begin{aligned} B_y &= (1/2)B_2\left(x^2 - y^2\right), \\ B_x &= B_2 xy. \end{aligned} \tag{13}$$

The locations and strengths of these various magnets determines what is commonly referred to as the accelerator 'lattice'. The origin of the name is obscure but is presumably related to the high degree of periodicity in the arrangement of the magnets around the circumference. 'Designing a lattice' refers

to the process of arranging the positions and strengths of these components in such a way as to optimize accelerator performance, as measured by emittance, beta functions, free space for undulators, beam lifetime, ease of beam injection and insensitivity to collective instabilities, among other things. A host of practical considerations must also be incorporated, such as magnet strength limits, synchrotron radiation heat load management, vacuum system performance, assembly methods, alignment accuracy, vibration, *etc*.

If low emittance were the only consideration, the circumference of the accelerator would be as tightly packed as possible with dipoles, quadrupoles and sextupoles. However, a vital feature of third-generation sources is a large number of 'straight sections', which are free of accelerator magnets and largely set aside for insertion devices (IDs), *i.e.* undulators and wigglers. These devices, *i.e.* periodic arrays of magnets typically producing a sinusoidally varying field, create much higher radiation flux and, in the case of undulators, extremely high brightness compared with what is possible from bending magnet sources.

A typical ring will have several tens of straight sections, a fraction of which are devoted to special requirements for accelerator operation that are not part of the lattice *per se*. These include radio frequency (RF) accelerating cavities to restore to the beam the energy it loses to synchrotron radiation, as well as hardware required to inject beam into the ring.

3.2. Single-particle motion and beam sizes

As mentioned, the storage-ring lattice consists of a typically periodic array of magnets, including quadrupole magnets that provide focusing. In some cases, the bending magnets also have a quadrupole gradient and hence provide significant focusing as well. This periodic focusing structure results in periodic 'lattice functions', which govern particle trajectories and help determine the beam size and divergence. Transverse particle motion in a circular accelerator may be broken into two components, expressed for the horizontal plane as

$$x_i(s) = \left[2A_i\beta(s)\right]^{1/2}\cos\left[\varphi_x(s) + \Delta\varphi_i\right] + \delta_i\eta_x(s). \qquad (14)$$

In this equation, the i subscript refers to a specific particle, s is the distance around the circumference C of the accelerator, $\beta_x(s)$ is the beta function, $\varphi_x(s)$ is the betatron phase, $\delta = \Delta E/E_0$ is the fractional momentum offset, and $\eta_x(s)$ is the dispersion function, which determines how the closed orbit varies with electron energy deviation. The first term in this equation represents the rapid 'betatron oscillations' that particles execute about their closed orbit. The quantity A_i quantifies the amplitude of these oscillations and is known as the Courant–Snyder invariant for particle i (Courant & Snyder, 1958). In the APS, for example, the horizontal betatron phase advances by about 72π per revolution. The second term, in contrast, oscillates less rapidly, as δ varies relatively slowly. The differential equation for the betatron oscillations (*i.e.* ignoring the momentum offset) is a harmonic oscillator with periodic position-dependent focusing strength (Hill's equation),

$$x_\beta'' + K_1(s)x_\beta = 0, \qquad (15)$$

where $K_1 = B_1/H$ is the geometric quadrupole strength, $H = \gamma m_e c/e$ is the beam rigidity, and primes represent derivatives with respect to s. Similar equations hold for the vertical plane.

The r.m.s. beam size and divergence in the horizontal ('x') plane can be expressed by taking ensemble averages of the transverse coordinates over all the particles in a beam, yielding

$$\begin{aligned}\sigma_x(s) &= \left\{\varepsilon_x\beta_x(s) + \left[\sigma_\delta\eta_x(s)\right]^2\right\}^{1/2},\\ \sigma_{x'}(s) &= \left\{\varepsilon_x\frac{4+\beta_x'^2(s)}{4\beta_x(s)} + \left[\sigma_\delta\eta_x'(s)\right]^2\right\}^{1/2},\end{aligned} \qquad (16)$$

where $\varepsilon_x = \langle A\rangle$ is the horizontal emittance and $\sigma_\delta = (\langle\delta^2\rangle)^{1/2}$ is the fractional r.m.s. energy spread. (Again, similar equations hold for the vertical plane.)

In a circular accelerator, there are necessarily magnets that bend the beam in the horizontal plane, meaning that the beam must be dispersed with energy in some locations and η_x cannot be everywhere zero. However, in an ideal ring, $\eta_y = 0$, while in a well corrected ring we have $\eta_y \ll \eta_x$. Fig. 1 shows the lattice functions and beam sizes for the present APS ring, which has a fairly typical 'double-bend' or Chasman–Green (Chasman *et al.*, 1975) structure.

At a symmetry point, where the derivatives $\beta_x'(s)$ and $\eta_x'(s)$ are zero, the product of the beam size and divergence is particularly simple and gives the 'effective emittance'

$$\sigma_x\sigma_x' = \varepsilon_x\left\{1 + \frac{\left[\sigma_\delta\eta_x(s)\right]^2}{\beta_x^2}\right\}^{1/2}. \qquad (17)$$

This is particularly relevant for IDs because the center of a straight section is almost always a symmetry point. In this case, it would appear that the effective emittance can be minimized if we arrange for $\eta_x = 0$, which is called the achromatic

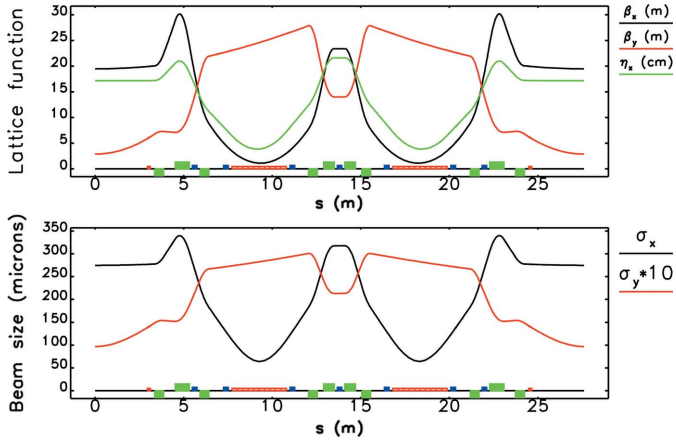

Figure 1
Lattice functions and beam sizes for one sector of the APS. Red blocks represent dipole or bending magnets, green blocks represent quadrupole magnets, and blue blocks represent sextupole magnets. The magnets are sliced into sections to provide a finer representation of the lattice functions.

condition. However, that is not necessarily true, because in double-bend lattices, such as the APS, ε_x can be made smaller by allowing η_x to be non-zero in the straight sections. In the case of APS, this results in an effective emittance of about 3 nm with 'distributed dispersion' compared with about 5 nm in the achromatic case (Farvacque *et al.*, 1994; Emery & Borland, 2002). Clearly, it is important to understand what determines the energy spread and emittance since these affect the contributions of the electron beam to the total photon beam phase space area.

3.3. Origin of energy spread and emittance

Energy spread and emittance have their origin in the quantum nature of radiation emitted by electrons in magnetic fields. Two electrons passing through the same magnetic field on identical trajectories will emit different numbers of photons, at different locations and with different energy. Clearly, this effect will directly increase the energy spread of the beam. In addition, it will increase the beam emittance, since electrons receiving different energy kicks while in a magnetic field will have divergent trajectories in the bending plane (typically the horizontal plane).

The dispersion function plays an important role in this process, since it defines the closed orbit for a particle of a given fractional energy error δ. When an electron emits a photon of energy e_p at position s, it suddenly has an additional energy error of $\Delta\delta = -e_\mathrm{p}/E_0$ and thus is an additional distance $\eta_x(s)\Delta\delta$ from its closed orbit. Referring to equation (14), we see that this increases the amplitude of the betatron oscillation by the same amount, thus increasing the amplitude A and, after ensemble averaging, the emittance of the beam.

If this 'quantum excitation' were the only phenomenon at work, we would expect the energy spread and emittance to increase monotonically, until the beam began to scrape on the vacuum chamber. However, a phenomenon know as radiation damping opposes the increase in energy spread and emittance, so that an equilibrium is reached. Damping of the energy spread occurs because the emission of synchrotron radiation depends on the particle energy. The storage ring RF systems restore the average energy U_0 lost each turn. However, particles with $\delta > 0$ ($\delta < 0$) lose more (less) energy than average. Together, this results in a damping of the energy spread increase.

When electrons radiate, the photons are emitted along the direction of travel of the electron. Hence, both the transverse and longitudinal momenta are decreased. The decrease in the transverse momentum is larger when the slope is larger, *i.e.* when the transverse momentum is larger. However, when reaccelerated in the RF cavities, only the longitudinal component is restored. This again results in a damping phenomenon, since we have reduction of transverse momentum that is greater when the magnitude of the transverse momentum is greater.

It can be shown that (Wiedemann, 2003*a*) the equilibrium energy spread and emittance can be expressed as

$$\sigma_{\delta,0} = \left[C_q \gamma^2 \frac{\langle |1/\rho^3| \rangle}{J_z \langle 1/\rho^2 \rangle} \right]^{1/2} \tag{18}$$

and

$$\varepsilon_0 = C_q \gamma^2 \frac{\langle \mathcal{H}/|\rho^3| \rangle}{J_x \langle 1/\rho^2 \rangle} \tag{19}$$

where γ is the relativistic factor, J_x (J_z) is the horizontal (longitudinal) damping partition number, $\rho(s)$ is the local bending radius of the ideal closed orbit, $C_q = 3.8319 \times 10^{-13}$ m is a combination of physical constants, and

$$\mathcal{H} = \beta_x \eta_x'^2 + 2\alpha_x \eta_x \eta_x' + \gamma_x \eta_x^2, \tag{20}$$

with $\alpha_x = -\beta_x'/2$ and $\gamma_x = (1 + \alpha_x^2)/\beta_x$. Angle brackets represent averages over the entire circumference of the ring.

We are now in a position to understand how to make the energy spread and emittance smaller. One obvious way is to make the beam energy (*i.e.* γ) smaller, but this has several pitfalls. First, the smaller the energy then the more difficult it becomes to make hard X-rays. Second, an intensity-dependent phenomenon known as intrabeam scattering, discussed below, becomes much more severe at low energy and will limit our ability to profit from the γ^2 factor in these equations.

The damping partition numbers J_x and J_z are related by $J_x + J_z = 3$, with $J_x = 1$ typical in rings that lack transverse field gradients in the dipoles. We can increase J_x by incorporating a horizontally defocusing gradient in the dipoles, which will help to reduce the emittance at the expense of energy spread. Usually, this is beneficial as long as not overdone.

Increasing the bending radius ρ may also be expected to have a beneficial effect. The energy spread will scale weakly with radius, *i.e.* $\sigma_\delta \simeq 1/(|\rho|)^{1/2}$, while the emittance will scale like $1/\rho^3$ (once we include the fact that $\mathcal{H} \simeq 1/\rho^2$). In the next section, we explore scaling in more detail.

Since the dipole magnets bend the beam path in the horizontal plane and thus create non-zero dispersion only in that plane, it would appear that the vertical emittance should be essentially zero. However, in all real storage rings there are rotational misalignments that result in non-zero vertical dispersion and also coupling of the motion between the x and y planes. In simple terms, this results in sharing of the natural emittance ε_0 between the x and y planes. It is common to express this with a parameter $0 \leq \kappa \leq 1$,

$$\begin{aligned} \varepsilon_x &= \varepsilon_0[1/(1+\kappa)], \\ \varepsilon_y &= \varepsilon_0[\kappa/(1+\kappa)]. \end{aligned} \tag{21}$$

Using these equations we see that κ is just the ratio $\varepsilon_y/\varepsilon_x$ and that $\varepsilon_x + \varepsilon_y = \varepsilon_0$. (These conclusions must be modified when the damping times are different in the two planes, but we will gloss over this detail in the interest of simplicity.)

3.4. Damping wigglers

As we mentioned before, the equilibrium emittance is a balance between quantum excitation and synchrotron radiation damping. The quantum excitation can be reduced by increasing the focusing of the arc optics, as we will describe

in §4. An alternative would be increasing the damping 'denominator' using a number of strong wiggler magnets in zero-dispersion straight sections. The alternating magnetic fields of wigglers, like undulators, do not produce a net deflection of the beam, but create a large powerful fan of dipole-magnet-like radiation. They are used in linear-collider damping ring design proposals and in several light source projects (Balewski *et al.*, 2004; Guo *et al.*, 2009). In order to gain a large fraction in horizontal emittance the wigglers would have to be specified to extract much more synchrotron radiation than the regular arc dipoles.

The main attraction of wigglers in the design of a low-emittance ring is that they allow the arc optics to have 'larger' quantum excitation than otherwise, thus reducing the overall optical aberrations or nonlinearities, which are considered the main design limitations in low-emittance rings. Another advantage of wigglers is that they are almost modular. That is, wiggler sections can be added or subtracted in operations without excessive beam dynamics impact. The impact of some collective effects will also be reduced compared with non-wiggler light sources because of the extra damping.

However, there are significant disadvantages to using wigglers. In light sources, space and straight sections are at a premium and thus devoting straight sections to wigglers is not always desirable. Since the energy loss per turn has been multiplied with the extra radiation produced by wigglers, the number of RF cavities and power sources may need to be increased, which is particularly a concern for higher energy rings and further reduces the amount of straight section length available for undulators.

Wigglers also introduce asymmetry into the lattice, which can cause beam dynamics problems, as described in the next section. This asymmetry may result from changes to the linear optics that are needed to optimize the performance of the wigglers and compensate for their strong vertical focusing. It also results from field roll-off in the wiggler coupled with the amplitude of the wiggle (Safranek *et al.*, 2002; Xiao *et al.*, 2013), both of which are a concern with strong long-period damping wigglers. Another negative effect of damping wigglers is that the energy spread increases, thus decreasing the brightness of undulators at high harmonics.

In practical terms, one can at most hope to obtain a factor of three or four reduction in emittance from the use of damping wigglers. As we will see, other approaches based on lattice modifications promise much larger factors and preserve more space for high-brightness radiation-producing devices (*i.e.* undulators instead of wigglers).

3.5. Tunes and resonances

Like any oscillatory motion, the motion described by equation (14) is subject to resonances, particularly when we consider that in a realistic accelerator the equation of motion, equation (15), has significant nonlinear terms on the right-hand side. Depending on which nonlinearities are present, motion at particular frequencies can grow without limit, resulting in particle loss. We define the 'tune' of the ring as the normalized betatron phase advance per turn,

$$\nu_q = \frac{\varphi_q(C) - \varphi_q(0)}{2\pi}, \tag{22}$$

where q represents x or y. It is fairly obvious that, when ν_q is an integer, any disturbance (*e.g.* a centroid kick) to the beam will accumulate turn after turn.

For simplicity, in most of this paper we have assumed that motion in the horizontal and vertical planes is separate, when in reality there is some coupling between the two planes. One implication is that betatron motion is subject to many resonances that involve both the horizontal and vertical tunes. These resonances are governed by the equation (Bruck, 1972)

$$n\nu_x + m\nu_y = p, \tag{23}$$

where n, m and p are abitrary integers. This equation represents a resonance of order $|n| + |m|$. When p is a multiple of N_s, the number of sectors in the ring, we speak of structural resonances. If the lattice is precisely symmetrical, these are the only resonances present and they are of high order (and therefore less serious). When the symmetry of the lattice is broken, we effectively have $N_s = 1$ and resonances appear at lower order.

There is a simple relationship between the order of a resonance and the type of magnetic field that can drive it. For example, integer resonances are driven by dipole errors, while half-integer resonances are driven by quadrupole errors. In general, a resonance labeled by (n, m) is driven by $2(|n| + |m|)$-pole magnets. Sextupoles, the need for which is described in the next section, can drive several resonances, *e.g.* $3\nu_x = p$ and $nu_x \pm 2\nu_y = p$.

A particular difficulty is that the electrons in the beam oscillate with a spread of frequencies, owing to the spread in energies and betatron amplitudes. As discussed in more detail in §5.1, one of the tasks of the lattice designer is to keep the beam away from such resonances, or at least ensure that the resonance is not driven strongly. Otherwise, particles can bleed out through a resonance, resulting in short lifetime or poor injection efficiency.

3.6. Chromatic effects

The beam motion described in equation (14) contains several simplifications, one of which is the assumption that the phase advance $\varphi_x(s)$ is independent of the energy deviation δ. However, focusing is provided by magnetic elements (quadrupoles) that of necessity affect particles of different energy differently. The focal length f of an ideal thin-lens quadrupole is $f = f_0(1 + \delta)$, where f_0 is the focal length when the energy deviation is zero. Put another way, the effective value of the geometric focusing strength is

$$K_{1,\mathrm{eff}} = K_1/(1 + \delta) \simeq K_1(1 - \delta). \tag{24}$$

This energy-dependent focusing results in an energy dependence of the tune, known as the chromaticity,

$$\xi_q = \mathrm{d}\nu_q/\mathrm{d}\delta. \tag{25}$$

Typically, when a lattice contains only dipole and quadrupole fields, the chromaticity is negative in both planes. This is referred to as the 'natural' chromaticity.

For example, for the APS we have $\nu_x = 36.1$ and a natural chromaticity $\xi_{x,\mathrm{nat}} = -90$. Thus, a particle with an energy deviation of 0.1/90 or about 0.11% would have a tune of 36, *i.e.* would sit on the integer resonance. Such a particle would certainly be lost. Indeed, since the r.m.s. energy spread of the APS electron beam is $\sigma_\delta = 0.096\%$, a steady stream of particles would be driven across the integer resonance by quantum excitation and be lost, resulting in an extremely short 'quantum' beam lifetime.

To prevent such problems, it is necessary to reduce the chromaticity to a much smaller value. Typically, one would like the crossing of significant resonances to occur at $10\sigma_\delta$, so that the impact on lifetime is negligible. In the APS example just given, this implies $|\xi_x| \lesssim 10$.

In fact, the quantum lifetime is only one consideration motivating a low value of chromaticity. As we will describe below, particles can be excited to momentum deviations of several percent by Touschek scattering events, leading to reduced lifetime for various reasons, *e.g.* scattered particles may damp across a resonance on their way back to the core of the beam, resulting in loss.

As mentioned above, sextupole magnets are introduced into storage rings for the purpose of controlling the chromaticity. To better understand how, we start by rewriting equation (14) as

$$x(s) = x_\beta(s) + \delta\eta_x(s), \tag{26}$$

where $x_\beta(s)$ represents the betatron oscillation. Referring to equation (13) and assuming $y = 0$ for simplicity, we obtain

$$B_y = (1/2)B_2\left(x_\beta^2 + 2\delta x_\beta\eta_x + \delta^2\eta_x^2\right). \tag{27}$$

Recalling that δ changes much more slowly than x_β, it is apparent that the second term of this expression acts as a δ-dependent gradient, so that equation (24) becomes

$$K_{1,\mathrm{eff}}(s) = K_1(s)(1-\delta) + 2\delta\eta_x(s)K_2(s), \tag{28}$$

where $K_2 = B_2/H$. In this equation, we see two energy-dependent gradients, the unavoidable one due to the quadrupoles as well as a term that is driven by the sextupole strength K_2 in dispersive areas (*i.e.* where $\eta_x \neq 0$). Without going into detail, we can see the possibility of 'correcting' the chromaticity by adjusting the location and strength of such 'chromatic' sextupoles. At minimum, two sets or 'families' of sextupoles are needed to allow correction of both the horizontal and vertical chromaticity to zero, which is, naively at least, the ideal value for maximizing the momentum acceptance. In modern light sources, additional families are used in order to allow correction of geometric and higher-order chromatic aberrations introduced by the chromatic sextupoles (Crosbie, 1987). This topic will be discussed in more detail in §5.1.

3.7. Longitudinal motion

As noted above, storage-ring light sources require RF cavities to replace the energy lost to synchrotron radiation. By virtue of the sinusoidal variation of the RF voltage $V(t) = V_0 \sin\omega_{\mathrm{RF}}t$, these RF cavities provide a longitudinal restoring or focusing force that is periodic in time. This results in the existence of an integer number h of stable regions, known as RF buckets, in phase and energy space, where the harmonic number h is the ratio of the RF frequency ω_{RF} to the revolution frequency ω_0. The focusing force further results in bunching of the beam near the center of each bucket at a particular RF phase φ_s known as the synchronous phase, given by

$$\sin\varphi_\mathrm{s} = U_0/V_0 = 1/q \tag{29}$$

where U_0 is the nominal energy loss per turn and $q > 1$ is known as the over-voltage factor.

Any restoring force results in oscillatory motion at a specific frequency. In the transverse planes, these are the two tunes ν_x and ν_y. Similarly, there is a longitudinal tune ν_s and a corresponding frequency Ω_s for 'synchrotron' oscillations, given by (Sands, 1970)

$$\Omega_\mathrm{s}^2 = \frac{\alpha_\mathrm{c}ceV_0\,\omega_{\mathrm{RF}}}{CE_0}\cos\varphi_\mathrm{s}. \tag{30}$$

The parameter α_c, known as the momentum compaction factor, relates the fractional change in orbit length to a change in fractional momentum error,

$$\frac{1}{\Delta\delta}\frac{\Delta L}{C} = \alpha_\mathrm{c} = \frac{1}{C}\int\frac{\eta_x(s)}{\rho(s)}\,\mathrm{d}s, \tag{31}$$

where $\rho(s)$ is the local bending radius of the design trajectory.

Given its role in linking energy deviation to path length, it is perhaps not surprising that α_c helps determine the r.m.s. electron bunch duration, which is proportional to the fractional energy spread σ_δ,

$$\sigma_t = (\alpha_\mathrm{c}/\Omega_\mathrm{s})\sigma_\delta. \tag{32}$$

Since α_c helps determine the bunch duration and thus the peak current within a bunch for a given bunch charge, its value is important for understanding single-bunch collective instabilities. The value of Ω_s, also influenced by α_c, affects both single- and multi-bunch instabilities by determining the rate at which particles in the head and tail of a bunch exchange positions.

Finally, α_c helps determine the RF acceptance $\varepsilon_{\mathrm{RF}}$, also known as the 'bucket height,' which defines the maximum fractional energy deviation a particle can have without being lost from its bucket (and eventually from the machine) (Sands, 1970),

$$\varepsilon_{\mathrm{RF}} = \pm\left\{\frac{2U_0}{\pi\alpha_\mathrm{c}hE_0}\left[\left(q^2-1\right)^{1/2} - \cos^{-1}(1/q)\right]\right\}^{1/2}. \tag{33}$$

Typically, $\varepsilon_{\mathrm{RF}}$ is at least 2%. Higher values may contribute to longer beam lifetime in the presence of inelastic scattering from residual gas atoms and Touschek scattering (see below),

provided that the energy acceptance is not limited first by other factors.

3.8. Collective effects

So far, we have described storage rings in terms of the behavior of particles under the influence of external fields from magnets and RF cavities. While this 'single-particle dynamics' is a rich and challenging subject, we must also be cognizant of 'collective effects,' in which particles interact with each other either directly, as in scattering or *via* space charge, or indirectly, typically *via* electromagnetic interaction with the vacuum chamber. Collective effects become more pronounced as the beam intensity increases, but also as the particle density increases when, for example, we attempt to decrease the beam emittance. If not properly managed, collective effects can result in intensity limitations, short beam lifetime, emittance growth and other undesirable consequences.

Two scattering phenomena are important in low-emittance storage rings. Both have their origins in collisions between pairs of particles within a single bunch. The first, commonly called 'intrabeam scattering' or IBS, refers to multiple small-angle scattering events, wherein each particle repeatedly collides with other particles in the bunch. Detailed treatment of IBS is available in the literature (Bjorken & Mtingwa, 1983). For present purposes, it suffices to note that the local collision rate is proportional to the bunch intensity and inversely proportional to the bunch volume, *i.e.*

$$R_{\rm IBS} \propto \frac{I}{N_{\rm b}\sigma_x\sigma_y c\sigma_t}, \tag{34}$$

where I is the total beam current and $N_{\rm b}$ is the number of bunches. During each collision, particles may experience relatively large transfers of momentum from betatron oscillations into synchrotron oscillations. When this happens in regions where η_x is nonzero, it may well induce a larger betatron oscillation in a similar fashion to quantum excitation. The result is that both the energy spread and the horizontal emittance grow. As (34) implies, the phenomenon is worse for low-emittance rings with relatively few high-current bunches. Intrabeam scattering is also worse for low-energy rings, which is important as it curtails our ability to take advantage of the $\varepsilon \propto \gamma^2$ dependence exhibited in equation (19). We will explore this in more detail below.

The Touschek effect is related to IBS but refers to relatively rare hard scattering within a bunch that results in sufficiently large momentum deviations that one or both particles are lost. As with IBS, the Touschek scattering rate is proportional to the bunch intensity and inversely proportional to the bunch volume. In low-emittance rings, the beam lifetime τ, defined as

$$\frac{1}{\tau} = \frac{1}{I}\frac{{\rm d}I}{{\rm d}t}, \tag{35}$$

is typically dominated by the Touschek scattering contribution. For third-generation rings, it is the deviation in the longitudinal momentum after a collision that matters. Thus having sufficient RF bucket height is essential in obtaining sufficiently long lifetime. In addition, the momentum acceptance of the lattice must be sufficient, *i.e.* transverse motion of the particles must be stable for large longitudinal momentum deviations. Hence, arrangement and adjustment of the sextupole magnets is an important technique for increasing the Touschek lifetime.

IBS and Touschek scattering are collective effects that result from the microscopic interaction of pairs of particles. An individual particle can also interact with the macroscopic field of the bunch in which it resides, or indeed of the entire set of bunches that compose a beam. This interaction includes direct space charge forces and, more importantly for high-energy beams, interaction mediated by the vacuum chamber walls. In particular, any irregularity or resistivity in the vacuum chamber walls can cause an electromagnetic 'wake' to form when a bunch passes, resulting in a time-dependent modulation of the longitudinal and transverse coordinates of the particles in the bunch. An equivalent way to characterize this interaction is with a frequency-dependent complex impedance, which characterizes the voltage produced by a given current distribution. As the bunch intensity increases, the effect of the wake or impedance increases, resulting variously in increased energy spread, increased emittance, motion of the beam centroid and beam loss. These effects can to some extent be mitigated by increased (positive) chromaticity and feedback.

Hence, it is common for storage rings to be tuned for positive chromaticity, particularly when running with high single-bunch current is desired. In the APS, for example, a chromaticity of about +7 is needed to stabilize a 5 mA bunch, while +11 is needed to stabilize a 20 mA bunch. As discussed above, when the chromaticity is too different from zero, it can adversely impact momentum aperture and hence beam lifetime. Use of bunch-by-bunch feedback allows reducing these values considerably (Revol & Nagaoka, 2001; Yao *et al.*, 2011). Benefits may also accrue from using a higher-harmonic RF cavity to stretch the bunches and introduce a spread in synchrotron frequencies, as reported, for example, for the Aladdin Light Source (Bosch *et al.*, 2001; Bosch, 2005), with theoretical grounding by Sagan (1994) and Krinsky & Wang (1985).

Some structures, such as the RF cavities used to replace the energy lost to synchrotron radiation, can trap electromagnetic energy from passing bunches in resonant modes. These resonant modes allow bunches to affect other bunches, and hence can cause a multibunch instability. Bunch-by-bunch feedback systems are quite effective in dealing with multibunch instabilities.

4. Scaling of ring performance

Understanding how ring performance scales with circumference and lattice type is very beneficial in guiding design efforts. If we imagine a ring of fixed circumference C divided into $N_{\rm s}$ identical sectors with $N_{\rm d}$ identical dipoles each, we know that the bending angle per dipole is $\theta_{\rm d} = 2\pi/(N_{\rm s}N_{\rm d})$. For small angles, the dispersion slope η_x' added by a single dipole is proportional to $\theta_{\rm d}$. In addition, the average beta function will

scale like $l_d = C\theta_d/(2\pi)$, since that characterizes the distance between focusing elements. Referring to equation (20), we can expect that $\mathcal{H} \simeq C\theta_d^3$ and thus, referring to equation (19), we anticipate that $\varepsilon_0 \simeq \mathcal{H}/\rho \simeq \theta_d^3 \simeq 1/(N_s N_d)^3$.

Indeed, for several specific lattice types (*e.g.* double-bend achromat), Wiedemann (2003*b*) shows that the minimum emittance scales in this fashion. The emittance also depends on the choice of lattice type and horizontal tune, so following Murphy (1989) we can write

$$\varepsilon_0 \propto F(\text{type}, \nu_x) \frac{\gamma^2}{J_x (N_s N_d)^3}. \quad (36)$$

Strictly speaking, this scaling formula should only be used for a fixed lattice type (*e.g.* a double-bend or triple-bend design), but in practical use we find it is reasonably accurate even when that condition is not met.

Hence, the 3.1 nm emittance of the APS may be expected to drop 40-fold if we replace the existing double-bend sectors design by a seven-bend design. As an indication of the importance of the $F(\text{type}, \nu_x)$ factor, we note that APS started operation with an emittance of about 8 nm, transitioning to 3.1 nm through adjustment of quadrupole strengths.

With such dramatic benefits apparently possible, it is worthwhile to understand the scaling of other storage-ring parameters. Rather than attempt to tease this out through analysis, we decided to simulate a model ring that allows more readily computing scaling of intricate parameters. This simplified model has no straight sections for IDs and consists of a simple repetitive cell with a 'theoretical minimum emittance' (TME) configuration, as shown in Fig. 2. This is very similar to the central cells of the MAX IV sector (Leemann *et al.*, 2009), for example.

For our model, we have chosen a fixed intermediate circumference of 600 m and an intermediate energy of 4.5 GeV. Because of the simplicity of the cell, the distinction between N_s (the number of sectors) and N_d is meaningless, so we just use the latter for simplicity of terminology. By varying N_d, we can examine the changes in various parameters as our

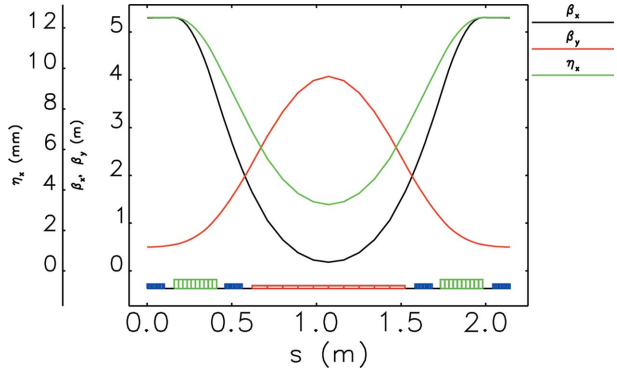

Figure 2
Lattice functions for one cell of the model ring for the case $N_d = 280$. Red blocks represent dipole or bending magnets (B1), green blocks represent quadrupole magnets (Q1), and blue blocks represent sextupole magnets (SF and SD).

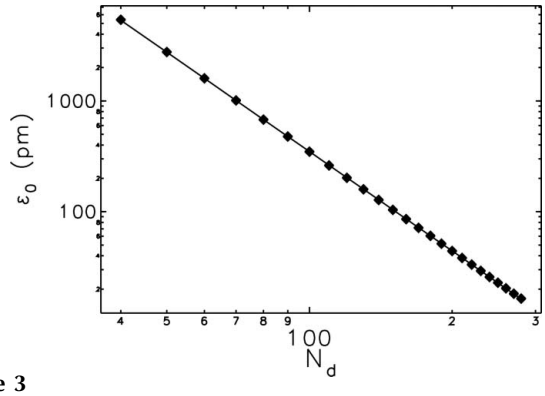

Figure 3
Scaling of natural emittance with number of dipoles in the model ring. The line shows a fit, giving $\varepsilon_0 \propto N^{-2.98}$.

hypothetical 600 m, 4.5 GeV ring is upgraded to lattices with different numbers of dipoles.

The modeling of the ring was performed with the program *elegant* (Borland, 2000), starting with $N_d = 40$. As N_d was increased, the previous solution was used as the starting point for the new solution. For each value of N_d, we minimized ε_0 subject to several constraints: 0.5 m $\leq \beta_{x,y} \leq$ 40 m, $J_{x,z} \geq 1$, and circumference equal to 600 m. The upper limitation on the beta function is needed to ensure that beam sizes, sensitivity to errors, and collective effects are reasonably constrained, while the lower limit was chosen as a practical value to prevent large changes in the character of the solutions. The lengths of the dipoles and quadrupoles was allowed to vary, as was their gradient and the distances between the magnets. The lengths of the sextupoles (a small fraction of the circumference) were kept fixed. Fig. 2 shows an example of the lattice functions for one cell of the ring with $N_d = 280$.

As Fig. 3 shows, the emittance does indeed scale as $1/N_d^3$. The fractional energy spread is found to scale very weakly, like $1/N_d^{0.06}$, which is understandable given that the radius of the ring is fixed.

As noted above, the decrease in emittance results from using strong focusing elements between the dipoles to reduce the magnitude of the dispersion function η_x. Not surprisingly then, the average value of η_x decreases rapidly, like $1/N_d^{1.9}$. Also not surprising is the rapid increase in the focusing gradients, shown in Fig. 4. The exponents differ for the gradient in the quadrupoles ('Q1') and dipoles ('B1'), and the scaling is not a pure power law, but in general the scaling is like N_d^2. The integrated strength of the focusing elements, defined as the product of the gradient and the length of the element, increases more slowly, like N_d, because the length of the elements must shrink like $1/N_d$ in order to fit the larger number of elements within the same circumference. That scaling is also in keeping with the focusing elements having to produce a focal length about equal to their separation.

Because of the increase in the number of focusing elements and the emphasis on focusing in the bending (dispersive) plane, the horizontal tune and natural chromaticity scale like N_d. Coupled with the $1/N_d^{1.9}$ drop in the dispersion function,

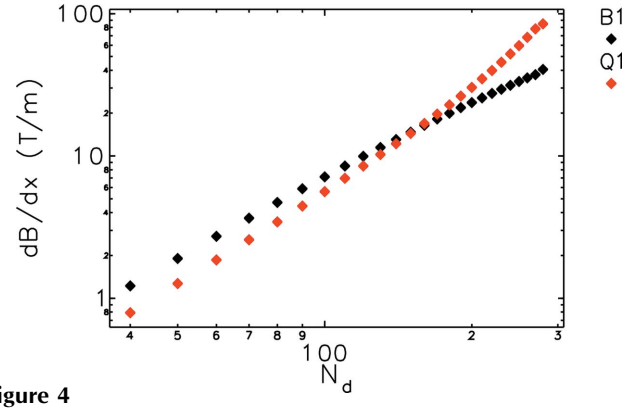

Figure 4
Scaling of focusing gradients in the quadrupoles ('Q1') and dipoles ('B1'). Fitting gives approximate scaling of $N_{\rm d}^{2.4}$ for the quadrupoles and $N_{\rm d}^{1.8}$ for the dipoles.

we expect a $\sim N_{\rm d}^3$ increase in the required sextupole strength, which is indeed seen in Fig. 5.

The rapid increase in the focusing and sextupole strengths has significant implications for magnet design. A good rule of thumb is that the field on the pole tips of quadrupoles and sextupoles should not exceed $B_t = 1$ T. For quadrupoles, this implies a pole tip radius of

$$R_{\rm q} = \frac{B_t}{{\rm d}B_y/{\rm d}x}, \tag{37}$$

while for sextupoles it implies a pole tip radius of

$$R_{\rm s} = \left(\frac{2B_t}{{\rm d}^2B_y/{\rm d}x^2}\right)^{1/2}. \tag{38}$$

As shown in Fig. 6, this implies a rapid reduction in the magnet bore radii. The conclusion is that for ultralow emittance we require bore radii of the order of 10 mm. This has significant repercussions for the vacuum systems, since the conductance of the vacuum pipe scales with the cube of the radius. In order to keep the average pressure fixed, the spacing between pumps should scale like R^3, *i.e.* like $1/N_{\rm d}^6$. Not surprisingly, a new approach is needed and has been found in the use of

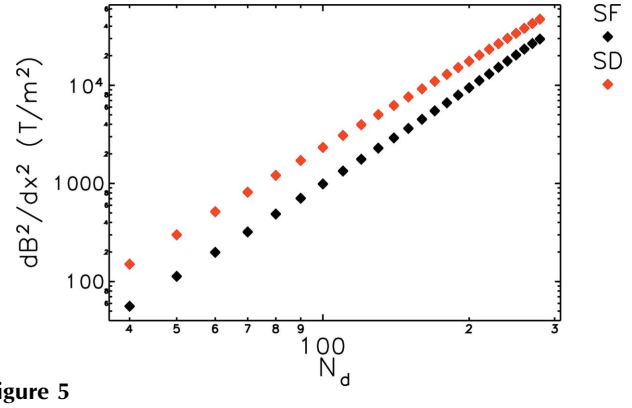

Figure 5
Scaling of sextupole strength required to correct the chromaticities to 0 for the two sextupole families. Fitting gives approximate scaling of $N_{\rm d}^{3.2}$ for the SF and $N_{\rm d}^{2.9}$ for the SD sextupoles.

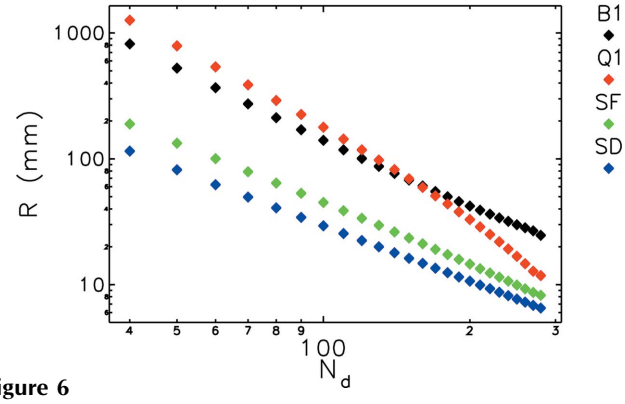

Figure 6
Scaling of dipole, quadrupole and sextupole bore radii assuming the pole tip field is 1 T. The dipole and quadrupole values scale approximately as $N_{\rm d}^{-2}$, while the sextupole values scale approximately as $N_{\rm d}^{-1.5}$.

NEG-coated chambers, which have continuous pumping along the circumference (Kersevan, 2000).

The strong variation of fields with transverse position implied by the high quadrupole and sextupole strengths implies greater sensitivity to alignment errors and vibration. Misaligned quadrupoles kick the beam orbit, while misaligned sextupoles introduce focusing errors and cross-plane coupling. For the horizontal (x) plane, the orbit is primarily affected by misalignments of the focusing quadrupoles (Q1), while for the vertical (y) plane the orbit is primarily affected by misalignments of the gradient dipoles (B1). For a displacement Δx of a focusing single quadrupole, the orbit displacement is (Sands, 1970)

$$x_1(s) = K_1 L \Delta x \, \frac{\left[\beta_x(0)\beta_x(s)\right]^{1/2}}{2\sin\pi\nu_x} \cos\left[\varphi_x(s) - \varphi_x(0) - \pi\nu_x\right], \tag{39}$$

where $K_1 = (1/H){\rm d}B_y/{\rm d}x$ is the geometric strength of a quadrupole of length L at $s = 0$, $H = \gamma m_{\rm e}c/e$ is the beam rigidity, $\varphi_x(s)$ is the betatron phase advance, and $\nu_x = \varphi_x(C)/(2\pi)$ is the horizontal tune. For $2N_{\rm d}$ independent displaced quadrupoles, we find the r.m.s. horizontal orbit amplification factor at the Q1 locations

$$F_x \simeq \frac{x_{\rm rms}}{\Delta x_{\rm rms}} = (2N_{\rm d})^{1/2} \, \frac{K_1 L \beta_{x,\max}}{2\sqrt{2}\sin\pi\nu_x} \tag{40}$$

where we have made use of the fact that β_x is approximately maximal at the Q1 locations (see Fig. 2). Similarly, the r.m.s. vertical orbit at the B1 locations is

$$F_y \simeq \frac{y_{\rm rms}}{\Delta y_{\rm rms}} = (N_{\rm d})^{1/2} \, \frac{K_1 L \beta_{y,\max}}{2\sqrt{2}\sin\pi\nu_y}, \tag{41}$$

where we have made use of the fact that β_y is maximal at the B1 locations. Assuming for simplicity and uniformity that $\sin\pi\nu_x = \sin\pi\nu_y = 1/\sqrt{2}$, we can estimate the amplification factors for our model lattice, as shown in Fig. 7. The dependence is relatively weak compared with magnet strength, say, a result of the significant drop in the maximum beta functions in both planes as $N_{\rm d}$ is increased.

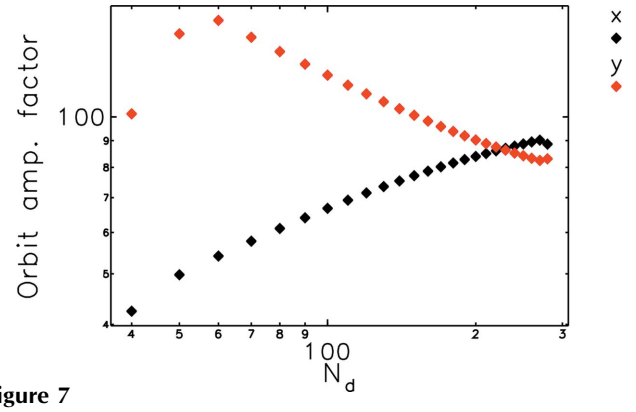

Figure 7
Scaling of orbit amplification factors for horizontal offsets of Q1 quadrupoles and vertical offsets of B1 dipoles. The values scale approximately as $N_d^{0.3}$ and $N_d^{-0.4}$, respectively.

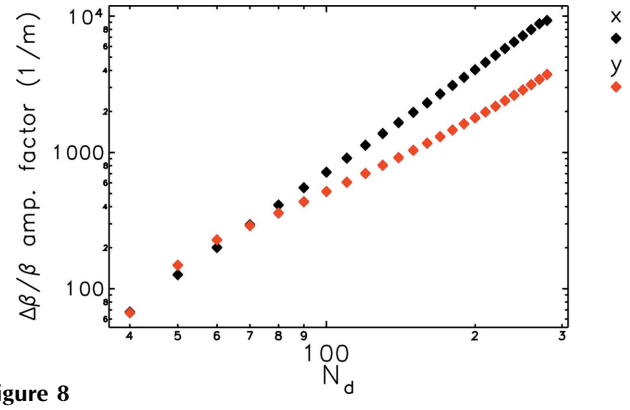

Figure 8
Scaling of fractional beta function error amplification factors for horizontal offsets of sextupoles. The values scale approximately as $N_d^{2.7}$ for the x plane and and $N_d^{1.8}$ for the y plane.

We can also estimate the alignment requirements for sextupoles using the relationship between a beta function error and the focusing error $(d^2B_y/dx^2)L\Delta x$ for a single sextupole displaced by Δx,

$$\frac{\Delta\beta_x(s)}{\beta_x(s)} = \frac{\beta_x(0)}{2\sin 2\pi\nu_x} K_2 L\Delta x \cos 2\left[\varphi_x(s) - \varphi_x(0) - \pi\nu_x\right], \quad (42)$$

where $K_2 = (1/H)\,d^2B_y/dx^2$ is the normalized strength of the sextupole of length L at location $s = 0$.

Consistent with our assumption above for the orbit amplification estimate, we will assume $\sin 2\pi\nu_x = \sin 2\pi\nu_y = 1$. From Fig. 2, we see that the horizontal beta function at the SF sextupoles (at the ends of the cell) is approximately maximal, while the vertical beta function is minimal. In contrast, for the SD sextupole both beta functions are close to the average value for the cell. Thus, for the horizontal plane we can write

$$\left(\frac{\Delta\beta_x}{\beta_x}\right)_{\rm rms} \simeq \frac{\beta_{x,\max}\left|K_2L\right|_{\rm SF} + \sqrt{2}\beta_{x,\rm ave}\left|K_2L\right|_{\rm SD}}{4} \times (N_d)^{1/2}(\Delta x)_{\rm rms}, \quad (43)$$

while for the vertical plane we have

$$\left(\frac{\Delta\beta_y}{\beta_y}\right)_{\rm rms} \simeq \frac{\sqrt{2}\beta_{y,\rm ave}\left|K_2L\right|_{\rm SD}}{4}(N_d)^{1/2}(\Delta y)_{\rm rms}. \quad (44)$$

Normalizing to the displacement $(\Delta x)_{\rm rms}$ or $(\Delta y)_{\rm rms}$ gives the amplification factors shown in Fig. 8. The scaling is quite rapid, particularly for the horizontal plane, where it is close to N_d^3. For the APS double-bend lattice ($N_sN_d = 80$), the alignment tolerances for sextupoles were 150 µm. This analysis implies that for a seven-bend lattice ($N_sN_d = 280$) the tolerance would be 5–15 µm, which is close to state of the art (Jain, 2011).

As we saw above, the strength of the sextupoles increases roughly as N_d^3. This has significant implications for nonlinear dynamics, which we can explore using scaling arguments (Emery, 1990). Looking at only the horizontal plane, and introducing the functions $k(s)$ and $m(s)$ giving, respectively, the distribution of K_1 and K_2 as a function of s, the equation of motion for a general lattice with N_d cells is

$$\frac{d^2x}{ds^2} + k(s)x = -\frac{1}{2}m(s)x^2, \quad (45)$$

where the s dependence of focusing strength for quadrupoles and sextupoles are made explicit in order to find a scaling law with N_d. For simplicity, we assume that the sextupole magnet length shrinks with the cell length, whereas above the sextupole length was fixed. This implies that m will scale as N_d^4 (not $\sim N_d^3$ as found above), in order to give the correct $\sim N_d^3$ scaling of the integrated strength of each sextupole. We start with the equation of motion for a reference value of N_d, namely N_{d0},

$$\frac{d^2x}{ds^2} + k_0(s)x = -\frac{1}{2}m_0(s)x^2, \quad (46)$$

for which, we assume, we know the stability boundary and instability growth times. Obviously the differences between (45) and (46) are in the focusing and sextupole terms, which are different in magnitude and spatial frequency. We will rewrite the terms of (46) to relate to those of (45). Using the scaling found above, the quantity $k(s)$ can be written as $n_d^2k_0(n_ds)$ where $n_d = N_d/N_{d0}$ is the normalized number of cells. The factor n_d in the argument n_ds indicates that magnet strengths vary more rapidly with S as we pack more magnets into the fixed-circumference ring. The n_d^2 factor indicates that the quadrupole gradient increases as the square of number of cells, roughly as found above. Similarly we have $m(s) = n_d^4m_0(n_ds)$ for the general sextupole distribution. Rewriting the double derivative d^2x/ds^2 as $n_d^2d^2x/d(n_ds)^2$, (45) becomes

$$n_d^2\frac{d^2x}{d(n_ds)^2} + n_d^2k_0(n_ds)x = -\frac{1}{2}n_d^4m_0(n_ds)x^2. \quad (47)$$

Letting $\chi = n_d^2x$ and $\sigma = n_ds$, then

$$\frac{d^2\chi}{d\sigma^2} + k_0(\sigma)\chi = -\frac{1}{2}m_0(\sigma)\chi^2, \quad (48)$$

which exhibits the same structure as (46), for which we already know the solution. The stability limit of χ (for a general n_d) in (48) has the same value as that of x (for $n_d = 1$) in (46). Thus the stability limit for a general N_d goes as $1/N_d^2$. The time

evolution of the instability goes as $1/N_d$, which means that the instability follows with the number of cells or number of sextupoles. The same result is found for the vertical-plane equation of motion.

Hence, if a double-bend lattice is replaced by a seven-bend lattice with the same number of cells, we expect the stability limit, or 'dynamic aperture', to be reduced to about 10% of the original value. This seems to imply that conventional injection, which requires a dynamic aperture of 10 mm or more, is impossible, a subject to which we return in §6.

Another aperture limit with which we must contend results from the variation in tune (oscillation frequency) with momentum offset. This influences the momentum acceptance, which in turn influences the beam lifetime *via* Touschek scattering (Bernardini *et al.*, 1963). Extending the description in §3.6, we can write

$$\nu_q(\delta) = \nu_q(0) + \delta\xi_q + \delta^2\xi_{q,2} + \delta^3\xi_{q,3} \ldots . \qquad (49)$$

Typically, we correct ξ_q to a small value and attempt to do likewise with the higher-order chromaticities using additional families of sextupoles. Since our model has only two families, we can only set $\xi_x = \xi_y = 0$. The resulting $\sim N^{1.6}$ increase in the higher-order chromaticities gives us an indication of the challenges faced in optimizing a workable lattice using additional sextupoles or other methods (described in more detail in §5).

These rough estimates give somewhat dire conclusions and are introduced primarily to provide an understanding of why designing ultralow-emittance light sources is so difficult. In reality, clever schemes and better optimization tools can significantly improve upon these scaling-based aperture expectations, as we will see below.

So far, we have primarily looked into scaling effects in the transverse planes. However, the motion in the longitudinal plane is also affected. In particular, the all-important parameter α_c scales like $1/N_d^{1.9}$. As mentioned above, the RF bucket height needs to be sufficiently large regardless of N_d, in order to maintain good lifetime, which becomes easier as α_c shrinks [see equation (33)]. If we maintain a typical bucket height of $\pm 3\%$, we can also determine the scaling of related quantities. For example, the required RF voltage scales like $1/N_d$, primarily a result of the change in α_c, since the other quantities going into equation (33) change only weakly. As a result, the synchrotron tune Ω_s decreases as $1/N_d^{1.5}$ and the bunch length scales as $1/N_d^{0.4}$. One might hope to counteract this reduced bunch length, which is undesirable from the standpoint of collective instabilities and Touschek scattering, by reducing the RF voltage, but that would result in reduced RF acceptance. This is not a wise choice when the emittance is shrinking rapidly, since small emittance leads to more rapid Touschek scattering and thus a strong requirement for large energy acceptance.

Having determined the scaling of the beam dimensions in all three planes, we are now in a position to appreciate the scaling of collective effects. Fig. 9 shows the scaling of the initial horizontal IBS growth rate, computed with the program *ibsEmittance* (Xiao, 2008; Borland *et al.*, 2009*b*) for a 1 nC bunch, exhibiting a N_d^5 dependence. As N_d gets larger, the initial IBS growth rate exceeds the synchrotron radiation damping rate, indicating that the beam will expand in volume until the rates are balanced. The computations are performed for two emittance ratios κ, from which we see that having larger κ results in significant suppression of the IBS growth rate. Not surprisingly, when the equilibrium emittances are computed, the case with larger κ shows significantly less impact from IBS, as shown in Fig. 10. Running with large κ is usually avoided because of problems it can cause with traditional accumulation-based injection, a subject to which we will return later.

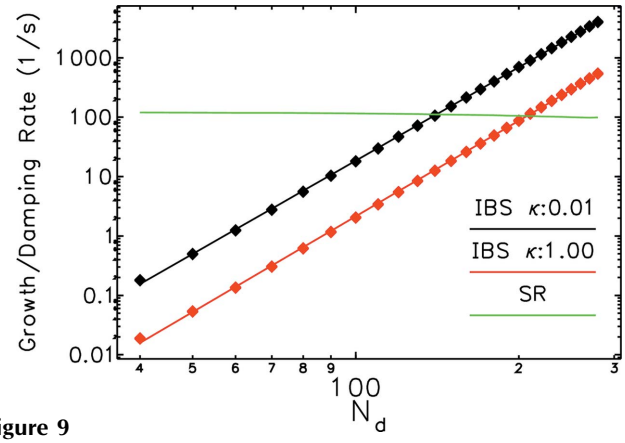

Figure 9
Comparison of initial intrabeam scattering growth rates in the horizontal plane to the synchrotron radiation damping rate, for a 1 nC bunch with $\kappa = 0.01$ and $\kappa = 1.00$. The fit shows a dependence of $\sim N_d^5$.

Another method of mitigating the effects of IBS is to artificially stretch the bunch, which can be accomplished using a higher-harmonic cavity or lower RF frequency. Such a cavity works by reducing the slope of the net RF voltage at the center of the bunch, without significantly impacting the bucket half height. It is practical to attempt to lengthen the bunch by a factor of about three, so we repeated our calculations assuming a fixed initial bunch r.m.s. length of 12 mm (40 ps),

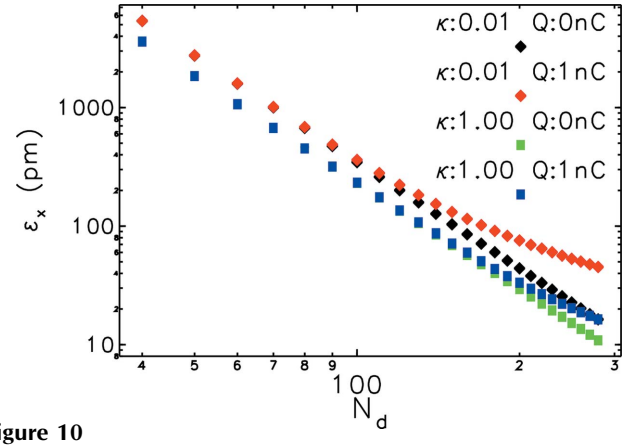

Figure 10
Comparison of equilibrium horizontal emittances in the presence of IBS for a 1 nC bunch with $\kappa = 0.01$ and $\kappa = 1.00$. The results for vanishingly small charge, labeled '0 nC,' show the limit as IBS effects become negligible.

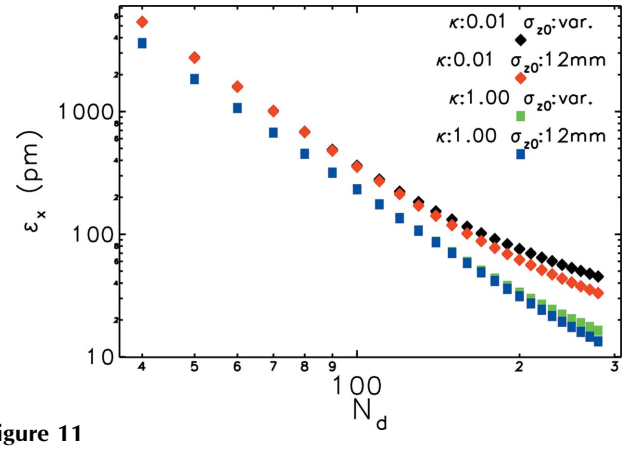

Figure 11
Comparison of equilibrium horizontal emittances in the presence of IBS for a 1 nC bunch with κ = 0.01 and κ = 1.00, for two cases: normally varying bunch length and with σ_z = 12 mm due to use of a bunch-lengthening cavity.

which is approximately three times the minimum bunch length seen in the simulations. The bunch may lengthen further under the influence of IBS-induced energy spread. As Fig. 11 shows, this helps to reduce the effect of IBS, but is not a complete cure by itself.

We noted above that IBS is much worse for lower energy. Using the model ring with N_d = 150, we scanned the beam energy for a 10 nC bunch charge while keeping σ_z = 12 mm and κ = 0.01. As shown in Fig. 12, the IBS prevents benefitting from the $1/\gamma^2$ dependence of the natural emittance on energy, which can only be fully eliminated by running with very low charge bunches. The impact can be mitigated by running with a high emittance ratio κ, but that also impacts brightness by increasing the vertical emittance. Evidently, if we wish to obtain the lowest emittance, we must consider several methods of controlling IBS, namely, use of many low-charge bunches, use of relatively large emittance ratios, use of a bunch-lengthening cavity, and use of reasonably high beam energy.

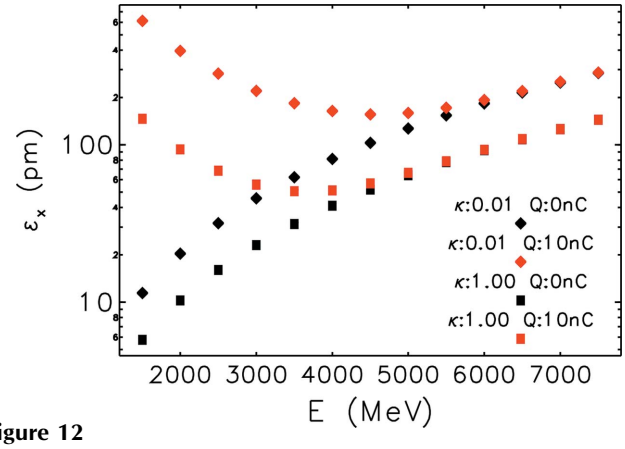

Figure 12
Comparison of equilibrium horizontal emittances in the presence of IBS for a 10 nC bunch with two values of κ, as a function of the beam energy for a ring with N_d = 150. The r.m.s. bunch length is assumed to be constant at σ_z = 12 mm.

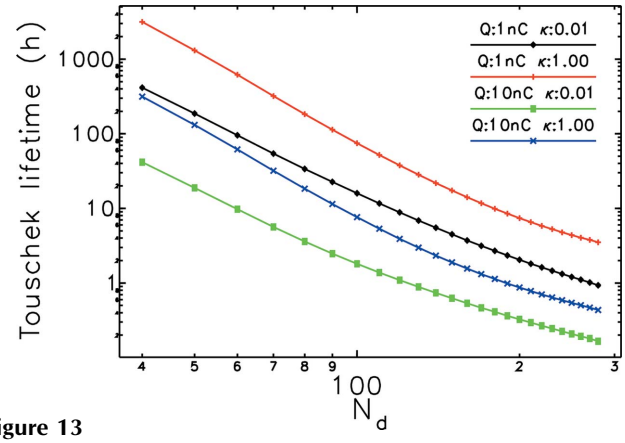

Figure 13
Comparison of Touschek lifetime for 1 and 10 nC bunches with κ = 0.01 and κ = 1.00, assuming the use of a bunch-lengthening cavity to stretch the bunch to σ_z = 12 mm.

We next turn our attention to Touschek lifetime, which we anticipate to also be a significant challenge for ultralow emittances. Let us assume that the momentum aperture can be maintained, through the methods detailed in the next section, at a nominal value of ±3%. We use the beam parameters computed in the presence of IBS with bunch lengthening, since this will enlarge the bunch volume and hence reduce the Touschek scattering rate. The Touschek lifetime was computed with the program *touschekLifetime* (Xiao & Borland, 2007), which is based on Piwinski (1998). As Fig. 13 shows, the Touschek lifetimes for large N_d can be quite short unless we use low-charge bunches and large κ. Another way to view these data, shown in Fig. 14, is to plot Touschek lifetime *versus* $\varepsilon_x\varepsilon_y\sigma_\delta$, which gives a rough gauge of potential X-ray brightness. If we take 1 h as a practical lower limit for the lifetime, we conclude that the best approach is to run with many weak bunches (*e.g.* 1 nC per bunch) and low κ. The graphs somewhat overstate the advantage of very low κ for increasing the X-ray brightness, since we are ignoring the considerable contribution of the intrinsic radiation emittance and energy spread. A more definitive statement requires selecting a

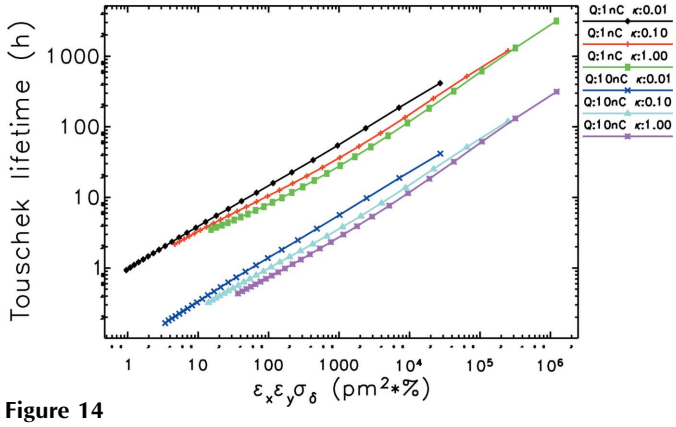

Figure 14
Touschek lifetime *versus* product of the emittances and energy spread for 1 and 10 nC bunches with κ = 0.01 and κ = 1.00, assuming the use of a bunch-lengthening cavity to stretch the bunch to σ_z = 12 mm. Roughly speaking, low values of $\varepsilon_x\varepsilon_y\sigma_\delta$ correspond to higher X-ray brightness.

photon energy, undulator and knowing the beta functions in the ID straight. However, the general conclusion is valid.

As mentioned in §3.8, there are other collective effects that result from the interaction of electrons with the macroscopic field of the bunch or of the entire beam, as modified or trapped by the vacuum chamber and RF cavities that are a necessary part of any accelerator. The thresholds for some of these collective instabilities can be roughly estimated and it is worth exploring how they might be expected to scale. One difficulty is that the vacuum chamber impedance will vary in a non-trivial way as the vacuum chamber radius R is varied. Typical contributions to the longitudinal impedance scale as $1/R$ or $1/R^2$, while typical contributions to the transverse impedance scale as $1/R^3$ to $1/R^4$. Even taking the smallest exponents indicates a serious issue with collective instabilities.

Two important instabilities in modern storage rings are the microwave instability (MWI), which results in inflation of the energy spread, and the transverse mode-coupling instability (TMCI), which limits the single-bunch current. The threshold for MWI is given by the Boussard criterion (Zotter, 2013), which, to avoid unnecessary details, we write as

$$I_{\rm MWI} \propto \frac{\alpha_c \sigma_z \sigma_\delta^2}{|Z_\parallel/n|}, \tag{50}$$

where $|Z_\parallel/n|$ is the broadband longitudinal impedance. The threshold for TMCI can similarly be characterized as (Zotter, 2013)

$$I_{\rm TMCI} \propto \frac{\Omega_s}{\langle\beta_y\rangle |Z_y|}, \tag{51}$$

where $|Z_y|$ is the vertical transverse impedance (which typically dominates).

Figs. 15 and 16 show the scaling of the thresholds under two conditions: the optimistic assumption of constant impedance and the more pessimistic assumption that $|Z_\parallel/n| \simeq 1/R$ and $|Z_y| \simeq 1/R^3$. Clearly, there is a huge variation in the scaling depending on what we assume for the scaling of the impedances. If the impedances do, in fact, scale rapidly with R, then we can expect to encounter serious issues as we move toward multi-bend lattices. Fortunately, there are steps one can take to mitigate the increase in impedance, including making transitions in the chamber smoother, using higher-conductivity material (*e.g.* aluminimum or copper in place of stainless steel), reducing beta functions at the locations with the smallest gap or most rapid variation in gap, and lengthening the bunch. One can also employ better feedback systems (*e.g.* with lower noise and higher power) and use positive chromaticity to stabilize TMCI to some degree. We will touch upon this issue again in §6.

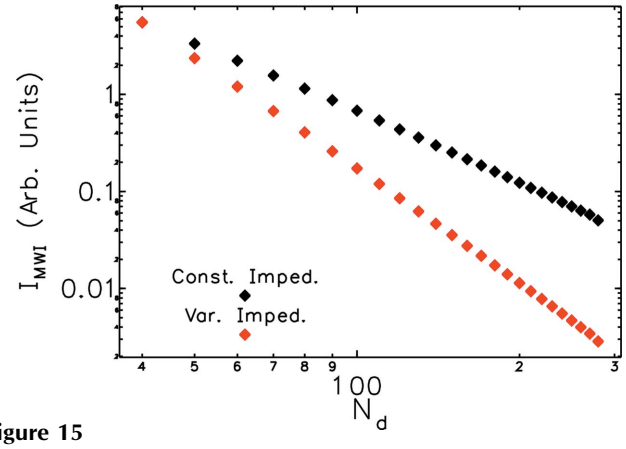

Figure 15
Scaling of microwave instability threshold for the case of constant impedance and impedance that varies like $1/R$, where R is the vacuum chamber radius. In the former case, the threshold scales like $1/N^{2.5}$, whereas in the latter case it scales like $1/N^{3.9}$.

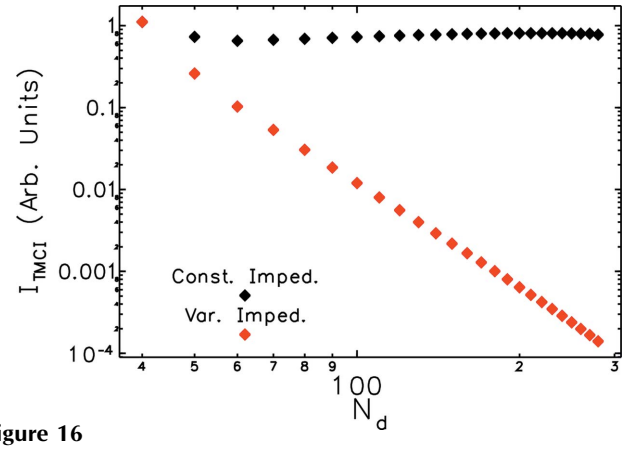

Figure 16
Scaling of transverse mode coupling instability threshold for the case of constant impedance and impedance that varies like $1/R^3$, where R is the vacuum chamber radius. In the former case, the threshold is roughly constant, whereas in the later case it scales like $1/N^{4.4}$.

5. Lattice optimization techniques

We have seen using scaling arguments that nonlinear dynamics is expected to be very challenging for next-generation rings. Because of the complexity of the full equations of motion, we assess the quality of a lattice design using particle 'tracking', propagating simulated particles for many turns in the ring by integrating the equations of motion through the various accelerator components. Particle tracking allows determining, for example, the dynamic and momentum acceptances.

The region in x–y space where motion of particles with initial coordinates (x, y) stays stable is called dynamic aperture or dynamic acceptance (DA). Typically the DA is considered most relevant to beam capture from the injector and is determined by tracking particles starting at the azimuthal location of beam injection. The DA is also important in determining the elastic gas scattering lifetime, as it determines the maximum deflection angle a particle may suffer without being lost from the beam.

The local momentum acceptance (LMA) (Steier *et al.*, 2002; Belgroune *et al.*, 2003) is the maximum negative and positive momentum deviation that is stable as a function of azimuthal coordinate s. Typically, the LMA will be wide at locations where $|\eta_x|$ is small and narrow at locations where $|\eta_x|$ is large. The LMA is relevant to Touschek lifetime calculations, since

it determines the maximum energy kick that a particle may experience without being lost. For a similar reason, it is involved in computations of gas bremsstrahlung lifetime.

In this section, we delve in more detail into the various techniques used for detailed optimization of DA and LMA. We begin with a discussion of the more traditional approach, which is based on perturbation theory and attempts to linearize the motion of particles over as wide a range of transverse amplitudes and momentum deviations as possible. We then describe more recent developments involving direct simulation and optimization of nonlinear dynamics.

5.1. Traditional approach to nonlinear dynamics optimization

The stability of particle motion depends on how well the nonlinear dynamics is optimized and controlled. To analyze particle motion in the presence of sextupoles, a perturbative approach is often used. In this method, the Hamiltonian for particle motion is split into linear and nonlinear parts, and the nonlinear part is assumed to be a small perturbation to the linear motion. Then, the motion can be analyzed analytically, and various expressions for tune shifts with amplitude, resonance driving terms, higher-order chromaticities, *etc.* can be obtained [see Bengtsson (1997) as an example of Hamiltonian dynamics and symplectic maps with normal form].

Simply speaking, to improve the nonlinear behavior of the lattice, one needs to minimize tune shifts with amplitude and momentum, as well as the strength of driving terms for nearby resonances. This can be done by varying the strengths and positions of different sextupole families, adding higher-order multipole magnets (*e.g.* octupoles), changing the fractional betatron tunes, and by varying the lattice functions at locations of nonlinear magnets. In general, there are two types of schemes to improve nonlinear beam dynamics: global cancellation schemes and local cancellation schemes.

A global cancellation scheme employs many repetitive cells with specially chosen phase advance in each cell. When properly constructed, these many repetitive cells, which are designed to have integer total phase advance in both planes, become an achromat to some order, where by 'order' we refer to the highest power of the particle coordinates. First-order achromats have been applied in particle accelerators for a long time. Brown (1979) developed a systematic matrix-based approach to designing second-order achromats. It adopts at least four identical cells with dipoles, quadrupoles and sextupoles and can eliminate all geometric and chromatic aberrations up to second order. After that, similar third-order achromat design approaches were developed analytically and numerically (Dragt, 1987), again using the concept of identical cells and integer phase advance. Wan & Berz (1996) developed a general method with Lie algebra to design achromats to arbitrary order, taking advantage of mirror symmetry and using multipole magnets for each specified order (for example, octupoles for a third-order achromat). The latest PEPX design utilizes a third-order geometric achromat which is achieved with the assistance of harmonic sextupoles (Cai *et al.*, 2012). It can eliminate all the betatron resonance driving terms below fourth order. A quasi-achromat scheme was proposed to achieve a similar performance in a more general case (Sun & Borland, 2013). As observed in Fig. 17, there are far fewer resonance lines showing up in the frequency map (Laskar, 1990) of the quasi-achromat scheme than a nominal scheme which has arbitrary phase advance.

The local cancellation scheme, as the name suggests, works by achieving cancellation of nonlinearities within a cell. One outstanding example is the famous idea proposed by Brown (1979), employing $-I$ transport between non-interleaved sextupoles in several FODO cells to cancel higher-order geometric aberrations. This idea is further developed in the hybrid multibend achromat concept developed at ESRF (Farvacque *et al.*, 2013) where three nearby sextupoles (two families) located in a dispersion bump are each separated by $-I$ from a matching magnet on the mirror side (see §6.1 for further discussion). The higher-order geometric aberrations cancel to a certain degree between the sextupole pairs.

Because approximations or impractical assumptions are often made in development of such schemes, they are rarely sufficient in themselves to ensure good performance. Octupole magnets may be included to directly minimize transverse amplitude-dependent detuning terms, thus avoiding major

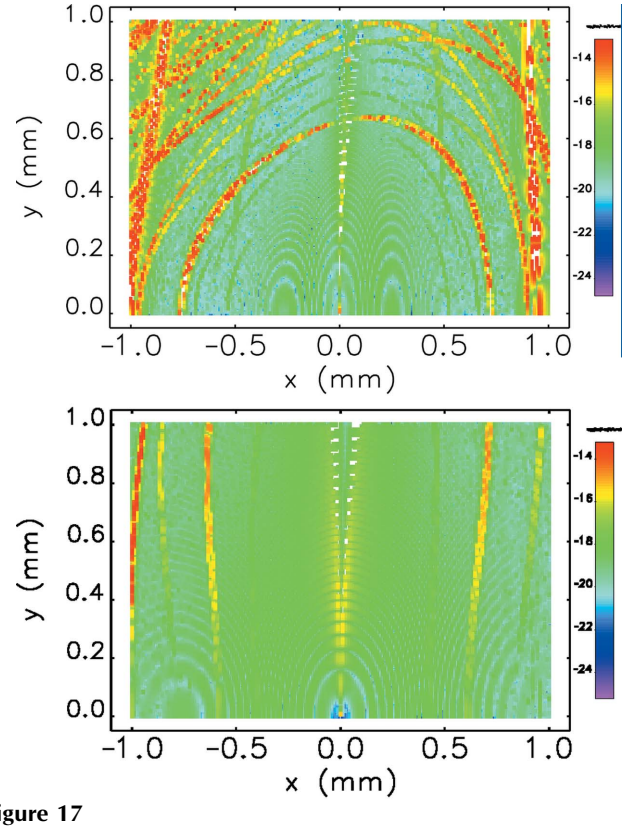

Figure 17
Comparison of frequency map analysis (FMA) shows that the quasi-third-order achromat scheme (bottom) has better performance than a similarly optimized lattice that uses no particular cancellation scheme. The FMA for the quasi-achromat shows far fewer resonance lines, which is generally indicative of a better solution. The frequency maps are evaluated at locations in the two lattices where the beta functions are roughly equivalent.

resonance crossings. More families of sextupoles may also be adopted to minimize higher-order chromaticities or geometrical aberrations, at the cost of reduced periodicity and symmetry. Placement and adjustment of sextupoles and octupoles is guided by analytical expressions that relate the various detuning terms and resonance driving terms to the strength of these magnets and the lattice functions at their locations. Although the expressions have a closed form, they are complex and numerous enough that the optimization must be performed with the aid of a computer program such as *OPA* (Streun, 1999) or *elegant*.

5.2. Direct approach to nonlinear dynamics optimization

In the process of optimizing a lattice as described in the previous section, it is essential to check the performance of the lattice using particle tracking. When properly performed, tracking can include effects to higher order than is possible in pertubative approaches and with full coupling of motion in all three planes. It can also easily incorporate errors, such as misalignments, rotations, strength errors and field nonuniformities, as well as physical apertures and radiation damping. For these reasons, it is more definitive than perturbative approaches, but is also far more demanding in computational resources.

The traditional approach will usually yield good results in a perfect machine, but the solution may fail to perform well in the presence of realistic imperfections or 'errors'. Hence, iteration between tracking and traditional optimization is usually required. In light of this and recognizing the vast improvement in computational capabilities seen in recent years, a number of groups have developed optimization based directly on tracking, without the use of the perturbative approach.

The first such work (Shang & Borland, 2005) known to the authors used a parallel simplex optimizer and the tracking program *elegant* to improve the simulated dynamic acceptance of the APS. In the same year, Bazarov & Sinclair (2005) described the first application of multi-objective genetic algorithms or MOGA (Srinivas & Deb, 1995) to tracking-based accelerator optimization, albeit for a photoinjector rather than a storage ring. [An overview of MOGA is presented below. See also Hofler (2013) and Hofler *et al.* (2013).] This was followed several years later (Borland *et al.*, 2009*c*) by an exploration of several tracking-based techniques to optimize both dynamic and momentum acceptances, and by work of Yang *et al.* (2009) in which MOGA was used to optimize storage-ring linear optics. The former included both a grid scan technique and the use of single-objective genetic optimization, including the first use of a tracking-based optimizer to improve the performance of an operating storage-ring light source. It also included the discovery that breaking the symmetry of the sextupole distribution, traditionally considered unwise or at best pointless, could in fact result in better performance.

This was followed shortly by publicaton of the results of several applications of MOGA to tracking-based optimization of storage rings. Borland *et al.* (2009*a*) demonstrated the use of MOGA to simultanelously and directly optimize both DA and LMA, and later extended this (Borland *et al.*, 2010) to directly optimize Touschek lifetime computed from the LMA. In addition to variation of sextupoles, variation of the linear lattice design was part of the MOGA process. Yang *et al.* (2011) demonstrated the use of MOGA to simultaneously optimize the on-momentum DA plus the off-momentum DA for two particular momentum deviations, which is a proxy for the LMA. This work used a combination of tracking results and low-order resonant driving terms, providing faster convergence. Gao *et al.* (2011) was a similar effort, but without inclusion of off-momentum analysis. Also of interest is Sun *et al.* (2012), who included minimization of the diffusion rate (a measure of resonance strengths) within the DA as one optimization goal.

MOGA, as the name suggests, relies on two concepts: multi-objective comparison and genetic, or evolutionary, improvement of solutions. Imagine that one needs to adjust an accelerator or other system to minimize the values of n goal functions $g_i(\mathbf{s})$, where $i = 1 \ldots n$ and $\mathbf{s}$ represents the collection of accelerator settings available for adjustment. For example, g_0 might be related to the DA and g_1 might be related to the Touschek lifetime, while $\mathbf{s}$ could be the sextupole strengths. A common and often effective approach is to form a new goal function that is a weighted sum of the individual goals, *e.g.*

$$G = \sum_{i=1}^{n} w_i g_i(\mathbf{s}). \tag{52}$$

However, this involves deciding ahead of time on the relative importance of the different goals, which may be difficult when there are many goals. It may also unnecessarily restrict the search space in the beginning stages, resulting in a failure to find better solutions. Multi-objective comparison sidesteps this problem by use of non-dominated sorting (Deb *et al.*, 2002), which allows ranking solutions based on multiple criteria. The first-rank solutions are those solutions for which no other solution exists that is superior in all objectives. Depending on the relative importance of the different objectives, any one of these solutions might be considered the 'best solution'. Once the first-rank solutions are known, they can be set aside and the analysis repeated to find the second-rank solutions, and so on until all solutions have been ranked.

Once solutions are ranked, the next step is to determine how to generate new trial values of the variables. Many methods exist, but the most common approach appears to be an 'elitist' method that selects a subset of the top-ranked prior solutions, then 'breeds' them to create new trial values. The method used at APS involves selecting the top few ranks, then breeding new trial configurations from pairs of these solutions. The breeding algorithm typically used blends each of the attributes of the two parents using a probability distribution that is roughly uniformly distributed between the values from the parents, with Gaussian tails that extend outside the interval defined by the parents.

It is common for MOGA to be performed in discrete generations, where all trials from a previous generation must

complete before new trial configurations are generated. However, this can leave computational resources idle, particularly when there is a spread in the length of time required to evaluate the configurations. This is not uncommon when performing searches for dynamic or momentum acceptance. Hence, the APS algorithm starts new trials continuously based on the highest-ranked solutions available at any time.

We noted above that one advantage of tracking is that it can include effects that are not readily included in pertubative methods. This provides an important advantage during optimization, since it results in more robust solutions. Toward this end, the optimization method employed at APS includes errors, radiation damping, RF cavities and physical apertures.

Errors are included because these break the symmetry and drive resonances. Optimizing in the presence of representative errors yields a more robust result, avoiding the collapse of the acceptances that can sometimes be experienced when errors are added to solutions optimized for a perfect machine. The most important errors in storage rings are residual orbit, misalignment and magnet strength errors. Similar effects can be introduced when insertion devices are included, particularly in low-energy storage rings. While one would, ideally, simulate all of these in a realistic way, to make this truly realistic involves simulation of correction schemes that are not necessarily simple to automate in a reliable fashion. As a result, we use a simplified set of errors intended to emulate the conditions that are typically achieved after correction. This involves adding only strength errors and roll errors to quadrupoles and sextupoles. This produces 'beating' (*i.e.* non-periodic variation) in the beta functions and dispersion, such as would be observed due to horizontal orbit error in sextupoles and strength error in quadrupoles. It also produces cross-plane coupling, such as would be observed due to vertical orbit error in sextupoles and roll errors in quadrupoles.

Including radiation damping is important for several reasons. First, radiation damping can stabilize particles that might otherwise be lost. Second, radiation damping causes particle oscillation amplitudes in both the transverse and longitudinal planes to vary, in the course of which the tunes will vary. This can result in resonance crossings that affect particle stability. For both reasons, excluding radiation damping can potentially give misleading results.

Another reason to include radiation damping is that it permits one to include RF cavities in a more realistic way, *i.e.* with realistic voltage and synchronous phase. This in turn results in an accurate simulation of synchrotron oscillations, which modulate the tune (*via* the chromaticities) and sweep the particle across resonances.

6. Possible APS upgrade lattice

Using the techniques described in the previous section, we have developed a seven-bend achromat lattice that could be used as a possible upgrade for the APS storage ring. The lattice is based on a particular MBA concept developed at the European Synchrotron Radiation Facility (ESRF) and

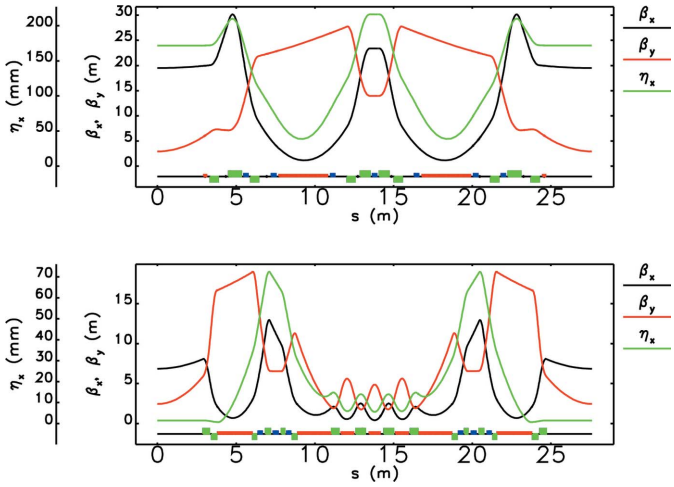

Figure 18
Comparison of lattice functions for present-day APS (top) and a possible HMBA replacement lattice (bottom). Red blocks represent dipole or bending magnets, green blocks represent quadrupole magnets, and blue blocks represent sextupole magnets.

is known as a 'hybrid' multibend achromat (Farvacque *et al.*, 2013) or HMBA lattice. Fig. 18 compares the lattice functions for the present-day APS with the lattice functions for a possible HMBA design. We can clearly see that the dispersion function is much smaller in the HMBA lattice compared with the existing lattice. In addition, the average beta functions are much smaller, both of which indicate the greater average strength of the focusing elements. These factors combine to reduce the magnitude of $\mathcal{H}$ [see equation (20)] and thus the natural emittance. We fully expect that such a lattice, if workable, would provide significant increases in X-ray beam quality. However, from the discussion on ring scaling, we also expect to encounter beam dynamics challenges. After a more detailed review of the concept and constraints on its implementation at APS, we will describe how our particular lattice was optimized and how it would perform if built.

Significant and interesting work along similar lines is ongoing at a number of other facilities world-wide. Among these are possible upgrades of ALS (Tarawneh *et al.*, 2013), Diamond (Bartolini *et al.*, 2013), ESRF (Farvacque *et al.*, 2013), SOLEIL (Nagaoka *et al.*, 2013) and SPring-8, as well as a possible new MBA-based ring in Beijing (Wang, 2013).

6.1. Lattice concept

An interesting feature of the HMBA lattice is that the dispersion η_x is not uniformly small, but has two symmetric bumps, one between the first and second dipoles, and another between the sixth and seventh dipoles. The sextupole magnets are located within these bumps, since the relatively large value of η_x allows the sextupoles to be more effective and thus weaker. This in turn reduces the strength of geometric aberrations introduced by the sextupoles.

Another noteworthy feature of the lattice is that the sextupoles are arranged in such a way that their leading geometric effects cancel within one sector. This is because the

optics between the two dispersion bumps in any sector are such as to give betatron phase advance of $\Delta\varphi_x = 3\pi$ and $\Delta\varphi_y = \pi$ between corresponding sextupoles (*i.e.* between the first and fourth, second and fifth, and third and sixth) in any sector. We can easily see how this works using the impulse approximation for two thin sextupoles with integrated strengths $m = K_2 L$. Consider an on-momentum ($\delta = 0$) particle entering the first sextupole with coordinates $\mathbf{x}_0 = (x_0, x_0', y_0, y_0')$. At the exit of the first sextupole, the coordinates become

$$\mathbf{x}_1 = \begin{cases} x_1 &= x_0, \\ x_1' &= x_0' - (1/2)m(x_0^2 - y_0^2), \\ y_1 &= y_0, \\ y_1' &= y_0' + m x_0 y_0. \end{cases} \tag{53}$$

Because of the phase advance $\Delta\varphi_x = 3\pi$ and $\Delta\varphi_y = \pi$, the coordinates at the entrance to the second sextupole are simply inverted, *i.e.* $\mathbf{x}_2 = -\mathbf{x}_1$. Since both x and y have changed sign, the changes in x' and y' from the second sextupole are identical to those from the first sextupole. Hence, the coordinates at the exit of the second sextupole are $-\mathbf{x}_0$, *i.e.* we have the very desirable result that the sextupoles have no effect for an on-momentum particle. This conclusion is only exactly true when the sextupoles can be treated in the impulse approximation and when the sextupoles are not interleaved (*i.e.* only two sextupoles per sector). Although neither of these conditions is actually met in this lattice, the configuration still provides a very good starting point for further optimization.

Another noteworthy feature of the ESRF concept is the use of longitudinal gradients in the first, second, sixth and seventh dipoles. The first (seventh) dipole has most of the bending concentrated at the entrance (exit), which helps to kick the dispersion up to larger values, thus making the sextupoles weaker. The central section of the lattice comprises three additional dipoles with transverse gradients. This part of the lattice is very similar to the cells for the MAX IV design (Leemann *et al.*, 2009) and those used for our model lattice in the scaling study presented in §4. The transverse gradient dipoles not only serve to make the lattice more compact, they also increase J_x and thus decrease the emittance [see equation (19)]. As we will show below, the gradients in these dipoles and the surrounding quadrupoles are very strong, helping to reduce $\mathcal{H}$ and providing the necessary phase advance for cancellation of geometric sextupole kicks.

6.2. Constraints

Because we are contemplating upgrading an existing ring, we have several geometrical constraints that should be satisfied to minimize other required changes. Of course, we must keep 40 sectors as now, with 35 of these dedicated to insertion device beamlines supporting 4.8 m-long IDs. The ID straight sections would ideally be exactly where they are now, which is a difficult constraint to satisfy without a significant change in circumference. Instead, we require them to move transversely by no more than 70 mm, in order that the X-ray beam pipe will fit within the existing shield-wall penetrations (Ramanathan, 2003). It is also desirable that the circumference of the new ring closely matches that of the existing ring, in order to avoid issues with the booster synchrotron in the existing injector, which is locked in RF frequency to the ring. At issue is that changing the common RF frequency will result in running off-momentum in the booster, according to [compare equation (31)]

$$(\Delta\delta)_{\rm booster} = -\frac{1}{\alpha_{\rm c,booster}}\left(\frac{\Delta f}{f}\right)_{\rm RF}, \tag{54}$$

where $\alpha_{\rm c,booster} = 7.1 \times 10^{-3}$ is the momentum compaction factor of the booster and $(\Delta f/f)_{\rm RF}$ is the fractional change in the common RF frequency. The change in circumference should ideally be less than 18 p.p.m. in order to avoid an energy offset of more than 0.25%.

In addition to ID beamlines, APS operates a number of bending magnet beamlines, which impose additional requirements. In order to reduce the difficulty of building the accelerator magnets, it is helpful to reduce the beam energy E from the present value of 7 GeV to 6 GeV. However, the critical photon energy $e_{\rm c}$ from a dipole scales like E^3B, where B is the bending field. This implies a significant reduction in $e_{\rm c}$ unless B is made 60% larger, which drives the emittance higher and makes the transverse gradient dipoles much more difficult to design. Following ESRF, we elected instead to make space for a 15 cm-long three-pole wiggler downstream of the middle dipole in each sector. The only requirement is then to provide sufficient space and ensure that the radiation source point has the same angle and is within 70 mm of the existing transverse position.

Another choice to be made in optimizing the lattice is the assumed injection method (see §6.5), *i.e.* whether to assume on-axis swap-out injection (Emery & Borland, 2003) or traditional accumulation-based injection. This choice significantly impacts the optimization because traditional injection requires five to ten times larger dynamic acceptance. Although studies are on-going, swap-out appears to provide a significant advantage in terms of emittance and also relaxes magnet field quality requirements, so here we present results for swap-out only and reserve a detailed comparison for a future publication.

APS presently operates in one of two timing fill patterns about 80% of the time. The prevalent pattern has 24 equispaced 16 nC bunches, while the 'camshaft' pattern has an isolated 59 nC bunch, with many small bunches crowded on the opposite side of the ring. It is highly desirable to maintain a strong timing capability in any upgrade, which entails maintaining a high single-bunch stability limit. For present purposes, we have targeted 16 nC per bunch as in the present 24-bunch mode, but with a total of 48 bunches in order to deliver twice the average current (200 mA instead of 100 mA). Detailed simulation of impedance and instabilities (Chae, 2014) indicate that this is possible provided the chromaticity in both planes is +5, so we have targeted this value. (The value should be less when bunch-by-bunch feedback is included, which is the subject of on-going analysis.)

6.3. Lattice optimization

Nominally, the sextupoles in the HMBA lattice are powered in just two families, with symmetry about the center of the sector and about the center of the dispersion bump. More explicitly, within each dispersion bump we have an SD–SF–SD arrangement. Although we will have to break this symmetry in order to compensate for the thickness and interleaving of the sextupoles, it is the natural starting point. Using this configuration and the assumption that $\xi_x = \xi_y = 2$ is sufficient (see below), we performed wide-ranging scans of working point, as described by Borland *et al.* (2013). This led us to use integer tunes of $\nu_x = 95$ and $\nu_y = 36$ as our starting point, which we have retained even though chromaticity requirements have increased to $\xi_x = \xi_y = 5$.

As described above, the MOGA process simultaneously runs many 'function evaluations' with different values for the control variables, including the strength of the sextupoles. Based on a suggestion from P. Raimondi of ESRF, we elected to split the sextupoles into 12 families, with odd sectors having the first six families and even sectors the other six. This provides a great deal of freedom in adjusting the nonlinear dynamics effects of the sextupoles.

Other control variables include goals for the horizontal and vertical phase advance between sextupoles, the fractional tunes, the beta functions at the IDs, the maximum beta functions and the maximum dispersion. These are described as goals since it is not guaranteed that an acceptable linear optics solution will exist for any combination of values. Instead, as part of each function evaluation, a lattice-matching step is performed that attempts to satisfy these goals, along with others such as obtaining an emittance below a fixed target, obtaining sufficiently small dispersion in the straight section, satisfying geometrical constraints, providing sufficient space between magnets for diagnostics and other equipment, and providing sufficient margin between magnet engineering designs and the fields used in the lattice. Although it varied with different stages of the optimization, at times we allowed the matching stage to vary essentially every property of the lattice, including the relative strength and length of all magnets and dipole segments, as well as the distances between magnets. As the design developed, we reduced the number of matching variables, typically by removing the ability to vary lengths and angles, thus fixing the geometry. Because of the complexity of the matching problems and in order to make good use of parallel resources allocated to each function evaluation, we used the parallel hybrid simplex optimizer (Wang *et al.*, 2011) in the parallel version of *elegant* (Wang *et al.*, 2009).

The next stage of function evaluaton, following matching, was tracking with parallel *elegant* to obtain the dynamic and local momentum acceptances. Symplectic integration was used for all magnetic elements, with drift spaces treated as a second-order matrix. Physical apertures were included, consisting of a 10 mm-radius round aperture in the arcs and a 20 mm by 6 mm ellipse in the IDs. Lumped radiation damping was included along with a thin-lens RF cavity set for ±4% bucket height. A common set of strength and roll errors was used for these computations, sufficient to give 2–4% beta function beats. For quadrupole and sextupole roll errors, we used a rather large r.m.s. value of 1 mrad in order to provide some assurance that coupling resonances were driven.

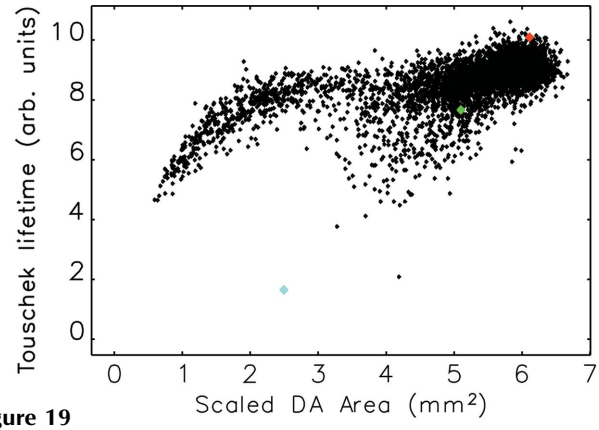

Figure 19
Example of MOGA optimization results, showing the scaled DA area and the Touschek lifetime (in arbitrary units) for several thousand evaluations across two MOGA runs. The cyan dot shows the result for symmetric sextupoles, for comparison. The green dot is the starting point, obtained from a previous MOGA run, while the red dot is the 'best' configuration.

Because of the iterative nature of the work, including repeated iteration between the physics design and engineering, it is not possible to show a pristine example of an optimization from a symmetric starting point to a final lattice. However, Fig. 19 illustrates the character of the results, obtained in this case by starting from a previous MOGA solution. Also shown is the performance with symmetric sextupoles, which is presumably poor in this case because of the large chromaticity, as better results were obtained for symmetric sextupoles with a chromaticity of +2.

Table 1 summarizes properties of the best solution obtained to date, the lattice functions for which are shown in Fig. 18. The effective emittance is under 66 pm, a 47-fold reduction from the effective emittance of APS today. (Naive application of $\varepsilon_0 \propto E^2/N_\mathrm{d}^3$ would have led us to expect a 58-fold reduction.) The circumference is matched to within 6 p.p.m., with a 42 mm offset of the IDs and a 23 mm offset of the BM lines.

The magnet strengths are quite high, with quadrupole gradients of 47 to 80 T m^{-1} and maximum transverse gradients in dipoles of 46–47 T m^{-1}. We anticipate building dipoles as transversely offset quadrupole magnets, in order to obtain the desired strength and good field quality. The SD sextupoles have strengths $|\mathrm{d}B_y^2/\mathrm{d}x^2|$ between 2.8 and 4.7 kT m^{-2}, while the SF sextupoles have strengths between 4.1 and 5.1 kT m^{-2}. Even assuming a pole tip radius of 13 mm, the use of vanadium permendur pole tips is required (Jaski, 2014) for all sextupole magnets and the strongest quadrupole magnets (specifically, those situated between the second and sixth dipole). The variation in the strengths among sextupoles that are nominally in equivalent locations is quite large, both within a sector and between the odd and even sectors.

As expected, the variation in field in the longitudinal gradient dipoles is largest for the first and seventh dipoles in

Table 1
Lattice parameters for a possible APS hybrid seven-bend achromat upgrade.

Tunes and chromaticities		
ν_x	95.107	
ν_y	36.144	
Natural ξ_x	−137.10	
Natural ξ_y	−109.88	
Lattice functions		
Average β_x	4.22	m
Average β_y	7.93	m
Average η_x	0.028	m
Radiation-integral-related quantities at 6 GeV		
Natural emittance	65.47	pm
Energy spread	0.095	%
Horizontal damping time	12.16	ms
Vertical damping time	19.75	ms
Longitudinal damping time	14.35	ms
Energy loss per turn	2.24	MeV
ID straight sections		
β_x	6.85	m
β_y	2.44	m
η_x	1.45	mm
$\varepsilon_{x,\mathrm{eff}}$	65.6	pm
Miscellaneous parameters		
Momentum compaction	5.86×10^{-5}	
$\Delta f_{\mathrm{RF}}/f_{\mathrm{RF}}$	5.54×10^{-6}	
Damping partition J_x	1.62	
Damping partition J_δ	1.38	

each sector, where the ratio is less than five-to-one, with a modest maximum field of 0.63 T at the upstream (downstream) end of the first (seventh) dipole. The ratio is less than three-to-one within the second and sixth dipoles, with a maximum field of only 0.35 T.

6.4. Lattice evaluation

Although the MOGA process described above includes optical errors as part of the modeling, it is essential to evaluate the solution with a number of error ensembles in order to verify robustness. This not only includes many instances of optical errors of the type introduced during optimization but also allows addition of other types of errors and exploration of their effects.

Because of the strong quadrupoles and sextupoles, the orbit correction and lattice correction procedure is somewhat involved and will be the subject of a future publication. For the present, the generation of optical error ensembles is similar to what was described in §5.2. That is, rather than simulate the process of orbit correction and lattice correction, we generate many (*e.g.* 12000) error ensembles, perform a simple tune correction by changing the two quadrupole families in the ID straights, then select only those configurations that exhibit characteristics consistent with more thorough correction. In generating the initial ensembles, we include transverse gradient errors as well as roll errors on quadrupoles and sextupoles. We computed the average vertical-to-horizontal emittance ratio at the IDs for each ensemble, then selected only those ensembles with values between 9 and 11%. We have chosen ensembles that exhibit r.m.s. beta beats between 5 and 8%, which left 117 of the original 12000 ensembles. The selected ensembles are intended to represent the results of reasonably thorough but not final correction of the lattice, which motivates selecting relatively large but not extreme values for the emittance ratio and beta beats.

For each ensemble, we computed the DA and LMA by 500-turn tracking with lumped radiation damping and an RF cavity set for a ±4% bucket. These simulations included systematic and random multipole errors in the dipoles, quadrupoles and sextupoles. Lacking completed magnet designs, we made some simplifying assumptions about the character of the systematic errors. For dipoles, we included sextupole through 18-pole (the highest multipole *elegant* can include in a dipole). For quadrupoles, we included allowed multipoles, in particular, the 12-, 20-, 28- and 36-pole. For sextupoles, we also included allowed multipoles, in particular, the 18-, 30- and 42-pole. Again for simplicity and in the interest of being conservative, each systematic multipole error was given the same magnitude as a fraction of the main harmonic at a reference radius of $r = 10$ mm.

Random multipole errors were added to quadrupoles and sextupoles based on Halbach (1969), which connects the r.m.s. value of such errors to r.m.s. values of construction errors. The construction errors included in this treatment, as coded in the program *sddsrandmult* (Borland *et al.*, 2003), are uncorrelated displacement of the poles, errors in the top/bottom and right/left symmetry, errors in the radius, and relative rotation of the top and bottom. We assumed 30 μm r.m.s. for each positional error and 150 μrad r.m.s. (30 μm over 20 cm) for the rotational error. Halbach's theory provides values for the random quadrupole through 20-pole. We elected to suppress the low-order contributions (quadrupole and sextupole) on the grounds that these are either already included in the generation of the optical errors, or else subject to correction using magnetic measurements. If, for example, we included the quadrupole errors, it would defeat the previously performed tune correction, increase the level of lattice function beating, and also inflate the vertical emittance.

Fig. 20 shows the DA for three different levels of systematic errors. We see that, for 0.03% and 0.1% systematic multipole errors (per component at 10 mm), the tenth percentile contours of the DA are nearly identical. On this basis, there is no reason to require errors below 0.1%. Fig. 21 shows a more complex picture, in that the LMA is smallest when the systematic error levels are smallest. This indicates that some component of the systematic errors is helpful to the dynamics, though exactly which is yet to be understood. Note that looking at the median LMA shows no such effect, indicating that the benefit accrues only for the worst-performing ensembles. It will be important to revisit this phenomenon once detailed magnet designs are available and detailed lattice correction has been implemented.

As discussed above, intrabeam scattering (IBS) can have a significant negative effect on emittance and energy spread, which can be mitigated by bunch lengthening. The zero-current r.m.s. bunch length is about 4 mm, but this can be

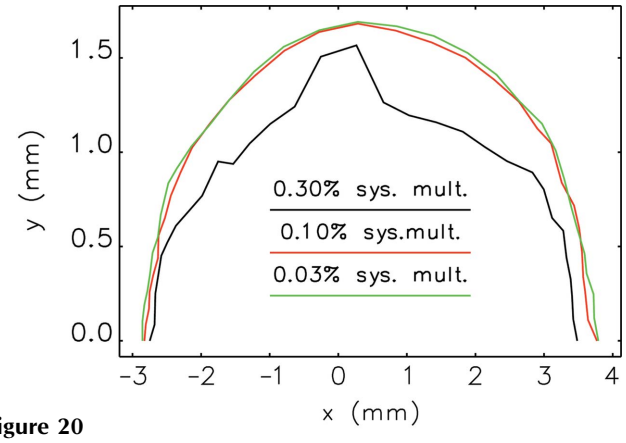

Figure 20
Tenth-percentile dynamic acceptance contours from ensemble evaluation of 117 error ensembles with random and systematic multipole errors. The level of systematic multipoles was varied between 0.03% and 0.3% of the main harmonic at a radius of 10 mm, for each multipole.

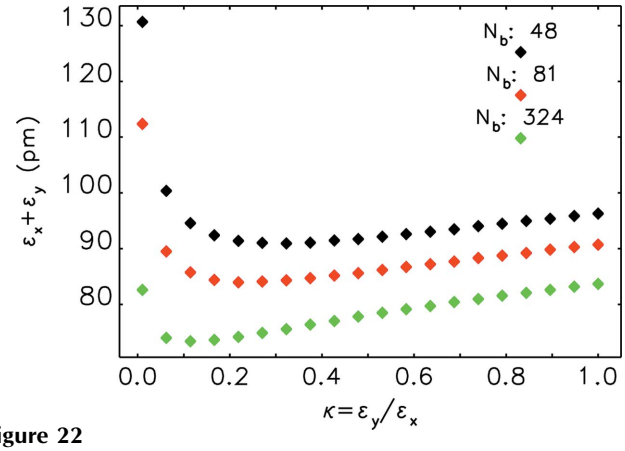

Figure 22
Sum of the horizontal and vertical emittances including IBS as a function of the emittance ratio κ, for 200 mA stored in various numbers of bunches, assuming a 15 mm initial r.m.s. bunch length.

lengthened considerably by use of a third-harmonic cavity. To make a conservative estimate of the required voltage, we assumed a (low) longitudinal impedance of $|Z/n| = 0.1\ \Omega$ and solved the Haissinski equation (Haissinski, 1973) using the program *haissinski* (Emery & Borland, 2000) for a 200 mA total beam current divided into various numbers of bunches. This allowed determining the harmonic cavity voltage required to lengthen the bunch to 15 mm r.m.s., which turns out be less than 1.05 MV. Since this seems practical, we then assume a 15 mm r.m.s. bunch length as the starting point for subsequent calculations. More detailed simulations (Chae, 2014) indicate that attempting to lengthen the bunch further may result in bifurcation of the bunches.

We next used *ibsEmittance* to compute IBS effects as a function of the number of bunches and the emittance ratio κ. One way to qualitatively assess the importance of IBS is to look at $\varepsilon_x + \varepsilon_y$ as a function of κ, since we naively expect this to be constant when IBS is unimportant. As shown in Fig. 22, this is nearly the case when $\kappa > 0.05$, in that the variations are under 10%. Fig. 23 shows the effect on the energy spread. For both emittance and energy spread, the effects for are only significant for low κ in few-bunch modes of operation. The slight increase in the emittance sum for large κ results from the inequality of the damping times in the horizontal and vertical planes.

Combining the IBS results with the LMA results allows us to compute the Touschek lifetime for each of the error ensembles as a function of κ and the number of bunches, which we did using the program *touschekLifetime*. These results are then analyzed statistically to determine the likely Touschek lifetime, as shown in Fig. 24. For 324 bunches, the Touschek lifetime is long even for low κ, but, for 48 and 81 bunches, increasing κ to at least 0.5 seems advisable.

Because of the limited DA and LMA, we may expect the gas scattering lifetime to be shorter than is common in today's storage rings. We have looked at the two principal components of gas scattering (Le Duff, 1985). The first of these is inelastic gas scattering, in which a particle receives an angular kick in a collision with a gas molecule and is lost if as a result it is outside the DA. The second is bremsstrahlung, in which a

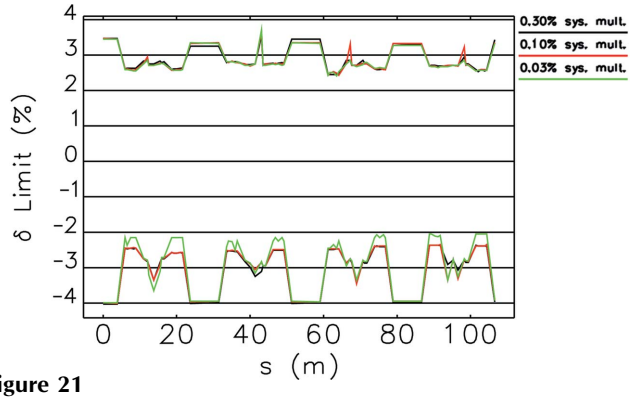

Figure 21
Tenth-percentile local momentum acceptance contours from ensemble evaluation of 117 error ensembles with random and systematic multipole errors. The level of systematic multipoles was varied between 0.03% and 0.3% of the main harmonic at a radius of 10 mm, for each multipole. In order to conserve computing resources, computations were peformed for the first four sectors only. Tests indicate that this is sufficient to characterize performance.

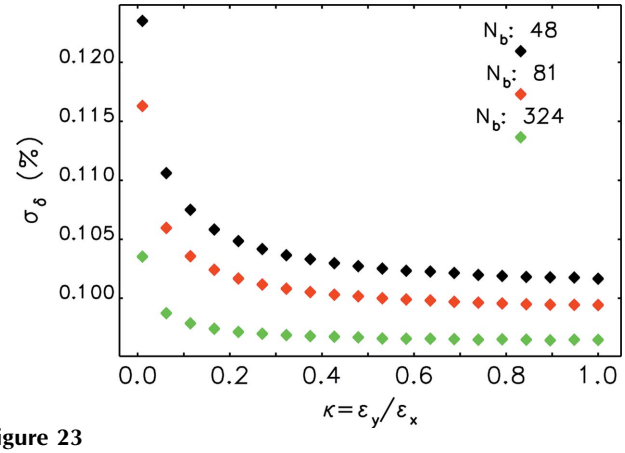

Figure 23
RMS fractional energy spread including IBS as a function of the emittance ratio κ, for 200 mA stored in various numbers of bunches, assuming a 15 mm initial r.m.s. bunch length.

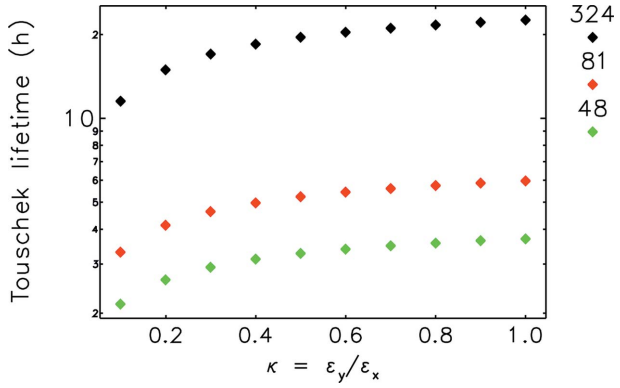

Figure 24
Tenth-percentile Touschek lifetime as a function of the emittance ratio κ, for 200 mA stored in various numbers of bunches, assuming a 15 mm initial r.m.s. bunch length and including IBS inflation of the emittances and energy spread.

particle emits a photon in a collision with a gas molecule and is lost if as a result it is outside the LMA. The usual method of computing the lifetime contributions from gas scattering relies on average values of the lattice functions and the momentum aperture. In calculating these lifetimes, we employed a more accurate method that uses the DA and LMA results from tracking (Borland, 2014).

Assuming 2 nT average pressure with the same fractional pressures (66% H_2, 2% H_2O, 7% CH_4, 21% CO and 4% CO_2) as in the APS today, we computed the gas scattering lifetime for each of the ensembles. The elastic gas scattering lifetime for the 0.1% systematic multipole error level is in excess of 140 h for all ensembles, while the corresponding gas bremsstrahlung lifetime is in excess of 89 h, giving a total gas scattering lifetime of greater than 54 h.

We are now in a position to compute the total lifetime and the injection interval. Because of the small DA, on-axis swap-out injection is the method of choice. For a beam with N_b bunches and an injection interval ΔT_i, the fractional droop D in the population of any bunch is

$$D = \Delta T_i N_b/\tau, \tag{55}$$

where we assume that the total lifetime $\tau \gg N_b \Delta T_i$. We assume $D = 0.1$, *i.e.* that a bunch must be replaced when it has decayed to 90% of its original intensity. Note that the droop in the total beam current is D/N_b, a much smaller value.

Because of the existing (very conservative) shielding calculations, we wish to maintain when possible a lifetime τ of greater than 5.7 h at 200 mA. For any particular N_b, this determines the minimum value of κ using the data in Fig. 24 combined with the gas scattering lifetime, and thus the injection interval. Table 2 lists the total lifetimes and injection intervals for various fill options. The 48 bunch mode will require a new assessment of the shielding. While the injection intervals and average injector current requirement should not cause difficulties for accelerator operation with the existing injector, the high charge per bunch in some modes will require upgrades to the injection systems (Yao *et al.*, 2013).

Table 2
Beam parameters for several bunch patterns including IBS.

The tenth-percentile lifetime includes contributions from Touschek, elastic nuclear and brehmsstrahlung scattering.

N_b	Q_b (nC)	κ	ε_x (pm)	ε_y (pm)	σ_δ (10^{-4})	τ_{10th} (h)	ΔT_{inj} (s)
48	15.33	1.00	48.16	48.16	10.2	3.47	26.0
81	9.09	1.00	45.36	45.36	9.94	5.39	23.9
162	4.54	0.12	70.25	8.08	10.0	5.70	12.7
216	3.41	0.10	69.18	6.87	9.93	6.91	11.5
324	2.27	0.10	66.92	6.65	9.81	9.55	10.6

6.5. Injection scheme and performance

The conventional, off-axis, accumulation-based injection scheme usually requires a dynamic aperture of $\sim$10 mm, which, as indicated in §4 and §6.3, is extremely difficult if not impossible to achieve for the very low emittance storage rings discussed in this paper. If the lattice is driven to sufficiently low emittance, we expect that on-axis 'swap-out' injection (Emery & Borland, 2003), in which the weakest bunch is kicked out and replaced with a fresh high charge injector bunch, is the only choice. Hence, we present that scheme for the APS upgrade lattice described above.

Another consequence of a small dynamic aperture is that perturbations to the stored beam from the injection elements must be kept at a low level. With the close bunch spacing that results from the adverse scaling of beam instabilities, IBS and Touschek lifetime, this means that the injection kickers have to have very fast rise and fall times. As a result, stripline kickers are the best choice for beam injection.

The kick angle from a stripline of length L with gap d between the electrodes is given by

$$\Delta\theta = 2g\,\frac{eV}{E}\frac{L}{d}, \tag{56}$$

where V is the voltage different between the two electrodes, E is the beam energy, $g \leq 1$ is a geometrical factor, and the factor 2 includes the force from both electric and magnetic fields. The timing requirements on the stripline kicker pulser are determined by the length of the stripline (L) and bunch spacing. The desired 324-bunch fill pattern with its 11.3 ns bunch spacing puts the tightest requirement on the timing. The timing requirements can be understood with reference to Fig. 25, which shows a sketch of the stripline pulser waveform and related stored and injected (extracted) bunches.

In designing the injection system, we must ensure that the incoming beam does not suffer significant losses, either as a result of hitting injection components (*e.g.* the septum blade or kicker electrodes) or as a result of being outside the DA. In general, we may choose to inject in the horizontal or vertical plane. Letting q stand for x (horizontal injection) or y (vertical injection), the required minimum beam separation Δq_{min} at the septum is given by

$$\Delta q_{min} = q_{inj} + q_{err} + \Delta q_a + \Delta q_d, \tag{57}$$

where q_{inj} is the injected beam size (normally $3\sigma_{inj}$), q_{err} is the allowed error from all sources (for example, trajectory jitter),

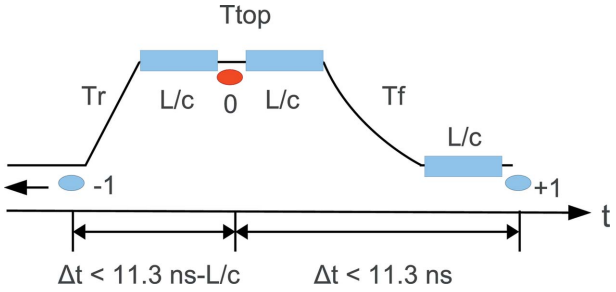

Figure 25
Timing relation among stripline kicker, stored and injected bunches. Legend: '−1': leading stored bunch; '0': injection bunch; '+1': following stored bunch; 'L': stripline length; 'Tr', 'Tf', 'Ttop': rise, fall and flattop of the pulser waveform.

Δq_a is the minimum septum aperture, and Δq_d is the septum thickness.

The beam separation has to be provided by a kicker element (stripline in our case), and is given by

$$\Delta q = (\beta_k \beta_s)^{1/2} \sin \Delta\varphi\, \theta_k, \tag{58}$$

where $\beta_{k,s}$ are beta functions at the kicker and septum location, respectively, $\Delta\varphi$ is the phase advance between the kicker and septum, and θ_k is the kick strength. One common choice is to design a lattice such that $\Delta\varphi \simeq \pi/2$, with β_s and β_k enlarged to further amplify the effect of the kick. This typically entails putting the kickers and septum in two different straight sections, but at a minimum requires several quadrupoles separating the kickers. In the simplest case, where the kickers and septum are all located in the same straight section, equation (58) simplifies to

$$\Delta q = D\theta_k \tag{59}$$

where D is the drift distance between the effective center of the kicker and the septum.

The choice between distributed injection [equation (58)] and a single sector injection [equation (59)] depends in part on a comparison of $(\beta_k\beta_s)^{1/2} \sin\Delta\varphi$ and D, which depends on the actual optical design. Even though $(\beta_k\beta_s)^{1/2} \sin\Delta\varphi$ can be made larger through a special optical design in the injection region, while D is more or less limited by the available space (the same as that preserved for the insertion devices), there are many disadvantages to using distributed injection:

(i) A special injection section would break the high symmetry of the lattice, which makes the already challenging nonlinear dynamic issues worse.

(ii) The beam separation achieved depends strongly on the optics design, which puts constraints on the optics optimization and future operational adjustment, giving less flexibility.

(iii) Large β_k implies a large beam size inside the stripline kicker, which conflicts with the desire to make a narrow-gap high-field device. Thus, the advantage of large β_k may be limited.

(iv) Larger β_k and β_s means the impedance contributions from the kicker and septum are also larger, which is unfavorable for high single-bunch current.

Table 3
Main parameters of injection element.

Title	Description	Value	Unit
Stripline	Length	0.72	m
	Gap	9	mm
	Pulser voltage	±15	kV
	Kick angle	0.72	mrad
	$\Delta t_{10\%-90\%}$	4.5	ns
	Δt_{top}	5.9	ns
	$\Delta t_{90\%-10\%}$	4.5	ns
	Reproducibility (top)	≤1%	
	Tail amplitude	≤3%	
Lambertson	Length	1.8	m
	Thickness	2	mm
	Field strength	1	T
	Bend angle	90	mrad
	Tilt angle	98.2	mrad
	Half-aperture to axis	3.5	mm

Another choice we have is between injection in the horizontal and vertical plane. The APS booster provides a beam with much smaller emittance in the vertical than in the horizontal plane. In addition, we anticipate future evolution of the MBA lattice might include reducing the horizontal beta function in the straight sections. Hence, it makes sense to inject in the vertical plane. Based on these arguments, which admittedly are not conclusive at this stage, we have selected vertical injection in a single straight section as our default option.

A conceptual injected beam orbit layout is shown in Fig. 26. Due to the limited available straight section length, the injected beam cannot avoid passing through the upstream quadrupole doublet. In order for the injected beam to stay near the accelerator midplane, a slightly tilted Lambertson septum magnet is used, giving a small vertical bending angle in addition to a relatively large horizontal bend. The main parameters of the injection components are listed in Table 3.

To ensure low losses, we have designed the injection system such that the injected beam envelope will freely pass through the kicker and septum apertures, using $\pm 3\sigma$ for the transverse planes and assuming 0.5% r.m.s. energy spread (which includes a large allowance for energy jitter). The coordinates

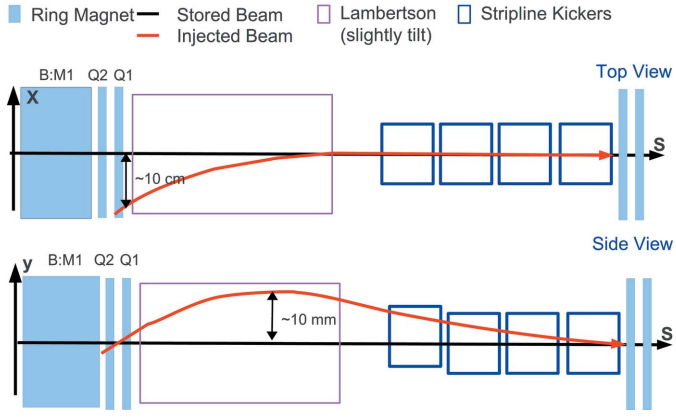

Figure 26
Schematic layout of vertical injection beam orbit (not to scale).

Table 4
Injected beam envelope ($\pm 3\sigma$) at specified locations for different injection configurations.

Name	Location	x^- (mm)	x^+ (mm)	y^- (mm)	y^+ (mm)
B:M1	Exit	−182.6	−175.7	−4.08	−1.77
Q2	Entrance	−177.3	−170.5	−3.72	−1.43
	Exit	−144.4	−138.3	−1.48	0.64
Q1	Entrance	−127.6	−121.9	−0.33	1.70
	Exit	−97.9	−92.9	1.69	3.58
Lambertson	Entrance	−82.9	−78.1	2.71	4.53
	Exit	−1.8	1.8	5.76	6.97
Kick-1	Entrance	−1.8	1.8	4.03	5.13
	Exit	−1.8	1.8	2.21	3.32

of the beam edges at the entrance and exit of the upstream B:M1, B:Q2 and B:Q1 magnets, and at the first stripline, are listed in Table 4. We also assume the injected beam is shielded from the magnetic field of B:Q1 and B:Q2 by using an appropriate material choice for the relevant section of the vacuum chamber.

To confirm the design, the injection efficiency for injected particles (*i.e.* not including transport loss from the booster to injection kicker) has been simulated using *elegant* with various injection errors. As in the lattice evaluation reported above, we simulated with 100 sets of optical errors and multipole errors. For each error set, we further included +2% or −2% systematic (*i.e.* the same on all kickers) strength errors together with 10000 instances of 5% r.m.s. random errors for the stripline kicker and 0.01% r.m.s. random errors for the Lambertson septum. The injected bunch ($\pm 3\sigma$ with ε_x = 60 nm, ε_y = 16 nm, $\delta p/p = 0.5\%$) has been tracked through the ring for 500 turns and particle loss is recorded. Longitudinal motion (RF) and synchrotron radiation damping are included in the tracking. The average injection loss distribution (average over 10000 shots) over 100 error ensembles is shown in Fig. 27; the average injection efficiency is better than 99%. For the optical error ensemble which gives the highest average loss, we find that only six of the 10000 shots have injection efficiency less than 95%. This indicates that the injection scheme should perform very well.

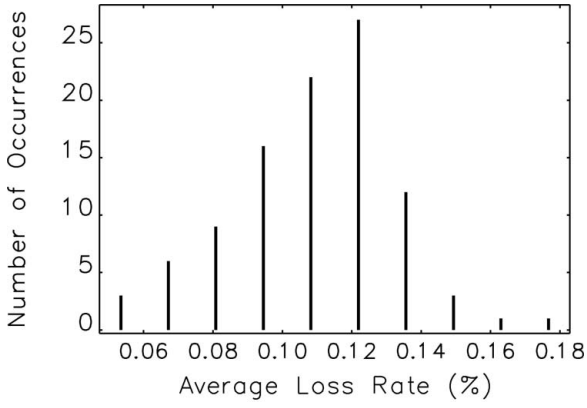

Figure 27
Average injection loss distribution (average over 10000 shots) over 100 optical errors.

6.6. X-ray performance

Using the electron beam parameters for various fill patterns allows us to compute the X-ray performance from undulator sources in the lattice just presented and compare it with the present-day APS. For the MBA lattice, we looked at the performance of 4.8 m-long hybrid-permanent magnet (HPM) undulators and 3.7 m-long superconducting undulators (SCUs), both with 8.5 mm magnetic gaps. The SCUs are shorter to allow space for warm-to-cold transitions at both ends.

To obtain a very broad idea of performance, instead of picking specific devices we varied the undulator periods from 12 to 33 mm in 0.5 mm steps. The maximum K value for the HPM devices was computed from logarithmic extrapolation of existing APS undulator measurements, which gives results within 10% of a simulation-derived model (Dejus *et al.*, 2009). For the SCUs, we used the scaling derived by Kim (2005), which is based on the use of NbTi superconductor. Having computed radiation properties for each device, we found the envelopes over all devices, which makes for a very straightforward comparison.

Fig. 28 compares the average brightness of the MBA lattice to the existing performance, showing increases of two to three orders of magnitude. Not surprisingly, the flux improvements, shown in Fig. 29, are less dramatic, but still significant. They result from a combination of higher beam current and shorter-

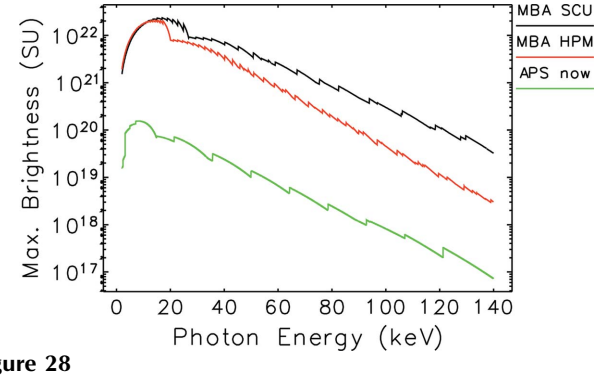

Figure 28
Comparison of envelopes of average brightness for the present-day APS with the MBA replacement lattice, in units of photons s^{-1} mm^{-2} $mrad^{-2}$ (0.1% bandwidth)$^{-1}$.

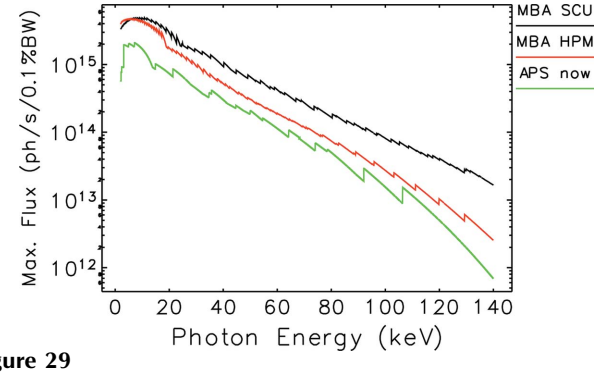

Figure 29
Comparison of envelopes of central cone flux for the present-day APS with the MBA replacement lattice.

period undulators, which allows squeezing more undulator periods into the same space.

7. Conclusion

Storage-ring light sources have made dramatic progress in the decades since the first parasitic experiments were performed. Third-generation light sources based (largely) on the Chasman–Green double-bend configuration and supporting large numbers of undulators provide excellent performance to an ever-growing community of scientists in diverse fields of research. Even as these machines began to come into operation, ideas for more compact higher-brightness sources were beginning to be explored (Einfeld & Plesko, 1996; Kaltchev *et al.*, 1996). These ideas, in the form of the multi-bend achromat lattice, have now blossomed into a subject of intense interest, with the first rings based on these concepts now under construction (Leemann *et al.*, 2009; Liu *et al.*, 2013). These rings and those that seem likely to follow will usher in a fourth generation of storage-ring light sources, with increases of brightness of several orders of magnitude. The beam dynamics and engineering challenges of these rings are significant, but modern theoretical and simulation techniques provide the tools necessary to meet those challenges. As an example, we presented an exploratory study of a possible APS upgrade lattice, which promises to increase the brightness by two or three orders of magnitude compared with the present-day APS.

Thanks to P. Raimondi, L. Farvacque, N. Carmignani and others at ESRF for providing a copy of an earlier version of their lattice, which helped in the development of the APS HMBA lattice. Parts of this work were performed using the Blues cluster at Argonne's Laboratory Computing Resource Center and the Mira cluster at Argonne's Advanced Leadership Computing Facility. This work was supported by the US Department of Energy, Office of Science, under Contract No. DE-AC02-06CH11357.

References

Balewski, K. *et al.* (2004). *Proceedings of EPAC 2004*, pp. 2302–2304.
Bartolini, R., Hammond, N., Kay, J., Pulampong, T. & Walker, R. P. (2013). *Proceedings of NA-PAC 2013*, pp. 24–26.
Bazarov, I. & Sinclair, C. (2005). *Phys. Rev. ST Accel. Beams*, **8**, 034202.
Belgroune, M. *et al.* (2003). *Proceedings of PAC 2003*, pp. 896–898.
Bengtsson, J. (1997). *The Sextupole Scheme for the Swiss Light Source (SLS): an Analytic Approach.* Technical Report SLS Note 9/97. Paul Scherrer Institut, Villigen, Switzerland.
Bernardini, C., Corazza, G. F., Giugno, G. D., Chigo, G., Haissinski, J., Marin, P., Querzoli, R. & Touschek, B. (1963). *Phys. Rev. Lett.* **10**, 407–409.
Bjorken, J. D. & Mtingwa, S. K. (1983). *Part. Accel.* **13**, 115.
Borland, M. (2000). *elegant: A Flexible SDDS-Compliant Code for Accelerator Simulation.* Technical Report LS-287. Advanced Photon Source, Argonne, IL, USA.
Borland, M. (2014). In preparation.
Borland, M., Emery, L., Sajaev, V., Xiao, A. & Guo, W. (2009*a*). *ICAP 2009*, p. 255.
Borland, M., Emery, L., Shang, H. & Soliday, R. (2003). *Proceedings of PAC 2003*, pp. 3461–3463.
Borland, M., Sajaev, V., Emery, L. & Xiao, A. (2010). *Multi-objective Direct Optimization of Dynamic Acceptance and Lifetime for Potential Upgrades of the Advanced Photon Source.* Technical Report LS-319. Advanced Photon Source, Argonne, IL, USA.
Borland, M., Sajaev, V., Shang, H., Soliday, R., Wang, Y., Xiao, A. & Guo, W. (2009*b*). *Proceedings of ICAP 2009*, pp. 111–116.
Borland, M., Sajaev, V. & Sun, Y. (2013). *Proceedings of PAC 2013*, pp. 249–251.
Borland, M. *et al.* (2009*c*). *Proceedings of 2009 PAC*, pp. 3850–3852.
Bosch, R. A. (2005). *Phys. Rev. ST Accel. Beams*, **8**, 084401.
Bosch, R. A., Kleman, K. J. & Bisognano, J. J. (2001). *Phys. Rev. ST Accel. Beams*, **4**, 074401.
Brown, K. (1979). *IEEE Trans. Nucl. Sci.* NS-**26**, 3490–3492.
Bruck, H. (1972). *Circular Particle Accelerators.* LA-TR. Los Alamos Scientific Laboratory, USA.
Cai, Y., Bane, K., Hettel, R., Nosochkov, Y., Wang, M. & Borland, M. (2012). *Phys. Rev. ST Accel. Beams*, **15**, 054002.
Chae, Y.-C. (2014). Private communication.
Chasman, R., Green, G. K. & Rowe, E. M. (1975). *Proceedings of PAC 1975.*
Courant, E. & Snyder, H. (1958). *Ann. Phys.* **3**, 1–48.
Crosbie, E. A. (1987). *Proceedings of 1987 PAC*, pp. 443–445.
Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. (2002). *IEEE Trans. Evolut. Comput.* **6**, 182–197.
Dejus, R., Jaski, M. & Kim, S. (2009). *On-Axis Brilliance and Power of In-Vacuum Undulators for the Advanced Photon Source.* Technical Report LS-314. Advanced Photon Source, Argonne, IL, USA.
Dragt, A. J. (1987). *Nucl. Instrum. Methods Phys. Res. A*, **258**, 339–354.
Einfeld, D. & Plesko, M. (1996). *Proceedings of EPAC 1996.*
Elleaume, P. (2003). *Undulator Radiation.* Taylor and Francis.
Emery, L. (1990). PhD thesis, Stanford University, Stanford, USA.
Emery, L. & Borland, M. (2000). Unpublished program.
Emery, L. & Borland, M. (2002). *Proceedings of 2002 EPAC*, pp. 218–220.
Emery, L. & Borland, M. (2003). *Proceedings of PAC03*, pp. 256–258.
Farvacque, L., Carmignani, N., Chavanne, J., Franchi, A., Bec, G. L., Liuzzo, S., Nash, B., Perron, T. & Raimondi, P. (2013). *Proceedings of 2013 PAC*, pp. 79–81.
Farvacque, L., Laclare, J. L., Nghiem, P., Payet, J., Ropert, A., Tanaka, H. & Tkatchenko, A. (1994). *Proceedings of EPAC 1994*, pp. 612–614.
Gao, W., Wang, L. & Li, W. (2011). *Phys. Rev. ST Accel. Beams*, **14**, 094001.
Guo, W., Kramer, S., Krinsky, S., Li, Y., Nash, B. & Tanabe, T. (2009). *Proceedings of PAC 2009*, pp. 1102–1104.
Haissinski, J. (1973). *Il Nuovo Cimento*, **18**, 72–82.
Halbach, K. (1969). *Nucl. Instrum. Methods Phys. Res. A*, **74**, 147–164.
Hofler, A. S. (2013). *Proceedings of PAC13*, pp. 1111–1115.
Hofler, A. S., Terzic, B., Kramer, M., Zvezdin, A., Morozov, V., Roblin, Y., Lin, F. & Jarvis, C. (2013). *Phys. Rev. ST Accel. Beams*, **16**, 010101.
Jain, A. (2011). *17th International Magnetic Measurement Workshop*, Barcelona, Spain.
Jaski, M. (2014). Private communication.
Kaltchev, D., Servranckx, R. V. & Craddock, M. K. (1996). *Proceedings of EPAC 1996.*
Kersevan, R. (2000). *Proceedings of EPAC 2000*, pp. 2289–2291.
Kim, S. H. (2005). *Nucl. Instrum. Methods Phys. Res. A*, **546**, 604–619.
Krinsky, S. & Wang, J. M. (1985). *Part. Accel.* **17**, 109.
Laskar, J. (1990). *Icarus*, **88**, 266–291.

Le Duff, J. (1985). *Nucl. Instrum. Methods Phys. Res. A*, **239**, 83–101.
Leemann, S. C., Andersson, A., Eriksson, M., Lindgren, L.-J., Wallen, E., Bengtsson, J. & Streun, A. (2009). *Phys. Rev. ST Accel. Beams*, **12**, 120701.
Liu, L., Milas, N., Mukai, A. H. C., Resende, X. R., Rodgrigues, A. R. D. & Sa, F. H. (2013). *Proceedings of IPAC 2013*, pp. 1874–1876.
Murphy, J. (1989). *Synchrotron Light Source Data Book*. Technical Report BNL-42333. BNL, Brookhaven, USA.
Nagaoka, R., Brunelle, P., Nadolski, L. S., Nadji, A., Gavalda, X. N., Klein, M., Loulergue, A. & Tordeux, M. A. (2013). *Proceedings of IPAC 2013*, pp. 76–78.
Piwinski, A. (1998). *The Touschek Effect in Strong Focusing Storage Rings*. Technical Report DESY 98–179. DESY, Germany.
Ramanathan, M. (2003). Private communication.
Revol, J. & Nagaoka, R. (2001). *Proceedings of PAC 2001*, pp. 1930–1932.
Safranek, J., Limborg, C., Terebilo, A., Blomqvist, K., Elleaume, P. & Nosochkov, Y. (2002). *Phys. Rev. ST Accel. Beams*, **5**, 010701.
Sagan, D. (1994). *Am. J. Phys.* **62**, 450.
Sands, M. (1970). *The Physics of Electron Storage Rings*. Technical Report SLAC-121. SLAC, USA.
Shang, H. & Borland, M. (2005). *Proceedings of 2005 PAC*, pp. 4230–4232.
Srinivas, N. & Deb, K. (1995). *Evol. Comput.* **2**, 221–248.
Steier, C., Robin, D., Nadolski, L., Decking, W., Wu, Y. & Laskar, J. (2002). *Phys. Rev. E*, **65**, 056506.
Streun, A., (1999). *Opa lattice design code*, http://ados.web.psi.ch/opa/.
Sun, C., Robin, D. S., Nishimura, H., Steier, C. & Wan, W. (2012). *Phys. Rev. ST Accel. Beams*, **15**, 054001.
Sun, Y.-P. & Borland, M. (2013). *Proceedings of NA-PAC 2013*.
Tarawneh, H., Steier, C., Robin, D., Nishimura, H., Sun, C. & Wan, W. (2013). *Proceedings of NA-PAC 2013*, pp. 288–290.
Wan, W. & Berz, M. (1996). *Phys. Rev. E*, **54**, 2870.
Wang, J. Q. (2013). *BAPS accelerator design and RandD*, presented at *Workshop on Facility Upgrades and High Brightness Sources*, SRI 2013.
Wang, Y., Borland, M. & Sajaev, V. (2011). *Proceedings of PAC 2011*, pp. 787–789.
Wang, Y., Borland, M., Shang, H. & Soliday, R. (2009). *Proceedings of ICAP 2009*, pp. 355–358.
Wiedemann, H. (2003*a*). *Particle Accelerator Physics*, Vol. 2, 2nd ed. Berlin: Springer.
Wiedemann, H. (2003*b*). *Particle Accelerator Physics*, Vol. 1, 2nd ed. Berlin: Springer.
Xiao, A. (2008). *Proceedings of Linac 08*, pp. 296–299.
Xiao, A. & Borland, M. (2007). *Proceedings of PAC 2007*, pp. 3453–3455.
Xiao, A., Borland, M., Emery, L. & Sajaev, V. (2013). *Proceedings of PAC 2013*, pp. 1073–1075.
Yang, L., Li, Y., Guo, W. & Krinsky, S. (2011). *Phys. Rev. ST Accel. Beams*, **14**, 054001.
Yang, L., Robin, D., Sannibale, F., Steier, C. & Wan, W. (2009). *Nucl. Instrum. Methods Phys. Res. A*, **609**, 50–57.
Yao, C.-Y., Borland, M., Emery, L., Donley, L., Lenkszus, F. & Yang, B.-X. (2013). *Proceedings of PAC 2013*.
Yao, C.-Y., Sajaev, V. & Monte, N. D. (2011). *Proceedings of IPAC 2011*, pp. 2803–2805.
Zotter, B. (2013). *Handbook of Accelerator Physics and Engineering*, 2nd ed., edited by A. W. Chao & M. Tigner, §2.4.9, pp. 147–149. Singapore: World Scientific.