where f is a dimensionless ω-dependent numerical factor, obtainable from Eq. (6.2.42). For $f \approx 0.5$, a typical value, $f/\gamma^2 \approx 0.5 \times 10^{-8}\,\mathrm{m}$. In the last step a "low emittance" lattice has been assumed, which amounts to neglecting electron angular spreads relative to photon angular spreads. This approximation is not very bad even for a machine with gigantic emittances, such as CESR.

Using Eq. (8.3.1) is tantamount to assuming that the detection apparatus accepts and utilizes the full photon cone angle. In practice, the cone angle would be reduced by collimation. According to Eq. (6.3.7), this would reduce the solid angle (and hence photon rate) and energy spread more or less proportionally. The brilliance, a ratio of photon rate to solid angle, would therefore tend to be more or less independent of collimation angle, except that, by convention, one is to count only photons in the tenth percent bandwidth. Because Eq. (8.3.1) has already incorporated the undulator-effect reduction of fractional energy spread to be of the order of $1/N_w$, this requirement has already been built into Eq. (8.3.1). As a result there is really a unique angular collimation compatible with this formula—that which limits the fractional energy spread to $\Delta\nu = 10^{-3}$. Since this formalism is almost certainly different from that employed by others, I will use $\mathcal{B}'$ as its symbol. $\mathcal{B}'$ is obtained by combining Eqs. (8.2.12) and (8.3.1);

$$\mathcal{B}' = \frac{\mathcal{F}'}{\pi^2 \sigma_{\gamma,x}\sigma_{\gamma,x'}\sigma_{\gamma,y}\sigma_{\gamma,y'}} = \frac{1}{\pi^2}\,\frac{4}{\pi^{5/2}}\,\frac{L_w\,R_0}{ab}\,\frac{U_0}{h\nu}\,\frac{I}{e}\,\frac{P_{\text{ext.}}\,(MW)}{693.1\,\mathrm{MW}}\,\frac{P}{P_{\text{ext.}}}\,\frac{N_w}{10^3}\,\frac{1/f}{\sqrt{\beta_x\epsilon_x}\,\sqrt{\beta_y\epsilon_y}} \tag{8.3.2}$$

From here on $\mathcal{B}'$ will not be distinguished from brilliance $\mathcal{B}$. As has been emphasized repeatedly, since "all" the energy is concentrated in the main peak, the total power of the beam is trivially small.

There are too many uncertain factors to give a definitive numerical value to $\mathcal{B}$. The main factors multiplying the flux $\mathcal{F}'$ to give brilliance $\mathcal{B}$, are

- $1/\pi^2 \approx 10^{-1}$, a factor that accounts for the fact that the bunch σ's are one-sided measures.
- $\gamma^2 \overset{\text{e.g.}}{=} 10^8$. As mentioned below Eq. (6.3.7), since the flux has not been de-rated to account for the angular spread, the brilliance must not be re-inflated to account for an angular aperture much less than $1/\gamma$.

- $(f\sqrt{\beta_x\beta_y})^{-1} \stackrel{\text{e.g.}}{=} 1/\text{m}$, because $f \approx 0.5$ and $\sqrt{\beta_x\beta_y} \approx 2\,\text{m}$.‡
- $1/\sqrt{\epsilon_x\epsilon_y} = 10^{10}\,(\text{m-rad})^{-1}$ for an exceedingly low, but probably achievable emittance, some 4 times greater (i.e. smaller emittance) than the corresponding product at ESRF or APS.
- $10^{-12}\,(\frac{\text{m}}{\text{mm}})^2\,(\frac{\text{r}}{\text{mr}})^2$ because $\mathcal{B}$ is conventionally quoted in terms of millimeters and milliradians, rather than meters and radians.

Combining these values with the flux value given in Eq. (8.2.13) for a 10 m undulator yields

$$\mathcal{B}_{0.1\text{A}} = 3 \times 10^{19}\,\frac{\text{photons/s/0.1\%BW/mm}^2\text{/mr}^2}{\text{MW}}\ . \tag{8.3.3}$$

For quantitative comparison with the brilliance achieved or advertised by other storage rings, one should be sure the assumptions going into the definition of brilliance are consistent. Various factors have been ignored, some making Eq. (8.3.3) too pessimistic (e.g. the factor 3 in Eq. (6.3.6)) others too optimistic (e.g. the treatment of the collimator, ignoring the elliptical shape of the beams, and oversimplifying the shape of the upper end of the undulator energy spectrum.) These uncertainties cause Eq. (8.3.3) to be little better than an order-of-magnitude estimate.

According to Suller, one of the magnetic undulators at ESRF, operating at 6 GeV.[14] produces brilliance 1.6×10^{19} photons/s/0.1%BW/mm^2/mr^2 at $h\nu = 7.4$ keV; this would be reduced by roughly a factor of ten in extrapolating to 12.4 keV. These numbers correspond to emittances $\epsilon_x = 4 \times 10^{-9}$ m and $\epsilon_y = 4 \times 10^{-11}$ m. The advertised brilliance at $h\nu =$ 12.4 keV, for an undulator in "Protein Crystallography Beamline" of SLS (The Swiss Light Source) is $\mathcal{B}_{0.4\text{A}} = 0.5\times10^{19}$photons/s/0.1%BW/mm^2/mr^2. This line will be known as the "Protein Crystallography Beamline". The undulator gap height is (planned to be) 4 mm, beam energy is 2.6 GeV(?), and $K_{\max} = 1.65$.

The brilliance increases only as $(\sqrt{\epsilon_x\epsilon_y})^{-1}$. The reason for this is that the radiation cone has been assumed to be large compared to the (elliptical) cone of electron angles. It is necessary to check that this is valid. Because low emittance storage ring designs lead to

‡ There is "headroom" for increasing $\mathcal{B}$ by decreasing either or both of β_x and β_y. Furthermore, since the waveguide can pass through the center of focusing magnets, it may be practical to keep β_x and β_y small along the length of even a long undulator.

$\langle\beta_x\rangle << \langle\beta_y\rangle$ and $\epsilon_x >> \epsilon_y$ the only condition that needs to be checked is

$$\frac{\epsilon_x}{\beta_x} << \frac{1}{\gamma^2} \ . \tag{8.3.4}$$

The same x, y asymmetry may influence the choice of waveguide dimensions. To increase X-ray intensity at fixed microwave power there is a premium on reducing the transverse waveguide dimensions. In this paper we have been using width/height $= a/b = 4\,\mathrm{cm}/2\,\mathrm{cm} = 2$, and will continue to do this. But, in principle the waveguide height could be reduced without violating condition (8.3.4) or clipping the vertical tails of the beam distribution.

Another condition to be satisfied is that the spread of "searchlight angles" of the electron beam passing through the undulator (about $K/(3\gamma)$) should roughly match the cone angle of a collimator whose purpose is to limit the spread of X-ray energies (about $\sqrt{\Delta\nu}/\gamma \approx 1/(30\gamma)$. This sets a limit $K < 0.1$. Since flux and brilliance are proportional to K^2, this consideration helps to make the brilliance from the microwave undulator (where achieving high K is difficult) competitive with the brilliance from a magnetic undulator (where achieving high K is easy.)

Reducing the horizontal emittance ϵ_x improves the X-ray beam brightness, but there is a value below which diminishing returns set in. A matching condition based on considerations similar to the previous paragraph is that the spread of "searchlight angles" of the electron beam passing through the undulator (about $K/(3\gamma)$) should not exceed the spread of electron angles (about $\sqrt{\epsilon_x/\beta_x}$.) Accepting this as a strict inequality yields

$$K < 3\gamma\sqrt{\frac{\epsilon_x}{\beta_x}} \ , \quad \text{or} \quad \epsilon_x > \frac{K^2\beta_x}{10\gamma^2} \ , \tag{8.3.5}$$

depending on whether the "pinch" comes from K or ϵ_x. For a long undulator β_x may have to be 10 m or greater, in which case, ϵ_x has to be K^2/γ^2 m or greater. For the Energy Recovery Linac X-ray source that has been under discussion recently, the proposed emittance is $\epsilon_x \approx 10^{-10}$. With γ being 10^4, the value of K^2 should not exceed 0.01. The flux and brilliance would be therefore be down from the $K = 1$ values by two orders of magnitude.

Appendix A.
Trajectory of electron in electromagnetic wave

This paper requires the description of charged particle orbits in a traveling electromagnetic wave and, for comparison with conventional undulators, the motion also in a periodic magnetic field. Unfortunately, though the orbits are very similar, the analysis for a periodic magnetic field cannot be subsumed into the analysis for a traveling wave, even by going to the zero frequency limit, because there is no frame of reference in which the electric field of a traveling wave vanishes. This correlates with the fact that energy transfer between particle and field is possible for a traveling wave, but not for a pure magnetic field. Such energy transfer is fundamental to the operation of free electron lasers, but is inessential to the operation of the microwave undulator being discussed in this paper.

Since the electron is highly relativistic, it is essential for relativistically valid formulas to be used. Fortunatately, exact equations of motion are known[†] for motion of a charged particle in an electromagnetic wave.

For an electromagnetic plane wave traveling in direction $\hat{\mathbf{n}}$, the electric and magnetic fields are related by

$$\mathbf{B} = \frac{1}{c}\,\hat{\mathbf{n}} \times \mathbf{E}\ , \quad \text{and} \quad \mathbf{B}\cdot\hat{\mathbf{n}} = \mathbf{E}\cdot\hat{\mathbf{n}} = 0\ . \tag{A.1}$$

For a monochromatic wave of frequency $\omega_{\rm rf}$, dependencies on both position $\mathbf{r}$ and time t can be expressed in terms of a single independent (phase) variable

$$\Phi = \omega_{\rm rf}\,(t - \hat{\mathbf{n}}\cdot\mathbf{r}/c)\ . \tag{A.2}$$

For the special case in which the wave is traveling parallel to the positive z-axis, $\Phi = \omega_{\rm rf}(t - z/c) = \omega_{\rm rf}t - kz$, where k is the photon wave number. Later, $\mathbf{r}$ will be taken to be the position of a particle with velocity $\mathbf{v} = d\mathbf{r}/dt$, and then

$$\frac{d\Phi}{dt} = \omega_{\rm rf}\,(1 - \hat{\mathbf{n}}\cdot\mathbf{v}/c)\ . \tag{A.3}$$

[†] Clemmow, P.C. and Dougherty, J.P., *Electrodynamics of Particles and Plasmas,* ascribe this theory to Kolomenskii A.A. and Lebedev A.N., *Sov. Phys. Doklady,* **7**, *745 (1963)* and *Sov. Phys. JETP,* **17**, *179 (1963).*

In effect, the variable Φ locates the particle by giving the instantaneous longitudinal projection (onto the wave normal) relative to some standard wavefront of the wave. We are primarily interested in the case in which the electron travels almost anti-parallel to the wave at almost the speed of light, $z = -|\overline{v_z}|t$. Then $\Phi = \omega_{\rm rf} t(1 + |\overline{v_z}|/c) \approx 2\omega_{\rm rf} t$.

The mechanical energy γmc^2 of an electron of velocity $\mathbf{v}$ is governed by

$$\frac{d\gamma}{dt} = \frac{e}{mc^2}\,\mathbf{E}\cdot\mathbf{v}, \tag{A.4}$$

and its equation of motion is

$$\frac{d}{dt}\,\left(\gamma\frac{\mathbf{v}}{c}\right) = \frac{e}{mc}\,(\mathbf{E}+\mathbf{v}\times\mathbf{B}) = \frac{e}{mc}\,(1-\hat{\mathbf{n}}\cdot\mathbf{v}/c)\,\mathbf{E} + \frac{e}{mc^2}\,(\mathbf{E}\cdot\mathbf{v})\,\hat{\mathbf{n}}. \tag{A.5}$$

By combining these equations one shows that the quantity

$$\mathcal{L} = \gamma\,(1-\hat{\mathbf{n}}\cdot\mathbf{v}/c)\ , \tag{A.6}$$

is a constant of the motion. For exactly anti-parallel motion $\mathcal{L} \approx 2\gamma$.

We now set about changing independent variable from t to Φ in the electron's equation of motion, using primes to indicate $d/d\Phi$. Then, using Eq. (A.3),

$$\mathbf{r}' = \frac{d\mathbf{r}/dt}{d\Phi/dt} = \frac{\mathbf{v}}{\omega_{\rm rf}\,(1-\hat{\mathbf{n}}\cdot\mathbf{v}/c)}\ . \tag{A.7}$$

Differentiating again yields

$$\mathbf{r}'' = \frac{1}{{\omega_{\rm rf}}^2\,(1-\hat{\mathbf{n}}\cdot\mathbf{v}/c)}\,\frac{d}{dt}\,\left(\frac{\mathbf{v}}{(1-\hat{\mathbf{n}}\cdot\mathbf{v}/c)}\right)\ . \tag{A.8}$$

Using this, after first substituting from Eq. (A.6), the left hand side of Eq. (A.5) can be rewritten as

$$\frac{d}{dt}\,(\gamma\,\mathbf{v}/c) = \mathcal{L}\,\frac{d}{dt}\,\left(\frac{\mathbf{v}/c}{1-\hat{\mathbf{n}}\cdot\mathbf{v}/c}\right) = {\omega_{\rm rf}}^2\mathcal{L}\,(1-\hat{\mathbf{n}}\cdot\mathbf{v}/c)\,\mathbf{r}''\ . \tag{A.9}$$

and, again using Eq. (A.7), the right hand side is

$$(1-\hat{\mathbf{n}}\cdot\mathbf{v}/c)\,\frac{e}{m}\,\left(\mathbf{E}+\frac{\omega_{\rm rf}}{c}\mathbf{E}\cdot\mathbf{r}'\,\hat{\mathbf{n}}\right)\ . \tag{A.10}$$

These manipulations have permitted the common factor $1-\hat{\mathbf{n}}\cdot\mathbf{v}/c$ to be cancelled, and the equation of motion becomes

$$\mathbf{r}'' = \frac{e}{\mathcal{L}\,m\,{\omega_{\rm rf}}^2}\,\left(\mathbf{E}+\frac{\omega_{\rm rf}}{c}\mathbf{E}\cdot\mathbf{r}'\,\hat{\mathbf{n}}\right)\ . \tag{A.11}$$

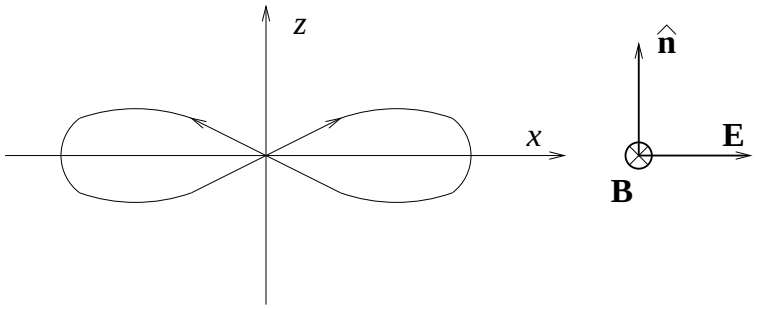


Figure A.1: Motion of electron, relative to its average motion, in a plane polarized, electromagnetic plane wave.

This equation is exact relativistically, except for not including the radiation reaction force. We know that the only important effect of this force is the "slowing down" required by energy conservation.

For the special case in which the wave is traveling parallel to the positive z-axis, $\hat{\mathbf{n}} = \hat{\mathbf{z}}$, with electric field directed along the x-axis,

$$\mathbf{n} = (0,0,1)\,, \quad \mathbf{E} = E\,\cos\Phi\;(1,0,0)\,, \quad \mathbf{B} = \frac{E}{c}\,\cos\Phi\;(0,1,0)\,, \tag{A.12}$$

and the equations of motion are

$$x'' = \frac{e\,E}{\mathcal{L}\,m\,{\omega_{\rm rf}}^2}\,\cos\Phi, \quad y'' = 0, \quad z'' = \frac{e\,E}{c\,\mathcal{L}\,m\,{\omega_{\rm rf}}^2}\,x'\,\cos\Phi. \tag{A.13}$$

The solution to the first of these equations is

$$x' = \frac{e\,E}{\mathcal{L}\,m\,{\omega_{\rm rf}}^2}\,\sin\Phi, \quad x = -\frac{e\,E}{\mathcal{L}\,m\,{\omega_{\rm rf}}^2}\,\cos\Phi, \tag{A.14}$$

where integration constants have been dropped since we now assume the particle is, on the average, moving with no transverse drift along the z-axis. Then the third of Eqs. (A.13) becomes

$$z'' = \frac{1}{2\,c}\left(\frac{e\,E}{\mathcal{L}\,m\,{\omega_{\rm rf}}^2}\right)^2\,\sin\,2\Phi, \tag{A.15}$$

with solution

$$\begin{aligned} z' &= -\frac{1}{4\,c}\left(\frac{e\,E}{\mathcal{L}\,m\,{\omega_{\rm rf}}^2}\right)^2\,\cos\,2\Phi + \overline{z'}, \\ z &= -\frac{1}{8\,c}\left(\frac{e\,E}{\mathcal{L}\,m\,{\omega_{\rm rf}}^2}\right)^2\,\sin\,2\Phi + \overline{z'}\,\Phi\;, \end{aligned} \tag{A.16}$$

where a constant contribution to z has been dropped.

Energy transfer. Because the motion has a transverse component, parallel to the electric field, the particle energy varies;

$$\frac{d\gamma}{d\Phi} = \frac{e}{mc^2} \cdot \mathbf{v}' = \frac{1}{\mathcal{L}} \left(\frac{e\,E}{m\,c\,\omega_{\rm rf}} \right)^2 \sin\Phi, \quad \gamma = -\frac{1}{\mathcal{L}} \left(\frac{e\,E}{m\,c\,\omega_{\rm rf}} \right)^2 \cos\Phi + \gamma_0 \ . \qquad (A.17)$$

Any decrease of electron energy must correspondingly increase the beam energy. This is known as free electron laser radiation. If the wave increases the electron energy (at the expense of its own intensity) it is known as an inverse free electron laser. Whether the electron gains or loses energy depends on where it rides on the wave (i.e. its phase). Of course the laser cannot be arbitrarily long, since the phase varies monotonically, and the energy change eventually averages to zero. The considerations of this aside are not really relevant if the beam and wave are approximately anti-parallel, as is the case in the microwave undulator, since the phase Φ varies rapidly, and this averaging takes place almost instantly.

Comparison with undulator. Making the approximations $\Phi \approx 2\omega_{\rm rf} t$ and $\mathcal{L} \approx 2\gamma$, Eq. $(A.16)$ yields

$$\frac{dx}{d\,(ct)} = \frac{c\,E}{\gamma\ (m\,c^2/e)\,\omega_{\rm rf}} \sin\Phi \ . \qquad (A.18)$$

For comparison with undulator formulas, it is convenient to define an "effective K value"

$$K_{\rm eff.} = \frac{E}{m\,c^2/e} \frac{c}{\omega_{\rm rf}} \ , \qquad (A.19)$$

which is the maximum angle, expressed in units of $1/\gamma$.

Connection to microwave power. According to Eq. (8.2.2), the maximum electric field in a microwave beam is related to the beam power and other guide parameters by (factor of two standing wave, traveling wave ambiguity here)

$$P = \frac{|E_{\rm max}|^2}{Z_0} \sqrt{1 - \left(\frac{\lambda_{\rm rf}}{2a} \right)^2} \frac{ab}{2} \ . \qquad (A.20)$$

Combining Eqs. $(A.19)$ and $(A.20)$, $K_{\rm eff.}$ can be expressed as a product of dimensionless factors

$$K_{\rm eff.} = \frac{c}{\omega_{\rm rf}} \sqrt{\frac{2}{ab}} \left(1 - \left(\frac{\lambda_{\rm rf}}{2a} \right)^2 \right)^{-1/4} \sqrt{\frac{P}{(mc^2/e)^2\,/Z_0}} \approx \frac{1}{\sqrt{2}\,\pi} \frac{\lambda_{\rm rf}}{\sqrt{ab}} \sqrt{\frac{P\,({\rm MW})}{693.1\,{\rm MW}}} \ . \qquad (A.21)$$

Appendix B.
Relativistically Invariant Treatment of the Microwave Undulator

Though the magnetic field of an undulator has been described as made up of photons, it is even more natural to treat a microwave beam this way, especially because the photons are real, not virtual. On the other hand, according to Eq. (8.2.3), the microwave photons are directed at angles $\pm\cos^{-1}\sqrt{1-(\lambda_{\rm rf}/(2a))^2}$ relative to the waveguide, and therefore also relative to the electron beam. Because of this, the electron-photon system has transverse momentum $p_{\perp,\gamma}$. This can be compared to the typical transverse momentum $p_{\perp,e}$ an electron has because of its betatron oscillation amplitude;

$$\frac{p_{\perp,\gamma}}{p_{\perp,e}} \approx \frac{h/\lambda_w}{\gamma m_e c\sqrt{\epsilon_x/\beta_x}} \quad \left(\stackrel{\text{e.g.}}{=} \frac{12.4\,\text{keV}}{2\gamma^3\times 0.511\,\text{MeV}}\sqrt{\frac{\beta_x}{\epsilon_x}}\,. \right) \tag{B.1}$$

This ratio is negligibly small for any conceivable electron beam. It is therefore legitimate to treat photon and electron as traveling on anti-parallel tracks. It may be important, later, to account for the spread of electron directions.

In this paper the polarization of the outgoing photons has been treated carelessly so far—only the in-plane polarization component has been retained, and only approximately at that. We are now in a position to tidy up this treatment, since the polarization dependence of Compton scattering is well documented. In the remainder of this section formulas will be copied from Berestetskii, Lifshitz, and Pitaevski, (BLP)[15] with considerably less than full understanding on my part. Feynman diagrams for the process are shown in Fig. B.1(a). Most formulas will be specialized to a frame of reference in which the accelerator is at rest.

Traditional discussions of Compton scattering have employed either electron rest system (sometimes called the "laboratory system), or a "center of mass" system in which the electron and photon momenta are equal and opposite. We will work in a different "laboratory system"; in it the photon is incident with four-momentum $\underline{k}$, traveling in the negative z-direction along the positive z-axis, and the electron is incident with four-momentum $\underline{p}$ traveling in the positive z-direction along the negative z-axis. With $\hbar = c = 1$, and

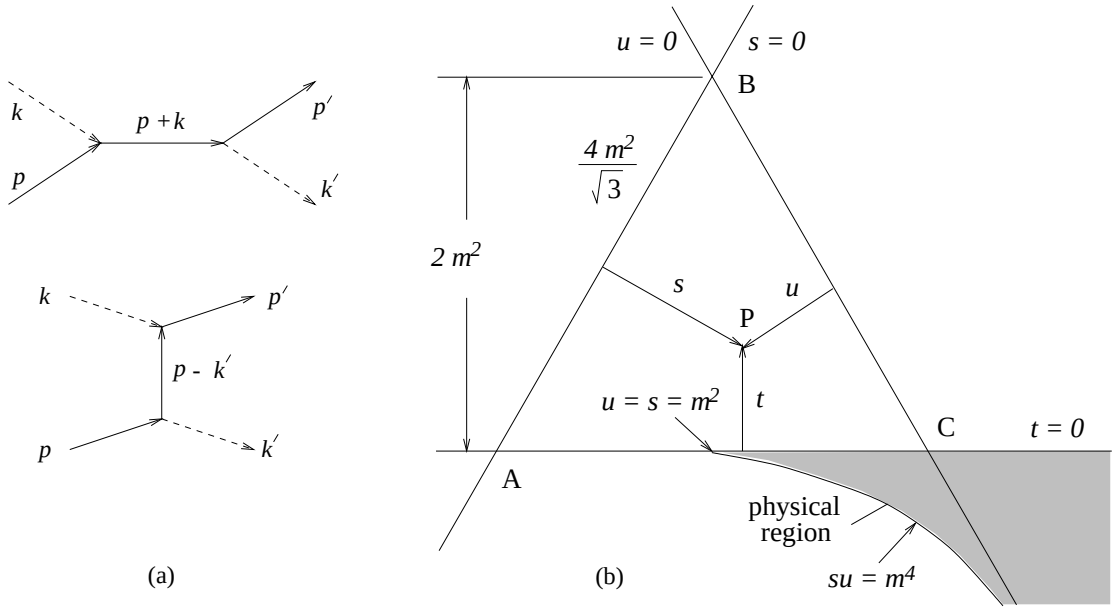


Figure B.1: (a) Feynman diagrams for Compton scattering. (b) Kinematic variables. Triangles APB, APC, and BPC have areas $(2/\sqrt{3})m^2 s$, $(2/\sqrt{3})m^2 t$, and $(2/\sqrt{3})m^2 s$, and triangle ABC has area $(4/\sqrt{3})m^4$ and is equal to their sum. This assures that Eq. $(B.6)$ is satisfied.

$E = \sqrt{p^2 + m^2}$, the initial four-momenta are†

$$\underline{k} = (\omega, -\omega, 0, 0) = (\omega, \mathbf{k}) \ , \quad \underline{p} = (E, p, 0, 0) = (E, \mathbf{p}) \ , \tag{B.2}$$

and the final four-momenta are

$$\underline{\omega'} = \left(\omega', k'_x, k'_y, k'_z\right) = \left(\omega', \mathbf{k}'\right) \ , \quad \underline{p'} = \left(E', p'_x, p'_y, p'_z\right) = \left(E', \mathbf{p}'\right) \ . \tag{B.3}$$

Representing invariant scalar products of 4-vectors a and b by

$$(\underline{a}, \underline{b}) \equiv a_0 b_0 - \mathbf{a} \cdot \mathbf{b} \ , \tag{B.4}$$

standard kinematic variables s, t, and u are defined by

$$\begin{aligned} s &= \left(\underline{p} + \underline{k}, \underline{p} + \underline{k}\right) = m^2 + 2\left(\underline{p}, \underline{k}\right) = m^2 + 2\omega\left(E + p\right) \ , \\ t &= \left(\underline{p} - \underline{p'}, \underline{p} - \underline{p'}\right) = -2\left(\underline{k'}, \underline{k}\right) = -2\omega\omega'\left(1 + \cos\vartheta\right) \ , \\ u &= \left(\underline{p} - \underline{k'}, \underline{p} - \underline{k'}\right) = m^2 - 2\left(\underline{p}, \underline{k'}\right) = m^2 - 2\omega'\left(E - p\cos\vartheta\right) \ . \end{aligned} \tag{B.5}$$

In the last steps, working in the laboratory system, with the photon scattering angle defined to be ϑ, the substitution $k'_x = \omega' \cos\vartheta$ has been performed.

† For consistency with Berestetskii, Lifshitz, and Pitaevski, the notation for the rest of this appendix is inconsistent with notation in the rest of the paper. The main inconsistencies are that the photon is incident from the right and the produced photon frequency, previously called ω, will, from here on, be ω'.

As illustrated in Fig. B.1, these variables satisfy

$$s + t + u = 2m^2 \quad \text{or} \quad \frac{\omega}{\omega'} = \frac{E - p\cos\vartheta}{E + p} + \frac{\omega\,(1 + \cos\vartheta)}{E + p} \approx \frac{1 - (p/E)\cos\vartheta}{1 + p/E} \ . \tag{B.6}$$

Dropping the last term will always be valid in the present context. To confirm this, even at $\vartheta = 0$, a useful number is

$$\frac{\omega E}{m^2} = \frac{\omega\gamma}{m}, \quad \left(\stackrel{\text{e.g.}}{=} \frac{1.24 \times 10^4\,\text{eV-Å}}{4 \times 10^8\,\text{Å}} \frac{10^4}{0.511 \times 10^6\,\text{eV}} = 0.6 \times 10^{-6} \ . \right) \tag{B.7}$$

For further simplifying Eq. (*B*.6), one can employ $p/E \approx 1 - \gamma^2/2$ and $\cos\vartheta \approx 1 - \vartheta^2/2$, so that

$$\omega'\,(\vartheta) \approx 2\omega \left(\frac{1}{2\gamma^2} + \frac{\vartheta^2}{2} \right)^{-1} \ . \tag{B.8}$$

Recognizing from Eq. (8.1.1) that $\lambda_w \approx \lambda_{\text{rf}}/2$, apart from notational differences, this relation is equivalent to Eq. (6.2.33) with $n = 1$ and $\Delta\Theta = 0$. One also obtains the approximation

$$t \approx -4\omega\omega' \approx -2\omega^2\,(1 - (p/E)\cos\vartheta) \ . \tag{B.9}$$

But it is not safe to evaluate u using the final (approximate) form of Eq. (*B*.6), since this would yield the result that u is independent of ϑ, (and we will need a formula for $du/d\cos\vartheta$.) From the exact form of Eq. (*B*.6) we have

$$\frac{d\omega}{d\cos\vartheta} = \frac{\omega'^2}{\omega} \frac{p - \omega}{E + p} \ . \tag{B.10}$$

Differentiating the last of Eqs. (*B*.5) then yields

$$\begin{aligned} \frac{du}{d\cos\vartheta} &= 2\omega' p - 2\omega'^2 \left(\frac{p}{\omega} - 1 \right) \frac{E - p\cos\vartheta}{E + p} \\ &= 2\omega' p - 2\omega'^2 \left(\frac{p}{\omega} - 1 \right) \left[\frac{\omega}{\omega'} - \frac{\omega\,(1 + \cos\vartheta)}{E + p} \right] \\ &\approx \omega'^2\,(1 + \cos\vartheta) \ . \end{aligned} \tag{B.11}$$

BLP also define

$$\begin{aligned} x &= \frac{s - m^2}{m^2} = -\frac{2\omega\,(E + p)}{m^2} \ , \\ y &= \frac{m^2 - u}{m^2} = \frac{2\omega'\,(E - p\cos\vartheta)}{m^2} \ . \end{aligned} \tag{B.12}$$

For calculating the differential cross section (per solid angle element $d\Omega'$), the following relations follow from approximation $(B.9)$:

$$\begin{aligned} d\Omega' &= \sin\vartheta\, d\phi\, d\vartheta = -d\cos\vartheta\, d\phi \ , \\ dt\, d\phi &\approx 2\omega^2\, d\cos\vartheta\, d\phi = -2\omega^2\, d\Omega' \ , \\ dy\, d\phi &\approx -\omega'^2\,(1+\cos\vartheta)\; d\cos\vartheta\, d\phi/m^2 = \omega'^2\,(1+\cos\vartheta)\; d\Omega'/m^2 \ . \end{aligned} \tag{B.13}$$

Here ϕ is the azimuthal angle, which is the same in laboratory system, electron rest system, and center of mass system. For azimuthally symmetric cross sections, $d\phi$ can be replaced by 2π to obtain a cross section differential in $d\vartheta$.

For the case of all incident particles being unpolarized and summing over final state polarizations, BLP give the Compton differential cross section to be

$$\frac{d\overline{\sigma}}{d\Omega'} = -8r_e^2\,\frac{\omega^2}{m^2x^2}\left(\left(\frac{1}{x}-\frac{1}{y}\right)^2+\left(\frac{1}{x}-\frac{1}{y}\right)+\frac{1}{4}\left(\frac{x}{y}+\frac{y}{x}\right)\right) \ . \tag{B.14}$$

In this formula, dt can be accurately approximated using Eq. $(B.9)$.

To analyse the polarization properties of the radiation one introduces photon polarization vector

$$\underline{e} = e_1\underline{e}_1 + e_2\underline{e}_2 \tag{B.15}$$

where $\underline{e}_1$ and $\underline{e}_2$ are four-vectors with vanishing time parts and with the unit vectors $\mathbf{e_1}$ and $\mathbf{e_2}$ as spatial parts. These vectors are orthogonal to the photon direction and to each other, $\mathbf{e_1}\cdot\mathbf{e_2^*}=0$, and $|e_1|^2$ and $|e_2|^2$ are probabilities that the photon has polarization $\overline{e}_1$ and $\overline{e}_2$ respectively. The state of polarization can be described by density matrix

$$\rho_{\alpha\beta} = e_\alpha e_\beta^* = \frac{1}{2}\begin{pmatrix} 1+\xi_3 & \xi_1-i\xi_2 \\ \xi_1+i\xi_2 & 1-\xi_3 \end{pmatrix} \ , \tag{B.16}$$

where the ξ_1, ξ_2, ξ_3 are "Stokes parameters" that take values in the range from -1 to $+1$. It is unclear to me at the moment how to calculate the Stokes parameters corresponding to a mode propagating in a waveguide though, because the wave is completely polarized $\xi_1^2+\xi_2^2+\xi_3^2=1$. The simplest possibility (it seems to me) has, for the density matrix of the initial photon state,

$$\mathbf{e}=\hat{\mathbf{x}}, \quad \rho_{\alpha\beta}=\begin{pmatrix}1&0\\0&0\end{pmatrix}, \quad (\xi_1,\xi_2,\xi_3)=(0,0,1) \ . \tag{B.17}$$

This would correspond to a pure, linearly polarized wave propagating parallel to the waveguide axis, with electric field horizontal.

For unpolarized incident electrons, and summing over polarizations of the final electron, the polarization dependent cross section (which is the cross section we are interested in) BLP give

$$\begin{aligned}\frac{d\sigma}{d\Omega'} &= \frac{1}{2}\frac{d\overline{\sigma}}{d\Omega'} + 2r_e^2\,\frac{\omega'^2\,(1+\cos\vartheta)}{m^2x^2}\,\Bigg(\,(\xi_3+\xi_3')\left(-\left(\frac{1}{x}-\frac{1}{y}\right)^2-\left(\frac{1}{x}-\frac{1}{y}\right)\right)\\ &\quad+\xi_1\xi_1'\left(\frac{1}{x}-\frac{1}{y}+\frac{1}{2}\right)+\frac{\xi_2\xi_2'}{4}\left(\frac{x}{y}+\frac{y}{x}\right)\left(1+\frac{2}{x}-\frac{2}{y}\right)\\ &\quad+\xi_3\xi_3'\left(\left(\frac{1}{x}-\frac{1}{y}\right)^2+\left(\frac{1}{x}-\frac{1}{y}\right)+\frac{1}{2}\right)\Bigg)\\ &\equiv G_0+G_3\left(\xi_3+\xi_3'\right)+G_{11}\xi_1\xi_1'+G_{22}\xi_2\xi_2'+G_{33}\xi_3\xi_3'\ .\end{aligned}\tag{B.18}$$

With coefficients defined this way, the Stokes parameters of the emitted photons are given by

$$\xi_1^{(f)}=\frac{G_{11}}{G_0+\xi_3G_3}\,\xi_1,\quad \xi_2^{(f)}=\frac{G_{22}}{G_0+\xi_3G_3}\,\xi_2,\quad \xi_3^{(f)}=\frac{G_3+G_{33}\xi_3}{G_0+\xi_3G_3}\ .\tag{B.19}$$

According to this $\xi_1=\xi_2=0$ implies $\xi_1^{(f)}=\xi_2^{(f)}=0$ which, with $\xi_3=1$, implies $\xi_3^{(f)}=(G_3+G_{33})/(G_0+G_3)$.

Chapter 9.
High Brilliance Circular Rings

9.1. Choice of Beam Energy

For centuries optical physicists had to be satisfied with the modest 3:1 dynamic wavelength range from red to violet, because light outside that range was invisible or their equipment was opaque. For some reason X-ray physicists are reluctant to accept a similar restriction, say to the range from 5 keV to 20 keV, even though their equipment tends to be opaque at energies below 5 keV and energies above 20 keV are hard to produce, hard to focus, and hard to handle. For designing a storage ring to be used as a highly selective instrument, it seems sensible, to me, to adopt a nominal energy in this "useful" range and to optimize the accelerator design for this energy.† Since $\lambda_\gamma = 1\mathring{A} \rightarrow E_\gamma = 12.4\,\mathrm{keV} \approx 12345\,\mathrm{eV}$ is easy to remember, and falls within the useful range, I suggest it be adopted as the "nominal" X-ray energy.

Since X-ray beams having good cleanliness can be produced only by undulators, it is essential to fold their inherent properties into the storage ring design. Apart from number of poles N_w, the key undulator/wiggler parameter is

$$K = \frac{\text{maximum deflection angle in wiggler}}{\text{half-angle of radiation cone}} \ . \tag{9.1}$$

Because the denominator has only a semi-quantitative significance, the same can be said for K, but it is the best we have. Because the deflection angle is proportional to the wiggler field B_0 and the radiated power is proportional to B^2, the radiated power is proportional to K^2. Until recent, third generation, light sources, this dependence has always caused K to be run to high values (making the ID a "wiggler") in order to obtain high flux (at great cost to cleanliness.)

For "clean" operation, one wants the lion's share of the power radiated from the undulator to be contained in the lowest order interference peak. In terms of wiggler period

† Since any electron storage ring produces X-rays at essentially all energies, even though the storage ring has been optimized for a particular energy, it can be used for X-ray experiments at any energy—though perhaps at unattractively low flux values.

λ_w, its position is given by

$$\lambda_\gamma = \frac{\lambda_w}{2\gamma^2}\left(1 + \frac{K^2}{2} + \text{angular fall-off factor}\right) \ . \tag{9.2}$$

The spectrum (integrated over "out-of-plane" angle) is shown in Fig. 6.3.2, from Jackson. As a compromise between too-dirty operation at too-great wavelength shift, on the one hand, and too-low flux, on the other, the value $K = 1$ will be adopted, (as is conventional.)

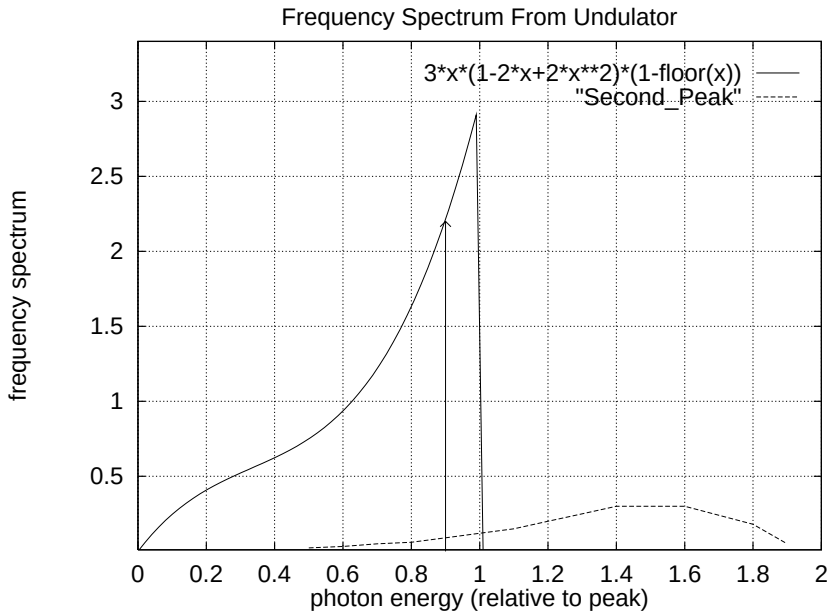


Figure 9.1: Undulator frequency spectrum plots copied (combined and somewhat garbled) from figures in Jackson, *Classical Electrodynamics*. The spectra plotted assume the "out-of-plane" angle has been integrated over. The $K = 0$ functional form can be read from the key in the upper right corner. Preceeding the apparatus by a horizontal slit that stopped out-of-plane angles greater than one third of the cone angle of the radiation, would allow only the narrow energy band above the arrow to be transmitted. This would be an extremely "clean" beam. The second interference maximum is shown corresponding to $K = 0.5$.

The gap height g of the undulator turns out to be the most important parameter in the whole facility. One visualizes a magnetic field pattern that is independent of x and y and has a perfect square-sawtooth z-dependence, but the reality is far different. No matter what contortions one goes through in undulator design, because of the need to reduce the period λ_w, the on-axis magnetic field always ends up essentially sinusoidal; $B(x, y = 0, z) \sim \cos(2\pi\, z/\lambda_w)$, (independent of x because the poles are wide.) Using

Maxwell's equations to match the off-axis behavior yields

$$B\,(x,y,z) = B_0\,\cosh\left(2\pi\,y/\lambda_w\right)\cos\left(2\pi\,z/\lambda_w\right)\ . \tag{9.3}$$

Though this formula is not strictly valid for $y \approx g/2$, it can be used to estimate the peak pole-tip field $B_{\rm p,max}$;

$$\frac{B_0}{B_{\rm p,max}} = \frac{1}{\cosh\left(\pi\,g/\lambda_w\right)}\ . \tag{9.4}$$

This function is plotted in Fig. 9.2. Though one might tolerate an appreciable reduction from the pole tip of the on-axis field, the presence of rapid dependence on y is undesirable, since it makes the radiated output depend sensitively on steering through the undulator. For this reason, a requirement such as $\lambda_w \geq 4g$ is conservative.

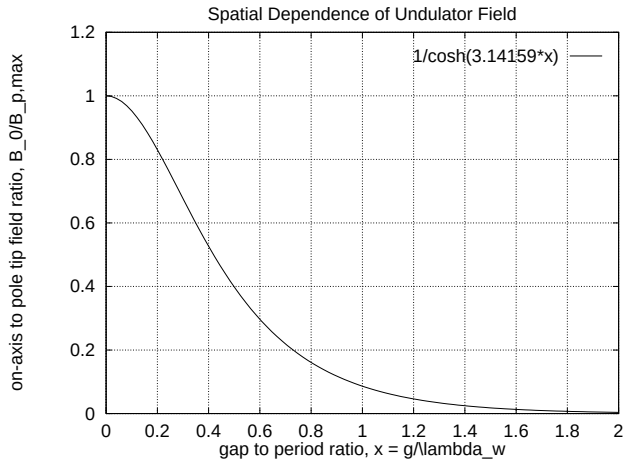


Figure 9.2: On-axis magnetic field relative to pole tip field as a fraction of gap-height/wiggler-period, g/λ_w.

Combining the values $K = 1$ and $\lambda_w = 4g$ with Eq. (9.2) yields $\lambda_\gamma = 3g/\gamma^2$ or

$$E_0 = \gamma\,mc^2 = mc^2\sqrt{\frac{3g}{\lambda_\gamma}} = \sqrt{g\,[\mathrm{mm}]}\,2.80\,\mathrm{GeV}\ . \tag{9.5}$$

This function is plotted in Fig. 9.3

Though this analysis may not seem very fundamental, I believe that it is, since the gap height g is certain to be limited by factors that have not yet been considered. Curiously, the usual $10\,\sigma$ or $15\,\sigma$ beam stay-clear requirement, traditional in the design of electron rings, is not one of the determining factors—for the shockingly low emittances to be studied in

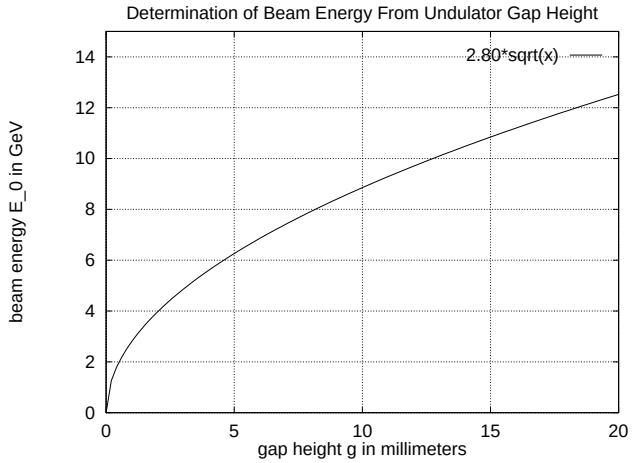


Figure 9.3: Having chosen λ_γ, K and λ_w/g, the beam energy is determined as a function of g, which is limited (from below) by beam steering and current dependent considerations.

the following sections, values of g as little as 1 mm are thinkable. But considerations of beam-wall interaction and beam steering accuracy are certain to force g to exceed a few millimeters—an exceedingly small "hoop to jump through". This leads to values of energy E_0 in the range 5-7 Gev, that seems to match CESR capabilities naturally. One knows, from experience with electron accelerators at Wilson Lab, that relaxing the need for small g by increasing E_0 beyond, say, 7 GeV, is unlikely to lessen the overall pain. It is probably no coincidence that the ESRF and APS rings lie in the 5-7 Gev range.†

9.1. The Trbojevic-Courant Minimum Emittance Cell

In a paper titled *Low Emittance Lattices for Electron Storage Rings Revisited*, D. Trbocevic and E. Courant give a lattice design which minimizes the emittance (under their assumed hypotheses.) Their cell configuration is shown in Fig. 9.1.1. There is a single dipole magnet, but it is convenient to regard the cell ends as occurring at the dipole center, as shown in the lower part of the figure.

† The "microwave undulator" or "back-scattered Compton" scattering source, described in CBN 00-8, may permit g to be increased somewhat. Since the beam tube could be small enough to block the 15 GHz microwaves required, the microwave beam could be provided by a superconducting cavity with modest power loss. Thanks to Maury Tigner for pointing this out.

The main results of the T-C paper can be distilled into the following few formulas. The horizontal emittance is given by

$$\epsilon_x = \frac{C_q \gamma^2}{J_x \rho} \mathcal{H}, \tag{9.1.1}$$

where $C_q = 3.84\ 10^{-13}$, $\gamma = E_0/(mc^2) = 0.978 \times 10^4$ for $E_0 = 5\,\mathrm{Gev}$, J_x, the partition number, can, for present purposes, be taken to be 1, and ρ is the bend radius in bending magnets. The factor $\mathcal{H}$, known as Sands's "curly H" is given, in terms of Twiss functions and dispersion functions, by

$$\mathcal{H} = \gamma_x \eta^2 + 2\alpha_x \eta\eta' + \beta_x \eta'^2. \tag{9.1.2}$$

The dependences on s of β and η within the dipole are[‡]

$$\beta = \beta_0 + \frac{s^2}{\beta_0}, \quad \alpha \equiv -\frac{\beta'}{2} = -\frac{s}{\beta_0}, \quad \eta = \eta_0 + \frac{s^2}{2\rho}, \tag{9.1.3}$$

and the mean value of curly-H is given by

$$\begin{aligned} \langle \mathcal{H} \rangle &= \frac{1}{L} \int_{-L/2}^{L/2} \frac{\eta^2 + (\eta\alpha + \eta'\beta)^2}{\beta}\, ds \\ &= \frac{\eta_0^2 - \theta\eta_0 L/12 + \theta^2 L^2/320}{\beta_0} + \frac{\theta_0^2 \beta_0}{12}, \end{aligned} \tag{9.1.4}$$

where L is the length of the bending magnet and $\theta = L/\rho$ is the bend angle in that magnet. The T-C mimimization procedure is to select values of β_0 and η_0 that minimize $\langle \mathcal{H} \rangle$ with L and θ fixed. The result is

$$\beta_0 = \frac{L}{2\sqrt{15}}, \quad \eta_0 = \frac{\theta L}{24}, \quad \langle \mathcal{H} \rangle = \frac{\sqrt{15}}{180} L\theta^2 = \frac{\sqrt{15}}{180} \rho\theta^3\ . \tag{9.1.5}$$

All that remains is to adjust lattice parameters to match these values of β_0 and η_0. These conditions fix the lattice design, modulo a certain amount of freedom in choosing element lengths and locations. Without providing details, Trbojevic and Courant exhibit one cell design that meets the requirements. In the next section I give a more detailed design prescription for design of individual cells and for the lattice as a whole.[†]

[‡] The formula for η was missing a factor of 2 in an earlier version, with a corresponding error in Eq. (9.2.16).

[†] In preparing this report I have profited from the work of Carl Smolenski, which he presented as both an oral and a written final assignment in Physics 689.

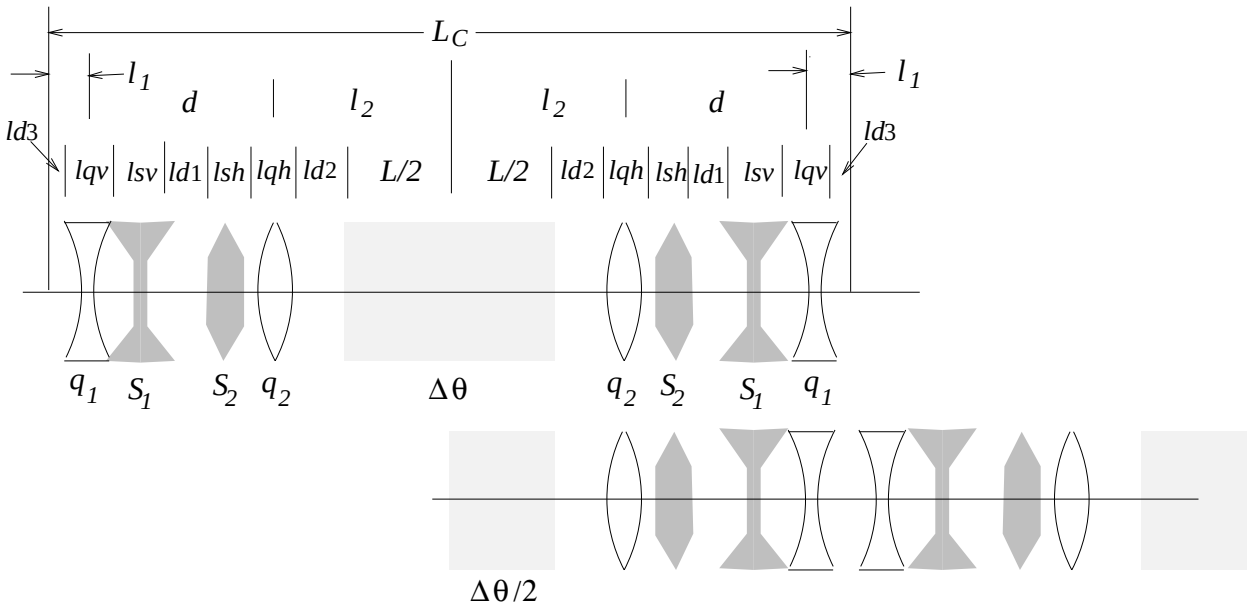


Figure 9.1.1: Thick and thin lens lengths and strengths for Trbojevic-Courant minimum emittance cell. The lower element ordering is convenient for relating design values of β_0 and η_0 to the Twiss functions of sectors consisting of repeated cells. The rectangular shaded region represents a bend magnet and the angular figures represent sextupoles. As the lenses are drawn, it is horizontal optics being illustrated. The parameter $ld3$, though tentatively set to zero, is available for flexibility.

9.2. Design Formulas For Trbojevic-Courant Lattice in CESR Tunnel

Though the T-C prescription certainly minimizes $\langle\mathcal{H}\rangle$ under the assumed hypotheses, this is not quite equivalent to minimizing ϵ_x in a practical lattice that would fit in an existing circular tunnel, such as CESR's, having circumference $\mathcal{C}$. To avoid worrying about sections needed for RF accelerating structures, wigglers or undulators, instrumentation, etc., let us introduce an "arc circumference" $\mathcal{C}_a \overset{\text{e.g.}}{=} 660\,\text{m}$, that leaves length $\mathcal{C} - \mathcal{C}_a$ available for miscellaneous elements. By far the most important design choice is that of cell length L_C or, equivalently, the number of cells N. These are related by

$$\theta = \frac{2\pi}{N}, \quad \text{and} \quad L_C = \frac{\mathcal{C}_a}{N}. \tag{9.2.1}$$

From this and the last of Eqs. (9.1.5) the emittance apparently varies as N^{-3} but, of course, considerations not included in the discussion so far will set a set a practical upper limit on N or, equivalently, a lower limit on L_C.

Another effect worth mentioning is that, because of the factor ρ in the denominator of Eq. (9.1.1), minimizing $\mathcal{H}$ is not equivalent to minimizing ϵ_x. In terms of quantities defined so far, ρ is given by†

$$\rho = \frac{\mathcal{C}_a}{2\pi}\frac{L}{L_C}, \tag{9.2.2}$$

To the extent that L fills most of the cell length L_C, minimizing $\mathcal{H}$ is equivalent to minimizing ϵ_x but, especially as L_C is reduced, an increasingly large fraction of the cell may be required for the quadrupoles and sextupoles that are needed to achieve the optimal values of β_0 and η_0. This leads to difficult dependencies, even in designing the linear optics, but the ultimate limit will be set by nonlinear effects that can only be studied by simulation using particle tracking. (It should not be surprising if the length required for sextupoles exceeds the quadrupole length allotment. In this case, a possibility worth considering is that one of the sextupole types be built into the dipoles, by shaping their poles. This would bring in nonlinear complications even in the analysis of the effects of quantum fluctuations in the synchrotron radiation, but the formalism for treating such effects is fairly well established.) Obviously a final design will have to be the result of iterating these steps. In this paper I make only a single, tentative, choice of element locations, and study the performance as a function of L_C. I do not attempt to iterate beyond this first cut design, and I ignore complications such as fringe fields or the circumference allotment required for magnet ends.

Yet another, and likely even more important, phenomenon militating against arbitrarily high values of N, is a beam current limitation. Even if the nonlinear optics can be designed to hold individual particles stably, for increasingly large values of N the magnet apertures must decrease correspondingly. At some point this is certain to limit the beam current, and hence the X-ray beam brightness. Similar comments apply to the Touschek effect.

For linear lattice design, following Trbojevic and Courant, though the bend region is crucial to evolution of the dispersion function, its "optical" effect on $\beta_x(s)$ is ignored. Also the quadrupoles are to be treated as thin lenses. These assumptions are retracted in the final design step using the TEAPOT program to study the dynamic acceptance of the lattice.

† Following Sands originally, as do T-C, this paper assumes an "isomagnetic" lattice in which the magnetic field has the same value everywhere except where it vanishes.

Trbojevic and Courant are correct in asserting that there are enough free parameters to allow the β_x and η_x-function values to be achieved. They do not, however, analyse the restriction that follows from the need for the vertical motion to be stable. In fact, the simplest solutions have both q_1 and q_2 horizontally focusing, which is obviously unsatisfactory. Hence, in seeking solutions, one should certainly impose the condition $q_1 q_2 < 0$. After a certain amount of trial and error, it appeared that the best place for the vertical focusing quad was as far from the dipole as possible. This suggested the virtue of pushing it to the very end of the cell, as shown in Fig. 9.1.1 for $ld3 = 0$. This has the further virtue of reducing the number of vertical focusing quads by a factor of two, by permitting elements from adjacent cells to be combined. The T-C conditions can be met in this case, while retaining vertical stability. There are indications though, that the "window" is very small. For example, the lattice is stable for q_1 being a thick lens, but it becomes unstable in the limit that the q_1 lenses on the boundaries between cells are combined into a shared thin lens. It remains to be seen whether this "flaky behavior" makes the nonlinear optics and dynamic acceptance unsatisfactory. It may be that the T-C lattice is "unnecessarily good" and that accepting somewhat higher emittance would lead to more robust optics. Though this direction seems, to me, to make sense, I have not been able to identify a promising way in which to relax the conditions.

Transfer matrix analysis of the thin lens doublet configuration shown in the left half of the upper lattice cell shown in Fig. 9.1.1 yields

$$\begin{aligned}
\begin{pmatrix} X_x & X_f \\ F_x & F_f \end{pmatrix} &= \begin{pmatrix} 1 & l_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -q_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -q_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & l_1 \\ 0 & 1 \end{pmatrix} \ , \\
\begin{pmatrix} Y_y & Y_g \\ G_y & G_g \end{pmatrix} &= \begin{pmatrix} 1 & l_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ q_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ q_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & l_1 \\ 0 & 1 \end{pmatrix} \ .
\end{aligned} \tag{9.2.3}$$

Notice that *positive* q values correspond to *horizontal focusing*, so $q_1 < 0$ and $q_2 > 0$. Completing the multiplications yields, in the horizontal plane,

$$\begin{aligned}
X_x =& 1 - (d + l_2)\, q_1 - l_2 q_2 + d l_2 q_1 q_2 \\
X_f =& l_1 + d + l_2 - l_1 (d + l_2)\, q_1 - (l_1 + d)\, l_2 q_2 + l_1 d l_2 q_1 q_2 \\
F_x =& - q_1 - q_2 + d q_1 q_2 \\
F_f =& 1 - l_1 q_1 - (l_1 + d)\, q_2 + l_1 d q_1 q_2,
\end{aligned} \tag{9.2.4}$$

and in the vertical plane, all negative signs are replaced by positive signs;

$$\begin{aligned} Y_y =& 1 + (d + l_2)\, q_1 + l_2 q_2 + d l_2 q_1 q_2 \\ Y_g =& l_1 + d + l_2 + l_1\,(d + l_2)\, q_1 + (l_1 + d)\, l_2 q_2 + l_1 d l_2 q_1 q_2 \\ G_y =& q_1 + q_2 + d q_1 q_2 \\ G_g =& 1 + l_1 q_1 + (l_1 + d)\, q_2 + l_1 d q_1 q_2. \end{aligned} \tag{9.2.5}$$

Letting $\mathbf{X}$ be the 4×4 transfer matrix of the left half cell in the upper part of Fig. 9.1.1, the input and output coordinates of a general trajectory are related by the usual relations,

$$\mathbf{x}_{\rm out} = \mathbf{X}\mathbf{x}_{\rm in}, \text{ and } \mathbf{x}_{\rm in} = \mathbf{X}^{-1}\mathbf{x}_{\rm out}. \tag{9.2.6}$$

The second half of the cell consists of the same elements, but encountered in the opposite order. Let $\breve{\mathbf{X}}$ stand for the transfer matrix of this sequence. Symmetry under this reflection is closely related to time reversal invariance that relates forward and backward evolving trajectories. To express this mathematically, introduce a matrix $\sigma_{\mathbf{3}}$ that reverses the slopes,

$$\begin{pmatrix} x \\ -x' \\ y \\ -y' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ x' \\ y \\ y' \end{pmatrix} \equiv \sigma_{\mathbf{3}}\mathbf{x}, \tag{9.2.7}$$

From the vector $\mathbf{x}_{\rm out}$ appearing in Eq. (9.2.6), form the vector $\sigma_{\mathbf{3}}\mathbf{x}_{\rm out}$ and take it as input to the reversed-order half-cell. By symmetry, after propagating through the half-cell, one should recover $\sigma_{\mathbf{3}}\mathbf{x}_{\rm in}$, where $\mathbf{x}_{\rm in}$ appears in Eq. (9.2.6). That is

$$\sigma_{\mathbf{3}}\mathbf{X}^{-1}\mathbf{x}_{\rm out} = \sigma_{\mathbf{3}}\mathbf{x}_{\rm in} = \breve{\mathbf{X}}\sigma_{\mathbf{3}}\mathbf{x}_{\rm out}, \tag{9.2.8}$$

and it follows from this that

$$\breve{\mathbf{X}} = \sigma_{\mathbf{3}}\mathbf{X}^{-1}\sigma_{\mathbf{3}}. \tag{9.2.9}$$

Because the determinants are equal to 1, the inverse matrices are given by

$$\begin{pmatrix} X_x & X_f \\ F_x & F_f \end{pmatrix}^{-1} = \begin{pmatrix} F_f & -X_f \\ -F_x & X_x \end{pmatrix}, \quad \begin{pmatrix} Y_y & Y_g \\ G_y & G_g \end{pmatrix}^{-1} = \begin{pmatrix} G_g & -Y_g \\ -G_y & Y_y \end{pmatrix}. \tag{9.2.10}$$

As mentioned before, it is most convenient to select the center of the dipole magnet as origin. The transfer matrix $\mathbf{M}$, through the full cell from this point, is given by

$$\mathbf{M} = \mathbf{X}\breve{\mathbf{X}} = \begin{pmatrix} X_x F_f + X_f F_x & 2X_x X_f & 0 & 0 \\ 2F_x F_f & X_x F_f + X_f F_x & 0 & 0 \\ 0 & 0 & Y_y G_g + Y_g G_y & 2Y_y Y_g \\ 0 & 0 & 2G_y G_g & Y_y G_g + Y_g G_y \end{pmatrix}. \tag{9.2.11}$$

If the cell is part of a periodic lattice, its transfer matrix can be written in the form

$$\mathbf{M} = \begin{pmatrix} \cos\mu_x & \beta_x \sin\mu_x & 0 & 0 \\ -\sin\mu_x/\beta_x & \cos\mu_x & 0 & 0 \\ 0 & 0 & \cos\mu_y & \beta_y \sin\mu_y \\ 0 & 0 & -\sin\mu_y/\beta_y & \cos\mu_y \end{pmatrix}. \tag{9.2.12}$$

Note that the vanishing of the α's is consistent with Eq. (9.2.11) because the 11,22 and 33,44 element are pairwise equal. The reason for this is that the full cell is invariant to element reversal, so Eq. (9.2.9) implies that $\mathbf{M} = \sigma_{\mathbf{3}}\mathbf{M}^{-1}\sigma_{\mathbf{3}}$. We have therefore obtained formulas for the Twiss parameters at the cell ends;

$$\begin{aligned} &\cos\mu_x = X_x F_f + X_f F_x, \quad \sin^2\mu_x = -4X_x X_f F_x F_f, \quad \beta_x^2 = -\frac{X_x X_f}{F_x F_f}, \\ &\cos\mu_y = Y_y G_g + Y_g G_y, \quad \sin^2\mu_y = -4Y_y Y_g G_y G_g, \quad \beta_y^2 = -\frac{Y_y Y_g}{G_y G_g}. \end{aligned} \tag{9.2.13}$$

Because of the multiple-valued nature of inverse trigonometric functions, a certain amount of care is required in extracting the Twiss parameters from these relation. For the numerical values given in the table:

$$\begin{aligned} Q_x &= 1 - \frac{\cos^{-1}\left(X_x F_f + X_f F_x\right)}{2\pi} = 1 - \frac{\sin^{-1}\left(\sqrt{-4X_x X_f F_x F_f}\right)}{2\pi}, \\ Q_y &= \frac{\cos^{-1}\left(Y_y G_g + Y_g G_y\right)}{2\pi} = 0.5 - \frac{\sin^{-1}\left(-\sqrt{-4Y_y Y_g G_y G_g}\right)}{2\pi}. \end{aligned} \tag{9.2.14}$$

The T-C requirement on β_0 can be expressed by combining this result with the first of Eqs. (9.1.5) to yield the significant design formula;

$$\sqrt{-\frac{X_x X_f}{F_x F_f}} = \frac{L}{2\sqrt{15}}. \tag{9.2.15}$$

If we regard all lengths as fixed, then this equation provides one of the two conditions needed to fix the adjustable parameters q_1 and q_2.

Except within the dipole region, the dispersion function has the same s-dependence as does a particle trajectory. At the dipole end, using Eq. (9.1.3), we have $\eta(L/2) = \eta_0 + L\theta/4$ and $\eta'(L/2) = \theta$. Since $\eta(s)$ must have even symmetry about the cell center, and $\eta'(s)$ is odd, we must have

$$\begin{pmatrix} \eta_0 + L\theta/8 \\ -\theta/2 \end{pmatrix} = \mathbf{M}_{l_2 \to l_2 - L/2} \begin{pmatrix} \eta_0 + L\theta/8 \\ \theta/2 \end{pmatrix}, \tag{9.2.16}$$

because the transfer matrix through the bend-less section differs from the full transfer matrix only by the reduced entrance and exit distances indicated by the subscript on $\mathbf{M}$. This equation, along with Eq. (9.2.15), fix both q_1 and q_2. (The redundancy resulting because Eq. (9.2.16) contains two equations provides a consistency check.)

Numerical values of the length and strength parameters of a first cut design are shown in Table 9.2.1. The parameter of greatest interest, the horizontal emittance ϵ_x, is simple to derive, using Eqs. (9.1.1) and (9.1.5);

$$\epsilon_x = \frac{C_q \gamma^2}{J_x} \frac{\sqrt{15}}{180} \left(\frac{2\pi}{N}\right)^3 \quad \left(\stackrel{\text{e.g.}}{=} 1.96 \times 10^{-10} \left(\frac{100}{N}\right)^3 \text{ m-rad} \right) \ . \tag{9.2.17}$$

To obtain the results in Table 9.2.1, after having fixed all length parameters, Eqs. (9.2.15) and (9.2.16), constrained by the requirement $q_1 q_2 > 0$, were solved using MAPLE. Because they are polynomial equations, MAPLE is able to solve the equations "exactly" without using numerical methods. This is not just apple-polishing since the desired solution can be picked automatically, circumventing the need to manually scrutinize all of the numerous solutions, some of which are complex, others unstable vertically.

Since E, B and ρ values are the same for all cases, so are the radiation quantities,

$$\begin{aligned} U_0 &= \text{energy radiated per turn} = 0.885 \times 10^{-4}\,\text{m/GeV}^3\, \frac{5^4}{42} = 1.32\,\text{MeV} \ , \\ u_c &= \text{critical energy} = 6.6\,\text{keV} \ , \\ \sigma_\delta &= \text{fractional energy spread} = 0.00066 \ , \\ P_{\text{tot}} &= \text{total radiated power} = 1.32\ (\text{MW/A})\ I_{\text{tot}} \ . \end{aligned} \tag{9.2.18}$$

Table 9.2.1: Sample Numerical Values

parameter	unit	x-value	y-value
E_0	GeV	5.0	
$\mathcal{C}_a$	m	660.0	
ρ	m	36.765	
B	T	0.4536	
L_C	m	6.0	
N		110	
L	m	2.1	
θ	rad	.05712	
lqh	m	0.33	
lqv	m	0.33	
lsh	m	0.48	
lsv	m	0.30	
$ld1$	m	0.33	
$ld2$	m	0.18	
$ld3$	m	0.0	
l_1	m	0.165	
l_2	m	1.395	
d	m	1.44	
q_1,q_2	1/m	-0.5104	1.2511
β_x,β_y	m	0.2711	0.4097
η_0	m	.00500	
Q_x,Q_y	one cell	0.7902	0.4334
Q_x,Q_y	total	86.92	47.68
$\langle\mathcal{H}\rangle$	10^{-4} m	1.474	
ϵ_x	10^{-10} m	1.474	
Thick Lens Analysis (param.'s as above)			
Q_x,Q_y	total	77.80	31.20
S_1,S_2	$1/m^2$	-27.15	35.02
η_0	m	0.00771	
β_x,β_y	m	0.3345	1.5348
q/S	mm	18.8	35.7

9.3. Thick Lens and Nonlinear Optics

Thick lens optical functions are shown in Fig. 9.3.1 and Fig. 9.3.2 and also in the form of numbers at the bottom of the table. The η_0 thick lens values are quite close to the design values. The β_x values are less good, typically too large by 50%, so the emittances would be increased by about the same factor. The values of β_y are even worse, too great by roughly a factor of two.

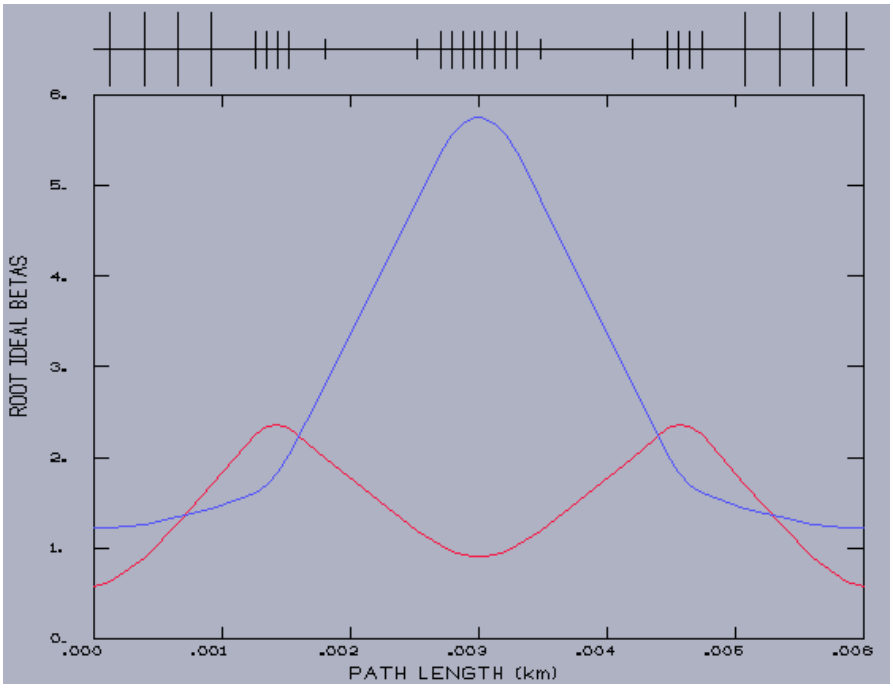


Figure 9.3.1: Functions $\sqrt{\beta_x(s)}$ and $\sqrt{\beta_y(s)}$ for the $L_C = 6\,\mathrm{m}$ case.

Anyone who has struggled to make CESR function properly with tunes greater than 10 is sure to doubt the feasability of running with $Q_x = 135$. The problem is that the strength of sextupoles (needed to compensate chromaticity) increase rapidly with tune. Typical (electrical) quad strengths (inverse focal lengths) in CESR are $q\tilde{<}0.25/\mathrm{m}$ for quads of length 0.6 m, bore diameter 80 m. Gradients in the existing permanent magnet quads are perhaps twice as great? For the proposed accelerator the bores have to be, and can be, reduced by perhaps a factor of 5. This should make the quad strengths achievable, at least for the $L_C \geq 6\,\mathrm{m}$. Sextupole strengths in CESR are perhaps $S \approx 0.3/\mathrm{m}^2$. Even

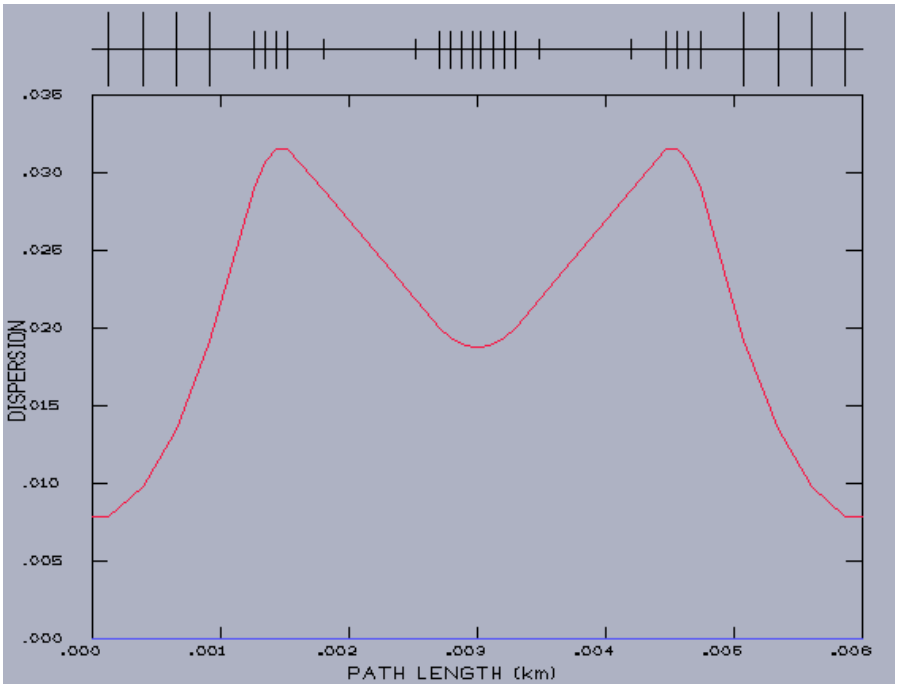


Figure 9.3.2: Dispersion function $\eta(s)$ for the $L_C = 6\,\mathrm{m}$ case.

though sextupole strengths vary inversely with bore diameter squared, it may be sextupole strengths that limit the cell length.

A dimensionally consistent (over-)estimate of dynamic aperture is q/S which, for CESR would yield about 0.5 m. To estimate the dynamic acceptance of the proposed ring, sextupoles S_1 and S_2 have been adjusted to set both chromaticities to zero. The resulting sextupole strengths are shown near the bottom of the table.† Values of the ratio q/S are also listed. From these numbers we have to expect dynamic apertures smaller than a centimeter. (As mentioned before, superimposing sextupole on the dipole may give an improvement.) In any case, when measured in units of beam σ's, the dynamic acceptance may be quite satisfactory because of the exceedingly low value of ϵ_x, since the needed aperture has a factor scaling as $\sqrt{\epsilon_x}$. (Possible injection difficulties are being ignored.)

The results of numerical tracking are shown in Fig. 9.3.3 for on-momentum, and Fig. 9.3.5 and Fig. 9.3.4 for off-momentum. What is plotted are boundaries, in the x-y plane, inside which the motion is stable (for at least 256 turns) and outside which it is

† One must beware a conventional factor $n! = 2$ that may or may not enter the definition of sextupole strength.

unstable. Since, in most cases, motion is stable out to almost $50\,\sigma$, all cases appear to be acceptable. An example may be helpful in interpreting these graphs: consider the barely stable $\Delta p/p = 0$, $L_C = 6\,\mathrm{m}$ point, with coordinates $x = 0$, $y/\sqrt{\beta_y} = 0.77 \times 10^{-3}\,\sqrt{\mathrm{m}}$. From Table 9.2.1 one has $\beta_y = 4.1\,\mathrm{m}$, with the result that $y = 1.5\,\mathrm{mm}$. This point sets a vertical limit, at the dipole center, for a beam with large vertical motion (present, for example, because of coupling.) However, one is more concerned with the aperture at the vertical focusing quads where $\beta_y \approx 25\,\mathrm{m}$, and $y \approx 4\,\mathrm{mm}$. Even ignoring the factor of two emittance reduction accompanying full coupling, this is roughly a "$45\,\sigma$" point. For a beam chamber with gap of, say, 8 mm, the physical and magnetic apertures would be comparable, with both comfortably large relative to the equilibrium beam size.

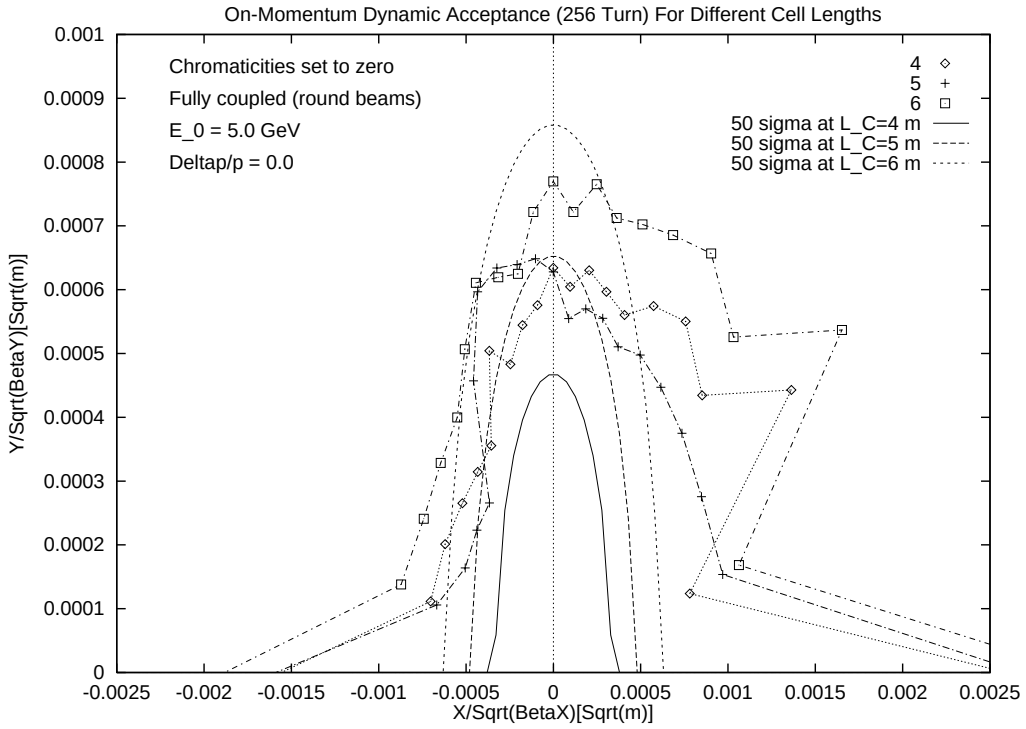


Figure 9.3.3: On-momentum, $\Delta p/p = 0$, transverse acceptance for $L_C = 4, 5$ and $6\,\mathrm{m}$. With all particles being launched from a dipole center, with $x' = y' = 0$, points are plotted at those points closest to the origin on rays emanating from the origin. Being plotted at $x/\sqrt{\beta_x}, y/\sqrt{\beta_y}$, they would lie on an ellipse $x^2/\beta_x + y^2/\beta_y = N_\sigma^2 \epsilon_x$ if the aperture limits were "isotropic". (There is no particular reason for this to be true.) Values for ϵ_x are taken from Table 9.2.1 and $N_\sigma = 50$. Since $\beta_x \approx 1\,\mathrm{m}$ the axes are not very different from transverse displacements measured in meters.

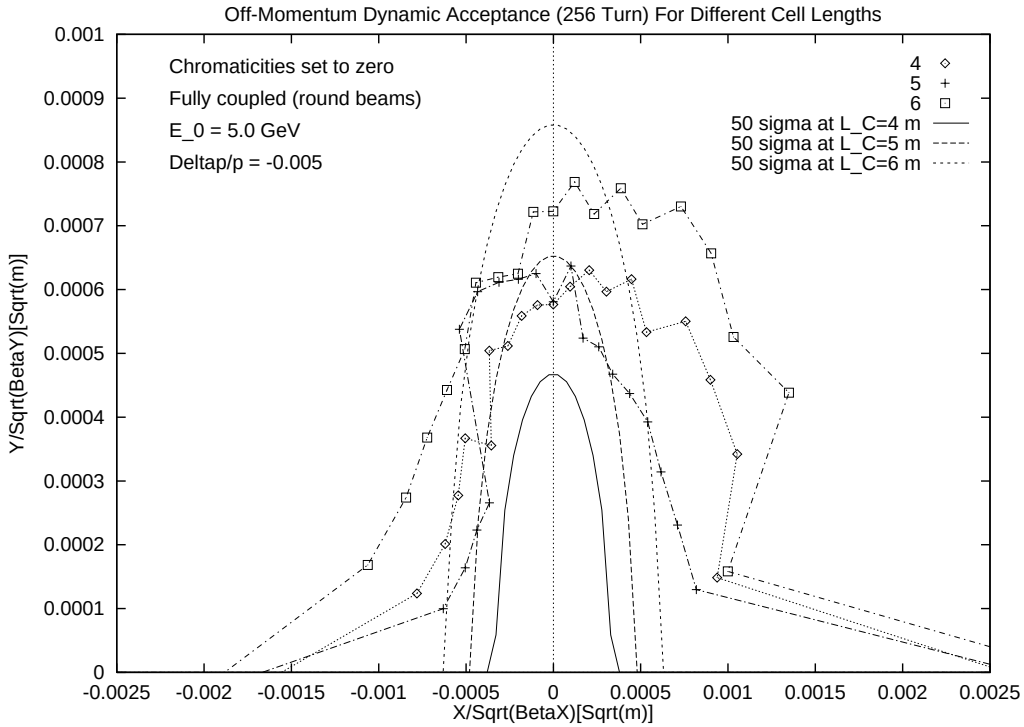


Figure 9.3.4: Off-momentum $\Delta p/p = -0.005$, transverse acceptance for $L_C = 4, 5$ and 6 m.

9.4. Dispersion Suppression

Within a pure dipole magnet the equation satisfied by the dispersion function is

$$\frac{d^2\eta}{ds^2} = \frac{1}{\rho} \quad \text{and hence} \quad \Delta\eta' = \frac{\Delta s}{\rho} \ . \tag{9.4.1}$$

In one T-C cell η' increases from $-\theta/2$ to $\theta/2$ in a magnet of length L. Correspondingly, η' would increase from 0 to $\theta/2$ in a magnet of length $L/2$. One therefore envisages a dispersion matching cell that is identical to the T-C cell, with the exception that the magnet is only half-length, and is also appropriately located, with its entrance position at $s = \tilde{s}$ so that the dispersion function $\tilde{\eta}(L/2)$ (in the special cell) matches $\eta(L/2)$ (of the regular cell.) The condition for match is

$$\frac{(L/2)^2}{2\rho} + \frac{\theta}{2}\left(\frac{L}{2} - \tilde{s} - \frac{L}{2}\right) = \eta_0 + \frac{L\theta}{8} \ . \tag{9.4.2}$$

In this equation one could substitute for η_0 using Eq. (9.1.5) but, since the thick lens and thin lens optics do not agree very well, we leave the condition as

$$\tilde{s} = -\frac{2\eta_0}{\theta} \ , \tag{9.4.3}$$

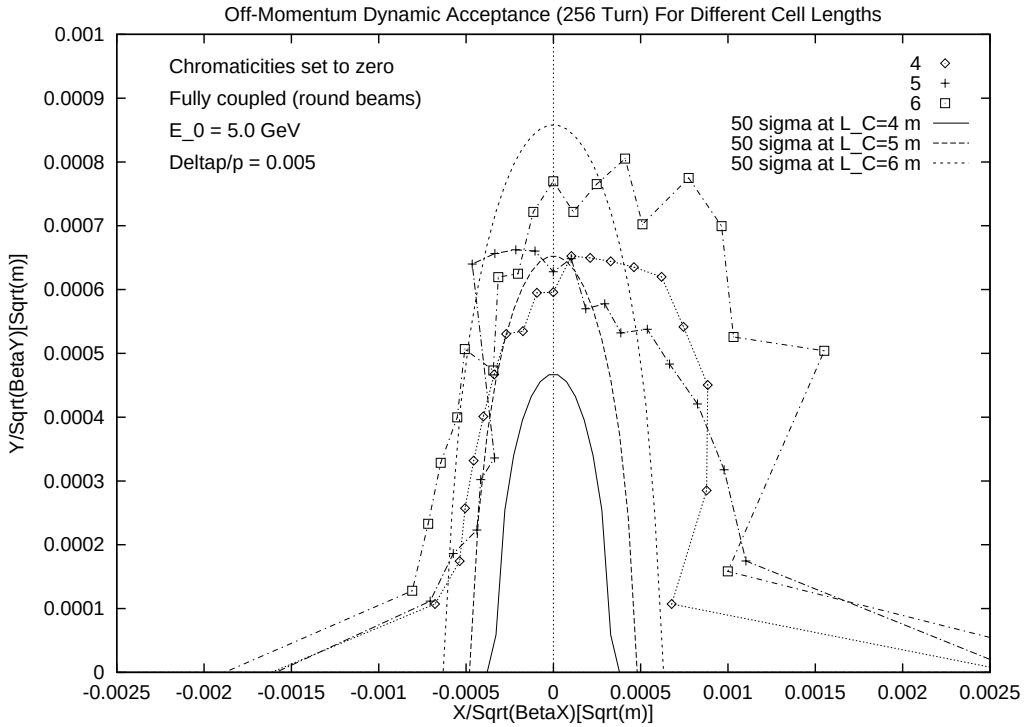


Figure 9.3.5: Off-momentum $\Delta p/p = 0.005$, transverse acceptance for $L_C = 4, 5$ and 6 m.

and anticipate fixing $\tilde{s}$ empirically, in the thick lens optics, to give perfect dispersion suppression. The optics for a sequence of six T-C cells, with dispersion suppressed in the central two, is shown in Fig. 8.3.1.

9.5. Touschek-Dominated Operation

Though the Touschek effect is well understood, it is not well understood by me. I have copied formulas more-or-less mindlessly, evaluated them, and plotted the results in Fig. 9.5.1. The formula labeled "Bruck" (theory due to Haissinski) comes from H. Bruck, *Accélérateurs Circulaires de Particules.* The formula labeled "LeDuff" comes from C. Bocchetta, *Lifetime and Beam Quality* in CERN 98-04. The theories apparently differ primarily in their treatments of dispersion. Since the LeDuff theory is more recent, it may be more reliable. It will be necessary to "pull out all the stops" to ameliorate the Touschek effect. The bunch length has been taken as 0.1 m, which would only be possible with a third-harmonic cavity (because large overvoltage of the fundamental RF is also required.) A peripheral benefit of bunch lengthening is that beam-chamber higher mode loss is likely

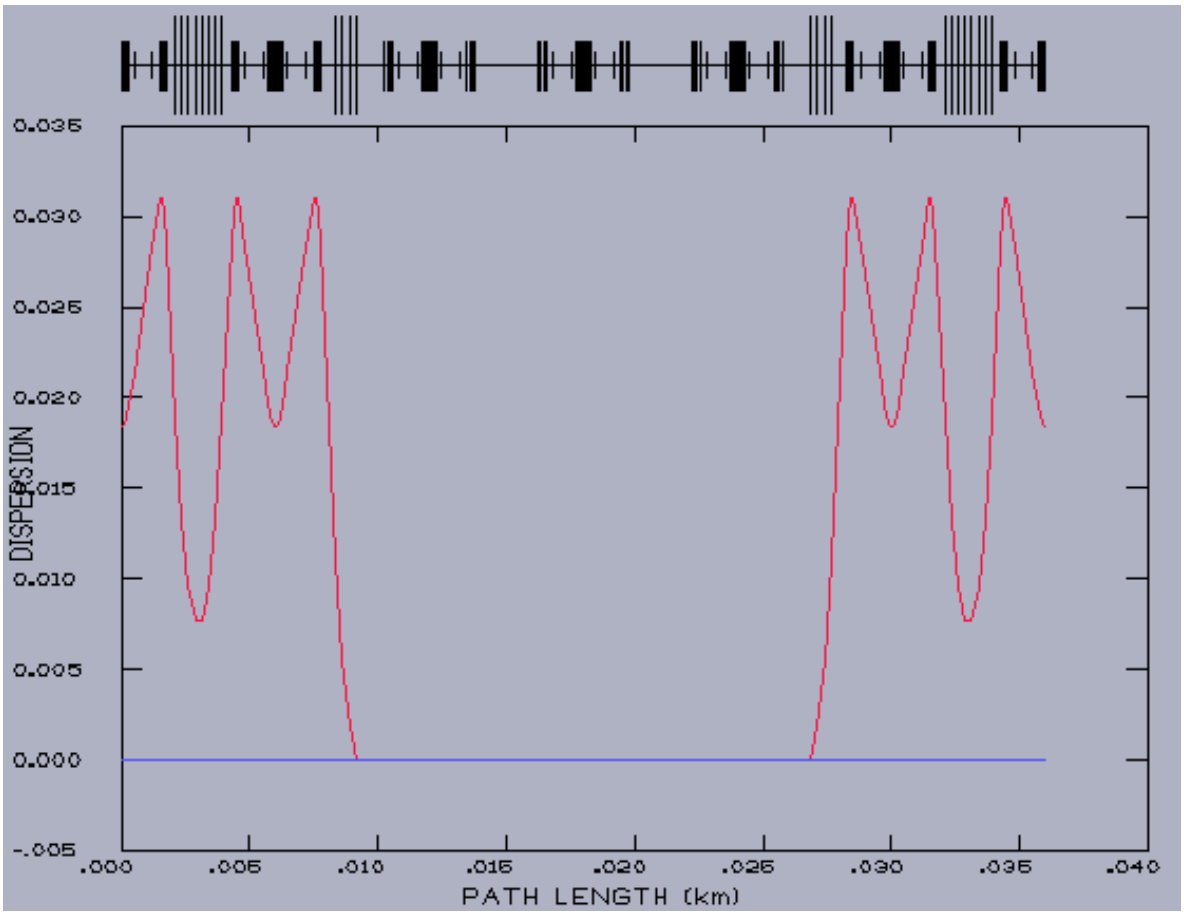


Figure 9.4.1: Six consecutive T-C cells, with the second and fifth cells modified to provide zero dispersion in the third and fourth cells.

to be unimportant, even with the chamber bore being tiny.† It is assumed that all buckets are filled, $N_B = 1280$, which will be possible only with feedback (even though the total current is modest.) Increasing the energy, say from 5 GeV to 7 GeV would also give a big improvement. The lifetime increases proportional to γ^4, or to an even higher power of γ, depending on assumptions concerning the beam σ's. [Y. Miyahara, IEEE Trans. Nucl. Sc., **32**, 3821 (1985).]

† Beam-chamber higher mode loss may be a significant problem in the ERL scheme, where short bunches are essential.

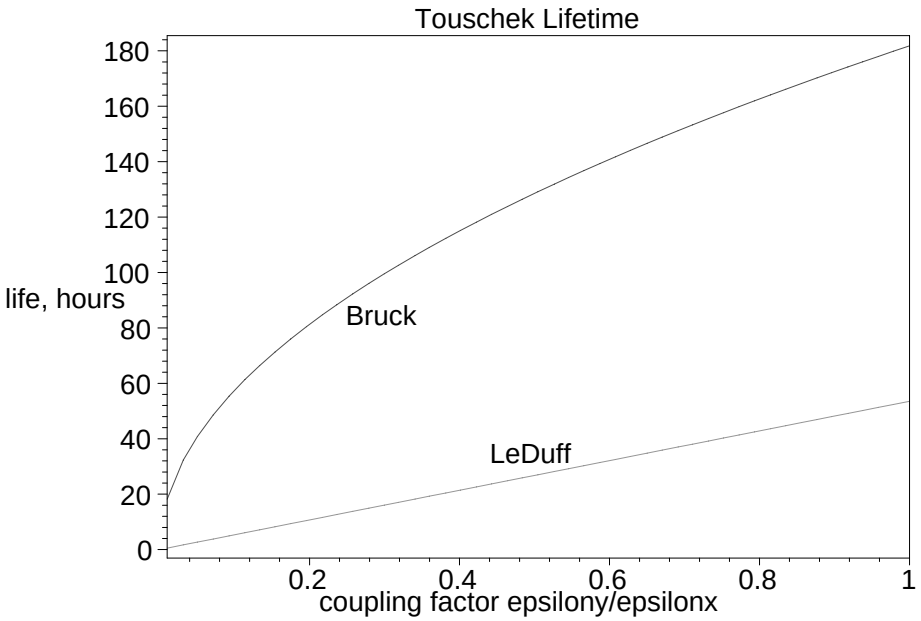


Figure 9.5.1: Beam lifetime caused by the Touschek effect. $L_C = 6\,\mathrm{m}$, $\epsilon_x = 4.0 \times 10^{-10}\,\mathrm{m}$, $E_0 = 5.0\,\mathrm{GeV}$, $\sigma_z = 0.1\,\mathrm{m}$, $N_B = 1280$.

9.6. Conclusions

From the calculations that have been performed, including Fig. 9.5.1, the beam conditions listed in the caption to that figure, with a coupling factor of perhaps 0.1, seem likely to be achievable, provided the hard probems that have been identified can be solved. Compared to the energy recovery linac (ERL) source that is currently under study, the beam current is the same, and so is the area $\epsilon_x \epsilon_y$. One therefore expects the achievable X-ray beam brilliance of this machine to be comparable to that of the ERL source. The present design (it seems to me) relies only on known, non-contraversial technology. It will be cheap to build and cheap to operate, and will be "world class" in performance. By specializing to a limited, but prime, spectral range, it should be able to out-perform existing machines in flux, more so in brilliance, and most of all in cleanliness. Since other sources, such as ESRF and APS, are (well)-designed for general purpose use, it is not realistic to suppose that a new (and inexpensive) ring can be superior in all respects, but it can be superior in some.

Still, there can really be no conclusions as yet, since this paper is so preliminary. It will be extremely difficult to decide how small the chamber height g can be made and,

without that, the choice of maximum beam energy cannot be made. (The same comment applies to ERL.) The reviews have been mixed on the performance of the all-permanent-magnet ring at Fermilab but, for the small bore required, such a design is attractive, unless giving up beam energy adjustability is too great a sacrifice. (Handing the responsibility for wavelength tuning to the undulators greatly complicates and compromises their design and performance.) The brilliance that can be achieved will depend on the design of undulators and their influence on the machine.

Chapter 10.
Some Ingredients of FEL Theory

Abstract: The free electron laser depends on many ingredients, some of which have been discussed already in the course. I am not qualified (at least not yet) to give a comprehensive yet simple overall explanation of how the FEL works, but in this packet I include discussion of some of the essential ingredients that have not yet been covered, and reminders of things that have.

The essential ingredients include:

- Beam parameters, σ_x, $\sigma_{x'}$, σ_y, $\sigma_{y'}$, σ_s, σ_δ. See packet 5.
- Undulator radiation. See Finkelstein lecture notes and packet 3.
- Mirror properties and limitations. See Gruner lecture notes.
- Interpretation of undulator radiation as Compton scattering.
- Spontaneous and stimulated photon emission, and photon absorbtion.
- Applicability condition for semi-classical treatment of FEL.
- Optical resonator.
- Density of states and stored energy in laser resonator.
- Trajectory and energy dependence of electron in electromagnetic wave.
- Optical klystron

10.1. The Free Electron Laser (FEL)

The general operation of the free electron laser can be inferred from Fig. 10.1.1. A circulating bunch of electrons is arranged to pass through the electromagnetic wave in an optical resonator in such a way that energy is extracted from the electron beam in the form of radiation. Some of the radiation passes through the partial mirror for its intended use. The rest replenishes the resonator energy. For the device to be useful the energy radiated per passage has to be greater than the energy lost in the resonator between passages.

The radiation can be analysed (semi-classically using conservation of energy) as the radiation that accompanies deceleration of the electrons by the electromagnetic wave, or (quantum mechanically) as Compton scattering of the virtual photons of the undulator, as

stimulated by the pre-existing photons trapped in the optical resonator. Historically the quantum picture came first, but the classical analysis is far more elementary and, as far as I know always valid in practical devices.

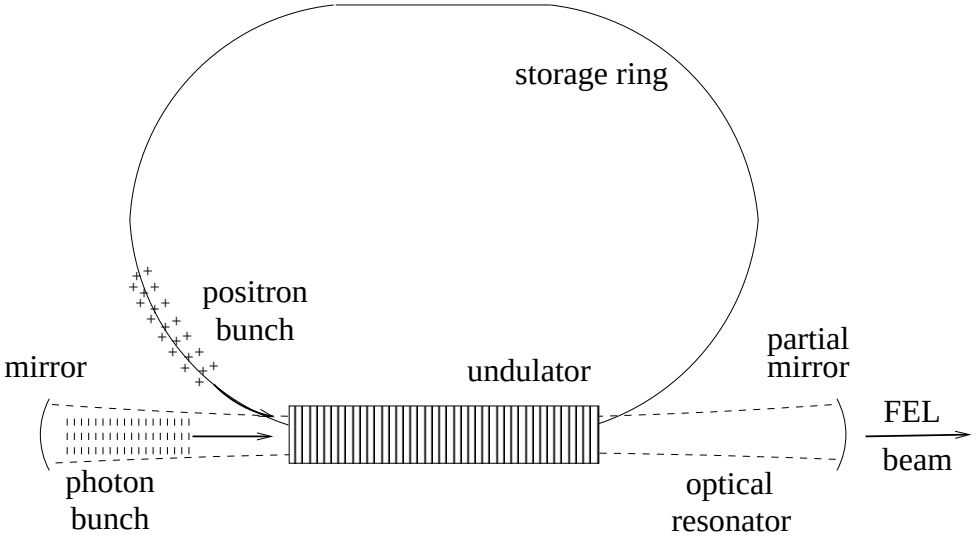


Figure 10.1.1: One long straight section of a racetrack-shaped storage ring is shared with an optical resonator and a magnetic wiggler to form a free electron laser.

The so-called "optical klystron" uses two undulators, separated by a dispersive region. The radiation from the first wiggler introduces energy oscillation along the electron bunch, which the dispersive section converts to a kind of "population inversion" in the form of particle density bunching. This bunching can be synchronized into optimal phase relationship with the pre-existing radiation, to yield greater radiation gain.

10.2. Interpretation of Undulator Radiation as Compton Scattering

For a horizontal-bending undulator aligned with the z-axis, the only non-vanishing electric or magnetic field component is $B_y = B_0 \cos(k_w z)$. With electrons propagating at speed v along the positive z-axis, it is useful to transform both the wiggler field and the electron coordinate into the rest frame of the electron. The result is

$$\mathbf{E}' = -\gamma\, v\, B_y \cos\left(k_w \gamma \left(z' + v t'\right)\right) \hat{\mathbf{x}}, \quad \text{and} \quad \mathbf{B}' = -\frac{\mathbf{v}}{v^2} \times \mathbf{E}' \ . \tag{10.2.1}$$

This is very nearly the same as the field of a plane polarized plane wave, propagating in free space. In fact, in the limit $v \to c$ the correspondance becomes exact. Making the replacement $v = c$ yields what is known as the Weizsäcker-Williams approximation.

The wave just derived is said to be made up of "virtual" photons, and these photons can Compton scatter off the electrons. The (magnitude of) the rest energy of one of these virtual photons (calculated most easily in the laboratory frame) is given by

$$|m_\gamma c^2| = \left| \sqrt{(\hbar\omega)^2 - (\hbar k_w)^2 c^2} \right| = \hbar k_w c \ , \tag{10.2.2}$$

since the frequency in the laboratory vanishes.

Next consider the situation in the rest frame of the electrons. In this frame the photon energy is $E'_\gamma = \hbar k_w \gamma v$, since $k_w \gamma v$ is the (frequency) factor multiplying t' in the argument of the cosine factor in Eq. (10.2.1). If this energy is small compared to the electron rest energy,

$$E'_\gamma = \hbar k_w \gamma v << mc^2 \ , \tag{10.2.3}$$

(as will always be true for cases of interest to us) the incident and scattered photon energies are the same. It will be valid to neglect the the virtual photon mass (calculated in Eq. (10.2.2)) if it is small compared to this energy;

$$\hbar k_w c \ \stackrel{?}{<<} \hbar k_w \gamma v \ , \tag{10.2.4}$$

which reduces to $\gamma >> 1$, and will be abundantly true in practice.

Condition (10.2.3) is also the condition for the validity of treating Compton scattering as Thomson scattering, for which the total cross section is

$$\sigma = \frac{8\pi}{3} r_e^2 = 0.665 \times 10^{-28}\,\mathrm{m}^2 \ . \tag{10.2.5}$$

where $r_e = 2.81784 \times 10^{-15}\,\mathrm{m}$. (See Eq. (2.21).) Though this cross section was calculated in the electron rest frame, the lab frame value is the same.

To calculate the radiation pattern in the laboratory it is necessary to write the angular distribution in the rest system of the electron, and then to transform it into the laboratory system. Though the scattered photons are mono-energetic in the electron rest system, this will no longer be true in the laboratory system. We can write down the maximum lab

energy, since it corresponds to pure back-scattering. The result of Lorentz transforming the photon four-vector, $(\hbar\gamma k_w c, 0, 0, \hbar\gamma k_w)$, back to the lab frame, is a photon of energy

$$E_\gamma = \left(1 + \frac{v}{c}\right) \gamma^2 \, \hbar k_w \ . \tag{10.2.6}$$

Since $v \approx c$, the back-scattered wave length is less than the wiggler period λ_w by the factor $2\gamma^2$. This agrees with the undulator peak calculated using classical electrodynamics.

There is still quite a bit of work to do to show that the radiation calculated this way is identical to that calculated previously using classical electrodynamics. But we have gone far enough to be able to introduce the original FEL motivation. When the cross section just calculated is calculated quantum mechanically (and called the Compton cross section) the emission probability into each state is enhanced proportionally to the pre-existing population of the state. (This "fundamental source of laser action" is discussed further in the next section.) This is the role of the photons trapped in the optical resonator. The stimulated emission enhances the intensity of this radiation. Unfortunately, there is a competing process, namely absorption, that reduces the intensity. To a first approximation (or rather, on the average) these intensity contributions are equal and opposite, and there is no appreciable intensity enhancement.

The trick to enhancing the emission relative to the absorption is to make the electrons have energy less than the "resonant energy" (the energy for which the laser wavelength is $\lambda_w/(2\gamma^2)$ since the absorption is then weakened relative to the emission. However, too great an offset, leads to reduced radiation as the electron loses synch with the wiggler—an effect that becomes increasingly important as N_w is increased. The combination of these factors leads to the "gain" dependence shown (arbitrary units) in Fig. 10.2.1.

10.3. Absorption, Spontaneous Emission, and Stimulated Emission of Photons

According to statistical mechanics, there is an important relation governing emission and absorption of radiation. Let the rate of spontaneous emission of a photon by an electron be given by A_{21} and assume that the rates of absorption and stimulated emission are both proportional to $\rho(\nu)$, the photon energy density per unit frequency range at the position of the electron. Transition rates for emission of a photon, W_{21}, or absorption a photon,

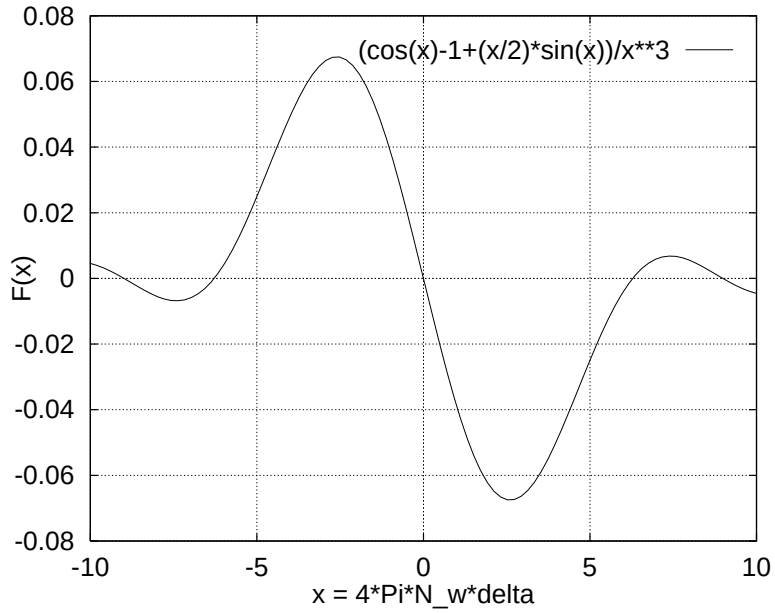


Figure 10.2.1: Dependence of intensity gain of FEL as a function of momentum offset δ, relative to the central momentum, in an undulator with N_w periods. The vertical scale is arbitrary. The horizontal variable is $4\pi N_w \delta$. The analytical formula is shown in the caption.

W_{12}, can then be expressed as

$$\begin{aligned} W_{21} &= B_{21}\rho\left(\nu\right) + A_{21}, \\ W_{12} &= B_{12}\rho\left(\nu\right), \end{aligned} \tag{10.3.1}$$

where B_{21} and B_{12} are the coefficients of stimulated emission and absorbtion.

There is no reason to suppose the electrons and photons we are dealing with are anywhere near thermal equilibrium, but if they were, $\rho(\nu)$ would be given by

$$\rho\left(\nu\right) = \frac{8\pi h\nu^3}{c^3}\,\frac{1}{e^{h\nu/kT}-1}\ , \tag{10.3.2}$$

and the number densities of electrons, N_2 and N_1, with and without the energy corresponding to that of one photon, would be in the ratio

$$\frac{N_2}{N_1} = e^{-h\nu/kT}\ , \tag{10.3.3}$$

The reason these relations are germane is that the coefficients A_{21}, B_{21} and B_{12}, themselves independent of the local distribution of radiation, must be consistent with this equilibrium. This would imply

$$N_2\left[B_{21}\rho\left(\nu\right) + A_{21}\right] = N_1\, B_{12}\rho\left(\nu\right)\ . \tag{10.3.4}$$

Solving Eqs. (10.3.4) for N_2/N_1 and substituting the result into (10.3.3) yields

$$\rho\left(\nu\right) = \frac{A_{21}}{B_{12}e^{h\nu/kT} - B_{21}} \ . \tag{10.3.5}$$

For this equation to be consistent with Eq. (10.3.2) requires

$$B_{12} = B_{21} = \frac{c^3}{8\pi h\nu^3}\, A_{21} \ . \tag{10.3.6}$$

The first equation expresses the equality of absorption and induced emission. The second expresses the stimulated enhancement over and above spontaneous emission.

As stated before, FEL operation depends on this stimulated emission. But, as with any laser, to get more emission than absorption, it is necessary for the ratio N_2/N_1 to be altered artificially from its thermal equilibrium value.

10.4. Applicability Condition for Semi-Classical Treatment of Undulator Radiation

It is sometimes convenient to treat analyse wiggler/undulator radiation in the rest frame of the electron. In this frame, because the wiggler is rushing past the electron, its spatial period is Lorentz contracted to λ_w/γ, and its overall length is $N_w\lambda_w/\gamma$. As the electron oscillates (transversely, parallel to the electric field in this frame) it emits radiation at frequency $\omega' = 2\pi\, c\,\gamma/\lambda_w$. This dipole radiation is directed more or less isotropically, but with a minimum along the electric field direction. When this radiation is transformed back into the laboratory frame, the radiation is Doppler shifted to far shorter wavelengths and is the strongly forward-peaked "searchlight beam" we have derived repeatedly.

Since atomic particles and photons are involved in this process, one can inquire whether quantum mechanical considerations are important. For a start, one can observe that it is inconsistent to start with an electron at rest. It is not possible to localize the electron better than describing it by a Gaussian wave packet, with spreads Δx_0 and Δp_x that satisfy the Heisenberg uncertainty condition,

$$\Delta x_0\, \Delta p_x \geq \hbar \ . \tag{10.4.1}$$

As the packet spreads with time, let its spatial spread be $\Delta x(t)$. Accepting the equality condition as determining Δp_x, and treating the electron motion non-relativistically, the

spatial width evolves with time according to

$$\Delta x^2\,(t) = \Delta x_0^2 + \left(\frac{\hbar\, t}{m\,\Delta x_0}\right)^2 \ . \tag{10.4.2}$$

The total time it takes for the wiggler to pass the electron is $N_w\,\lambda_w/\gamma\,c$. To minimize the importance of spreading, we can choose the value for Δx_0 that minimizes $\Delta x(t)$ after this time, namely

$$\Delta x_0^2\,(t) = \frac{2\hbar\, t}{m} = 4\pi\,\frac{N_w \lambda_w}{\gamma}\,\frac{h\,c}{mc^2} \ . \tag{10.4.3}$$

The final factor, $\lambda_C = hc/(mc^2) = 2.43 \times 10^{-12}\,\mathrm{m}$, is known as the electron Compton wavelength. To maintain coherence over all wiggler radiation, one requires the electron's Heisenberg motion to be small compared to the wavelength of its radiation. Requiring $\Delta x_0(t)$ to be much less than λ_w/γ, the wavelength of the field in the electron's rest frame, yields

$$N_w << \frac{1}{4\pi}\,\frac{1}{\gamma}\,\frac{\lambda_w}{\lambda_C} \quad \left(\overset{\text{e.g.}}{=} \frac{1}{4\pi}\,10^{-4}\,\frac{10^{-2}}{2.4\times 10^{-12}} \approx 10^5\right) \ . \tag{10.4.4}$$

By this estimate, as long as the number of poles is less than ten thousand or so, quantum effects can be neglected.

In striving for ever more brilliant beams one reduces the emittance of the electron beam, say to $\epsilon_e = \sigma_x^2/\beta_w = 10^{-10}\,\mathrm{m}$, where β_w, say equal to $1\,\mathrm{m}$, is the lattice β-function at the wiggler location. It is appropriate to compare ϵ_e to the quantity (from Eq. (10.4.3))

$$\frac{\Delta x_0^2\,(t)}{\beta_w} \approx N_w\,\frac{\lambda_w}{\beta_w}\,3\times 10^{-15}\,(\mathrm{m}) \ , \tag{10.4.5}$$

since it would be inconsistent to demand the dimensions of the packets describing individual electrons to exceed the beam dimensions. For the numerical values we have been using, and $N_w < 10^4$, this condition is comfortably met, provided $\beta_w > \lambda_w$.

10.5. Optical Resonator

This section is largely unchanged from an appendix to an earlier handout, that was never discussed in class. It is reproduced here for your convenience in not having to recover those notes, and for my convenience in being free to make additions and corrections.

One purpose is to further develop the analogy between beams of light and beams of particles. This analysis is essential to the understanding of optical lasers—in particular free electron lasers.†

At the cost of suppressing polarization properties, one can treat light as a scalar wave. Let $E(t;x,y,z) \sim \exp(-i\omega t)$ be such a wave, which we assume to be monochromatic and traveling more or less parallel to the z axis in a "focusing medium" such as an optical fiber. Paraxial approximations will be assumed to be valid. The wave equation satisfied by E is

$$\nabla^2 E + \left(k_0^2 - k_0 k_{2,x} x^2 - k_0 k_{2,y} y^2\right) E = 0. \tag{10.5.1}$$

It is the quadratic terms $-k_0 k_{2,x} x^2 - k_0 k_{2,y} y^2$ that cause the wavelength to depend on position and this is what causes focusing. Before solving this equation, we embark on a digression to further correlate wave and particle descriptions.

10.5.1. Wave particle duality

A closely analogous mechanical system has a single particle of kinetic energy $T = (p_x^2 + p_y^2 + p_z^2)/(2m)$ traveling more or less parallel to the z axis in a "potential well" such that its potential energy is $V = k'_{2,x} x^2/2 + k'_{2,y} y^2/2$, with Hamiltonian $H(x,y;p_x,p_y,p_z) = T + V$ and total energy $\mathcal{E}$. The Schrödinger equation for such a system is

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2 \Psi + V(\mathbf{r})\,\Psi. \tag{10.5.2}$$

In the classical limit one neglects $\hbar$, for example by substituting

$$\Psi(\mathbf{r},t) = A e^{iS(\mathbf{r},t)/\hbar}, \tag{10.5.3}$$

and then setting $\hbar = 0$. The result is

$$\frac{1}{2m}(\nabla S)^2 + V(\mathbf{r}) = -\frac{\partial S}{\partial t}. \tag{10.5.4}$$

† Much of this appendix is drawn from A. Yariv, *Optical Electronics*, Holt, Rinehart and Winston, 1976. There may be "factor of two" or sign errors, since I have transcribed the formulas from 2D cylindrically symmetric waves, dependent on $r = \sqrt{x^2+y^2}$, to 1D waves, dependent on y.

This is the Hamilton-Jacobi equation for the analogous mechanical system. In this formalism $\mathbf{p} = \nabla S$ and $\mathcal{E} = -\partial S/\partial t$, so Eq. (10.5.4) amounts to conservation of energy. Because the gradient operator picks out the normal to a surface of constant $\mathcal{S}$, the equation $\mathbf{p} = \nabla S$ relates the particle direction to surfaces of constant S in the same way that rays are related to wavefronts in optics—which is that rays are everywhere normal to wavefronts. The trajectories derived from this equation are necessarily the same as the solutions of a classical harmonic oscillator equation like Eq. (1.2).

The connections between particle quantities and wave quantities are the deBroglie and Planck relations,

$$\mathbf{k}\left(\mathbf{r}\right) = \frac{\mathbf{p}\left(\mathbf{r}\right)}{\hbar}, \quad \text{and} \quad \omega = \frac{p_0^2}{2m\hbar} = \frac{\hbar k_0^2}{2m}. \tag{10.5.5}$$

Let us then introduce a new unknown $\mathbf{k}(\mathbf{r})$, related to $S(\mathbf{r})$ by

$$\frac{S\left(\mathbf{r}\right)}{\hbar} = \mathbf{k}\left(\mathbf{r}\right)\cdot\mathbf{r} - \frac{p_0^2}{2m\hbar}t, \quad \text{so} \quad \Psi\left(\mathbf{r},t\right) = Ae^{i\mathbf{k}(\mathbf{r})\cdot\mathbf{r}}e^{-i\frac{\hbar k_0^2}{2m}t}. \tag{10.5.6}$$

The frequency ω introduced here is rather artificial in classical mechanics and has no significance other than being $\mathcal{E}$ is different units. In any case, $\hbar k_0$ is a constant equal to the momentum a particle of this frequency has when, because it is traveling precisely along the z-axis, it has neither potential energy nor transverse kinetic energy.

With Ψ having the time dependence shown in Eq. (10.5.6), and after relating the parameters $k'_{2,x}$ and $k'_{2,y}$ to $k_{2,x}$ and $k_{2,y}$, one sees that Eq. (10.5.1) for E is the same as (time-independent) Schrödinger equation for the particle wave function Ψ. In the short wavelength, geometric optics limit then, the rays of a solution to the wave equation are trajectories of a material particle propagating in the corresponding potential.

10.5.2. Gaussian beam in a focusing medium

To emphasize the wave picture, for later application to lasers, we return to the task of solving Eq. (10.5.1). For consistency with the rest of the packet of notes we will simplify a bit by assuming that $k_{2,x} = 0$ and that E is independent of x. However all the formulas generalize naturally to simultaneous x and y dependence and (especially) to azimuthally-symmetric systems for which $k_{2,x} = k_{2,y}$, in which case the motion depends only on $r = \sqrt{x^2+y^2}$. We also simplify the notation slightly $k_0 \to k$, $k_{2,y} \to k_2$:

$$\frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} + \left(k^2 - k\,k_2\,y^2\right)E = 0. \tag{10.5.7}$$

Since we will study only paraxial beams parallel to the z-axis we represent E as

$$E(y,z) = \psi(y,z)\, e^{ikz}, \tag{10.5.8}$$

where $\psi(y,z)$ is assumed to be a slowly varying "modulating" factor, satisfying the inequalities $\partial^2\psi/\partial z^2 << k^2\psi$ and $\partial^2\psi/\partial z^2 << k\partial\psi/\partial z$. Substituting into (10.5.7), we find that $\psi(y,z)$ must satisfy

$$\frac{\partial^2\psi}{\partial y^2} + 2ik\,\frac{\partial\psi}{\partial z} - kk_2y^2\psi = 0. \tag{10.5.9}$$

With a view towards obtaining two conditions, one for motion for which y is negligible, and another where y motion is important, we seek a solution of this equation in the form

$$\psi = \exp\left(iP(z) + ik\frac{y^2}{2}\frac{1}{q(z)}\right). \tag{10.5.10}$$

This substitution will lead to a Gaussian beam if $q(z)$ is allowed to be complex and have appropriate z-dependence. For reasons to be explained later q will be called the "complex radius of curvature". When substituting from Eq. (10.5.10) into Eq. (10.5.9) one can separate terms proportional to y^2 from terms having coefficients independent of y. To be valid for all y both coefficients must vanish. This leads to

$$\left(\frac{1}{q}\right)^2 + \frac{d}{dz}\left(\frac{1}{q}\right) + \frac{k_2}{k} = 0, \quad \text{and} \quad \frac{dP}{dz} = \frac{i}{2q}. \tag{10.5.11}$$

10.5.3. Gaussian beam in free space

In free space Eqs. (10.5.11) simplify to

$$\left(\frac{1}{q}\right)^2 + \frac{d}{dz}\left(\frac{1}{q}\right) = 0, \quad \text{and} \quad \frac{dP}{dz} = \frac{i}{2q}. \tag{10.5.12}$$

From the first of these $q(z) = az + b$, where a and b are arbitrary constants. Being free to choose these constants, we define

$$\frac{1}{q} = \frac{1}{z - iz_0} = \frac{z + iz_0}{z^2 + z_0^2} = \frac{1}{z\left(1 + z_0^2/z^2\right)} + \frac{i}{z_0\left(1 + z^2/z_0^2\right)}. \tag{10.5.13}$$

Then the second equation yields

$$P(z) = \frac{i}{2}\ln\left(1 + \frac{z}{z_0}\right), \tag{10.5.14}$$

where the constant of integration has been chosen as $-(i/2)\ln(-iz_0)$. Substituting these expressions into Eq. (10.5.10) yields

$$\psi = \exp\left(-\ln\sqrt{1+i\frac{z}{z_0}} + k\frac{y^2}{2}\left(\frac{i}{z\left(1+z_0^2/z^2\right)} - \frac{1}{z_0\left(1+z^2/z_0^2\right)}\right)\right). \tag{10.5.15}$$

Using the relation $\ln(a+ib) = \ln\sqrt{a^2+b^2} + i\tan^{-1}(b/a)$ the first term in the exponent becomes

$$\exp\left(-\ln\sqrt{1+i\frac{z}{z_0}}\right) = \frac{1}{\sqrt{1+\left(z/z_0\right)^2}}\exp\left(-i\,\tan^{-1}\frac{z}{z_0}\right). \tag{10.5.16}$$

To cast Eq. (10.5.15) into a form appropriate for comparison with Figure 1.6 and Eq. (), we define

$$w_0^2 = \frac{z_0}{k}, \quad w^2\left(z\right) = w_0^2\left(1+\frac{z^2}{z_0^2}\right), \quad \text{and} \quad R\left(z\right) = z\left(1+\frac{z_0^2}{z^2}\right). \tag{10.5.17}$$

Finally, by Eq. (10.5.8), the electric field is given

$$E\left(y,z\right) = E_0\,\frac{w_0}{w\left(z\right)}\exp\left(ikz - i\,\tan^{-1}\left(z/z_0\right) + \frac{y^2}{2}\left(\frac{ik}{R\left(z\right)} - \frac{1}{w^2\left(z\right)}\right)\right). \tag{10.5.18}$$

In laser terminology this is known as the "fundamental mode", though that terminology refers to a cylindrically symmetric beam while we are dealing with only one transverse dimension. The "lowest mode" has resulted because only terms of order y^2 have been retained.

10.5.4. The ABCD Law

The reason for having chosen the symbol $R(z)$ is that in a pure cylindrical wave, the spatial dependence near the z-axis is as

$$\exp ikR = \exp ik\sqrt{z^2+y^2} \approx \exp ikz + ik\frac{y^2}{2R}. \tag{10.5.19}$$

Also the symbol q in Eq. (10.5.10) was chosen with this dependence in mind. Combining the various formulas we have

$$\frac{1}{q\left(z\right)} = \frac{1}{R\left(z\right)} + \frac{i}{kw^2\left(z\right)}. \tag{10.5.20}$$

This is the basis for calling q the "complex radius of curvature" of the wave. Note though, that it includes the phase evolution corresponding tho the fact that the wave is spreading.

Though Eq. (10.5.18) was derived to describe free space evolution of the wave, we need not demand that $E(y,z)$ as given by Eq. (10.5.18) came from that source. Rather we can regard $R(z)$ and $w(z)$ as local parameters describing the field at point z in an arbitrary Gaussian optical line. $R(z)$ is the radius of curvature of the local wave front and $w(z)$ is a measure of the local beam width (actually height since y is a vertical coordinate.) For this representation to be useful we must be able to calculate the evolution of $R(z)$ and $w(z)$ through arbitrary optical elements.

According to Eq. (10.5.11), the differential equation satisfied by q, in order for Eq. (10.5.18) to properly describe field evolution in a focusing medium is

$$\left(\frac{1}{q}\right)^2 + \frac{d}{dz}\left(\frac{1}{q}\right) + \frac{k_2}{k} = 0. \tag{10.5.21}$$

This can be converted into a linear equation by defining

$$\frac{1}{q} = \frac{s'}{s}, \quad \text{so} \quad s'' + \frac{k_2}{k}\, s = 0. \tag{10.5.22}$$

where d/dz is indicated by a prime. The general solution of this equation is

$$\begin{aligned} s\,(z) &= a\,\sin\sqrt{\frac{k_2}{k}}z + b\,\cos\sqrt{\frac{k_2}{k}}z \\ s'\,(z) &= a\,\sqrt{\frac{k_2}{k}}\cos\sqrt{\frac{k_2}{k}}z - b\,\sqrt{\frac{k_2}{k}}\sin\sqrt{\frac{k_2}{k}}z. \end{aligned} \tag{10.5.23}$$

This solution subsumes the free space solution obtained earlier, as well as propagation through a thin lens (by an appropriate limiting process.) Reexpressing Eq. (10.5.23) to obtain $q(z)$ in terms of $q(0) = q_0$, we obtain

$$q\,(z) = \frac{q_0\cos\sqrt{\frac{k_2}{k}}z + \sqrt{\frac{k}{k_2}}\sin\sqrt{\frac{k_2}{k}}z}{-q_0\sqrt{\frac{k_2}{k}}\sin\sqrt{\frac{k_2}{k}}z + \cos\sqrt{\frac{k_2}{k}}z}. \tag{10.5.24}$$

For the evolution $z_0 \to z_1$ this transformation is traditionally written as

$$q_1 = \frac{A_{10}\,q_0 + B_{10}}{C_{10}\,q_0 + D_{10}}. \tag{10.5.25}$$

The coefficients can be read off from Eq. (10.5.24). Subsequent evolution $z_1 \to z_2$ is given by

$$q_2 = \frac{A_{21}\,q_1 + B_{21}}{C_{21}\,q_1 + D_{21}}. \tag{10.5.26}$$

Defining the concatenated transformation by

$$q_2 = \frac{A_{20}\, q_0 + B_{20}}{C_{20}\, q_0 + D_{20}}. \tag{10.5.27}$$

Direct substitution shows that this concatenation requires

$$\begin{pmatrix} A_{20} & B_{20} \\ C_{20} & D_{20} \end{pmatrix} = \begin{pmatrix} A_{21} & B_{21} \\ C_{21} & D_{21} \end{pmatrix} \begin{pmatrix} A_{10} & B_{10} \\ C_{10} & D_{10} \end{pmatrix}. \tag{10.5.28}$$

By a miracle of the sort that makes physics so satisfactory, the matrices introduced here are none other than the transfer matrices introduced in section 1.2. The reason for this is that Eqs. 1.2 and the second of Eqs. (10.5.22) are the same. As a result the coefficients in Eq. (10.5.24) are the same as the transfer matrix elements for Eq. 1.2. Since all linear optical elements are special (or limiting) cases of a uniform focusing medium, and since the same matrix concatenation holds, the result is true for arbitrary beam lines. This rule governing the evolution of $q(z)$ (or equivalently $R(z)$ and $w(z)$) is known as the "ABCD law". It applies equally to photon and electron beams though, as far as I know, it has never been applied to electron beams. In the world of lasers there is a formalism that treats $1/q$ as a complex impedance and develops analogs between optical lines and lumped constant electrical circuits.†

10.5.5. Optics using mirrors

Designing optical beamlines for X-rays is greatly hampered by the absence of lossless refractive media at short wavelengths. There is however an appreciable range of wavelengths for which soft X-rays can be reflected and this makes spherical or cylindrical mirrors practical. The formalism just derived can be used for the analysis of optics based on mirrors. Of course, mirrors can also be used in optical systems for visible light, with laser resonators being the most important example. see Fig. 10.5.1.

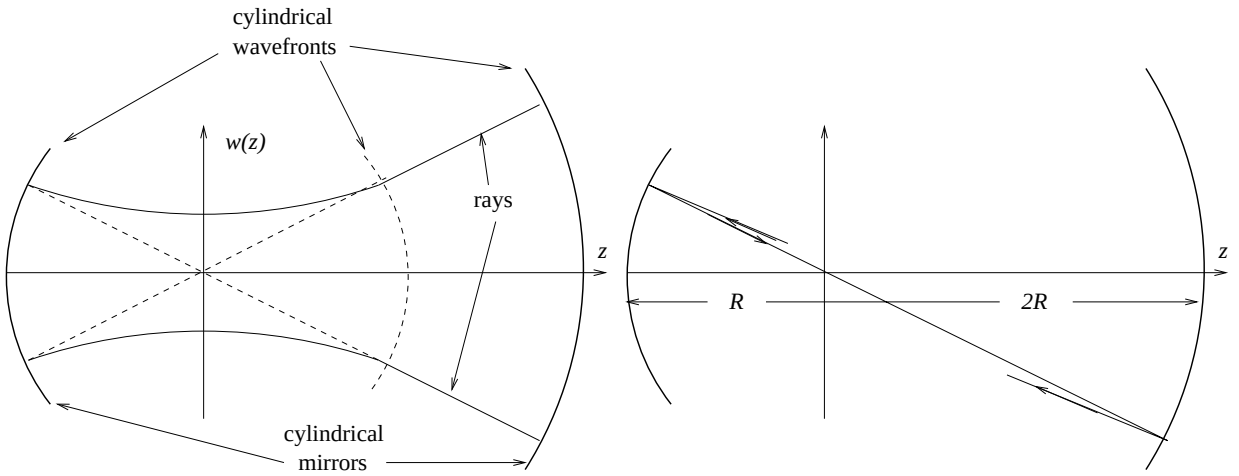
Since the transfer matrix for a mirror is

$$\begin{pmatrix} 1 & 0 \\ -2/R & 1 \end{pmatrix}, \tag{10.5.29}$$

Eq. (10.5.25) becomes

$$\frac{1}{q_1} = -\frac{2}{R} + \frac{1}{q_0}. \tag{10.5.30}$$

† Kogelnik, H., *On the propagation of Gaussian beams through lenslike media*, Applied Optics, vol. 4, p. 1562, 1965. Gerrard, A. and Burch, J.M., *Introduction to Matrix Methods in Optics, Dover, 1975.*



(a) "realistic" ray picture (a) "high school optics" ray trace

Figure 10.5.1: Optical resonator defined by two cylindrical mirrors. (a) More or less realistic rays and wavefronts. (b) Elementary ray tracing using focal length equals radius/2.

Substituting from Eq. (10.5.20), and ignoring imaginary parts, (signs???)

$$\frac{1}{R_1} = -\frac{2}{R} + \frac{1}{R_0}. \tag{10.5.31}$$

When this formula is applied to the optical system shown in Fig. 10.5.1 one sees that it partially validates an analysis using the "high school optics" formula

$$\frac{1}{\text{object-dist.}} + \frac{1}{\text{image-dist.}} = \frac{1}{\text{focal-length}}, \tag{10.5.32}$$

where the focal length of a mirror is half its radius. One sees though that the actual rays are not at all what one draws with classical ray tracing. The actual beam envelope shown is the same as would be obtained using the β-function formalism of accelerator physics. Of course the beam divergences shown in the figure are unrealistically great. Also, for soft X-rays, because reflection is only appreciable for glancing reflections, practical geometries are far less favorable than that shown.

10.6. Density of States and Stored Energy in Laser Resonator

The reason an FEL deserves to have "laser" as part of its name is that its functioning depends on stimulated emission. The phenomenon of stimulated emission is due to the Bose nature of photons. The spontaneous radiation of photons into any particular state (or mode) is augmented by stimulated emission that is proportional to the number of photons already in that state. These pre-existing photons have energy, and to get a substantial enhancement may make this energy large. If the laser resonator does not have a very high Q-value, the power loss may not be tolerable. In this section we estimate the stored energy. The absolute value of the energy will not be especially important or reliable, but the dependence on radiation wavelength λ should give a "back of the envelope estimate" of how the difficulty of obtaining laser operation depends on λ.

In an RF cavity, designed for operation, say, at a frequency of $\nu = 500\,\mathrm{MHz}$, or $\lambda = 60\,\mathrm{cm}$, there is only one, or perhaps several, modes of oscillation at the wavelengths for which there is appreciable excitation. But at shorter wavelengths, say nanometer scale, in a resonator having macroscopic dimensions, there are far more modes. There is an easily derived relation between resonator volume V and n, the number of states per unit frequency interval. It is

$$n = \frac{8\pi\,\nu^2\,V}{c^3}\ . \tag{10.6.1}$$

Quantum mechanically, this result corresponds to the standard rule of thumb that there is approximately one state per volume of six dimensional phase space equal to h^3. (Phase space volume of spatial volume V and a spherical shell of radius (momentum) $h\nu/c$ and thickness $h\Delta\nu/c$ is $(4\pi)(h\nu/c)^2(h\Delta\nu/c)V$, and perhaps the other factor of 2 comes from the two polarization states.)

To minimize the total energy, only states in an appropriately narrow frequency band $\Delta\nu$ should be populated. For FEL's operating with $\lambda \approx 1\,\mathrm{micron}$, the ratio $\Delta\nu/\nu$ has been in the range from 10^{-4} to 10^{-6}. Let us accept the more optimistic value, $\Delta\nu/\nu \approx 10^{-6}$, and assume it to be valid for all λ, (which is probably way too optimistic, as far as pushing to short wavelengths is concerned.) The total stored energy W is then given by

$$W = \frac{8\pi\,\nu^3\,V}{c^3}\,\frac{\Delta\nu}{\nu}\,h\nu\,\frac{\Delta\Omega}{4\pi} = 8\pi\,\frac{V}{\lambda^3}\,\frac{\Delta\nu}{\nu}\,\frac{hc}{\lambda}\,\frac{\Delta\nu}{\nu}\,h\nu\ . \tag{10.6.2}$$

Note that this varies as the inverse fourth power of λ. If V is held fixed, to go from micron scale to Angstrom scale entails a 16 orders of magnitude increase in stored energy. (Of course the absence of mirrors at Angstrom scale wavelengths, all by itself prevents the design that has been discussed here from working.)

Fig. 10.5.1 suggests that, rather than filling the resonator uniformly, it is appropriate to build a wave packet that overlaps the electron bunch as perfectly as possible. This permits us to set V to the volume of the electron bunch. For CESR the dimensions length/width/height are $4 \times 10^{-2}/4 \times 10^{-3}/10^{-4}$ m or $V \approx 10^{-8}$ m, and $\frac{\Delta\Omega}{4\pi} \approx 10^{-9}$

$$W = 8\pi \frac{V}{\lambda^3} \frac{\Delta\nu}{\nu} \frac{hc}{\lambda} \frac{\Delta\Omega}{4\pi} = 8\pi \times 10^{-6} \frac{10^{-8}}{\lambda^3} \frac{hc/e}{\lambda} \frac{\Delta\Omega}{4\pi} \ . \tag{10.6.3}$$

For operation at $\lambda = 1\,\mathring{A}$, this yields

$$W = 8\pi \frac{10^{-8}}{10^{-30}} \times 10^{-6} \times 10^{-9} \times 1.24 \times 10^{4} \approx 3 \times 10^{13}\,\text{Joules}. \tag{10.6.4}$$

This suggests that an X-ray FEL will not be operating anytime soon at CESR.

Chapter 11.
Accelerator Correction

11.1. Introduction

It is essential that the X-ray beams emanating from a storage ring light source be extremely stable, and this requires the electron beams to be similarly stable. Because of the numerous external lines, accelerator correction algorithms may be more complicated than is required for colliding beam operation. One mathematical technique that has been developed especially by the light source accelerator community is known as "singular value decomposition" (SVD) and this technique will be emphasized. Orbit smoothing in one plane is emphasized as the prototypical application.

Orbit correction is accomplished by calculating and using a set of magnet steering strengths that minimize (in a sense to be spelled out) the deviation of the closed orbit from an ideal orbit. In early accelerators the main purpose of steering such as this was to prevent beam particles from being lost by striking physical obstacles such as the beam pipe. For high energy hadron colliders, because the aperture (the "dynamic aperture") is determined by magnetic fields there is a more stringent requirement placed on the steering capability. The effect of off-axis passage of the closed orbit through a nonlinear element (present for example in superconducting dipoles due to systematic, hysteric, persistent currents) is the introduction of "feed-down". When this feed-down is randomized by the more-or-less random, closed orbit deviations, the result is a random multipole of the sort known to reduce the dynamic aperture. Because of this it is necessary to pay more attention to closed orbit deviations in the dipoles than is customary with smaller accelerators.

For the algorithm described in this paper it is assumed that commissioning of the ring has proceeded to a point where first turn injection and a closed orbit have been achieved. In the process of achieving this it is necessary to "thread" the beam through "open sectors" such as transfer lines or sectors of the ring. Mathematically this means that boundary conditions are to be satisfied at the beginning and end of the sector. But once a closed orbit has been found, it is necessarily periodic, even with arbitrary errors present. From then on the problem becomes mathematically one of satisfying periodic

boundary conditions. While open sector methods tend to be especially sensitive to errors near the ends, all elements are of comparable importance once a closed orbit has been established. Mentioned briefly below is a method for concentrating attention to a limited sector of a machine, such as an intersection region. This method can also be applied to steering through transfer lines and establishing a first turn.

To be able to compensate the closed orbit one must have beam position monitors (BPM's) distributed more or less regularly around the ring; their purpose is to measure local deviations. In describing the algorithm it is not necessary for the detectors to be in particular locations; only that they are N_d in number (where the index d (for detector) with $d = 1, 2, ...N_d$ is used to label them), that the betatron phases ϕ_d and the Twiss parameters α_d and β_d are known at these locations, and that the longitudinal coordinates, measured along the ideal orbit, are s_d. It is also necessary to have N_a elements (a for adjuster) whose strengths are to be set based on the detector readings to give a "best" compensation. The standard lattice functions are presumably also known at these adjuster locations. Though it is not the only possibility, we will describe only methods for which "best" means a least-squares minimum solution. There will be at least as many, and typically many more, detectors than there are adjusters. ($N_d \geq N_a$)

For smoothing the orbit it is extremely convenient to be able to assume that the motion is decoupled; i.e. one of the two normal modes of transverse motion is approximately horizontal everywhere and the other is approximately vertical everywhere. Unfortunately, with coupling errors of typical magnitude, this is only achievable by the use of skew quadrupole correctors. These are to be adjusted using an algorithm much like the orbit smoothing algorithm. Since the performance of this decoupling algorithm is also impaired by closed-orbit errors, it is necessary to improve the situation by successive approximation, alternating decoupling and orbit smoothing operations.

In the presence of errors the lattice functions will differ from the theoretical values on which the orbit correction algorithm is based. To reduce the sensitivity to this effect in our simulation we proceed the same way as is done in the control room; we set the tunes back to their nominal values after every iteration. With the tune being given by an integral over the inverse β-function, this places a reliable constraint on the lattice functions. With errors of the magnitude expected for high energy hadron colliders, the convergence of the

closed-orbit-smoothing algorithms proves not to be seriously impaired by the use of ideal rather than actual lattice functions and phase advances.

11.2. The Closed-Orbit Influence Function.

We define the closed orbit influence function as the closed orbit deviation at s_d (where there is a detector) per unit angular deflection at s_a (where there is an adjuster). To derive it one can first write the closed orbit deviation, at the center of a thin deflecting element located at s_a in an otherwise ideal lattice. The deflection causes a horizontal kink $\Delta x'(s_a)/2$ just before the thin element, and an equal kink just after it. In phase space (with $x' \equiv dx/ds$) the local closed orbit deviation is given by

$$\begin{aligned} x_{co}(s_a) =& \beta_{xa}\frac{S_x/2}{1-C_x}\Delta x'(s_a) \\ x'_{co}(s_a) =& -\alpha_{xa}\frac{S_x/2}{1-C_x}\Delta x'(s_a) \end{aligned} \tag{11.2.1}$$

Here α_{xa} and β_{xa} are Twiss functions, Q_x is the horizontal tune, $S_x = \sin 2\pi Q_x$, and $C_x = \cos 2\pi Q_x$. In passing, it can be observed that the denominator factor $1-C_x$ appearing in these expressions causes blow-up for integer tune values. This orbit sensitivity is due to an "integer resonance". This not only prevents operation at integer tune values, but favors operation well away from integers in order for the lattice to exhibit reduced sensitivity to bend errors.

With the displacement x being continuous across the thin bend element and taking account of the kink, the phase space coordinates just after the thin element are given by

$$\begin{aligned} x_{co}(s_a+) =& \beta_a\frac{S_x/2}{1-C_x}\Delta x'(s_a) \\ x'_{co}(s_a+) =& -\alpha_a\frac{S_x/2}{1-C_x}\Delta x'(s_a) + \frac{\Delta x'(s_a)}{2} \end{aligned} \tag{11.2.2}$$

Both of these are proportional to $\Delta x'(s_a)$. Since it has been assumed that the rest of the lattice is ideal, the rest of the closed orbit is simply the pure betatron oscillation starting with these coordinates. At location s_d, the coordinates, divided by $\Delta x'(s_a)$, are given by

$$\begin{aligned} \begin{pmatrix} x_{co}(s_d)/\Delta x'(s_a) \\ x'_{co}(s_d)/\Delta x'(s_a) \end{pmatrix} =& \begin{pmatrix} \beta_{xd}^{1/2} & 0 \\ -\alpha_{xd}\beta_{xd}^{-1/2} & \beta_{xd}^{-1/2} \end{pmatrix} \times \\ \times& \begin{pmatrix} C_x(d,a) & S_x(d,a) \\ -S_x(d,a) & C_x(d,a) \end{pmatrix} \begin{pmatrix} \beta_{xa}^{-1/2} & 0 \\ \alpha_{xa}\beta_{xa}^{-1/2} & \beta_{xa}^{1/2} \end{pmatrix} \begin{pmatrix} \beta_{xa}\frac{S_x/2}{1-C_x} \\ -\alpha_{xa}\frac{S_x/2}{1-C_x} + \frac{1}{2} \end{pmatrix} \end{aligned} \tag{11.2.3}$$

where $C_x(d,a)$ stands for the cosine of the phase advance ϕ^X_{da} from a to d and similarly $S_x(d,a) = \sin\phi^X_{da}$. This formula can be understood as a sequence of three linear transformations: first is a transformation from the initial Cartesian phase space frame to a reference frame in which betatron motion is pure rotation, next the rotation through angle ϕ^X_{da}, and finally transformation back to the Cartesian phase space coordinates. Completing the matrix multiplications in Eq. (11.2.3) yields the result

$$\begin{aligned} T^X_{da} \equiv \frac{x_{co}\,(s_d)}{\Delta x'_a} &= \left[\frac{S_x}{1-C_x} C_x\,(d,a) + S_x\,(d,a)\right] \frac{\sqrt{\beta_{xa}\beta_{xd}}}{2} \\ &= \frac{\cos\left(\mu_x/2 - \phi^X_{da}\right)}{2\sin\frac{\mu_x}{2}} \sqrt{\beta_{xa}\beta_{xd}}. \end{aligned} \tag{11.2.4}$$

As written, the adjuster is regarded as preceeding the detector, even in cases where $s_d < s_a$. This requires passing the "starting point" in the lattice in evaluating ϕ^X_{da}. If one prefers one can avoid this by the replacement $\cos(\mu_x/2 - \phi^X_{da}) \to \cos(\mu_x/2 - |\phi^X_{da}|)$ in Eq. (11.2.4). There is a similar relation for x'_d which will not be useful since the detector at d measures x_d not x'_d. The function T^X_{da} is the desired influence function. Though the indices a and d appear symmetrically, they will be treated rather differently in subsequent formulas. In particula d is never a free index—it is always summed over.

The entire discussion of smoothing the orbit horizontally can be take over to vertical smoothing with just a change of symbols;

$$T^Y_{da} = \frac{\cos\left(\mu_y/2 - \phi^Y_{da}\right)}{2\sin\frac{\mu_y}{2}} \sqrt{\beta_{ya}\beta_{yd}}. \tag{11.2.5}$$

11.3. Improvement of the Closed Orbit Using Steering Correctors.

Formula (11.2.4), as well as applying to the adjuster elements (whose strengths are under our control), can also be used to calculate the influence of the (unknown) steering errors present in the lattice. All these deflections can be added to obtain an expression (valid to lowest order in the angular deflections $\Delta x'_i$) for the closed orbit deviation at s_d. There is no point, however, in writing out the error terms explicitly as they are unknown. Rather all of the unknowns can be combined into a single term Δx_d, the total error contribution to the closed orbit deviation. Applying this formulation to write the closed orbit deviation

at each of the N_d detectors yields

$$x_d = \Delta x_d + \sum_{a=1}^{N_a} T^X_{da} \Delta x'_a \qquad d = 1, 2, ..., N_d \tag{11.3.1}$$

In this equation the quantities μ_x, ϕ^X_{da}, β_{xa}, and β_{xd} are known from the ideal lattice model, Δx_d is measureable, and the deflections $\Delta x'_a$ are controllable quantities, available for improving the x_d values.

To express Eq. (11.3.1) in matrix form we introduce a row vector of displacements $\mathbf{X} = (x_1, x_2, \ldots, x_{N_d})$ and a column vector $\mathbf{\Delta X'} = (\Delta x'_1, \Delta x'_2, \ldots, \Delta x'_{N_a})^T$ of adjustments. With the influence functions arrayed in a matrix $\mathbf{T}$ having one column for every adjuster and one row for every detector, Eq. (11.3.1) becomes

$$\mathbf{X} = \mathbf{\Delta X} + \mathbf{T}\,\mathbf{\Delta X'}. \tag{11.3.2}$$

The simplest (but unrealistic) closed orbit adjustment would have precisely one deflection error in the lattice, and an adjuster at the same location. In that case Eq. (11.3.1) could be used with just one detector, placed almost arbitrarily, to determine the value of $\Delta x'_a$ needed to adjust x_d to zero. Because the compensating element is superimposed on the error the orbit will then be correct everywhere. But, if the compensating element were not superimposed on the error then the closed orbit, though correct at the detector location, could not be expected to be improved elsewhere.

Orbit correction can also be based on localized "n-bumps" where n, typically 2, 3, or 4, is the number of steering elements in the bump. If two adjusters are located at points separated by multiples of π in betatron phase, the kink due to the first element can be cancelled by the second, thereby causing a central orbit excursion that vanishes except between the two elements. Such a localized 2-bump can be used for local orbit smoothing. Far more flexible, because the betatron phase intervals can be almost arbitrary, are "three-bumps" made with three steering elements. It is possible to formulate a steering algorithms that is based entirely on three-bumps. If the bumps are labeled 1,2, and 3, the error $x_2^{(0)}$ at the center element can be canceled, while at the same time leaving the orbit unchanged outside the three-bump region. These "few bump" correction schemes correct the orbit locally, without running the risk of damaging the orbit elsewhere. This type of correction

can be regarded as a special case of the algorithm that is to be described here, which applies to any number of adjusters and detectors.

Realistically, the detectors are situated at locations unrelated to the locations of errors. One possibility (still impractical) would be to have exactly the same number of detectors and adjusters. In that case, from the known measurements $\mathbf{\Delta X}$, $N_a = N_d$ in number, the deflection $\mathbf{\Delta X'}$ can be adjusted by solving the equations

$$\mathbf{T}\,\mathbf{\Delta X'} = -\mathbf{\Delta X}^T. \tag{11.3.3}$$

One reason this method was said to be "impractical" is that it is not known whether these equations are solvable. That is, the existence of $\mathbf{T}^{-1}$ is uncertain.

A more general, and more robust, compensation procedure is to permit $N_d \geq N_a$—at least as many, and typically far more, detectors than adjusters. In that case, since there are more equations than there are unknowns, a procedure other than solving Eqs. (11.3.3) is required. We follow a routine procedure for such cases by recasting the problem as a minimization problem, minimizing for example $\sum_{d=1}^{N_d} x_d^2$. (In the special case that $N_a = N_d$ this leads to the solution of the previous paragraph.) We will call the target positive-definite function to be minimized a "badness function", B;

$$B\left(\Delta x'_1, ..., \Delta x'_{N_a}\right) = \sum_{d=1}^{N_d} x_d^2 = \mathbf{X}\mathbf{X}^T = \left(\mathbf{\Delta X} + \mathbf{T}\,\mathbf{\Delta X'}\right)\left(\mathbf{\Delta X} + \mathbf{T}\,\mathbf{\Delta X'}\right)^T. \tag{11.3.4}$$

The functional dependence $(\Delta x'_1, ..., \Delta x'_{N_a})$ ascribed here to B emphasizes the fact that, since we have control over the adjustments, we have (some) control over the badness.

11.4. Possible refinements

Another candidate badness function is $\sum_{d=1}^{N_d} \epsilon_d$ where $\epsilon = [x^2 + (\alpha x + \beta x')^2]/\beta$ is the Courant-Snyder (C.S.) invariant. This weighting can be characterized as giving equal weight to equal C.S. deviations. This choice cannot be used in any straightforward operational way however, since the slope x'_d is measureable only indirectly. Next best might be x^2/β. At least for orbit smoothing of the long arcs of the accelerator this is unnecessary since most of the horizontal BPM's are situated at points where the spread of beta function values is not great. But special weighting can be applied to particular locations in the lattice, and the subsequent formulas then acquire consistent factors.

At this point it is appropriate to introduce some practical considerations and definitions and to modify the interpretation of the formulas accordingly. In a computer model of the lattice there is an "ideal" or "reference" orbit, having all local and global parameters exactly correct. If errors are introduced into this model, there will be a distorted closed orbit deviating from the reference orbit. The physical elements of the actual accelerator, to be called the "as-built" accelerator, may exhibit large deviations (perhaps a centimeter) from this globally perfect reference lattice. Locally, however, on the scale of several cell lengths, the practicalities of survey and alignment will permit an orbit that is very smooth, showing deviations that are considerably less than a millimeter from element centers. One can visualize this curve being extended into a perfectly smooth, globally-slightly-distorted curve passing close to the centers of all the magnet elements. This can be regarded as a "sufficiently-ideal" orbit and will be called the as-built reference orbit. Provided the magnet elements do not shift spatially it stays constant for all time, independent of magnet powering. For accelerator operations it is deviation from this orbit that matters. For better or for worse this must be lived with until another global survey or magnet repositioning takes place. When simulating operations in the computer it would be possible by introducing a random-but-smooth distortion, to model such locally small but globally large deviations from the design reference lattice. Being a nuisance, this is not usually done. Rather, we measure deviations from the as-built reference orbit in the accelerator and refuse to distinguish them from the deviations from the design reference orbit in the computer model. This accuracy of this procedure can be investigated and defended provided a smoothness, or adiabatic condition is satisfied. From here on, when referring to the as-built accelerator, we will skip all this qualification and simply use one of the synonomous terms ideal orbit, or reference orbit, or central orbit. This means there is a certain fuzziness in relationship between the actual orbit and its computer model; this is an inevitable accompaniment to the realities of global equipment positioning. To model performance degradation due to this in a computer simulation it would be necessary to apply globally correlated errors.

In modeling random location errors it is necessary to specify the probability distributions of the errors. Except for random deviations depending on the anticipated achievable

adjustment precision (say ±0.5mm) the lattice quadrupoles will lie precisely on the just-defined ideal orbit, and so will the BPM's. Since quadrupoles and BPM's are located side-by-side, the practicalities of equipment positioning result in the relative position uncertainty of quads and BPM's being less yet (say ±0.1mm). If the quads and BPM's are assigned random, uncorrelated, position errors at the ±0.5mm level then this correlation will not be properly modeled. For this reason the distributions of BPM displacements are better regarded as being centered on the adjacent quads. Other correlations are common. For example the "girder" design may be such that the chromaticity compensating sextupoles and the BPM's are mounted on the same beam and hence are accurately related. Simulation codes should have the capability of modeling such correlations.

In specifying a badness function "the proof of the pudding is in the eating". It is not something that can be "wrong"; rather it is something that gives better or worse performance, where performance refers to some measure of deviation. Different measures of performance can be expected to favor different badness functions, and that is one basis for choosing a particular function. As mentioned previously it is necessary to impose the operational requirement that the badness function be measureable during actual accelerator operations, using the digitized outputs of the BPM's. Letting x_d^{BPM} be the displacement of the d'th detector from the ideal orbit, the measurable quantities are $\mathcal{X}_d = x_d - x_d^{BPM}$. We insist that the badness depend only on these quantities. In this paper we will hide this complexity by simply redefining the symbol x_d to stand for $\mathcal{X}_d$ from this point on.

11.5. Least squares compensation

The conditions for minimizing the badness defined in Eq. (11.3.4) are

$$\frac{\partial B}{\partial \Delta x'_a} = 0; \quad a = 1, ..., N_a \tag{11.5.1}$$

To perform the differentiation we spell out B in terms of components (using the summation convention)

$$\frac{\partial B}{\partial \Delta x'_a} = \frac{\partial}{\partial \Delta x'_a}\left(\Delta x_i + T_{ij}\Delta x'_j\right)\left(\Delta x_i + T_{ik}\Delta x'_k\right) = 2T_{ia}\Delta x_i + 2T_{ia}T_{il}\Delta x'_l. \tag{11.5.2}$$

To convert this back into a matrix equation it is necessary to treat rows and columns consistently, as defined above. The result is

$$\mathbf{T}^T\mathbf{T}\,\boldsymbol{\Delta}\mathbf{X}' = -\mathbf{T}^T\boldsymbol{\Delta}\mathbf{X}^T. \tag{11.5.3}$$

This equation is curiously similar to Eq. (11.3.3)—in fact it results from left multiplying that equation by $\mathbf{T}^T$. But we will see shortly that Eq. (11.5.3) is much more manageable. Proceeding optimistically, we can "solve" Eq. (11.5.3) to obtain

$$\mathbf{\Delta X}' = -\left(\mathbf{T}^T\mathbf{T}\right)^{-1}\mathbf{T}^T\mathbf{\Delta X}^T. \tag{11.5.4}$$

Assuming the inverse $(\mathbf{T}^T\mathbf{T})^{-1}$ exists, this amounts to having "solved" Eq. (11.3.3). This motivates the terminology "generalized inverse of $\mathbf{T}$" for the combination

$$\mathbf{G} = \left(\mathbf{T}^T\mathbf{T}\right)^{-1}\mathbf{T}^T. \tag{11.5.5}$$

$\mathbf{G}$ is also known as the "Moore-Penrose inverse" of $\mathbf{T}$. In case $\mathbf{T}$ should itself be invertable we have

$$\left(\mathbf{T}^T\mathbf{T}\right)^{-1}\mathbf{T}^T = \mathbf{T}^{-1}; \tag{11.5.6}$$

so the generalized inverse and the regular inverse are identical in this case.

junk junk

Depending, as it does, only on the influence function, the elements of this matrix can be calculated for the ideal orbit. They do not depend on errors or closed-orbit deviations and can be calculated once-and-for-all in the beginning from the design lattice and used in subsequent iterations even if the errors change. We also define a vector containing the inhomogeneous terms in Eq. (11.5.1)

$$\mathbf{V}^X = \left(V_a^X\right) = \left(-\sum_{d=1}^{N_d}\mathcal{X}_d^{(0)}T_{da}^X\right). \tag{11.5.7}$$

This depends on the measured deviations; it is through this vector that the measurements influence the corrections. Then, suppressing superscripts, Eq. (11.5.1) and its solution can be written

$$\mathbf{T}^T\mathbf{T}\mathbf{Q} = \mathbf{V}, \quad \mathbf{Q} = \left(\mathbf{T}^T\mathbf{T}\right)^{-1}\mathbf{V}. \tag{11.5.8}$$

One sets the adjusters according to this equation with the expectation that the orbit will then be smoother. If the lattice is purely linear, success is assured and there is no need to iterate as the solution (11.5.8) yields the "best-fit". With nonlinearity and cross-coupling one or two more iterations usually yield noticeable improvement. The possibility of large deviations in regions with no detectors, unfortunately, always remains. Most

accelerator correction algorithms are analogous to, or a variant of, or an improvement upon, Eq. (11.5.8)

The use of Eq. (11.5.8) in setting adjuster values involves the inversion of the matrix $\mathbf{M}^X$. Such an approach was once thought impractical for large accelerators due to numerical difficulties or lack of computer resources. But for the SSC, the inversion of the appropriate 420×420 matrix proved to be straightforward and quick (a few minutes on a modest scientific workstation, 1990 vintage).

This "brute force" mathematical approach, the inversion of a large matrix, can be contrasted with "physics-motivated" iterated bump approaches in which many few-dimensional local steering operations are performed. At some basic level it can probably be shown that the final results of these approaches are mathematically equivalent as long as the same figures of merit are employed, and the iterative procedures converge. Some considerations that might favor one or the other approach are numerical robustness and relative immunity to malfunctioning of a few detectors. It is found empirically that the brute force matrix inversion approach is reasonably unaffected by a few dead detectors (which report unrealistically small orbit deviations.) This result seems obvious to me, based on the general formulation, with numbers of detectors and their location being arbitrary. Clearly, though, there can be anomalously large excursions near broken detectors. It is also pretty clear that malfunctioning adjustors need to be identified and removed from the algorithm. BPM wiring errors that are equivalent to reversing the sign of the measured deflection cause serious degradation of the closed orbit, and reversed adjusters are worse yet.

Within the formalism of using equations Eq. (11.5.7) and Eq. (11.5.8), considerable flexibility remains since the detectors and adjustors can be allotted arbitrarily. For sector compensation, say during first-turn steering, one can use only those detectors and adjustors in the sector that the beam has successfully traversed. In TEAPOT, sub-families can be defined to achieve this functionality. These sub-families can also be used to emphasize the steering in critical sectors such as intersection regions. In this case it is useful to include at least a few detectors from elsewhere in the ring in order to keep the closed orbit there somewhat under control.

In planning for large accelerators various measures of orbit smoothness are possible. One measure is the r.m.s. deviation of the closed-orbit from the ideal orbit. Another is

the r.m.s. deviation of the dipoles from the closed-orbit. This can be used in providing a measure of multipole feed-down in the dipoles—an effect whose importance has already been mentioned. Yet another measure is the mean square deviation of the closed-orbit from the ideal orbit, measured at the positions of the BPM's. Except for a constant factor this last measure is the same as the badness function B defined in Eq. (11.3.4). This figure of merit will be misleadingly good, especially if there are equal numbers of adjusters and detectors, in which case a zero badness value can be obtained.

We have been assuming, and will continue to do so for the time being, that the BPM measurements themselves are free of errors. This is not necessarily realistic, especially for low intensity beams.

11.6. Singular Value Decomposition

For homogeneous sections of an accelerator, with sensibly placed adjusters and correctors, the procedure described so far is usually free of serious numerical complications. But it is not hard to visulize ways it can fail and, even if that were not the case, it is appropriate to apply sophisticated methods of numerical analysis in order to make the procedure as robust as possible.

If there are more adjusters than detectors then the minimization equations are underdetermined and cannot be solved. Even if there are more detectors than adjusters it is possible for the detectors to be placed such that there are "too many" in some areas and "too few" in others. Evidently two detectors at the same location are redundant, and two detectors close together may be essentially redundant.

It is also possible for the equations to be undetermined because of degenerate adjuster locations. Two superimposed adjusters obviously "fight each other". Since it is only their sum that affects the orbit, a correction algorithm, by making their signs opposite, can try to turn them both on very hard. This can happen even when the adjusters are not side-by-side. For example, two adjusters separated by an exact multiple of half betatron wavelengths will appear degenerate if there are no detectors between them.

These sorts of misdesign can usually be diagnosed using "physics" when the algorithm fails on their account. But one can be more confident if there are "mathematical" tricks

that reduce the seriousness of degenerate detectors or adjusters and provide diagnostic tools for identifying problems of this sort.

11.7. References

1. S. Chandrasekhar, *Stochastic Problems in Physics and Astronomy*, R.M.P., **15**, (1943).
2. R. L. Stratonovich, *Topics in the Theory of Random Noise*, Gordon and Breach, 1963.
3. S. O. Rice, Mathematical Analysis of Random Noise, Bell System Technical Journal, Vols. 23 and 24.
4. K. Kim, *Characteristics of Synchrotron Radiation*, in *Physics of Particle Accelerators*, M. Month and M. Dienes, editors, AIP 184, 1988.
5. A. Hoffmann, *Quasi-Monochromatic Synchrotron Radiation From Undulators*, NIM **152**, 17 (1978).
6. D. Attwood, K. Halbach, and K. Kim, *Science*, **228**, 1265 (1985).
7. J.D. Jackson, *Classical Electrodynamics,* Third Edition, John Wiley, 1999.
8. M. Sands, *The Physics of Electron Storage Rings,* in International School of Physics, "Enrico Fermi", Academic Press, 1971.
9. R. Coisson, *On Synchrotron Radiation in Nonuniform Magnetic Fields*, Optics Communications, **22**, 135 (1977).
10. Bossart, *Observation of Visible Synchrotron Radiation Emitted by a High-Energy Proton Beam at the Edge of a Magnetic Field*, NIM, **164**, 375 (1979).
11. J. Als-Nielsen and D. McMorrow, *Elements of Modern X-Ray Physics*, John Wiley, 2001..
12. N. Marcuvitz, *Waveguide Handbook,* p. 58, McGraw-Hill, New York, 1951.
13. R.P. Walker, *Insertion Devices: Undulators and Wigglers*, in CERN 98-04, p. 156.
14. V.F. Suller, *Introduction to Current and Brightness Limits*, CERN 98-04, p. 89.
15. V. Berestetskii, E. Lifshitz, and L. Pitaevskii, *Quantum Electodynamics, Vol. 4 of Landau-Lifshitz*, Pergamon Press, 1982.